# WiMAX, New Developments

# WIMAX, New Developments

Edited by
Dr Ms Upena D Dalal and Dr Y P Kosta

*In-Tech*
*intechweb.org*

Published by In-Teh

**In-Teh**
Olajnica 19/2, 32000 Vukovar, Croatia

Abstracting and non-profit use of the material is permitted with credit to the source. Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. Publisher assumes no responsibility liability for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained inside. After this work has been published by the In-Teh, authors have the right to republish it, in whole or part, in any publication of which they are an author or editor, and the make other personal use of the work.

© 2009 In-teh
www.intechweb.org
Additional copies can be obtained from:
publication@intechweb.org

First published December 2009
Printed in India

Technical Editor: Goran Bajac

WIMAX, New Developments,
Edited by Dr Ms Upena D Dalal and Dr Y P Kosta

p. cm.
ISBN 978-953-7619-53-4

# Preface

WiMAX (Worldwide Interoperability for Microwave Access) is a wireless broadband access network, named by industry group called the WiMAX forum formed in June 2001. It is Wireless MAN with IEEE 802.16 family standards. Loosely, WiMAX is a standardized wireless version of Ethernet or enabling the last mile, intended primarily as an alternative to wire technologies (such as Cable Modems, DSL and T1/E1 links ) to provide broadband access to customer premises. Mission of the WiMAX forum is to promote and certify compatibility and interoperability of broadband wireless products. WiMAX operates similar to WiFi but at higher speeds, over greater distances and for a greater number of users. It provides wireless data over long distances in a variety of ways, from point–to-point links to full mobile cellular type access in support with soft or hard handovers between base stations. It can connect Wi-Fi hotspots with each other. A scalable (Adaptive) wireless platform with OFDM technology is selected for constructing alternative and complementary broadband networks. WiMAX standardize essentially two aspects of air interface: The physical and MAC layer.

The current trend is towards "anywhere, anytime" wireless communication for voice, video as well as data that is with on-foot to vehicular mobility. Hence, wireless networks facilitating resource sharing are required which should be converged with Internet. This has become possible with WiFi and its follower-WiMAX broadband systems. The work is also in progress to integrate UMTS (voice services) and WiMAX (data services) to offer significant benefits by assuring the global delivery of Wireless Broadband Internet services through voice networks at the lowest cost. The need of time is (1) high speed communication of data, image, audio and video, (2) quality reception of information and (3) efficient utilization of spectrum to accommodate maximum number of users as the mobiles user-mass is increasing day by day. All three are still the open problems to the researchers at physical layer which must be supported by the network management layers above it.

In general, the design of wireless networks/systems (WiMAX as well) is difficult as the wireless channel problems are very serious especially due to multipath fading in addition to white Gaussian noise, Delay spread and Doppler spread. Receiver vehicular mobility undergoes even worst condition of fast fading channel environment. These all causes bit errors which require sturdy methods of channel coding/decoding, sturdy wideband digital modulation methods like Orthogonal Frequency Division Multiplexing/Multiple Access (OFDM/OFDMA), sturdy methods of pilot transmission supporting dynamic channel estimation and also the handovers, scheduling and mobility management. The performance of the same system can vary significantly depending on the terrain, area-rural/urban, type of propagation, weather, frequency of operation, velocity of the mobile terminal, interference sources, and other dynamic factors. Sometimes, the adaptive receiver designs must be adopted

to cop up with the channel variations. At the same time cost and complexity of the handset is also a constraint. The handset designs should be compact with minimum processing so that the battery power can be saved. All these constraints are equally applied to WiMAX supported devices. Hence, though the WiMAX protocols are standardized specifically for fixed and mobile wireless access, the developments are still going on to optimize one or other associated issue. VLSI and ULSI designs for WiMAX also have a lot of scope for research.

This book touches most of the above issues in form of 22 individuals' papers containing research work in WiMAX domain in particular. WiMAX has two important standards/usage models: A fixed usage model IEEE 802.16-2004 for Fixed Wireless Broadband Access (FWBA) and a portable usage model IEEE 802.16e-2005 (Which is mainly concentrated) for Mobile Wireless Broadband Access (MWBA). Both are released standards and amendments are available in form of drafts. Higher data rate transmissions (@ 100 Mbps) are achieved in IEEE 802.16-2004 WiMAX through LOS communications which incorporate a stationary transmitter and receiver but IEEE 802.16e supporting NLOS communication is much complicated and little less bit rate is achieved. 2-11 GHz licensed band is the range of frequencies with TDD and FDD supports.

The order of the papers/chapters is decided specifically to describe the developments in physical layer to management issues. Initially, in first two chapters, the general introduction to WiMAX and its applications are given. WiMAX is a network provides last mile broadband access within around 35-50 km area. Thereafter the physical layer design issues like, handset front end design and challenges, Companding in WiMAX, Channel coders with VLSI aspects, WiMAX transceiver design, adaptive mechanisms, filter applications, adaptive antenna system, beamforming etc. are dealt with upto chapter 11. This follows the interference and coexistence studies with other radio systems, which is partly related to physical layer and forms the basis for heterogeneous network environment.

After that the management issues and special issues of considerations are in queue including handovers. Active scheduling, packetization issue, throughput and QoS performance analysis, handovers management are described in order. Techniques for the next generation heterogeneous networks including WiMAX are studied in the end chapters, giving the scenario of multihop relay network as well as suitability of the systems to high mobility conditions.

The book will provide a wide horizon to visualize the WiMAX technology and its developments leading towards 4G systems. It will provide the good platform to the researchers with the clues to the innovative ideas in WiMAX domain. I wish all the best to the authors and readers of this book for their successful research in WiMAX technology.

Dr Ms Upena D Dalal
Dr Y P Kosta

# Contents


Preface V

1. WiMAX: Appropriate Technology to Provide Last Mile Access to ICTs Infrastructure and Services in Rural Areas 001
   Frank Ibikunle

2. The Role of WiMAX Technology on Broadband Access Networks: Economic Model 017
   João Paulo Ribeiro Pereira

3. Mobile WiMAX Handset Front-end: Design Aspects and Challenges 047
   Václav Valenta, Geneviève Baudoin, Antoine Diet, Roman Maršálek, Fabien Robert, Martha Suarez and Martine Villegas

4. The Application of μ-Law Companding to Mobile WiMax 081
   Brian G Stewart and Athinarayanan Vallavaraj

5. VLSI Architectures for WIMAX Channel Decoders 107
   Maurizio Martina and Guido Masera

6. LDPC Decoders for the WiMAX (IEEE 802.16e) based on Multicore Architectures 133
   Gabriel Falcão, Vitor Silva, José Marinho and Leonel Sousa

7. Design and Implementation WiMAX Transceiver on Multicore Platform 151
   Qing Wang, Yonghua Lin, Jianwen Chen, Da Fan and Ling Shao

8. Advanced Filter Development for WiMAX Applications 167
   Kaixue Ma, Leyu Zhang and Kiat Seng Yeo

9. Adaptive Parameters Based Transmission Control and Optimization in Mobile WiMAX at Physical Layer 183
   U. D. Dalal and Y. P. Kosta

10. Smart Antenna Systems for WiMAX Radio Technology 201
    Yikun Huang

11. Enabling WiMAX Video Streaming 213
    M. Fleury, R. Razavi, S. Saleh, L. Al-Jobouri and M. Ghanbari

12. Achieving Frequency Reuse 1 in WiMAX Networks with Beamforming 239
Masood Maqbool, Marceau Coupechoux, Philippe Godlewski and Véronique Capdevielle

13. WiMax Interference and Coexistence Studies with Other Radio Systems 253
A. Rahim Biswas and A. Finger

14. Resource Management Framework for QoS Scheduling in IEEE 802.16 WiMAX Networks 275
HuaWang and Lars Dittmann

15. Analyzing the Throughput and QoS Performance of WiMAX Link in an Urban Environment 307
Faqir Zarrar Yousaf, Kai Daniel and Christian Wietfeld

16. WiMAX-WiFi Synergy for Next Generation Heterogynous Network 321
Dr. Rashid A. Saeed, Amran Naemat, Azrin Aris and Mat Kamil Awang

17. System-Level Modelling of Heterogeneous System-on-Chip Architectures Targeting Multi-Standard Wireless Applications 343
Ali Ahmadinia, Balal Ahmad and Tughrul Arslan

18. WiMAX Cross-Layer System for Next Generation Heterogeneous Environments 357
Pedro Neves, Susana Sargento, Kostas Pentikousis and Francisco Fontes

19. Intelligent WiMax Vertical Handovers 377
Nadine Akkari Adra

20. Signal Metrics for Vertical Handoff Towards (Cognitive) WiMAX 387
François-Xavier Socheleau, Sébastien Houcke, Philippe Ciblat and Abdeldjalil Aïssa-El-Bey

21. Multi-hop Relay Networks 407
David Soldani

22. Broadband communication in the high mobility scenario: the WiMAX opportunity 429
M. Aguado, E. Jacob, M.V. Higuero, P. Saiz and Marion Berbineau

# WiMAX: Appropriate Technology to Provide Last Mile Access to ICTs Infrastructure and Services in Rural Areas

Frank Ibikunle
*Covenant University, Ota*
*Nigeria*

**Abstract**
WiMAX is a broadband wireless standard that enjoys widespread support from both the computer and telecommunications industries worldwide, making this technology particularly cost effective. It is engineered to deliver significant business benefits to operators and users in diverse environments (enterprise, consumer, emerging, public service), geographies and demographies (urban, suburban, rural), both over the short and long terms. The paper highlights some points that make "**WiMAX"-** the World Interoperability for Microwave Access, the most appropriate technology solution to favor the broadband wireless access to ICTs infrastructure and services in rural areas. And also present the various usage scenarios of WiMAX to addressing different applications.
**Keywords:** Broadband wireless access, Last-mile, Developing countries, WiMAX, ICTs

## 1. Introduction

The idea that access to information opens doors to wider economic and social development opportunities is not new. In 1984, the "Missing Link Report'' pointed to the fact that the lack of telecommunication infrastructure in developing countries impedes economic growth, but with a scope limited to access to telephones rather than today's wider concept of Information and Communications Technologies **(ICTs)** access and usage. In 1996, the International Telecommunications Union (**ITU**) initiated a United Nations project for the "Right to Communicate'' aimed at providing access to basic ICTs for all, with motivation to reduce information poverty for developing countries. Thus, during the first World Summit on the Information Society (WSIS) held in Geneva in December 2003, the 'Digital Divide' was defined as the unequal access to ICT. Although this unequal access usually apply to differences between countries (the international digital divide) e.g. comparing developed and developing countries or regions; within countries (the domestic digital divide); and most importantly the divide between rural and urban, well educated or poorly educated populations or poor and rich citizens. Despite the various parameters and selected criteria (e.g. Internet host and/or users, fixed and mobile telephones) that can be considered or use to express an overall trend of growing ICTs disparities between and within countries, the

availability and the quality of the access (i.e. the physical telecommunication infrastructure) is the key to a quick and reliable development of ICTs in the countries. Looking closer to the statistics as published by the different official bodies such as ITU or the World Bank in Figure 1, it is clear that the gap in ICTs access between developed and developing countries do exist (ITU World Telecommunications Development report 2003).

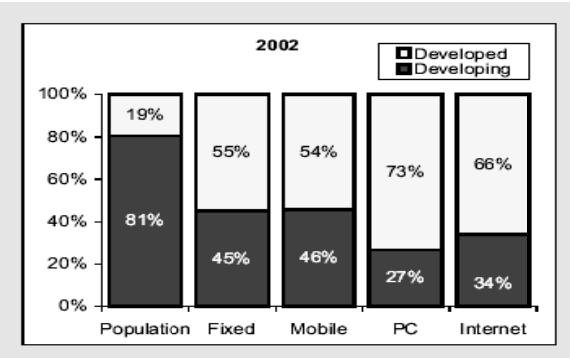



Fig. 1. Gap in ICTs access between developed and developing countries.
**Source:** ITU World Telecom indicator database

There are many reasons for the barriers why, until now, broadband ICT access was mainly deployed in developed countries and more precisely in urban areas as shown in Figure 2. However it can easily be explained that economics and existing technologies are the main drivers (and barriers in developing countries) to ICTs. *Economics* in the first case because, any entrepreneur in this new business, operators and service providers will naturally inclined to first serve the most populated areas (i.e. dense, rich cities and suburbs) where most of the potential customers are located instead of less populated areas such as remote and rural areas.

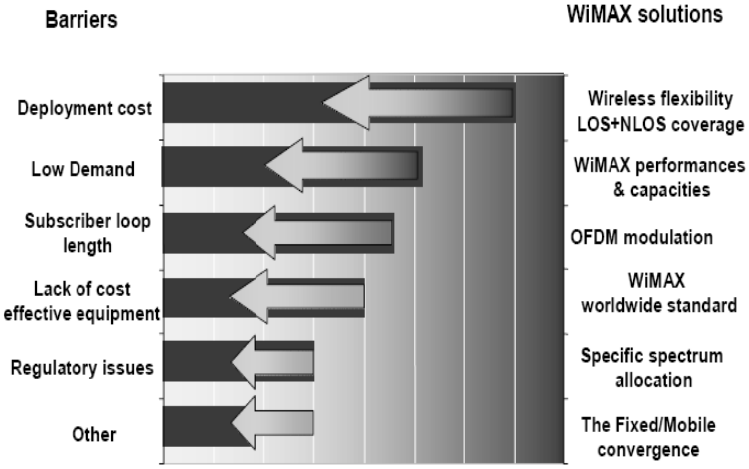



Fig. 2. Major Barriers to ICT and WiMAX Solutions Source: ITU 2003

In the same way, while developed countries and urban areas are requiring higher and higher bit rates for multimedia applications, rural areas in developing countries are still at first favoring voice communication with a slow evolution towards ICTs. Moreover, while developed countries already have existing telecommunication infrastructure ready to evolve and the financial resources to invest and pay for new services, developing countries still suffer from the lack of basic infrastructures (not only telecommunication infrastructures, but also power supply, roads), and more crucial, have great difficulty to mobilize the necessary financial resources. This is what we call the "demand factor''. *Technologies* in the second case because, until now, existing wired or wireless technologies have inherent limitations either in performances or in capacity (e.g. the 6 Km maximum distance from the exchange for digital subscriber line (DSL)) or the Line-of-Site (LOS) Customer Premise Equipment (CPE) location from the base station for wireless access. And even if such limitations can be overcome using other complementary backbones or equipments (optical fibers or microwaves links, remote switching units.etc), the extra cost of these new equipments, of their deployment and their operation directly impacts the business model. This is what we call the "cost factor''. So, both the "Demand and Cost'' factors are the major barriers to the broadband access as identified by ((ITU World Telecommunications Development report 2003).
The paper is organized as follows. Section II defines WiMAX, the Forum and Profiles. Section III describes why and how WiMAX will be a key element in this new important worldwide objective to provide an equitable and affordable access to ICTs infrastructure and services. The various usage scenarios to illustrate the ability of WiMAX to address different applications are given in sections IV and V. And the paper ended with the conclusion.

## 2. What is WiMAX?

WiMAX stands for "World Interoperability for Microwave Access''. It is a broadband wireless technology that supports fixed, nomadic, portable and mobile access. WiMAX is largely supported by the computer and the telecommunications industry, cost-effective and standard base. It is engineered to deliver the latest type of ubiquitous fixed and mobile services such as Voice 0ver Internet Protocol (VoIP), Information Technology and Video at very low cost. WiMAX systems are able to cover a large geographical area, up to 50 km and to deliver significant bandwidth to end-users up to 72 Mbps. To meet the requirements of different types of access, two versions of WiMAX have been defined. The first is based on IEEE 802.16-2004 and is optimized for fixed and nomadic access. The initial WiMAX Forum CERTIFIED products will be based on this version of WiMAX. The second version is designed to support portability and mobility, and will be based on the IEEE 802.16e amendment to the standard.

### 2.1 IEEE 802.16-2004 Standard

The IEEE 802.16 standard which includes Medium Access Control (MAC) and physical (PHY) layer specifications, aims at supporting Internet services over wireless metropolitan area networks (WMAN). It is also an alternative to traditional wired networks, such as Asymmetric Digital Subscriber Line (ADSL) and cable-modem. There are two modes (two different air-interfaces) defined in WiMAX networks: Point-to-Multi-Point (PMP) and Mesh modes. In PMP mode, two SSs (Subscriber Stations) can only communicate through BS (Base Station), while in Mesh mode, two SSs can communicate directly.

In the *PHY layer*, the IEEE 802.16 standard adopts the orthogonal frequency division multiplexing (OFDM), which is a multicarrier modulation scheme. The IEEE 802.16 standard has two OFDM-based modes: OFDM and orthogonal frequency division multiplexing access (OFDMA). Both of these technologies allow subcarriers to be adaptively modulated (e.g., QPSK, 16-QAM, and 64-QAM), depending on transmission distance and noise. Moreover, OFDMA has scalability to provide efficient use of bandwidth. *The MAC layer* of IEEE 802.16 standard was originally designed for the PMP mode. On the later amendments of the IEEE 802.16a and the IEEE 802.16d, the mesh mode was included. The IEEE 802.16a adopts OFDM to provide greater spectral efficiency and to mitigate interference. IEEE 802.16d covers most of the quality of service (QoS) aspects. The IEEE 802.16e introduces scalable OFDMA into the standard, and supports mobile communications. With handover mechanisms, WiMAX is thus able to support mobile communications at vehicular speeds.

### 2.2 IEEE 802.16e -2005 Standard

On July 2002, a study group called IEEE 802.16 Mobile WirelessMAN Task Group was initiated to produce an amendment covering the PHY and MAC layers for combined, fixed, and mobile operations in the licensed band range. The amendment was approved in December 2005 and the new standard called IEEE 802.16e-2005 was published in February 2006. The scope of this standard is to provide mobility enhancement support for SS moving at the vehicular speed, in addition to corrections to 802.16-2004 fixed operation that was developed as IEEE 802.16-2004/Cor1-2005 and published along with IEEE 802.16e-2005. 802.16e introduces many changes to PHY and MAC layer protocols owing to mobility support, which required addressing new issues that were not required in 802.16-2004, such as handoff and power management.

The mobile WiMAX PHY layer is based on OFDMA technology. The network is an IP end-to-end conventional architecture that provides high-speed broadband. With a modification from fixed WiMAX to mobile WiMAX, the PHY layer also supports BWs from 1.25 to 20 MHz. The standard is designed to accommodate either TDD or FDD (Time / Frequency Division Duplexing) deployments, allowing for both full- and half-duplex terminals in the FDD case.

The MAC layer was initially designed specifically for the point-to-point wireless access environment. It supports higher layer or transport protocols such as asynchronous transfer mode (ATM), Ethernet, or IP, and is designed to easily accommodate future protocols that have not yet been developed. MAC layer specification practices considerable departures from 802.16-2004 to provide support for mobility. It adds support for handoff and power management.

***QoS Support-*** 802.16e defines new scheduling mechanisms: the extended real-time polling service (ErtPS), which is based on two services defined in 802.16-2004; the unsolicited grant service (UGS); and the real-time polling service (rtPS). ErtPS is similar to UGS in providing unicast grants, thus saving the delay incurred for requesting the bandwidth. However, ErtPS allocations are dynamic as rtPS while UGS allocations are fixed. The ErtPS is introduced to support real-time service flows that generate periodical variable sized data packets. Thus, ErtPS is especially important to support VoIP, since it allows for managing traffic rates and improves latency and jitter.

***Handover Support-*** 802.16e includes new MAC-level request/grant mechanisms to achieve similar seamless mobility as that provided for cellular users. 802.16e includes fast base station switching and hard handoff mechanisms for inter-cell and inter-sector handover. In

802.16e, handoff process may be triggered for two reasons. One is due to fading of the signal, interference level, etc. within the current cell or sector. The other is due to the fact that another cell can provide a higher level of QoS for the mobile station (MS). Furthermore, 802.16e supports macro diversity handovers and inter-technology roaming. Macro-diversity handovers support handoffs between different sized cells, while inter-technology roaming addresses MS handoffs from BS to backhaul or wired network by providing roaming authentication mechanisms.

**Power management** is a critical process for mobile applications to enable efficient operation of the MS. 802.16e defines two power management operations, sleep mode and idle mode. Idle mode operation is carried out by MS when the MS does not intend to register to a specific BS as the MS traverses a region covered by multiple BS. The advantage of idle mode for the BS is to avoid multiple handoffs and other normal operations while the SS is traversing the region, and for the BS and network is to avoid unnecessary handoffs from an inactive MS. When the MS enters the idle mode, it needs to periodically check for broadcast messages sent by the BS to see if new downlink frames have been sent to it. Sleep mode operation is a state in which MS sends a request to be unavailable to the BS. If the BS responds with approval, the MS is provided with a sleep interval time vector that determines the length of the sleep mode period. The benefit of the sleep mode operation is to minimize MS power usage and utilization of the air interface resources of the BS. While the MS is in the sleep mode, the MS scans other BSs to collect information required for handover during the sleep mode.

## 3. WHY WiMAX AS SOLUTION?

Compared with other wired solution such as ADSL, or any other wireless or satellite system, WiMAX based access networks will enable operators and service providers to cost-effectively reach million of new potential customers providing them with broadband ICTs access. This is even true for developing countries and rural areas for which the cost/profitability and the demand factors are essential. This obviously includes adequate coverage, reliability, performances (throughput), capacity and applications. Table 1 shows how WiMAX supports different types of access and their requirement.

| Definition | Devices | Location /Speed | Handoffs | 802.16-2004 | 802-16e |
|---|---|---|---|---|---|
| **Fixed** | Outdoor and indoor | Single/stationary | No | Yes | Yes |
| **Nomadic access** | Indoor CPEs, PCMCIA cards | Multiple/stationary | No | Yes | Yes |
| **Portability** | Laptop PCMCIA or Mini-cards | Multiple/walking Speed | Hard handoff | No | Yes |
| **Simple mobility** | Laptop PCMCIA or Mini cards PDAs or smartphone | Multiple/low Vehicular Speed | Hard handoff | No | Yes |
| **Full mobility** | Laptop PCMCIA or Mini cards PDAs or smartphone | Multiple/high Vehicular Speed | Soft handoff | No | Yes |

Table 1. Types of Access to a WIMAX Network

The WiMAX standard has been developed with many objectives in mind. These objectives are as follows:

**Flexible Architecture:** WiMAX supports several system architectures, including Point-to-Point, Point-to-Multipoint and ubiquitous coverage. The WiMAX MAC (Media Access Control) supports Point-to-Multipoint and ubiquitous service by scheduling a time slot for each Subscriber Station (**SS**). If there is only one SS in the network, the WiMAX Base Station (**BS**) will communicate with the SS on a Point-to- Point basis. A BS in a Point-to-Point configuration may use a narrower beam antenna to cover longer distances. Wireless is more flexible and thus easier to deploy according to the market demand. Although most of the existing wireless

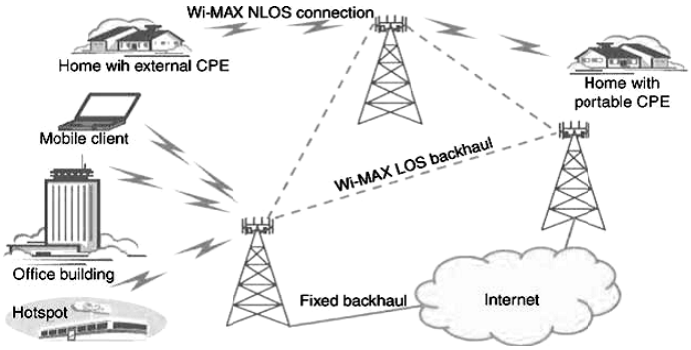



Fig. 3. A WiMAX typical scenario

technologies suffer from limited range and coverage (usually a few hundred meters around the base station) resulting in very costly combination of technologies (wired and/or wireless), WiMAX technology benefits in a wide coverage and can be deployed as a Point-to-Multipoint "last mile'' connection but also as part of the backhaul to the private switched telephone network (PSTN) and Internet access points. The WiMAX role in an access network is illustrated in Figure 3. With potential range of 30 to 50 kilometers in line-of-site conditions, WiMAX offer a huge improvement over all existing broadband wireless technology.

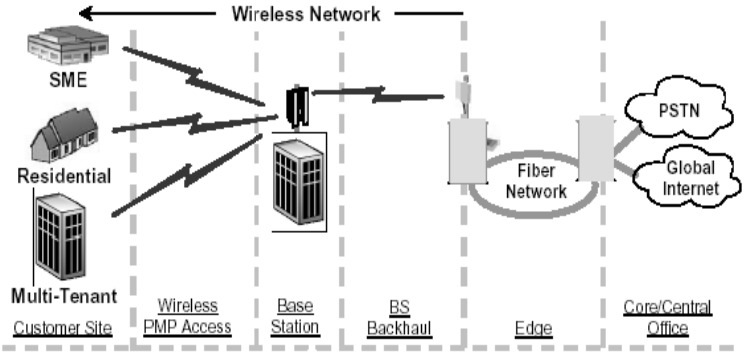



Fig. 4. WiMAX Network Description

**Quick Deployment:** Compared with the deployment of wired solutions, WiMAX requires little or no external plant construction. Ease of installation is one of the key issues to lower

deployment costs in developing countries. In rural areas, the consequences of the long distances from the core network access point and the scattered location of villages, farms in the countryside makes any deployment very costly. In developing countries, the lack of main infrastructure (electricity, roads), and environmental condition (temperature, humidity) adds on the difficulty. Thanks to the non-line-of-sight (NLOS)/LOS coverage advantage, the service provider can easily plans a 95% predictability coverage ensuring high installation success rates and controls deployment costs. A quicker and simpler installation with a much greater rate of success means operators spend less money rolling out their networks. WiMAX NLOS capability also allows indoors self install CPEs within several kilometers radius.

**Worldwide Standardization:** Developed and supported by the WiMAX forum (more than 300 members), WiMAX will become the worldwide technology based standard for broadband and will guaranty *interoperability* (i.e., multivendor CPEs), *reliability* and evolving technology, but also will ensure equipments with very low cost. With low CPEs cost as one of the first objective, business model can easily be profitable even in developing countries.

**Mobility:** The IEEE 802.16e amendment has added key features in support of mobility. Improvements have been made to the Orthogonal Frequency Division Multiplexing (OFDM) and OFDMA physical layers to support devices and services in a mobile environment. These improvements, which include Scalable OFDMA, Multiple input, multiple Output (MIMO), and support for idle/sleep mode and hand-off, will allow full mobility at speeds up to 160 km/hr. The WiMAX Forum-supported standard has inherited OFDM's superior NLOS performance and multipath-resistant operation, making it highly suitable for the mobile environment.

**Wider Coverage:** Even more important than the range limitation, the coverage (i.e. the capability to reach any potential customer within the base station covering area) is essential for the operator/service provider. While many currently available wireless broadband solutions can only provide line-of-site coverage, WiMAX due to its OFDM technology, has been optimized to provide excellent non-line-of-site coverage (up to 15Km around the base station) and long range transmission up to 50Km in LOS conditions. Combining both LOS and NLOS coverage, WiMAX is the ideal solution for getting the requested coverage in the most economical way.

**High Capacity:** Using higher modulation (64-QAM) and channel bandwidth (currently 7 MHz, with planned evolution towards the full bandwidth specified in the associated IEEE and ETSI standards), WiMAX systems can provide significant bandwidth to end-users.

**Spectrum Flexibility:** In line with the objective to become the worldwide standard based technology for broadband, WiMAX use single radio covering all licensed and unlicensed frequency bands allocated by the ITU for such services. These are:

- The two licensed bands 3.3-3.8 GHz and 2.3-2.7 GHz
- One license exempt band 5.725-5.85 GHz.

In addition to the flexibility offered to address all national spectrum situations, this single radio, will makes base stations and customer premises equipment costs very attractive.

**Multi-application Technology:** Following the normal trend of digitalization and packet transmission and switching, WiMAX uses the Internet protocol and thus supports all multimedia services from Voice over IP to high speed internet and video transmission. WiMAX allows service providers to offer all the latest generation of services and beyond, thanks to a throughput up to tens of Mbps. With regard to the potential users, this means that WiMAX have the capacity to deliver services from households to small and medium

enterprises (SMEs), small office home office (SOHO), Cybercafés, Multimedia Telecentres, Schools and Hospitals. See Table 2 for more WiMAX range of applications.

## 4. WiMAX Applications

WiMAX was developed to become a last mile access technology comparable to DSL, cable and $E_1/T_1$ technologies. It is a rapidly growing technology that is most viable for backhauling the rapidly increasing volumes of traffic being generated by Wi-Fi hotspots. WiMAX is a MAN technology that fits between wireless LANs, such as 802.11, and wireless wide-area networks (WANs), such as the cellular networks. WiMAX can serve in applications such as cellular backhaul systems, in which microwave technologies dominate, backhaul systems for Wi-Fi hot spots and most prominently as residential and business broadband services. WiMAX is billed to support many types of wireless broadband connections including but not limited to the following: high-bandwidth MANs, cellular backhaul, clustered Wi-Fi hotspot backhaul, last-mile broadband, cell phone replacements and other miscellaneous applications such as automatic teller machines (ATMs), vehicular data and voice, security applications and wireless VoIP.
Today, wherever available, these applications use expensive, proprietary methods for broadband access. WiMAX was developed to provide low-cost, high-quality, flexible, BWA using certified, compatible and interoperable equipments from multiple vendors. As WiMAX is based on interoperability-tested systems that were built using the IEEE 802.16 standard-based silicon solutions, WiMAX will reduce costs. WiMAX is well placed to address challenges associated with traditional wired access deployment types such as:

i. Large area coverage access, covering a large area (also referred to as hot zones) around the base station and providing access to 802.16 clients using point-to-multipoint topology;
ii. Last-mile access, connecting residential or business subscribers to the base station using point-to-multipoint topology;
iii. Backhaul, connecting aggregate subscriber sites to each other and to base stations across long distances using point-to-point topology.

In summary, the WiMAX standard has been developed to address a wide range of applications as shown in Table 2.

| **Class description** | **Real Time?** | **Application type** | **Bandwidth** |
|---|---|---|---|
| **Interactive Gaming** | Yes | Interactive gamig | 50 – 85 Kbps |
| **VoIP,Video conference** | Yes | -VoIP<br>-Video phone | 4 - 64 Kbps<br>32 – 384Kbps |
| **Streaming media** | Yes | -Music/Speech<br>-Video clips<br>-Movies streaming | 5 – 128Kbps<br>20 - 384Kbps<br>>2 Mbps |
| **Information Technology** | No | -Instant messaging<br>-Web browsing<br>E-mail(with attachment) | <250byte messages<br>>500 Kbps<br>>500 Kbps |
| **Media Content Download Store and forward** | No | Bulk data, Movie Download | >1Mbps<br>>500 Kbps |

Table 2. WiMAX Ranges of Applications

Some more disruptive applications of WiMAX can be:

- Remote monitoring of patients' vital signs in health-care facilities to provide continuous information and immediate response in the event of a patient crisis.
- Mobile transmission of maps, floor layouts and architectural drawings to assist fire-fighters and other response personnel in the rescue of individuals involved in emergency situations.
- Real-time monitoring, alerting and control in situations involving handling of hazardous materials.
- Wireless transmission of fingerprints, photographs, warrants and other images to and from law-enforcement field personnel.

## 5. Usage Scenarios

Based on its technical attributes and service classes, WiMAX is suited to supporting a large number of usage scenarios. WiMAX technology will revolutionize the way we communicate. It will provide total freedom to people who are highly mobile, allowing them to stay connected with voice, data and video services. It will allow people to go from their homes to their cars, and then travel to their offices or anywhere in the world, all seamlessly. To illustrate the ability of WiMAX to address the applications outlined in Table 2, several representative usage scenarios are outlined in Table 3.

| Technical Attribute / Usage Scenarios | Flexible Arch | High Security | Wimax QoS | QuickDeployment | Interoperability | Portability | Mobility | Cost Effecttive | Wider Coverage | NLOS | High Capacity |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Cellular Backhaul** | | | | * | | | | * | | | * |
| **Wireless Service Provider** | | | | * | | | | * | | | * |
| **Banking Network** | * | * | * | | | | | * | | * | |
| **Education Network** | * | | * | | | | | * | * | | |
| **Public Safety** | * | * | * | * | | * | * | | | * | |
| **Offshore Comms.** | * | | * | | | * | * | | * | * | |
| **CampusConnectivity** | * | * | * | | | | | | | | * |
| **Theme Parks** | * | | * | | | * | * | | | * | |
| **Rural Connectivity** | | | * | | * | | | * | * | | |

Table 3. WiMAX usage scenarios

### 5.1 Cellular Backhaul

Majority of backhaul is done by leasing $E_1$ services from incumbent wire-line operators. With the WiMAX technology, cellular operators will have the opportunity to lessen their independence on backhaul facilities leased from their competitors. Outside the United States, the use of point-to-point microwave is more prevalent for mobile backhaul, but WiMAX can still play a role in enabling mobile operators to cost-effectively increase backhaul capacity using WiMAX as an overlay network as depicted in Figure 5. This overlay

approach will enable mobile operators to add the capacity required to support the wide range of new mobile services they plan to offer without the risk of disrupting existing services.
Some salient points about WiMAX use as cellular backhaul are:

- high-capacity backhaul;
- multiple cell sites are served;
- there is capacity to expand for future mobile services;
- It is a lower cost solution than traditional landline backhaul.

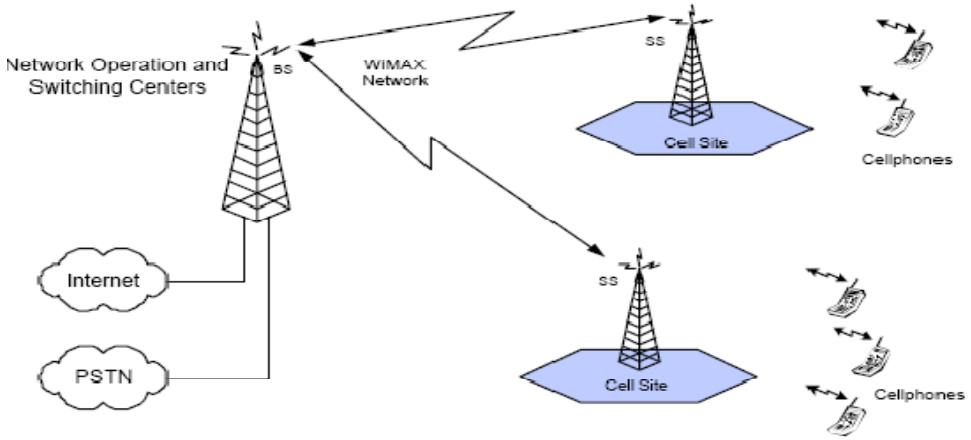



Fig. 5. Cellular backhaul with WiMAX

### 5.2 Education Networks

School boards can use WiMAX networks to connect schools and school board offices within a district, as shown Figure 6. Some of the key requirements for a school system are NLOS, high bandwidth (>15 Mbps), Point-to-Point and Point-to-Multipoint capability, and a large coverage footprint. WiMAX-based education networks, using QoS, can deliver the full range of communication requirements, including telephony voice, operating data (such as student records), email, Internet and intranet access (data), and distance education (video) between the school board office and all of the schools in the school district, and between the schools themselves. The WiMAX solution provides broad coverage, making it very cost-effective, particularly for rural schools, which may have little or no communications infrastructure, and which are widely dispersed. When school boards own and operate their own network, they can be responsive to changes in the location and layout of their facilities. This will significantly reduce the annual operating cost of leased lines. Wired solutions cannot offer a quickly deployable, low-cost solution, and most versions of DSL and cable technology do not have the throughput required by these education networks.

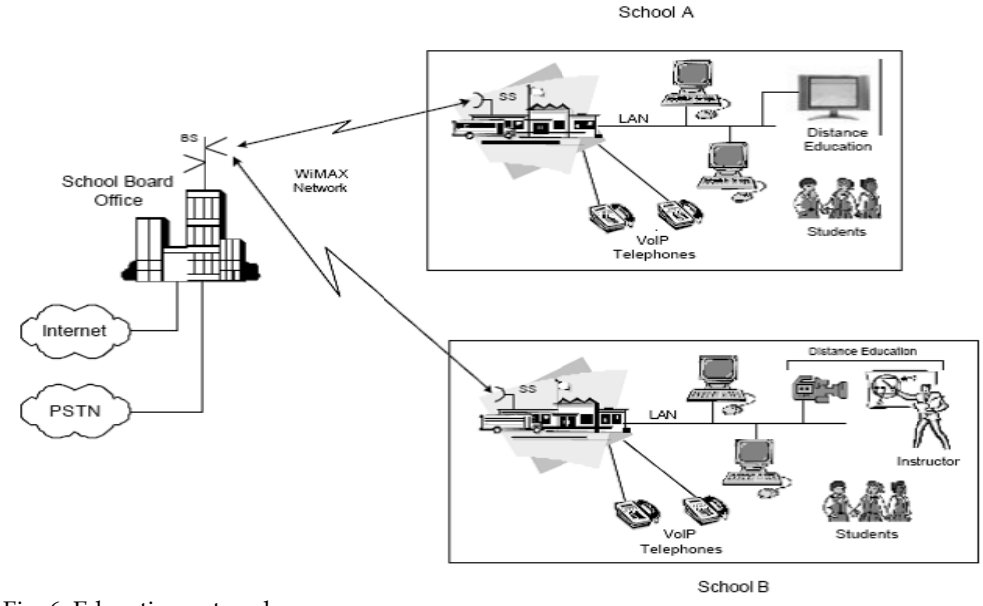



Fig. 6. Education networks

### 5.3 Campus Connectivity

Government agencies, large enterprises, universities, and colleges, can use WiMAX networks to connect multiple locations, sites and offices within their campus, as shown in Figure 7. Campus systems require high data capacity, low latency, a large coverage footprint, and high security. Campus networks carry a mix of voice, data, and video, which the WiMAX QoS helps prioritize and optimize. It takes less time and resources to interconnect a campus through a WiMAX network, since excavation and external construction are not required. Some campuses have been around for a long time, and digging trenches for cable may not be permitted. In such cases, WiMAX solutions may be one of the most effective ways to interconnect campus buildings. Even if wired installations are permitted, the lead-time to deploy a wired solution is much longer than the lead-time to deploy a WiMAX solution, without offering any accompanying benefits.

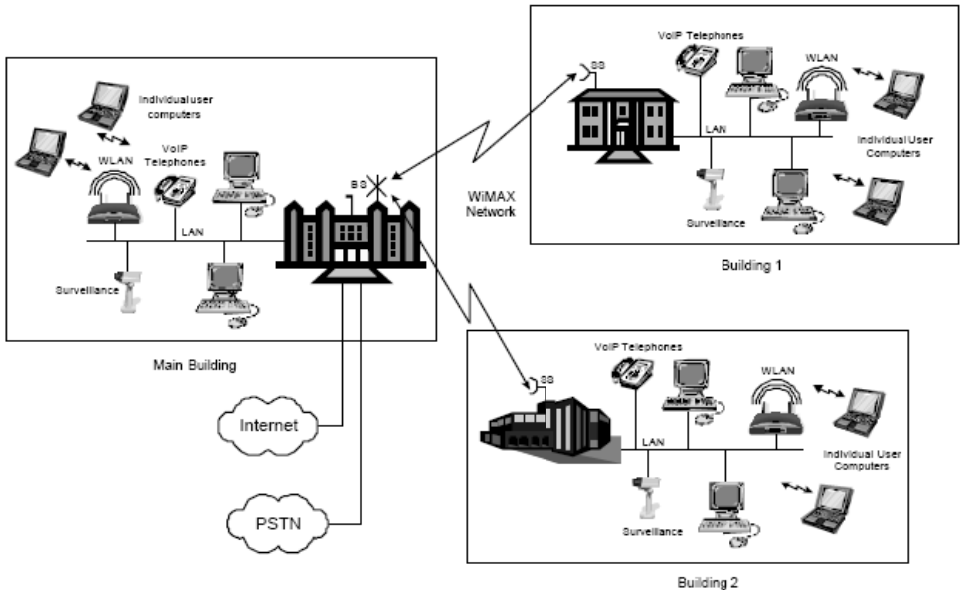



Fig. 7. Campus connectivity

### 5.4 Rural Connectivity

Service providers use WiMAX networks to deliver service to underserved markets in rural areas and the suburban outskirts of cities as shown in Figure 8. The delivery of rural connectivity is critical in many developing countries and underserved areas of developed countries, where little or no infrastructure is available. Rural connectivity delivers much-needed voice telephony and Internet service. Since the WiMAX solution provides extended coverage, it is a much more cost-effective solution than wired technology in areas with lower population densities. It solutions can be deployed quickly, providing communication links to these underserved areas, providing a more secure environment, and helping to improve their local economies.

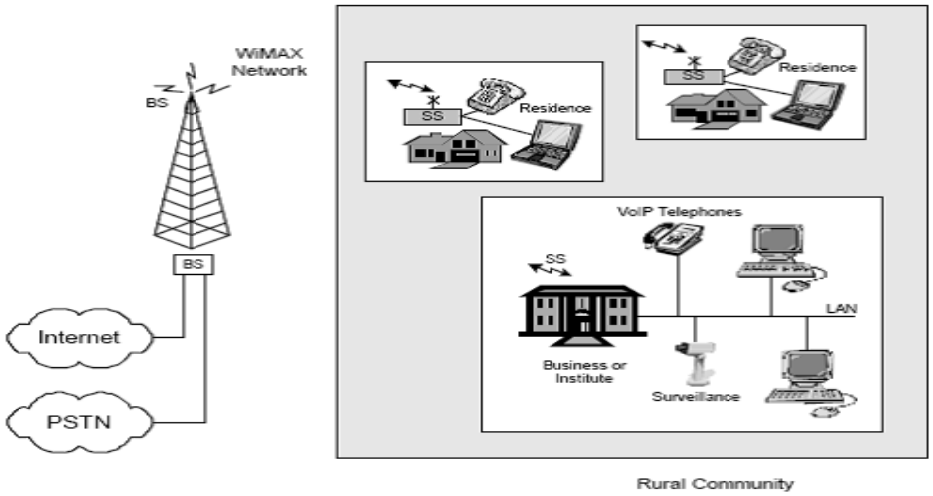



Fig. 8. Rural connectivity

### 5.5 Wireless Service Provider Access Network

Wireless Service Providers (WSPs) use WiMAX networks to provide connectivity to both residential (voice, data and video) and business customers (primarily voice and Internet), as illustrated in Figure 9. The WSP could be a CLEC (Competitive Local Exchange Carriers) that

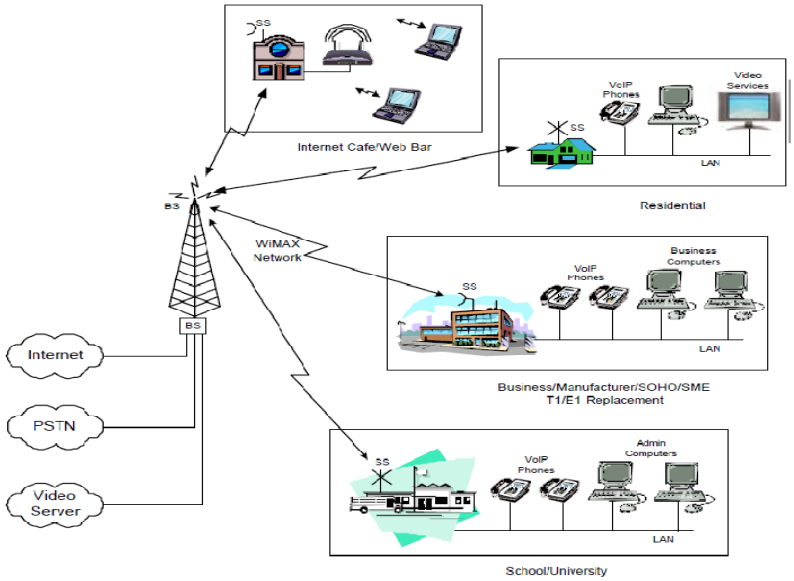



Fig. 9. Wireless Service Provider Access Network

is starting its business with little or no installed infrastructure. Since WiMAX is easy to deploy, the CLEC can quickly install its network and be in position to compete with the ILEC (Incumbent Local Exchange Carrier). A common network platform, offering voice, data and video, is highly attractive to end customers, because it presents a one-stop shop and a single monthly bill. Support for multiple service types allows for different revenue streams, yet it reduces customer acquisition cost, and increases ARPU (Average Revenue Per User). The WSP needs only one billing system and one customer database. Cellular operators may also be interested in applying WiMAX in their networks. These operators already have towers, billing infrastructure and a customer base in place, but the deployment of a WiMAX solution will expand their market presence in their service area.

### 5.6 Theme Parks

Theme park operators can use WiMAX to deliver a broad range of communication services for their amusement parks, expositions, hospitality and operation centers, and buses and service vehicles, as shown Figure 10. The network can support a wide range of communications traffic, including two-way dispatch from a control center, video surveillance throughout the park, reservation data, inventory database access and update, site status monitoring, video on demand, and voice telephony. Some of the key requirements for a system like this are support for fixed and mobile operations, high security, scalable architecture and low latency. The broad coverage range of WiMAX means an entire park can be covered from only a few numbers of Base Stations, scalable upwards as capacity requirements increase.

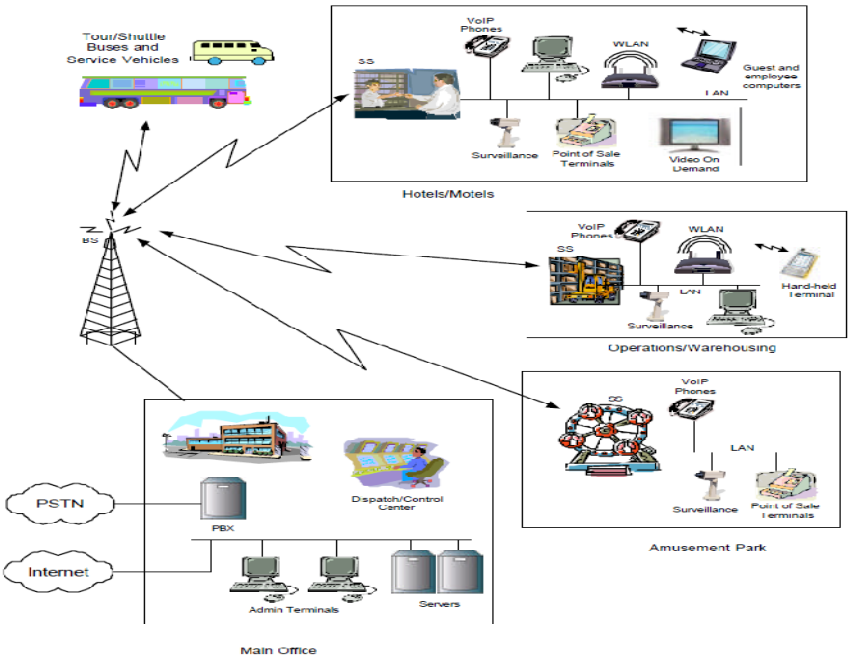



Fig. 10. Theme Parks

Re-deployment of the network, in response to changes in theme park facilities, is straightforward and simple, unlike the changes that would be required had the park been served by wired facilities, such as DSL or cable. WiMAX mobility capability will support two-way voice and data communications to the theme park's tour buses and service vehicles. Real-time video can be broadcast to tour buses, providing tourist information, promotions, and weather to passengers.

## 6. Conclusion

Demand for wireless broadband access is growing fast and embracing an ever-widening range of applications that encompass fixed, nomadic, portable and mobile data access as well as fixed and mobile voice services, and content streaming. Looking back to the barriers as summarized by ITU in Figure 2 and section 3, WiMAX appears clearly as the solution to favor the broadband ICTs access in developing countries. Undoubtedly, WiMAX is a new powerful broadband wireless technology aiming at providing a universal ubiquitous and equitable and affordable access to ICTs infrastructure and services, and thus highly contributing to bridge the "Digital Divide".

## 7. References

[1] IEEE Std. 802.16-2004, IEEE Standard for Local and Metropolitan Area Networks, part 16, Air Interface for Fixed Broadband Wireless Access Systems, *IEEE Press*, 2004.

[2] IEEE Std. 802.16e-2005, IEEE Standard for Local and Metropolitan Area Networks, part 16, Air Interface for Fixed and Mobile Broadband Wireless Access Systems, *IEEE Press*, 2005.

[3] IEEE 802.16 2004: "Air Interface for Fixed Broadband Wireless Access Systems," IEEE STD 802.16 – 2004, October, 2004.

[4] IEEE 802.16e 2005: "Air Interface for Fixed and Mobile Broadband Wireless Access Systems," IEEE P802.16e/D12, February, 2005.

[5] ITU World Telecomunication Development report, 2003.

[6] 802.16.3c-01/29r4, "Channel Models for Fixed Wireless Applications," http://www.ieee802.org/16.

[7] LAN/MAN Standards Committee, *IEEE* standard *802.16e-2005 and 802.16- 2004*/Cor1-2005, IEEE Computer Society and IEEE Microwave Theory and Techniques Society.

[8] Spanning the Digital Divide, bridges.org, Washinton DC.

[9] World Summit on the Information Society (WSIS) Declaration of Principles and Plan of Action, Geneva, December 2003.

[10] WiMAX Forum white paper: "WiMAX's technology for LOS and NLOS environments".

[11] WiMAX Forum white paper: "Fixed, nomadic, portable and mobile applications for 802.16-2004 and 802.16e WiMAX networks", November 2005.

[12] WiMAX Forum white paper: "Can WiMAX Address your Applications?" *Westech press,* on Behalf of the WiMAX Forum. October, 2005.

[13] D. Gu and J. Zhang, "QoS enhancement in IEEE 802.11 wireless local area networks," *IEEE Commun. Mag.*, vol. 41, pp. 120–124, June 2003.

[14] "Wireless LAN medium access control (MAC) and physical layer (PHY) specifications-amendment 8: Medium access control (MAC) quality of service enhancements," IEEE Std. 802.11e-2005, 2005.
[15] Yoshio Utsumi. June 2005. Press conference of the ITU Secretary-General in Geneva, Swiss.
[16] Pareek, D., 2005. WiMAX: Taking Wireless to the Maximum, CRC Press.
[17] WiMAX Forum. 2004. Business Case Models for Fixed Broadband Wireless Access based on WiMAX Technology and the 802.16 Standard.

# The Role of WiMAX Technology on Broadband Access Networks: Economic Model

João Paulo Ribeiro Pereira
*Computer Science Department, Polytechnic Institute of Bragança*
*Portugal*

## 1. Introduction

The increasing demand of quadruple-play services, including video, voice, data and mobility, has created new challenges to the modern broadband wireless/wired access networks (Pereira & Ferreira, 2009). Moreover, with the proliferation of broadband applications, the need for comprehensive broadband infrastructure becomes critical (Fong & Nour, 2004).
Broadband in the OECD is still dominated by DSL (Digital Subscriber Line) but there is an obvious trend emerging to upgrade last-mile access networks to support the new services requirements (OECD, 2008). To address these network requirements, many carriers in emerging markets will have to move from legacy platforms toward next-generation solutions with a combination of wireless and wireline technologies, such as worldwide interoperability for microwave access (WiMAX), IP-Ethernet, new forms of DSL technology and fiber.
In this context, several broadband access technologies are being deployed to address the bandwidth bottleneck for the "last mile," the connection of homes and small businesses to this infrastructure. There are many competing technologies which can provide the bandwidth required to deliver broadband services, but eacch technology has its limits in terms of bandwidth, reliability, cost or coverage (Corning, 2005). Some of the most important last-mile solutions include xDSL, HFC, FTTH, PLC, Satellite, and the fixed broadband wireless access network.
In general, broadband access technologies can be classified by the physical medium in two major groups, namely wired and wireless technologies (Corning, 2005;Fernando, 2008;Ims *et al.*, 2004).

### 1.1 Broadband Wired access technologies

Dominant broadband wired access technologies today are DSL and Hybrid Fiber Coaxial (HFC) networks. Nevertheless, copper based access networks still represent the overwhelming share of fixed telecommunications access network infrastructures around the world (Fernando, 2008;Sigurdsson, 2007).
DSL networks have been gradually developing over the past several decades but all share the same principle of connecting households within a given area to a central office with an

individual pair of Unshielded Twisted Pair (UTP), called the local loop (Figure 1). DSL technology uses existing twisted-pair phone lines for broadband connection. The main limitation of this technology is that the transmission speeds decreases when copper local loop length increases.

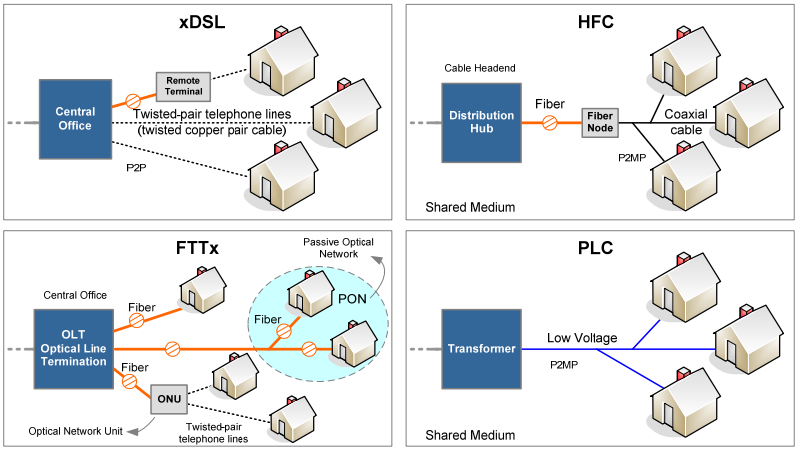



Fig. 1. Wired access networks

An HFC network is normally a cable television network that includes a combination of optical fiber (between head end and the fiber node) and coaxial cables (from fiber nodes to CPE). Initially, the original HFC networks were designed to support only downstream broadcasts. Currently, upstream traffic must also be supported using Cable Modems (CM) for full Internet access. The bandwidth provided by cable networks will allow for 160 Mbps downstream and 120 Mbps upstream for end-users (using Docsis 3.0 standard). HFC uses a shared medium, therefore the bandwidth will also have to be shared by end-users. Typically there are 500-1000 subscribers in a single local distribution point, which can be brought down to about 250 on average.

The FTTH (PON) technology connects Optical Line Terminal (OLT) at the Central Office (CO) to Optical Network Units (ONUs) at subscribers homes and businesses over one wavelength in the downstream and another wavelength in the upstream (from ONUs to OLT). This technology provides much higher bandwidth than DSL and Cable Modem. A PON can cover a maximum distance of 20 km from the OLT to the ONU. Under normal conditions, PON allows 16 ONUs at a maximum distance of 20 km from the OLT and 32 ONUs at a maximum distance of 10 km from the OLT supporting up to 1 Gbps.

Power Line Communications is a broadband access technology that uses the low and medium voltage electricity grid to provide telecommunication services (Little, 2004). The Low Voltage (LV) electrical power distribution network represents the most attractive medium for high-speed digital communication purposes due to an ever-increasing demand caused by the advances in communication and information technologies (Lim *et al.*, 2000). Notwithstanding the benefits that the availability of an extensive infrastructure can allow, for the moment the service provision is far from standardized, and the capacity of bandwidth provided through BPL is still being questioned (Sarrocco & Ypsilanti, 2008).

### 1.2 Broadband Wireless Access Networks

Generally, wireless broadband refers to technologies that use point-to-point or point-to-multipoint microwave in various frequencies between 2.5 and 43 GHz to transmit signals between hub sites and an end-user receiver. At network level, they are suitable for both access and backbone infrastructure so, it is in the access network where wireless broadband technology is actually proliferating. As a consequence, the terms "wireless broadband" and "wireless broadband access" are used interchangeably (Corning, 2005). However, current wireless access networks show limits in terms of data rate and QoS (El Zein & Khaleghi, 2007).

Wireless broadband is the "great equalizer" of broadband technologies. It enables even those without twisted copper, coaxial cable, fiber, or satellite to enter competition for the broadband dollar. It therefore expands the definition of who can be a carrier (WiMAX Forum, 2005b). Unlike fixed technologies, wireless solutions are flexible and scalable, both in terms of their physical layout and their logical architecture. Due to the cost advantages associated with wireless and the use of unlicensed spectrum bands, infrastructure investments are now within the reach of local organizations which are able to better understand local conditions. Covering sparsely populated areas is not economically viable for the use of conventional techniques. The newest wireless technologies encourage a different model of infrastructure development, better suited to the challenges of extending Internet connectivity to rural areas, and thus becoming a solution that offers wide-area coverage well suited to small villages (Sarrocco & Ypsilanti, 2008) (Fernando, 2008). The infrastructure can therefore expand from the bottom-up, without a preconceived plan, linked to the needs and attributes (geographical, demographic, and economic) of local communities. The distribution of telecommunications infrastructures has traditionally been associated with hefty investment programs undertaken by large entities such as telecommunications operators and government agencies (Galperin, 2005).

Wireless networking technologies can be divided into three main categories (Carcelle *et al.*, 2006). Each addresses specific requirements and purposes in point-to-point and point-to-multipoint communication (see Fig. 2).

The range performances for a WPAN is typically from 1 meter to a few dozens meters, and is designed for low data rate (usually 100-200 kbps). WPAN technologies comprise: ZigBee, Bluetooth and UWB. In WLAN technologies the range performances are typically from a few dozen meters indoors to a few hundred meters outdoors, and are designed for high data rate (usually 1 to 20 Mbps). One technology of this family is Wi-Fi. The WWAN is mainly focused for long-distance point-to-point connections with high data rate.

Wireless broadband access systems have been deployed at frequencies ranging from 400 MHz to about 30 GHz (Lehr *et al.*, 2004). Different parts of the spectrum have dramatically different physical properties. For example, at 30 GHz, signals attenuate rapidly with distance, they are limited by rainfall, and to Line-of-Sight (LOS) deployments. These frequencies have been used most often as a substitute for high-capacity point-to-point links such as those used by service providers and large enterprise customers to connect backbone nodes or buildings. In contrast, at lower frequencies, below say 10GHz, the spectrum is much more encumbered by legacy incumbents and licensees are typically limited to smaller bandwidth channels. However, the spectrum is less susceptible to interference from rainfall and can operate in Non-Line-of-Sight (NLOS) situations (e.g. it reaches inside buildings to a desktop or mobile antenna), and requires less costly technology. The lower the frequency,

the better the penetration of buildings or of foliage, besides immunity to rainfall, but there is less bandwidth available.

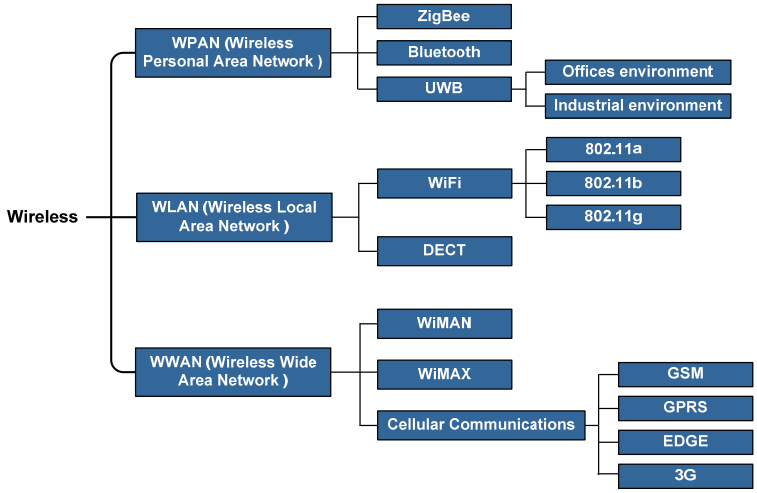



Fig. 2. Wireless technologies taxonomy (Carcelle *et al.*, 2006)

If we look from the line of sight perspective, wireless technologies can be broadly categorized into those requiring Line-of-Sight (LOS) and those that do not (NLOS) (Corning, 2005): Line of sight means that there is an unobstructed path from the CPE antenna to the access point antenna. If the signal can only go from the CPE to the access point by being reflected by objects, such as trees, the situation is called non-line of sight. NLOS systems are based on OFDM, which combats multipath interference, thereby permitting the distance between the CPE and the access point to reach up to 50 kilometers in the MMDS band. However, NLOS systems are more expensive than LOS systems (Ibe, 2002).

## 2. WiMAX

WiMAX (Worldwide Interoperability for Microwave Access) is a standardized form of wireless metropolitan area network (WMAN) technology that has historically been based on proprietary solutions, such as MMDS and LMDS. The first version of the IEEE 802.16 standard was completed in October 2001 and defines the air interface and medium access control (MAC) protocol for a wireless metropolitan area network, intended to provide high-bandwidth wireless voice and data for residential and enterprise use (Ghosh *et al.*, 2005). This standard was followed by the 802.16a standard in early 2003. Both standards support peak data rates up to 75 Mbps and have a maximum range of about 50 km. Because WiMAX systems have the capability to address broad geographic areas without the costly infrastructure requirement to display cable links to individual sites, the technology may prove less expensive to expand and should lead to more ubiquitous broadband access (Peng & Wang, 2007).

Wireless broadband promises to bring high-speed data to multitudes of people in various geographical locations where wired transmission is too costly, inconvenient, or unavailable

(Salvekar *et al*., 2004). The 802.16 standard uses Orthogonal Frequency Division Multiple Access (OFDMA), which is similar to OFDM in the way that it divides the carriers into multiple sub-carriers. OFDMA, however, goes a step further by then grouping multiple sub-carriers into sub-channels. A single client or subscriber station might thus transmit using all of the sub-channels within the carrier space, or multiple clients might also transmit with each using a portion of the total number of sub-channels simultaneously (Konhauser, 2006). In the RF front-end, WiMAX uses OFDM, which is robust in adverse channel conditions and enables NLOS operation. This feature simplifies installation issues and improves coverage, while maintaining a high level of spectral efficiency. Modulation and coding can be adapted per burst, ever striving to achieve a balance between robustness and efficiency in accordance with prevailing link conditions.
Service providers will operate WiMAX both on licensed and unlicensed frequencies. The technology enables long distance wireless connections with speeds up to 75 Mbps. This can provide very high data rates and extended coverage. However:

- 75 Mbps capacity for the base station is achievable with a 20 MHz channel at best propagation conditions. But regulators will often allow only smaller channels (10 MHz or less) reducing the maximum bandwidth.
- Even though 50 km is achievable under optimal conditions and with a reduced data rate (a few Mbps), the typical coverage will be around 5 km with indoor CPE (NLOS) and around 15 km with a CPE connected to an external antenna (LOS).
- To keep from serving too many customers and thereby greatly reducing each user's bandwidth, providers will want to serve no more than 500 subscribers per 802.16 base station (Vaughan-Nichols, 2004).

One of the main advantages of this technology is the capacity to deploy broadband services in large areas without physical cables. These characteristics give to telecommunication supplier the capacity to implement new broadband telecommunication infrastructures very quickly, and with a lower cost than the wired networks.
To sum up, the main advantages of the WiMAX technology in relation to other connection technologies are: it does not need cable installation, which can solve the access problem to remote places; it is rather quick to deploy. This technology could have an access velocity which is 30 times higher than basic ADSL technology. Besides frequency range is between 2 and 11 GHz, with the maximum range of 50 km from the base station, and data transmission to 70 Mbps. So, one BS sector can serve different businesses or many homes with DSL-rate connectivity. Another advantage is the high capacity to service modulation (data and voice), to perform symmetric transmission (the same velocity to send and receive data) and the use of QoS.

### 2.1 System Architecture

A fixed broadband wireless access network is essentially a sectorized network, composed of two key elements: base station (BS) and customer premises equipment (CPE). The BS connects to the network backbone and uses an outdoor antenna to send and receive high-speed data and voice to subscriber equipment, thereby eliminating the need for extensive and expensive wireline infrastructure and providing highly flexible and cost-effective last-mile solutions.
FWA base station equipment multiplexes the traffic from multiple sectors and provides an interface to the backbone network. For each sector, a radio transceiver module and a sector antenna is also required. The multiplexer (such as a switch) aggregates the traffic from the

different sectors and forwards it to a router that is connected to the service provider's backbone IP network (Ibe, 2002). The backbone connection can be provided with a point-to-point radio link or a fiber cable, and can be either IP or ATM-based. The distance between the CPE and the BS depends on how the system is designed and the frequency band in which it operates. The CPE with an indoor antenna can be installed by the customers themselves, whereas the outdoor antenna requires a technician to install it (Smura, 2004).
When we need to define a point-to-multipoint wireless system, several parameters are very important: the characteristics of the geographical area (for example, mountains), the subscriber density, the bandwidth required, QoS, the number of cells, etc. In areas with a low traffic demand and/or low subscriber density, the most important factor is the radio coverage whereas in areas with a high traffic demand and/or high subscriber density, capacity becomes a more important issue. Through a careful selection of network design parameters, tradeoffs can be made between coverage and capacity objectives to best serve the end users within the service area (Wanichkorm, 2002).

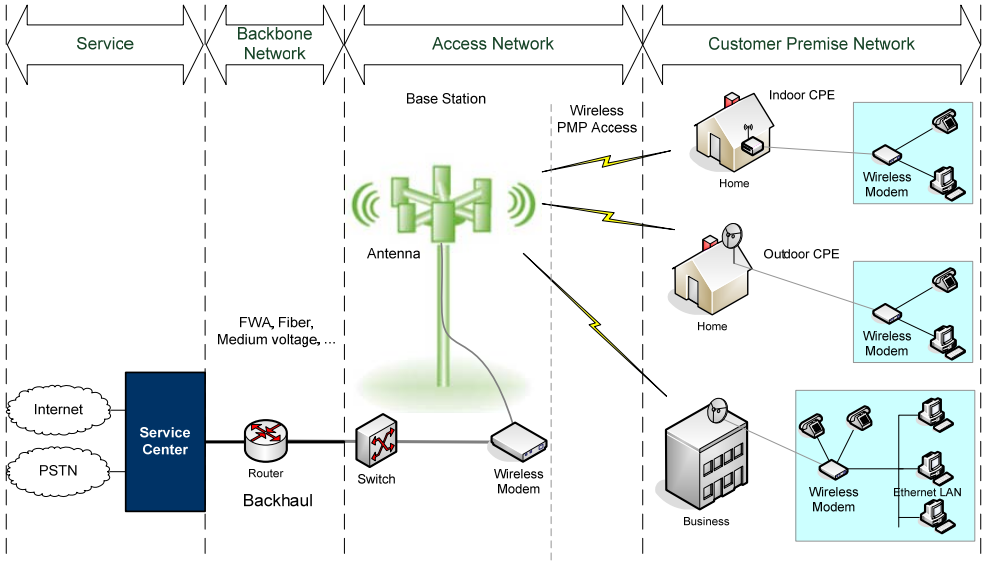



Fig. 3. WiMAX System Architecture

The WiMAX wireless link operates with a central BS through a sectorized antenna that is capable of handling multiple independent sectors simultaneously.

### 2.2 System Components

As previously referred to, base station equipment and customer premise equipment are the two main components of WiMAX architecture for the access network. The CPE enables a user in the customer's network to access Wide Area Network (WAN). The BS controls the CPEs within a coverage area, and consists of many access points or wireless hubs, each of which control the CPE in one sector. The following figure shows the basic components of a radio communication system.

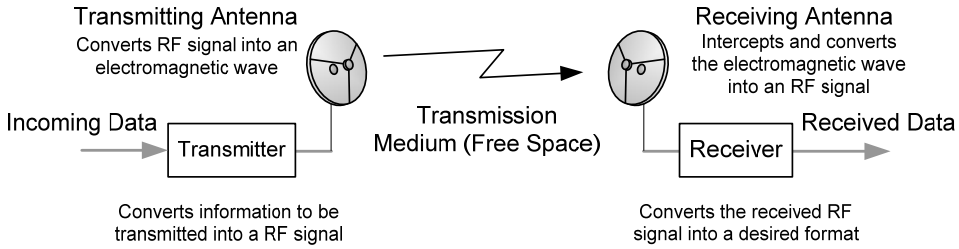



Fig. 4. Components of a radio communication system (Ibe, 2002)

### 2.2.1 Customer Premise Equipment – CPE

Residential CPEs are expected to be available in a fully integrated indoor self-installable unit as well as indoor/outdoor configuration with a high-gain antenna for use on customer sites with lower signal strength (Ohrtman, 2005). In most cases, a simple plug and play terminal, similar to a DSL modem, provides connectivity. For customers located several kilometers away from the WiMAX base station, an outdoor antenna may be required to improve transmission quality. To serve isolated customers, a directive antenna pointing to the WiMAX base station may be required.

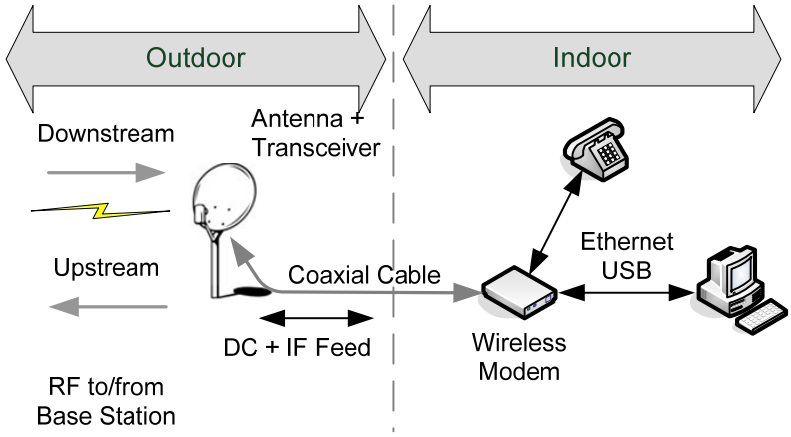



Fig. 5. FWA Subscriber Configuration (Outdoor CPE)

CPE or terminals are expected to be available in a number of configurations for customer specific applications and for different types of customers. Households in multi-tenant buildings can be served by installing a high throughput WiMAX outdoor unit with a low to medium capacity DSLAM (Digital Subscriber Line Access Multiplexer) as an in-building access device utilizing the in-building telephone wiring to reach individual apartments or by installing an individual WiMAX terminal in each household (WiMAX Forum, 2005a). These units are priced higher for the business case, consistent with the added performance (WiMAX Forum, 2004).

FWA CPE is often divided into three main components parts (Fig. 5): the modem, the radio, and the antenna. The modem device provides an interface between the customer's network and the fixed broadband wireless access network, while the radio provides an interface

between the modem and the antenna. As a matter of fact, some vendors integrate these two components to form a compact CPE, while others have the three units as standalone systems (Ibe, 2002). The CPE antenna type depends on the Non-Line-of-Sight capabilities of the system. In a Line-of-Sight FWA network, the CPE antennas are highly directional and installed outdoors by a professional technician. In Non-Line-of-Sight systems, the beamwidth of the CPE antenna is typically larger, and in the case of user-installable CPE's the antenna should be omnidirectional (Smura, 2004).

### 2.2.2 Base Station Equipment

The capacity of a single FWA base station sector depends on the channel bandwidth and the spectral efficiency of the utilized modulation and coding scheme. WiMAX systems take advantage of adaptive modulation and coding, meaning that inside one BS sector each CPE may use the most suitable modulation and coding type irrespective of the others (Smura, 2006).

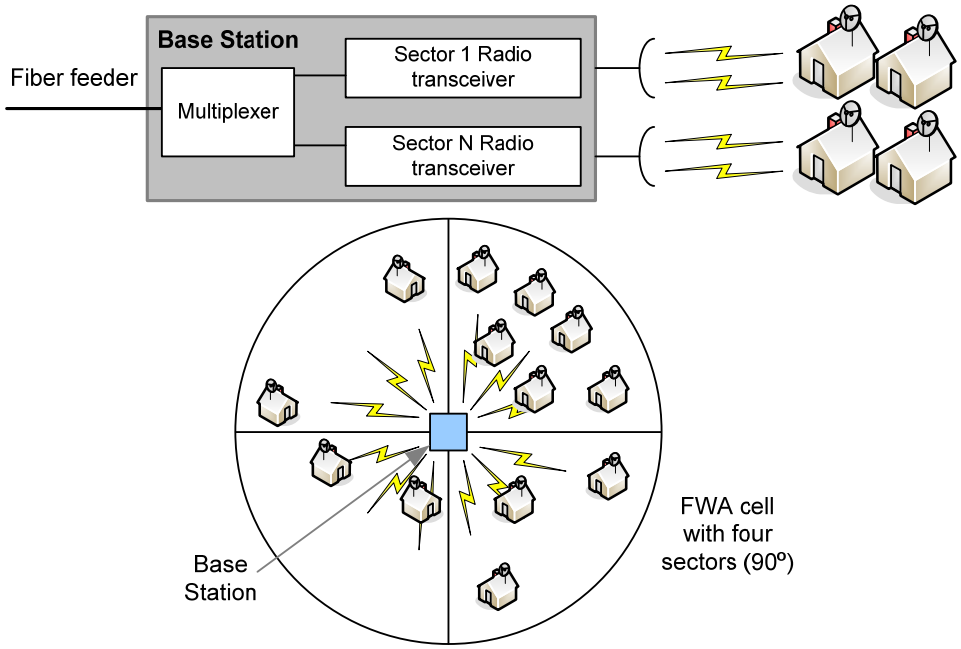



Fig. 6. Base Station components (Ufongene, 1999)

The base station equipment, like CPE, consists of two main building blocks: The antenna unit and the modulator/demodulator equipment (see Fig. 6 and Fig. 7). The antenna unit represents the outdoor part of the base station, and is composed of an antenna, a duplexer, a radio frequency (RF), a low noise amplifier and a down/up converter. The choice of antennas has a great impact on the capacity and coverage of fixed wireless systems.

The BS consists of one or more radio transceivers, each of which connects to several CPEs inside a sectorized area. In the BS one directional sector antenna is required for each sector.

Sector antennas are directional antennas and the beamwidth depends both on the service area and capacity requirements of the system. A BS with one sector using an omnidirectional antenna has a quarter of the capacity of a four-sector system (Anderson, 2003). The modem equipment modulates and mixes together each flow over the IF cable which is connected to the antenna unit.

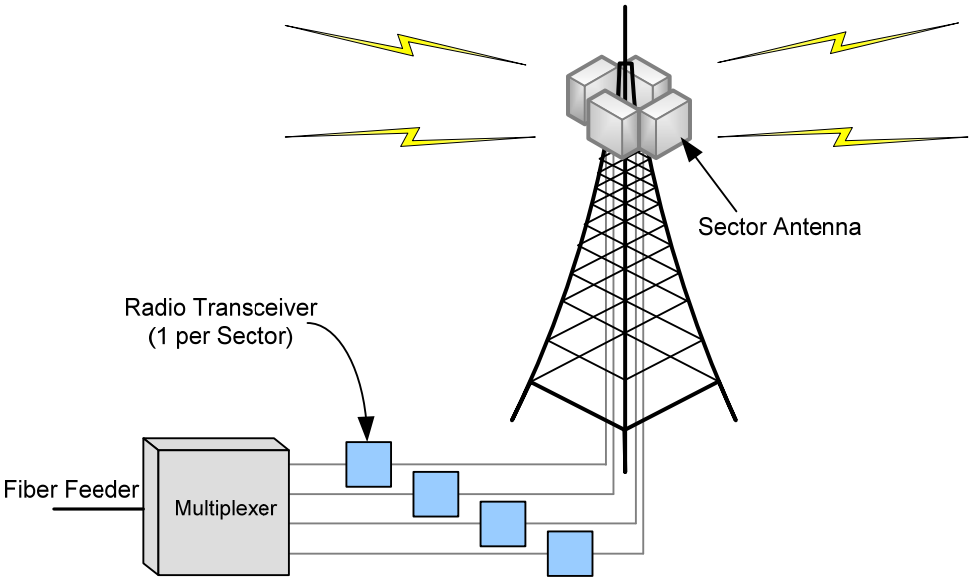



Fig. 7. Base Station components

As we can see in Fig. 7, each FWA base station consists of a number of sectors. The traffic capacities of these sectors depend most importantly on the modulation and coding methods, as well as on the bandwidth of the radio channel in use. The sector capacity is divided between all the subscribers in the sector's coverage area (Smura, 2004).

## 3. Techno-Economic Model

To support the new needs of the access networks (bandwidth and mobility), the proposed framework (Fig. 8) is divided into two perspectives (static and nomadic) and three layers. In the static perspective, users are stationary and normally require data, voice, and video quality services. These subscribers demand great bandwidth. In the nomadic/mobility perspective, the main preoccupation is mobility, and normally, the required bandwidth is smaller than the static layer (Pereira & Ferreira, 2009).

Focus of the wireless networks was to support mobility and flexibility while that of the wired access networks is bandwidth and high QoS. However, with the advancement of technology wireless networks such as WiMAX also geared to provide wideband and high QoS services competing with wired access networks recently (Fernando, 2008). The proposed model divides the area into several access networks (the figure is divided into 9

sub-areas, but the model can divide the main area between 1 and 36). The central office is located in the center of the area, and each sub-area will have one or more Aggregation Nodes (AGN) depending on the technology in use.

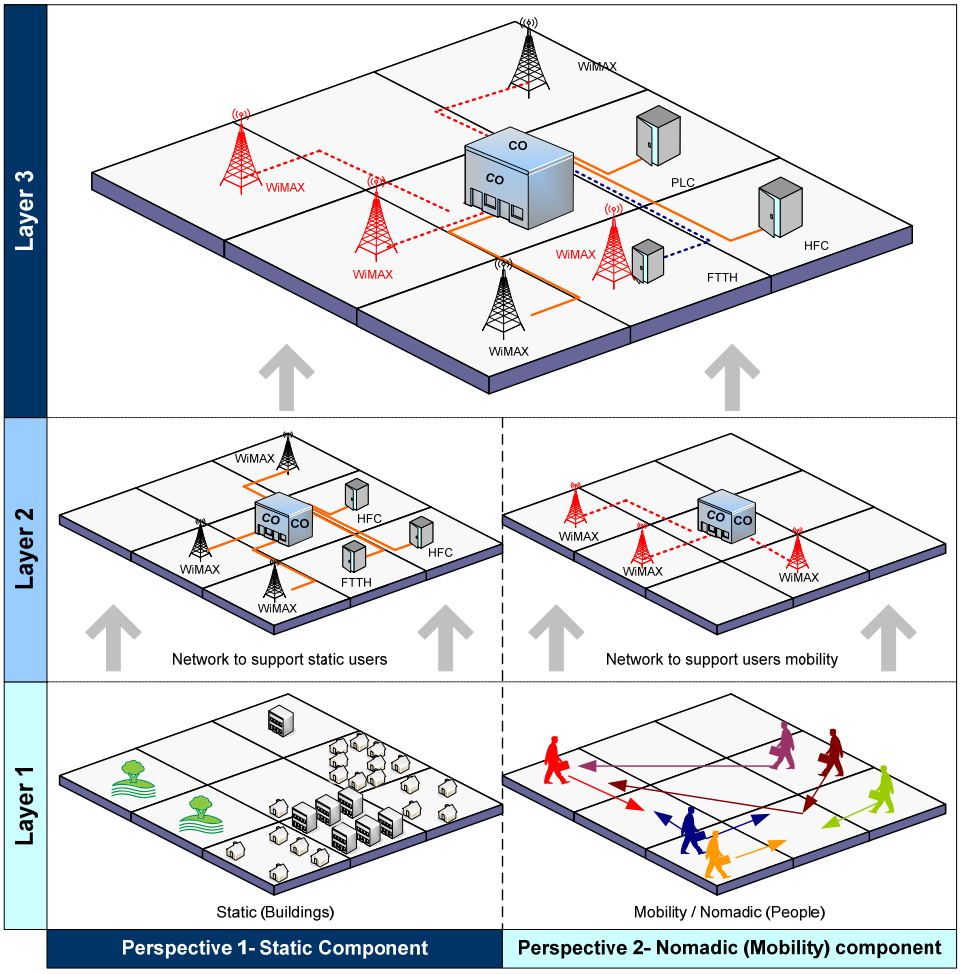



Fig. 8. Cost model framework architecture

As we can see in Fig. 8, the framework is separated into three main layers (Pereira, 2007a): (Layer A) Firstly, we identify the total households and SMEs (Static analysis) for each sub-area, as well as the total nomadic users (Mobility analysis). The proposed model initially separates these two components because they have different characteristics. In layer B, the best technology is analyzed for each Access Network, the static and nomadic components. For the static analysis we consider the following technologies: FTTH (PON), DSL, HFC, and WiMAX PLC. For the nomadic analysis we use the WiMAX technology. The final result of this layer is the best technological solution to support the different needs (Static and

nomadic). The selection of the best option is based on four output results: NPV, IRR, Cost per subscriber in year 1, and Cost per subscriber in year n. The next step (Layer C) is to create a single infrastructure that supports the two components. Bearing this in mind, the tool analyzes each Access Network which is the best solution (based on NPV, IRR, etc). Then, for each sub-area we verify if the best solution is: a) the use of wired technologies (FTTH, DSL, HFC, and PLC) to support the static component and the WiMAX technology for mobility; or b) the use of WiMAX technology to support the Fixed and Nomadic component.

### 3.1 Cost Model Structure

The structure of a network depends on the nature of the services offered and their requirements including bandwidth, symmetry of communication and expected levels of demand.

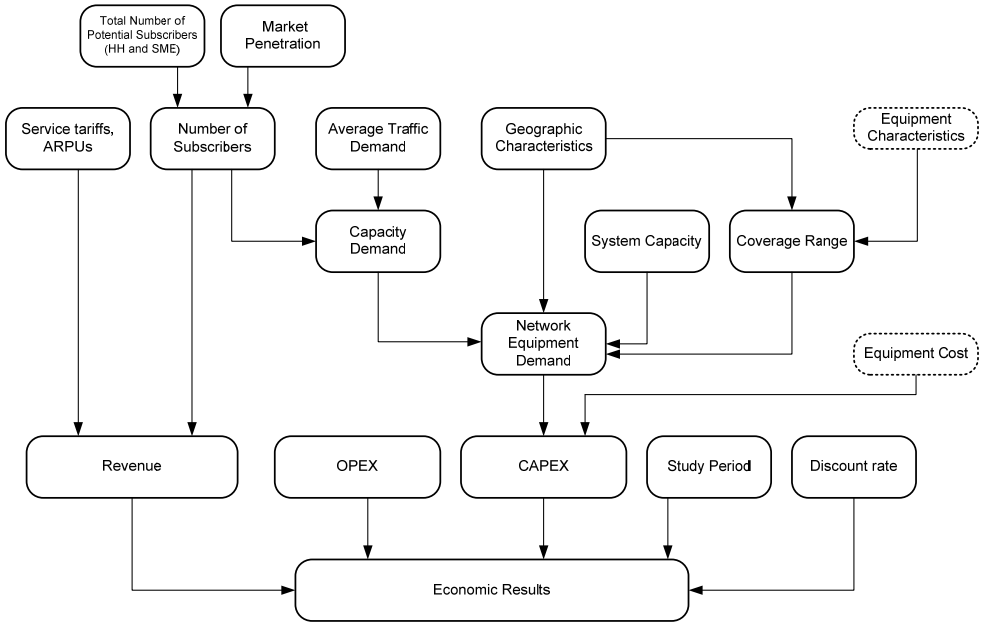



Fig. 9. Techno-economic parameters

As shown in Fig. 9, the techno-economic framework basically consists of the following building blocks (Montagne *et al.*, 2005): Area definition (geography and existing network infrastructure situation); Service definitions for each user segment with adoption rates and tariffs, such as network dimensioning rules and cost trends of relevant network equipment; cost models for investments (CAPEX) and operation costs (OPEX); Discounted cash flow model; Output metrics to be calculated.

The model analyzes several technical parameters (distances, bandwidth, equipment performance, etc.) as well as economic parameters (equipment costs, installation costs, service pricing, demographic distribution, etc.). The model simulates the evolution of the

business from 5 to 25 years. This means that each parameter can have a different value each year, which can be useful for reflecting factors that evolve over time.

### 3.1.1 General Model Assumptions

Our model framework defines the network starting from a single central office (or head-end) node and ending at a subscriber CPE. At the CO, we consider only the devices that support the connection to the access network (OLT).
Users are usually classified in four main categories: Home (residential customers), SOHO (Small Offices and Home Offices), SME (Small- to Medium-size Enterprises) and LE (Large Enterprises). The tool implements a methodology for the techno-economic analysis of access networks for residential customers and SME.

| Network Component | Component Costs | Description |
|---|---|---|
| Physical Plant component costs | Housing | The housing cost is the cost of building any structures required (e.g., remote terminal huts and CO buildings), and includes the cost of permits, labor, and materials. |
| | Cabling | The cabling cost is the cost of the materials (i.e., the cost of the necessary fiber optic, twisted pair, or coax cables). |
| | Trenching | The trenching cost is the cost of the labor required to install the cabling either in underground ducts (buried trenching) or on overhead poles (aerial trenching). |
| Network Equipment | Equipment needed between CO and subscriber house | The electronic switches and/or optical devices (e.g., splitters) needed to carry the traffic over the physical plant. |
| | Subscriber Equipment | The price and other properties of the Access node, as well as the nature of the CPE unit, depend strongly on the access technology. |

Table 1. General Model Assumptions

Access networks (for Wired technologies) have two separate but related components (Weldon & Zane, 2003): physical plant and network equipment (see Table 1). The physical plant includes the locations where equipment is placed and the connections between them. The physical plant costs depend primarily on the labor and real estate costs associated with the network service area, rather than on the specific technology to expand.
Access network costs can be grouped into two categories (Baker *et al.*, 2007): the costs of building the network before services can be offered (homes passed), and the costs of building connections to new subscribers (homes connected). More specifically, the homes passed portion of costs consists of exchange/CO fit out, feeder cables and civil works, cabinet and splitters, and distribution cables and civil works. The deployment cost calculations assumptions suppose that all construction work required to provide service to all homes passed takes place during the first year (deployment phase). However, only the necessary electronic equipments are deployed in the CO as well as the aggregation nodes to accommodate the initial assumption for the take rate.

### 3.1.2 Input Parameters

As mentioned beforehand, the definition of the input attributes is fundamental to obtain the right outputs. The model divides the inputs into two main categories: general and specific input parameters. General parameters are those that describe the area and service characteristics and are common to all the technologies. The specific parameters are those that characterize each solution, in technological terms.

These parameters are divided into three main groups: Equipment Components; Cable Infrastructure and Housing. The housing cost is the price to build any structures required in the outside plant (Cabinets, closures, etc.) This plant corresponds to the part between CO and the subscriber house. With regard to the cable infrastructure, the percentage of new cable corresponds to the need of the new cable required, and the percentage of new conduit parameter takes into account both underground and aerial lines. The civil work cost is based on the above mentioned parameters (for example: % of new conduit (Underground/Aerial), etc.) and on the Database cost. The cost of the labor also takes into account the cabling either in underground ducts (buried trenching) or on overhead poles (aerial trenching).

To build a new network or upgrade an existing one, an operator has to choose from a set of technologies. The cost structure may vary significantly from one technology to the other in terms of up-front costs, variable cost and maintenance costs. Each technology type has elements which are dedicated, like modems and shared elements (shared by many users) such as cabinets, optical network units, base stations and cables.

While some costs like equipment pricing, are easy to compute given the data in the Cost Database, because they do not depend on network topography, the per subscriber cabling costs (i.e. trenching and fiber) and equipment housing costs (which depend on distance and density) do, so they require optimization (Weldon & Zane, 2003).

A number of choices, assumptions, and predictions have to be made before proceeding to the techno-economic analysis of a broadband access network. These include the selection of the geographical areas and customer segments to be served, the services to be provided, and the technology to be used to provide the services (Smura, 2006). As we have seen above, the definition of the input attributes is fundamental to obtain the right outputs. Then, we define three main activities: Area Definition (Area parameters), Requested Services (Service parameters), Commercial Parameters and Type of Access.

### 3.1.3 Output Results

The financial analysis requires several outputs from the tool. The financial analysis is basically focused on the following steps: to compute the amount of equipment that needs to be installed each year for providing the service; to compute the amount of money spent on operational costs (Operations and Maintenance, Customer Support, Service Provisioning, Marketing); to compute the yearly income, taking into account that existing customers pay for 12 months; to compute the net profit obtained each year; and the NPV (Net Present Value) of the yearly profits. The calculated outputs are presented in Table 2:

| Output | Description |
|---|---|
| Cost per subscriber | Average cost divided by all subscribers reachable with the system. |
| Cost per home passed | Average cost divided by all homes reachable with the system. The cost per home passed will include both the up front costs of equipment and installation and the ongoing costs of maintaining and managing the network. |
| CAPEX | Investments costs |
| OPEX | Operation costs |
| Installation cost | Costs for equipment installation |
| Total expenses | CAPEX + OPEX |
| Total revenue | The total amount customers will pay for their telecommunications services. |
| Life Cycle Cost | Is defined as the sum of global discounted investments and global discounted running costs. This gives the total costs for constructing and running the network over the study period. |
| Profit per year (cash flow) | The Cash Balance (accumulated discounted Cash Flow) curve generally goes deeply negative because of high initial investments (Monath *et al.*, 2003). Once revenues are generated, the cash flow turns positive and the Cash Balance curve starts to rise. |
| Ending Cash Balance (or Cumulated Cash Flow) | The balance in the Cash Account at the end of the reporting period and, therefore, on the ending balance sheet. |
| Payback Period (Months) | First year with positive |
| Net Present Value (NPV profit) | The NPV is today's value of the sum of resultant discounted cash flows (annual investments and running costs), or the volume of money which can be expected over a given period of time. |
| Internal Rate of Return (IRR) | IRR is the discount rate at which the NPV is zero. If the IRR is higher than the opportunity cost of money (that is, interest of an average long term investment), the project is viable. |

Table 2. Output Results

### 3.2 Access Network Architecture

Our model studies the access part of the network, starting at the central office and ending at the subscriber's CPE (see Fig. 10). The cost model is based on a single central office, connecting the subscribers through several aggregation nodes.

The goal is to optimize the network in order to minimize the cost for a given performance criterion. The network is sized for the total number of Homes Passed. Consequently, all infrastructure costs (trenches, housing, electronics and fiber deployment) are incurred for all Homes Passed. Despite he costs of the CPE's, ports in the fiber node are only incurred when a home subscribes.

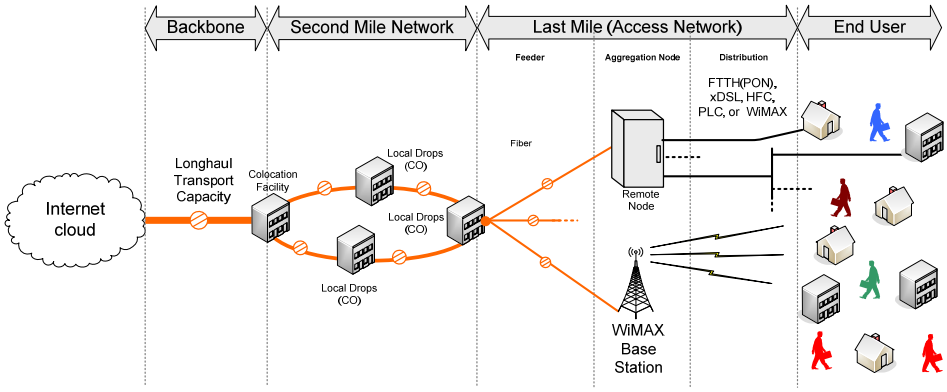



Fig. 10. Network architecture (Pereira, 2007a)

The access network architecture used in our model is divided into three main segments (Fig. 11): Inside, Outside, and End User. In the CO the different traffic flows are multiplexed/demultiplexed for further uplink connection to metropolitan and transport networks or, when it concerns local traffic, switched or routed back to the access network. For the CO we consider the following components: OLT ports, OLT chassis and passive splitters.

The outside segment is divided into three main parts: the feeder, aggregation Nodes and distribution (for HFC technology the distribution segment is divided into distribution and drop). Feeder segment comprise the network between the CO and the aggregation nodes. The model includes not only the cost of equipment (Fiber repeaters), but also the optical fiber cables, installation, trenches, and housing (street cabinets) costs. The ducts can be shared by several optical fiber cables. The aggregation nodes are located in access areas street cabinets. The components of these nodes depend on the technology. In the next section we will present the elements for the five technologies in study. The distribution network links the aggregation nodes with CPE. Like feeder networks, in distribution, the model includes not only the cost of equipment (copper, coax, and LV grid repeaters), but also the cables, installation, and trenches costs.

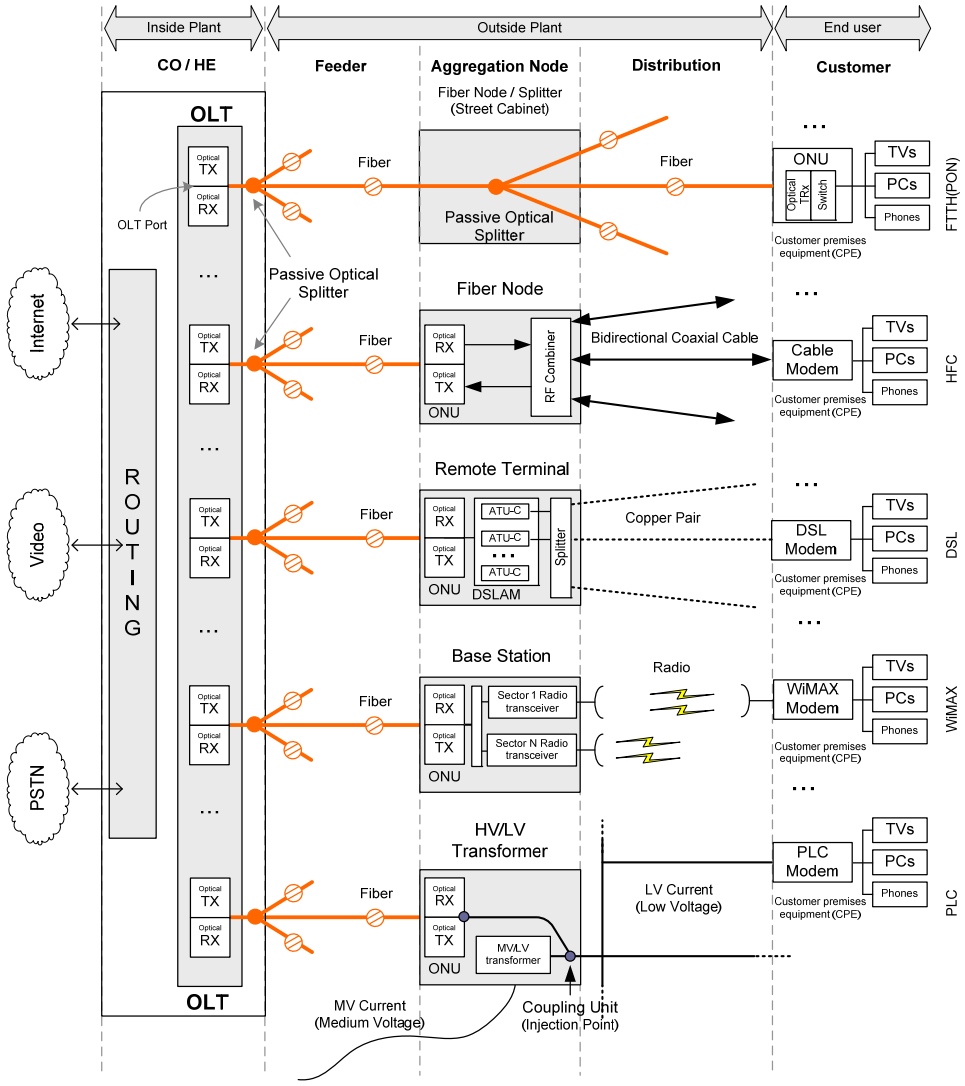



Fig. 11. Block diagram for Access Technologies (Pereira & Ferreira, 2009)

### 3.2.1 Access Network Components

Table 3 show the components used in our analysis. The components are divided into five segments (see Fig. 11). The inside plant and feeder segment components are common to all solutions. Optimally, there would eventually be 32 fibers reaching the ONTs of 32 homes (Pereira, 2007b). For example if the primary split is 1x4 and the secondary split is 1x8, then the network splitting ratio (or split scenario) will be 32. This means that a single feeder network supports 32 subscribers.

| Inside Plant | Outside Plant | | | End User | |
|---|---|---|---|---|---|
| | Feeder | Aggregation Node | Distribution | | |
| 1) OLT ports<br>2) Chassis<br>3) Splitter (Primary Split)<br>4) Installation: Ports, chassis, and split. | 1) Optical repeater<br>2) Repeater installation<br>3) Aerial/Buried trenches/ducts (Trenching costs)<br>4) Fiber Cable (cable cost)<br>5) Cable Installation | 1) Splitter (Secondary Split)<br>2) Splitter Installation<br>3)Housing: Street Cabinet | 1) Optical repeater<br>2) Repeater installation<br>3) Aerial/Buried trenches/ducts (Trenching costs)<br>4) Fiber Cable (cable cost)<br>5) Cable Installation | 1) ONU<br>2) Fiber Modem<br>2) Installation | FTTH(PON) |
| | | 1) Node Cabinet equipment: ONU; DSLAM; Splitter; Line-cards; Chassis; Racks<br>2) equipment Installation<br>3)Housing: Street Cabinet | 1) Copper regenerator / repeater<br>2) Repeater installation<br>3) Aerial/Buried trenches (Trenching costs)<br>4) Copper Cable (cable cost)<br>5) Cable Installation | 1) xDSL Modem<br>2) Splitter<br>3) Installation | xDSL |
| | | 1) Fiber Node Cabinet equipment: O/E converter (ONU); RF combiner<br>2) equipment Installation<br>3)Housing: Street Cabinet | 1) RF amplifier<br>2) Amplifier installation<br>3) Aerial/Buried trenches (Trenching costs)<br>4) Coaxial Cable (cable cost)<br>5) Cable Installation | 1) Cable Modem<br>2) Splitter<br>2) Installation | HFC |
| | | 1) Local MV/LV Transformer Station equipment (TE equipment): O/E converter; Coupling unit (injection point)<br>4)Housing: Street Cabinet<br>2) Transformer Station equipment<br>3) equipment Installation | 1) Repeater for LV network<br>2) Installation | 1) PLC Modem<br>2) Installation | PLC |

Table 3. Components used for wired technologies

The aggregation node, distribution and end user segments have different components, depending on each technology. In this table the components for the four wired technologies used in the model are presented.
The components for WiMAX technology are presented in the next section (see Table 4). However, the inside plant and feeder components are the same as the wired technologies.

### 3.2.2 Access Network Architecture for WiMAX

**a) System Architecture**

Fig. 12 shows the WiMAX system architecture used in our model. The "air" segments can replace the distribution and drop segment presented in Table 3.

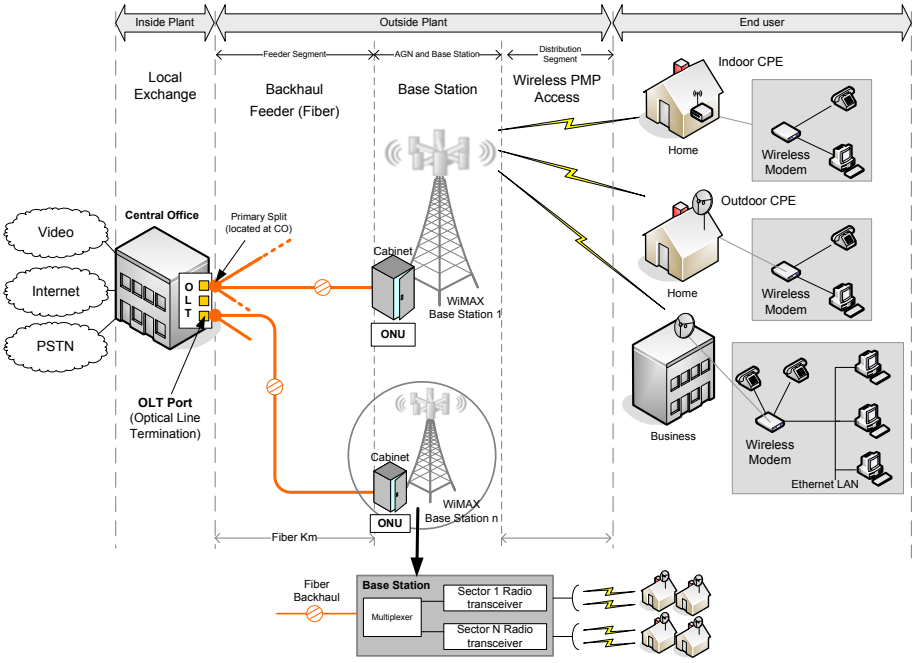



Fig. 12. Block Diagram of the baseline WiMAX system architecture

**b) Components**

According to the network architecture, the radio access network basically includes base stations, sites and "last mile" transmission (Wang, 2004). We therefore assume that the architecture is composed of a BS in the central end, a station in the subscriber side, and a PMP topology (between BS and CPE).

For capacity limited deployment scenarios it is necessary to position base stations with a BS to BS spacing sufficient to match the expected density of end customers. Data density is an excellent metric for matching capacity to market requirements. Demographic information including population, households and businesses per sq Km is readily available from a variety of sources for most metropolitan areas. With this information and the expected services to be offered along with an expected market penetration, data density requirements are easily calculated (WiMAX Forum, 2005b).

Base stations (towers) and base station equipment does not need to be installed in totality during the first year, but can be displayed over a period of time to address specific market segments or geographical areas of interest for the operator. However, in an area with a high number of potential subscribers, it is desirable to install a sufficient number of base stations to cover an addressable market large enough to quickly recover the fixed infrastructure cost (WiMAX Forum, 2004).

As presented in the previous figure, the common cost elements assumed in our model are in terms of Base Station the upfront costs; sector costs (including transducer and antenna); and Installation cost (co-siting, new site); Customer Premise Equipment (CPE): Indoor/Outdoor CPE; and Installation of the CPE, besides Operation, Administration and Maintenance (OAM).

| Outside Plant | | |
|---|---|---|
| Aggregation Node (Base Station) | Distribution (Wireless PMP Access) | End User |
| 1) Site acquisition<br>2) Site lease<br>3) Civil works BS<br>4) Housing Cabinet / Closures for each BS<br>5) PMP equipment (multiplexer + cost sector X # sectors per BS)<br>6) BS installation Cost (including sectors)<br>7) ONU (BS) and Installation | air | 1) WiMAX terminal (include: Antenna, Transceiver, Radio Modem)<br>2) Installation |

Table 4. WiMAX architecture components

It is rather important to calculate the required number of FWA base stations and sectors to fulfill the traffic capacity demands of all the subscribers in a given service area (Smura, 2004). The first step is the prediction of aggregate subscriber traffic in the service area. The number of the required BS is calculated as a function of the demand specified by the service area to be covered; average capacity required per user during busy hour; and number of subscribers within the coverage area (Johansson *et al.*, 2004). When radius of service-area cell is small, there are many cells of total service area. When the radius of the service-area cell is large, the number of cells is smaller in a total service area. This is the reason why total construction relative cost is decreasing when radius of service area is increasing.

### 3.3 Geometric Model Assumptions

The definition of the geometric model is required to calculate the length of trenches, ducts and cables. Some of the construction techniques are aerial, using string along utility poles (mostly rural areas); trench, digging up earth and then lay a new conduit and fiber (used in urban areas); and pull-through, running through existing underground conduits. Since each technology has different characteristics, the model assumes different assumptions for the several access technologies, which are described in the previous sections.

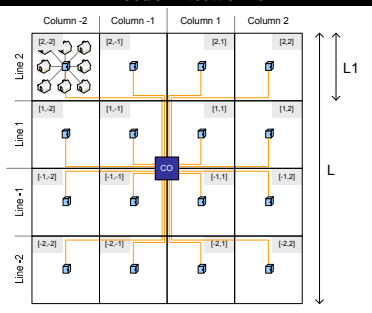
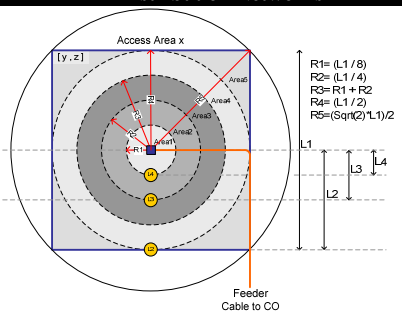
| Feeder Networks | Distribution Networks |
|---|---|
|  |  |



Table 5. Geometric model for Feeder and Distribution networks

In our work we consider that trench length represents the civil work required for digging and ducting – the model does not make distinction between aerial (overhead poles) and

buried (underground ducts). However, the costs are more significant where infrastructure must be buried than where it can be installed on existing poles (usually, aerial installation is almost twice as inexpensive as when the infrastructure is buried).

## 4. Results

### 4.1 Scenario description

| | | Value | Trend (% per year) |
|---|---|---|---|
| **Years (Study Period)** | | 15 | |
| **Geographical Area Description** | | Urban | |
| **Total Access Networks (Sub-areas)** | | 4 | |
| Area Characteristics | Area Size (Km2) | 47 | 0,00% |
| | Access Network area (Km2) | **11,75** | |
| | **Residential** | | |
| | Total Households (potential subscribers) | 11510 | 1,10% |
| | Households Density (Households / Km2) | **245** | |
| | Population Density (people/Km2) | 250 | 3,80% |
| | Population | **11.750** | |
| | Inhabitants per household | **1,02** | |
| | Technology penetration rate (expected market penetration) | 40,00% | 8,00% |
| | Number of subscribers | **4.604** | |
| | Average Households per building | 6 | |
| | Number of buildings in serving area (homes/km2) | **1918** | |
| | **SME (small-to-medium sized enterprises)** | | |
| | Total SME in Area | 2502 | 1,50% |
| | Technology penetration rate (expected market penetration) | 30,00% | 5,00% |
| | Total SME (customers) | **751** | |
| | **Nomadic Users** | | |
| | Total Nomadic Users | 1950 | 15,00% |
| Service Characteristics | **Residential** | | |
| | Required Downstream bandwidth (Mbps): Avg data rate | 8 | 1,2% |
| | Required Upstream bandwidth (Mbps): Avg data rate | 0,512 | 1,2% |
| | **SME** | | |
| | Required Downstream bandwidth (Mbps): Avg data rate | 12 | 1,2% |
| | Required Upstream bandwidth (Mbps): Avg data rate | 0,512 | 1,2% |
| | **Nomadic Users** | | |
| | Required Downstream bandwidth (Mbps): Avg data rate | 2 | 2,0% |
| | Required Upstream bandwidth (Mbps): Avg data rate | 0,512 | 2,0% |

| | | | |
|---|---|---|---|
| Pricing | **Residential** | | |
| | One-time Activation/connection fee (€) | 100 | 0,15% |
| | Subscription fee (€ / month) | 50 | 0,15% |
| | **SME** | | |
| | One-time Activation/connection fee (€) | 150 | 0,15% |
| | Subscription fee (€ / month) | 75 | 0,15% |
| | **Nomadic Users** | | |
| | One-time Activation/connection fee (€) | 75 | 0,15% |
| | Subscription fee (€ / month) | 45 | 0,15% |
| **Discount Rate (on cash flows)** | | 0% | |

Table 6. General Input Parameters for Access Network

The three main activities for the scenario description are: area definition, definition of the set of services to be offered, and the pricing. Table 6 shows the general input parameters used in our model and tool.
The trends for each parameter are presented in the last column. This scenario is defined for a study period of 15 years and for an urban area (City located in a remote area). The definition of the area type is essential because several costs are influenced by the fact that it is either an urban or a rural area.
Other important parameter is the definition of the number of access networks in which we want to divide the studied area (between 1 and 36). This scenario assumes the division of the area into 4 sub-areas (or access networks). Next, the definition of the number of households (HH), SMEs and nomadic users is also required for each access network (Table 7).

| | Access Network 1 | Access Network 2 | Access Network 3 | Access Network 4 | Total Area (Year 1) |
|---|---|---|---|---|---|
| HH: | 9000 | 2000 | 500 | 10 | **11510** |
| SME: | 1000 | 5000 | 1000 | 2 | **7002** |
| Nomadic Users: | 100 | 850 | 0 | 1000 | **1950** |

Table 7. Input Parameters: Total subscribers for each Access Network

As mentioned before (see Table 5), the access areas can be divided into five circular areas (between 1 and 5). This way, we can distribute the users in each access area, and calculate the trenches and required cable for the wired technologies (Table 8). When using wireless technologies, this structure is a good option to manage the required base stations for each access area more effectively.

| | Access Network 1 | | Access Network 2 | | Access Network 3 | | Access Network 4 | |
|---|---|---|---|---|---|---|---|---|
| | HHs | SMEs | HHs | SMEs | HHs | SMEs | HHs | SMEs |
| Area1 | 50% | 10% | 20% | 20% | 20% | 20% | 5% | 20% |
| Area2 | 20% | 20% | 20% | 20% | 20% | 20% | 10% | 20% |
| Area3 | 15% | 40% | 20% | 20% | 20% | 20% | 15% | 20% |
| Area4 | 10% | 20% | 20% | 20% | 20% | 20% | 20% | 20% |
| Area5 | 5% | 10% | 20% | 20% | 20% | 20% | 50% | 20% |

Table 8. Input Parameters: Subscribers localization for each Access Network

### 4.1.1 Feeder and Distribution Network Parameters

As described in Fig. 11, the feeder segment corresponds to the network part between CO and the aggregation node. In the CO, our model considers the OLT equipment, and we assume that the primary splits are at the CO. Table 9 shows the parameters used for the feeder and distribution segment of the network.

**Feeder Network Parameters**

| | |
|---|---|
| **Technology** | |
| Primary Split (located at CO) | 04 |
| OLT Chassis | |
| Number of OLT card slots per OLT Chassis | 16 |
| OLT Cards (only for Subs not for HP) | |
| Number of OLT ports per Card | 08 |
| Max. ONU's per OLT Port | 64 |
| Downstream Rate (Mbps) per OLT port | 622 |
| Upstream Rate (Mbps) per OLT port | 155 |
| Optical repeater and Copper regenerator | |
| Distance between Optical Repeater (km) | 30 |
| **Trench Parameters** | |
| Total Trench Lenght (Km) | **10,28** |
| % of new trenches | 65% |
| **Street Cabinet Parameters** | |
| % of new Street Cabinets/Closures | 60% |
| Max. # of Users per Street Cabinet(AGN) | 2048 user |
| **Cable Parameters for feeder network** | |
| % of new cable | 60% |
| Cable type | Optical Cable 48Fiber |
| Capacity per fiber (Gb/s) | 1 |

**Distribution Network Parameters**

| | |
|---|---|
| **Technology** | |
| **FTTH(PON)** | |
| Secondary Split (Street Cabinet) | 08 |
| Split Ratio: Subscribers per OLT port | 32 |
| **xDSL** | |
| xDSL technology | ADSL |
| ONU | |
| Maximum DS Capacity per ONU (Mbps) | 2000 |
| Maximum US Capacity per ONU (Mbps) | 1000 |
| Remote Terminal DSLAM | |
| DSLAM Units (Chassis) | 6 Line Cards |
| Number of Line Cards (ATU-Cs) per DSLAM unit: Slots | 6 |
| DSLAM Line Card (only for Subs not for homes passed) | Line_card_ports_48 |
| Number of port per line card (Max. subs per line card) | 48 |
| Downstream Rate (Mbps) per DSLAM line card port | 8,00 |
| Upstream Rate (Mbps) per DSLAM line card port | 2,00 |
| DSLAM Line Card Splitter | |
| DSLAM Splitter Card | 48 ports |
| Remote cabinet capacity | |
| Max. Number of DSLAMs per cabinet | 5 |
| Copper regenerator /repeater | |
| Distance Between Copper Repeater (km) | 5 |
| **HFC** | |
| ONU | |
| Maximum DS Capacity per ONU (Mbps) | 2000 |
| Maximum US Capacity per ONU (Mbps) | 1000 |
| Number of Ports per ONU (for RF modem) | 8 |
| RF node modem | |
| Maximum DS Capacity per RF Node Modem (Mbps) | 100 |
| Maximum US Capacity per RF Node Modem (Mbps) | 100 |
| RF amplifiers (2-way) | |
| Distance Between RF Amplifiers (2-way): km | 2 |
| TAP equipment | |
| Number of Drops per TAP | 2 |
| **WiMAX** | |
| ONU | |
| Maximum DS Capacity per ONU (Mbps) | 2000 |
| Maximum US Capacity per ONU (Mbps) | 1000 |
| Base Station | |
| Downstream Sector capacity (Mbps) | 36 |
| Upstream Sector capacity (Mbps) | 10 |
| Maximum sector throughput (Mbps): Capacity per sector | 46 |
| Maximum number of sectors per base station | 4 |
| Max. Base Station range - radius (km) | 3 |
| CPE (only for Subs not for homes passed) | |
| % of Indoor CPE | 60% |
| % of Outdoor CPE | 40% |
| **PLC** | |
| ONU | |
| Maximum DS Capacity per ONU (Mbps) | 2000 |
| Maximum US Capacity per ONU (Mbps) | 1000 |

| LV Transformer equipment | |
|---|---|
| Number of MV/LV transformers per Transformer Substation | 1 |
| AVG Number of feeders per MV/LV transformer | 8 |
| Downstream LV TE capacity (Mbps):link between TE-CPEs) | 25 |
| Upstream LV TE capacity (Mbps) | 25 |
| Max. # HH per MV/LV transformer: (# Subsc per LV network) | 24 |
| **PLC Repeater for all LV networks (all homes passed)** | |
| AVG Length of the LV lines (m) | 0,3 |
| Maximum repeater reach (m) | 0,325 |
| Average number of repeaters in Single house | 0 |
| Average # repeaters in building (repeater in the meter room) | 1,27 |

Table 9. Feeder and Distribution Network Parameters

The distribution segment links the aggregation nodes to the customers. The technologies used in our work are: FTTH(PON), xDSL, HFC, PLC, and WiMAX. As each solution embraces different characteristics, the previous table shows the parameters used for each one of them.

### 4.2 Results

This section presents the final results to support the new requirements of broadband access (fixed and nomadic users). Table 10 shows the results for the use of the several technologies to support the static layer (HH and SMEs). Each column corresponds to an access network. The output variables are represented in the lines: Payback period, NPV, IRR, Cost per subscriber in year 1, and cost per subscriber in year n.

| | | Access Network 1 | Access Network 2 | Access Network 3 | Access Network 4 | Total Area | |
|---|---|---|---|---|---|---|---|
| | # Fixed Users | 10000 | 2500 | 1500 | 12 | 14012 | |
| | # Nomadic Users | 100 | 850 | 0 | 1000 | 1950 | |
| FTTH | Payback Period | 14 | 21 | 22 | 63 | 33 | (Average) |
| | NPV | 11.965.382 € | - 2.957.765 € | - 2.390.925 € | - 165.567 € | 3.136.297 € | (Average) |
| | IRR | 3,06% | -2,42% | -3,43% | -13,19% | -4,52% | (Average) |
| | Cost Subc Y1 | 10.852 € | 16.153 € | 18.444 € | 48.135 € | 25.811 € | (Average) |
| | Cost Subc Y15 | 153 € | 189 € | 249 € | 207 € | 203 € | (Average) |
| | CAPEX | 52.220.733 € | 18.694.854 € | 11.437.130 € | 237.165 € | 82.589.882 € | (Sum) |
| | OPEX | 9.016.542 € | 2.407.516 € | 1.403.593 € | 15.745 € | 12.843.396 € | (Sum) |
| **WIMAX** | Payback Period | 14 | 14 | 13 | 55 | 27 | (Average) |
| | NPV | 3.821.570 € | 978.222 € | 890.269 € | - 140.025 € | 1.523.938 € | (Average) |
| | IRR | 2,33% | 2,33% | 3,32% | -12,04% | -2,13% | (Average) |
| | Cost Subc Y1 | 4.964 € | 5.245 € | 6.414 € | 42.018 € | 17.799 € | (Average) |
| | Cost Subc Y15 | 453 € | 458 € | 484 € | 287 € | 408 € | (Average) |
| | CAPEX | 53.374.912 € | 13.198.012 € | 7.290.309 € | 195.945 € | 74.059.178 € | (Sum) |
| | OPEX | 16.006.174 € | 3.968.371 € | 2.269.220 € | 31.423 € | 22.275.190 € | (Sum) |
| DSL | Payback Period | 20 | 34 | 38 | 61 | 40 | (Average) |
| | NPV | - 11.094.161 € | - 11.716.167 € | - 7.892.363 € | - 131.557 € | - 6.372.694 € | (Average) |
| | IRR | -2,21% | -7,53% | -9,02% | -12,74% | -7,99% | (Average) |
| | Cost Subc Y1 | 15.968 € | 23.909 € | 27.281 € | 38.812 € | 27.354 € | (Average) |
| | Cost Subc Y15 | 173 € | 230 € | 321 € | 260 € | 252 € | (Average) |
| | CAPEX | 74.679.305 € | 27.215.883 € | 16.787.143 € | 204.107 € | 118.886.438 € | (Sum) |
| | OPEX | 9.617.513 € | 2.644.889 € | 1.555.018 € | 14.794 € | 13.832.213 € | (Sum) |
| HFC | Payback Period | 15 | 23 | 26 | 69 | 37 | (Average) |
| | NPV | 4.715.962 € | - 5.051.179 € | - 3.789.621 € | - 180.505 € | 248.612 € | (Average) |
| | IRR | 1,09% | -3,82% | -5,04% | -13,89% | -5,95% | (Average) |
| | Cost Subc Y1 | 12.606 € | 18.167 € | 20.812 € | 50.991 € | 28.137 € | (Average) |
| | Cost Subc Y15 | 155 € | 194 € | 264 € | 214 € | 211 € | (Average) |
| | CAPEX | 59.287.426 € | 20.721.462 € | 12.789.876 € | 251.477 € | 93.050.241 € | (Sum) |
| | OPEX | 9.199.269 € | 2.474.322 € | 1.449.543 € | 16.371 € | 13.139.505 € | (Sum) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| PLC | Payback Period | 30 | 46 | 51 | 51 | 44 | (Average) |
| | NPV | - 37.581.298 € | - 18.359.891 € | - 11.979.733 € | - 133.062 € | - 16.564.697 € | (Average) |
| | IRR | -6,36% | -10,39% | -11,81% | -12,93% | -10,37% | (Average) |
| | Cost Subc Y1 | 21.257 € | 29.714 € | 34.382 € | 38.622 € | 31.420 € | (Average) |
| | Cost Subc Y15 | 221 € | 267 € | 351 € | 220 € | 264 € | (Average) |
| | CAPEX | 98.485.189 € | 33.056.567 € | 20.378.811 € | 202.299 € | 152.122.866 € | (Sum) |
| | OPEX | 12.298.765 € | 3.447.930 € | 2.050.719 € | 18.107 € | 17.815.521 € | (Sum) |
| **Best Solution** | Payback Period | **FTTH** | **WIMAX** | **WIMAX** | **PLC** | | |
| | NPV | **FTTH** | **WIMAX** | **WIMAX** | **DSL** | | |
| | IRR | **FTTH** | **WIMAX** | **WIMAX** | **WIMAX** | | |
| | CostSubsc Y1 | **WIMAX** | **WIMAX** | **WIMAX** | **PLC** | | |
| | CostSubsc Yn | **FTTH** | **FTTH** | **FTTH** | **FTTH** | | |

Table 10. Broadband Access General Results

With these results we can identify the best solution for each access network. In general, the WiMAX technology is the best option for areas with fewer subscribers (area 3 and 4).

### 4.3 Sensitivity analysis

A sensitivity analysis is a systematic study of how an output result changes as the assumptions are varied. For the sensitivity analysis we use the tornado diagrams (graphical sensitivity analysis technique). A tornado diagram provides a graphical display of the sensitivity of some system responses to uncertainties in the various inputs of that system. The diagrams show the effects of uncertainties in each input variables on the output of the analysis. The following table shows the effect of these input variables (first column: coverage area, potential HH, etc.) on the output variables (Cost per subscriber, payback period, NPV, IRR, CAPEX, and OPEX).

| | | | Cost Per Subscriber | | PaybackPeriod | | NPV | |
|---|---|---|---|---|---|---|---|---|
| **Parameters Names** | **Low Parameter Values** | **High Parameter Values** | **Variation (Low Value)** | **Variation (High Value)** | **Variation (Low Value)** | **Variation (High Value)** | **Variation (Low Value)** | **Variation (High Value)** |
| Coverage Area | -50% | 50% | -0,1% | 0,1% | 0,0% | 0,0% | 1,3% | -1,3% |
| Potential HH | -55% | 55% | 4,9% | -1,8% | 0,0% | 7,1% | -31,3% | 31,0% |
| Potential SMEs | -65% | 65% | -2,9% | 2,5% | 7,1% | 0,0% | -30,2% | 29,9% |
| TakeRate HH | -50% | 50% | 4,0% | -1,7% | 0,0% | 7,1% | -28,0% | 27,9% |
| TakeRate SME | -55% | 55% | -2,5% | 2,1% | 7,1% | 0,0% | -25,6% | 25,5% |
| Required DS Bandwidth HH | -60% | 60% | -46,4% | 46,2% | -57,1% | 264,3% | 1366,1% | -1366,1% |
| Required DS Bandwidth SME | -55% | 55% | -7,8% | 7,6% | -7,1% | 28,6% | 152,3% | -152,3% |
| Activation Fee HH | -50% | 50% | -0,1% | 0,1% | 7,1% | 0,0% | -23,8% | 23,8% |
| Month Fee HH | -40% | 40% | -0,4% | 0,4% | 264,3% | -35,7% | -951,4% | 951,4% |
| Activation Fee SME | -40% | 40% | 0,0% | 0,0% | 0,0% | 0,0% | -2,2% | 2,2% |
| Month Fee SME | -50% | 50% | -0,1% | 0,1% | 35,7% | -7,1% | -163,3% | 163,3% |
| CAPEX: Equipment | -50% | 50% | -21,3% | 21,3% | -21,4% | 64,3% | 334,9% | -334,9% |
| CAPEX: Equip. Installation | -55% | 55% | -4,0% | 4,0% | -7,1% | 28,6% | 127,9% | -127,9% |
| CAPEX: Housing | -50% | 50% | -20,6% | 20,6% | -28,6% | 257,1% | 680,0% | -680,0% |
| CAPEX: Cable | -55% | 55% | 0,0% | 0,0% | 0,0% | 0,0% | 0,2% | -0,2% |
| CAPEX: Civil Works | -50% | 50% | -0,5% | 0,5% | 0,0% | 7,1% | 14,5% | -14,5% |
| OPEX: Network Operations | -55% | 55% | -0,8% | 0,8% | -7,1% | 35,7% | 168,5% | -168,5% |
| OPEX: Equipment | -50% | 50% | -1,5% | 1,5% | 0,0% | 7,1% | 22,9% | -22,9% |
| OPEX: Equip. Installation | -55% | 55% | -0,2% | 0,2% | 0,0% | 0,0% | 7,7% | -7,7% |
| OPEX: Housing | -50% | 50% | -1,0% | 1,0% | 0,0% | 7,1% | 34,0% | -34,0% |
| OPEX: Civil Works | -50% | 50% | 0,0% | 0,0% | 0,0% | 0,0% | 0,4% | -0,4% |
| OPEX: Lease | -55% | 55% | -0,5% | 0,5% | -7,1% | 28,6% | 130,8% | -130,8% |

| | | | IRR | | CAPEX | | OPEX | |
|---|---|---|---|---|---|---|---|---|
| Coverage Area | -50% | 50% | 1,4% | -1,4% | -0,2% | 0,2% | -0,1% | 0,1% |
| Potential HH | -55% | 55% | 28,3% | -10,5% | -46,5% | 46,3% | -46,6% | 46,4% |
| Potential SMEs | -65% | 65% | -22,5% | 18,2% | -9,7% | 9,7% | -9,8% | 9,8% |
| TakeRate HH | -50% | 50% | 24,6% | -10,0% | -42,4% | 42,2% | -42,4% | 42,3% |
| TakeRate SME | -55% | 55% | -18,9% | 15,8% | -8,3% | 8,1% | -8,3% | 8,2% |
| Required DS Bandwidth HH | -60% | 60% | 1687,2% | -100,0% | -47,0% | 46,8% | -38,3% | 38,2% |
| Required DS Bandwidth SME | -55% | 55% | 158,6% | -151,4% | -7,9% | 7,7% | -6,4% | 6,3% |
| Activation Fee HH | -50% | 50% | -23,6% | 23,5% | 0,0% | 0,0% | -1,1% | 1,1% |
| Month Fee HH | -40% | 40% | -100,0% | 711,0% | 0,0% | 0,0% | -5,4% | 5,4% |
| Activation Fee SME | -40% | 40% | -2,2% | 2,2% | 0,0% | 0,0% | -0,2% | 0,2% |
| Month Fee SME | -50% | 50% | -167,2% | 150,5% | 0,0% | 0,0% | -1,2% | 1,2% |
| CAPEX: Equipment | -50% | 50% | 401,7% | -314,5% | -23,1% | 23,1% | 0,0% | 0,0% |
| CAPEX: Equip. Installation | -55% | 55% | 124,6% | -128,5% | -4,3% | 4,3% | 0,0% | 0,0% |
| CAPEX: Housing | -50% | 50% | 657,9% | -854,0% | -22,4% | 22,4% | 0,0% | 0,0% |
| CAPEX: Cable | -55% | 55% | 0,2% | -0,2% | 0,0% | 0,0% | 0,0% | 0,0% |
| CAPEX: Civil Works | -50% | 50% | 14,4% | -14,4% | -0,6% | 0,6% | 0,0% | 0,0% |
| OPEX: Network Operations | -55% | 55% | 152,4% | -174,0% | 0,0% | 0,0% | -10,3% | 10,3% |
| OPEX: Equipment | -50% | 50% | 23,9% | -23,5% | 0,0% | 0,0% | -18,7% | 18,7% |
| OPEX: Equip. Installation | -55% | 55% | 7,6% | -7,6% | 0,0% | 0,0% | -3,0% | 3,0% |
| OPEX: Housing | -50% | 50% | 33,3% | -33,6% | 0,0% | 0,0% | -13,1% | 13,1% |
| OPEX: Civil Works | -50% | 50% | 0,4% | -0,4% | 0,0% | 0,0% | -0,2% | 0,2% |
| OPEX: Lease | -55% | 55% | 118,9% | -132,8% | 0,0% | 0,0% | -6,4% | 6,4% |

Table 11. Sensitivity analysis for WiMAX technology

The tornado diagrams are a series of horizontal bars (one for each variable) around the base value result. The big bars mean high impact and are on the top of the diagram. The bars decline in size to the smallest at the bottom, representing the parameter that causes least change to the base value. The red bars represent the Output for Low Value (negative variation in parameter), and the blue bars represents the Output for High Value (positive variation in parameter).

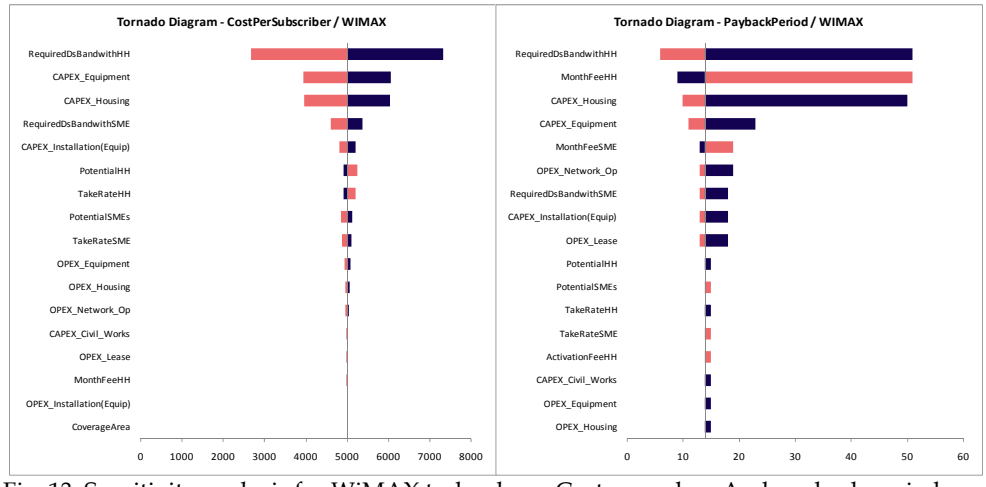



Fig. 13. Sensitivity analysis for WiMAX technology: Cost per subsc. And payback period

As we can see in graph 1 (Fig. 13), the three input variables which influence more the cost per subscriber are the required downstream bandwidth; equipment costs and housing costs. The three more critical variables that affect the payback period (graph 2) are bandwidth

(the increase of 60% in bandwidth implies an increase of 264% in payback period), households month fee (a decrease of 40% in this fee implies an increase of 264%), and housing costs (an increase of 50% of this feature leads to an increase of 257% in the payback period).

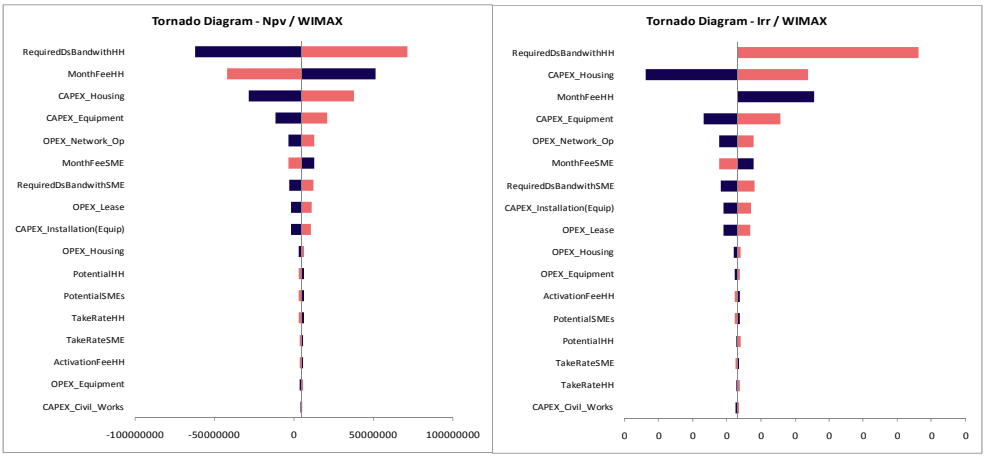



Fig. 14. Sensitivity analysis for WiMAX technology: NPV and IRR

For NPV and IRR (Fig. 14), the three most critical input variables remain the same as cost per subscriber and payback period (bandwidth, housing and equipment costs, and month fee).

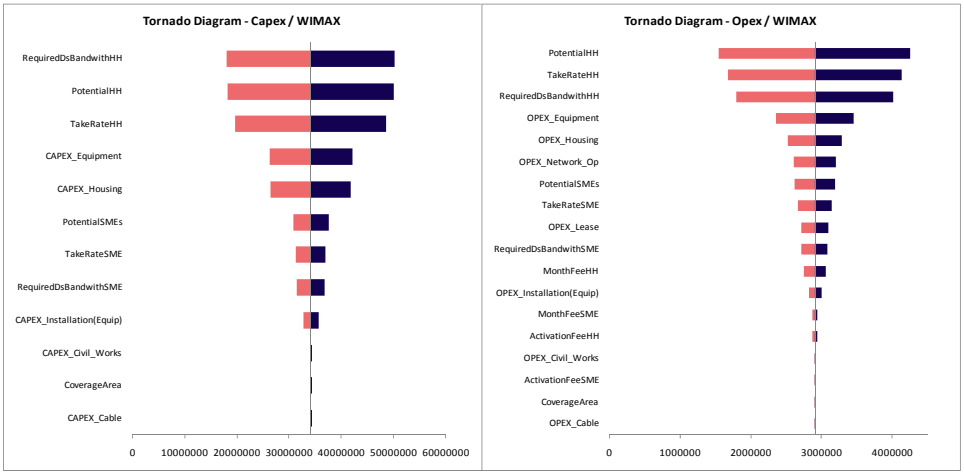



Fig. 15. Sensitivity analysis for WiMAX technology: CAPEX and OPEX

## 5. Conclusion

Nowadays access networks face two main challenges: the increasing bandwidth demand and mobility trends. All this will require fundamental changes to the operations of access networks, the functionality of network nodes and the architecture itself. The initial focus of wireless networks was to support mobility and flexibility, while for the wired access networks it was both bandwidth and high QoS. However, with the advances in technology, wireless solutions (such as WiMAX) have the capacity to provide both wideband and high QoS. WiMAX technology can also offer very high data rates, extended coverage and quickly deployable alternative to cabled access networks, such as fiber optic links, coaxial systems using cable modems, and Digital Subscriber Line (DSL) links. At present, WiMAX systems have the capability to address broad geographic areas without the costly infrastructure requirement of deploying cable links to individual sites. Besides, the technology may prove less expensive to disseminate and should lead to more ubiquitous broadband access.
In this context, we present a techno-economic model framework to support the bandwidth and mobility trends of access networks. The proposed tool performs a detailed comparison of WiMAX technology with different broadband access technologies (FTTH: PON. xDSL, HFC and PLC). For that, we identify the critical components of the WiMAX architecture. The produced results can analyze how the costs vary, calculating the cost per user, cost per homes passed, payback period, NPV, IRR, end cash balance, CAPEX, OPEX, and so on. For each sub-area, we describe the best solution, based on the output results.
Finally, we present the results of the sensitivity analysis. This analysis shows the effect of the input parameters (Coverage Area, Potential HH/SMEs, Take Rate, Required Downstream Bandwidth, Required Upstream Bandwidth, Activation Fee, Month Fee, etc.) on the output parameters (Cost per subscriber, Cost per Homes Passed, End cash balance, Payback period NPV, IRR, CAPEX, and OPEX.) and also identifies the critical parameters for several technologies. With this information it is possible to define better strategies for building broadband access networks.

## 6. References

1. Anderson,H.R. (2003) *Fixed Broadband Wireless System Design*. John Wiley & Sons.
2. Baker,J., Cagenius,T., Goodwin,C., Hansson,M. & Hatas,M. (2007) Deep-fiber broadband access networks. *ERICSSON REVIEW,* **84**, 4-8.
3. Carcelle,X., Dang,T. & Devic,C. (2006) Wireless Networks in industrial environments: State of the art and Issues. In: *IFIP Interactive Conference on Ad-Hoc Networking,* pp. 141-156. Springer Boston.
4. Corning (2005) Broadband Technology Overview: Optical Fiber. In: Corning.
5. El Zein,G. & Khaleghi,A. (2007) FTTH Explained: Delivering Efficient Customer Bandwidth and Enhanced Services. In: *New Technologies, Mobility and Security,* Labiod,H. & Badra,M. (eds.), pp. 271-179. Springer Netherlands.
6. Fernando,X. (2008) Broadband Access Networks. In: *IEEE - International Conference on Signal Processing, Communications and Networking (ICSCN 08),* pp. 380-383. IEEE International Conference on Signal Processing, Communications and Networking.
7. Fong,G.L. & Nour,K. (2004) Broadband and the Role of Satellite Services. In: pp. 1-19.

8. Galperin,H. (2005) Wireless Networks and Rural Development: Opportunities for Latin America. *Information Technologies and International Development,* **2**, 47-56.
9. Ghosh,A., Wolter,D.R., Andrews,J.G. & Chen,R. (2005) Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential. *Communications Magazine, IEEE,* **43**, 129-136.
10. Ibe,O.C. (2002) *Fixed Broadband Wireless Access Networks and Services*. John Wiley & Sons, Inc..
11. Ims,L.A., Bhatnagar,A., Østlyngen,E. & Bozorgebrahimi,K. (2004) Towards the next generation broadband network platform. *Telektronikk,* **4**, 107-125.
12. Johansson,K., FuruskÄar,A. & Karlsson,P. (2004) Relation Between Base Station Characteristics and Cost Structure in Cellular Systems. In: *IEEE PIMRC 2004,* IEEE.
13. Konhauser,W. (2006) Broadband wireless access solutions - Progressive challenges and potential value of next generation mobile networks. *Wireless Personal Communications,* **37**, 243-259.
14. Lehr,W., Sirbu,M. & Gillett,S. (2004) Municipal Wireless Broadband: Policy and Business Implications of emerging Access Technologies. In.
15. Lim,C.K., So,P.L., Gunawan,E., Chen,S., Lie,T.T. & Guan,Y.L. (2000) Development of a test bed for high-speed power line communications. In: pp. 451-456.
16. Little,A.D. (2004) White Paper on Power Line Communications (PLC) - 2004. In: PLC Utilities Alliance (PUA) (ed.), PLC Utilities Alliance (PUA).
17. Monath,T., Elnegaard,N.K., Cadro,P. & Katsianis,D. (2003) Economics of Fixed Broadband Access Network Strategies. *IEEE Communications Magazine,* **163**, 132-139.
18. Montagne,R., Causse,A., Elnegaard,N., Ryan,D. & Bråten,L. (2005) Broadband access roadmap based on market assessment and technical economic analysis. In: pp. 1-130. Broadwan.
19. Ohrtman,F. (2005) *WiMAX Handbook: Building 802.16 Wireless Networks*, 1 edn. McGraw-Hill Professional.
20. Peng,M. & Wang,W. (2007) A Unified Architecture and Key Techniques for Interworking between WiMAX and Beyond 3G/4G Systems. *Wireless Personal Communications,* **43**, 1-24.
21. Pereira,J.P. (2007a) A Cost Model for Broadband Access Networks: FTTx versus WiMAX. In: *2007 Second International Conference on Access Networks,* pp. 1-8. Second International Conference on Access Networks.
22. Pereira,J.P. (2007b) Broadband Access Technologies Evaluation Tool (BATET). *Technological and Economic Development of Economy,* **XIII**, 288-294.
23. Pereira,J.P. & Ferreira,P. (2009) Access Networks for Mobility: A Techno-Economic Model for Broadband Access Technologies. In: *TRIDENTCOM 2009 - The 5th International Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities,* IEEE.
24. Salvekar,A., Sandhu,S., Li,Q., Vuong,M.-A. & Qian,X. (2004) Multiple-Antenna Technology in WiMAX Systems. *Intel Technology Journal,* **8**, 229-240.
25. Sarrocco,C. & Ypsilanti,D. (2008) Convergence and Next Generation Networks. In: OECD (ed.), OECD.
26. Sigurdsson,H.M. (2007) Techno-Economics of Residential Broadband Deployment (PhD Thesis). In: Technical University of Denmark - Center for Information and Communication Technologies.

27. Smura,T. (2006) Competitive Potential of WiMAX in the Broadband Access Market: A Techno-Economic Analysis. *Helsinki University of Technology,* 1-21.
28. Smura,T. (2004) Techno-Economic Analysis of IEEE 802.16a-Based Fixed Wireless Access Networks. In: Helsinki University of Technology.
29. Ufongene,C.M. (1999) A model for the cost analysis of wireless access architectures. *Bell Labs Technical Journal,* **4**, 134-154.
30. Vaughan-Nichols,S.J. (2004) Achieving wireless broadband with WiMax. *Computer,* **37**, 10-13.
31. Wang,J. (2004) Will WiMAX+WLAN Constitute a Substitute to 3G? - A Techno-Economic Case Study. In: KTH Center for Wireless Systems., Stockholm.
32. Wanichkorm,K. (2002) The role of fixed wireless access networks in the deployment of broadband service and competition in local telecommunications markets. In: Carnegie Mellon University.
33. Weldon,M.K. & Zane,F. (2003) The economics of fiber to the home revisited. *Bell Labs Technical Journal,* **8**, 181-206.
34. WiMAX Forum (2005a) WiMAX: The Business Case for Fixed Wireless Access in Emerging Markets. In.
35. WiMAX Forum (2005b) WiMAX Deployment Considerations for Fixed Wireless Access in the 2.5 GHz and 3.5 GHz Licensed Bands. In: WiMAX Forum (ed.).
36. WiMAX Forum (2004) Business Case Models for Fixed Broadband Wireless Access based on WiMAX Technology and the 802.16 Standard. In: WiMAX Forum (ed.), WiMAX Forum.

3

# Mobile WiMAX Handset Front-end: Design Aspects and Challenges

Václav Valenta[1, 2], Geneviève Baudoin[2], Antoine Diet[3], Roman Maršálek[1], Fabien Robert[2], Martha Suarez[2] and Martine Villegas[2]
[1]*Brno University of Technology, Department of Radioelectronics*
[2]*Université Paris-Est, ESIEE Paris*
[3]*Université Paris-Sud 11, UMR 8506*
[1]*Czech Republic,* [2,3]*France*

## 1. From the Standard and Regulations to Front-End Specifications

WiMAX standards and regulations specify the minimum requirements for RF transceivers. They are expressed by parameters, which can take values that impose more or less stringent constraints on the RF system blocks such as the Power Amplifier (PA), synthesizer, filters and Low Noise Amplifiers (LNA). Among the important parameters are:

Signal parameters:

- Peak to Average Power Ratio (PAPR) of waveforms
- Peak transmitted power or output power level
- Transmit Power Control (TPC): precision and range
- Signal bandwidth

Performance criteria for transceivers:

- Noise: Noise Factor (NF), phase noise, etc.
- Linearity: AM-AM and AM-PM characteristics of power amplifiers, compression points at 1dB, interception point, noise power ratio etc.
- Efficiency: drain efficiency, power added efficiency, global efficiency
- Signal to Noise plus Distorsion Ratio (SNDR)

System performance:

- Bit Error Rate (BER) and Error Vector Magnitude (EVM)
- Adjacent Channel Interference (ACI): Adjacent Channel Leakage Ratio (ACLR) or Adjacent Channel Power Ratio (ACPR)
- Spectrum mask

System parameters or modes:

- Frequency bands and channel frequency step
- Channel bandwidth and data rate
- Duplexing mode: in time TDD, frequency FDD or half duplex FDD HFDD
- Multi-antenna MIMO and smart antenna possibility

**Peak to Average Power Ratio (PAPR) of Waveforms**

OFDM modulation has many well-known interests for mobile communications: robustness to multi-path, simple equalization and good spectral efficiency. Unfortunately, OFDM waveforms are characterized by high PAPR. It is sometimes noted Peak to Mean Power Ratio (PMPR). This power ratio measures signal amplitude fluctuation before the PA. It can be defined on sampled or continuous signals and estimated on RF or baseband signals with a 3 dB difference between both values. For an OFDM signal, its value depends on the number of carriers ($N$), constellation size ($M$), the shaping filter and the oversampling ratio ($L$). For $L$=1 and $M$=$2^{2n}$, the baseband PAPR is:

$$PAPR = 10\log(N) + 10\log\left(3\frac{\sqrt{M}-1}{\sqrt{M}+1}\right). \quad (1)$$

For $N$=1024 carriers this value is greater than 30 dB. But this maximum is not used in practice for transmitter design. For a random data signal, the PAPR, calculated on each OFDM symbol, is a random value and its value depends on the statistical distribution of signal amplitudes. It can be analyzed with its Complementary Cumulative Distribution Function (CCDF) that describes the probability that the PAPR of an OFDM symbol exceeds a given threshold. The PAPR value that is used in practice is an Effective Power Ratio (EPR) associated with a given probability $a$ (typically $a$ =$10^{-2}$ or $10^{-3}$). EPR is the threshold that is exceeded with the probability $a$:

$$CCDF(EPR) = \Pr(PAPR \geq EPR) = \alpha . \quad (2)$$

Many expressions have been proposed to approximate the probability distribution of the PAPR of OFDM signals. For example Zou's expressions:

$$\Pr(PAPR > x) \approx \sqrt{\frac{\pi}{3}} N\sqrt{x}\, e^{-x}. \quad (3)$$

It can be observed in Figure 1 that the probability that the power ratio is greater than 12 dB is approximately 10-3 for N=1024 carriers. For a full OFDM WiMAX signal with a 1024 FFT and a 16-QAM mapping, the PAPR (or effective power ratio with a probability of $10^{-3}$) is approximately 12 dB. For the transmitter design this value of 12 dB is used instead of the theoretical maximum of 30 dB that has a very low probability of occurrence.

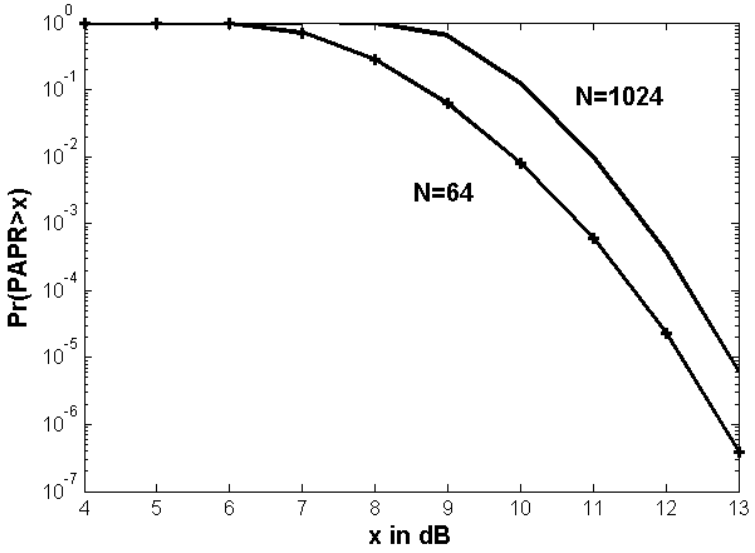



Fig. 1. Distribution for *N*=64 and *N*=1024.

**Peak Transmitted Power or Output Power Level and Transmit Power Control (TPC): Range and Precision**

The Peak transmit power is typically 23 dBm for WiMAX subscriber stations.

The power control of the transmitted signal compensates for variations in signal strength (due to distance variation for example). For WiMAX the TPC must be monotonic and able to cover a range of at least 45 dB by steps of 1 dB with a relative accuracy of 0.5 dB.

**Error Vector Magnitude (EVM)**

The Error Vector Magnitude or EVM represents the average deformation of a constellation after compensation of distortions due to rotation, translation and gain. The relative constellation error and ratio of error magnitude (or power) on constellation points magnitude (or power), can be expressed in percentage or in dB.

The WiMAX 802.16e standard specifies the EVM requirements. The relative constellation error is given for the different QAM mapping and coding rate: QPSK(1/2), QPSK(3/4), 16-QAM(1/2), 16-QAM(3/4), 64-QAM(1/2), 64-QAM(2/3), 64-QAM(3/4) with respective EVM values in dB equal to -15, -18, -20.5, -24, -26, -28 and -30.

For larger constellations such as 64-QAM, the distance between the constellation points is reduced for a given average power. Therefore the constraints on EVM are more severe in order to maintain an acceptable BER.

For the transmitter, the EVM value is one of the parameters (with ACLR) that specify the linearity requirements of the PA. It measures the inband distortion generated by PA non-linearity and it is more or less difficult to fulfil depending on the signal PAPR.

For the receiver, the EVM value specification has consequences on the acceptable ratio of the noise floor to peak signal. For an EVM of -30 dB, this ratio should be greater than 42 dB if we consider a PAPR of 12dB. It is also necessary to add some margin to take blocking signals, for example, into account.

**Adjacent Channel Leakage Ratio (ACLR) or Adjacent Channel Power Ratio (ACPR)**
The ACLR (or equivalently ACPR) parameter is the ratio of the power in the signal channel to the power in an adjacent channel. The precise definition can vary from one standard to another. The bandwidths in which the powers are measured can be different for the signal channel and for the adjacent channel.
The ACLR impacts the allowed channel spacing and adjacent channel interference ACI performance. For WiMAX, the ACLR requirements are given by national regulatory bodies (such as FCC, ETSI, TTA) and network service providers. It depends on regions, frequency bands, and channel bandwidths.
The ACLR and EVM requirements of WiMAX combined with a PAPR value of 12 dB for the signals, are very challenging specifications for linearity and efficiency of the transmitter. To fulfil linearity requirements, it is possible to use the PA with a high level of back-off. However, high PA back-off has two drawbacks. The first is the necessary over-sizing of the PA: to transmit a WiMAX signal with a maximal power of 23 dBm using a PA with a back-off of 8 dB, the saturation power of the PA must be superior to 31 dBm. The second drawback is the resulting poor efficiency, since the linear class power amplifiers have a much better efficiency for large values of mean input power than for low values. The common efficiency obtained with a class AB PA for a WiMAX signal is smaller than 20%, which is much smaller than what could be achieved for constant envelope signals such as GSM signals.

**Frequency Bands**
As indicated in Table 1, there are several possible frequency bands whose bandwidths are wide: 100 MHz or 200 MHz. The number and wideness of the frequency bands have consequences on the RF filtering (spurious signals are poorly filtered) and on the necessary tuning range of the RF blocks: broadband PA, LNA and synthesizer. For mobile terminals and ESTI specifications, the level of spurious signals should be below -30 dBm for a measurement bandwidth equal to 10 kHz, 100 kHz or 1 MHz depending on the frequency range spacing from the carrier.

**Channel Bandwidth and Channel Frequency Step**
The mobile terminal must be able to deal with the different possible channel bandwidths (see Table 1).
For Mobile WiMAX channel profiles, frequency synthesizers must be able to generate frequencies with steps of half the frequency raster (see Chapter 3).

**MIMO and smart antenna possibility**
The WiMAX standard includes the possibility to use MIMO (Multiple-Input Multiple-Output) technology and specifies the support for multi-antenna transceivers. MIMO technology takes advantage of diversity and spatial multiplexing in order to improve the quality of the communication and to increase the data rate in multi-path environments. Implementing MIMO technology in a mobile WiMAX terminal is a challenge because of the size of antennas in the considered frequency bands. A WiMAX transceiver can contain several transmit and receive channels: typically 1 transmit and 2 receive channels.

### 1.1 Regulations Characteristics

The WiMAX standard includes two sub-standards. The first standard is 802.16-2004 or Fixed WiMAX, which is designated for fixed Line-Of-Sight (LOS) point-to-multipoint wideband communications within the radio band of 10–66 GHz. The Fixed WiMAX employs a Wireless MAN Single Carrier (SC) transmission with QPSK, 16-QAM or 64-QAM modulation schemes. Either the Time Division Duplexing (TDD) or the Frequency Division Duplexing (FDD) method can be used.
The second standard is 802.16e-2005 or Mobile WiMAX, which operates in multiple licensed bands within 2-6(11) GHz and enables full mobile broadband access in a cellular network (for fixed and mobile Non-LOS (NLOS) applications). As in the Fixed WiMAX standard, both duplexing methods (TDD and FDD) can be used. Air-interface designations of the Mobile WiMAX are following:

- Wireless MAN-SCa
- Wireless MAN-OFDM
- Wireless MAN-OFDMA

Single carrier based Wireless MAN-SCa employs Time Division Multiple Access (TDMA) using modulation techniques such as BPSK, QPSK, 16-QAM, 64-QAM and 256-QAM.
Orthogonal Frequency Division Multiplexing (OFDM) and OFDM Access (OFDMA) based WMAN use 256 OFDM or a scalable OFDM with 128, 512, 1024 or 2048 carriers modulated with BPSK, QPSK, 16-QAM or 64-QAM. The channel bandwidth is variable in all of the above mentioned modes and it can take different values between 1 MHz and 20 MHz.
Due to a very large variety of multiple variables that can be employed, the WiMAX Forum organisation has established certification profiles, in order to promote the compatibility and interoperability of wireless communication products. The basic radio channel features are summarized in Table 1 and will be considered in the RF design.
Data flow rates depend on the modulation, access mode and the channel bandwidth. For example, in WMAN-OFDM and OFDMA, it is possible to achieve a theoretical throughput of 73.19 Mbit/s for a band of 20 MHz and 256 carriers using a modulation type 64-QAM. As for power requirements, the minimum power is: – 50 dBm for WMAN-SCa and -45 dBm for WMAN-OFDM and OFDMA. Maximum power at emission must comply with the classification as follows (transmit power for QPSK), (WiMAX Forum, 2008):

- Class 1: $20 \leq P(dBm) \leq 23$
- Class 2: $23 \leq P(dBm) \leq 27$
- Class 3: $27 \leq P(dBm) \leq 30$
- Class 4: $30 \leq P(dBm)$

The signal integrity can be characterized by the EVM. For example, WiMAX OFDMA requires an EVM less than 3.16% (for 64-QAM(3/4)) (IEEE, 2005).

| Frequency Range (GHz) | Channel Frequency Step (kHz) | Channel Bandwidth (MHz) | FFT Size | Duplexing Mode |
|---|---|---|---|---|
| 2.3-2.4 | 250 | 5 | 512 | TDD |
| | | 10 | 1024 | |
| | | 8.75 | 1024 | |
| 2.305-2.320<br>2.345-2.360 | 250 | 3.5 | 512 | TDD |
| | | 5 | 512 | |
| | | 10 | 1024 | |
| 2.496-2.69 | 250 (200) | 5 | 512 | TDD |
| | | 10 | 1024 | |
| 3.3-3.4<br>3.4-3.8<br>3.4-3.6<br>3.6-3.8 | 250 | 5 | 512 | TDD |
| | | 7 | 1024 | |
| | | 10 | 1024 | |

Table 1. Mobile WiMAX profiles defined by the WiMAX Forum (WiMAX Forum, 2008).

## 2. Front-End Architecture Challenges

### 2.1 Architectures for WiMAX

Wireless communications require highly efficient and compact transceivers, whatever the signal characteristics are. In this section we focus only on the transmitter where design challenges are more critical in terms of power, size and consumption. A WiMAX transmitter architecture must meet design constraints of: providing high efficiency and linearity for a wideband OFDM (the bandwidth can be up to 100 MHz) and high PAPR signal (or high dynamic range) in the typical range of 20 dB (29 dB theoretical maximum). The linearization of the transmitter is mandatory because power amplification of the WiMAX signal introduces Non-Linearities (NLs) in amplitude and phase, as illustrated in Figure 2.

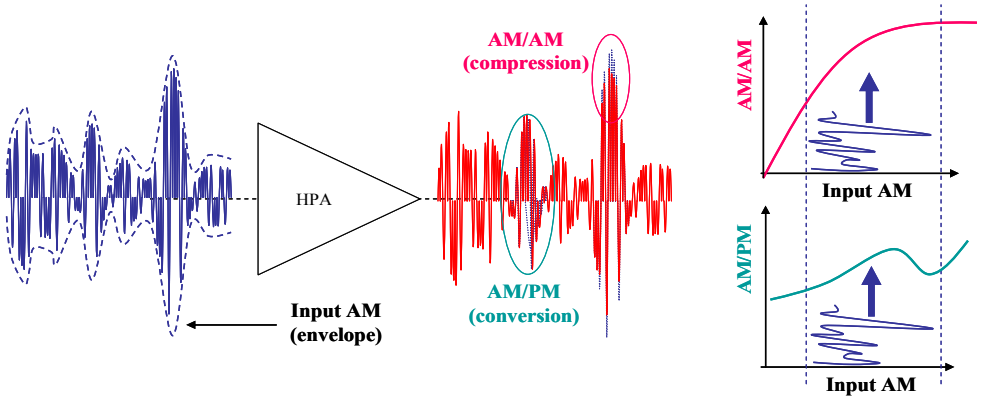



Fig. 2. Non Linearity effects of compression and conversion in high power amplification.

Identifying a type of architecture for WiMAX requires a careful study of linearization techniques and their performance with wideband and high dynamic range signals. There are several linearization techniques depending on the PAPR of the signal, the added

complexity and the increase in size and consumption of the system that can be accepted by the designers (Villegas et al., 2007). Many criteria characterize the linearization techniques such as static/dynamic processing, adaptability, frequency (digital, baseband, IF or RF), memory effect correction, complexity, stability, resulting efficiency, size increase etc. Herein, for WiMAX application, we basically classify these techniques in three groups: (i) correction techniques, (ii) anticipation of NL and (iii) those based on a decomposition and recombination of the signal, often dedicated to wideband signals.
Examples of correction techniques are feed-back (*A*), feed-forward (*B*) and the anticipation technique principle of pre-distortion (*C*) (see Figure 3). Their common point is to add a modification/correction to the modulated signal (before or after) at the PA stage. The architecture considerations here do not include the modulator nor the baseband signal processing. This needs a carefully derived model of the PA non-linear effects (Volterra series, Wiener or Saleh model etc.). Adaptability to the signal amplitude can be introduced in order to compensate for the lack of accuracy in the NL effects model and memory effects of the PA (also temperature drift compensation can be considered) (Baudoin et al., 2007).

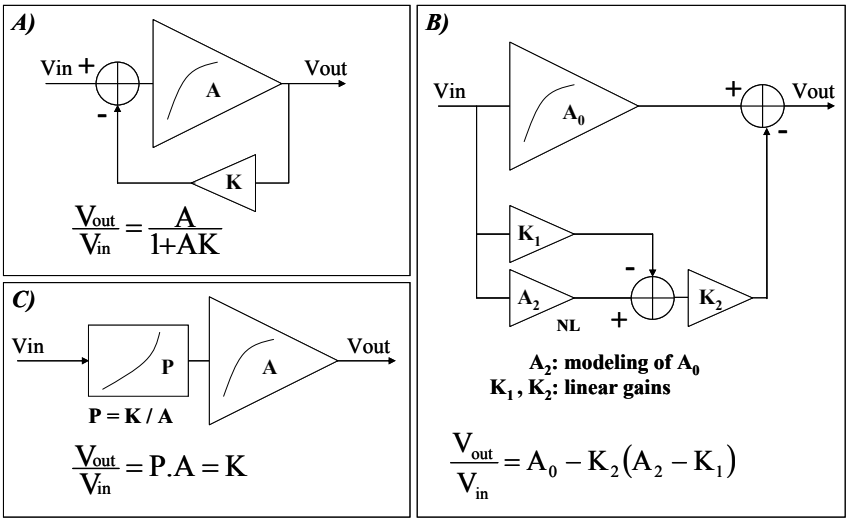



Fig. 3. Principles of feed-back (*A*), feed-forward (*B*) and pre-distortion (*C*).

Each structure contains a major defect. Feed-back (*A*) reduces the gain of the amplification and introduces a bandwidth limitation due to the transfer function of the loop (stability and dynamic response). The feed-back can be realised on the amplitude (Polar feed-back) or on I and Q quadrature components of the signal (Cartesian feed-back) and both are dedicated to narrowband signal linearization. Feed-forward (*B*) requires a significant increase in signal processing and RF blocks in the transmitter, with the hypothesis of a precise matching between NLs and reconstructed transfer functions. The improvement in linearity will be expensively paid for in terms of consumption and size (integration). Advantages are the stability and possibility to process wideband signals. The most interesting is pre-distortion (*C*) because of its flexibility: the anticipation can be done in the digital part and so provide adaptability of the technique, but this needs a feed-back loop. The digital pre-distortion

represents a non-negligible additional consumption of a Digital Signal Processor (DSP) and often requires a look up table. The signal is widened in frequency because of the expensive non-linear law of the pre-distorter (as for IPx theory on a modulated signal spectrum), requiring baseband and RF parts to be wideband designed. Interesting improvements of pre-distortion have been made with OFDM signals in (Baudoin et al., 2007).
Others techniques presented are based on a vectorial decomposition of the signal in order to drive high efficiency switched mode RF PAs with constant envelope (constant power) signals, avoiding AM/AM and AM/PM (Raab et al., 2003); (Diet et al., 2003-2004). These techniques are dedicated to correcting strong NL effects. We consider the problem of linearization in the communication chain from the digital part to the emission. This drives for a complete modification of the architecture and its elements' specifications in baseband, RF and power RF. After amplification of constant envelope parts of the signals, the difficulty is to reintroduce the variable envelope information with lower NL than in a direct amplification case, while maintaining the efficiency of the architecture. Basic examples of these techniques are the LInearization with Non Linear Components (LINC) and the Envelope Elimination and Restoration (EER) methods (and theirs recent evolutions) (Cox, 1974); (Kahn, 1952); (Diet et al., 2004).
The LINC principle relies on a decomposition of the modulated signal into two constant envelope signals as is shown in Figure 4. The decomposition can be computed by a DSP or by combining two Voltage Controlled Oscillators (VCOs) in quadrature locked-loop configuration (CALLUM). CALLUM is an interesting architecture but presents a possibility of instability and additional costs of realisation. The amplification of these two constant envelope signals drives for the design of two identical High Power Amplifiers (HPAs) at the RF frequency, and often causes signal distortion due to imbalance mismatch. Also the HPA has to be wideband because the signal decomposition is a non-linear process, and the phase modulation ratio is increased.

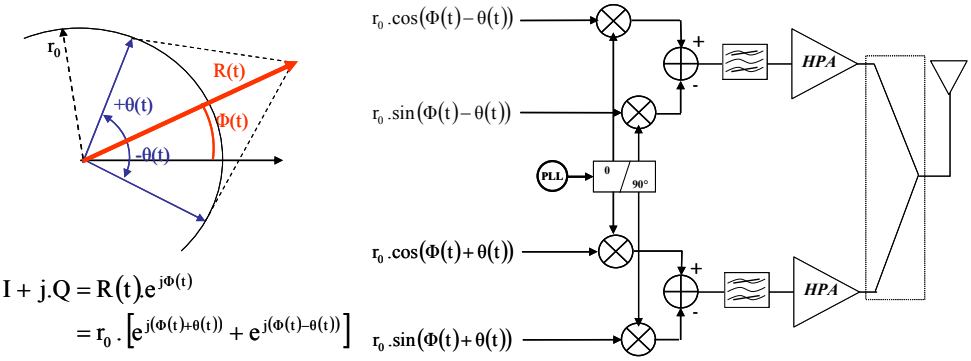


Fig. 4. LINC decomposition and recombination at RF power amplification.

Whatever the decomposition technique is (LINC/CALLUM), the default is that efficiency is directly determined by the recombination sum operation. It is very difficult to avoid losses at high frequency while designing an RF power combiner.

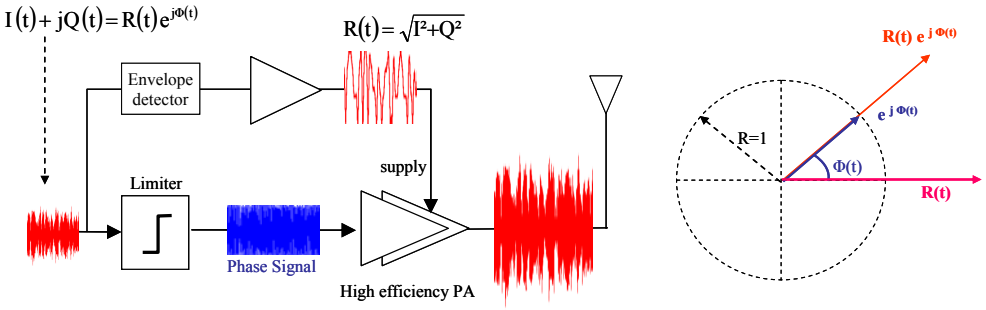



Fig. 5. Principle of the EER technique (Kahn, 1952).

Another decomposition technique was proposed by Kahn in 1952 and is basically an amplitude and phase separation (EER). This method was first proposed for AM signals as represented in Figure 5. The advantage of EER is to drive the RF PA with a constant envelope modulated signal (carrying the phase information), enabling the use of a switched and high efficiency amplifier (Raab et al., 2003); (Sokal & Sokal, 1975); (Diet et al., 2005-2008). The difficulty is to reintroduce the amplitude information using the variation of the PA supply voltage. This implies a power amplification of the envelope signal, at the symbol rate frequency. The recombination can be done with high efficiency switched (saturated) class PA because their output voltage is linearly dependant on the voltage supply. Synchronisation between the phase and the amplitude information and linear amplification of the amplitude before the recombination are the two major difficulties of such a linearization technique as reported in (Diet, 2003-2004), where a maximum delay of 3 nanoseconds during the recombination of a 20 MHz OFDM 802.11a signal causes spectral re-growth of more than 40 dBc (standard limit) at 30 MHz from the carrier frequency (5 GHz). Recently, a lot of work has been done on the EER based architectures, often classified as polar architectures (Nielsen & Larsen, 2007); (Choi et al., 2007); (Suarez et al., 2008); (Diet et al., 2008-2009). The generation of the amplitude and phase components can be expected to be done numerically thanks to the power of DSPs, as is shown in Figure 6. As was previously discussed in (Diet et al., 2003), the bandwidths of the envelope and phase signals are widened and make it necessary to design the circuit for three to four times the symbol rate. For a WiMAX signal, such a technique requires bandwidth on the phase and amplitude baseband paths in the range of a hundred MHz (as for LINC technique and any other non-linear decomposition method). Since a clipping in frequency and on the envelope is possible, they are suited for new high data rate standards such as WiMAX, where efficiency of the emitter and linearization are mandatory. Also, the multi-standards and multi-radio concepts evolved the polar architectures in multiple ways (Diet et al., 2008). For example, the recombination on the drive signal of the PA is possible because the amplitude information modulates the phase RF signal and is restored by the band-pass shape function of the following blocks: PA + emission filter + antenna. The emitted spectrum is the criterion of quality to be considered carefully, because the Pulse-Width Modulation (PWM) or ΣΔ envelope coding are the source of useless and crippling spectral re-growth. The efficiency is also penalized by the power amplification of such useless components but this is balanced by the advantage of high flexibility of this architecture (Robert et al., 2009). Actual work is focused on the front end design to provide the highest efficiency possible with the PA

driven signal composed by the phase and amplitude coded information. The digitally controlled PA and the mixed-mode digital to RF converter performance are key parameters in the evolution of polar architectures (Suarez et al., 2008); (Robert et al., 2009); (Diet et al., 2008).

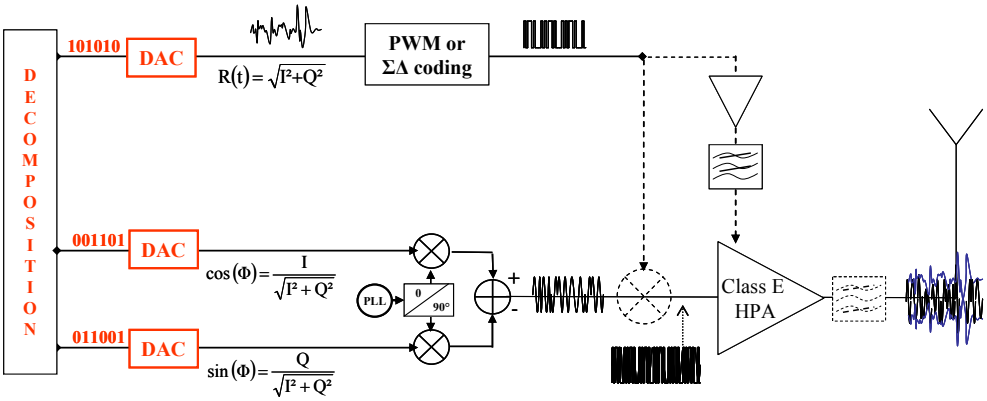



Fig. 6. Recent improvements of EER/polar architectures for wideband OFDM signals.

To summarize, WiMAX architecture is expected to be wideband and high efficiency due to the signal characteristics. Polar based architectures seem to be the best candidate at the moment and corresponds to the actual focus in radio-communication research topics. The high performance of digital adaptive pre-distortion techniques on OFDM signals should be noted. It is later of interest in a polar/pre-distorted mixed linearized technique.

### 2.2 PAPR Reduction

As mentioned above, the OFDM signals suffer from the high envelope dynamics, characterized by the PAPR. The PAPR reduction methods can be categorized into two general groups - the methods introducing signal distortion and the distortionless methods. The simplest method from the former group is the clipping. In its basic form, this method simply clips the signal magnitude to the desired maximal level. In this case the PAPR is reduced at the expense of signal distortions resulting in out of band emissions and BER degradation. In order to suppress these effects, the research was directed towards more advanced methods like repeated clipping and filtering etc.

Many techniques that do not introduce the distortions have been proposed in the past. The Selective Mapping (SLM) is based on the creation of several alternative signal representations of the original input data, bearing the same information (LIM et al., 2005). The one with the lowest PAPR is selected for transmission. In the Partial Transmit Sequences (PTS) (Mäuller & Huber, 1997) method, the IFFT input symbols are divided into several frequency disjoint sub-blocks. The output of each sub-block is multiplied by the complex rotation factor. These factors are optimized in order to find the variant with the lowest PAPR. Moreover, methods like the tone reservation or tone injection, belonging to the group of so-called additive methods (Tellado, 2000), can be used as well.

The potential of PAPR improvement can be illustrated on the example resulting from the use of PTS (4 sub-blocks, rotation factors being either 1 or -1) and SLM (6 alternatives) methods shown in Figure 7 for the case of FFT size equal to 256.

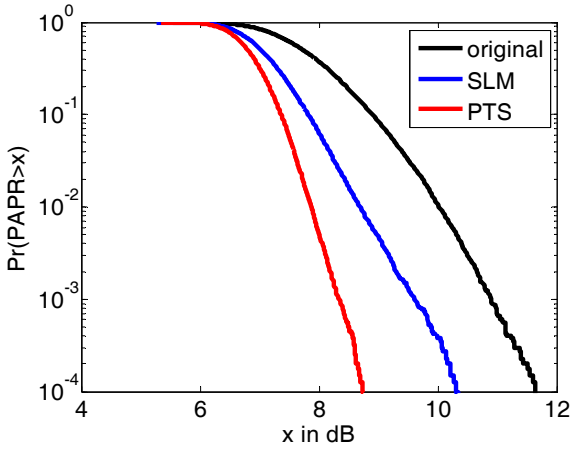



Fig. 7. The PAPR improvement due to use of PAPR reduction for FFT size equal to 256

### 2.3 Integration Technologies: System-in-Package (SiP) and System-on-Chip (SoC)

Due to the increasing complexity of integrated circuits and systems, new packaging technologies are developed. Their choices shape the design of elementary circuits.
Three constraints are essential to take into account in the design and integration of architecture transceiver:

- Improving performances
- Circuit size and costs
- Energy consumption

To tackle these challenges, the silicon integration and the packaging must be taken into account when designing systems.
SoC have different digital and analog functions live on the same chip. It is a full integration on silicon. SoC technology combines collectively processed components on the same chip, and using compatible processes. It is more complex to design but can achieve more competitive prices for major production.
The overall reliability is also improved. But today, some functions such as filtering and power amplification can not employ the SoC technology, particularly for WiMAX systems.
The system must be divided into functions. This technique requires co-design. The chips are then assembled using a specific connection technology. The design constraints are released but the unit cost is higher. The reliability is lower and the operating speed of circuits may decrease depending on the chosen connection technology.

## 3. Frequency Synthesis for Mobile WiMAX Radios

Unlike 2G and 3G wireless communication systems with a fixed channel bandwidth, the Mobile WiMAX standard enables a variable channel size and multiple frequency allocation, and thereby offers very flexible deployment. As a result of multi-wideband operation, a high performance RF front-end has to be employed. One of the most challenging RF blocks

to design in a radio front-end is undoubtedly the frequency synthesizer. The frequency synthesizer acts as a Local Oscillator (LO) generation unit that is used to translate baseband and RF signals by means of mixing and it frequently determines the overall performance of a radio communication system. Evaluation of the frequency synthesizer's complexity and requirements relates to the WiMAX RF architecture. WiMAX considers three RF architectures, TDD, FDD and Half FDD (HFDD). Since the TDD mode utilizes a single frequency band for the uplink and downlink direction, only one local oscillator is required. Contrary to the TDD, the FDD mode requires two separate synthesizers for RX and TX due to the full duplex nature of the architecture and hence, switching requirements can be relaxed. However, the overall complexity of the FDD front-end leads to larger RFIC due to additional circuitry, higher power consumption and most importantly, to higher costs of implementation. Therefore, the TDD architecture is more suitable for the mobile version of the WiMAX standard.

Multi-carrier OFDM scheme, which is employed in WiMAX, has become very popular in wideband communication systems as it offers very high spectral efficiency and can efficiently combat selective fading resulting from multipath propagation (Cimini, 1985). High spectral efficiency is achieved by parallel low-rate narrowband modulation of orthogonal sub-carriers equally spaced by $\Delta f=1/T_u$, where $T_u$ is the useful symbol length (length of the FT interval). In the ideal case, all sub-carriers are orthogonal, in other words, they don't interfere with each other. Therefore, an inter-carrier guard interval commonly used in frequency-division multiplexing is not required. However, as the sub-carriers are very close together, the OFDM signal becomes less resilient to frequency errors, namely to inaccuracies such as Doppler shift and LO phase noise (Muschallik, 1995); (Zou et al., 2007). The phase noise introduced by the frequency synthesizer can be interpreted as a parasitic phase modulation of the LO carrier that is subsequently superimposed on individual OFDM sub-carriers as a result of the frequency translation (up and down conversions). This parasitic phase modulation causes significant degradation of the OFDM signal and may lead to the loss of orthogonality as a result of Inter-Carrier Interference (ICI). The effect becomes even more harmful as the number of sub-carriers increases in a given bandwidth, because the length of the OFDM symbol $T_u$ gets longer, making the system more sensitive to the LO phase noise. However, longer symbols increase spectral efficiency of an OFDM system, and therefore, these parameters have to be compromised.

There exist two different effects of the LO phase noise on the OFDM signal: Common Phase Error (CPE) resulting from the LO noise contribution present within the sub-channel. CPE affects all sub-carriers in the same manner and can be observed as a common rotation of all constellation points in the I/Q plane (Muschallik, 1995). Other phenomenon is ICI, which results from the adjacent LO noise contribution.

Correlation between the integrated phase noise (phase jitter) and the BER degradation has been reported in (Muschallik, 1995); (Herzel et al., 2005); (Armada, 2001). In order to eliminate this impact, phase noise of the OFDM synthesizer has to be optimized according to the maximum allowed phase noise or phase jitter (for a given BER).

### 3.1 Frequency Synthesizer Requirements

The frequency synthesizer has to provide the range of all necessary frequencies with proper channel spacing that corresponds to the channel raster. WiMAX channel profiles are summarized in Table 1 (WiMAX Forum, 2008). The channel frequency step given by the

WiMAX standard is 250 kHz, however, the LO has to deliver the required frequency with a resolution of 125 kHz as a result of raster overlapping for different channel bandwidths (WiMAX Forum, 2008). Another directive parameter for design is the tuning range and the frequency settling time. Frequency switching time between RX and TX in the HFDD mode has to be performed agilely, with respect to settling time requirements of the standard, which is to be less than 50 μs. Moreover, the local frequency synthesizer has to fulfil the tightest signal purity requirements that can be expressed in terms of the integrated phase noise and the spurious output. The integrated phase noise is to be less than 1 deg rms within an integration frequency of 1/20 of the tone spacing (modulated carrier spacing) to ½ the channel bandwidth (Eline et al., 2004). Thus for smaller channel bandwidths the integration of the phase noise can start from as low as a few hundred Hertz.
To satisfy very high requirements imposed by the WiMAX whilst at the same time benefiting from the ease of integration and flexibility, fractional-*N* Phase Locked Loop (PLL) based architecture becomes the appropriate design option. Fractional-*N* PLL can achieve very small frequency resolution equal to the fractional portion of the reference frequency and hence improve the phase noise performance by a factor of 20log (*N*) compared to the conventional integer-*N* PLL (Keliu & Sanchez-Sinencio, 2005).

### 3.2 High-Speed Frequency Synthesizer Design Example

Fractional-*N* synthesizers have become very popular and widely used in a range of RF applications because they allow the comparison Phase Frequency Detector (PFD) frequency $f_{PFD}$ to be significantly higher than the required frequency resolution (Keliu & Sanchez-Sinencio, 2005). Higher PFD frequency automatically leads to wider loop bandwidth (as the loop bandwidth is to be < $0.1f_{PFD}$) and therefore significantly improves settling time performance compared to the integer-*N* PLL. Moreover, since the Mobile WiMAX standard defines the channel raster resolution as 125 kHz, the PFD frequency of an integer-*N* synthesizer would have to be as low as 125 kHz and hence, the division factor *N* would reach up to 30,400 in order to generate the highest frequency of 3.8 GHz (Valenta et al., 2008). This would, in turn, deteriorate the in-band phase noise contribution by up to 48 dB, compared to the fractional-*N* synthesizer with a reference frequency of 32 MHz (where $N \approx 119$, assuming the same phase noise performance of both reference clocks).
In the next example, we consider a fractional-*N* Charge Pump (CP) frequency synthesizer as an application for the 3.4 – 3.8 GHz channel profile. A simplified model is depicted in Figure 8. This model includes a tri-state PFD that produces output *up* and *down* signals, proportional to the phase and frequency difference between the reference and the feedback signal. PFD employs two positive edge-triggered resetable FF (flip-flop) to detect the phase and frequency difference and one AND gate to monitor the *up* and *down* signals. The upper FF is clocked by $f_{ref}$, the lower by $f_{div}$. Signals *up* and *down* are used to switch the current sources in the CP. These CP current pulses change the voltage drop on the loop impedance and tune the VCO with tuning gain of 125 MHz/V and tuning range of 3.4-3.88 GHz. The fractional division is achieved by altering the division value between two integer values *N* and *N*+1, hence the average division becomes a fraction. However, the periodic switching between two division values leads to a sawtooth phase error, which can create spurious fractional tones. This problem is solved with help of ΔΣ modulator, which randomize the switching between *N* and *N*+1, but on the other hand, it introduces quantization noise into the loop.

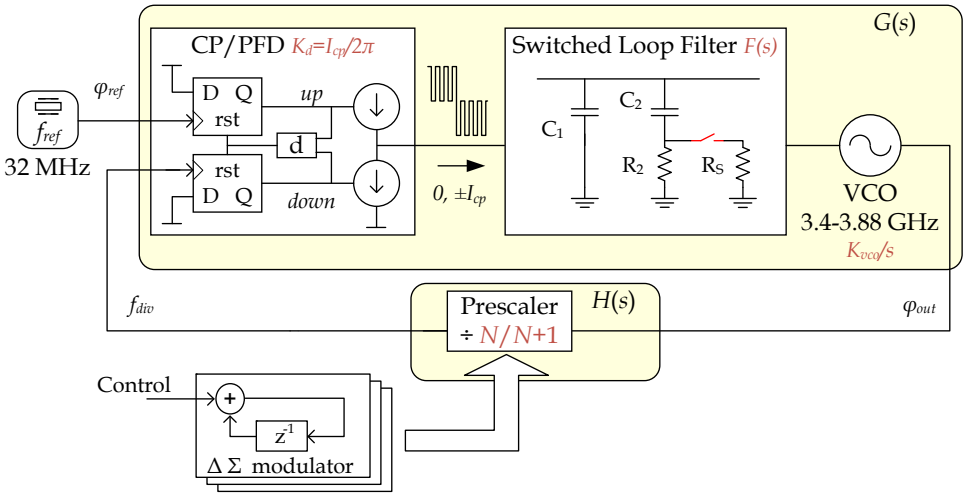



Fig. 8. General model of the fractional-*N* charge pump synthesizer.

The loop filter is the key component of the PLL and it characterizes the dynamic performance of the synthesizer. Loop filter design involves choosing the proper loop filter topology, loop filter order, phase margin and loop bandwidth. Due to low phase noise requirements set by the Mobile WiMAX standard, a passive filter has been chosen. The trade-off between the minimum loop bandwidth and the settling time has to be taken into account for optimal loop filter design. A simplified trade-off presented by (Crawford, 1994) is $B_{PLL} = 4/t_{lock}$, where $t_{lock}$ is the settling time. Applied to WiMAX, the PLL would call for at least 200 kHz loop bandwidth in order to settle within 20 μs. However, such a wide PLL bandwidth would result in a very high in-band phase noise integration and phase jitter deterioration.

A significant settling time improvement can be achieved by means of loop filter switching (Crowley, 1979). The loop filter is switched to the fast wideband mode during the frequency transition and then, after a certain programmable period, is shifted back to the normal narrowband value. To understand the switching principle, let us have a look at the PLL control theory and the PLL linearized model. The effect of a closed feedback loop on the input reference signal $\varphi_{in}$ can be described by the closed loop transfer function $T(s)$ as:

$$T(s) = \frac{\varphi_{out}(s)}{\varphi_{in}(s)} = \frac{G(s)}{1 + G(s)H} = \frac{\frac{K_d K_{vco}}{s} F(s)}{1 + \frac{K_d K_{vco}}{s} F(s) \frac{1}{N}} \tag{4}$$

where the $G(s)$ represents the open loop transfer function and $H$ corresponds to the division factor $1/N$. $K_d$ is the gain of the CP/PFD detector and equals to $I_{cp}/2\pi$, $K_{vco}$ is the VCO gain in MHz/V and $F(s)$ refers to the transimpedance of the second order loop filter (see Figure 8.).

$$F(s) = \frac{1 + sC_2R_2}{s(C_1 + C_2)\left(1 + s\dfrac{C_1C_2R_2}{C_1 + C_2}\right)} \tag{5}$$

The angular open loop crossover frequency $\omega_c$ and the phase margin $\theta_c$ are defined at the point where the magnitude of the loop gain reaches unity. This can be expressed as $\|G(s)H\|=1$ (0 dB), where

$$G(s)H = \frac{K_dK_{vco}F(s)}{sN} = \frac{I_{cp}K_{vco}F(s)}{2\pi sN}, \tag{6}$$

$$G(s)H\big|_{s=j\omega_c} = -\frac{I_{cp}K_{vco}}{2\pi\omega_c^2N}\frac{1 + j\omega_cT_2}{1 + j\omega_cT_1}\frac{1}{C_1 + C_2}. \tag{7}$$

And then, the open loop phase margin $\theta_c$ at the crossover frequency $\omega_c$ reads

$$\theta_c[rad] = \pi + \arctan(\omega_cT_2) - \arctan(\omega_cT_1). \tag{8}$$

$T_2$ and $T_1$ correspond to time constants of zero and the pole in the loop filter transfer function respectively ($T_2=C_2R_2$, $T_1=C_1C_2R_2/(C_1+C_2)$).
Let us consider a situation, where the crossover frequency $\omega_c$ is increased by factor $a$ in order to increase the loop bandwidth and thereby decrease the settling time. This adjustment is applied only during the frequency transition. To ensure the loop stability at $a \cdot \omega_c$, the phase margin $\theta_c$ defined by the equation (8) has to remain constant. This can be done by means of reducing the value of $T_2$ and $T_1$ by the factor $a$ with help of a parallel resistor $R_s = R_2/(a-1)$. as displayed in Figure 8. Moreover, the product of all elements in (7) has to be increased by factor of $a^2$ as the angular frequency $\omega_c$ in (7) is in the power of two. This can be done by

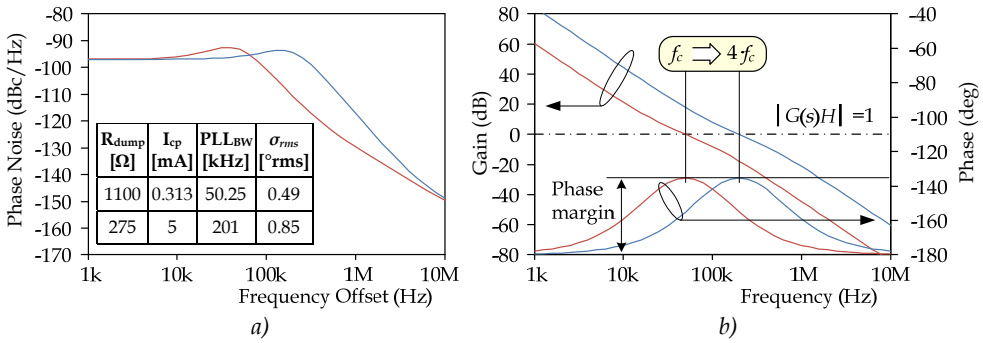



Fig. 9. Phase noise performance (*a*) and the open loop gain with the phase (*b*) at 3.59 GHz. The blue line corresponds to behaviour during the frequency transition (wideband mode).

means of increasing the charge pump current $I_{cp}$ by factor $a^2$. In this example, we consider $a$ =4 and therefore, to activate the fast wideband mode, the CP current is increased by a

factor of 16 ($I_{cp} \rightarrow 16 \cdot I_{cp}$), while reducing the dumping resistance by a factor of 4 (by using a parallel resistor $R_s$). The PLL open-loop crossover frequency, the zero and pole frequency ($1/R_2C_2$ and $1/[R_2C_1C_2/(C_1+C_2)]$) are all increased by a factor of 4 while the loop stability remains unaffected (see the constant phase margin in Figure 9 *b)*. Figure 9 *a)* shows the phase noise performance and the resulting phase jitter of the dual bandwidth adjustment. Phase jitter of both loop configurations has been calculated as follows:

$$\sigma_{rms}[^\circ] = \frac{180}{\pi}\sqrt{2\int_{f1}^{f2} S\,(f)df}\ , \tag{9}$$

where $S(f)$ is the total phase noise at the PLL output. $f_1$ and $f_2$ correspond to the integration borders (from 1/20 of the carrier spacing to half the bandwidth). Figure 10 presents the settling time improvement that has been achieved by means of loop filter switching. The settling time has dropped from 88 µs to 32 µs (settling time within 100 Hz).

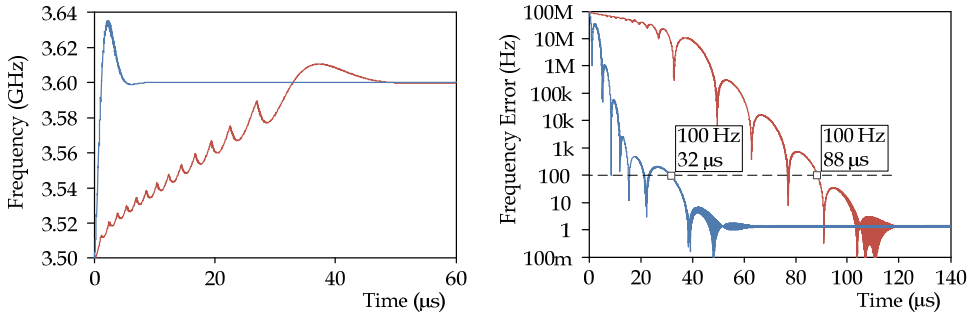



Fig. 10. Transient responses of the fast PLL for two cases: wideband mode enabled/disabled (blue/red line respectively). Plot *b)* displays the absolute frequency error to 3.6 GHz.

### 3.3 Hybrid PLL Approach for Frequency Synthesis

The product of all elements in equation (7) can be changed not only by increasing the $I_{cp}$, but also by simultaneous altering of the feedback division factor $N$ and the $I_{cp}$. However, by altering the division factor in the feedback path, the output frequency will shift as well. To keep the output frequency constant either during the transition, or once in the stable mode, the hybrid fractional/integer PLL approach has to be considered (Memmler et al., 2000); (Kyoungho et al., 2008). The fractional-$N$ wideband mode is enabled along with the loop switching during the frequency transition and then, after settling, the integer-$N$ narrowband mode is turned on. This approach brings a new degree of flexibility and alleviates CP requirements in terms of the $I_{cp}$ current range by reducing the division ratio $N$. In a very particular situation, the bandwidth can be switched only by altering the feedback division factor and the dumping resistance while keeping the $I_{cp}$ constant. In the fast wideband fractional-$N$ mode, the additional dividers ($x$–labelled blocks in Figure 11) are not employed and the reference frequency as well as the feedback signal is applied directly to the PFD. After settling, the integer-$N$ mode is enabled by switching the loop bandwidth to the normal

narrowband value and by activating two additional dividers. The additional division ratio $x$ is chosen such that the PFD frequency corresponds to the required frequency resolution. Applied to the previous example where a 32 MHz reference is used, the division $x$ has to be 256, in order to achieve 125 kHz resolution. The wideband fractional mode is then enabled by switching the resistance $R_s$ and by disabling both dividers. This adjustment results in boosting the loop bandwidth by factor 16 while keeping the $I_{cp}$ constant. The stability is not affected since $R_s=R_2/(a\text{-}1)$, where $a=\sqrt{256}$ and at the same time the ratio $I_{cp}/N$ is increased by factor $a^2$; $\{I_{cp}/(N/a^2)\}$.

The hybrid approach inherits the speed performance from the fractional-$N$ PLL and at the same time the design simplicity of the integer-$N$ PLL (the $\Delta\Sigma$ modulator can be replaced by an accumulator because the fractional spur reduction is not needed during the transient period). Moreover, a higher loop bandwidth enlargement is possible compared to the conventional bandwidth switching (where only $I_{cp}$ and $R$ are altered). On the other hand, the main drawback of the hybrid architecture is evident: the phase noise performance of such a synthesizer corresponds to the performance of an integer-$N$ synthesizer, which has inherently worse phase noise performance compared to the fractional-$N$ architecture.

Due to the high degree of flexibility and integrability, the hybrid PLL approach is a very promising choice for multi-standard and multi-band transceivers, where different standards impose different requirements in terms of phase noise, settling time or channel raster (Valenta et al., 2009).

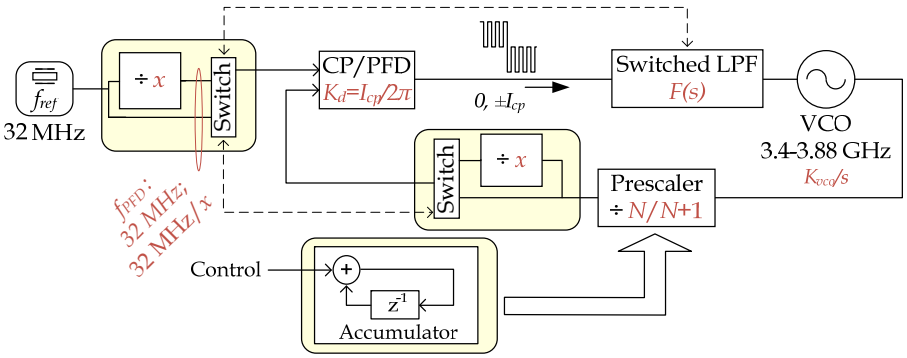



Fig. 11. Functional scheme of the hybrid PLL synthesizer.

## 4. High Performance Power Amplifier Design for Mobile WiMAX Radios

### 4.1 Transistor Technology

The technologies available to realize integrated circuits in the WiMAX bands are as follows. Gallium Arsenide (GaAs), which allows high operating frequency and high output power to be achieved. This technology is generally used for power amplifiers and switches. The manufacturing cost is higher compared to silicon. Using this technology requires a SiP packaging.

Silicon technologies are numerous: silicon bipolar transistor, CMOS, BiCMOS etc. Under certain conditions, they can achieve good performances, up to tens of GHz. A final interesting technology is the SiGe that may eventually replace the GaAs for some functions.

Figures of merit are classically used to describe RF and microwave transistors characteristics. They are used by the designer to compare the performance of technologies. Two frequency characteristics are used: the cutoff frequency $f_t$, and the maximum oscillation frequency $f_{max}$. These frequencies define the upper limits of component operation. They are particularly important for low noise or power devices.
The cut-off frequency $f_t$ is defined when the amplitude of the current gain ($H_{21}$) is equal to 1 (0 dB) and the output is short circuited. The maximum oscillation frequency $f_{max}$ is defined when the unilateral power gain is equal to 1 (0 dB).
For RF or microwave integrated circuit design, $f_t$ and $f_{max}$ must be as high as possible. A rule of thumb is to give an operating frequency equal to 10% of the maximum $f_t$.
There are other electrical characteristics used depending on the application. The minimum noise factor, $N_{fmin}$, given in dB at the working frequency, it is important for all low-noise components. The output power $P_{out}$, given in dBm or W, is important for power components. For such power applications, it is also important to define the output power density, expressed in watts per millimetre gate development in the case of field-effect transistors, and emitter surface in the case of bipolar transistors.

### 4.2 Continuous Wave (CW) Classes

Power Amplifiers in RF applications such as WiMAX are designed to linearly amplify high dynamic signals with high efficiency, due to the consideration of the battery lifetime. They are loaded by the emission filter and/or the antenna and deliver a RF power that can reach 27 to 30 dBm. In that context, the performance evaluation criteria are the spectral re-growths as a form of measure of the linearity (linked ACPR and to the ACLR), the factor of use *Fu* as a technological aspect for the design, see (10), and the efficiency of operation; see (11) for drain/collector $\eta$ and (12) for added efficiency $\eta_{AE}$. For simplicity, PA topologies are presented here for common emitter RF MESFETs and can be easily adapted to BJT, HBT, (LD)MOS or x-HEMT transistors.

$$Fu = \frac{V_{\max} \times I_{\max}}{P_{RF_out}} . \tag{10}$$

$$\eta = \frac{P_{RF_out}}{P_{DC}} . \tag{11}$$

$$\eta_{AE} = \frac{P_{RF_out} - P_{RF_in}}{P_{DC}} . \tag{12}$$

PAs are grouped in several classes of operation. A class is determined by (i) the polarisation of the transistor, (ii) the wanted shape of the output signal (voltage and current) and (iii) the hypothesis on transistor saturation (current source or switch behaviour). There are two families of PA classes: the switched class (SW), discussed in the following section, and the CW or biased class. A CW class is characterized by the polarisation point of the transistor, implying the conduction angle ($\theta$) of the PA, as illustrated in Figure 12.

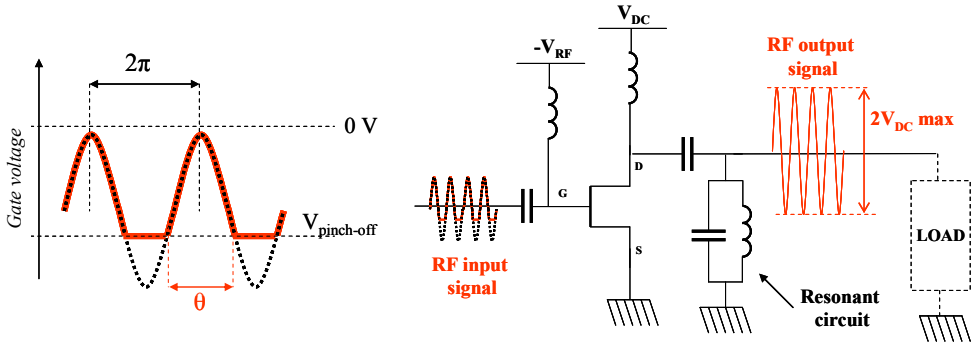



Fig. 12. Angle of conduction and polarization of a CW MESFET PA.

The PA classes with $\theta \geq \pi$, are class A ($2\pi$), AB ($2\pi > \theta > \pi$,) and B ($\pi$). Their main characteristics are that the PA dissipates power whatever the amplitude of the signal, and therefore the maximum efficiency is obtained at maximum linear amplitude of the output voltage. For class C ($< \pi$) the dissipated power is non-linearly proportional to the input signal amplitude and increases with the reduction of $\theta$. Equations of the maximum efficiency as a function of the conduction angle $\theta = \theta_c$ are based on the analysis of (Krauss et al., 1980) and are summarised by (13), (14), (15) and in Figure 12.

$$V_{RF} = \frac{V_p}{\cos\left(\frac{\theta_c}{2}\right) - 1} \tag{13}$$

$$\eta = \frac{P_{RF}}{P_{DC}} = \frac{\theta_c - \sin(\theta_c)}{4.\left(\sin\left(\frac{\theta_c}{2}\right) - \frac{\theta_c}{2}.\cos\left(\frac{\theta_c}{2}\right)\right)} \tag{14}$$

$$Fu = \frac{V_{\max}.i_{RF}\left(1-\cos\left(\frac{\theta_c}{2}\right)\right)}{P_{RF}} = \frac{8\pi\left(1 - \cos\left(\frac{\theta_c}{2}\right)\right)}{\theta_c - \sin(\theta_c)} \tag{15}$$

It is important to consider that the efficiency is given as a peak value for CW classes. For an AM signal, the average efficiency will rely on the amplitude information statistic, if there is no saturation. An improvement in efficiency is gained if saturation/clipping on the peak values is introduced in order to increase the average power of the output signal for the same power dissipated by the amplifier. The amplifier gain depends on the load line and so also on $\theta$. Gain and linearity reduce while the efficiency increases. The WiMAX signal presents such a high PAPR that the amplification by a CW class PA requires a low efficiency back off or a linearization technique to reduce the non-linear effects of compression and conversion introduced by the saturation. Techniques interesting for wideband and high PAPR signal are those using the highest efficiency PAs. A way to improve CW efficiency is to saturate

class B or class C PAs as described in (Krauss et al., 1980) as the mixed mode class C/C*. In fact, the saturation of the PA corresponds to the use of the transistor as a switch and not as a classical current source. The amplification of amplitude information by the driving input signal becomes impossible in this mode. SW class PAs are based on this principle.

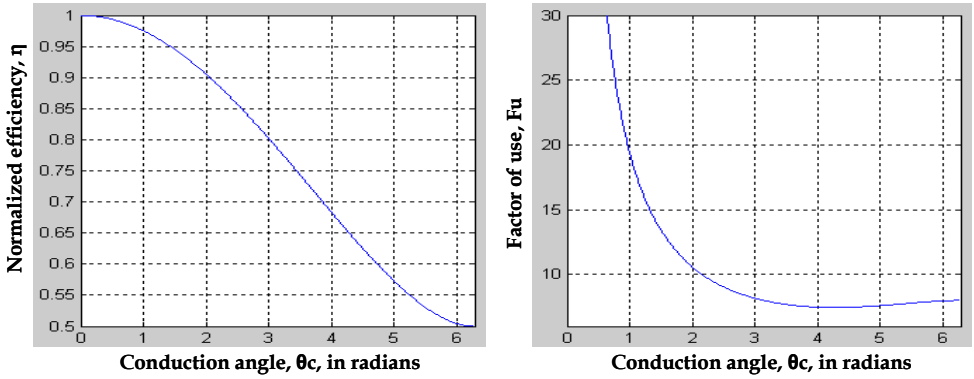



Fig. 13. Efficiency and Factor of use of CW classes.

It is important to consider that the efficiency is given as a peak value for CW classes. For an AM signal, the average efficiency will rely on the amplitude information statistic, if there is no saturation. An improvement in efficiency is gained if saturation/clipping on the peak values is introduced in order to increase the average power of the output signal for the same power dissipated by the amplifier. The amplifier gain depends on the load line and so also on $\theta$. Gain and linearity reduce while the efficiency increases. The WiMAX signal presents such a high PAPR that the amplification by a CW class PA requires a low efficiency back off or a linearization technique to reduce the non-linear effects of compression and conversion introduced by the saturation. Techniques interesting for wideband and high PAPR signal are those using the highest efficiency PAs. A way to improve CW efficiency is to saturate class B or class C PAs as described in (Krauss et al., 1980) as the mixed mode class C/C*. In fact, the saturation of the PA corresponds to the use of the transistor as a switch and not as a classical current source. The amplification of amplitude information by the driving input signal becomes impossible in this mode. SW class PAs are based on this principle.

### 4.3 SW Classes

SW classes are based on the hypothesis that the transistor switches perfectly. This latter is supposed to act as a current source and as a voltage source alternately with non-overlapping, which would result in dissipated power (Krauss et al., 1980); (Raab et al., 2003), see Figure 14. The filtering is mandatory in RF applications, except if the SW PA is dedicated to the amplification of a square signal, which is class D. When the square signal is low-pass filtered in order to recover a PWM or other pulse-coded information (ΣΔ), the amplifier is often named a class S amplifier. The load line of every SW PA class tends to be that of an ideal switch, which is impossible in practice due to the real characteristics of a transistor. These imperfections are mainly represented by resistive and capacitive channel effects,

which induce current and voltage overlapping and reduce the ideal 100% efficiency. This is illustrated in Figure 14. Although the switching cannot be perfect, the SW class presents higher efficiency than the CW class at peak amplitude signals. Average efficiency is higher for AM signals if the AM information is recombined. The problem is to reintroduce the amplitude information of AM signals efficiently. One of the solutions in the case of SW PAs is that the transistor has an output voltage linearly dependant on the drain (for a MESFET) voltage due to the saturation, and this variation does not theoretically affect the high efficiency. This enables a supply modulation possibility in order to amplify non-constant modulated signals.

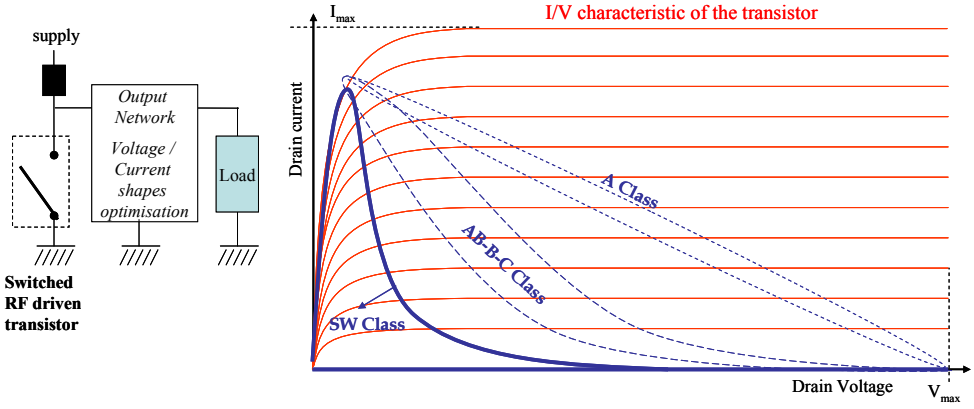



Fig. 14. Principle of SW classes (left) and Load line for CW and SW PA Classes (right).

Some SW classes can be considered for the emission of RF signals: class F and E. A class F PA is a saturated transistor, switched at the centre frequency, with additional resonating tanks in order to produce a sum of odd harmonics on the current and even harmonics on the voltage. This results in the generation of RF power only at the fundamental frequency, restoring the switching signal. It mathematically provides a null dissipated power (100% efficiency) but in practice the class F performance is limited by the non infinite number of resonating tanks. Another limitation of this PA class is the large number of shunt and series reactive elements necessary, all optimised for one frequency of operation. The tolerances of the inductors and capacitors limit the performance of the PA due to its increase in sensitivity to the values of the reactive elements.

Class E PA design is different from class F. The number of shunt and series reactive elements (2 to 3) of the output network is limited and the calculus of optimum values is far more complex because it results from several hypotheses on the drain voltage and drain current shapes. These are: (i) perfect switching of the transistor at the centre frequency (duty cycle 50%) implying no overlapping of the drain voltage and drain current, (ii) perfect filtering of the fundamental frequency signal delivered to the load, (iii) null drain current in the transistor off state and null voltage in the transistor on state and (iv) null value of the drain voltage derivate when the transistor is switched on. Computations of these hypotheses discussed in (Sokal & Sokal, 1975); (Krauss et al., 1980); (Raab et al., 2003); (Diet et al., 2004-2008); (Robert et al., 2009) are summarized in Figure 15, where $\theta = \omega t$.

Theoretical equations determine particular shapes of the drain voltage and drain current for the class E operation. This is the typical reference for the PA designer. The large value of the factor *Fu*, particular to the class E, results in the PA being large, which is costly when integrating. Despite these design difficulties, a class E PA is very attractive because it provides very high efficiency regarding the number of reactive elements necessary for the output network. Some parasitic effects due to the transistor technology such as the drain capacitance (NL) and the drain resistance can be introduced in the determination of the (inductive) optimal impedance of the output network $Z_{in}(\omega)$. Works like those of (Diet et al., 2008) show that it is possible to determine some different output networks providing the same $Z_{in}(\omega)$. This was discussed for the first parallel inductance class E topology in (Grebennikov, 2002). Two class E topologies where compared in (Diet et al., 2008) for their

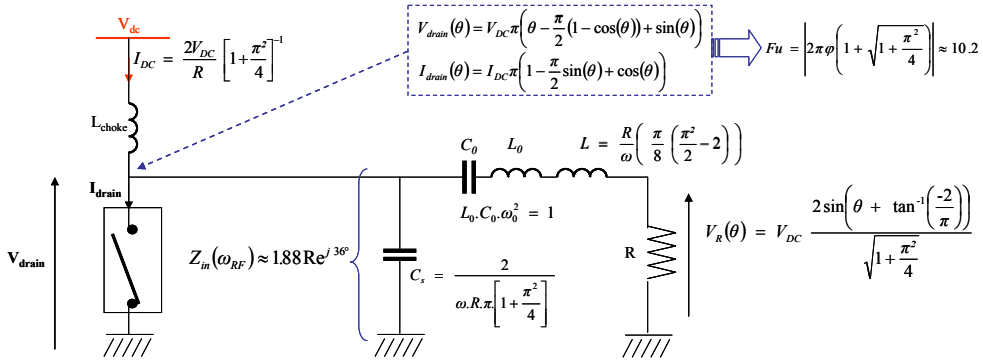



Fig. 15. Class E basic design equations (Sokal &Sokal, 1975); (Kraus and al., 1980).

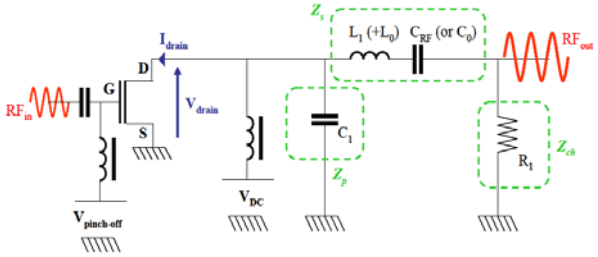
potential use in high efficiency RF polar architecture, showing the keen interest in class E PAs for high data rate applications such as WiMAX.

To conclude the section on PAs for RF applications, one has to consider that the resulting efficiency in a given architecture is the criterion in the choice of a PA class. In the context of WiMAX transmitters, the linearization is mandatory whichever the class. High efficiency (SW) classes are favoured due to their high efficiency but the challenge shifts to the linearity of the amplified signal, expressed by the EVM, the ACPR and whether the emitted signal spectrum respects the standard mask. Class E topology in polar architecture is currently the most popular solution.

### 4.4 Example of a Design for Mobile WiMAX

This sub-chapter presents and explains an example of a Mobile WiMAX transmitter. In the past years many transmitter architectures have been conceived (Masse, 2004); (Liu et al., 2005); (Pozsgay et al., 2008); (Yamazaki et al., 2008). All these transmitters are based on the direct conversion principle that offers a high degree of integration. Therefore, these architectures are preferred for Mobile WiMAX applications. Most of the time they successfully fill specifications of the standard (noise, linearity and emitted power), but have reduced performance in terms of global efficiency. As explained in (Lloyd, 2006), WiMAX PAs are mainly in CW classes. They operate with 7 or 8 dB back-off to guarantee good linearity. The efficiency of such amplifiers used in Mobile WiMAX applications is about 18%

(Eline et al., 2004). To address solutions to this issue, polar architecture associated with a class E amplifier seems to be a good trade-off between linearity and efficiency. The example presented here was extracted from (Robert et al., 2009). This transmitter is designed for WiMAX applications at 3.7 GHz. The first step is to conceive the Class E power amplifier, able to work at such a frequency. Figure 16 shows a serial inductor Class E topology (Sokal &Sokal, 1975). It is possible to calculate the optimum value of the Class E output network elements. In this case, the output of the transistor is supposed to be a short or an open circuit and the quality factor ($Q$) of $C_0$-$L_0$ is set to $Q$=5 for good switching.




$$C_0 = \frac{1}{(Q - 1.15249) \times R_1 \times \omega_0}$$

$$C_1 = \frac{0.1836}{R_1 \times \omega_0} - C_{DS}$$

$$L_1 = \frac{R_1 \times 1.15249}{\omega_0} \qquad L_0 = \frac{1}{C_0 \omega_0^2}$$

Fig. 16. Class E serial inductor topology and corresponding equations (Sokal & Sokal, 1975); (Robert et al., 2009).

The transistor that has been used is a GaAs E-PHEMT, Avago ATF50189. The model introduces a drain to source capacitance $C_{DS}$. $C_1$ has to be calculated considering it absorbs the value of $C_{DS}$. An input matching network is designed in such a way that it should not filter the wide spectrum signal provided by the architecture (see Figure 17), to preserve a constant envelope property. The amplifier offers drain efficiency ($\eta$) of 86.4% and 10.1 dB gain when the input is an 8 dBm 1-tone signal. As the amplifier will work at 3.7 GHz, a WiMAX signal is then injected at the polar architecture input (Suarez et al., 2008). Simulations parameters are fixed as: 10 MHz channel, 841 used subcarriers (1024 FFT size). The raw symbol rate is calculated using 64-QAM modulation to achieve the highest PAPR. Through the architecture, the signal provided to the amplifier is an 8 dBm constant envelope signal. First we observe the filtering behaviour of the amplifier (see Figure 17). The overall gain is 3.5 dB whereas at the section with a 20 MHz bandwidth around the carrier the gain is 8.3 dB. The signal has been amplified more at frequencies close to the carrier frequency than in the rest of the spectrum, due to the Class-E output network.

The amplifier drain efficiency obtained using a high PAPR WiMAX (10 MHz) signal through a polar ΣΔ architecture at 3.7 GHz is 46.2%. The overall efficiency reaches 36.8%. This example stops at the frontier between amplifier and filter. It can be observed in Figure 17 that filtering is still necessary after the amplifier. As the class E output network is dependant on the impedance presented to it (by filter and antenna), a co-design between theses three elements should be done to guaranty best performance (Diet et al., 2008).

This example shows that direct conversion architecture is not the only solution to Mobile WiMAX signal emission. A trade-off between integration, cost and performance, and thus efficiency and consumption needs to be considered.

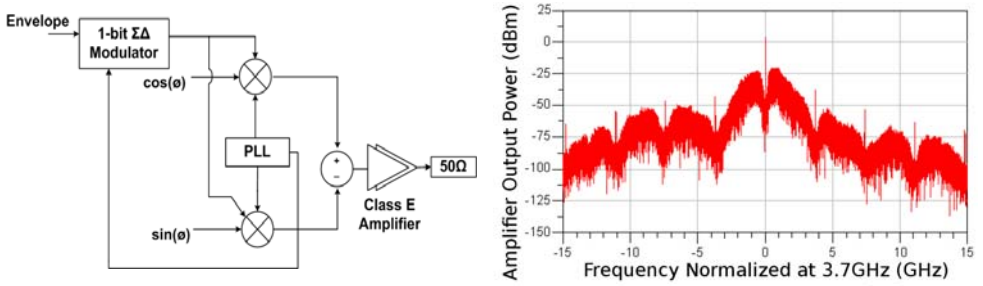



Fig. 17. ΣΔ Polar architecture with class E PA output spectrum.

## 5. Filter Design Aspects and Solutions for Mobile WiMAX Radios

This section is dedicated to the band-pass filters required in WiMAX transmitters at the end of the transmission chain. The filter's requirements are related to the WiMAX RF architecture as was the case with the frequency synthesizer and the power amplifier. As presented in section 2.1, architectures based on decomposition and recombination of the signal offer high efficiency and become good candidates for WiMAX transmitters. In particular, polar architectures for wideband OFDM signals require a band-pass filter after the PA to filter the quantization noise introduced by envelope coders and to guarantee that the output signal respects the spectrum power mask (see Figure 18).

### 5.1 WiMAX Filtering Requirements

Mobile WiMAX profiles (Table 1) are defined in terms of the frequency bands and the TDD mode which is most suitable. The RF filter in a TDD system is not required to attenuate the TX noise as severely as in FDD systems. The TDD mode prevents the TX noise from self jamming the RX since only one is *on* at any time (Eline et al., 2004). Furthermore, TDD mode implies that all the frequencies of the frequency range can be used for transmission at a given time. It is required that one RF band-pass filter be designed for the whole allocated frequency range. Within this range, any channel can be selected (3.5, 5, 7, 8.75 or 10 MHz). According to Table 1 and depending on the frequency range, the bandwidth of the RF filter can be 15 MHz, 100 MHz, 190 MHz or 200 MHz.

The ratio (*f*) between the filter bandwidth and its central frequency serves to compare filter requirements and is also a useful criterion to choose the filter's technology. Table 2 summarizes the bandwidth and the fractional bandwidth (*f*) for each frequency range.

| **Freq. Range GHz** | 2.3-2.4 | 2.305-2.32 / 2.345-2.36 | 2.496-2.69 | 3.3-3.4 / 3.4-3.8 / 3.4-3.6 / 3.6-3.8 |
|---|---|---|---|---|
| **Bandwidth-MHz** | 100 | 15 / 15 | 194 | 100 / - / 200 / 200 |
| *f* - % | 4.26 | 0.64 | 6.96 | 2.99 / - / 5.71 / 5.40 |

Table 2. Bandwidths and fractional bandwidths calculated from Mobile WiMAX profiles

The WiMAX standard establishes the maximum EVM accepted for WiMAX transmitters depending on the modulation and coding rate. The most stringent EVM is -30 dB (3.16%)

corresponding to a 64-QAM (3/4) modulation (IEEE, 2005). The EVM is calculated observing all the imperfections of the transmission chain blocks. Therefore, the maximum in-band ripple and group delay of the filter depend on this EVM value and on the imperfections generated by the other blocks of the architecture.
The filter's out-of-band rejection is determined from the ACLR requirements which are set by national or regional regulatory bodies. This emission mask is often designed as a transmit spectrum mask. ACLR specifications may vary by region, band and channel bandwidth. For the particular case of Europe, the mask proposed by the European Telecommunications Standards Institute (ETSI) for the last WiMAX frequency range (> 3 GHz) is used as reference (ETSI, 2003). Figure 18 presents the masks for a high complexity modulation format (such as 64-QAM) which leads to the most stringent filtering constraints.
Since the filtering is carried out after the power amplification, the transmission filter must offer high power handling capability (high power dynamics and maximum power levels up to 23 dBm (Lloyd, 2006)). The transmission filter for the WiMAX transmitter should also

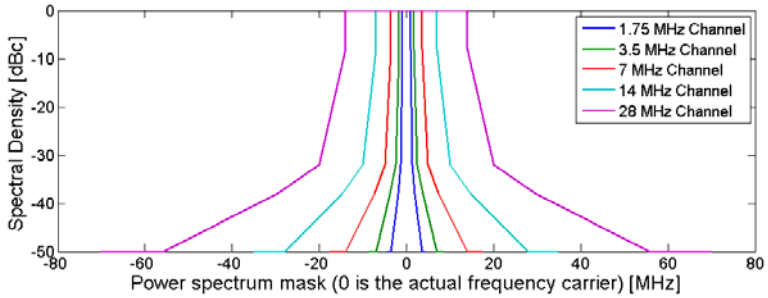


Fig. 18. Power Spectrum mask for a high complexity modulation format (ETSI, 2003).

present low Insertion Losses (IL) to increase the whole architecture power efficiency. Moreover, as size and cost are critical parameters for manufacturers, it is very often required to use a filtering technology that enables integration.

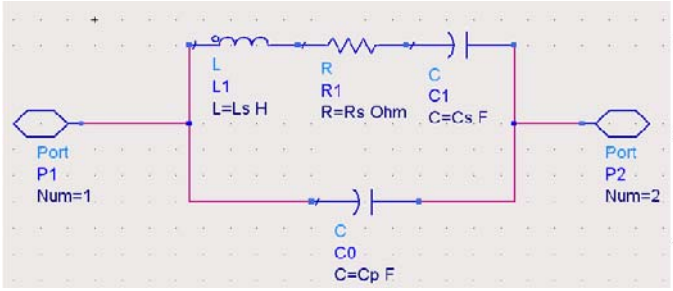
### 5.2 Filtering Technologies

The most notable RF filtering technologies include LC Filters, Ceramic Filters, Surface Acoustic Wave (SAW) Filters, Bulk Acoustic Wave (BAW) Filters and Low Temperature Co-Fired Ceramic (LTCC) Filters. LC filters can support high frequencies and can be integrated as a SoC. However, their main drawback is that they require too much area and can offer only a limited quality factor ($Q$). Ceramic filters offer low IL (about 1.5 - 2.5 dB), high out-of-band rejection (> 35 dB) and low cost. On the other hand the large size of ceramic filters significantly penalizes the integration.
SAW filters are smaller than LC and Ceramic filters, but have limitations in the frequency domain (up to 3 GHz). The power limitation of 1 W is acceptable for Mobile WiMAX devices. Typical IL varies between 2.5 and 3 dB and out-of-band rejection can reach up to 30 dB. The main drawback is that SAW filters cannot be IC integrated.
LTCC technology offers integration of high $Q$ passive components along with low IL, high maximal operation frequency and acceptable out-of-band rejection. LTCC filters are smaller than LC and ceramic filters and can be integrated as SOP.

BAW filters use Film Bulk Acoustic Resonators (FBAR) that are characterized by a high quality factor *Q*. Moreover, they have low IL (1.5 – 3 dB), significant out-of-band rejection (≈ 40 dB) and high maximal operation frequency (up to 15 GHz). BAW filters can also deal with high output power (3 W). They are CMOS compatible and can be integrated "above IC", which results in size reduction and package simplicity (Lakin, 2004). Besides, by employing wafer scale manufacturing using IC processing, BAW techniques offer high potential for low cost manufacturing.

### 5.3 Recent WiMAX Filters: BAW filters

BAW technology using Aluminium Nitride (AlN) piezoelectric material allows frequency operation up to 12 GHz. Each BAW filter is implemented with different sizes of FBARs, which are connected in Ladder or Lattice topologies (Lakin, 2004). The frequency response of this particular type of resonator depends on the thickness of a thin piezoelectric film. Ladder filters are the most straight forward to implement because resonators may be individually optimized for reactance and central frequency during the design and




$$C_p = \frac{\varepsilon_0 \varepsilon_r A}{d}; \quad C_s = \frac{8k_t^2 C_p}{\pi^2}$$

$$L_s = \frac{\pi^3 V_a}{8 w_s^3 \varepsilon_0 \varepsilon_r A k_t^2}; \quad R_s = \frac{L_s w_s}{Q}$$

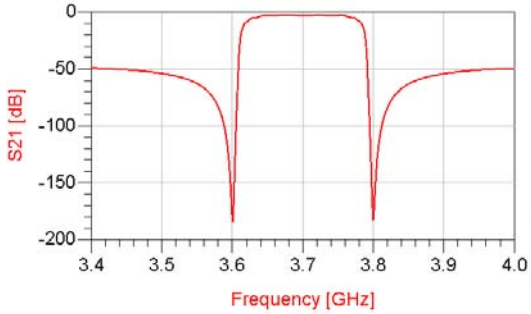
$$f_s = \frac{V_a}{2d_1}; \quad f_p = f_s \sqrt{1 + \frac{8k_t^2}{\pi^2}}$$

Fig. 19. FBAR resonator – BVD model.

manufacturing process. Ladder filters are single ended, while lattice filters are double ended. The filter output is at the antenna input and therefore, the ladder topology is preferred to the lattice topology due to signal unbalance effects and the ease of tuning. Filters are preceded by unbalanced output of the HPA as well. The Butterworth Van Dyke (BVD) model is an electric circuit model that characterises FBAR resonators. The BVD equivalent circuit of the crystal resonator is shown in Figure 19.

The resonator is in the form of a simple capacitor, having a piezoelectric material as the dielectric layer and suitable top and bottom metal electrodes. The simplified equivalent circuit of the piezoelectric resonator has two arms. $C_p$ is the geometric capacitance of the structure. The $R_s$, $L_s$, $C_s$ branch of the circuit is called "motional arm," which arises from mechanical vibrations of the crystal. The series elements $R_s$, $L_s$, $C_s$ are controlled by the acoustic properties of the device and they model the motional loss, the inertia and the elasticity respectively. These parameters can be calculated from equations presented in Figure 19. $\varepsilon_r$ is the material's relative permittivity (10.59 for the AlN), $k_t^2$ is the electromechanical coupling constant (6% for the AlN), $V_a$ is the acoustic material velocity (10937 for the AlN), $A$ is the surface area of the electrodes, $d$ is the thickness of the piezoelectric material, and $Q$ is the quality factor. $w_s$ and $w_p$ correspond to $2\pi$ multiple of the resonance ($f_s$) and anti-resonance ($f_p$) frequencies of the resonator.

Thickness of series and shunt resonators can be different; $d_1$ and $d_2$ refer to the thickness of the series and the shunt resonator respectively. Only two parameters need to be optimized in order to design a band-pass filter. These parameters are the area $A$ (expressed as $l \times l$) and




| Stage | $L_{serie}$ [µm] | $L_{shunt}$ [µm] |
|---|---|---|
| 1 | 122 | 87 |
| 2 | 102 | 98 |
| 3 | 113 | 107 |
| 4 | 124 | 117 |
| 5 | 136 | 128 |
| 6 | 150 | 142 |
| 7 | 168 | 132 |

| | |
|---|---|
| **Insertion Loses** | 3.06 dB |
| **Rejection** | 49.03 dB |

Fig. 20. [$S_{21}$] parameter of a WiMAX RF filter using BAW technology (7th order ladder).

the resonator thickness ($d_1$ and $d_2$). The series and the shunt resonator form a stage. In order to achieve the required frequency response and out-of-band rejection, particular stages are put together to build cascades. Each additional stage (two series-shunt resonators) increases the filter order by one. Therefore, a six resonator ladder filter is a third order filter.
An example of a WiMAX filter in the 3.6 – 3.8 GHz frequency band was proposed in (Suarez et al., 2008). The emission filter in this case has a bandwidth of 200 MHz. Figure 18 shows that the out-of-band rejection must be 50 dB at twice the channel bandwidth (20 MHz from the edge for a 10 MHz channel).
Resonator manufacturing constraints impose the thickness to be between 1 µm and 6.5 µm and surface areas between 10 µm * 10 µm and 600 µm * 600 µm. Thickness of the series and shunt resonators is 2.79 µm and 2.88 µm respectively. Figure 20 presents the frequency response of the WiMAX filter along with dimensions, IS and out-of-band rejection.

### 5.4 Recent WiMAX Filters: Other Examples.

Some other examples of band-pass filters for WiMAX using LTCC technologies have been published recently. For example, (Kim et al., 2008) proposes a Quad band module for Wi-Fi/WiMAX applications. This module provides insertion loss of less than 4.6 dB at 3.5 GHz and the attenuation is more than 26 dB. A LTCC filter for the same frequency band has been proposed by (Accute). It provides bandwidth of approximately 200 MHz, 2.7 dB IL, ripple of less than 1 dB and more than 30 dB of attenuation.
A single band WiMAX SiP solution based on the LTCC technology has been recently proposed by (Heyen et al., 2008). It comprises the complete passive and active RF front-end plus the transceiver RFIC for up and down conversion to base band. The RF transmission filter provides harmonic and spurious rejection better than 40 dB.

## 6. Other Design Aspects

In previous sections the performance of the baseband and radiofrequency parts of a WiMAX transmitter were discussed. These are the parts of the transmitter between the digital

electronics and the antenna. The importance of the choice of architecture has been demonstrated, as have the impacts of key elements such as frequency synthesizers, power amplifiers and emission filters. This section points out the considerations to add for a global design of such a transceiver regarding the performance of Digital to Analogue Converters (DACs) and antennas at these frequencies.

### 6.1 Digital and Analogue

The DAC enables baseband signal generation after the shaping filter. It should have low distortion, sufficient bandwidth and low consumption. DACs are used in conventional architecture for I and Q paths generation and in polar architecture for phase (I and Q) and envelope paths. In polar architecture there is one DAC more and the required bandwidth is extended due to the non-linear processing when generating the "phase" and the "magnitude/envelope" of the signal. Also, the coding of the envelope is an additional restriction in terms of speed for the ΣΔ. As the Signal to Noise Ratio (SNR) of the signal is admitted to grow with the number of bits and bandwidth, these specifications are mandatory limiting factors. Nowadays, some converters work in the range of several bits near a GHz and around 12 to 20 bits near a MHz.

Due to the conclusions of previous sections, the example presented here is the simulation of a polar architecture for an OFDM signal with 64 sub-carriers (typically IEEE802.11a). The symbol rate is 20 MHz and the carrier frequency is 5.2 GHz but can be shifted to 3.7 GHz without altering the observations because the DAC influences are introduced on the baseband processing. Figure 21 presents the emitted spectrum in an ideal polar/EER transmitter simulation with limitation of the bandwidth on the envelope and phase signals. Limits are three times the symbol rate for the envelope and seven times for the phase ones. The mask of the IEEE 802.11a standard is added on the same figure and it is noticeable that the emitted spectrum is not far from the limit. The most limiting parameters are the phase signals.

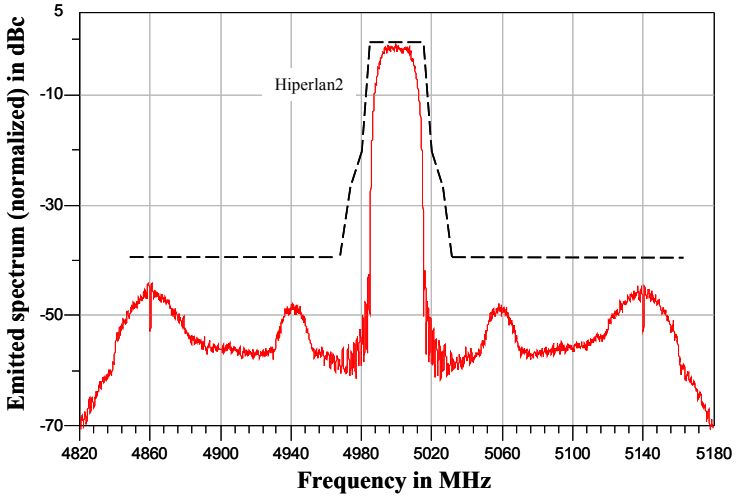



Fig. 21. Emitted spectrum of a 20 MHz OFDM Hiperlan 2 signal with band width limitations of 60 MHz (envelope signal) and 140 MHz (I and Q phase signals). The EVM rms is 0.2%.

This implies a high bandwidth for the baseband signals generation. The resolution will therefore be strongly impacted because the higher the bandwidth, the lower the resolution (without consideration of power consumption). The second step of our example is to limit the number of bits for the signal representation. Here the envelope is coded in either signed or unsigned format (depending on the specification/complexity of the hardware part of the system) and without a clipping that could have reduced the needed dynamic for the envelope but at the cost of an EVM increase. Results in the classical architecture case and polar one are presented on Figure 22.

| EVM / EVM max (% rms) | 7 bits | 6 bits | 5 bits | 4 bits | 6 bits (envelope unsigned) | 8 bits (2'complement) |
|---|---|---|---|---|---|---|
| I and Q paths for OFDM sig. | 0.3 / 0.5 | 0.6 / 0.9 | 1.2 / 1.8 | 2.8 / 5 | | |
| I and Q paths OFDM phase sig. | 0.1 / 0.2 | 0.4 / 0.7 | 0.7 / 1.3 | 1.6 / 2.5 | 0.9 / 1.8 | 0.4 / 1.1 |
| Envelope default is unsigned format | 0.2 / 0.3 | 0.4 / 0.5 | 0.6 / 0.9 | 1.3 / 2.1 | | |

Fig. 22. Results of resolution limitation for an OFDM Hiperlan 2 transmitter.

The limitation of the resolution with an acceptable EVM of 0.5% rms (without any other architecture imperfection) is at the edge of the actual DACs performance, which is 8 bits with a supposed bandwidth of tens of MHz. To realistically illustrate the influence of both parameters introduced in the simulation, we show in Figure 23 the simulation of the polar/EER architecture with DACs of 8 bits resolution and with the same bandwidth limitations as shown in Figure 21. The emitted spectrum is compared with the same mask and the constellation and EVM are presented. The results show an acceptable EVM below 0.5% rms and the spectrum is, in conclusion, the main criterion for characterizing the DAC impact in the architecture.

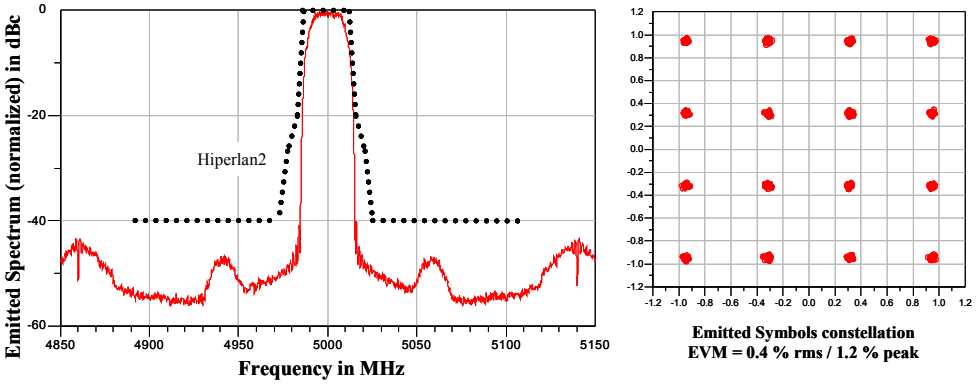



Fig. 23. Emitted spectrum and constellation for an OFDM Hiperlan 2 transmitter with bandwidth limitations of 60 MHz (envelope signal) and 140 MHz (I and Q phase signals), and an 8 bit DAC.

### 6.2 Antennas

Antennas for handsets have to be adapted to the difficult environment of indoor mobility (omni-directivity or wide radiation lobe, polarization) while maintaining a small size and cost. Solutions are, for example, helicoidal antennas, patch or planar antennas with tuned slot; often with a ground reflector in the case of mobile phone application to avoid radiations toward the user and coupling to the circuit (in this case the ground plane is a kind of "shield"). The use of antenna diversity or Multiple Input Multiple Output (MIMO) benefits the receiver and significantly increases its performance, but this is a challenge in terms of power consumption for a battery operated system (additional RF sub-systems). In the case of the integration of multiple wireless systems, it is important to focus on antenna integration and especially multi-band or wideband antennas. Whatever the standards considered, diversity of antennas and antennas for multiple standards are research topics for systems offering mobile communications and connectivity (such as WiMAX). In conclusion, integrated low cost antennas are to be investigated for this type of system with regards to the standards specifications (bandwidths, propagation environment) and with architectural considerations (size, cost, consumption in the case of MIMO).

## 7. Conclusion

As a result of flexible and multi-band radio operation, the Mobile WiMAX standard presents a challenge for every stage of the RF front-end. Promising techniques and mechanisms for linear and high efficient transmission have been discussed, along with their advantages and limitations. The ultimate goals are high degree of RF integration into cheap CMOS technology and high power efficiency along with linearity. At this point, the polar based architecture seems to offer high performance solutions for high PAPR wideband signals, while providing high efficiency due to switched mode amplification.
It has been shown that the RF filtering, which is required after the power amplifier presents a significant challenge for RF designers. Appropriate filtering technologies have been presented, including current examples of WiMAX filters.
Moreover, signal deterioration resulting from the frequency synthesizer's phase noise contribution has been discussed as well, along with solutions for low noise high speed synthesis.

## 8. Acknowledgement

The research has received funding from the European Community's Seventh Framework Programme under grant agreement no. 230126 and partially by the Czech science foundation projects 102/09/0776, 102/08/H027, 102/07/1295 and research programme MSM 0021630513.

## 9. References

Accute Microwave. Specification of LTCC Filter - LF43B3500P34-N42.

Armada, G. A. (2001). Understanding the effects of phase noise in orthogonal frequency division multiplexing. IEEE Trans. Broadcast., Vol. 47, No. 2, pp. 153-159, June 2001.

Baudoin, G.; Bercher, JF.; Berland, C.; Brossier, JM.; Courivaud, D.; Gresset, N.; Jardin, P.; Bazin-Lissorgues, G.; Ripoll, C.; Venard, O.; Villegas. M. (2007). Radiocommunications Numériques : Principes, Modélisation et Simulation. Dunod, EEA/Electronique, 672 pages, 2ème édition 2007.

Choi, J.; Yim, J.; Yang, J.; Kim, J.; Cha, J.; Kang, D.; Kim, D. ; Kim, B. (2007). A ΣΔ digitized polar RF transmitter. IEEE Trans. on Microwave Theory and Techniques, Vol. 52, No. 12, 2007, pp 2679-2690.

Cimini, L. J. (1985). Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. IEEE Trans. Commun., Vol. 33, No. 7 (July 1985), pp. 665–675.

Cox, D. C. (1974). Linear amplification with non-linear components, LINC method. IEEE transactions on Communications, Vol. COM-23, pp 1942-1945, December 1974.

Crowley et al. (1979). Phase locked loop with variable gain and bandwidth. U.S. Patent 4,156,855, May 29, 1979.

Diet, A.; Berland, C.; Villegas, M.; Baudoin, G. (2004). EER architecture specifications for OFDM transmitter using a class E power amplifier. IEEE Microwave and Wireless Components Letters (MTT-S), Vol.-14 I-8, August 2004, pp 389-391, ISSN 1531-1309.

Diet, A.; Robert, F.; Suárez, M.; Valenta, V.; Andia Montes, L.; Ripoll, C.; Villegas, M.; Baudoin. (2008) G. Flexibility of class E HPA for cognitive radio, Proceedings of IEEE 19th symposium on Personal Indoor and Mobile Radio Communications, PIMRC 2008, 15-18 September, Cannes, France. CD-ROM ISBN 978-1-4244-2644-7.

Diet, A.; Villegas, M.; Baudoin, G. (2008). EER-LINC RF transmitter architecture for high PAPR signals using switched Power Amplifiers. Physical Communication, ELSEVIER, ISSN: 1874-4907, V-1 I-4, December 2008, pp. 248-254.

Eline, R.; Franca-Neto, L.M.; Bisla, B. (2004). RF System and circuit challenges for WiMAX. Intel Technology Journal, Vol. 08, Issue 03 2004

ETSI. (2003). European Standard, Telecommunications Series, ETSI 301021 V1.6.1, 2003.

Grebennikov, A. (2002). Class E high efficiency PAs : Historical aspect and future prospect. Applied Microwave and Wireless, July 2002, pp 64-71.

Herzel, F.; Piz, M. and Grass, E. (2005). Frequency synthesis for 60 GHz OFDM systems, Proceedings of the 10th International OFDM Workshop (InOWo'05), Hamburg, Germany, pp. 303–307, 2005.

Heyen, J.; Yatsenko, A.; Nalezinski, M.; Sevskiy, G.; Heide, P. (2008). WiMAX System-in-package solutions based on LTCC Technology, Proceedings of COMCAS 2008.

IEEE Standard 802.16e. (2005). Air interface for fixed and mobile broadband wireless access systems amendment 2: physical and medium access control layers for combined fixed and mobile operation in licensed bands, 2005.

Kahn, L. R. (1952). Single sideband transmission by envelope elimination and restoration, Proceedings of the I.R.E., 1952, pp. 803-806.

Keliu, S. and Sanchez-Sinencio, E. (2005). CMOS PLL Synthesizers: Analysis and Design, Springer, 0-387-23668-6, Boston.

Kim, D.; Dong Ho Kim; Jong In Ryu; Jun Chul Kim; Chong Dae Park; Chul Soo Kim; In Sang Song. (2008). A quad-band front-end module for Wi-Fi and WiMAX applications using FBAR and LTCC Technologies, Proceedings of APMC 2008.

Krauss, H. C.; Bostian, C. W. and Raab, F. H. (1980). Solid State Radio Engineering, Wiley, 047103018X, New York.

Kyoungho W.; Yong L.; Eunsoo N.; Donhee, H. (2008). Fast-lock hybrid PLL combining fractional-N and integer-N modes of differing bandwidths. IEEE Journal of solid state circuits, Vol. 43, No. 2, pp. 379-389, Feb. 2008

Lakin, K. (2004). Thin film BAW filters for wide bandwidth and high performance applications, IEEE MTT-S 2004.

LIM, D.-W, et al. (2005). A new SLM OFDM Scheme With Low Complexity for PAPR Reduction. IEEE Signal Processing Letters, Vol. 12, No. 2, February 2005, pp. 93-96.

Liu, H.; Chin, H.; Chen, T.; Wang, S.S. Lu. (2005). A CMOS transmitter front-end with digital power control for WiMAX 802.16e applications, Microwave Conference Proceedings, APMC 2005. Asia-Pacific Conference Proceedings, Vol. 3.

Lloyd, S. (2006). Challenges of mobile WiMAX RF transceivers. Solid-State and Integrated Circuit Technology, 2006. ICSICT '06. 23-26 Oct. 2006.

Masse, C. (2006). A 2.4 GHz direct conversion transmitter for WiMAX applications, Radio Frequency Integrated Circuits Symposium, 11-13 June 2006.

Mäuller, H. S. & Huber, J.B. (1997). A novel peak power reduction scheme for OFDM, Proc. of the Int. Symposium on Personal, Indoor and Mobile Radio Communications PIMRC'97, Sept. 1997, Helsinki, Findland, pp. 1090-1094.

Memmler, B.; Gotz, E.; Schonleber, G. (2000). New fast-lock PLL for mobile GSM GPRS applications, Solid-State Circuits Conference, ESSCIRC 2000.

Muschallik, C. (1995). Influence of RF oscillators on an OFDM signal. IEEE Trans. Consumer Electronics, Vol. 41, No. 7, pp. 592–603, Aug 1995.

Nielsen, M.; Larsen, T. (2007). Transmitter architecture based on ΔΣ modulation and switch-mode power amplification. IEEE Trans. on Circuits and Systems II, 2007, Vol. 54, No. 8, pp. 735-739.

Pozsgay, A.; Zounes, T.; Hossain, R.; Boulemnakher, M.; Knopik, V.; Grange, S.; A fully digital 65nm CMOS transmitter for the 2.4-to-2.7 GHz WiFi/WiMAX bands using 5.4 GHz ΔΣ RF DACs, Proceedings of ISSCC 2008, pp: 360-619.

Raab, F. et al. (2003). RF and microwave PA and transmitter technologies. High Frequency Electronics, May-November 2003, pp 22-49.

Robert, F.; Suarez, M.; Baudoin, G.; Villegas, M.; Diet, A. (2009). Analyse de l'influence du codage d'enveloppe sur les performances de l'amplificateur classe E d'une architecture polaire, XVI Journées Nationales Micro-ondes, JNM, mai 2009, Grenoble, France.

Robert, F.; Suarez, M.; Diet, A.; Villegas, M.; Baudoin, G. (2009). Study of a polar sigma-delta transmitter associated to a high efficiency switched mode power amplifier for mobile WiMAX, Proceedings of IEEE WAMICON 2009, 20-21 Apr.2009

Sokal, N. and Sokal, A. (1975). Class E, a new class of high efficiency tuned single ended switching PAs. IEEE journal of Solid State Circuits, Vol. 10, No. 3, Juin 1975, pp 168-176.

Suarez, M.; Villegas, M.; Baudoin, G. (2008). Front end filtering requirements on a mobile cognitive multi-radio transmitter, Proceedings of the 11th International Symposium on wireless Personal Multimedia Communications, 8-11 Sept. 2008, Saariselka, Finlande.

Suarez, M.; Valenta, V.; Baudoin, G.; Villegas, M. (2008). Study of a modified polar sigma-delta transmitter architecture for multi-radio applications, Proceedings of EuMW, European Microwave Week, 27-31 Oct. 2008, Amsterdam, Netherlands.

Tellado, J. (2000). Multicarrier Modulation with Low PAR, Kluwer Academic Publishers, 2000

Valenta, V.; Villegas, M.; Baudoin, G. (2008). Analysis of a PLL based frequency synthesizer using switched loop bandwidth for mobile WiMAX, Proceedings of the 18th International Conference Radioelektronika 2008, pp. 127-130. ISBN: 978-1-4244-2087-2.

Valenta, V.; Marsalek R.; Villegas, M.; Baudoin, G. (2009). Dual mode hybrid PLL based frequency synthesizer for cognitive multi-radio applications, to appear in WPMC'09.

Villegas, M.; Berland, C. ; Courivaud, D. ; Bazin-Lissorgues, G. ; Picon, O. ; Ripoll, C. ; Baudoin, G. (2007). Radiocommunications Numériques : Conception de circuits intégrés RF et micro-ondes. Dunod, EEA/Electronique, 464 pages, 2ème édition 2007.

Yamazaki, D.; Kobayashi, N.; Oishi, K.; Kudo, M.; Arai, T.; Hasegawa, N.; Kobayashi, K. (2008). 2.5-GHz fully-integrated WiMAX transceiver IC for a compact, low-power consumption RF module, Radio Frequency Integrated Circuits Symposium, RFIC 2008, June 17 2008-April 17 2008.

Qiyue Zou, Tarighat, A. and Sayed, A.H. (2007). Compensation of phase noise in OFDM wireless systems. IEEE Trans. Signal Processing, Vol. 55, No. 11, pp. 5407-5424, Nov 2007.

WiMAX Forum™ Mobile System Profile 3 Release 1.0 Approved Specification 4 (Revision 1.7.1: 2008-11-07).

4

# The Application of μ-Law Companding to Mobile WiMax

Brian G Stewart[1] and Athinarayanan Vallavaraj[2]
[1]*Glasgow Caledonian University*
*Scotland, UK*
[2]*Caledonian College of Engineering*
*Sultanate of Oman*

## 1. Introduction

The IEEE802.16e mobile WiMax standard employs Orthogonal Frequency Division Multiplexing (OFDM) principles in the transmission of data (IEEE802.16e, 2005). Within multicarrier systems, like WiMax and other OFDM technologies, a major problem relates to issues associated with instantaneous values of the peak transmission output power. At some instant in time, the subcarriers of an OFDM signal may add coherently producing a very high peak power that may reach a maximum value of the number of subcarriers times the average power. The peak power can be expressed in relation to the average power, referred to as the Peak-to-Average Power Ratio (PAPR), which is defined as the ratio of the peak of the instantaneous envelope power to the average power of the OFDM signal. One of the main drawbacks of WiMax systems is the high value of PAPR often encountered, typically around levels of 12dB to 13dB or even higher (e.g. Lloyd, 2006). A high PAPR necessitates that the A/D and D/A converters used in the communication system have a higher level of bit conversion to accommodate the peaks. In addition it requires the OFDM power amplifiers to remain linear over an extended region above the average power value to include the peak amplitudes. Also, if there are any regulatory or application constraints on the extent of peak power, a high PAPR would require the average power of the signal to be reduced, thus reducing the range of transmission of OFDM signals (Han & Lee, 2005). The nonlinearity of any power amplifiers also introduces in-band and out-of-band radiation or spectral splatter, increasing the Bit-Error-Rate (BER) and causing interference with neighbouring frequency channels.

A variety of techniques have been published in the literature which attempt to reduce the PAPR in OFDM signals. These techniques can be classified into three broad categories as, signal pre-distortion techniques, coding techniques and scrambling techniques (Van Nee & Prasad, 2000). There are also techniques that combine either two or more of these techniques in order to improve the PAPR reduction. Though many solutions have been proposed to deal with the high value of PAPR existing in random data transmission within OFDM systems, one method of PAPR reduction which has received little critical attention in this area is the application of companding. In an attempt to address this weakness, this chapter

presents a thorough investigation of the performance of μ-Law companding to mobile WiMax and in particular to the Down Link Partially Used Subcarrier (DL PUSC) mode of operation. Parameters investigated and quantified as a function of various μ-Law companding profiles include the Power Spectral Density (PSD), BER, PAPR reduction, and the influence of mobility on performance when WiMax multipath mobile channels are considered. Many of these results are new and have never been investigated for companding or specifically evaluated in relation to WiMax architectures. One further aspect presented in this chapter, which is often neglected in the literature, is the comparison and evaluation of companding in regard to equalised symbol power for all companding situations. It is well known that companding naturally increases the average power of OFDM symbol transmissions. However, equalised symbol power transmission performance requires to be quantified fully to allow a complete understanding of the limitations of companding within WiMax systems. Results will show that companding does have potential for application to mobile WiMax, but there are limitations in relation to PSD, BER, PAPR reduction and mobility, and these will be discussed within the relevant sections.
The structure of the chapter is as follows. Section 2 introduces the concepts and definitions associated with PAPR. Section 3 briefly discusses the general techniques which are currently employed to reduce the PAPR of OFDM data symbols; Section 4 introduces the principles associated with companding and in particular μ-Law companding; Section 5 presents details of the mobile WiMax physical layer model used for the simulations and investigations; Section 6 discusses the issues of PSD related to WiMax companding; Section 7 investigates the BER performance; Section 8 presents the PAPR improvements and Section 9 investigates the influence of mobility for companded WiMax within two common multipath channels. Section 10 is a conclusions section and summarises the main points from the chapter.

## 2. The PAPR of an OFDM Signal

The instantaneous amplitude of a baseband OFDM signal can be written as

$$x(t) = \sum_{n=0}^{N-1} X_n \exp\, j(\Phi_n + 2\pi nt / T) \tag{1}$$

where $X_n$ exp j(Φ) is the complex baseband modulated symbol, and $N$ is the number of subcarriers. The instantaneous envelope power of an OFDM signal, assuming a unity impedance load, is evaluated through

$$P(t) = \left|x(t)\right|^2 = \sum_{n=0}^{N-1} {X_n}^2 + 2\sum_{n=0}^{N-2}\sum_{m=n+1}^{N-1} X_n X_m \cos \Psi_{nm} \tag{2}$$

where $\Psi_{nm} = (\Phi_n - \Phi_m + 2\pi(n-m)t / N)$. The average envelope power is calculated through

$$P_{avg} = E\left\{\left|x(t)\right|^2\right\} \tag{3}$$

where $E\{.\}$ is defined as the expectation value. Using equation (1), the expression for the average power becomes

$$P_{avg} = E\left\{\left|x(t)\right|^2\right\} = \sum_{n=0}^{N-1}\left|X_n\right|^2 + \sum_{n=0}^{N-2}\sum_{m=n+1}^{N-1} E(X_n X_m^*)\exp j\left(2\pi(n-m)t/T\right) \tag{4}$$

The symbols on different subcarriers within OFDM may be assumed to be independent, and hence, $E(X_nX_m^*) = E(X_n)E(X_m^*)$. Since the signals are orthogonal, then the second term in (4) is zero thus the average power reduces to

$$P_{avg} = E\left\{\left|x(t)\right|^2\right\} = \sum_{n=0}^{N-1} {X_n}^2 \tag{5}$$

Since PAPR is defined as the ratio of the maximum (peak) instantaneous envelope power to the average power, then the PAPR may be expressed as

$$PAPR = \max\left\{\frac{P(t)}{P_{avg}}\right\} = \max\left\{\frac{\left|x(t)\right|^2}{E\left\{\left|x(t)\right|^2\right\}}\right\} \tag{6}$$

Substituting (2) and (5) into (6) results in the general formula for the PAPR of a general MQAM OFDM transmission, i.e.

$$PAPR = \max\left\{1 + \frac{2}{\sum_{n=0}^{N-1} {X_n}^2}\sum_{n=0}^{N-2}\sum_{m=n+1}^{N-1} X_n X_m \cos\Psi_{nm}\right\} \tag{7}$$

To help appreciate PAPR, consider MPSK modulation where the amplitudes of all the baseband signals are equal. In this situation equation (7) reduces to

$$PAPR = \max\left\{1 + \frac{2}{N}\sum_{n=0}^{N-2}\sum_{m=n+1}^{N-1}\cos(\Phi_n - \Phi_m + 2\pi(n-m)t/T)\right\} \tag{8}$$

If the data symbols are presumed to be identical on all subcarriers, then when $N$ subcarriers are added together with the same phase, they sum up coherently and produce a peak power that is $N$ times the average power. Figure 1 illustrates the ratio of the instantaneous envelope power to the average power of a single OFDM symbol transmission of period $T$ which comprises 16 QPSK subcarriers all carrying the same data. For this situation, the output from the IFFT produces a single peak at the first and last of the 16 time sampled points of the symbol with zero at all other time samples. The maximum value of the envelope power to the average power (i.e. the PAPR) is 16 (=12.04dB), indicating that the peak power is 16 times greater than the average power. In most cases the PAPR situation to be addressed relates to random data and methods used to reduce PAPR in these situations are briefly discussed in the next section.

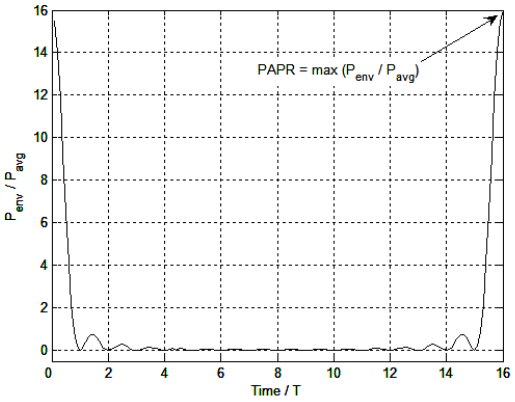



Fig. 1. The normalised instantaneous power transmission for a 16-subcarrier QPSK OFDM symbol when the data on each subcarrier is identical

## 3. Reducing the PAPR of OFDM Signals

### 3.1 Methods of PAPR Reduction

A number of PAPR reduction techniques that attempt to reduce the maximum PAPR of random data within OFDM signals exist (see for example the reviews by Han & Lee, 2005, and Wang & Tellambura, 2006). The most popular of these are signal pre-distortion techniques such as clipping, peak windowing and peak cancellation which aim to reduce the peak amplitudes of the transmitted signals by non-linearly distorting the OFDM signal at or around the peak values (e.g. O'Neill & Lopes, 1995; De Wild, 1997; Li & Cimini, 1997; Pauli & Kuchenbecker, 1997; May & Rohling, 1998; Van Nee & De Wild, 1998; Van Nee & Prasad, 2000; Armstrong, 2001, 2002; Wang & Tellambura, 2005)

A second category of PAPR reduction techniques relates to probabilistic and scrambling methods comprising phase modification techniques, amplitude modification techniques and scrambling and interleaving techniques. These are becoming perhaps the most popular methods of reducing the PAPR in data transmissions within OFDM systems. All these techniques modify the phase, amplitude or subcarrier position of input symbols, thus creating several OFDM signals representing the same information. The OFDM signal with the lowest PAPR is then selected for transmission (e.g. Boyd, 1986; Bäuml et al. 1996; Van Eetvelt et al., 1996; Müller & Huber, 1997a, 1997b; Cimini & Sollenberger, 2000; Hill et al. 2000; Jayalath & Tellambura, 2000; Breiling et al., 2001; Tellambura & Jayalath, 2001; Han & Lee, 2005; Wang & Tellambura, 2006). In most cases extra overhead or side band information is also required to be sent to allow recovery of the original information at the receiver. Perhaps the best known and most popular of these techniques are called Selected Mapping (SLM) and Partial Transmit Sequences (PTS).

A third category of PAPR reduction methods relates to coding techniques. Block and channel coding, or specialised codewords with particular and special autocorrelation properties are employed in an attempt to reduce the PAPR. One of the additional advantages of these techniques is that improved BER as well as reduced PAPR can ensue though at the cost of increased redundancy (e.g. Golay, 1961; Jones et al., 1994; Jones &

Wilkinson, 1995, 1996; Davis & Jedwab, 1999; Paterson & Tarokh, 200; Tarokh & Jafarkhani, 2000; Breiling et al., 2001; Yang & Chang, 2003, Han & Lee, 2005; Kang, 2006). Often though, significantly reduced PAPR cannot always be guaranteed for all symbol transmissions using these techniques. However, developments and refinements of these techniques are constantly being investigated and reported.

### 3.2 Choice of PAPR Reduction Techniques

Pre-distortion techniques like clipping and filtering are the simplest to implement and do not require any side information to be transmitted, however they result in a distorted signal which produces in-band and out-of-band signal splatter. Peak cancellation, however, does not result in any frequency signal splatter. Scrambling and probabilistic techniques, such as SLM and PTS, are distortionless methods. The complexity however of these techniques is increased in that the number of IFFT operations increases in proportion to the number of scrambled sequences used to produce a reduced PAPR. In addition, these techniques, in general, need side information and as a result the data rate is decreased. There may also be a small compromise on the PAPR due to the transmission of this side information. Coding techniques increase the complexity of the PAPR reduction solution with an additional requirement of encoding and decoding at the transmitter and receiver. As encoding increases the number of bits in the transmitted signal, the data rate is therefore reduced. There is no distortion or signal splatter as in clipping, and encoding can also serve the dual purpose of BER reduction and PAPR reduction.

Clearly there are a variety of PAPR reduction techniques available with each one claiming to have some advantages over the other. The choice of a particular technique depends on a number of factors, for example, PAPR reduction capability required, PSD distortion, acceptable BER at the receiver, signal power requirements, data rate employed, implementation complexity, consideration of the effect of the components in the transmitter, etc. Han & Lee (2005) have outlined a brief description of these criteria. The quest for inventing new PAPR reduction techniques has not come to an end. With the increasing use of OFDM in mobile broadband applications, the necessity for PAPR reduction has gained critical importance since an increased PAPR means an increased envelope power and thus a reduction in battery standby and battery life time.

One other method of reducing PAPR is called companding. This method falls best under the category of pre-distortion technique. A limited number of publications exist on companding. These publications indicate that companding may have potential in reducing PAPR, but this potential has still to be fully explored and quantified for OFDM type systems. In this regard, an evaluation of companding for Mobile WiMax forms the main thrust of this chapter. The method of μ-Law companding will be introduced in the next section.

## 4. Companding of OFDM Signals

Companding is fundamentally the process of compressing amplitude signals at a transmitter and expanding them at a receiver. A number of authors have advocated the use of companding techniques to OFDM systems to improve the PAPR. Wang et al. (1999) introduced companding as a potential PAPR reduction technique and provided the transmitted waveforms of 16QAM based 256-subcarrier OFDM signals before and after

companding. The symbol-error-rate (SER) was also shown to vary with the companding coefficients. However, no quantified results in terms of precise PAPR reduction or SER improvement as a function of companding parameters were detailed. Huang et al. (2001) demonstrated that a non-linear-quasi-symmetrical μ-Law companding transform can outperform a clipping-filtering scheme by 4.6 dB in relation to SNR for a BER of $10^{-4}$ in an additive white Gaussian noise channel, and a PAPR reduction of 4.1 dB could be achieved for QPSK based 128 subcarrier OFDM signals. Companding profiles considered have included traditional μ-Law and A-Law, as well as exponential type forms (e.g. Jiang & Song, 2005). Companding of OFDM has also normally been restricted to situations where no pilots are included, one type of modulation is employed, and smaller numbers of subcarriers are considered (e.g. Vallavaraj et al., 2004). The literature therefore demonstrates that companding may be considered to have some validity in relation to possible PAPR reduction. However, one of the main drawbacks of companding is that as a consequence of the non-linear companding profile, PSD distortion occurs resulting in frequency splatter where residual frequency power is "splattered" out with the transmission bandwidth causing inter channel interference. Spectral re-growth also occurs within the OFDM channel bandwidth as a consequence of increased power arising from the companding process. This increased power is also considered to provide an advantage of improved BER due to the effective increased SNR naturally arising from the direct application of companding (Mattsson et al., 1999). However, the question of how spectral re-growth and distortion effects precisely influence the quantification of the performance of an OFDM system has still to be fully considered for OFDM architectures. These issues will be explored in more detail for companded WiMax in the following sections.

### 4.1 The Concepts of Companding

Companding is a very popular technique in communication engineering, especially in voice communication systems using Pulse Code Modulation (PCM) (e.g. Lathi, 1998). A PCM block consists of a signal sampler, an amplitude quantization unit and an encoder. The quantization process leads to the approximation of the amplitudes of the samples. Considering a quantization step size of Q, the amplitude of a sampled signal that falls into this particular step size level will be represented by the quantization value of this level irrespective of the actual amplitude of the sample. This process introduces a maximum quantization error of ±Q/2.

The quantization error is the difference between the quantized output value and the true value of the sample. Quantization error adds noise to the signal, known as quantization noise. Generally, as the size of all quantization steps is the same, the quantization error will be constant for all steps thus the quantization noise is constant, while the signal amplitude can vary. This results in a varying signal-to-quantization noise power ratio, SQR, given by

$$SQR = \frac{Signal\ Power}{Quantisation\ Noise\ Power} = \frac{P_s}{P_{qn}} \tag{9}$$

As $P_{qn}$ is constant, the SQR is directly proportional to the signal power $P_s$, which means that the large signals will have a higher SQR and hence a better quality than the small signals. The SQR can be maintained constant if $P_{qn}$ is decreased or increased in the same proportion as the decrease or increase of $P_s$.

The process known as companding in PCM systems is used to maintain a constant SQR. The signal to be transmitted is passed through an amplifier with a non-linear transfer characteristic that favours amplification of the small-amplitude signals. The signal then appears large during quantization and hence the effect of quantization noise upon smaller signals is reduced. The correct amplitude relations are restored by the reciprocal expander in the receiver.

### 4.2 μ-Law Companding Profiles

The μ-Law compander, introduced by Bell Systems, is perhaps the most popular compander in relation to PCM systems and is widely used in North America. The input-output transfer characteristics of a μ-Law compander are described by the formula

$$y = x_{peak} \frac{\ln\left(1+\mu\frac{|x|}{x_{peak}}\right)}{\ln(1+\mu)}\mathrm{sgn}(x) \tag{10}$$

where $x$ is the instantaneous input signal, $y$ is the companded output signal, $x_{peak}$ is the maximum input/output signal and sgn is the signum function. The parameter μ determines the companding profile. The standard value for μ is 255 and this is normally used with an 8-bit converter (e.g. Sklar, 2001). Figure 2 shows the μ-Law compander input-output characteristics for μ = 0 (linear) and for μ varying from 0.1 to 1000.

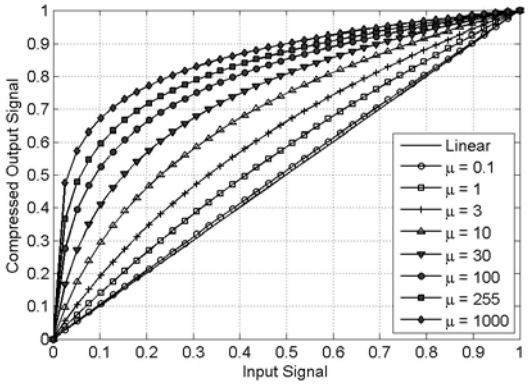



Fig. 2. The μ-Law compander profile for values of μ from 0 to 1000

### 4.3 μ-Law Companding and PAPR Reduction

From the transfer characteristics of the μ-Law compander, the signals with lower amplitudes are amplified with greater gain than the higher amplitudes signals which are amplified with lower gain. In OFDM systems, the occurrence of subcarriers having very large peak amplitudes is less frequent, while most of the subcarriers have lower peak amplitudes. Because of the less frequent high amplitude subcarriers, the average power is low, resulting in a high PAPR. The high PAPR can be reduced if one of the following is

possible – either a decrease in the peak amplitudes of the subcarriers or an increase in the average power of the transmitted OFDM signal. If an OFDM signal is passed through a μ-Law compander which is designed to cover the range of all amplitudes encountered, the subcarrier with the highest peak amplitude will remain relatively unaltered if near the higher levels of the transfer characteristic, while all other subcarriers with lower peak amplitudes will be amplified with varying but larger gains. Thus, the peak power remains relatively unaltered while the average power of the signal is increased due to the companding process. As a result, there is potential reduction in PAPR.

The mathematical expression for the PAPR of general OFDM was given in equation (7). The PAPR for companded OFDM can now be derived. Companding the subcarriers using a μ-Law profile outlined through equation (10), then it can be shown that the PAPR formula for companded OFDM is given by

$$PAPR_{comp} = \max\left(\frac{N\ln(1+\mu[P_{peak}]^{0.5})}{\sum_{n=1}^{N}\ln(1+\mu[P_n(t)]^{0.5})}\right) \tag{11}$$

where $P_n(t)$ is the normalised instantaneous power, i.e. $(|x|/x_{peak})^2$, of the $n^{th}$ subcarrier and $P_{peak}$ is the normalised peak power from the compander. The maximum PAPR for the companded OFDM transmission occurs for the situation when the data on each subcarrier is the same. In such a situation, the IFFT of the data results in a peak power at one IFFT point and zero power at all other points, similar to the situation outlined in Figure 1 for OFDM with 16 subcarriers carrying the same data and employing QPSK. Thus, the instantaneous power of the first subcarrier, $P_1(t)$ will be the same as the peak envelope power, $P_{peak}$, and the power at all other subcarriers will be zero. Therefore, the maximum $PAPR$ is given by

$$PAPR_{comp_\max} = \max\left(\frac{N\ln(1+\mu[P_{peak}]^{0.5})}{\ln(1+\mu[P_{peak}]^{0.5})}\right) = N \tag{12}$$

In this situation, the PAPR of both general OFDM and companded OFDM signals is identical, indicating that companding does not reduce PAPR for this situation. However, when the input data is random, which is presumed to be the normal situation for most applications, then equation (12) indicates that the companding process may be very effective in reducing PAPR. Equation (12) also helps in validating the formulation of PAPR of companded OFDM by showing that the maximum PAPR is the same as the maximum theoretical value of OFDM given by $N$ or $10log_{10}(N)$ in dB.

Even though the average and peak envelope power for MPSK and MQAM signal transmissions will be different, it may be shown that the PAPR distributions are in fact identical for all MPSK and MQAM OFDM modulations (e.g. Vallavaraj, 2008). The amplitude levels of MQAM are different from MPSK, which will increase the peak envelope power. However, the average power will also increase in MQAM, thereby making MQAM no different from MPSK as far as the PAPR is concerned. This can be established easily through analytical or simulation studies.

To demonstrate the PAPR reduction arising from companding, QPSK OFDM transmissions can be numerically computed and compared for random data transmission for general

OFDM and μ-Law companded OFDM with μ = 255. Figure 3 shows the conceptual block diagram of the basic Matlab/Simulink simulation model. Random binary sequence data is mapped onto subcarriers using QPSK and then oversampling by a factor of 8 is performed by padding zeros to the baseband modulated signals. The OFDM signals are then obtained by taking the IFFT, from which the instantaneous peak power and the average power over each OFDM symbol can be calculated in order to evaluate the instantaneous PAPR of each OFDM symbol. Simulated results can then be compared with analytical calculations derived from equations (7) and (11).

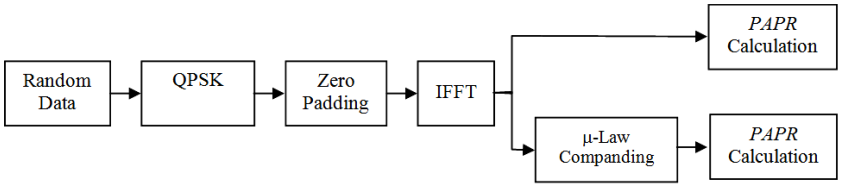



Fig. 3. Block diagram of the basic OFDM model used for the PAPR measurement

Figure 4 displays the maximum PAPR obtained over 1000 OFDM symbols as a function of the number of subcarriers for both normal OFDM and companded OFDM.

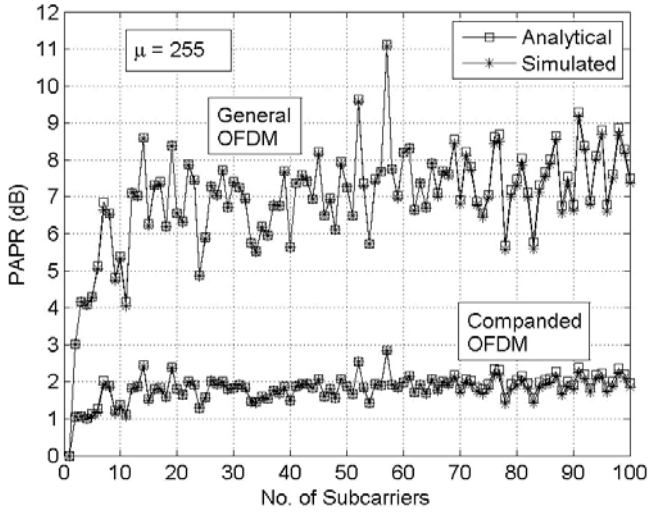



Fig. 4. PAPR as a function of the number of subcarriers for general OFDM and companded OFDM with μ = 255

The results show almost identical agreement between all analytical and simulated results. It can also be seen that μ-Law companding with μ = 255 provides significant improvement in PAPR over general OFDM transmissions. Specifically, the companding scheme produces a much lower PAPR even for a small number of subcarriers and approaches about 2 dB as the number of subcarriers increases towards 100 and beyond.

## 5. Companding the WiMax IEEE802.16e DL PUSC

The investigation into the application of companding specifically to Mobile WiMax with the inclusion of modulated pilots has received only little attention in the literature (Stewart & Vallavaraj, 2009). To help address this topic the rest of this chapter endeavours to provide a comprehensive simulation investigation of the application of μ-Law companding to one implementation of mobile WiMax where the influence on PSD, BER, PAPR and mobile multipath channel as a function of μ and modulation is investigated and quantified.
Mobile WiMax uses the principle of OFDMA (OFDM Access) in which a number of subchannels are assigned to different subscribers. The mobile WiMax implementation used in the simulation studies undertaken here is the DL PUSC (Down Link Partially Used Sub-Carrier). This downlink mode has been selected because it assumes all subchannels can be utilised for information transmission from a base station. Table 1 outlines the parameters of the IEEE802.16e OFDMA configuration employed for the simulation studies. This represents one particular embodiment of the IEEE802.16e standard and is used to demonstrate the application of companding to the system. In the configuration chosen, there are 720 data subcarriers, 184 null subcarriers, 30 subchannels, an IFFT size of 1024, and a guard interval of 1/8 of the 1024 IFFT OFDMA period. The modulations employed are QPSK, 16QAM and 64QAM. Each subchannel comprises 12 subcarriers and 2 pilots.

| Parameter | Value |
|---|---|
| System Bandwidth | 10 MHz |
| IFFT Size | 1024 |
| Modulation | QPSK, 16QAM, 64QAM |
| Subchannels | 30 |
| Null subcarriers | 184 |
| Data subcarriers | 720 |
| Cluster size | 28 subcarriers (14 even symbol + 14 odd symbol) |
| Clusters per subchannel | 2 |
| Data subcarriers per subchannel | 48 |
| Pilot subcarriers per subchannel | 8 |
| Pilot insertion even symbol | $5^{th}$ and $9^{th}$ subcarriers |
| Pilot insertion odd symbol | $1^{st}$ and $13^{th}$ subcarriers |
| Pilot values | 4/3 BPSK Modulated |
| Guard Interval $T_g$ (1/8) | 12.8μs |
| OFDMA Symbol period $T_s$ | 115.2μs |

Table 1. WiMax DL PUSC simulation parameters

Each pilot has a magnitude of 4/3 and is BPSK modulated to conform to the IEEE802.16e standard. The modulation of the pilots uses -1 and +1 PN sequences. Other orthogonal pilot modulation code sequences may be simulated if desired but these can be shown to provide exactly the same results as presented below for the PN sequences. As per the standard, the pilot assignments in the even and odd symbols of clusters in the DL PUSB are shown in Figure 5. These pilots are used for channel estimation and correction. A number of techniques such as piecewise linear interpolation across subcarriers within each cluster, or more advanced techniques such as Least Squares (LS), or Minimum Mean Squares Error

(MMSE) may be employed (e.g. Hanzo et al., 2003). In the work presented here, when no mobile channel is considered, then the pilots are not utilised for channel estimation or data correction. For simplicity, when a mobile channel is considered in this work a piecewise linear interpolation is applied to correct the channel across the subcarriers independently within the odd and even symbols of each cluster.

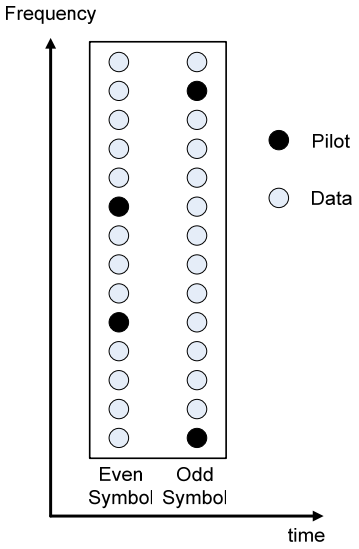
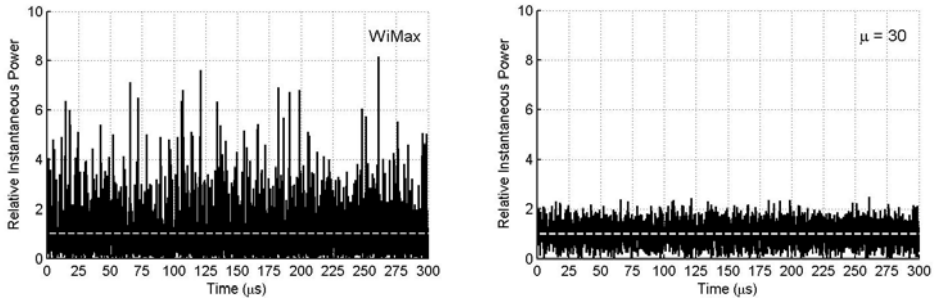



Fig. 5. Pilot assignment for individual OFDMA clusters

In order to set the peak level of the compander in the simulations, 100,000 standard WiMax symbols were transmitted for each of QPSK, 16QAM and 64QAM modulations. The maximum peak amplitude of the set was then chosen as $x_{peak}$ in the companding equation (10). It may be commented that the peak amplitudes for each of QPSK, 16QAM and 64QAM were within approximately 1.0% of each other. All modelling and simulation of the WiMax system was again carried out using Matlab/Simulink. To simplify the evaluation of PAPR and BER in the simulation studies of the DL PUSC, the preamble, Reed-Solomon encoding, convolution coding or Turbo coding have not been considered.

To appreciate the general influence of μ-Law companding, Figure 6 compares the relative instantaneous power transmission signals for the same random data over a period of 300μs for 16QAM modulated general WiMax and companded WiMax employing $\mu = 30$. In Figure 6, the instantaneous power is presented in relative terms with respect to the average symbol power of the transmissions. The impact of companding can be immediately appreciated as the PAPR has been significantly reduced. For example, over the 300μs time period in Figure 6(a), the peak power for the general WiMax system is approximately 8.2 times larger than the average power producing a PAPR of 9.1 dB. In Figure 6(b) the peak power is approximately 2.6 times larger than the average power for companded WiMax resulting in a PAPR of 4.15 dB. This reduction is a very attractive feature of the companding process. However, the drawback is that when companding is directly applied, the average power of each OFDMA symbol is increased.




(a) WiMax (b) Companded WiMax with μ = 30

Fig. 6. Comparison of relative instantaneous power transmissions for (a) standard WiMax, and (b) companded WiMax using μ = 30

In order to appreciate this increase in average power, Figure 7 shows the relative average symbol power gain as a function of μ with respect to the average uncompanded WiMax symbol power for each OFDMA symbol transmission. As the companding profiles are independent of the data modulation employed then it is easy to demonstrate (through simulation) that Figure 7 is also independent of the modulation.

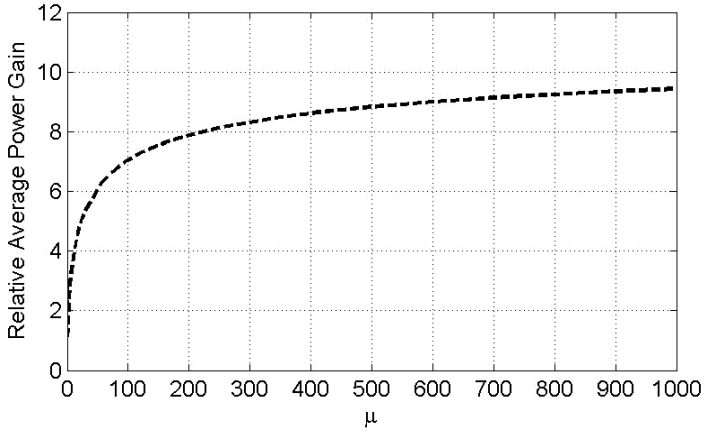
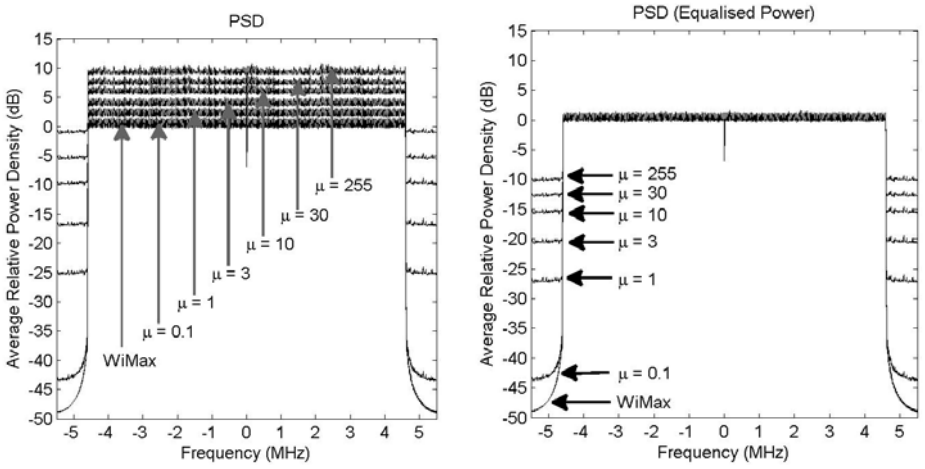



Fig. 7. Relative average power gain of companded WiMax as a function of μ

Clearly as μ increases from μ = 0, i.e. general WiMax, the average power increases rapidly but then becomes asymptotic as μ tends to higher values. For the PAPR transmission example shown in Figure 6, with μ = 30, the average relative power gain is 5.37 times higher than standard WiMax even though the PAPR has dramatically reduced. The issue of increased power due to companding, though recognised, is not often quantified or evaluated in relation to a complete assessment of the significance of companding. The following chapters will attempt to investigate the assessment of both straight companding and also the consideration of companding when the average symbol power of companded transmissions is equalised to the uncompanded WiMax symbol power for each value of μ.

## 6. Spectral Issues Associated With μ-Law Companding

The averaged PSD of WiMax and companded WiMax as a function of μ between 0.1 and 255 is shown in Figures 8(a) and (b). These figures compare the PSD for both the direct application of companding and also the equalised power companded situations. The averages were performed over 2000 symbol transmissions to provide accuracy in the results.




(a) Companded (b) Companded equalised power

Fig. 8. Averaged PSDs of general WiMax and companded WiMax as a function of μ between 0.1 and 255 for (a) μ-Law companding, (b) μ-Law companding employing equalised averaged symbol power

The presence of increased in-band spectral power as a function of increasing values of μ is clearly evident in Figure 8(a), where spectral power density re-growth has occurred within the signal bandwidth. This phenomenon is appreciated in the literature and as discussed above is an artefact of the raw companding process. As an example, for μ = 255, the in-band PSD level has increased by around 9dB from the normal uncompanded WiMax situation. This is in general agreement with the power gain curve in Figure 7 for μ = 255. However, out-of-band spectral splatter is also introduced and increases with increasing μ. It can be seen that the rapid or sharp roll-off of the PSD outside the signal bandwidth for μ values still exists, but the base of the bandwidth skirt rises significantly as μ increases. For μ = 255, it can be seen that the base of the skirt is only around 10 dB lower than the in-band PSD. In the equalised power situation, it is easier to appreciate that small values of μ introduce lower levels of out-of-band energy and will therefore potentially cause less interference.

Internal spectral re-growth will assist in improving BER when raw campanding is applied and this may be considered a natural advantage of companding. In reality, if power is to be equalised for all values of μ, then a critical assessment of how BER variations with equalised power must be understood. However, a further important issue in relation to PSD is that the presence of any increased out-of-band spectral power will cause inter channel interference if not addressed. Clearly choice of very small values of μ will assist in reducing this

interference. Application of suitable digital filtering may also be employed but this in itself has to be carefully considered as phase variations associated with filter roll-off must be appreciated around the bandwidth edges otherwise inherent BER will occur at the receiver if not properly corrected.

## 7. BER Evaluation as a Function of μ

The BER for companded WiMax and equalised power companded WiMax has been evaluated as a function of μ over 0.1 to 1000 for QPSK, 16QAM and 64QAM modulations. The results are shown in Figures 10(a) –10(f) (see next page) where the BER probability for each situation is plotted as a function of the SNR. These graphs demonstrate two main aspects related to companding. Firstly, for the straight companding situations (i.e. Figures 10(a), (c) and (e)), as μ increases from zero (i.e. from standard WiMax), the BER performance starts to improve as a direct consequence of the increased power. However, after a certain value of μ the BER performance starts to degrade and the curves gradually move outwards towards higher values of SNR.

The reason for this can be understood as follows. As μ increases from 0 towards larger values, the companding profiles presented in Figure 2 indicate that for larger range input signals there is less range variation on the companded output signals. The signals are of course expanded through decompanding at the receiver. When noise is present for larger μ values, even a small noise variation on a large amplitude signal can produce significant variations on the decompressed amplitude signal at the receiver thus causing larger bit errors to be produced. To understand the behaviour more accurately, and to determine optimized μ values which produce best or optimised BER performance for companding it is possible to plot the SNR value which produces a BER of 0.001 as a function of μ and each modulation. These plots are shown in Figure 9.

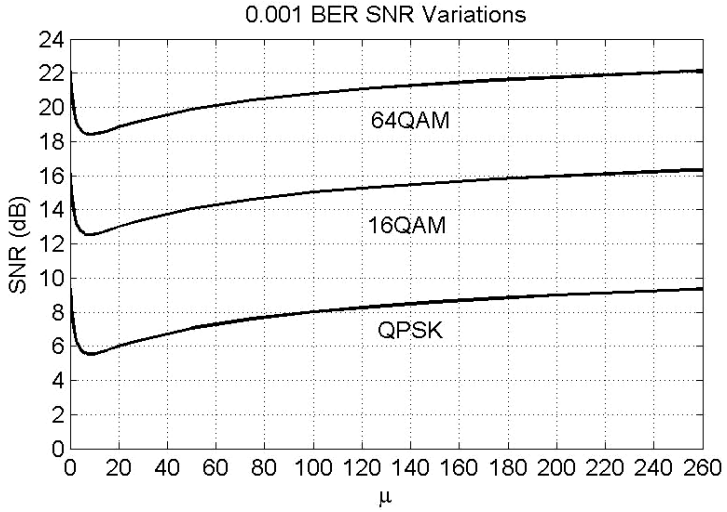
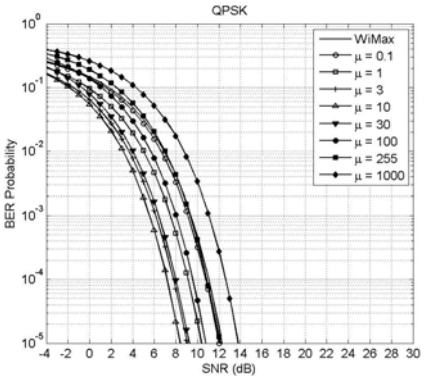
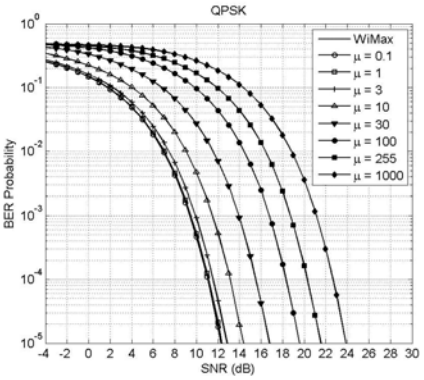


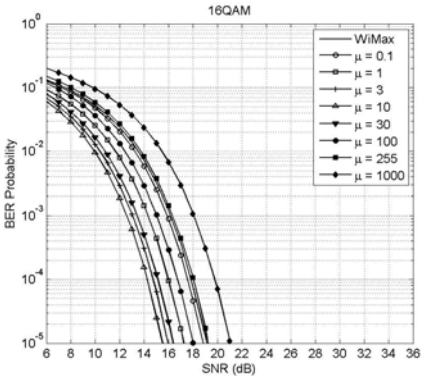

Fig. 9. The variation in SNR as a function of μ and modulation for companded WiMax for a BER probability of 0.001



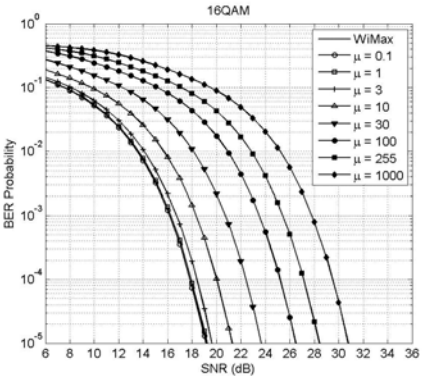

(a) Companded QPSK



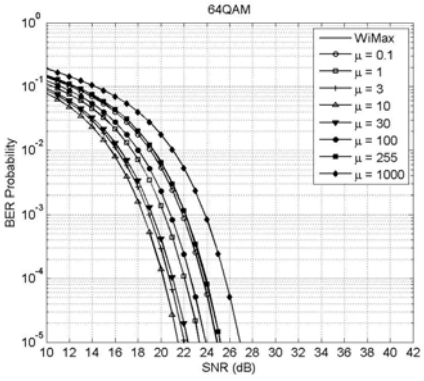

(b) Companded QPSK Equalised Power



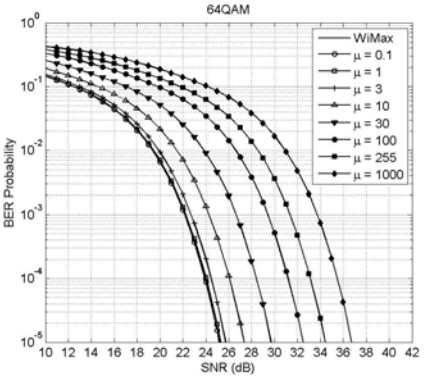

(c) Companded 16QAM




(d) Companded 16QAM Equalised Power




(e) Companded 64QAM




(f) Companded 64QAM Equalised Power

Fig. 10. Evaluation of BER probabilities against SNR as a function of μ for companded WiMax and also companded WiMax with equalised symbol power - QPSK BER curves are shown in (a) and (b), 16QAM in (c) and (d), and 64QAM in (e) and (f)

Figure 9 demonstrates that the BER improves quickly (due to the decrease in SNR) for low values of μ but then gradually degrades as the companding profile changes for larger μ values. The change in SNR is also seen to be similar for each modulation. The optimised μ value which achieves best BER performance (i.e. achieves the best SNR) for each modulation is shown in Table 2.

| Modulation | Optimised μ |
|---|---|
| QPSK | μ = 8 |
| 16QAM | μ = 8 |
| 64QAM | μ = 8 |

Table 2. Optimised BER μ values for companded WiMax

Table 2 demonstrates that the optimized μ value is independent of the modulation used in WiMax. From Figure 9 there is also very little variation in BER around the optimized μ values i.e. around μ = 7 to 9. The BER curves for QPSK, 16QAM and 64QAM modulations for optimised μ = 8 are displayed in Figure 11 alongside the uncompanded WiMax BER curves. These curves allow the optimized BER performance of companded WiMax to be made with standard WiMax.

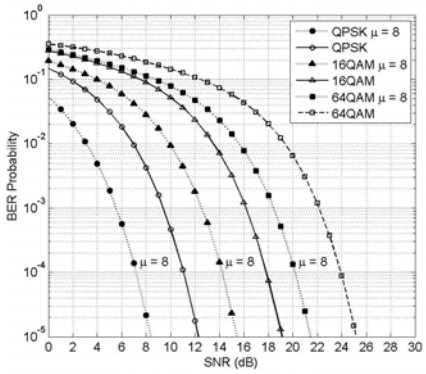
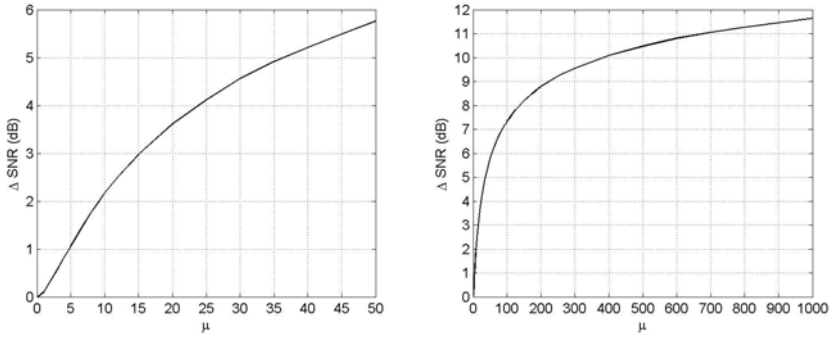



Fig. 11. BER probability comparison between optimised companding with μ = 8, and WiMax as a function of the modulation

However, the BER graphs of Figures 10 (a), (c) and (e) may be misleading as any WiMax transmitter must still operate under the constraints of actual real output power requirements. In this situation, the equalised power BER graphs represented in Figure 10 (b), (d) and (f) should be consulted. These curves indicate that when equalised symbol power is considered for each value of μ, then as μ is increased the BER actually degrades for all SNR situations. This indicates, as expected, that in terms of output power requirements there is in reality a reduction in BER arising as an artefact of the companding process. However, it should be noted that the degradation in BER is not too significant for reduced values of μ indicating that for companding curves associated with lower values of μ, the decrease in BER performance is not severe. Simulations demonstrate that the cost in terms of increased SNR when companding is applied is nearly independent of whether QPSK,

16QAM or 64QAM modulation is utilised. The difference in increased SNR cost across all simulations was in agreement to approximately ±0.1dB for each value of μ. This is expected as the increased SNR companding power increase also appears to be independent of the modulation from the optimised results shown in Figure 9 and discussed above. The cost in terms of increased SNR, denoted by Δ SNR, as a function of μ for a BER probability of 0.001 is plotted in Figure 12(a) for the range of μ = 0 to 50 to show how variations in Δ SNR occur for smaller μ values, and also in Figure 12(b) for the general range between μ = 0 to 1000.




(a) μ = 0 to 50 (b) μ = 0 to 1000

Fig. 12. The variations in Δ SNR as a function of μ for (a) μ values in the range 0 to 50, and (b) μ values in the general range 0 to 1000

Figures 12(a) and (b) are important results as they allow a designer to choose a suitable companding profile knowing the BER degradation or reduction in SNR expected as a function of μ. Ultimately this degradation in BER performance can be chosen in conjunction with the desired PAPR required when companding is considered. Quantification of the reduction in PAPR as a function of μ will now be presented in the next section.

## 8. PAPR Reduction as a Function of μ

It has been shown previously that the PAPR is independent of the modulation employed in the data transmissions of OFDM systems. This is also easily verified for the WiMax simulations as well. Through simulation, the PAPR values of 100,000 DL PUSB WiMax symbols were evaluated for QPSK, 16QAM and 64QAM. The conventional PAPR Complementary Cumulative Distribution Function (CCDF) distributions were then constructed. The CCDFs represent the PAPR probability distributions and allow the probability of a PAPR value of any WiMax symbol exceeding a certain value (denoted by $PAPR_0$) to be determined. In agreement with previous general investigations of companded OFDM systems, the PAPR CCDF distributions of both non-companded and companded OFDM signals is dependent only on the number of subcarriers and is independent of the type of modulation used (e.g. Vallavaraj, 2008; Stewart and Vallavaraj, 2008). For the CCDFs in each of the modulation situations for WiMax, the same values within PAPR tolerances of approximately ±0.1dB were obtained at the 0.001 probability level. The PAPR CCDFs as a function of μ are shown in Figure 13. Also shown in Figure 13 is the CCDF curve for the

optimised non-equalised power value of μ = 8. It can be seen that as μ increases, the PAPR can be significantly reduced. Though it is attractive to reduce the PAPR to very low levels, the disadvantage of companding is the inherent compromise in BER for equalised power situations and spectral splatter as outlined in the previous sections. However, it is worthy to note that for small values of μ, the PAPR reduction can still be significant without severe compromise to the BER. For example, even with μ = 1, at the 0.001 probability level the PAPR reduces from around 11.7dB to 9.9dB. For the optimized non-equalised power situation, μ = 8, the PAPR is reduced from around 11.7dB to 6.6dB, a reduction of 5.1dB.

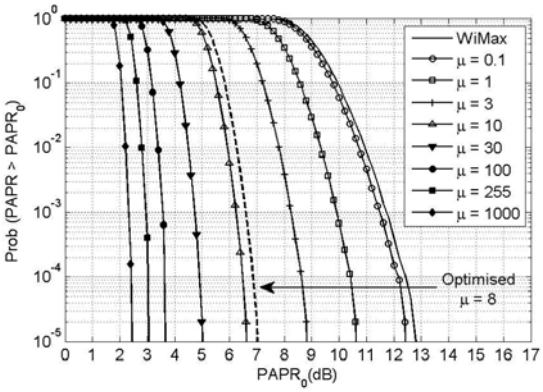
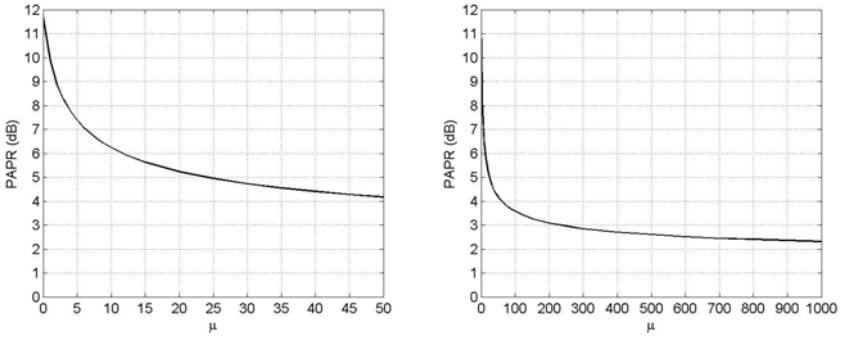



Fig. 13. Companded WiMax CCDF curves as a function of μ

Figure 14 quantifies the PAPR reduction at the CCDF 0.001 probability level as a function of μ. Two ranges of μ are presented to allow a more detailed appreciation of the PAPR levels achieved for lower values of μ as well as the general reduction in PAPR up to μ = 1000.




(a) μ = 0 to 50 (b) μ = 0 to 1000

Fig. 14. WiMax PAPR levels at the CCDF 0.001 probability level for (a) the range μ = 0 to 50 and (b) the general range μ = 0 to 1000

For situations of improved PAPR a compromise is required to be struck between and desired PAPR and any acceptable degradation in BER. The BER and CCDF figures

presented in this chapter will therefore allow decisions to be made between PAPR reduction and expected BER levels for specific application requirements.

## 9. The Influence of Mobility on Companding

The influence of companding on the BER performance of OFDM systems when mobility is present has also received little general attention in the literature. The effect of companding within mobile environments for WiMax is investigated here for two specific standard channel models, viz. the WiMax ITU fixed Vehicular A (Veh A) channel and the WiMax ITU mobile Pedestrian B (Ped B) channel. Veh A represents mobility in a vehicular environment while Ped B represents pedestrian mobility. The specific relative multipath delays and average powers associated with these channels are outlined below in Table 3.

| Veh A | | Ped B | |
|---|---|---|---|
| **Delay (ns)** | **Power (dB)** | **Delay (ns)** | **Power (dB)** |
| 0 | 0 | 0 | -3.9 |
| 310 | -1 | 200 | -4.8 |
| 710 | -9 | 800 | -8.8 |
| 1090 | -10 | 1200 | -11.9 |
| 1730 | -15 | 2300 | -11.7 |
| 2510 | -20 | 3700 | -27.8 |

Table 3. Channel parameters for the ITU Veh A and Ped B simulations

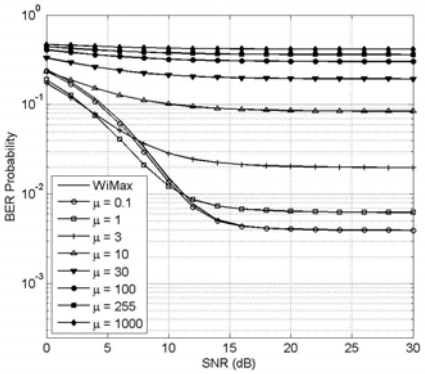
The channel correction method employed in this study is a simple piecewise linear interpolation between pilots in each individual symbol within each cluster. More advanced channel correction methods have not been considered in this study as the purpose is to establish the general influence of mobility on companded WiMax.

Speeds of 60kmh$^{-1}$ for Veh A and 3kmh$^{-1}$ for Ped B have been considered for evaluation of the influence of companding on the BER performance. Assuming a carrier frequency of approximately 2.4GHz, then the appropriate maximum Doppler frequencies related to mobility can be incorporated in the Matlab/Simulink channel models. Table 4 shows the adopted mobile speeds and the associated Doppler frequencies used for the simulations models.

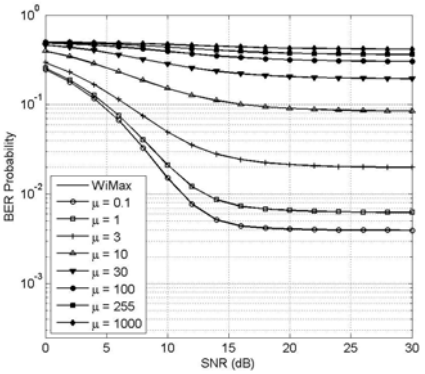
| Veh A | | Ped B | |
|---|---|---|---|
| **Speed** | **Doppler Frequency** | **Speed** | **Doppler Frequency** |
| 60kmh$^{-1}$ | 133.33Hz | 3kmh$^{-1}$ | 6.67Hz |

Table 4. Speeds and associated maximum Doppler frequencies used in the WiMax channel simulation models for Veh A and Ped B channels

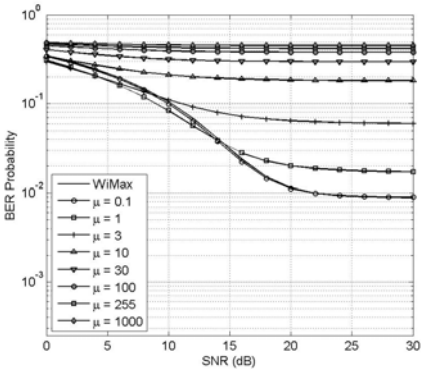
Figures 15(a)-(f) display the results for mobility at 60kmh$^{-1}$ in the Veh A channel as a function of QPSK, 16QAM and 64QAM modulation for general WiMax, companded WiMax and equalised power companded WiMax over $\mu = 0.1$ to 1000. Figures. 16(a)-(f) display the results for mobility at 3kmh$^{-1}$ in the Ped B channel for the same parameters evaluated for Figure 15.



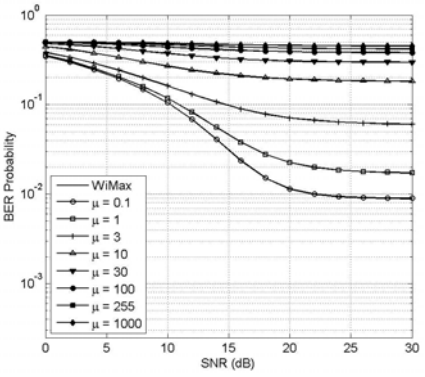

(a) QPSK Veh A



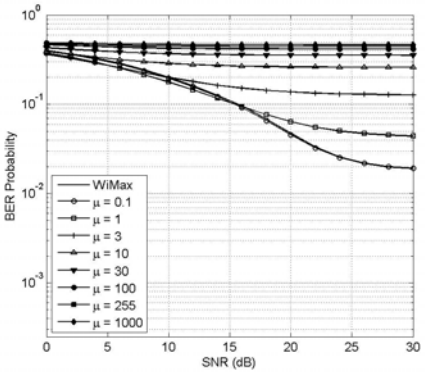

(b) QPSK Veh A Equalised Power



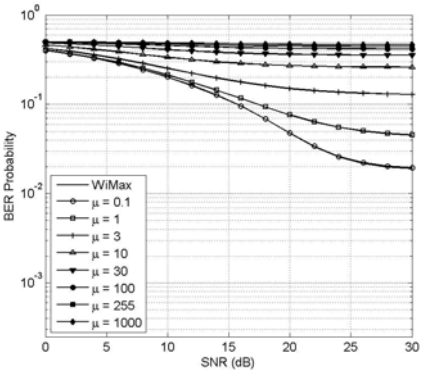

(c) 16QAM Veh A




(d) 16QAM Veh A Equalised Power

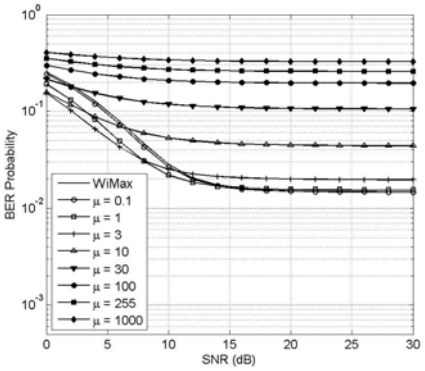



(e) 64QAM Veh A

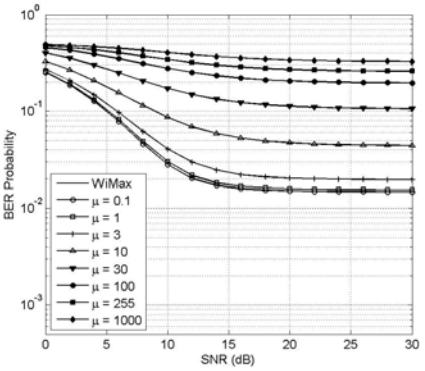



(f) 64QAM Veh A Equalised Power

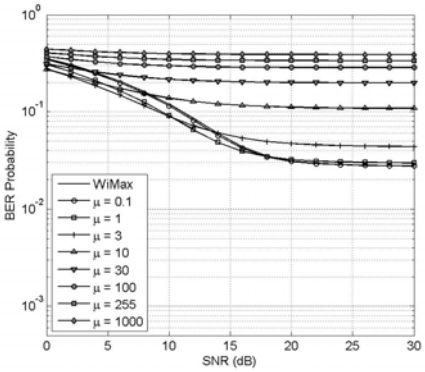
Fig. 15. QPSK, 16QAM and 64QAM Veh A BER probability curves as a function of μ for situations of companded and equalised power companded WiMax



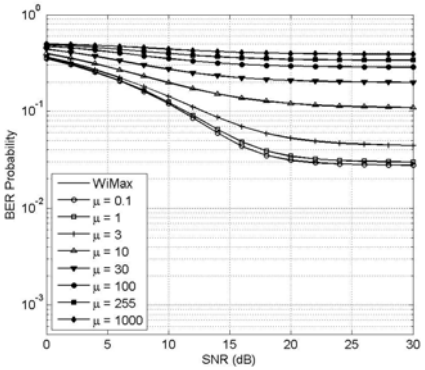

(a) QPSK Ped B



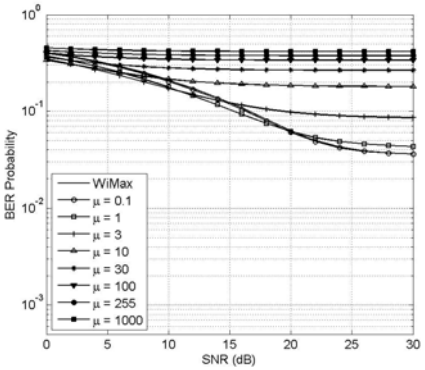

(b) QPSK Ped B Equalised Power



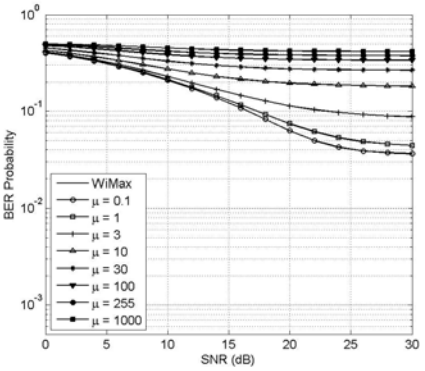

(c) 16QAM Ped B




(d) 16QAM Ped B Equalised Power




(e) 64QAM Ped B




(f) 64QAM Ped B Equalised Power

Fig. 16. QPSK, 16QAM and 64QAM Ped B BER probability curves as a function of μ for situations of companded and equalised power companded WiMax

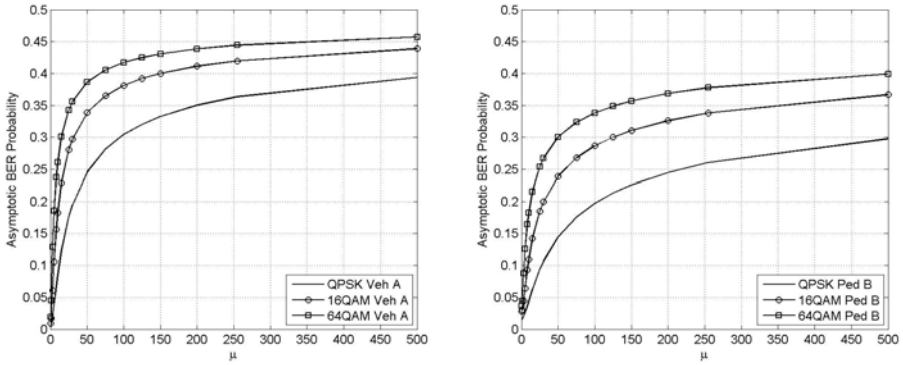
The significant observations regarding mobility are as follows. The BER depreciates significantly for Veh A and Ped B channels as μ increases for both companded and equalised power companded situations. It can also be seen that for each value of μ, as the SNR increases, the BER flattens off to an asymptotic optimum value. This asymptotic value decreases with increasing μ, and also with increased data modulation on the subcarriers i.e. the performance is best for QPSK, deteriorates for 16QAM and further deteriorates for 64QAM. Thus a general conclusion is that increased companding will always degrade the performance of WiMax systems for larger SNR in the mobile channels considered. It may also be noted that for very small values of μ, the BER performance in the asymptotic region is comparable to the asymptotic value associated with standard WiMax. The main reason for the depreciated BER performance is clearly a combination of the companding profile, the modulation and also the affects of the channel.

Interestingly, for the direct companding situations, there is a marginal improvement in BER rate over WiMax at lower SNR values across a range of μ values. An improvement in BER with companding is expected due to the inherent increased average power provided through the companding process itself. However, the BER is still poor over the regions where the improvement over WiMax occurs. For the Veh A scenarios in Figure 15(a), (b) and (c), the value of μ which optimises the BER varies over the lower SNR range under consideration. The optimised μ values over the lower SNR range is also nearly independent of the modulation employed. For example, for QPSK, 16QAM and 64QAM, for SNR < 4dB the curve for $\mu \approx 3$ provides the best BER performance. For the approximate range 4dB < SNR < 11dB, the curve for $\mu \approx 1$ is best, and for 11dB < SNR < 16dB, $\mu \approx 0.1$ is optimum. Above 16dB WiMax provides the best BER performance, although there is minimal difference for values of μ around 1 or less than 1 and WiMax as the BER levels off.

For the Ped B channel in Figures 16(a), (c) and (e), the best performance of companding for lower values of SNR appears to be more dependent on the modulation. For the QPSK BER curves evaluated, for SNR < 8dB, $\mu \approx 3$ is preferred, for SNR > 8, $\mu \leq 1$ is best, though values of μ around 1 provide similar results to WiMax in this situation. For 16QAM, $\mu \approx 3$ is preferred for SNR < 10dB, and for 10 dB < SNR > 18dB $\mu \approx 1$ is preferred and for SNR > 18dB $\mu \approx 0.1$ is best. For 64QAM, $\mu \approx 3$ is preferred for SNR < 11dB, whilst for the range 11dB < SNR < 20dB $\mu \approx 1$ produces the best BER, and for SNR > 20dB, $\mu \leq 0.1$ is the best. Again for increasing SNR WiMax produces the best asymptotic BER performance, though there is little difference in BER performance between WiMax and very small values of μ as the BER levels off. Clearly, the BER performance in lower SNR values, when mobility is present, depends not just on the companding profile, but on the modulation and the nature of the multipath channel.

As discussed previously, the raw companding BER curves may be slightly misleading due to the fact that real transmitters may be required to work on power limitations in which case the equalised symbol power curves are important. For the equalised power situations, as expected, as μ increases, the BER performance depreciates. However, for very small values of μ, in all situations, the companded performance is similar or close to the general WiMax situation. The reason for the rapid deterioration in BER with increasing μ can be explained again as a consequence of the nature of the companding profiles, i.e. large peak amplitude signals can have significant decompanding bit errors at a receiver for larger μ values when noise is present. This, mixed with the problems of a mobile channel, accentuates the deterioration in BER. However, for some situations the increased BER may be acceptable

within some mobile channels when a significant improvement in PAPR is desired. But perhaps the most important result is that the asymptotic BER values for the equalised power companded situations are nearly identical to the raw companded asymptotic BER values. These asymptotic values are plotted in Figure 17 and indicate that for large SNR values, when μ is applied, the influence of the multipath channel is the overwhelming limiting factor on the BER performance. Figure 17 is therefore useful to precisely quantify the optimum BERs achievable for the Veh A and Ped B channels when companding is applied.




(a) Veh A 60kmh$^{-1}$ (b) Ped B 3kmh$^{-1}$

Fig 17. Variation of the asymptotic BER values as a function of μ for QPSK, 16QAM and 64QAM for (a) Veh A 60km$^{-1}$ and (b) Ped B 3kmh$^{-1}$

## 10. Conclusions

This chapter has presented and discussed the principles of PAPR reduction and the principles of μ-Law companding. The application of μ-Law compounding was applied to one implementation of mobile WiMax using an FFT/IFFT of 1024. The main conclusions are as follows. Companding using μ-Law profiles has the potential to reduce significantly the PAPR of WiMax. For straight companded WiMax the average power increases and as a consequence the BER performance can be improved. For direct companding the optimum BER performance occurs for μ = 8, which produces a PAPR of approximately 6.6dB at the 0.001 probability level, i.e. a reduction of 5.1 dB. However an increase in frequency spectral energy splatter occurs which must be addressed to minimise inter channel interference. For equalised symbol power companded transmissions, the BER performance is actually shown to deteriorate for all values of μ. However, for small values of μ, the BER degradation is not severe. This is advantageous as a balance between cost in terms of BER and PAPR reduction can now be quantified along with the expected out-of-band PSD for any chosen value of μ. The figures produced in this chapter will allow an engineer to take informed decisions on these issues. In relation to mobility, the influence of companding on performance is more complex and appears to depend on the modulation, mobile speed and more importantly the nature of the channel itself. It was shown that for straight companding the optimum BER performance at low values of SNR was dependent on the value of μ as well as the nature of the channel. Different ranges of lower SNR values defined different optimum values of μ.

Generally, for larger SNR values the BER performance degraded as μ was increased and became asymptotic with increasing SNR. For the equalised power companding situation, WiMax always produces best BER performance. However, for very small values of μ, there is very little difference between companded WiMax and WiMax. A compromise may also be reached in relation to a reduced BER performance in mobility versus a required PAPR level. It was also discovered that the companded and equalised power companded BER optimised asymptotic values for mobility were approximately the same indicating that the best BER performance for the minimum SNR requirements can be quantified for any design value of μ. This is also helpful in understanding the anticipated best BER performance available in mobile channels when companding is chosen to provide a reduced PAPR level.
Further work in relation to the results presented in this chapter may be carried out. This includes an investigation of the BER performance for companded WiMax when channel coding is incorporated, i.e. convolution and turbo coding, when Reed-Solomon coding is employed, and when other more advanced channel estimation techniques are considered. The importance of filtering or innovative techniques for reducing the spectral splatter should also be explored. Other areas for investigation also include quantifying the influence on BER for a larger range of different mobile channels as a function of μ.

## 11. References


Armstrong, J. (2001). *New OFDM Peak-to-Average Power Reduction Scheme, Proc. IEEE, VTC2001 Spring*, Rhodes, Greece, pp. 756-760

Armstrong, J. (2002). *Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering*, *Electronics Letters*, Vol.38, No.5, pp.246-247, Feb. 2002.

Bäuml, R.W.; Fisher, R.F.H. & Huber, J.B. (1996). *Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping, IEE Electronics Letters*, Vol.32, No.22, pp. 2056-2057

Boyd, S. (1986). *Multitone Signal with Low Crest Factor, IEEE Transactions on Circuits and Systems, Vol. CAS-33*, No.10, pp. 1018-1022

Breiling, M.; Müller-Weinfurtner, S.H. & Huber, J.B. (2001). *SLM Peak-Power Reduction Without Explicit Side Information, IEEE Communications Letters*, Vol.5, No.6, pp. 239-241

Cimini, L.J.,Jr.; & Sollenberger, N.R. (2000). *Peak-to-Average Power Ratio Reduction of an OFDM Signal Using Partial Transmit Sequences, IEEE Communications Letters*, Vol.4, No.3, pp. 86-88

Davis, J.A. & Jedwab, J. (1999). *Peak-to-Mean Power Control in OFDM, Golay Complementary Sequences, and Reed-Muller Codes, IEEE Transactions on Information Theory*, Vol. 45, No.7, pp. 2397-2417

De Wild, A. (1997). *The Peak-to-Average Power Ratio of OFDM, MSc Thesis*, Delft University of Technology, Delft, The Netherlands, 1997

Golay, M. (1961). Complementary Series, IEEE Transactions on Information Theory, Vol.7, No.2, pp. 82-87

Hanzo, L.; Münster, M; Choi, B.J. & Keller, T. (2003). *OFDM and MC-CDMA for Broadcasting Multi-User Communications, WLANS and Broadcasting*, Wiley-IEEE Press, ISBN 0470858796

Han, S.H. & Lee, J.H. (2005). *An Overview of Peak-to-Average Power Ratio Reduction Techniques for Multicarrier Transmission,* IEEE Wireless Communications, Vol.12, Issue 2, pp. 56-65, April 2005

Hill, G.R.; Faulkner, M. & Singh, J. (2000). *Reducing the peak-to-average power ratio in OFDM by cyclically shifting partial transmit sequences,* IEE Electronics Letters, Vol.33, No.6, pp. 560-561

Huang, X.; Lu, J., Chang, J. & Zheng, J. (2001). *Companding Transform for the Reduction of Peak-to-Average Power Ratio of OFDM Signals, Proc.* IEEE Vehicular Technology Conference 2001, pp. 835-839

IEEE Std. 802.16e. (2005). *Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*, IEEE, New York, 2005.

Jayalath, A.D.S. & Tellambura, C. (2000). *Reducing the Peak-to-Average Power Ratio of Orthogonal Frequency Division Multiplexing Signal Through Bit or Symbol Interleaving*, IEE Electronics Letters, Vol.36, No.13, pp. 1161-1163

Jiang, T. & Song, Y-H. (2005). *Exponential Companding Technique for PAPR Reduction in OFDM Systems, IEEE Trans*. Broadcasting, Vol. 51(2), pp. 244-248

Jones, A.E, & Wilkinson, T.A. (1995). *Minimization of the Peak to Mean Envelope Power Ratio in Multicarrier Transmission Schemes by Block Coding*, Proc. IEEE VTC'95, Chicago, pp. 825-831

Jones, A.E. & Wilkinson, T.A (1996). *Combined Coding for Error Control and Increased Robustness to System Nonlinearities in OFDM*, Proc. IEEE VTC'96, Atlanta, GA, pp. 904-908

Jones, A.E.; Wilkinson, T.A. & Barton, S.K. (1994). *Block coding scheme for the reduction of peak to mean envelope power ratio of multicarrier transmission schemes*, Electronics Letters, Vol.30, No.25, pp. 2098-2099

Kang, S.G. (2006). *The Minimum PAPR Code for OFDM Systems*, ETRI Journal, Vol.28, No.2, pp. 235-238

Lathi, B.P. (1998). *Modern Digital and Analog Communication Systems*, 3rd Ed., pp. 262-278, Oxford University Press, ISBN 0195110099

Li, X. & Cimini Jr, L.J. (1997). *Effects of Clipping and Filtering on the Performance of OFDM*, Proc. IEEE VTC 1997, pp. 1634-1638

Lloyd, S. (2006). Challenges of Mobile WiMAX RF Transceivers, *Proceedings of the 8th International Conference on Solid-State and Integrated Circuit Technology*, pp. 1821–1824, ISBN 1424401607, October, 2006, Shanghai

May, T. & Rohling, H. (1998). *Reducing the Peak-to-Average Power Ratio in OFDM Radio Transmission Systems,* Proc. IEEE Vehicular Technology Conf. (VTC'98), pp.2774-2778

Mattsson, A.; Mendenhall, G. & Dittmer, T. (1999). Comments on "*Reduction of peak-to-average power ratio of OFDM systems using a companding technique*", IEEE Transactions on Broadcasting, Vol. 45, No. 4, pp. 418-419

Müller, S.H. & Huber, J.B. (1997a). *OFDM with Reduced Peak-to-Average Power Ratio by Optimum Combination of Partial Transmit Sequences*, Electronics Letters, Vol.33, No.5, pp.368-369

Müller, S.H. & Huber, J.B. (1997b). *A Novel Peak Power Reduction Scheme for OFDM, Proc. IEEE PIMRC '97*, Helsinki, Finland, pp.1090-1094

O'Neill, R. & Lopes, L.B. (1995). *Envelope variations and Spectral Splatter in Clipped Multicarrier signals*, Proc. IEEE PIMRC '95, Toronto, Canada. pp. 71-75

Paterson, G.K. and Tarokh, V. (2000). *On the Existence and Construction of Good Codes with Low Peak-to-Average Power Ratios*, IEEE Transactions on Information Theory, Vol.46, No.6, pp. 1974-1987

Pauli, M & Kuchenbecker, H.P. (1996). *Minimization of the Intermodulation Distortion of a Nonlinearly Amplified OFDM Signal, Wireless Personal Communications*, Vol.4, No.1, pp. 93-101

Sklar, B. (2001). *Digital Communications – Fundamentals and Applications*, 2nd Ed, Pearson Education, pp. 851-854

Stewart, B.G. & Vallavaraj, A. 2008. *The Application of μ-Law Companding to the WiMax IEEE802.16e Down Link PUSC,* 14th IEEE International Conference on Parallel and Distributed Systems, (ICPADS'08), pp. 896-901, Melbourne, December, 2008

Tarokh, V. & Jafarkhani, H. (2000). *On the computation and Reduction of the Peak-to-Average Power Ratio in Multicarrier Communications*, IEEE Transactions on Communications, Vol.48, No.1, pp. 37-44

Tellambura, C. & Jayalath, A.D.S. (2001). *PAR reduction of an OFDM signal using partial transmit sequences*, Proc. VTC 2001, Atlanta City, NJ, pp.465-469

Vallavaraj, A. (2008). *An Investigation into the Application of Companding to Wireless OFDM Systems*, PhD Thesis, Glasgow Caledonian University, 2008

Vallavaraj, A.; Stewart, B.G.; Harrison, D.K. & McIntosh, F.G. (2004). *Reduction of Peak-to-Average Power Ratio of OFDM Signals Using Companding*, 9th Int. Conf. Commun. Systems (ICCS), Singapore, pp. 160-164

Van Eetvelt, P.; Wade, G. & Tomlinson, M. (1996). *Peak to average power reduction for OFDM schemes by selective scrambling,* IEE Electronics Letters, Vol.32, No.21, pp. 1963-1964

Van Nee, R. & De Wild, A. (1998). *Reducing the peak-to-average power ratio of OFDM, Proc. IEEE Vehicular Technology Conf.* (VTC'98), pp. 2072–2076

Van Nee, R. & Prasad, R. (2000). OFDM *for Wireless Multimedia Communications*, Artech House, London, pp. 241-243

Wang, L. & Tellambura, C. (2005). *A Simplified Clipping and Filtering Technique for PAR Reduction in OFDM Systems*, IEEE Signal Processing Letters, Vol.12, No.6, pp. 453-456

Wang, L. & Tellambura, C. (2006). *An Overview of Peak-to-Average Power Ratio Reduction Techniques for OFDM Systems*, Proc. IEEE International Symposium on Signal Processing and Information Technology, ISSPIT-2006, pp. 840-845

Wang, X., Tjhung, T.T. and Ng, C.S. (1999). *Reduction of Peak-to-Average Power Ratio of OFDM System Using a Companding Technique*, IEEE Transactions on Broadcasting, Vol.45, No.3, pp. 303-307

Yang, K. & Chang, S.-II. (2003). *Peak-to-Average Power Control in OFDM Using Standard Arrays of Linear Block Codes*, IEEE Communications Letters, Vol.7, No.4, pp. 174-176

5

# VLSI Architectures for WIMAX Channel Decoders

Maurizio Martina and Guido Masera
*Politecnico di Torino*
*Italy*

## 1. Introduction

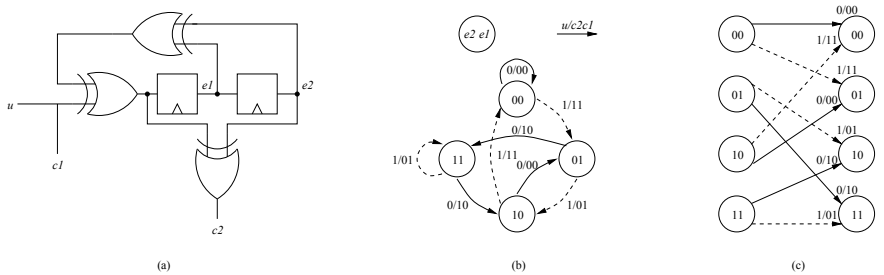
WIMAX has gained a wide popularity due to the growing interest and diffusion of broadband wireless access systems. In order to be flexible and reliable WIMAX adopts several different channel codes, namely convolutional-codes (CC), convolutional-turbo-codes (CTC), block-turbo-codes (BTC) and low-density-parity-check (LDPC) codes, that are able to cope with different channel conditions and application needs.
On the other hand, high performance digital CMOS technologies have reached such a development that very complex algorithms can be implemented in low cost chips. Moreover, embedded processors, digital signal processors, programmable devices, as FPGAs, application specific instruction-set processors and VLSI technologies have come to the point where the computing power and the memory required to execute several real time applications can be incorporated even in cheap portable devices.
Among the several application fields that have been strongly reinforced by this technology progress, channel decoding is one of the most significant and interesting ones. In fact, it is known that the design of efficient architectures to implement such channel decoders is a hard task, hardened by the high throughput required by WIMAX systems, which is up to about 75 Mb/s per channel. In particular, CTC and LDPC codes, whose decoding algorithms are iterative, are still a major topic of interest in the scientific literature and the design of efficient architectures is still fostering several research efforts both in industry and academy.
In this Chapter, the design of VLSI architectures for WIMAX channel decoders will be analyzed with emphasis on three main aspects: performance, complexity and flexibility. The chapter will be divided into two main parts; the first part will deal with the impact of system requirements on the decoder design with emphasis on memory requirements, the structure of the key components of the decoders and the need for parallel architectures. To that purpose a quantitative approach will be adopted to derive from system specifications key architectural choices; most important architectures available in the literature will be also described and compared.
The second part will concentrate on a significant case of study: the design of a complete CTC decoder architecture for WIMAX, including also hardware units for depuncturing (bit-deselection) and external deinterleaving (sub-block deinterleaver) functions.

## 2. From system specifications to architectural choices

The system specifications and in particular the requirement of a peak throughput of about 75 Mb/s per channel imposed by the WIMAX standard have a significant impact on the decoder architecture. In the following sections we analyze the most significant architectures proposed in the literature to implement CC decoders (Viterbi decoders), BTC, CTC and LDPC decoders.

### 2.1 Viterbi decoders

The most widely used algorithm to decode CCs is the Viterbi algorithm [Viterbi, 1967], which is based on finding the shortest path along a graph that represents the CC trellis. As an example in Fig. 1 a binary 4-states CC is shown as a feedback shift register (a) together with the corresponding state diagram (b) and trellis (c) representations.



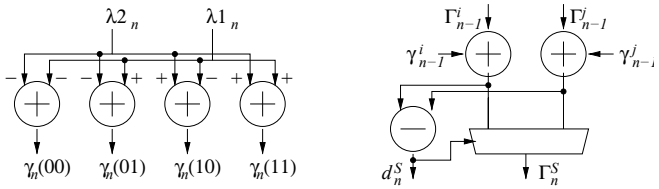

(a) (b) (c)

Fig. 1. Binary 4-state CC example: shift register (a), state diagram (b) and trellis (c) representations

In the given example, the feedback shift register implementation of the encoder generates two output bits, $c_1$ and $c_2$ for each received information bit, $u$; $c_1$ is the systematic bit. The state diagram basically is a Mealy finite state machine describing the encoder behaviour in a time independent way: each node corresponds to a valid encoder state, represented by means of the flip flop content, $e_1$ and $e_2$, while edges are labelled with input and output bits. The trellis representation also provides time information, explicitly showing the evolution from one state to another in different time steps (one single step is drawn in the picture).
At each trellis step $n$, the Viterbi algorithm associates to each trellis state $S$ a state metric $\Gamma^S_n$ that is calculated along the shortest path and stores a decision $d^S_n$, which identifies the entering transition on the shortest path. First, the decoder computes the branch metrics ($\gamma_n$), that are the distances from the metrics labelling each edge on the trellis and the actual received soft symbols. In the case of a binary CC with rate 0.5 the soft symbols are $\lambda 1_n$ and $\lambda 2_n$ and the branch metrics $\gamma_n(c2,c1)$ (see Fig. 2 (a)). Starting from these values, the state metrics are updated by selecting the larger metric among the metrics related to each incoming edge of a trellis state and storing the corresponding decision $d^S_n$. Finally, decoded bits are obtained by means of a recursive procedure usually referred to as trace-back. In order to estimate the sequence of bits that were encoded for transmission, a state is first selected at the end of the trellis portion to be decoded, then the decoder iteratively goes backward through the state history memory where decisions $d^S_n$ have been previously stored: this allows one to select, for current state, a new state, which is listed in the state

history trace as being the predecessor to that state. Different implementation methods are available to make the initial state choice and to size the portion of trellis where the trace back operation is performed: these methods affect both decoder complexity and error correcting capability. For further details on the algorithm the reader can refer to [Viterbi, 1967]; [Forney, 1973]. Looking at the global architecture, the main blocks required in a Viterbi decoder are the branch metric unit (BMU) devoted to compute $\gamma_n$, the state metric unit (SMU) to calculate $\Gamma^S{}_n$ and the trace-back unit (TBU) to obtain the decoded sequence. The BMU is made of adders and subtracters to properly combine the input soft symbols (see Fig. 2 (a)). The SMU is based on the so called add-compare select structure (ACS) as shown in Fig.2 (b). Said $i$ the $i$-th starting state that is connected to an arriving state $S$ by an edge whose branch metric is $\gamma^i{}_{n-1}$, then $\Gamma^S{}_n$ is calculated as in (1).

$$\Gamma^S{}_n = \max_i \{\Gamma^i{}_{n-1} + \gamma^i{}_{n-1}\} \tag{1}$$




(a) (b)

Fig. 2. BMU and ACS architectures for a rate 0.5 CC

As it can be inferred from (1) $\Gamma^S{}_n$ is obtained by adding branch metrics with state metrics, comparing and selecting the higher metric that represents the shortest incoming path. The corresponding decision $d^S{}_n$ is stored in a memory that is later read by the TBU to reconstruct the survived path. Due to the recursive form of (1), as long as $n$ increases, the number of bits to represent $\Gamma^S{}_n$ tends to become larger. This problem can be solved by normalizing the state metrics at each step. However, this solution requires to add a normalization stage increasing both the SMU complexity and critical path. An effective technique, based on two complement representation, helps limiting the growth of state metrics, as described in [Hekstra, 1989].

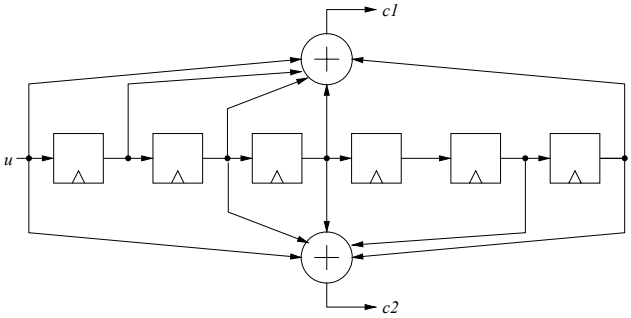



Fig. 3. WIMAX binary 64-state CC with rate 0.5 shift register representation

The WIMAX standard specifies a binary 64 states CC with rate 0.5, whose shift register representation is shown in Fig. 3. Usually Viterbi decoder architectures exploit the trellis intrinsic parallelism to simultaneously compute at each trellis step all the branch metrics and update all the state metrics. Thus, said $n$ the number of states of a CC, a parallel architecture employs a BMU and $n$ ACS modules. Moreover, to reduce the decoding latency, the trace-back is performed as a sliding-window process [Radar, 1981] on portions of trellis of width $W$. This approach not only reduces the latency, but also the size of the decision memory that depending on the TBU radix requires usually $3W$ or $4W$ cells [Black & Meng, 1992].
To improve the decoder throughput, two [Black & Meng, 1992] or more [Fettweis & Meyr, 1989]; [Kong & Parhi, 2004]; [Cheng & Parhi, 2008] trellis steps can be processed concurrently. These solutions lead to the so called higher radix or $M$-look-ahead step architectures. According to [Kong & Parhi, 2004], the throughput sustained by an $M$-look-ahead step architecture, defined as the number of decoded bits over the decoding time is

$$T = \frac{k \cdot N_T \cdot f_{clk}}{N_T / M + W} \approx f_{clk} \cdot M \cdot k \tag{2}$$

where $f_{clk}$ is the clock frequency, $N_T$ is the number of trellis steps, $k$=1 for a binary CC, $k$=2 for a double binary CC and the right most expression is obtained under the condition $W << N_T$ that is a reasonable assumption in real cases.
Thus, to achieve the throughput required by the WIMAX standard with a clock frequency limited to tens to few thousands of MHz, $M$=1 (radix-2) or $M$=2 (radix-4) is a reasonable choice.
However, since CCs are widely used in many communication systems, some recent works as [Batcha & Shameri, 2007] and [Kamuf et al., 2008] address the design of flexible Viterbi decoders that are able to support different CCs. As a further step [Vogt & When, 2008] proposed a multi-code decoder architecture, able to support both CCs and CTCs.

### 2.2 BTC decoders

Block Turbo Codes or product codes are serially concatenated block codes. Given two block codes $C_1=(n_1,k_1,\delta_1)$ and $C_2=(n_2,k_2,\delta_2)$ where $n_i$, $k_i$ and $\delta_i$ represent the code-word length, the number of information bits, and the minimum Hamming distance, respectively, the corresponding product code is obtained according to [Pyndiah, 1998] as an array with $k_1$ rows and $k_2$ columns containing the information bits. Then coding is performed on the $k_1$ rows with $C_2$ and on the $n_2$ obtained columns with $C_1$. The decoding of BTC codes can be performed iteratively row-wise and column-wise by using the sub-optimal algorithm detailed in [Pyndiah, 1998]. The basic idea relies on using the Chase search [Chase, 1972] a near-maximum-likelihood (near-ML) searching strategy to find a list of code-words and an ML decided code-word $\underline{d}=\{d_0,\ldots, d_{n-1}\}$ with $d_j \in \{-1,+1\}$. According to the notation used in [Vanstraceele et al., 2008], decision reliabilities are computed as

$$\lambda(d_j) \approx \frac{|\underline{r} - \underline{c}^{-1(j)}|^2 - |\underline{r} - \underline{c}^{+1(j)}|^2}{4} \tag{3}$$

where $\underline{r}=\{r_0,\ldots r_{n-1}\}$ is the received code-word and $c^{-1(j)}$ and $c^{+1(j)}$ are the code-words in the Chase list at minimum Euclidean distance from $\underline{r}$ such that the $j$-th bit of the code-word is -1 and +1 respectively. Then one decoder sends to the other the extrinsic information

$$w_j^{out} = \lambda(d_j) - r_j \tag{4}$$

If the Chase search fails the extrinsic information is approximated as

$$w_j^{out} = \beta \cdot d_j \tag{5}$$

where $\beta$ is a weight factor increasing with the number of iterations.
The decoder that receives the extrinsic information uses an updated version of $\underline{r}$ obtained as

$$r_j^{new} = r_j^{old} + \alpha \cdot w_j^{in} \tag{6}$$

where $\alpha$ is a weight factor increasing with the number of iterations. A scheme of the elementary block turbo decoder is shown in Fig. 4 where the block named "decoder" is a Soft-In-Soft-out (SISO) module that performs the Chase search and implements (3), (4) and (5). An effective solution to implement the SISO module is based on a three pipelined stage architecture where the three stages are identified as reception, processing, and transmission units [Kerouedan & Adde, 2000]. As detailed in [LeBidan et al., 2008], during each stage, the $N$ soft values of the received word $r$ are processed sequentially in $N$ clock periods. The reception stage is devoted to find the least reliable bits in the received code-word. The processing stage performs the Chase search and the transmission stage calculates $\lambda(d_j)$, $w_j$ and $r_j^{new}$. Another solution is proposed in [Goubier et al. 2008] where the elementary decoder is implemented as a pipeline resorting to the mini-maxi algorithm, namely by using mini-maxi arrays to store the best metrics of all decoded code-words in the Chase list.

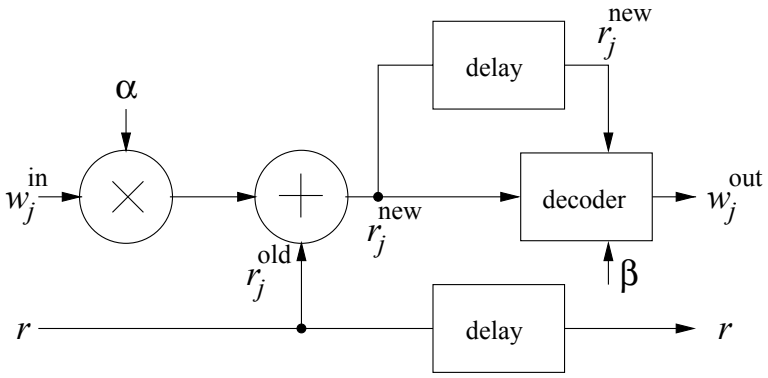



Fig. 4. Elementary block turbo decoder scheme

Several works in the literature deal with BTC complexity reduction. As an example [Adde & Pyndiah, 2000] suggests to compute $\beta$ in (5) on a per-code-word basis, whereas in [Chi et al.,

2004] the dependency on $\alpha$ in (6) is solved by replacing the term $\alpha \cdot w_j$ with $\tanh(w_j/2)$. In [Le et al. 2005] both $\alpha$ in (6) and $\beta$ in (5) are avoided by exploiting Euclidean distance property.
Due to its row-column structure, the block turbo decoder can be parallelized by instantiating several elementary decoders to concurrently process more rows or columns, thus increasing the throughput. As a significant example in [Jego et al., 2006] a fully parallel BTC decoder is proposed. This solution instantiates $n_1+n_2$ decoders that work concurrently. Moreover, by properly managing the scheduling of the decoders and interconnecting them through an Omega network intermediate results (row decoded data or column decoded data) are not stored.
A detailed analysis of throughput and complexity of BTC decoder architectures can be found in [Goubier et al. 2008] and [LeBidan et al., 2008]. In particular, according to [Goubier et al. 2008] a simple one block decoder architecture that performs the row/column decoding sequentially (interleaved architecture) requires $2(n_1+n_2)$ cycles to complete an iteration; as a consequence it achieves a throughput

$$T = \frac{k_1 \cdot k_2 \cdot f_{clk}}{I \cdot 2(n_1 + n_2)} \tag{7}$$

where $I$ is the number of iterations and $f_{clk}$ is the clock frequency. The BTC specified for WIMAX is obtained using twice a binary extended Hamming code out of the ones show in Table 1

| $N$ | $k$ |
|---|---|
| 15 | 11 |
| 31 | 26 |
| 63 | 57 |

Table 1. WIMAX binary extended Hamming codes ($H(n,k)$) used for BTC

Considering the interleaved architecture described in [Goubier et al. 2008] where a fully decoded block is output every 4.5 half iterations, we obtain that 75 Mb/s can be obtained with a clock frequency of 84 MHz, 31 MHz and 14 MHz for $H(15,11)$, $H(31,26)$ and $H(63,57)$ respectively.

### 2.3 CTC decoders

Convolutional turbo codes were proposed in 1993 by Berrou, Glavieux and Thitimajshima [Berrou et al., 1993] as a coding scheme based on the parallel concatenation of two CCs by the means of an interleaver (Π) as shown in Fig. 5 (a). The decoding algorithm is iterative and is based on the BCJR algorithm [Bahl et al., 1974] applied on the trellis representation of each constituent CC (Fig. 5 (b)). The key idea relies on the fact that the extrinsic information output by one CC is used as an updated version of the input a-priori information by the other CC. As a consequence, each iteration is made of two half iterations, in one half iteration the data are processed according to the interleaver ($\Pi$) and in the other half iteration according to the deinterleaver ($\Pi^{-1}$). The same result can be obtained by implementing an in-order read/write half iteration and a scrambled (interleaved) read/write half iteration. The basic block in a turbo decoder is a SISO module that

implements the BCJR algorithm in its logarithmic likelihood ratio (LLR) form. If we consider a Recursive Systematic CC (RSC code), the extrinsic information $\lambda_k(u;O)$ of an uncoded symbol $u$ at trellis step $k$ output by a SISO is

$$\lambda_k(u;O) = \max_{e:u(e)=u}{}^{*}\{b(e)\} - \max_{e:u(e)=\tilde{u}}{}^{*}\{b(e)\} - \lambda_k(u;I) - \pi_k(c^u;I) \tag{8}$$

where $\tilde{u}$ is an uncoded symbol taken as a reference (usually $\tilde{u}$=**0**), $e$ represents a certain transition on the trellis and $u(e)$ is the uncoded symbol $u$ associated to $e$. The max* function is usually implemented as a max followed by a correction term [Robertson et al., 1995]; [Gross & Gulak, 1998]; [Cheng & Ottosson, 2000]; [Classon et al., 2002]; [Wang et al., 2006]; [Talakoub et al. 2007]. A scaling factor can also be applied to further improve the max or max* approximation [Vogt & Finger, 2000]. The correction term, usually adopted when decoding binary codes, can be omitted for double binary turbo codes [Berrou et al. 2001] with minor error rate performance degradation. The term $b(e)$ in (8) is defined as

$$b(e) = \alpha_{k-1}[s^S(e)] + \gamma_k[e] + \beta_k[s^E(e)] \tag{9}$$

$$\alpha_k[s] = \max_{e:s^E(e)=s}\{\alpha_{k-1}[s^S(e)] + \gamma_k[e]\} \tag{10}$$

$$\beta_k[s] = \max_{e:s^S(e)=s}\{\beta_{k+1}[s^E(e)] + \gamma_k[e]\} \tag{11}$$

$$\gamma_k[e] = \pi_k[u(e);I] + \pi_k[c(e);I] \tag{12}$$

where $s^S(e)$ and $s^E(e)$ are the starting and the ending states of $e$, $\alpha_k[s^S(e)]$ and $\beta_k[s^E(e)]$ are the forward and backward state metrics associated to $s^S(e)$ and $s^E(e)$ respectively (see Fig. 5 (b)) and $\gamma_k[e]$ is the branch metric associated to $e$. The $\pi_k[c(e);I]$ term is computed as a weighted sum of the $\lambda_k[c;I]$ produced by the soft demodulator as

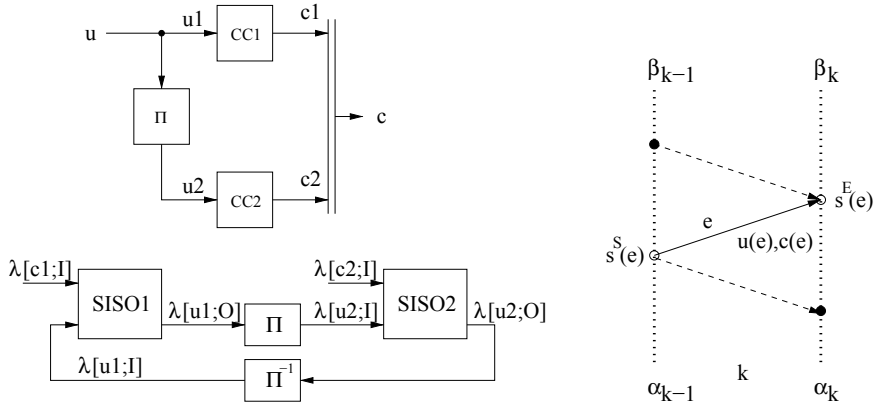
$$\pi_k[c(e);I] = \sum_{i}^{n_c} c_i(e)\lambda_k[c_i(e);I] \tag{13}$$

where $c_i(e)$ is one of the coded bits associated to $e$ and $n_c$ is the number of bits forming a coded symbol $c$ and $\pi_k[c^u(e);I]$ in (8) is obtained as $\pi_k[c(e);I]$ considering only the systematic bits corresponding to the uncoded symbol $u$ out of the $n_c$ coded bits. The $\pi_k[u(e);I]$ term is obtained combining the input a-priori information $\lambda_k(u;I)$ and for a double binary code can be written as in (14), where $A$ and $B$ represent the two bits forming an uncoded symbol $u$.

The CTC specified in the WIMAX standard is based on a double binary 8-state constituent CC as shown in Fig. 6, where each CC receives two uncoded bits ($A$, $B$) and produces four coded bits, two systematic bits ($A$,$B$) and two parity bits ($Y$,$W$). As a consequence, at each trellis step four transitions connect a starting state to four possible ending states. Due to the trellis symmetry only 16 branch metrics out of the possible 32 branch metrics are required at each trellis step. As pointed out in [Muller et al. 2006] high throughput can be achieved by

exploiting the trellis parallelism, namely computing concurrently all the branch and state metrics.

$$\pi_k(u;I) = \begin{cases} 0 & if \quad u(e) = (A = '0', B = '0') \\ \lambda_k^{\bar{A}\bar{B}} & if \quad u(e) = (A = '0', B = '1') \\ \lambda_k^{A\bar{B}} & if \quad u(e) = (A = '1', B = '0') \\ \lambda_k^{AB} & if \quad u(e) = (A = '1', B = '1') \end{cases} \tag{14}$$




(a) (b)

Fig. 5. Convolutional turbo code: coder and iterative SISO based decoder (a), notation for a trellis step in the SISO (b)

The 16 branch metrics are computed by a BMU that implements (12) as shown in Fig. 7. To reduce the latency of the SISO, usually the decoding is based on a sliding-window approach [Benedetto et al., 1996]. As a consequence, at least two BMUs are required to compute the two recursions (forward and backward) according to the BCJR algorithm. However, since β metrics require to be trained between successive windows, usually a further BMU is required. A solution based on the inheritance of the border metrics of each window [Abbasfar & Yao 2003] requires only two BMUs. Furthermore, this strategy reduces the SISO latency to the sliding window width *W*. The state metrics are updated according to (10) and (11) by two state metric processors, each of which is made of a proper number of processing elements (PE). As shown in Fig. 7 for the WIMAX CTC 8 PEs are required. It is worth pointing out that the constituent codes of the WIMAX CTC use the circulation state tailbiting strategy proposed in [Weiss et al. 2001] that ensures that the ending state of the last trellis step is equal to the starting state of the fist trellis step. However, this technique requires estimating the circulation state at the decoder side. Since training operations to estimate the circulation state would increase the SISO latency, an effective alternative [Zhan et al. 2006] is to inherit these metrics from the previous iteration.

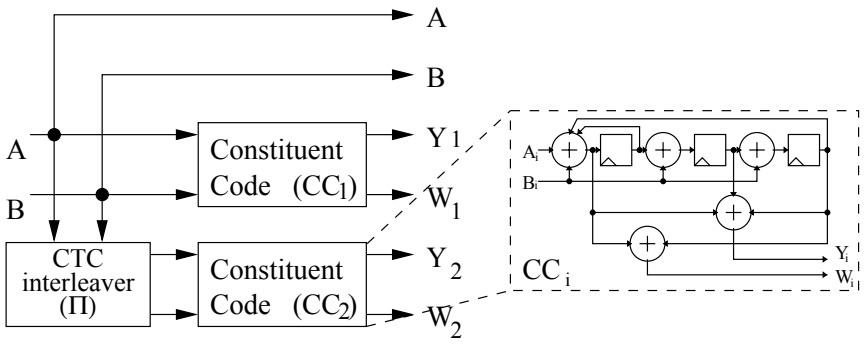



Fig. 6. WIMAX CTC: encoder and constituent CC structures

As in Viterbi decoder architectures often in CTC decoders the state metrics are computed by means of the "wrapping" representation technique proposed in [Hekstra, 1989]. This solution requires a normalization stage, depicted in Fig. 7, when combining $\alpha$, $\beta$ and $\gamma$ metrics to compute the extrinsic information as in (8). The last stage of the output processor, that computes the output extrinsic information, is a tree of max blocks for each component of the extrinsic information and few adders to implement (8). As highlighted in Fig. 7 this scheduling requires a buffer to store input LLRs that are used to compute the backward recursion (BMU-MEM). Since the output extrinsic information is computed during the backward recursion, forward recursion metrics are stored in a buffer (α-MEM). Further memory is required to implement the border metric inheritance, α-EXT-MEM, β-EXT-MEM and β-LOC-MEM.

The throughput sustained by the CTC decoder, defined as the number of decoded bits over the time required for their computation, is

$$T = \frac{k \cdot N_T \cdot f_{clk}}{2I \cdot (N_{cyc}^{SISO} + N_{cyc}^{ID})} = \frac{k \cdot N_T \cdot f_{clk}}{2I \cdot N_{cyc}^{dec}} \tag{15}$$

where $f_{clk}$ is the clock frequency, $N_T$ is the number of trellis steps, $k$=1 for a binary CTC, $k$=2 for a double binary CTC, $2I$ is the number of half iterations, $N_{cyc}^{SISO}$ and $N_{cyc}^{ID}$ represent the number of clock cycles required by one SISO and by the interleaving/deinterleaving structure. Since both $N_{cyc}^{SISO}$ and $N_{cyc}^{ID}$ are a function of $N_T$ they can be rewritten as $N_{cyc}^{SISO}=N_T \cdot SP+SISO_{cyc}^{lat}$ and $N_{cyc}^{ID}=N_T \cdot SP+ID_{cyc}^{oh}$ where $SP$ is the sending period, namely the rate sustained by the decoder to output two consecutive valid output data ($SP$=1 means at each clock cycle new valid output data are ready), $SISO_{cyc}^{lat}$ is the decoder latency, namely the number of clock cycles spent to produce the first valid output data, and $ID_{cyc}^{oh}$ is the interleaver/deinterleaver architecture overhead expressed in clock cycles. Usually, resorting to pipelining, $N_{cyc}^{SISO}$ and $N_{cyc}^{ID}$ can be partially overlapped; thus, the number of cycles required by one SISO decoder is $N_{cyc}^{dec}=N_T \cdot SP+SISO_{cyc}^{lat}+ID_{cyc}^{oh}$. Using the sliding window technique with the border metric inheritance strategy [Abbasfar & Yao 2003]; [Zhan et al. 2006] we obtain $SISO_{cyc}^{lat} \approx SP \cdot W$ and so (15) can be rewritten as (16), where the rightmost expression is obtained considering $W << N_T$ and $ID_{cyc}^{oh} << SP \cdot N_T$ that is a reasonable assumption in real cases.

$$T = \frac{k \cdot N_T \cdot f_{clk}}{2I \cdot [SP \cdot (N_T + W) + ID_{cyc}^{oh}]} \approx \frac{k \cdot f_{clk}}{2I \cdot SP} \tag{16}$$

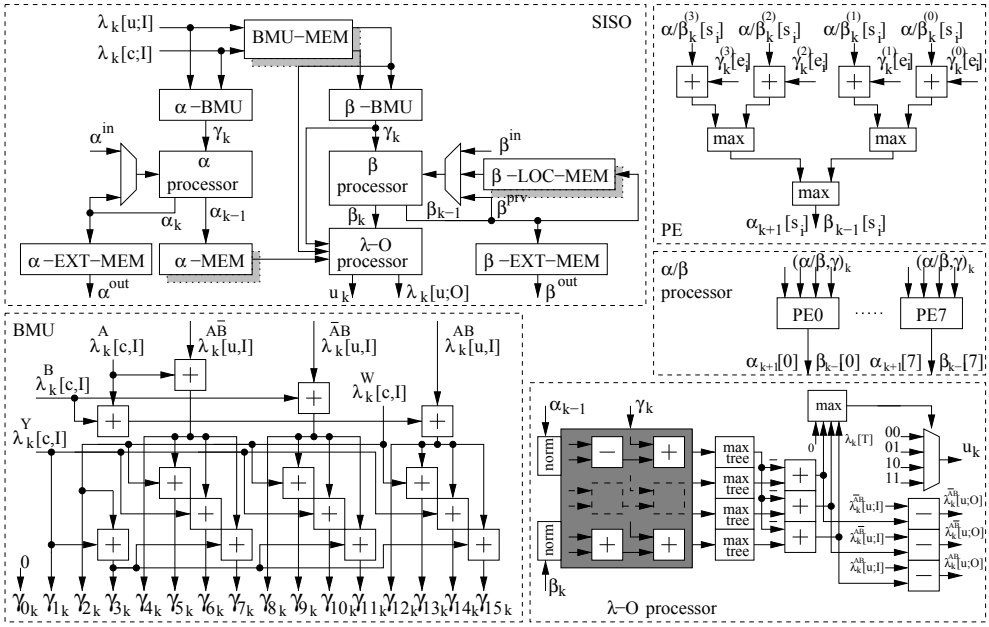



Fig. 7. WIMAX SISO block scheme

Usually optimized architectures [Masera et al., 1999]; [Bickerstaff et al., 2003]; [Kim & Park, 2008] are obtained with *SP*=1, whereas flexible architectures have higher *SP* values [Vogt & Wehn, 2008]; [Muller et al., 2009]. However, even with *SP*=1, a double binary turbo decoder architecture that achieves the throughput imposed by WIMAX with eight iterations (*I*=8), would require $f_{clk}$=600 MHz. A possible solution to improve the throughput by a factor that ranges in [1.2, 1.9] is the based on decoder level parallelism [Muller et al. 2006] and is usually referred to as "shuffling" [Zhang & Fossorier, 2005]. However, to further improve the throughput a parallel decoder made of *P* SISOs working concurrently is required. As a consequence, a parallel architecture achieves a throughput

$$T = \frac{k \cdot N_T \cdot f_{clk}}{2I \cdot [SP \cdot (N_T / P + W) + ID_{cyc}^{oh}]} \approx \frac{k \cdot P \cdot f_{clk}}{2I \cdot SP} \tag{17}$$

Thus, setting *P*=4, *I*=8 and *SP*=1, the WIMAX throughput is obtained with $f_{clk}$=150 MHz. It is worth pointing out that a *P*-parallel CTC decoder is made of *P* SISOs connected to *P* memories devoted to store the extrinsic information. However, in a parallel decoder during the scrambled half iteration collisions can occur, namely more SISOs could need to access

the same memory during the same cycle. Since the collision phenomenon increases $ID_{cyc}^{oh}$, several algorithmic approaches to design collision free interleavers [Giulietti et al. 2002]; [Kwak & Lee, 2002]; [Gnaedig et al., 2003]; [Tarable et al., 2004] have been proposed. On the other hand, architectures to manage collisions in a parallel turbo decoder have also been proposed in the literature [Thul et al., 2002]; [Gilbert et al., 2003]; [Thul et al., 2003]; [Speziali & Zory, 2004]; [Martina et al. 2008-a]; [Martina et al., 2008-b], in particular [Martina et al. 2008-b] deals with the parallelization of the WIMAX CTC interleaver and avoids collision by the means of a throughput/parallelism scalable architecture that features $ID_{cyc}^{oh}$=0.
It is worth pointing out that parallel architectures increase not only the throughput but also the complexity of the decoder, so that some recent works aim at reducing the amount of memory required to implement SISO local buffers. In [Liu et al., 2007] and [Kim & Park, 2008] saturation of forward state metrics and quantization of border backward state metrics is proposed. Further studies have been performed to reduce the extrinsic information bit width by using adaptive quantization [Singh et al., 2008], pseudo-floating point representation [Park et al., 2008] and bit level representation [Kim & Park, 2009].

### 2.4 LDPC code decoders

LDPC codes were originally introduced in 1962 by Gallager [Gallager, 1962] and rediscovered in 1996 by MacKay and Neal [MacKay, 1996]. As turbo codes, they achieve near optimum error correction performance and are decoded by means of high complexity iterative algorithms.
An LDPC code is a linear block code defined by a $C \times B$ parity check matrix $H$, characterized by a low density of ones: $B$ is the number of bits in the code (block length), while $C$ is the number of parity checks. A one in a given cell of the $H$ matrix indicates that the bit corresponding to the cell column is used for the calculation of the parity check associated to the row. A popular description of an LDPC code is the bipartite (or Tanner) graph shown in Figure 8 for a small example, where $B$ variable nodes (VN) are connected to C check nodes (CN) through edges corresponding to the positions of the ones in $H$.
LDPC codes are usually decoded by means of an iterative algorithm variously known as sum-product, belief propagation or message passing, and reformulated in a version that processes logarithmic likelihood ratios instead of probabilities. In the first iteration, half variable nodes receive data from adjacent check nodes and from the channel and use them to obtain updated information sent to the check nodes; in the second half, check nodes take the updated information received from connected bit nodes and generate new messages to be sent back to variable nodes.
In message passing decoders, messages are exchanged along the edges of the Tanner graph, and computations are performed at the nodes. To avoid multiplications and divisions, the decoder usually works in the logarithmic domain.

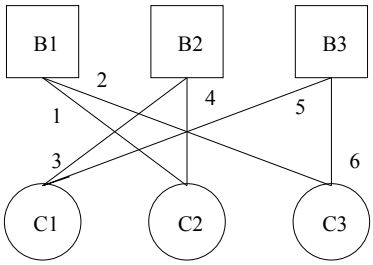



Fig. 8. Example Tanner graph

The message passing algorithm is described in the following equations, where $k$ represents the current iteration, $Q_{ji}$ is the message generated by VN $j$ and directed to CN $i$, $R_{ij}$ is the message computed by CN $i$ and sent to VN $j$. $C[j]$ is the whole set of incoming messages for VN $j$ and $R[i]$ is the whole set of the incoming messages for CN $i$.
Each variable node is initialized with the log-likelihood ratio (LLR) $\lambda_j$ associated to the received bit. Next, messages are propagated from the variable nodes to the check nodes along the edges of the Tanner graph. At the first iteration, only $\lambda_j$ are delivered, while starting from the second iteration VNs sum up all the messages $R_{ij}$ coming from CNs and combine them with $\lambda_j$ according to

$$Q_{ji}^{k} = \lambda_j + \sum_{\alpha \in C[j]/i} R_{\alpha j}^{k-1} \tag{18}$$

The check node computes new check to variable messages as

$$R_{ij}^{k} = \Psi^{-1}\left[\sum_{\beta \in R[i]/j} \Psi\left[Q_{\beta i}^{k}\right]\right] \cdot \delta_{ij} \qquad \text{with} \qquad \delta_{ij} = (-1)^{|R[i]|} \prod_{j} \operatorname{sgn}\left(Q_{ji}\right) \tag{19}$$

where $|R[j]|$ is the cardinality of the CN and $\Psi(x)$ is a non linear function defined as

$$\Psi[x] = -\ln\left(\tanh\left|\frac{x}{2}\right|\right) \tag{20}$$

After a number of iterations that strongly depends on the addressed application and code rate (typically 5 to 40), variable nodes compute an overall estimation of the decoded bit in the form

$$\Lambda_j^{k} = \lambda_j + \sum_{\alpha \in C[j]} R_{\alpha j}^{k-1} \tag{21}$$

where the sign of $\Lambda_j$ can be understood as the hard decision on the decoded bit.

A large implementation complexity is associated to (19), which is simplified in different ways. First of all, function $\Psi(x)$ can be obtained by means of reduced complexity estimations [Masera et al., 2005]. Moreover sub-optimal, low complexity algorithms have been successfully proposed to simplify (19), such as for example the normalized Min-Sum algorithm [Chen et al., 2005] where only the two smallest magnitudes are used.
A further change is usually applied to the scheduling of variable and check nodes in order to improve communications performance. In the two-phase scheduling, the updating of variable and check nodes is accomplished in two separate phases. On the contrary, the turbo decoding message passing (TDMP) [Mansour & Shanbhag, 2003], also known as layered or shuffled decoding, allows for overlapped update operations: messages calculated by a subset of check nodes are immediately used to update variable nodes. This scheduling has been proved to be able to reduce the number of iterations by up to 50% at a fixed communications performance.
The required number of functional units in a decoder can be estimated based on the concept of processing power $P_c$ [Gouillod et al., 2007], which can be evaluated on the basis of the rate $R_c$ of the code, the number $K$ of information bits transmitted per codeword, the block size $N=K/R_c$, the required information throughput $D$, the operating clock frequency $f_{clk}$, the maximum number of iterations $i_{MAX}$ and the total number of edges to be processed per iteration $\varepsilon$. This relation is expressed as

$$P_c = \frac{\varepsilon \cdot D \cdot i_{MAX}}{K \cdot f_{clk}} \tag{22}$$

As two messages are associated with each edge (to be sent from the CN to the VN and vice versa), $2P_c$ gives the number of messages that must be concurrently processed at each decoding iteration in order to achieve the target throughput $D$. Equation (22) does not consider the message exchange overhead: yet it assumes that all messages dispatched during a cycle are delivered simultaneously during the same cycle. The $P_c$ value must then be assumed as a lower bound and the actual degree of parallelism strongly depends on both the structure of the $H$ matrix [Dinoi et al., 2006] and the adopted interconnect architecture among processing units [Quaglio et al., 2006] [Masera et al., 2007].
Actually, most of the implementation concerns come from the communication structure that must be allocated to support message passing from bit to check nodes and vice versa. Several hardware realizations that have been proposed in the literature are focused on how efficiently passing messages between the two types of processing units.
Three approaches can be followed in the high level organization of the decoder, coming to three kinds of architectures.

- Serial architectures: bit and check processors are allocated as single instances, each serving multiple nodes sequentially; messages are exchanged by means of a memory.
- Fully parallel architectures: processing units are allocated for each single bit and check node and all messages are passed in parallel on dedicated routes.
- Partially parallel architectures: more processing units work in parallel, serving all bit and check nodes within a number of cycles; suitable organization and hardware support is required to exchange messages.

For most codes and applications, the first approach results in slow implementations, while the second one has an excessive cost. As a result the only general viable solution is the third partially parallel approach, which on the other hand introduces the collision problem, already known in the implementation of parallel turbo decoders. Two main approaches have been proposed to deal with collisions:

- To design collision free codes [Mansour & Shanbhag , 2003], [Hocevar, 2003],
- To design decoder architecture able to avoid or at least mitigate collision effects [Kienle et al., 2003], [Tarable et al., 2004].

Even if the first approach has proven to be effective, it significantly limits the supported code classes. The second approach, on the other hand, is well suited for flexible and general architectures. An even more challenging task is the design of LDPC decoders that are flexible in terms of supported block sizes and code rates [Masera et al., 2007].

In partially parallel structures, permutation networks are used to establish the correct connections between functional units. However, structured LDPC codes, such as those specified in WIMAX, allow for replacing permutation networks by low complexity barrel shifters [Boutillon et al., 2000]; [Mansour & Shanbhag, 2003].

Early terminal schemes can be adopted to improve the decoding efficiency by dynamically adjusting the iteration number according to the SNR values. The simplest approach requires that decoding decisions are stored and compared across two consecutive iterations: if no changes are detected, the decoding is terminated, otherwise it is continued up to a maximum number of iterations. More sophisticated iteration control schemes are able to reduce the mean number of iterations, so saving both latency and energy [Kienle & When, 2005]; [Shin et al., 2007].

## 3. Case of study: complete WIMAX CTC decoder design

The WIMAX CTC decoder is made of three main blocks: symbol deselection (SD), subblock deinterleaver and CTC decoder as highlighted in Fig. 9 where $N$ represents the number of couples included in a data frame. SD, subblock deinterleaver and CTC decoder blocks are connected together by means of memory buffers in order to guarantee that the non iterative part of the decoder (namely SD and subblock deinterleaver) and the decoding loop work simultaneously on consecutive data frames. Since the maximum decoder throughput is about 75 Mb/s and the native CTC rate is 1/3 (two uncoded bits produce six coded bits), at the input of the decoding loop the maximum throughput can rise up to 225 millions of LLRs per second. The same throughput ought to be sustained by the subblock deinterleaver, whereas even higher throughput has to be sustained at the SD unit in case of repetition.

### 3.1 Symbol deselection

Depending on amount of data sent by the encoder (puncturing or repetition), the throughput sustained by the symbol deselection (SD) can rise up to 900 millions of LLRs per second (repetition ×4). When the encoder performs repetition, the same symbol is sent more than once. Thus, the decoder combines the LLRs referred to the same symbol to improve the reliability of that symbol. As shown in Fig. 9 this can be achieved partitioning the symbol deselection input buffer into four memories, each of which containing up to $6N$ LLRs.

Since the symbol deselection architecture can read up to four LLRs per clock cycle, it reduces the incoming throughput to 225 millions of LLRs per second. However, the symbol

deselection has to compute the starting location and the number of LLRs to be written into the output buffer. The number of LLRs and the starting location are obtained as in (23) and (24) respectively, where $N_{SCHk}$, $m_k$ and $SPID_k$ are parameters specified by the WIMAX standard for the $k$-index subpacket when HARQ is enabled, namely $N_{SCHk}$, is the number of concatenated slots, $m_k$ is the modulation order and $SPID_k$ is the subpacket ID.

$$L_k = 48 \cdot m_k \cdot N_{SCHk} \tag{23}$$

$$F_k = (SPID_k \cdot L_k) \bmod 6N \tag{24}$$

Since $N_{SCHk} \in [1, 480]$ and $m_k \in \{2, 4, 6\}$ we can rewrite (23) as

$$L_k = \begin{cases} (2N_{SCHk} + N_{SCHk}) \cdot 2^5 & when \quad m_k = 2 \\ (2N_{SCHk} + N_{SCHk}) \cdot 2^6 & when \quad m_k = 4 \\ (8N_{SCHk} + N_{SCHk}) \cdot 2^5 & when \quad m_k = 6 \end{cases} \tag{25}$$

The efficient implementation of (25) is obtained with an adder whose inputs are $N_{SCHk}$ and the selection between two hardwired left shifted versions of $N_{SCHk}$ (one position and three positions), followed by a programmable left shifter (five-six positions). Similarly, since $SPID_k \in \{0, 1, 2, 3\}$, the multiplication in (24) is avoided as

$$F_k = \begin{cases} 0 & when \quad SPID_k = 0 \\ L_k \bmod 6N & when \quad SPID_k = 1 \\ 2L_k \bmod 6N & when \quad SPID_k = 2 \\ (2L_k + L_k) \bmod 6N & when \quad SPID_k = 3 \end{cases} \tag{26}$$

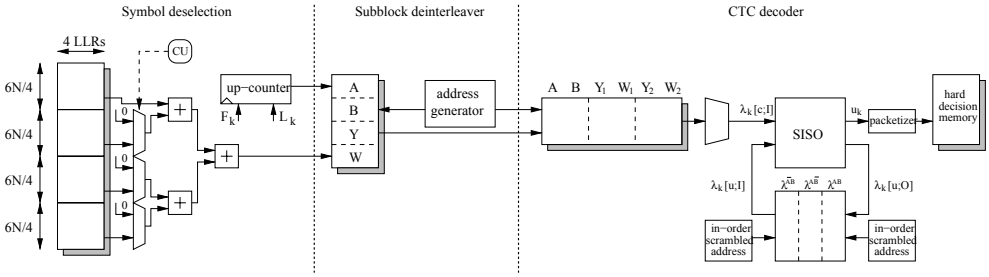



Fig. 9. Complete CTC decoder block scheme

A block scheme of the architecture employed to compute $F_k$ and $L_k$ is depicted in Fig. 10 (a). Furthermore, in order to support the puncturing mode, the output memory locations corresponding to unsent bits must be set to zero. To ease the SD architecture implementation, all the output memory locations are set to zero while $L_k$ and $F_k$ are

computed. As a consequence, about two clock cycles per sample are required to complete the symbol deselection, namely 6N LLRs are output in 12N clock cycles. So that the symbol deselection throughput can be estimated as

$$T_{SD} = \frac{6N}{12N} f_{clk} = \frac{f_{clk}}{2} \tag{27}$$

As it can be observed, to sustain 225 millions of LLRs per second a clock frequency of 450 MHz is required. To overcome this problem we impose not only to partition the input buffer into four memories, but also to increase the memory parallelism, so that each memory location contains $p$ LLRs. Thus, we can rewrite (27) as (28) and by setting $p$ to a conservative value, as $p$=4, the SD architecture processes simultaneously up to sixteen LLRs with $f_{clk}$=113 MHz.

$$T_{SD} = \frac{6N}{12N/p} f_{clk} = \frac{p \cdot f_{clk}}{2} \tag{28}$$

### 3.2 Subblock deinterleaver

The received LLRs belong to six possible subblocks depending on the coded bits they are referred to ($A$, $B$, $Y_1$, $W_1$, $Y_2$, $W_2$) and each subblock is made of $N$ LLRs. The subblock deinterleaver treats each subblock separately and scrambles its LLRs according to Algorithm 1, given below, where $m$ and $J$ are constants specified by the WIMAX standard and $BRO_m(y)$ is the bit-reversed $m$-bit value of $y$.

```
1: k←0
2: i←0
3: while i<N do
4:    T_k←2^m(k mod J)+BRO_m(⌊k/J⌋)
5:    if T_k<N then
6:       i←i+1
7:    else
8:       discard T_k
9:    end if
10:   k←k+1
11: end while
```

Algorithm 1. Subblock deinterleaver address generator

As a consequence, the number of tentative addresses generated, $N_M$, can be greater than $N$. Exhaustive simulations, performed on the possible $N$ specified by the standard, show that the worst case is $N_M$=191 that occurs with $N$=144. Since 191/144=1.326, a conservative approximation is $N_M$=4$N$/3. The whole subblock deinterleaver architecture is obtained with one single address generator implementing Algorithm 1 to simultaneously write one LLR from each of the six subblock memories. In particular, as imposed by the WiMax standard, the interleaved LLRs belonging to the A and B subblocks are stored separately, whereas the

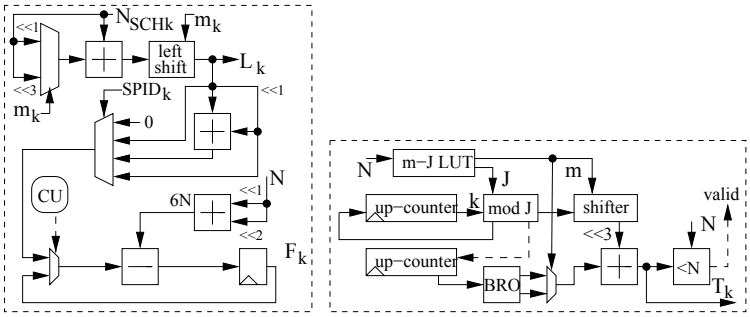
interleaved LLRs belonging to $Y_1$ and $Y_2$ are stored as a symbol-by-symbol multiplexed sequence, creating a "macro-subblock" made of 2N LLRs. Similarly a macro-subblock made of 2N LLRs is generated storing a symbol-by-symbol multiplexed sequence of interleaved $W_1$ and $W_2$ subblocks.
Since all the subblocks can be processed simultaneously, this architecture deinterleaves six LLRs per clock cycle. As a consequence, the subblock deinterleaver sustains a throughput

$$T_{SubDein} = \frac{6N}{4N/3} f_{clk} = 4.5 f_{clk} \tag{29}$$

Thus, a throughput of 225 Millions of LLRs per second is sustained using $f_{clk}$=50 MHz.
To implement line 4 and 5 in Algorithm 1, three steps are required, namely the calculation of $k$ mod $J$ and $\lfloor k/J \rfloor$, the calculation of $2^m(k \bmod J)$ and $BRO_m(\lfloor k/J \rfloor)$, the generation of $T_k$ while checking $T_k$<$N$. It is worth pointing out that $k$ mod $J$ can be efficiently implemented as an up-counter followed by a mod $J$ block. Moreover, each time the mod $J$ block detects $k$=$J$, a second counter is incremented: the final value in the second counter is $\lfloor k/J \rfloor$. Since $m \in [3, 10]$, the $2^m(k \bmod J)$ term is implemented as a programmable shifter in the range [0, 7] followed by a hardwired three position left shifter. The $BRO_m(\lfloor k/J \rfloor)$ term is obtained by multiplexing eight hardwired bit reversal networks. Finally, a valid $T_k$ address is obtained with an adder and is validated by a comparator. The address generation architecture is shown in Fig. 10 (b).

### 3.3 CTC decoder

As detailed in section 2.3 to sustain the throughput required by the WIMAX standard a parallel decoder architecture is required. To that purpose we set $SP$=1, $I$=8, and $f_{clk}$=200 MHz, then from (17) we analyze the throughput as a function of $N$ for $W$=32. As shown in Fig. 11, only $P$=4 allows to achieve the target throughput (horizontal solid line) for $N \geq 480$.
Moreover, the window width impacts both on the decoder throughput and on the depth of SISO local buffers. So that a proper $W$ value for each frame size must be selected. In particular if $N/(P \cdot W) \in \mathbb{N}$ SISOs synchronization is simplified. However, the choice of $P$ should minimize collisions in memory access.
Exhaustive simulations show that collisions occur for $P$=2 and $P$=4 only with $N$=108. As a consequence, we select $P$ as a function of $N$ to simultaneously obtain a monotonically increasing throughput as a function of $N$ and to avoid collisions. It is worth pointing out that, when collisions are avoided, the resulting parallel interleaver is a circular shifting interleaver: the address generation is simplified with all SISOs simultaneously accessing the same location of different memories.
Said $idx^0_t$ the memory accessed by SISO-*0* at time $t$ during a scrambled half iteration, the memory concurrently accessed by SISO-$k$ is $idx^k_t = (idx^0_t \pm k) \bmod P$.




(a) (b)

Fig. 10. Symbol deselection starting address and number of elements generation block scheme (a), subblock deinterleaver address generation block scheme (b).

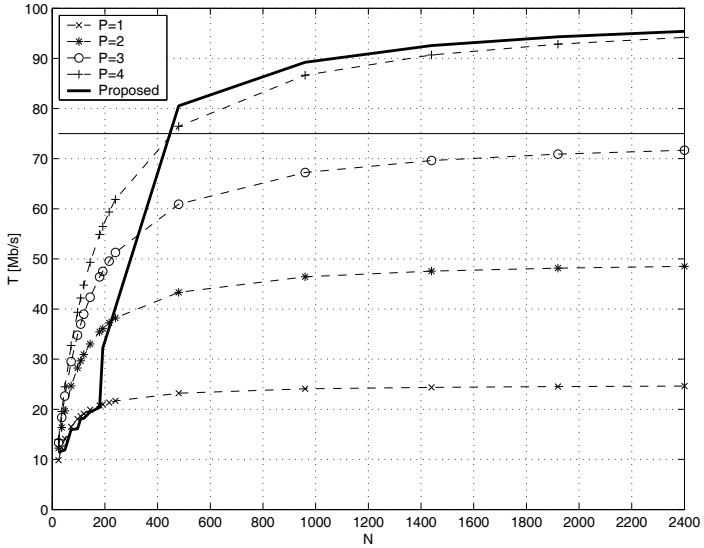



Fig. 11. Parallel CTC decoder throughput as a function of the block size N for different parallelism degree values P. The horizontal line represents the target throughput.

Thus, the parallel CTC interleaver-deinterleaver system is obtained as a cascaded two stage architecture (see Fig. 12). The first stage efficiently implements the WIMAX interleaver algorithm, whereas the second one extracts the common memory address $adx_t$ and the memory identifiers $idx^k_t$ from the scrambled address *i*.

The CTC interleaver algorithm specified in the WIMAX standard is structured in two steps. The first step switches the LLRs referred to *A* and *B* that are stored at odd addresses. The second step provides the interleaved address *i* of the *j*-th couple as

$$i = (P_0 \cdot j + P_j^{'}) \bmod N \quad j = 0, \cdots, N-1 \tag{30}$$

where $P_0$ and $P_j'$ are constants that depend only on $N$ and are specified by the standard. It is worth pointing out that the two steps can be swapped, as a consequence the first step can be performed on-the-fly, avoiding the use of an intermediate buffer to store switched LLRs. A simple architecture to implement (30) can be derived by rewriting (30) as

$$i = \left[ i_j' + \left( P_j' \bmod N \right) \right] \bmod N \tag{31}$$

where

$$i_j' = \begin{cases} i_0' = 0 & when \quad j = 0 \\ \left( i_{j-1}' + P_0 \bmod N \right) \bmod N & when \quad j = 1, \cdots, N-1 \end{cases} \tag{32}$$

A small Look-Up-Table (LUT) is employed to store $P_0$ mod $N$ and $P_j'$ mod $N$ terms; then (31) is implemented by two parts as depicted in Fig. 12. The first part accumulates $P_0$ to implement the $P_0 \cdot j$ term and the mod $N$ block produces the correct modulo $N$ result. The second part employs the two least significant bits of a counter ($j$−cnt) to select the proper $P_j'$ mod $N$ value, which is added to the ($P_0 \cdot j$) mod $N$ term. A further modulo $N$ operation is performed at the output. Since in this architecture both the first and the second part work on data belonging to [0, 2$N$−1], all the mod $N$ operations are implemented by means of a subtracter and a multiplexer.

The second stage of the parallel CTC interleaver-deinterleaver architecture works as follows. Since $adx_t \in [0, N/P\text{-}1]$, it can be obtained from the scrambled address $i$ produced by the first stage as

$$adx_t = \begin{cases} i & when \quad i \in \left[ 0, \dfrac{N}{P} - 1 \right] \\ i - \dfrac{N}{P} & when \quad i \in \left[ \dfrac{N}{P}, \dfrac{2N}{P} - 1 \right] \\ \cdots & \cdots \quad \cdots \\ i - (P-1)\dfrac{N}{P} & when \quad i \in \left[ (P-1)\dfrac{N}{P}, N-1 \right] \end{cases} \tag{33}$$

The straightforward implementation of (33) needs to calculate $N/P$ and to allocate $P$−2 multipliers, $P$−1 subtracters, a $P$-way multiplexer and few logic for selecting the proper $adx_t$ value. The $N/P$ division can be simplified by choosing the possible $P$ values as powers of two. Thus, we obtain a CTC decoder architecture that exploits throughput/parallelism scalability to avoid collisions, namely we employ: $P$=1 when $N \leq 180$, $P$=2 when $192 \leq N \leq 240$ and $P$=4 when $480 \leq N \leq 2400$. Moreover, as it can be inferred from Fig. 12, multiplications are avoided resorting to simple shift operations ($x >> i = x/2^i$). The sign of the subtractions (dashed lines in Fig. 12) allows not only to select the proper $adx_t$ but also to find $idx^0{}_t$ . Then, with $P$−1 modulo $P$ adders the other $idx^k{}_t$ values are straightforwardly generated. As it can be

observed, choosing $P$ as a power of two reduces the modulo $P$ adders to simpler, binary adders. The actual throughput sustained by the described throughput/parallelism scalable architecture is represented by the bold line in Fig. 11.

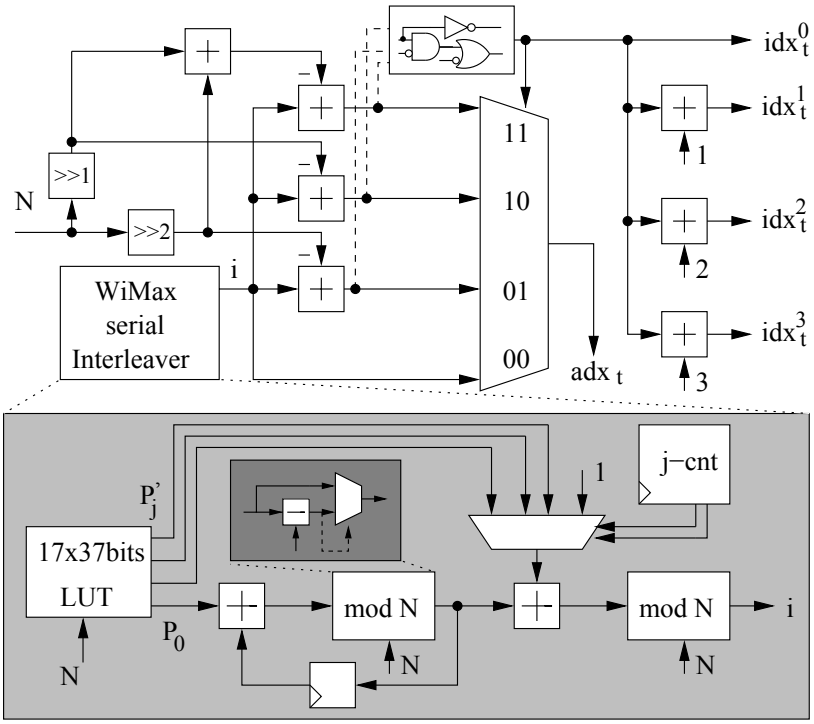



Fig. 12. Parallel CTC address generator

The global architecture of the designed parallel SISO is given in Fig. 13 where each SISO contains the processors devoted to compute the different metrics required by the BCJR algorithm as detailed in section 2.3. A simple network is used to properly connect the SISOs according to the current value of $P$ by setting the signal last_SISO. Furthermore, one address crossbar-switch (radx-switch) is used to implement the reading operation, a LIFO stores the address and makes them available for the writing phase, two data crossbar-switches (rdata-switch/wdata-switch) are used to properly send (receive) the data to (from) the memory (EI-MEM) according to the parallel interleaver $idx^k_t$ values.

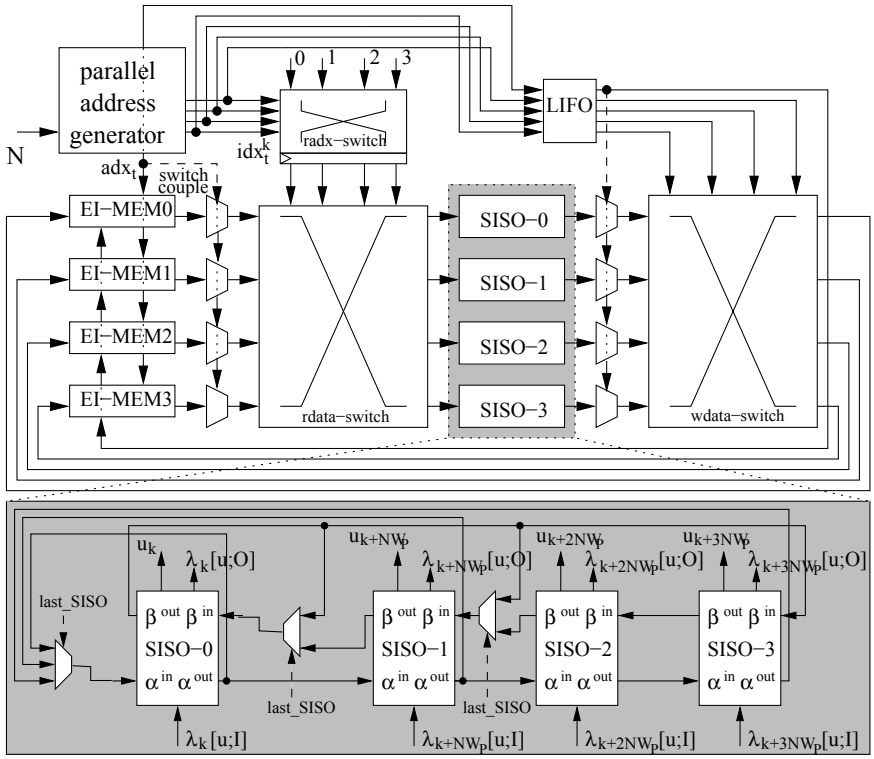



Fig. 13. Parallel CTC decoder architecture

In Table 2 the complexity of all the blocks for a 130 nm standard cell technology is reported. The bit-width is: 6 bit for *λ[c;I]*, 8 bit for *λ[u;I]*, and 12 bit for the state metrics. For further details the reader can refer to [Martina et al., 2009].

| *Architecture* | *SD* | *Subblock Deinterl.* | *SISOx1* | *Parallel Interl.* |
|---|---|---|---|---|
| Logic [kgate] | 11 | 1.7 | 37 | 2.8 |
| Memory [kbit] | 0 | 0 | 14.2 | 59 |

Table 2. Complexity of the whole receiver

## 4. Acknowledgements

This work is partially supported by the WIMAGIC project funded by the European Community.

## 5. References


Abbasfar, A. ; Yao, K. (2003). An Efficient and Practival Architecture for High Speed Turbo Decoders, *Proceedings of the IEEE Vehicular Technology Conference*, pp 337-341, Orlando, USA, Oct. 2003

Adde, P & Pyndiah, R. (2000). Recent Simplifications and Improvements in Block Turbo Codes, *Proceedings of the 2nd International Symposium on Turbo Codes & Related Topics*, pp. 133-136, Brest, France, Sep. 2000

Bahl, L. R.; Cocke, J.; Jelinek, F. & Raviv, J. (1974). Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate, *IEEE Trans. on Information Theory*, Vol. 20, No. 2, Mar. 1974, pp. 284-287

Batcha, M. F. N. & Shameri, A. Z. (2007). Configurable, Adaptive Viterbi Decoder for GPRS, EDGA and WIMAX, *Proceedings of the IEEE International Conference on Telecommunications and Malaysia International Conference on Communications*, pp. 237-241, Pennang, Malaysia, May 2007

Benedetto, S.; Montorsi G.; Divsalar, D. & Pollara, F. (1996). Algorithm for Continuous Decoding of Turbo Codes, *IET Electronics Letters*, Vol. 32, No. 4, Apr. 1996, pp 314-315

Berrou, C.; Glavieux, A. & Thitimajshima P. (1993). Near Shannon Limit Error-Correcting Codes : Turbo Codes, *Proceedings of the IEEE International Conference on Communications*, pp. 1064-1070, Geneva, Switzerland, May 1993

Berrou, C.; Jezequel, M.; Douillard, C. & Kerouedan S. (2001). The Advantages of non-binary turbo codes, *Proceedings of the IEEE Information Theory Workshop*, pp. 61-63, Cairns, Australia, Sep. 2001

Bickerstaff, M.; Davis, L.; Thomas, C.; Garret, D. & Nicol, C. (2003). A 24 Mb/s Radix/4 LogMAP Turbo Decoder for 3GPP-HSDPA Mobile Wireless, *Proceedings of the IEEE International Solid State Circuits Conference,* Section 8 – paper 8.5, San Francisco, USA, Feb. 2003

Black, P. J. & Meng, T. H. (1992). A 140-Mb/s, 32-State, Radix4 Viterbi Decoder, *IEEE Journal of Solid State Circuits,* Vol. 27, No. 12, Dec. 1992, pp. 1877-1885

Boutillon, E.; Castura, J. & Kschischang, F. (2000). Decoder-first code design, *Proceedings of the 2nd International Symposium on Turbo Codes & Related Topics*, pp. 459–462, Brest, France, Sep. 2000

Chase, D. (1972). A Class of Algorithms for Decoding Block Codes with Channel Measurement Information, *IEEE Trans. on Information Theory*, Vol. IT-19, No. 1, Dec. 1972, pp.170-182

Chen, J.; Dholakia, A.; Eleftheriou, E.; Fossorier, M. P. C. & Hu, X. Y. (2005). Reduced-Complexity Decoding of LDPC Codes, *IEEE Trans. on Communications*, Vol. 53, No. 8, Aug. 2005, pp. 1288–1299

Cheng, C. & Parhi, K. K. (2008). Hardware Efficient Low-Latency Architecture for High-Throughput Rate Viterbi Decoders, *IEEE Trans. on Circuits and Sytems II,* Vol. 55, No. 12, Dec. 2008, pp. 1254-1258

Cheng, J. F. & Ottosson, T. (2000). Linearly Approximated Log-MAP Algorithm for Turbo Decoding, *Proceedings of the IEEE Vehicular Technology Conference*, Tokio, Japan, May 2000, pp. 2252-2256

Chi, Z.; Song, L. & Parhi, K. K. (2004). On the Performance/Complexity Tradeoff in Block Turbo Decoder Design, *IEEE Trans. on Communications*, Vol. 52, No. 2, Feb. 2004, pp. 173-175

Classon, B.; Blankenship, K. & Desai, V. (2002). Channel Coding for 4G Systems with Adaptive Modulation and Coding, *IEEE Wireless Communications Magazine*, Vol. 9, No. 2, Apr. 2002, pp. 8-13

Dinoi, L.; Martini, R.; Masera, G.; Quaglio, F. & Vacca, F. (2006). ASIP design for partially structured LDPC codes, *Electronics Letters,* Vol. 42, No. 18, Aug. 2006, pp.1048 - 1049

Fettweis, G. & Meyr, H. (1989). Parallel Viterbi algorithm implementation: Breaking the ACS-bottleneck, *IEEE Trans. on Communications,* Vol. 37, No. 8, Aug. 1989, pp. 785-790

Gallager R. G. (1962), Low-Density Parity-Check Codes, *IRE Trans. on Information Theory*, Vol. 8, No. 1, Jan. 1962, pp. 21–28

Gilbert, F.; Thul, M. J. & Wehn, N. (2003). Communication Centric Architectures for Turbo-Decoding on Embedded Multiprocessors, *Proceedings of Design Automation and Test in Europe Conference and Exhibition*, pp. 356-361, Munich, Germany, Mar. 2003

Giulietti, A.; van der Perre, L. & Strum, M. (2002). Parallel Turbo Coding Interleavers: Avoiding Collisions in Accesses to Storage Elements, *IET Electronics Letters*, Vol. 38, No. 5, Feb. 2002, pp. 232-234

Gnaedig, D.; Boutillon, E.; Jezequel, M.; Gaudet. V. C. & Gulak, P. G. (2003). Multiple Slice Turbo Codes, *Proceedings of the 3rd International Symposium on Turbo Codes & Related Topics*, pp. 343-346, Brest, France, Sep. 2003

Goubier, T.; Dezan, C.; Pottier, B. & Jego, C. (2008). Fine Grain Parallel Decoding of Product Turbo Codes: Algorithm and Architecture, *Proceedings of the 5th International Symposium on Turbo Codes & Related Topics*, pp. 90-95, Lausanne, Switzerland, Sep. 2008

Gross, W. J. & Gulak, P.G. (1998). Simplified MAP Algorithm Suitable for Implementation of Turbo Decoders, *IET Electronics Letters*, Vol. 34, No. 16, Aug. 1998, pp. 1577-1578

Guilloud, F.; Boutillon, E.; Tousch, J. & Danger, J. (2007). Generic description and synthesis of LDPC decoders, *IEEE Trans. on Communications*, Vol. 55, No. 11, Nov. 2007, pp. 2084-2091

Hekstra, A. P. (1989). An Alternative to Metric Rescaling in Viterbi Decoders, *IEEE Trans. on Communications,* Vol. 37, No. 11, Nov. 1989, pp. 1220-1222

Hocevar, D. E. (2003). LDPC code construction with flexible hardware implementation, *Proceedings of the IEEE International Conference on Communications*, pp. 2708-2712, Anchorage, USA, May 2003

Jego, C.; Adde, P. & Leroux, C. (2006). Full-Parallel Architecture for Turbo Decoding of Product Codes, *IET Electronics Letters*, Vol. 42, No. 18, Aug. 2006, pp. 1052-1053

Kamuf, M.; Owall, V. & Anderson, J. B. (2008). Optimization and Implementation of a Viterbi Decoder Under Flexibility Constraints, *IEEE Trans. on Circuits and Systems I,* Vol. 55, No. 9, Sep. 2008, pp. 2411-2422

Kerouedan, S. & Adde, P. (2000). Implementation of a Block Turbo Decoder on a Single Chip, *Proceedings of the 2nd International Symposium on Turbo Codes & Related Topics*, pp. 133-136, Brest, France, Sep. 2000

Kienle, F.; Thul, M. J. & Wehn, N. (2003). Implementation issue of scalable LDPC-decoders, *Proceedings of the 3rd International Symposium on Turbo Codes & Related Topics*, pp. 291-294, Brest, France, Sep. 2003

Kienle, F. & Wehn, N. (2005). Low Complexity Stopping Criterion for LDPC Code Decoders, *Proceedings* of the IEEE *Vehicular Technology Conference* , pp. 606-609, Stockholm, Sweden, May 2005

Kim, J. H. & Park, I. C. (2008). Double Binary Circular Turbo Decoding Based on Border Metric Encoding, *IEEE Trans. on Circuits and Systems II*, Vol. 55, No. 1, Jan. 2008, pp. 79-83

Kim, J. H. & Park, I. C. (2009). Bit-Level Extrinsic Information Exchange Method for Double-Binary Turbo Codes, *IEEE Trans. on Circuits and Systems II*, Vol. 56, No. 1, Jan. 2009, pp. 81-85

Kong, J. J. & Parhi, K. K. (2004). Low-Latency Architectures for High-Throughput Rate Viterbi Decoders, *IEEE Trans. on VLSI,* Vol. 12, No. 6, Jun. 2004, pp. 642-651

Kwak, J. & Lee, K. (2002). Design of Dividable Interleaver for Parallel Decoding in Turbo Codes, *IET Electronics Letters*, Vol. 38, No. 22, Oct. 2002, pp. 1362-1364

Le, N.; Soleymani, M. R. & Shayan, Y. R. (2005). Distance-Based Decoding of Block Turbo Codes, *IEEE Communications Letters*, Vol. 9, No. 11, Nov. 2005, pp. 1006-1008

LeBidan, R.; Leroux, C.; Jego, C.; Adde, P. & Pyndiah, R. (2008). Reed-Solomon Turbo Product Codes for Optical Communications: From Code Optimization to Decoder Design, *Journal on Wireless Communications and Networking*, Vol. 2008, Article ID 658042, 14 pages

Liu, H.; Diguet, J. P.; Jego, C. ; Jezequel, M. & Boutillon, E. (2007). Energy Efficient Turbo Decoder with Reduced State Metric Quantization, *Proceedings of the IEEE Workshop on Signal Processing Systems*, pp. 237-242, Shanghai, China, Oct. 2007

Mansour, M. M. & Shanbhag, N. R. (2003). High throughput LDPC decoders, *IEEE Trans. on VLSI.*, Vol. 11, No. 6, Dec. 2003, pp. 976–996

Martina, M.; Nicola, M. & Masera, G. (2008-a). A Flexible UMTS-WiMax Turbo Decoder Architecture, *IEEE Trans. on Circuits and Systems II*, Vol. 55, No. 4, Apr. 2008, pp. 369-373

Martina, M.; Nicola, M. & Masera G. (2008-b). Hardware Design of a Low Complexity, Parallel Interleaver for WiMax Duo-Binary Turbo Decoding, *IEEE Communications Letters*, Vol 12, No. 11, Nov. 2008, pp. 846-848

Martina, M.; Nicola, M. & Masera, G. (2009). VLSI Implementation of WiMax Convolutional Turbo Code Encoder and Decoder, Journal of Circuits, Systems, and Computers, Vol. 18, No. 3, May 2009, pp. 535-564

Masera, G.; Piccinini, G.; Ruo Roch, M. & Zamboni, M. (1999). VLSI Architectures for Turbo Codes, *IEEE Trans. on VLSI*, Vol. 7, No. 3, Sep. 1999, pp. 369-379

Masera, G.; Quaglio, F. & Vacca, F. (2005). Finite precision implementation of LDPC decoders, *IEE Proceedings - Communications*, Vol. 152, No. 6, Dec. 2005, pp. 1098 - 1102

Masera, G.; Quaglio, F.& Vacca, F. (2007). Implementation of a Flexible LDPC Decoder, *IEEE Trans. on Circuits and Systems II*, Vol. 54, No. 6, Jun 2007, pp. 542 – 546

Muller, O.; Baghdadi, A. & Jezequel, M. (2006). Exploiting Parallel Processing Levels for Convolutional Turbo Decoding, *Proceedings of the IEEE International Conference on Information and Communication Technologies*, pp. 2353-2358, Damascus, Syria, Apr. 2006

Muller, O.; Baghdadi, A. & Jezequel, M. (2009). From Parallelism Levels to a Multi-ASIP Architecture for Turbo Decoding, *IEEE Trans. on VLSI*, Vol. 17, No. 1, Jan. 2009, pp. 92-102

Park, S. M.; Kwak, J. & Lee, K. (2008). Extrinsic Information Memory Reduced Architecture for Non-Binary Turbo Decoder Implementation, *Proceedings of the IEEE Vehicular Technology Conference,* pp. 539-543, Marina Bay, Singapore, May 2008

Pyndiah, R. M. (1998). Near-Optimum Decoding of Product Codes: Block Turbo Codes, *IEEE Trans. on Communications*, Vol. 46, No. 8, Aug. 1998, pp. 1003-1010

Quaglio, F.; Vacca, F.; Castellano, C.; Tarable, A. & Masera, G. (2006). Interconnection framework for high-throughput, flexible LDPC decoders, *Proceedings of Design, Automation and Test in Europe Conference and Exhibition,* pp. 1-6, Munich, Germany, Mar. 2006,

Radar, C. M. (1981), Memory Management in a Viterbi Decoder, *IEEE Trans. on Communications*, Vol. COM-29, No. 9, Sep. 1981, pp. 1399-1401

Robertson, P.; Villebrun E. & Hoeher P. (1995). A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain, *Proceedings of the IEEE International Conference on Communications*, pp. 1009-1013, Seattle, USA, Jun. 1995

Shin, D.; Heo, K.; Oh, S.; & Ha, J. (2007). A Stopping Criterion for Low-Density Parity-Check Codes, *Proceedings of the IEEE Vehicular Technology Conference*, pp. 1529-1533, Dublin, Ireland, Apr. 2007.

Singh, A.; Boutillon, E. & Masera, G. (2008). Bit-width Optimization of Extrinsic Information in Turbo Decoder, *Proceedings of the 5th International Symposium on Turbo Codes & Related Topics*, pp. 134-138, Lausanne, Switzerland, Sep. 2008

Speziali, F. & Zory, J. (2004). Scalable and Area Efficient Concurrent Interleaver for High Throughput Turbo-Decoders, *Proceedings of Euromicro Symposium on Digital SystemDesign,* pp. 334-341, Rennes, France, Sep. 2004

Talakoub, S.; Sabeti, L.; Shahrrava, B. & Ahmadi, M. (2007). An Improved Max-Log-MAP Algorithm for Turbo Decoding and Turbo Equalization, *IEEE Trans. on Intrumentation and Measurement*, Vol. 56, No. 3, Jun. 2007, pp 1058-1063

Tarable, A.; Benedetto, S. & Montorsi, G. (2004). Mapping interleaver laws to parallel turbo and LDPC decoders architectures, *IEEE Trans. on Information Theory*, Vol. 50, No. 9, Sep. 2004, pp. 2002–2009

Thul, M. J.; Wehn, N, & Rao L. P. (2002). Enabling High-Speed Turbo-Decoding through Concurrent Interleaving, *Proceedings of the IEEE International Symposium on Circuits and Systems,* pp. 897-900, Scottsdale, USA, May 2002

Thul, M. J.; Gilbert, F. & Wehn, N. (2003). Concurrent Interleaving Architectures for High-Throughput Channel Coding, *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, pp. 613-616, Hong Kong, Apr. 2003

Vanstraceele, C.; Geller, B.; Brossier, J. M. & Barbot, J. P., A Low Complexity Block Turbo Decoder Architecture, *IEEE Trans. on Communications*, Vol. 56, No. 12, Dec. 2008, pp. 1985-1987

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. on Information Theory,* Vol. IT-13, Apr. 1967, pp. 260-269

Vogt, J. & Finger, A. (2000). Improving the Max-Log-MAP Turbo Decoder. *IET Electronics Letters,* Vol. 36, No. 23, Nov. 2000, pp. 1937-1939

Vogt, T. & When, N. (2008). A Reconfigurable ASIP for Convolutional and Turbo Decoding in an SDR Environment, *IEEE Trans. on VLSI,* Vol. 16, No. 10, Oct. 2008, pp. 1309-1320

Wang, H.; Yang H. & Yang, D. (2006). Improved Log-MAP Decoding Algorithm for Turbo-like Codes, *IEEE Communications Letters*, Vol. 10, No. 3, Mar. 2006, pp. 186-188

Weiss, C.; Bettstetter, C. & Riedel, S. (2001). Code Construction and Decoding of Parallel Concatenated Tailbiting Codes, *IEEE Trans. on Information Theory*, Vol. 47, No. 1, Jan. 2001, pp. 366-368

Zhan, C.; Arslan, T.; Erdogan, A. T. & MacDougall, S. (2006). An Efficient Decoder Scheme for Double Binary Circular Turbo Codes, *Proceesings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 229-232, Toulouse, France, May 2006

Zhang, J. and Fossorier, M. P. C. (2005). Shuffled iterative decoding, *IEEE Trans. on Communications*, Vol. 53, No. 2, Feb. 2005, pp. 209-213

6

# LDPC Decoders for the WiMAX (IEEE 802.16e) based on Multicore Architectures

Gabriel Falcão, Vitor Silva, José Marinho and Leonel Sousa
*Instituto de Telecomunicações/ FCTUC/University of Coimbra*
*INESC-ID/IST/Technical University of Lisbon*
*Portugal*

## 1. Introduction

The WiMAX is a wireless communications standard (IEEE 802.16e) used in small to medium distances in urban areas, that uses Low-Density Parity-Check (LDPC) codes, which are powerful error correcting codes very demanding from a computational perspective [Gallager, 1962; Mackay & Neal, 1996]. The throughput requirements imposed by the WiMAX standard define a maximum of 75 Mbps [IEEE P802.16e/D12, 2005], which typically demands a high number of arithmetic operations and memory accesses per second. In order to accommodate such requirements hardware-dedicated solutions were investigated and developed to deliver such computational power.
With the number of transistors on a die approximately doubling every 18 months, we have seen in last years processors scaling up to hundreds of millions of gates. To overcome power and memory wall constraints, the industry of processors has introduced a new trend: increasing the number of cores per processor rather than using higher clock speeds. Associated to this new paradigm, new kinds of different homogeneous and heterogeneous multicore architectures have been proposed. Initially developed essentially for rendering purposes in the industry of games, recently we have seen multicores start offering new possibilities to support general purpose computing. Nowadays largely disseminated worldwide and supported by appropriate tools, they can be exploited by convenient parallel programming models delivering unmatched performances. This reality introduced the massive generalization of general purpose processing in these parallel architectures.
So far, dedicated VLSI was the only solution capable of decoding LDPC codes at acceptable rates [Brack et al., 2006; Liu et al., 2008]. In this chapter we present a novel, programmable/flexible and scalable parallel LDPC decoding approach for the WiMAX standard based on multicore accelerators such as the Cell/B.E. architecture from Sony-Toshiba-IBM (STI) [Hofstee, 2005; International Business Machines Corporation, 2007; Falcão et al., 2008]. Moreover, we exploit parallel programming models and present parallel algorithms for this architecture [Falcão et al., 2009a]. We also report experimental results and compare them with state-of-the-art LDPC hardware decoding solutions, based on Application Specific Integrated Circuits (ASIC).

Besides the introductory and closure sections, we propose to organize this chapter around five other main sections. The second and third sections introduce, respectively, WiMAX LDPC codes and the algorithms used to perform LDPC decoding. The main characteristics of the Cell/B.E. multicore architecture are presented in section four. The fifth section describes the parallel programming models and the parallel algorithms developed to efficiently implement LDPC decoding. Finally, the experimental results section shows that the obtained throughputs compare fairly against state-of-the-art ASIC LDPC decoders.

## 2. WiMAX LDPC codes

The WiMAX standard (IEEE 802.16e) works in distances typically below the 10 Km range [Falcão et al., 2008], and uses LDPC codes, whose decoders can be very demanding from a computational perspective. For this reason, they are still implemented using dedicated hardware based on ASIC solutions. LDPCs are linear (*n*, *k*) block codes [Lin & Costello, 2004] defined by sparse binary parity-check **H** matrices of dimension (*n-k*) x *n*, and represented by bipartite graphs also called Tanner graphs [Tanner, 1981]. The Tanner graph is formed by Bit Nodes (BNs) and Check Nodes (CNs) linked by bidirectional edges. An example is shown in figure 2.

The Forward Error Correcting (FEC) system of the WiMAX standard is based on a special class of LDPC codes [IEEE P802.16e/D12, 2005] characterized by a sparse binary block parity-check matrix **H** of the form:

$$\mathbf{H}_{(n-k)\times n} = \left[\mathbf{H}1 \mid \mathbf{H}2\right] = \left[\begin{array}{cccc|c|ccccc} \mathbf{P}_{0,0} & \mathbf{P}_{0,1} & \cdots & \mathbf{P}_{0,k/z-1} & \mathbf{Pb}_0 & \mathbf{I} & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{P}_{1,0} & \mathbf{P}_{1,1} & \cdots & \mathbf{P}_{1,k/z-1} & & \mathbf{I} & \mathbf{I} & & & \vdots \\ \vdots & \ddots & & \vdots & & \mathbf{0} & \mathbf{I} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \mathbf{I} & \mathbf{0} \\ \mathbf{P}_{n-k/z-2,0} & \mathbf{P}_{n-k/z-2,1} & \cdots & \mathbf{P}_{n-k/z-2,k/z-1} & \vdots & & \ddots & & \mathbf{I} & \mathbf{I} \\ \mathbf{P}_{n-k/z-1,0} & \mathbf{P}_{n-k/z-1,1} & \cdots & \mathbf{P}_{n-k/z-1,k/z-1} & \mathbf{Pb}_{n-k/z-1} & \mathbf{0} & & \cdots & \mathbf{0} & \mathbf{I} \end{array}\right], \quad (1)$$

where **H1** is sparse and adopted special periodicity constraints introduced in the pseudo random design of the matrix [IEEE P802.16e/D12, 2005; Brack et al., 2006] and **H2** is a sparse lower triangular block matrix with a staircase profile. The periodic nature of these codes defines **H1** based on permutation sub-matrices $\mathbf{P}_{i,j}$, which are: (1) quasi-random circularly shifted right identity sub-matrices (**I**) (as depicted in figure 1) , with dimensions $z \times z$ ranging from $24 \times 24$ to $96 \times 96$ and incremental granularity of 4 (table 1); or (2) $z \times z$ null sub-matrices. The periodic nature of such codes allowed simplifying the architecture of the system and storage requirements without code performance loss [IEEE P802.16e/D12, 2005]. Also, the right sub-matrix **H2** is formed by identity: (1) **I** sub-matrices of dimension $z \times z$ ; or (2) null sub-matrices of dimension $z \times z$ .

Therefore, the LDPC codes adopted by the WiMAX standard (IEEE 802.16e) support 19 different codeword sizes with 4 distinct code rates (*rate*=*k*/*n*) and 6 different class codes (distinct distributions of the number of BNs per column or CNs per row). They are depicted

in table 1. Class $2/3A$ defines codes having {2, 3, 6} BNs per row and {10} CNs per column, while class $2/3B$ codes have {2, 3, 4} BNs and {10, 11} CNs. Also, class $3/4A$ has {2, 3, 4} BNs and {14, 15} CNs, and class $3/4B$ has {2, 3, 6} BNs per row and {14, 15} CNs per column.

| Code | Codeword bits ($n$) | $z \times z$ factor | Information bits ($k$) | | | |
|---|---|---|---|---|---|---|
| | | | *(rate)* | *(rate)* | *(rate)* | *(rate)* |
| | | | 1/2 | 2/3 | 3/4 | 5/6 |
| 1 | 576 | $24 \times 24$ | 288 | 384 | 432 | 480 |
| 2 | 672 | $28 \times 28$ | 336 | 448 | 504 | 560 |
| 3 | 768 | $32 \times 32$ | 384 | 512 | 576 | 640 |
| 4 | 864 | $36 \times 36$ | 432 | 576 | 648 | 720 |
| 5 | 960 | $40 \times 40$ | 480 | 640 | 720 | 800 |
| 6 | 1056 | $44 \times 44$ | 528 | 704 | 792 | 880 |
| 7 | 1152 | $48 \times 48$ | 576 | 768 | 864 | 960 |
| 8 | 1248 | $52 \times 52$ | 624 | 832 | 936 | 1040 |
| 9 | 1344 | $56 \times 56$ | 672 | 896 | 1008 | 1120 |
| 10 | 1440 | $60 \times 60$ | 720 | 960 | 1080 | 1200 |
| 11 | 1536 | $64 \times 64$ | 768 | 1024 | 1152 | 1280 |
| 12 | 1632 | $68 \times 68$ | 816 | 1088 | 1224 | 1360 |
| 13 | 1728 | $72 \times 72$ | 864 | 1152 | 1296 | 1440 |
| 14 | 1824 | $76 \times 76$ | 912 | 1216 | 1368 | 1520 |
| 15 | 1920 | $80 \times 80$ | 960 | 1280 | 1440 | 1600 |
| 16 | 2016 | $84 \times 84$ | 1008 | 1344 | 1512 | 1680 |
| 17 | 2112 | $88 \times 88$ | 1056 | 1408 | 1584 | 1760 |
| 18 | 2208 | $92 \times 92$ | 1104 | 1472 | 1656 | 1840 |
| 19 | 2304 | $96 \times 96$ | 1152 | 1536 | 1728 | 1920 |

Table 1. Properties of LDPC codes used in the WiMAX IEEE 802.16e standard.

To illustrate the periodic nature introduced in the design of the **H** matrix, the generic structure of a sample parity-check matrix used in WiMAX, with $n$=2304 and *rate*=1/2, is depicted in figure 1.

## 3. Algorithms for LDPC decoding (SPA and MSA)

Computationally intensive message-passing algorithms can be used for LDPC decoding, varying from the well known belief propagation, or Sum-Product algorithm (SPA) [Falcão et al., 2009a] to the more efficient but yet still intensive Min-Sum algorithm (MSA) [Falcão et al., 2008]. The intensive message passing procedure is illustrated in figure 2. We will discuss the parallelization of the SPA and MSA, and show how data dependencies can be manipulated in order to allow the implementation of parallel decoders.

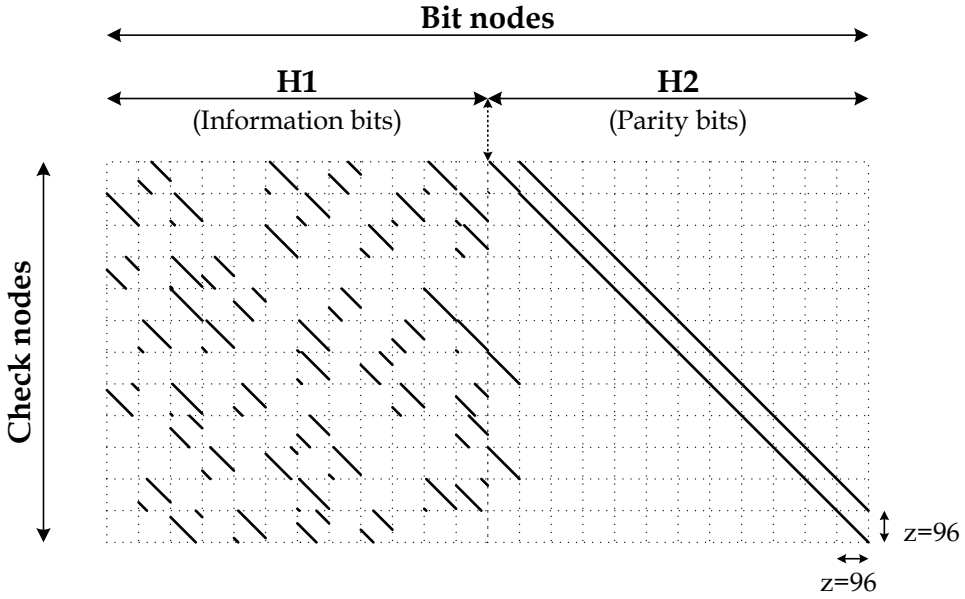
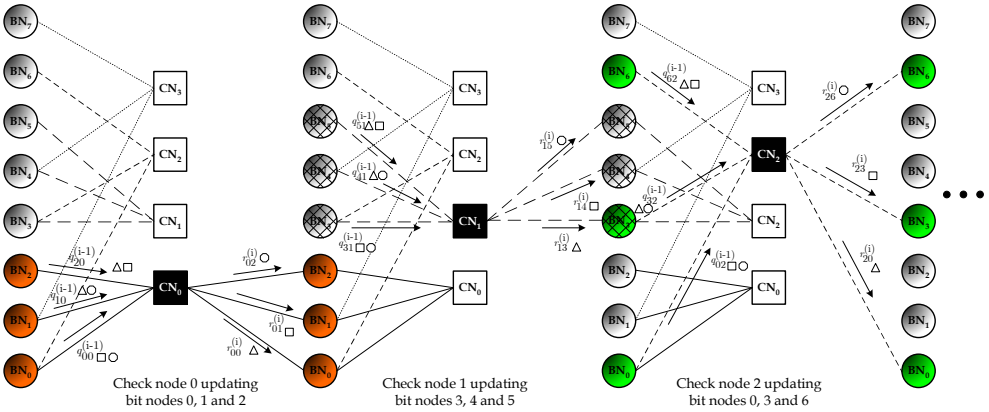



Fig. 1. Periodicity $z \times z = 96 \times 96$ for a matrix with $n$=2304, rate=1/2 and {3, 6} CNs per column.

### 3.1 The Sum-Product Algorithm (SPA)

The SPA is a very efficient algorithm [Lin & Costello, 2004] used for LDPC decoding and it is based on the belief propagation between adjacent nodes connected as indicated by the Tanner graph edges (figure 2). As proposed by Gallager, the SPA operates on probabilities [Gallager, 1962; Mackay & Neal, 1996; Lin & Costello, 2004]. Given a (n, k) LDPC code, we assume Binary-Phase Shift Keying (BPSK) modulation which maps a codeword c = ($c_1$, $c_2$, ... , $c_n$) into the sequence x = ($x_1$, $x_2$, ... , $x_n$), according to $x_i = (-1)^{c_i}$. Then, x is transmitted through an additive white Gaussian noise (AWGN) channel originating the received sequence y = ($y_1$, $y_2$ , ... , $y_n$) on the decoder side, with $y_i = x_i + n_i$, where $n_i$ is a random variable with zero mean and variance $\sigma^2$.

The SPA is depicted from (2) to (8) [Lin & Costello, 2004]. It is mainly described by two horizontal and vertical intensive processing blocks, respectively defined by equations (3), (4) and (5), (6). The first two calculate messages moving from each $CN_m$ to $BN_n$, considering accesses to **H** on a row basis. It indicates the probability of $BN_n$ being 0 or 1. Figure 2 exemplifies, for a particular 4 x 8 **H** matrix, $BN_0$, $BN_1$ and $BN_2$ being updated by $CN_0$, then $BN_3$, $BN_4$ and $BN_5$ updated by $CN_1$ and finally $BN_0$, $BN_3$, $BN_6$ by $CN_2$. Similarly, the vertical processing block computes messages sent from $BN_n$ to $CN_m$, assuming accesses on a column basis. The iterative procedure is stopped if the decoded word $\hat{\mathbf{c}}$ verifies all parity-check equations of the code $\hat{\mathbf{c}}.\mathbf{H}^T = \mathbf{0}$, or a maximum number of iterations $I$ is reached, in which case no codeword is detected.




(8 bit nodes checked by 4 check node equations)

$$\mathbf{H} = \begin{array}{c} \begin{matrix} \mathbf{BN_0} & \mathbf{BN_1} & \mathbf{BN_2} & \mathbf{BN_3} & \mathbf{BN_4} & \mathbf{BN_5} & \mathbf{BN_6} & \mathbf{BN_7} \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array} \begin{matrix} \mathbf{CN_0} \\ \mathbf{CN_1} \\ \mathbf{CN_2} \\ \mathbf{CN_3} \end{matrix}$$

Fig. 2. A 4x8 **H** matrix and corresponding Tanner graph representation. The example shows messages being exchanged from $CN_m$ to $BN_n$. A similar representation applies for messages going in the opposite direction.

$$q_{nm}^{(0)}(0) = 1 - p_n;\ q_{nm}^{(0)}(1) = p_n;$$
$$p_n = \frac{1}{1 + e^{\frac{2y_n}{\sigma^2}}};\ \frac{E_b}{N_0} = \frac{N}{2K\sigma^2}\{Initialization\} \tag{2}$$

**while** ($\hat{\mathbf{c}}.\mathbf{H}^T \neq \mathbf{0} \cap \mathbf{i} < \mathbf{I}$) { $\hat{\mathbf{c}}$-decoded word; **I**-max. number of iterations.}
**do**
{For each node pair ($BN_n$, $CN_m$), corresponding to $\mathbf{H_{mn}} = \mathbf{1}$ in the parity-check matrix **H** of the code **do**:}
{1. **Horizontal Processing (kernel 1)** – Compute messages sent from $CN_m$ to $BN_n$, that indicate the probability of $BN_n$ being **0** or **1**:}

$$r_{mn}^{(i)}(0) = \frac{1}{2} + \frac{1}{2}\prod_{n' \in N(m)\backslash n}\left(1 - 2q_{n'm}^{(i-1)}(1)\right) \tag{3}$$

$$r_{mn}^{(i)}(1) = 1 - r_{mn}^{(i)}(0) \tag{4}$$

{where N(m)\n represents BNs connected to $CN_m$ excluding $BN_n$.
{2. **Vertical Processing (kernel 2)** – Compute messages sent from $BN_n$ to $CN_m$:}

$$q_{nm}^{(i)}(0) = K_{nm}(1-p_n) \prod_{m' \in M(n) \setminus m} r_{m'n}^{(i)}(0) \tag{5}$$

$$q_{nm}^{(i)}(1) = K_{nm} p_n \prod_{m' \in M(n) \setminus m} r_{m'n}^{(i)}(1) \tag{6}$$

{where $k_{nm}$ are chosen to ensure $q_{nm}^{(i)}(0) + q_{nm}^{(i)}(1) = 1$, and M(n)\m is the set of CNs connected to $BN_n$ excluding $CN_m$.}
{3. Compute *a posteriori* pseudo-probabilities:}

$$\begin{aligned} Q_n^{(i)}(0) &= K_n(1-p_n) \prod_{m \in M(n)} r_{mn}^{(i)}(0) \\ Q_n^{(i)}(1) &= K_n p_n \prod_{m \in M(n)} r_{mn}^{(i)}(1) \end{aligned} \tag{7}$$

{where $k_n$ are chosen to guarantee $Q_n^{(i)}(0) + Q_n^{(i)}(1) = 1$.}
{Perform hard decoding} $\forall n$,

$$\hat{c}_n^{(i)} = \begin{cases} 1 \Leftarrow Q_n^{(i)}(1) > 0.5 \\ 0 \Leftarrow Q_n^{(i)}(1) < 0.5 \end{cases} \tag{8}$$

**end while**

### 3.2 The Min-Sum Algorithm (MSA)

The MSA [Guilloud et al., 2003] is a simplification of the well-known SPA in the logarithmic domain. It is also based on the intensive belief propagation between nodes connected as indicated by the Tanner graph edges, but that uses only comparison and addition operations. Being one of the most efficient algorithms used for LDPC decoding [Liu et al., 2008; Seo et al., 2007], even so, the MSA still requires intensive processing.
Let us denote the log-likelihood ratio LLR of a random variable as *L(x) = ln (p(x = 0)/p(x = 1))*. Also, considering the message propagation from nodes $CN_m$ to $BN_n$ and vice-versa, the set of bits that participate in check equation *m* with bit *n* excluded is represented by *N(m)\n* and, similarly, the set of check equations in which bit *n* participates with check *m* excluded is *M(n)\m*. $LP_n$ designates the *a priori* LLR of $BN_n$, derived from the values received from the channel, and $Lr_{mn}$ the message that is sent from $CN_m$ to $BN_n$, computed based on all received messages from BNs *N(m)\n*. Also, $Lq_{nm}$ is the LLR of $BN_n$, which is sent to $CN_m$ and calculated based on all received messages from CNs *M(n)\m* and the channel information $LP_n$. For each node pair ($BN_n$, $CN_m$) we initialize $Lq_{nm}$ with the *a priori* log-likelihood ratio (LLR) information received from the channel, $LP_n$. Then, we proceed to the iterative body of the algorithm by calculating (9) and (10), respectively, where $Lr_{mn}$ denotes the message sent from $CN_m$ to $BN_n$, and $Lq_{nm}$ the message sent from $BN_n$ to $CN_m$.

{1. **Horizontal Processing (kernel 1) –** Compute messages sent from $CN_m$ to $BN_n$:}

$$Lr_{mn}^{(i)} = \prod_{n' \in N(m) \setminus n} sign\left(Lq_{n'm}^{(i-1)}\right) \min_{n' \in N(m) \setminus n} \left|Lq_{n'm}^{(i-1)}\right|. \tag{9}$$

{2. **Vertical Processing (kernel 2) –** Compute messages sent from $BN_n$ to $CN_m$:}

$$Lq_{nm}^{(i)} = LP_n + \sum_{m' \in M(n) \setminus m} Lr_{m'n}^{(i)}. \tag{10}$$

{3. Finally, we calculate the *a posteriori* pseudo-probabilities and perform hard decoding: }

$$LQ_n^{(i)} = LP_n + \sum_{m' \in M(n)} Lr_{m'n}^{(i)}. \tag{11}$$

## 4. The Cell/B.E. multicore architecture

The required computational power and the irregular memory access patterns for LDPC decoding define hard challenges that we propose to tackle for the heterogeneous Cell/B.E. multicore [Hofstee, 2005] shown in figure 3. The Cell/B.E. processor provides a set of cores that include one main 64-bit PowerPC Processor Element (PPE) and eight Synergistic Processor Elements (SPEs) under a vectorized 128-bit wide Single Instruction Multiple Data (SIMD) oriented architecture [Sony Computer Entertainment Incorporated, 2005]. Data transfers between the main memory and each SPE's local memory (256KByte) are performed by using efficient Direct Memory Access (DMA) mechanisms that offload the processors from the expensive task of moving data. Each SPE, by its turn, exploits quite efficiently a dual pipeline mechanism executing independently: one supports arithmetic operations; while the other performs load and store memory operations.

The memory in the CELL/B.E. is organized as a distributed memory system. Data is loaded from the main memory into the SPE's local storage and vice-versa, allowing each processor to exploit data locality individually. As in opposition to architectures based on shared memory models, here the programmer is free from having to deal with strategies to avoid memory access conflicts.

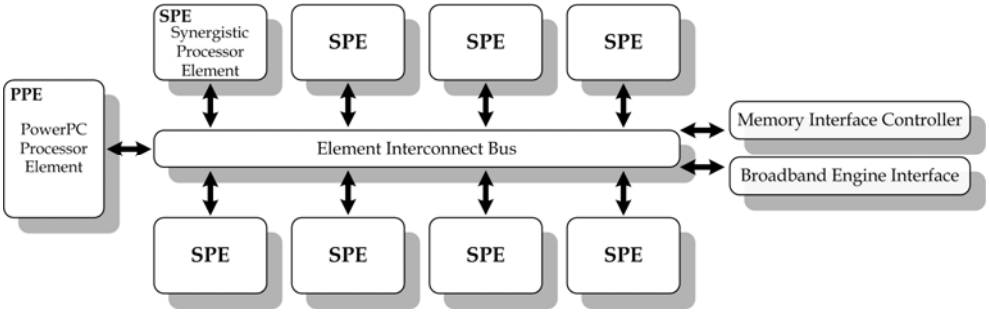



Fig. 3. Generic overview of the Cell/B.E. architecture.

## 5. The parallel programming model and parallel algorithms for LDPC decoding on multicores

The Single Program Multiple Data (SPMD) and the SIMD programming models are adopted to exploit data parallelism and to develop the parallel methods and algorithms for LDPC decoding on the Cell/B.E. A vectorized SIMD-based multicore approach that exploits data locality and fast data-block transfers associated to a powerful dual pipeline mechanism, allowed us to efficiently implement the concept of simultaneous multicodeword LDPC decoding on the Cell/B.E. Each SPE decodes several complete codewords and a total of 24 to 96 codewords, depending if we use 32- or 8-bit data precision, are decoded in parallel and at the same time in all the 6 SPEs available on the PlayStation 3 platform used in this work.

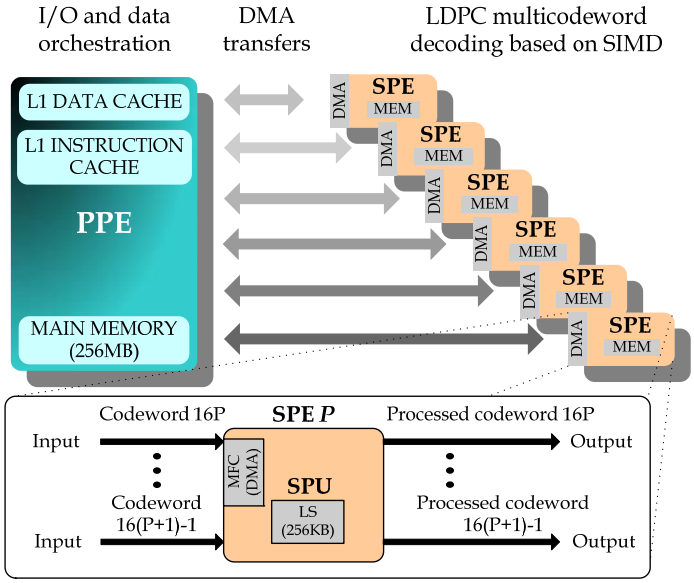



Fig. 4. Parallelization model for an LDPC decoder on the Cell/B.E. architecture.

The parallel LDPC decoder exploits SIMD data-parallelism by applying the same algorithm to different codewords on each SPE. As the Tanner graph is common to all codewords under decoding, these data structures can be shared allowing multicodeword decoding simultaneously in all SPEs. These features are illustrated in figure 4.

The irregular memory access patterns common in LDPC decoding represent a challenge to the efficiency of the algorithm as illustrated in figure 5 for the example shown in figure 2. The access to different nodes in the Tanner graph is defined by the **H** matrix and should favor randomization in order to allow good coding gains. For that reason, the data structures developed and represented in figure 6 try to minimize that effect, by grouping contiguously in memory associated data computed in the same kernel. A global irregular access pattern is translated into several partial regular access patterns. Moreover, only non-null elements of the **H** matrix are stored which represents savings in terms of memory usage.

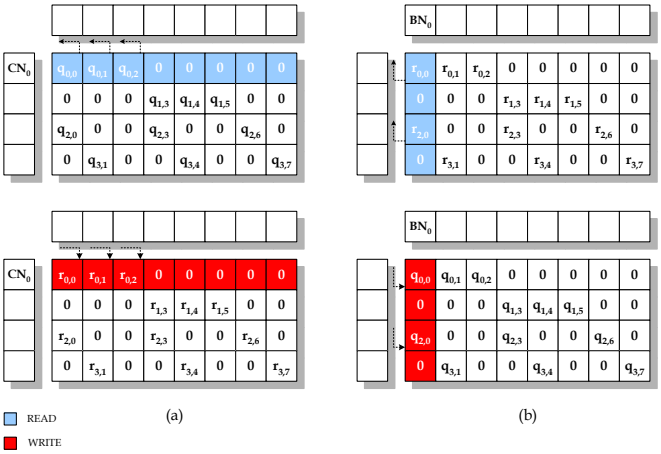



Fig. 5. Illustration of irregular memory accesses for the example shown in figure 2.

### 5.1 The parallel algorithm on the Cell/B.E.

The parallelization approach proposed for developing an LDPC decoder is explained in the context of the Cell/B.E. architecture. The data structures that define the Tanner graph and the program as well are loaded into the local storage on the SPEs where the processing is performed. The LDPC decoder processes on an iterative basis. The PPE reads information $y_n$ from the input channel and produces the probabilities $p_n$ as indicated in (2). The PPE controls the main tasks, offloading the intensive processing to the SPEs, where the processing is then distributed over several threads. Each SPE runs independently of the other SPEs. After receiving the $p_n$ values associated to the corresponding codewords, each SPE performs two steps: *(i)* computes kernel 1 and kernel 2 alternately using SIMD instructions; and *(ii)* sends the final results back to the PPE, which concludes the computation of the current codewords and starts new ones by replacing data to be sent to the SPEs.

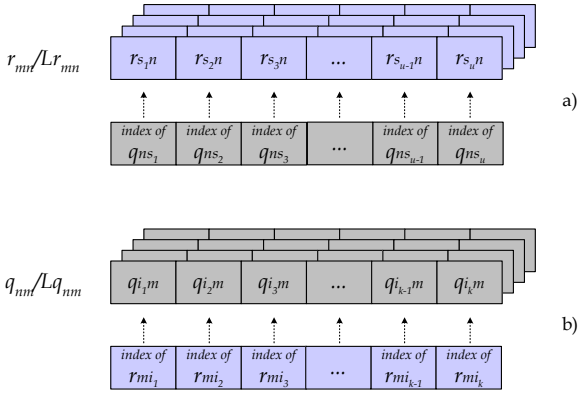



Fig. 6. Segment of the SIMD vectorized data structures in memory to represent: a) $r_{mn}/Lr_{mn}$ messages associated to BNs; and b) $q_{nm}/Lq_{nm}$ messages associated to CNs.

The parallel LDPC decoder explores data-parallelism by applying the same algorithm to several codewords simultaneously on each SPE (as shown in figure 4). Data is represented as 32-bit precision floating-point or 8-bit integer elements, depending on the algorithm used – SPA or MSA. The proposed LDPC decoder suits scalability and for that reason it can be easily adopted by future generations of the architecture with a higher number of SPEs. In that case, it will be able of decoding more codewords simultaneously, increasing the efficiency and aggregate throughput of the decoder.

Processing a complete iteration inside the SPE is performed in two phases: *(i)* kernel 1 computes data according to (3) and (4) for the SPA or (9) for the MSA, where the horizontal processing (row-major order) is performed. The data structure designed to represent $r_{mn}$/$Lr_{mn}$ in order to perform this task is depicted in figure 6 a). It can be seen in this figure that data related to BNs common to a CN equation is stored in contiguous memory positions to optimize processing; and *(ii)* kernel 2 processes data according to (5) and (6) for the SPA or (10) for the MSA, performing the vertical processing (column-major order). The $q_{nm}$/$Lq_{nm}$ data structure used in this case is depicted in figure 6 b). The SPE accesses data in a row- or column-major order, depending on the kernel that is being processed at the time. Also, the parallel algorithm implemented exploits the double buffering technique by overlapping processing and data accesses in memory.

Kernel 1 performs the horizontal processing according to the Tanner graph, and the $r_{mn}^{(i)} / Lr_{mn}^{(i)}$ data is updated for iteration $i$. The data is initially transferred to the local storage of the SPE by performing a DMA transaction, and its access organization maximizes data reuse, because a CN updating BNs reads common information from several BNs that share data among them. Figure 6 b) shows the data structures that hold the $q_{nm}$/$Lq_{nm}$ values to be read and also the corresponding indexes of the $r_{mn}$/$Lr_{mn}$ elements in figure 6 a) that they are going to update. As depicted in figure 6, BNs and CNs associated with the same $r_{mn}$/$Lr_{mn}$ or $q_{nm}$/$Lq_{nm}$ equation are represented in contiguous blocks of memory.

In kernel 2 data is processed in a column-major order. According to the Tanner graph, each BN updates all the CNs connected to it and holds the addresses necessary to complete the update of all $q_{nm}^{(i)} / Lq_{nm}^{(i)}$ data for iteration $i$. Once again, maximum data reuse is achieved, but this time among data belonging to the same column of the **H** matrix, as depicted in figures 2 and 6 b).

The computation is performed in the SPE for a predefined number of iterations. One of the purposes of this work is to assess the performance of the proposed solutions in terms of throughput. Pursuing this goal, we decided to develop a solution where the number of iterations is fixed to allow a fair comparison between different approaches, where the processing workload is known *a priori* and the same for all environments. This is why, at the end of an iteration, we don't check if the decoder produces a valid codeword, which could cause the decoding process to stop. Nevertheless, this operation represents a negligible overhead. The iterative updating mechanism applied to BNs and CNs is performed in a sequence of pairs (BN, CN) and tested for a number of iterations ranging from 10 to 100. When all the BNs and CNs are updated after the final iteration, the SPE activates a DMA transaction and sends data back to the main memory, signalizing the PPE to conclude the processing. As the DMA finishes transferring data, synchronization points are introduced to allow data buffers reuse.

**Algorithm 1** PPE side of the algorithm

**for** $th_ctr = 1$ *to* $NSPEs$ : **do**
  *Create* $th_ctrt$ *hread*
**end for**

**repeat**
  *Receive* $y_n$ *from the channel and calculate* $p_n$ *probabilities*
  *Send msg NEW_WORD to MASTER SPE*
**Ensure :** *Wait until mail* is *received* $(SPE[i].mailboxcount > 0)$ *from MASTER SPE*
  $msg = SPE[i].mailbox(received\ msg\ END_DECODE\ from\ MASTER\ SPE)$
**until** *true*

Synchronization between the PPE and the SPEs is performed using mailboxes. This approach tries to exploit data-parallelism and data locality by performing the partitioning and mapping of the algorithm and data structures over the multiple cores, while at the same time minimizes delays caused by latency and synchronization.

### 5.2 Processing on the PPE

The part of the algorithm that executes on the PPE side is presented in Algorithm 1. We force the PPE to communicate with only one SPE, called MASTER SPE, which performs the control over the remaining SPEs. This is more efficient than putting the PPE controlling all the SPEs.

**Algorithm 2** MASTER SPE side of the algorithm

**repeat**
**Ensure :** *Read mailbox* $(waiting$ a $NEW_WORD\ mail\ from\ PPE)$
  *Broadcast msg NEW_WORD to all other SPEs*
  *Get* $p_n$ *probabilities*
  **for** $i = 1$ *to* $N_Iter$ : **do**
    *Compute* $r_{mn} / Lr_{mn}$
    *Compute* $q_{nm} / Lq_{nm}$
  **end for**
  *Put final* $Q_n$ *values on the PPE*
**Ensure :** *Read mailbox* $(waiting$ an $END_DECODE\ mail\ from\ all\ other\ SPEs)$
  *Send msg END_DECODE to PPE*
**until** *true*

The PPE receives the $y_n$ information from the channel and calculates probabilities $p_n$, after which it sends a *NEW_WORD* message to the MASTER SPE. Then, it waits for the download of all $p_n$ probabilities to the SPEs and for the processing to be completed in each one of them.

**Algorithm 3** SLAVE SPE side of the algorithm

**repeat**
**Ensure :** *Read mailbox* $\left(waiting\ \text{a}\ NEW_WORD\ mail\ from\ MASTER\ SPE\right)$
*Get* $p_n$ *probabilities*
**for** $i = 1$ *to* $N_Iter$ : **do**
*Compute* $r_{mn} / Lr_{mn}$
*Compute* $q_{nm} / Lq_{nm}$
**end for**
*Put final* $Q_n$ *values on the PPE*
*Send msg END_DECODE to MASTER SPE*
**until** *true*

Finally, when all the iterations are completed, the MASTER SPE sends an *END_DECODE* message to the PPE to conclude the current decoding process and get ready to start processing a new set of codewords.

### 5.3 Processing on the SPE

The SPEs are used in the intensive task of updating all BNs and CNs by executing kernels 1 and 2 (either for the SPA or for the MSA), in each decoding iteration. Each thread running on the SPEs accesses data in the main memory by using DMA and computes data according to the Tanner graph, as defined in the **H** matrix (figure 2). The MASTER SPE side of the procedure is described in Algorithm 2. The *Get* operation is adopted to represent a communication PPE → SPE, while the *Put* operation is used for communications in the opposite direction.

We initialize the process and start an infinite loop, waiting for communications to arrive from the PPE (in the case of the MASTER SPE), or from the MASTER SPE (for all remaining SPEs). In the MASTER SPE, the only kind of message expected from the PPE is a *NEW_WORD* message. When a *NEW_WORD* message is received, the MASTER SPE broadcasts a *NEW_WORD* message to all other SPEs and loads corresponding $p_n$ probabilities. After receiving these messages, each one of the other SPEs gets its own $p_n$ values.

The processing starts and terminates when the number of iterations is reached and an *END_DECODE* mail is sent by all SPEs to the MASTER SPE, which immediately notifies the PPE with an *END_DECODE* message, as described in Algorithms 2 and 3.

The intensive part of the computation in LDPC decoding on the Cell/B.E. architecture takes advantage of the processing power and SIMD instruction set available on the SPEs, which means that several codewords are decoded in parallel.

## 6. Experimental evaluation

To evaluate the performance of the proposed LDPC decoder, the Cell/B.E. was programmed using: *(i)* the PPE alone which is denoted by serial mode in figure 7; and *(ii)* the complete set of PPE and 6 SPE processors denoted by parallel mode in the same figure.

The Cell/B.E. under test is included in a PlayStation 3 (PS3) platform, which restricts the number of available SPEs to 6, from a total of 8. The experimental setup of the PS3 platform is presented in table 2.

| | Serial mode | Parallel mode | |
|---|---|---|---|
| Platform | PPE | STI Cell/B.E. | |
| Language | C | C | |
| OS | Linux (Fedora) kernel 2.6.16 | | |
| | | PPE | SPE |
| Clock frequency | 3.2GHz | 3.2GHz | 3.2GHz |
| Memory | 256MB | 256MB | 256KB |

Table 2. Experimental setup.

### 6.1 Serial versus parallel execution modes

The serial mode depicted in figure 7 uses a dual thread approach and exploits SIMD instructions. It should be noted that by performing the comparison based on the time per bit decoded, the serial solution that uses only the PPE is slower than the execution on a single SPE, because the PPE accesses slow main memory, while the SPE accesses faster local storage memory.

| Code (n, k) | Edges | Occupancy of data structures on the local storage of a SPE (Bytes) |
|---|---|---|
| (504, 252) | 1512 | 70560 |
| (1024, 512) | 3072 | 143360 |

Table 3. Size of data structures used in each SPE in parallel decoding mode.

On the parallel approach the experimental results were also obtained using SIMD instructions on the SPEs, which are responsible for executing the intensive decoding part of the algorithm. In this case the PPE orchestrates the execution on the SPEs as explained before, while inside each SPE several codewords are being simultaneously decoded in parallel. All the processing times were measured for a number of iterations ranging from 10 to 100. In literature, the average number of iterations considered for WiMAX LDPC decoders to work under realistic conditions is typically below 20 iterations [Brack et al., 2006; Seo et al., 2007; Liu et al., 2008].

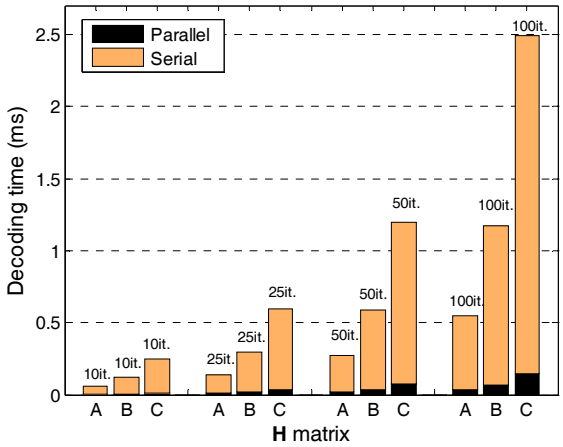



Fig. 7. Comparing LDPC decoding times for serial and parallel modes, where codes A, B and C represent, respectively, matrices (256, 128), (504, 252) and (1024, 512).

Table 3 shows the dimensions of two regular LDPC codes with *rate*=1/2, the corresponding number of edges and the size of data structures used to represent the Tanner graph. The local memory of the SPE is limited to 256 KByte. It should be taken in consideration that the SPE's local memory should hold both data structures and also the program.

### 6.2 LDPC decoding using the SPA

The first parallel approach mentioned in 6.1 uses the SPA. Data elements have 32-bit floating-point precision, and 4 floating-point elements are packed and operated on a single instruction, making it possible to decode 4 codewords in parallel on each SPE. Then, with 6 SPEs available, the global architecture can decode 24 codewords in simultaneous. We assessed the results for regular codes, which typically execute faster than irregular ones. The average throughput obtained is presented in table 4, and it ranges from 69.1 to 69.5 Mbps, when decoding regular codes (504, 252) and (1024, 512) in 10 iterations. It should be noticed that the decoding time per bit and per iteration remains approximately constant. Although real-life performances demand throughputs which can be typically in the order of 40 Mbps per channel, the theoretical maximum required by the WiMAX standard can go up to approximately 75 Mbps per channel. The throughputs reported in table 4 are inferior to 70 Mbps and do not guarantee such requirements. Also, adapting the algorithm to support the necessary irregular codes used in the WiMAX would produce even worst results, because in that case accesses to memory depend on a variable number of edges per row/column. Therefore, optimizing the LDPC decoding algorithm to make it execute in a shorter period of time became mandatory. One possible solution consisted of exploiting the computationally less demanding MSA described in section 3.2.

| Code ($n$, $k$) | *rate* | 10 iter. | 25 iter. | 50 iter. |
|---|---|---|---|---|
| (504, 252) | 1/2 | 69.1 | 28.3 | 14.2 |
| (1024, 512) | 1/2 | 69.5 | 28.4 | 14.3 |

Table 4. Throughput (Mbps) obtained in the parallel mode for the SPA.

### 6.3 LDPC decoding using the MSA

To increase the efficiency of the LDPC decoder we implemented the MSA on the Cell/B.E. It requires less computation, based essentially in addition and comparison operations [Falcão et al., 2009b]. Additionally, we also adopted the Forward-and-Backward simplification of the algorithm [Mackay, 1999] that avoids redundant computation and eliminates repeated accesses to memory. In the MSA data elements have 8-bit integer precision, which allows packing 16 data elements per 128-bit memory access. This increases the arithmetic intensity of the algorithm, here defined as the number of arithmetic operations per memory access, which favors the global performance of the LDPC decoder. The instruction set of the Cell/B.E. architecture supports intrinsic instructions to deal efficiently with these parallel 128-bit data types. Moreover, because there are 6 SPEs available, the algorithm now supports the simultaneous decoding of 96 codewords in parallel. However, the set of 8-bit integer intrinsic parallel instructions of the Cell/B.E. is more limited than those of the 32-bit floating-point family of instructions. This explains that the speedup obtained when changing from the SPA to the MSA is lower than we would expect. Table 5 shows the throughputs obtained for some example codes used in the WiMAX IEEE 802.16e standard. For 10 iterations, in some cases they approach quite well while in others they even surpass the 75 Mbps required by the standard to work in (theoretical) worst case conditions.

| Code ($n$, $k$) | *rate* | 10 iter. | 25 iter. | 50 iter. |
|---|---|---|---|---|
| (576, 288) | 1/2 | 79.8 | 32.7 | 16.5 |
| (576, 432) | 3/4 | 73.1 | 29.9 | 15.1 |
| (576, 480) | 5/6 | 79.3 | 32.5 | 16.4 |
| (672, 448) | 2/3 | 74.8 | 30.6 | 15.4 |
| (672, 504) | 3/4 | 72.6 | 29.7 | 15.0 |
| (672, 560) | 5/6 | 78.5 | 32.2 | 16.2 |
| (960, 480) | 1/2 | 79.6 | 32.6 | 16.4 |
| (960, 640) | 2/3 | 74.7 | 30.6 | 15.4 |
| (960, 720) | 3/4 | 72.6 | 29.7 | 15.0 |
| (960, 800) | 5/6 | 78.4 | 32.1 | 16.2 |
| (1152, 576) | 1/2 | 79.6 | 32.6 | 16.4 |
| (1152, 768) | 2/3 | 74.6 | 30.5 | 15.4 |
| (1152, 864) | 3/4 | 72.6 | 29.7 | 15.0 |
| (1152, 960) | 5/6 | 78.4 | 32.1 | 16.2 |
| (1248, 624) | 1/2 | 79.6 | 32.6 | 16.4 |
| (1248, 832) | 2/3 | 78.5 | 32.2 | 16.2 |
| (1248, 936) | 3/4 | 72.7 | 29.7 | 15.0 |
| (1248, 1040) | 5/6 | 78.4 | 32.1 | 16.2 |

Table 5. Throughput (Mbps) obtained in the parallel mode for the MSA.

For codes with $n$=576 running 10 iterations, it can be seen that throughputs range from 73.1 to 79.8 Mbps. For codes with n=1248 and for the same number of iterations, they vary from 72.7 to 79.6 Mbps. All codes in table 5 were tested for the MSA and approach quite well the maximum theoretical limit of 75 Mbps. They all show better performances than those obtained with the SPA. Furthermore, if we consider a lower number of iterations, the

decoder's throughput may rise significantly. For example, if we consider an LDPC decoder running 5 iterations, the throughput will approximately double to values above 145Mbps.

### 6.4 Discussion

In multicore architectures, efficient parallel programming, both in terms of computation and memory accesses, represent a significant challenge. The Cell/B.E. is based on a distributed memory model where the problem of data collisions can decrease when properly handled by the programmer. The reported experimental results allow assessing the performance of LDPC decoders based on multicores. We have shown that for LDPC decoders running the SPA on the Cell/B.E., throughputs can range from 68 to nearly 70 Mbps [Falcão et al., 2009a]. Concerning the MSA, a more efficient solution is achieved producing throughputs that range from 72 to 80 Mbps [Falcão et al., 2008]. Regarding to non-scalable hardware dedicated ASIC solutions, which typically adopt 5 to 6-bit precision arithmetic [Liu, 2008], the parallel programmable architecture here proposed allows using 8-bit data precision or even more, which produces lower Bit Error Rates (BER) and superior coding gains as depicted in figure 8. The adoption of specific parallelization techniques on a low-cost multicore platform allowed us to achieve throughputs that approach well those obtained with ASIC-based solutions for WiMAX [Brack, 2006; Seo, 2007; Liu, 2008;]. They also guarantee enough bandwidth for LDPC codes used in the WiMAX standard to work in worst case conditions.

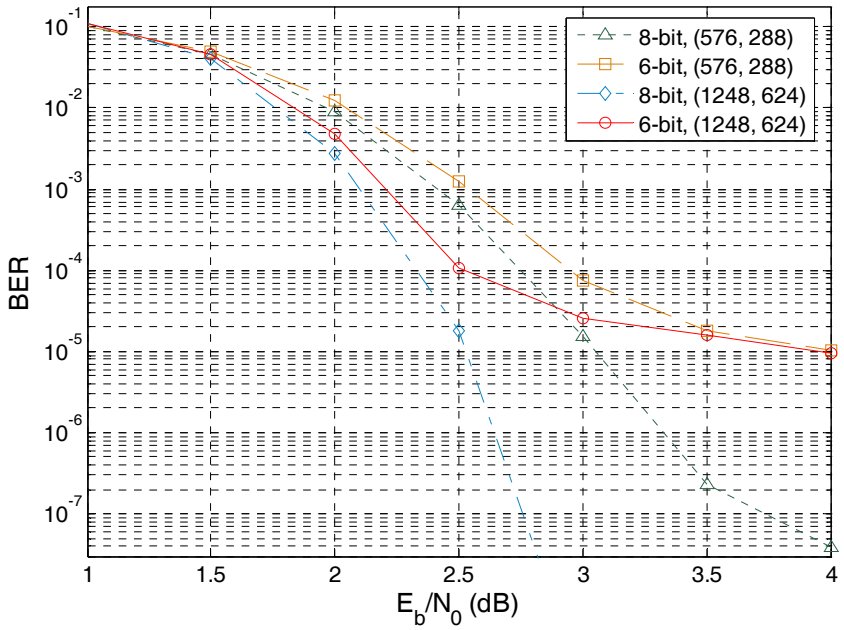



Fig. 8. BER curves for WiMAX codes (576, 288) and (1248, 624), considering both 6- and 8-bit data precision representations.

## 7. Conclusions

The advent of inexpensive multicore architectures has allowed to develop a novel programmable LDPC decoding solution for the WiMAX standard, with excellent throughputs, on the Cell/B.E. architecture. The LDPC decoder here presented exploits parallelism and data locality and is scalable to future generations of the Cell/B.E. architecture that are expected to have more SPEs, and should therefore improve the performance even further, processing more channels/subcarriers per second. The proposed decoder compares well with non-scalable and hardware-dedicated typical ASIC LDPC decoding solutions, reporting superior BER performances and throughputs above 72 Mbps. On going additional work related with LDPC decoders running on alternative parallel architectures can be found in [Falcão et al., 2009b; Seo et al., 2007].

## 8. References

Brack, T.; Alles, M.; Kienle, F. & Wehn, N. (2006). A Synthesizable IP Core for WIMAX 802.16E LDPC Code Decoding, *Proceedings of the IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications,* pp. 1—5, September 2006.

Falcão, G.; Silva, V.; Sousa, L. & Marinho, J. (2008). High coded data rate and multicodeword WiMAX LDPC decoding on the Cell/BE. *IET Electronics Letters*, Vol. 44, No. 24, November 2008, pp. 1415–1417.

Falcão, G.; Sousa, L.; Silva, V. & Marinho, J. (2009a). Parallel LDPC Decoding on the Cell/B.E. Processor, *Proceedings of the 4th International Conference on High Performance and Embedded Architectures and Compilers (HiPEAC 2009)*, pp. 389–403, Paphos, Cyprus, January 2009, Lecture Notes in Computer Science, Springer, Vol. 5409.

Falcão, G.; Silva, V. & Sousa, L. (2009b). How GPUs Can Outperform ASICs for Fast LDPC Decoding, *Proceedings of the 23rd International Conference on Supercomputing (ICS)*, New York, USA, June 2009, pp. 390–399

Gallager, R. (1962). Low-Density Parity-Check Codes. *IRE Transactions on Information Theory*, Vol. 8, No. 1, January 1962, pp. 21–28.

Guilloud, F.; Boutillon, E. & Danger, J.-L. (2003). λ-min decoding algorithm of regular and irregular LDPC codes. *Proceeding of the 3rd Int. Symp. Turbo Codes Relat. Topics*, pp. 1–4, September 2003.

Hofstee, H. (2005). Power Efficient Processor Architecture and the Cell Processor, *Proceedings of the 11th International Symposium on High-Performance Computer Architectures (HPCA)*, pp. 258–262, San Francisco, CA, USA, February 2005.

IEEE P802.16e/D12 (2005). Draft IEEE Standard for Local and Metropolitan Area Networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, October 2005.

International Business Machines Corporation (2007). *Synergistic Processor Unit Instruction Set Architecture*, IBM Systems and Technology Group.

Lin, S. & Costello, D. (2004). *Error Control Coding*, Prentice Hall, second edition, ISBN-13: 978-0130426727.

Liu, C.-H.; Yen, S.-W.; Chen, C.-L.; Chang, H.-C.; Lee, C.-Y.; Hsu, Y.-S. & Jou, S.-J. (2008). An LDPC Decoder Chip Based on Self-Routing Network for IEEE 802.16e Applications. *IEEE Journal of Solid-State Circuits*, Vol. 43, No. 3, March 2008, pp. 684–694.

Mackay, D. & Neal, R. (1996). Near Shannon Limit Performance of Low Density Parity Check Codes. *IEE Electronics Letters*, Vol. 32, No. 18, August 1996, pp. 1645–1646.

Mackay, D. (1999). Good Error-Correcting Codes Based on Very Sparse Matrices. *IEEE Transactions on Information Theory*, Vol. 45, No. 2, March 1999, pp. 399–431.

Seo, S.; Mudge, T.; Zhu, Y. & Chakrabarti, C. (2007). Design and Analysis of LDPC Decoders for Software Defined Radio, *Proceedings of the IEEE Workshop on Signal Processing Systems,* pp. 210–215, October 2007.

Sony Computer Entertainment Incorporated (2005). *SPU C/C++ Language Extensions*, Sony Corporation.

Tanner, R. (1981). A Recursive Approach to Low Complexity Codes. *IEEE Transactions on Information Theory*, Vol. 27, No. 5, September 1981, pp. 533–547.

7

# Design and Implementation WiMAX Transceiver on Multicore Platform

Qing Wang[1], Yonghua Lin[1], Jianwen Chen[1],
Da Fan[1,2] and Ling Shao[1]
1. IBM *China Research Lab, 2. Tsinghua University*
*Beijing, P.RChina, 100193*

Before we design the WiMAX Base Station (BS), three questions we should answered. The first is what kind of system parameters will be selected? The second is which platform will be used for the BS design. And the third one is which algorithm will be selected for some modules that are not defined in 16e standard, especially for the receiver modules, such as synchronization, channel estimation, and STC (Space Time Coding) decoder, etc. When all the above questions obtain proper answers, we can start the BS design and implementation on specific platform to achieve aimed system performance. In this chapter, we will focus on the PHY (physical Layer) design of WiMAX BS.

## 1. System Parameters

The setting of system parameters depend on the system requirements, such as the system bandwidth, the coverage radius and the users supported, etc.. WiMAX forum proposes some profiles for system specifications' setting [1]. For the WiMAX BS we designed, the system profile is defined as Table 1. We only set the profile under UL_PUSC zone as an example. That is we implement the BS receiver and MS transmitter to illustrate the design process.

| Parameters | Values |
|---|---|
| System Parameters | |
| System Frequency Band | 3.3GHz |
| System Channel Bandwidth (BW in MHz) | 10 |
| Sampling Frequency (Fs in MHz) | 11.2 (n=28/25) |
| Channel Model | SUI3 and AWGN |
| OFDMA Symbol Parameters | |
| FFT Size (NFFT) | 1024 |
| Sub-Carrier Frequency Spacing (kHz) (Fs/NFFT) | 10.94 |
| Useful Symbol Time (Tb = 1/f) | 91.4 μs |
| Guard Time (Tg =Tb/8) | 11.4 μs |
| OFDMA Symbol Duration (Ts = Tb + Tg) | 102.9 μs |

| | | |
|---|---|---|
| UL PUSC | One Slot | 1 subchannel<br>3 OFDMA symbols |
| | Null Subcarriers | 184 (92 left, 91 right, 1 DC) |
| | Pilot Subcarriers (symbol) | [420; 0; 420] |
| | Data Subcarriers (symbol) | [420; 840; 420] |
| | Subchannels | 35 |
| | Pilot Subcarriers (subchannel) | [12; 0; 12] |
| | Data Subcarriers (subchannel) | [12; 24; 12] |
| | Number of tiles | 210 |
| | Subcarriers per tile | 4 |
| | Tiles per subchannel | 6 |
| **Link Parameters** | | |
| Link Direction | | uplink |
| Frame duration (T_frame) | | 5 ms |
| Number of Frames (per second) | | 200 |
| Number of OFDMA Symbols in one frame | | 48 |
| **Channel Coding** | | |
| FEC | | CC: 1/2 |
| Modulation | | 16 QAM |
| Sub-Channels | | Full |
| **STBC Parameters** | | |
| Tx number | | 1&2, single channel, STC Matrix A and Matrix B |
| Rx number | | 1&2, , single channel, STC Matrix A and Matrix B |
| STC Decoding | | SFBC |

Table 1. Parameters of WiMAX BS PHY (base band without Ranging, UL_PUSC Zone)

## 2. Platform Selection

In traditional, the base station (BS) market is dominated by the proprietary hardware platforms, which are composed of DSPs and FPGA, or CPU and DSPs. Although there are ASIC solutions for BS applications, most of them are used in picocell or microcell. For the macrocell application, processing capability is always a key issue that should be considered during the BS platform selection. For the WiMAX base station (BS) design, the traditional platforms can be classified as [2]:

- CPU+DSP array+FPGAs

  The CPU is used for control and system management. It can be a simple processor, such as ARM or MIPS chip. DSP array is responsible for complex calculation and FPGAs for acceleration. Currently, most of BS are based on this kind of platform. The advantages of this kind of platform are low power consumption, good development ecosystem, and rich libraries and tools. For example, TI and ADI can provide high performance DSPs for wireless communication. And Xilinx and Altera have their own solution for WiMAX. The disadvantages of the platform are less scalability and flexibility.

- CPU+ASIC

  The solution based on ASIC is easy for development and implementation. However, cost should be considered and the system is difficult to upgrade due to the appropriative chip.
- Customerized Chipset

  It's proprietary platform, generally with high cost, less flexibility and interoperability.

Processing capability, power consumption and cost are the three main factors when considering a BS design. From the traditional view, general IT (information technology) platform, such as blade servers, is not suitable for wireless BS applications. One reason is the processing capability. Compared with DSP or FPGA, the general purpose processor (GPP) has lower performance for complex computation. The other main reason is the power consumption. However, the situation has changed for the occurrence of multicore processor. With the specified architecture, specially designed instruction set, and optimized compiler, the general purpose processor with multicore has powerful processing capability in many applications, including wireless application. For example, Cell Broadband Engine (BE) has 256GFlops processing capability at 3.2GHz [3]. And the Integrated Performance Primitives (IPP) provided by Intel (multicore ready) achieves good performance for multimedia and data processing applications [4]. Considering the total power consumption of the system, the power consumed by baseband is a small portion. Therefore, the general IT platform with multicore or multithread is a good candidate for open wireless architecture (OWA) for its powerful processing capability, flexibility, scalability and interoperability.
In this Chapter, we select the Cell BE as the platform to design and implement the WiMAX BS (base band).

## 3. System Structure and Functions

### 3.1 System Structure

The system structure of the proposed BS transceiver (baseband) is described as Fig. 1 [2].
Fig. 2 depicts the uplink subframe with partially used subchannel (PUSC) subcarrier assignment scheme. The uplink subframe is shown in Fig. 2a. Each uplink transmitter is assigned several subchannels to transmit its burst in a timedivision- multiplexing (TDM) manner. Note that the ranging subchanel is not considered in our design. The uplink supports 35 subchannels where each transmission uses 48 data carriers as the minimal block of processing. A slot in the uplink is composed of three symbols and one subchannel, within each slot, there are 48 data subcarriers and 24 fixed-location pilot subcarrier. Multiple subchannels (slots) can be allocated to each user, with one subchannel being the minimum resource that can be allocated to a user. In the frequency domain, a subchannel is constructed from six uplink tiles, each tile has four successive active subcarriers and modulated with a mix of data and pilots over three OFDMA symbols. The configuration of a tile is illustrated in Fig. 2b. Here, we only consider one user and the user occupies the entire system bandwidth.

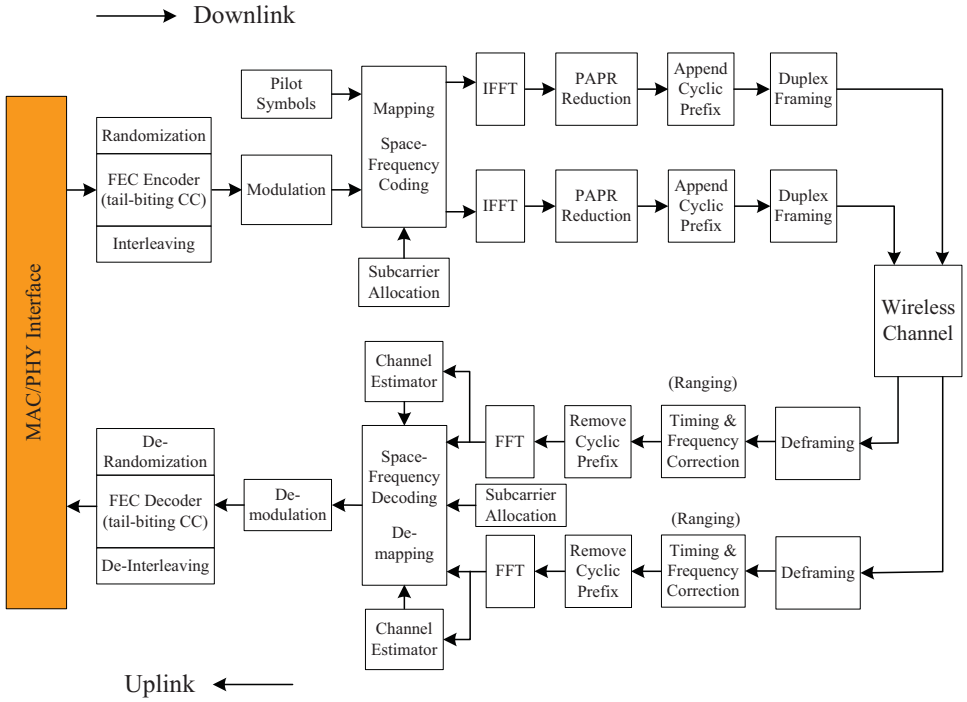



Fig. 1. WiMAX BS PHY System Structure (baseband)

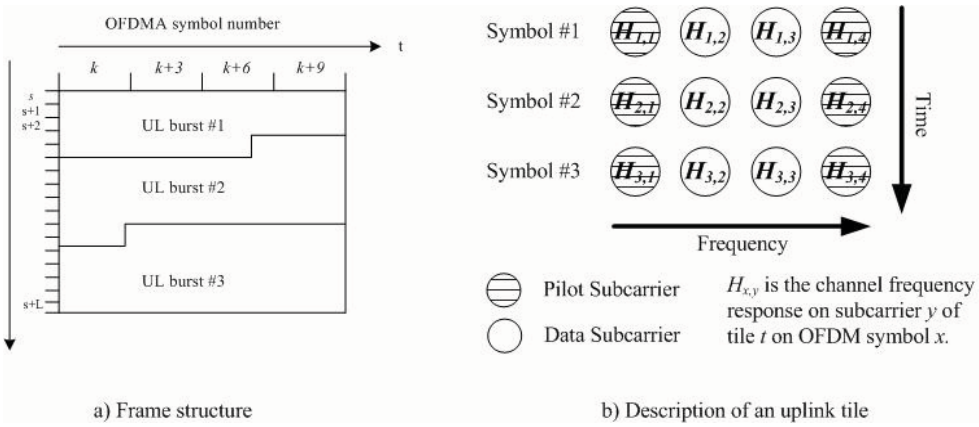



Fig. 2. Frame and Tile Structure

### 3.2 Signal Model

In the downlink shown in Fig.1, after FEC (Forward Error Control) coding, modulation, zone permutation, OFDMA modulation and cyclic prefix (CP) insertion, the time-domain samples of an OFDM symbol can be obtained from frequency-domain symbols as

$$x(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X(k) e^{j2\pi nk/N} \qquad -N_{CP} \le n \le N-1 \tag{1}$$

where X (k ) is the modulated data on the kth subcarrier of one OFDM symbol, N is the number of subcarriers and $N_{CP}$ is the length of cyclic prefix.
The impulse response of multi-path channel can be approximately denoted as:

$$h(\tau,t) = \sum_{l=1}^{L} \alpha_l(t)\delta(\tau - \eta) \tag{2}$$

where L is the total number of paths, and $\alpha_l$ and $\tau_l$ are the complex gain and time delay of the lth path. It supposes that the signals are transmitted over a quasi-static multipath fading channel, that is to say, the channel varies much slowly and the fading coefficients can be assumed to be constant during the OFDM block [5].
Assuming perfect time and frequency synchronization, the model of received signal at the BS after removal of the CP can be written as

$$y(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} H(k)X(k)e^{j2\pi nk/N} + z(n) \qquad 0 \le n \le N-1 \tag{3}$$

where H(k) is the channel frequency response at the kth subcarrier and z(n) is the additive white complex Gaussian noise (AWCGN).

### 3.3 Algorithm Selections

For the transmitter of WiMAX PHY, the algorithm of each module, such as FEC coding, modulation, map constellation, etc., is mature relatively. Thus we focus on the discussion of algorithm selection for receiver block in this section.

#### 3.3.1. Synchronization

Timing and frequency synchronization are two important tasks needed to be performed by the receiver. Through the timing and frequency offset estimation and correction, the effects of ISI (inter symbol interference) and ICI (inter-carrier interference) can be reduced.
In the presence of symbol timing offset (STO) and carrier frequency offset (CFO), equation (3) should be modified as follows:

$$y(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} H(k)X(k)e^{\frac{j2\pi k(n-\theta)}{N}} e^{\frac{j2\pi\varepsilon n}{N}} + z(n) \qquad 0 \le n \le N-1$$

where $\theta$ is the normalized STO with respect to the sample duration, and $\varepsilon$ denotes the normalized CFO as a fraction of the intercarrier spacing.
A number of approaches to estimate timing and frequency offset in OFDM systems have been presented in the literature. Some operate in the time domain [6][7], while others use the cycle prefix or the cyclostationarity of OFDM transmissions (e.g., Van de Beek algorithm [8])

to gain information about the symbol timing and frequency offset. As theWiMAX standard, the preamble in OFDMA-mode does not have the repeating pattern similar to that in OFDM-mode. And only uplink subframe is considered in our design. Therefore, in this paper, ML algorithm based on the CP [8] is chosen to achieve the symbol timing and carrier frequency synchronization.
Through the algorithm introduced in [8], we can obtain the estimation of $\mu$ and " according to the following two equations:

$$\hat{\theta}_{ML} = arg\max_{\theta}\{|\gamma(\theta)| - \phi(\theta)\} \tag{4}$$

$$\hat{\varepsilon}_{ML} = -\frac{1}{2\pi}\gamma(\hat{\theta}_{ML}) \tag{5}$$

where $\gamma(m)$ is a sum of L consecutive correlations between pairs of samples spaced $N$ samples apart. The term $\phi(m)$ is an energy term, independent of the frequency offset ε .
Once the STO and CFO are estimated, the received time samples can be corrected as follows:

$$y(n)_{corrected} = y(n+\hat{\theta}_{ML})e^{\frac{-j2\pi n\hat{\varepsilon}_{ML}}{N}} \tag{6}$$

### 3.3.2. Channel Estimation

It is well known that it is necessary to remove the amplitude and phase shift caused by the channel.
Based on the uplink tile structure, shown as Fig.2b, the pilot-aided channel estimation methods can be employed, which consist of algorithms to estimate the channel at pilot frequencies and to interpolate the channel. The estimation of the channel at the pilot frequencies can be based on least square (LS), minimum mean-square (MMSE) or least mean-square (LMS). Though MMSE has been shown to perform much better than LS, it needs knowledge of the channel statistics and the operating SNR [9]. The interpolation of the channel can depend on linear interpolation, second order interpolation, low-pass interpolation, spline cubic interpolation, and time domain interpolation. Considering the tradeoff between feasibility of implementation and system performance, we choose linear interpolation in time and frequency on a tile-by- tile basis for each subchannel.
When the data and pilot information has been assembled as shown in Fig. 2b, it is possible to calculate $H_{11}$, $H_{14}$, $H_{31}$ and $H_{34}$ using the equation:

$$\hat{H}_p(t,m) = \frac{Y_p(t,m)}{S_p(t,m)} \tag{7}$$

for the mth OFDMA symbol of the tth tile where:
$Y_P(t,m)$ is the pth received pilot subcarrier
$S_P(t,m)$ is the pth transmitted pilot subcarrier.
We omit the index of receive antenna here, since channel estimation for each receive antenna is performed independently. Subsequently, frequency domain linear interpolation is performed to calculate channel estimates using the following equations:

$$\hat{H}_{1,2}=\frac{1}{3}(\hat{H}_{1,4}-\hat{H}_{1,1})+\hat{H}_{1,1} \qquad \hat{H}_{1,3}=\frac{2}{3}(\hat{H}_{1,4}-\hat{H}_{1,1})+\hat{H}_{1,1}$$
$$\hat{H}_{3,2}=\frac{1}{3}(\hat{H}_{3,4}-\hat{H}_{3,1})+\hat{H}_{3,1} \qquad \hat{H}_{3,3}=\frac{2}{3}(\hat{H}_{3,4}-\hat{H}_{3,1})+\hat{H}_{3,1} \tag{8}$$

where $H_{m,k}$ is the channel frequency response at the kth subcarrier of the mth OFDM symbol and $\hat{H}_{m,k}$ Hm k is the estimation of $H_{m,k}$ .

Finally, time domain linear interpolation is achieved as follows:

$$\hat{H}_{2,1}=\frac{1}{2}(\hat{H}_{1,1}-\hat{H}_{3,1}) \qquad \hat{H}_{2,2}=\frac{1}{2}(\hat{H}_{1,2}-\hat{H}_{3,2})$$
$$\hat{H}_{2,3}=\frac{1}{2}(\hat{H}_{1,3}-\hat{H}_{3,3}) \qquad \hat{H}_{2,4}=\frac{1}{2}(\hat{H}_{1,4}-\hat{H}_{3,4}) \tag{9}$$

When all of the channel estimates have been formed, these estimated values are transmitted to the space-frequency decoding module for the data detection using ML method.

### 3.3.3. SFBC

A user-supporting transmission using transmit diversity configuration in the uplink, shall use a modified uplink tile. The pilots in each tile shall be split between the two antennas and the data subcarriers shall be encoded in pairs after constellation mapping, as depicted in Fig. 3. Because this is applied in the frequency domain (OFDM carriers) rather than in the time domain (OFDM symbols), we note it as space-frequency block coding (SFBC) [10].

Defined $H_{m,k}^{(i,j)}$ as the channel frequency response at the $k$ th subcarrier of the $m$ th OFDM symbol corresponding to the $i$ th transmit and the $j$ th receive antenna pairs, and $Z_{m,k}^{j}$ as the frequency response of the AWCGN on the $k$ th subcarrier of the $m$ th OFDM symbol at antenna $j$ respectively, on the assumption that the neighboring subcarriers have the same frequency response, the estimation of $X_1$ and $X_2$ are:

$$\hat{X}_1=\arg\min_{\hat{X}_1\in X}\left[\left(\sum_{i=1}^{2}\sum_{j=1}^{2}\left|\hat{H}_{1,2(3)}^{(i,j)}\right|^2-1\right)\left|\hat{X}_1\right|^2+d^2(\tilde{X}_1,\hat{X}_1)\right]$$

$$\hat{X}_2=\arg\min_{\hat{X}_2\in X}\left[\left(\sum_{i=1}^{2}\sum_{j=1}^{2}\left|\hat{H}_{1,2(3)}^{(i,j)}\right|^2-1\right)\left|\hat{X}_2\right|^2+d^2(\tilde{X}_2,\hat{X}_2)\right] \tag{10}$$

where

$$\tilde{X}_1=\sum_{i=1}^{2}\sum_{j=1}^{2}\left|\hat{H}_{1,2(3)}^{(i,j)}\right|^2X_1+\sum_{j=1}^{2}\left(\hat{H}_{1,2}^{(1,j)}\right)^*Z_{1,2}^{(j)}+\sum_{j=1}^{2}\hat{H}_{1,3}^{(2,j)}\left(Z_{1,3}^{(j)}\right)^*$$
$$\tilde{X}_2=\sum_{i=1}^{2}\sum_{j=1}^{2}\left|\hat{H}_{1,2(3)}^{(i,j)}\right|^2X_2+\sum_{j=1}^{2}\hat{H}_{1,2}^{(2,j)}\left(Z_{1,2}^{(j)}\right)^*+\sum_{j=1}^{2}\left(\hat{H}_{1,3}^{(1,j)}\right)^*Z_{1,3}^{(j)} \tag{11}$$

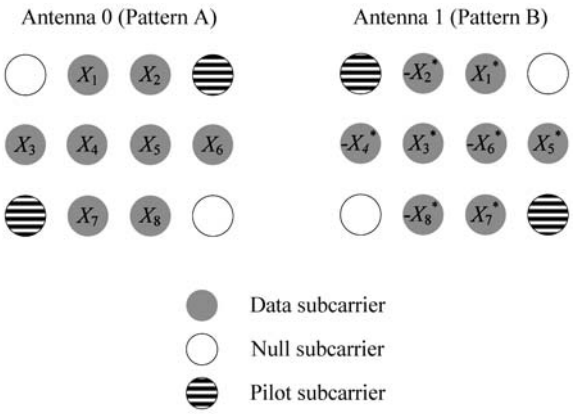



Fig. 3. Pilots and Data Subcarriers in SFBC Mode

## 4. Implementation on Cell BE

### 4.1 Cell Processor

Cell processor is proposed and designed as the engine of the PlayStation 3 of Sony initially. But as a powerful, all-purpose multiprocessor, Cell can be expect to be much potential in other areas. A single chip Cell processor contains one PowerPC Processor Element (PPE) and eight Synergistic Processor Elements (SPE). The PPE unit on Cell is a general purpose 64-bit RISC core with 2-way hardware multithreading, used for operating systems and system control, and 8 SPE cores are optimized for compute-intensive, single-precision, floating-point workloads. These units are interconnected with a coherent on-chip element interconnect bus (EIB). The system frequency of Cell is 3.2GHz and the computation capability is 256GFlops [3][11].

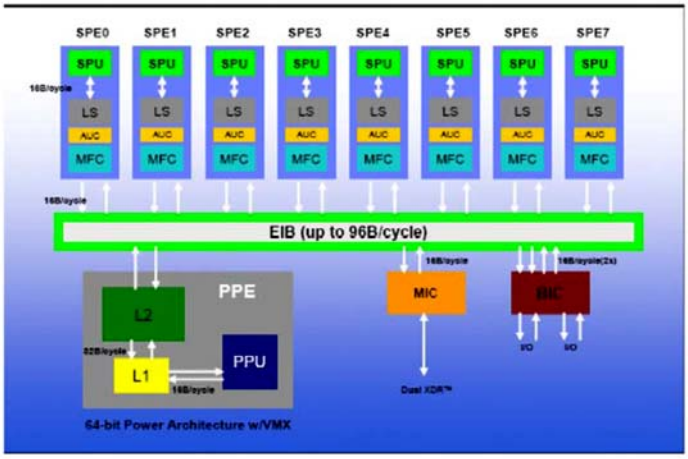



Fig. 4. Block Diagram of Cell Processor and Workload Partition (source: cell diagram from [11])

### 4.2 Programming on Cell

Cell processor is a kind of heterogeneous multicore processor. Its programming model is novel, see [12] in details. In summary, programming on Cell includes two main points. One is the programming on SPE, especially optimization on SPU since SPE acts as computation accelerator. It has special chip architecture and instruction sets to support such acceleration. The other is the communication between PPE and SPE, and communications among multiple SPEs. In this section, we will focus on the discussion about optimization on SPE (SPU). The communication mechanism of PPE and SPE will be contained in the introduction about software framework design.

For the optimization on Cell, it includes two aspects. One is the processing speed, evaluated by the number of cycle. The other is the local store consuming since each SPU only have 256KB local store. We should make balance between these two factors during optimization. If the computation capability is critical for one component while the buffer and code size are small, we can scarify some local store for achieving high computation performance and vice versa. In our case, for most components, limited local store is more troubled than computation capability. In general, it can solved by good coding design, optimization and local store overlay. Some general optimization techniques on Cell are listed as follows[17][18]:

- Reduce Branch

  Branch can significantly influence the efficiency of the SPU since SPU is an in-order processor with no branch prediction, any judgment will result in the SPU stall. Using the compare-select function instead of short judgment function is a good optimization method for most branches.
- Access Local Store pattern

  The best assess pattern for SPU is data and structure aligned with vector operation. The Scalar and unaligned access will result in many additional instructions for data aligned and scalars extracted from vectors. In some case, we can operate the scalar as the vector. This method solves the data access problem of the SPU which can not be made as SIMD pattern.
- SIMD Accelerating

  SIMD (single instruction multiple data) is a very useful accelerating technique for SPU. It generally has 4—8 times speed-up rate.
- Pipeline and Dual-issue

  Each instruction has its latency and Stall cycles which will influence the efficiency of the SPU due to the dependency. If two conjoint instructions can be placed in the different pipeline with no dependency, the two instructions can be dual-issue.

### 4.3 Workload Analysis and Optimization

#### 4.3.1 Workload Analysis

From the theoretical analysis, we know the modules of uplink, such as channel decoding, channel estimation and SFBC, consume most of the computation resource. They are the modules with heavy workload. This conclusion is also verified by workload test on Cell. Table 2 shows the workload of each module of uplink for processing 3 OFDMA symbols. The test runs on Cell BE simulator-Mambo with Cell SDK2.1. The cycle numbers of "CP remove" module and "channel estimation" module are for one antenna. The "viterbi" module

is 1/2 data rate and the constraint length is 7. We note that the Viterbi, deinterleave and SFBC are the top three modules with heavy workload. And the other modules, such as channel estimation, derandomize and demodulation modules, do not match the throughput requirement without optimization. Thus we need to optimize those modules to meet the targeted 20Mbps throughput.

| Module | Original Cycles | Optimized Cycles | Speed-up Rate |
|---|---|---|---|
| CP remove | 93261 | 10919 | 8.54 |
| Channel est. | 260812 | 19420 | 13.43 |
| SFBC | 1413688 | 162179 | 8.72 |
| Demodulation | 267175 | 41159 | 6.49 |
| Deinterleave | 3622165 | 154506 | 23.44 |
| Viterbi | 4728180 | 343227 | 13.78 |
| Derandomize | 278893 | 8031 | 34.73 |

Table 2. Workload for Modules of Receiver

For the modules of downlink, the result of workload testing depicts as Table 3. The test environment and data length are the same as that of uplink. The initial length of data is 3354 bits, containing 3 OFDMA symbols.

| Module | Original Cycles | Optimized Cycles | Speed-up Rate |
|---|---|---|---|
| Randomize | 278893 | 6139 | 45.43 |
| Convolutional coding | 190816 | 79979 | 2.39 |
| Interleave | 3729682 | 146975 | 25.38 |
| Modulation (16QAM) | 144543 | 76137 | 1.90 |
| Zone Permutation | 372386 | 187054 | 1.99 |
| CP Insert | 209997 | 24551 | 8.55 |

Table 3. Workload for Modules of Transmitter

We use convolutional code (data rate =1/2, constraint length = 7) for channel coding and the modulation is 16QAM. Except the interleave module, the other modules of downlink have the same level workload before optimization. Compared with the workload of uplink, the modules of downlink consume less computation resource. For the FFT and IFFT used for system, we will use the library provided by Cell SDK. There is no optimization work on these two modules. Hence we did not list their workloads here.

#### 4.3.2 Workload Optimization

Based on the workload analysis, we optimize each module to meet the throughput requirement we pre-set. That is 20Mbps processing capability for both downlink and uplink. In our application, each technique mentioned above is used and the speed-up rate of each module is shown in Table 2 and Table 3 for uplink and downlink respectively.
During the optimization, we should tradeoff between computation performance (cycles) and

local store consumption. For the computation critical module, such as Viterbi decoding, we will scarify the local store to obtain the smaller cycles; While for the local store critical module, we will try to save buffers instead of achieving highest performance. Therefore, when we refer to the performance of each module, we should consider both the number of cycles and the consuming of local store, which is very important for workload partition on different SPEs. The optimized results shown in Table 2 and Table 3 are not the best one. We just optimize them till they can meet our design requirements. They still have potential optimization space.

Based on the optimization results and local store consumptions, the workload can be partitioned to five SPEs, in which two SPEs for downlink and three SPEs for uplink. PPE is responsible for SPE control and management. So one Cell BE chip can process both uplink and downlink with 20Mbps throughput in theory. Figure 5 depicts the workload partition of Cell.

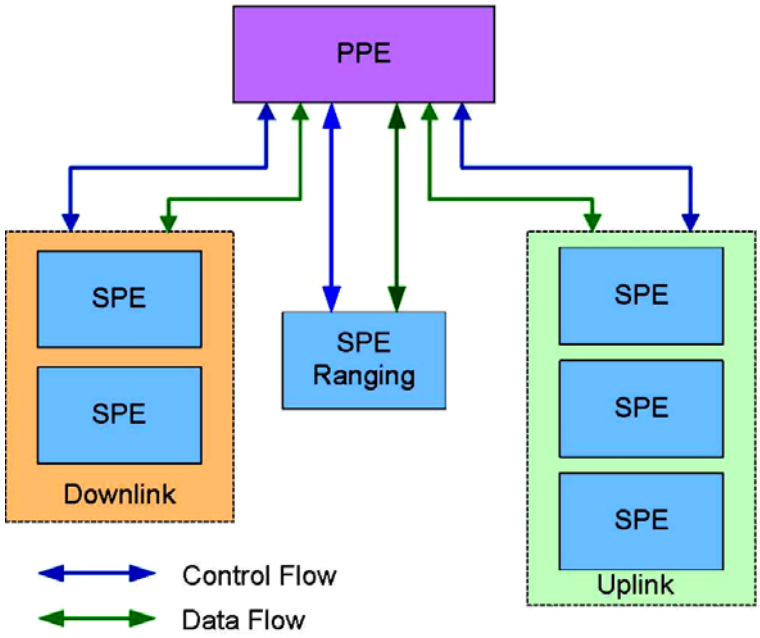



Fig. 5. Workload Partition of Cell

### 4.4 Framework Design

For the software framework design, we consider three scenarios: sequential framework, PPU synchronization and SPU synchronization. In the sequential framework, PPU is used as a controller for SPU control and data management. SPU is used for data processing. The data is stored in the main memory. SPU fetches the data from the main memory, calculates them and then sends the computation results back. Sequential framework is the simplest one with low efficiency. It can not satisfy the 20Mbps throughput requirement. Therefore,

we only use this framework to verify the system correctness at the beginning of system integration. For the PPU synchronization framework, PPU is used to manage the synchronization of SPUs. This results in the PPU to take heavy workload. If the system (Cell blade server, named as QS20, containing two Cell Processor) wants to support 3 sectors, PPU becomes the bottleneck of system. Therefore, we do not adopt this framework. SPU synchronization is the framework we used in the current system, shown as Fig. 6.
In this design, different modules will work in parallel. SPUs will manage their synchronization through messages passing. Since there is no feedback path in the data flow of both uplink and downlink, pipeline can be used in the framework design. There are two different levels of pipeline:

- SPU Level Pipelining. This level pipelining can be realized by double the input and output buffers.The double buffers are allocated on main memory.
- Functional Level Pipelining. The functional units in one SPU can also work in pipelining, but it is heavily dependent on the algorithms and local store limitation. Only when the local store can support double buffer for both input and output, the pipelining can be used. Functional level pipelining can overlap the time consumption of DMA tasks and computation tasks.

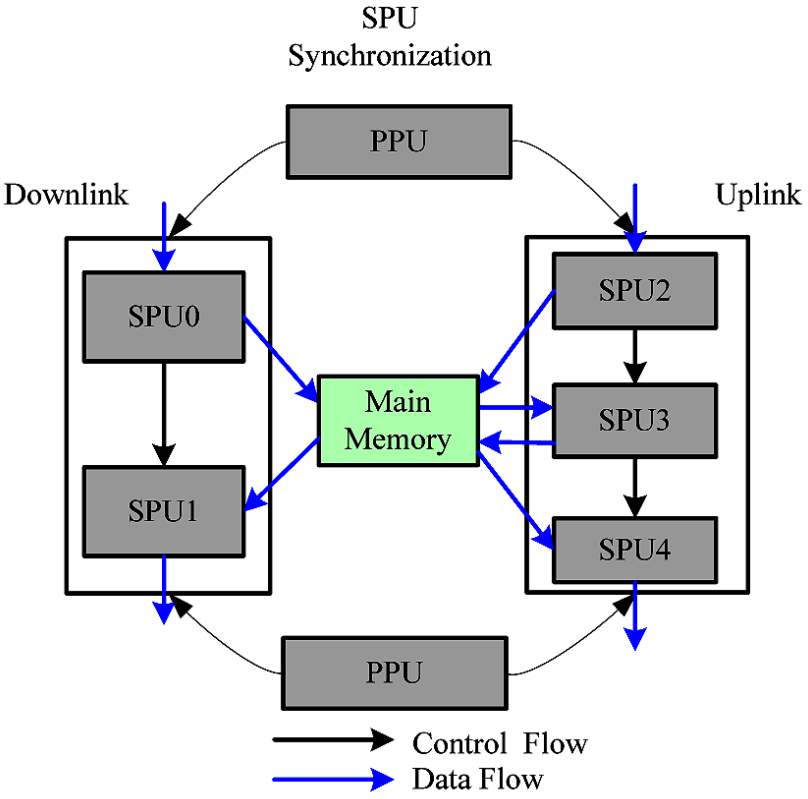



Fig. 6. Software Framework for One Sector

## 5. Simulation Results and System Performance

The system is implemented on IBM Cell blade server, named QS-20, which has two Cell B.E. processors (a 2-way SMP) operating at 3.2 GHz. We use $2R_x$ X $2T_x$ MIMO technique and the system parameters are set as Table 1. The uplink bandwidth is 10MHz, the subcarrier frequency spacing $\Delta f$ is 10.94kHz, N=1024, and $N_{CP}$ =128. The following parameters are also assumed: 1/2 convolutional coding with constraint length of 7 and generator polynomial matrix of [133 171]. A discrete channel model based on the Stanford University Interim 3 (SUI-3) [13] model is used, which represents a low delay spread case with $\tau_{rms} = 0.264\mu s$ (low frequency selectivity). The bit-error rate (BER) performance is evaluated by averaging over 200 frames, and each frame has 3 OFDMA symbols. Figure 7 is the simulation results at different stages of the system level simulator.

We evaluate the system performance from two aspects. One is the throughput of uplink and downlink, the other is the system BER. The throughput demonstrates the system processing capability. Table 4 shows the throughput test results. Each sector can achieve 20Mbps throughput whether for downlink or uplink. The total throughput of one QS20 will exceed 60Mbps.

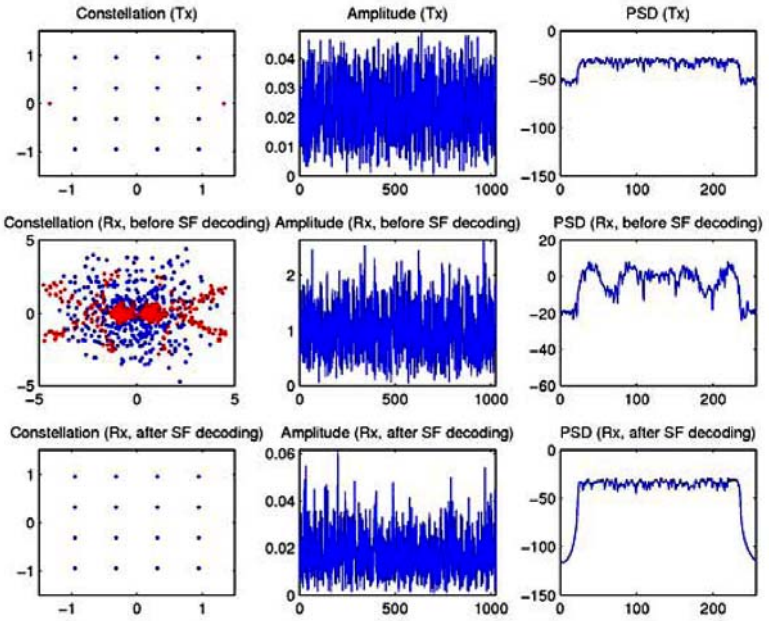



Fig. 7. Simulation Results at Different Stages of the System Level Simulator

| Throughput | Downlink (Mbps) | Uplink (Mbps) |
|---|---|---|
| Sector1 | 24.409414 | 20.970757 |
| Sector2 | 25.042559 | 21.517656 |
| Sector3 | 24.442323 | 21.473296 |

Table 4. Throughput of Three Sectors on One QS20

BER performance reflects the correctness of system design and the system precision. Figure 8 is the BER results tested on QS20 and X86 processor (Intel Xeron@2.8GHz) respectively. We tested both AWGN channel and Rayleigh channel on X86 and Cell platform. The results indicate that the BER performances are almost the same for X86 platform and cell platform whether under AWGN channel or Rayleigh channel.

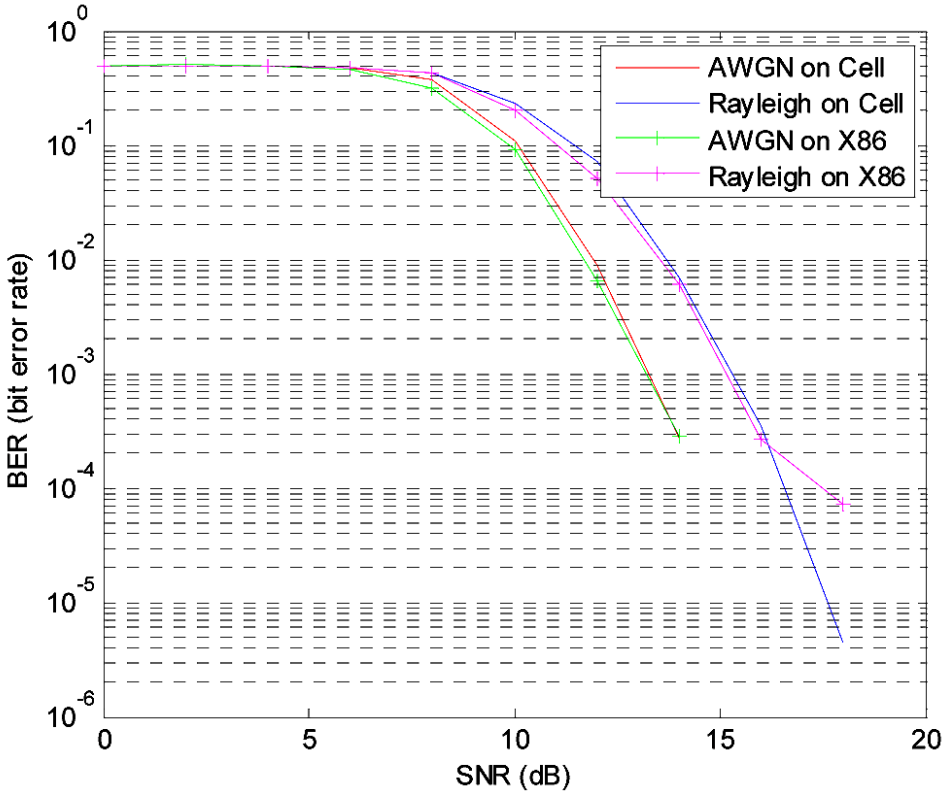



Fig. 8. BER Performance for AWGN and Rayleigh Channel under Different Platforms

## 6. Summary

In this chapter, we propose the possible solutions for the issues during WiMAX BS implementation, such as the platform selection, algorithm selection, and performance optimization. And we design and implement a WiMAX BS (PHY, baseband) on Cell processor as an example for illustration. The system requirements decide the platform selection, and the system processing capability and system performance requirements are the main factors considered during the BS design. The performance optimization can be classified as individual module optimization and system framework optimization. Both of them heavily depend on system hardware structures. Although different platforms have their specific optimization methods according to the system structures, efficient

communications between each modules and acceleration for some key modules with heavy workloads are general methods that should be considered.

## 7. Reference

[1] WiMAX Forum™ Mobile System Profile 3-- Release 1.0 Approved Specification (Revision 1.7.1: 2008-11-07), WiMAX Forum.

[2] Qing Wang, Da Fan1, Jianwen Chen, Yonghua Lin and Zhenbo Zhu (2008). WiMAX BS Transceiver Based on Cell Broadband Engine, Proceedings of IEEE International Conference on Circuits & Systems for Communications, May 2008.

[3] J. A. Kahle, M. N. Day, H. P. Hofstee, C. R. Johns, T. R. Maeurer, and D. Shippy (2005). Introduction to the cell multiprocessor, IBM Journal of Research and Development, vol. 49, no. 4/5, 2005.

[4] Intel (2003). Performance benchmarks for intel integrated performance primitives-white paper, 2003.

[5] Qing Wang, Da Fan, YongHua Lin, Jianwen Chen, Zhenbo Zhu (2008). DESIGN OF BS TRANSCEIVER FOR IEEE 802.16E OFDMA MODE, Proceedings of ICASSP08, April 2008. [6] P. Moose (1994). A technique for orthogonal frequency division multiplexing frequency offset correction, IEEE Trans. on Communications, vol. 42, pp. 2908-2914, 1994. [7] T. Schmidl and D. Cox (1997). Robust frequency and timing synchronization for ofdm, IEEE Trans. on Communications, vol. 45, pp. 1613-1621, 1997.

[8] M. Sandell J. van de Beek and P. Borjesson (1997). ML estimation of time and frequency offset in ofdm systems, IEEE Trans. Signal Processing, vol. 45, pp. 1800-1805, July 1997. [9] J.-J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson (1995). On channel estimation in OFDM systems, Proc. IEEE 45th Vehicular Technology Conf., vol. 45, pp. 815-819, Chicago, IL,July 1995.

[10] K. F. Lee and D. B. Williams (2000). A space-frequency transmitter diversity technique for OFDM systems, Proc. IEEE GLOBECOM, pp. 1473-1477, San Francisco, CA, Nov.2000.

[11] IBM(2006). Cell broadband engine processor based systems-white paper, p. 3, Sep. 2006. [12] IBM(2007). Software Development Kit for Multicore Acceleration Version 3.0--Programming Tutorial.

[13] V. Erceg, K.V.S. Hari, and et al. M.S. Smith (2003). Channel models for fixed wireless applications, Contribution IEEE 802.16a-03/01, Jun. 2003.

[14] IEEE Std 802.16-2004. Part 16: Air Interface for Fixed Broadband Wireless Access Systems, ," Oct. 2004.

[15] IEEE Std 802.16e-2005. Part 16: Air Interface for Fixed, Mobile Broadband Wireless Access Systems Amendment2: Physical, Medium Access Control Layers for Combined Fixed, and Mobile Operation in Licensed Bands., ," Feb. 2006.

[16]Hassan Yagoobi (2004). Scalable OFDMA physical layer in IEEE 802.16 wirelessman, Intel Technology Journal, vol. 8, Aug. 2004.

[17]Daniel A. Brokenshire (2006). Maximizing the power of the Cell broadband engine processor: 25 tips to optimal application performance, IBM developerWorks, June 2006.

[18] IBM(2006). Cell broadband engine programming handbook v1.0, pp. 681-708, April 2006.
[19] M. Hsieh, C. Wei (1998). Channel Estimation for OFDM Systems Based on Comb-type Pilot Arrangement in Frequency Selective Fading Channels, IEEE Trans. Consumer Electron., vol. 44, no. 1, Feb. 1998.

8

# Advanced Filter Development for WiMAX Applications

*Kaixue Ma, Leyu Zhang and Kiat Seng Yeo
*School of Computer Science & Technology; Tianjin University, China
Nanyang Technological University (NTU), Singapore

## 1. Introduction

Wireless communication continues to bring increased opportunities for the radio frequency (RF) component and test equipment manufacturers. For example, according to the IEEE.802.16-2004 standard [1] [2], frequency band of 2GHz~11GHz (see Figure 1 a)) is split into three different radio frequency bands, 2.4 GHz and 3.5 GHz for licensed bands and 5.8GHz for unlicensed, each of which has unique processing requirements that are incompatible with the other frequency bands. The result will allow manufacturers of RF components and test equipment to have their products used for mass deployment. The high linearity requirements of the WiMAX communication, for example 2.4 GHz and 3.5 GHz WiMAX communication, urge the test equipment can not only acquire the fundamental signal but also its' harmonics. High performance Bandstop filter, which needs high rejection of fundament signal $f_s$ and ultra-wide passband across the harmonic up to $5f_s$ as shown in Figure 1 b), is required according to the WiMAX equipment scheme, which will be introduced in Section II.

Bandstop filter is important in the filter family [3] [4]. It is used to reject some particularly strong interfering frequency and while passing through the rest. It is widely used in cable televisions, satellite communication systems, and other communication systems. In the family of bandstop filters, lumped elements or resistor-inductor- capacitor bandstop filter is commonly used. The disadvantages including relative large size, large power consumption and possible parasitic effect limit the utilizations above the multi-gigahertz range. In fact, most of the bandstop filters operating above multi-GHz range are built by distributed transmission lines [5-14]. The bandstop filter architectures of a transmission line coupled to a grounded bandstop resonator by capacitive gaps and by parallel-line coupling are given in [5] and [6] respectively. In [7], Bell uses the quarter-wavelength L-shape resonator configuration to build up the bandstop filters. Qian and Zhuang [8] introduce a complementary relationship between a dual–mode bandpass and bandstop waveguide filter and verified the relationship by experiments. Superconducting bandstop filters using the open-ended half wavelength resonator are reported in [9] and [10]. More recently, in [11], the traditional bandstop filter using two open-ended stubs [3] is modified to bended square shape and 50% size reduction is achievable. The compact size microstrip interdigital bandstop filter [12], which is realized by loading the interdigital capacitors on the

transmission line, is presented. A close loop ring coupling to CPW transmission line by transverse slot is used for bandstop filter by using micro-electro-mechanical systems (MEMS) technology as demonstrated in [13]. In [14], the synthesizing of bandstop filter based on transmission line model is demonstrated.

In this Chapter, the new application of the bandstop filter for the nonlinear measurement for WiMAX application is introduced. A physical based lumped circuit model introduced in [16] is used for the bandstop filter design and analysis. Bandstop filters operate in 2.5 GHz and 3.5 GHz licensed WiMAX bands are developed and the physical parameters of these filters are given in this Chapter. As shown in Figure 1 b), the high performance bandstop filter with the rejection of $f_s$ more than 60dB and passband up to the fifth of the fundamental frequency are achieved simultaneously. The important leakage issue due to packaging is discussed and a solution for the packaging to avoid the leakage and further improve the performance is introduced. Additional work on WiMAX bandpass filter is also introduced.

The Chapter is arranged as follows: Firstly, the application and the requirements of the bandstop filter in WiMAX test equipment based on nonlinear measurement scheme are introduced in Section II. In Section III, the step impedance resonator section with different electric parameters and the resonant characteristics are investigated by using the circuit model. In Section IV, the step impedance resonator sections are used to build the 2.4 GHz and 3.5 GHz bandstop filter. 2.4 GHz and 3.5 GHz filters with good performance for WiMAX equipment are verified by experiments. In Section V, the leakage issues are investigated and the packaging free of leakages is designed and implemented. In section VI, the recent development work on 2.4GHz and 3.5GHz WiMAX bandpass filters is illustrated for WiMAX application. Finally, the conclusion is given in Section VII.

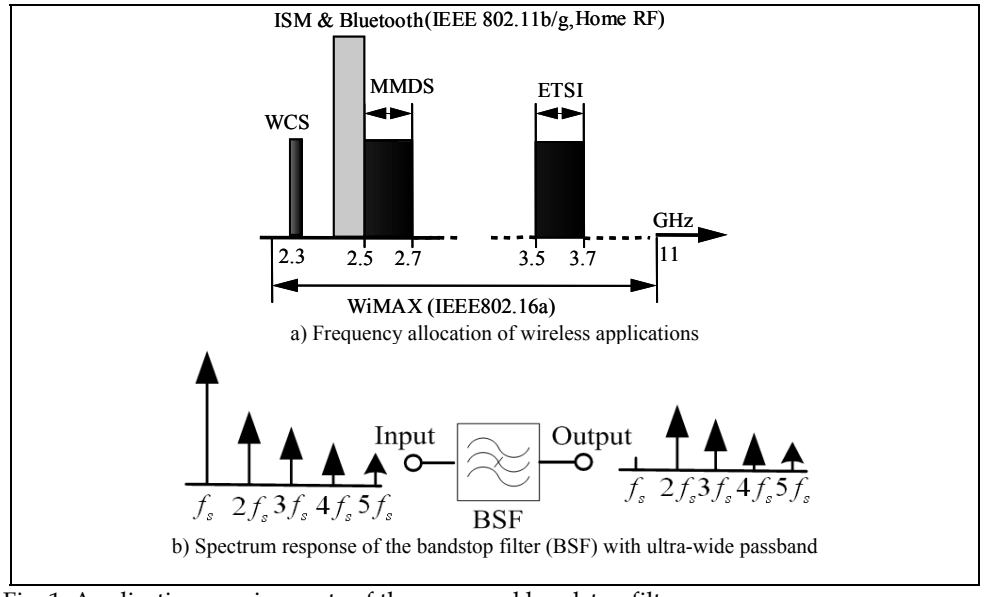



Fig. 1. Application requirements of the proposed bandstop filter

## 2. bandstop Filter Application in WiMAX Equipment

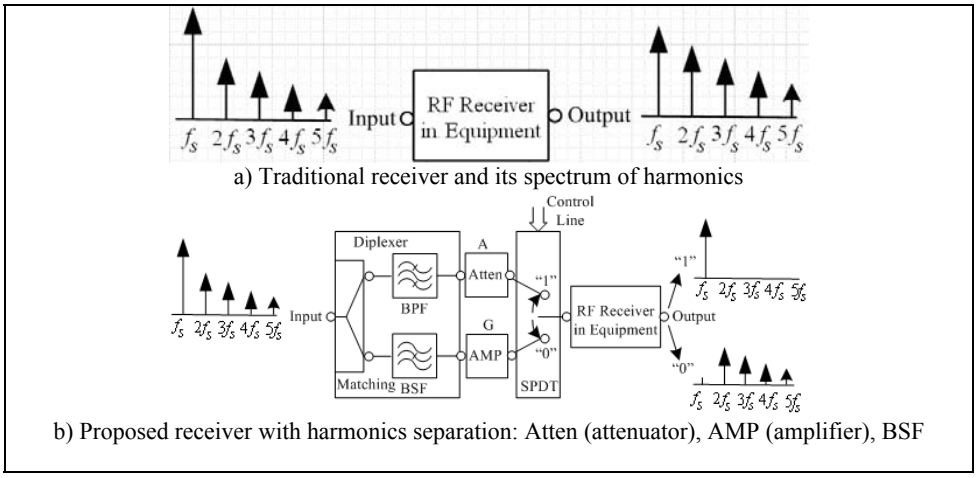



Fig. 2. Receiver front-end with/without harmonics separation

In nonlinear measurements such as measurement of WiMAX power amplifier or communication system, the equipment receiver is important to accurately capture the information of the device under testament (DUT). To avoid the nonlinear effects generated from the equipment receiver itself, which will increase the errors in measurement, both the fundamental signal and its harmonics should be within the linear dynamic range of the of the equipment receiver. However, the fundamental signal has higher power level while its harmonics has relative lower power level. The limited dynamic range of the equipment receiver becomes challenge to handle these two extreme power level conditions simultaneously. As shown in Figure 2 a), when the power level difference of the fundamental signal and its harmonics is larger than the dynamic range of broadband equipment receiver, the results from the output of equipment receiver will be either one of these two cases: case 1, only the fundamental signal is obtained and the harmonics information cannot be captured accurately due to the sensitivity of the receiver. Case 2, the fundamental signal will be suppressed and the harmonic will be increased due to the compression point of the equipment receiver itself).

In order to solve the above problems, the system scheme shown in Figure 2 b) is introduced to separate the fundamental signal and it harmonics. The fundamental signal, which has higher power level, goes to the bandpass filter and then compresses the power level of A dB by an attenuator. On the other hand, the harmonics with lower power level below the minimum receivable power level go through the bandstop filter and then are amplified by a low noise amplifier with a gain of G dB. Finally, the attenuated fundamental signal and the amplified harmonics are selected by the single-pole-double- through (SPDT) switch and send to equipment receiver (or are combined by a power combiner and send these signals to the equipment receiver). By doing so, the theoretical dynamic range of the testament receiver can be increased by around A plus G dB. As in the system shown in Figure 2 b), to

achieve accurate measurements of the nonlinear characteristics, the performance of the bandstop filter is very important. Thus, the high rejection stopband, to avoid the fundamental signal entering into the harmonic channel, and ultra-wide passband, to capture the harmonics up to fifth for nonlinear measurement, are required.

## 3. Step Impedance Resonator Section Model

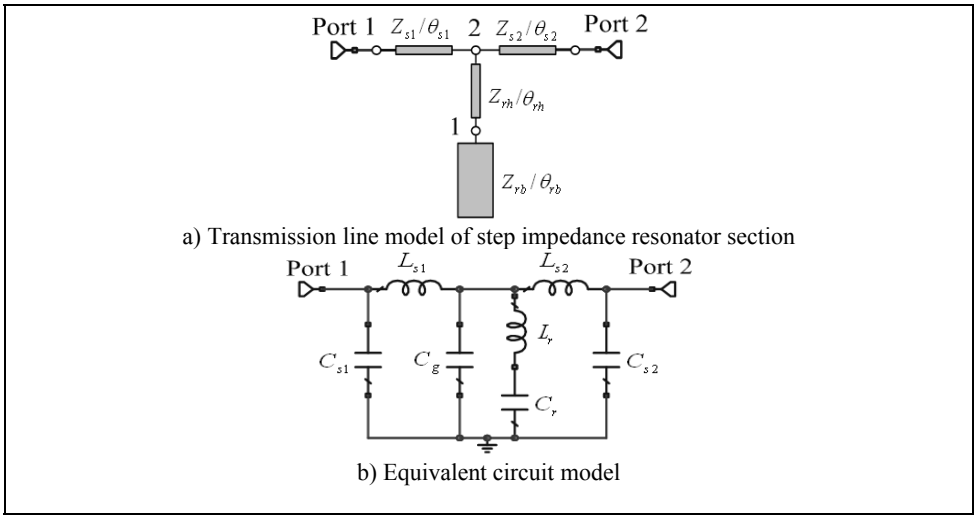



Fig. 3. Transmission line model and equivalent circuit of step impedance resonator section

The fundamental prototype block named step impedance resonator section for the bandstop filter is shown in Figure 3. It consists of a step impedance resonator and two matching section lines. The step impedance resonator is formed by an open `stub (a characteristic impedance of $Z_{ru}$ and electric length of $\theta_{ru}$) connected in serial with a high impedance section (a characteristic impedance of $Z_{rb}$ and an electric length of $\theta_{rb}$). The two matching section lines with characteristic impedance and electric length of ($Z_{s1}$, $\theta_{s1}$) and ($Z_{s2}$, $\theta_{s2}$) respectively are used for source and load (source impedance $Z_s$ and load impedance $Z_L$) matching at node "1" and "2" respectively. The input impedance of the step impedance resonator at node "2" can be given as:

$$Z_{SIR} = \frac{1 + s^2 L_r C_r}{s(C_r + C_g + s^2 L_r C_r)} \tag{1}$$

$$C_r = C_o + C_h \tag{2}$$

$$C_g = C_{s1} + C_{s2} + C_h \tag{3}$$

$$C_h = j(1 - \cos\theta_{rh})/(sZ_{rh}\sin\theta_{rh}) \tag{4}$$

$$L_r = jZ_{rh}\sin\theta_{rh}/s \tag{5}$$

$$C_{s1} = j(1 - \cos\theta_{s1})/(sZ_{s1}\sin\theta_{s1}) \tag{6}$$

$$C_{s1} = j(1-\cos\theta_{s2})/(sZ_{s2}\sin\theta_{s2}) \tag{7}$$

$$C_o = jZ_{rb}\tan\theta_{rb}/s \tag{8}$$

$$L_{s1} = jZ_{s1}\sin\theta_{s1}/s \tag{9}$$

$$L_{s2} = jZ_{s2}\sin\theta_{s2}/s \tag{10}$$

where $s = j\omega$ ($\omega$ is the angular frequency). The fundamental resonance condition is when the input impedance is equal to zero. Thus the resonance is constrained by the equations

$$f_r = 1/(2\pi\sqrt{L_rC_r}) \tag{11}$$

$$f_p = 1/2\pi\sqrt{(C_g+C_r)/(L_rC_r)} \tag{12}$$

The above model-based derivation is convenient for the initial design of the filter. To get the frequency response for the whole frequency range, the two-port network parameters are used and the transfer matrix of the model in Figure 3 b) can be calculated by

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ sC_{s1} & 1 \end{bmatrix} \begin{bmatrix} 1+\dfrac{sL_{s1}}{Z_{sir}} & \dfrac{s^2L_{s1}L_{s2}}{Z_{sir}}+s(L_{s1}+L_{s2}) \\ \dfrac{1}{Z_{sir}} & 1+\dfrac{sL_{s2}}{Z_{sir}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ sC_{s2} & 1 \end{bmatrix} \tag{13}$$

and

$$T_{11} = 1+s^2C_{s2}(L_{s1}+L_{s2})+(sL_{s1}+s^3C_{s2}L_{s1}L_{s2})/Z_{sir} \tag{14}$$

$$T_{12} = s(L_{s1}+L_{s2})+s^2L_{s1}L_{s2}/Z_{sir} \tag{15}$$

$$T_{21} = sC_{s1}+(1+s^2L_{s1}C_{s1})/Z_{sir} \tag{16}$$

$$T_{22} = 1+s^2C_{s1}(L_{s1}+L_{s2})+(sL_{s2}+s^3C_{s1}L_{s1}L_{s2})/Z_{sir} \tag{17}$$

The transmission and reflection characteristics can be determined by

$$S_{21} = \frac{2}{2T_{11}+T_{12}/Z_0+T_{21}Z_0} \tag{18}$$

$$S_{11} = \frac{T_{12}/Z_0-T_{21}Z_0}{2T_{11}+T_{12}/Z_0+T_{21}Z_0} \tag{19}$$

When $\theta_{ru} = \theta_{rb} = \theta_r$, the second harmonic or spurious resonance occurs with following constraint

$$f_{r2}/f_{r1} = (180-\theta_r)/\theta_r \tag{20}$$

where $f_{r1}$ and $f_{r2}$ are the fundamental resonant frequency and the second harmonic respectively.

It is found that the two matching section lines mainly affect the matching condition and have little effect on the resonant frequencies. Thus, the two section lines are fixed ( $Z_{s1} = Z_{s2} = 50\ \Omega$, $\theta_{s1} = \theta_{s2} = 12\ °$). The simulation results of five different conditions are compared in Figure 4 (All of the impedances are in Ω and all of the electric lengths are in degrees). As illustrated in Figure 4, the fundamental resonant frequency $f_{r1} = 1$ GHz for the cases of a), b) c), d), e) is chosen for the convenience of comparison. It is clear that the smaller the ratio of $Z_{ru} / Z_{rb}$, the shorter the electric length is needed for the same operating frequency of 1 GHz (i.e. for the same operating frequency, the size of the step impedance resonator becomes compact with reduced $Z_{ru} / Z_{rb}$). On the other hand, the first harmonic $f_{r2}$ depends on both the impedance ratio $Z_{ru} / Z_{rb}$ and the electric lengths $\theta_{rb}$ and $\theta_{ru}$. For the one special case, when the $Z_{ru} / Z_{rb} = 1$, the fundamental resonant frequency occurs at $\theta_{ru} = \theta_{rb} = 45$ and the first harmonic occurs at three times of the fundamental operating frequency. For another special case $\theta_{ru} = \theta_{rb} = \theta_r$, the relationship between the fundamental operating frequency and its first harmonic is constrained by equation (19).

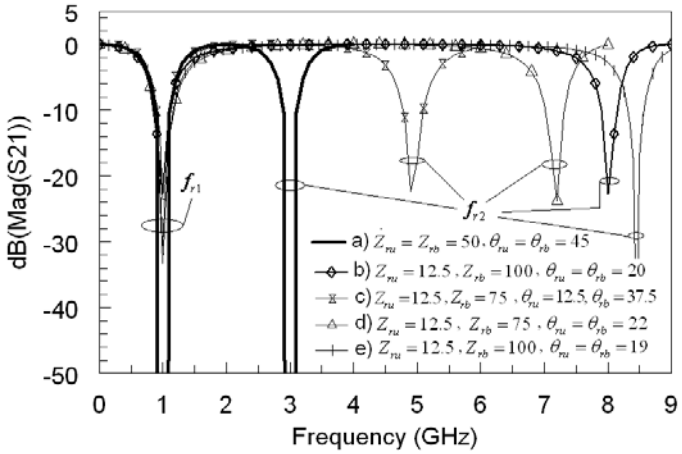



Fig. 4. Simulated results of the SIRS with different electric parameters

## 4. Design of filters for WiMAX equipment

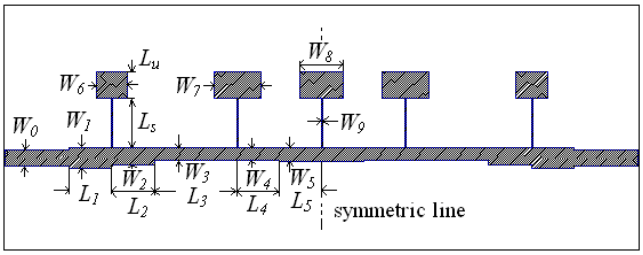



Fig. 5. Configuration of the proposed Microstip bandstop filter

By understanding the characteristic of the step impedance resonator section, the multi-stage cascaded step impedance resonator sections can be used to build the bandstop filters shown in Figure 5 for WiMAX application introduced in Section II. According to the application requirements of the WiMAX equipment in Figure 1 and Figure 2, two bandstop filters are built for 2.4 GHz and 3.5 GHz Wimax bands respectively. The fifth-order bandstop filter is formed by cascading five step impedance resonator sections using the cascade procedure in [15]. The electric parameters according to different step impedance resonator sections are synthesized and presented in [16] with consideration of the required total performance. The Rogers substrate RO3003 with dielectric constant of 3.0 has a substrate thickness of 20mil. The planar microstrip line is chosen for the design. The circuit simulator and Moment EM simulator in ADS 2005A are used to optimize the physical parameters for the 2.4 GHz and 3.5 GHz WiMAX filters. The dimensions of 2.4 GHz filter are: W0=1.26 mm, W1=1.49 mm, L1=4.77 mm, W2=1.23 mm, L2=5.24 mm, W3 = 0.75 mm, L3=10.9 mm, W4=0.88 mm, L4=4.64 mm, W5=0.91mm, L5=6.1mm, W6=2.72, Lu=3.59 mm, Ls=6.3 mm, W7=3.91 mm, W8=2.75 mm and W9=0.1mm. The dimensions of 3.5 GHz filter are: W0=1.26 mm, W1=1.6 mm, L1=3.31 mm, W2=1.4, L2=3.31 mm, W3=1.0 mm, L3=6.7 mm, W4=1.0 mm, L4=3.83 mm, W5=1.04 mm, L5=3.54 mm, W6=2.51 mm, Lu=2.32 mm, Ls=4.07 mm, W7=3.61 mm, W8=3.74 mm and W9=0.1mm.

In Figure 6 and Figure 7, the simulated and the measured results of the 2.4 GHz and 3.5 GHz filters without packaging housing are compared respectively. Both of the results agreed well in the passband as well as in the stopband. Both the 2.4 GHz and 3.5 GHz bandstop filters have the 3dB rejection band of 1.8~3.3 GHz and 3~5 GHz respectively, pass band ranges of DC~13.5 GHz and DC~21 GHz (except the stopband range) respectively. In the stopbands of 2.3-2.85 GHz and 3.4~4.1 GHz, the rejections are better than 60dB for both filters. The center frequency of the second stopband appears at as high as 6.2 times (15 GHz) and 6.3 times (22 GHz) of that of the fundamental stopband respectively. The board sizes of the 2.4 GHz and 3.5 GHz filters are only 74 mm by 35 mm (0.6λ0 × 0.28λ0, λ0 is the free space wavelength at center operating frequency) and 52 mm by 30 mm (0.6λ0 × 0.34λ0) respectively.

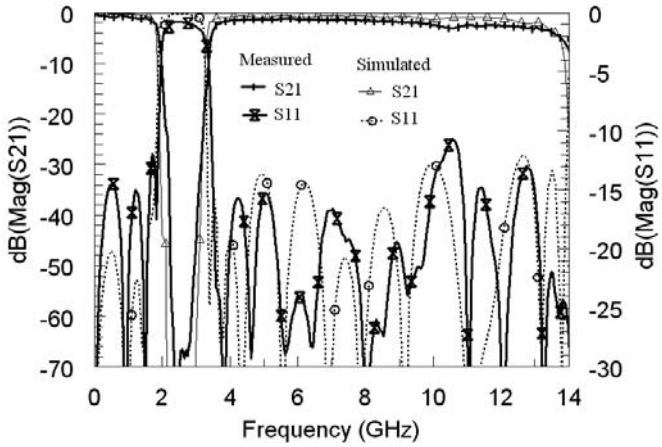



Fig. 6. Results of 2.4 GHz filter without metal housing

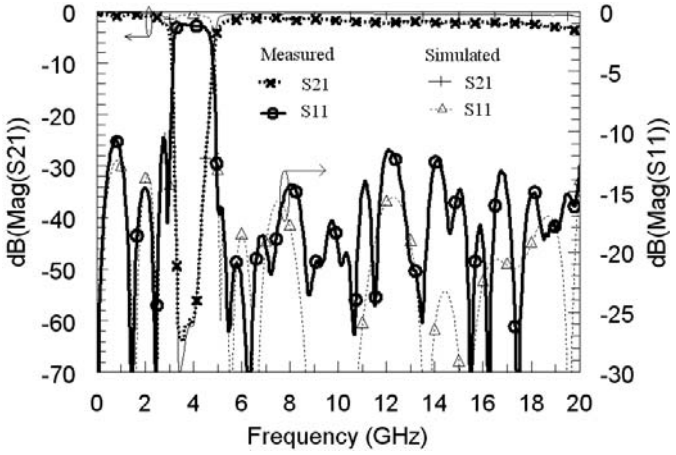



Fig. 7. Results of 3.5 GHz filter without metal housing

## 5. Packaging Related Issues

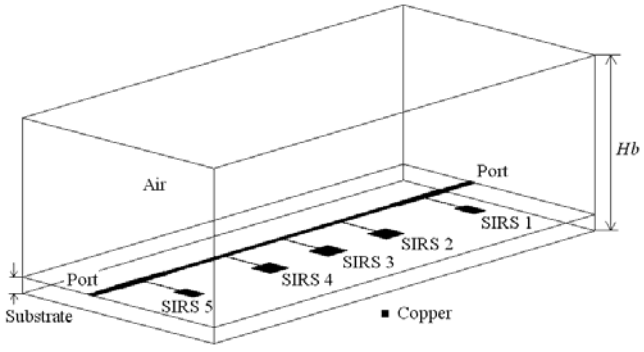



Fig. 8. Perspective view of the bandstop filter in metal housing SIRS (step impedance resonator section)

Electromagnetic compatibility/Electromagnetic Interference (EMC/EMI) control in the test equipment is important. Therefore, the modules in the equipment such as mixer, filter and amplifier, are required to be well shielded to avoid unwanted radiation and interference. The WiMAX filters introduced in Section IV has to be packaged for the equipment use. The perspective view of a 2.4 GHz WiMAX filter in the 3D metal housing of using the Sonnet EM simulator environment is shown in Figure 8. The metal house has a length of 74 mm, a width of 35 mm and a height of Hb. The 2.4 GHz WiMAX filter with five step impedance resonator section is denoted in the dark area. The leakage modes exist in the passband when the metal shielding house is introduced. The dip points (DPs) due to leakages in the metal housing appear in the pass band as the three cases of Hb=5 mm, 12 mm, and 15 mm shown in Figure 9. The dip points in the loss factor curves ($1-|S_{11}|^2-|S_{21}|^2$) in Figure 8 are formed by leakages due to the metal housing. When the multiple step impedance resonator sections from 1 to 5 are cascaded, the multiple conductors in parallel may generate unwanted

leakage wave as in [17-21]. The leakage wave excites the waveguide cavity modes [17] of the metal housing and causes the dip points in the passband of the filter. Compared to the results of the same filter without metal house (which is free of leakage in the passband) in Figure 6 and Figure 7, the packaging of the filter through the metal housing deteriorates the performance of the pass band and causes the filter unsuitable for wide passband operation to be used in WiMAX test equipment. On the other hand, if the unpackaged filters are used, EMI/EMC requirements for the WiMAX test equipment may not be met.

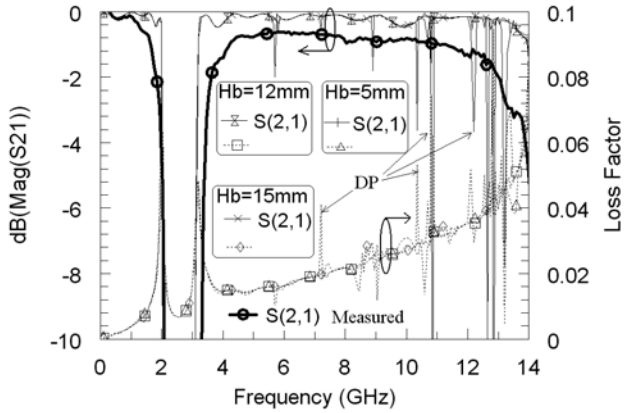



Fig. 9. Results of 2.4 GHz filter with/without metal housing

Through the experimental investigation, it is found that the issue can be solved by using the absorber material SS-6M from Emerson & Cuming. The absorber material attached inside metal cover changes the boundary condition of the metal housing. It is interesting that after packaging using the metal housing with absorber, the dip points in the passband no longer exist as measured result shown in Figure 9 Meanwhile, insertion loss in filter passband is also improved dramatically. The measured results of the 2.4 GHz and 3.5 GHz bandstop filters with and without housing are compared in Figure 10 and Figure 11 respectively. For the 2.4 GHz filter, the insertion loss including the SMA connector loss is less than 2.0 dB (most of the frequency range is less than 1 dB). Compared to the results of the 2.4 GHz filter without packaging, the proper packaging can reduce the insertion loss up to 0.5dB at higher passband. For the 3.5 GHz filter, the improvement of the insertion loss with proper packaging is up to 1dB at certain frequency range. In general, the proper packaging not only overcomes the leakage issue but also improve the passband performance. The packaging doesn't degrade the performance of the stopband as the measured group delays (GD) of the fifth-order filters are shown in Figure 12. The group delays in the passband are almost constant. The resonances occur in the stopband and the passband is a traveling-wave. The group delay variations of ± 0.5 ns are good for communication systems with high data rate or high linearity requirement. The photographs of the 2.4 GHz and 3.5 GHz WiMAX filter samples are shown in Figure 19. These two types of filter are now applied to the Agilent Technologies WiMAX test equipment.

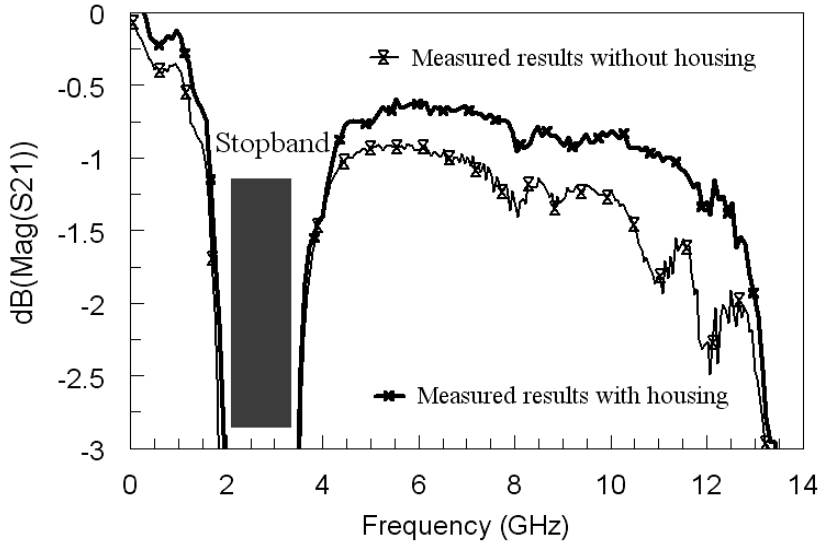



Fig. 10. Comparison of 2.4 GHz filter with/without metal housing

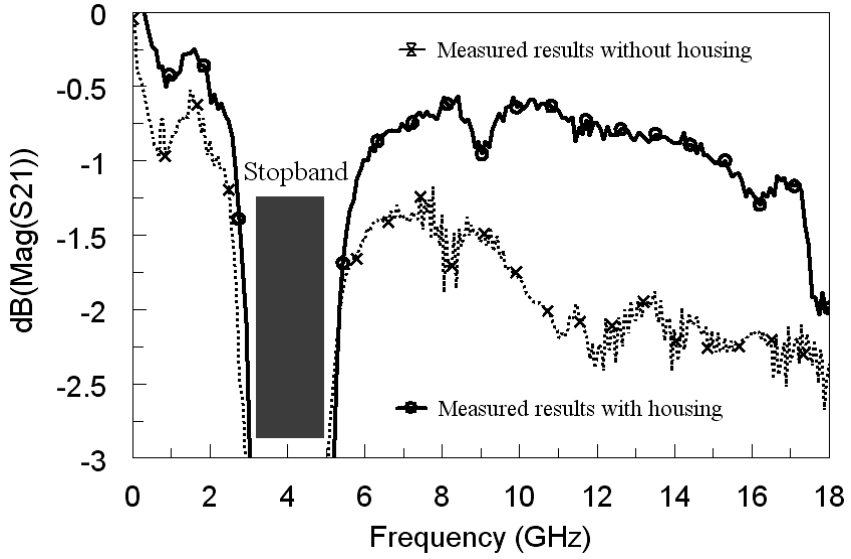



Fig. 11. Comparison of 3.5 GHz filter with/without metal housing

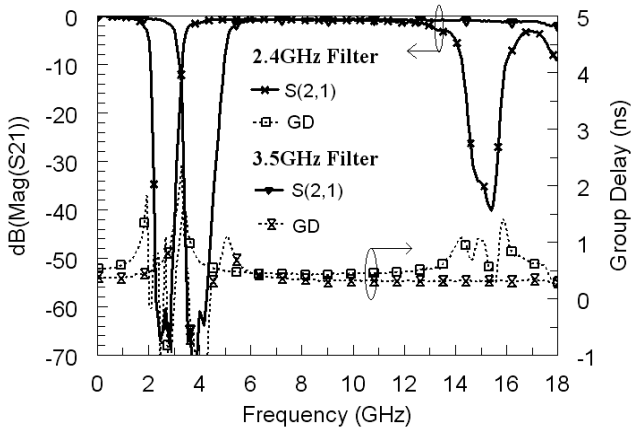



Fig. 12. Measured results of 2.4 GHz and 3.5 GHz filters with proper packaged metal housing

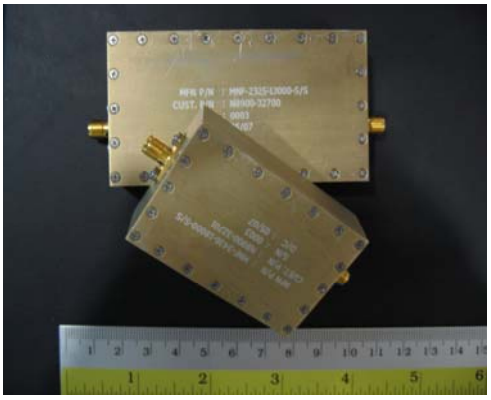
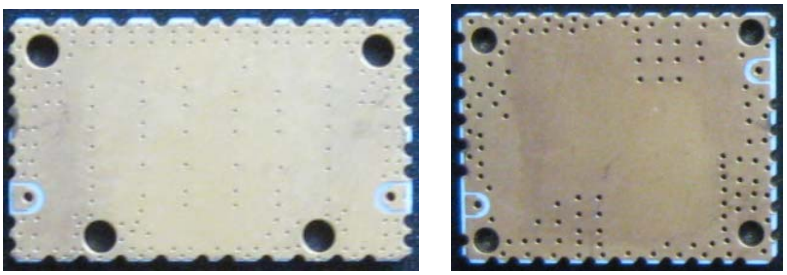


Fig. 13. Photographs of 2.4 GHz and 3.5 GHz WiMAX filter samples

## 6. Miniaturized Surface Mounted Bandpass Filter

In the application of WiMAX frontend, a bandpass filter is required to define the communication channel inside of the frontend and avoid the interferences from and to the other wireless communication applications. The portability and high linearity of the WiMAX terminal requires bandpass filter has a compact size, low cost and high performance. It is known that the miniaturization is the trend and designing compact passive devices is the challenging. Obviously, it would be impossible to face the challenging without novel design methodologies. Mainly, the electromagnetic (EM) characteristics determine the performance and size of the distributed transmission line filters. By investigating the EM characteristics of the circuits, different approaches, which form the basis of bandpass filters, need to be used for this target. Below shows parts of bandpass filters with high power handling, miniaturized size and high performance for WiMAX application multilayer through innovate use of multilayer technology. The focus is still on 2.4GHz and 3.5GHz WiMAX band.

a) 2.4GHz WiMAX bandpass filters



a) High rejection band design b) Low insertion loss design

Fig. 14. Photographs of 2.4 GHz WiMAX filter samples

The band pass filter in Figure 14 a) has unique performances with pass band from 2.2GHz~2.6GHz with minimum insertion loss less than 2dB and ultra wide stop band from 2.9GHz to 10GHz as the measured results shown in Figure 15. The stop band rejection is better than 50dB from 3.7GHz to 7.7GHz. The band pass filter has much compact size of 29mm X 19mm X 1.5 $mm^3$. The other strength is the packaging format of SMT which is convenient for integration.

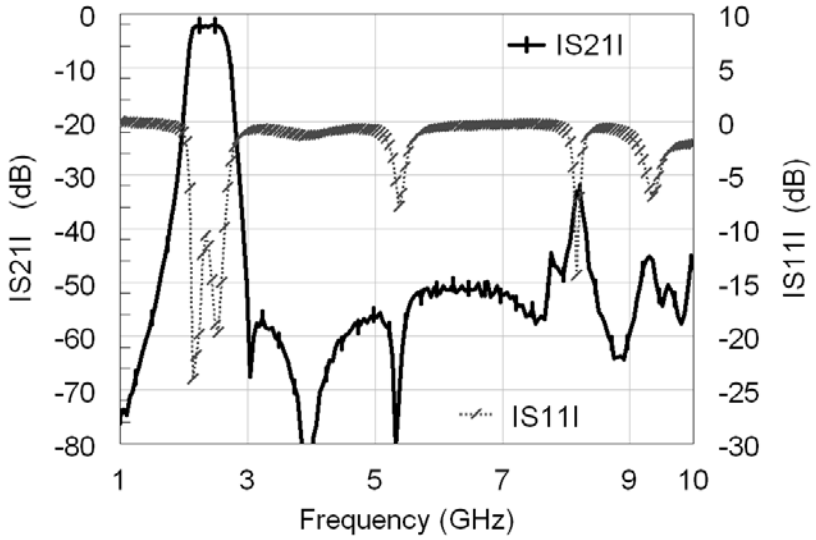



Fig. 15. Measured results of 2.2~2.6GHz high rejection WiMAX bandpass filter

The band pass filter in Figure 14 b) has unique performances with pass band from 2.2GHz~2.7GHz with minimum insertion loss less than 1dB and ultra wide stop band from 3GHz to 7.7GHz as the measured results shown in Figure 16. The stop band rejection is better than 50dB from 3.7GHz to 7.7GHz. The band pass filter has much compact size of 24.5mm X 19.5mm X 1.5 $mm^3$. The other strength is the packaging format of surface mounting which is convenient for integration and free of RF power leakage.

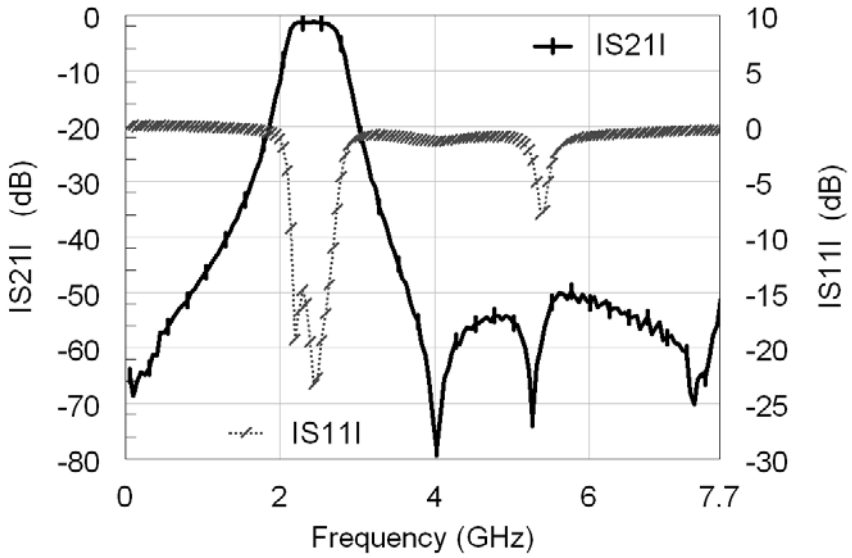
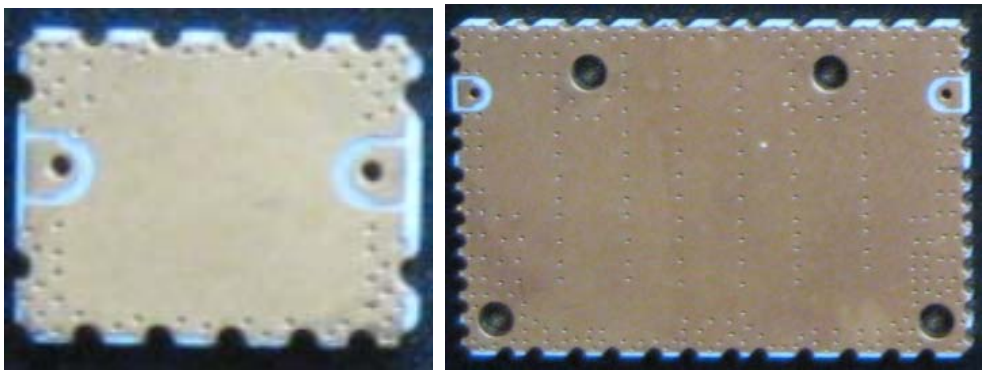



Fig. 16. Measured results of 2.2~2.7GHz low loss WiMAX bandpass filter

b) 3.5GHz WiMAX bandpass filters



a) Mini design b) high performance design

Fig. 17. Photographs of 3.5 GHz WiMAX filter samples

The band pass filter in Figure 17 a) has performances with pass band from 3.3GHz~3.6GHz with minimum insertion loss less than 2dB stop band from 3.7GHz to 7GHz as the measured results shown in Figure 18. The stop band rejection is better than 25dB from 3.7GHz to 7GHz. The band pass filter has much compact size of 11mm X 8.5mm X 1.5 $mm^3$. The other strength for this type filter is surface mounting and self-packaging with low RF radiation.

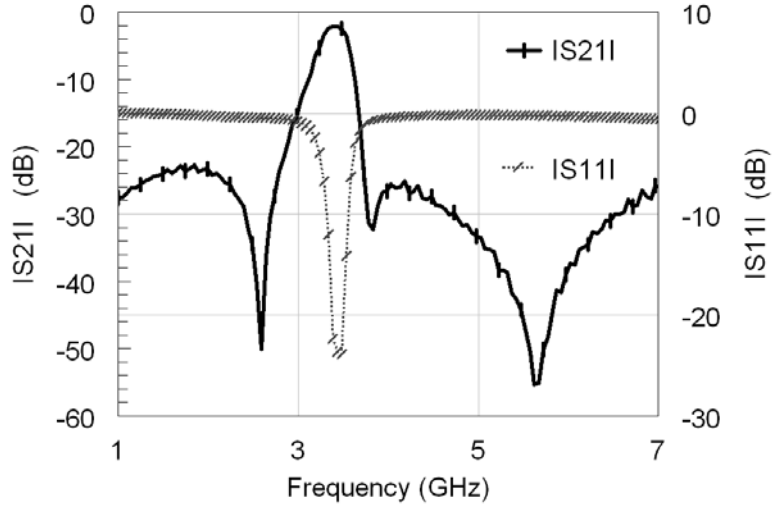



Fig. 18. Measured results of 3.3~3.6GHz mini WiMAX bandpass filter

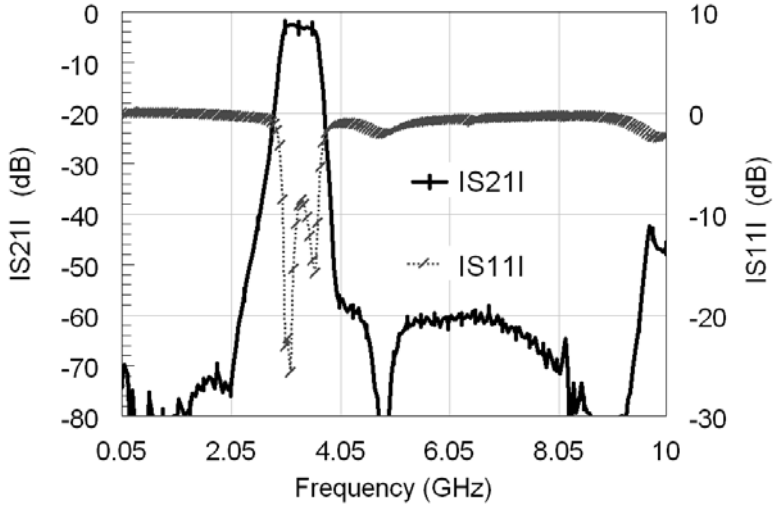



Fig. 19. Measured results of 3.0~3.6GHz high performance WiMax bandpass filter

The band pass filter in Figure 17 b) has performances with pass band from 3.0GHz~3.6GHz with minimum insertion loss of less than 2.5dB and ultra wide stop band from 3.8GHz to 10GHz as shown in Figure 19. The stop band rejection is better than 55dB from 4GHz to 9.5 GHz. The band pass filter has compact size of 25mm X 15mm X 1.5 mm. The other strength is the self-packaging format of surface mounting which is convenient for integration and free of RF leakage.

## 7. Conclusion

In this Chapter, bandstop filter is investigated based on proposed equivalent circuit model. The 2.4 GHz and 3.5 GHz bandstop filters for WiMAX application are designed and implemented theoretically and experimentally. The important issue of packaging is investigated and an approach to overcome the leakages to improve the performance is adopted. The implemented step impedance resonator section filters with compact size, high stopband rejection, flat group delay and ultra-wide passband bandwidth are used in the new application scenario of WiMAX equipment for nonlinear measurement. It should be addressed that although the filters are designed for WiMAX test equipment, it can also be adopted for other wireless communications with the requirement of rejection in certain range of the frequency 2.3GHz~2.85GHz such as WCS, ISM equipment and MMDS. The recent work on the series of WiMAX bandpass filter developments in 2.4 and 3.5GHz WiMAX bands is also demonstrated for WiMAX applications.

## 8. Acknowledgment

The authors would like to thank Dr. Dylan Williams for his kind review and constructive advices which make the manuscript more readable. The authors also thank Mr. Yap Kai Meng Leong from ST electronic (Satcom & Sensor) for his help in grammar checking.

## 9. References

[1] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems", *IEEE Std. 802.16-2004*.

[2] WiMAX Forum. IEEE 802.16a standard and WiMAX igniting broadband wireless access. Whitepaper, May 2004. http://www.wimaxforum.org.

[3] Matthaei G. L., Young L. and Jones E. M. T.. *Microwave Filters, Impedance-matching Networks, and Coupling Structures*. Artech House books · Dedham, MA 1964.

[4] Hong J.-S. and Lancaster M. J., *Microstrip Filters for RF/Microwave Applications*, John Wiley & Sons, Inc. 2001.

[5] Young L., Matthaei G. L., and Jones E. M. T., "Microwave band-stop filters with narrow stop bands," *IEEE Trans. Microwave Theory Tech.* vol. 12, pp. 416–427, Nov. 1962.

[6] Schiffman B. M. and Matthaei G. L., "Exact design of band-stop microwave filters," *IEEE Trans. Microwave Theory Tech.* vol. 10, pp. 6–15, Jan. 1964.

[7] Bell H. C., "L-resonator bandstop filters," *IEEE Trans. Microwave Theory Tech.* vol. 10, pp. 2669–2672, Dec. 1996.

[8] Qian J.-R. and Zhuang W.-C., "New narrow-band dual-mode bandstop waveguide filters," *IEEE Trans. Microwave Theory Tech.*, vol.31 pp.1045-1050, Dec. 1983.

[9] Saenz E. M., Subramanyam G., Keuls F. W.V., Chen C., and Miranda F. A., "Fixed-frequency and frequency-agile (Au, HTS) microstrip bandstop filters for L-band applications," *IEEE Trans. Appl. Supercond.*, vol.11 pp.395-398, Mar. 2001.

[10] Wallage S., Tauritz J. L., Hadley P., Lander L., and Mooij J. E., "High Tc superconducting CPW bandstop filters," *IEEE Microwave and Guided Wave Lett.*, vol. 6 pp.292-294, Aug. 1996.

[11] Gőrűr A. and Karpuz C., "Uniplanar compact wideband bandstop filter," *IEEE Microwave and Wireless Components lett.*, vol. 13, pp. 114-116, Mar. 2003.

[12] Yang R.-Y., Weng M.-H., Hung C.-Y., Chen H.-J., and Houng M.-P.,"Novel compact microstrip interdigital bandstop filters," *IEEE Trans. On Ultronsonic., Ferrelectrics, and Frequency Control.*, vol. 51 pp.1022-1025, Aug. 2004.

[13] Zhang X. J., Liu A.Q., Karim M. F., Yu A. B. and Shen Z. X., "MEMS-based photonic bandgap (PBG) bandstop filter," *IEEE MTT-S Int. Symp. Dig.*, pp. 1463–1466. 2004.

[14] Levy R., Snyder R. V. and Shin S.,"Bandstop Filters With extended Upper Passbands," *IEEE Trans. Microwave Theory Tech*, vol.54, pp.2503-2515, June, 2006,.

[15] Ma K., Xiao S.-Q., Chan K.-T., Yeo K. S., Ma J.-G., and Do M. A., "Characterizing and modeling conductor-backed CPW Periodic band stop filter with miniaturized size" *IEEE MTT-S Int. Microwave Symp. Dig.*, pp.983~986, 2007.

[16] Ma K., Jayasuriya R. M. and Chan K. T., "High Performance Bandstop Filter Design and Investigation Using Physical Model for WiMAX Measurement Equipment," *IEEE MTT-S Int. Microwave Symp. Dig.*, 2008.

[17] Kim S.-J., Yoon H.-S. and Lee H.-Y., "Suppression of leakage resonance in coplanar MMIC packages using a Si sub-mount layer," *IEEE Trans. Microwave Theory Tech.*, vol. 48, pp. 2664-2669 Dec. 2000.

[18] Lo Y. T., Tzuang C. C., Peng S. T., Tien C. C., Chang C. C., and Huang J. W., "Resonance phenomena in conductor-backed coplanar waveguides (CBCPW's)," *IEEE Trans. Microwave Theory Tech.*, vol. 41, pp. 2099-2108, DEC. 1993.

[19] Raskin J.- P., Gauthier G., Katehi, L. P. and Rebeiz G. M., "Mode conversion at GCPW-to-microstrip-line transitions," *IEEE Trans. Microwave Theory Tech.*, vol. 48, pp. 158-161 Jan. 2000.

[20] Ma K., Ma J.-G., Do M. A., and Yeo K. S., "Experimentally Investigating Slow-Wave Transmission Lines and Filters Based on Conductor-Backed CPW Periodic Cells," *IEEE MTT-S Int. Microwave Symp. Dig*, pp.1653-1657, 2005.

[21] Ma K., Yeo K. S., Ma J.-G., and Do M. A., "An Ultra-Compact Planar Band Pass Filter with Open-ground Spirals for Wireless Application" *IEEE Transaction on advanced packaging*, Vol. 31, NO. 2, pp.285-291, May 2008

9

# Adaptive Parameters Based Transmission Control and Optimization in Mobile WiMAX at Physical Layer

U. D. Dalal[a] and Y. P. Kosta[b]
*a – Sr. Lecturer, Electronics Department, SVNIT, Surat*
*b - Professor, Electronics Department, CIT, Changa*

## 1. Introduction

WiMAX is the system for wireless broadband access. It is based on IEEE 802.16 standards which are mainly based on Orthogonal Frequency Division Multiplexing (OFDM) technology. OFDM is a wideband modulation scheme using multicarrier digital communication [Bahai & Saltzberg, 1999][Molisch, 2002]. During the communication there is an uplink and a downlink between base station transmitter and mobile receiver. While communicating, the wireless channel effects on the received OFDM signal differ due to selection of different parameter values as well as existing Signal to Noise Ratio (SNR) condition at that time. SNR variations may arise due to distance variations between transmitter and receiver and fast or slow mobility, exhibiting time varying channel [Lowray, 2001]. Different allowable bandwidths with same number of subcarriers, different number of subcarriers with same bandwidth, different modulation mapping schemes [Hole & Qien, 2001], variation in pilot power [Alsusa; Baidas & Lee, 2005] and pilot positions, pilot sequences, variations in cyclic prefix interval etc., are considered as important OFDM parameters requiring critical selection, whose effect is directly reflected in the performance. Up to certain extent adaptive nature is adopted in mobile WiMAX standard 802.16e in terms of scalable OFDMA [IEEE 802.16e, 2005] and linkage of modulation mapping scheme with channel coding [Hole and Qien, 2001] which can be extended further in terms of few more parameters. Such parameters are identified and reinvestigated. Few more possibilities are described here in a comprehensive manner. These parameters can be made adaptive with SNR conditions of the channel in both uplink as well as downlink. This chapter investigates those parameters which can be varied and on the basis of that adaptive OFDM transmission link control concept is developed, which can be applied in practice, maybe in vehicular mobility up to certain extent. Indirectly, effective and efficient Quality of Service (QoS) control can be achieved. Such control can be adopted in any frame based systems in general with at least half duplexity between transmitter and receiver. This adaptive nature may reduce the wastage of unnecessary energy utilized for the users who are very near to the transmitter, bringing the optimum solutions. Of course, the discussion reflects multiuser scenario.

After identifying such parameters we have simulated the mobile WiMAX for OFDM-256 case using different channel types (as this scheme is common to both fixed as well as mobile

WiMAX) and BER-SNR plots for different such parameters, with given conditions, are represented in the chapter at the end to approximate the optimum SNR conditions for various such parameters to maintain the targeted BER.
Adaptive algorithms are also presented along with adaptive user allocation. These techniques utilize knowledge obtained by dynamically tracking the radio channel response, to optimize the user bandwidth and subcarrier modulation. Adaptive modulation independently optimizes the modulation scheme applied to each subcarrier so that the spectral efficiency is maximized, while maintaining a target BER [Hole and Quien, 2001]. The performance of this technique is dependent on the correlation of the frequency selective fading and how fast the fading changes with position of the transceiver.

## 2. WiMAX Scenario

The demand for high-speed mobile wireless communications and use of the radio spectrum is rapidly growing with terrestrial mobile communication systems being just one of many applications vying for suitable bandwidth. These applications require the system to operate reliably in non-line-of-sight environments with a propagation distance of 0.5 - 30 km, and at velocities up to 100 km/hr or higher. This operating environment limits the maximum RF frequency to 5 GHz, as operating above this frequency results in excessive channel path loss, and excessive Doppler spread at high velocity. This limits the spectrum available for mobile applications, making the value of the radio spectrum extremely high [Wu and Lin, 2006]. The Mobile WiMAX standards IEEE 802.16e onwards are developed by keeping in mind the above scenario. Obviously, mobile WiMAX defit in terms of bit rate compared to fixed WiMAX. To meet the demand of speed and spectrum the physical layer becomes very important. Indirectly, by physical layer optimization using parameter control, one can get the required Quality of Service (QoS). Physical layer is based on scalable OFDM in case of mobile WiMAX [IEEE 802.16e, 2005], where, OFDM technology promises to be a key technique for achieving the high data capacity and spectral efficiency requirements for wireless communication systems of even 4G. For the visualization of physical layer design and transmission control, one must know the architecture. WiMAX architecture is shown in Figure 1, which includes line of sight (LOS) and non line of sight (NLOS) communication links. Figure 2 shows the development of WiMAX system over cellular infrastructure, and adaptive modulation requirement in a cell.

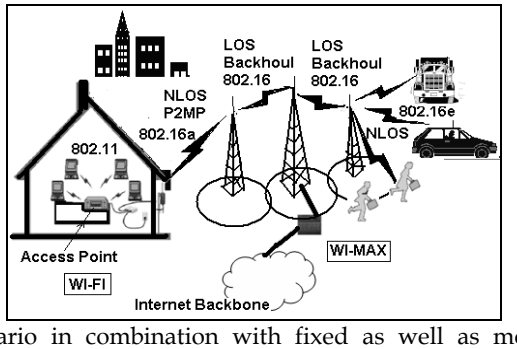



Fig. 1. WiMAX scenario in combination with fixed as well as mobile wireless access interworking with WiFi

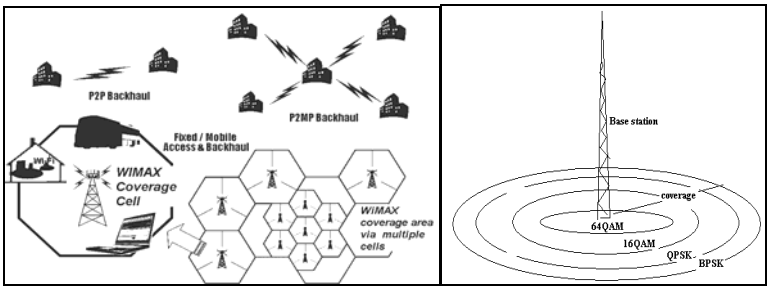



Fig. 2. WiMAX topology examples with cellular structure and adaptive modulation concept

Instead of fixed allocations of physical parameters over such architecture, adaptive nature in multiuser diversity has become a topic of recent interest and is mainly required in multiuser scenario. Adaptive user allocation exploits the difference in frequency selective fading between users, to optimize user subcarrier allocation. In a multipath environment the fading experienced on each subcarrier varies from user to user, thus by utilizing user/subcarrier combinations that suffer the least fading, the overall performance is maximized. Of course there must be the considerations for the minimum calculation complexity to meet the real time requirements.

## 3. OFDM vs Scalable OFDMA in WiMAX

OFDM and scalable OFDMA (SC-OFDMA) work slightly differently but both have the feasibility of adaptive nature. SC-OFDMA is akin to mobile version of WiMAX only, while former is in both the fixed as well as mobile WiMAX cases. In OFDM when we increase the bandwidth, we increase the channel bandwidth of each tone as the number of tones remain constant. But in SC-OFDMA we instead increase the number of Fast Fourier transform (FFT) points increasing the channel bandwidth, bandwidth of the tones is kept constant in the mobile environment. The FFT size and the number of carriers are equal in both fixed and mobile WiMAX, based on OFDM256, but they are different in SC-OFDMA. The fixed bandwidth is a compromised solution in mobile environment [though with such simpler case of 256 FFT point based OFDM simulations are done for time varying channels and results are given later on].
In SC-OFDMA, Scaling of FFT to the channel bandwidth is in order to keep the carrier spacing constant across different channel bandwidths (typically 1.25 MHz, 5 MHz, 10 MHz or 20 MHz). Constant carrier spacing results in higher spectrum efficiency in wide channels, and a cost reduction in narrow channels. Other bands not multiples of 1.25 MHz are defined in the standard, but because the allowed FFT subcarrier numbers are only 128, 512, 1024 and 2048, other frequency bands will not have exactly the same carrier spacing, which might not be optimal for implementations [IEEE 802.16e, 2005].
SC-OFDMA (used in 802.16e-2005) and OFDM256 (802.16d) are not compatible thus most equipment will have to be replaced if an operator wants or needs to move to the later standard. However, some manufacturers are planning to provide a migration path for older equipment to SC-OFDMA compatibility which would ease the transition for those networks which have already made the OFDM256 investment. Intel provides a dual-mode 802.16-2004 802.16-2005 chipset for subscriber units. This affects a relatively small number users and

operators.With the advent of mobile WiMAX, there is an increasing focus on portable subscriber units. 802.16e-2005 has been accepted as IP-OFDMA for inclusion as sixth wireless link system under IMT-2000. So it is very important in the present scenario.

## 4. Adaptive Parameters Identification

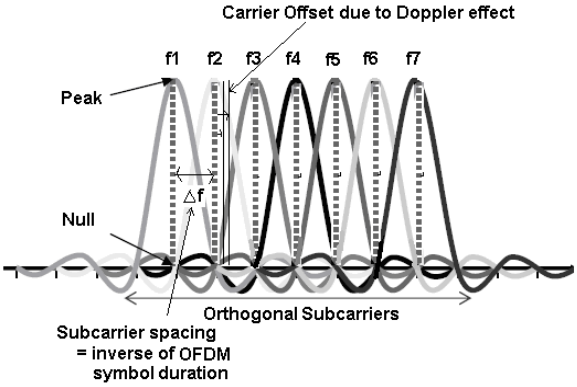



Fig. 3. OFDM spectral setting

Figure 3 shows ideal OFDM spectrum setting along with few subcarriers adjusted in a particular transmission bandwidth maintaining spacing $\Delta f$ and orthogonality among the subcarriers. Frequency offset due to Doppler effect destroys the orthogonality [Bahai & Saltzberg, 1999]. Thus, OFDM based system has to satisfy four requirements in general while designing:

*1) Available bandwidth*: The bandwidth limit will play a significant role in the selection of number of subcarriers along with spacing. Large bandwidth will allow obtaining a large number of subcarriers with reasonable Cyclic Prefix (CP) length avoiding multipath.

*2) Required bit rate*: The size of the frame must be decided on the basis of symbol mapping scheme and number of subcarriers (hence spacing $\Delta f$) to be assigned to it. This will decide bit rate indirectly.

*3) Tolerable delay spread based on terrain and distance:* A user environment specific maximum tolerable delay spread should be known beforehand in determining the CP length.

*4) Doppler spread based on velocity support:* The effect of Doppler shift due to user movement should be taken into account for allowable subcarrier spacing.

The design parameters which can be applied adaptation are derived according to the system requirements. The transmission parameters are adjusted to provide an acceptable level of performance to the most impaired link. This approach then limits the performance that might be offered to subscribers with less impaired channels. Clearly this method results in sub-optimal utilization of the total channel capacity. Hence, the identified design parameters for an adaptive OFDM system are as follows with the opportunities for optimizing the overall system performance (For which the simulation is done):

*1) Number of subcarriers and subcarrier spacing:* It is stated earlier that the selection of large number of subcarriers (Of course within specified bandwidth) will help to combat multipath effects. But, at the same time, this will increase the synchronization complexity at the

receiver side as well as increase problems due to Doppler spread. So, allocate user subcarriers so as to minimize Signal to Interference Ratio (SIR) in cellular systems and to minimize the effects of frequency selective fading.

*2) Symbol duration and CP length:* For specified delay spread perfect ratio between the CP length and symbol duration should be selected, so that multipath effects are combated and significant amount bandwidth is not lost due to CP.

*3) Modulation type per subcarrier:* The performance requirement will decide the selection of modulation scheme. Adaptive modulation can be used to support the performance requirements in changing environment. So, dynamically allocate the modulation scheme on an individual subcarrier basis to match the current channel conditions. In our simulations, by varying the SNR conditions the BER is tested for various modulation schemes to find out the limiting conditions. Of course, it is necessary to adopt appropriate channel coding in accordance with as given in Table 2 and 3.

*4) Bandwidth:* Dynamically change the bandwidth of each user based on the link quality. This allows the bandwidth of weak users to be reduced so that their energy spectral density remains sufficiently high to maintain communications. The concept of SNR-Bandwidth trade-off can be exploited.

There are many papers in which the adaptive allocations of pilot is proposed. Pilot management is necessary because it is an extra investment of power and that should be optimized. Here are few possibilities summarized.

- Superimposed pilots over data symbols
- Variation in Pilot positions (randomly hopping or sequentially varied), same way at the receiver.
- Variation in the length of training sequence
- Variation in the training sequence (pattern) itself. There is a separate research area in which sequences and their properties are analyzed and genetic algorithms are there to develop sequences with desirable properties.
- Variation in the pilot power—maintaining the Peak to Average Power Ratio (PAPR)
- Variation in the number of pilots- Again PAPR comes in picture

Apart from these, Forward Error Correction, adaptive power control and adaptive user allocation are important issues. Table 1 contains a synthetic view of some adaptive techniques used nowadays in broadband multicarrier wireless systems, including WiMAX, together with the benefits they bring.

| Adaptive Techniques | Improvements |
|---|---|
| Adaptive modulation | 12-16dB SNR increase |
| Adaptive power control | Efficient use of radio resources |
| Adaptive subcarrier allocation | Link availability and robustness through allocation of a lower modulation scheme for the low SNR subcarriers |
| Adaptive bandwidth | QoS increase, BER decrease<br>A 10 times smaller bandwidth is needed, and this results in a received SNR increase of 10 dB |
| Adaptive user allocation | Average signal power increase of 3-5dB<br>Bite rate and SNR increase, BER decrease<br>Higher capacity achieved by exploitation of frequency dispersive fading, by allocating the peaks in frequency response |

Table 1. Benefits of adaptive radio techniques

## 5. Radio Resource Management by Adaptive Features

Advanced radio resource algorithms in broadband wireless systems enable service providers to maximize subscriber throughput and overall coverage while maintaining QoS. Techniques to optimize the use of available radio resources include power control, rate adaptation, automatic repeat requests, channel quality indication, scheduling, and admission control. WiMAX with its OFDMA-based structure provides a means to balance the effects of these techniques to provide an optimal tradeoff between throughput and link quality.

*Power Control*

Adaptive power control is an important function for ensuring link quality. In the upstream direction, adaptive transmit power control is used to maximize the usable modulation level, which achieves the highest throughput, while at the same time controlling interference to adjacent cells. In the downstream direction, different power allocations for specific subchannels can be used to provide better service to subscribers at the edge of the cell while providing sufficient signal levels to subscribers in closer proximity to the base station.

*Improved Power Consumption*

The mobile WiMAX standard incorporates mechanisms that enable subscriber terminals to be active only at certain times as negotiated with the base station. When no data is to be transmitted or received, the subscriber terminal can move to 'sleep' or 'idle' modes to minimize power consumption. The base station scheduler is kept aware of every sleep or idle subscriber terminal and has the ability to switch the terminal to transmit or receive mode whenever required. In the subscriber terminal transmit mode the use of subchannels ensures that the transmit power is no greater than what is necessary to maintain sufficient link quality consistent with the traffic being transmitted, thus further reducing power consumption in the subscriber terminal.

*Rate Adaptation*

In any terrestrial multi-cellular network, mobile subscribers will experience transmission path conditions that vary with relative location and time. With OFDMA the specific modulation and coding scheme can be adapted on a per subscriber basis dependent on path conditions to maximize channel throughput while maintaining link quality to each subscriber. With OFDMA systems, the subcarriers are modulated with either the more robust QPSK or the higher order, more efficient QAM modulations – with the more sophisticated modulation schemes having higher throughput but being much more susceptible to interference and noise. This rate adaptation, through adaptive modulation and error coding schemes ensures that the number of bits conveyed by each subcarrier is optimized relative to the CINR required to ensure a reliable air link connection. OFDMA systems can also increase throughput to individual subscribers by increasing the number of allocated subchannels at any given time. Both of these concepts are included in the mobile WiMAX specification.

*Hybrid Automatic Repeat Request*

Automatic repeat request (ARQ) algorithms are well known in wireless, and wireline, networks for retransmitting failed transmissions. The effective use of ARQ however, requires precise selection of both transmit power and data rate for the retransmissions, otherwise the link becomes underutilized or experiences excessive packet errors. Since it is challenging to maintain these optimal settings in the time varying environment of mobile broadband services, a significantly more robust mechanism called Hybrid- ARQ (H-ARQ)

was developed. With H-ARQ, which is part of the mobile WiMAX specification, the receiver combines the information from a faulty packet with the re-transmissions of the same packet until enough information is gathered to retrieve the packet in its entirety.

*Channel Quality Indication*

Timely channel quality indication (CQI) messages at the receiver are essential for adaptive power and rate control and H-ARQ to be effective. The support of high mobility services requires that fast corrective actions be taken at the transmitter to ensure the link is operating optimally at all times. Mobile WiMAX specifies a compact size (4-6 bits each) CQI messages, resulting in lower delay and greater reliability than regular control messages. This ensures that the CQI messages provide fast and reliable feedback of path conditions to the base station while maintaining low overhead.

*Scheduling Control*

Scheduling control is a mechanism, located in the base station, for managing upstream and downstream packet allocations based on traffic requirements and channel conditions at any given moment. The scheduler allocates radio resources in frequency and time, based on considerations such as; QoS parameters for the specific traffic-type, individual subscriber service level agreements (SLA), and connection-by-connection path conditions. Since data-oriented traffic can vary considerably between uplink and downlink, asymmetric capacity allocation is also supported in time division duplex (TDD) implementations with appropriate radio resources and packet assignments done on a per-sector basis for a variable duration based on actual demand. These basic scheduling control mechanisms are part of the mobile WiMAX standard.

*Admission Control*

Admission control is the process of determining whether or not to allow a new connection to be established based on: current traffic conditions, available resources, and cumulative QoS requirements. Excessive traffic in a cell increases the amount of interference to adjacent cells thus reducing cell coverage. Admission control is used to accept or reject the connection requests so as to maintain the cell load within acceptable limits. The admission control function is located in either the WiMAX base station or the access service network (ASN) gateway where the load information for several base stations can be monitored.

## 6. Trade off Between Allocated Bandwidth and Adaptive Bandwidth

In most cases user is allocated a fixed amount of bandwidth, regardless of the received signal power. But in mobile WiMAX like environment or in multiuser scenario, this may lead to problems for users that have low received signal strength [Lowray, 2001]. The SNR of these users may be insufficient to support communications even using BPSK. The SNR seen at the receiver is dependent on the signal bandwidth, and so reducing the bandwidth while using the same transmitter power increases the SNR of the signal. For example, reducing the signal bandwidth by 5 times, allows the full transmitter power to be concentrated into one fifth the bandwidth, increasing the transmitted power spectral density by 5 fold, resulting in an improved received SNR of 5 dB.

The main aim of adaptive bandwidth allocation is to maintain communications with users that have low received signal strength due to far distance with respect to the base station transmitter. This is achieved by reducing their bandwidth to the point where the transmitted power spectral density is high enough to support communications at a low data rate. This

can be used as a method for improving the quality of service (i.e. decreasing the outage probability).
Adaptive bandwidth by itself will not be suitable for all applications, especially those that required a fixed data rate such as streaming video and audio. In these applications a joint optimization of bandwidth and modulation scheme could be performed to maintain a fixed data rate, while minimizing the amount of bandwidth used at any one time. This could be achieved by allocating both the user bandwidth and modulation scheme so that the spectral efficiency multiplied by the user bandwidth results in the required data rate. This way, as the signal strength becomes weaker, the amount of bandwidth allocated to that user increases to compensate. This fixed data rate optimization is not included in the simulations and could be researched.

## 7. Trade off between Fixed Modulation and Adaptive Modulation

Adaptive modulation has not been used extensively in wireless applications due to the difficulty in tracking the radio channel effectively. Work has been done studying the use of adaptive modulation in single carrier systems, however not many works have been published on use of adaptive modulation in OFDM systems.
Most OFDM systems use a fixed modulation scheme over all subcarriers for simplicity. However each subcarrier in a multiuser OFDM system can potentially have a different modulation scheme depending on the channel conditions. Any coherent or differential, phase or amplitude modulation scheme can be used including BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM, etc, each providing a trade off between spectral efficiency and the bit error rate. The spectral efficiency can be maximized by choosing the highest modulation scheme that will give an acceptable BER.

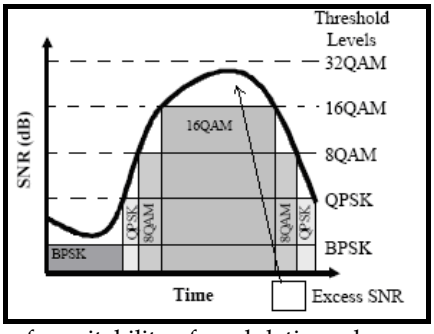



Fig. 4. General comparisons for suitability of modulation scheme with SNR

In a multipath radio channel, frequency selective fading can result in large variations in the received power of each subcarrier. For a channel with no direct signal path this variation can be as much as 30 dB in the received power resulting in a similar variation in the SNR. In addition to this, interference from neighboring cells can cause the SNR to vary significantly over the system bandwidth.
To cope with this large variation in SNR over the system subcarriers, it is possible to adaptively allocate the subcarrier modulation scheme, so that the spectral efficiency is

maximized while maintaining an acceptable BER. Figure 4 shows a reference plot for applying adaptive modulation to an individual subcarrier as the channel SNR varies with time.

Using adaptive modulation has a number of key advantages over using static modulation. In systems that use a fixed modulation scheme the subcarrier modulation must be designed to provide an acceptable BER under the worst channel conditions. This results in most systems using BPSK or QPSK. However these modulation schemes give a poor spectral efficiency (1 - 2 b/s/Hz) and result in an excess link margin most of the time. Using adaptive modulation, the remote stations can use a much higher modulation scheme when the radio channel is good. Thus as a remote station approaches the base station the modulation can be increased from 1 b/s/Hz (BPSK) up to 4 - 8 b/s/Hz (16-QAM – 256-QAM), significantly increasing the spectral efficiency of the overall system. Using adaptive modulation can effectively control the BER of the transmission, as subcarriers that have a poor SNR can be allocated a low modulation scheme such as BPSK, or none at all, rather than causing large amounts of errors with a fixed modulation scheme. This significantly reduces the need for Forward Error Correction

There are several limitations with adaptive modulation. Overhead information needs to be transferred, as both the transmitter and receiver must know what modulation is currently being used. Also as the mobility of the remote station is increased, the adaptive modulation process requires regular updates, further increasing the overhead.

There is a trade off between power control and adaptive modulation. If a remote station has a good channel path the transmitted power can be maintained and a high modulation scheme used (i.e. 64-QAM), or the power can be reduced and the modulation scheme reduced accordingly (i.e. QPSK). Distortion, frequency error and the maximum allowable power variation between users limit the maximum modulation scheme that can be used. The received power for neighboring subcarriers must have no more than 20 - 30 dB variation at the base station, as large variations can result in strong signals swamping weaker subcarriers. Inter-modulation distortion results from any non-linear components in the transmission, and causes a higher noise floor in the transmission band, limiting the maximum SNR to typically 30 - 40 dB. In our simulations the SNR is limited to 25 dB. Frequency errors in the transmission due to synchronization errors and Doppler shift result in a loss of orthogonality between the subcarriers. A frequency offset of only 1 - 2 % of the subcarrier spacing results in the effective SNR being limited to 20 dB.

### 7.1 Adaptive Modulation and Coding Support in Mobile WiMAX

Schemes offering varied modulation and coding for different classes of channel propagation conditions by assigning user classes to different physical channels are a step in the direction of improved channel utilization, but do not completely address the dynamic nature of network traffic and its impact on required link gains. The PHY layer described as per adaptive modulation and coding optimizes channel utilization by permitting dynamic adaptation of both modulation format and Forward Error Correction (FEC) rate on a subscriber-by-subscriber basis. Thus, if the channel characteristics vary over time, say seasonal variations in path loss caused by the presence or absence of foliage, that subscriber's modulation and coding parameters can be adjusted to compensate for these changes.

| | Downlink | Uplink |
|---|---|---|
| **Modulation** | BPSK, QPSK, 16 QAM, 64 QAM; BPSK optional for OFDMA-PHY | BPSK, QPSK, 16 QAM; 64 QAM optional |
| **Coding** | **Mandatory:** convolutional codes at rate 1/2, 2/3, 3/4, 5/6<br><br>**Optional:** convolutional turbo codes at rate 1/2, 2/3, 3/4, 5/6; repetition codes at rate 1/2, 1/3, 1/6, LDPC, RS-Codes for OFDM-PHY | **Mandatory:** convolutional codes at rate 1/2, 2/3, 3/4, 5/6<br><br>**Optional:** convolutional turbo codes at rate 1/2, 2/3, 3/4, 5/6; repetition codes at rate 1/2, 1/3, 1/6, LDPC |

Table 2. Modulation and coding schemes for WiMAX

WiMAX supports a variety of modulation and coding schemes and allows for the scheme to change on a burst-by-burst basis per link, depending on channel conditions. Using the channel quality feedback indicator, the mobile can provide the base station with feedback on the downlink channel quality. For the uplink, the base station can estimate the channel quality, based on the received signal quality. In conditions of good signal, a highly efficient 64QAM coding scheme is used, whereas when signal is poorer, a more robust BPSK scheme is used. QPSK and 16QAM are for intermediate conditions. Following is a list of the various modulation and coding schemes supported by WiMAX-802.16e as well.

| Modulation | Uncoded Block Size (bytes) | Coded Block Size (bytes) | Overall coding rate | RS code | CC Code rate | Puncture pattern |
|---|---|---|---|---|---|---|
| BPSK | 12 | 24 | 1/2 | (12,12,0) | 1/2 | 1 |
| QPSK | 24 | 48 | 1/2 | (32,24,4) | 2/3 | [1011] |
| QPSK | 36 | 48 | 3/4 | (40,36,2) | 5/6 | [10101] |
| 16QAM | 48 | 96 | 1/2 | (64,48,8) | 2/3 | [1011] |
| 16QAM | 72 | 96 | 3/4 | (80,72,4) | 5/6 | [10101] |
| 64QAM | 96 | 144 | 2/3 | (108,9,6) | 3/4 | [110] |
| 64QAM | 108 | 144 | 3/4 | (120,108,6) | 5/6 | [10101] |

Table 3. Channel coding per modulation

The key parameter for traffic modeling (maybe using adaptive antenna) from a planning tool point of view is the bit rate that can be offered for a given received signal level. This characteristic itself depends on the modulation used, hardware algorithms, transmitting and receiving antenna specifications.

| Channel Bandwidth | 3.5MHz | | 1.25MHz | | 5MHz | | 10MHz | |
|---|---|---|---|---|---|---|---|---|
| PHY mode | 256 OFDM | | 128 OFDMA | | 512 OFDMA | | 1,024 OFDMA | |
| Oversampling | 8/7 | | 28/25 | | 28/25 | | 28/25 | |
| **Modulation & Code Rate** | **PHY-Layer Data Rate (kbps)** | | | | | | | |
| | DL | UL | DL | UL | DL | UL | DL | UL |
| BPSK, 1/2 | 946 | 326 | **Not applicable** | | | | | |
| QPSK, 1/2 | 1,882 | 653 | 504 | 154 | 2,520 | 653 | 5,040 | 1,344 |
| QPSK, 3/4 | 2,822 | 979 | 756 | 230 | 3,780 | 979 | 7,560 | 2,016 |
| 16 QAM, 1/2 | 3,763 | 1,306 | 1,008 | 307 | 5,040 | 1,306 | 10,080 | 2,688 |
| 16 QAM, 3/4 | 5,645 | 1,958 | 1,512 | 461 | 7,560 | 1,958 | 15,120 | 4,032 |
| 64 QAM, 1/2 | 5,645 | 1,958 | 1,512 | 461 | 7,560 | 1,958 | 15,120 | 4,032 |
| 64 QAM, 2/3 | 7,526 | 2,611 | 2,016 | 614 | 10,080 | 2,611 | 20,160 | 5,376 |
| 64 QAM, 3/4 | 8,467 | 2,938 | 2,268 | 691 | 11,340 | 2,938 | 22,680 | 6,048 |
| 64 QAM, 5/6 | 9,408 | 3,264 | 2,520 | 768 | 12,600 | 3,264 | 25,200 | 6,720 |

Table 4. Achievable data rates for various cases in IEEE 802.16e

**PHY-Layer Data Rates:**

Because the physical layer of WiMAX is quite flexible, data rate performance varies based on the operating parameters. Parameters that have a significant impact on the physical-layer data rate are channel bandwidth and the modulation and coding scheme used. Other parameters, such as number of subchannels, OFDM guard time, and oversampling rate also have an impact. Operating over maximum range (50 km) increase bit error rate and thus must use a lower bit rate. Lowering the range allows a device to operate at higher bit rates. Following is the PHY-layer data rate at various channel bandwidths, as well as modulation and coding schemes with scalable OFDM cases.

Further, the OFDM symbol is one block of frame expanded to frame duration during serial to parallel conversion and assigned to a subcarrier (usually a block is referred to as an "OFDM symbol") A 'block adaptive' PHY is proposed for 802.16.3 in which modulation format and coding rate can be adjusted on a block by block basis to provide optimum channel utilization for the widest range of channel conditions. This proposal is offered as a PHY layer for the 802.16.3 Task Group [IEEE 802.16.3c-00/39, Oct 2000] The 'Block-Adaptive' PHY is based upon a Frequency Division Duplex (FDD) approach employing Downstream-Upstream access methods of DOCSIS like TDM data distribution in the downstream path and Time Division Multiple Access (TDMA) in the upstream direction. Although it is envisioned that many traffic models will require asymmetric data rates in the downstream and upstream directions, the PHY layer proposed here allows for symmetric data rates if so desired. Further, the proposed PHY can support a wide range of channel bandwidths and assignment plans. In the "Block-Adaptive" structure the data is partitioned into Blocks delimited by a Block Identifier Word. Each Block represents a unit of data to be sent to an individual subscriber.

### 7.2 Adaptive Modulation Based TCP Aware Uplink Scheduling in IEEE 802.16 Netwrks

There are many schemes at higher layers utilizing the adaptive allocations of PHY. One of the scheme is highlighted here as a case study. The primary contribution of paper by Hemant Kumar Rath , Abhay Kharandikar and Vishal Sharma is to propose a fair adaptive modulation-based uplink scheduling scheme for applications based on TCP in IEEE 802.16. Since, the TCP congestion window size (*cwnd*) changes only after one RTT, *cwnd* is an indication of the number of time slots required per Round Trip Time (RTT). Hence, instead of assigning equal number of slots to all users, we argue that the BS should assign slots in proportion to their *cwnd*, i.e., as per the flow's requirement. Assigning time slots based only on *cwnd* will result in unfairness among the TCP flows, since flows with smaller RTT s will have larger window size as compared to the flows with larger RTT. To avoid this unfairness, they introduced a credit-based approach that ensures fairness among the flows. More slots are assigned to the flows which are closer to their TCP timeout, thereby preventing their congestion window from dropping to one due to timeout. By introducing adaptive modulation, fairness measure that only considers slots assigned becomes irrelevant, rather, fairness in terms of amount of data transmitted in a frame should be considered. Hence, they measured fairness on the amount of data transmitted by SSs.

## 8. Tracking Schemes for Experimentation [Lowrey, 2001]

*a. Adaptive Subcarrier Allocation Algorithms:*
There are several methods for allocating subcarriers to users in a multiuser OFDM system on adaptive basis. The main five schemes are to use a group of subcarriers with

1) *A fixed frequency grouped subcarriers*
2) *Randomly hopped subcarriers*
3) *TDMA allocation*
4) *Spread out subcarriers in a comb pattern*
5) *Adaptive user allocation*

These methods are described here to have experimentation. For performing such experiment, set up with multiple OFDM transceivers is necessary for measurements of path loss as a function of frequency and distance. This data can be used for the simulations. Perfect channel estimation is necessary to assume in this simulation. All five simulations can use adaptive modulation in order to determine the modulation for each subcarrier in each time slot, with the only difference being the subcarrier frequencies allocated to each user. The adaptive modulation and the subcarrier frequency allocation can be updated at a rate of 250 Hz, which corresponds to a distance traveled between updates of 4% of wavelength, for the carrier frequency of 1000 MHz used. The frequency allocations can be subdivided into 256 or less subcarriers over the system bandwidth of 1.25-3.5 MHz.
In practice the number of subcarriers would be more than this, however the number should be kept reasonably low so that the frequency allocations could be seen in the simulation plots. A BER threshold of $1x10^{-5}$ can be used to decide on the SNR thresholds for the adaptive modulation allocation. The choice of this BER threshold is arbitrary and dependent on requirements of the final application of the system. Changing this threshold would result in a slight change in data throughput,
In each of these simulations multiple users are transmitting at the same time in the same frequency band. These multiple transmissions form a single OFDM signal at the base

station's receiver. In order for these signals to remain orthogonal to each other, they must all be frequency and time synchronized with each other. The time synchronization must be accurate to within the effective guard period length (shorter then the actual guard period length due to channel delay spread), while frequency synchronization must be sufficiently accurate as to maintain a sufficiently high effective SNR as to use the modulation schemes used in the simulation.

*b. Adaptive Bandwidth Algorithm:*

*1)* Initially allocate all users an equal number of subcarriers, using the *adaptive user allocation* algorithm (described next)

*2)* Calculate the SNR of all the subcarriers for a given user, based on the bandwidth allocated and the particular subcarriers that were allocated in step 1.

*3)* Find the minimum SNR of the subcarriers allocated to each user.

*4)* For each user check that every subcarrier allocated to that user, has a SNR greater than a given threshold. This is done to ensure that each subcarrier has a sufficiently high SNR to support at least BPSK modulation. A threshold of 5-6 dB can be used in the simulations from the observations. If a user has any subcarriers below this threshold then reduce the BW (number of subcarriers) allocated to that user. This frees up BW, making it available to other users. Make sure that each user is allocated at least one subcarrier. This prevents the degenerate case where a user is allocated zero subcarriers, which in turn makes the estimated SNR infinite (The SNR is inversely proportional to the signal bandwidth).
If all subcarriers meet the minimum SNR threshold, then the user allocation is complete, so apply adaptive modulation to each subcarrier and exit. If the SNR of one or more subcarriers is below the threshold then continue onto step 5.

*5)* Redistribute the free BW by allocating it to users that have a SNR significantly greater than the minimum SNR threshold. Transmitting over a wider BW results in a lower transmitted spectral density and received SNR. Thus there is no use reallocating the extra BW to a user that is just above the SNR threshold, as this would result in their SNR dropping below the threshold, defeating the purpose of the reallocation. To help with data load leveling, the extra BW should be allocated to users that have a low SNR first, as users with a high SNR will already have a comparatively high data rate and thus shouldn't need more BW.

*6)* Based on the new user BW allocations, reallocate the subcarriers to users by repeating the *adaptive user allocation* algorithm

*7)* Repeat from *(2)* until all allocated subcarriers have a SNR above the required threshold.

*c. Adaptive User Allocation Algorithm:*

*1)* Find the mean SNR over the entire system BW for each user.

*2)* Perform allocation of subcarriers to users in-order, from, lowest mean SNR, to highest mean SNR. This helps to ensure that weak users get access to the best subcarriers.

*3)* Sort the SNR response for the user being allocated, removing any subcarriers that have already been allocated to other users. The SNR response is the SNR of each subcarrier as seen by that user. This will be different for each user, due to propagation variations.

*4)* Allocate from the sorted SNR response subcarriers in descending order from best SNR to worst SNR subcarriers to meet the BW required for the user. Repeat from step 2 until all users have been allocated.

## 9. Simulation and Results

| | |
|---|---|
| Modulation/mapping | BPSK, QPSK, 16QAM, 64QAM |
| IFFT | 256 points |
| Max. speed of mobile | 120 km/h |
| Max. distance | 30 km |
| Scalable channel bandwidth | 1.25 to 3.5 MHz |
| Imposed comb Pilots | 8 carrying training sequence partly along with data |
| Pilot Positions in FFT bin | 41, 66, 91, 116, 142, 167, 192, 217 |
| Data carriers | 192 + 8 |
| Total carriers, $N_c$ | 192 + 8 + guard(28+28) = 256 |
| Channel models | SUI 1-6, ITU and MIMO |

Table 5. Simulation parameters

The OFDM modulation scheme based WiMAX system simulation is done using MATLAB and the following results are obtained for finding the limiting conditions to eliminate outages. From the plots the concept of feasibility of adaptive allocations is clear.

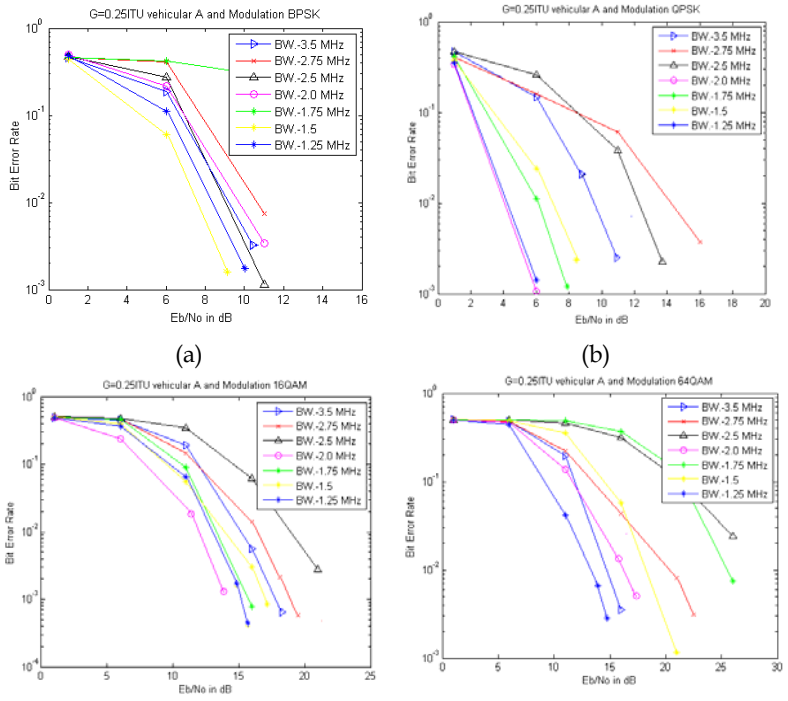
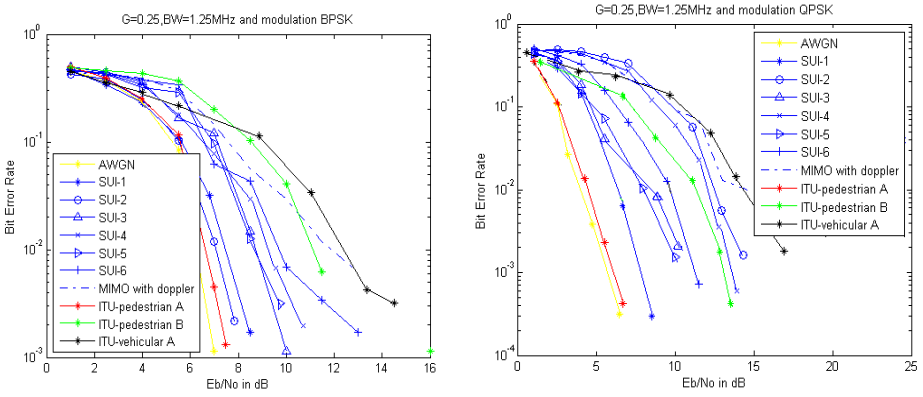
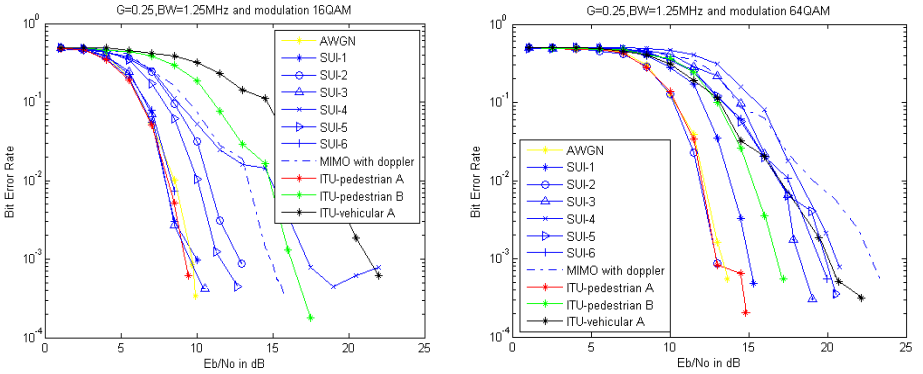



Fig. 5. SNR response in case of various bandwidths for vehicular A channel with 120 km/hr speed for different modulation schemes

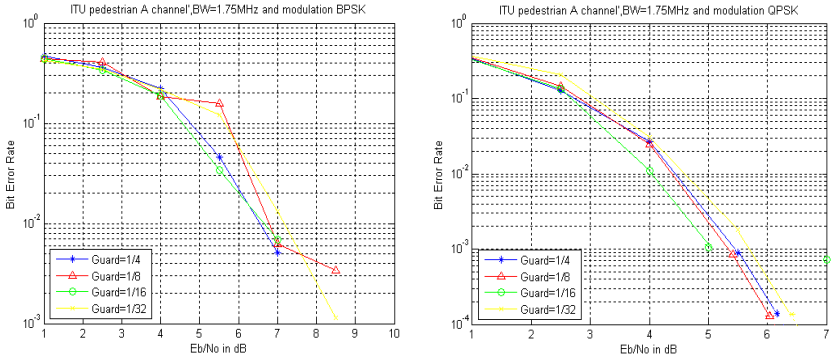



(a) (b)

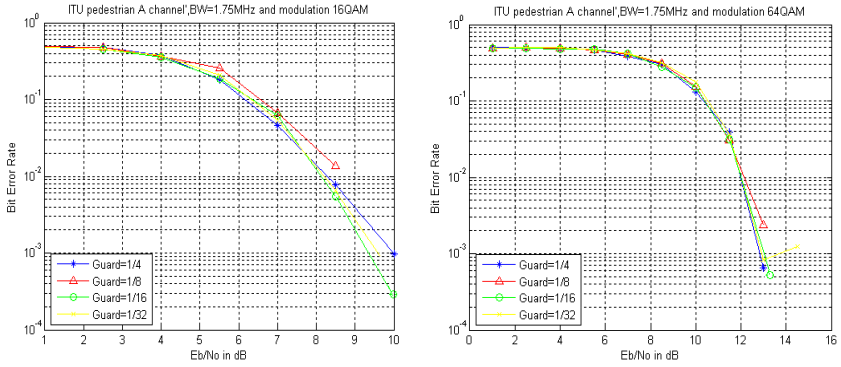



(a) (b)

Fig. 6. Performance of modulation schemes in different channel environment




(a) (b)




(c) (d)

Fig. 7. Performance due to various guard intervals for different modulation schemes and ITU pedestrian A channel

| Doppler | Low Delay spread | Moderate Delay Spread | High Delay Spread |
|---|---|---|---|
| Low | SUI-1,2(High *K*-factor) SUI-3, ITU pedestrian A | | SUI-5 |
| High | | SUI-4 | SUI-6 |
| AWGN –no delay spread and no Doppler spread<br>ITU pedestrian B-30 km/hr channel with 10 Hz Doppler spread<br>ITU Vehicular A -120 km/hr channel with 400 Hz Doppler spread<br>MIMO is Concatenated 2 x 2 channel<br>Vehicular Probability Urban: 30 km/hr 60% of the time, 80 km/hr 30% and 120 km/hr 10%<br>Vehicular Probability Suburban: 30 km/hr 20%, 80 km/hr 20%, 120 km/hr 60% | | | |

Table 6. Characteristic summary of channels used in simulations

Same guard ratio ¼ but different bandwidth (Figure 5) – here due to cyclic prefix addition the IFFT points increases and hence the spectrum widens slightly. More the bandwidth more effect of noise, so the probability of errors increases and hence with wider bandwidth, the performance degrades. At the same time Doppler margin will also come in picture. When the bandwidth is increased with the same number of carriers and FFT points, effectively the Doppler margin between carriers increases. This is rather advantageous in heavy Doppler condition. After addition of CP the original spectral setting does not get affected. With the above discussion it can be summarized that the bandwidth related performance is a tradeoff. It can be identified from the graphs that bandwidth 1.25 give the best performance in all mapping schemes. That is why the majority of the previous plots are taken with 1.25 MHz bandwidth with and ¼ guard ratio. Figure 6. gives the different SNR requirements for various types of channels. SNR requirement increase in this order: AWGN, SUI 1, ITU pedestrian A, SUI2-3, SUI 4, MIMO with Doppler, ITU pedestrian B, SUI 5-6 and ITU vehicular A. From the results of Figure 7, it can be commented that reducing the guard interval spectrum width reduces so effect of noise can reduce however, if the CP duration is less than CIR than complete mulitpath effect elimination is not possible. So depending upon

the situation the performance improves or degrades. Consistency may not be there due to time varying nature of the channel, estimation and interpolation errors, random nature of the data etc. However, it is sure that if channel conditions are not known, the performance with guard ratio ¼ is optimum in most of the cases.

## 10. References

Bahai A.R.S.; Saltzberg B. R. (1999) Multi-carrier digital communications: Theory and applications of OFDM, Kluwer Academic/Plenum.

Molisch A. (2002), Wideband Wireless Digital Communication, Pearson Education, LPE.

E. Alsusa; M. Baidas & Y. Lee (2005), "On the impact of efficient power allocation in pilot based channel estimation techniques for multicarrier systems", IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, pp 706-710.

Hole K. J. & Øien G. E., "Spectral Efficiency of Adaptive Coded Modulation in Urban Microcellular Networks", IEEE Transactions on Vehicular Technology, Vol. 50, No.1, pp. 205 – 222.

Aguayo-Torres M. C.; Entrambasaguas, J. & Ruiz T. F. (1998), "Variable rate multitone system for multipath fading channels", Electronic Letters, , Vol. 34, No. 12, pp. 1203 – 1204.

Mohamed-Slim Alouini (1998) "Adaptive and Diversity Techniques for Wireless Digital Communications over Fading Channels", Ph.D Thesis, Department of Electrical Engineering, California Institute of Technology, Pasadena, California,.

Cheong Yui Wong, Roger S. Cheng, Khaled Ben Letaief, Ross D. Murch (1999), "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation", IEEE Journal on Selected Areas in Communications, Vol. 17, No. 10, pp. 1747-1757.

Carl Magnus Frodigh; Perols Leif Mikael Gudmundson (1998), "Adaptive channel allocation in a frequency division multiplexed system", US. Patent 5,726,978, Filed: June 22,1995.

Jen-Ming Wu, Wen-Bin Lin (2006), "Channel estimation for Non line-of –sight WiMax communication system", Vehicular technology conference, VTC 2006, spring, IEEE 63rd, Vol.6 , pages 2671-75.

Wong C. Y.; R. S. Cheng; Letaief K. B., & Murch R. D. (1999), "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," IEEE Journal on Selected Areas in Communications, vol. 17, no. 10, pp. 1747–1758.

B. Makarevitch (2007), "Adaptive resource allocation for WiMAX," in Proceedings of the 18th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '07), pp. 1–6, Athens, Greece.

L. B. Le & E. Hossain (2007), "Joint rate control and resource allocation in OFDMA wireless mesh networks," Proceedings of IEEE Wireless Communications and Networking Conference (WCNC '07), pp. 3041–3045, Hong Kong.

Lee K.-D. & Leung V. C. M. (2006), "Fair allocation of subcarrier and power in an OFDMA wireless mesh network," IEEE Journal on Selected Areas in Communications, vol. 24, no. 11, pp. 2051–2060.

Z. Shen J.; Andrews G., & Evans B. L. (2005), "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," IEEE Transactions on Wireless Communications, vol. 4, no. 6, pp. 2726–2737.

L. Reggiani, L. G. Giordano, and L. Dossi (2007), "Multi-user sub-channel, bit and power allocation in IEEE 802.16 systems," in Proceedings of the 65th IEEE Vehicular Technology Conference (VTC '07), pp. 3120–3124, Dublin, Ireland.

IEEE standard drafts and WiMAX forum literature.

10

# Smart Antenna Systems for WiMAX Radio Technology

Yikun Huang
*Montana State University*
*U.S.A.*

## 1. Introduction

In this chapter we present our design and implementation of smart antenna systems for WiMAX radio technology in Montana State University. WiMAX offers the potential of long range (up to 50 miles) and high bandwidth (up to 50 Mb/s) radio links[Eklund 2002, Nuaymi 2007]. This new radio technology vastly exceeds the capabilities of widely used wireless such as Wi-Fi, offering features not supported by current wireless systems. The new WiMAX standard offers the potential for using adaptive antenna system (AAS), but this functionality has not been widely implemented. One notibale example is Smart WiMAX[Hedayat et al. (2007)], however no technical details have been reported. This chapter services as an introduction to the reader with a perspective on historical and recent progress in this rapidly involving field of communications research.

Smart antennas for fixed and mobile wireless communications have received enormous interest worldwide in recent decades, and a wide variety of approaches for beamforming antenna design and application have been described over the last five decades [Liberti and Rappaport 1999, Van Veen, and Buckle 1988, Widrow and Stearns 1996, Krim and Viberg 2003, Loadman et al. 2003, Godara 1997]. Unlike a conventional omni-directional antenna that wastes most of its energy in directions devoid of users, a smart antenna can form one or more beams in the direction of fixed or mobile users and create nulls toward interferences, thus greatly improving system performance. In addition, since the radiation power of a beamforming antenna is focused toward the user, a beamforming antenna equipped radio system will have a much larger range than that of a conventional omni-directional antenna equipped system. Also, this efficiency increase, in most cases, will save energy. While adaptive smart antennas have been used in the past for other applications, the technology has seen limited use due to high costs and poor integration with other radio system elements.

There are two types of smart antennas: switched beam and adaptive array. For a switched beam system, a set of specific, sometimes predetermined, beam patterns are formed with the main lobe toward the mobile node. The antenna system monitors the signal strength and switches among the lobes periodically to update beam selection. This antenna design improves performance by increasing signal strength and suppressing interferences that are not in the same direction as the signal. However, if interference is within the same lobe as

the signal, the interference will not be suppressed. This is a major disadvantage of the switched beam approach.

A phased array (or steeable array) can be viewed as a special type of switched beam. With predefined weight (array element current), a fixed beam can be formed to any desired pointing angle. In reality, due to the limitation of hardware, only limited pointing angle can be achieved.

An adaptive array utilizes sophisticated signal processing algorithms to continuously distinguish among the desired signal and interferences and can form an unlimited number of beam patterns to optimally improve signal strength and suppress interferences. The adaptive array offers higher gain, because of the precise pointing ability, than the switched beam array and greater interference rejection. Adaptive arrays may require longer computational time to converge to optimal patterns, and thus may not be suitable for real time high data rate communications where there are a large number of highly mobile nodes and interferences. Also, they will consume a greater amount of power than a switched beam system. In a system where there is low interference, a switched beam array may be adequate because it is less costly and it can produce a signal gain comparable to an adaptive array. In a system where there is considerable interference, tracking the exact location of the nodes is an integral part of increasing system performance and therefore the adaptive array may be a better choice.

This chapter is organized as follows. Section 2 presents the switched beam system. It details the function and operation of the system. Section 3 describes a phased array system architecture. The procedures used to test the smart antenna performance with a WiMAX radio and the results generated by these tests are presented. Section 3 presents an adaptive array testbed design while detailing the function and operation of the system and its modules. The type of algorithms used in the system will be discussed along with the challenges involved in their implementation. The design of the PC interface which consists of the brains operating the system. The conclusions drawn from the results are presented in the section 4.

## 2. Switched Beam System

A switched beam antenna is the simplest form of AAS. It creates a group of overlapping beams that together result in orminidirectional coverage. In a switched beam system, predefined beam pattern is formed with the main beam towards the user. In general an $m$-element array may generate an arbitrary number of beam patterns. However it is much simpler to form $qm$ beam patterns, where $q$=1,2,…$Q$, with $360°/Qm \geq 1/10$ of the half-power beamwidth. The beam pattern is generated using specific weights, which translates into specific amplitude adjustments and phase shifts for each of the array elements.

The $m$-element circular array receives signals from several spatially separated users. The received signals usually contain both direct path and multipath signals, which are mostly likely from different directions of arrival (DOA). Assume that the array response vector (also called steering vector) to a transmitted signal $s_1(t)$ from a DOA $\varphi$ is $a(\varphi)=[1, a_1(\varphi), a_2(\varphi),\ldots, a_{m-1}(\varphi)]^T$, where $a_i(\varphi)$ a complex number denoting the amplitude gain and phase shift of the signal at the ($ith$+1) antenna relative to the first antenna and superscript $T$ is the transpose operator. For an $m$-element uniform circular array of radius $\rho$,

$$\boldsymbol{a}(\varphi)=\left[1,e^{j\beta\rho\cos\left(\varphi-\frac{2\pi}{m}\right)},\ldots,e^{j\beta\rho\cos\left(\varphi-\frac{(m-1)\times 2\pi}{m}\right)}\right]^T \tag{1}$$

where $\beta = 2\pi/\lambda$ is the wave number and the superscript T is the transpose operator. In a typical open space, we can ignore the multi-path signals. Thus the total signal vector received by the array can be written as:

$$\boldsymbol{x}(t)=\boldsymbol{a}(\varphi)s_1(t)+\boldsymbol{n}(t) \tag{2}$$

where $\boldsymbol{n}(t)$ is interference and noise. If there are $K$ sources that share the same frequency and time slot, then the signal received by the array is:

$$\boldsymbol{x}(t)=\sum_{k=1}^{K}\boldsymbol{a}_k(\varphi)s_k(t)+\boldsymbol{n}(t) \tag{3}$$

The practical switched beam operation process is based upon a simple 3-stage switched beam mechanism:

*A. Target searching*

After identification of the received signals as targets of interest, they are averaged over several sets of consecutive phase delays, and a beam that has the largest outcome is selected. Before the operation, we have a predetermined $N$ set of spatial signatures corresponding to the fixed beams saved in the system. Each beam $m$ has a specific spatial signature $\boldsymbol{a}_n(\varphi)$, $n$ = 1, 2, …, $N$. For an m-element circular array, $N$ may be chosen as $qm$, where $q$=1,2,…$Q$, with $360°/Qm \geq 1/10$ of the beamwidth. The switched beam array output vector

$$\boldsymbol{y}_n(t)=\left(a_1,a_2\ldots,a_N\right)^H x(t) \tag{4}$$

where the superscript H is the conjugate transpose operator. Assume there is only one target, when an $\boldsymbol{a}_n^*(\varphi)$ is equal or very close to the signal spatial signature $\boldsymbol{a}_k^*(\varphi)$, the $n^{\text{th}}$ element in output vector will be equal or very close to the signal strength received: $\boldsymbol{y}_n(t)\cong\left(0,0\ldots,\sqrt{P_n},\ldots 0\right)^T$. Thus the target is in the region of the $n^{\text{th}}$ beam. We will need to routinely update the DOA of the user.

Beam switching algorithms will determine when a particular beam should be selected or rejected to maintain the highest quality signal. The system continuously updates beam selection to ensure the quality of the communication. The antenna system switches through the outputs of each beam and selects the beam with maximal signal strength as well as suppressing interference arriving from the direction away from the active beam's center.

*B. Beamforming*

This process is based on the knowledge of the direction of the target. Once the desired direction is known, the smart antenna system (described here mainly as a receiving antenna system) will choose one sector in active mode and a properly selected phase delay is applied. Thus the signal from that specific direction will have the maximal gain. The

direction of the target is updated as required. In a TDD system where the uplink and downlink share the same carrier, we can design and keep a weight vector of the smart antenna system based on the spatial signature received at $i^{th}$ time slot such that $w_i = \boldsymbol{a}_i^*$ for the downlink. At the $j^{th}$ slot, the signal received by the mobile user will be $\left(\boldsymbol{a}_i^* \boldsymbol{a}_j\right) s(t)$, where $a_i$ and $\boldsymbol{a}_j$ are normalized vectors. If the update rate is fast enough so that the relative change ≈ 0, the mobile user will receive maximal signal power. However if the update rate is slow so that $\left|\boldsymbol{a}_i^* \boldsymbol{a}_j\right| \approx 0$ or the relative change ≈ 100%, the mobile user will not receive any signal power. In practice, we need to set a threshold to determine which beam should be active. For communicating with more than one user, and to save energy a beam will stay at a direction as long as possible.

To reduce the side lobes of a particular beam pattern, the channel signals are shaped by a window function, e.g. Chebyshev, Hamming, Hanning, Cosine, triangular, etc. This is the simplest way to beam form in order to maximize the signal to interference ratio of a switched beam array. By carefully controlling the side lobes in a non-adaptive windowed array, most interference can be reduced to an acceptable level by as little as 6 dB or as much as 12 dB.

*C. Update tracking*

When there is only one target, this task is very simple since the target will not change its location dramatically (i.e. the direction of arrival from the desired target will not change much at moderate distances during the communication). Assume that at the time, beam *m* is chosen. To update, we may compare the signal strength from beam *m* and its neighbor beams: beam *m*-1 and beam *m*+1, and choose the strongest one as the updated beam. The tracking cycle is properly chosen so that the target will not travel out of the small range(between beam *m*-1 to beam *m*+1) . When there are more mobile users, the communication system needs to have a table to record the location of each user. The table will update periodically.

## 3. Phased array system

Phased array system can be viewed as a special case of switched beam system. The algorithm is as follows. Assume that a fixed-frequency incident plane wave projects onto the array from a particular DOA. The wave crests of the signal will arrive at one end of the array first and will arrive at each successive array element at a progressively later times. The output of the array can be analyzed in such a manner that the signals arriving at the array from that specific DOA can be enhanced, and all those from other DOAs can be attenuated. This is achieved by adding carefully-chosen phase delays to the signals arriving at the analyzer from the different array elements: if the phase delays are chosen to compensate exactly for the time delays of the wave crests from one specific DOA, then summation of all outputs at the analyzer will maximize the component from that DOA. Waves with different DOAs will arrive at each array element with different set of time delays, and summation of the outputs with a set of delays inappropriate for that DOA will result in a much smaller summed signal. A single antenna array can be used to differentiate signals from a variety of DOAs simultaneously. This is achieved by configuring the analyzer to carry out simultaneous but independent sets of processing operations, each one corresponding to

"tuning" the analyzer to signals from a different DOA. Each of these independent processes is characterized by the addition of a different (and unique) set of phase delays to the signals arriving from the array elements, corresponding to the phase delays of signals arriving from a specific DOA.

We have recently designed and developed an 8-element uniform circular array (UCA) antenna prototype operating at 5.8GHz that can cover a 360° field of view and beam form toward multi-users as shown in Figure 1 [Huang 2008, Huang et al. 2009]. The core algorithms have been derived and preliminary simulations of the approach have been carried out. Simulation studies, as well as lab and initial field test results show that the smart antenna is capable of tracking mobile targets, communicating directionally with desired users, suppressing interference and jamming, and covering a long range with high throughput and reliable connection.

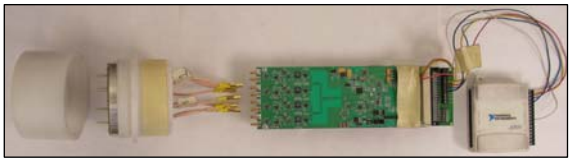


Fig. 1. A photo of the phased array antenna system

There are basically two types of approaches in phased array antenna systems. The analog architecture is usually based on RF phase shifters, while the digital systems commonly consist of a digital signal processor (DSP) that calculates and applies weight vectors to each sampled data at base band frequencies. The digital beamforming (DBF) architecture offers several fascinating functionalities, including simpler architecture, programmable control of antenna radiation pattern, direction-of-arrival (DOA) estimation, and adaptive steering of its beams and nulls to enhance the signal-to-interference-noise ratio (SINR). It is generally recognized that these advantages can only be carried out by digital technology. The analog approach, on the other hand, is re-emerging to create an alternative architecture of adaptive array antennas. Analog approaches are expected to offer dramatically low power consumption in smart antennas, especially for battery-operated wireless terminal devices. Analog approaches are capable of adapting to existing single channel radios. Figure 2 shows the diagram of the analog beamforming architecture we have used for our design. The concept of analog beamforming itself was proposed more than forty years ago, but it is considered practically impossible for analog systems to provide the smart functionalities that DBF does. If analog beamforming is available in the RF stage of adaptive antenna arrays, it should be able to provide drastic improvement in both DC power dissipation and fabrication costs since it could eliminate the need for frequency converters and analog to digital converters by the number of array branches. The initial RF prototype control uses one USB DAQ card to control a Xilinx Spartan FPGA which delivers 96 control lines to 16 phase and attenuation ICs on a RF board.

In the array illustrated in Figure 1, the operation frequency is at 5.8GHz. The array electric size $\beta r$ is 3.0, where $r$ is the array radius. The smart antenna unit is composed of three functional blocks: the monopole array, the ground plane with ground skirt, and the beamformer. The former two blocks provide an essential mechanical function as well as a

passive electrical function to properly transmit and receive the signal into/from the air with minimal attenuation and distortion, particularly at the proper frequency. The ground skirt provides a virtual infinite ground plane for the array as well as a mechanical rigidity. The infinite virtual ground plane is crucial in forming a uniform and almost level vertical beam. The array gain is 13-15 dBi depending on the beamforming algorithm used. Our operating frequency range (200MHz bandwidth centered at 5.8GHz) will be in a relatively flat gain range. The antenna easily accommodates WiMAX channels, which range up to 20MHz in bandwidth.

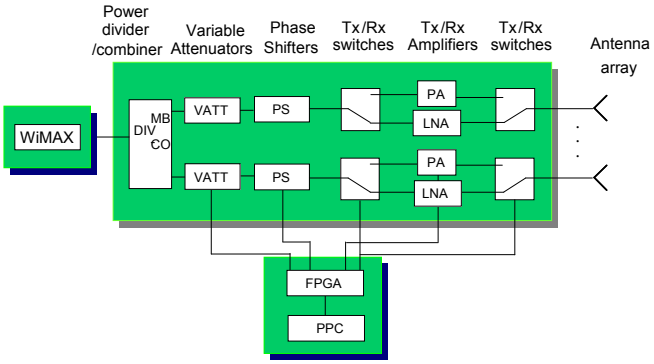



Fig. 2. Analog beamforming system diagram

The beamforming board (in Figure 2) is a digitally controlled analog beamforming system. The beamforming system consists of five functional blocks: the controlling unit, the analog beamforming circuitry, the radio interface, power detection, and power regulation. The controlling unit provides precise, digital, beamforming control via FPGA. The FPGA constantly delivers 96 control lines to eight separate phase shifters and eight attenuators in no more than 100 nanoseconds of combined propagation and delay time. The analog beamforming circuitry consists of an 8-way power divider/combiner, a 6-bit attenuator, a 6-bit phase shifter, two T/R switches, an amplifier stage for both transmit and receive, and filtering. The active circuitry after the power divider/combiner is duplicated eight times to provide the exact analog path for each of the eight channels. The radio interface is composed of circuitry for conditioning and current driving capabilities of a radio's T/R signal for proper synchronization. Low voltage control logic gates with high current driving capabilities are used to condition the signal and direct the signal to drive the switches and amplifiers with a delay no greater than 100 ns. When operating as a switched beam system it is able to form predefined beams in both Tx and Rx mode using co-phasal excitation or several windowing functions, such as chebyshev. The RF processing gain in receive mode is about 8 dB and in transmission mode 0 dB.

Local power regulation is kept to a minimum since a majority of the regulation is done by the Power Distribution System (PDS) system, however; there are crucial Low Drop-out (LDO) regulators and energy storage components which make up the power regulation block. A dual-output, power sequencing LDO supplies regulation for the FPGA and the radio interface circuitry. A very low-noise LDO supplies power for the power detection

circuitry because of its need for a clean reference. Also, a bank of low ESR, high energy, tantalum caps provide energy storage at the power input connector to mitigate the high current transients caused by the amplifiers.

Co-phasal excitation is the simplest of all spatial filtering schemes for beamforming. If a desired signal from a known DOA is chosen then the main beam of the antenna array can be steered towards this direction. Simply multiply each element by a complex weight, corresponding to a phase delay, so that when the signals are combined the signal from the desired direction at each element add completely in phase. Figure 3 shows the simulated beam pattern and lab measured beam pattern using the co-phasal excitation beamforming algorithm.

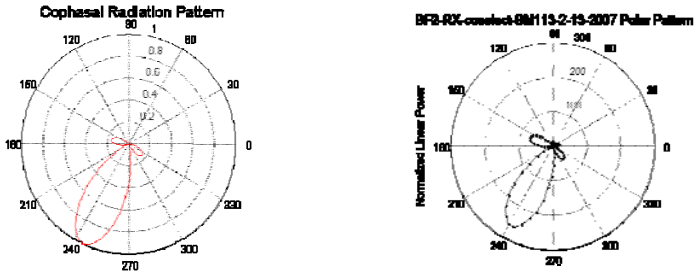



Fig. 3. Co-phasal beam pattern. Left: Matlab simulation; Right: lab measured pattern

Many lab tests and field tests have been carried out. Two cases have been tested: 1) new beamforming antenna at one side and a directional antenna (G=17dBi) at the other side and 2) new beamforming antennas at both sides. Both scenarios incorporated two Airspan Radio Units, WiMAX compliant to the IEEE 802.16d-2004 standard. The results are summarized in Tables 1 and 2. The field test results show that we can achieve high throughput at very long distance with our new beamforming antenna.

| Stationary Tests: | Gain Ant1/Ant2 (dBi) | Distance (Miles) | Throughput (Mbps) |
|---|---|---|---|
| 1 | 17/15 | 11 | 10.9 |
| 2 | 17/15 | 11 | 9.7 |
| 3 | 17/15 | 11 | 9.6 |

Table 1. Beamformer To Fixed Directional Antenna

| Stationary Tests: | Gain Ant1/Ant2 (dBi) | Distance (Miles) | Throughput (Mbps) |
|---|---|---|---|
| 1 | 15/15 | 11 | 17 |
| 2 | 15/15 | 11 | 16.2 |
| 3 | 15/15 | 24 | 2.5-4 |
| Mobile Tests: | Gain Ant1/Ant2 (dBi) | Distance (Miles) | Throughput (Mbps) |
| 1 | 15/15 | 4.5-6 | 10-12 |

Table 2. Beamformer to Beamformer

## 4. Adaptive Array Testbed

An adaptive array utilizes sophisticated signal processing algorithms to continuously distinguish among the desired signal and interferences and can form an unlimited number of beam patterns to optimally improve signal strength and suppress interferences.
An open-loop adaptive smart antenna testbed was developped to explore the feasibility of using AAS with high bandwidth radio systems[Panique 2008, Huang 2008, Khallaayoun 2009]. Despite the fact that closed loop systems are the most popular methods, an open loop approach was chosen though more complex to implement. Closed loop systems performance functions do not have unique optima and might converge to a local one, or even worst the algorithm might diverge, in addition, and as any in closed loop system instability becomes a concern [Widrow and Stearns, 1985]. Contrary to the closed loop approach that broadcasts to be able to detect sources, the open loop approach is solely based on the receive signal which makes the system very secure which is a wanted attribute in army related communication intelligence. In addition, the DOA estimation computational burden in the open loop scheme is done away from traffic which improves on throughput.
The adaptive array testbed consists of an eight element UCA operating at 5.8GHz, a translation board that down converts 5.8GHz to the baseband and a/d convertion, a beamforming board, a Data AcQuisition (DAQ) card and a PC interface that uses LaBVIEW PC interface for the testbed control.
In order to appropriately test a particular system, the different parts constituting the system should be tested individually and collectively. The testbed was designed to provide a flexible testing environment. For example, the testbed allows one to test the DOA estimation along with the beamforming modules individually to make sure that they function properly, one can then use the same testbed to examine the performance of the system utilizing both modules simultaneously.
The UCA array was designed to operate at a center frequency of 5.8 GHz. Because UCA offer a 360 degree beamsteering without a significant effect on the beamshape, along with the fact that no angular estimation ambiguity is inherent in the system, the circular geometry was chosen instead of the linear one.
An automatic calibration system was developped [Weber and Huang 2009]for multi-channel beamforming board calibration, which includes correction of RF circuit errors and mutual coupling effects for operating in both transmission and receiving modes.
The translation board represents the hardware responsible for taking the 5.8 GHz signal to baseband and delivering the information with minimal phase and magnitude jitter to the DAQ card. The translation board uses a single stage image reject mixing to achieve frequency translation to baseband. The RF signal is amplified, filtered and mixed using a distributed local oscillator; the oscillator can be tuned to any desired frequency. The baseband frequency is solely dependent on the speed at which one can acquire the signals. The Manual Gain control settings are used to provide an acceptable level to the DAQ card. The RF from the antenna array is fed onto the 8 top inputs shown in the snap shot. The lower SMAs are the baseband output with the LO fed at the center SMA. The Manual gain control settings are located on the back of the board. The WiMAX radios used are a product the "Airspan" company, the BS and SS units used fall under the MicroMAXd family which are based on the IEEE802.16-2004 standard.
The AAS starts by locating the bearings of the users and interference sources using the DOA estimation module. Once the impinging signals are acquired the processing is done via the

PC interface which exploits a variety of algorithms to estimate the bearings of users and interference sources. The algorithms are implemented in MATLAB and vary from spectral-based (Bartlett [Bartlett 1950] and Capon [Capon 1969]) to subspace-based (Multiple SIgnal Classification (MUSIC) [Schmidt 1986]). The next step consists of calculating the appropriate weights necessary to form a beam toward the wanted users and null the directions from which interference signals are originating from. The beamforming and nullsteering are achieved by translating the calculated weights into phase and magnitude settings which are sent to another DAQ card to the CPLD. The beamforming algorithms used are based on cophasal excitation and nullsteering. The WiMAX incoming or outgoing signals are fed to the beamforming board and become subject to spatial multiplexing.
In DOA estimation, the MUSIC algorithm was implemented to determine the bearing of the impinging signals. A detailed simulation study of different DOA estimation algorithms have been carried out, showing that MUSIC offers the best resolution but is computationally expensive. An alternative algorithm, namely the spatial-selective MUSIC[Khallaayoun 2007], has been developed that use greatly reduces the computational cost while not affecting the resolution. By splitting the space into sections using a switched beam method (computationally inexpensive) and by using the symmetric property of the array, one can reduce the number of element needed in performing DOA estimation.
The PC interface is used to process the incoming signals after acquisition in order to determine DOA, second it is used to beamform towards the users and null the interferences in addition to tracking mobile users [Panique 2008]. LabVIEW is used to control all the instrumentation and hardware. A snapshot of the PC interface front panel is shown in Figure 4.
The interface GUI contains the needed inputs needed to run the system in a flexible manner. The system characteristics, RF board control, beamforming parameters along with calibration data are shown in the GUI and can be changed as the user desires.

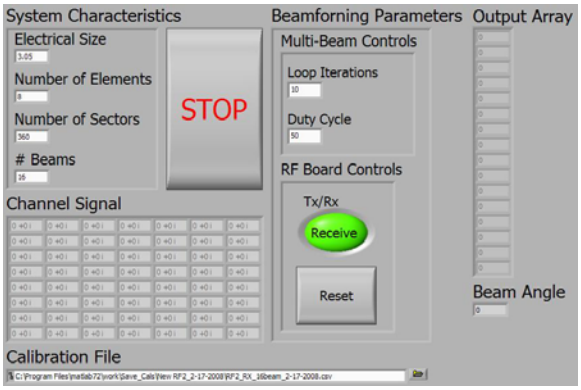


Fig. 4. PC interface front panel

Figure 5 illustrates the software conceptual flow diagram used in the PC interface. Once the signal is acquired, it goes though the DOA block determining the bearing of impinging sources which utilizes MUSIC algorithm. The sources are analyzed to determine if they are wanted users or interference sources. If interferences are detected nullsteering is used to null those bearings, if multi-users are detected the multibeam block activates and calculates

the appropriate weights. The tracking is also activated to lock to mobile users or moving interference sources.

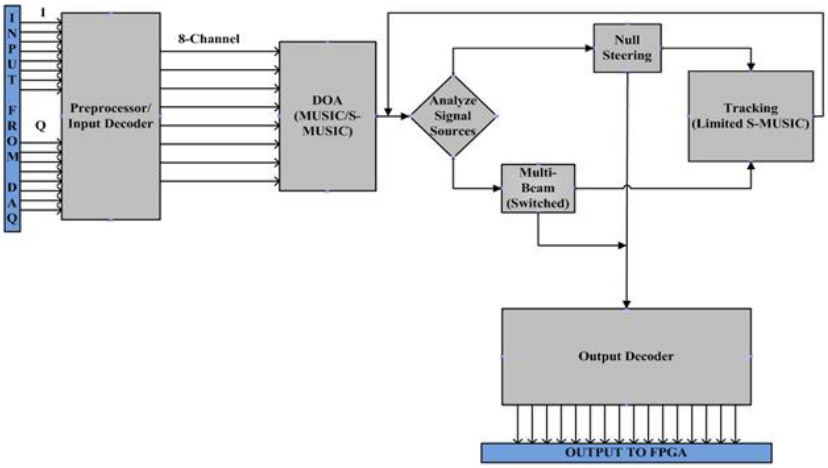



Fig. 5. Conceptual software diagram

The performance of the beamforming portion of the testbed was based on comparing the measured accuracy of the pointing angle, the height of the sidelobes and the depth the nulls with simulation results.
Co-phasal excitation and several window beamforming algorithms, including Chebyshev window filter have been tested. Figure 6 shows a comparison of a measured beam pattern with the simulated pattern with co-phasal beamforming.

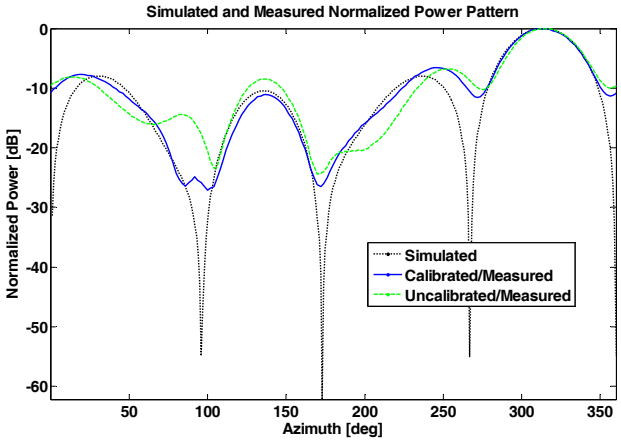



Fig. 6. Comparison of simulation with measured results for beamforming

The simulated beam and measured beam show very similar behavior, the measured beam point to within a few degrees of the expected bearing. The sidelobes measured where at the same location and just a few dB higher than the simulated results. The beamforming hardware and algorithms performed very well and almost matched the simulation results.
When there is interfernce signal, the system will operate in nullsteering mode. We have analyzed a nullsteering algorithm that can form deep null even in the main beam [Huang 2007]. Figure 7 depicts the simulation vs. the measured results with a desired target at 90 degrees and an interference source at 180º. The results indicated that the null in the measured pattern is 3º away from the interference location. The depth of the null was measured at -22 dB. Since the tunring table in our poor man anechoic chamber has 2 degrees error, the actual angle error may be smaller than 3º.

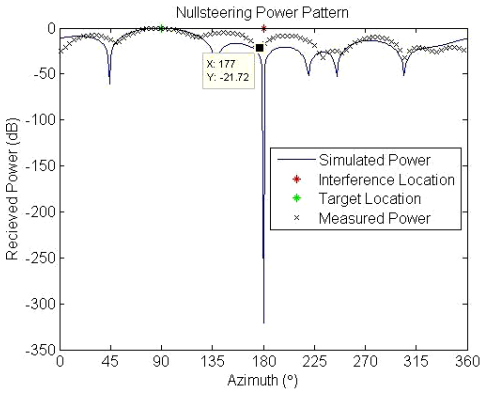



Fig. 7. Nullsteering pattern comparison.

## 5. Conclusion

This chapter reviews our design of smart antennas for WiMAX radio system. Our studies of smart antenna technologies indicate that compact, low cost, lightweight, highly directional antennas are feasible for mobile applications and can be inter-worked with emerging WiMAX chip sets following the IEEE 802.16d-2004 standard. Our design will be resulting in commercially available compact smart antennas for a WiMAX radio system. This will enable effective solutions to major problems in wireless networks in remote areas where a complete path from a source to a destination does not exist or where such a path is highly unstable. The smart antenna will also be effective in dense areas where interference mitigation is a critical consideration. This will also be the first commercial technology to consolidate several advanced techniques in one small size, light weight, multi-function, low cost communication unit providing users easy access to a communication network. The furure work will include digital beamforming system design, FPGA based DOA estimator and phase shifterless phased array development.

## 6. References


Bartlett, M.S. (1950) Periodogram Analysis and Continuous Spectra. *University of Manchester*

Bellofiore, S. , et al. (2002) Smart antenna system analysis, integration and performance for mobile ad-hoc networks. *IEEE Trans. Antennas and Propag.* Vol. 50, pp. 571-581.

Capon, J. (1969) High Resolution Frequency-Wavenumber Spectrum Analysis. *Proceeding of IEEE*, Vol. 57, NO. 8.

Compton, R. T. , et al. (1976) Adaptive arrays for communication-systems - overview of research at Ohio-State-University. *IEEE Trans. Antennas and Propag.* vol. 24, pp. 599-607.

Cooper, M. and Goldburg, M.( 1996). Intelligent antennas: spatial division multiple access. *Annual review of Communications*, pp 999-1002.

Eklund, C., et al. (2002) IEEE Standard 802.16: A technical overview of the wireless MAN™ Air Interface for Broadband Wireless Access. *IEEE Commun. Mag.,* pp. 98-107.

Godara, L. C. (1997). Application of antenna arrays to mobile communications, part II: beam-forming and direction-of-arrival considerations. *Proc. IEEE,* vol. 85, No. 8.

Hedayat, A.R. et al. (2007). Smart WiMAX Delivering personal broadband. *http://navini.com/assets/pdfs/White_Papers/2007_08_06_Smart_WiMAX.pdf*

Huang, Y. and Panique, M.,(2007). Performance analysis of a null steering algorithm. *IEEE APS International Symposium*. Honolulu, USA.

Huang, Y. (2008). Design of a Dynamic Beamforming Antenna for Wimax Radio Systems. *2008 IEEE Aerospace conference*. Big Sky, USA.

Huang, Y. et al. (2009). A Compact Smart Antenna for WiMAX Radio,", *IEEE Mobile WiMAX International Symposium*. Napa Valley, USA.

Khallaayoun, A. and Huang, Y. (2007). Spatial selective MUSIC for direction of arrival estimation with uniform circular array. *IEEE APS International Symposium*. Honolulu, USA.

Khallaayoun, A. et al. (2009). An Adaptive Smart Antenna Testbed for WiMAX Radio," *IEEE Mobile WiMAX International Symposium*. Napa Valley, USA.

Krim, H. and Viberg, M.,(1996). Two decades of array signal processing research," *IEEE Signal Processing Mag,* vol. 13, pp. 67–94.

Lehne, P. H. , and Pettersen, M. (1999). An overview of smart antenna technology for mobile communications systems. *IEEE Communication Surveys*, vol.2, pp.2-13.

Liberti, J. C. and Rappaport, T. S.( 1999) . Smart Antennas for Wireless Communications: IS-95 and Third-Generation CDMA Applica –tions. *Prentice Hall*. NJ, USA.

Loadman, C. et al, (2003). An overview of adaptive antenna technologies for wireless communications," Communication networks and services research conference. A3, pp.15- 19.

Nuaymi, L. (2007). *WiMAX Technology For Broadband Wireless Access.* John Wiley & Sons Ltd.

Panique, M, (2008). Design and evaluation of test bed software for a smart antenna system supporting wireless communication in rural areas. *Master degree thesis*, Montana State University, Bozeman, USA.

Schmidt, R. O. (1986). Multiple Emitter Location and Signal Parameter Estimation. *IEEE transactions on antennas and propagation*, vol. AP-34, No.3

Widrow, B and Stearns, S. D. (1985). *Adaptive Signal Processing,* Prentice Hall, Englewood Cliffs, New Jersey, USA.

Weber, Raymond and Huang, Y. (2009) An Automatic Calibration System for Beamforming Boards. *IEEE APS International Symposium*. Charleston, USA.

# 11

# Enabling WiMAX Video Streaming

M. Fleury, R. Razavi, S. Saleh, L. Al-Jobouri and M. Ghanbari
*University of Essex*
*United Kingdom*

## 1. Introduction

IEEE 802.16 (WiMAX—Worldwide Interoperability for Microwave Access) (Andrews, 2007; Nuaymi, 2007) is a broadband wireless technology that allows rapid deployment of video streaming services, including Video-on-Demand (VoD), multicast varieties of IPTV (TV over the Internet Protocol) and broadcast, digital TV (Meloni. 2008). Streaming assumes limited buffering of data for video-rate display, rather than video file download, video messaging services, or hybrids for video clip download in which display is overlapped with download of video chunks. In these services, delay is important but not as important as in interactive, conversational services such as video-phoning and video conferencing (Wenger, 2006), in which one-way delay should ideally approach (but not reach in practice) 100 ms. To illustrate the issues, the Chapter considers the possibility of a simple but effective adaptive modulation scheme for WiMAX offering unequal error protection (UEP) of video according to frame type (Ghanbari, 2003). To allow channel swapping when broadcasting TV channels in a multiplexed stream and to emulate VCR functionality in VoD, spatially-coded (Intra-frame) I-frames are often inserted around every half-second of video. I-frames also act to anchor the decoder in case of loss of synchronisation and consequent propagation of distortion. Other frame types include Predicted (P-) frames, which are backwardsly predicted from I- or other P-frames, and Bi-predicted (B-) frames, which as the name suggests can be predicted from I- and P-frames in either temporal direction. B-frames are not employed in the Baseline profile of the state-of-the-art H.264 codec (Richardson, 2003) in order to reduce computation levels on mobile devices but can be employed in streaming,when latency is less of an issue.

In the UEP scheme, by monitoring the input video bitrate, a burst profile is selected that offers a lower rate than the arrival rate to more important anchor frames. If I-frame data is protected then the less important P-and B-frame data rates must compensate, which occurs in the scheme through selection of one or more WiMAX burst profiles offering higher data rates. Allocation depends on the congestion level indicated by transmit buffer fullness. For example, low buffer fullness may allow some P-frame data to be protected while high buffer fullness requires some P-frame data to be allocated to a still higher data rate.

An important consideration for WiMAX is the effect of propagation conditions in the outdoor environment. The Stanford University Interim (SUI) channel model adopted in the IEEE 802.16 standardisation is an extensive and complete channel model, implemented by the authors within the well-known ns-2 simulator, and used along with the NIST WiMAX

simulation module. The behavior of the WiMAX modulation modes along with physical layer characteristics under the SUI model will be discussed herein. An effective UEP scheme is one that combines simplicity of implementation with compelling gains in video quality. One such algorithm is proposed in the Chapter. Applying an additive-increase and multiplicative-decrease algorithm to the choice of burst profile, results in many decibels gain in delivered video quality and significant gains in delay/jitter reduction over a non-adaptive modulation scheme.

Considerable research has taken place into bandwidth allocation, packet scheduling and queuing management for WiMAX service classes. The Chapter evaluates Active Queuing Management techniques for uplink scheduling, as this topic has received comparatively limited attention in the research community. In the WiMAX standard, queuing discipline and queue scheduling are left open to the developer for reasons of commercial differentiation. Unlike work conserving queuing mechanisms, AQM drops packets in anticipation of growing congestion. The Chapter will examine the most effective strategy for WiMAX uplink management.

The current H.264 codec outputs compressed data encapsulated as Network Adaptation Layer Units (NALUs) which are mapped onto MAC Packet Data Units (MPDU), before transmission over the WiMAX link. However, there are different ways to aggregate or segment the H.264 NALUs according to the MPDU size, which may be fixed or variable. As an alternative, UEP can also be based upon H.264 data partitioning or differential error concealment can be applied. In data partitioning, the more important parts of the compressed bitstream are collected in separate NALUs, so that in H.264 up to three different types of NALU can be generated and the more important partition protected with increased application layer Forward Error Correction (FEC). Our research is at an exploratory stage for this topic, and Section 5 in conjunction with Section 2 introduces the topic. Section 3 reports research on the UEP scheme, whereas Section 4 reports investigations into WiMAX queue management with video transport in mind. Section 6 concludes this Chapter.

## 2. WiMAX background

This Section briefly introduces WiMAX in preparation for descriptions of recent developments in video streaming in later Sections. It is assumed that general descriptions of WiMAX and its applications will be available to the reader from other sources, some of which appear in the references to this Chapter (Andrews et al. 2007, Nuaymi, 2007). For coverage of standard video codecs in general, we can but recommend one of the contributors to this Chapter's textbook (Ghanbari, 2003), while for H.264 in particular (Richardson, 2003) is helpful. For video transport over wireless (Sadka, 2002; Schaar & Chou, 2007 ) are recommended.

WiMAX is a competitive solution for `last-mile' broadband access compared to cable, digital subscriber line (DSL) or fiber, because of the low cost of deploying WiMAX over a wide area. As such WiMAX may be suitable for depolyment in rural areas (Hillestad et al., 2006) in which there is limited previous communication infrastructure, particularly within the developing world. Moreover, rural areas do not present the same problems of wireless propagation as exist in urban areas. There are two main standards for WiMAX: the first one is IEEE 802.16-2004 for fixed wireless connections and the second one is IEEE 802.16-2005 for mobile wireless connections (Yan & Kavehard, 2006). WiMAX operates within two licensed

spectrum ranges: 1) from 2.5 to 2.7 GHz in the USA and 2) a range from 3.4 to 3.7 GHz in Europe and the rest of the world. Fixed WiMAX (IEEE 802.16-2004) is suitable for the frequency band 2-11 GHz and the theoretical range is up to 30 miles for line-of-sight (LOS) communication with data rates up to 75 Mbps and for non-line-of-sight (NLOS) the theoretical range is 5 miles with data rates up to 5 Mbps. WiMAX supports different duplex modes (Tariq et al., 2007): 1) time division duplex (TDD), and 2) frequency division duplex (FDD) and 3) half-frequency division duplex (H-FDD). For TDD, the system employs the same frequency to transmit over the downlink (DL) and uplink (UL) but within different time slots. Projections normally assume (Yun & Kavehard 2006) that the DL will have the predominance of the time slots (see Section 2.1 for sub-frame structure), though the allocation could vary dynamically. In FDD, the system requires two different frequencies, while for H-FDD the system uses two different frequencies with different time slots. FDD is most suitable for symmetric loads such as voice channels, while TDD is more suitable for asymmetric loads such as video streaming.
WiMAX has two main components, Base Stations (BSs) and Subscriber Stations (SSs), BSs uses downlink to transmit while the SSs use the UL to transmit. The 802.16 standard can work in a point–to-point (PTP), point-to-multipoint (PMP), and mesh topology modes. By using omni-directional or directional antenna (Andrews et al., 2007) the effective range of the system can be increased.

### 2.1 WiMAX frame structure

The key to understanding how video streams may map onto the WiMAX frame structure is knowledge of the WiMAX frame structure. However, those readers familiar with this background material may wish to proceed to Section 3.
The structure takes different forms across the protocol stack layers. The physical layer of WiMAX is based on the physical layer of Orthogonal Frequency Division Multiplexing (OFDM), which gives WiMAX immunity to frequency selective fading and intersymbol interference. Assuming the more spectrally efficient TDD mode, for the fixed standard (IEEE 802.16-2004) the size of FFT is limited to 256 subcarriers, 192 of the being subcarriers used for data carrying, with 8 subcarriers used as pilot subcarriers for synchronization and channel estimation, while the rest of the subcarriers are used as guard subcarriers.
Figure 1 shows the TDD frame of fixed WiMAX (Andrews, 2007). This frame is divided into a DL sub-frame and UL sub-frame. The frames are separated by a receive-transmit transition gap (RTG) and the sub-frames are separated by transmit-receive transition gap. The DL sub-frame starts with a preamble and then a Frame Control Header (FCH). The FCH contains the DL- Mobile Application Part (MAP), UL- MAP, downlink channel descriptor (DCD), uplink channel descriptor (UCD) and one or more DL-bursts. The other DL-bursts are used for data .The UL sub-frame starts with ranging slots, Contention for Bandwidth Request (CBR) and UL-bursts.
One of the attractive features of the Medium Access Control (MAC) layer of WiMAX that it is able to construct and transmit fixed and variable-sized frames. The MAC layer has three sublayers; these sublayers interact with each other via service access point (SAPs). The service-specific convergence sublayer achieves the mapping or transformation of external network data by the help of the SAP. The common part of the SAP receives information from the MAC Service data units (MSDUs), which are packed into the payload field to form MAC protocol data units (MPDUs).

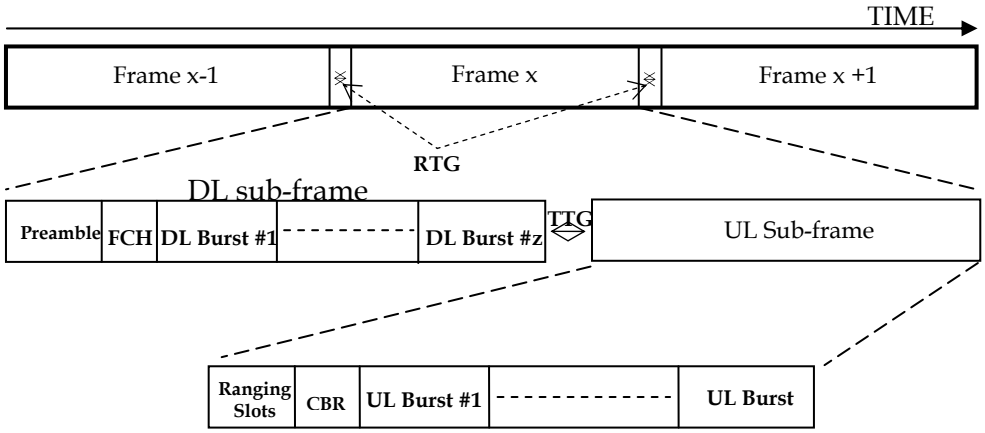



Fig. 1. Fixed WiMAX OFDM-TDD frame stucture

The common part sub-layer is able to pack MSDU into MPDU in different ways (more than one complete or partial). This ability is called aggregation. Figure 2 shows the ability of the MPDU to accommodate more than two MSDUs but not three. As a result, the remaining part of a third MSDU is packed with the previous two MSDUs to fill the remaining payload field to prevent wastage of resources. Feedback from an SS and on-air timing slots decide the payload size. Aggregation in the payload of MPDU is indicated by 'type' bit in the generic MAC frame header (GMH). Subheaders (Ahson & Ilyas, 2008) consist of two bits of fragment control (FC), three bits of fragment sequence number (FSN) when the MSDU is fragmented and three bits of length field. To indicate the start of the next subheader in the payload. FC is set to 00 when the MSDU is not fragmented and if the MSDU is fragmented, FC is set to 10, 01 or 11.

The common part sublayer also has the ability (Ahson & Ilyas, 2008) to fragment an MSDU across multiple MPDUs. Figure 3 shows how a portion of a single MSDU taking up the entire payload of an MPDU. When the size of the transmitted packet of MAC payload is too small to accommodate a complete MSDU, the MSDU is fragemented and a fragment is placed in another MPDU. Here, the FC bit is set to 01 if it is the last fragment of the MSDU; 10 if it is the first fragment of the MSDU; and if it is the MSDU is fragmented, FC is set to 10, 01 or 11.

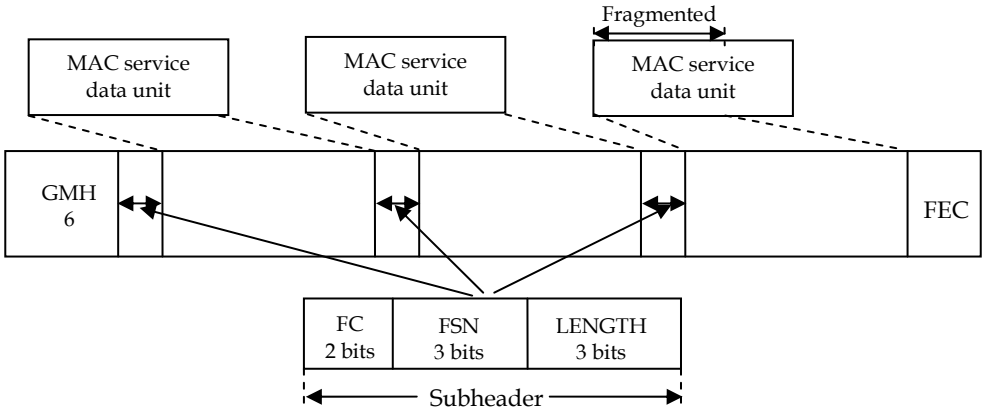



Fig. 2. Multiple MSDUs forming one MPDU

## 3. Unequal Error Protection

### 3.1 WiMAX modulation

Fixed WiMAX (IEEE 802.16d) employs OFDM to counter the effect of inter-symbol interference (ISS) when channel delay spread, τ, is large relative to symbol time $T_s$, due to high data rates and multipath fading. OFDM achieves this by dividing the input into $n$ sub-carriers such that that $nT_s >> \tau$ . To avoid any ISI through sub-carrier roll-off a cyclic prefix (CP) acts as a guard interval between bursts greater than τ. In fact, addition of a CP allows the application of an Inverse Fast Fourier Transform (IFFT) to act upon the data stream as if performing a cyclic convolution to form the $n$ sub-carriers.

As described in Section 2, through OFDM, IEEE 802.16d NLOS supports 256 sub-carriers in a TDMA/TDD system, with scheduling of the sub-carriers at a BS, with notification of scheduling to SSs in the frame control header. Appropriate symbol mapping to the complex frequency domain allows selection of the modulation mode prior to IFFT. Within a DL sub-frame, data bursts are ordered according to the robustness of the burst profile. Therefore, depending on the adaptive modulation mode chosen, data can be mapped onto one or more DL-bursts, as detailed in Section 2.1. Table 1 shows the 7 burst profiles, with change of Forward Error Coding (FEC) rate achieved by puncturing upon an inner non-recursive convolution code (CC) (with symbol-level protection from a fixed outer Reed-Solomon code).

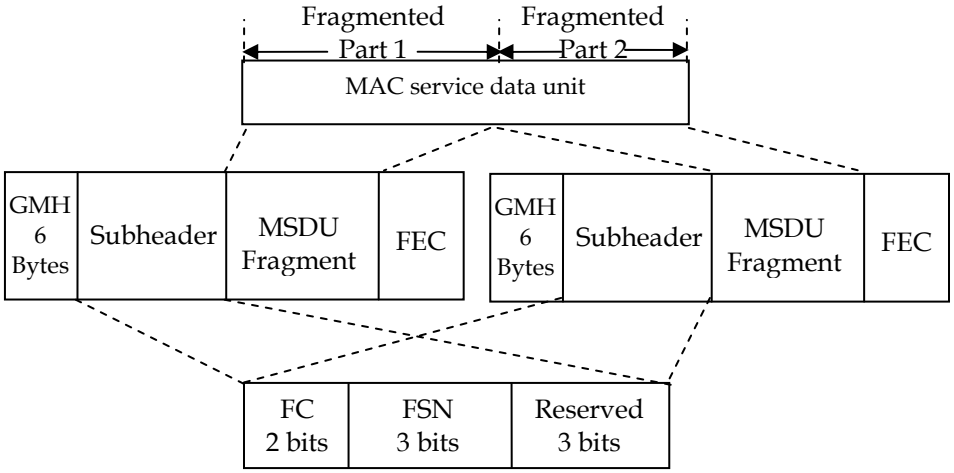



Fig. 3. Single MSDU fragmented across more than one MPDU

Selection of the appropriate WiMAX modulation mode in practice very much depends on channel estimation. WiMAX provides sub-carrier pilot tones for this purpose, as well as optional data burst preamables and/or midambles. Various schemes exist in WiMAX for placement of pilot tones, but other than noting this is important for channel estimation (Negi & Cioffi, 1998), the topic is beyond the scope of this Chapter. It is also possible (Juan et al., 2007) to monitor the downlink channel at the receiver and return the results in the UL sub-frame.

| Profile | Modulation mode | Coding rate |
|---|---|---|
| 1 | BPSK | 1/2 |
| 2 | QPSK | 1/2 |
| 3 | | 3/4 |
| 4 | 16QAM | 1/2 |
| 5 | | 3/4 |
| 6 | 64QAM | 2/3 |
| 7 | | 3/4 |

Table 1. Modulation profiles under fixed WiMAX, BPSK = Binary Phase Shift Keying, QPSK = Quadrature PSK, 16QAM = 16 Quadrature Amplitude Modulation

### 3.2 WiMAX propagation model

Broadband wireless propagation in general is introduced in (Anderson, 2003). There are a number of empirical models such as IEEE 802.16 (SUI), COST-231 Hata model, MMDS Band Empirical Path Loss. The channel model set selected by us were the IEEE 802.16 Stanford University Interim (SUI) models (Ercge et al., 2001; Hoymann, 2005). From these, channel two out of six empirical channels designed to cover terrain in the US. SUI-2 models a flat

terrain with light tree density and low Doppler spread. The cell radius is assumed to be 6.4 km. Antenna height at the BS is 15.24 m, beamwidth 120° and at the SS is 3.048 m, beamwidth 50°, with vertical polarization only. The SUI models provide generic system development modelling, as does COST-231. Studies such as (Milanovic, 2007) have concluded that the SUI model provides good accuracy for NLOS propagation but should not be relied upon for LOS propagation, when other models may be preferable.
Table 2 summarizes settings in the Standard for a 7 MHz licensed bandwidth, for which we chose to simulate the 5 ms frame duration. Notice that profile 1 is reserved for burst mapping data. The SSTTG/SSRTG are Subscriber Station Transmit (Receive) Transition Gaps. The achievable throughputs, with hard CC decoding, according to burst profile are shown in Figure 4. For SUI-2, a WiMAX CP of 1/32 is most appropriate, yielding a relatively limited delay spread of 0.2 µs. Figure 5 shows the bit error rate (BER) response of the differing burst profiles for SUI-2. For a given burst profile, there is a rapid fall-off in acceptable BER, once outside a narrow Signal-to-Noise (SNR) range. There are also uneven jumps in SNR and, thus, selecting a burst profile to achieve a target BER is difficult.

| Parameter | Value(s) |
|---|---|
| OFDM symbol duration | 34 µs |
| Channel bandwidth | 7.0 MHz |
| Frame duration | 5 ms, 10 ms, or 20 ms |
| Maps modulation | BPSK |
| FEC code | RS-CC |
| SSTTG/SSRTG duration | 1 OFDM symbol |

Table 2. Standardized system parameters

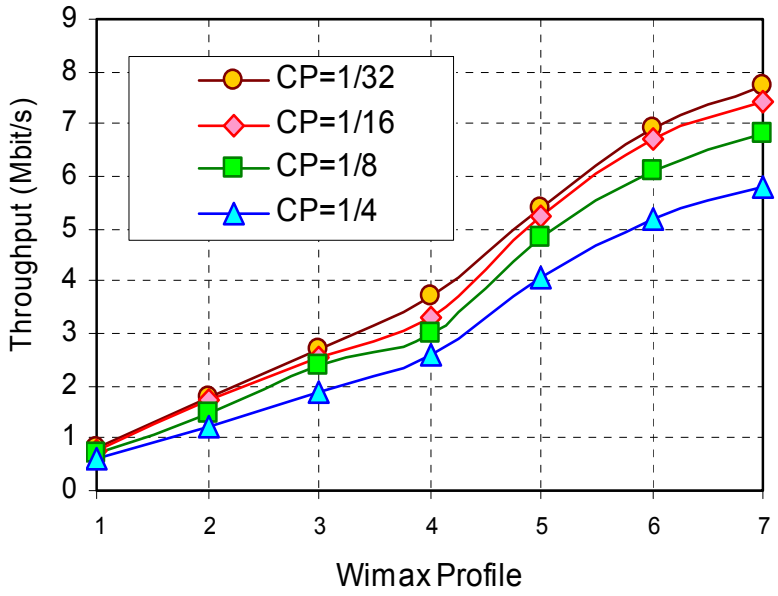



Fig. 4. Throughput for various CP, with 7 MHz bandwidth and 7:3 ratio of DL:UL

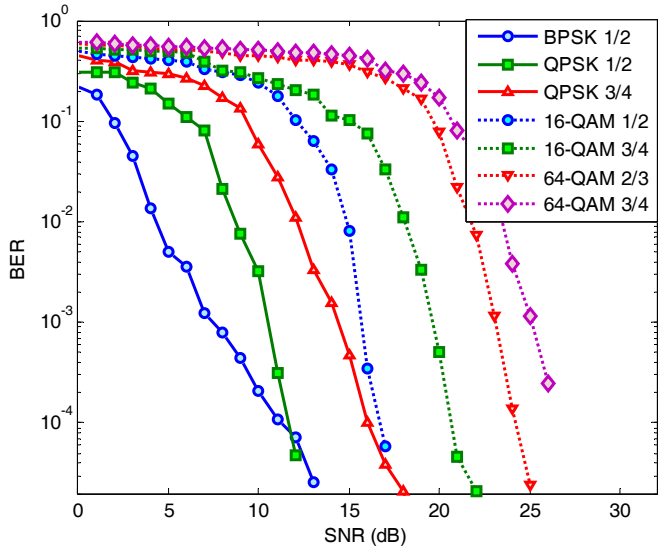



Fig. 5. BER over an SUI-2 channel for CP = 1/32 with 7 MHz bandwidth and various burst profiles.

### 3.3 Adaptive modulation

Adaptive modulation schemes for video UEP purposes have been considered in general (Wenger, 2003) or for particular wireless technologies (Razavi, 2008). Frequently, video UEP is combined with cross-layer techniques to vary error protection, for example in (Chen et al., 2001)). However, WiMAX already supports variable error protection through changes in the coding rate at the data-link layer. Examples of video UEP for WiMAX are apparently few. In (Juan et al., 2007) for a scalable video stream, the stream was partitioned between different connections to the same station, using the mobile version of WiMAX, IEEE 802.16e. This method did employ different burst profiles within the same UL sub-frame. However, the wireless channel model was the well-known Gilbert-Elliott, two-state Markovian model, which though suitable for modelling 'bursty' errors in indoor environments, fails to take into account the impact of terrain, precipitation, or attenuation in outdoor environments. For data transport in general, adaptive modulation is an important technique with recent attention (Hanzo & Choi, 2007) concentrated on adaptive OFDM to support constant error rates.

### 3.4 Unequal error protection with adaptive modulation

The general adaptive modulation procedure that we employ, which is repeated over time, is summarized in Figure 6. Arriving compressed video data is monitored to form a time-averaged estimate of the arrival rate. By means of this estimate the lowest modulation profile that can satisfy the rate at that time is determined. The protection policy across the video packet frame types is updated according to send buffer fullness at the WiMAX BS. All I-frame packets are protected with a robust profile but, to compensate, other packets are accorded less protection according to a policy, which evolves with buffer fullness.

The first step in our UEP algorithm is to apply a moving average filter to the arriving video data in order to form an estimate of the required WiMAX burst profile, *M,* that fulfils the input data rate. The input video rate is monitored through a linear predictive filter (LPF) (Honig & Messerschmitt, 1990) over 50 Group of Pictures[1] (GOPs). For example from Figure 1, for an average 2.5 Mbps input rate, *M* would be set to profile three, as this most closely meets the data rate requirements. Notice that in this instance it is possible to choose the profile without additional channel estimation, avoiding inaccuracies in estimation.
Upon determination of *M*, the profile *M*-1 is invariably selected for all arriving IP packets bearing I-frame data. Relative allocation of P- and B-frame packets to the different burst profiles is now decided by the normalized buffer fullness, *B*. The fullness is updated after an interval, empirically determined to avoid excessive fluctuations in burst profile selection policy. Assume an available capacity index (ACI) initially set at 1 within a range [0, 1], where 1 signifies an empty buffer. Changes to the ACI are only made if buffer fullness either exceeds a high threshold or drops below a low threshold. If *B* falls below the low threshold ACI is additively increased by $\alpha$, which has the effect of increasing the number of packets sent at a lower data rate. If *B* rises above the high threshold then ACI is multiplicatively reduced by factor $\beta$ which correspondingly backs off the sending rate to avoid packet loss. This is achieved by favoring higher data rates through higher burst profiles.
Transmit buffer fullness is a general indicator of congestion from input traffic arriving at the BS and of channel conditions. As such the algorithm can be applied across the burst profiles and according to traffic sharing the TDMA frame.

### 3.5 Experimental verification of UEP adaptive modulation

The well-known ns-2 network simulator augmented with the NIST WiMAX module was employed to model the system. The simulations were repeated and averaged over fifty runs to ensure convergence in the face of a high channel error rate. WiMAX's frame duration was set to 5 ms. The video-bearing DL sub-frame was allocated 70% of the frame duration. In order to isolate the impact of the video stream upon buffer fullness, the arriving video stream was the sole input into the BS send buffer. Thus, the DL sub-frame was occupied by a set of bursts according to the allocation by frame type packet in the UEP algorithm. Data is always assumed to be available to fill the remaining duration of the DL sub-frame as well as the UL sub-frame, though as mentioned, it does not occupy the send buffer. Table 3 reports other settings for the UEP algorithm during simulations. Buffer size was set to 50 packets as also used in (Li & Schaar, 2004) for video streaming over wireless.

[1] In this Chapter, the terms 'picture' and 'frame' are employed interchangeably as we assume progressively scanned video throughout.

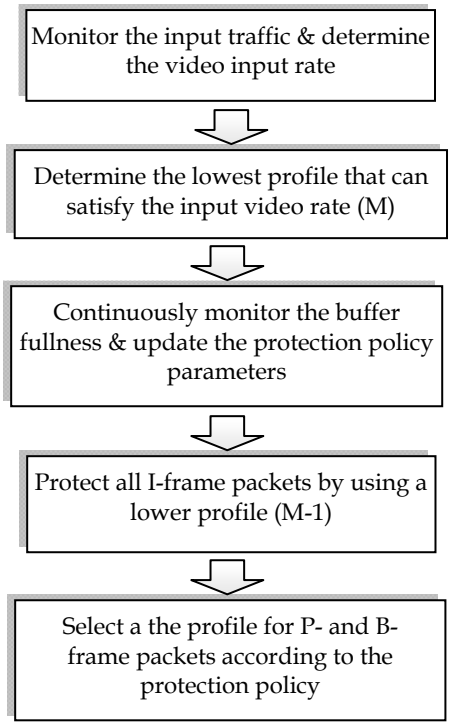



Fig. 6. Overview of the UEP scheme

| Parameter | Setting |
|---|---|
| Buffer capacity | 50 IP packets |
| Update interval | Every 50 IP packets |
| Low threshold | 0.20 |
| High threshold | 0.80 |
| $\alpha$, $\beta$ | 0.05, 1.25 |

Table 3. Parameter settings for the UEP algorithm.

To establish the efficiency of the algorithm, a non-adaptive scheme was compared with the UEP scheme. The LPF was again used to select the appropriate profile from Table 1. However, unlike in the UEP algorithm, the same profile was selected for all arriving video-bearing IP packets.

Three MPEG-2 video clips appear in the experiments. All were encoded at a variable bit-rate (VBR) with a target rate of 2.5 Mbps for 40 s. Clip 1, 'News', contains some motion, showing a newsreader and changing backdrop. Clip 2, 'Friends', an excerpt from the well-known situational comedy, is more complex, with scene cuts and moderate motion. Finally, clip 3, 'Football', contains considerable motion and is of higher complexity. Most legacy video is encoded in MPEG-2 and it remains widely used in the broadcasting industry. Peak Signal-

to-Noise Ration (PSNR)[2] was found by reconstructing with a reference MPEG-2 decoder. The display rate was 25 Hz, resulting in 1000 frames in each simulation run (averaged over fifteen runs). The source video format was European Source Input Format (SIF) $352 \times 288$ pixel/frame, with a GOP structure of N = 12, and M = 3, i.e. an IBBPBBP…structure. Error concealment was by previous frame replacement, which is a simple but standard scheme. The scenario modelled assumed arriving video-bearing IP packets at the BS containing, for error resilience purposes, one slice per packet. A slice in MPEG-2 takes-up a row of macroblocks.

Figure 7 shows packet loss rates for the 'News' clip under the SUI-2 channel model at several different average SNR settings[3]. Due to the adaptive modulation scheme, not only does the total packet loss rate considerably reduce but the proportion of lost B-frame packets increases, resulting in preservation of the vital I-frame packets. The loss rate of I- frames is always less than 1% with the adaptive scheme. As an example, at 17 dB SNR, the loss rate of B- and P-frames combined is 24.68% without adaptation and 15.76% with. Because the adaptive scheme allocates I-frame packets to a more robust profile, the result is a general reduction in the number of lost packets. The presumably increased loss rate of packets allocated to a less robust profile does not match the gain from allocation to the more robust profile.

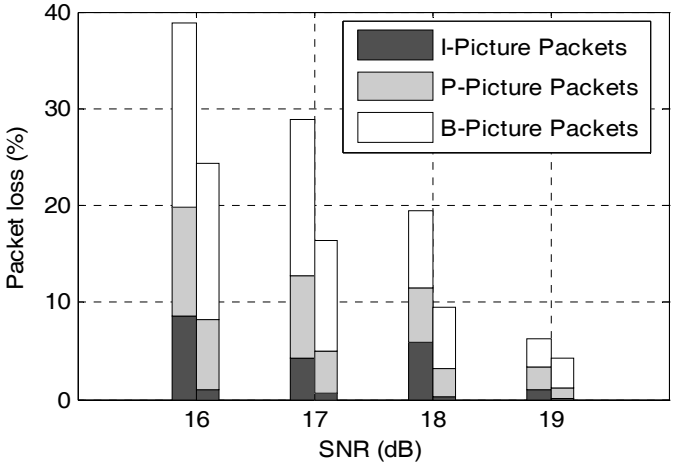



Fig. 7. Comparison of IP packet loss rates for (lefthand) non-adaptive modulation; and (righthand) adaptive modulation for the 'News' clip under the SUI-2 channel model.

Figure 8 compares the resulting video quality (luminance PSNR) between non-adaptive and adaptive modulation with UEP. The effect of UEP is to accentuate the gain from the reduction in packet losses from adaptive modulation. At 16 dB, with non-adaptation all clips would be unwatchable, while the UEP scheme quality verges on the acceptable by virtue of an almost 10 dB improvement. At SNR of 17 dB, the video quality is good when UEP is

[2] PSNR = 10 log ($MAX^2$/MSE), where MAX is the maximum intensity value possible for a pixel, and ME is the pixel-wise mean square error between a reference frame and the frame under test.

[3] Lower SNRs result under this model result in unacceptable video quality and are not presented.

applied but is poor otherwise. When SNR improves, as might be expected the relative advantage of UEP declines. However, UEP at 19 dB SNR still contributes to several dBs improvement in video quality, which is now excellent at 40 dB.

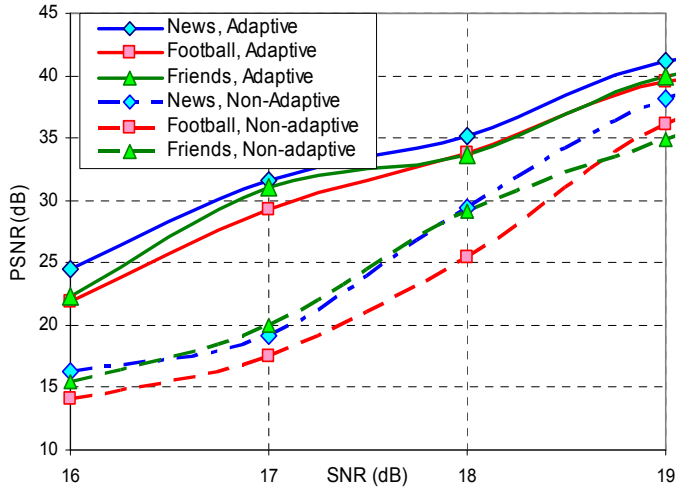



Fig. 8. Comparison of video quality for three video clips of increasing complexity using adaptive and non-adaptive modulation under the SUI-2 channel model

Figure 9 shows that delivered video quality is somewhat sensitive to selection of α and β in the adaptive algorithm. The Figure shows variation of the chosen parameters with a drop in PSNR if these parameters are not chosen well. Because the additional plots in Figure 7 occur with larger values for α and β then the changes in rate would tend to be more dramatic, with more of a risk from packet loss or under- utilisation of the available bandwidth. However, no dramatic variation in quality occurs, indicating the robustness of the algorithm.

The impact of the adaptation technique on the delay and delay jitter experienced by the application is now considered. If not probably addressed, the delay and delay jitter (variance of delay) could degrade the performance of an application. Click-on-Web-clip users are especially delay intolerant, as are interactive applications, when the delay should be no more than 100 ms in total (Wenger, 2003). Longer start-up delay requires larger playout buffers and possible overflow if not provided or underflow if excessive jitter occurs. Unfortunately, adaptive buffers sensitive to changes in network traffic can result in disconcerting playout discontinuities. By utilizing WiMAX's ARQ provision, IP packets contained in a burst received in error after application of WiMAX FEC are retransmitted in subsequent DL sub-frames up to a pre-determined retransmission limit (RL), after which the data is discarded if not successfully received.

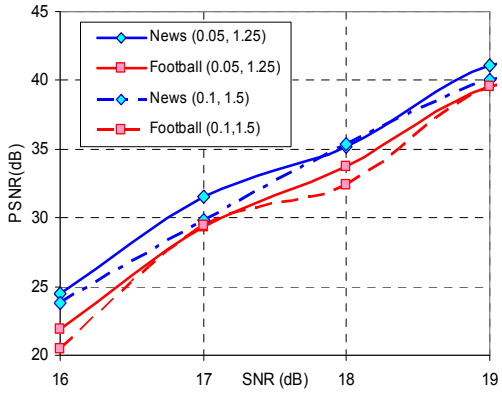



Fig. 9. Comparison of video quality for two video clips of increasing complexity using adaptive modulation under the SUI-2 channel model with variation of α and β parameters.

In Figure 10 (for the 'News' clip) the RL is varied and a comparison is made of the resulting delay, with and without the presence of UEP. In all cases, delay is significantly greater without adaptive modulation than it is with. Particularly for RL = 1, average SNR, though the delay of all schemes significantly improves for SNRs over 18 dB. From Figure 7, it was observed that the packet loss figures were generally reduced for the adaptive modulation scheme, irrespective of the impact of UEP. Though throughput is reduced for those packets transmitted under a robust profile, this delay is more than compensated by the reduction in packet loss, which also reduces the need to retransmit.

As will be noticed from Figure 11, the levels of delay are quite significant for all schemes at lower SNR, especially when it is considered that there will be additional delay across the core network prior to final hop WiMAX delivery, as well as any delay at the codec for real-time applications. Again adaptive modulation always alleviates jitter, Figure 11, and this is as well because levels of jitter are relatively high.

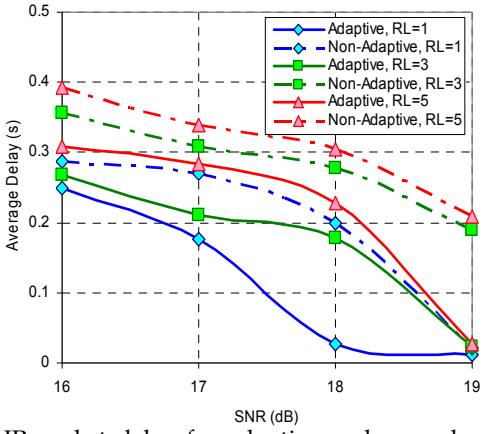



Fig. 10. Comparison of IP packet delay for adaptive and non-adaptive schemes across a WiMAX link

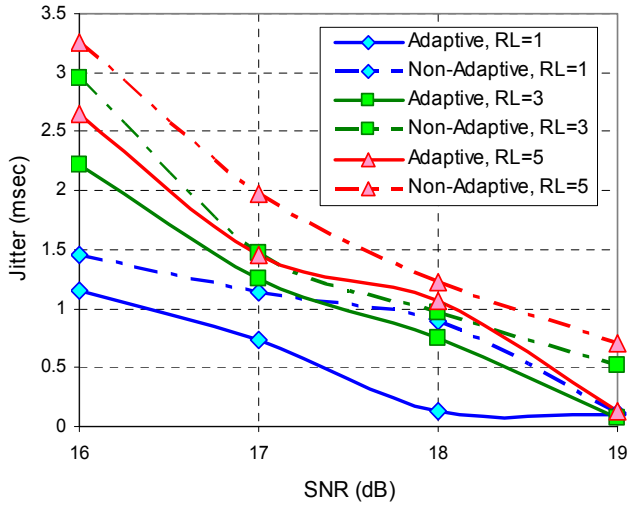



Fig. 11. Comparison of IP packet jitter for adaptive and non-adaptive schemes across a WiMAX link

## 4. Queue management for video streaming

### 4.1 WiMAX uplink scheduling and management

WiMAX in all its versions (IEEE 802.16 parts a–e) does not specify queue management algorithms at the BS or SS, allowing these, along with bandwidth allocation at the BS and packet scheduling at the SS, to be designed by vendors as a means of commercial differentiation. Choice of queuing discipline is sensitive to application type and service class whether it is implemented on the BS or the SS. WiMAX networks support multiple service classes to accommodate heterogeneous traffic with varying requirements. In particular, these are Unsolicited Grant Service (UGS), extended real-time Polling Service (ertPS), real-time Polling Service (rtPS), non-real-time Polling Service (nrtPS), and Best effort (BE). UGS is designed for synchronous services such as legacy voice without silence suppression. The ertPS class (Zhang et al., 2006) is intended for voice with silence suppression, when a SS may not have data available upon polling by the BS. The rtPS class is most suitable for real-time video services, particularly for Variable Bitrate Video (VBR), which is employed to maintain delivered video quality but may lead to 'bursty' arrival rates.

Figure 12 shows how an SS communicates with a BS and how the UL packet scheduling works across different service classes. An initial request by the SS results in a bandwidth allocation by the BS to that SS. The SS packet scheduler is able dynamically to request a modification to its bandwidth allocation in a piggybacked requested to the BS. However, for simplicity of analysis in this paper it is assumed that the allocation remains static. As an SS may not have data available, it is polled by the BS. Polling may be on a group basis or in unicast fashion. As group polling may result in access delay at the SS, this paper assumes unicast polling, though there are UL sub-frame utilisation implications (Chang, 2008). The BS passes video and other data to the SS over the DL sub-frame whereas the UL sub-frame accommodates its traffic according to service class and burst slot availability.

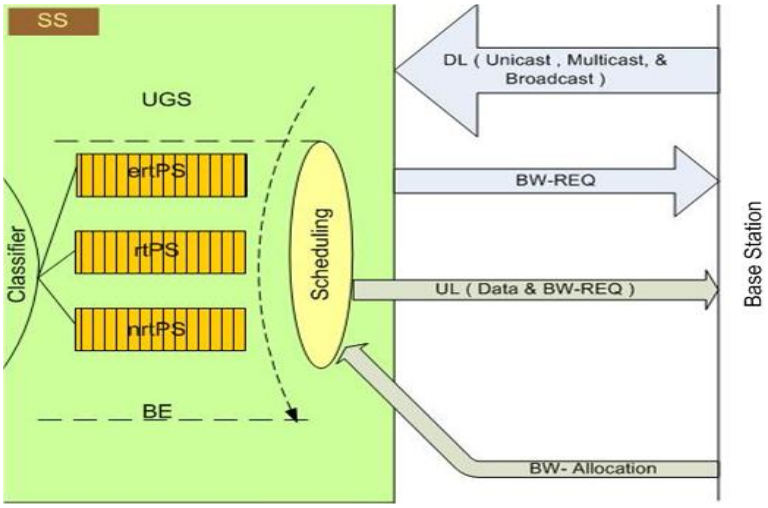



Fig. 12. IEEE 802.16 uplink scheduling system

The UGS is allocated bandwidth as of priority, while by default SS servicing of polled queues takes place in round-robin fashion (though many other possibilities exist [16]). In Figure 12 the queues are serviced by the packet scheduler in the order indicated with BE traffic considered last. In this Chapter, we assume an exhaustive service with the default simple round-robin access by the SS packet scheduler.

This Chapter concentrates on a particular class of SS UL queue discipline, Active Queue Management (AQM), as these have at least the potential to benefit real-time services by favoring traffic with urgent deadlines or with differing packet priorities (Orlov & Necker, 2007). Within the fixed Internet, AQM techniques (Koo et al., 2004) are intended to achieve high link utilisation without introducing delays. The most well-known AQM technique is Random Early Detection (RED) (Floyd & Jacobsen, 1993) in which packets in the queue are intentionally dropped or explicitly marked as a signal to the source to reduce its rate. For compliant sources, notably TCP, a dropped packet prompts the congestion control mechanism to reduce its sending rate. For a non-compliant source, one that does not reduce its rate, RED (Floyd & Jacobsen, 1993) also keeps the queue size low but does not discriminate against 'bursty' traffic sources. Dropping occurs in a probabilistic fashion according to the time-wise, exponentially-averaged queue size and no record is kept of the state of each connection to avoid unreasonable overhead. By random dropping, rate synchronization amongst compliant sources is avoided.

RED has a number of helpful features but it is known (Rahmani et al., 2008) to underperform when there are few co-existing flows in a queue because buffer occupancy fluctuates rapidly. In an SS, the number of coexisting flows could well be limited. RED maintains two thresholds: a low threshold in which all arriving packets are accepted and a high threshold when all arriving packets are rejected. When buffer fullness is between the thresholds, RED is in the congested state in which packets are randomly dropped. However, RED may not be able to react quickly enough to 'bursty' traffic operating within this region. Unfortunately, if buffer size is increased to compensate, apart from the drain on energy at

any mobile device arising from the increased buffer size, interactive multimedia applications may suffer from a large buffer as a result of increased waiting times.
Random Exponential Marking (REM)'s packet marking probability (Athuraliya, 2001) grows considerably more aggressively than RED's. Its congestion measure is based on any mismatch between input rate and link capacity and between queue length and target length, though in practice rate change is measured by the rate of change of the queue length. The aim of REM is to stabilize the queue length to a target which is independent of traffic arrival intensity and of RED's queue thresholds. Like REM the BLUE algorithm (Feng, et al., 2002) also decouples queue length from congestion management. Rather than queue occupancy, BLUE employs a history of the current buffer packet overflow rate and link utilization to form the packet dropping (marking) probability. By updating its estimates over time, BLUE is able to learn the correct form of queue regulation. BLUE can also randomize times between marking probability updates to avoid source synchronization.
Drop-tail (FIFO) queue management (Clark et al., 1992), though not a form of AQM is ubiquitous across Internet routers and has the advantage that it scales well and shares delay between different connections. However, drop-tail may lock-out some connections to the advantage of others occupying the buffer. It also may maintain buffer fullness for longer periods of time, as it only signals congestion when the buffer is full.

### 4.2 Video streaming from an SS

In this Chapter, we consider two alternative ways of transporting video across an UL, the first through a basic UDP stream, representing a non-compliant source and the second through the TCP-Friendly Rate Control (TFRC) (Handley et al., 2003), representing a compliant source. TFRC is intended to provide a less aggressive congestion control than TCP's 'sawtooth' response to network congestion, resulting in a smoother sending rate more suitable for video streams. As a TFRC responds to packet loss, as well as round-trip time and packet size, it will change its rate if packets are actively dropped from an SS queue. TFRC has been widely studied in the fixed Internet and there is some interest in its use within wireless networks (Chen & Zakhor, 2006). Because our interest is in queue management issues, the wireless channel is assumed to be error-free, while a detraction of TFRC for wireless networks is that it will also change its sending rate in response to packet losses on the channel, not just those resulting from congestion. However, in the event of channel loss, packet retransmission without acknowledgment to the TFRC application can avoid this problem, as it has for TCP in the SNOOP approach (Balkrishnan et al., 1997). Alternatively, in (Chen & Zakhor, 2006) multiple TFRC connections are employed to dissipate channel packet loss on any one of the wireless connections.
In the simulations reported in Section 4.3, the reference 'Paris' clip was H.264 VBR-encoded at Common Intermediate Format (CIF) (352×288 pixel/frame) with 4:2:0 sampling and target bit-rate of 1.5 Mbps. The display rate was 30 Hz. Paris consists of two figures seated around a table in TV studio format, with high spatial coding complexity. The Baseline profile was selected, as this is more easily supported by mobile devices because of reduced computational overhead. The Intra-refresh rate was every 15 frames with IPPP… structure. 1000 frames were transmitted.

### 4.3 Assessing uplink queuing disciplines

WiMAX operating in point-to-multipoint mode was again simulated using the ns-2 simulator but this time through a WiMAX module from a Taiwanese research group (Tsai, 2006) that has proved an effective way of modeling WiMAX's behavior. The PHY settings selected for WiMAX simulation are given in Table 4, with the MAC settings defaulted from (Tsai et al., 2006). The DL/UL ratio is not intended to be realistic but to aid in testing the AQM algorithms, as in practice the DL would be allocated the majority of the bandwidth. The simulated configuration is shown in Figure 13. There were three SS communicating to the BS, with one of the SS sending VBR video encoded with the H.264/Advanced Video Codec (AVC). A trace file is input to ns-2 and packet losses recorded in the output. The output is employed to form the PSNR as a measure of video quality. Video quality comparisons were made under the EvalVid environment (Klaue et al., 2004).

| Parameter | Value |
|---|---|
| PHY | OFDMA |
| Duplexing mode | TDD |
| Frame length | 5 ms |
| Bandwidth | 6 MHz |
| FFT size | 1024 |
| DL/UL ratio | 1:3 |
| Fragmentation | yes |

Table 4. Simulated WiMAX settings.

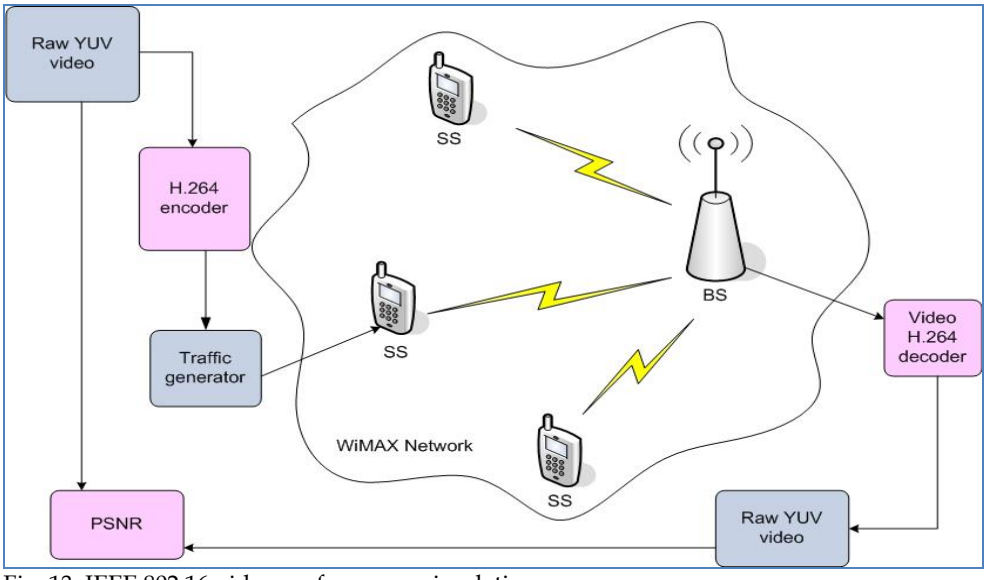



Fig. 13. IEEE 802.16 video performance simulation

Table 5 records the simulated traffic characteristics. As mentioned in Section 4.2, the video source was transported in two different ways: with simple UDP packetization and using

TFRC. Network Adaptation Layer Units (NALUs) from H.264 were encapsulated with Real Time Protocol (RTP) headers on a single slice per frame basis. RTP includes a frame creation time-stamp allowing end-to-end delay to be estimated (see Section 5). However, the packet trace created by EvalVid (Klaue et al., 2003) from the H.264 output ensured packets did not exceed 1000 B. This implies that there is some risk of decoder de-synchronization if a slice is split into several packets, causing slice de-synchronization markers to be missing from some packets. Under UDP transport, packets were transmitted with an inter-packet gap of 0.03 s. For TFRC, the H.264 packet sizes were the same as for UDP but the inter-packet gap was varied according to the TFRC equation. As described in (Handley et al., 2003), TFRC is a receiver-based system in which the packet loss rate is found at the receiver and fed-back to the sender in acknowledgment messages. The sender calculates the round-trip delay from the acknowledgment messages and updates the packet sending rate. An equation that models TCP New Reno's sending rate is employed to find the sending rate. In a variant to standard TFRC, the packet size in the TFRC equation was dynamically altered according to the EvalVid-created trace file sizes. The underlying TFRC protocol is also UDP.

| SS-UL | Service type | Traffic type | Protocol | Packet size (B) |
|---|---|---|---|---|
| 1 | rtPS | VBR (video) | UDP/ TFRC | Variable |
| | | CBR | UDP | 1000 |
| | nrtPS | FTP | TCP | |
| 2 | rtPS | CBR | UDP | 1000 |
| | nrtPS | FTP | TCP | |
| 3 | rtPS | CBR | UDP | 1000 |
| | nrtPS | FTP | TCP | |
| SS-DL | | | | |
| 1,2, 3 | rtPS | CBR | UDP | 1000 |
| | nrtPS | FTP | TCP | |

Table 5. Simulated traffic characteristics

Coexisting rtPS queue CBR sources were all sent at 1500 kbps, i.e. the same target rate as the video source. The inter-packet gap was also 0.03 s for the CBR traffic. The FTP applications were set up out of convenience as a way of occupying the nrtPS queues; otherwise a BE queue might be more appropriate. Likewise, the DL traffic is simply selected to fully occupy the DL link capacity.

The buffer size was set to fifty packets. This is not so large as to cause long queuing delays. Packets were dropped rather than explicitly marked at the queues, which will affect non-compliant applications, i.e. UDP, as well as compliant ones, i.e. TFRC. For RED, the dropping thresholds were set to 20% and 80% of the buffer size.

Tables 6 and 7 present mean video quality and end-to-end delay for UDP and TFRC transport. PSNR and delay were recorded on a per frame basis. From these Tables, it is apparent that TFRC definitely responds to packet drops at the rtPS queue and consequently improves the overall video quality. The mean delay for TFRC is also a little reduced compared to UDP transport, though the maximum delay is higher, due no doubt to TFRC managing its rate. REM's mean delay is less than the other queuing disciples suggesting a

more aggressive regime which is confirmed by the lower video quality whether under UDP or TFRC transport, implying more packet drops. For TFRC, drop-tail queuing is actually preferable to the AQM queuing, while there is some advantage if UDP transport were to be used. From the Tables it is apparent that there is considerable variation in video quality on a frame-by-frame basis. The standard deviation (s.d.) is highest for TFRC under REM. The minimum per-frame PSNR values are unacceptable for viewing but the mean quality can be described as good, especially if received on a mobile station. However, our results do not take account of channel error, which would certainly reduce quality or delay (if some form of error control were applied).

| Queue discipline | PSNR (dB) | | | End-to-end delay (s) | |
|---|---|---|---|---|---|
| | Mean | s.d. | Min. | Mean | Max. |
| Drop Tail | 26.7 | 8.2 | 9.8 | 0.035 | 0.17 |
| RED | 29.8 | 9.1 | 10.7 | 0.035 | 0.17 |
| REM | 24.2 | 8.9 | 101 | 0.032 | 0.16 |
| BLUE | 29.8 | 9.1 | 10.7 | 0.036 | 0.16 |

Table 6. Overall video quality and latency for UDP

| Queue discipline | PSNR (dB) | | | End-to-end delay (s) | |
|---|---|---|---|---|---|
| | Mean | s.d. | Min. | Mean | Max. |
| Drop Tail | 31.6 | 8.6 | 12.1 | 0.032 | 0.23 |
| RED | 31.4 | 8.8 | 12.2 | 0.033 | 0.24 |
| REM | 29.7 | 10.6 | 11.5 | 0.023 | 0.19 |
| BLUE | 31.4 | 8.8 | 12.2 | 0.033 | 0.24 |

Table 7. Overall video quality and latency for TFRC

The packet drop figures were also compared, refer to Figure 14. As the total number of video packets sent was 4530, the dropped video packet numbers are around 10% of the whole. 10% packet loss is normally taken as recoverable from in video streaming (Agboma & Liotta, 2008). The number of packets lost by the UDP stream was found to be greater than UDP for all queuing disciplines except REM. REM also balances its dropping rate between CBR sharing the rtPS queue and the video sources. However, REM's packet losses are relatively high. Other queuing disciplines have similar behaviors to each other, whether AQM or drop-tail.

Because encoded I-frames tend to be larger than encoded P-frames, after packetization to restrict the maximum packet size to 1000 B, I-frames tend to form longer bursts of packets. From Figure 15, the percentage of I-frame packets dropped is larger than the percentage of P-frame packets (though the total of these is larger because there are more P-frames). Clearly this is a problem for REM and employing TFRC tends to result in a greater percentage of I-frame packet drops. This suggests that a drop policy based on preserving I-frame packets would benefit TFRC transport and could be applied using either RED or BLUE.

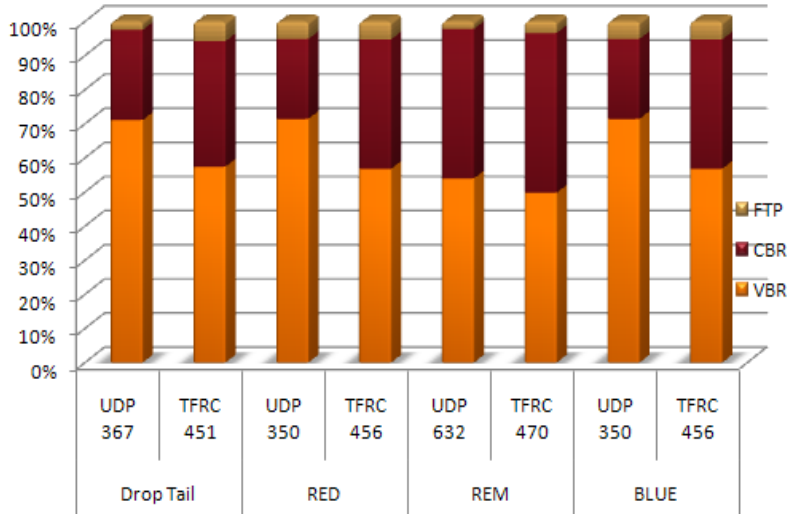



Fig. 14. Percentage packet drops according to traffic type and transport for differing queuing disciplines.

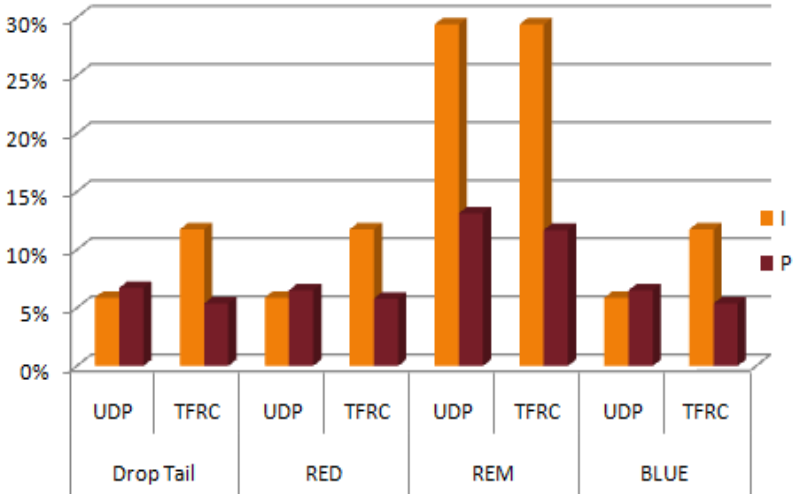



Fig. 15. Percentage of I- and P-frame packets dropped relative to numbers of I- and P-frame packets respectively

Apart from REM, drop-tail queuing and AQM for TFRC transport are equivalent in terms of the resulting mean video quality. Despite the detractions of using congestion controllers originally designed for the wired Internet, some form of congestion control is needed for WiMAX queues. REM is too aggressive, resulting in too many packet drops for 'bursty' traffic during periods of congestion. Because I-frame packets tend to arrive in bursts, there is a benefit from a dropping policy that favors I-frame packets. However, dropping predictive

P-frame packets also has an enduring impact and given that B-frame packets are not available for this purpose in the low-complexity Baseline profile, some other content dependent queue management is required.

## 5. H. 264 Packetization Issues

H.264/Advanced Video Codec (AVC) (Richardson, 2003) is a compression tool that is able to offer better compression ratios for low-bit rate video than prior codecs. The reader is referred to ( Wiegand et al, 2003), and other papers in the Special Issue on H.264 in which that paper appears for further information about H.264 itself, while in this Section we consider issues connected with packetisation.

Standardisation of H.264 was initiated by the Video Coding Experts Group (VCEG) which is a working group of the International Telecommunication Union (ITU-T). The Joint Video Team carried out the final work of developing the H.264, a co-operative effort between the Moving Picture Experts Group (MPEG) and VCEG in 2003. H.264 is divided into two parts: between a Video Coding Layer (VCL) and a Network Abstraction Layer (NAL). This allows VCL and NAL to be separately processed both when initially constructed and also when transported across a heterogeneous network. In the latter, the NAL could be reformed without any processing of the VCL content. For example, in wireline networks an Ethernet frame size of around 1500 B is typical of the Maximum Transport Unit (MTU) size, though this my vary across the public Internet but not within managed all-IP networks, whereas frame sizes nearer to 100 B are more typical of the MTU in wireless networks.

Briefly, the VCL represents the signal-processing functionality of the codec, with mechanisms motion compensation, frequency transform, loop filter and quantization. It is a macroblock-based coder that uses motion prediction between frames, with transform coding and motion compensation of the residual signal. Encoded macroblocks are collected into error resilience units called slices. A slice contains decoder resynchronisation markers and also indicates which control information is relevant to decoding the slice (parameter set). The output of the VCL consists of a bit-stream that containss the slice header (containing the spatial address of the first macroblock in the slice, the initial quantization parameter and similar information) and the macroblock data of an integer number of macroblocks. In H.264/AVC there are two types of slices: intra-frame coded and inter-frame coded. In the case of intra-frame coding, no previous or later frames are needed for encoding the present one, while in inter-frame coding previous and/or later frames are needed. Because H.264 allows prediction reference outside the boundary of a GOP, an Instantaneous Decoder Refresh (IDR) slice fulfills the role of an I-slice in previous codecs, in terms of a random access point for decoding, whereas an I-slice does not force the removal of prior references from the codecs' buffers.

VLC slices are encapsulated (Wenger et al., 2005) into Network Abstraction Layer Units (NALUs). Each NALU consists of a one-byte header and the payload bit string. The header refers to the type of the NAL unit, the (potential) appearance of bit errors or syntax violations in the NALU payload and other information related to the relative importance of the NALU for the decoding process. By identifying the importance of a NAL to error propagation avoidance, transmission software could judge the level of application-layer FEC to include. If transmission errors are possible (in a wireless network) then a `forbidden' bit within the NAL to indicate to a decoder to apply error correction procedures (as most

decoders contain some error detection and/or correction capability). An additional Real-Time Protocol (RTP) (Schulzrinne et al., 1996) header is assumed, which is why the NALU is referred to as the RTP-payload.

The RTP header is 12 B long and conatins a sequence number, an application-specific timestamp such as the sampling instance, application-specific payload type, a marker bit for signalling the end of related packets such as those emitted at the same time. The interpretation of application-specific data may be negotiated through contraol data at the start of a streaming session. FEC data can additionally be sent in media-unaware RTP packets. Though consideration is beyond the scope of this Chapter, there will normally also be an outer transport-layer UDP packet of 8 B, containing sequence number and a simple checksum. Congestion control for video streams is normally managed by an application-layer congestion controller, such as (refer to Section 4.2) TCP-Friendly Rate Control (TFRC) (Handley et al, 2003; Chen & Zakhor, 2006). TFRC controls the rate at which UDP packets are emitted through a feedback mechanism form the receiver. Increasingly, and certainly for all-IP managed networks, a 20 B IP header is also included to bring the total non-video header overhead to 40 B.

In the default case, one NALU contains a single coded slice. However, for very large slices such as might occur in high definition digital cinema, there is an option to sequence NAL fragments eah within its own RTP packet. Conversely, some NALUs are only a few bytes in size and may be aggregated within a single RTP packet, either all with the same timestamp or with differing timestamps, which must be coded relative to the RTP header's timestamp.

Addtionally it is possible to partition a slice. When using partitioning of data, a single slice of coded data can be divided in up to three parts, with each part put in a separate NALU. The encoder can be set so that each partiction does not exceed the MTU. The definition of the partitioning for a P- slice is as follows:

1- *A partition* has a slice header, quantization parameter, macroblock type, mode of prediction and motion vectors.
2- *B partition* contains the intra-coded macroblocks residual information.
3- *C partition* has the inter-frame coded macroblocks residual information.

Notice that some macroblocks in a pre-dominantly inter-coded slice may be predicted in an intra-coded (spatially predicted) manner in order to correct error propagation but as these are usually few it may be preferable to merge B- and C-partitions to avoid small NALUs. However, if a encoder is designed for wireless channels then more intra-coded macroblocks may be inserted and in which case separate B and C NALUs may be more appropriate (Stockhammer & Bystrom, 2004). The 3 partition arrangement differs from MPEG-4 and H-263, which only support two partitions. As I-slices in H.264 do not contain inter-frame coded blocks, C partitions do not occur, while B-slices have only partition.

The ability to use information from a correctly received partition when one of the other partitions is lost is the main idea of data partitioning. To be capable of doing this, it is important to be aware of the differing dependencies between the partitions. Each of partitions B and C cannot be decoded when partition A is lost, because without the motion vectors and prediction mode, the decoder cannot decode macroblocks. On the other hand, partition A information, especially the motion vectors, could be helpful to the decoder to compensate for the loss of either or both of the corresponding B and C partitions. We can conclude that partition A is independent of both partitions B and C, while the reverse is not correct. Therefore, the best achievable quality of decoded video arises if the proabability of

losing partition A is smaller than the probability of losing both partitions B and C. This is can be effected (Wenger, 2003; Stockhammer & Bystrom, 2004) by either using UEP or by sending partition A by a side channel.
In (Ferre et al., 2008), a packetization method was presented for robust H.264 video transmission over IEEE 802.11 wireless local area network (WLAN) configured as a home network. This research, though not applied to WiMAX does illustrate how careful handling of H.264's NALUs can improve video quality. Video robustness was enhanced by using small NALUs each of which was encapsulated (after addition of RTP header) into an IP packet. Then several NALU-bearing IP packets were mapped into a larger MAC frame. By retrieving error-free IP packets from within the received MAC frame greater robustness to error was achieved. The NAL aggregation mechanism with packet recovery was evaluated via simulation. The system provided a 2.5-dB gain in video quality for similar throughput and an 80% improvement in throughput efficiency was achieved for a similar video quality (PSNR) video performance. However, this method seems less efficient than the NAL aggregation method which requires a single IP header.
The research reported in (Chatterjee, 2007) used FEC and ARQ techniques to support streaming over WiMAX without violating the standard. The channel state information was employed to dynamically construct the MPDU. The sizes of these units were determined such that the packet dropping probability was minimized without compromising the goodput. Simulation results showed the performance enhancements achieved by the proposed ARQ-enabled adaptive algorithm for streaming data.
A Digital Cinema video distribution architecture with the motion JPEG codec is discussed in (Micanti et al., 2008). Though the issue of the need to avoid retransmission in the packetization strategy is discussed, this is on a general level and tests appear to only be for 802.11 rather than 802.16 as the title suggests. The scheme suggests careful mapping of the codestream structure onto the underlying transport packets.
Finally in (Stockhammer & Baystrom, 2004), the researchers applied data partitioning under H.264/AVC for conversational applications (e.g. videophone, videoconferencing) over mobile wireless channels. The experimental results showed that the probability of entirely lost frames was lowered and the probability of poor quality decoded video was reduced by the data-partitioning scheme. In this scheme, if FEC fails to protect an A partition then previous frame replacement is applied, while if a B- but not an A-partition is lost temporal error concealment is employed.

## 6. Conclusion

This Chapter has presented techniques to achieve successful video streaming over WiMAX networks. Because WiMAX queue management and scheduling are in the gift of the developer, it is possible to devise schemes that are either multimedia-friendly or do not discriminate against the requirements for video streaming. Because WiMAX supports OFDM, UEP through adaptive modulation is possible and, because there are many flexibilities in the WiMAX frame and MPDU, it is possible to may the video stream output onto these structures. The Chapter described research in adaptive modulation based on prioritising between frame types and uplink SS queue management. The necessary background for manipulation of H.264 NALUs, especially through data-partitioning was presented, though investigations are at an early stage. It is possible that all these techniques

can be combined and there are other possibilities supported by H.264 such as smooth switching between streams of different quality according to channel conditions. As in all applications of video streaming over wireless, WiMAX technology represents a rich field of investigation.

## 7. Acknowledgment

We gratefully thank B. Tanoh for conducting simulations for R. Razavi to verify the results in Section 3 of this Chapter.

## 8. References

Agnoma, F.; & Liotta, A. (2008) QoE Analysis of a Peer-to-Peer Television System. *In Proceedings of IADIS Int. Conf. on Telecomms, Networks and Systems*

Ahson, A. & Ilyas, M. (eds.) (2008) *WiMAX Applications*, Taylor & Francis Group, ISBN 978-1-4200-4547-5, Boca Raton, FL

Ali, N.A., Dhrona, P. & Hassanein, H. (2009) A Performance Study of Uplink Scheduling Algorithms in Point-to-Multipoint WiMAX Networks, *Computer Comms*., Vol. 32, No. 3, pp. 511-521

Anderson, H. R. (2003) Fixed Broadband Wireless System Design, Wiley & Sons, ISBN 0-470-84438-8, Chichester, UK

Andrews, J. G.; Ghosh, A. & Muhamed, R. (2007) *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*, Prentice Hall, ISBN 0-13-222552-2, Upper Saddle River, NJ

Athuraliya, S.; Li, V.H.; Low, S.H. & Yin, Q. (2001) REM: Active Queue Management. *IEEE Network*, Vol. 15, No. 2, (48-53)

Balkrishnan, H.; Padmanabhan, V.; Seshan, S. & Katz, R. (1997) A Comparison of Mechanisms for Improving TCP Performance over Wireless Links, *IEEE/ACM Trans. On Networking,* Vol. 5, No. 6, pp. 756-769

Chang, B.-J.; Chou C.-M.; & Liang, Y.-H. (2008) Markov Chain Analysis of Uplink Subframe in Polling-based WiMAX Networks, *Computer Comms*., Vol. 31, No. 10, pp. 2381-2390

Chatterjee, M.; Sengupta, S. & Ganauly, S. (2007) Feedback-Based Real-Time Streaming over WiMAX, *IEEE Wireless Commun*., Vol. 14, No. 1, pp. 64-71

Chen, C.-M.; Lin, C. –W. & Chen, Y.-C. (2006) Unequal Error Protection for Video Streaming Over Wireless LANs using Content-Aware Packet Retry Limit, *In Proceedings of IEEE Int. Conf. on Multimedia and Expo*, pp. 1961-1964

Chen, M. & Zakhor, A. (2006) Multiple TFRC Connections Based Rate Control for Wireless Networks. *IEEE Trans. On Multimedia*, Vol. 8, No. 5, pp. 1045-1061

Clark, D.D.; Shenker, S. & Zhang, L. (1992) Supporting Real-time Applications in an Integrated Services Packet Network: Architecture and Mechanism, *SIGCOMM '92*, pp. 14-26

Ercge, V.; Hari, K. V. S. et al., (2001) Channel Models for Fixed Wireless Applications, Contribution IEEE 802.16.3c-01/29rl Feb. 2001

Feng, W.-C.; Shin, K. G.; Dilip, D. K; & Saha, D. (2002) Active Queue Management Algorithms, *IEEE/ACM Trans. on Networking*, Vol. 10, No. 4, pp. 513-528

Ferre, P.; Doufexi, A.; Chung-How, J.; Nix, A. R. & Bull, D. R. (2008) Robust Video Transmission over Wireless LANs, *IEEE Trans. on Vehicular Technol.*, Vol. 57, No. 4, pp. 2596- 2602

Floyd, S. & Jacobson, V. (1993) Random Early Detection Gateways for Congestion Avoidance, *IEEE/ACM Trans. on Networking*, Vol. 1, No. 4, pp. 397–413

Ghanbari, M. (2003), *Standard Codecs: Image Compression to Advanced Video Coding*, IET Press, Stevenage, ISBN 0-85296-710-1, UK

Handley, M.; Pahdye, J.; Floyd, S. & Widmer, J. (2003) TCP-Friendly Rate Control (TFRC): Protocol Specification. RFC 3448.

Hanzo, L. & Choi, B.-J. (2007) Near-Instantaneously Adaptive HSDPA-Style OFDM Versus MC-CDMA Transceivers of WiFI, WiMAX, and Next-Generation Cellular Systems, *Proceedings of the IEEE*, Vol. 95, No. 12, pp. 2368-2392

Hillestad, O. J.; Perkis, A.; Genc, V.; Murphy, S. & Murphy, J. (2006) Delivery of On-Demand Video Services in Rural Areas via IEEE 802.16 Broadband Wireless Access Networks, *In Proceedings of 2nd ACM Workshop on Wireless Multimedia Networking and Performance Modeling*, pp. 43-51

Honig, M. K. & Messerschmitt, D. G. (1990) *Adaptive Filters Structures, Algorithms, and Applications*, Kluwer, ISBN 978-0-898-38163-4, Boston, MA

Hoymann, C. (2005) Analysis and Performance Evaluation of the OFDM-based Metropolitan Area Network IEEE 802.16, *Computer Networks*, Vol. 49, pp. 341-363

Juan, H. H.; Huang, H.-C.; Huang, C.-Y. & Chiang, T. (2007) Scalable Video Streaming over Mobile WiMAX, IEEE Int. Symposium on Circuits and Systems, pp. 3463-3466.

Klaue, J.; Rathke, B. & Wolisz, A. (2003) EvalVid - A Framework for Video Transmission and Quality Evaluation, *In Proceedings of Int. Conf. on Modeling Techniques and Tools for Computer Performance*, pp. 255-272

Koo, J.; Ahn, S.; & Chung, J. (2004) Performance Analysis of Active Queue Management Schemes for IP Network, *In Proceedings of Int. Conf. on Computational Science*, pp. 349-356

Li, Q. & Schaar, M. van der (2004) Providing QoS to Layered Video Over Wireless Local Area Networks Through Real-Time Retry Limit Adaptation, *IEEE Trans. on Multimedia*, Vol. 6, No. 2, pp. 278-290

Meloni, L. G. P. (2008) A New WiMAX profile for DTV Return Channel and Wireless Access., pp. 291-392, In Mobile WiMAX, Chen, K.-C. and de Marca, J. R. B. (eds.), Wiley & Sons, ISBN 978-0-470-51941-7, Chichester, UK

Micanti, P.; Baruffa, G. & Fabrizio Frescura, F. (2008) A Packetization Technique for D-Cinema Contents Multicasting over Metropolitan Wireless Networks, pp. 313-328, In Mobile WiMAX, Chen, K.-C. and de Marca, J. R. B. (eds.), Wiley & Sons, ISBN 978-0-470-51941-7, Chichester, UK

Milanovic, J.; Rimac-Drlje, S. & Bejuk, K. (2007) Comparison of Propagation Models Accuracy for WiMAX on 3.5 GHz, *In Proceedings of IEEE Int. Conf. on Electronics, Circuits and Systems*, pp. 1111-1114

Negi, R.; & Cioffi, J. (1998) Pilot Tone Selection for Channel Estimation in a Mobile OFDM System, *IEEE Trans. On Consumer Electron.*, Vol. 44, No. 3, pp. 1122-1128

Nuaymi, L. (2007) *WiMAX: Technology fo Broadband Wireless Access*, Wiley & Sons, ISBN 978-0-470-02808-7, Chichester, UK

Orlov, Z. & Necker, M. C. (2007) Enhancement of video streaming QoS with Active Buffer Management in Wireless Environments, *In Proceedings of 13th European Wireless Conf.*

Rahmani, R., Hjelm, M. and Aklund, C. (2008) Active Queue Management for TCP Friendly Rate Control Traffic in Heterogeneous Networks, *In Proceedings of Int. Conf. on Telecoms.*, pp. 1-7

Richardson, I. E. (2003) *H.264 and MPEG-4 Video Compression,* Wiley & Sons, ISBN 0-470-84837-5, Chichester, UK

Razavi, R., Fleury, M. & Ghanbari, M. (2008) Unequal Protection of Video Streaming through Adaptive Modulation with a Tri-Zone Buffer over Bluetooth EDR, *EURASIP J. on Wireless Comms. And Networking*, 16 pages, online vol.

Sadka, A. (2002) *Compressed Video Communications*, Wiley & Sons, ISBN 0-470-84312-8, Chichester, UK

Schaar, M. van der; Chou, P. A. (eds.) (2007) *Multimedia over IP and Wireless Networks*, Academic Press, ISBN 13-978-0-12-088480-3, Burlington, MA

Schulzrinne, H.; Casner, S.; Frederick, R. & Jacobson, V. (1996) RTP: A Transport Protocol for Real-Time Applications, RFC 1889, 1996

Stockhammer, T. & Baystrom, M. (2004) H.264/AVC Data Partitioning for Mobile Video Communication, *In Proceedings of Int. Conf. on Image Processing,* pp. 545-548

Tariq, U.; Jilani, U. N. & Siddiqui, T.A. (2007) Analysis on Fixed and Mobile WiMAX, MSc Thesis Report, Blekinge Institute of Technology, Sweden

Tsai, F. C.-D., et al., (2006) The Design and Implementation of WiMAX Module for ns-2 Simulator. *Workshop on ns2*, article no. 5

Wenger, S. (2003) H.264/AVC Over IP, *IEEE Trans. on Circuits and Syst. for Video Technol.*, Vol. 13, No. 7, pp. 645-656

Wenger, S.; Hannuksela, H. H.; Stockhammer, T.; Westerlund, M. & Singer, R. (2005) RTP Payload Format for H.264 Video, RFC 3984

Wiegand, T.; Sullivan, G. J.; Bjontegaard, G. & Luthra, A. (2003) Overview of the H.264/AVC Video Coding Standard, *IEEE Trans. on Circuits and Syst. for Video Technol.*, Vol. 13, No. 7, pp. 560-576

Yun, J. & Kavehard, M. (2006) PHY/MAC Cross- Layer Issues in Mobile WiMAX, Bechtel Telecomm. Techn. Journal, Vol. 4, No. 1, pp. 45-56

Zhang, H.; Li, Y.; Feng, S. & Wu, W. (2006) A New Extended rtPS Scheduling Mechanism Based on Multi-polling for VoIP Service in IEEE 802.16e System, *In Proceedings of Int. Conf. on Communication Technol.*, pp. 1-4

12

# Achieving Frequency Reuse 1 in WiMAX Networks with Beamforming

Masood Maqbool, Marceau Coupechoux and Philippe Godlewski
*TELECOM ParisTech & CNRS LTCI, Paris*
*France*

Véronique Capdevielle
*Alcatel-Lucent Bell Labs, Paris*
*France*

In this chapter, we examine the performance of adaptive beamforming in connection with three different subcarrier permutation schemes (PUSC, FUSC and AMC) in WiMAX cellular network with frequency reuse 1. Performance is evaluated in terms of radio quality parameters and system throughput. We show that organization of pilot subcarriers in PUSC Major groups has a pronounced effect on system performance while considering adaptive beamforming. Adaptive beamforming per PUSC group offers full resource utilization without need of coordination among base stations. Though FUSC is also a type of distributed subcarrier permutation, its performance in terms of outage probability is somewhat less than that of PUSC. We also show that because of lack of diversity, adjacent subcarrier permutation AMC has the least performance as far as outage probability is concerned. Results in this chapter are based on Monte Carlo simulations performed in downlink.

## 1. Introduction

Network bandwidth is a precious resource in wireless systems. As a consequence, reuse 1 is always cherished by wireless network operators. The advantage of reuse 1, availability of more bandwidth per cell, is jeopardized by increased interference because of extensive reutilization of spectrum. However, the emergence of new technologies like WiMAX, characterized by improved features such as advance antenna system (AAS), promises to overcome such problems.

Mobile WiMAX, a broadband wireless access (BWA) technology, is based on IEEE standard 802.16-2005. Orthogonal frequency division multiple access (OFDMA) is a distinctive characteristic of physical layer of 802.16e based systems. The underlying technology for OFDMA based systems is orthogonal frequency division multiplexing (OFDM).

In OFDM, available spectrum is split into a number of parallel orthogonal narrowband subcarriers. These subcarriers are grouped together to form subchannels. The distribution of subcarriers to subchannels is done using three major permutation methods called: partial usage of subchannels (PUSC), full usage of subchannels (FUSC) and adaptive modulation and coding (AMC). The subcarriers in a subchannel for first two methods are distributed throughout the available spectrum while these are contiguous in case of AMC. Resources of an OFDMA

system occupy place both in time (OFDM symbols) and frequency (subchannels) domains thus introducing both the time and frequency multiple access (Kulkarni et al., 2005).
Adaptive beamforming technique is a key feature of mobile WiMAX. It does not only enhance the desired directional signal but also its narrow beamwidth may reduce interference caused to the users in the neighboring cells. Resultant increase in signal to interference-plus-noise ratio (SINR) offers higher capacity and lower outage probability, which is defined as the probability that a user does not achieve minimum SINR level required to connect to a service. Adaptive beamforming can be used with PUSC, FUSC and AMC (refer Tab. 278 of IEEE standard 802.16-2005).
Network bandwidth is of high value for mobile network operators. It is always desired to get the maximum out of an available bandwidth by implementing frequency reuse 1 (network bandwidth being re-utilized in every sector see Fig. 1). However, with increased frequency reuse, radio quality of the users starts to deteriorate. Hence outage probability becomes more significant. To combat this problem, the conventional solution, in existing literature, is partial resource utilization or base station coordination to achieve frequency reuse 1.
Authors of (Porter et al., 2007) study the power gain, because of adaptive beamforming, of a IEEE 802.16e based system. Results presented by authors are based on measurements carried out in one sector of a cell with no consideration of interference. Measurements are carried out using an experimental adaptive beamforming system. Reference (Pabst et al., 2007) discusses the performance of WiMAX network using beamforming in conjunction with space division multiple access (SDMA). The simulations are carried out for OFDM (not OFDMA). Hence frequency diversity, because of distributed subcarrier permutations, is not taken into account. In (Necker, 2006) and (Necker, M. C., 2007), author has analyzed the performance of beamforming capable IEEE 802.16e systems with AMC. Unlike distributed subcarrier permutations (PUSC and FUSC), subcarriers in an AMC subchannel are contiguous on frequency scale. Hence PUSC/FUSC offer more frequency diversity as compared to AMC. Suggested interference coordination technique allows reuse 1 at the cost of reduced resource utilization. In (Maqbool et al., 2008a), we have carried out system level simulations for WiMAX networks. The analysis was focused on comparison of different frequency reuse patterns. Adaptive beamforming gain was also considered. We have shown that reuse 1 is possible with partial loading of subchannels.
In (Maqbool et al., 2008b), however, we have shown that by employing beamforming per PUSC group, the antenna-plus-array gain can be diversified and as a result reuse 1 is possible

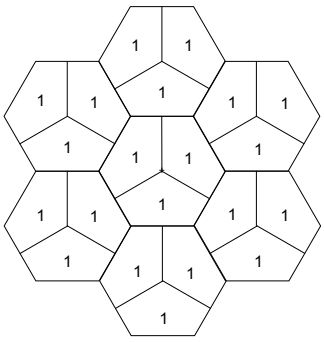



Fig. 1. Frequency Reuse Pattern 1x3x1.

without even partial loading of subchannels or base station coordination. In this chapter, we present results from (Maqbool et al., 2008b). We also extend those results by giving a comparison of system performance with all three subcarrier permutation types (PUSC, FUSC and AMC). The performance is analyzed in terms of cell throughput, $SINR_{eff}$ and probability of outage. Monte Carlo simulations are carried out in downlink (DL) for this purpose.
Rest of the chapter is organized as follows: section 2 gives an introductory account of subcarrier permutation types to be analyzed in this chapter. Possibility of beamforming with different subcarrier permutation types is discussed in section 3. SINR, beamforming, physical abstraction model MIC, modulation and coding scheme (MCS) and simulator details are introduced in section 4. Simulation results have been presented in section 5. Finally section 6 discusses the conclusion of this analysis.

## 2. Subcarrier Permutation Types

In this section, we present the salient features of subcarrier permutation with PUSC, FUSC and AMC in DL. In Tab. 1, values of various parameters for each permutation scheme are listed. These values correspond to 10 MHz bandwidth.
A detailed account can be found in (Maqbool et al., 2008c) where permutation method has been explained with the help of examples.

### 2.1 Partial Usage of Subchannels (PUSC)

One slot of PUSC DL is two OFDM symbols by one subchannel while one PUSC DL subchannel comprises 24 data subcarriers. Subchannels are built as follows:

1. The used subcarriers (data and pilots) are sequentially divided among a number of physical clusters such that each cluster carriers twelve data and two pilot subcarriers.
2. These physical clusters are permuted to form logical clusters using the renumbering formula on p. 530 in IEEE standard 802.16-2005. This process is called outer permutation. This permutation is characterized by a pseudo-random sequence and an offset called $DL_PermBase$.
3. Logical clusters are combined together in six groups called the Major Groups. The even groups possess more logical clusters as compared to odd Major Groups. Throughout this chapter, we shall refer these Major Groups as groups only.
4. The assignment of subcarriers to subchannels in a group is obtained by applying Eq. 111 of IEEE standard 802.16-2005. This process is known as inner permutation. The assignment in inner permutation is also controlled by $DL_PermBase$. Pilot subcarriers are specific to each group. Since number of logical clusters is different in even and odd groups, the number of their respective subchannels is also different.

### 2.2 Full Usage of Subchannels (FUSC)

The slot in FUSC mode is one OFDM symbol by one subchannel. Since slot in each permutation mode has same number of subcarriers, unlike in PUSC, the subchannel in FUSC comprises 48 data subcarriers. Subcarriers are assigned to subchannels in the following manner:

1. Before subcarriers are assigned to subchannels, pilot subcarriers are first identified (subcarrier positions for pilot subcarriers are given in section 8.4.6.1.2.2 of IEEE standard 802.16-2005) and are separated from others. These pilot subcarriers are common to all subchannels.

| Subcarrier Permutation | Parameter | Value |
|---|---|---|
| PUSC | No. of subchannels $N_{Sch}$ | 30 |
| | No. of subchannels per even group $N_e$ | 6 |
| | No. of subchannels per odd group $N_o$ | 4 |
| | No. of PUSC groups | 6 |
| | No. of total data subcarriers | 720 |
| | No. of total pilot subcarriers | 120 |
| | No. of available slots in DL (considering 30 OFDM symbols in DL) | 450 |
| FUSC | No. of subchannels $N_{Sch}$ | 16 |
| | No. of total data subcarriers | 768 |
| | No. of total pilot subcarriers | 82 |
| | No. of available slots in DL (considering 30 OFDM symbols in DL) | 480 |
| AMC | No. of subchannels $N_{Sch}$ | 48 |
| | No. of total data subcarriers | 768 |
| | No. of total pilot subcarriers | 96 |
| | No. of available slots in DL (considering 30 OFDM symbols in DL) | 480 |

Table 1. PUSC/FUSC/AMC parameters for 1024 FFT IEEE standard 802.16-2005.

2. In next step, the remaining subcarriers are divided among 48 groups.
3. Using Eq. 111 of IEEE standard 802.16-2005, a particular subcarrier is picked up from each group and is assigned to a subchannel. Similar to inner permutation of PUSC, this assignment is also controlled by $DL_PermBase$.

In PUSC and FUSC, by using different $DL_PermBase$ in network cells, subcarriers of a given subchannel are not identical in adjacent cells. In this case, it has been shown in (Ramadas & Jain, 2007) and (Lengoumbi et al., 2007), that the above process is equivalent to choosing subcarriers using uniform random distribution on the entire bandwidth in every cell. During our simulations, we consider the same assumption.

### 2.3 Adaptive Modulation and Coding (AMC)

In adjacent subcarrier permutation mode AMC, a slot is defined as $N_b$ bins $\times$ $M$ OFDM symbols, where ($N_b \times M = 6$). All available subcarriers (data+pilot) are sequentially grouped into bins. A bin is composed of nine contiguous subcarriers such that eight are data and one is pilot subcarrier. Though not exclusively specified in IEEE standards 802.16-2004 and 802.16-2005, but in consistent with nomenclature of PUSC and FUSC, we call ensemble the bins in a slot as subchannel. Out of possible combinations, we choose 2 bins $\times$ 3 OFDM symbols in our simulations.

## 3. Subcarrier Permutation and Beamforming

Pilot subcarriers are required for channel estimation. In case of beamforming, dedicated pilots are required for each beam in the cell. For PUSC and FUSC, there is a common set of pilot sub-

carriers for a number of subchannels while in AMC mode, each subchannel has its own pilot subcarriers. Hence, the number of possible orthogonal beams in a cell (of cellular network) depends upon the distribution of pilot subcarriers and hence the subcarrier permutation type. In PUSC, subchannels are put together in six groups. Each group has its own set of pilot subcarriers and hence, beamforming can be done per PUSC group. As subcarriers of a subchannel are chosen randomly, each subcarrier may experience the interference from different beams of a given interfering cell. In this way, subcarriers of a subchannel will not experience the same interference. The value of interference will dependent upon array-plus-antenna gain of the colliding subcarrier that may belong to any of six interfering beams in neighboring cell.
Pilot subcarriers in FUSC are common to all subchannels. Hence a single beam is possible in every cell. In contrast to PUSC, all subcarriers of a subchannel experience the same interference. This is due to the fact that every colliding subcarrier will have the same array-plus-antenna gain since there is only one beam per interfering cell.
When we consider AMC for beamforming, there can be as many orthogonal beams as the number of subchannels since every subchannel has its own pilot subcarriers. Due to similar assignment of subcarriers to subchannels in neighboring cells, all subcarriers will experience the same amount of interference because of an interfering beam in the neighbouring cell. Colliding subcarriers in a beam will have same array-plus-antenna. In addition, unlike PUSC and FUSC, since subcarriers of a subchannel are contiguous in AMC, no diversity gain is achieved.

## 4. Network and Interference Model

### 4.1 Subcarrier SINR

SINR of a subcarrier $n$ is computed by the following formula:

$$SINR_n = \frac{P_{n,Tx} a^{(0)}_{n,Sh} a^{(0)}_{n,FF} \frac{K}{d^{(0)\alpha}}}{N_0 W_{Sc} + \sum_{b=1}^{B} P_{n,Tx} a^{(b)}_{n,Sh} a^{(b)}_{n,FF} \frac{K}{d^{(b)\alpha}} \delta^{(b)}_n}, \quad (1)$$

where $P_{n,Tx}$ is the per subcarrier power, $a^{(0)}_{n,Sh}$ and $a^{(0)}_{n,FF}$ represent the shadowing (log-normal) and fast fading (Rician) factors for the signal received from serving BS respectively, $B$ is the number of interfering BS, $K$ is the path loss constant, $\alpha$ is the path loss exponent and $d^{(0)}$ is the distance between MS and serving BS. The terms with superscript $b$ are related to interfering BS. $W_{Sc}$ is the subcarrier frequency spacing, $N_0$ is the thermal noise density and $\delta^{(b)}_n$ is equal to 1 if interfering BS transmits on $n^{th}$ subcarrier and 0 otherwise.

### 4.2 Effective SINR

Slot is the basic resource unit in an IEEE 802.16 based system. We compute $SINR_{eff}$ over the subcarriers of a slot. The physical abstraction model used for this purpose is MIC (Ramadas & Jain, 2007) and is explained hereafter.
After calculating SINR of $n^{th}$ subcarrier, its spectral efficiency is computed using Shannon's formula:

$$C_n = \log_2 (1 + SINR_n)[bps/Hz],$$

MIC is computed by averaging spectral efficiencies of $N'$ subcarriers of a slot:

$$MIC = \frac{1}{N'} \sum_{n=1}^{N'} C_n[bps/Hz],$$

at the end $SINR_{eff}$ is obtained from MIC value using following equation:

$$SINR_{eff} = 2^{MIC} - 1.$$

For computation of $SINR_{eff}$, log-normal shadowing is drawn randomly for a slot and is same for all subcarriers of a slot. In presence of beamforming, it is essential to know the exact location of MS in the cell. For that purpose, line of sight (LOS) environment has been considered in simulations. Hence for fast fading, Rice distribution has been considered. Rician K-factor has been referred from (D.S. Baum et al., 2005) (scenario C1). Since in PUSC and FUSC, subcarriers of a subchannel (hence a slot) are not contiguous, fast fading is drawn independently for every subcarrier of a slot (Fig.2). On the other hand, the subcarriers in an AMC slot are contiguous and hence their fast fading factor can no longer be considered independent and a correlation factor of 0.5 has been considered in simulations. Coherence bandwidth is calculated by taking into account the powers and delays of six paths of vehicular-A profile with speed of MS equal to 60 Kmph (Tab. A.1.1 of (Ramadas & Jain, 2007)) and is found to be 1.12 MHz.

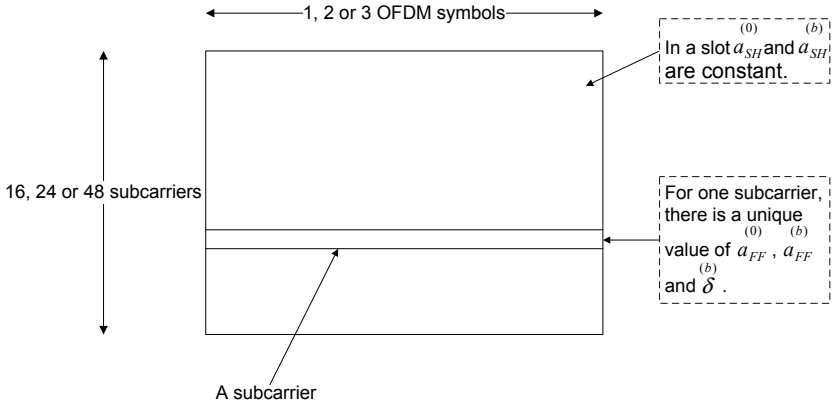



Fig. 2. Shadowing and fast fading over a PUSC/FUSC/AMC slot.

### 4.3 Beamforming Model

The beamforming model considered in our simulation is the delay and sum beamformer (or conventional beamformer) with uniform linear array (ULA). The power radiation pattern for a conventional beamformer is a product of array factor and radiation pattern of a single antenna. The array factor for this power radiation pattern is given as (Tse & Viswanath, 2006):

$$AF(\theta) = \frac{1}{n_t}\left|\frac{\sin(\frac{n_t\pi}{2}(\cos(\theta)-\cos(\phi)))}{\sin(\frac{\pi}{2}(\cos(\theta)-\cos(\phi)))}\right|^2, \tag{2}$$

where $n_t$ is the number of transmit antennas at BS (with inter-antenna spacing equal to half wavelength), $\phi$ is the look direction (towards which the beam is steered) and $\theta$ is any arbitrary direction. Both these angles are measured with respect to array axis at BS (see Fig.3).
The gain of single antenna associated with array factor is given by Eq.3 (Ramadas & Jain, 2007):

$$G(\psi) = G_{max} + \max\left[-12\left(\frac{\psi}{\psi_{3dB}}\right)^2, -G_{FB}\right], \tag{3}$$

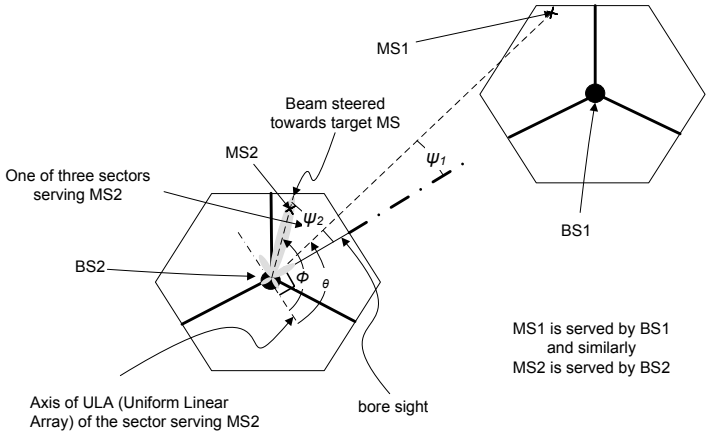



Fig. 3. Example showing beamforming scenario.

where $G_{max}$ is the maximum antenna gain in boresight direction, $\psi$ is the angle MS subtends with sector boresight such that $|\psi| \leq 180^{\circ}$, $\psi_{3dB}$ is the angle associated with half power beamwidth and $G_{FB}$ is the front-to-back power ratio.

### 4.4 Path Loss Model

Line-of-sight (LOS) path loss ($PL$) model for suburban macro (scenario C1) has been referred from (D.S. Baum et al., 2005). It is a three slope model described by the following expressions:

$$PL(d) = \begin{cases} free\ space\ model & \text{if } d \leq 20m; \\ C(f_c) + 23.8log_{10}(d) & \text{if } 20m < d \leq d_{BP}; \\ C(f_c) + 40log_{10}(d/d_{BP}) & \text{if } d > d_{BP}, \\ +23.8log_{10}(d_{BP}) & \end{cases}$$

where $f_c$ is the carrier frequency in Hz, $C(f_c)$ is the frequency factor given as: $33.2 + 20log_{10}(f_c/2 \cdot 10^9)$, $d_{BP}$ is the breakpoint distance and $\sigma_{Sh}$ is the standard deviation of log-normal shadowing. The breakpoint distance is computed as: $d_{BP} = 4h_{BS}h_{MS}/\lambda_c$, with $h_{BS}$ and $h_{MS}$ being the heights of BS and MS respectively. The value of $\sigma_{Sh}$ associated with above model is 4 $dB$ for $d \leq d_{BP}$ and is equal to 6 $dB$ beyond $d_{BP}$.

### 4.5 Modulation and Coding Scheme (MCS)

One of the important features of IEEE 802.16 based network is assignment of MCS type to a user depending upon its channel conditions. We have considered six different MCS types in our simulation model: QPSK-1/2 (the most robust), QPSK-3/4, 16QAM-1/2, 16QAM-3/4, 64QAM-2/3 and 64QAM-3/4 (for the best radio conditions). SINR threshold values for MCS types are given in Tab.2 and have been referred from *WiMAX Forum Mobile System Profile* (2007). If SINR of a mobile station (MS) is less than the threshold of the most robust MCS (i.e., less than 2.9 dB), it can neither receive nor transmit anything and is said to be in outage.

| **Index** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| **MCS** | QPSK 1/2 | QPSK 3/4 | 16QAM 1/2 | 16QAM 3/4 | 64QAM 2/3 | 64QAM 3/4 |
| $SINR_{eff}$ **[dB]** | 2.9 | 6.3 | 8.6 | 12.7 | 16.9 | 18 |

Table 2. Threshold of $SINR_{eff}$ values for six MCS types *WiMAX Forum Mobile System Profile* (2007).

### 4.6 Simulator Details

The frequency reuse pattern considered in simulations is 1x3x1 (Fig.1). The number of cells in the network is nineteen (i.e., eighteen interfering BS). To speed up the simulation process and to include the effect of an infinite network, wraparound technique has been employed. A significant number of snapshots are being carried out for Monte Carlo simulations. Locations of MS in a sector are drawn using uniform random distribution and beams are steered according to these locations. At BS, four transmitting antennas have been considered while MS is supposed to possess one receiving antenna. All simulations are carried out with full loading of subchannels.

As explained earlier, when PUSC is used, there can be up to six beams per sector i.e., one beam per group. For simulations with PUSC, we have considered three different cases with 1, 3 and 6 adaptive beams respectively. For the first case, all six PUSC groups are used by one beam. In the second case, each beam uses one odd and one even group. In the last case, each beam uses a distinct group. It is to be noted that number of channels per even and odd group are different (see Tab.1). To find the direction of adaptive beams, equivalent number of MS are drawn in a cell using spatial uniform distribution.

For the first case, one MS is drawn per sector and all subcarriers of a slot experience the same interfering beam pattern from a neighboring sector. On the other hand, in the second case, three MS are dropped in a sector and hence there are three interfering beams per sector. For each subcarrier used by a MS, the interfering beam is chosen with equal probability.

When there are six beams in a sector, the selection of interfering beam per subcarrier is no more equally probable. The reason being that beams are associated to even or odd groups and thus have different number of subchannels. Hence, for a subcarrier, the probability of interfering with an even beam is given as:

$$p_e = \frac{N_e}{N_{Sch}},$$

and with an odd beam it is:

$$p_o = \frac{N_o}{N_{Sch}}.$$

Considering a subcarrier, six MS are drawn per interfering sector. Respective beams are steered, three of them are odd and the others three are even. In a given interfering sector, the chosen beam is drawn according to the above discrete distribution.

In case of FUSC and AMC, one MS is drawn per sector and all subcarriers of a slot experience the same interfering beam pattern from a neighboring sector.

During every snapshot, $SINR_{eff}$ of a MS is calculated using MIC model. Cell space around BS is divided into twenty rings. Since MS is dropped using uniform random distribution,

| Parameter | Value |
|---|---|
| Carrier frequency $f_c$ | 2.5 GHz |
| BS rms tansmit power $P_{Tx}$ | 43 dBm |
| Subcarrier spacing $\triangle f$ | 10.9375 kHz |
| No. of DL OFDM Symbols $N_S$ | 30 |
| Thermal noise density $N_0$ | -174 dBm/Hz |
| One side of hexagonal cell $R$ | 1.5 Km |
| Height of BS $h_{BS}$ | 32 m |
| Height of MS $h_{MS}$ | 1.5 m |
| Antenna Gain (boresight) $G_{max}$ | 16 dBi |
| Front-to-back power ratio $G_{FB}$ | 25 dB |
| 3-dB beamwidth $\psi_{3dB}$ | 70° |
| No. of transmitting antennas per sector for beamforming $n_t$ | 4 |

Table 3. Parameters of simulations (Ramadas & Jain, 2007).

during a snapshot, it might be located in any of the twenty rings. $SINR_{eff}$ and throughput are averaged over each of these rings and over complete cell as well. The former is used to study the effect of change in the values of $SINR_{eff}$ and throughput w.r.t. distance from the BS. Throughput of a MS during a snapshot, depends upon the MCS used by it.
Simulation parameters are given in Tab.3. The parameter values are mainly based on (Ramadas & Jain, 2007).

## 5. Simulation Results

In this section we present the simulation results. Since PUSC has three possibilities for implementation of beamforming (cf. section 4.6), we first present results for three possible cases of PUSC. We compare these results with a case when beamforming is not considered. We call it without beamforming case. In addition, a scenario assuming beamforming only in the serving cell is also presented. Average $SINR_{eff}$ and average global throughput with respect to distance from BS are presented in Fig.4 and 5 respectively.
A clear difference can be observed between beamforming and without beamforming cases. We can observe about 7 to 8 dB gain. The gain for “beamforming in the serving cell only" scenario is about 2 dB less. The difference shows the effect of beamforming on interference reduction. The difference in terms of $SINR_{eff}$ and global throughput is not much with varying number of interfering beams.
However, it can be clearly seen in Fig.6 that outage probability significantly decreases when we take full advantage of diversity offered by PUSC. When increasing the number of beams, outage probability decreases from an unacceptable 9% (with one beam) to a reasonable 2% (with six beams). It is interesting to note that average throughput and $SINR_{eff}$ are not affected by the gain in outage probability. It can also be noticed that outage probability of “beamforming in the serving cell only" scenario is quite small. The reason being, the signal strength in the serving cell is increased because of beamforming while absence of beamforming in interfereing cells keeps the interference strength unchanged.

Next we compare the results of three subcarrier permutation types. In this comparison, PUSC has been considered with six interfering beams. In Fig. 7, average values of effective SINR ($SINR_{eff}$) are plotted as a function of distance from base station (BS). As can be noticed, there is almost no difference between values of $SINR_{eff}$ with PUSC, FUSC and AMC. On the other hand, when we look at MCS probabilities in Fig. 9, PUSC outclasses the other two (FUSC and AMC) in terms of outage probabilities. Though average $SINR_{eff}$ are same for all, only PUSC offers an outage probability in an acceptable range (less than 5%). Since subcarriers in a PUSC subchannel experience variable interference gains, it average outs the possibility of all subcarriers suffering from same and high interference. That is why outage probability is reduced. At the same time, it also reduces the probability that all coliding subcarriers have low power. This effect can be noticed while looking at probabilities of high rate MCS. For example, with PUSC, probability to transmit with 64QAM-3/4 is less as compared to FUSC and AMC.

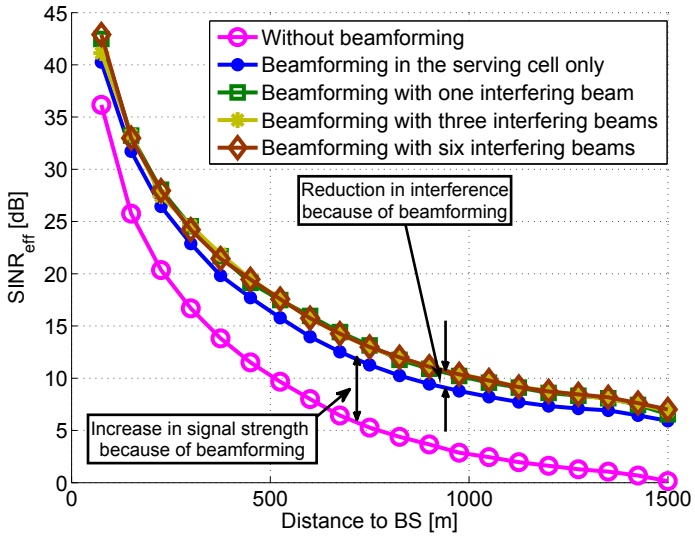



Fig. 4. Average $SINR_{eff}$ versus distance to base station for PUSC.

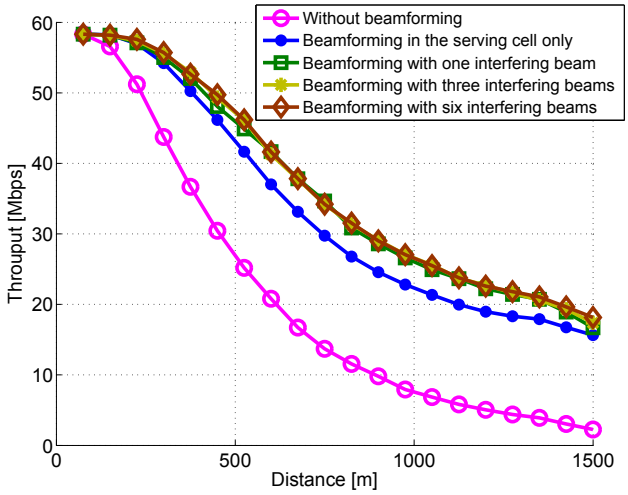



Fig. 5. Average cell throughput versus distance to base station for PUSC.

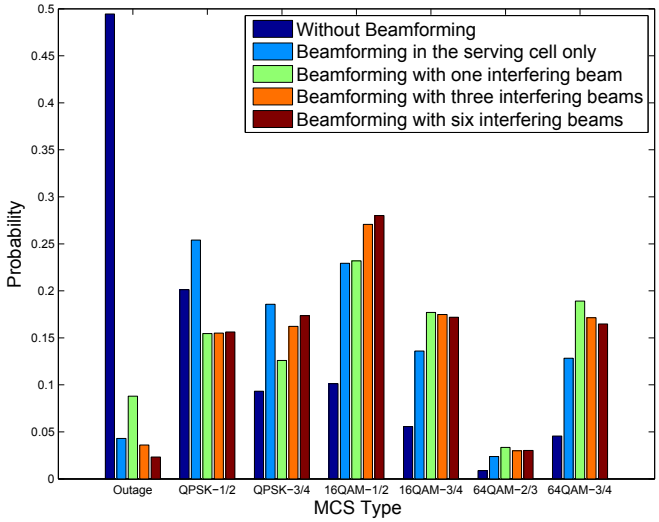



Fig. 6. MCS distribution for PUSC.

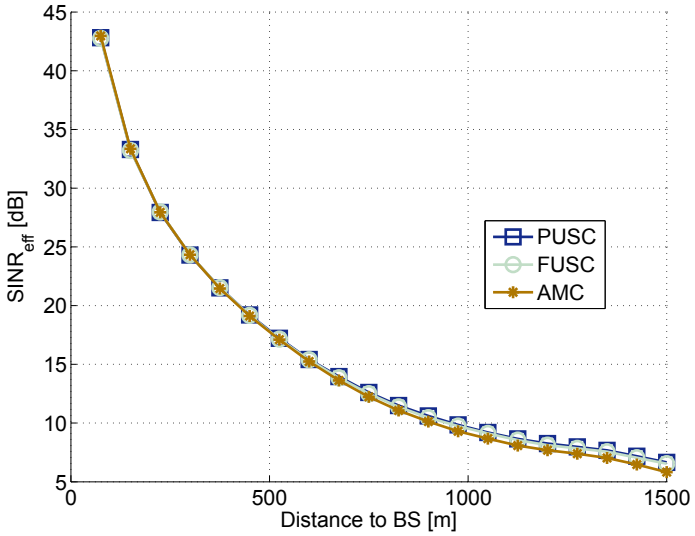



Fig. 7. Average $SINR_{eff}$ versus distance to base station for PUSC/FUSC/AMC with beamforming.

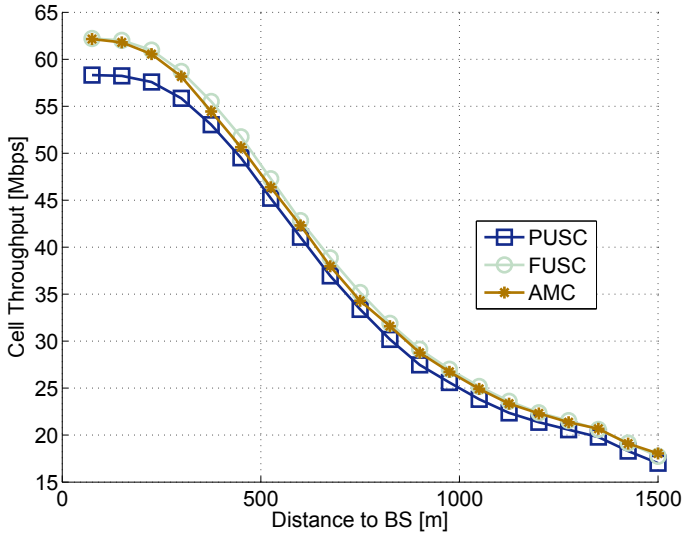



Fig. 8. Average cell throughput versus distance to base station for PUSC/FUSC/AMC with beamforming.

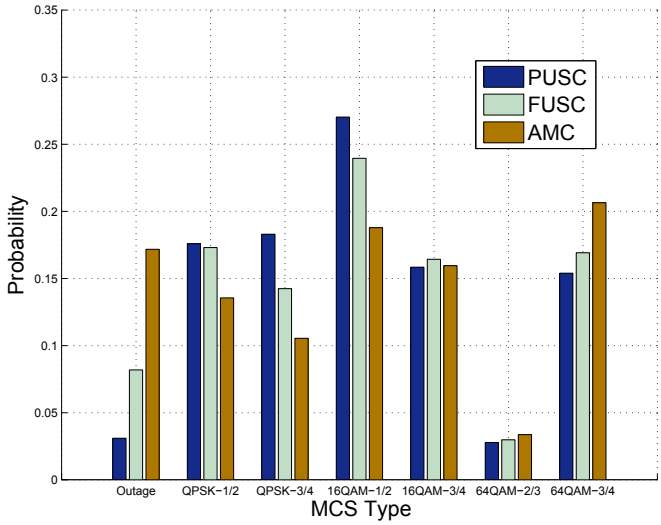



Fig. 9. MCS distribution for PUSC/FUSC/AMC with beamforming.

If we look at average values of cell throughput (w.r.t. distance from BS) in Fig. 8, it can be noticed that in the region close to BS, PUSC is somewhat less performing than FUSC and AMC. This result can be justified in light of probabilities of MCS in Fig. 9 where stationary probabilities of the best MCS (64QAM-3/4) are higher with FUSC and AMC. Owing to strong signal strength in the region close to base station, probability for a MS to achieve better MCS is more. At about 350 m and onward (from base station), throughput with PUSC is around 1 Mbps less than that of FUSC and AMC even if PUSC has better performance in terms of radio quality. This is because of the fact that with PUSC, number of available slots are lesser (see Tab. 1).

## 6. Conclusion

Currently, WiMAX networks are going through trial and deployment phase. Therefore, it is important at this stage to analyze various features of WiMAX. In this chapter, we have studied the possibility of adaptive beamforming in connection with three subcarrier permutation types of WiMAX. We have shown that beamforming per PUSC group offers a low outage probability as compared to FUSC and AMC. FUSC and AMC have more number of data subcarriers and hence the resultant throughput with the two is slightly more than that of PUSC. At the same time, outage probabilities for FUSC and AMC are more than 5%. Hence, adaptive beamforming per PUSC group can be exploited to achieve acceptable radio quality without need of partial loading of subchannels or base station coordination.

## 7. References

D.S. Baum et al. (2005). IST-2003-507581, D5.4 Final Report on Link and System Level Channel Models. WINNER.
**URL:** *https://www.ist-winner.org/DeliverableDocuments/D5.4.pdf*

Kulkarni, G., Adlakha, S. & Srivastava, M. (2005). Subcarrier Allocation and Bit Loading Algorithms for OFDMA-Based Wireless Networks, *IEEE Trans. on Mobile Computing, Vol. 4*.

Lengoumbi, C., Godlewski, P. & Martins, P. (2007). Subchannelization Performance for the Downlink of a Multi-Cell OFDMA System, *Proc. of IEEE WiMob*.

Maqbool, M., Coupechoux, M. & Godlewski, P. (2008a). Comparison of Various Frequency Reuse Patterns for WiMAX Networks with Adaptive Beamforming, *Proc. of IEEE VTC Spring*.

Maqbool, M., Coupechoux, M. & Godlewski, P. (2008b). Effect of Distributed Subcarrier Permutation on Adaptive Beamforming in WiMAX Networks, *Proc. of IEEE VTC Fall*.

Maqbool, M., Coupechoux, M. & Godlewski, P. (2008c). Subcarrier Permutation Types in IEEE 802.16e, *Technical report*, TELECOM ParisTech.

Necker, M. C. (2006). Towards Frequency Reuse 1 Cellular FDM/TDM Systems, *ACM MSWiM*.

Necker, M. C. (2007). Coordinated Fractional Frequency Reuse, *ACM MSWiM*.

Pabst, R., Ellenbeck, J., Schinnenburg, M. & Hoymann, C. (2007). System Level Performance of Cellular WiMAX IEEE 802.16 with SDMA-enhanced Medium Access, *Proc. of IEEE WCNC*, pp. 1820–1825.

Porter, J. W., Kepler, J. F., Krauss, T. P., Vook, F. W., Blankenship, T. K., Desai, V., Schooler, A. & Thomas, J. (2007). An Experimental Adaptive Beamforming System for the IEEE 802.16e-2005 OFDMA Downlink, *Proc. of IEEE Radio and Wireless Symposium*.

Ramadas, K. & Jain, R. (2007). WiMAX System Evaluation Methodology, *Technical report*, Wimax Forum.

Tse, D. & Viswanath, P. (2006). *Fundamentals of Wireless Communications*, Cambridge University Press.

*WiMAX Forum Mobile System Profile* (2007).

13

# WiMax Interference and Coexistence Studies with Other Radio Systems

A. Rahim Biswas and A. Finger
*Technical University Dresden*
*Germany*

## 1. Introduction

The first version of WiMax standard was focused merely on the fixed subscriber applications and the next generation mobile WiMax technologies are designed enabling both fixed and mobile subscribers with various form factors like: personal digital assistance, notebook PCs, handsets, and consumer electronics (WiMaxForum, 2005). The air interface of mobile standard is licensed for using 2.3 GHz, 2.5 GHz, 3.3 GHz and 3.5 GHz frequency bands (IEEE, 2005). However, the allocation of these frequency bands are varying from one country to another or one geographical region to another for example the band 3.3 GHz and 3.5 GHz are reserved, for fixed and mobile version, in Europe as well as some other parts in the world except USA. Hence, the WiMax interference and coexistence scenario in those region will not be identical like in Europe where WiMax will necessary to coexist with Ultra-Wide Band (UWB), since in Feb 2002 FCC (FCC, 2002) approved an unlicensed frequency band from 3.1 to 10.6 GHz with maximum transmission power of -41.3 dBm/MHz for short range, high and low data rate applications on notebook PCs, headset and consumer electronics. Therefore, the usage models of both WiMax and UWB systems will be in very close proximity to desktop PC/laptop, handset, etc. Consequently, the interference from UWB radio can raise the Noise Floor (NF) by an amount sufficient to cause performance degradation at the WiMax subscriber receiver (Rahim, A. & Zeisberg, S. 2007), (Rahim et. at 2007). Therefore, the regulation authorities in Europe (ECC, 2006), Japan, Korea, etc had to define the UWB mask approximately 35~40 dB lower than the FCC limits over 3.1 to 3.8 GHz without applying mitigation technique. In addition, a mitigation technique called Detect and Avoid (DAA) have been well studied to increase the UWB power spectral density while protecting WiMax services; this mitigation technique is based on the detection of the WiMax signal and once the signal is detected then UWB devices require to give sufficient protection to the WiMax services (Mubaraq, 2007), (Rahim et. at 2008), (Kim et. Al 2007). The coexistence studies among different systems are referred to (Indepen & Quotient, 2005), (Snow et. al, 2007), (Nader et. al 2007).

In line with the above descriptions, in this chapter the interference and coexistence issues between WiMax and UWB will be discussed. First, the WiMax receiver performance parameters like the noise floor, sensitivity level, antenna effect will be characterized and the maximum interference limits from UWB to WiMax receiver will be estimated. Depending on

the required protection level at the WiMax receiver, the maximum interference zone that a UWB can cause harmful interference to WiMax system will be identified. Since the interference from UWB devices may appear as an increase of the noise floor and sensitivity level of the WiMax terminal receiver, it can cause an impact on the WiMax system performance. Therefore, the UWB interference impact on WiMax cell coverage and an outage of active users will be studied considering different acceptable nose raise level at the WiMax receiver.

## 2. WiMax Receiver Characteristics and its Protection Limits

At the early stage of the coexistence studies, the victim receiver characteristics (e.g. noise floor, receiver sensitivity, bandwidth, acceptable interference criteria, etc) are necessary to discuss, since the interference impact from incumbent UWB devices entirely depends on those parameters. In generally, the receiver is designed by means of current market demands in developing a product being low cost, small size and low power consumption while compromising the system performance. A system contains two different kind of receivers namely Base Station Receiver (BSR) and Subscriber Station Receiver (SSR); the target quality and performance is fairly better in BSR compared to the SSR. It is true and well known for any product "Low cost trade-off low quality". In order to achieve such low cost several compromises are made on the receiver design, especially in the adjustment of high Noise Floor (NoF) as well as compensation of some losses. The design of low power and low size receiver is also strongly influenced by several factors like radio architecture, integrated circuit process technology, etc. They may encounter additional losses in the devices. Moreover the NF and losses within the SSR can not be replicated with the BSR.

### 2.1 Characterization of WiMax Noise Density

The Noise Floor (NoF) is the integration of noises, losses and errors present over the effective operating bandwidth at the WiMax receiver. Thermal Noise (TN) is the main part of the NoF and remaining noises like noise figure (LNA and mixer noises), synthesizer noise and several errors such as estimation error, tracking error etc has also a significant contribution in the NoF estimation. We define the Noise Density (ND) when the NoF is normalized with 1 MHz of channel bandwidth and given by

$$ND\,\big|_{dBm\ /\ MHz} = TN\,\big|_{dBm\ /\ MHz} + NF\,\big|_{dB} + IL\,\big|_{dB} \tag{1}$$

TN is computed base on, the kTBw formula, Boltzmann's constant ($k = 1.38 \times 10^{-23}$), the noise temperature T (in degree Kelvin $0^{o}C+273$), and the effective channel bandwidth of the receiver, Bw (in MHz). For example, considering the temperature of T=290K results in a TN of -114 dBm/MHz (-174 dBm/Hz normalized with 1 Hz), which is in fact adopted in most applications. IL is the implementation loss including the non-ideal receiver effects such as channel estimation errors, tracking errors, quantization errors, and phase noise.

NF and IL are the critical parameters of the receiver performance in presence or in absence of any interference because the normalized TN is fairly constant. The increase of NF or IL obviously reduces the receiver activities which results in a higher level received signal requirement. In critical situations, a fraction of dB can make a huge difference of

interference impact to the receiver from incumbent systems and can play a big part in winning or losing contracts. Then predicting accurate NF and IL is of paramount importance. The (IEEE, 2004) standard limited these values with 7 dB of NF and 5 dB of IL but some vendors can reduce the values by means of some sophisticated techniques. The standard does not separate the BSR and SSR in terms of NF or IL, but in realistic applications both receivers may not be replicated. In addition, the worst case values are always recommended for performance evaluation. Hence, we have considered the NF of 5 dB and IL of 0 dB at the SSR. In this paper, we have investigated only the impact on the SSR but it is also possible to consider the NF of 3 dB and IL of 0 dB for BSR in further investigations.

### 2.2 WiMax OFDM Receiver Sensitivity

The sensitivity, $R_{Sen}$ level of a receiver is an equivalent to the NoF plus the minimum acceptable Signal to Noise Ratio (SNR) at the receiver. The type of Modulation and Coding (MC) used in the system usually determines the SNR if the standard Bit Error Rate (BER) is met. The (IEEE, 2004) standard derives the $R_{Sen}$ based on the SNR of an ideal receiver required to achieve a BER of $10^{-6}$ in Additive White Gaussian Noise (AWGN). But some other wireless standards may define the reduced BER of $10^{-6}$ in the system for the purpose to achieve the required system demands by accomplished better quality, reliability and robustness. One possibility to do this is to select a low data rate modulation and coding in the system which gives a lower SNR and robust link. WiMax systems use adaptive modulation and coding scheme and the corresponding SNR can be found in (IEEE, 2004) but this value has been corrected in (C802, 2005) and a new SNR ratio was proposed which is approximately 3.4 dB lower than the previous one.

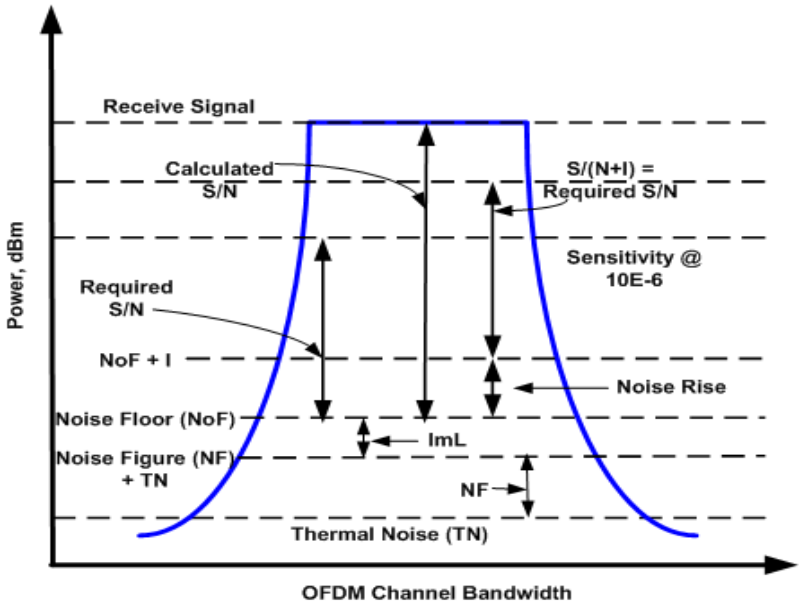



Fig. 1. Relationship between WiMax receiver parameters

The receiver $R_{Sen}$ in dBm is given by

$$R_{Sen} = ND + SNR + 10 \log_{10}(Bw) \quad (2)$$

Where Bw is the effective channel bandwidth of the receiver in MHz. OFDM scheme in IEEE 802.16 standard does not allocate the entire channel bandwidth for information transmission. There are some significant factors like FFT, sampling, DC and guard band, etc which partially reduces the effective channel bandwidth and it can be straightforwardly computed as follows (IEEE, 2004).

$$Bw = F_s N_{used} / RN_{FFT} \quad (3)$$

Finally, substituting (3) into (2), the receiver $R_{Sen}$ for OFDM physical layer is give by

$$R_{Sen} = ND + SNR + 10 \log_{10}(F_s N_{used} / RN_{FFT}) \quad (4)$$

Whereas $F_s$ is the sampling frequency, $N_{used}$ is the number of subcarriers used, and $N_{FFT}$ is the length of FFT in OFDM physical and R is the repetition rate. Figure 1 depicts the graphical relationship among TN, NoF and $R_{sen}$.

### 2.3 WiMax Antenna Characterizations

The interference influences are also modelled and characterized by the antenna mounted in the victim receiver. The antenna parameters (e.g directivity, gain, radiation patterns, height, etc.) can lead to a significant interference mismatch on the victim receiver. The interference impact on MSR and SSR will differ in terms of these factors because the BSR in the WiMax system contains high gain directional antennas but the SSR may contain directional (FWA subscribers) or omni-direction (nomadic/mobility subscribers) antenna depending on the services. For an equal separation distance between fixed or nomadic victim SSR and UWB interferer, the fixed directional receiver will experience greater impact than nomadic one, if both antennas (UWB and SSR) are located in the boresight. Therefore, in order to investigate the impact on the high gain directional receiver, the antenna parameters need to be taken into account carefully. However, we have limited the evaluation within indoor nomadic SSR, i.e. the interference signal received by an omni-directional antenna is with 0 dB gain. In addition, the performances of BSR and SSR can not be replicated against interference because of the robust antenna used in the BSR. In generally, BSR gives better performance but it can also involve a feeder loss that is not being seen in SSR. Therefore the selection of antennas for SSRs is a mature of science and keeping the following characteristics in mind: i) small size ii) low cost iii) good gain characteristics and iv) a good omni-directional radiation patterns.

### 2.4 Estimation of Maximum Interference Limits

Maximum permissible noise raise at the WiMax receiver from a UWB transmitter is an important consideration. Starting from this limit the maximum allowable aggregate interference effects from that UWB terminal are computed. They strongly depend on the location of the WiMax terminal in the cell. In general, the maximum permissible noise limit

is a function of the distance between the WiMax client and the Base Station (BS). Therefore, the permissible level varies from one client to another client and those are located far from the cell edge can be operated with higher acceptable level of noise raise than clients close to the BS because of stronger desired received signal. Theoretically, the interference impact on the victim receiver is negligible when the received interference power is much below the NoF. For the terminals located near cell edge, the maximum interference power can be equal to the NoF, which corresponds to 3 dB noise raise limit at the victim receiver (Noise Floor, NoF (mW) + maximum UWB interference power, $I_{UWB}$ = NoF(mW) + NoF(mW) = 2NoF(mW) = NoF(dBm) + 3 dB. In (Giuliano, et. at 2005), UWB interference is considered to be harmful when the noise raise is more than 3 dB. However, we limited our interference analyse for not only 3 dB; it is assumed that the terminals can accept noise raise of 1-3 dB. For the generic WiMax victim receiver, the maximum permissible interference due to UWB interference can be accounted as follows (Sarfaraz, et. al, 2005):

$$I_{UWB} = ND + 10\log_{10}(10^{N_r/10} - 1) \quad (5)$$

Here $N_r$ is the maximum allowable noise raise in the WiMax client receiver in dB.

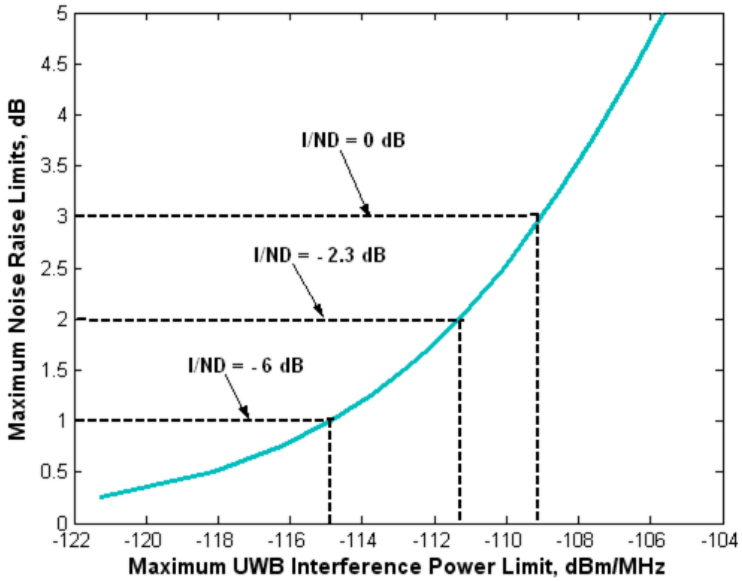



Fig. 2. Noise raise vs. UWB permissible interference power

Figure 2 shows the maximum allowable interference input power corresponding to the permissible noise raise at the WiMax client. The results are computed by (1) with applied the NF of 5 dB and IL of 0 dB. One can see that at $I_{UWB}$/ND of -6 dB the corresponding noise raise limit is 1 dB. In similar way, when the $I_{UWB}$/ND is -2.35 dB the noise limit is 2 dB and at $I_{UWB}$/ND of 0 dB the maximum noise raise is 3 dB. Finally, at $I_{UWB}$/ND of 0 dB and the NF of 5 dB the maximum interference limit of -109 dBm/MHz is computed. The results are summarized in the table 1.

| Noise Raise | 1 dB | 2 dB | 3 dB |
|---|---|---|---|
| $I_{UWB} / ND$ | - 6 dB | - 2.35 dB | 0 dB |
| $I_{UWB}$ | - 115 dBm/MHz | -111.35 dBm/MHz | -109 dBm/MHz |

Table 1. Maximum UWB permissible interference power at SSR

## 3. SEMCAT Analysis of Maximum Possible UWB PSD at 3.5 GHz band

The analytical and simulation results for only one interfering signal by a UWB system have been taken into account to evaluate the impact of interference to a WiMax receiver. In the analytical part, we examined the results in the following order: assess a realistic interference scenario to find out the true separation distance between UWB transmitter and WiMax receiver, to determine the minimum coupling loss between the interferer and victim by using the minimum permissible interference power input, to translate the minimum coupling loss into a minimum interference range for a single UWB interferer by means of a free space propagation model, and to calculate the probability of UWB transmitter be located inside the interference range. In the simulation part, the Monte-Carlo based SEAMCAT® tool is used to implement the interference scenario and to evaluate the cumulative probability of interference. The most important parameters used for subsequent analysis are indicated in Table 2.

| Parameters | Values |
|---|---|
| UWB Tx Power | -65 dBm/MHz to -85 dBm/MHz |
| UWB Tx Gain | 0 dB |
| UWB RF Loss | I dB |
| Channel Model | Free Space |
| Frequency | 3.5 GHz |
| Channel BW | 500 MHz (UWB) |
| Separation | 0.35 m to 2 m |
| Victim Rx Gain | 0 dBm |
| Channel BW | 10 MHz (WiMax) |
| Noise Figure | 5 dB to 7 dB |
| Imp: Loss | |

Table 2. SEMCAT simulation parameters

### 3.1 Coexistence Scenario in SEMCAT

The scenario addressed the case when the UWB transmitter is "non co-located" with WiMax system (see Figure 3). Non collocated means UWB device is not included within the same hardware platform as the WiMax terminal. Hence, the scenario is considered where UWB will be operating in close vicinity to a WiMax terminal connected to the PC (TG3, 2006). Air interference exists between both UWB transmitter and WiMax terminal.

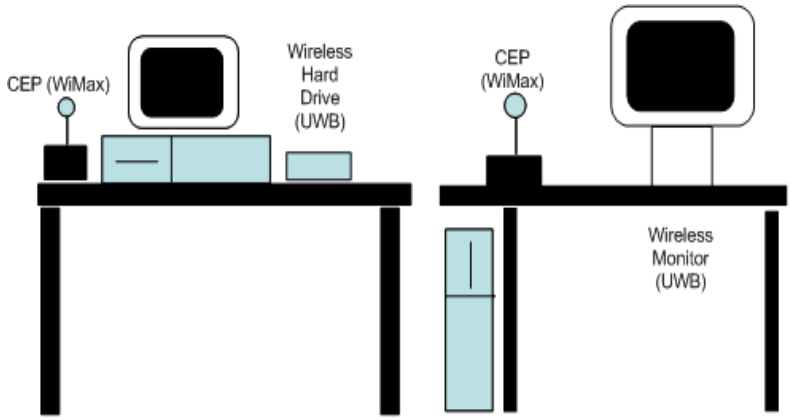



Fig. 3. Coexistence scenarios in the office desktop environment

The main idea behind the interference scenario analysis is to find out a realistic distance between the UWB transmitter and the victim receiver. We consider an office desk scenario where the UWB device is located together with a WiMax system on the same desk or in very close proximity. The separation distances between the UWB and WiMax receiver in the office desk are depending on the desk arrangement and desk size. Two types of layout are often observed; a long main desk and main desk with a small desk placed at the one corner of main desk. The small desk is mainly used to place PC, PDA, etc. The main desk will generally be 1.8 m to 2 m long. A WiMax terminal will be placed centrally or opposite corner of the desk, distance from the UWB to a WiMax terminal would be 0.5 m to 2 m (TG3, 2006). Alternative case, a WiMax terminal will often be placed at one of the back corners of the main desk and a wireless monitor placed centrally on the main desk, distances from the monitor to a WiMax terminal would be 0.35 m to 0.5 m. The UWB devices which are associated with a PC (such printers, remote hard drives, etc.) will be normally placed between 0.5 m to 2m. Therefore, the minimum separation between UWB and WiMax is assumed to be 0.35 m to 2 m.

### 3.2 Minimum Coupling Loss between UWB Tx and WiMax Rx

The Minimum Coupling Loss (MCL) characterizes the minimum protection distance required between the UWB transmitter and WiMax client to ensure that there is no such interference coming from interferer. In other words, it determines the interference zone surrounding of victim receiver and received interference signal if UWB transmitter appears in the zone. The MCL can be derived as follows:

$$MCL = P_t^{UWB} + G_t^{UWB} - L_{RF}^{UWB} + G_r^{WTS} - I \tag{6}$$

Where: $P_t^{UWB}$- UWB transmitter conducted power in dBm/MHz, $G_t^{UWB}$ - UWB transmitter antenna gain in dB, $L_{RF}^{UWB}$ - UWB transmitter RF Loss in dB, $G_r^{WTS}$- WiMax client antenna gain in dB, I - maximum permissible interference input at WiMax terminal in dBm/MHz. The above parameters are defined in the Table 2, except I. We assume RF loss in the UWB transmitter because it is commercialized as a low cost product. In order to achieve an

aggressive low cost goal several compromises are made particularly on fundamental receiver and transmitter parameters, which normally resulting in RF loss and high noise figure.

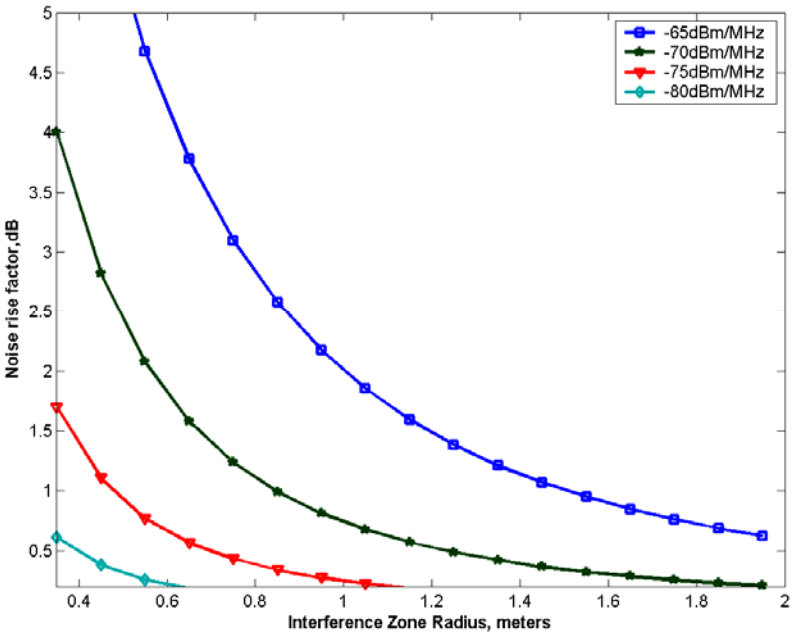



Fig. 4. Maximum interference zone radius vs. maximum $N_r$

### 3.3 Interference Zone Radius Center at WiMax Receiver

If we assume that the victim receiver has a certain noise figure and is within the line of sight of the UWB transmitter such that the free space path loss equation applied, then we can calculate the distance (interference zone radius, $r_i$ ) at which the received power will equal to the permissible interference input of the incumbent receiver, for a given UWB transmitter power.

$$PL\ (dB\ ) = 20\ \log\ 10\ (\frac{4\pi}{c}) + 20\ \log\ 10\ (\frac{f}{10^{-6}}) + 20\ \log\ 10\ (r_i) \tag{7}$$

Here, f is the center frequency in MHz, c is the light speed $3x10^{-8}$ m/s and $r_i$ is the interference zone radius in meter.

Figure 4 depicts the radius of interference zone which is computed for noise figure of 5 dB and implementation loss of 2 dB. It is represented that the WiMax client receiver will not see the UWB interferer at any distance beyond 0.43 m, even assuming free space propagation when transmit power of -70 dBm/MHz and noise raise of 3 dB performed. This distance is independent of receiver bandwidth as long as the UWB system has a flat power spectral density across the incumbent's bandwidth.

| $P_t$ | Nr = 3 dB | | | Nr = 2 dB | | | Nr = 1 dB | | |
|---|---|---|---|---|---|---|---|---|---|
| dBm/ MHz | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB |
| -65 | 0.76 m | 0.68 m | 0.60 m | 1.0 m | 0.89 m | 0.79 m | 1.50 m | 1.34 m | 1.19m |
| -70 | 0.43 m | 0.38 m | 0.34 m | 0.56m | 0.50 m | 0.44 m | 0.84 m | 0.75 m | 0.67m |
| -75 | 0.24 m | 0.21 m | 0.19 m | 0.31m | 0.28 m | 0.25 m | 0.47 m | 0.42 m | 0.37m |
| -80 | 0.13 m | 0.12 m | 0.10 m | 0.17m | 0.15 m | 0.14 m | 0.26 m | 0.23 m | 0.21m |
| -85 | 0.08 m | 0.07 m | 0.06 m | 0.1 m | 0.09 m | 0.07 m | 0.15 m | 0.13 m | 0.11m |

Table 3. Interference zone radius

Table 3 represents the interference zone radius for different combinations of noise figure, transmit power and receiver noise raise. It is expected that the very weak interference signal may reach to the victim receiver if the interferer is located beyond the interference zone in case both systems agreed with the corresponding criteria. For example, a limited noise raise of 2 dB and noise figure of 6 dB, the WiMax receiver can operate without impact of UWB interference if the interferer is located beyond 0.15 m for -80 dBm/MHz transmit power and 0.5 m for -70 dBm/MHz.

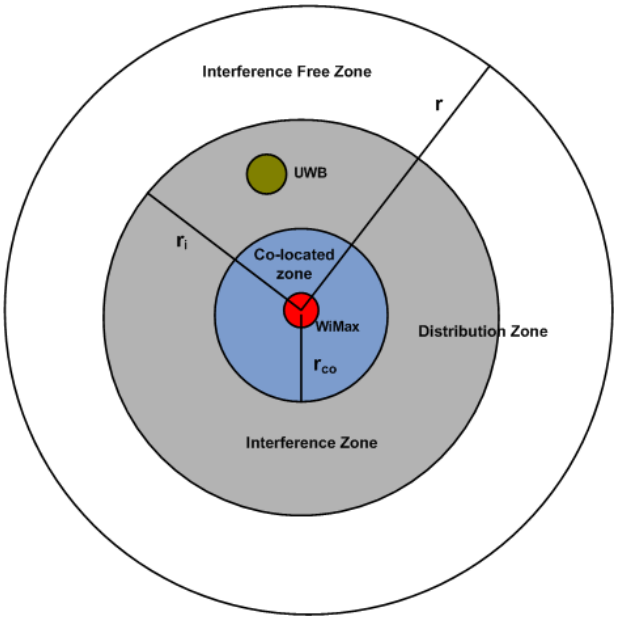



Fig. 5. SEMCAT simulation scenario

### 3.4 Probability of Single UWB to be Inside Interference Zone

We assess the probability that a UWB transmitter is located inside the interference zone. The aggregation interference input to the victim receiver is negligible when the UWB device is placed outside of the interference zone. In Figure 5, the interference range is shown by the symbol of $r_i$. A WiMax victim receiver with omni-directional antenna is located in the origin of the circle. A UWB terminal is randomly distributed over a concentric ring, centred at the

victim receiver, with inner and outer radii $r_{co}$ and r. Therefore, the area of distribution of $\pi(r^2-r^2_{co})$ and area of collocated zone of $\pi r^2_{co}$ was computed. The cumulative probability of UWB device is located in $\pi(r^2_i-r^2_{co})$ area can be expressed by the following equation:

$$p_i = \int_{r_{co}}^{r_i} P(r_i)dr_i \tag{8}$$

Where $P(r_i)$ is the probability density function of a UWB terminal that is inside the interference zone with radius of $r_i$. Since the terminal is distributed in the circle area of $\pi(r^2-r^2_{co})$ then the $P(r_i)$ is given by

$$p(r_i) = \frac{2r_i}{r^2 - r_{co}^2} \quad r_{co} \leq r_i \leq r \tag{9}$$

The radius $r_{co}$ and r are examined in the scenario section and these are 0.35 m and 2 m respectively.

| $P_t$ | Nr = 3 dB | | | Nr = 2 dB | | | Nr = 1 dB | | |
|---|---|---|---|---|---|---|---|---|---|
| dBm/ MHz | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB |
| -65 | 0.117 | 0.087 | 0.061 | 0.226 | 0.172 | 0.129 | 0.548 | 0.431 | 0.333 |
| -70 | 0.016 | 0.005 | 0 | 0.049 | 0.032 | 0.018 | 0.150 | 0.113 | 0.084 |
| -75 | 0 | 0 | 0 | 0 | 0 | 0 | 0.025 | 0.013 | 0.003 |
| -80 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 4. Probability of UWB to be inside interference zone

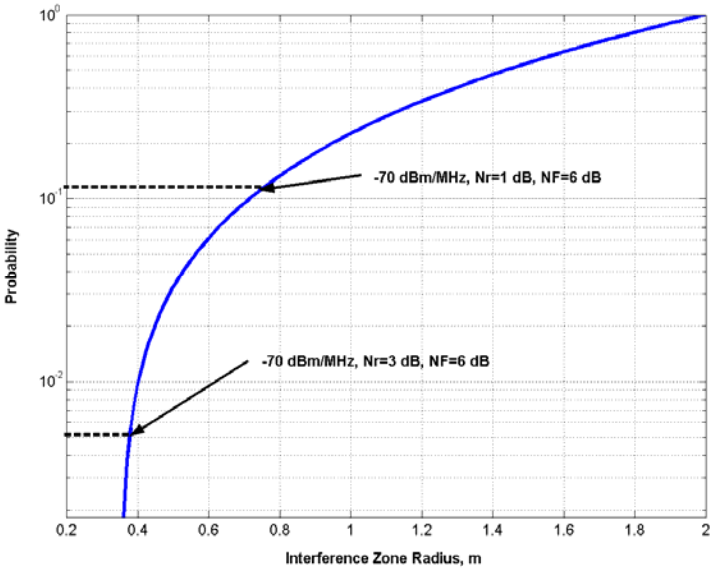



Fig. 6. Probability vs. interference zone radius

Figure 6 shows the probability of UWB device is located inside the interference zone based upon the victim receiver and interferer parameters. When we take the interference zone radius from the Table 3 and set it to the Figure 6 it will give us the probability of interference power arrived to the victim receiver. For example, at UWB transmit power of -70 dBm/MHz, noise figure of 6 dB and noise raise limit of 3 dB interference zone radius of 0.38 m is determined. That means the interference impact on victim is negligible when the interferer is located outside of the zone. The probability of UWB devices being located inside this are is 0.5% . If the receiver noise raise is 1 dB then the probability reaches to 11%. However for transmit power of -80 dBm/MHz the examined probability is negligible for any distance even at the receiver limited noise raise of 1 dB. Table 4 represents this probability for various combinations of receiver and transmitter parameters.

### 3.5 Probability of Interference for Different UWB Power Emission Levels

We have performed a system level simulation using SEAMCAT® (Spectrum Engineering Advance Monte Carlo Analysis Tool) (SEMCAT) software tool in order to compute more precise result of probability of interference from UWB transmitter to WiMax victim client receiver. It is a tool developed by the group of CEPT administrations, ETSI members and international scientific bodies to study the coexistence problem between radio systems. It is an implementation of Monte Carlo methodology whose main principle is taking samples of random variables from their probability density functions defined by user and then using those samples to calculate the probability of interference. The parameters presented in the Table 2 are used to perform the simulation. A uniform polar distribution is carried out to distribute the UWB transmitter over the area between two circles with radius of 0.35 m and 2 m, respectively. In each trial, SEAMCAT® calculates the interference power from randomly distributed UWB devices over the distribution area. The resulting interference power is calculated by

$$iRSS = 10 \log_{10}\left(\sum_{j=1}^{n} 10^{\frac{iRSS_j}{10}}\right) \tag{10}$$

with iRSS= interfering Received Signal Strength in dBm; n= number of trails.

The probabilities of interference for different UWB transmit power levels are depicted in the Figure 7, Figure 8 and Figure 9 for noise figure of 5 dB, 6 dB and 7 dB, respectively. The results are compared for three dissimilar maximum noise raise limit of 1 dB (I/N= -6), 2 dB (I/N = - 2.35) and 3 dB (I/N= 0 dB) respectively. It is observed that for PSD of -80 dBm/MHz the probability of interference is zero even if low noise raise limit and high noise figure are taken into account. The maximum probability of interference of 15% is found when the PSD of -70 dBm/MHz and the receiver is satisfied with noise raise limit of 1 dB and noise figure of 5 dB. But it is negligible if the target noise raise limit is 2 dB or 3 dB. For a PSD of -65 dBm/MHz, the probability of interference mostly was found below of 20% if the noise raise limit of 2 dB or 3 dB is considered. The results show that the interference effects from a -70 dBm/MHz UWB transmitter to a WiMax client are negligible.

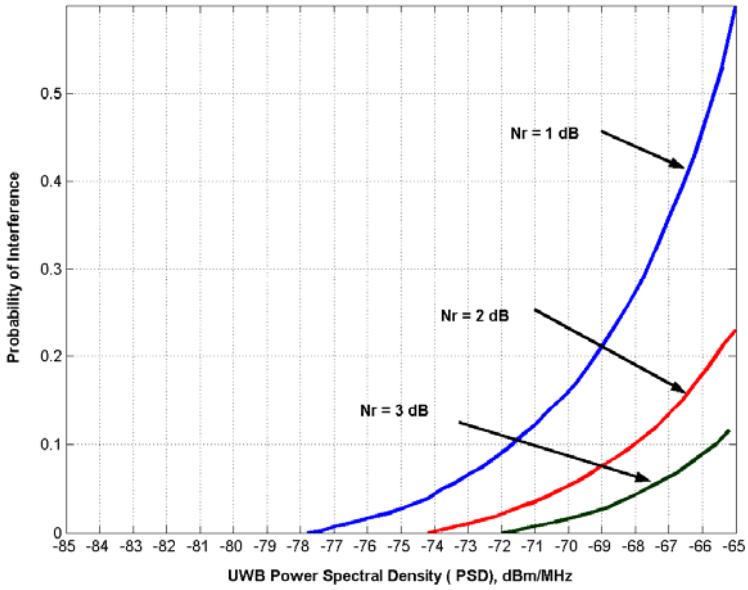



Fig. 7. Probability of interference for noise figure of 5 dB

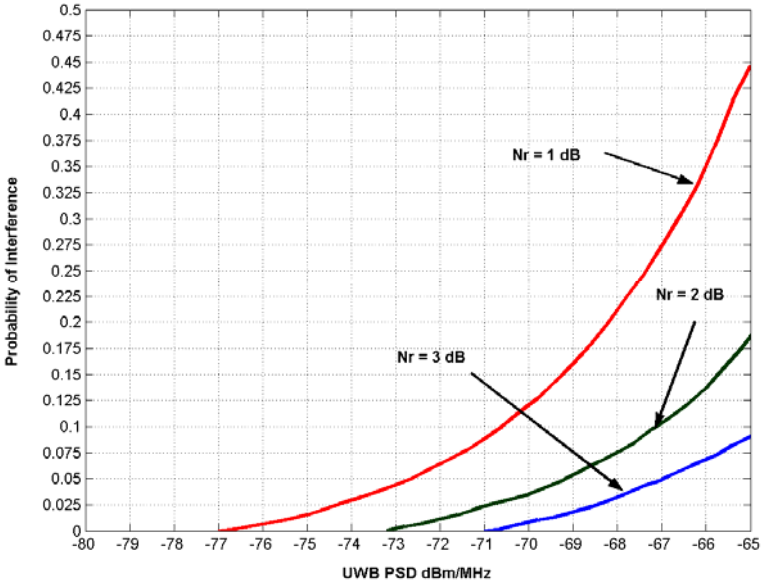



Fig. 8. Probability of interference for noise figure of 6 dB

The presented simulation results agreed with the analytical results specified in Table 4. Hence, the probability of UWB device being located inside the interference zone is equal to the probability of interference.

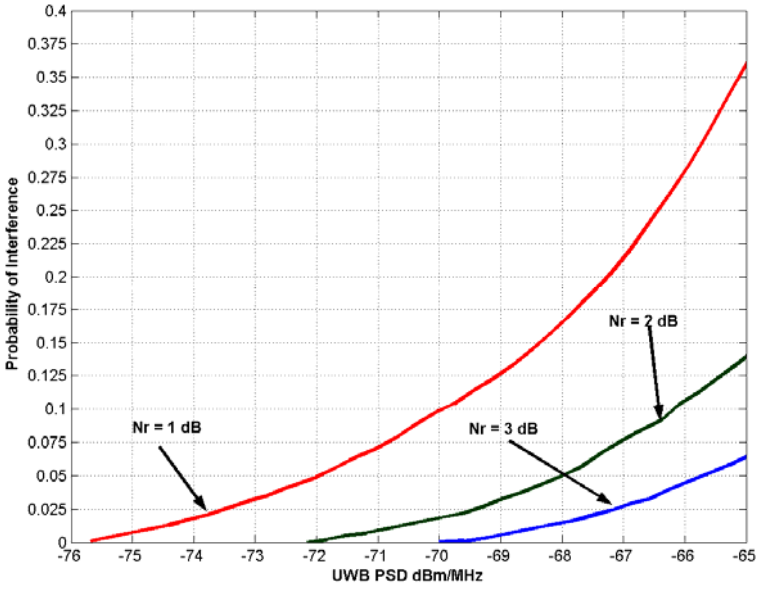



Fig. 9. Probability of interference for noise figure of 7 dB

### 3.6 Interference Evaluation in presence of inter-cell interference

Due to the inter-cell interference, the permissible noise raise at the WiMax receiver will be increased if such interference itself becomes equal to or higher than nose floor. If we consider the inter-cell interference then we rewrite equation (5) as follows:

$$N_r = 10\log_{10}\left(\frac{I_{\text{inter}} + I_{UWB} + N}{I_{\text{inter}} + N}\right) \tag{11}$$

Here, $I_{inter}$ is the inter-cell interference.

| $P_t$ | Nr = 3 dB | | | Nr = 2 dB | | | Nr = 1 dB | | |
|---|---|---|---|---|---|---|---|---|---|
| dBm/ MHz | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB |
| -65 | 0.098 | 0.0074 | 0.055 | 0.191 | 0.150 | 0.113 | 0.466 | 0.377 | 0.297 |
| -70 | 0.009 | 0.001 | 0 | 0.038 | 0.025 | 0.013 | 0.125 | 0.098 | 0.074 |
| -75 | 0 | 0 | 0 | 0 | 0 | 0 | 0.018 | 0.007 | 0.001 |
| -80 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 5. Probability of interference presence of inter-cell interference of -115 dBm/MHz

Table 5 presents the probability of interference when the inter-cell interference power of -115 dBm/MHz is considered. It is show that the probability of interference is reduced from 15% to 9.8% if noise raise limit of 1 dB and noise figure of 5 dB were assumed.

### 3.7 Interference Evaluation for Random path Loss Exponent

A free space path loss between the UWB transmitter and WiMax receiver has been used to evaluate the above interference results. Since the separation distance is about 2 meters, therefore, it is reasonable to consider the free space path loss. However, the path loss is not only depended on the separation distance rather on the environment conditions. The office desk may scatter with many small objects like books, files, monitor, etc which results of reflection, scattering of the signals. In addition, antennas might not be line-of-sight as it is integrated on the devices. It is assumed that due to multipath the path loss may decrease about 1 dB while the path loss exponent varying from 2 to 2.5. Therefore, the probability of free space path loss between these systems is low. In the following, we study the probability of interference considering the free space path loss is being 80% cases (see table 6).

| $P_t$ | Nr = 3 dB | | | Nr = 2 dB | | | Nr = 1 dB | | |
|---|---|---|---|---|---|---|---|---|---|
| dBm/ MHz | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB | $N_F$ 5dB | $N_F$ 6dB | $N_F$ 7dB |
| -65 | 0.078 | 0.059 | 0.044 | 0.153 | 0.120 | 0.090 | 0.373 | 0.302 | 0.238 |
| -70 | 0.007 | 0.001 | 0 | 0.030 | 0.020 | 0.011 | 0.100 | 0.078 | 0.059 |
| -75 | 0 | 0 | 0 | 0 | 0 | 0 | 0.014 | 0.006 | 0.001 |
| -80 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 6. Probability of Interference in presence of inter-cell interference of -115 dBm/MHz and 80% free space cases

## 4. Evaluation of UWB Interference Impact on WiMax System Performance

Since the interference from UWB devices may appear as an increasing of the NoF and $R_{sen}$, the tolerable interference levels at the receiver for the WiMax services required to be defined very carefully. Depending on its dimension, the link degradation may lead to decrease the quality of service in a certain degree. It will have a possible impact on the WiMax system in terms of loss of capacity, coverage reduction, outage of users, loss of link availability, etc. The remaining part of this chapter will investigate some of the feasible impact of the UWB emission on WiMax system, by means of loss of coverage and outage of the active users. In order to evaluate those impact, it is initially needed an estimation of cell radius using appropriate propagation model. The impacts have been studied when the receiver tolerable interference levels are limited with 1 dB, 2 dB and 3 dB of noise raise.

### 4.1 WiMax Cell Edge Reliability and Cell Radius

In the following, we present the relevant procedures and techniques to estimate the radius of a single cell. The initial approach is to select a proper channel model which is agreed with the geographical and environmental conditions on the planning areas. The IEEE 802.16 standard proposed to use Erceg propagation model for a WiMax system coverage prediction (Erceg, et. at, 1999). We used category B and category C of the Erceg path loss model with

the frequency and antenna height correction factors. The other two common factors which also indeed influence the cell radius evaluations are: CER and Fade Margin (FM). The CER refers to the probability that the RF signal strength on a circular contour at the cell edge will meet or exceed the quality threshold (e.g. -98 dBm for QPSK 1/2). However, the cell coverage reliability can be also used instead of CER, since for a given propagation environment, CER and cell area reliability are deterministically related and easily transformable.

A FM is calculated to ensure the desired CER and it is relied on the actual signal variation within each cell. If CER is increased the FM will be also increased relatively. The FM is computed on the basis of predetermined target CER figure and the said shadow fading, σ in dB. The σ is usually modelled as a lognormal distribution that describes the variation of the decibel value of the mean signal as a normal or Gaussian distribution. FM is usually given by (Bernardin, 1989)

$$F_M(\sigma, \text{cov } \%) = z\sigma \tag{12}$$

wherein z may be calculated from the defined cell edge reliability, then CER(z) is calculated as follows

$$CER(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-(t^2/2)} dt \tag{13}$$

For example, a cell edge reliability, CER(z) of 90% estimate the FM of 1.282σ. Similarly CER(z) of 75% compute FM of 0.675σ.

| Parameters | Values |
|---|---|
| Bandwidth, $B_W$ | 3.5 MHz |
| BS Power, $P_t$ | 35 dBm |
| BS Antenna Gain, $G_{BS}$ | 16 dB |
| Channel Model | Erceg Cat.B and Cat. C |
| Shadow, σ | 9.6 dB (Cat. C) 8.2 dB (Cat.B) |
| Penetration Loss, $L_{Wall}$ | 12 dB |
| SS Antenna Gain, $G_{SS}$ | 0 dB |
| Frequency | 3.5 GHz |
| BS Height | 30 m |
| SS Height | 6 m |
| Noise Figure, NF | 5 dB |
| Implementation Loss, IL | 0 dB |
| Sensitivity, $R_{Sen}$ | -98 dBm (QPSK ½)<br>-91 dBm (16QAM ½)<br>-85 dBm (64QAM 2/3) |

Table 7. WiMax system parameters for simulation

A calculation of path loss is an essential case in the cell planning and in determining the cell radius. The maximum path loss, $L_{path}$ (path attenuation) between BS and Subscriber Station (SS) can be found using a practical power budget. It is based on the computed FM, both

antennas characteristics, BS transmit power ($P_t$), SS receiver $R_{sen}$ level, and outdoor to indoor penetration losses $L_{wall}$. Then $L_{path}$ can be expressed by the following equation,

$$L_{path} = P_t + G_{BS} - FM - L_{Wall} + G_{SS} - R_{Sen} \quad (14)$$

Where, $G_{BS}$ and $G_{TS}$ are antennas gains at BS and at SS respectively. The assumed values of those parameters except FM can be taken from the table 7.
Finally, the $L_{path}$ is applied in the Erceg path loss equation in order to extract the cell radius *R*. The Erceg path loss model can be given by (Erceg, et. at, 1999)

$$L_{path} = A + 10\gamma \log_{10}(\frac{R}{d_0}) + X_f + X_h + \sigma \quad (15)$$

Where A is the free space path loss at a reference distance of $d_0$=100 m, *R* is the distance from BS to the cell edge point and $X_f$, $X_h$ are the correction factors of the operating frequency and the receiver antenna respectively. σ is the path loss exponent, which is computed according to the considered terrain type. σ is omitted in this equation because this term is already included in the FM.

| CER | FM | Cell Radius Cat. B (km) | | | Cell Radius Cat. C (km) | | |
|---|---|---|---|---|---|---|---|
| | | QPSK 1/2 | 16QAM 1/2 | 64QAM 2/3 | QPSK 1/2 | 16QAM 1/2 | 64QAM 2/3 |
| 75% | 0.675σ | 1.456 | 1.007 | 0.735 | 2.322 | 1.570 | 1.122 |
| 80% | 0.841σ | 1.338 | 0.926 | 0.675 | 2.151 | 1.454 | 1.039 |
| 85% | 1.036σ | 1.213 | 0.839 | 0.612 | 1.967 | 1.330 | 0.951 |
| 90% | 1.282σ | 1.072 | 0.741 | 0.541 | 1.758 | 1.188 | 0.850 |
| 95% | 1.645σ | 0.892 | 0.617 | 0.450 | 1.488 | 1.006 | 0.719 |
| 99% | 2.327σ | 0.632 | 0.437 | 0.319 | 1.089 | 0.736 | 0.526 |

Table 8. Estimated cell radius for Cat. B and Cat. C in km

Table 8 shows the calculated FM and cell radius for the corresponding CER. The radius is calculated for the category B and category C of the Erceg model. Type C is associated with the minimum path loss for flat terrain with light tree densities. On the other hand type B is mostly for flat terrains with moderate to heavy tree densities or hilly terrains with light tree density. For more details please refer to (Erceg, et. at, 1999). The WiMax system adopted adaptive modulation and the upper boundary of the cell coverage is determined by the robustness QPSK ½ modulation scheme. Since, it gives lowest $R_{sen}$ level, the low power signal can be feasible to receive. The cell radius is represented in the table 8 and seems slightly smaller than other literatures. The reason is the SSR antenna gain and the penetration loss. Most of the studies have taken into account the SSR antenna gain of 16 dB and penetration loss 0 dB. That means 28 dB (12 dB + 16 dB) additional path attenuation is considered in our study which results in a smaller cell radius in comparison to the previous one.

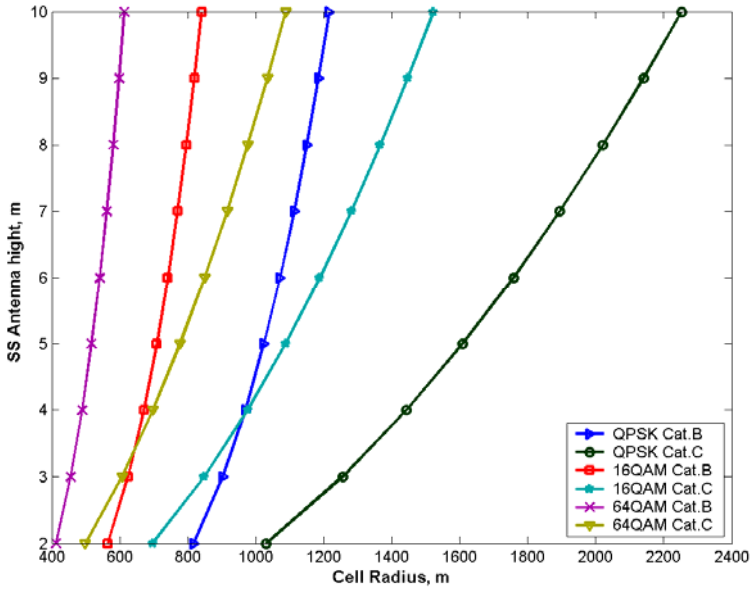



Fig. 10. Subscriber station height vs. cell radius

### 4.2 UWB Impact on the WiMax Cell Coverage

The potential UWB interference impact on WiMax cell coverage in terms of coverage reduction or cell radius reduction is estimated in the following part. The network provider may be affected economically because the reduction of cell coverage can increase the instalments cost and reduce the net profit. The provider will need to expand the number of BS or cell to cover the same area. The process to compute the reduction of cell radius can be conveniently considered in two steps:

i) The first step is to define the tolerable noise raise limits which will present a given level of UWB signal at the WiMax SSR.

ii) The second step is to compute the reduction of cell radius with introducing the noise raise limits. The decreased of the NoF will need a compensation of the $R_{sen}$ level in order to meet the minimum signal level. It must be received with a certain acceptable BER or necessary SNR for a particular modulation and coding scheme in order to decode correct the data stream. Since the cell radius is computed with respect to the $R_{sen}$ level, it will reduce with reduction level of the $R_{sen}$. At the end the percentage of cell radius reduction is calculated.

| Noise Rise | 1 dB | 2 dB | 3 dB |
|---|---|---|---|
| $I_{UWB}$/ND | -6 dB | -2.35 dB | 0 dB |
| Cat. B | 5.12% | 9.98% | 14.61% |
| Cat. C | 5.44% | 10.58% | 15.44% |

Table 9. Estimated reduction of cell radius

Table 9 shows the cell radius reduction with respect to the noise raise limits of 1 dB, 2 dB and 3 dB for the category B and category C channel model. It is found that the percentage of reduction slightly depended on the channel model. The reduction seems unacceptable when the tolerable link degradation of 3 dB is applied at SSR. For example it is about 15% when the noise raise limit is of 3 dB. On the other hand around 5% of cell radius reduction is observed if the noise increased of 1 dB is considered. In principle the 10% of cell reduction is well acceptable.

### 4.3 Interference Impact of the Active Users (Outage of Users)

WiMax SSR will be suffered by UWB interference that results of outage if it is located near the cell edge. The receiver can experience on outage when it does not meet the required SNR. Those terminals are operating very close to cell edge can goes to outage because they are running with few dB of SNR margin. The users are situated far from the cell edge will be effected on the capacity not on the outage because they usually run with enough SNR margin. Our investigation following two categories: one is to determine the percentage of the devices are situated in the 1 dB, 2 dB and 3 dB zone and other is to find out the total number of outage corresponding to the noise raise limits.

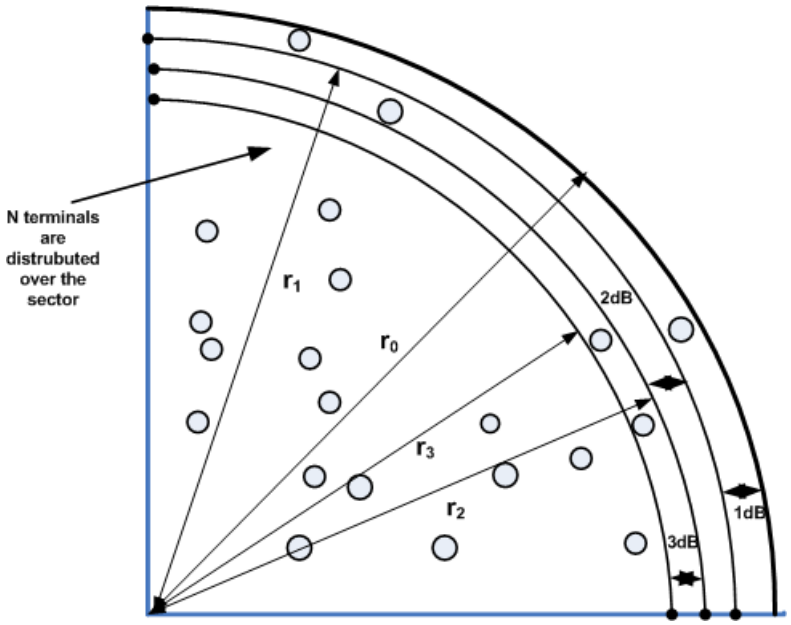



Fig. 11. 1 dB, 2 dB and 3 dB zones in the cell planning

The representation of three zones is shown in the Figure 11. The possible number of devices are located in the zone is expressed by the following equation,

$$P(x) = N.\left(\frac{r_i^{\,2} - r_{i+1}^{\,2}}{r_i^{\,2}}\right) CER \quad (z) \tag{16}$$

where N is the total number of active users distributed uniformly in the cell, $r_i$ and $r_{(i+1)}$ are the inner and outer radius of the zone and i=0,1,2.

| | Cat.B | | | Cat.C | | |
|---|---|---|---|---|---|---|
| CER | 3dB Zone | 2dB Zone | 1dB Zone | 3dB Zone | 2dB Zone | 1dB Zone |
| 75% | 6.07% | 6.74% | 7.49% | 6.35% | 7.10% | 7.94% |
| 80% | 6.47% | 7.19% | 7.99% | 6.77% | 7.57% | 8.50% |
| 85% | 6.90% | 7.64% | 8.50% | 7.19% | 8.04% | 9.00% |
| 90% | 7.28% | 8.09% | 9.00% | 7.62% | 8.51% | 9.52% |
| 95% | 7.69% | 8.54% | 9.50% | 8.04% | 9.00% | 10.05% |
| 99% | 8.00% | 8.90% | 9.91% | 8.38% | 9.37% | 10.48% |

Table 10. Victim users located in the zones at QPSK ½

Table 10 represents the percentage of users being located in the noise raise limits of 1 dB, 2 dB and 3 dB zone. It is obvious that the percentage of users in 3 dB zones will be less compared to 1 dB zone because of the less area. It is also shown that around 7.28% of victims are placed in the 3 dB zones if CER of 90% is considered.

| | Cat.B | | | Cat.C | | |
|---|---|---|---|---|---|---|
| CER | 3dB Zone | 2dB Zone | 1dB Zone | 3dB Zone | 2dB Zone | 1dB Zone |
| 75% | 20.3% | 14.2% | 7.49% | 21.3% | 15.0% | 7.94% |
| 80% | 21.6% | 15.2% | 7.99% | 22.8% | 16.1% | 8.50% |
| 85% | 23.0% | 16.1% | 8.50% | 24.2% | 17.0% | 9.00% |
| 90% | 24.4% | 17.1% | 9.00% | 25.6% | 18.0% | 9.52% |
| 95% | 25.7% | 18% | 9.50% | 27.1% | 19.00% | 10.05% |
| 99% | 26.8% | 18.8% | 9.91% | 28.2% | 19.8% | 10.48% |

Table 11. Outage users for noise raise of 1 dB, 2 dB and 3 dB at QPSK 1/2

Table 11 shows the computed number of active devices which are suffered by outage if the corresponding noise raise limits is allowed in the SSR. It is found that for 3 dB of noise raise limits above 20% of active devices are experienced outage. If it was assumed the total numbers of active devices are 30 then about 5-6 devices will be gone on outage in the case of 3 dB noise raise. Similarly about 2-3 users for 2 dB and about 1-2 users for 1 dB of noise raise limits.

## 5. Conclusion

In this chapter, the interference effect and coexistence of UWB system with WiMax has been analysed. Results have been investigated by the analytical and simulation studies. A SEAMCAT tool based on Monte-Carlo simulation methodology is used to determine the maximum possible power spectral density at the 3.5 GHz band by limiting the maximum acceptable interference level at the WiMax receiver. Also SEMCAT is used to evaluate the probability of interference by implementing a realistic interference scenario where UWB and WiMax are operating in linking with desktop PC. It is found that UWB interference impact on WiMax is harmful if UWB conducted transmit power is of more than -70 dBm/MHz.

Then, the possible UWB interference impact on the WiMax cell coverage and on outage of users has computed by considering the maximum allowable noise raise level at the receiver or vice versa. This evaluation was important to investigate how severe is UWB interference for WiMax system. At prior, the realistic cell radius by considering cell edge reliability and the practical WiMax system parameters have been calculated. It is found that cause of interference the nose raise of 1 dB, 2 dB and 3 dB at the WiMax receiver, the cell radius can be reduced about 5%, 10% and 15%, respectively.

## 6. References

Bernardin, P.; Yee, M.F. and Ellis, T. (1989), "Cell Radius Inaccuracy: A new measure of coverage reliability", IEEE Tran. on Vehicular technology, November, 1989

C802 (2005),"Correction to Rx SNR, Rx Sensitivity, and Tx Relative Constellation Error for OFDM and OFDMA systems" C80216maint-05-112r8, September, 2005

ECC (2006), "ECC Decesion of 24 March 2006 on the harminised Conditions for Devices using UWB Technologies in Bands below 10.6 GHz", Doc.ECC/DEC/(06)(04).

Erceg, V.; Greenstein, L.J.; Tjandra, S.Y.; Parkoff, S.R.; Gupta, A.; Kulic, B.; Julius, A. A.; and Bianchi, R., (1999), "An Empirically Based Path Loss Model for Wireless Channels in Suburban Environments", Vol. 17, July, 1999

FCC (2002) ,"Revision of Part 15 of the Communications Rules regaring UWB Transmission Systems", First Report and Order, ET-Docket 98-153, Feb, 2002

Giuliano, R. and Mazzenga, F., (2005), "On the Coexistence of Power-Controlled Ultrawide-Band System with UMTS, GPS, DCS180, and Fixed Wireless Systems", IEEE Trans. on Vehicular Technology, pp. 505-510, Vol. 54, 2005

IEEE (2005), "Air Interference for Fixed and Mobile Broadband Wireless Access Systems", IEEE p802.16e/D12

IEEE (2004), "Part 16: Air Interface for Fixed broadband Wireless Access systems", IEEE Std 802.16-2004.

Indepen & Quotient, (2005),"A Technical Evaluation of the Effect of UWB on Broadband Fixed Wireless Access in the 3.4 GHz Band", An investigation undertaken by Indepen and Quotient, August 2005, www.ofcom.org.uk

Kim, K.; Park, J.; Cho, J.; Lim, K.; Razzell, C. J.; Kim, K.; Lee, C.; Kim. H.; Laskar, J. (2007) "Interference Analysis and Sensing Threshold of Detect and Avoid (DAA) for UWB Coexistence with WiMax", *IEEE International Conference on UWB*, September, 2007

Mubaraq, S.; Mishra, (2007) "Detect and Avoid: An UWB/WiMax Coexistence Mechanism,"IEEE Com. Magazine, June 2007.

Nader, G. & Annamalai, A. (2007) "A Methodology for the Analysis of the Coexistence between UWB Systems and UMTS Networks", 65th VTC-Spring, April 2007.

Rahim, A. & Zeisberg, S. (2007), "Evaluation of UWB Interfernce on 3.5 GHz Fixed Terminal", *IST Mobile Summit, 2007*.

Rahim, A.; Zeisberg, S.; & Finger, A. (2007) Coexistence Study between UWB and WiMax at 3.5 GHz Band", *ICUWB 2007*.

Rahim, A.; Zeisberg, S.; Idriss, A.; and Finger, A. (2008),"The Impact of UWB Interference on WiMax Client Receiver: Detect and Avoid", ICTTA 2008

SEMCAT, "http://www.seamcat.org"

Sarfaraz, K.; Ghorashi, S.A.; Ghavami, M.; and Aghvami, A.H. (2005)," Performance of WiMax receiver in presence of DS-UWB system", IEEE Electronics Letters, December, 2005.

Snow, C.; lampe, L. and Schober, R. (2007)"Analysis of the Impact of WiMax-OFDM Interference on Multiband OFDM", *IEEE International Conference on UWB*, September, 2007

TG3 (2006)," Draft report on FWA, Annex3", 17th TG3 meeting, December, 2006

WiMaxForum (2005), "Mobile WiMax- Part I- A Technical Overview and Performance Evaluation", WiMax Forum

# 14

# Resource Management Framework for QoS Scheduling in IEEE 802.16 WiMAX Networks

Hua Wang and Lars Dittmann
*Networks Technology and Service Platforms*
*Department of Photonics Engineering*
*Technical University of Denmark, Lyngby, Denmark*
*hwan@fotonik.dtu.dk, ladit@fotonik.dtu.dk*

**Abstract**
IEEE 802.16, also known as WiMAX, has received much attention recently for its capability to support multiple types of applications with diverse Quality-of-Service (QoS) requirements. Beyond what the standard has defined, radio resource management (RRM) still remains an open issue, which plays an important role in QoS provisioning for different types of services. In this chapter, we propose a downlink resource management framework for QoS scheduling in OFDMA based WiMAX systems. Our framework consists of a dynamic resource allocation (DRA) module and a connection admission control (CAC) module. A two-level hierarchical scheduler is developed for the DRA module, which can provide more organized service differentiation among different service classes, and a measurement-based connection admission control strategy is introduced for the CAC module. Through system-level simulation, it is shown that the proposed framework can work adaptively and efficiently to improve the system performance in terms of high spectral efficiency and low outage probability.
***Keywords:*** WiMAX OFDMA radio resource management QoS scheduling

## 1. Introduction

Over the last decade, the rapid growth of high-speed multimedia services for residential and small business customers has created explosive demand for last mile broadband access. Currently, most broadband access is offered through wired lines, such as xDSL, cable or T1 networks. However, there are still a large number of areas where wired infrastructures are difficult to be deployed because of technical or commercial reasons. Broadband Wireless Access (BWA) systems are gaining extensive interests from both industry and research communities due to the advantages of rapid deployment, lower maintenance and upgrade costs, and granular investment to match market growth (1). Among the emerging technologies for BWA, IEEE 802.16 based technology, also known as Worldwide Interoperability for Microwave Access (WiMAX), is one of the most promising and attractive alternatives for last mile broadband wireless access. As expected, IEEE 802.16 standard and its evolutions have been developed to deliver a variety of multimedia applications with different Quality-of-Service (QoS) requirements, such as throughput, delay, delay jitter, fairness and packet loss rate. The physical layer specifications and MAC signaling protocols have been well defined in the standard (2),

however, radio resource management (RRM), i.e., scheduling and call admission control, still remains as an open issue, which plays an important role in QoS provisioning for different types of service.

Orthogonal Frequency Division Multiple Access (OFDMA) is a physical layer specification for IEEE 802.16 systems. OFDMA builds on Orthogonal Frequency Division Multiplexing (OFDM), which is immune to intersymbol interference and frequency selective fading, as it divides the frequency band into a group of mutually orthogonal subcarriers, each having a much lower bandwidth than the coherence bandwidth of the channel. In multi-user environment, OFDMA provides another degree of freedom by enabling dynamic assignment of subcarriers to different users at different time instances, to take advantage of the fact that at any time instance channel responses are different for different users at different subcarriers (3). Thus, dynamic subcarrier assignment (DSA) and adaptive power allocation (APA) to multiple users can be employed to improve the system performance significantly.

Recently, radio resource management for OFDMA systems has attracted enormous research interests. Many scheduling algorithms have been proposed which can adapt to changes in users' channel conditions and QoS requirements. In the literature, the resource allocation problem can be divided into two categories with different objectives. The objective of the first category is to minimize the total transmit power subject to individual data rate constraints, see (7)-(9). The objective of the second category aims at maximizing the overall (weighted) transmission rate subject to power constraints, see (10)-(12). In either case, the optimal resource allocation solutions are difficult to get due to high computational complexity. Instead, suboptimal solutions based on relaxation, problem splitting, or heuristic algorithms are proposed to reduce computational complexity (4). Such algorithms are often refereed to as *loading algorithms*.

In most loading algorithms, the QoS requirement of each user is usually defined in terms of a fixed data rate per frame. However, in practical communication systems, it is neither sufficient nor efficient to represent different QoS requirements by a fixed data rate per frame. The resource allocation problem for systems supporting both real-time (RT) and non-real-time (NRT) multimedia traffic becomes much more complicated when diverse QoS requirements have to be considered. The transmission of RT packets can be delayed as long as the delay constraint is not violated, and the transmission of NRT packets can be more elastic. Furthermore, most loading algorithms assume that users always have data to transmit, which is not the case in real systems. Instead, appropriate traffic models should be taken into account in the design of scheduling algorithms. Therefore, efficient *packet-based scheduling algorithms* are of interest. Many packet scheduling algorithms with different design objectives have been proposed in (13)-(22).

With respect to packet-based scheduling algorithms, most of the existing literature focuses on the design of one-level flat scheduler. In such approach, each connection is assigned a priority value based on some criterion and the connection with the highest priority value is scheduled for transmission. This approach has the advantage of low implementation complexity. However, due to different traffic patterns and diverse QoS requirements among different service classes, it is hard to well define a unified priority criterion that can work well for all service classes. Thus, it is desirable to individually design the scheduling algorithm for each service class and separate the resource allocation from the packet scheduling. The first paper proposing the idea of a two-level hierarchical scheduler is in (23). Performance comparisons between one-level and two-level schedulers are done in (24). In (25), Chang et al. proposed an adaptive hierarchical polling approach to minimize the average polling delay and band-

width used for polling. However, so far little work has been done in the design of an efficient bandwidth distribution algorithm for the aggregate resource allocator, which is critical on the overall performance of a two-level hierarchical scheduler and therefore should be carefully designed.

In this chapter, we present a downlink resource management framework for QoS scheduling in OFDMA based WiMAX Systems. The framework consists of a dynamic resource allocation (DRA) module and a connection admission control (CAC) module. DRA emphasizes on how to share the limited radio resource in terms of subchannels and time slots among subscriber stations with the objective of increasing the spectral efficiency while satisfying the diverse QoS requirements in each service class. CAC highlights how to limit the number of ongoing connections preventing the system capacity from being overused. The major contributions of this chapter include:

1. A two-level hierarchical scheduler is employed to split the resource allocation problem into two subproblems: a bandwidth distribution problem in the aggregate resource allocator and a scheduling problem in class schedulers. Yet there is sufficient coupling between the allocator and the schedulers as the allocator is aware of the performance of the schedulers.
2. A novel priority-based scheduling algorithm is proposed for rtPS and nrtPS class schedulers, which tightly couples the packet scheduling and subcarrier allocation together through in integrated cross-layer approach to take advantage of the inter-dependencies between the PHY and MAC layers.
3. An adaptive estimation-based bandwidth distribution scheme is proposed for the aggregate resource allocator. The proposed scheme first estimates the required amount of bandwidth in each class scheduler based on the backlogged traffic and the average modulation efficiency. Then an exponentially smoothed curve with respect to QoS satisfaction is applied to adjust the estimation in order to increase the spectral efficiency while maintaining a guaranteed QoS performance.
4. An effective measurement-based connection admission control policy is proposed for the CAC module, which takes the current state of the network and class priority into consideration when admission decisions are made.

Through the detailed system-level simulations, the study shows that the proposed resource management framework can significantly increase the spectral efficiency while ensuring the QoS requirements of each service class.

The rest of the chapter is organized as follows: We first give a brief introduction of IEEE 802.16 in Section 2. Then, the structure of the proposed downlink resource management framework is described in Section 3, followed by the design of hierarchical resource allocation algorithms and connection admission control policies in Section 4 and Section 5, respectively. In Section 6, simulation environments and results are outlined and discussed. Finally, conclusions are drawn in Section 7.

## 2. Overview of IEEE 802.16 Networks

IEEE 802.16/WiMAX technology supports both mesh and point-to-multipoint (PMP) networks (2). In the PMP mode, the network has a cellular structure where base station (BS) governs all the communications in the network and the subscriber stations (SSs) cannot communicate with each other directly. In contrast, in the mesh mode, traffics can be exchanged

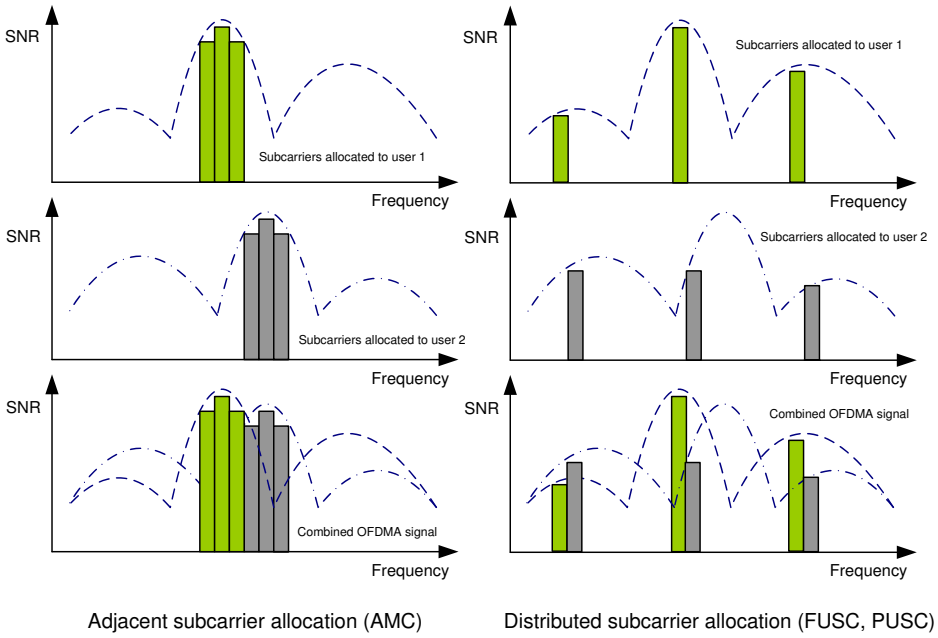



Fig. 1. Adjacent and distributed subcarrier allocation

directly between SSs. In this chapter, we only concentrate on the WiMAX PMP network. The physical layer of the IEEE 802.16 air interface operates either at 10-66 GHz for line-of-sight (LOS) communications or 2-11 GHz for non-line-of-sight (NLOS) communications, and it supports data rates in the range of 32-130 Mbps depending on the transmission bandwidth as well as the modulation and coding schemes used (5). Adaptive modulation and coding scheme (AMC) is supported in the standard. The design objective of AMC is to maximize the data rates by adjusting the transmission parameters according to time-varying channel conditions, while maintaining a prescribed target packet error rate (PER).

As specified in the standard, IEEE 802.16 employs OFDM in the physical layer. In particular, two different air interfaces based on OFDM can be used: WirelessMAN-OFDM and WirelessMAN-OFDMA. The first option employs fast Fourier transform (FFT) of size 256 (subcarriers). Time-division multiplexing (TDM) and time-division multiple access (TDMA) are used for downlink and uplink transmission respectively. The second option employs a larger FFT space (2048 and 4096 subcarriers) which are further grouped into subchannels. The subchannels are assigned to different subscriber stations and may employ different modulation and coding schemes to exploit frequency diversity as well as time diversity. The subchannels are also used for multiple access, namely, orthogonal frequency division multiple access (OFDMA). There are two approaches of allocating subcarriers to form a subchannel: distributed subcarrier permutation and adjacent subcarrier permutation. The two approaches are shown in Fig. 1. In distributed subcarrier permutation, a subchannel is formed with different subcarriers randomly distributed across the channel spectrum. This approach maximizes

the frequency diversity and averages inter-cell interference. It is suitable for mobile environment where channel characteristics change fast. Both partial usage of subchannels (PUSC) and full usage of subchannels (FUSC) schemes employ distributed subcarrier permutation. In adjacent subcarrier permutation, a subchannel is formed by grouping adjacent subcarriers. This approach creates a 'loading gain' and is easy to use with beam-forming adaptive antenna system (AAS). It is suitable for stationary or nomadic environment where channel characteristics change slowly. The AMC scheme employs adjacent subcarrier permutation.
IEEE 802.16 standard supports both frequency-division duplex (FDD) and time-division duplex (TDD) transmission modes. For FDD scheme, distinct frequency channels are assigned for uplink and downlink transmissions. In contrast, TDD scheme uses a single frequency channel for uplink and downlink transmissions by dividing the MAC frame into uplink and downlink subframes. The length of these subframes are determined dynamically by the BS and are broadcasted to the SSs through downlink and uplink MAP messages (DL-MAP and UL-MAP) at the beginning of each frame.
Four types of services are defined in the standard, each of which has different QoS requirements (2):

- Unsolicited grant service (UGS): This type of service is designed to support real-time service flows, with strict delay requirement, which generate fixed-size data packets periodically, such as T1/E1.
- Real-time polling service (rtPS): This type of service is designed to support real-time service flows, with less stringent delay requirements, which generate variable-size data packets at periodic intervals, such as VoIP with silence suppression.
- Non-real-time polling service (nrtPS): This type of service is designed to support delay-tolerant data streams which are more bursty in nature, such as FTP. In general, the nrtPS can tolerate longer delays and is insensitive to delay jitter, but requires a minimum throughput.
- Best-effort service (BE): This type of service is designed for traffic with no QoS requirements, such as email, and therefore may be handled on a resource-available basis.

## 3. Downlink Resource Management Framework

### 3.1 Assumptions and Preliminaries

In this chapter, we only investigate the WiMAX PMP network with OFDMA-TDD operation. We assume that subscriber stations are stationary or nomadic users with slowly varying channel conditions. Adjacent subcarrier permutation strategy is employed to support AMC. In OFDMA, radio resource is partitioned in both frequency domain and time domain, which results in a hybrid frequency-time domain resource allocation. It provides an added dimension of flexibility in terms of higher granularity compared to OFDM/TDM systems.
We consider the downlink scenario of an infrastructure-based OFDMA system with $U_s$ subcarriers and $K$ users. At the physical layer, the time axis is divided into frames with fixed length, each of which consists of a downlink (DL) and an uplink (UL) subframe to support TDD operation. In each DL subframe, there are $U_t$ time slots available for downlink transmissions, each of which may contain one or several OFDM symbols. To reduce the resource addressing space, channel coherence in frequency and time is exploited by grouping $I_s$ adjacent subcarriers and $I_t$ time slots to form a basic resource unit (BRU) for resource allocation. A BRU is the minimum resource allocation unit as shown in Fig. 2. The size of a BRU is adjusted so that the channel experiences flat fading in both frequency and time domain. Thus in each

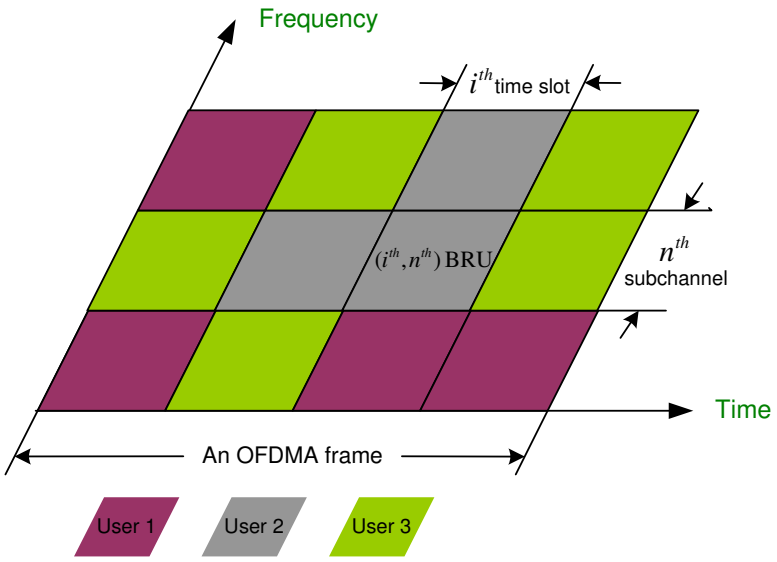



Fig. 2. Frequency-time domain radio resource allocation in OFDMA systems

DL subframe, there are $S = U_s / I_s$ subchannels in frequency domain and $N = U_t / I_t$ slots in time domain, which corresponds to a total of $S * N$ BRUs available in frequency-time domain for DL transmissions. Each BRU can be assigned to different users and be independently bit and power loaded. In principle, adaptive power allocation in each BRU can improve the system performance. However, some studies show that performance improvements are only marginal over a wide range of SNRs due to the statistical effects (3). Therefore, we assume that the total transmission power is equally distributed among all subchannels.
We further assume that in each frame the base station has perfect knowledge of channel state information (CSI) for each subchannel of each user. This can be obtained by piggybacking such information in each uplink packet, which is suitable for slowly varying channels. Based on CSI, adaptive modulation and coding scheme is employed to adjust the transmission mode dynamically according to the time-varying channel conditions. Multiple transmission modes are available, with each mode representing a pair of specific modulation format and a forward error correcting code. The transmission mode is determined by the instantaneous signal-to-noise ratio (SNR). To utilize the PHY layer resources more efficiently, fragmentation at the MAC layer is enabled. A separate queue with a finite queue length of $L$ MAC protocol data units (PDUs) is maintained for each connection at the base station. We assume that the MAC PDUs are of fixed size, each of which contains $d$ information bits.

### 3.2 Structure of the Resource Management Framework

The proposed downlink resource management framework consists of a dynamic resource allocation (DRA) module and a connection admission control (CAC) module. CAC is responsible for preventing the system capacity from being overused by limiting the number of ongoing connections. DRA aims at an efficient usage of the scarce radio resource, while maintaining

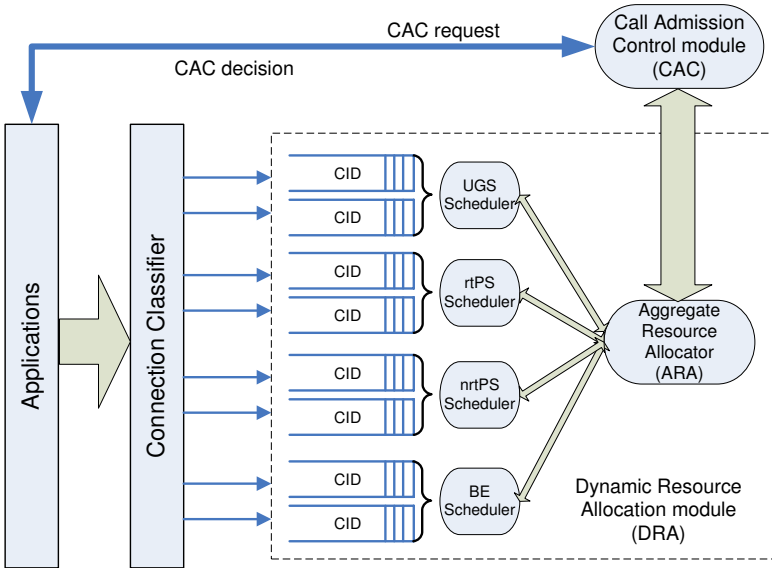



Fig. 3. Structure of the proposed downlink resource management framework for IEEE 802.16 systems

a guaranteed QoS performance among all admitted users. From a cross-layer perspective, resource management should be a joint optimization of a large number of variables ranging from application layer to physical layer. That is, which user should be scheduled for transmission in which BRU, under certain QoS constraints and time-varying channel conditions. However, this would lead to a very complex algorithm design since four types of services with different QoS requirements are defined in the standard. This approach would also lose flexibility if new traffic requirements or different optimization goals were to be considered.
Therefore, we adopt a two-level hierarchical scheduler for the DRA module, a loosely cross-layer approach trying to strike a balance between flexibility and modularity. Then the resource allocation problem can be split into two subproblems, i.e., a bandwidth distribution problem in the aggregate resource allocator and a scheduling problem in class schedulers. Yet there is sufficient coupling between the allocator and the schedulers as the allocator is aware of the performance of the schedulers. An advantage of this two-level hierarchical resource allocation architecture is that the algorithms for the allocator and the schedulers can be developed independently of each other. As an example, if the scheduling algorithm of a class scheduler is changed from, let's say, maximum SNR to proportional fairness, this will affect the way of BRU assignment in that class scheduler, but the bandwidth distribution algorithm in the aggregate resource allocator can be kept the same, as long as the design objective remains unchanged.
Fig. 3 depicts the proposed downlink resource management framework for IEEE 802.16 systems. When an application initiates a connection, it sends the connection request to the CAC module with connection type, traffic parameters, and QoS requirements. Then the CAC module interacts with the DRA module to get the current network state and commits admission

decisions. All arriving packets from the application layer are classified by the connection classifier according to their connection identifications (CID) and traffic types, and are sent to the corresponding service class and get queued. The DRA module is responsible for scheduling packets of all admitted connections. It consists of an aggregate resource allocator (ARA) and four class schedulers. The ARA distributes bandwidth to each class scheduler based on some criterion. Once the class scheduler receives bandwidth from the ARA, it schedules packets in its queues. In each class scheduler, because the incoming flows have similar traffic patterns and QoS requirements, the class scheduler has the freedom to independently choose its own scheduling algorithm which can best meet the QoS requirements. Therefore, this two-level hierarchical resource allocation module can have multiple scheduling criteria and better schedule packets in each service class than its one-level flat counterpart.

## 4. Hierarchical Resource Allocation

In this section, we first describe the scheduling algorithms employed in each class scheduler, then an adaptive estimation-based bandwidth distribution algorithm for the aggregate resource allocator is proposed.

### 4.1 Scheduling Algorithms for Class Schedulers

Class scheduler in each service class receives bandwidth from the ARA and involves in the allocation of subchannels and time slots among different users in its service queues. Scheduling algorithms designed for class schedulers should have the goal of maximizing the spectral efficiency with satisfied QoS performance. In this section, we apply the appropriate packet scheduling algorithm to each class scheduler.

#### 4.1.1 Scheduling UGS connections

The scheduling of UGS connections is well defined by the standard. In UGS, the transmission mode at the PHY layer is fixed during the whole service time (2). The AMC is not employed for UGS connections. The time slots allocated for UGS connections per frame are fixed, based on their constant bit-rate requirements negotiated in the initial service access phase.

#### 4.1.2 Scheduling rtPS and nrtPS connections

The rtPS connection is delay-sensitive and has strict delay requirement. The nrtPS connection can tolerate longer delays, but requires a minimum throughput. We propose a novel priority-based scheduling algorithm for rtPS and nrtPS class schedulers. The basic idea behind the proposed algorithm is that the transmission is scheduled in a packet-by-packet fashion based on its priority value. Specifically, at each scheduling interval, if a PDU was scheduled for transmission on a specific subchannel, it is assigned a priority value based on the instantaneous channel condition (PHY layer issue), as well as the QoS constraint (MAC layer issue). Then we can formulate the scheduling problem into a mathematical optimization problem with the objective to maximize the total achievable priority values.

We apply an extended EXP algorithm as our priority function for both rtPS and nrtPS connections. The EXP rule was proposed to provide QoS guarantees over a shared wireless link in terms of the average packet delay for RT traffic and a minimum throughput for NRT traffic (19).

For rtPS connections, if the $i^{th}$ PDU from the $k^{th}$ connection is scheduled for transmission on subchannel $n$, its priority is calculated as:

$$\mathbf{P}(k,i,n) = a_k \cdot \frac{\mu_{k,n}(t)}{\overline{\mu}_k(t)} \cdot \exp\left(\frac{a_k W_{k,i}(t) - \overline{aW}}{1+\sqrt{\overline{aW}}}\right) \tag{1}$$

where $\overline{aW} = \frac{1}{K}\sum_k a_k W_{k,1}(t)$, and $a_k = -\log \delta_k / T_{k,\max}$. $W_{k,i}(t)$ is the $i^{th}$ PDU delay of connection $k$ at time $t$, $T_{k,\max}$ is the maximum allowable delay of connection $k$, $\delta_k$ is the maximum outage probability of connection $k$, $\mu_{k,n}(t)$ is the instantaneous channel rate with respect to the signal-to-noise ratio (SNR) and a predetermined target error probability if subchannel $n$ is assigned to connection $k$ at time $t$, and $\overline{\mu}_k(t)$ is the exponential moving average (EMA) channel rate of connection $k$ with a smoothing factor $t_c$, calculated as:

$$\overline{\mu}_k(t) = (1-\frac{1}{t_c})\overline{\mu}_k(t-1) + \frac{1}{t_c}\mu_k(t) \tag{2}$$

where $\mu_k(t) = \sum_{n=1}^{N} c_{k,n} \cdot \mu_{k,n}(t)$ is the total channel rate of connection $k$ at time $t$. If subchannel $n$ is assigned to connection $k$, $c_{k,n} = 1$, otherwise $c_{k,n} = 0$.

For nrtPS connections, the extended EXP algorithm is used in conjunction with a token bucket control to guarantee a minimum throughput (19). We associate each nrtPS queue with a virtual token bucket. Tokens in each bucket arrive at a constant rate $r_{k,\text{req}}$, which is the required minimum throughput of connection $k$. After a PDU is scheduled for transmission, the number of tokens in the corresponding token queue is reduced by the actual amount of data transmitted. The calculation of the priority for an nrtPS PDU is similar to Exp. (1), with the exception that $W_{k,i}(t)$ in nrtPS is defined as the virtual waiting time of the $i^{th}$ PDU from connection $k$:

$$W_{k,i}(t) = \frac{\max\left\{0, Q_k(t) - (i-1)\cdot d\right\}}{r_{k,\text{req}}} \quad k \in \text{nrtPS} \tag{3}$$

where $Q_k(t)$ is the number of tokens associated with connection $k$ at time $t$, and $d$ is the fixed size of a MAC PDU.

Let $\mathbf{u}(k,i,n)$ be defined as a binary random variable indicating subchannel allocation. That is, $\mathbf{u}(k,i,n) = 1$ means that the $i^{th}$ PDU from connection $k$ is allocated for transmission on subchannel $n$, and $\mathbf{u}(k,i,n) = 0$ otherwise. Also let us define $\mathbf{m}(k,i,n)$ be the number of time slots occupied on subchannel $n$ if the $i^{th}$ PDU from connection $k$ is scheduled for transmission on subchannel $n$, calculated as:

$$\mathbf{m}(k,i,n) = \lceil \frac{d}{\mu_{k,n}(t)} \rceil \tag{4}$$

where $\lceil x \rceil$ denotes the smallest integer larger than $x$.

Then, the scheduling problem can be mathematically formulated as follows:

$$\arg\max_{\mathbf{u}(k,i,n)} \sum_{k=1}^{K}\sum_{i=1}^{L}\sum_{n=1}^{S} \mathbf{u}(k,i,n)\cdot\mathbf{P}(k,i,n) \tag{5}$$

subject to:

$$\sum_{k=1}^{K}\sum_{i=1}^{L} \mathbf{u}(k,i,n)\cdot\mathbf{m}(k,i,n) \le N \quad \forall n \tag{6}$$

$$\sum_{n=1}^{S} \mathbf{u}(k,i,n) \leq 1 \quad \forall k,i \tag{7}$$

$$\mathbf{u}(k,i,n) \in \{0,1\} \quad \forall k,i,n \tag{8}$$

where $S$ denotes the total number of subchannels, $N$ denotes the total number of time slots, $K$ denotes the total number of connections, and $L$ denotes the maximum queue size.
The first constraint ensures that the allocated bandwidth does not exceed the total available bandwidth in terms of time slots on each subchannel. The second constraint says that a PDU can only be transmitted via one subchannel. The instantaneous channel conditions and the QoS related parameters are embodied into the priority function $\mathbf{P}(k,i,n)$ with the objective of maximizing the total achievable priority values, thus improving the spectral efficiency while maintaining QoS guarantees.
The above optimization problem can be solved by determining the values of binary variable $\mathbf{u}(k,i,n)$ through standard linear integer programming (LIP)[1]. The solution to the problem provides an optimal resource allocation. However, the computation complexity of the optimal solution is too high to be applied in practical systems. To reduce the computational complexity, we propose a suboptimal algorithm with low complexity.
In the suboptimal algorithm, we allocate radio resources on a packet-by-packet basis. The general idea is that, at each scheduling interval, the packet with the highest priority value from all queues is scheduled for transmission, and this procedure continues until either there is no radio resource left or there is no packet remaining unscheduled in the queue. A detailed description of the proposed scheduling algorithm is listed in pseudocode 1, where $\Omega_s^k$ is the set of subchannels that are available for data transmission of connection $k$, $t_n$ is the number of residual time slots on subchannel $n$, $q_k$ is the current queue size of connection $k$, and $i_k$ is a pointer to the next PDU to be scheduled of connection $k$.

[1] The optimal solution of the LIP problem formulated in this chapter is obtained by using the General Algebraic Modeling System (GAMS).

**Algorithm 1** Suboptimal Scheduling Algorithm for rtPS and nrtPS Class Schedulers

1: Set $t_n \leftarrow N$ for $\forall n$ {initialize $t_n$}
2: Set $i_k \leftarrow 1$ for $\forall k$ {initialize $i_k$}
3: Get $q_k$ for $\forall k$ {get the queue size of connection $k$}
4: **for** $k = 1$ to $K$ **do**
5: **if** $q_k > 0$ **then**
6: Set $\Omega_s^k \leftarrow \{1, \cdots, S\}$ {initialize $\Omega_s^k$}
7: **else**
8: Set $\Omega_s^k \leftarrow \phi$ {set $\Omega_s^k$ to be null}
9: **end if**
10: **end for**
11: **while** $\exists x, \Omega_s^x \neq \phi$ **do**
12: **for** $k = 1$ to $K$ **do**
13: **while** $\Omega_s^k \neq \phi$ **do**
14: Select $n \leftarrow \arg\max_{n \in \Omega_s^k} \mu_{k,n}(t)$ {assign the best subchannel from the available subchannel set}
15: **if** $t_n \geq \lceil \frac{d}{\mu_{k,n}(t)} \rceil$ **then**
16: Calculate $\mathbf{P}(k, i_k, n)$ in Exp. (1)
17: BREAK
18: **else**
19: $\Omega_s^k \leftarrow \Omega_s^k - \{n\}$ {remove $n$ from the available subchannel set if there is not enough capacity left}
20: CONTINUE
21: **end if**
22: **end while**
23: **end for**
24: Schedule the $i_{k^*}$th PDU of connection $k^*$ on subchannel $n^*$, where $(k^*, i_{k^*}, n^*) \leftarrow \arg\max \mathbf{P}(k, i_k, n)$
25: $t_{n^*} \leftarrow t_{n^*} - \lceil \frac{d}{\mu_{k^*,n^*}(t)} \rceil$ {update the residual time slots}
26: **if** $i_{k^*} = q_k$ **then**
27: $\Omega_s^{k^*} \leftarrow \phi$ {set $\Omega_s^{k^*}$ to be null when all pending PDUs of connection $k^*$ have been scheduled for transmission}
28: **else**
29: $i_{k^*} \leftarrow i_{k^*} + 1$ {point to the next pending PDU}
30: **end if**
31: **end while**

It works as follows: If connection $k$ has pending traffic in the queue, the proposed algorithm first pre-allocates the best subchannel $n$ in terms of the instantaneous channel quality to connection $k$ from its available subchannel set $\Omega_s^k$ (see Step 14). If there is not enough capacity left on the best subchannel $n$ to accommodate one PDU from connection $k$'s queue, subchannel $n$ will be removed from connection $k$'s available subchannel set, and the second best subchannel $n'$ will be selected. This procedure continues until a best possible subchannel is pre-allocated to connection $k$ (see Step 13-22). Otherwise, connection $k$ is removed from the scheduling list. After the subchannel pre-allocation process for all connections is complete, the algorithm calculates the priority value of the head-of-line (HOL) PDU in each non-empty queue, and

schedule the PDU with the highest priority value for transmission on subchannel $n^*$ (see Step 16 & 24). The scheduled PDU is removed from the corresponding queue and the consumed radio resources in terms of time slots are subtracted on subchannel $n^*$ (see Step 25 & 26-30). Then it starts from the beginning and continues until either there is no radio resource left or there is no PDU pending in the queue. A detailed description of the proposed suboptimal algorithm is listed in pseudocode 1.

#### 4.1.3 Scheduling BE connections

Since there are no QoS guarantees for BE connections, we simply apply the Proportional Fair (PF) algorithm to schedule BE traffic. The PF algorithm attempts to serve each user at its peak channel condition. Hence the PF algorithm can utilize the radio resource efficiently and give proportional fairness among users (20). At each scheduling point, the PF algorithm selects connection $k$ for transmission on subchannel $n$ as follows:

$$(k, n) = \arg\max_{k} \frac{\mu_{k,n}(t)}{\overline{\mu}_k(t)} \quad \forall n \tag{9}$$

where $\mu_{k,n}(t)$ and $\overline{\mu}_k(t)$ are defined as the same in (1).

### 4.2 Scheduling Algorithm for the Aggregate Resource Allocator

The aggregate resource allocator (ARA) is responsible for distributing the total available bandwidth among class schedulers. If the ARA does not allocate enough bandwidth to the class scheduler, the QoS requirements in the corresponding service class may not be guaranteed. On the other hand, if the ARA allocates too much bandwidth to the class scheduler, the allocated radio resource may not be utilized efficiently or even be wasted. Therefore, the resource distribution algorithm of ARA is a critical factor on the performance of class schedulers and has to be carefully designed.

#### 4.2.1 Conventional Bandwidth Distribution Algorithms

One possible solution is that the ARA distributes bandwidth among service classes following strict class priority, from highest to lowest, i.e., UGS, rtPS, nrtPS and BE. After all connections in high priority class have been served, connections in low priority class are scheduled for transmission. By doing so, the ARA can differentiate different service classes based on their class priority. The priority-based scheme is simple, but one disadvantage of this algorithm is that higher priority classes may starve the bandwidth for lower priority classes.

To overcome this problem, the ARA may partition the total bandwidth into several portions to satisfy proportional fairness among service classes. This method can prevent the starvation of low priority classes. There are static and dynamic bandwidth distribution schemes in this method. In the static scheme, the ARA distributes a fixed amount of bandwidth to each class scheduler at every scheduling interval. This approach has the advantage of simplicity and works well when the traffic pattern in each service class is regular and stable, which unfortunately is not always the case in data communications. Therefore, a dynamic bandwidth distribution scheme which can adapt to the traffic pattern and the performance of class schedulers is believed to be a better solution.

#### 4.2.2 Proposed Adaptive Bandwidth Distribution Algorithm

The design objective of our proposed resource allocation algorithm is to adaptively allocate bandwidth to each service class in order to increase the spectral efficiency while satisfying the

diverse QoS requirements. In designing the proposed adaptive resource allocation algorithm, we have taken the following aspects into account: (i) the backlogged traffic; (ii) the average modulation efficiency; (iii) the QoS satisfaction. The general idea is that, in the scheduling interval, the ARA first estimates the amount of bandwidth required in each class scheduler based on the backlogged traffic and the average modulation efficiency. Then depending on the QoS performance in each class scheduler, the estimation is further increased or decreased to maintain the guaranteed performance.

We separate the bandwidth allocation of UGS class from the others as it has been defined by the standard. At the beginning of each frame, the ARA allocates a fixed amount of time slots $N_{\text{UGS}} = \sum_{i\in\{\text{UGS}\}} \theta_i$ to UGS connections based on their constant bit-rate requirements negotiated in the initial service access phase, where $\theta_i$ is the number of time slots required by UGS connection $i$. Let $N_{\text{total}}$ be the total number of time slots in each frame, then the residual time slots after serving UGS class $N_{\text{rest}} = N_{\text{total}} - N_{\text{UGS}}$ are distributed among rtPS, nrtPS and BE classes, which employ AMC scheme at the PHY layer.

For rtPS class, since each packet has a rigid delay requirement, the total sum of the current queue size in rtPS class is an appropriate measure for the backlogged traffic $B_{\text{rtPS}}(t) = \sum_{i\in\{\text{rtPS}\}} q_i(t)$, where $q_i(t)$ is the number of bits in queue $i$ at time $t$. The average modulation efficiency $\overline{\mu}_{\text{rtPS}}(t)$ is defined as the average number of bits carried per OFDM symbol over a sliding time window $t_c$. $\gamma$ is a QoS related parameter (i.e., maximum allowable delay in rtPS) representing the proportion of backlogged traffic that has to be transmitted within each frame. Then the estimated number of time slots required for rtPS class can be expressed as follows:

$$E_{\text{rtPS}}(t) = \alpha(t) \cdot \frac{\gamma B_{\text{rtPS}}(t)}{\overline{\mu}_{\text{rtPS}}(t)} \tag{10}$$

where $\alpha(t)$ is a QoS-aware adjustment factor that is updated according to the performance of the class scheduler on a frame by frame basis. The basic idea is that when the class scheduler experiences good QoS satisfaction, the value of $\alpha(t)$ is decreased to save bandwidth for other classes. Otherwise, the value of $\alpha(t)$ is increased to guarantee the required QoS. Towards this end, an exponentially smoothed curve is applied to adjust the value of $\alpha(t)$. The adjustment, which is $|\Delta\alpha(t)| = |\alpha(t) - \alpha(t-1)|$, is minor if the QoS outage probability is around a predefined target threshold. Otherwise, $|\Delta\alpha(t)|$ is exponentially increased to either increase or decrease the allocated bandwidth to the class scheduler, as illustrated in Fig. 4. The calculation of $\Delta\alpha(t)$ is specified as follows:

$$\Delta\alpha(t) = \begin{cases} \xi_{\max} \cdot \frac{\exp(\beta \cdot d(t)) - 1}{\exp(\beta \cdot D_{\max}) - 1} & \text{if } \mathrm{P_r}(t) \geq T_h \\ -\xi_{\max} \cdot \frac{\exp(\beta \cdot d(t)) - 1}{\exp(\beta \cdot D_{\max}) - 1} & \text{if } \mathrm{P_r}(t) < T_h \end{cases} \tag{11}$$

where $d(t)$ is the truncated difference between the current outage probability and the outage probability threshold:

$$d(t) = \min\{|\mathrm{P_r}(t) - T_h|, D_{\max}\}$$

where $\mathrm{P_r}(t)$ is the delay outage probability at time $t$, $T_{\text{h}}$ is the outage probability threshold, $D_{\max}$ is the truncated maximum value of $|d(t)|$, $\beta$ is a shape factor which is used to tune the adaptation degree, and $\xi_{\max}$ is the maximum value of $|\Delta\alpha(t)|$. Term $(\exp(\beta \cdot d(t)) - 1)/(\exp(\beta \cdot D_{\max}) - 1)$ is a normalization function of $(\mathrm{P_r}(t) - T_{\text{h}})$. When $\mathrm{P_r}(t)$ is close to $T_{\text{h}}$, the normalized value is close to zero. Otherwise it increases exponentially to one. The overall bandwidth estimation procedure for rtPS class can be described as follows:

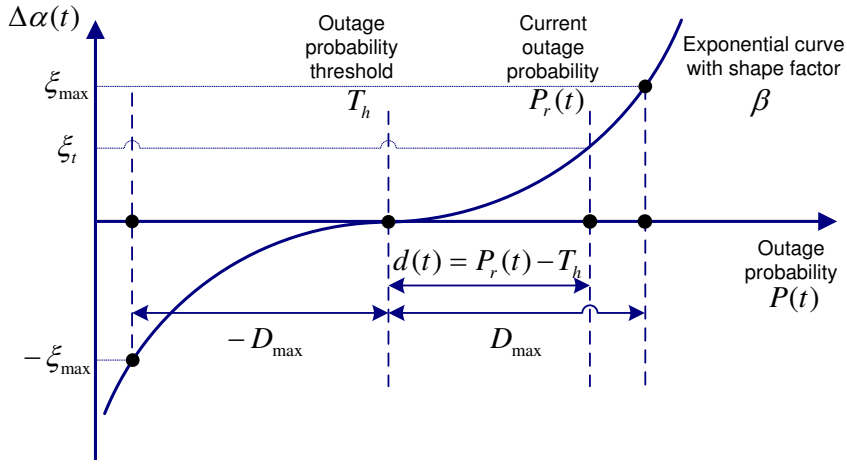



Fig. 4. An exponentially smoothed curve with respect to QoS satisfaction is applied to adjust the estimated amount of bandwidth in the class scheduler

- **Step 1:** At each scheduling instant, calculate the backlogged traffic $B_{\text{rtPS}}(t)$, the average modulation efficiency $\overline{\mu}_{\text{rtPS}}(t)$, and the current delay outage probability $\text{P}_\text{r}(t)$. Update the value of $\alpha(t)$:

$$\alpha(t) = \begin{cases} \min\{\alpha(t-1) + \Delta\alpha(t), \alpha_{\max}\} & \text{if } \text{P}_\text{r}(t) \geq T_\text{h} \\ \max\{\alpha(t-1) + \Delta\alpha(t), \alpha_{\min}\} & \text{if } \text{P}_\text{r}(t) < T_\text{h} \end{cases} \tag{12}$$

  where $\alpha_{\max}$ and $\alpha_{\min}$ are the maximum and minimum values of $\alpha(t)$, respectively.
- **Step 2:** Calculate the estimated bandwidth for rtPS class according to Exp. (10).

For nrtPS class, the bandwidth estimation procedure is the same as in rtPS class, except the definition of the backlogged traffic and the outage probability. Here we take the total number of virtual tokens associated with each queue as a measure for the backlogged traffic $B_{\text{nrtPS}} = \sum_{i\in\{\text{nrtPS}\}} v_i(t)$, where $v_i(t)$ is the number of virtual tokens in bucket $i$ at time $t$. $\text{P}_\text{r}(t)$ in nrtPS is the throughput outage probability.

The proposed adaptive bandwidth distribution algorithm works as follows: At each scheduling instant, the ARA first allocates the amount of required bandwidth to UGS class ($N_{\text{UGS}}$) and a minimum amount of bandwidth to BE class ($N_{\text{BE}}^{\min}$). Once the ARA has estimated the amount of required bandwidth for rtPS and nrtPS classes (i.e., $E_{\text{rtPS}}$ and $E_{\text{nrtPS}}$), it checks the remaining bandwidth. If the remaining bandwidth is larger than the estimated sum of $E_{\text{rtPS}}$ and $E_{\text{nrtPS}}$, the ARA first allocates $E_{\text{rtPS}}$ and $E_{\text{nrtPS}}$ to rtPS and nrtPS class schedulers respectively. Then the residual bandwidth is distributed among rtPS, nrtPS and BE class schedulers proportional to their queue size $Q_{\text{rtPS}}$, $Q_{\text{nrtPS}}$, and $Q_{\text{BE}}$. Otherwise, if the remaining bandwidth is smaller than the estimated sum of $E_{\text{rtPS}}$ and $E_{\text{nrtPS}}$, the ARA first allocates $E_{\text{rtPS}}$ to rtPS class, then the residual bandwidth is allocated to nrtPS class. It is worth mentioning that in order to satisfy proportional fairness among class schedulers, each class scheduler is reserved a minimum amount of bandwidth. A detailed description of the proposed algorithm is listed in pseudocode 2.

**Algorithm 2** Adaptive bandwidth distribution algorithm for the Aggregate Resource Allocator (ARA)

Set initial $N_{\text{total}}$ at the beginning of each frame
$N_{\text{UGS}} \leftarrow \sum_{i \in \{\text{UGS}\}} \theta_i$
Set $N_{\text{BE}}^{\min}$
$N_{\text{residual}} \leftarrow N_{\text{total}} - N_{\text{UGS}} - N_{\text{BE}}^{\min}$
Estimate the number of time slots required in rtPS class scheduler $E_{\text{rtPS}}$ by Exp.(10)
Estimate the number of time slots required in nrtPS class scheduler $E_{\text{nrtPS}}$ by Exp. (10)
**if** $N_{\text{residual}} \geq (E_{\text{rtPS}} + E_{\text{nrtPS}})$ **then**
    $N_{\text{residual}} \leftarrow N_{\text{residual}} - E_{\text{rtPS}} - E_{\text{nrtPS}}$
    $N_{\text{rtPS}} \leftarrow E_{\text{rtPS}} + N_{\text{residual}} \cdot \frac{Q_{\text{rtPS}}}{Q_{\text{rtPS}}+Q_{\text{nrtPS}}+Q_{\text{BE}}}$
    $N_{\text{nrtPS}} \leftarrow E_{\text{nrtPS}} + N_{\text{residual}} \cdot \frac{Q_{\text{nrtPS}}}{Q_{\text{rtPS}}+Q_{\text{nrtPS}}+Q_{\text{BE}}}$
    $N_{\text{BE}} \leftarrow N_{\text{BE}}^{\min} + N_{\text{residual}} \cdot \frac{Q_{\text{BE}}}{Q_{\text{rtPS}}+Q_{\text{nrtPS}}+Q_{\text{BE}}}$
    **if** $N_{\text{rtPS}} < N_{\text{rtPS}}^{\min}$ or $N_{\text{nrtPS}} < N_{\text{nrtPS}}^{\min}$ **then**
        Adjust the values of $N_{\text{rtPS}}$, $N_{\text{nrtPS}}$ and $N_{\text{BE}}$ so that $N_{\text{rtPS}} \geq N_{\text{rtPS}}^{\min}$ and $N_{\text{nrtPS}} \geq N_{\text{nrtPS}}^{\min}$
    **end if**
**else**
    $N_{\text{rtPS}} \leftarrow \min\{E_{\text{rtPS}}, N_{\text{residual}}\}$
    $N_{\text{nrtPS}} \leftarrow N_{\text{residual}} - N_{\text{rtPS}}$
    $N_{\text{BE}} \leftarrow N_{\text{BE}}^{\min}$
    **if** $N_{\text{rtPS}} < N_{\text{rtPS}}^{\min}$ or $N_{\text{nrtPS}} < N_{\text{nrtPS}}^{\min}$ **then**
        Adjust the values of $N_{\text{rtPS}}$ and $N_{\text{nrtPS}}$ so that $N_{\text{rtPS}} \geq N_{\text{rtPS}}^{\min}$ and $N_{\text{nrtPS}} \geq N_{\text{nrtPS}}^{\min}$
    **end if**
**end if**

## 5. Connection Admission Control

Connection admission control is a key component of QoS provisioning for wireless systems supporting multiple types of applications. It aims at maintaining the delivered QoS to different users at the target level by limiting the number of ongoing connections in the system. In this chapter, we propose a measurement-based approach as our CAC policy, of which a CAC decision is made depending on the current resource utilization of the network. When a new connection request is initiated, it informs the CAC module of the connection type (i.e., rtPS or nrtPS), the traffic parameters (i.e., arrival rate and service rate), and the QoS requirements (i.e., maximum delay or minimum throughput). Then the CAC module estimates the required amount of bandwidth $\Delta N$ to accommodate the incoming connection and performs a CAC decision based on the following conditions.

For UGS connections, as the transmission mode and the number of time slots allocated per connection per frame are negotiated in the initial service access phase and are fixed during the whole service time, a simple threshold-based CAC is applied:

$$N_{\text{UGS}}^{\text{current}} + \Delta N_{\text{UGS}} \leq N_{\text{UGS}}^{\max} \tag{13}$$

where $N_{\text{UGS}}^{\text{current}}$ is the number of time slots occupied by the ongoing UGS connections, and $N_{\text{UGS}}^{\max}$ is the maximum number of time slots that can be allocated to the UGS class scheduler. If this condition is satisfied, the incoming connection is accepted; otherwise, it is rejected.

For rtPS and nrtPS connections, when a new connection arrives, the CAC module interacts with ARA in the DRA module and gets the current estimated bandwidth occupied by the ongoing rtPS and nrtPS connections, $\overline{E}_{\text{rtPS}}(t)$ and $\overline{E}_{\text{nrtPS}}(t)$, which are the exponential moving average of $E_{\text{rtPS}}(t)$ and $E_{\text{nrtPS}}(t)$ mentioned in Eqn. (10). If the sum of the estimated bandwidth used by the ongoing rtPS and nrtPS connections ($\overline{E}_{\text{rtPS}}(t)$, $\overline{E}_{\text{nrtPS}}(t)$) and the estimated bandwidth to be used by the incoming connection ($\Delta N_{\text{rtPS}}$ or $\Delta N_{\text{nrtPS}}$) is larger than a predefined upper threshold, the incoming connection is rejected; otherwise, the connection is accepted with certain probability depending on the estimated bandwidth usage and the connection priority. Specifically, when the estimated bandwidth occupancy is high or the priority of the incoming connection is low, the acceptance probability is small, and vice versa. A detailed description of the proposed CAC algorithm for rtPS connections is listed in pseudocode 3, where $N_{\text{th}}^{\max}$ and $N_{\text{th}}^{\min}$ are the maximum and minimum capacity threshold respectively, and $\rho_{\text{rtPS}} \in (0, 1]$ is a parameter that is used to differentiae class priorities. The same CAC algorithm is applied for nrtPS connections.

**Algorithm 3** Connection admission control algorithm for rtPS connections

**if** $\overline{E}_{\text{rtPS}}(t) + \overline{E}_{\text{nrtPS}}(t) + \Delta N_{\text{rtPS}} > N_{\text{th}}^{\max}$ **then**
  Reject the incoming connection
**else if** $\overline{E}_{\text{rtPS}}(t) + \overline{E}_{\text{nrtPS}}(t) + \Delta N_{\text{rtPS}} < N_{\text{th}}^{\min}$ **then**
  Accept the incoming connection with probability $\rho_{\text{rtPS}}$
**else**
  Accept the incoming connection with probability $\rho_{\text{rtPS}} \cdot \frac{N_{\text{th}}^{\max} - (\overline{E}_{\text{rtPS}}(t) + \overline{E}_{\text{nrtPS}}(t) + \Delta N_{\text{rtPS}})}{N_{\text{th}}^{\max} - N_{\text{th}}^{\min}}$
**end if**

For BE connections, they are always accepted since they do not impose any QoS constraints.

## 6. Simulation Results and Discussions

To evaluate the performance of the proposed downlink resource management framework for QoS scheduling in OFDMA based WiMAX networks, a system-level simulation is performed in OPNET.

### 6.1 System Model

We consider the downlink of a single-cell IEEE 802.16 system with OFDMA TDD operation. The cell radius is 2 km, where subscriber stations are randomly placed in the cell with uniform distribution. The total bandwidth is set to be 5 MHz, which is divided into 10 subchannels. The BS transmit power is set to 20W (43 dBm) which is evenly distributed among all subchannels. The duration of a frame is set to be 1 ms so that the channel quality of each connection remains almost constant within a frame, but may vary from frame to frame. The propagation model is derived from IEEE 802.16 SUI channel model (30). Path loss is modeled according to terrain Type A suburban macro-cell. Large-scale shadowing is modeled by log-normal distribution with zero mean and standard deviation of 8 dB. Small-scale shadowing is modeled by Rayleigh fading.

Table 1 summarizes the system parameters used in the simulation. We assume that all MAC PDUs are transmitted and received without errors and the transmission delay is negligible.

| Parameters | Value |
|---|---|
| System | OFDMA/TDD |
| Central frequency | 3500 MHz |
| Channel bandwidth | 5 MHz |
| Number of subchannels | 10 |
| User distribution | Uniform |
| Beam pattern | Omni-directional |
| Cell radius | 2 km |
| Frame duration | 1 ms |
| BS transmit power | 20 W |
| Thermal noise density | −174 dBm/Hz |
| Propagation model | 802.16 SUI-5 Channel model |
| Maximum MAC PDU size | 256 bytes |

Table 1. A summary of system parameters

| Modulation scheme | Coding rate | bits/symbol | Target SNR for 1% PER (dB) |
|---|---|---|---|
| BPSK | 1/2 | 0.5 | 1.5 |
| QPSK | 1/2 | 1 | 6.4 |
| QPSK | 3/4 | 1.5 | 8.2 |
| 16QAM | 1/2 | 2 | 13.4 |
| 16QAM | 3/4 | 3 | 16.2 |
| 64QAM | 1/2 | 4 | 21.7 |
| 64QAM | 3/4 | 4.5 | 24.4 |

Table 2. Modulation and coding schemes for 802.16 (27)

The modulation order and coding rate in AMC is determined by the instantaneous SNR of each user on each subchannel. We follow the AMC table shown in Table 2, which specifies the minimum SNR required to meet a target packet error rate, e.g., 1%.

### 6.2 Traffic Model

In the simulation, different types of traffic sources are generated: VoIP, videoconference, and Internet traffic. VoIP and videoconference are served in UGS class and rtPS class, respectively. Internet traffic is served in nrtPS class and BE class. Each user alternates between the states of idle and busy, which are both exponentially distributed, and generates one or several traffic types independently during the busy period. VoIP traffic is modeled as a two-state Markov ON/OFF source (16). A videoconference consists of a VoIP source and a video source (16). Internet traffic can be web browsing that requires large bandwidth and generates bursty data of variable size. We apply the Web browsing model for the Internet traffic (17). A summary of traffic parameters for different traffic types are listed in Table 3.

### 6.3 Performance Evaluation

Since the performance of fixed bandwidth allocation for UGS connections is well defined by the standard and BE connections do not have any specific QoS requirements, here we only focus on the performance evaluation of rtPS and nrtPS connections. The delay constraint for

| Type | Characteristics | Distribution | Parameters |
|---|---|---|---|
| VoIP | ON period | Exponential | Mean = 1.34 sec |
| VoIP | OFF period | Exponential | Mean = 1.67 sec |
| VoIP | Packet size | Constant | 66 bytes |
| VoIP | Inter-arrival time between packets | Constant | 20 ms |
| Video | Packet size | Log-normal | Mean = 4.9 bytes<br>Std. dev. = 0.75 bytes |
| Video | Inter-arrival time between packets | Normal | Mean = 33 ms<br>Std. dev. = 10 ms |
| Web | Reading time between sessions | Exponential | Mean = 5 sec |
| Web | Number of packets within a packet call | Geometric | Mean = 25 packets |
| Web | Inter-arrival time between packets | Geometric | Mean = 0.0277 sec |
| Web | Packet size | Truncated Pareto | $k$ = 81.5 bytes<br>$\alpha$ = 1.1<br>$m$ = 2 M bytes |

Table 3. A summary of traffic parameters

rtPS class is 50 ms and the minimum throughput constraint for nrtPS class is 100 Kbits/sec. The outage probabilities for both rtPS and nrtPS classes should be less than 3%.

#### 6.3.1 Effectiveness of the proposed resource management framework

We start the analysis by evaluating the effectiveness of the proposed downlink resource management framework when both rtPS and nrtPS classes are employed. We first set up a scenario with a fixed number of 45 SSs, each of which has a connection pair consisting of a rtPS connection and a nrtPS connection. Fig. 5 shows the performance of the rtPS class scheduler over time, where the first subplot indicates the estimated and the granted bandwidth to rtPS class scheduler, the second and third subplots depict packet delay and delay outage probability respectively. It is clearly shown in the figure that the estimated bandwidth adaptively and closely follows the pattern of the delay outage probability. This is in accordance with our expectation and can be explained as follows: In designing the bandwidth distribution algorithm of the ARA, the proposed algorithm first pre-estimates the required amount of bandwidth based on the backlogged traffic and the average modulation efficiency. Then an adjustment of pre-estimation is performed with respect to QoS satisfaction (e.g., delay outage probability in rtPS class). Specifically, when the class scheduler experiences good QoS satisfaction, i.e., the delay outage probability is below the threshold, the amount of the estimated bandwidth is decreased to save for other classes. Otherwise, the amount of the estimated bandwidth is increased to guarantee the required QoS. We can see from the figure that the delay outage probability fluctuates around the threshold. One advantage of the proposed bandwidth distribution algorithm is that instead of allocating the available bandwidth to each class scheduler in a static manner, the proposed algorithm adaptively allocates a "necessary" amount of bandwidth to each class scheduler to intentionally keep its outage probability around a predefined threshold, which is 3% in our case. Similar phenomenon can be observed in Fig. 6, where

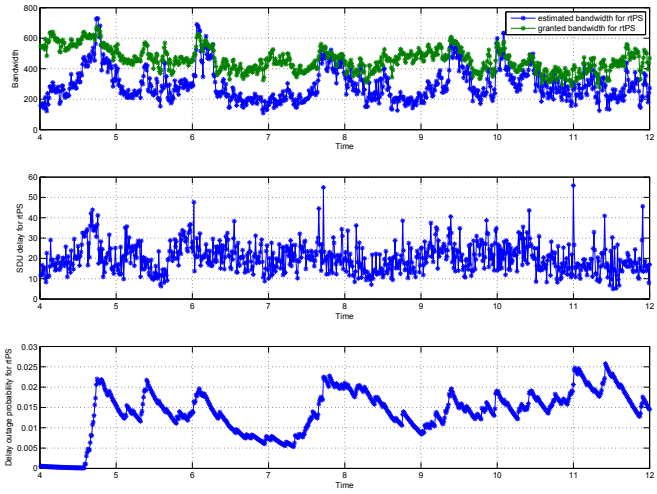



Fig. 5. Performance of rtPS class scheduler over time (estimated bandwidth & granted bandwidth, packet delay, and delay outage probability), with 45 SSs

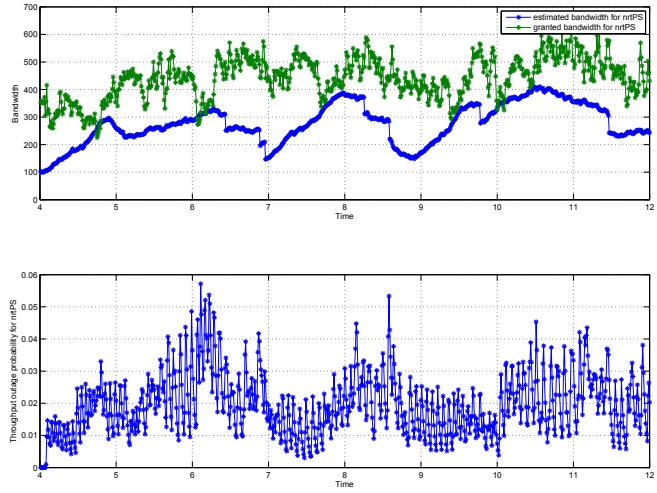



Fig. 6. Performance of nrtPS class scheduler over time (estimated bandwidth & granted bandwidth, and minimum throughput outage probability), with 45 SSs

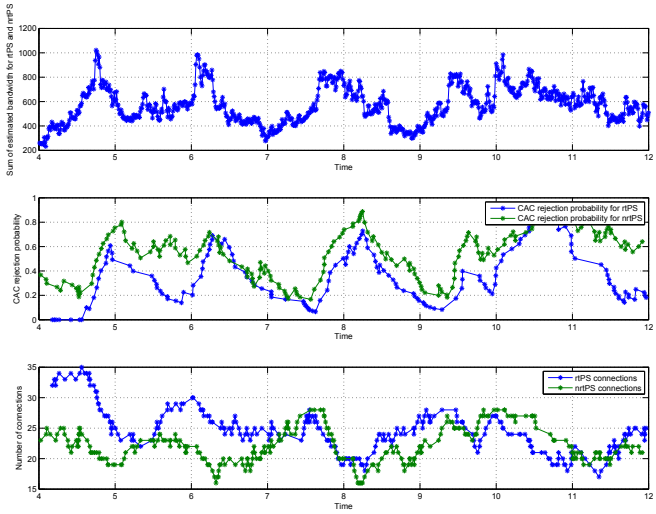



Fig. 7. CAC performance for rtPS and nrtPS connections over time, with 45 SSs

the first subplot indicates the estimated and the granted bandwidth to nrtPS class scheduler and the second subplot depicts the throughput outage probability. Again, the estimated bandwidth shows a strong correlation with the throughput outage probability in nrtPS class. Furthermore, the curve of the granted bandwidth in both rtPS and nrtPS classes follows the pattern of the estimated bandwidth, but is higher than the estimated bandwidth. This is because in the proposed bandwidth distribution algorithm, when the available bandwidth is larger than the estimated sum of $E_{\text{rtPS}}$ and $E_{\text{nrtPS}}$, the ARA first allocates $E_{\text{rtPS}}$ and $E_{\text{nrtPS}}$ to rtPS and nrtPS class schedulers respectively. Then the residual bandwidth is distributed among rtPS, nrtPS and BE class schedulers proportional to their queue size. To summarize, the proposed two-level hierarchical scheduler can distribute the available bandwidth among class schedulers and schedule packets in each service class efficiently and adaptively.

Fig. 7 shows the performance of CAC for rtPS and nrtPS classes over time, where the first subplot indicates the sum of the estimated bandwidth for rtPS and nrtPS classes, the second subplot depicts the connection rejection probability for rtPS and nrtPS classes, and the third subplot shows the number of ongoing rtPS and nrtPS connections. It is obvious from the figure that the CAC decision of an incoming connection attempt depends on the current state of the network (i.e., the total amount of bandwidth used by the ongoing connections plus the estimated bandwidth to be used by the incoming connection). The rejection probability is proportional to the resource usage by the ongoing connections. The number of ongoing rtPS and nrtPS connections are inverse-proportional to the CAC rejection probability as expected. Furthermore, both the acceptance probability and the number of ongoing connections for rtPS class are relatively higher than those for nrtPS class. This is because rtPS connections are given a higher priority value than nrtPS connections in our scenario, which results in higher acceptance probability. In general, the simulation results confirmed that the proposed

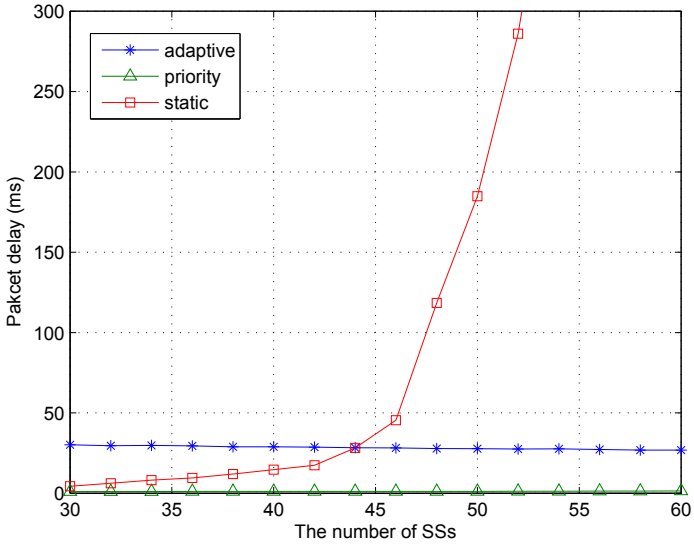



Fig. 8. Average packet delay in rtPS under different bandwidth distribution schemes in the ARA

measurement-based CAC strategy can effectively limit the number of ongoing rtPS and nrtPS connections, thus preventing the system capacity from being overused.

#### 6.3.2 Comparisons of different bandwidth distribution algorithms in ARA

Since the bandwidth distribution algorithm in ARA is critical on the performance of class schedulers, next we evaluate and compare the performance of the proposed adaptive estimation-based bandwidth distribution scheme with two other conventional bandwidth distribution schemes, e.g., strict priority-based scheme and static scheme. For comparisons, the CAC module in the framework is disabled so that all connection requests will be accepted. Some of the parameters used in the proposed adaptive scheme are set as follows: $\gamma = 0.02$, $\beta = 80$, $T_{\mathrm{h}} = 0.03$, $D_{\max} = 0.015$, $\xi = 0.02$, $\alpha_{\min} = 0.5$, and $\alpha_{\max} = 1.5$. In the static scheme, the proportions of the total available bandwidth allocated to rtPS and nrtPS class schedulers are set to be 50% and 50%, respectively.

Fig. 8 shows the average packet delay in rtPS class versus the number of SSs under different bandwidth distribution schemes. The average packet delay of the proposed adaptive scheme and the priority-based scheme remains almost constant regardless of the number of SSs in the system, while in the static scheme, the average packet delay increases sharply when the number of SSs is above 45. Similar phenomenon can be observed for the delay outage probability shown in Fig. 9. From Fig. 8 & 9, we can obviously see that both the priority-based scheme and the proposed adaptive scheme can meet the QoS requirements in rtPS class in terms of the average packet delay and the delay outage probability. On the other hand, the static scheme can not adapt to the traffic load, thus is not suitable for load varying systems.

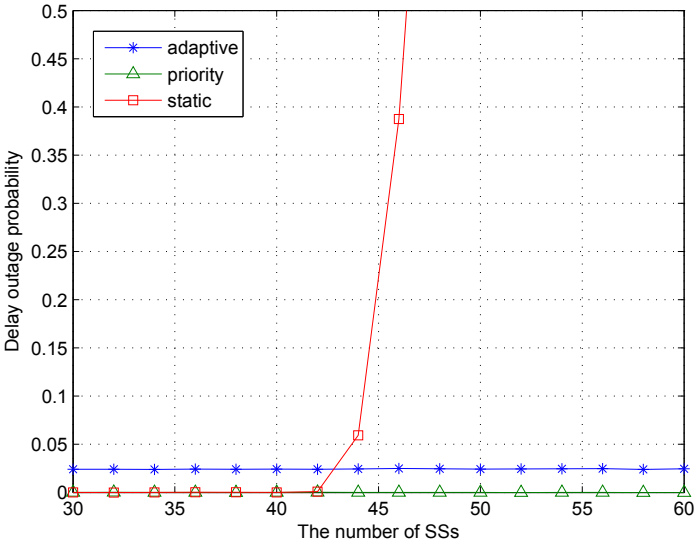



Fig. 9. Delay outage probability in rtPS under different bandwidth distribution schemes in the ARA

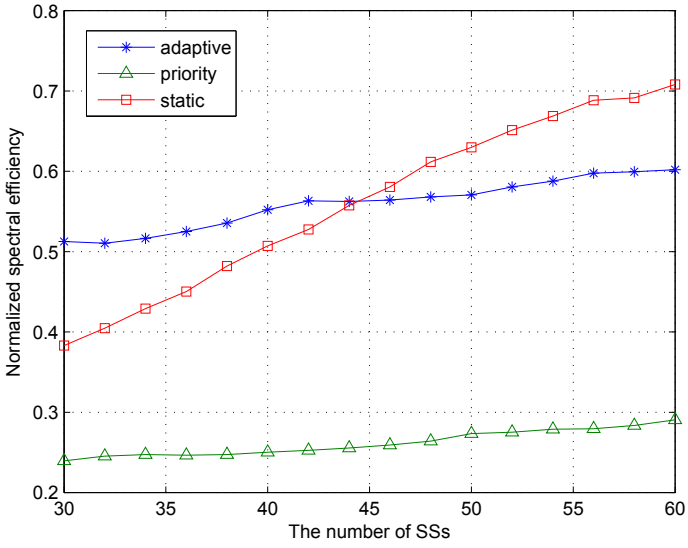



Fig. 10. Normalized spectral efficiency in rtPS under different bandwidth distribution schemes in the ARA

The advantage of the proposed adaptive scheme over the priority-based scheme is depicted in Fig. 10, which shows the normalized spectral efficiency, defined as the ratio between the achieved modulation order and the highest modulation order, in rtPS class under different bandwidth distribution schemes. It is shown that the normalized spectral efficiency of the adaptive scheme is about two times than that of the priority-based scheme. This is because in the priority-based scheme, the ARA distributes the available bandwidth to each class scheduler following strict class priorities (e.g., rtPS→nrtPS in our scenario). Therefore, high priority class may get more bandwidth than it is actually needed. As a consequence, much of the allocated bandwidth is utilized with low spectral efficiency. However, in the proposed adaptive scheme, the ARA only allocates a "necessary" amount of bandwidth to each class scheduler with the objective to keep its outage probability around a predefined threshold, so that the performance of the class scheduler can be maximized. By doing so, the channel and QoS aware class scheduler has more chance to serve a user in good channel condition without sacrificing the QoS requirement (as we can see in Fig. 8 that the average packet delay in the adaptive scheme is higher than that of the priority-based scheme, but is well kept below a threshold), thus significantly increases the spectral efficiency. We also notice in Fig. 10 that the spectral efficiency of the static scheme becomes higher than the adaptive scheme when the number of SSs is more than 44. This is because in the static scheme, the ARA allocates a fixed amount of bandwidth to each class scheduler. While in the adaptive scheme, the bandwidth is allocated dynamically according to the actual needs. When the number of SSs is less than 44, the bandwidth allocated to rtPS class scheduler in the static scheme is more than needed, thus the allocated bandwidth is utilized with low spectral efficiency. However, when the number of SSs is more than 44, the bandwidth allocated to rtPS class scheduler in the static scheme is less than the adaptive scheme, which results in a higher modulation efficiency due to the scheduling mechanism of EXP algorithm.

Fig. 11 shows the throughput in nrtPS class versus the number of SSs under different bandwidth distribution schemes. When the system is underloaded (the number of SSs is less than 34), the throughput in all three schemes increases proportional to the number of SSs. On the other hand, there is a big difference among different schemes when the system is overloaded. The throughput of the priority-based scheme is significantly lower than the other two schemes. Specifically, in the priority-based scheme, the throughput is inverse proportional to the number of SSs. In the static scheme, the throughput increases proportional to the number of SSs when there are less than 42 SSs. After that point, the throughput remains on a steady level regardless of the number of SSs. While in the adaptive scheme, the throughput keeps increasing proportional to the number of SSs when there are less than 52 SSs. After that point, the throughput decreases as the number of SSs increases. It can be explained as follows: In the adaptive scheme, the ARA tries to balance the bandwidth distribution among different class schedulers to increase the spectral efficiency while satisfying the QoS requirements. When the system is unsaturated, i.e., the total available bandwidth is larger than the estimated sum (the number of SSs is less than 52), the ARA first allocates the estimated amount of bandwidth to each class scheduler, then the residual bandwidth is distributed among all class schedulers according to their queue size. Therefore, the throughput in each service class increases proportional to the number of SSs. When saturation occurs, i.e., the total available bandwidth is smaller than the estimated sum (the number of SSs is larger than 52), the ARA allocates the estimated amount of bandwidth to the class schedulers from high priority to low priority, which means that rtPS class gets the estimated amount of bandwidth first, followed by nrtPS class. Thus the throughput in service class with low priority (nrtPS) decreases as the band-

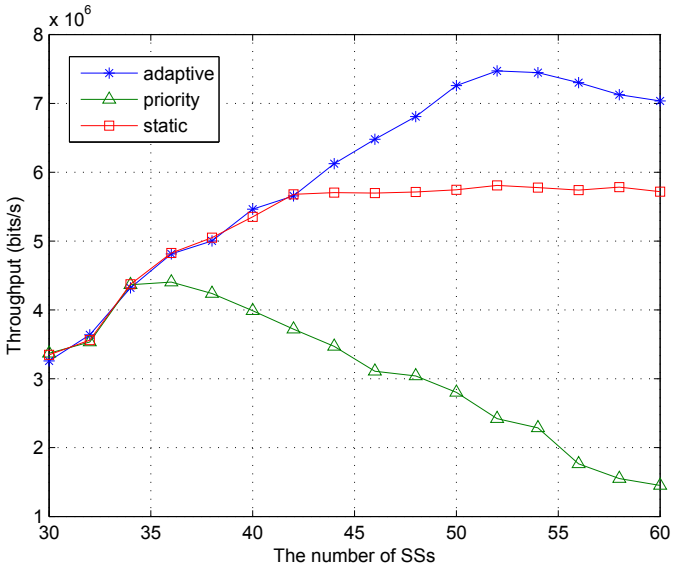



Fig. 11. Average throughput in nrtPS under different bandwidth distribution schemes in the ARA

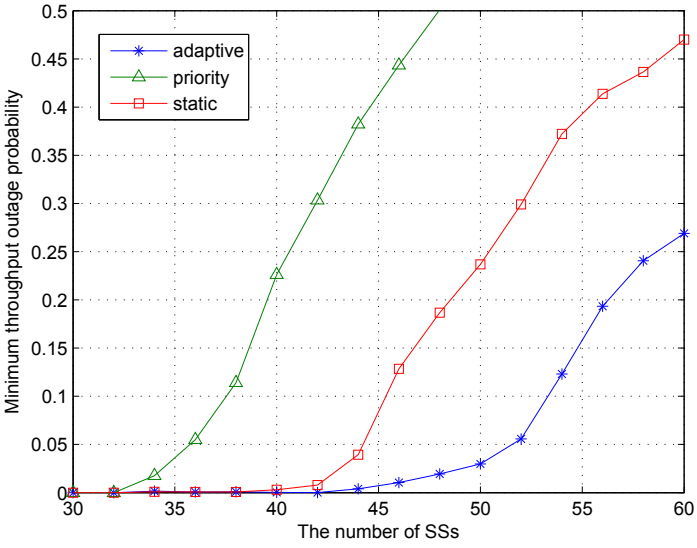



Fig. 12. Throughput outage probability in nrtPS under different bandwidth distribution schemes in the ARA

width balancing scheme favors service class with high priority (rtPS) when congestion occurs. However, in the priority-based scheme, nrtPS class experiences severe bandwidth starvation due to the reason that much of the bandwidth allocated to rtPS class is utilized with low spectral efficiency, thus the residual bandwidth allocated to nrtPS class is not sufficient to serve nrtPS connections. While in the static scheme, the amount of bandwidth allocated to each class is fixed, therefore the throughput in each class reaches at a steady level when the system is overloaded.

Fig. 12 shows the throughput outage probability in nrtPS class versus the number of SSs under different bandwidth distribution schemes. It is obvious that the proposed adaptive scheme outperforms over the other two schemes. The maximum number of supportable SSs under a predefined 3% outage probability in priority-based, static and adaptive scheme a re 34, 44 and 50 respectively.

#### 6.3.3 Comparisons of different scheduling algorithms in class schedulers

In the results discussed so far, we have seen that the proposed adaptive bandwidth distribution scheme for the ARA has better performance than the priority-based scheme and static scheme. In order to evaluate the performance of the proposed priority-based scheduling algorithm for rtPS and nrtPS class schedulers, we have included the simulation results of two other conventional scheduling algorithms, e.g., maximum SNR and proportional fair (PF). For comparisons, the static bandwidth distribution scheme is employed in the ARA. The proportions of the total available bandwidth allocated to rtPS and nrtPS class schedulers are set to be 60% and 40%, respectively. The CAC module is disabled so that all connection requests will be accepted.

Fig. 13 shows the average packet delay in rtPS class versus the number of SSs under different scheduling algorithms. When the number of SSs is below 48, the average packet delay of the proposed priority-based scheme increases marginally and it is well kept below the maximum allowable delay, which is 50 ms in our scenario. After that point, the system is overloaded and the average packet delay increases sharply. Similar phenomenon of the proposed scheme can be observed for the delay outage probability shown in Fig. 14. However, the average packet delay of the PF scheme and the MAX-SNR scheme is much larger compared to our proposed scheme, which consequently results in a higher delay outage probability when the number of SSs is below 48. Furthermore, it can be seen from Fig. 14 that when the number of SSs is above 48, the delay outage probability of the proposed scheme increases rapidly to one, which means that the system is overloaded and almost no rtPS connections can maintain the required delay constraint. On the other hand, some rtPS connections in the PF and MAX-SNR schemes can still maintain the required delay constraint as the delay outage probabilities in these two schemes increase steadily with respect to the number of SSs. This is because in the proposed scheme, it not only takes the instantaneous channel conditions, but also the delay requirement into consideration when scheduling packets. rtPS connections with larger packet delay are assigned higher priorities in an effort to average out the packet delay among all rtPS connections. As a result, each rtPS connection will have similar average packet delay regardless of its channel conditions. When the system is overloaded, congestion occurs and all rtPS connections will experience bandwidth starvation, which results in a sharp increase of the average packet delay and the delay outage probability. However, in the PF and MAX-SNR schemes, the scheduler selects a connection for transmission only based on instantaneous channel conditions. As a consequence, connections with good channel conditions will always experience very short delay at the cost of bandwidth starvation for connections with poor

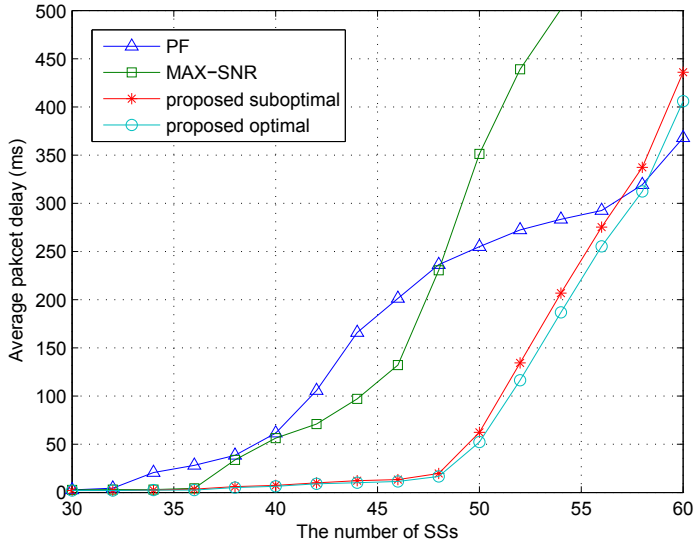



Fig. 13. Average packet delay in rtPS under different scheduling algorithms, with static bandwidth distribution in the ARA

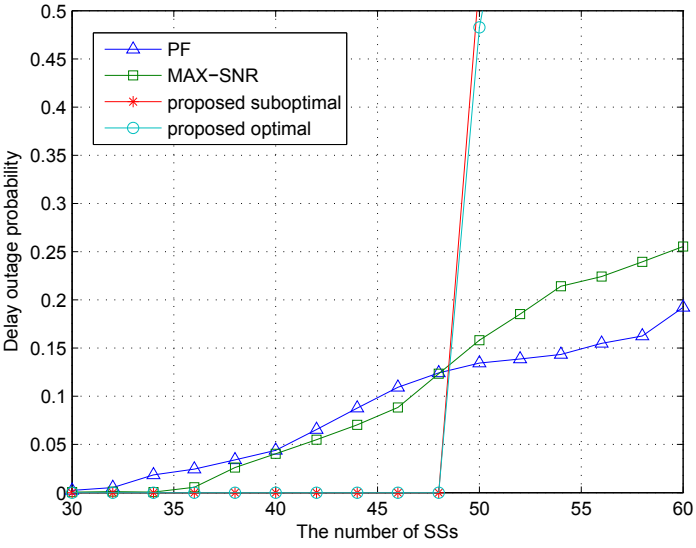



Fig. 14. Delay outage probability in rtPS under different scheduling algorithms, with static bandwidth distribution in the ARA

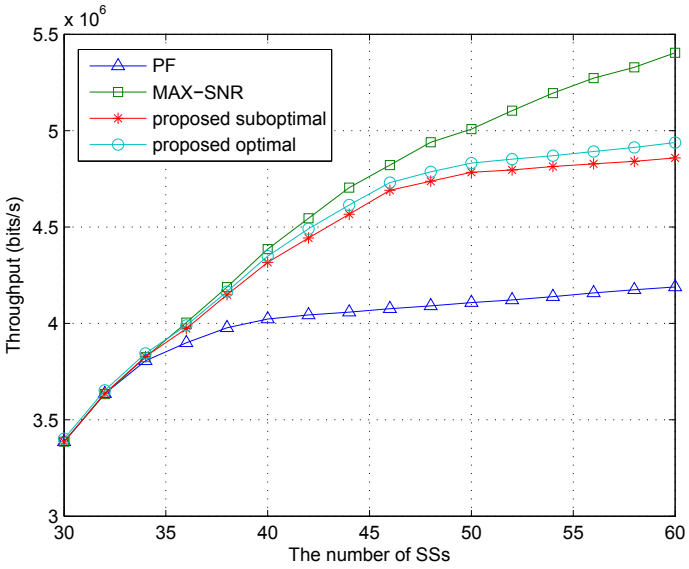



Fig. 15. Average throughput in nrtPS under different scheduling algorithms, with static bandwidth distribution in the ARA

channel conditions. Therefore, the delay outage probability in the PF and MAX-SNR schemes increases much more smoothly compared to the proposed scheme when the number of SSs is above 48.

Fig. 14 shows the delay outage probability in rtPS class versus the number of SSs under different scheduling algorithms. It is obvious that the proposed scheme outperforms over the other conventional schemes. The maximum number of supportable SSs under a predefined 3% outage probability in PF, MAX-SNR and the proposed scheme are 38, 38, and 48 respectively.

Fig. 15 shows the throughput in nrtPS class versus the number of SSs under different scheduling algorithms. The throughput of the MAX-SNR scheme achieves the highest value among all schemes. It increases proportional to the number of SSs. In the proposed scheme, the throughput increases proportional to the number of SSs when there are less than 50 SSs. After that point, the throughput remains on a steady level regardless of the number of SSs. While in the PF scheme, the throughput is significantly lower than the other schemes. It can be explained as follows: In the MAX-SNR scheme, the scheduler simply selects the connection with the best CSI for transmission. When the number of SSs increases, the scheduler has more chance to serve a SS in good channel conditions (multi-user diversity gain) which results in a high throughput. That's why the spectral efficiency of the MAX-SNR scheme increases with respect to the number of SSs as shown in Fig. 17. In the proposed scheme, both the CSI and the QoS constraints are taken into account to guarantee the required QoS performance (i.e., a minimum throughput of 100Kbpbs for each nrtPS connection). When the system is underloaded (the number of SSs is less than 50), the bandwidth is large enough to satisfy the QoS requirements of all connections and the scheduling criterion mainly concerns with the CSI of

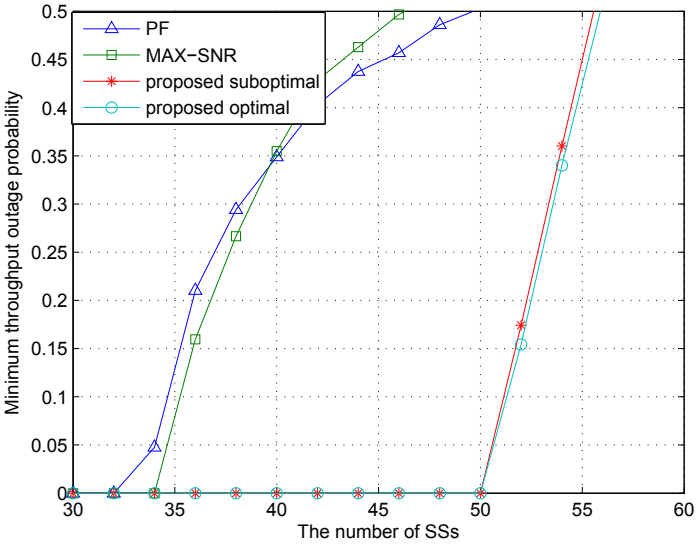



Fig. 16. Throughput outage probability in nrtPS under different scheduling algorithms, with static bandwidth distribution in the ARA

each connection. As a result, the throughput as well as the spectral efficiency increases proportional to the number of SSs. However, when the system is overloaded (the number of SSs is above 50), the bandwidth is not sufficient to satisfy the QoS requirements of all connections. Thus congestion occurs and the throughput reaches at a steady level. When congestion occurs, the proposed algorithm tends to put more weight on the QoS constraint than the CSI in an effort to provide equal opportunities of QoS satisfaction among all nrtPS connections. In other words, the throughput of each nrtPS connection in the proposed scheme decreases proportionally to the number of SSs when the system is overloaded. That explains a sharp increase of the throughput outage probability shown in Fig. 16. In the PF scheme, the throughput is relatively low due to the reason that the spectral efficiency is significantly lower than the MAX-SNR and the proposed schemes.

Fig. 16 shows the throughput outage probability in nrtPS class versus the number of SSs under different scheduling algorithms. It is obvious that the proposed scheme outperforms over the other conventional schemes. The maximum number of supportable SSs under a predefined 3% outage probability in PF, MAX-SNR and the proposed scheme are 34, 34, and 50 respectively.

Fig. 17 depicts the normalized spectral efficiency under different scheduling algorithms. It can be seen that the MAX-SNR scheme achieves the highest spectral efficiency due to the reason that in the the MAX-SNR, the connection with the best CSI is selected for transmission. The proposed scheme can also achieve a relatively high spectral efficiency as it takes both the channel condition as well as the QoS constraints into account when scheduling packets. While

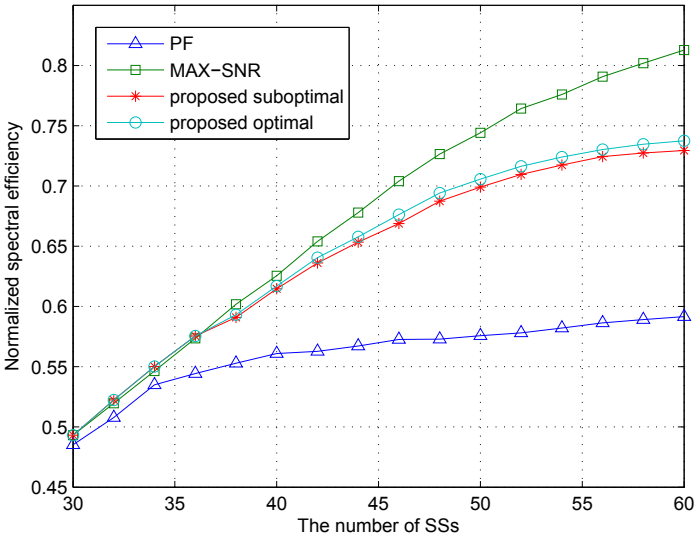



Fig. 17. Normalized spectral efficiency in rtPS and nrtPS under different scheduling algorithms, with static bandwidth distribution in the ARA

the spectral efficiency in the PF scheme is relatively low compared to the MAX-SNR and the proposed schemes.
From the above figures, we can also see that the performance of the proposed suboptimal scheme is close to the optimal scheme, but with considerably low computation complexity.

## 7. Conclusions

This chapter addresses the problem of downlink radio resource management in OFDMA-based WiMAX systems. Specifically, we have proposed a downlink resource management framework which consists of a DRA module and a CAC module. We developed a two-level hierarchical scheduler for the DRA module, which is comprised of an aggregate resource allocator and four class schedulers. The advantage of this two-level hierarchical resource allocation architecture is that the algorithms for the allocator and the schedulers can be developed independently, yet there is still sufficient coupling between them as the allocator is aware of the performance of the schedulers. A priority-based scheduling algorithm for rtPS and nrtPS class schedulers and an adaptive bandwidth distribution algorithm for the aggregate resource allocator are proposed. A measurement-based admission control strategy is proposed for the CAC module, which works cooperatively with the DRA module when admission decisions are made. Simulation results show that the proposed framework can work efficiently and adaptively to various kinds of traffic loads, and can satisfy the diverse QoS requirements in each service class with high spectral efficiency.

## 8. References


[1] Bolcskel H., Paulraj A.J., Hari K.V.S., Nabar R.U., and Lu W.W.: *Fixed broadband wireless access: state of the art, challenges, and future directions*, IEEE Communications Magazine, Vol.39 Issue.1 , pp. 100–108, May. 2001.

[2] IEEE 802.16-2004, *IEEE standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, 2004

[3] Ali S.H., Ki-Dong Lee, and Leung V.C.M.: *Dynamic resource allocation in OFDMA wireless metropolitan area networks*, IEEE Wireless Communications, Vol.14 Issue.1, pp. 6–13, 2007.

[4] Bohge M., Gross J., Wolisz A., and Meyer M.: *Dynamic resource allocation in OFDM systems: an overview of cross-layer optimization principles and techniques*, IEEE Network, Vol.21 Issue.1, pp. 53–59, 2007.

[5] Niyato D., and Hossain E.: *Radio resource management games in wireless networks: an approach to bandwidth allocation and admission control for polling service in IEEE 802.16*, IEEE Wireless Communications, Vol.14 Issue.1, pp. 27–35, 2007.

[6] Badia L., Baiocchi A., Todini A., Merlin S., Pupolin S., Zanella A., and Zorzi M.: *On the impact of physical layer awareness on scheduling and resource allocation in broadband multicellular IEEE 802.16 systems*, IEEE Wireless Communications, Vol.14 Issue.1, pp. 36–43, 2007.

[7] Ying J. Z., and Khaled B.: *Energy-Efficient MAC-PHY Resource Management with Guaranteed QoS in Wireless OFDM Networks*, ICC 2005, Vol.5, pp. 3127–3131, May. 2005.

[8] Ying J. Z., and Khaled B.: *Adaptive Resource Allocation and Scheduling for Multiuser Packet-based OFDM Networks*, ICC 2004, Vol.5, pp. 2949–2953, June. 2004.

[9] Alfedo T., Marco M., Andrea V. and Andrea B.: *A modular cross-layer scheduling and resource allocation architecture for OFDMA systems*, GLOBECOM 2006.

[10] Xing Z., En Z., Renshui Z., Shiming L., and Wenbo W.: *Adaptive Multiuser Radio Resource Allocation for OFDMA Systems*, GLOBECOM 2005, Vol.6, Dec. 2005.

[11] Xing Z., and Wenbo W.: *Multiuser frequency-time domain radio resource allocation in downlink OFDM systems: Capacity analysis and scheduling methods*, Computers and Electrical Engineering, Vol.32 Issue.1-3, pp. 118–134, 2006.

[12] Sang S. J., Dong G. J., and Wha S. J.: *Cross-layer Design of Packet Scheduling and Resource Allocation in OFDMA Wireless Multimedia Networks*, VTC 2006, Vol.1, pp. 309–313, 2006.

[13] Qingwen L., Xin W., and Georgios B.G.: *Cross-Layer Scheduler Design with QoS support for Wireless Access Networks*, 2nd International Conference on QoS in Heterogeneous Wired/Wireless Networks, 2005.

[14] Flavio D. A., Ibrahim H., Giovanni G. and Samuele G.: *Scheduling for Differentiated Traffic Types in HSDPA Cellular Systems*, Global Telecommunications Conference, Vol.1, 2005 Dec.

[15] Muhammad K., and Niclas W.: *Scheduling Algorithms for HS-DSCH in a WCDMA Mixed Traffic Scenario*, PIMRC, Vol.2, 2003 Sep.

[16] Claudio C., Luciano L. and Enzo M.: *Quality of Service Support in IEEE 802.16 Networks*, IEEE Network, Vol.20 Issue.2, pp. 50–55, 2006.

[17] Dong H. K., Byung H. R., and Chung G. K.: *Packet Scheduling Algorithm Considering a Minimum Bit Rate for Non-real-time Traffic in an OFDMA/FDD-Based Mobile Internet Access System*, ETRI Journal, Vol.26, Issue.1, pp. 48–52, 2004

[18] Matthew A., Krishnan K., Kavita R., Alexander L. S., and Phil W.: *Providing Quality of Service over a Shared Wireless Link*, IEEE Communications Magazine, Vol.39 Issue.2, pp. 150–154, 2001.

[19] Sanjay S., and Alexander L. S.: *Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR*, Proceedings of International Teletraffic Congress (ITC), 2001.

[20] Jalali A., Padovani R., and Pankaj R.: *Data Throughput of CDMA-HDR a High Efficiency-High Data Rate Personal Communication Wireless System*, Vehicular Technology Conference Proceedings, Vol.3, pp. 1854–1858, 2000.

[21] Niyato D., and Hossain E.: *A Queuing-Theoretic and Optimization-Based Model for Radio Resource Management in IEEE 802.16 Broadband Wireless Networks*, IEEE Transactions on Computers, Vol.55 Issue.11, pp. 1473–1488, 2006.

[22] Rong B., Qian Y., and Lu K.: *Integrated Downlink Resource Management for Multiservice WiMAX Networks*, IEEE Transactions on Mobile Computing, Vol.6 Issue.6, pp. 621–632, 2007.

[23] Kitti W., and Aura G.: *Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems*, International Journal of Communication Systems, Vol.16 Issue.1, pp. 81–96, 2003.

[24] Woo J. K., Joo Y. B., Sun D. L., Young J. S., Yun S. K., and Jin A. K.: *Efficient Uplink Scheduler Architecture of Subscriber Station in IEEE 802.16 System*, Lecture Notes in Computer Science, Vol.3823, pp. 734–743, 2005.

[25] Chang B. J., and Chen Y. L.: *Adaptive hierarchical polling and Markov decision process based CAC for increasing network reward and reducing average delay in IEEE 802.16 WiMAX networks*, Computer Communications, Elsvier, Vol.31 Issue.10, pp. 2280–2292, 2008.

[26] Majid G., and Raouf B.: *Call admission control in mobile cellular networks: a comprehensive survey*, Wireless Communications and Mobile Computing, Vol.6 Issue.1, pp. 69–93, 2006.

[27] Christian H.: *Analysis and performance evaluation of the OFDM-based metropolitan area networks IEEE 802.16*, Computer Networks, Vol.49 Issue.3, pp. 341–363, 2005.

[28] Wang H.: *Priority-based Resource Allocation for RT and NRT Traffics in OFDMA Systems*, The 3rd IEEE International Conference on Wireless Communications, Networking and Mobile Computing (IEEE WiCOM), Vol.1, pp. 791–794, 2007.

[29] Wang H., and Dittmann L.: *Adaptive Radio Resource Allocation in Hierarchical QoS Scheduling for IEEE 802.16 Systems*, 50th IEEE Global Telecommunications Conference (IEEE GLOBECOM), pp. 4769–4774, 2007.

[30] IEEE 802.16.3c-01/29r4, *Channel Models for Fixed Wireless Applications*, IEEE 802.16 Broadband Wireless Access Working Group, July 2001.

15

# Analyzing the Throughput and QoS Performance of WiMAX Link in an Urban Environment

Faqir Zarrar Yousaf, Kai Daniel and Christian Wietfeld
*Communication Networks Institute,*
*Dortmund University of Technology (TU Dortmund),*
*Germany*

## 1. Introduction

WiMAX was formally introduced in 2004, with the publishing of the IEEE 802.16-2004 standard (IEEE 802.16-2004, 2004), which specifies the air interface, including the medium access control (MAC) layer and multiple physical (PHY) layer specifications for Fixed Broadband Wireless Access systems WiMAX claims to provide not only higher data rates over larger geographical areas but also claims to provide sophisticated QoS support capability matching its wired counterparts. However, due to lack of deployment there is a need to bridge the gap between the performance perception and the actual performance limitations of WiMAX. In this chapter we intend to present experimental results that would give an insight into the stability of the WiMAX link in comparison to its performance claims. Some of the performance results presented in this chapter is based on our previous efforts that has been recorded and published in (Yousaf et al., 2007).

### 1.1 Motivation and Objectives

Despite the high performance claim of WiMAX, it is yet desired to see a high deployment density. Due to this lack of deployment not enough off-the-field empirical data is available that would provide quantitative and qualitative attestation of the performance and operational capabilities and efficiencies of the WiMAX link. Such a provision of performance data will enable network service providers, planners and designers to make informed choices when deploying and integrating WiMAX into their network infrastructure. Availability of such a data could also be potentially used as a benchmark by telecommunication engineers and researchers to develop accurate mathematical and simulation models that could compare more realistically to the actual performance behaviour of the WiMAX link. This would thus enable them to develop and recommend better and efficient algorithms and solutions in terms of Next Generation Wireless Access Networks.

This lack of availability of real empirical data was the main motivation behind conducting field experiments over an actual 3.5 GHz WiMAX test-bed established at Communication

Networks Institute (CNI), Technische Universität Dortmund. The tests were designed so as to give a better understanding and insight into the performance of the WiMAX link in terms of data throughput, link stability and QoS support for various services and traffic types at specified distances from the Base Station (BS). This experimental link was stressed in both uplink and downlink direction to determine the performance boundaries in an urban environment.

### 1.2 Related Work

Although some performance data is available but these are mostly based on analytical and/or simulation models (Hoymann, 2005). Also some experimental results are provided in (Schwengler & Pendharkar, 2005), but the experiments were conducted by *emulating* the WiMAX link using different channel models. The experiments performed in (Schwengler & Pendharkar, 2005) were conducted with multi-rate support enabled in which the Subscriber Station (SS) is allowed to choose a specific modulation scheme according to its SNR and the experiments were conducted for a limited range of modulation schemes. There is no mention of the transmit power of the base station or the type of the antennas used, a factor crucial to the correct understanding of the system's scope, and the maximum distance over which the tests were conducted was around 3000 meters (around 2 miles).

In our experiments, the *multi-rate support* (or the Adaptive Modulation & Coding (AMC) feature) was disabled to correctly ascertain the throughput and link quality for each of the eight available modulation schemes (please see table 1) with minimum transmit power of 13dBm over distances ranging from 220 meters to 9400 meters (9.4 km) from the BS. The AMC feature enables the system to dynamically adapt the modulation scheme and forward error correction (FEC) coding to actual link conditions for each uplink and downlink direction.

The link was tested for both the TCP and UDP traffic generated in both uplink and downlink direction, thereby emulating data relating to a case of real customer deployment.

In terms of testing the QoS support in WiMAX, experimental measurements have been made in (Scalabrino et al., 2006) using VoIP traffic and targeting the residential broadband access scenario, in which multiple SS, each having the same QoS configuration, communicate with a single base station in a point to multipoint (PMP) configuration. In our tests we have targeted a SOHO/SME broadband access scenario in which a single SS will be providing services to hundreds of users and/or multiple departments, each with a possibly unique QoS requirements, in a Virtual LAN (VLAN) configuration environment, where subscribers sharing the same QoS class are mapped onto the same service pipe, a virtual connection between a subscriber's application and the network resources, over the WiMAX link.

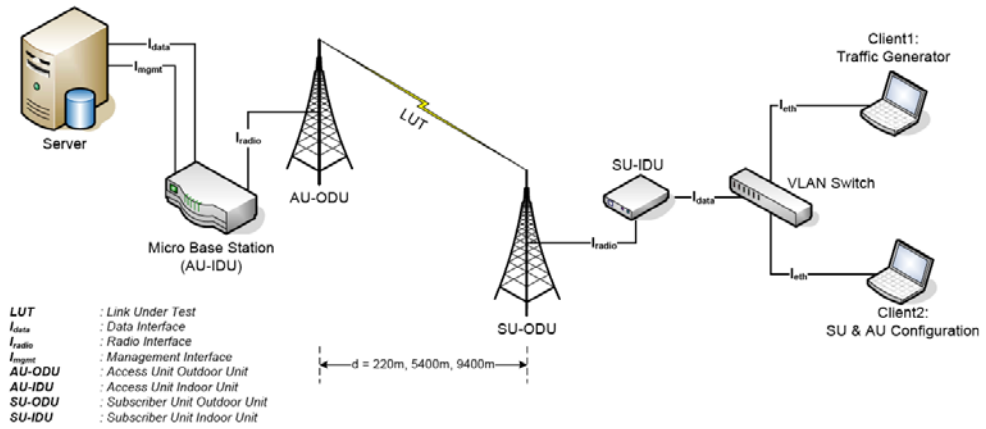



Fig. 1. Generic Test Bed Setup

## 2. Test Bed Arrangement and Setup

For the field experiments an IEEE 802.16d (IEEE 802.16-2004, 2004) compliant equipment operating in a 3.5 GHz licensed band was selected.

As the WiMAX link is tested in terms of its capacity and its ability to provide QoS to multiple simultaneous applications services, two different test-beds were composed and configured. Both the test-beds had the same generic setup as shown in figure 1 but differed in terms of the equipment configuration.

On the subscriber side, the *Client1* laptop is used for generating user traffic in the uplink direction and captures and analyses data in the downlink direction whereas *Client2* laptop is running scripts for remote access and configuration of the WiMAX base station (AU-IDU) and the Customer Premises Equipment (CPE) (SU-IDU). The two client laptops and the Subscriber Unit Indoor Unit (SU-IDU) are connected via an Ethernet Interface ($I_{eth}$) to the VLAN switch, where the VLAN feature is disabled except for the QoS testing, described later. The SU-IDU in turn is connected to the Subscriber Unit Outdoor Unit (SU-ODU), which is a vertically polarized directional antenna with an 18dBi gain.

On the BS side, a Server is configured for generating traffic in the downlink, over the LUT (Link Under Test), and capturing and analyzing data in the uplink. The Access Unit Indoor Unit (AU-IDU) is connected via an IF cable ($I_{radio}$) to the Access Unit Outdoor Unit (AU-ODU), which consists of a 10.5 dBi gain omni-directional antenna and a high-power, full-duplex multi-carrier radio unit. The AU-IDU consists of two RJ45 ports namely; management port and data port, for micro base station remote configuration and the latter for data communication. These two ports are connected to the Server via Ethernet (i.e., $I_{data}$ & $I_{mgmt}$).

The technical specification of the test equipment is illustrated in figure 2 while the detailed radio specifications of the IDU and the ODU are given in table 1.

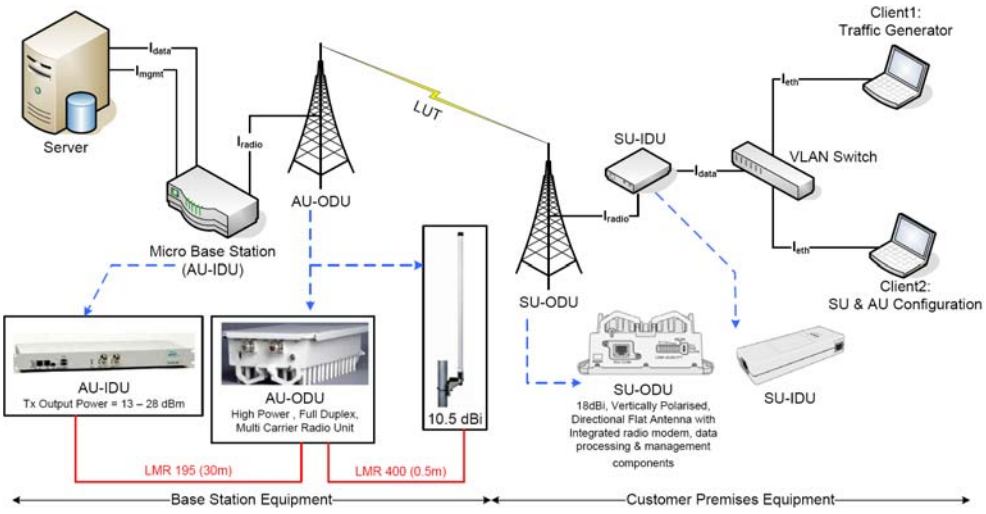



Fig. 2. Technical Specification of the Test Equipment

| | | | |
|---|---|---|---|
| Frequency | Unit Type | Uplink (MHz) | Downlink (MHz) |
| | AU- ODU | 3399.5 – 3435.5 | 3499.5 – 3553.5 |
| | SU-ODU | 3399.5 - 3500 | 3499.5 - 3600 |
| Channel Bandwidth | 3.5 MHz | | |
| Operation Mode | AU-ODU | FDD, Full Duplex | |
| | SU-ODU | FDD, Half Duplex | |
| Modulation | OFDM Modulation, 256 FFT points;<br>BPSK, QPSK, QAM16, QAM64 | | |
| Transmit Power | AU | 13dBm (20 mW)<br>28 dBm (631 mW) | |
| | SU | 20 dBm (100 mW) | |
| Bit Rate | Modulation & Coding | Net Physical<br>Bit Rate (Mbps) | |
| | BPSK ½ | 1.41 | |
| | BPSK ¾ | 2.12 | |
| | QPSK ½ | 2.82 | |
| | QPSK ¾ | 4.23 | |
| | QAM16 ½ | 5.64 | |
| | QAM16 ¾ | 8.47 | |
| | QAM64 ²/₃ | 11.29 | |
| | QAM64 ¾ | 12.71 | |

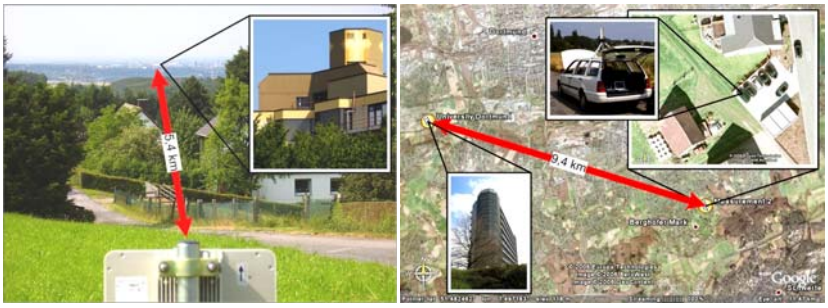
Table 1. Radio Specification of the Test Equipment

## 3. Measurement Methodology

The tests were divided into two categories:

- ***Link Capacity/Throughput Testing,*** in which the ability, reliability and the robustness of the WiMAX link was tested in both the uplink and downlink direction for all eight modulation schemes by transmitting TCP and UDP traffic and real time video streams at different distances from the BS, and
- ***QoS Testing,*** in which the inherent QoS support feature of the WiMAX was tested.

To attain reliable results with an appreciable level of confidence, each test run was conducted ten times and the duration of each test run was 60 seconds, accounting to a total approximate test duration of 260 minutes excluding the time for initializing the test scripts and making necessary configurations on the equipment. The throughput of the uplink and the downlink were measured separately in order to exclude interference effects.




(a) (b)

Fig. 3. Test Locations in Dortmund, Germany (a) Auf dem Schnee, and (b) Radio Tower, Schwerte

The amount of traffic generated was always equal to the net physical data rate for each respective modulation scheme (see table 1) and corresponding packet losses were obviously measured.
The link capacity was tested for both Line Of Sight (LOS) and Near-LOS scenarios, but Non-LOS (NLOS) communication was not possible. The CPE were stationed away from the BS at the following distances and locations:

- d = 220 meters; North Campus, University Dortmund
- d = 5400 meters (5.4 km); Auf dem Schnee, Dortmund. LOS (Figure 3(a))
- d = 9400 meters (9.4 km); Radiotower Schwerte, Dortmund. Near-LOS (Figure 3(b))

Table 2 summarizes the relevant test configuration of the AU.

| | |
|---|---|
| Multi Rate Support (BS) | Disable |
| ATPC Support (BS) | Enable / Disable |
| Transmit Power (at BS antenna port) | Minimum = 13 dBm (20 mW)<br>Maximum = 20 dBm (100 mW) |
| Transmit Frequency (BS) | Downlink = 3501.25 MHz<br>Uplink = 3401.25 MHz |
| VLAN Support | Yes |
| QoS Profile | Real Time (RT) Non Real Time (NRT), Best Effort (BE) |

Table 2. Access Unit (AU) Configuration

To stress-test the link the transmit power of the AU was set to the minimum of 13dBm and the Automatic Transmit Power Control (ATPC) feature was enabled for the SU and AU. Although the maximum transmit power of the BS is 28dBm (6.918 Watts EIRP *incl. 10.5dBi antenna gain*, but the limit of 20dBm (1.096 Watts EIRP *incl. 10.5dBi antenna gain*) was imposed by the terms and conditions of the 3.5GHz research license. As a QoS-Profile a non real-time setting with a Committed Information Rate (CIR) of 12Mbps is chosen. The *Multirate* support in the AU was disabled, as mentioned earlier, in order to set the modulation scheme manually for each measurement to validate the link behaviour for each specific modulation scheme.
The outdoor tests were carried out by mounting the antenna on the roof of a car and powering the whole SU by a portable DC to AC power supply, which also powers the switch and the client PC.

### 3.1 Uplink and Downlink Test Configurations

For the uplink and downlink tests, TCP and UDP traffic was generated using IPerf (Iperf), which reported bandwidth, delay jitter and datagram loss, and captured on the receiver end by Wireshark (Wireshark) to capture and analyze the received data. In case of *uplink capacity tests*, the *Client1* runs the IPerf server instance whereas the *Server* runs the IPerf client instance and Wireshark application; and this arrangement is reversed for the *downlink capacity tests*. The *Client2* machine is used primarily to manage and configure the AU remotely by establishing an SSH session and is also used to monitor the configuration and connection statistics of the SU (for example, RSSI, SNR, modulation scheme, transmit and receive power level of the SU). This automatic management, configuration and monitoring of the AU and the SU is controlled by custom test processing scripts developed in Perl.
The captured data is then fed to the processing scripts where it is analysed for various parameters and mean values are hence calculated. The results are then graphically depicted using the GNU-Plot/Excel application.

### 3.2 QoS Test Configurations

In order to test the inherent ability of the WiMAX protocol to support QoS over the broadband wireless links the general set up is the same as in figure 1 with the exception that the VLAN feature in the switch is now enabled and the two client machines (Client1 and Client2) are placed in two separate VLANs. Client1 is generating UDP traffic with transmission parameters emulating a *real time* VoIP service, whereas Client2 is generating TCP traffic streams emulating a *non-real time* service application such as HTTP etc. The two VLANs are mapped onto two different service pipes, where the service pipe designated to carry the real time application data is configured with a *Real Time Polling Services (rtPS)* service class and the second service pipe carrying the non-real time application data is configured with a *Best Effort (BE)* service class. The two clients generate simultaneous traffic using the QAM64 ¾ modulation and coding scheme over the maximum bandwidth of the link so that the effect of these two interfering traffic streams over the QoS enabled WiMAX link can be duly analyzed.
The QoS tests were then performed for all the eight available modulation schemes.

## 4. Analysis and Evaluation

This section will provide the measurement results for the test configurations discussed above and also discuss the effect of transmit power on the overall link quality.

### 4.1 Throughput Tests

Table 3 shows the TCP and UDP average throughput measured, which can be compared to the net physical bit rates for the corresponding modulation scheme given in table 1, for the eight modulation schemes at reference distances of 220 meters, 5400 meters and 9400 meters and the transmit power was set at a constant minimum of 13dBm. In the *downlink direction* (see figure 4) TCP throughput remains almost stable but shows greater inconsistency especially at higher modulation schemes (QAM64). It has been experimentally verified and observed that the TCP throughput stability is dependent on distance and transmit power. In terms of distance, the higher modulation schemes shows greater consistency at 220m (figure 4 (a)) than at 5400m (figure 4 (b)) or 9400m (figure 4 (c)). The effect of transmit power on the TCP downlink throughput is subsequently discussed.

| Modulation Schemes | Downlink Throughput in Mbps | | | | | | Uplink Throughput in Mbps | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 220 meters | | 5400 meters | | 9400 meters | | 220 meters | | 5400 meters | | 9400 meters | |
| | TCP | UDP | TCP | UDP | TCP | UDP | TCP | UDP | TCP | UDP | TCP | UDP |
| *BPSK 1/2* | 1.08 | 1.13 | 1.08 | 1.12 | 1.08 | 1.12 | 0.98 | 0.98 | 0.97 | 1.03 | 0.97 | 1.03 |
| *BPSK 3/4* | 1.65 | 1.74 | 1.64 | 1.73 | 1.65 | 1.73 | 1.51 | 1.54 | 1.46 | 1.57 | 1.5 | 1.58 |
| *QPSK 1/2* | 2.21 | 2.32 | 2.2 | 2.32 | 2.2 | 2.32 | 2.03 | 2.14 | 2.03 | 2.13 | 2.03 | 2.14 |
| *QPSK 3/4* | 3.37 | 3.53 | 3.35 | 3.53 | 3.35 | 3.53 | 3.09 | 3.24 | 3.08 | 3.23 | 3.08 | 3.23 |
| *QAM16 1/2* | 4.5 | 4.74 | 4.5 | 4.73 | 4.48 | 4.74 | 4.16 | 4.33 | 4.14 | 4.33 | 4.14 | 4.33 |
| *QAM16 3/4* | 6.8 | 7.11 | 6.28 | 7.11 | 6.41 | 7.1 | 6.24 | 6.52 | 6.22 | 6.52 | 6.22 | 6.52 |
| *QAM64 2/3* | 8.88 | 9.4 | 8.19 | 9.51 | 8.31 | 9.51 | 8.28 | 8.67 | 8.22 | 8.64 | 8.25 | 8.65 |
| *QAM64 3/4* | 9.58 | 10.55 | 9.13 | 10.71 | 8.34 | 10.69 | 9.25 | 9.67 | 9.22 | 9.69 | 9.22 | 9.67 |

Table 3. Average Throughputs for Uplink and Downlink TCP & UDP Traffic

The TCP throughput in the *uplink* direction is more stable at all the three reference distances and for all the modulation schemes and is similar, in performance, to figure 4 (a). This may be due to the quality, high sensitivity and the signal processing capabilities of the BS receiver.

In contrast to TCP downlink traffic, UDP has shown more throughput stability for all the modulation schemes in both the uplink and downlink direction at all three reference distances and the stability in throughput is similar to figure 4 (a)). The average UDP throughputs are given in table 3.

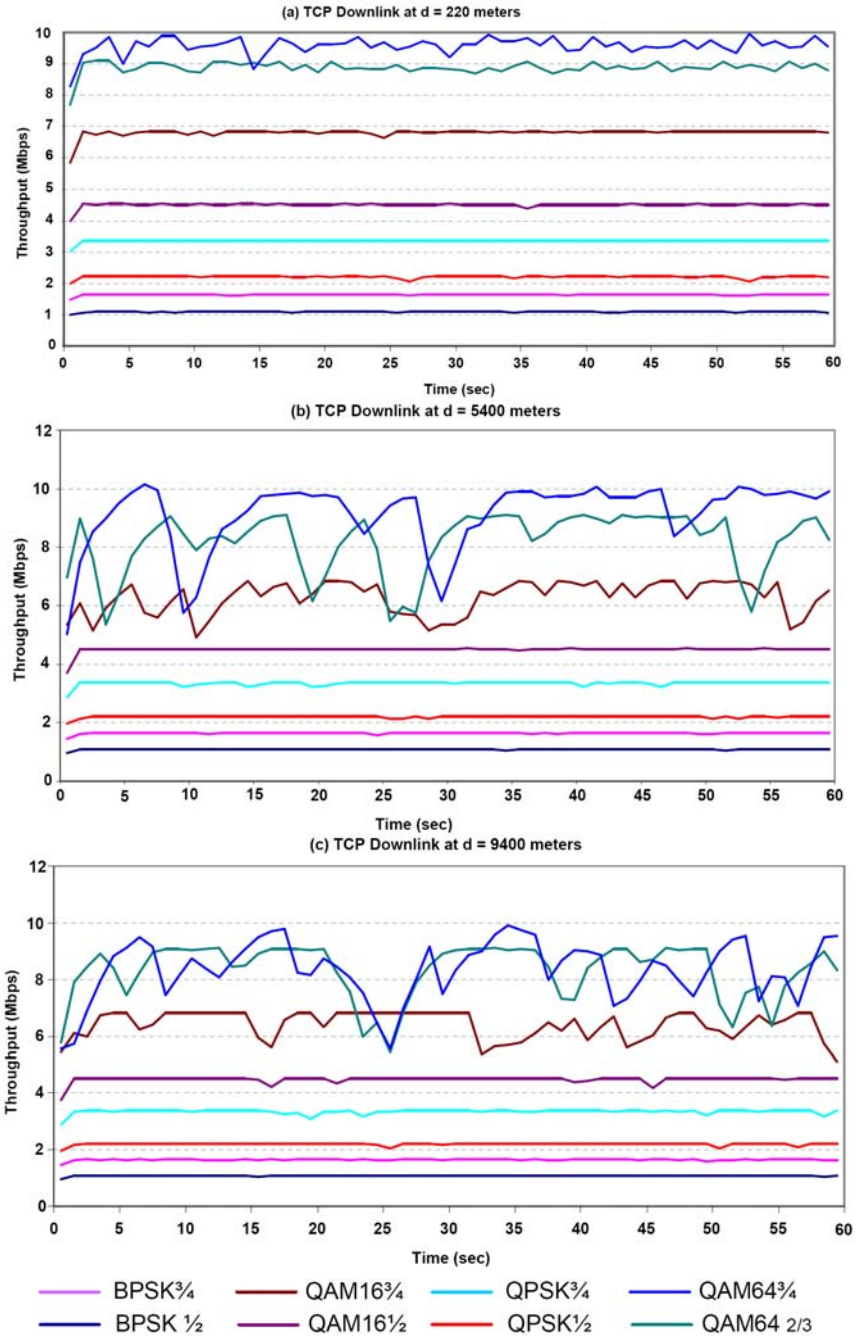



Fig. 4. TCP Downlink Throughput for the Different WiMAX Modulation Schemes (Yousaf et al., 2007)

### 4.2 Effect of the Transmit Power and Link Layer ARQ on the TCP Throughput Performance

In the previously discussed results, the link was stress tested by keeping the transmit power of the base station to the minimum value of 13dBm at all distances and for all tests. In order to evaluate the effect of transmit power on the overall link quality, a second set of tests at 20dBm transmit power was carried out at d=9400m by transmitting TCP traffic in the downlink direction using QAM64 ¾, as that was more critical in terms of throughput stability (as evident from figure 4 (c)). As seen in figure 5, there is marked improvement in terms of link and throughput stability at 20dBm, especially for TCP traffic, which earlier had shown considerable instability at 13dBm.

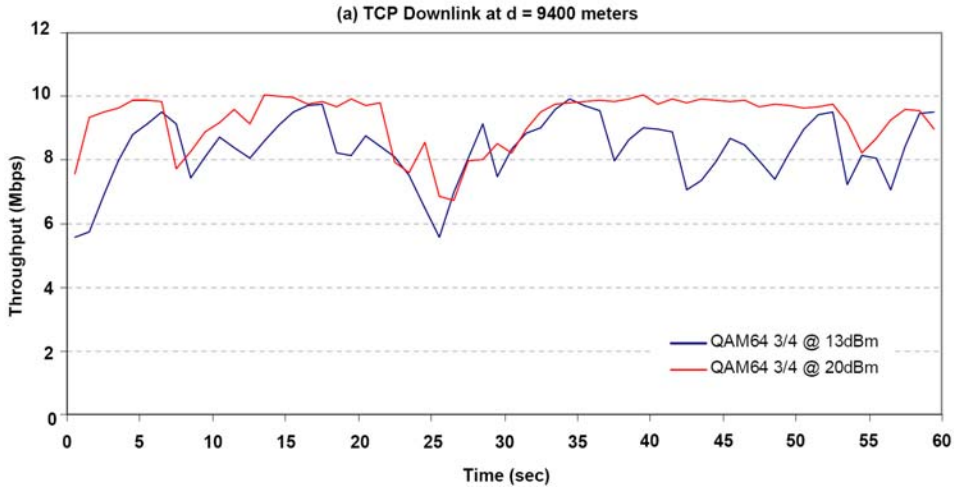



Fig. 5. Effect of Transmit Power on TCP Throughput

However, even at 20dBm there are still fluctuations in the throughput which may not suit throughput sensitive applications. It is suggested that for TCP sessions using higher modulation schemes, the link layer (L2) ARQ algorithm in the BS must be enabled so that the errors can be detected and retransmissions requested at the link level. This will have a significant impact on the throughput stability, especially for the more susceptible higher modulation schemes.

The effect of ARQ on the TCP throughput for the QAM64 ¾ modulation scheme for power levels of 13dBm and 20dBm is shown in figure 6 (a) and 6 (b) respectively. As depicted in figure 6, when the ARQ feature is *disabled*, there are fluctuations in the TCP throughput, even at higher power levels (20dBm in our case). With the ARQ feature enabled, the TCP throughput shows an equally stable throughput performance for both the reference transmit-power levels of 13dBm and 20dBm.

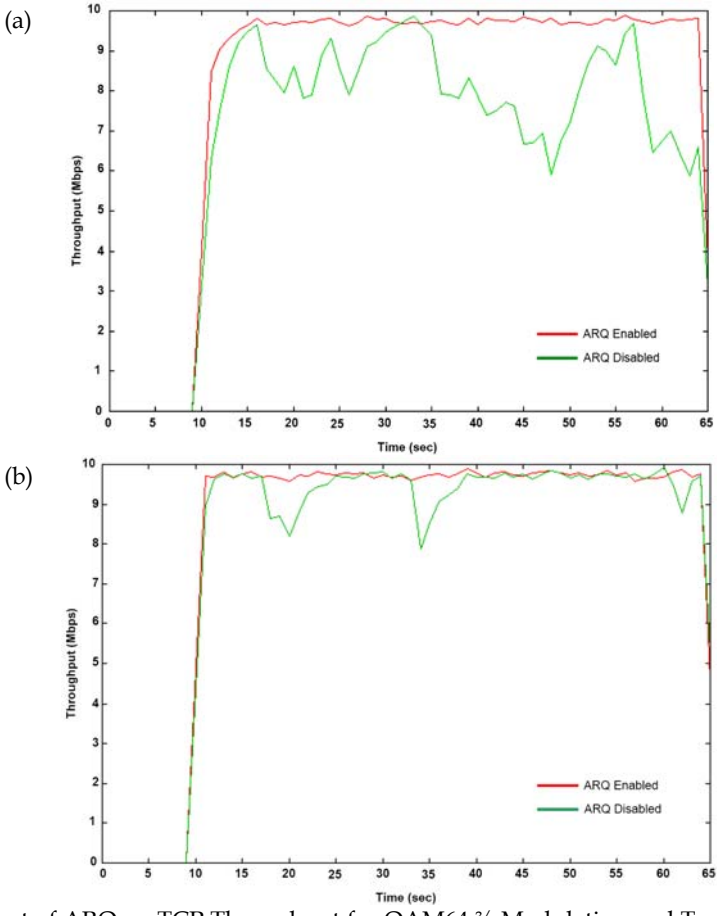



Fig. 6. Effect of ARQ on TCP Throughput for QAM64 ¾ Modulation and Transmit Power of (a) 13dBm, and (b) 20dBm

### 4.3 TCP Window Size for Optimum Performance

During TCP connection, setting the right TCP Receive Window (RWIN) size is crucial from the perspective of throughput performance. The size of the RWIN is negotiated at the start of the connection during the TCP handshake process. The transmitter will stop sending data after it has transmitted data equal to the negotiated window size and will wait for an acknowledgement before resuming transmission.

Smaller values of RWIN will have adverse affect on the session quality as packets will be frequently dropped by the overflowing buffer. This dropping of packets will result in data retransmissions and will not only result in reduced average throughput but also affect the QoS of delay and throughput sensitive applications. This is depicted in figure 7, which shows the variation in the throughput of the TCP traffic in the downlink direction using the QAM64 ¾ modulation scheme for the RWIN size set to 30KB. For this test the Automatic Transmit Power Control (ATPC) feature is enabled which will dynamically adapt the

transmit power of the SS so that transmissions can be received by the BS at optimal levels. The test is conducted at a distance of 220m. As can be seen from the figure 7 a 30KB RWIN size results in an *unstable low throughput* due to packet losses and repeated retransmissions.

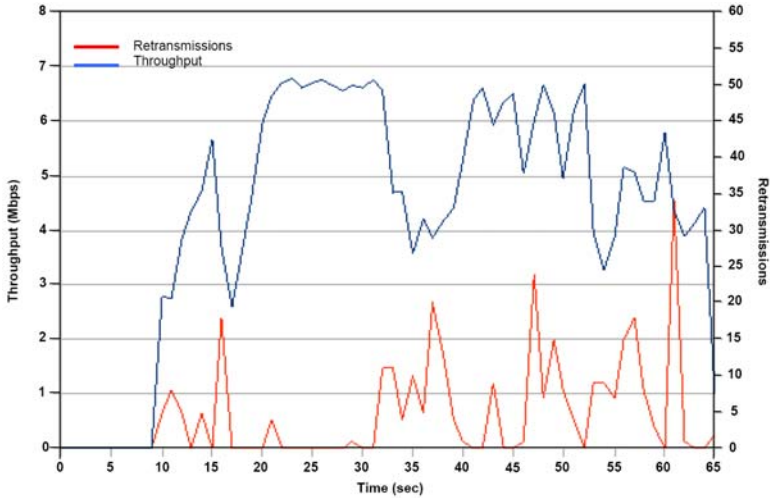



Fig. 7. Effect of Buffer Overflows During a TCP Transmission over WiMAX link

To find the optimal boundaries for RWIN, it is observed that the RWIN should be between 85KB (default setting in Linux) and 256KB. At RWIN 256KB, the TCP session throughput not only increases to 10Mbps but shows a *stable and constant* throughput performance as depicted in figure 8.

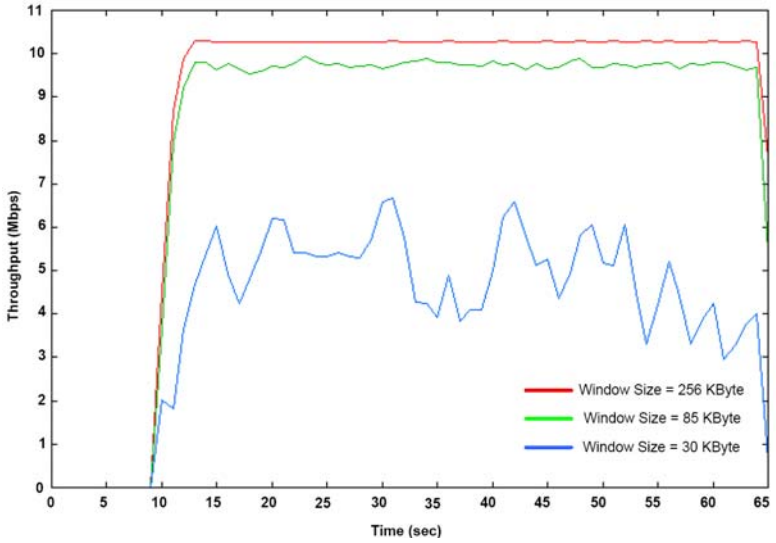



Fig. 8. Effect of Window Size on TCP Throughput for QAM64 ¾

### 4.4 QoS Tests

To ascertain the degree of QoS support, the tests were conducted by sending TCP background traffic, using QAM64 ¾ modulation, through a *non real-time* connection at full capacity of the bandwidth throughout the 60 seconds test duration. A 12 Mbps UDP traffic is sent in periodic durations of 10 seconds through a *real-time* connection with a committed bit rate of 5 Mbps and 10 Mbps the result of which is shown in figure 9 (a) and 9 (b) respectively.

From figure 9 it is seen that the throughput of the TCP data stream is exactly maximal (i.e, it is occupying the full available bandwidth spectrum) in periods where there is no UDP traffic. As soon as the UDP traffic is generated, the system will guarantee the required bandwidth as UDP is sent on a link which is assigned a higher QoS priority. This availability of bandwidth to the high priority UDP stream will be made at the cost of hogging the required bandwidth from the TCP session.

Thus, it is evident from the figure 9 (a) and 9 (b) that the TCP background traffic will occupy only that portion of the bandwidth which is not utilized by the UDP traffic and upon greater demand for bandwidth by the real-time traffic, the system ensures the guaranteed provisioning of the demanded bandwidth by claiming it from non-real time traffic, which in turn will experience bandwidth reduction, and this behaviour is consistent with the overall QoS philosophy.

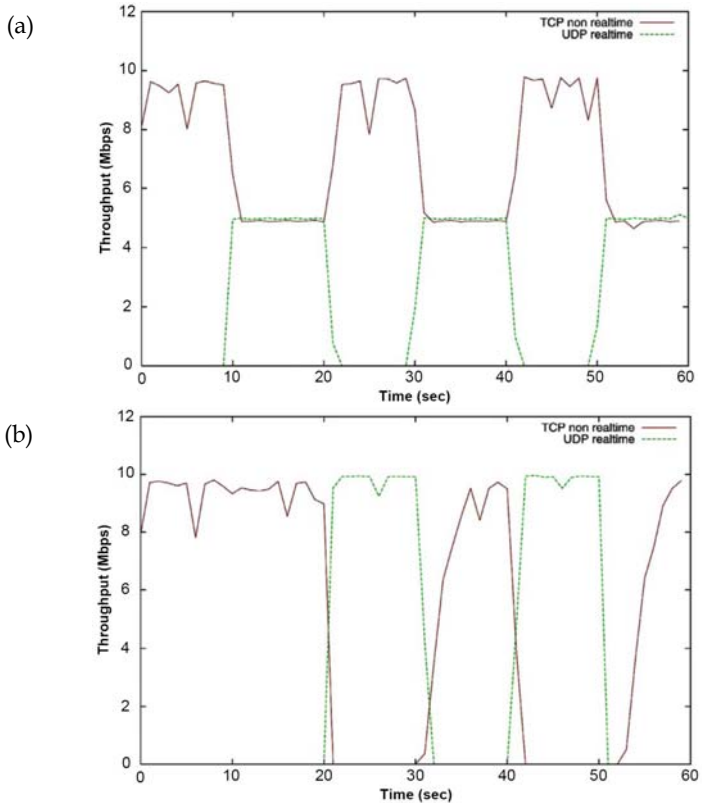



Fig. 9. QoS Measurement at d = 220m with (a) 5Mbps, and (b) 10Mbps UDP Traffic and TCP Background Traffic

Similar QoS tests were conducted for all the 8 modulation schemes and it was observed that the QoS paradigm applies consistently across all modulation schemes. As an example, figure 10 below shows the QoS tests for BPSK ½, QAM16 ½, and QAM64 ¾ for a real time UDP stream of 5Mbps.

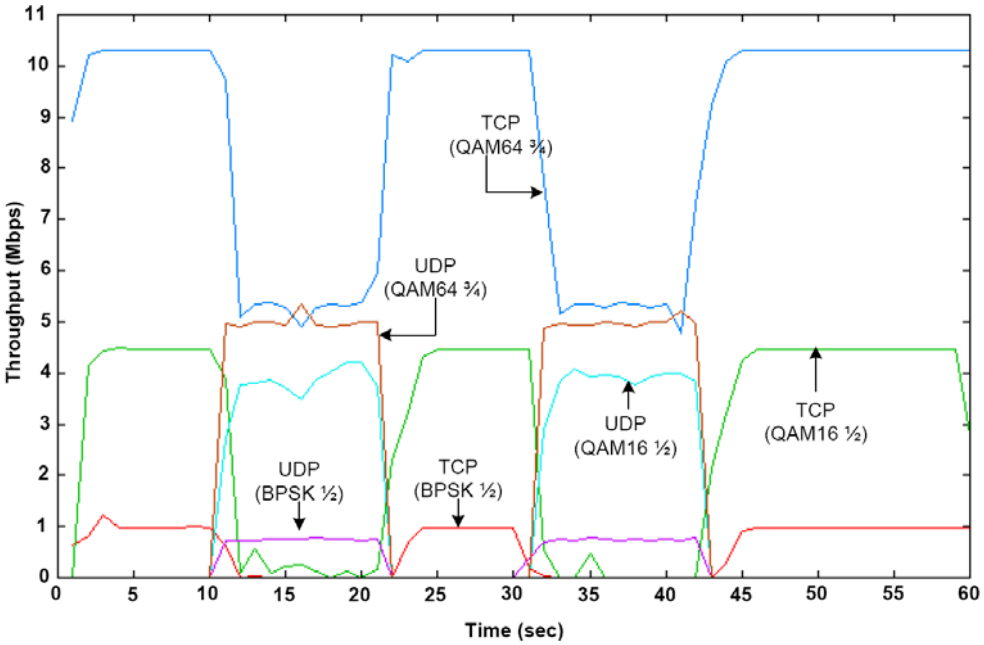



Fig. 10. QoS Measurement at d = 220m for BPSK ½, QAM16 ½ and QAM64 ¾

## 5. Conclusions and Future Work

With the results gained from this pilot project, and within the given limits, resources and distance limitations, the field tests conducted at the CNI's WiMAX test-bed can be regarded as meeting the functional specification and requirements of a wireless broadband access network as per the IEEE 802.16-d standard. The measurements and results gathered through stress testing the WiMAX link demonstrates satisfactory throughput and level of services at different distances for both LOS and Near-LOS, even at the lowest transmission power (13dBm) at distances as far as 9.4km. An important observation is that Non Line of Sight (NLOS) operation was not possible in our tests, unless relays are used as demonstrated in (Marques et al., 2007), and the correct operation depends on the accurate adjustment of the CPE in order to get the best possible receive signal. It has been observed that even one centimetre change of the vertical or horizontal orientation could have a strong influence on the RSSI and the SNR and with this the quality of the link, adversely affecting the throughput.

Although the transmit power does influence the throughput stability, but it was observed that enabling ARQ at the link layer will not only stabilise the throughput but it will also increase the throughput and will compensate the effect of low transmit power on average

throughput. Also important is to ensure the correct setting of the RWIN size for TCP traffic as smaller window sizes will adversely impact not only the average throughput but the link stability as well.
The WiMAX also has an effective QoS mechanism that applies equally well across all 8 modulation schemes.
Considerable degradation of service was observed during adverse weather conditions, such as during rain fall, resulting in delay, jitter and inconsistent service due to packet losses. Such weather conditions will prohibit the use of bandwidth intensive streaming multimedia services over the WiMAX link.
As part of the future work, more field tests are being planned at distances beyond 9.4 km and by introducing a second SS. Further investigation into the effect of transmit power on the link throughput and extensive testing of the QoS feature is also planned, particularly in terms of interference of different traffic types through traffic flow using all available service classes simultaneously. Similar tests are planned to be executed using the IEEE 802.16e; the mobility variant of WiMAX, also called the Mobile WiMAX.

## 6. References

Hoymann, C. (2005). Analysis and performance evaluation of the OFDM-based metropolitan area network IEEE 802.16, *Computer Networks,* Volume 49, Issue 3, October 2005, pp. 341-363, ISSN 1389-1286.

IEEE 802.16-2004 (2004). IEEE standard for Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed Broadband Wireless Access Systems, October 2004.

Marques, M. et al. (2007). Design and Implementation of IEEE 802.16 Radio Links with Relays, *6th Conference on Telecommunications (Conftele07)*, Portugal, May 9-11, Peniche.

Schwengler, T.; Pendharkar, N. (2005). Testing of fixed broadband wireless systems at 5.8 GHz, *Technical, Professional and Student Development Workshop*, pp. 32-38, IEEE Region 5 and IEEE Denver Section, April 2005.

Scalabrino, F.; Pellegrini, I.; Chlamtac A.; Ghittino & S. Pera (2006). Performance Evaluation of WiMAX Testbed under VoIP Traffic, *Proceedings of the 1st International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, pp. 97-98, ISBN:1-59593-540-0, USA, September 2006, ACM, Los Angeles.

Yousaf, F. Z.; Daniel, K. & Wietfeld, C. (2007). Performance Evaluation of IEEE 802.16 WiMAX Link With Respect to Higher Layer Protocols, *Proceedings of IEEE International Symposium on Wireless Communication Systems (ISWCS)*, pp. 180-184, ISBN: 978-1-4244-0979-2, Norway, November 2007, IEEE, Trondheim.

IPerf. http://www.noc.ucf.edu/Tools/Iperf/. Accessed on May 12, 2009.

Wireshark. http://www.wireshark.org/. Accessed on May 12, 2009.

16

# WiMAX-WiFi Synergy for Next Generation Heterogynous Network

Dr. Rashid A. Saeed, Amran Naemat, Azrin Aris and Mat Kamil Awang
*Telekom Malaysia, TMRND innovation centre, 63000 Cyberjaya, Malaysia*

## 1. Introduction

Networks integration is crucial criterion for next generation wireless networks, where the diverse of the technologies available has been optimized for different usage models. WiMAX and WiFi are the most promising techniques for future wireless networks; interworking between these technologies is inevitable for better usability of the networks infrastructure and support for seamless mobility and roaming. The unique similarities between WiMAX and WiFi networks that make the proposed synergy promising and meaningful that both technologies are fully packet switching uses IP-based technologies to provide connection services to the Internet. This standards- and IP-based network approaches provide compelling benefits to service providers to collaborate between these technologies.

By distributing high-speed Internet access from cable, Digital Subscriber Line (DSL), and other fixed broadband connections within wireless hotspots, WiFi has dramatically increased productivity and convenience. Today, there are nearly pervasive WiFi delivers high-speed Wireless Local Area Network (WLAN) connectivity to millions of offices, homes, and public locations, such as hotels, cafés, and airports. The integration of WiFi into notebooks, handhelds and Consumer Electronics (CE) devices has accelerated the adoption of WiFi to the point where it is nearly a default feature in these devices (Xu and Saadawi 2001). On the other hand, WiMAX takes wireless Internet access to the next level, and over time, could achieve similar rates to devices as WiFi. WiMAX can deliver Internet access miles from the nearest WiFi hotspot and blanket large areas called wide area networks (WANs), be they metropolitan, suburban, or rural with multi-megabit per second mobile broadband Internet access (Sundaresan et al., 2004), also see table 2. Although the wide area Internet connectivity offered by 2.5 and 3G cellular data services are mobile, these services do not provide the broadband speeds to which users have become accustomed and that WiMAX can deliver. In the last few years, WiMAX has established its relevance as an alternative to wired DSL and cable, providing a competitive broadband service offering that can be rapidly and cost effectively deployed (Gunasekaran and Harmantzis, 2005).

The unique similarities between WiMAX and WiFi networks that make the proposed synergy is promising is both technologies are fully packet switching uses IP-based technologies to provide connection services to the Internet. This standards- and IP-based network approach provides compelling benefits to service providers to collaborate between these technologies (Niyato, Hossain, 2004):

- A common user experience for wireless broadband services, which is a critical enabler in attaining rapid user adoption.
- An open network philosophy where any WiMAX or WiFi device is able to connect to any WiMAX or WiFi network that supports the same network setting, improving today's business models for delivering mobile broadband services.
- Vendor agreed-upon certification profiles, facilitating volume production and global economies of scale. Wireless client and network equipment subjected to extensive interoperability and conformance testing, enabling an open and competitive multi-vendor environment.
- An all-IP based network infrastructure, enabling cost-effective deployments for operators and open Internet services for users

| Feature | WiFi 802.11 a/b/g | WiMAX |
|---|---|---|
| Rang | Very small<br>Optimized for 100 m | Considerably larger<br>Optimized for 7 – 10 km<br>Up to 50 km range |
| Indoor vs. outdoor | Preferably Indoor<br>PHY tolerates only<br>low multi-path delay spreads | Indoor (16a) Outdoor (16, 16e)<br>PHY tolerates high multi-path delay spread (10 µsec = Trees, buildings, persons |
| Scalability | Poor<br>Fixed bandwidth of 20 MHz<br>Fixed channels<br>Certain degree of channel overlap | Excellent Variable bandwidth 1.5 MHz – 20 MHz<br>Only limit is available spectrum |
| Bandwidth | Unlicensed bands | Licensed and unlicensed bands |
| Sensing | Low to medium<br>1 Mbps - 11 Mbps - 54 Mbps - 108 Mbps. Efficiency 2,7 bps / Hz peak | High 75 - 100 Mbps – 134 Mbps<br>Efficiency 5 bps / Hz |
| Layer | Contention based MAC: CSMA / CA | Grant based MAC (TDM/TDMA) |
| Security | Mostly questionable<br>Various attempts on improvements<br>WEP, WPA | Supposedly good<br>Nevertheless flawed in principle<br>3DES, AES |
| QoS | poor QoS, 802.11e using the Hybrid Coordination Function (HCF). | Good QoS 4 QoS classes for voice & video |

Table 1. Comparison of WiFi, WiMAX and Cellular IP

This chapter explores the complementary nature of WiMAX and WiFi, as well as illustrates how service providers can leverage these technologies to offer wireless broadband Internet connectivity and compelling new services at affordable prices and in more locations. It also focuses on the synergies between the IEEE 802.11n and IEEE 802.16e-2005. The draft IEEE 802.11n standard is a new high-throughput enhancement designed for digital home and office applications based on orthogonal frequency division multiplexing (OFDM) modulation. IEEE 802.16e-2005 is the mobile enhancement to IEEE 802.16-2004 and is designed to support wide area mobility via scalable OFDMA technology. Both of these technologies leverage OFDM and advanced antenna innovations to attain high-broadband data rates and improved signal reception.

## 2. Interoperability Opportunities and Challenges

The industrial implementation of IEEE 802.11 (WLAN) and IEEE 802.16 (WirelessMAN) standards are referred to as WiFi and WiMAX, respectively (Dhawan, (2007). The protocols of 802.16 and 802.11 have fundamental differences in their MAC layers: While 802.16 is a frame-based, centrally coordinated MAC protocol, 802.11 allows distributed control and a contention-based medium-access. In addition, 802.11 also realize a contention-free, centrally controlled access to the channel. Both 802.16 and 802.11 have a similar OFDM-based transmission scheme and channelization which facilitates their interworking (Man et al., 2007). The integration of 802.16 and 802.11 implies interworking between similar and different types of devices in a common protocol. The network is coordinated by central coordinator, which combines the 802.16 BS with the point coordination function (PCF) of 802.11 (a.k.a. infrastructure mode) and is thus referred to as WiWi BS or WiWi module in the MS. WiWi is abbreviation for WiMAX/WiFi project running in MIMOS Berhad since 2007 (MIMOS, 2009). WiWi is capable of operating in both 802.16 and 802.11 mode. In (Berlemann et al., 2006) the authors called it as Base Station Hybrid Coordinator (BSHC) where the interworking is based on an integration of 802.11 transmission sequences into the MAC frame structure of 802.16. Note that the distributed coordination function (DCF) is difficult to be implemented in BSHC and WiWi due to its random behaviour. In BSHC, an optional period for contention-based access maybe placed between two consecutive 802.16 MAC frames. While WiWi based on protocol encapsulation.

The integration of 802.11 and 802.16 into one WiFi/WiMAX module has been discussed extensively in the literature (Behmann, 2005); (Jong-Ok, 2004); (IEEE Std 802.16g, 2009), which all are work for the realization of an interworking between these two standards. In (Behmann, 2005) a common framework was introduced, that allows the operation of 802.11 and 802.16 with optimal bandwidth sharing. Game theory and genetic algorithm have been used to analyze and obtain the optimal pricing for bandwidth sharing between a WiMAX BS and WiFi APs, taking into account the bandwidth demand of the WiFi users. In (Jong-Ok, 2004) airtime-based link aggregation for WiFi and WiMAX was discussed. The airtime cost was used to measure the available resource of heterogeneous wireless links and it calculated on a packet basis for single user.

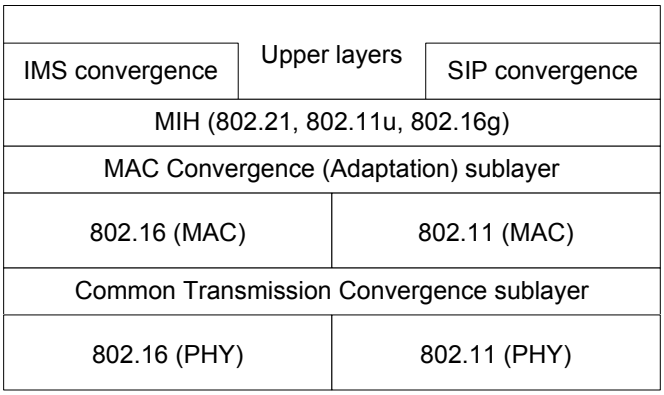



Fig. 1. The potential protocol stack WiMAX/WiFi integration

Figure 1 illustrates the integration of WiMAX and WiFi from OSI layered architecture perspective from upper layers down to physical layer.

The integration and convergence can be done in many places in the OSI model. The easiest way is to go for upper layer, where all the implementation will be based on the software. Software design is more flexible and robustness than hardware design. However, more delay and jitter will be experimented. On the other hand, the fastest integration solution is working at lower layers (i.e. MAC and PHY), but at the expense of complexities to the system. One of the most popular solutions now a day is implementing IMS (IP multimedia Subsystem) at the application layer. The IP Multimedia Subsystem (IMS) (3GPP, 2007) defined by 3GPP provides an enabling architecture that is access independent which is considered central in the move towards convergence. At the application layer SIP (session initiation protocol) sublayer from IETF under the Request for Comments (RFC) 3261 (Gaitan et al., 2005); (Gaitan et al., 2005), provides converged and unified communication services, such as voice and video conversations.

## 3. PHY Layer Integration

There are many similarities between WiFi and WiMAX at the PHY and MAC layers perspectives which both are IEEE 802® LAN/MAN standards, where there are many same design concepts also some features are polled from 802.11 to 802.16. At the MAC layer, the Media Independent Handover (MIH) protocol developed by IEEE 802.21 group enables the handover of IP sessions from one layer 2 access technology to another, to achieve mobility of end user devices (Lampropoulos et al., 2007). MIH is discussed in section 3.

At the physical layer, both 802.16 and 802.1 are uses the orthogonal frequency division multiplexing (OFDM) transmission concepts i.e. mobile WiMAX is using orthogonal frequency division multiplexing access (OFDMA) techniques (Surducan et al., 2007). The convergence at physical layer will reduce the base-station cost significantly where the base station can use the same IF, RF, and antenna parts for both technologies. This concept is illustrated in Figure 2. However, the integration at PHY level needs a change on the silicon

chip which increases the complexity of the baseband chip (Zlydareva, Sacchi, 2008). This can be realized by implementing Software defined radio (SDR) technique to switch between the two techniques at base band (BB) and RF levels where components that have typically been implemented in hardware (i.e. mixers, filters, amplifiers, modulators/demodulators, detectors. etc.) are instead implemented using software on the base-station baseband. Implementing cognitive radio can help the integration at the RF level. In this chapter we focused and proposed a new handover algorithm that works above the MAC layer. The realization of the interworking of the algorithm is discussed and evaluated in the next two sections.

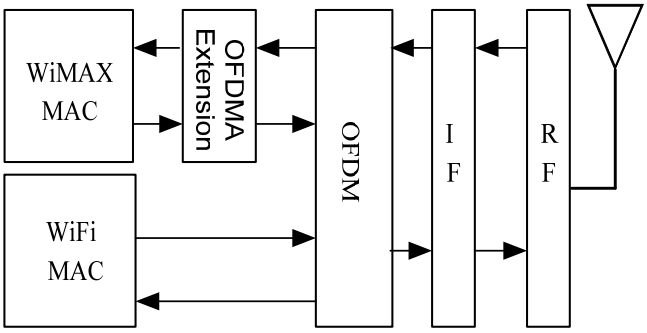



Fig. 2. the integration of WiMAX and WiFi PHY layer at the baseband level

Software Define Radio, SDR system is a radio communication system where components that have typically been implemented in hardware (e.g. mixers, filters, amplifiers, modulators/demodulators, detectors. etc.) are instead implemented using software on a personal computer or other embedded computing devices. Which is enables feature-rich services deployment for the increasing flexibility and mobility demands of enterprise and residential broadband users. SDR technology provides highly granular control over all radio system domains, including frequency, time, modulation, power, space, and coding. The result is a deployment and user experience agile enough for the rapidly changing broadband landscape, and connectivity powerful and intelligent enough to support flexible and profitable high value broadband IP services anywhere, anytime. While SDR is applied for based band (BB) the SR (software radio) is introduced for the whole system baseband, IF and RF. Software radio is the art and science of building radios using software. Given the constraints of today's technology, there is still some RF hardware not involved, but the idea is to get the software as close to the antenna as is feasible. Ultimately, we're turning hardware problems into software problems.

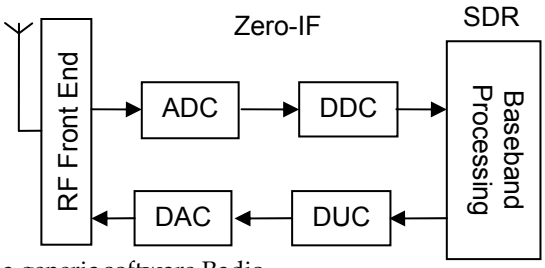



Fig. 3. Block diagram of a generic software Radio

The software radio (SR) contains a number of basic functional blocks. The radio can be split into three basic blocks, namely the front end, the IF section and the base-band section as shown in Figure 3. Each of the sections undertakes different types of functions and therefore is likely to use different circuit technologies.

The front end section uses analogue RF circuitry and it is responsible for receiving and transmitting the signal at the operational frequency, coupling the radio to the antenna or its feeder. It also changes the signal to or from the intermediate frequency (IF). Thus on the receive path the front end serves is connected to the antenna input using matching circuitry to ensure the optimum signal transfer. It then amplifies the signal uses low noise amplifier (LNA) and applies it to a mixer with a signal from a local oscillator to down-convert it to the intermediate frequency.

The IF section performs the digital to analogue conversion and vice versa. It also contains the processing that undertakes what may be thought of as the traditional radio processing elements, including filtering, modulation and demodulation and any other signal processing that may be required. On the receive path the signal enters the DAC where it is digitized and enters the DDC, the Digital Down Converter, where the signal is processed and demodulated to provide the baseband signal for the baseband processor. Similarly on the transmit side the signal arrives from the baseband processor and is modulated onto the carrier and conditioned as required. It is then converted from its digital format to analogue using a digital to analogue converter.

The DDC and DUC require significant levels of processing. This is required to perform all the processing on the actual signals in digital format. This processing must be achieved in real time for te system to be able to operate satisfactorily. As a result the processors are implemented in either stock DSPs or ASICs. In fact to achieve the full programmability and reconfigurability needed for a software defined radio the signal processors may be implemented as FPGAs. In this way the circuit can be totally reconfigured if needed. The final stage of the radio is the baseband processor. It is at this point that the digital data is processed, with protocols being accommodated and the data payload assembled or disassembled from the datastream. Although not as demanding as the DDC and DUC areas, with protocols becoming ever more complicated and demanding, the level of processing is increasing in these areas as well.

## 4. IEEE 802 Standards for Integration

### 4.1 802.11u (interworking with external networks)

IEEE 802.11u (IEEE Std 802.11u, 2009) is a proposed amendment to the IEEE 802.11standard to add features that improve interworking with external networks. The interworking with external networks includes other 802 based networks such as 802.16 and 802.3 and non-802 networks as 3GPP based IMS (IP multimedia subsystem) networks through ssubscription service provider network (SSPN). In this case, interworking refers to MAC layer enhancements that help selection of a network and allow higher layer functionality to provide the overall end to end solution. 802.11u is also to permit an emergency service i.e. E911.

### 4.2 802.16.4 WirelessHUMAN

WirelessHUMAN is an IEEE802.16 study group stand for wireless high-speed unlicensed metropolitan area networks (IEEE Std 802.16.4, 2009). The objective was to investigate the feasibility of providing High-speed Unlicensed MAN access (focus on UNII bands i.e. 5-6 GHz). This standard specifies the medium access control layer and physical layers of the air interface of interoperable fixed point-to-multipoint broadband wireless access systems. The specification enables transport of data, video, and voice services with quality of service. Physical layers are specified for both licensed and license-exempt bands designated for public network access (see Figure 4). The standard will be based on modifications of the IEEE 802.16 medium access control layer, while the physical layer will be based on the OFDM mechanism of IEEE 802.11a and similar standards.

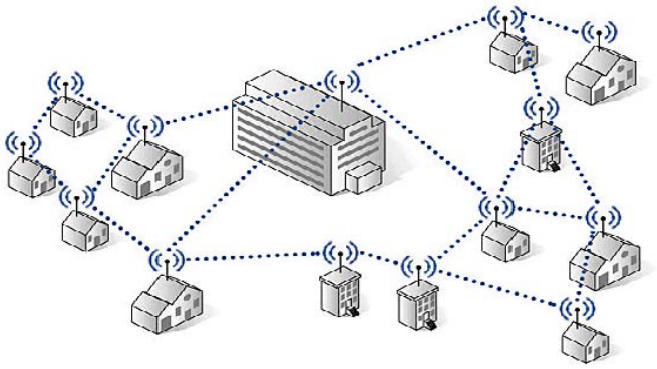


Fig. 4. WirelessHUMAN system architecture.

### 4.3 IEEE 802.21 Media Independent Handover (MIH)

IEEE 802.21 standard -recently approved at the IEEE-SA (IEEE Std 802.21, 2008)- is specifies procedures that facilitate handover decision making, providing link layer state information to MIH users and enabling low latency handovers across multi-technology access networks. Using IEEE 802.21 adds mobility between IEEE802 standards and other cellular networks and creates a single integrated system as shown in Figure 5. IEEE 802.21 provides the glue at

layer 2 (or layer 2.5) to make the any two radio technologies work together as one. Figure 5 illustrates 802.21 services in the protocol stack in a multimode client (e.g., a laptop or handset). 802.21 WG intends to start two amendments study groups for an emergency service i.e. 802.21.1 and multi-radio power Management i.e. 802.21c. IEEE 802.21c amendment shall define mechanisms to reduce power consumption of multi-radio mobile devices on heterogeneous IEEE 802.21 compliant networks.

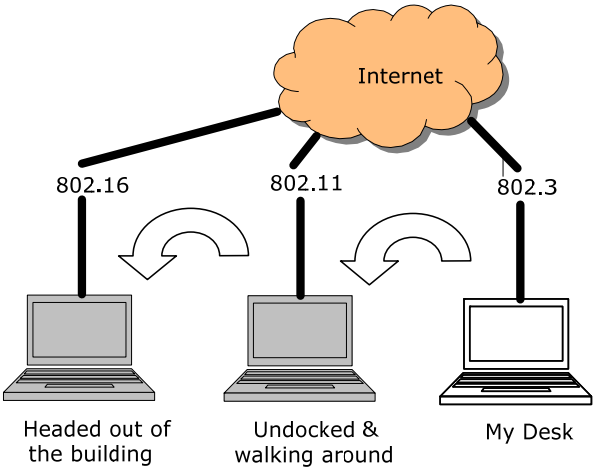



Fig. 5. 802.21 use case for 802.11 and 802.16

## 5. WiFi/WiMAX Integrated Service

The IEEE 802.11s and IEEE 802.16j groups are specifying standards for WiFi WLAN mesh networks and WiMAX MMR networks, respectively. We truly believe that one strong candidate for the next-generation FMC network architecture consists of optical networks, WiMAX relay networks, and WiFi mesh networks. An example of such architecture is shown in Figure 6.

At the top level, a dense wavelength division multiplexing (DWDM) optical ring forms the core of the metropolitan area network (MAN). At the middle level, WiMAX base stations (BSs) and relay stations (RSs) form a WiMAX MMR network that relays traffic from the lower level to the core network. The bottom level consists of WiFi mesh networks that provide high data rate connections directly to the end users. A transition point between different levels serves as a bridge between networks. Traffic is aggregated and disseminated through the transition points. Especially, the optical switch nodes on the WDM optical ring, which are also portal nodes in WiMAX MMR networks, transfer traffic between wireless networks. Similarly, each WiMAX BS/RS also serves as a portal node in WiFi mesh networks.

IEEE 802.11 WiFi currently has a wide deployment base. The emerging 802.16 broadband access therefore needs to be integrated with existing WiFi technology. There are two types of

integrated WiFi/WiMAX architectures. In the first architecture (Figure 7), WiFi and WiMAX provide complementary wireless access. Dual-mode WiFi/WiMAX users roam between WiFi hotspots and WiMAX base stations. Seamless handoff between WiFi and WiMAX is the main design issue in this integrated environment. In the second scenario, WiMAX is served as a backbone connection for WiFi access points (Figure 8). The wide bandwidth and long transmission range make WiMAX a flexible and cost-effective backbone solution for WiFi access points. Several WiFi access points are connected to the backbone network via a WiMAX base station.

The advantages of multi-hop wireless mesh networks are as follows. First, the connectivity range of the core wireline networks is extended. The concept of multi-hop relay not only extends the communication range of a portal node beyond the single-hop coverage but also relaxes the ties between mesh nodes and the infrastructure. The wireline infrastructure is replaced by wireless backhauls in wireless mesh networks. Second, the deployment of mesh nodes becomes easier and more flexible, and the deployment cost is much lower than the cost of building a wireline connection.

Third, the mesh networks are robust because the networks are interconnected by multiple links. In this section, we introduce the proposed channel-assignment traffic engineering methods for WiFi and WiMAX mesh networks. The differences between WiFi and WiMAX networks are in the Media Access Control (MAC) and the spectrum determine the mechanisms of channel assignment. The allocated spectrum for the WiFi system is completely operated in the license-exempt (unlicensed) band, while most WiMAX systems use the licensed band. As a result, WiMAX and WiFi have different MAC designs. The MAC in WiFi is the contention-based CSMA/CA, while the MAC defined in WiMAX is contention-free. WiMAX BSs schedule time slots to subscriber stations (SSs) (i.e., in the time domain). In addition, with Orthogonal Frequency Division Multiple Access (OFDMA) physical layer (PHY), BSs can allocate a subset of subcarriers to each SS (i.e., in the frequency domain).

Channel assignment can be executed in a centralized or distributed manner. In the centralized channel assignment approach, the channels are allocated by a controller that periodically collects the topology and traffic information from the mesh nodes. Then, the results of channel assignments are disseminated to all mesh nodes for further adjustment.
In the distributed approach, each mesh node decides which channels to use based on the local information, including the topology, channel conditions, and traffic conditions. The distributed approach is simpler and more robust than the centralized approach. Nevertheless, the low channel utilization and lower controlled usage of resources might degrade system performance.

Consequently, distributed approaches are suited to the license-exempt band because the same channel can be used by other systems. Immediate monitoring of variations in channel conditions is more effective than attempting to control the resources. For the licensed spectrum, the centralized approach is recommended simply because it has better performance. In the next two subsections, we present the proposed centralized and distributed channel assignment schemes for WiMAX MMR and WiFi mesh networks.

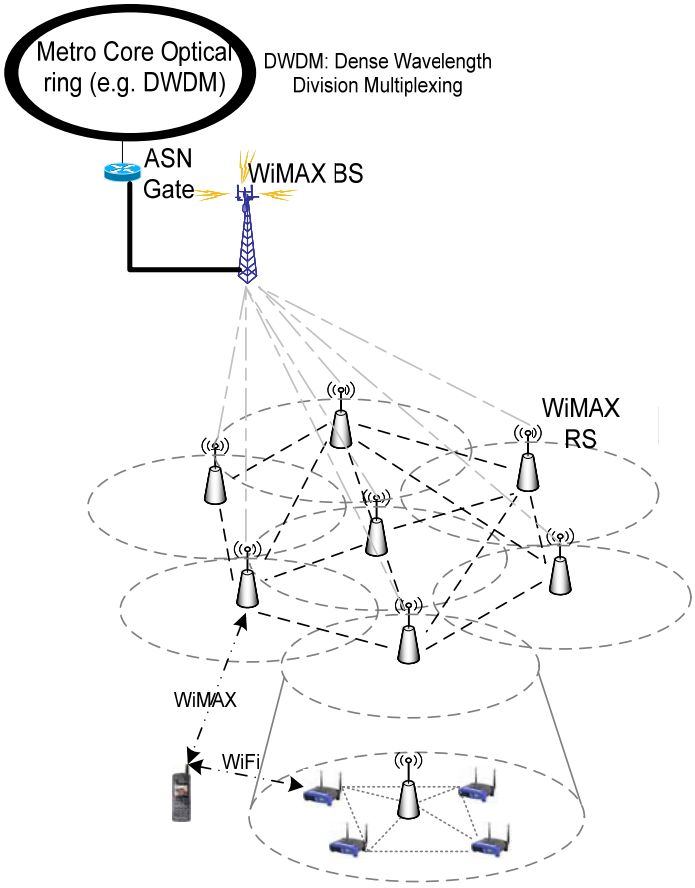



Fig. 6. Heterogeneous network architecture of FMC consisting of optical ring, WiMAX, and WiFi mesh networks.

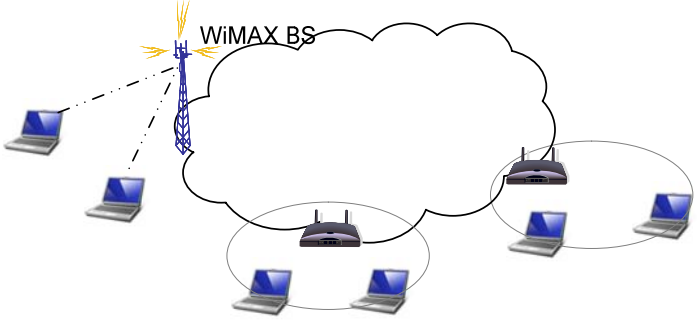



Fig. 7. Integrated WiMAX/WiFi wireless access.

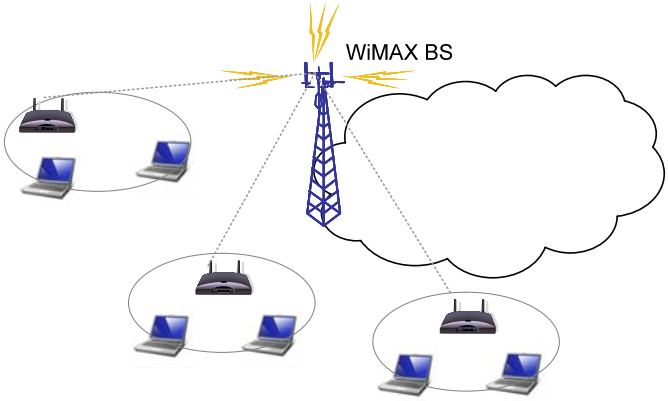



Fig. 8. WiMAX as backbone connection for WiFi access points.

## 6. WiMAX/WiFi Vertical Handover (VHO)

Together, WiMAX and WiFi are ideal partners for service providers to deliver convenient, affordable mobile broadband Internet services in more places. Both are open IEEE wireless standards built from the ground up for Internet Protocol (IP)-based applications and services. However, both techniques have their own sets of advantages and disadvantages. While IEEE 802.11 has accelerated the network deployment for providing high transmission rate in limited geographical coverage, while IEEE 802.16 offers more flexibility in while maintaining the technology's data rate and transmission range. The limited coverage range of WiFi makes it difficult to meet the future ubiquitous networks need while IEEE 802.16 can provide high speed Internet access in wide area. In this chapter we introduce a new combination of WiMAX and WiFi (called WiWi) (Rashid et al., 2008) to create a complete wireless solution for delivering high speed Internet access to businesses, homes and hotspots. This integration is to avoid the disadvantages of each technique by the other and using the advantages optimally. For example, WiFi may offer a high data rate (up to 500Mb/s is envisaged), but it is power limited due to the use of unlicensed band and are therefore much more confined in coverage, while on the other hand, even though WiMAX is data rate limited (up to 70Mb/s fixed), it can provide extensive coverage much like the cellular systems. Because of that it is instructive to employ both technologies in the laptop/mobile for ubiquitous connection with high data rates (Inayat et al., 2007).

As shown in Figure 9, the WiWi module comprised of physical and MAC layers for both WiMAX and WiFi. For the backhaul the signal is uses WiMAX PHY and MAC protocols, while in the access the signal is uses WiFi PHY and MAC protocols. For the local traffic the signal do not need to pass through the WiMAX protocols. For the Internet traffic the signal need to pass through both WiFi and WiMAX protocols from the access to backhaul and vice versa. The traffic need switch function to recognize the local and Internet traffic and reroute the traffic according to the switch decision, this will be discussed in section 5.1.

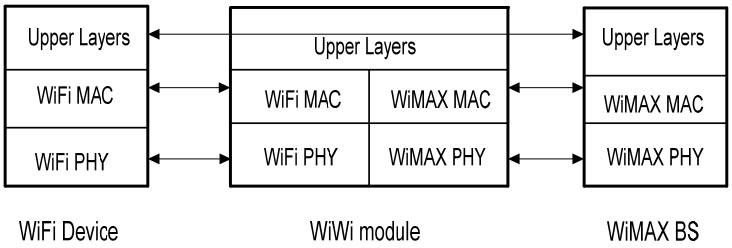



Fig. 9. The WiWi protocol Stack with the WiMAX and WiFi components.

## 7. WiWi Deployment Models and Scenarios

### 7.1 WiWi Module Structure

The needs of Internet for rural communities and remote sites using the available broadband access technologies, motivate for deploy new systems design and applications. The widely available, and highly cost-reduced, WiFi hardware meets the cost target for rural applications. However, the current 802.11 MAC protocol has fundamental shortcomings when used over long distances under the power limit enforcement by the local regulators.
The wireless long distance has been studied based on WiFi extensively in WiFiRe (Krishna et al, 2007) and WiLDNet (Robin et al, 2006).

Krishna et al. introduce a new system for rural broadband voice and data access, based on the WiFi PHY, and a new single channel multisector TDM MAC using directional antennas called wireless fidelity - rural extension (WiFiRe) where the concept of wireless communication over WiFi IEEE 802.11b physical layer (PHY) and WiMAX IEEE 802.16 MAC layer is discussed and implemented. As well known, IEEE 802.11b PHY has better availability of low cost chip sets which can operate on unlicensed 2.4GHz frequency band while WiMAX has potential to work over larger distances of 30-40km range. WiFiRe is support maximum of 25Mbps data rate includes both UL and DL and two parallel transmissions by opposite sectors. The design was tested using VOIP simultaneous calls in all STs the capacity was improved about 3.5 times the conventional IEEE 802.11b.

WiLDNet is WiFi-based Long Distance (WiLD) networks with mesh link configuration have the potential to provide connectivity at substantially lower costs than traditional approaches based on IEEE 802.11 with several essential changes to the 802.11 MAC protocol, but continues to rely on standard WiFi network cards. These changes are to overcome the short distance and the random contention-based access of the IEEE802.11. WiLDNet used an adaptive loss recovery mechanism using FEC and bulk acknowledgements.

In this chapter we discuss wireless broadband system using 802.11 in the access and point-to-multipoint (PtMP) WiMAX-based TDD/TDMA technique on each (downlink) carrier at the backhaul in addition to signaling information to control the allocation of upstream traffic channels. The discussion includes the key issues of MAC system and architecture design and performance analysis for the protocol conversion from WiMAX/TDMA frames to WiFi/contention-based packets and vise versa, and the delay/jitter associated.

The Broadband Access PtMP System provides wireless broadband access to the operator service nodes (POPs) from remote customer locations. As shown in Figure 10, each remote location is served by a peripheral station that is able to offer a variety of service interfaces to end users. The AP has a two-way full duplex radio link to the BS, which is connected to the core network with different types of trunk interfaces. Subscribers can have access to the full range of services by means of various standardized user-network interfaces (i.e. Ethernet, WiFi, dial-up, etc). The BS provides also standard core network interfaces, to transparently connect subscribers to the appropriate service node. The system allows services to be provided to a number of subscribers, ranging from a few tens to many hundreds of users in the same area, and over a wide range of distances (depending on the RF frequency used). In this section we will concentrate on the design of the backhaul network interface which is combines of WiFi and WiMAX.

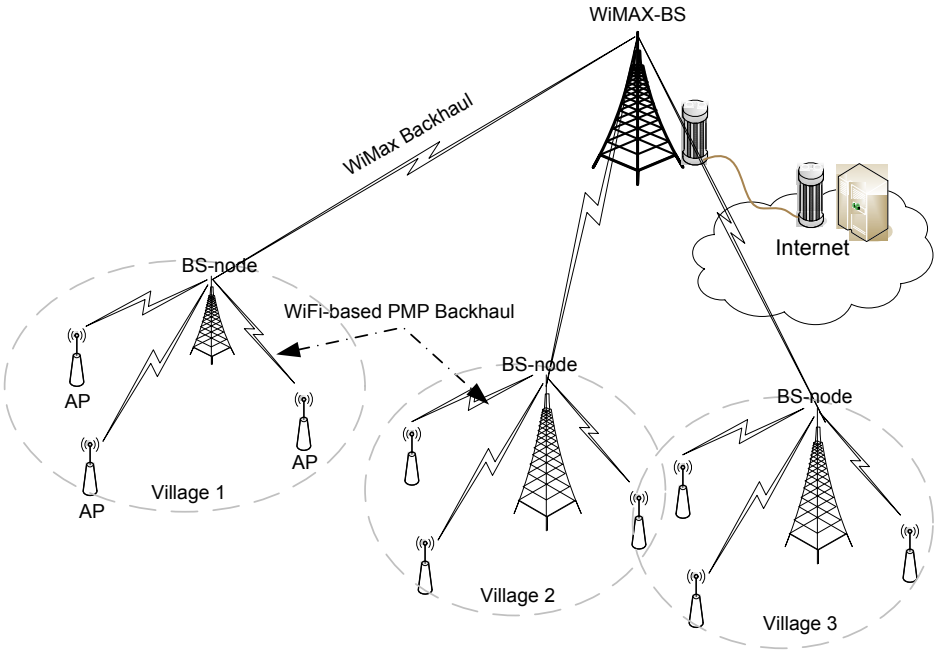



Fig. 10. wireless broadband network architecture

The system will allow a point-to-multipoint (PtMP) topology to cover a wide highly populated metropolitan area collecting various kinds of services. Moreover, with the PtMP scenario each cell, served by a central BS, can be split into a variable number of sectors using directional antenna, thus allowing efficient reuse of allocated spectrum and increasing system capacity.

Each AP is equipped with a high gain directional antenna in line or non-line of sight with the BS. It receives the multiplexed traffic channels and de-multiplexes the information directed to the user served by the AP (Koutsakis and Paterakis, 2005). In its turn it transmits back to the BS, on a separate frequency (uplink), traffic and signaling information. Uplink transmission is slot based TDMA that is each AP is assigned a time slot to transmit in the 802.16d frame. One BS can support up to 150 Mbps of traffic payload per direction per sector (Su et al., 2003). Figure 11 is shows the protocol architecture for the proposed network. The WiMAX BS is associated with gateway which is provides multi interfaces to the internet (FDDI, DSL, 3G, etc) (Tang, 2004). As shown in the figure, the BS is IEEE802.16 layer-2 device, PHY layer based OFDM modulation baseband and MAC based fixed-WiMAX/TDMA techniques to provide time slot sharing between the APs. The APs are also WiWi-based layer-2 devices, has two interfaces, the backhaul interface connect with the BS through the backhaul and interface to the end-user.

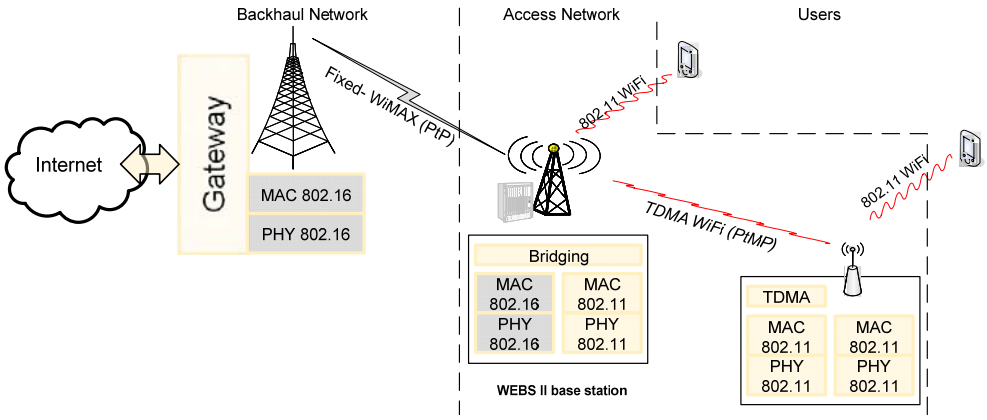



Fig. 11. wireless broadband system description and protocol Architecture

The access point (AP) need protocol converter/bridge to convey the data from the access to the backhaul. Figure 12 shows the proposed protocol converter (802.16/11 bridge), which is comprises of traffic classifier, scheduler and protocol encapsulation/de-encapsulation. Local traffic is the traffic from/to the access network under the same local community also called intra access point traffic. The external traffic is the traffic from/to Internet. First the traffic is classified; then the traffic is scheduled in to two queues based on the traffic types. A simple scheduler can be used here i.e. round robin or FIFO. Finally, the Internet traffic needs protocol encapsulation due to the different air interfaces. WiFi packets are encapsulated in the WiMAX frame for the upload link and vice versa for the download traffic.

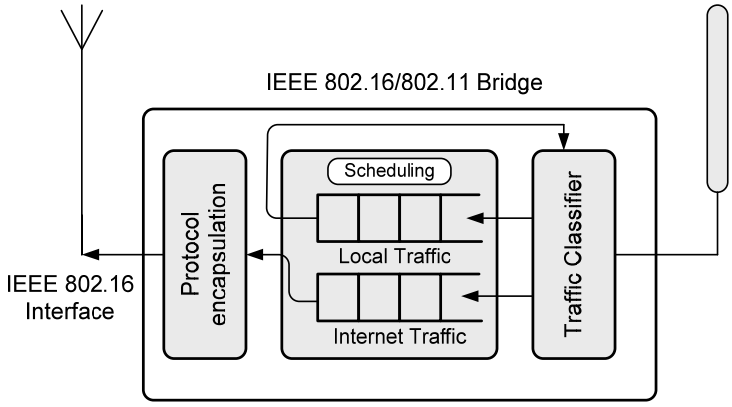



Fig. 12. the proposed protocol converter/bridge between IEEE 802.16/802.11

### 7.2 WiWi Mobile Multihop Module

In this section, we incorporate the WiWi module at the mobile station site. Where WiWi module is works as a relay module (a.k.a. mobile relay node). The mobile relay node (WiWi-based device) receives signal from the WiMAX network (i.e. BS) and convert it to WiFi signal and relay this signal to the WiFi-based devices (i.e. laptop). The scope of this section is using the WiWi module to increase the WiFi-based devices range when the user moves out of the WiFi coverage. The main advantage of the WiWi module is the WiFi-based devices are no longer need to have WiMAX SIM (Subscriber Identification module) module. Since, the WiMAX connection to the operator is not free, so equipped all the devices with WiMAX SIM module is economically not feasible. However, since WiFi is unlicensed and free access, so the better is to equip one device (WiMAX mobile) with WiWi module and relay the signal to the other WiFi-based devices (laptop, PDA, etc). This scenario can be in open/public mode which means the WiWi module will relay the signal to all other devices around; similar to the mobile ad-hoc network (MANET) architecture. Or can be in close/private mode which is offer services to its own devices using specific connection permission and identification. This will allow Internet connection continuity for WiFi-based devices out of WiFi service provider/operators hotspots coverage for free and the cost will be paid to the WiWi-based devices battery life.

In Figure 13, there are two possible scenarios; (a) the WiFi-based devices (i.e. laptop) within WiWi and WiFi networks coverage and (b) the WiFi-based devices out of WiFi network coverage. In the scenario (b) the process is straightforward in which the WiFi-based device sees only one network connection, which is provided by WiWi module. Directly the device will start network entry process to join WiMAX network using WiWi module installed at the WiWi-based device (i.e. WiMAX mobile). On the other hand, the scenario in Figure 13(a), there is two networks available, WiMAX (using WiWi module) and WiFi network. The WiFi-based devices should select only one of them to access the Internet based on the received signal strength indicator (RSSI) message received from the AP/BS. In this scenario using the conventional handover for network selection based on RSSI message is problematic, because WiFi-based devices sense two WiFi signals, the first signal from WiWi-based device and the other from WiFi access point (AP) (Li-Chun et al., 2007).

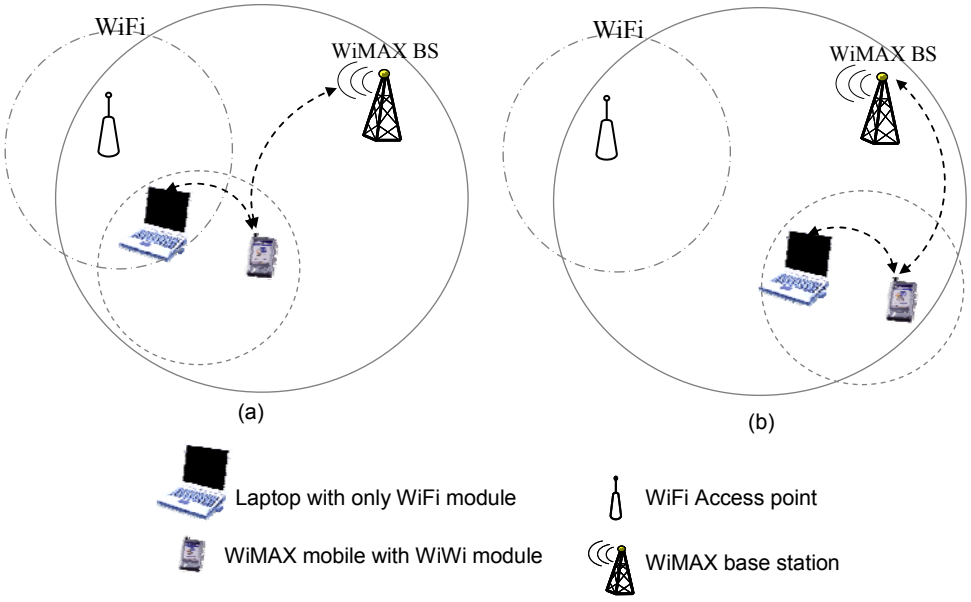



Fig. 13. WiMAX and WiFi coverage in indoor and outdoor (a) the laptop in the WiFi coverage (b) laptop is out of WiFi coverage.

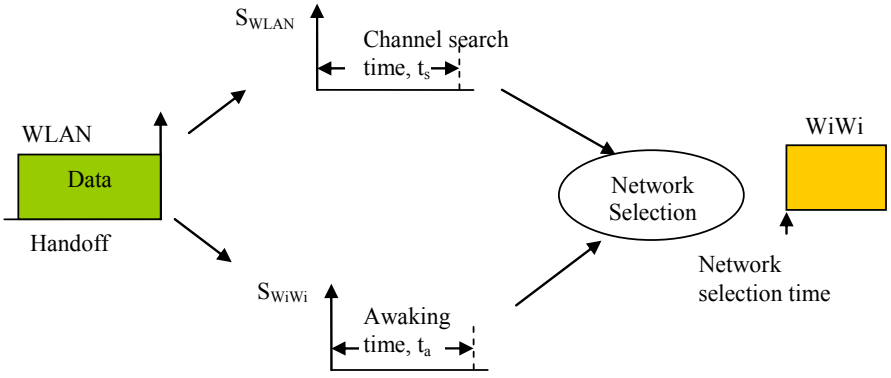



Fig. 14. System model for jointly vertical handoff (WiFi to WiWi).

The difference between this scenario and the conventional scenario is that the decision criteria are not the same due to the difference in the technology, in which the WiWi is virtual WiFi technology which only the interface is the WiFi, which is relay (regenerate) the WiMAX signal to WiFi format. So, the WiFi-based device by using the conventional handover/roaming protocols will lead to wrong decision and results of non-optimized throughput. Usually, the throughput of the WiWi network is lower than WiFi throughput, due to the degradation results from multihop relay in the WiWi network (Rashid et al., 2008). RSSI is one of the important criteria for AP selection and handover. So, in WiWi cased the RSSI value alone is not sufficient to be decision criteria for handover. For example in

close/private mode, all WiFi-based devices are closer to WiWi module than the AP. So, the RSSI from WiWi signal is usually greater than the signal from AP (WiFi network) signal. This is shown in Figure 14. In this chapter we propose new algorithm for the network selection based the total throughput on addition to RSSI criteria.

Figure 15 shows the proposed handover decision based on RSSI and throughput. The total throughput $\phi(\lambda_{WiWi})$ for WiWi network is calculated based on WiMAX throughput $\phi(\lambda_{WiMAX})$ and the effect of vertical handover within the WiWi module. This is done as follows:

$$\phi(\lambda_{WiWi}) = \sum_{n} \phi(\lambda_{WiMAX} - c_{WiMAX} / \sqrt{n} - \psi) \tag{1}$$

where $c_{WiMAX} / \sqrt{n}$ is the degradation of throughput due to multihop relay (Zhang et al. 2003). $c_{WiMAX}$ is the WiMAX capacity and $n$ is the number of hops (i.e. in WiWi, $n = 2$). $\psi$ is the total throughput loss due to protocol converter and interference between WiFi and WiMAX. Figure 16 shows the flow of the messages for connecting WiFi devices the WiWi network. Here we assume the WiWi in the closed mode so that there are no other nodes sharing the channel with the WiFi device. The $\phi(\lambda_{WiMAX})$ throughput depends on the WiMAX channel bandwidth (i.e. 1.25-20MHz) and total number of tones (256 for fixed and 1024 for mobile) and channel quality (Su Yi et al., 2003).

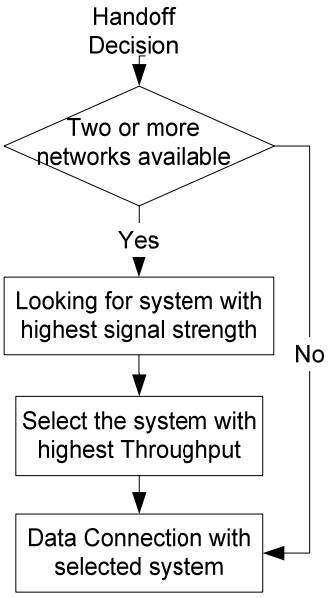



Fig. 15. Proposed network selection scheme in the WiFi to WiMAX handoff scenario

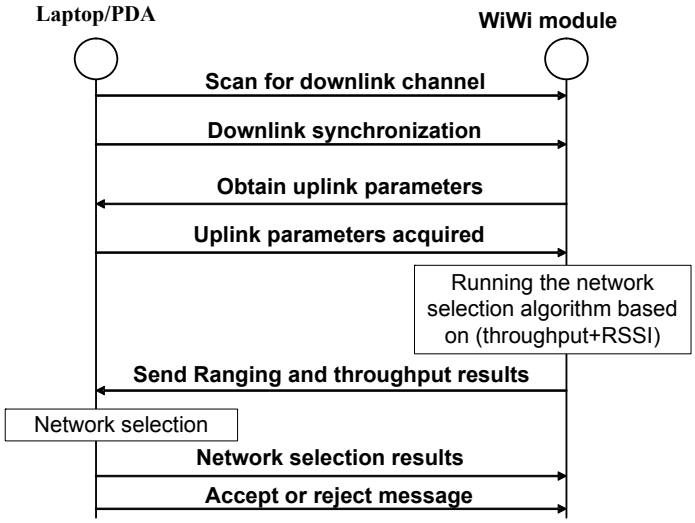



Fig. 16. Proposed signalling process in the WLAN to WiMAX handoff scenario

Figure 17 shows the time delay WiMAX and WiFi internetworking, which shows the delay time for the proposed method is less than using MIH and MIS.

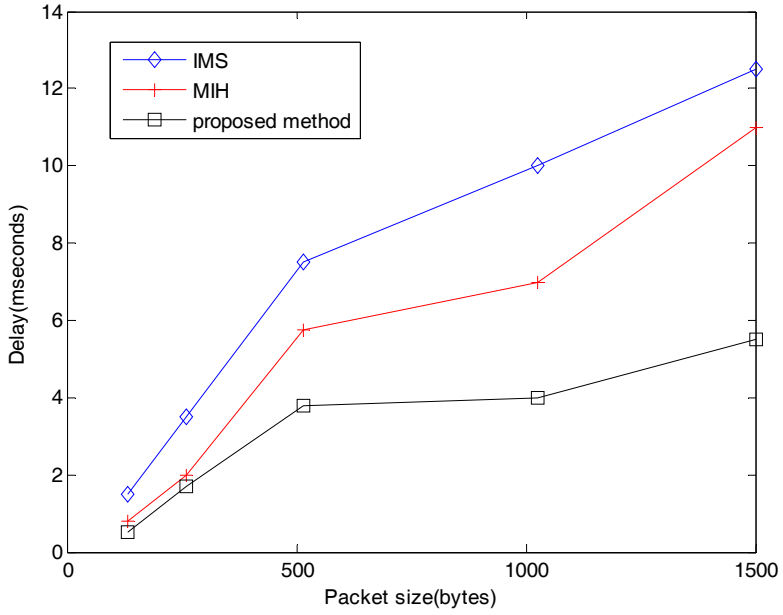



Fig. 17. Delay for the proposed vertical handover using WiWi module delay compared with MIH and IMS.

## 8. Conclusions

In this chapter the integration between WiFi and WiMAX networks is discussed. Some of potential technologies for integration include IEEE 802.11u (IEN), IEEE 802.21 (MIH) and IEEE 802.16.4 (WirelessHUMAN) are also presented. A new module called WiWi is proposed for integration and two scenarios are discussed. The first scenario is the infrastructure scenario which WiWi module allocated in the access point (AP). The second scenario is the mobile multihop scenario which the WiWi module allocated at the mobile station. Finally, a new vertical handover algorithm for WiWi network is proposed. The objective of the new vertical handover is to solve the virtual WiFi problem that introduced by the WiWi module. The proposed algorithm is performed at a new logical adaptation layer which allocated above the MAC layer. The results of the proposed algorithm shows less overhead in the protocol conversion compared with the other conventional convergence protocols IP multimedia subsystem (IMS) and media independent handover (MIH).

## 9. References

3GPP TS 23.206 (currently Release 7 draft), "Voice Call Continuity between CS and IMS."

Behmann F., (2005) "Impact of Wireless (Wi-Fi, WiMAX) on 3G and Next Generation: An Initial Assessment" Electro Information Technology, 2005 IEEE International pp: 1-6.

Berlemann, L., Hoymann, C., Hiertz, G.R., Mangold, S. (2006) "Coexistence and Interworking of IEEE 802.16 and IEEE 802.11(e) ", IEEE 63rd Vehicular Technology Conference, 2006. (VTC 2006-Spring). Volume 1, Page(s):27 – 31.

Dhawan S., (2007) "Analogy of Promising Wireless Technologies on Different Frequencies: Bluetooth, WiFi, and WiMAX" Wireless Broadband and Ultra Wideband Communications, 2007 (AusWireless2007). The 2nd International Conference on pp: 14 – 17.

F. Cacace and L. Vollero, (2006) "Managing Mobility and Adaptation in Upcoming 802.21 Enabled Devices," Proc. 4th Int'l. Wksp. Wireless Mobile Apps. and Svcs. on WLAN Hotspots, Los Angeles, CA, Sept. 29–29, pp. 1–10.

Gaitan O.S., Martins P., Tohme S., Demerjian J., (2005) "SIP Embedded Attribute Certificates for Service Mobility in Heterogeneous Multi-Operator Wireless Networks" Vehicular Technology Conference, 2007 (VTC2007), pp2000 – 2004.

IEEE Std 802.11u, (2009) "Information Technology — Telecommunications and Information Exchange between Systems — Local And Metropolitan Area Networks — Specific Requirements — Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Wireless Interworking with External Networks."

IEEE Std 802.16.4, (2008), "IEEE 802.21: Media Independent Handover Services", Approved in Nov. 2008.

IEEE Std 802.16.4, (2009), "IEEE 802.16 Task Group 4 (WirelessHUMANTM Task Group)"wirelessman.org/tg4/, accessed on March 2009.

IEEE Std 802.16g, (2009) "Information Technology — Telecommunications and Information Exchange between system- Local and Metropolitan Area Networks — Specific Requirements — Part 16: Air Interface for Fixed Broadband Wireless Access Systems: Management Plane Procedures and Services."

Inayat, R, Aibara, R, Nishimura, K, (2007) "A seamless handoff for dual-interfaced mobile devices in hybrid wireless access networks", Advanced Information Networking and Applications (AlNA2007), Volume: 1, pp 373-378.

Inayat, R, Aibara, R, Nishimura, K, (2007) "A seamless handoff for dual-interfaced mobile devices in hybrid wireless access networks", Advanced Information Networking and Applications, AlNA 2007, Volume: 1, Pages: 373-378.

Jong-Ok Kim, Shigeno H., Yamaguchi A., Obana, S., (2004) "Airtime-based Link Aggregation at the Co-existence of WiMAX and WiFi", Personal, Indoor and Mobile communication journal, pp. 51-58.

K. Sundaresan, H. Hsieh, and R. Sivakumar (2004) "IEEE 802.11 over Multihop Wireless Networks: Problems and New Perspectives" Ad Hoc Networks Journal.

Koutsakis, P., and Paterakis (2005), "On multiple traffic type integration over wireless TDMA channels with adjustable request bandwidth", International Journal of Wireless Information Networks 7, PP 55–68.

Krishna Paul, Anitha Varghese, Sridhar Iyer, Bhaskar Ramamurthi, Anurag Kumar, (2007) "WiFiRe: Rural Area Broadband Access using the WiFi PHY and a Multisector TDD MAC", IEEE Communications Magazine, special issue on Telecommunications in Emerging Economies, Vol 8 (10), pp-23-30.

Lampropoulos G, Salkintzis K, Passas N, (2007) "Media-independent handover for seamless service provision in heterogeneous networks" IEEE Communications Magazine, V(46), Issue: 1 pp 64-71.

Li-Chun Wang, Anderson Chen, Hung-Hsi Chen, (2007) "Network Selection with Joint Vertical and Horizontal Handoff in Heterogeneous WLAN and Mobile WiMAX Systems", Vehicular Technology Conference, 2007 (VTC2007), pp 794-798.

Man Cheuk Ng Vijayaraghavan, M. Dave, N. Arvind Raghavan, G. Hicks, J. (2007), "From WiFi to WiMAX: Techniques for High-Level IP Reuse across Different OFDM Protocols" 5th IEEE/ACM International Conference on Formal Methods and Models for Codesign, 2007 (MEMOCODE 2007) page(s): 71-80.

MIMOS (2009), www.mimos.my, accessed in Dec 2009.

Niyato D., Hossain E., (2004) "Wireless Broadband Access: WiMAX and Beyond - Integration of WiMAX and WiFi: Optimal Pricing for Bandwidth Sharing" IEEE Communications Magazine, pp: 140-146, V (45), Issue 5.

Pahlavan K, Krishnamurthy P, Hatami A; Ylianttila M, Makela J.P, Pichna R, Vallstron J, (2007) "Handoff in Hybrid Mobile Data Networks", Mobile and Wireless Communications Summit, V (7), Issue: 2, pp 34-47.

Q. Zhang et al. (2003), "Efficient Mobility Management for Vertical Handover between WWAN and WLAN," IEEE Commun. Mag., Vol. 41 (11), pp. 102- 108

Rabin Patra, Sergiu Nedevschi, Sonesh Surana, Anmol Sheth (2008) "WiLDNet: Design and Implementation of High Performance WiFi Based Long Distance Networks", NSDI, pp 86-91.

Rabin Patra, Sergiu Nedevschi, Sonesh Surana, Anmol Sheth, Lakshminarayanan Subramanian, Eric Brewer, (2007) "WiLDNet: Design and Implementation of High PerformanceWiFi Based Long Distance Networks" 26th IEEE International Conference on Computer Communications (INFOCOM 2007), Pp:312 – 320.

Rashid A. Saeed, Hafizal Mohamad, B.M. Ali, (2009) "Vertical Handover Evaluation for WiWi Network based on Overall Throughput", accepted in Praise Worthy Prize, Italy.

S. Xu and T. Saadawi (2001) "Does the IEEE 802.11 MAC Protocol Work Well in Multihop Wireless Ad hoc Networks" IEEE Communications Magazine, 39(6):130–137.

Su Yi, Yong Pei, Shivkumar K. (2003) "On the capacity improvement of ad hoc wireless networks using directional antennas" International Symposium on MobiHoc, pp 108–116.

Surducan, E. Iancu, D. Surducan, V. Glossner, J. (2007) "Miniature multiband antennas for hand held WIMAX and WiFi application" Electromagnetics in Advanced Applications, (ICEAA 2007), pp 13-16.

Tang C., (2004) "Evolving to wireless and wireline convergence - an overview of IMS" Wireless and Optical Communications, 2005. (WOCC2005), pp 27-29.

V. Gunasekaran and C. Harmantzis, (2005) "Affordable Infrastructure for Deploying WiMAX Systems: Mesh v. Non Mesh" In IEEE Vehicular Technology Conference (VTC), Stockholm, Sweden.

Zlydareva, O. Sacchi, C. (2008) "Multi-Standard WIMAX/UMTS System Framework SDR", IEEE Aerospace Conference, page(s): 1-13.

17

# System-Level Modelling of Heterogeneous System-on-Chip Architectures Targeting Multi-Standard Wireless Applications

Ali Ahmadinia[1], Balal Ahmad[2] and Tughrul Arslan[2]
[1]*Glasgow Caledonian University,* [2]*University of Edinburgh*
*United Kingdom*

## 1. Introduction

In this chapter, we propose a novel framework for system-level modelling of heterogeneous SoC (System-on-Chip) architectures with emphasis on reconfigurable components targeting multi-standard wireless applications including WiMAX (Worldwide Interoperability for Microwave Access) standard (IEEE 802.16e, 2005). System-level modelling for system-on-chip and hardware-software co-design has been gaining importance due to rising complexity of such systems, continual advances in semiconductor technology, and the productivity gap in design of complex systems.

The IEEE 802.16 standard, WiMAX technology, is a wireless broadband access standard which helps make the vision of pervasive connectivity a reality. WiMAX technology can provide up to tens of Mbps symmetric bandwidth over many kilometres. This gives WiMAX a significant advantage over other alternatives like Wi-Fi and DSL. The physical layer of WiMAX includes both downlink (channel coding and IFFT) and uplink (FFT, modulation and decoding) data processing. WiMAX standard is rapidly proving itself as a technology that will play a key role in fixed broadband wireless metropolitan area networks. The scalable architecture, high data throughput and low cost deployment make Mobile WiMAX a leading solution for wireless broadband services.

The physical layer of WiMAX provides the wireless means of transmitting raw bits among devices. At present, the WiMAX standard has been implemented on DSP or ASIC technologies such as the WiMAX solution based on the Freescale MSC8 126 DSP (Freescale, 2005) and Intel NetStructure WiMAX Baseband Card (Intel, 2006). These solutions cannot satisfy all the features required by current and future embedded systems, including higher performance, higher power efficiency and higher flexibility. Filling the gap between the high flexibility of DSPs and the high performance of ASICs, reconfigurable architectures offer an alternative solution for embedded systems. This chapter extends our work in (Ahmadinia et al., 2007) and presents specially tailored high performance and low power architectures, based on newly emerging custom reconfigurable cores and communication centric platforms, for the WiMAX physical layer. To address this, we have proposed a system level framework that enables automatic generation and optimization of heterogeneous SoC

architectures based on communication centric platforms. The WiMAX 802.16-2004 physical layer was ported on this platform.
In WiMAX physical layer, apart from the usual functions such as randomization, forward error correction (FEC), interleaving, and mapping to QPSK and QAM symbols, the standard also specifies optional multiple antenna techniques. This includes space time coding (STC), beam forming using adaptive antennas schemes, and multiple input multiple output (MIMO) techniques which achieve higher data rates. The OFDM modulation/demodulation is usually implemented by performing FFT and inverse FFT on the data signal.
Consequently, FFT and Viterbi are considered in this work are modelled in SystemC as hardware accelerators which are compute intensive modules in WiMAX receiver computing chain. The WiMAX application runs as a software core on the ARM7 (ARM, 1999) processor. A Reconfigurable FFT IP and a Viterbi decoder as hardware accelerators are integrated into the WiMAX chain, to validate their performance by co-simulating modules from different levels of abstraction.

## 2. Related Work

In (Rissa et al., 2002), a modelling technique for reconfigurable systems is proposed based on OCAPI-xl – a modelling language similar to SystemC. A set of processes is used to model a run time reconfigurable process which is supervised by a HardWare Scheduler (HWS). In (Noguera & Badia, 2003), a modelling methodology for reconfigurable architectures based on discrete event systems is presented. In an object-oriented manner discrete event classes and objects form a reconfigurable system, which introduces all necessary modelling capabilities to cover run time reconfiguration. In (Enzler et al., 2003), a co-simulation environment is unveiled enabling performance estimation for a dedicated architecture, which couples a CPU together with a coarse grained, multi-context reconfigurable unit (RU). Performance evaluation is based on a cycle-accurate simulation including execution and reconfiguration times. The CPU behaviour is modelled by an extension of a CPU simulator, while the RU itself is implemented in VHDL. The approach addresses the impact of multi-context RUs and focuses on the coupling of a multi-context device and a controlling CPU. In (Qu et al., 2005), a method is proposed to model dynamically reconfigurable hardware in SystemC. For this purpose, a special component is introduced called dynamically reconfigurable fabric (DRCF), which includes a configuration scheduler, an input splitter, and an associated configuration memory. After analyzing the SystemC modules, the DRCF code is constructed using a template. During this process, the reconfigurable functionality is identified and integrated into different configurations within a DRCF. The DRCF modification enables modelling of context switching and bus traffic. A system-level framework called, Perfecto, for rapid explorations of different reconfiguration alternatives and performance evaluation is presented in (Liao & Hsiung, 2005).
Through additional information, e.g. task execution time or configuration time, Perfecto is able to estimate aspects like total execution time and overall area usage. In (Brito et al., 2007) a technique is proposed for dimensioning dynamic and partially reconfigurable FPGA area by using a SystemC approach.
Unlike above approaches, our methodology is applied to a real world application, namely WiMAX. Furthermore, our modelling captures the reconfigurable features of cores, not just

instantiating different components. Also, our models have been verified in a hybrid simulation framework.

## 3. System-Level Reconfigurable Core Modelling

System-level modelling for system-on-chip and hardware and software co-design has been gaining importance due to rising complexity of such systems, continual advances in semiconductor technology, and the productivity gap in design of complex systems.
In the past few years, the introduction of many transaction level languages such as SystemC and SpecC has been seen (Open SystemC Initiative, 2008). Transaction level modelling (TLM) is a top-down approach to system design widely adopted in the verification phase of the design. In TLM, the system is first described at a high level of abstraction where communication details are hidden, and then the description is refined with the necessary details (Cai & Gajski, 2003).
The key concept of TLM is the separation of functionality definitions from communication details, by using the concept of channel. A channel allows communication through methods calling. Then, two functional components, interconnected with a channel, communicate calling the methods exported by the channel. In this way, the communication detail is hidden in the channel definition. SystemC enables TLM through *sc_interface*, *sc_channel*, and *sc_port*. A channel interface can be derived from *sc_interface*, then the *sc_interface* can be implemented in a *sc_channel*. At this point, a *sc_module* can communicate through a sc channel, using a sc port connected to the sc channel.
WiMAX is one of crucial wireless standards which requires SoC implementation platforms. In general, WiMAX as a whole needs reconfigurable architecture, as the system should adjust a significant number of parameters based on the volume of data, data rate, type of modulation, number of channels, etc.
In the proposed framework, the hardware component has been described, using the SystemC description language, as a C++ class derived from the sc module class. Due to the fact that no *sc_module* object, neither any objects derived from this class, can be initialized after the beginning of the simulation process, it is not possible to model a module reconfiguration process as a new sc module initialization during the simulation phase. The approach followed in the framework to model the system reconfiguration using SystemC is based on the fact that the functionalities of a *sc_module* could be implemented as *sc_methods/sc_threads*. Therefore by using one of them as a wrapper to call a function pointer, we are allowed to change the functionality of modules, by varying only this function pointer.
Different to other approaches, that implement the complete reconfigurable area or the reconfigurable modules as buses, the reconfigurable modules remain in their original scope and the surrounding design must be changed only a little. Obviously the topology of a static design remains basically unchanged if some parts are made reconfigurable. The straightforward way to achieve this is to intercept the communication between static and reconfigurable modules at the channels, which interconnect those parts. The current configuration is simulated by forwarding channel events only to the currently active modules by using dynamic switches. Additionally, all channel accesses of inactive modules are suppressed. If the modules are either sensitive to certain events or make use of blocking channel accesses, such a system will behave like a reconfigurable design during simulation.

Still this is not a completely satisfying solution, since for every channel type a new switch has to be built. This leads to the concept of a reconfigurable interface (*sc_rec_interface*), which provides a framework for building these switches.
A *sc_rec_interface* is a special switch, designed to connect a static channel with a port of a reconfigurable module, as shown in Figure 1. During the simulation, accesses to the corresponding port from within the reconfigurable module are then forwarded to the static channel. Additionally, any required events, the reconfigurable module is listening to (via sensitivity or dynamic *wait()* statements), is forwarded from the static channel to the module. Multiple reconfigurable modules can be bound to a single *sc_rec_interface*. During the simulation, the actual reconfiguration operations change the data-flow through the interface depending on the reconfiguration state of the connected modules. If all ports of a reconfigurable module are equipped with *sc_rec_interface*, no port can be triggered from the outside, if the module is inactive (not configured).

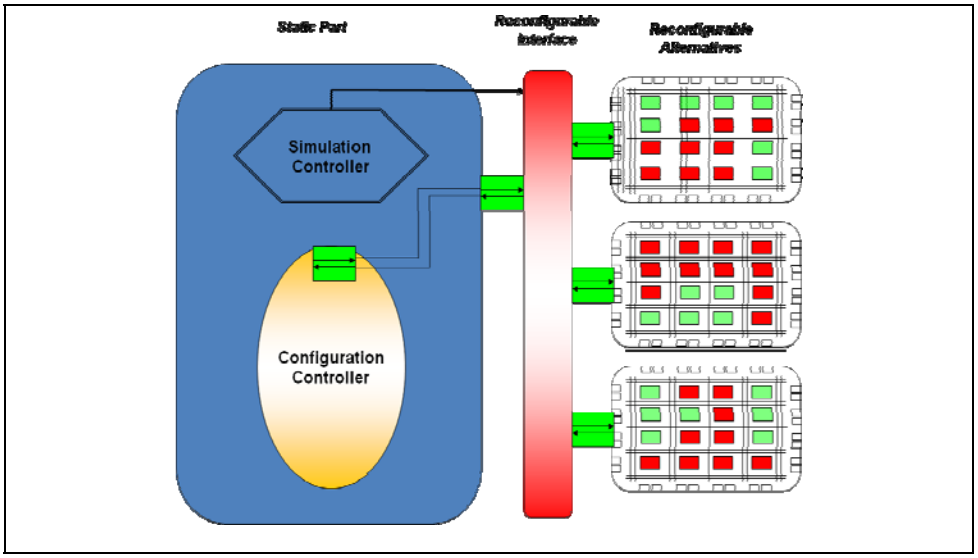



Fig. 1. A TLM-based Reconfigurable Design

Therefore, no outbound traffic should occur any longer, since the module's processes are no longer triggered. Nevertheless, technically it is possible that a module keeps on triggering itself (for instance by a member of type *sc_clock*). In this case outbound traffic is suppressed and a warning is reported to the designer.
For example, in the following code a manual generation of a reconfigurable module is shown. Module Smod reconfig is derived from reconfig module and static module Smod.
The constructor starts the finite state setup to machine that takes care of the reconfiguration state of the module and calls reconfig reset the module at start-up. In member function
reconfig setup a is registered to be reset to zero and b to be preserved during reconfiguration. Modelling of configuration times is split into loading and activation delay. Integrating a reconfigurable interface into a design is done analogously to the integration of a channel. Here the usage of a FIFO reconfigurable interface is shown.

```
class Smod_reconfig:
public reconfig_module,
public Smod {
protected:
inline void reconfig_setup();
public:
Smod_reconfig( sc_module_name _name ):
Smod(_name)
{
recofig_init();
//Initialize reconfiguration
//module functionality
reconfig_setup();
// Setting up reconfiguration
}
};

virtual inline
void Smod_reconfig::reconfig_setup() {
reconfig_reset<int>(a,0);
// reset a to 0
reconfig_store<sc_signal<int > > (b);
// store b
set_config_time(sc_time(25,SC_MS));
// module needs 25 ms to configure
set_exec_time((sc_time(1.5,SC_MS));
// and 2 ms to execute
set_removal_time((sc_time(2,SC_MS));
// storing variables
}

Smod_reconfig module1, module2, module3;
// instantiate reconfigurable modules
sc_fifo< int > fifo;
// instantiate (static) FiFo
sc_rec_interface<sc_fifo_out<int>>
rec_int1;
// instantiate reconfigurable interface
// for sc_fifo_out<int> port
rec_int1.static_port( fifo );
// connect the static channel
rec_int1.bind( module1.out );
// bind reconfigurable modules' ports
//to interface
rec_int1.bind( module2.in);
rec_int1.bind( module3.in);
```

## 4. WiMAX Application

To evaluate the proposed methodology and SystemC reconfigurable cores, WiMAX is chosen as a real application for the SoC design. WiMAX technology is standard widely known as worldwide wireless networking standard, the technology that addresses the interoperability issues proposed by the IEEE 802.16.

The WiMAX technology is globally adopted and it has unlimited demand in the new markets for wireless chips. In this work, the receiver section of the WiMAX system is the area of interest. The main constraint in the receiver section is the performance. There are upcoming trends to commercialize the technology, it evidently deserves attention while determining infrastructure of the receiver sections. The system designers are looking into the type of tools that would help them to design the WiMAX system (WiMAX Forum, 2008).

The typical WiMAX receiver measurements includes measurements of sensitivity, the maximum input level, adjacent- channel and alternate-channel rejection, reference timing accuracy, and SS uplink transmit time tracking accuracy (Heegard & Stephen, 1999). As the receiver section of the WiMAX system suffers from the challenges with performance, it is essential to obtain optimal system performance. This can be achieved by obtaining optimization in the individual modules. One such module dealt in the following sections is the de-interleaver module of eth WiMAX receiver section (Larsen, 1973).

The WiMAX system is capable of working on the voice, data and video applications. There could be various applications of the WiMAX, could be used for VoIP, radio, IPTV, Video Teleconferencing, Displays, Interactivity, social networking, public works, sensor networks and event-based media(WiMAX Forum, 2008).

Figure 2 shows block diagram of a simple WiMAX system. The orthogonal frequency division multiplexing (OFDM) is one of the modules in the WiMAX system which supports many modulation schemes. The OFDM modulation/demodulation is usually implemented by performing FFT and inverse FFT on the data signal.

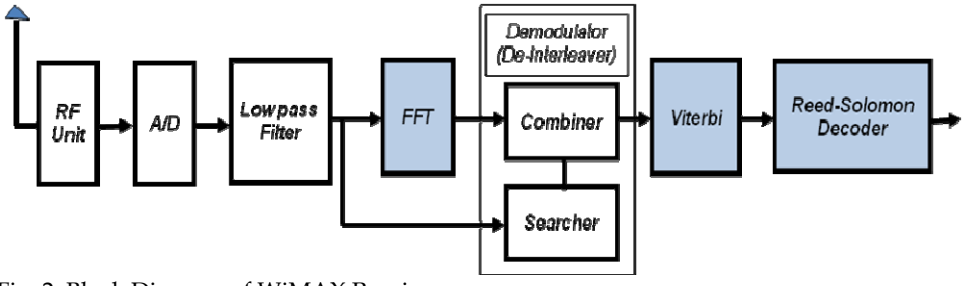



Fig. 2. Block Diagram of WiMAX Receiver

As a result, FFT and Viterbi are considered in this work as case studies which are computational intensive modules in WiMAX computing chain. A Reconfigurable FFT core and a Viterbi decoder as hardware accelerators are integrated into the WiMAX chain, to validate their performance by co-simulating modules from different levels of abstraction.

## 5. Reconfigurable FFT Implementation in SystemC

Discrete Fourier Transform (DFT) is one of the most widely used algorithms in digital signal processing for spectral analysis and for filter implementations. It operates on an N point sequence of numbers x(n). However, the DFT has limitation in speed for larger transform sizes. As a result, Fast Fourier Transform (FFT) is used, which is an efficient algorithm for computation of the DFT.

The implemented FFT block is based on a radix-2 algorithm and can be reconfigured for different FFT sizes from 16-points to 2048-points. The input data is read as a signed 16-bit fixed-point number with 10 fractional bits. Twiddle factors and output values use the same representation. Internally, computation is performed with fixed-point arithmetic. The input samples and output transformations are externally inferred as 16-bit integers. A fixed-point functional version of the FFT was developed to be compatible with the original synthesizable FFT core to verify the results, working at the highest level of abstraction.

Figure 3 shows block diagram of implemented reconfigurable FFT. The details of each block are discussed below:

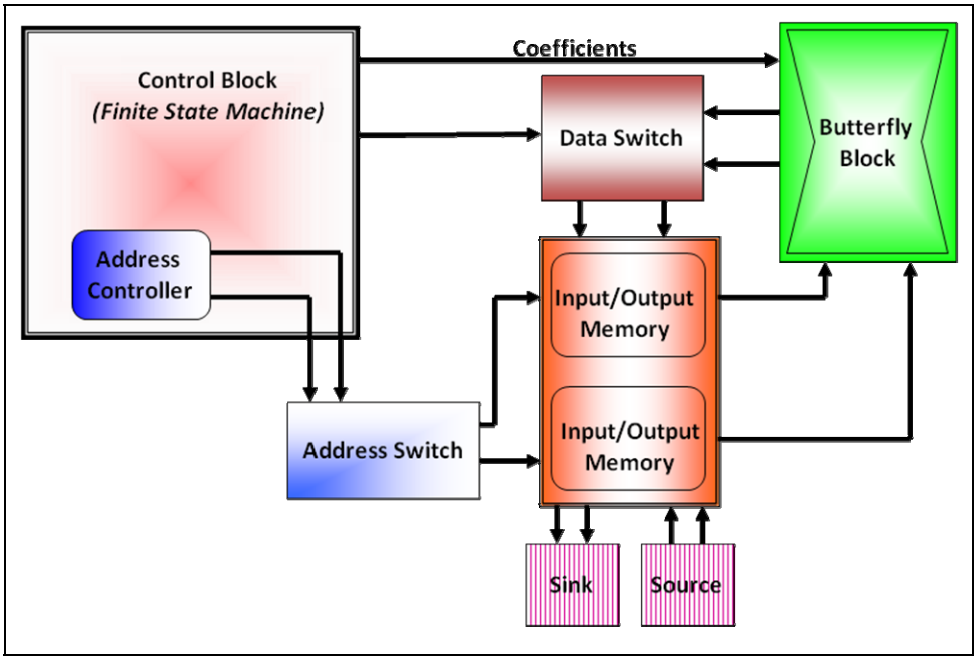



Fig. 3. Block Diagram of Implemented Reconfigurable FFT Processor

- **Control Block:** This is the main control unit of the FFT fabric. Two different counters are used for butterfly operations. First one is used to count the number of butterfly operations in each pass and then the second one is used to count the number of passes. It generates the inputs for Address Controller, the control signals for Data Memories and the configuration bits for data and address switching.

- **Butterfly Block:** This butterfly block has two parallel inputs and two outputs. Each butterfly computation consists of a number of additions, subtractions and multiplications.
- **Data Memory Interface:** As shown in Figure 3, this architecture has two data memories. Each divided into sixty-four 8x32-bit dual port memories. Therefore, each data memory has a capacity of 512x32-bit.
- **Address Controller:** This block is designed to generate addresses for data memories. Different from the address blocks in conventional fixed-point FFT fabrics, these reconfigurable ones can generate addresses for various FFT sizes. In (Ma & Wanhammar, 2000), an efficient coefficient address generation method was proposed by Cohen. According to his work, for a $N = 2n$ point FFT, at the $i$th butterfly calculation in the $j$th pass, $W_K^0 = (\exp(2\pi/N))^K$ is the coefficient needed. The $k$ can be obtained from $i$ by masking out its $(n-1-j)$ least significant bits.
- **Data and Address Switching:** The Data Switch receives the outputs of the butterfly module and routes them into the right data memory. For all the FFTs ranging from 16-2048 points, the address switch input width is 10 bits. Address switching works in parallel with data switching in order to allocate the right addresses for the data which are being processed through the data switch block.

## 6. Reconfigurable Viterbi Implementation in SystemC

All communication systems involve three main subsystems: the transmitter, the channel and the receiver. The channel medium in general attenuates the signal and as a result the noise level of the channel and the noise of the receiver will cause errors at the receiver. Here comes the task of the engineers to design communication systems with the least possible errors while satisfying the design constraints, such as allowable transmitted energy, allowable signal bandwidth and cost. To optimize the performance of the communication systems, channel coding is used. The term channel coding refers to the techniques performed on the information bearing digital signal, which requires the use of redundancy to detect and possibly correct errors. The main techniques of channel coding are: automatic repeat request (ARQ) and forward error correction (FEC).

One of the most efficient FEC techniques is the convolution encoding and Viterbi decoding. Convolution codes operate on serial data, one or few bits at a time, and they are usually described by the code rate and the constraint length. The code rate, $k/n$, is expressed as a ratio of the number of bits into the convolution encoder ($k$) to the number of channel symbols output by the convolution encoder ($n$), which gives a bandwidth expansion of $n/k$. The constraint length parameter is the length of the convolution encoder.

The implemented reconfigurable Viterbi module can decode for different constraint lengths ($K=3$, $K=5$, *and* $K=7$). It would be clocked with external clock and at the beginning of the burst sequence decoder would be synchronized with the input sequence (i.e. reset to the beginning known state). Two main modes of operation are implemented: normal and self-test modes. In normal mode, decoder is initialized to the known state, with *Synchronizer* pulse signal asserted while *Self Test* is de-asserted, one cycle prior to the beginning of data sequence. Eleven cycles after the *Synchronizer* pulse was asserted, *Data Valid* becomes active signalling that first valid bits have appeared at the *Decoded Bit* output of the decoder. In self-

test mode, the decoded sequence is compared with delayed original bits from pseudo random generator and in case of an error, the *Fail* is asserted.
The entire Viterbi decoder unit can be divided into three distinct functional units: Add Compare Select unit, Survivor Path unit, and Controller (See Figure 4). First, the operation of each unit will be explained, and then focus on the interface between the units.

- **Add, Compare, and Select Unit (ACS):** consists of 64 single-state ACS units and branch metric unit (BMU). BMU generates the Hamming distance metrics for the input encoded sequence, arriving from controller. The inputs of each single-state ACS unit are path metrics from the two states in the previous time (trellis) stage, and the two branch metrics each for one state to state transition. The single-state ACS unit adds corresponding branch metrics to path metrics to obtain the total path metrics for the resulting state and selects the smaller of the two total path metrics. The need for explicit metric rescaling is avoided by using implicit metric rescaling. In this way, the new path metric for the given state is calculated and latched in order to be fed back to the inputs of single-state ACS units in the next cycle (trellis stage). All single-state ACS units are interconnected in shuffle-exchange fashion.

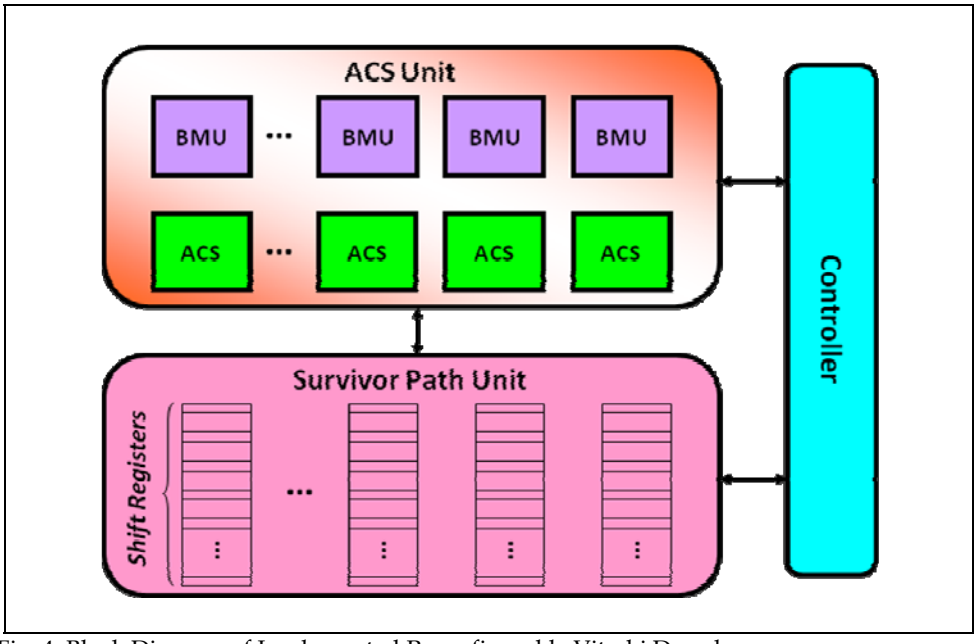



Fig. 4. Block Diagram of Implemented Reconfigurable Viterbi Decoder

- **Survivor Path Unit (SPU):** consists of a matrix of cells placed and interconnected according to the trellis diagram. Each cell consists of a latch cell and a combined latch multiplexing cell. Each row of cells corresponds to one of the states in the trellis diagram, and at each time instance holds the complete decoded history of the survivor path ending in that particular state.

As the top registers of the survivor paths change states at each new time instance, the complete history shifts, horizontally for one cell and vertically to the row corresponding to the new state of the survivor path. Each cell in a row has the same select input, which is the decision bit from the ACS unit for the corresponding state. Two data inputs are from the cells in the rows corresponding to two possible previous states in trellis diagram. The first column of cells has fixed inputs which correspond to the decoded bits for the given state and type of convolution code used. The last column contains the decoded bit. Ideally, all bits in decoded column would be the same, since all the paths converge to the common trunk, but sometimes that does not happen, and bits are not the same, thus raising the question of which of them to choose as the final result. The majority function is implemented to select the most represented value in the column, and thus minimize the decoding error.

- **Controller Unit:** consists of the self-test module, delay line, mode selection logic and majority function logic (See Figure 5). Self-test module consists of a convolution encoder and a pseudo-random generator. Mode selection logic selects between external encoded inputs and internally generated self test inputs, based on the value of the Self Test signal. Majority function logic selects the most represented value in decoded column, and outputs the decoded bit. Shift register is used both as a delay line, in self test mode, and as a counter for asserting the Data Valid signal.

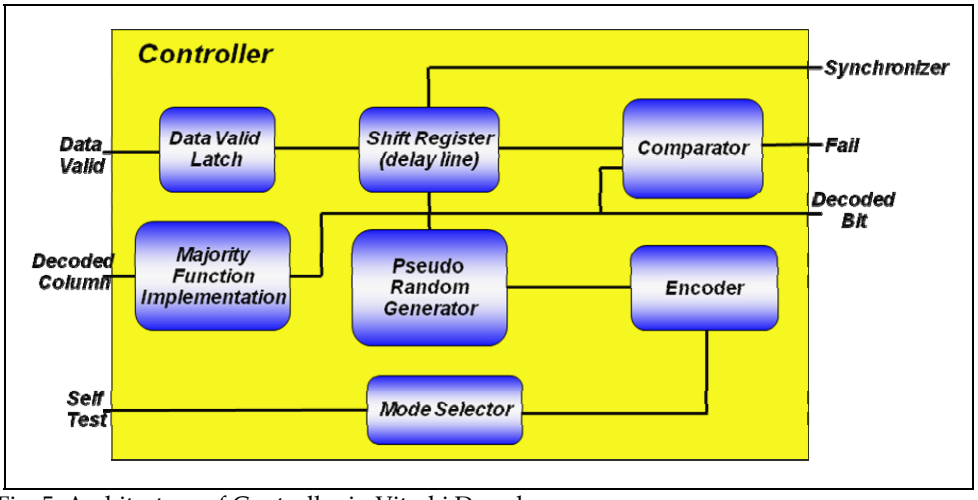



Fig. 5. Architecture of Controller in Viterbi Decoder

The implementation of the classical state machine controller is avoided, since it would just increase the number of latches used in the controller. Controller unit outputs Select Initial signal which loads the ACS units' path metric registers with the known initial metrics. Branch metric signals generated in controller are inputs to the corresponding single-state ACS units. Controller inputs are the decoded column signals from the survivor path unit. The decision bit outputs of the ACS units are wired to the inputs of the survivor path unit.

## 7. Result Analysis

A new design evaluation framework was constructed. This involved plugging the developed SystemC modules of the FFT and Viterbi into an industrial C-based implementation of the complete WiMAX telecommunication standard. Functionality evaluation was performed through the complete model. This is demonstrated as received bit rate performance of the overall receiver. This is shown in Figure 6, which demonstrates similar profiles between exact C-based model of the WiMAX for a range of FFT sizes: 128-2048 points. Since, in Mobile WiMAX (IEEE 802.16e, 2005), the FFT size of OFDM part is scalable from 128 to 2,048. As shown in Figure 6, the WiMAX bit rate is accelerated by plugging in SystemC cores, which gives more speed up for smaller FFT sizes due to decreasing FFT computation time.

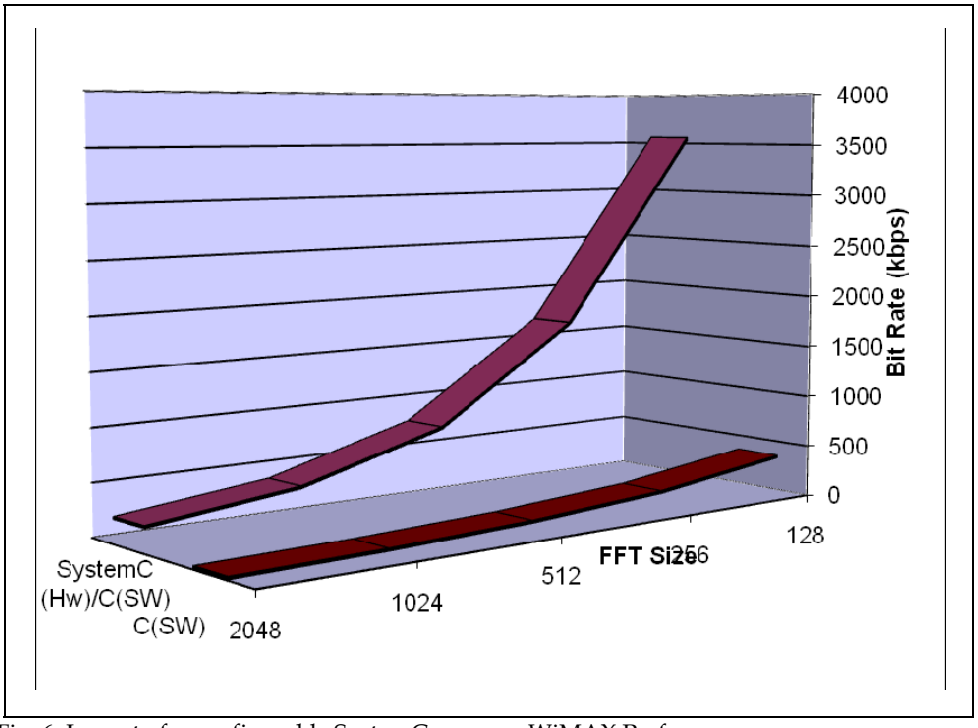



Fig. 6. Impact of reconfigurable SystemC cores on WiMAX Performance

To investigate power and area characteristics of the main blocks of WiMAX receiver, and their impact on the overall performance of system, we have used our RTL-version of these blocks. Figure 7 illustrates the change in power and area as FFT size increases. We have also investigated the impact of different code rates of Viterbi decoder compatible to (IEEE 802.16e, 2005), ranging from 12 to 78. It showed that the code rate has nearly no influence on area and power figures of the system, because by changing the code rate, the structure of Viterbi decoder remains the same, just one more code word has to be added in the same clock cycle.

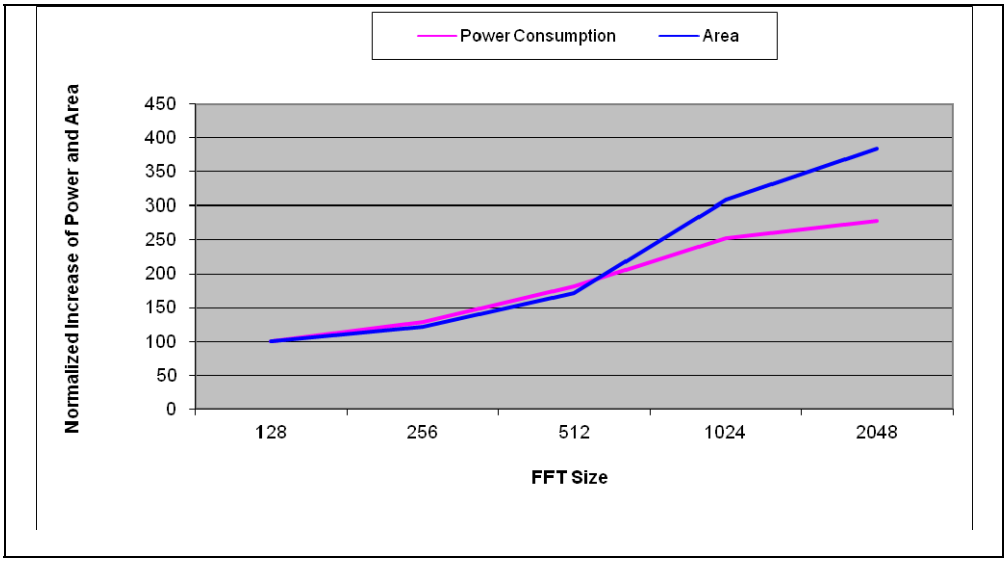



Fig. 7. Impact of reconfigurable FFT on Power and Area of WiMAX

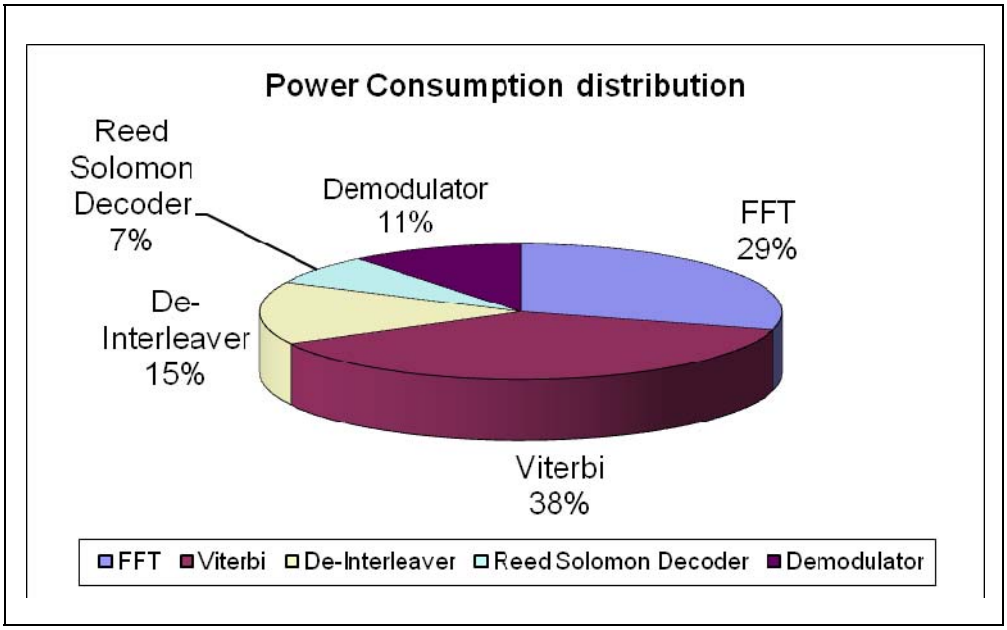



Fig. 8. Power Analysis of WiMAX Blocks

The power and area distribution of each module in WiMAX is demonstrated in Figure 8 and Gig. 9. Gate level simulation is carried out and Synopsys design compiler is used to calculate power and area results. A 0.13 micron technology library is used for ASIC synthesis. The

power consumption of Viterbi decoder is the highest among other blocks of the WiMAX receiver, and FFT gives the largest area overhead compared to others.

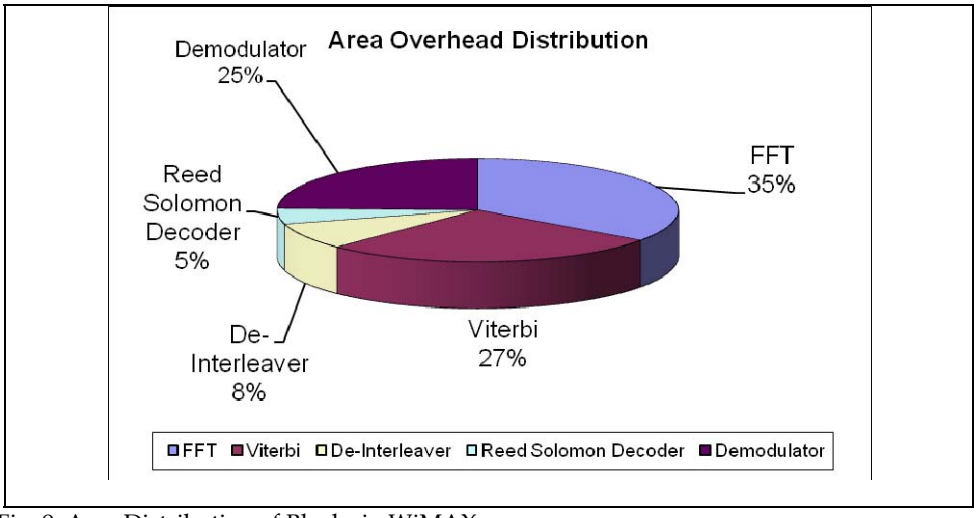



Fig. 9. Area Distribution of Blocks in WiMAX

## 8. Conclusion

In this paper, a new multi-context representation of system level modelling is proposed in order to capture the reconfigurable features within different embedded reconfigurable cores. A hybrid SystemC-HDL co-simulation framework was presented in order to demonstrate interoperability of the designed reconfigurable hardware block. As a case study, we applied our modelling strategy in order to design a complete WiMAX compliant receiver where two new reconfigurable fabrics are employed for the implementation of the performance bottlenecks within these standards, namely FFT and Viterbi processing tasks. Such a receiver needs to reconfigure the parameters and these tasks in order to balance performance against issues such as channel condition, spread delay, etc. We showed how our modelling approach can be utilised to capture complete system functionality and provide performance and power consumption figures for system design.

These reconfigurable SystemC models and their power and area attributes will be used in library of an SoC design and generation tool, for automatic parameterizable embedded core integration, functional verification at the system level, and optimization in terms of power and area.

## 9. References

Ahmadinia A., Ahmad B. & Arslan T., "System-level Modelling and Analysis of Embedded Reconfigurable Cores for Wireless Systems", In *Proceedings of International Conference on Field-Programmable Logic and Applications (FPL)*. 27-29 August 2007. Amsterdam, Netherlands, pp. 757-760.

ARM. ARM7 TDMI Datasheet. ARM. 1999. http://www.arm.com.

Brito A., Kuehnle M.,Huebner M., Becker J. & Melcher. U., Modelling and simulation of dynamic and partially reconfigurable systems using Systemc. In *Proceedings of ISVLSI*, pages 35–40, 2007.

Cai L. & Gajski. D., Transaction level modeling: an overview. In *Proceedings of the CODES+ISSS* conference, pages 19–24, 2003.

Enzler R., Plessl C. & Platzner M., Co-Simulation of a Hybrid Multi-Context Architecture. In *Proceedings of Engineering of Reconfigurable Systems and Algorithms Conference*, pages 174–180, 2003.

Freescale Semiconductor. Implementing WiMAX PHY Processing Using the Freescale MSC8126 Multi-Core DSP. Freescale Semiconductor. 2005. http://www.freescale.com.

Heegard, C. & Stephen B. W., *Turbo Coding*, Boston, Kluwer Academic Publishers, 1999.

IEEE 802.16e: Air interface for fixed and mobile broadband wireless access systems. 2005. http://standards.ieee.org/getieee802/802.16.html

Intel Corporation. Intel NetStructure WiMAX Baseband Card Product Brief. Intel Corporation. http://www.intel.com.

Larsen K. J., Short convolutional codes with maximal free distances for rate 1/2, 1/3 and 1/4, *IEEE Trans. Inf. Theory*, pp. 371-372, May 1973.

Liao C. & Hsiung. P., A SystemC-based Performance Evaluation Framework for Dynamically Reconfigurable SoC. In *Proceedings of the VLSI Design / CAD Symposium*, pages 444–447, Aug. 2005.

Ma Y. & Wanhammar L., A hardware efficient control of memory addressing for high-performance FFT processors. In *IEEE Transactions on Signal Processing*, 48(3):917–921, 2000.

Noguera J. & Badia R., System-level power performance trade-offs in task scheduling for dynamically reconfigurable architectures. In *Proceedings of the CASES conference*, pages 73–83, 2003.

Open SystemC Initiative, 2008, www.systemc.org.

Qu Y., Tiensyrja K. & Soininen J.-P., SystemC-based Design Methodology for Reconfigurable System-on-Chip. In *Proceedings of the 8th Euromicro Conference on Digital System Design*, pages 364–371, Los Alamitos, CA, USA, 2005.

Rissa T., Vasilko M., & Niittylahti. J., System-Level Modelling and Implementation Technique for Run-Time Reconfigurable Systems. In *Proceedings of FCCM Symposium*, pages 295–296, Los Alamitos, CA, USA, 2002.

WiMAX Forum, 2008, www.wimaxforum.org.

18

# WiMAX Cross-Layer System for Next Generation Heterogeneous Environments

Pedro Neves[1], Susana Sargento[2], Kostas Pentikousis[3] and Francisco Fontes[1]
[1] *Portugal Telecom Inovação*
*Portugal*
[2] *University of Aveiro / Institute of Telecommunications*
*Portugal*
[3] *VTT Technical Research Centre of Finland*
*Finland*

## 1. Introduction

The Access Network (AN), which is the important path between the operator Core Network (CN) and the Customer Premises Network (CPN), often has high requirements with respect to available bandwidth, and must be thoroughly planned. Current wired solutions, such as coaxial cable, Digital Subscriber Line (DSL) and fiber, are the main access technologies used for "last mile" connectivity. Nevertheless, due to the associated costs and difficult terrestrial conditions, important geographical areas are left without broadband connectivity. Often, lack of broadband access is the main reason for unequal access to Information and Communication Technologies (ICTs), a gap known as the Digital Divide. Previously, a significant amount of effort and resources have been invested in broadband wireless access technologies (e.g. Motorola Canopy) in order to fill this gap. Yet, since these were proprietary solutions, their adoption has been very low.
IEEE 802.16, a Broadband Wireless Access (BWA) solution for Wireless Metropolitan Area Networks (WMANs), covering large distances with very high throughputs, is an attractive BWA technology that can be used to overcome the previously mentioned limited access to the ICTs, for both fixed and mobile environments (IEEE 802.16, 2004; IEEE 802.16 a, 2005). However, the IEEE 802.16 standards do not define an end-to-end (E2E) system or architecture. They only define the PHY and MAC layers in the protocol stack. The Worldwide Interoperability for Microwave Access (WiMAX) Forum is currently defining a high-performance "All-IP" end-to-end (E2E) network architecture to support fixed and mobile users (WiMAX Forum a, 2008; WiMAX Forum b, 2008). The WiMAX Forum is also committed to perform an interoperability effort across different vendors. This opens a different set of business opportunities for telecom operators, turning WiMAX into a viable technology for the Next Generation Network (NGN) environments (ITU-T a, 2004; ITU-T b, 2004). In these environments, users wish for having ubiquitous Internet access with a wide range of possible services (e.g. "triple play") and assured QoS guarantees, even while moving. These characteristics will lead NGNs to the "Always Best Connected" (ABC)

paradigm, and as a consequence, operators have to cope with a whole new set of challenging requirements. Therefore, all the network design, including the WiMAX access network, should be able to address these aspects.
In this chapter we present a NGN architecture for WiMAX, able to support and differentiate real-time services. Special attention is given to the integration of QoS and mobility management functionalities, such as the IEEE 802.21 (IEEE 802.21, 2008) Media Independent Handover (MIH) framework, between the WiMAX technology and the higher layer entities of the designed architecture, through the specification and development of the WiMAX Cross-Layer (WXL) system. The WXL system comprises a middleware layer between the WiMAX technology and the network layer, hiding the WiMAX equipment specific functionalities from the network control plane, thus providing robustness, scalability and vendor independency. Following the recent trends in the WiMAX standardization field, the WXL interaction with the WiMAX system is established through standardized IEEE 802.16g (IEEE 802.16, 2007) primitives. In order to validate and evaluate the WXL system, a set of performance measurements were made, focusing on the resource management mechanisms. This evaluation was made in the EU IST WEIRD project (WEIRD, 2008). The results obtained from the development and experimental setup of this architecture show that the processing times for the resource reservations, both in the resource control modules and WiMAX equipments are small, even for a large number of simultaneous reservation requests, which enable the use of WiMAX in real-time environments with high mobility scenarios.
The chapter is organized as follows. Section 2 briefly describes both the IEEE 802.16 and the WiMAX Forum reference models and Section 3 presents the main characteristics of the IEEE 802.21 Media Independent Handover framework. Section 4 introduces the proposed WiMAX Cross-Layer system, including its interfaces and services. Section 5 describes a practical use case of the WXL system, developed in the WEIRD project, whereas Section 6 discusses the experimental results obtained, as well as our test methodology. Finally, Section 7 concludes this chapter.

## 2. WiMAX Overview

This section starts by briefly introducing the IEEE 802.16 reference model, including the data, control and management planes. Thereafter we present the network architecture defined by the WiMAX Forum standardization body to integrate the WiMAX access technology in a real network deployment.

### 2.1 IEEE 802.16 Reference Model

WiMAX standardization efforts are currently being made by the IEEE 802.16 working group and by the WiMAX Forum. The IEEE 802.16 defines the air interface specifications, namely the Physical (PHY) layer and the Medium Access Control (MAC) layer, for both fixed (IEEE 802.16, 2004) and mobile (IEEE 802.16 a, 2005) terminals. The MAC layer provides the interface with the higher layers through the Service Specific Convergence Sublayer (CS). Below the Service Specific Convergence Sublayer is the Common Part Sublayer (CPS), which is responsible for the core MAC functions. Finally, under the Common Part Sublayer, there is the Security Sublayer. Figure 1 illustrates the IEEE 802.16 reference model, comprising the data, management and control planes.

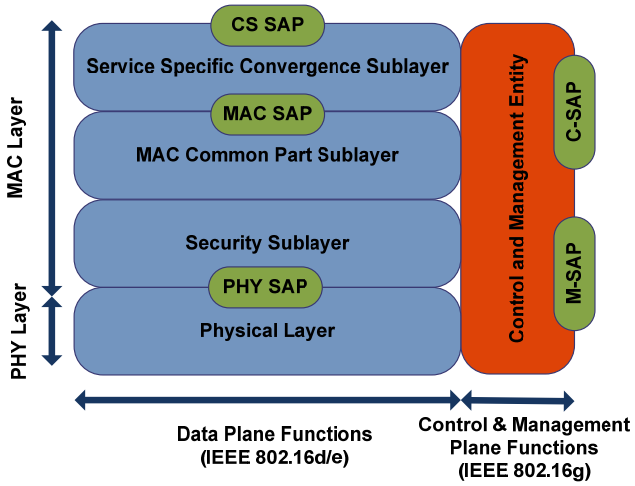



Fig. 1. IEEE 802.16 Reference Model

As presented in Figure 1, the CS is the first sublayer from the MAC layer. The sending CS accepts higher layer MAC Service Data Units (SDU) coming through the CS Service Access Point (SAP) and classifying them to the appropriate connection. The classifier is based on a set of packet matching criteria applied to each packet. It consists of some protocol-specific fields, such as IP and Ethernet addresses, a classifier priority and a reference to a specific connection. Each connection has a specific service flow associated providing the necessary QoS requirements for that packet. If no classifier is found for a specific packet, the packet can be discarded, sent on a default connection or a new dedicated connection can be established. Finally, the CS Protocol Data Unit (PDU) is delivered to the MAC CPS through the MAC SAP and delivered to the peer MAC CPS. Downlink classifiers are applied by the WiMAX Base Station (BS) and uplink classifiers are applied by the WiMAX Subscriber Station (SS) or Mobile Station (MS).

The CPS is the second sublayer from the MAC layer. It receives packets arriving from the CS and is responsible for the most important MAC functions, such as system access, construction and transmission of MAC PDUs, scheduling services management, bandwidth allocation and contention resolution. The 802.16 system is connection oriented since all services are mapped to a connection. Associated with each connection is a Service Flow (SF) and a scheduling service. Five scheduling services are supported by IEEE 802.16: Unsolicited Grant Service (UGS), real time Polling Service (rtPS), extended real time Polling Service (ertPS), non real time Polling Service (nrtPS) and Best Effort (BE).

Finally, the Security Sublayer is the last sublayer from the MAC layer, providing authentication and security mechanisms.

Since the IEEE 802.16 entities are part of a larger network environment, it is necessary to define control and management communication mechanisms between the WiMAX system and the higher layers. The IEEE 802.16g-2007 (802.16, 2007) standard has been defined to efficiently integrate the IEEE 802.16 entities with the higher layer control and management functionalities. In particular, 802.16g defines two Service Access Points (SAPs) – the Management Service Access Point (M-SAP) and the Control Service Access Point (C-SAP) –

to establish communication between an IEEE 802.16-based system and the control and management entities on the higher layers, respectively. The C-SAP and the M-SAP are depicted in Figure 1. The Control SAP (C-SAP) is oriented for more time sensitive control plane primitives, including:

- Handovers security management;
- Subscriber mode (idle, sleep, normal) management;
- Handover context management;
- Media Independent Handover management;
- Radio resources management;
- Network entry and exit management;
- Service flow management.

The Management SAP (M-SAP) is used for less time-sensitive management plane primitives, such as:

- Connections and subscribers accounting management;
- Subscriber / Mobile Station interface management;
- System configuration;
- Statistics monitoring;
- Notifications and triggers management.

Detailed information about each one of the aforementioned primitives can be found in the IEEE 802.16g standard (IEEE 802.16, 2007).

### 2.2 WiMAX Network Reference Model

Based on the PHY and MAC layer specifications provided by the IEEE 802.16 working group standards, the WiMAX Forum standardization body is designing an all-IP end-to-end network architecture for WiMAX environments. As a result, the WiMAX Forum has defined the WiMAX Network Reference Model (NRM), illustrated in Figure 2.

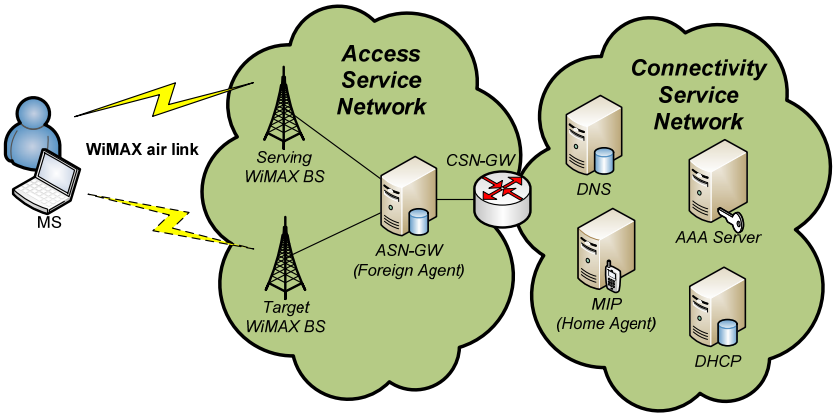



Fig. 2. WiMAX Network Reference Model

The WiMAX NRM is a logical representation of the WiMAX architecture, based on a set of functional entities and eight standardized interfaces, also known as Reference Points (RPs).

Using this model, multiple implementation options for each functional entity are allowed, while maintaining interoperability using the RPs. Three functional entities are defined: Connectivity Service Network (CSN), Access Service Network (ASN) and the Customer Premise Network (CPN). The CPN is the terminal equipment, responsible for establishing radio connectivity with the BS. For fixed WiMAX (IEEE 802.16, 2004), the CPE might be composed by single-user SSs or multi-users SSs (MSS) when a LAN/WLAN is connected to the SS. For mobile WiMAX (IEEE 802.16 a, 2005), a single-user Mobile Station (MS) is the envisioned scenario (see Figure 2). The ASN is composed by several BSs connected to several gateways (ASN-GWs), which establish connectivity with the CSN. The ASN includes a set of functionalities in order to provide radio connectivity to the WiMAX subscribers, including service flow and micro/macro mobility management. Additionally, it also performs relay functions to the CSN in order to establish IP connectivity and Authentication, Authorization and Accounting (AAA) mechanisms. Finally, the CSN holds the DHCP, DNS, MIP and AAA servers, as well the SIP Proxy. Moreover, the CSN is responsible for establishing connectivity with the IP backbone.

## 3. IEEE 802.21 – Media Independent Handover Services

The main goal of IEEE 802.21 (802.21, 2008) is to optimize mobility mechanisms in heterogeneous access environments. Towards this aim, it defines a Media Independent Handover (MIH) framework, which provides standardized interfaces between the access technologies and the mobility protocols from the higher layers in the protocol stack. IEEE 802.21 introduces a new entity called MIH Function (MIHF), which hides the specificities of different link layer technologies from the higher layers mobility entities. Several higher layer entities, known as MIH Users (MIHUs) can take advantage of the MIH framework, including mobility management protocols, such as Mobile IPv4 (MIPv4) (Perkins, 2002), Mobile IPv6 (MIPv6) (Johnson et al, 2004), Fast Mobile IPv6 (FMIPv6) (Koodli, 2005), Proxy Mobile IPv6 (PMIPv6) (Gundavelli et al, 2008) and Session Initiation Protocol (SIP) (Rosenborg & Camarillo, 2002), as well as the other mobility decision algorithms.

In order to detect, prepare and execute the handovers, the MIH platform provides three services: Media Independent Event Service (MIES), Media Independent Command Service (MICS) and Media Independent Information Service (MIIS). MIES provides event reporting such as dynamic changes in link conditions, link status and link quality. Multiple higher layer entities may be interested in these events at the same time, so these events may need to be sent to multiple destinations. MICS enables MIHUs to control the physical, data link and logical link layer. The higher layers may utilize MICS to determine the status of links and/or control the access to different networks. Furthermore, MICS may also enable MIHUs to facilitate optional handover policies. Events and/or Commands can be sent to MIHUs within the same protocol stack (local) or to MIHUs located in a peer entity (remote). Finally, MIIS provides a framework by which a MIHF located at the MS or at the network side may discover and obtain network information within a geographical area to facilitate handovers. The objective is to acquire a global view of all heterogeneous networks in the area in order optimize seamless handovers when roaming across these networks.

Figure 3 presents the 802.21 MIH framework, including the three aforementioned services – MIES, MICS and MIIS.

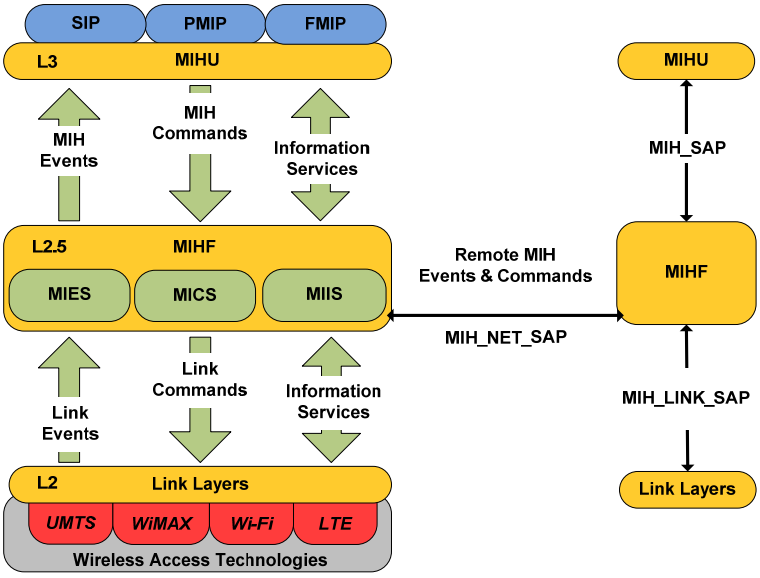



Fig. 3. IEEE 802.21 MIH Framework (Neves et al, 2009)

## 4. WiMAX Cross-Layer System Design

Nowadays, standardized mechanisms for both network layer and link layer are already in place. For example, on the network layer side, QoS signaling protocols such as NSIS (Hancock et al, 2005; Schulzrinne & Hancock, 2007) or Diffserv (Blake et al, 1998), as well mobility management protocols such as Mobile IP and Host Identity Protocol (HIP) (Moskowitz & Nikander, 2006) are well established. However, one of the major gaps is the inter-layer connectivity, also known as cross-layer, between the network and the link layer functions. One particular example is the WiMAX access technology, which requires an efficient communication with the network layer to exploit all its functionalities. To establish an efficient communication between the network and the WiMAX MAC layers, we propose the WiMAX Cross-Layer (WXL) system. The WXL is responsible for providing all the required cross-layer services between the WiMAX MAC layer and the network layers, such as QoS, mobility, AAA, security, multicast and broadcast. The WXL system comprises a middleware layer between the WiMAX technology and the network layer, hiding the WiMAX technology specific functionalities from the network control plane. To interact with the network layer, the WXL provides a set of dedicated interfaces, known as *WXL Upper Interfaces*, comprising, among others, support for QoS, mobility, accounting and security.
Likewise, the *WXL Lower Interfaces* also encompass several interfaces to interact with the link layer technologies. The preferred and most appropriated interface between the WXL system and the WiMAX BS is established using the IEEE 802.16g (802.16, 2007) standardized service primitives. As described in Section 2, 802.16g provides a set of primitives to address each one of the specific WiMAX services, and therefore it is the ideal solution to satisfy all the requirements for this type of interface. Nevertheless, we cannot assume beforehand that all WiMAX systems support the 802.16g standardized interface. For example, current

commercial WiMAX equipments available on the market, though already certified by the WiMAX Forum, only supports Simple Network Management Protocol (SNMP), Hypertext Transfer Protocol (HTTP) or a simple Command Line Interface (CLI) to manage the equipments. Hence, the WiMAX Cross-Layer system must be flexible to support any type of interface with the link layer as alternatives to the preferred IEEE 802.16g interface. Figure 4 illustrates the WiMAX Cross-Layer (WXL) system architecture.

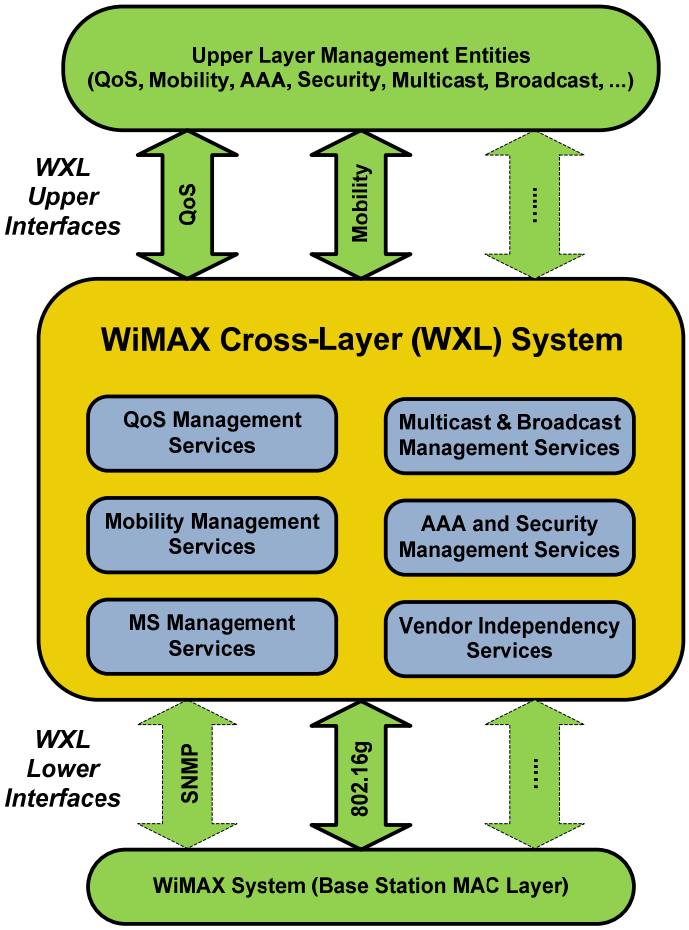



Fig. 4. WXL System Architecture

The functionalities provided by the WXL system are grouped on different management services. As illustrated in Figure 4 and described in Table 1, six services are exposed by the WXL to the upper layers, namely QoS, Mobility, Mobile Station, AAA & Security, Multicast & Broadcast and Vendor Independency Management Services.

| Management Service | Function |
|---|---|
| **QoS** | Service flows and admission control management for the WiMAX link. |
| **Mobility** | Control of the handover context between the serving and the target BSs, as well as translation of the 802.21 Media Independent Handover primitives to the WiMAX system. |
| **Mobile Station** | Management of the MS states (normal, sleep and idle modes), MS status (on, off) and the MS network entry, re-entry and exit procedures. |
| **AAA & Security** | Subscribers and connections accounting, initiated by the end of the subscribers registration procedure and by the service flow allocation, respectively. |
| **Multicast & Broadcast** | Management of multicast and broadcast related functionalities. |
| **Vendor Independency** | Management of WiMAX equipments from different vendors. |

Table 1. WXL Management Services

In this chapter we will focus on the WXL QoS and Vendor Independency Management Services from the WXL system.

### 4.1 WXL QoS Management Service

The **QoS Management Service (QMS)** controls the QoS operation of the WiMAX network, handling the creation, modification and deletion of service flows (Service Flow Manager), as well as providing full control of all the service classes (Service Class Manager) and the related convergence sublayer classifiers (Convergence Sublayer Manager). As defined in the IEEE 802.16d/e standard (802.16, 2004; 802.16 a, 2005), three different sets of QoS parameters might be associated with a particular SF such as:

- *ProvisionedQoSParamSet*: indicates a QoS parameter set provisioned via means outside of the scope of 802.16d/e standard, such as the operator network management system.
- *AdmittedQoSParamSet*: defines a set of QoS parameters for which the BS and the SS/MS are reserving resources. This QoS parameter set is the preparation to subsequently activate the flow.
- *ActiveQoSParamSet*: specifies the set of QoS parameters that are actually being provided to the SF.

Depending on the QoS parameters set, a specific service flow will have a status associated. The following service flow statuses are defined and supported in the WXL system:

- *Provisioned Service Flow*: created via provisioning by, for example, the network management system. Its *AdmittedQoSParamSet* and *ActiveQoSParamSet* are both null.

- *Admitted Service Flow*: indicates that the service flow has resources reserved by the BS for its *AdmittedQoSParamSet*, but these parameters are not yet active.
- *Active Service Flow*: resources are committed by the BS for its *ActiveQoSParamSet*, and as a result the BS is able to forward data packets through this service flow for the subscribers.

With respect to the supported convergence sublayers, the WXL system supports Ethernet, VLAN tags, IPv4 and IPv6 CSs for packet classification. To manage the service flows and their status, both static and dynamic models are defined. The static model is used for medium- and long-term QoS reservations, triggered by the Network Management System (NMS). In this case, the service flow must pass through all the existing status (Provisioned and Admitted) until it reaches the Active one. In the dynamic model, the service flow can be directly activated without traversing the remaining status. This model is suitable for real-time services that require instantaneous allocations of service flows on the WiMAX path. Figure 5 illustrates the WXL QoS Management Service (QMS) framework.

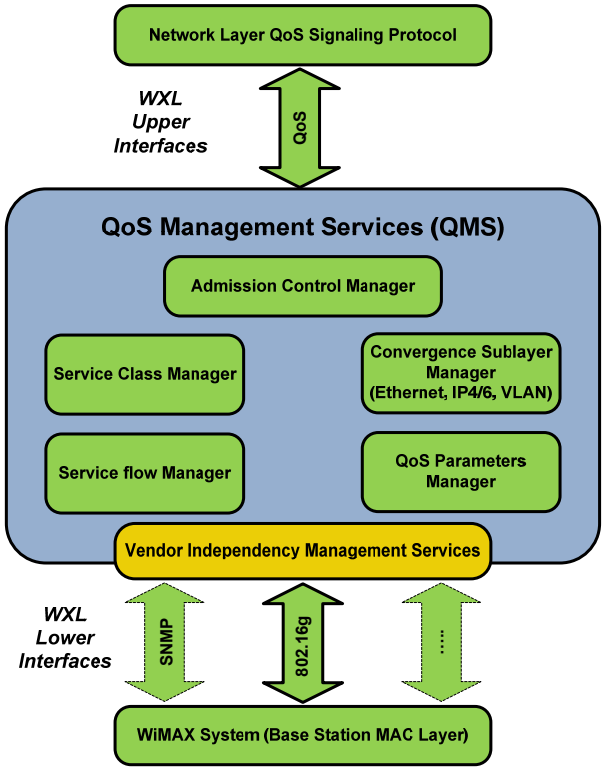



Fig. 5. WXL QoS Management Services Framework

Moreover, the QMS is also responsible for deciding on new QoS requests through the Admission Control Manager. The Admission Control Manager verifies if the available resources in the WiMAX link are enough to satisfy the requirements of the new QoS request. If enough resources are available, the service reservation is accepted. On the other hand, if

resources are not available to satisfy the requested QoS parameters, the Admission Control Manager will have to analyze the Service Class and the Priority parameters from the new service flow and compare it with the already existent service flows in the WiMAX link. Based on this evaluation, a lower priority service flow can be downgraded and hence the new QoS request is accepted, or the new QoS request is rejected.
Figure 5 also illustrates the QoS Parameters Manager, which is responsible for performing the translation of generic network QoS parameters (e.g. DiffServ) to WiMAX specific QoS parameters, described in IEEE 802.16d/e standard.

### 4.2 WXL Vendor Independency Management Service

To cope with the complexity that the increasing heterogeneity brings into future communication networks, we propose the **Vendor Independency Management Service (VIMS)**, a common solution that applies to different WiMAX vendor equipments. This service provides architecture independence, as much as possible, from the WiMAX equipments manufacturers. To achieve this level of independency permits that different WiMAX equipments can be seamlessly integrated and supported without requiring modifications over the remaining architecture modules and interfaces. Figure 6 presents the Vendor Independency Management Service (VIMS).

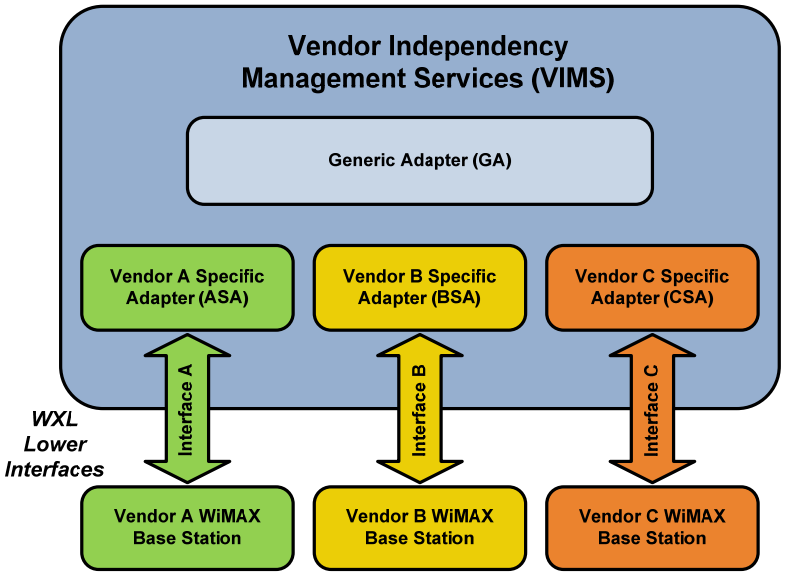



Fig. 6. Vendor Independency Management Services (VIMS) Framework

Internally, this service is composed by a Generic Adapter (GA) module and one or more Vendor Specific Adapter (VSA). The Generic Adapter (GA) allows the convergence between different Vendor Specific Adapters (VSAs) for the adaptive applications and processes. The Generic Adapter provides an abstraction of the hardware management functions. All details are hidden from the upper modules, allowing the use of common primitives. To allow this, the Generic Adapter converts the common primitives to more specific primitives, serving

the needs of the Specific Adapters. GA is able to communicate with local and remote VSAs modules using inter-process communication.
The Vendor Specific Adapter deals with the WiMAX equipment specificities. For each vendor specific equipment, a library is implemented providing a differentiated request processing, according to the supported protocol by the WiMAX system (802.16g, SNMP, HTTP, CLI, XML, …). Each VSA is responsible to fetch information from the WiMAX network elements (allowing important control functions, such as admission control and effective resources control) and to enforce the control decisions triggered by upper layer entities in the WiMAX segment. Hence, its functionalities assume considerable relevance in the global architecture.
Summarizing, in order to support an additional WiMAX vendor equipment in the WXL system, minor modifications are required. More precisely, it is only necessary to design and implement a new VSA, as well as the associated interface without modifying any other module or interface.

## 5. The WEIRD System: A Practical Use Case of the WXL System

This section presents a practical use case to show the implementation of the WXL for QoS Management (QSM) and Vendor Independency Management Services (VIMS). We consider and describe the proposed architecture in the WEIRD Project (WEIRD, 2008).
Following NGN trends (Knightson, 2005), the WEIRD system (Guainella, 2007) follows a multi-plane structure. Vertically it is composed by two stratums – the Applications and Service Stratum responsible for the management and control of the applications, and the Transport Stratum responsible for the resources management, as well as the data transport. Horizontally, the WEIRD architecture is split into three well-known parallel planes – Management, Control and Data Plane.
Figure 7 depicts the WEIRD Control Plane modules and interfaces. Although the WiMAX NRM is used as the basis of the architecture, new modules have been defined on the Control Plane to efficiently support real time services with QoS differentiation. Both SIP (Rosenberg, 2002) and legacy applications are supported. For SIP-based applications, the SIP User Agent (SIP UA) in the MS communicates directly with the SIP Proxy at the CSN. For legacy applications, a specific module is specified for the MS – the WEIRD Agent – which adapts and configures the QoS parameters as required by legacy applications.
The Connectivity Service Controller (CSC) modules, located at all NRM entities, include the most important functions of the system. Since WEIRD is focused on the ASN segment, the CSC at the ASN (CSC_ASN) is the main coordination point for QoS functions, such as resource allocation and admission control in the ASN and the WiMAX segments. For SIP applications, the SIP Proxy extracts the QoS parameters from the SIP/SDP messages, performs user authentication and authorization with the AAA server, and forwards the collected QoS information to the CSC_ASN using a Diameter (Gq/Gq') (Calhoun, 2003) interface.

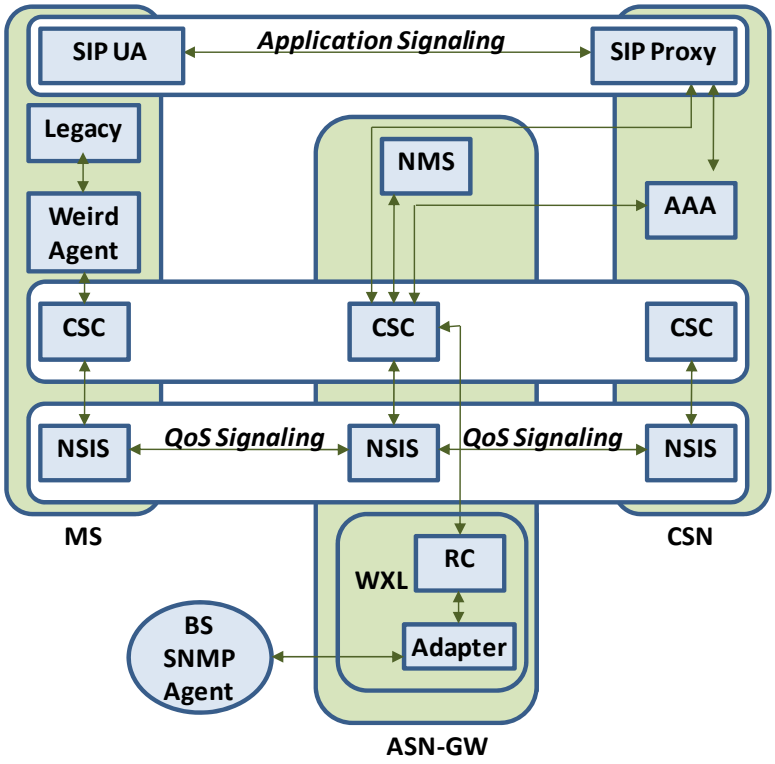



Fig. 7. WEIRD architecture – Control Plane

CSC_MS communicates with the WEIRD Agent to obtain the QoS parameters required by the legacy applications and provides this information to the main QoS coordination point (CSC_ASN). When the CSC_CSN receives the QoS reservations requests from the CSC_ASN, it establishes the QoS paths on the core network. Moreover, the CSC_ASN has an interface with the Network Management System (NMS) for medium- and long-term functions, such as QoS provisioning.

The communication between the several CSCs (MS, ASN and CSN) is performed through the usage of the Next Steps in Signaling (NSIS) QoS signaling protocol (Hancock, 2005). NSIS decomposes the overall signaling protocol suite into a generic (lower) layer and specific upper layers for each specific signaling application. At the lower layer, General Internet Signaling Transport (GIST) (Schulzrinne, 2007) offers transport services to higher layer signaling applications. Above this layer, the NSIS Signaling Layer Protocol (NSLP) (Manner, 2007) supports any protocol within the signaling application layer.

All functions related with the WiMAX system are managed and controlled by the WXL system described in Section 4. It contains the Resource Controller (RC) module for QoS Management Services (QSM), including service classes, service flows and convergence sublayer classifiers control (e.g. Ethernet, IPv4 and IPv6 classification), as well as for admission control tasks on the WiMAX link. Furthermore, the RC acts as an abstraction layer for QoS between the upper parts of the architecture and the lower level modules. It

hides all QoS WiMAX technology related functionalities from the upper layers, keeping them independent and oblivious of WiMAX-specific QoS characteristics. To enforce the QoS decisions on the WiMAX BS, the RC triggers the Vendor Independency Management Service (VIMS) through the Adapter module that will communicate the decisions to the WiMAX BS through an SNMP interface. The initial proposed model of the WiMAX Adapter is presented in (Nissilä, 2007). As previously described in Section 4, it is split into a Generic Adapter (GA) component and one or more Vendor-Specific Adapter (VSA) modules.
From all modules defined on the WEIRD architecture, VSAs are the only dependent on the specific WiMAX equipment used. As an example, Figure 8 illustrates the signaling between the different entities involved in the QoS reservation process. The signaling for QoS modifications and deletions is similar.

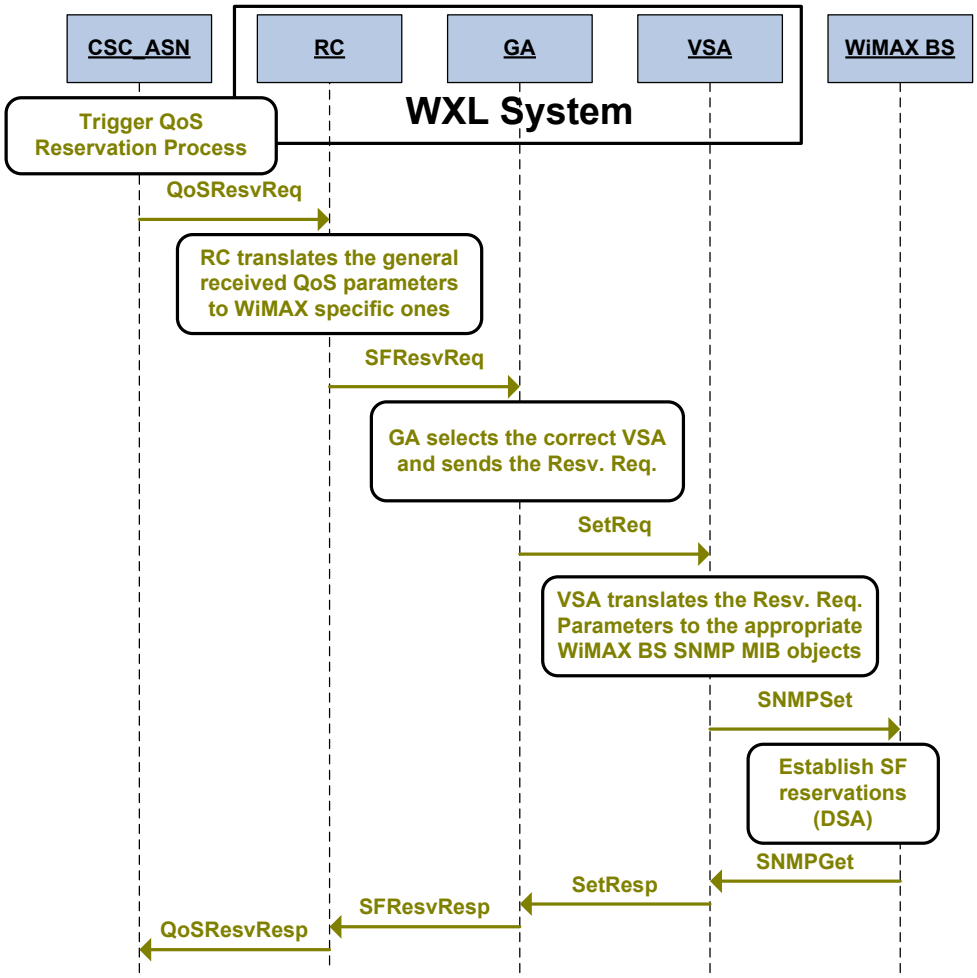



Fig. 8. QoS reservation signalling diagram

## 6. Performance Evaluation

We tested the proposed resource control framework, encompassing the WXL system for both QoS Management (QSM) and Vendor Independency Services (VIMS), in the Portuguese (for the Redline Communications WiMAX equipment) and Finnish (for the Airspan WiMAX equipment) testbeds, based on the WiMAX NRM, as illustrated in Figure 9. Particularly, the testbed is composed by the CSN, the ASN and the CPN. Under the ASN we have the BS directly connected to the ASN-GW. Two SSs are connected to the BS creating a Point-to-Multipoint (PMP) topology for the Radio Access Network (RAN).
Our tests focus on evaluating our system with respect to efficiently managing the WiMAX system and integrating it with a NGN IP architecture. Specifically, we evaluate the performance of the WXL system regarding QoS integration in a vendor independent WiMAX RAN. Each module in the chain has been evaluated, as well as the overall path towards the WiMAX system.

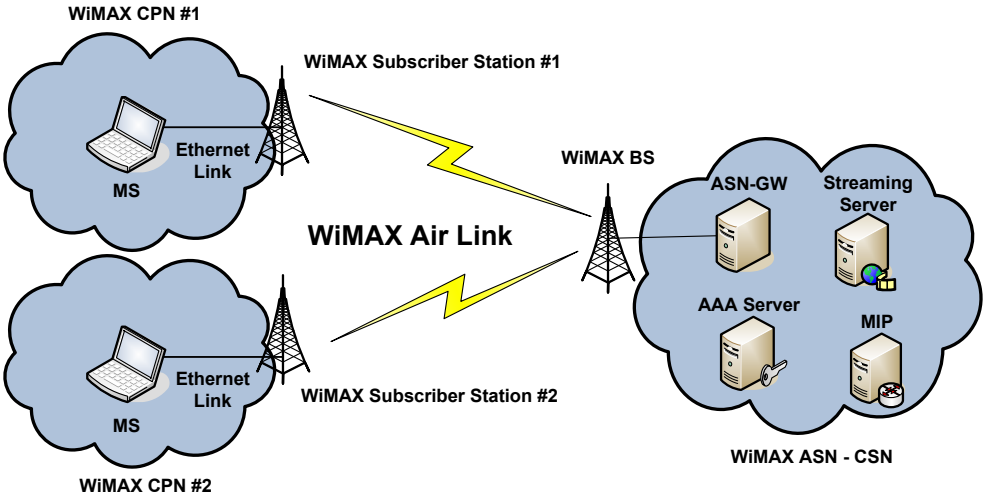


Fig. 9. WEIRD testbed

The Redline Communications certified WiMAX equipment has been used since it fully supports the evaluated QoS actions: reservations, modifications and/or deletions. The results obtained for every specific QoS action that has been evaluated for a specific number of service flows (2, 8, 32, 64, 128 and 256) in both uplink and downlink directions, comprise the mean of three runs.

### 6.1 Overall QoS Tests

This section presents the results of QoS sessions establishments, modifications and deletions. Figure 10 illustrates a stacked column graphic, where each column represents a specific module of the ASN-GW (CSC, RC, GA and the Redline Communications Specific Adatper – RSA).

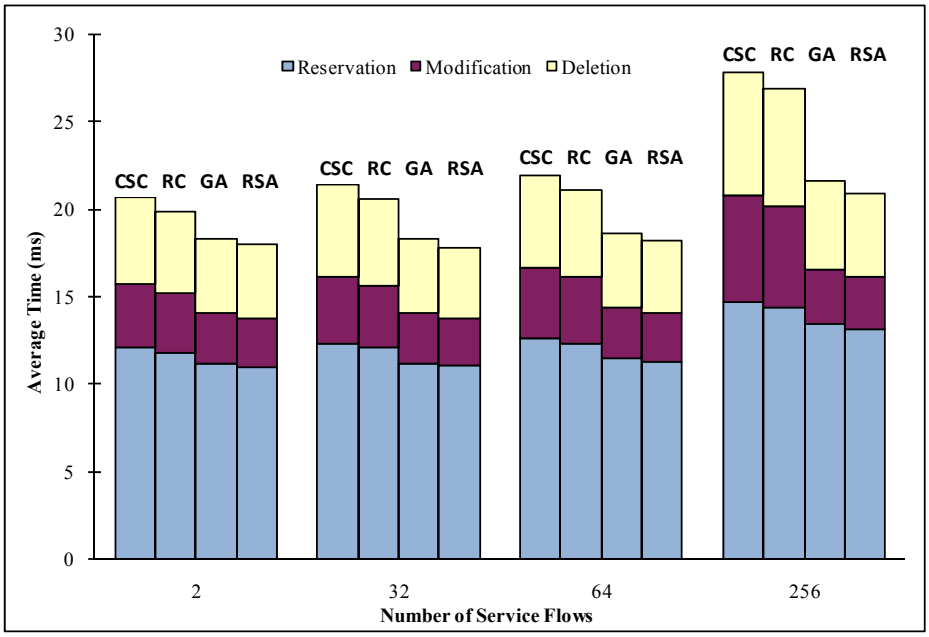



Fig. 10. QoS reservation, modification and deletion measurements

Each stack column is further split in three parts, each one corresponding to a specific action, namely, QoS session establishment (blue), modification (purple) and deletion (white). For each action, the vertical axis represents the cumulative average time (in milliseconds - ms) to enforce a specific action on the WiMAX system. As an example, the blue part of the CSC column represents the cumulative average time, composed by the CSC, RC, GA, RSA and WiMAX BS modules, to establish the SF reservations in the WiMAX system. The blue part of the RC column describes the QoS reservation cumulative average time, composed by the RC, GA, RSA and WiMAX BS, whereas the blue section of the GA column provides the cumulative average time, given by the GA, RSA and WiMAX BS modules. Finally, the RSA column represents the internal module processing time, as well as the SNMP management messages exchange with the WiMAX BS. On the horizontal axis is represented the number of SFs (2, 8, 32, 64, 128, 256) that have been used for each test. Therefore, each group of four columns represents a specific performance test.

Comparing the obtained results with respect to the type of actions that are being requested to the WiMAX system, we show that the required time to establish a QoS reservation (blue) is approximately 53 % of the total time consumed. The rationale for this behavior is due to the fact that the amount of operations required by the WiMAX system for a QoS reservation is higher when compared to the modification and deletion processes. Figure 10 also shows that the most time consuming module is the RSA, mainly because it includes the negotiation of the QoS parameters between the BS and the SS through the usage of the DSA/C/D (Dynamic Service Addition / Change / Deletion) messages, as defined in the IEEE 802.16 standard, which is the major time consuming process in the chain. The average time taken by the RSA to perform the three actions has a minimum of 17,9 ms for 2 SFs up to a

maximum of 20,9 ms for 256 SFs, with small impact on the increase of SFs. The time spent by the remaining modules is due to the message flow and internal processing. The average time measured on the CSC to reserve 256 SFs is 15 ms; moreover, 6 ms are required to modify the previously established SFs and then, finally, 6 ms are required to delete the 256 SF reservations at the end of the flow lifetime. In this case, the entire management of the 256 flows required approximately 27 ms.
In short, an increasing number of SFs slightly raises the time spent to establish, modify and delete the QoS sessions. This behavior occurs because the number of entries in the hash tables increases, as well as the interaction with the SNMP Management Information Base (MIB) (Case, 2002; McCloghrie, 1999) tables in the WiMAX BS. The overall elapsed time is very good for real-time applications and fast mobility environments, ensuring a desirable quick resource reallocation, from 21 ms for 2 SFs up to 27 ms for 256 SFs.

### 6.2 Redline Specific Adapter Tests

This section is devoted to the discussion and analysis of the RSA performance results. A considerable set of times was collected in order to analyze the time distribution along the different tasks accomplished by RSA. The RSA is the Vendor-Specific Adapter module from the WXL VIMS that interacts with the Redline WiMAX equipment. Thus, the total time spent from the reception of a request (SF Reservation / SF Modification / SF Deletion) to the sending of the correspondent answer to the GA was obtained; the partial times were also gathered, in order to distinguish between the internal processing time of the RSA module and the time it takes to set each SNMP MIB table. Figure 11 illustrates the RSA performance times. The RSA Resv/Mod/Del Total Path blocks show the total time used up by RSA to process each request, that is, the entire path since it receives a request from the GA, until it sends back the response to the latter. The RSA Resv/Mod/Del Total SNMPSET blocks represent the sum of all SNMPSET times needed, respectively, for a SF reservation, modification, or deletion. Finally, the RSA Resv/Mod/Del Internal Processing blocks represent the internal processing time of RA routines.
As can be observed in Figure 11, the RSA internal processing time is so small that the total path time is almost not affected, and therefore unseen on the graphic bars. In the worst case, the internal processing time is less than 55 us. This case occurs when 256 SFs are deleted, which causes the module to deal with considerable processing work when looking for the associated indexes to set the SNMP MIB tables (IEEE 802.16 b, 2005).
Another conclusion taken from the graphic is that QoS session's establishment takes more time than the deletion ones, and these, in turn, are more time consuming than the modification requests. Furthermore, we can also perceive that there is a slight increase of time when the number of SFs increases. Nevertheless, these values are kept stable, showing that the RSA and the WiMAX equipment are prepared to deal efficiently with a large number of sequential requests. Finally, the differences between the RSA Resv/Mod/Del Total SNMPSET processing times are due to the different amount of Object IDs (OIDs) that have to be set on the WiMAX MIB, which is larger for the reservation requests when compared to modifications or deletion ones.

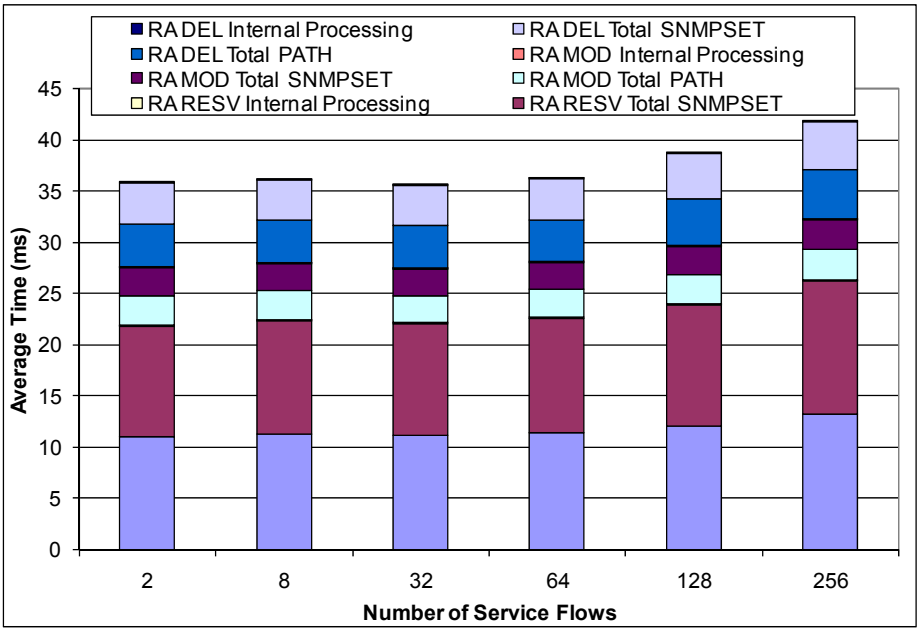



Fig. 11. RSA performance times

Table 2 details all the SNMP MIB Tables, and the correspondent OIDs number that are assigned when performing a SF Reservation, Modification and Deletion.

| SF Request: MIB Tables | Function | OIDs |
|---|---|---|
| **DEL : *ServiceClassTable*** | Contains the SF QoS parameters | 1 |
| **MOD : *ServiceClassTable*** | | 5 |
| **RESV : *ServiceClassTable*** | | 12 |
| **DEL : *ProvisionedSfTable*** | Contains the SF profiles provisioned by NMS | 1 |
| **RESV : *ProvisionedSfTable*** | | 5 |
| **RESV : *ProvisionedForSFTable*** | Maps the MAC addresses of SSs to the provisioned SFs | 1 |
| **RESV :*ClassifierRuleTable*** | Contains packet classifier rules associated with SFs | 14 |

Table 2. WiMAX QoS MIB Tables

Figure 12 shows the time spent by every single SNMP MIB table set. Once again, the inherent processing time is coupled with the increased number of SFs requests, increasing as we increase the number of sequential requests. Furthermore, the time taken to process each SNMPSET is also strictly related with the number of OIDs that must be set in each operation – the larger the number of OIDs to assign, the larger the time spent on performing this task by the equipment. The differences that occur when setting distinct SNMP MIB tables with the same OIDs number are due to the inherent processing time of WiMAX equipment for

those tasks. For instance, erasing an existing SF by the WiMAX equipment is more time consuming than just performing an amendment to the QoS parameters of an existing one. Moreover, as an example, when deleting a *ProvisionedSfTable* row entry, we are in fact deleting, at the same time, all the associated entries in *ProvisionedForSfTable* and *ClassifierRuleTable*. For that reason, the deletion process when erasing elements of *ServiceClassTable* is smaller than when deleting elements from *ProvisionedSfTable*.

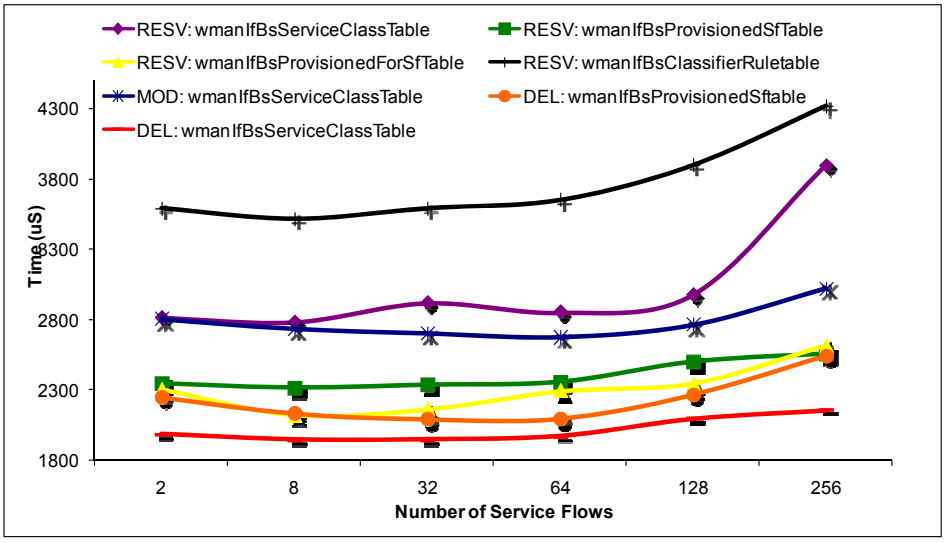



Fig. 12. Single SNMPSET performance times

## 7. Conclusions

This chapter presented a NGN architecture integrating WiMAX, able to support and differentiate real-time services in mobile environments. To provide several services in the network (such as mobility, QoS, network management, AAA and multicast/broadcast services) independently from the equipment specific functionalities, between the WiMAX technology and the higher layer entities of the designed architecture, a WiMAX Cross-Layer (WXL) system was specified and implemented. This system provides independency between the network control plane and the WiMAX system through standardized IEEE 802.16g functionalities. The presented system architecture, developed as a practical use case in the WEIRD project, comprised, integrated in the WXL, a Resource Controller and WiMAX Adapter modules for the support of QoS reservations, modifications and deletions.

In order to validate the WXL system, and the consequent support of QoS resource management of both Resource Controller and Adapters, QoS measurements have been performed in a WiMAX certified testbed. The results obtained from the real deployment of this architecture in the testbed have shown that the processing times for the QoS reservations, modifications and deletions are very small, even for a large number of simultaneous reservation requests, which enable the use of the WiMAX based architecture in NGN real-time and mobile environments without traffic disruptions.

## 8. References


Blake S., Black D., Carlson M., Davies E., Wang Z., Weiss W., *An Architecture for Differentiated Services*, IETF RFC 2475, December 1998.

Calhoun P., Loughney J., *Diameter Base Protocol*, IETF RFC 3588, September 2003.

Case, McCloghrie K., Rose M., *Management Information Base (MIB) for the Simple Management Protocol (SNMP)*, IETF RFC 3418, December 2002.

Guainella E., Borcoci E., Katz M., Neves P., Curado M., Andreotti F., Angori E., *WiMAX Extension to Isolated Research Data Networks*, IEEE Mobile WiMAX world summit, March 2007, Orlando, USA.

Gundavelli S., Leung K., Devarapalli V., Chowdhury K., Patil B., *Proxy Mobile IPv6*, IETF RFC 5213, August 2008.

Hancock R., Karagiannis G., Loughney J., Van den Bosch S., *Next Steps in Signaling (NSIS): Framework*, IETF RFC 4080, June 2005.

IEEE 802.16 WG, IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, *IEEE Std. 802.16-2004*, October 2004.

IEEE 802.16 WG a, IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems – Amendment 1: Management Information Base, *IEEE Std. 802.16f-2005*, December 2005.

IEEE 802.16 WG b, IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layer for Combined Fixed and Mobile Operation in Licensed Bands, *IEEE Std. 802.16e-2005*, December 2005.

IEEE 802.16 WG, IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems, Amendment 3: Management Plane Procedures and Services, IEEE 802.16g-2007, December 2007.

IEEE 802.21 WG, IEEE Standard for Local and Metropolitan Area Networks, Media Independent Handover Services, IEEE 802.21-2008, December 2008.

Knightson K.,Morita N., Towle T., *NGN Architecture: Generic Principles, Functional Architectures and Implementation*, IEEE Communications Magazine, p. 49-55, October 2005.

Manner J., Karagiannis G., *NSLP for Quality-of-Service Signaling*, IETF NSIS WG Internet-Draft, July 2007.

Nissilä T., Huusko J., Harjula I., Katz M., *Adapter Implementation between WiMAX Specific Layers and Network/Application Layers*, 1st BWA Workshop, Cardiff, Wales, September, 2007, pp. 328-333.

Perkins C., *IP Mobility Support of IPv4*, IETF RFC 3220, January 2002.

ITU-T a, General Principles and General Reference Model for Next Generation Networks, *Rec. Y.2011*, October 2004.

ITU-T b, General Overview of NGN, *Rec. Y.2001*, December 2004.

Johnson D., Perkins C., Arkko J., *Mobility Support in IPv6*, IETF RFC 3775, June 2004.

Koodli R., *Fast Handovers for Mobile IPv6*, IETF RFC 4068, July 2005.

McCloghrie K., Perkins D., Schoenwaelder J., *Structure of Management Information version 2*, IETF RFC 2578, April 1999.

Moskowitz R., Nikander P., *Host Identity Protocol (HIP) Architecture*, IETF RFC 4423, May 2006.

Neves P., Melo M., Sargento S., Pentikousis K., Sargento S., *Enhanced Media Independent Handover Framework*, IEEE Vehicular Technology Conference (VTC-Spring), May 2009.

Rosenberg J., Camarillo G., *SIP: Session Initiation Protocol*, IETF RFC 3261, June 2002.

Schulzrinne H., Hancock R., *GIST: General Internet Signalling Transport*, IETF NSIS WG Internet-Draft, July 2007.

WEIRD, WiMAX Extension to Isolated Research Data networks (WEIRD) project: http://www.ist-weird.eu

WiMAX Forum a, WiMAX End-to-End Network Systems Architecture Stage 2: Architecture Telnets, Reference Points, *WiMAX Forum Release 1, Version 1.2*, January 2008.

WiMAX Forum b, WiMAX End-to-End Network Systems Architecture Stage 3: Detailed Protocols and Procedures, *WiMAX Forum Release 1, Version 1.2*, January 2008.

19

# Intelligent WiMax Vertical Handovers

Nadine Akkari Adra
*King Abdul Aziz University,*
*Saudi Arabia*

## 1. Introduction

In next generation networks, tending to integrate different access technologies, the evolving WiMax 802.16 will play an important role in this integration for the purpose of offering the user the best possible service. WiMax will provide high bandwidth and high coverage, as opposed to current cellular systems providing high coverage and low bandwidth, and WiFi networks with limited coverage and high bandwidth. Hence, the vision of next generation networks is an all-IP network supporting heterogeneous access technologies such as WLAN, UMTS and WiMax. To enable inter- network access mobility known as vertical handover (VHO), a special mobility management is proposed in order for the handover to be performed with service continuity and minimum QoS degradation. The proposed intelligent vertical handover management consists of "guiding" the HO request to the best network capable of providing the required QoS and context parameters. The processing of the HO is based on an anticipated scenario where the HO decision parameters are compared to a pre-defined threshold in order for the HO to be guided accordingly. The handover could be guided to WiMax to benefit from high data rate and network availability to satisfy QoS requirements. The proposed handover management offers mobile users, roaming in next generation networks, service continuity by deciding on whether the handover is to be performed vertically or horizontally thus deciding on the destination network. In the context of the integration architecture, an Inter-Domain Management module IDM is introduced which is responsible of guiding the user to the next access network capable of offering the required QoS and context parameters. The mobility management is based on a "make-before-break" approach for a pre-network selection coupled with the Context Transfer Protocol CTP. In addition, when the horizontal handover could not be performed, it will be guided vertically. In this case, the user is guided to a different access technology, resulting in a vertical handover. This handover management will decrease the handover dropping probability and increase the HO performance in terms of less QoS degradation and best mapping of the HO requirements leading to "best" access network selection.

## 2. WiMax and Next Generation networks handovers

In the context of next generation networks enabling seamless handover for inter-technology access, a lot of work was conducted for wireless network integration architecture, mobility management and seamless handover. In [Wei, 2005], an overview of issues related to

horizontal and vertical handoffs were addressed with the architecture of integrating WLAN and WAAN networks based on Mobile IPv6. Bandwidth measurements of WLAN were used as the decision parameter taking into account QoS and user service requirement. In [Les, 2008], IEEE802.21 was proposed to provide the seamless handovers between two radio systems supporting the challenging decision and pre-execution phases of inter-technology handovers. In what we propose, WiMax will be the destination network that will be used to provide not only seamless handover, based on the specified decision parameters, but also the stand-by network for horizontal handovers that could not be performed and should be guided vertically. The evolving WiMax 802.16 will play a key role in next generation networks integrating different access technologies. Worldwide Interoperability for Microwave Access, or WiMax, intended mainly for the exchange of data at home or in the office, has the potential to provide a significant improvement in cost and performance compared to existing wireless broadband access systems. IEEE 802.16e will enable a new set of high-speed nomadic and mobile data services over a wide metropolitan area with lower-cost solutions, higher performance and reliability. Coverage will be based on large cells interconnected to provide the user with high data rate. The ability to maintain connection while moving across cell borders is a prerequisite for mobility and will be included as a requirement in 802.16e system profile [Finneran, 2004]. WiMax will be considered as the destination access network capable of providing higher data rates and higher network coverage hence providing seamless handover with the required QoS. Instead of dropping the handover call in case WLAN/UMTS handover or degrading its performance, the handover is guided with respect to the context of the application, the user profile as well as the current network parameters.

## 3. Intelligent Vertical Handovers

In order to manage the mobility of the mobile nodes moving from one access technology to another, known as vertical handover, a special mobility and domain management are needed in order for the session to be resumed on the new data path with minimum delay, packet loss, and QoS degradation. This is why FMIPv6 is adopted in our proposition as the mobility protocol to take advantage of the "Make-before-break" scheme through the proposed IDM. The mobile node could be connected to the New access router (NAR) due to a prior a stateless auto-configuration of the address Care-of-Address (COA) before the handover is established through the **IDM interference.** Relying on Fast MIPv6 (FMIPv6) principle with the IDM interfering in the fast handover operation, has the advantage of not only preparing the handover in advance but also the handover will not occur unless the required settings with the New Access Network (NAN) are made. The handover delay is minimized since the handover settings are done in advance specially in the case of the HO guided vertically, where on-line packets will be directly forwarded to the New Access Router (NAR) once the Mobile Node (MN) acknowledges its neighborhood by sending Fast Neighborhood Acknowledgment (FNACK) which are done within the WiMax network. In addition, the QoS degradation is minimized based on the IDM selection process where resources are checked in advance in order to avoid QOS deterioration of an on-going session. The Context Transfer Protocol messages are exchanged in order to send the current context of a mobile node to the IDM capable of selecting the required NAR in the new access network. The Context Transfer Protocol will transfer the context data of the current mobile

user through the IDM responsible of re-establishing the service with the same QoS parameters after the handover is performed. This transfer will eliminate the necessity of the MN to perform the AAA check and registration from scratch. Once the IDM is involved, re-authentication and re-establishment of the mobile host's authorization in the new network are transferred in the AAA context. This allows the mobile host to continue access in the new network once the handover is performed. In order for the IDM to select successfully the new access network, the main issue is that the handover should be performed after the context transfer messages are exchanged, so that the context data of the existing session will be transferred to the new network before the MN performs handover. In the proposed scenario, the mobile node "is guided" to the new access network capable of satisfying the required context parameters. Before describing the flow of messages as already briefly explained, figure 1 shows the WiMax vertical handover architecture with IDM interference.

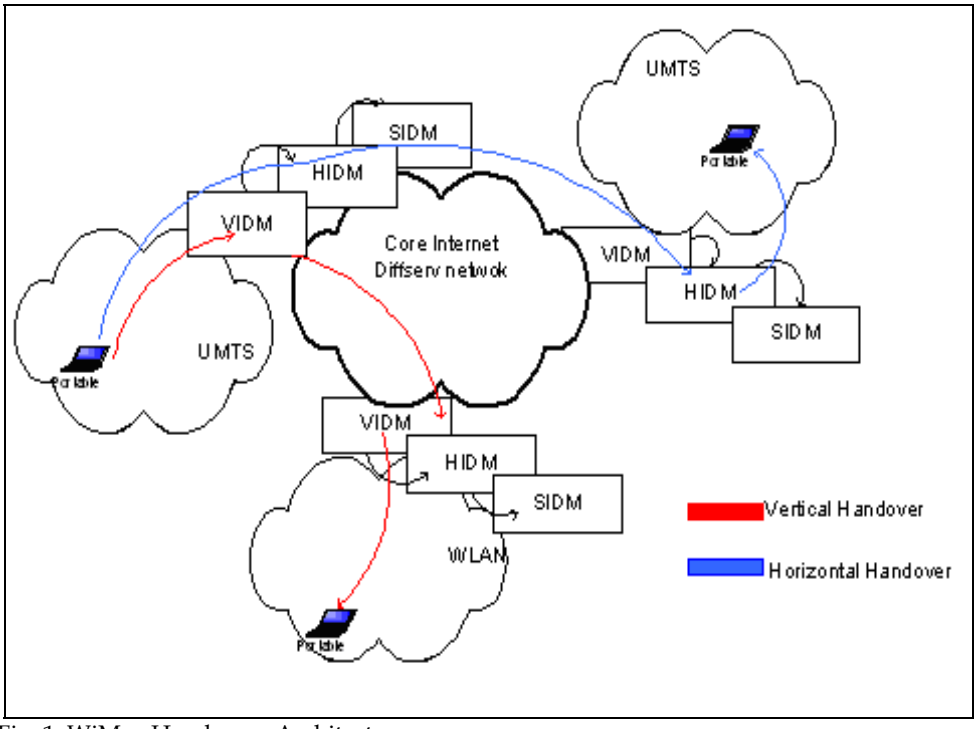



Fig. 1. WiMax Handovers Architecture

### 3.1 IDM Description

The Inter-Domain Management module IDM is responsible of mapping the mobile profile and the current context of a session to the Next IDM (NIDM) providing the necessary resources for the requested handover parameters. Each IDM keeps a table of the possible IDM's in the neighbored access network with their corresponding traffic load, network context and QoS parameters. The IDM will search accordingly for the available access routers: Each home-IDM scans its table to find the next IDM capable of providing the new

data path with the required context parameters processing the handover based on FMIPv6, context transfer functionalities and the WiMax vertical handover algorithm. IDM will perform the following tasks:

- Stores the QoS attributes of a current session based on the context feature sent in the context data message.
- Stores the profile of the neighbored IDM and their corresponding traffic parameters (data rate, traffic load, resource availability…).
- Selects the next access network based on the previously collected data.
- Treats and accepts (accordingly) the handover request based on the "WHO'S NEXT" phase.

To support this handover management strategy, three types of IDM are proposed: Vertical IDM (VIDM), Horizontal IDM (HIDM), and Stand-by IDM (SIDM). VIDM role is to accomplish successfully vertical handovers between heterogeneous access network UMTS to WLAN, for example. HIDM is concerned with the inter handovers or the mobility belonging to the same access technologies. SIDM is to guide the handover to another network with high capacity and capable of satisfying the users QoS requirements from WLAN to WiMax, for example.

### 3.2 WiMax Vertical Handover Algorithm

The mobile node can perform handover from WLAN to WiMax or from UMTS to WiMax since WiMax will enable higher throughput and higher coverage compared to UMTS and WLAN. In the WiMax Vertical Handover Algorithm, the handover is considered for both network-initiated handover and user initiated-handover. These two cases will be managed by the vertical handover based on the proposed scenario:

**Network-initiated handover:** Consider the case where the mobile is in a WLAN hotspot. In the case of network-initiated handover, two thresholds $T_1$ and $T_2$ are chosen to initiate the handover. These two thresholds are set below the real threshold T which makes the user loose network coverage. Hence, if the Received Signal Strength Indicator (RSSI), or S for simplicity, falls below $T_1$, $S<T_1$, then the handover could still to be handled locally within the same access network. Thus the WLAN network will be the destination network and the HO is to be performed horizontally. On the other side, If $T_1<S<T_2$, then the handover is guided to the WiMax access network. Since, in this case, S is approaching the real threshold T and the HO should be performed smoothly with minimum delay, the handover is guided to the WiMax network providing high coverage and capable of providing service continuity. In this context, the handover dropping probability will be decreased sine WiMax network provides higher network capacity and coverage compared to WLAN. If $S > T_2$, which will not be considered unless the HO was not performed earlier, high priority WiMax HO should be requested to enable the HO to be performed with no additional delay.

In the case of **user initiated-handover**, the handover is decided according to the user profile and application parameters in order not for the application to degrade during HO. This could be done according to the user preferences with respect to a predefined threshold. The user initiates the HO based on the user profile (user preferences, charges, type of service, etc), network profile such as network bandwidth, and application profile such as traffic classes and

types. The pre-defined degradation thresholds $D_1$ and $D_2$ are proposed based on accepted user profile and QoS requirements. Thus the handover is first initiated when the service degrades below the threshold $D_1$. If the degradation parameter D falls below $D_1$, then the handover is forwarded to another WLAN network capable of providing the required service. This could also be chosen if the service is tolerant to degradation according to the user profile and greater data rate is not required. The handover could be guided to another access network, WiMax network, if $D_1<D< D_2$. Thus, before reaching the real degradation threshold, the HO is guided to WiMax providing higher data rates and higher coverage. This could also be the case where the service could be no more served with the same QoS, and higher bandwidth is required according to the current application profile. The handover could also be guided to WiMax with higher priority request if $D>D_2$. In the last 2 cases, the handover is guided vertically to WiMax as a destination network. In all these cases, the IDM will exchange messages with the new IDM in the next selected network. These signaling messages are treated within the Diffserv core network as high priority traffic in order for the handover overhead to be minimized. The IDM traffic is mapped to the EF class providing low loss rate, low latency, assured bandwidth and low jitter [Mykoniati el al, 2003]. SIDM is responsible of scanning all the table entries to find the next IDM capable of providing the mobile with the required parameters ensuring a seamless handover to WiMax network. The flowchart of the proposed WiMax vertical handover scenario in case of user-initiated handovers as well as network-initiated handover is shown in figure 2.

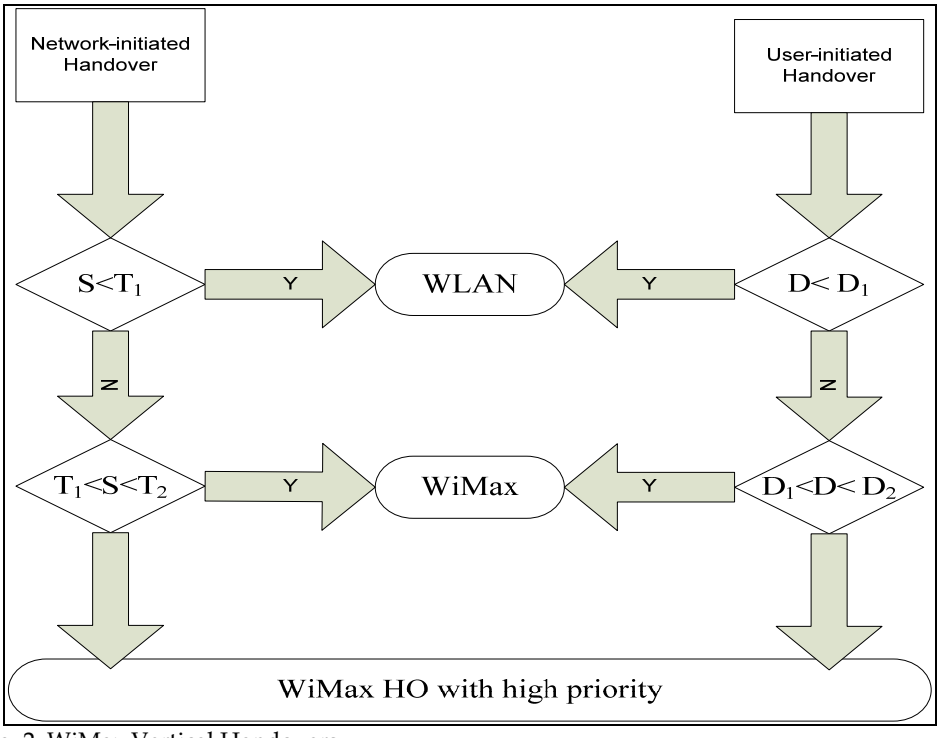



Fig. 2. WiMax Vertical Handovers

## 4. WiMax Vertical Handover processing

The main operations involved in the vertical handover from WLAN to WiMax are shown in figures 3 and 4. The context transfer messages used in the proposed scenario are similar to that described in [Kempf, 2002]. The transport of these messages is based on ICMPv6 protocol where the type and code are set to the specific type of CTP messages and the context data transfer is added in the data option field [Politis, 2004]. Contexts are identified by FPT (Feature Profile Type), which is a 16-bit unsigned integer. The context type numbers are assigned by IANA, and handled according to the message specifications [Politis, 2004]. FPT are transferred by data blocks used for transferring the actual feature context. Context Transfer Data (CTD) Message is sent from previous IDM (PIDM) to Previous Access Network (PAR) and includes feature data (CTP data). Once the required context parameters are mapped into the new IDM (NIDM) in the new network, the mobile user can perform handover without QoS degradation due to the appropriate selection of NIDM and accordingly the new access router in the new access network. The WiMax is composed of the Mobile Control Point (MCP) which provides the control and mobility anchor point for a mobile station (MS) as it moves between base stations (BSs) in the access network**.** Network Operations and Support Services (NOSS) includes functions required to operate and maintain the wireless access network such as authentication, authorization, and accounting (AAA) services; and configuration services. Figure 3 shows the WiMax network components role during the handover. The complete message exchange is done through the inter-Domain manager (IDM) which monitors and set the next access network within its domain based on FMIPv6 and context transfer protocol messages described in [Koodli, 2004; Kempf, 2002]. The scenario of the exchange is as follows:

- MN sends RTSOLPR to PAR to indicate HO initiation and to start the operation of acquiring a new address. The PAR will not reply directly (as in the normal FMIPv6 signaling flow) but it will wait for the IDM to select the appropriate NAR.
- MN sends Context Transfer Activation Request CTAR to PAR and PIDM prior to HO. this message contains MN IP address and PIDM IP address, and the entire context to be transferred. It also contains token to be used by NIDM for verification.
- Context Transfer Request CTREQ is sent from PIDM to PAR to update the context already available in the entry of the PIDM.
- PAR starts to send Context Transfer Data CTD to PIDM with the feature data context parameters.
- The "Who's next phase" is started. PIDM will start scanning (based on CTD) for a NIDM whose resources will satisfy the required context of MN asking for HO.
- Handover Initiation HI is sent from the PIDM to NIDM and NAR to configure the new CoA.
- All the FMIPv6 signaling flow will now proceed to perform the handover request to the selected access network.
- MN will finally send a Fast Neighbor Advertisement (FNA) to the NAR-MCP to notify it of its presence in the new access network.
- The packets destined to the old access router are no more buffered in PAR but in the PIDM which will forward these packets to the NIDM, as soon as the mobile node starts to communicate with the NAR.

The messages exchanged in the vertical handover scenario from WLAN to WiMax are shown in figure 3.

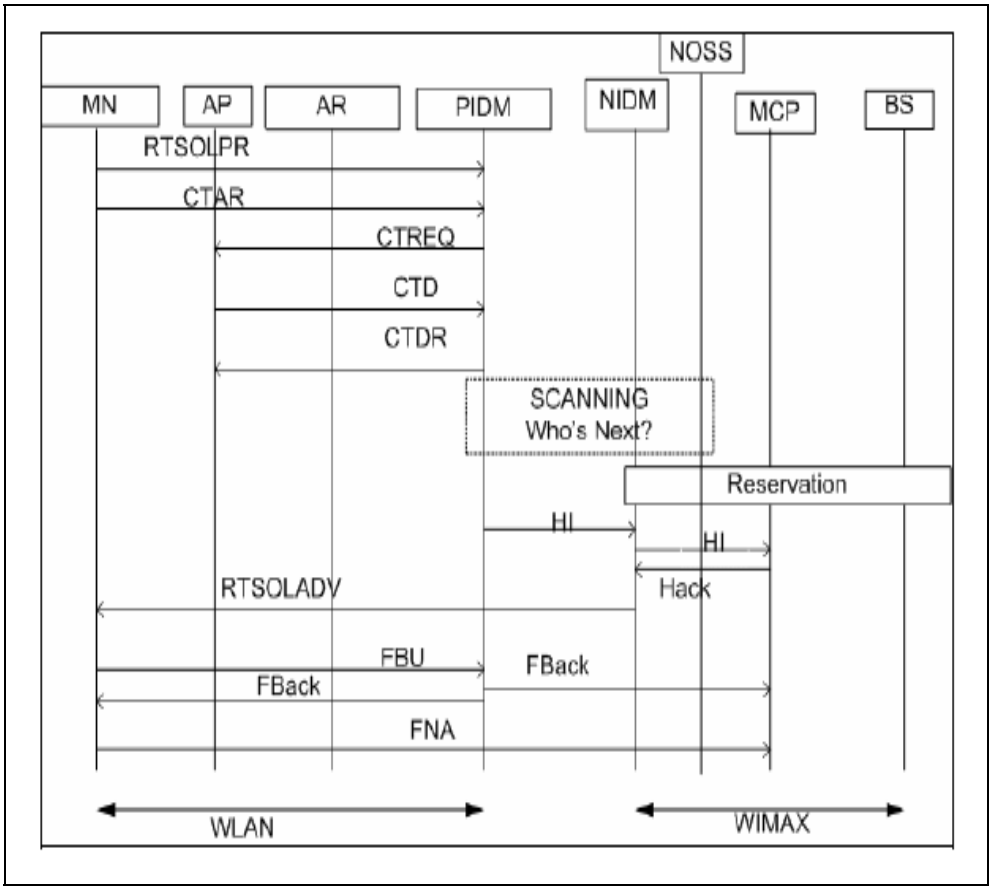



Fig. 3. WLAN to WiMax Handover

In the case of vertical handover, before the MN is out of coverage in the current network or before application degradation occurs, the handover operation is initiated. The handover could be requested by the user to satisfy the user policies and preferences. The handover could also be requested to benefit from a faster and high coverage network as from WLAN networks to WiMax. The handover is first requested before the real threshold is reached in order not to be disconnected. Once initiated, the HO request is treated by starting the context transfer exchange and mapping the requested QoS parameters into the next selected network. The request is classified into horizontal or vertical handover. If the application is tolerant to degradation, the HIDM will perform horizontal HO. On the other hand, if the application is of high sensitivity and not tolerant to degradation, the HO call will be guided to a SIDM which will perform vertical HO. This is done in the "WHO'S NEXT" phase. The mobility management will take place and the CoA is configured and updates to the IDM

and AR are done. Next, the mobile node will start to communicate with the New access router (NAR) on the new access network.

## 5. WiMax Vertical Handover Evaluation

The handover from WLAN to WiMax is modeled on the following basis: Since WiMax has more bandwidth and can accommodate more calls; the following Markovien model is used figure 4**.** In this model, the HO calls are privileged over the local calls. The local calls are accepted until the threshold T is reached. When the number of calls is greater then T, then the only the HO calls (vertical calls from WLAN) are accepted.
SIDM admission Model (UMTS, WLAN to WiMax)The following assumptions were made:

- The IDM is modeled as M/M/C/C queuing system.
- C-T interval channels are the guard band channels used only for handover calls.
- The local traffic is a poisson process of rate $\lambda_l$ and the handoff request is a poisson process of rate $\lambda_h$.
- The time of stay in IDM is exponentially distributed of rate $1/\mu$.

The blocking and dropping probabilities $P_B$ and $P_D$ in function of the threshold T are as shown in figures 5 and 6. The choice of the threshold will highly affects the handover dropping probability as well as the blocking probability of the incoming local calls. The variation of these two probabilities with respect to the threshold is given in figures 5 and 6. The maximum allowed threshold is when T=C. As T increases, the blocking probability decreases, since the calls (local requests) are accepted until T is reached; whereas, the dropping probability is increased since the probability of dropping the handover calls is increased with no priority scheme. Figure 4 shows the results with T and C difference from 0 to 50.

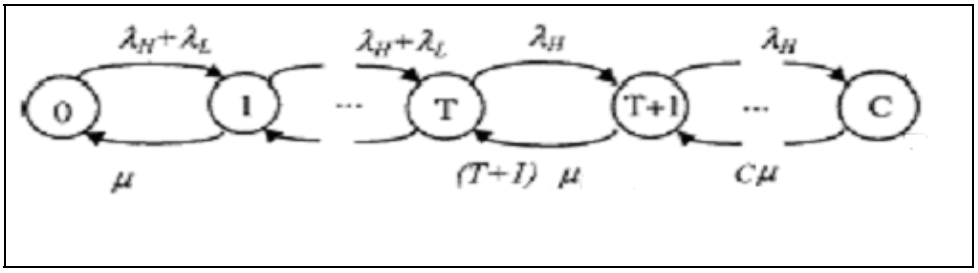



Fig. 4. SIDM model for WiMax Handovers

## 6. Conclusion

The horizontal and vertical handover considered are a complementary solution for a seamless handover. We considered that VHO could occur when the HHO could not be performed successfully or seamlessly. Thus, when no available resources are possible to perform the required parameters of the HHO, the VHO will take place. The proposed solution provides QoS provisioning and seamless handover between different access technologies. Integrating the emerging WiMax in the all-IP architecture is a vertical handover solution that decreases the Handover dropping probability and profits from high data rates provided by WiMax networks thus guiding the handover request to WiMax.

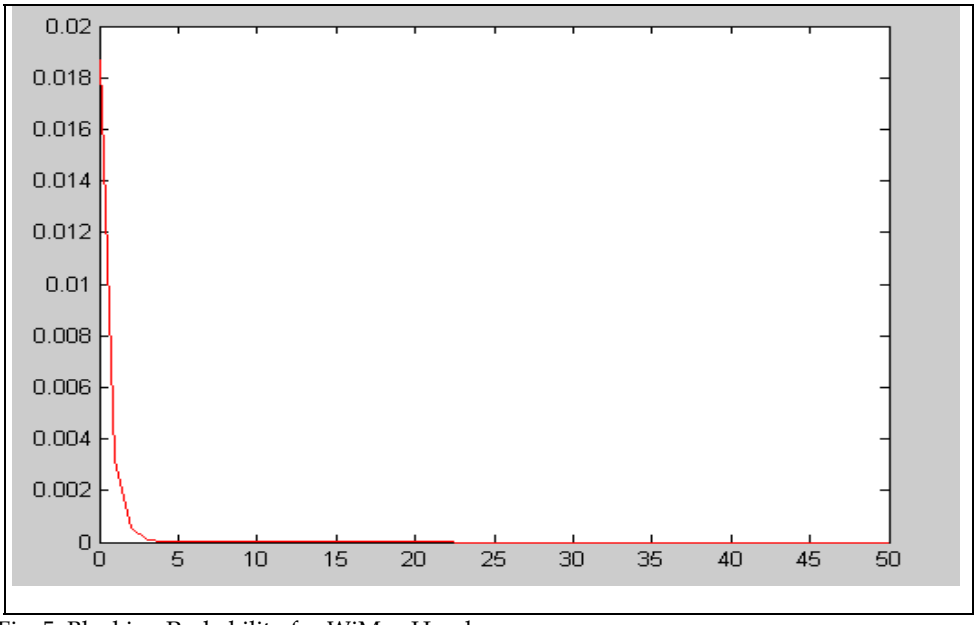



Fig. 5. Blocking Probability for WiMax Handovers

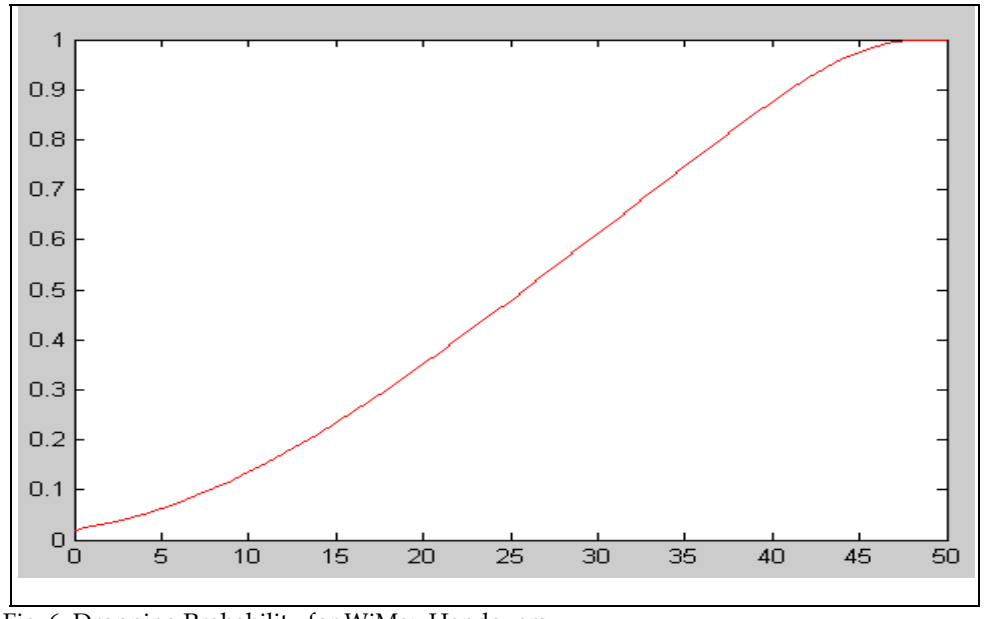



Fig. 6. Dropping Probability for WiMax Handovers

## 7. References

J. Kempf.(2002). Reasons For Doing Context Transfers Between Nodes in an IP Access Network, RFC 3374, Internet Engineering Task Force (IETF), Sep 2002

Wei, L.; Li, Ming. C.; Meng, Chang, C. & Yeali, S. A Framework of Handoffs in Wireless Overlay Networks Based on Mobile IPv6, Ieee Journal On Selected Areas In Communications, Vol. 23, No. 11, November 2005

R. Koodli. (2004). Fast Handovers for Mobile IPv6", draft-ietf-mipshop-fast-mipv6-03.txt, Nokia ReseCenter, October 2004.

M. Finneran. (2004). WiMax versus Wi-Fi, a comparison of technologies, markets and business plans, June 2004.

E. Mykoniati; C. Charalampous; P. Georgatsos; T. Damilatis; D. Goderis, P. Trimintzios, G. Pavlou & D. Griffin. (2003). Admission Control for Providing QoS in DiffServ IP Networks: The TEQUILA Approach. IEEE Communications Magazine, January 2003.

C. Politis, *et al*. (2004). Hybrid Multilayer Mobility Management with AAA Context Transfer Capabilities for All-IP Networks. IEEE Wireless Communications, August 2004.

Les, A.; Scott, M.; Qiaobing. X. & Vivek, G. (2008). Mobility Using IEEE 802.21 In A Heterogeneous IEEE 802.16/802.11-Based, IMT-Advanced (4g) Network, IEEE Wireless Communications, April 2008.

20

# Signal Metrics for Vertical Handoff Towards (Cognitive) WiMAX

François-Xavier Socheleau[1], Sébastien Houcke[1],
Philippe Ciblat[2] and Abdeldjalil Aïssa-El-Bey[1]
[1]*Institut TELECOM; TELECOM Bretagne; UMR CNRS 3192 Lab-STICC ;*
*Université européenne de Bretagne,*
[2]*Institut TELECOM; TELECOM ParisTech,*

## 1. Introduction

Mobile WiMAX is currently operating in several frequency bands ranging from 2.3GHz to 5.8GHz. To accommodate the anticipated growth in mobile services, there is ongoing pressure on regulators to make additional spectrum available for mobile applications. For instance, the WiMAX Forum is currently investigating new spectrum opportunities below 1GHz (WiMAX Forum, 2008) that would offer good mobile propagation conditions and larger coverage. As a consequence of the physical limits on the amount of available spectrum, it is very likely that regulatory policies in the some bands will evolve from the current fixed spectrum rules to opportunistic spectrum sharing models (U.S. FCC, 2002; 2008). In these new models, regulators allocate section of spectrum that can be used by several (or any) systems under a minimum set of restrictions called spectrum etiquette. Spectrum sharing will require cognitive radio technology (Akyildiz et al., 2006) with dynamic spectrum access (DSA) capabilities and a high level of flexibility to alter wireless system transmission parameters (i.e. adapt the carrier frequency, modulation parameters, power, etc.) according to the surrounding radio environment and the specified policy. Thanks to its adaptive PHY layer and specially to the OFDMA scalability, WiMAX is very well prepared to meet the spectrum sharing model requirements (Blaschke et al., 2008; Leu et al., 2009).
In parallel to the development of new mobile services, there is an emerging trend to provide users with ubiquitous (anywhere, anytime) seamless wireless access. This can be made possible by taking advantage of the coexistence of complementary heterogeneous networks such as 3G(LTE), WiMAX, Wifi etc. In such environment, a challenge is to develop multi-interface terminals able to smartly switch from one wireless interface to another while maintaining IP connectivity and required QoS. This switching process is known as *vertical handoff*. In contrast with horizontal handoff (handoff within a network made of homogeneous wireless interfaces), vertical handoff triggering is based on a multi-criteria decision involving signal quality, bandwidth, traffic load, price, battery status, latency etc. (McNair & Zhu, 2004).
Based on the current WiMAX PHY layer and especially on OFDM(A) properties, we present

original works and synthesise existing methods to estimate signal metrics relevant as inputs for any algorithm that has to decide whether to trigger a vertical handoff from any system to WiMAX. We consider the possibility that the WiMAX interface of interest may be cognitive so that the carrier frequency in use by a base station (BS) may evolve over the time. We assume that the section of spectrum allocated to cognitive operation is channelized so that BSs can only transmit on a finite set of possible carrier frequencies. The signal metrics estimation is sequenced in three steps.

1. WiMAX detection: First, upon start-up, pattern detection has to be performed by the terminal in order to identify the spectrum channels in use by a WiMAX BS. In a cognitive radio context, classical coherent frame or superframe preamble detection may not be relevant as it can be a very slow process if many channels have to be scanned. We present more efficient methods based on OFDM cyclostationarity (Bouzegzi et al., 2008) or WiMAX pilot tones properties (Socheleau et al., 2008a; 2009).
2. SINR estimation: Signal to Interference plus Noise Ratio is known to be a good indicator for communication link quality estimation. In the book chapter we discuss the relevance of data-aided SINR estimators for cognitive WiMAX and detail an innovative blind algorithm exploiting OFDMA cyclostationarity (Socheleau et al., 2008b) and a likelihood metric.
3. OFDMA slot activity rate estimation: the time-frequency activity of WiMAX signals is direclty proportional to the traffic load which represents an informative input to the handoff decision algorithm. We show that this rate can directly be deduced by decoding the downlink and uplink mapping (DL/UL-MAP) messages (Dai et al., 2008) but can also be blindly estimated. The latter method suiting cognitive WiMAX better.

This contribution is organised as follows. Cognitive WiMAX scenarios are discussed in section 2. Section 3 presents the signal model. Channels detection algorithms are then proposed in section 4. SINR and slot activity rate estimators are detailed in sections 5 and 6 respectively. Finally, conclusions are presented in section 7.

## 2. Cognitive WiMAX scenarios

On-going reforms (U.S. FCC, 2008; Wireless Innovation) to spectrum management now offer the opportunity to better exploit highly underutilised portions of spectrum. Regulatory bodies are indeed considering to extend the range and number of license-exempt bands as well as to authorise secondary usage of some licensed spectrum. These new policies open up the possibility to make a clever use (in term of spectral efficiency) of the radio resources by developping devices embedding cognitive radio (CR) technologies (Haykin, 2005). CR enables dynamic spectrum access (IEEE DySPAN) by sensing the electro-magnetic environment and specifically by detecting and then operating on idle frequency channels (or white spaces) at a particular time and place (Haykin, 2005).

From a service provider perspective, CR directly translates in spectral efficiency improvement (Haddad et al., 2007) and therefore in capacity increase. WiMAX could benefit from the CR technology in 3 main scenarios:

1. Today, WiMAX systems mainly operate as primary users in licensed bands (i.e. WiMAX service providers own and control their spectrum). In that context, the network operations could be simplified and the capacity increased by implementing what is called the *cognitive channel assignment* in (Leu et al., 2009). The idea is that frequency channels are not statically assigned to cells but BS are instead equipped with sensitive detectors and dynamically

assign channels to subscriber stations (SSs) based on spectrum availability. Power control is also employed to increase frequency reuse in conjunction with spectrum sensing.

2. In a near future, WiMAX providers may have an economical interest in operating in license-exempt bands (no license fee, economies of scale due to WiMAX devices profusion etc.). This is even more likely if the regulatory bodies open up more of these bands and if they are below 1GHz (which would increase the coverage and thus reduce the number of BS). In this scenario and still for capacity reasons, WiMAX will gain from CR technology as it will have to compete with other systems to get access to the spectrum.

3. In a longer term, we could also imagine WiMAX networks operating as secondary networks in frequency ranges under licenses owned by other systems. This typical CR scenario has already been suggested in (Blaschke et al., 2008) where WiMAX coexists with GSM.

These scenarios mainly differ in the kind of interference they have to deal with (self-interference or other system interference) and in the level of possible cooperation between the various systems sharing the same spectrum (cooperation is hardly possible in the last two scenarios).

Even if a few PHY or MAC layer modifications (sensing signalling (Blaschke et al., 2008), superframe encapsulation (Stevenson et al., 2009) etc.) may be needed for mobile WiMAX to operate in a cognitive context, it is globally well prepared to meet the CR requirements. It is indeed scalable by turning on or off some OFDMA subcarriers to fit into the white spaces, its supports adaptive modulation and coding as well as power control and it can perform spectrum sensing in the frequency domain thanks to its built-in FFT.

## 3. Signal model

Assuming that in a given frequency channel, a transmitted OFDMA symbol consists of $N$ subcarriers, the discrete-time baseband equivalent transmitted signal is given by [1]

$$x(m) = \sqrt{\frac{E_s}{N}} \sum_{k \in \mathbb{Z}} \sum_{n=0}^{N-1} \varepsilon_k(n) c_k(n) e^{2i\pi \frac{n}{N}(m-D-k(N+D))} g(m-k(N+D)), \tag{1}$$

where $E_s$ is the signal power; $c_k(n)$ are the transmitted symbols at the $n$-th subcarrier of the $k$-th OFDM block. These data symbols are assumed to be independent and identically distributed (i.i.d), $D$ is the CP length; $m \mapsto g(m)$ is the pulse shaping filter. $\varepsilon_k(n)$ represents a i.i.d sequence of random variables valued in $\{0,1\}$ that express the absence or presence of signal activity in a time-frequency slot $(k,n)$.

Let $\{h(l)\}_{l=0,\cdots,L}$ be the baseband equivalent discrete-time propagation channel impulse response of length $L+1$. We assume that D is chosen such that $D > L+1$. The received samples of the OFDMA signal are then expressed as

$$y(m) = e^{-i(2\pi\delta_f \frac{m-\tau}{N} + \theta)} \sum_{l=0}^{L-1} h(l) x(m-l-\tau) + \eta(m) + \gamma(m), \tag{2}$$

[1]Note that for signal metric estimation presented in this chapter there is no need to differentiate the OFDMA users.

where $\delta_f$ is the carrier frequency offset (normalised by the subcarrier spacing), $\theta$ the initial arbitrary carrier phase, $\tau$ the timing offset and $\eta$ a zero mean circularly-symmetric complex-valued white Gaussian noise of variance $\sigma_\eta^2$ per complex dimension. $\gamma$ denotes a possible interference with other systems.
At reception, the signal-plus-interference-to-noise ratio (SINR) is expressed as

$$SINR = \frac{S}{\sigma^2} \tag{3}$$

where

$$S = E_s \mathbf{E}[|\varepsilon_k(n) c_k(n)|^2] \sum_{l=0}^{L} |h(l)|^2 \tag{4}$$

and

$$\sigma^2 = \sigma_\eta^2 + \mathbf{E}[|\gamma(m)|^2]. \tag{5}$$

$\mathbf{E}[.]$ stands for the expectation operator.

## 4. Active channel detection

Since cognitive networks dynamically modify their operating frequency so as to transmit on unused channels, network entry for SSs may not be straightforward as upon star-up SSs are not necessarily aware of the channel(s) currently used by a BS. As an example, the TV band recently released by the FCC for unlicensed operation based on CRs can range up to several hundreds of megahertz which leads to several ten or so possible channels. Figure 1 shows a basic illustration of spectrum usage resulting from dynamic spectrum access where three systems share the same frequency range divided into 5 channels.

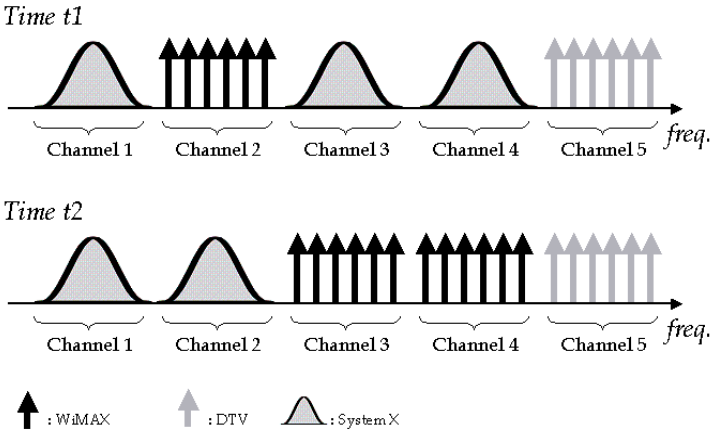



Fig. 1. Example of a channel occupation scenario at different time intervals.

Therefore, any multi-interface terminal with vertical handoff capability, sensing for a cognitive WiMAX BS, has first to detect the frequency channel(s) where the BS operates. In this section, we develop 3 different approaches to perform this detection

• Coherent detection: based on the knowledge of downlink training sequences or preamble, the terminal can reliably detect a BS by a simple cross-correlation. This method offers the best performance but can be time consuming.

• OFDM cyclostationary feature detection: as we will show, WiMAX signals are cyclostationary which can represent a signature useful for detection. Cyclostationarity is relatively easy to detect, not time consuming but has the main drawback of not differentiating downlink (DL) from uplink (UL) frames which may be a concern in FDD networks.

• OFDM pilot structure detection: Pilot tones are of great interest for detection since they are (almost) always present in the transmitted signals and therefore easy to intercept and can be used to discriminate DL from UL frames.

### 4.1 Coherent detection

To facilitate initial synchronisation, each WiMAX DL frame starts with a preamble (IEEE Std. 802.16, 2005) belonging to a finite set of sequences known a priori by SSs. In a cognitive context where the number and the frequency of the channels used for DL frame transmission are not known by SSs and where active channels are not necessarily contiguous, the current WiMAX preamble may not be relevant to enable initial synchronisation. It may be required to add a bit of signalling overhead. As a reference, for initial synchronisation, the IEEE 802.22 (Stevenson et al., 2009) has defined a superframe structure that encapsulates classical frames and that starts with a preamble duplicated on every channel used by a BS. When a terminal finds a superframe header on a particular channel it then obtains all the necessary information (number and frequency of all the channels used by the BS etc.) to demodulate the frames that follow. This mechanism is pretty efficient as channel detection can be performed by simple cross-correlation between the received signal and the known preamble sequence. In addition, a SS needs just to detect a single channel to get all the information to get connected to the network. However, the drawback of the superframe is that it decreases the network capacity so that to limit the overhead, the superframe header must not be sent frequently. As an example, the 802.22 superframe header is broadcasted every 160 ms which corresponds to 16 frames. In our context where a multi-interface terminal is sensing all the surrounding networks to decide which one is the most appropriate to get connected to, long delays in getting superframe synchronised with each network before deciding which one is the most suited to its needs may not be tolerated. This is even more critical if the set of possible active channels is large. To avoid time consuming process, two detection methods based on the WiMAX signal properties as opposed to frame structure are presented in the sequel.

### 4.2. OFDM cyclostationary feature detection

To limit the impact of the propagation channel time spread, a cyclic prefix (CP) is added before each WiMAX OFDM symbol (IEEE Std. 802.16, 2005). This CP induces periodic correlation on the OFDMA signals that can be used as a detection pattern (Bouzegzi et al., 2008 ; Ki et al., 2006 ; Ishii & Wornell, 2005 ; Yucek & Arslan, 2007 ; Oner & Jondral, 2007).

More precisely, the WiMAX signal is said to be cyclostationary. Using cyclostationarity for detection is appealing in contrast to energy detection since the noise is hardly never cyclostationary. WiMAX detection can thus be based on the cyclic autocorrelation defined as (Gardner et al. , 2006)

$$R_x^{\alpha}(u) = \lim_{M\to+\infty} \frac{1}{M} \sum_{m=0}^{M-1} \mathbf{E}\left[ x(m)x^*(m+u) \right] e^{-2i\pi m\alpha} \tag{6}$$

where $\alpha$ represents the cycle frequencies. From Eq. (1), the cyclic autocorrelation of an OFDMA signal verifies

$$\begin{aligned} R_x^{\alpha}(N) &= \lim_{M\to+\infty} \sum_{m=0}^{M-1} \sum_{k\in\mathbb{Z}} g(m-k(N+D)+N)g(m-k(N+D))e^{-2i\pi m\alpha} E_s \frac{\mathbf{E}\left[ |\varepsilon_k(n)c_k(n)|^2 \right]}{NM} \sum_{n=0}^{N-1} e^{2i\pi n} \\ &= E_s \mathbf{E}\left[ |\varepsilon_k(n)c_k(n)|^2 \right] \frac{\sin(\pi\alpha D)}{(N+D)\sin(\pi\alpha)} e^{-i\pi\alpha(D-1)} \sum_{q\in\mathbb{Z}} \delta\left( \alpha - \frac{q}{N+D} \right) \end{aligned} \tag{7}$$

where $\delta(.)$ is the Kronecker symbol. The WiMAX signature then corresponds to the set of powerful cycle frequencies. Figure 2 illustrates the the correlation properties induced by the OFDM CP.

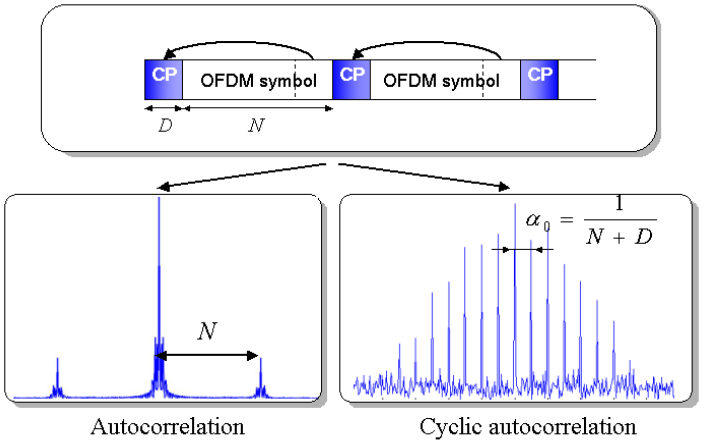



Fig. 2. Illustration of the correlation properties induced by the OFDM cyclic prefix.

As shown in (Jallon, 2008), the cyclic autocorrelation at reception writes

$$R_y^{\alpha}(N) = R_x^{\alpha}(N) \sum_{l=0}^{L-1} |h(l)|^2 \, e^{-2i\pi \frac{\alpha l}{N+D}}. \tag{8}$$

Detection can then be done by estimating the energy of the cycle frequencies $\alpha_q = \frac{q}{N+D}$ thanks to the following cyclic autocorrelation estimator

$$\hat{R}_y^{\alpha_q}(N) = \frac{\sum_{m=0}^{M(N+D)-1} y(m)y^*(m+N)e^{-2i\pi m\alpha_q}}{M(N+D)} \tag{9}$$

where $M(N+D)$ is the length of the observation window. We assume that $N$ and $D$ are known by the multi-interface terminal. The choice of $N$ results from the channel bandwidth defined by regulatory bodies and $D$ is standardised and does not change on the flight. Several structures of detectors have been proposed in the literature, refer to (Bouzegzi et al., 2008; Jallon, 2008; Oner, 2007) for instance.

As illustrated in subsection 4.4, cyclostationary feature detection shows excellent performance even for short CP. However, since the WiMAX DL and UL signals have the same cylostationary signature they cannot be distinguished. This can be a concern for the WiMAX FDD as the signal metrics used to status on a possible vertical handoff have to be estimated on DL subframes. Moreover, in a CR scenario, it is very likely that OFDM systems competing for the access to the same frequency band will have very close (or even the same) modulation parameters (subcarrier spacing, CP length etc.) and thus similar cylostationary signatures. The PHY layer design is indeed strongly driven by features related to the spectrum in which a system operate (propagation channel, available bandwidth etc.). Therefore, methods that involve more particular signatures to detect WiMAX BS may be required and very useful.

### 4.3. OFDMA pilots structure detection

Pilot tones are of great interest for WiMAX BS detection since (i) they enable to discriminate DL from UL frames as well as systems with similar modulation parameters (ii) they are always present in the transmitted signals and therefore easy to intercept (iii) they are power boosted. If we now consider the pilot tones, the signal model of Eq. (1) becomes

$$x(m) = \sqrt{\frac{E_s}{N}}[x_d(m) + x_t(m)], \tag{10}$$

where

$$x_d(m) = \sum_{k\in\mathbb{Z}} \sum_{\substack{n=0 \\ n\notin \mathrm{I}(k)}}^{N-1} \varepsilon_k(n)a_k(n)e^{2i\pi\frac{n}{N}(m-D-k(N+D))} g(m-k(N+D)), \tag{11}$$

and

$$x_t(m) = \sum_{k\in\mathbb{Z}} \sum_{n\in \mathrm{I}(k)} b_k(n)e^{2i\pi\frac{n}{N}(m-D-k(N+D))} g(m-k(N+D)). \tag{12}$$

$\mathrm{I}(k)$ denotes the set of pilot subcarrier indexes of the $k$ -th symbol, $a_k(n)$ and $b_k(n)$ are the data and pilot symbols respectively. Note that

$$c_k(n) = \begin{cases} b_k(n), & if\ n \in \mathrm{I}(k) \\ a_k(n) & otherwise. \end{cases} \tag{13}$$

OFDMA pilot symbols, used for channel estimation and/or synchronisation purposes, are often replicated according to a pre-defined time/frequency distribution. This property induces correlation between pilot subcarriers that can be exploited in conjunction with the periodicity of the time/frequency pilot mapping to perform WiMAX BS detection. The IEEE 802.16 standard defines several pilot tones structures that depend on the PHY layer (OFDM or OFDMA), the DL and UL frames and on the subcarrier permutation mode. For the OFDMA PHY layer (Mobile WiMAX), $b_k(n) = 8/3(1/2 - w_k)$ where $w_k$ is a pseudo-random binary sequence and the pilot mapping is periodic such that $\mathrm{I}(k+K) = \mathrm{I}(k)$, $K \in \mathbb{Z}$ and depends on the permutation mode ($K = 2$ in DL-PUSC and DL-FUSC and $K = 9$ in optional FUSC etc.). Such a periodicity is a useful property that induces cyclostationarity in OFDMA frames. In fact, as shown in (Socheleau et al., 2009), if the pilot tones are designed such that

$$b_k(p) = b_{k+d^{(p,q)}}(q)e^{i\phi}$$

with $d^{(p,q)} \in \mathbb{Z}$ and $\phi \in [-\pi;\pi[$, then the processes $\{c_k(p)\}_k$ and $\{c_k(q)\}_k$ are jointly cyclostationary with the set of powerful cycle frequencies given by $\mathrm{A}_{c_k(p)c_k(q)} = \left\{ \frac{m - \lfloor K/2 \rfloor}{K}, m \in \{0,1,\cdots,K-1\} \right\}$ where $\lfloor\ \rfloor$ stands for integer flooring. The WiMAX signature $S$ then corresponds to the combination of $p, q, d^{(p,q)}$ and $K$. It is defined as

$$S \overset{\Delta}{=} \left\{ \left( p, q, d^{(p,q)}, K \right) \mid \mathrm{A}_{(p,q)} \neq \varnothing \right\}. \tag{14}$$

For the DL-PUSC mode (the only mandatory permutation mode), $K = 2$, $d^{(p,q)} = 0$ or 1 and the set of $(p,q)$ is defined by all the possible combinations within $\bigcup_{k \in \mathbb{Z}} \mathrm{I}(k)$. The figure below shows the subcarrier allocation in a DL-PUSC cluster.

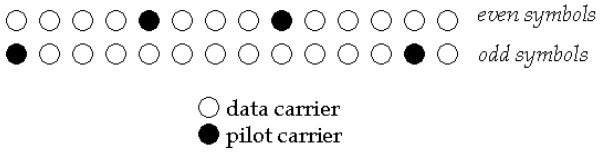



Fig. 3. DL-PUSC subcarrier allocation.

The detection can then be performed thanks to the cyclic cross-correlation energy estimation

$$J \overset{\Delta}{=} \sum_{(p,q) \in \xi} \left( \sum_{\alpha \in \mathrm{A}_{(p,q)}} \left| \hat{R}^{\alpha}_{Y(p,q)} \left( d^{(p,q)} \right) \right|^2 \right) \tag{15}$$

where

$$\hat{R}^{\alpha}_{y^{(p,q)}}\left(d^{(p,q)}\right) = \frac{1}{M-d^{(p,q)}}.\sum_{k=0}^{M-d^{(p,q)}-1} Y_k(p)Y^{*}_{k+d^{(p,q)}}(q)e^{-i2\pi\alpha k} \tag{16}$$

and $\xi = \{(p,q) \mid \mathrm{A}_{(p,q)} \neq \varnothing \text{ and } d^{(p,q)} + K \leq M\}$. $Y_k(n)$ are the observations expressed as

$$Y_k(n) \overset{\Delta}{=} \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} y[k(N+D)+D+m]e^{-2i\pi\frac{nm}{N}}. \tag{17}$$

A constant false alarm rate (CFAR) detector is suggested in (Socheleau et al. , 2009) where the detection threshold is given by a Laguerre expansion. As shown in Figure 4, the estimation of the cycle frequencies power deteriorates with timing missynchronization and/or frequency offset as inter-symbol (ISI) and inter-carrier (ICI) interferences occur. However, $J$ is maximum in the case of perfect synchronisation so that $\varepsilon$ and $\tau$ can be estimated as

$$[\hat{\varepsilon}, \hat{\tau}] = \underset{(\varepsilon,t)}{\operatorname{argmax}} J^{(\varepsilon,t)} \tag{18}$$

where $J^{(\varepsilon,t)}$ is defined as in Eq. (15) by replacing $Y_k(n)$ by

$$Y_k^{(\varepsilon,t)}(n) \overset{\Delta}{=} \frac{1}{\sqrt{N}} \sum_{m=0}^{N-1} y[k(N+D)+D+m+t]e^{-2i\pi\left(\frac{nm}{N}-\varepsilon\right)}. \tag{19}$$

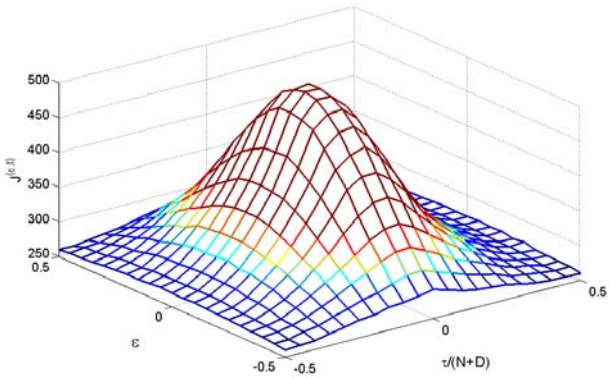



Fig. 4. Effect of synchronisation impairments on the cost function $J$ (SNR=0dB, $M$=24, $P_{fa} = 0.02$).

Detection is thus based on the knowledge of the pilot structure without the knowledge of pilot symbols so that the detection can be performed on every portion of the received signal. There is no need for frame or superframe synchronisation in contrast with coherent detection methods using known symbol sequences.

Note also that WiMAX pilot subcarriers present other features that can be exploited for

detection purposes. An example is given in (Socheleau et al., 2008) where a detector based on the third order statistics of the pseudo-random binary sequence $w_k$ is proposed.

### 4.4. Detection performance

To illustrate the performance of the detection algorithms, we here consider a Mobile WiMAX BS using 512-subcarriers per channel [2] and a DL-PUSC permutation mode. We recall that there are 60 pilot, 360 data, 91 guard and 1 DC subcarrier [13]. Unless otherwise stated, the cyclic prefix length $D$ is set to 64. The asymptotic false alarm probability $P_{fa}$ is set to $0.02$. The Signal-to-Noise Ratio (SNR) is defined as $SNR(dB) = 10log_{10}\left(E_s / \sigma_\eta^2\right)$. The propagation channel is a time-variant discrete-time channel $\{h_k(l)\}_{l=0,\cdots,L}$ with $L = D$ and an exponential decay profile for its non-null component (i.e., $\mathrm{E}[|h_k(l)|^2] = Ge^{-l/\beta}$ for $l = 0,\cdots,L$ and $G$ is chosen such that $\sum_{l=0}^{L} \mathrm{E}[|h_k(l)|^2] = 1$). The channel time variation is simulated using the Jakes model and the maximum Doppler frequency $f_d$ is set to 100Hz. For the simulations, uniformly distributed random $\varepsilon$ and $\tau$ were generated with $-0.5 \le \varepsilon \le 0.5$ and $-0.5(N+D) \le \tau \le 0.5(N+D)$.

Figure 5 shows the performance of a CFAR detector (see [10]) based on the cyclostationarity induced by the CP. Two observation windows as well as two CP lengths have been tested. The results indicate that only 24 OFDM symbols, which corresponds to half of a 5ms frame, are required to obtain excellent performance and this, even for very short CP (i.e. $N/D = 32$). In contrast with licensed user detection [30], the common framework in cognitive radio, WiMAX detection at negative SNR is not required here. Detectors must only guarantee good performance in SNR ranges where systems experience bit error rates low enough to operate. In our context, detection as such is not much of interest if the communication cannot be established afterwards.

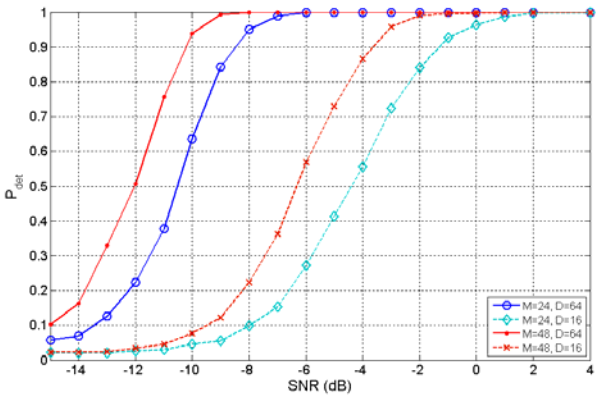



Fig. 5. Detection performance based on the cyclostationary feature induced by the CP ($\beta = 0.25D$).

[2]Note that it would nicely fit into a 6MHz TV channel [6]

The performance of the pilots structure detector is displayed on Figure 6. Similarly to the previous criterion, the detection rate is significantly improved as the number of available symbols increases and is still excellent for a short observation window.

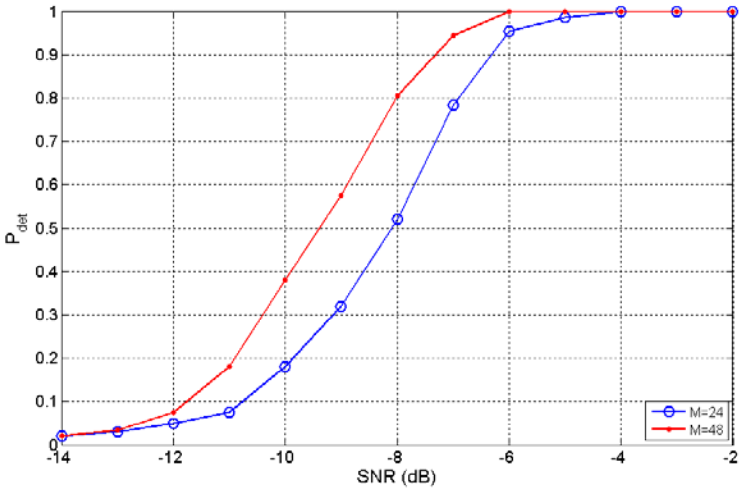



Fig. 6. Detection performance based on the pilot tones structure ( $\beta = 0.25D$ ).

## 5. SINR estimation

Upon detection of a wireless interface which it is compatible with, our terminal has then to estimate the link quality it can hope with this interface in order to decide if it meets its data rate and robustness requirements. SINR is a relevant indicator commonly used to evaluate this link quality. For the WiMAX interface, SINR is usually measured on preambles or specific broadcast DL zones (see Eq.(144) in (IEEE Std. 8022.16, 2005)). This kind of estimation method, based on the knowledge of pilot symbols, gives excellent performance since it is data-aided. However, for the same reasons as those discussed in subsection 4.1, this method can be time consuming as it requires to get (super)frame synchronised. An alternative approach, based on the correlation and the cyclostationarity induced by the OFDM CP, is presented in this section. This algorithm estimates the signal and the noise plus interference power independently and does not require the knowledge of pilots symbols so that it can be applied on any portion of WiMAX signals.

### 5.1. Noise plus interference power estimation

As in (Socheleau et al., 2008b), we here suggest to take advantage of OFDM signals particular structure to estimate the noise variance. More precisely, we show hereafter that the noise variance can be estimated thanks to the redundancy induced by the CP. In fact, the CP use leads to $x\left(k\left(N+D\right)+m\right)=x\left(k\left(N+D\right)+N+m\right)$, for any integer $k$ and any $m\in\left\{0,\cdots,D-1\right\}$. It is then straightforward to see that if we assume perfect synchronisation at reception (i.e $\tau=0$ and $\delta_f=0$) and a time-invariant channel over an OFDM symbol

duration, we can get $D-L$ noise plus interference power estimates defined as

$$\hat{\sigma}_u^2 = J(u),\ L \le u \le D-1 \tag{20}$$

with

$$J(u) = \frac{1}{2M(D-u)} \sum_{k=0}^{M-1} \sum_{m=u}^{D-1} \left| y\left(k\left(N+D\right)+m\right) - y\left(k\left(N+D\right)+N+m\right) \right|^2 \tag{21}$$

where $M$ denotes the number of OFDM symbols in the observation window. Note that non data-aided synchronisation can be done thanks to the detection algorithm of subsection 4.3 or to more general algorithms such as those discussed in (van de Beek et al., 1997; Park et al., 2001 ; Xiaoli et al., 2001).

It can be easily shown that the estimator with the smallest variance is found for $u = L$. The difficulty is then to estimate $L$. Cui et al. suggested an estimator in (Cui & Tellambura, 2006) but it has the major disadvantage of being based on a threshold level chosen arbitrarily. To overcome this limitation we hereafter propose a method inspired by maximum likelihood estimation.

From Eq. (21), $J(u)$ can be expressed as

$$J(u) = \left(1 - \frac{1}{D-u}\right) J(u+1) + \xi(u) \tag{22}$$

where $\xi(u)$ is a random variable given by

$$\xi(u) = \frac{1}{2M(D-u)} \sum_{k=0}^{M-1} \left| y\left(k\left(N+D\right)+u\right) - y\left(k\left(N+D\right)+N+u\right) \right|^2 . \tag{23}$$

For $L \le u \le D-1$ and $M$ large enough, $\xi(u)$ is Gaussian and verifies

$$\xi(u) \sim \mathcal{N}\left(\frac{\sigma^2}{D-u}, \frac{\sigma^4}{M(D-u)^2}\right). \tag{24}$$

$L$ is then estimated using the likelihood function $f\left(X_u \mid L=u\right)$ with $X_u$ the multivariate observation variables defined as $X_u = \left(\xi(u), \xi(u+1), \cdots, \xi(D-1)\right)$. The different $\xi(u)$ being independent, $\hat{L}$ is given by

$$\hat{L} = \underset{0 \le u \le D-1}{\operatorname{argmax}} \left[ \prod_{m=u}^{D-1} f\left(\xi(m) \mid L=u\right) \right]^{1/(D-u)}, \tag{25}$$

where $f\left(\xi(m) \mid L=u\right)$ is computed thanks to Eq. (24) by making the approximation that $\sigma^2 = J(u)$. Note that because the observations $X_u$ are of variable lengths, Eq. (25) is defined as an average likelihood which is the geometric mean of the individual likelihood elements.

### 5.2. Signal power estimation

Thanks to equations (7) and (9) it is straightforward to show that the signal power can be estimated as

$$\hat{S} = \frac{1}{2N_c+1}\left|\sum_{q=-N_c}^{N_c} \hat{R}_y^{q\alpha_0}(N)\frac{\sin(\pi q\alpha_0)}{\alpha_0\sin(\pi q\alpha_0 D)}e^{i\pi q\alpha_0(D-1)}\right| \tag{26}$$

where $\alpha_0 = 1/(N+D)$ and $N_c$ represents the number of considered cycle frequencies to estimate the signal power. We assume that the interference does not show the same cyclostationary properties as WiMAX. The choice of $N_c$ is a trade-off between the estimator bias and variance. From Eq. (26), it can be shown that the estimator asymptotic variance (i.e. for $M \to +\infty$) decreases as the number of cycle frequencies increases. However, Eq. (7) indicates that it may be judicious to choose cycle frequencies within the first lob of $\hat{R}_y^{q\alpha_0}(N)$ as for $q \geq N/D$, the power of this function is very small. In addition, to limit the estimator bias, $N_c$ has to be bounded by the channel coherence bandwidth. In fact, from Eq. (8) and thanks to Parseval's identity, $R_y^{q\alpha_0}(N)$ can be expressed as

$$R_y^{q\alpha_0}(N) = R_x^{q\alpha_0}(N)\int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{E}\left[H(\nu)H^*(\nu - q\alpha_0)\right]d\nu \tag{27}$$

where $H(\nu) = \sum_{l=0}^{L} h(l)e^{-2i\pi l\nu}$. Thus, from the definition of S (see Eq. (4)), the cycle frequencies used to estimate the signal power have to be limited to the case where $R_y^{q\alpha_0}(N) = R_x^{q\alpha_0}(N)\sum_{l=0}^{L}|h(l)|^2$ which is equivalent to choose $q\alpha_0$ within the coherence bandwidth where $\mathbf{E}\left[H(\nu)H^*(\nu - q\alpha_0)\right] \simeq \mathbf{E}\left[|H(\nu)|^2\right]$. The coherence bandwidth $B_c$ is usually defined thanks to the channel root mean square delay spread [25]. In our case, as the channel impulse response is unknown at reception, $B_c$ is approximated as $\hat{B}_c = 1/(\rho\hat{L})$ where $\rho$ is a coefficient expressing the desired correlation rate within $B_c$. Consequently, we choose $N_c = \min\left(\frac{N+D}{\rho\hat{L}}, \frac{N}{2D}\right)$. As shown in (Socheleau et al., 2008b), $\rho$'s choice has only a very little influence on the estimator performance.

### 5.3 SINR estimation performance

Using the same simulation context as in subsection 4.4, we plot on Figure 7 the Normalised Mean Square Error (NMSE) of the SINR estimation versus the true SINR for different $M$. $\text{NMSE} = \mathbf{E}\left[\left(\hat{S}/\hat{\sigma}^2 - S/\sigma^2\right)^2 \frac{\sigma^4}{S^2}\right]$, $D = 64$, $\beta = 16$ and $\rho$ is set to 5. As expected, the performance is significantly improved as the number of available OFDM symbols increases

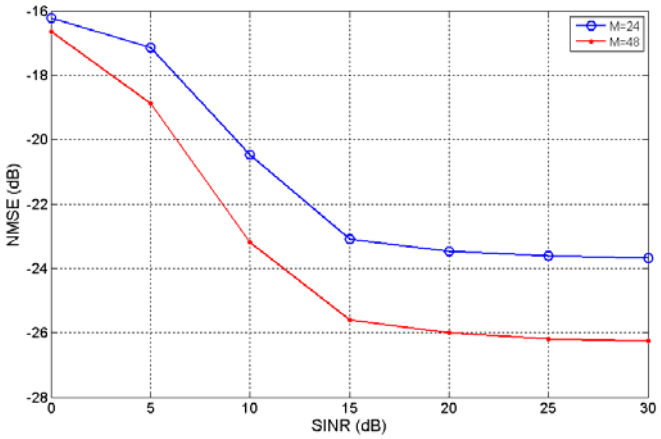



Fig. 7. NMSE of the SINR estimator.

As detailed in (Dai et al., 2008), the measured SINR of a WiMAX DL signal can be transformed into a data rate. WiMAX supports a large number of modulation and forward error correction coding schemes and allows the scheme to be changed based on the channel conditions. This is what is called adaptive modulation and coding (AMC). The objective of AMC is to maximise throughput in a time-varying channel. Since the adaptation algorithm typically calls for the use of the highest modulation and coding scheme that can be supported for the current SINR, it is possible to know the used data rate. For WiMAX, there is a modulation and coding scheme defined per SINR fluctuation of 2dB. Consequently, in Figure 8 we plot the probability of estimating the SINR within the range of +/-1 dB of the true value. It clearly indicates that our SINR estimator gives a reliable measure that can be used for vertical handoff decision. Note that this probability becomes greater than 97% for $M = 24$ and a SINR $\geq 0$ dB if the tolerated range is increased to +/-1.5 dB.

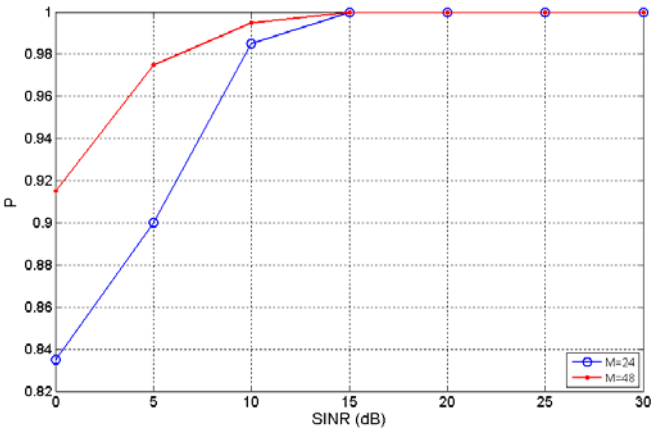



Fig. 8. Probability of estimating the SINR within +/-1 dB of the true value.

## 6. Slot activity rate estimation

In the context of signal metrics estimation for vertical handoff, the SINR knowledge of OFDMA signals is not fully informative without the knowledge of the time-frequency slots activity rate (SAR). This rate is defined as the probability $\mathrm{P}(\{\varepsilon_k(n)=1\})$ where $\varepsilon_k(n)$ is defined in Eq. (1). There can be scenarios (low network load, segmentation etc.) where a few slots are active within an OFDMA frame. In this case, even if each active slot is very powerful (which indicates a good quality of communication link), the SINR as defined in Eq. (3) and estimated in section 5 can be low. This finds an explanation in the SINR computation that is averaged over all (active or not) slots. Moreover, the main question a multi-interface terminal is trying to answer when measuring signal metrics is: what is amount of resources that will be allocated to me if I get connected to the wireless interface I am currently sensing? The SAR can be useful to answer this question since it indicates the part of the network resources that is already occupied by other users. However, for WiMAX networks, SAR has to be carefully interpreted as it is just an indicator of the traffic load and is mainly relevant when this load is low. In fact, when the SAR is low, it indicates that the sensed cell has not reached its maximum of capacity so that resources are available for new connections. But if the SAR is high (equal to 1 for instance), one can wrongly infer that no new connection will be allowed by the network. Since the 802.16 standard does not specify the resources allocation methods, there can be strategies that aim to maximise the use of physical resources whatever the number of open connections. This can lead to situations where even if the current SAR is equal to one, new connections will be allowed and resources dynamically redistributed among the set of users.
In the WiMAX system, the resource allocation is specified in the UL/DL MAP messages broadcasted at the beginning of each frame. These messages gives the number of slots allocated to each user in uplink and downlink respectively. By decoding the UL/DL MAP the SAR can be directly deduced. This operation, requiring frame synchronisation, can be once again time consuming for a multi-interface terminal. In addition, it is also power consuming since demodulation and decoding are required. In this section, an alternative approach is derived based on the estimation of all $\varepsilon_k(n)$ on the observation window.

### 6.1. Algorithm

SAR estimation is equivalent to differentiate active slots from inactive slots. Intuitively, by considering that noise plus interference power $\sigma^2$ is known, common detector structures can be used. For instance,

$$\widehat{\mathrm{P}}(\{\varepsilon_k(n)=1\}) = \frac{\sum_{k,n} I\left(|Y_k(n)|^2 > g(\sigma^2)\right)}{MN} \tag{28}$$

where $I(A)$ denotes the indicator function of a given event $A$, $g(\sigma^2)$ is a thresholding function like $g(\sigma^2)=\lambda\sigma^2$ for instance and $Y_k(n)$ is defined by Eq. (17). We here assume perfect synchronisation at reception.
The problem of estimators as the one defined in Eq. (28) is that the choice of the detection threshold $\lambda$ has a strong influence on the performance. To overcome this constraint, we

suggest an alternative method sequenced in two stages:

1. Sorting of the observed symbols $Y_k(n)$ based on a likelihood criterion. From the knowledge of the noise plus interference probability density function (pdf), the idea is to sort the $Y_k(n)$ according to their probability of being made of noise plus interference only.
2. Cost function minimisation. Once the $Y_k(n)$ are sorted, a breakdown point is sought in the ordered set in order to separate symbols of signal plus noise and interference from symbols of noise plus interference only.

Let $\Omega$ be the set of observed symbols defined as

$$\Omega = \bigcup_{k,n} Y_k(n),\ k \in \{0,1,\cdots,M-1\} \text{ and } n \in \{0,1,\cdots,N-1\}. \tag{29}$$

We define the relation of order $\mathbf{R}$ as

$$(x,y) \in \Omega^2 \mid f(x) \le f(y) \tag{30}$$

where $f(x)$ is the pdf of the noise plus interference. For the sake of simplicity, we assume that this pdf is Gaussian so that

$$f(x) = \frac{1}{\pi\sigma^2} e^{-|x|^2/\sigma^2}. \tag{31}$$

Note that this may not always be true, especially in the case of strong non Gaussian interference. In this particular case, $f(x)$ can be estimated thanks to the samples used to compute $\sigma^2$ in subsection 5.1.

$(\Omega, \mathbf{R})$ is then the ordered set of the $Y_k(n)$ sorted out by their crescent probability of being symbols of noise plus interference only. $(\Omega, \mathbf{R})$ is equivalent to sort out the $Y_k(n)$ by their decreasing energy. The elements that composed this set are written as

$$(\Omega, \mathbf{R}) = \{\overline{Y}_0, \overline{Y}_1, \cdots, \overline{Y}_{MN-1}\}. \tag{32}$$

Once the symbols are ordered, we suggest to work on the subset $(\Omega_u, \mathbf{R}) = \{\overline{Y}_u, \overline{Y}_{u+1}, \cdots, \overline{Y}_{MN-1}\}$ and to detect the first $u$ for which $(\Omega_u, \mathbf{R})$ is made of noise plus interference only. The approach consists for each $u \in \{0,1,\cdots,MN-1\}$, to estimate a parameter $\hat{\theta}_u$ of the pdf $f(x)$ from the elements of $(\Omega_u, \mathbf{R})$. Once these $\hat{\theta}_u$ are estimated, the breakdown point $p_r$ is expressed as

$$p_r = \arg\min_u (\hat{\theta}_u - \theta)^2. \tag{33}$$

The $\overline{Y}_0, \overline{Y}_1, \cdots, \overline{Y}_{p_r-1}$ are then considered as symbols made of signal plus noise and interference and the $\overline{Y}_{p_r}, \overline{Y}_{p_r+1}, \cdots, \overline{Y}_{MN-1}$ as symbols of noise plus interference only. It now remains to

choose the parameter $\theta$ as well as the estimation method associated to it. The $Y_k(n)$ being usually centred whatever the value of $\varepsilon_k(n)$, we suggest to choose $\theta = \sigma$. As for the estimation method, it will depend on two criteria:

1. The estimator efficiency. If $T$ is the estimator of $\sigma$, is efficiency is defined as

$$e(T) = \frac{1/\Im(\sigma)}{Var[T]} \tag{34}$$

where $\Im(\sigma)$ is the Fisher information. We seek to have the estimator with the highest efficiency. If $(\Omega_u, \mathbf{R})$ is made of noise plus interference only, then the estimation squared error $(\hat{\theta}_u - \theta)^2$ has to be as small as possible.

2. Robustness. Robustness translates the estimator resistance to outliers. It is measured by the proportion of incorrect observations (arbitrarily large) an estimator can accept before returning results that are also arbitrarily large. In contrast to what is usually quested, we here focus on estimators with the lowest possible robustness. If $(\Omega_u, \mathbf{R})$ includes at least a symbol of signal plus noise and interference, ideally we want to have $(\hat{\theta}_u - \theta)^2$ as large as possible.

The maximum likelihood estimator expressed as $\bar{\sigma}_u = \sqrt{\frac{1}{MN-u}\sum_{\ell=u}^{MN-1} |\bar{Y}_\ell|^2}$ is the optimal estimator according to the wanted efficiency and robustness criteria (efficiency=1 and robustness=0). Consequently, the breakdown point is expressed as

$$p_r = \underset{u}{\operatorname{argmin}} \left( \sqrt{\frac{1}{MN-u} \sum_{\ell=u}^{MN-1} |\bar{Y}_\ell|^2} - \sigma \right)^2 \tag{35}$$

and the slot activity rate is given by

$$\hat{\mathrm{P}}(\{\varepsilon_k(n) = 1\}) = \frac{p_r}{MN}. \tag{36}$$

### 6.2. Slot activity rate estimation performance

With a simulation context similar to the one depicted in the previous section, Figure 9 shows the NMSE of the SAR estimator for several SNR. To find the breakdown point of Eq. (35), we used either the perfect knowledge of $\sigma$ or its estimate presented in section 5. The results indicates that the proposed method is robust to estimation errors of $\sigma$ and not much dependent on the true SAR but very sensitive to the SINR.

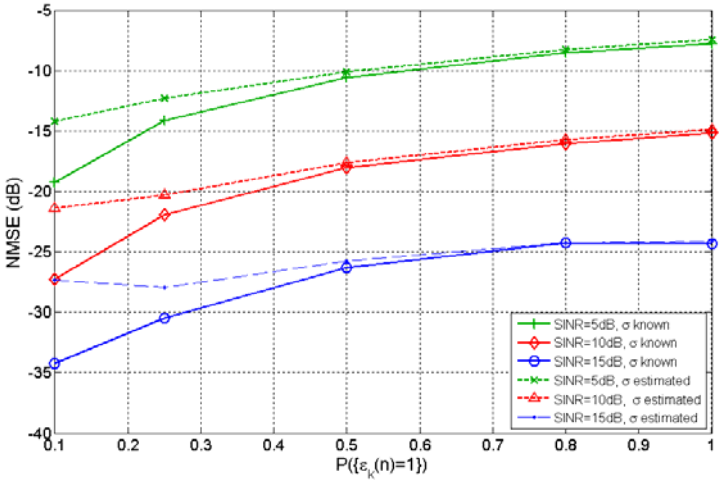



Fig. 9. Normalised mean square error of the slot activity rate estimator for various SNR.

Figure 10 compares the performance of our algorithm with the common constant false alarm rate (CFAR) detection method for a SINR of 10dB. the CFAR detector assume that $\varepsilon_k(n)=1$ when $P_{fa} > e^{-\frac{|Y_k(n)|^2}{\sigma^2}}$. The results highlight the limitations of detectors based on a threshold. It can be observed that the choice of the threshold has a strong influence on the performance in contrast to our method that offers better controlled results.

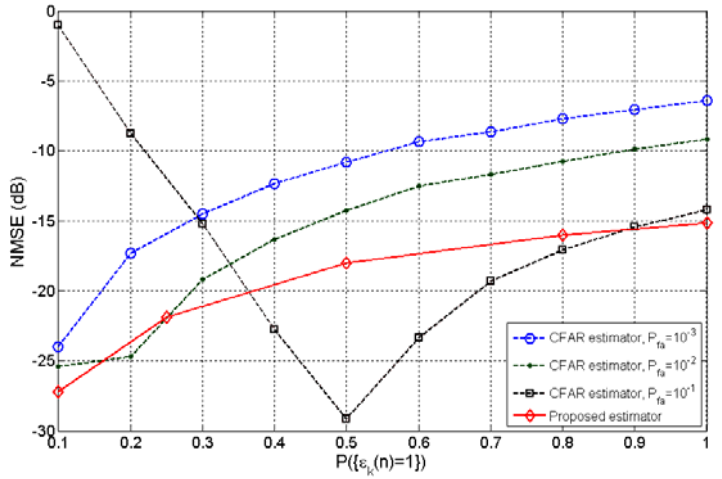



Fig. 10. Comparison of the proposed method with the CFAR detector (SNR=10dB, $\sigma$ known).

## 7. Conclusion

The main conclusion of this chapter is that WiMAX signal metrics measurement for vertical handoff decision is possible in a cognitive context without requiring any modifications of the current standard. The data-aided metrics measurement based on preamble detection or frame header decoding have been constrasted to non data-aided approaches that mainly rely on OFDM(A) specific features. We have shown that non data-aided DL subframe detection, SINR estimation and slot activity rate estimation perform well enough to be relevant inputs for algorithms that have to decide whether to trigger a vertical handoff from any system to WiMAX. In addition, the proposed metrics estimators suit well cognitive radio scenarios since they require only small portions of signal and therefore allow a fast decision making process on the WiMAX signal quality and traffic load.

## 8. References

Akyildiz, I. F. ; Lee, W-Y. ; Vuran, M. C. & Mohanty, S. (2006). NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer Networks*, pp. 2127-2159.

van de Beek, J. ; Sandell, M. & Borjesson, P.O. (1997). ML estimation of time and frequency offset in OFDM systems. IEEE Trans. on Signal Processing, Vol. 49, pp. 1800-1805.

Blaschke, V. ; Kloeck, C.; Weiss, J.; Renk, T. & Jondral, F.K. (2008). Opportunistic WiMAX-GSM coexistence. Communications, IET, Vol. 2, No. 6.

Bouzegzi, A.; Jallon, P. & Ciblat, P. (2008). A Second Order Statistics Based Algorithm for Blind Recognition of OFDM Based Systems. IEEE Globecom Conf.

Cordeiro, C.; Challapali, K.; Birru, D. & Shankar, S. (2006). IEEE 802.22: An Introduction to the First Wireless Standard based on Cognitive Radios. Journal of Communications, Vol. 1, No. 1, pp. 38-47.

Cui, T. & Tellambura, C. (2006). Power delay profile and noise variance estimation for OFDM. IEEE Communications Letters, Vol. 10 No. 1.

Dai, Z; Fracchia, R.; Gosteau, J.; Pellati, P. & Vivier, G. (2008). Vertical Handover Criteria and Algorithm in IEEE802.11 and 802.16 Hybrid Networks. IEEE Int. Conf. On Communications.

Gardner, W. A. ; Napolitano, A. & Paurac, L. (2006). Cyclostationarity: Half a century of research. Signal processing, Vol. 86, No. 4, pp. 639-697.

Haddad, M.; Hayar, A.M. & Debbah, M. (2007). Spectral Efficiency of Cognitive Radio Systems. IEEE GLOBECOM.

Haykin, S. Cognitive radio: brain-empowered wireless communications (2005). IEEE Journal Sel. Areas in Comm., Vol. 23, No. 2, pp. 201-220.

IEEE Dynamic Spectrum Access Networks Conference (DySPAN). http://www.ieee-dyspan.org/

IEEE Std. 802.16 (2005). Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operations in License Bands and Corrigendum 1.

Ishii, H. & Wornell (2005), G. W. OFDM blind parameter identification in cognitive radios. IEEE Conf. on Personal, Indoor and Mobile Radio Communications, pp. 700-705.

Jallon, P. (2008). An algorithm for detection of DVB-T signals based on their second order statistics. EURASIP Journal on Wireless Communications and Networking, article ID 538236.

Leu, A.E.; Mark, B.L. & McHenry, M. (2009). A Framework for Cognitive WiMAX with Frequency Agility. Proceedings of IEEE, Special Issue on Cognitive Radio, Vol. 97, No. 4.

Li, H.; Bar-Ness, Y.; Abdi, A.; Somekh, O. & Su, W. (2006). OFDM modulation classification and parameters extraction. IEEE Conf. on Cognitive Radio Oriented Wireless Networks and Communications, pp. 1-6.

McNair, J. & Zhu, F. (2004). Vertical handoffs in fourth-generation multinetwork environments. IEEE Wireless Communications, Vol. 11, No. 3.

Oner, M. & Jondral, F. (2007). On the Extraction of the Channel Allocation Information in Spectrum Pooling Systems. IEEE Journal Sel. Areas in Comm., Vol. 25, No. 3.

Park, B.; Ko, E.; Cheon, H.; Kang, C. & Hong, D. (2001). A blind OFDM synchronization algorithm based on cyclic correlation. IEEE Globecom Conf., pp. 3116-3119.

Rappaport, T.S. (2002). Wireless Communications: Principles and Practices . Prentice Hall.

Socheleau, F.-X.; Houcke, S.; Aïssa-El-Bey, A. & Ciblat, P. (2008a). OFDM system identification based on m-sequence signatures in cognitive radio context. IEEE PIMRC.

Socheleau, F-X; Aïssa-El-Bey, A. & Houcke, S. (2008b). Non Data-Aided SNR Estimation of OFDM Signals. IEEE Communications Letters, Vol. 12, No. 11.

Socheleau, F.-X.; Ciblat, P. & Houcke, S. (2009). OFDM System Identification for Cognitive Radio Based on Pilot Induced Cyclostationarity. IEEE WCNC.

Stevenson, C.; Chouinard, G.; Zhongding L.; Wendong, H.; Shellhammer, S. & Caldwell, W. (2009). IEEE 802.22: The First Cognitive Radio Wireless Regional Area Network Standard. IEEE Communications Magazine, Vol. 47, No. 1.

U.S. FCC (2002). Review of Spectrum Management Practices.

U.S. FCC (2008). Second Report and Order and Memorandum Opinion and Order, in the Matter of Unlicensed Operation in the TV Broadcast Bands Additional Spectrum for Unlicensed Devices Below 900 MHz and in the 3 GHz Band.

WiMAX Forum (2008). Spectrum Opportunities below 1GHz.

Wireless Innovation. http://www.wirelessinnovation.org/

Xiaoli, M.; Giannakis, G.B. & Barbarossa, S. (2001). Non-data-aided frequency-offset and channel estimation in OFDM and related block transmissions. IEEE ICC.

Yucek, T. & Arslan, H. (2007). OFDM Signal Identification and Transmission Parameter Estimation for Cognitive Radio Applications. IEEE Glob. Telecom., pp. 4056-4060.

Zeng, Y.; Liang, Y-C. & Zhang, R. (2008). Blindly Combined Energy Detection for Spectrum Sensing in Cognitive Radio. IEEE Signal Processing Letters, Vol. 15, pp. 649-652.

# 21

# Multi-hop Relay Networks

David Soldani
*Huawei Technologies, European Research Centre*
*Germany*

## 1. Introduction

The current outlook for the overall wireless broadband market worldwide shows that mobile data traffic will double every year through 2013, and mobile broadband handsets with higher than 3G speeds and laptops with air cards will drive over 80% of global mobile traffic by 2013. About 1.2 billions of new users are expected on top of the current 200 millions. The expansion of wireless broadband ubiquity will allow the convergence of the "3 Screens" (TV, PC & mobile handset) with exciting opportunities for always-on networking and an increased volume of consumers who will access and rely on mobile networks, thus creating a need for greater economies of scale and lower cost per-bit.
Today cost of 2G/3G network transmission accounts for approximately 19% of the overall cost for delivering a bit of data to an end user. Analyses of Tier 1 European mobile carriers suggest that, given the current backhaul architectures, which are based primarily on leased lines, operational expenditure (OPEX) for transmission would increase to around 80% of the spending per cell site, if the cell exploited its maximum capacity. Therefore, it is not feasible to offer mass market mobile data services using the existing backhaul networks, mainly based on leased lines.
The broadband radio interfaces for next generation mobile networks (NGMN), such as 3G Long Term Evolution (LTE) and mobile WiMAX 802.16e and beyond, will be characterized by a very limited range due to their high operating frequencies, which are expected to be in most of the cases above the 2 GHz band currently used for the 3G systems. For example, WiMAX 802.16e is expected to be the Time Division Duplex (TDD) technology for 2.5 GHz and 3.5 GHz bands (Andrews et al., 2007). In a suburban environment, a wireless solution for good quality indoor coverage would take nearly four times as many sites to deploy at 2 GHz than at 1 GHz, and ten times as many for 3.5 GHz. On the other hand, at a given power and carrier frequency, the available data rate decreases if the distance from the access node (AN) increases and Quality of Experience (QoE) must not depend on where the terminal is located in a cell. Hence, the deployment of a conventional cellular network would require a very high density of access nodes to achieve a satisfactory radio coverage, capacity and QoE, resulting in a very high costly solution per delivered bit.
Consequently, to meet the goal of low cost radio network deployment, for both, short-range and wide-area coverage, the deployment concept based on layer 2 (multi-hop) relay nodes is currently one of the most promising solutions (Soldani & Dixit, 2008).

Wireless relays help overcome the current dependencies on wired backbones and enable cost-effective enhancement of coverage, throughput and system capacity. In fact, relay nodes do not need a wired (e.g. copper or fibre) backbone access, reducing deployment costs, and offer high flexibility in placing the stations, allowing fast network rollout and adaptive traffic capacity engineering. Further, multi-hop relaying gives better trunking efficiency at aggregation points and site acquisition and antenna structures are much less expensive. A relay augmented network is thus expected to provide an improved Return on Assets (RoA), which means higher average revenue per user (ARPU) with superior grade of service at lower overall incremental cost. Such networks are well suited for deployments in emergency and disaster scenarios. Also, in rural areas, where the traffic density is low and population sparsely distributed, it may not be economically viable to build traditional cellular access networks with full fledged base stations; rather a network architecture with a single base station (BS) flanked by relay nodes to improve capacity and range extension may be a more economical and flexible approach.
This chapter presents an overview of the concept of relaying and specifically targets a system that utilizes the WiMAX technology and related standards.

### 1.1 Concept of Relaying

The classical *three-node relaying* model is a network consisting of only one *Source* (mobile multi-hop relay-base station, MMR-BS or MR-BS), one *Relay* (fixed, nomadic, or mobile relay station, F/N/M-RS) and one *Destination* (mobile station, MS, or user terminal, UT). (In the rest of the chapter, the nomenclature "UT" is used for both fixed and mobile station.) The model was originally introduced by Van der Meulen (1971).
The three-node relay network can be seen as a primitive building block for larger relaying systems. As illustrated in Figure 1, the destination can receive the signal from the source via two paths: the two-hop relaying link (first and second hop) and the one-hop *direct link* (third hop). In the two-hop relaying link, the physical channel between the source and the relay is called *relay link* and the physical channel between the relay and the destination is called *access link*.
A *relay path* is by definition a concatenation of consecutive relay links between the source and the designated access relay station.

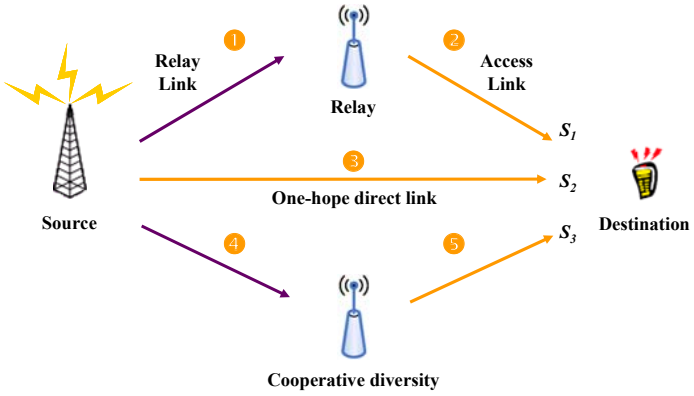



Fig. 1. Three node relay network and cooperative relaying (Soldani & Dixit, 2008)
Reproduced by permission of © IEEE 2008

The relay normally works in half-duplex mode, which means that the station does not receive and transmit using the same channel simultaneously. This is because it would be difficult to separate the received signal from the transmitted one. For this reason, the channel allocated for the relay operation consists of two (ideally) orthogonal sub-channels. For example, in TDD, the orthogonal sub-channels are two consecutive time slots: the relay receives during the first time slot and uses the second slot for retransmitting. Frequency or code division duplex is also possible (Pabst et al., 2004).

The relay may operate in three possible schemes, depending on how the received signal is processed: *Amplifying and Forwarding* (AF), *Decoding and Forwarding* (DF) or *Estimating and Forwarding* (EF).

In the AF scheme (*analogue repeaters*), the relay node just amplifies and retransmits the input symbols. The received signal is deteriorated by relay link fading and additive receiver noise. The degraded signal and noise are amplified and forwarded, thereby increasing the system noise level. This is demonstrated in Figure 2, where $S_S$ is the Source transmit power, $S_R$ is the Relay transmit power, $L_{S\text{-}R}$ the Source-Relay path loss (includes antenna gains), $L_{R\text{-}D}$ is the Relay-Destination path loss (includes antenna gains), $L_{S\text{-}D}$ is the Source-Destination path loss (includes antenna gains), $G_R$ is the Relay gain, $G_D$ is the Destination gain, $B_N$ is the equivalent Noise bandwidth, $T_X$ is the effective noise temperature referred to the input, $N_{AX}$ is the Antenna noise temperature (includes interference), $k$ is the Boltzmann constant (1.3806 x $10^{-23}$ J/K), $F_0$ is the noise figure (Noise Factor), i.e. $F_0 = (1 + T_X/T_A)$, when $T_A = T_0 = 290$ K (standard room temperature), and $\alpha$ is the Source-Relay transmit power ration, i.e. $\alpha = S_S / S_R$. From the basic theoretical analysis showed in the figure results that multi-hop techniques enable a link budget gain, which can provide an improvement in coverage, throughput and reduction in transmission power.

In the DF scheme (*digital repeaters*, or *layer 2 relays*), the relay demodulates and decodes the received signal before the retransmission. In this case, the forwarded signal does not contain additional degradation, but only symbol errors resulting from it.

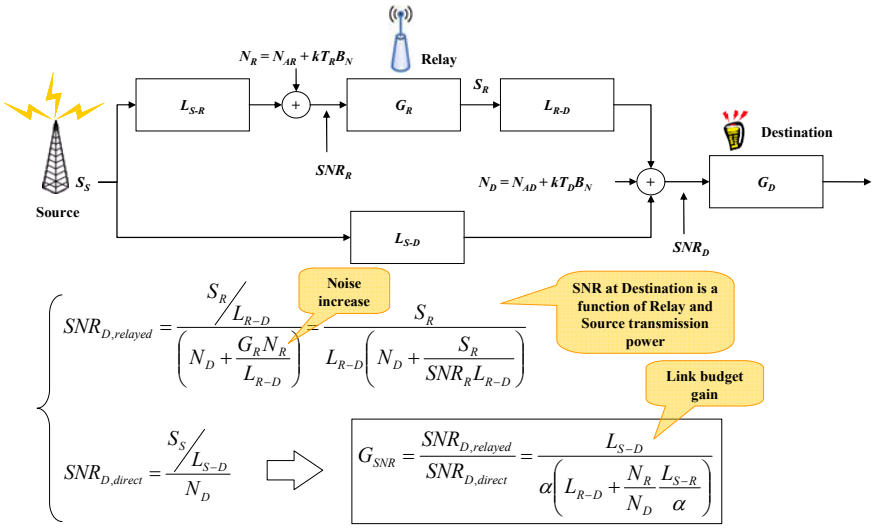



Fig. 2. Impact of relaying on received SNR for an Amplifying & Forwarding (AF) relay

In the EF scheme (*hybrid relays*), the relay does not decode the input data, but it source codes the received signal (after quantization) and transmits it to the destination. In this case, the forwarded signal contains possible estimation errors. The destination uses the relay estimation as side information while coding the actual direct link signal.
The concept of *cooperative relaying* is based on the fact that a signal, once transmitted, can be received (and usefully forwarded) by multiple terminals (Pabst et al., 2004). In general, cooperative relaying systems have a source node multicasting a message to a number of cooperative (helping) relays, for example, two in Figure 1, which in turn resend a processed version to the intended destination node. The destination node combines and exploits the inherent diversity of the signals received from the relays and, possibly, the source signal, as illustrated in Figure 1.
The coverage area of an MMR-BS cell and all its subordinate RS cells constitutes a *multi-hop relay cell* (MMR-Cell). The MMR-BS manages all resources within an MMR-cell through a *centralized* or *decentralized* control function. Resource management and control of UTs within an MMR-cell may be via direct radio links or via relayed messages.

### 1.2 The State of the Art

The DF concept has been studied widely and many results are available in the literature. The most relevant research contributions are from the Wireless World Initiative New Radio (WINNER) project, and its continuation (WINNER+), which is an EU-funded consortium with strong participation from the academia and the industry, working towards enhancing the performance of mobile communication systems (WINNER, 2008).
Wireless relays are also being specified in the standards. For instance, at the time of writing, relaying is studied as an enhancement of 3GPP LTE and planned to be standardized in LTE-Advanced by the end of 2010, with mesh extensions in later releases. The IEEE 802.16j amendment (IEEE 802.16j, 2009) provides specifications for mobile multi-hop relay features, functions, and interoperable relay stations, to enhance coverage, throughput and system capacity of WiMAX networks. The document specifies orthogonal frequency division multiple access physical layer (OFDMA-PHY) and medium access control protocol (MAC) enhancements to the IEEE 802.16e to enable the operation of relay stations for licensed bands. The standard does not require any modification to the mobile stations (including subscriber stations, SS) and provides full backward compatibility with 802.16-2004 (fixed) and 802.16e-2005 (mobile) WiMAX systems. Mesh extensions are specification items for IEEE 802.16m system (IEEE 802.16m, 2008). A key differentiator between 802.16m and 802.16j is that the former is not constrained by any legacy issue and hence is at liberty to design an entirely new radio access system within the ITU IMT-Advanced initiative.

## 2. Multi-hop Relaying for WiMAX Networks

This section provides a comprehensive description of the IEE 802.16j relays-based wireless access networks; it covers the different usage scenarios, network topologies, frame structure, radio interface protocols, relay path management and routing, radio resource management (RRM) functions, quality of service (QoS), and cooperative relaying. More information on the radio interface protocols and MMR frame structure can be found in the Multi-hop Relay Specification (IEEE 802.16j, 2009) and in (Chen & De Marca, 2008).

### 2.1 The 802.16j Specifications

In 802.16j, the RS is supposed to work in line of sight (LOS) and non-LOS propagation conditions and to support UT handover. The relay link can operate either in TDD or frequency division duplex (FDD) mode. The MMR-BS supports bandwidth requests and allocation mechanisms for RS and maintains full compliance with legacy standard. QoS for multi-hop is supported as defined for legacy 802.16 systems and delivery of unicast data, via RS and UT connections, is based on connection identifiers (CIDs). Security procedures are compulsory between MMR-BS and RS, between relay stations, and between RS and UT.
Both *in-band* and *out-of-band relaying* are possible. In *in-band relaying*, MMR uses the same RF channels on relay (MMR-BS-to-RS or RS-to-RS) and access links (MMR-BS-to-UT or RS-to-UT), and in *out-of-band relaying*, the MMR uses different RF channels on relay (i.e., MMR-BS-to-RS or RS-to-RS) and access links (i.e., MMR-BS-to-UT or RS-to-UT).
Other relevant functionalities, such as automatic repeat-request (ARQ) of UT via RS, RS with mobility and its subordinate UTs, multiple antennas (MIMO, beam forming, transmit diversity, etc.), RS to support feedback mechanisms and to participate in cooperative relaying, mechanism to support various forms of radio resource assignment, RS to process and forward the DL and UL control information, mechanisms for both MMR-BS and RSs to learn the topology of the MMR-cell in which they operate, are optional requirements (IEEE 802.16j, 2009).
The next sections give more insights into usage scenarios and basic functionalities of mobile multi-hop relay (MMR) networks.

### 2.2 Use Cases

The most important use cases for MMR are illustrated in Figure 3. The RSs can be owned by either the infrastructure provider or by the customer, depending on the usage scenario.
In the infrastructure with *fixed relay stations* (F-RSs), RSs and MMR-BSs are deployed to improve coverage, capacity, or user bit rate, in areas not sufficiently covered (e.g., indoor, in shadow, tunnels or underground), or to provide access for clusters of users outside the coverage area of the MMR-BS.
A *nomadic relay stations* (N-RS) can be temporarily deployed to provide additional coverage or capacity in an area where MMR-BSs and/or fixed RSs do not provide good coverage or capacity. For instance, temporary coverage could be required in emergency or disaster recovery situations, or in sporting occasions or fairs, where coverage is required only for the duration of that particular event. Furthermore, nomadic (portable) relay stations can provide access to subscribers within a room or to a large building, such as a multi-tenant dwelling or office building.
A *mobile relay station* (M-RS) can be mounted on a vehicle and connected to an MMR-BS or RS via a mobile link. In this case, the RS provides a fixed access link to terminals residing on the platform.
Enhancements in throughput, capacity and reliability can be achieved by providing higher signal interference noise ratio (SINR) to users at the edge of the cell, so that higher bit rates can be used, or by deploying RSs and MMR-BSs in a dense topology (small cells). The latter scenario may also provide a higher degree of routing diversity. For example, traffic from a given UT can be routed to avoid a congested relay link or to save power at a particular node.
For coverage and/or range extension the RSs can be deployed to provide higher signal level to users in low lying or isolated areas, beyond the MMR-BS coverage. Coverage can also be extended to UTs riding on mobile vehicles and to areas frequently travelled.

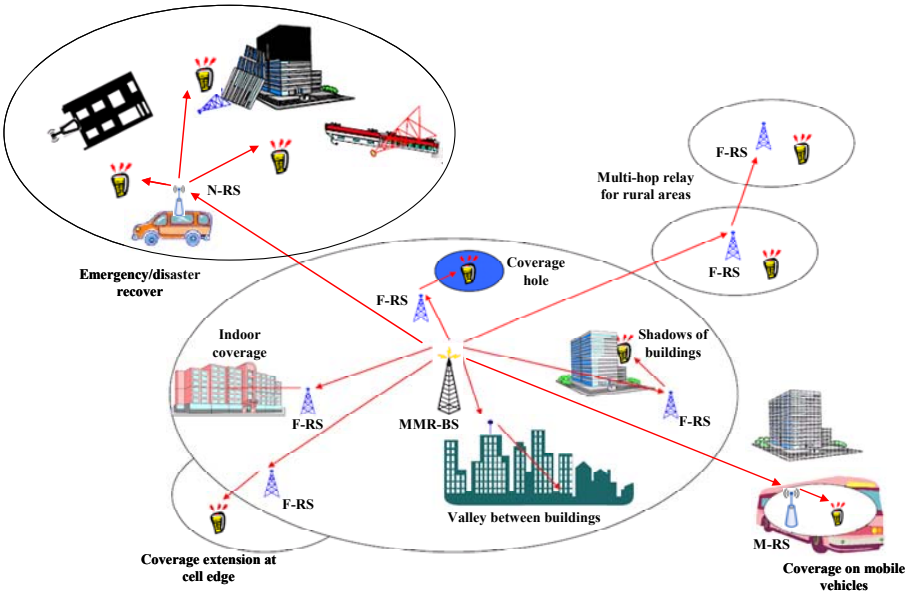



Fig. 3. Examples of usage scenarios for fixed, nomadic and mobile relay stations (Soldani & Dixit, 2008) Reproduced by permission of © IEEE 2008

### 2.3 Network Topology

The topology follows a *tree structure*, point to multipoint (PMP) compatible, neither ad hoc nor mesh. The interconnections between nodes include the following types of links: MMR-BS to UT, where the MMR-BS can connect to multiple UTs; RS to UT, where RSs can associate with multiple UTs; MMR-BS to RS, where the MMR-BS can connect to multiple RSs; and RS to RS, where RSs can associate among themselves.

For instance, in all usage models described in Section 2.2, all data transmission may occur between MMR-BS and mobile stations directly, or through one or more relay stations.

Downlink (DL) and uplink (UL) connectivity between the MMR-BS and UT can be provided via either symmetric or asymmetric routes. Multiple routes between one MMR-BS and one UT are also possible, e.g., as shown in Figure 4, where there are two routes (one through RS1 and the other through RS3) between the MMR-BS and UT5, which is served by RS4.

### 2.4 Radio Interface Protocols

An example of MMR data protocol stack for simple relay stations is depicted in Figure 4.

*MAC-Convergence Sublayer* (MAC-CS) constitutes the interfaces to upper layers (e.g. IP, Ethernet). This protocol is also responsible for: (i) classifying and mapping the MAC-Service Data Units (MSDUs) into appropriate CIDs for QoS management, routing and forwarding; (ii) processing (if required) the higher-layer PDUs based on their classification; and (iii) delivering CS PDUs to the appropriate MAC SAP and receiving the CS PDUs from the peer entity. An optional function of the CS is payload header suppression (PHS), the process of suppressing repetitive parts of payload headers at the sender and restoring these headers at the receiver.

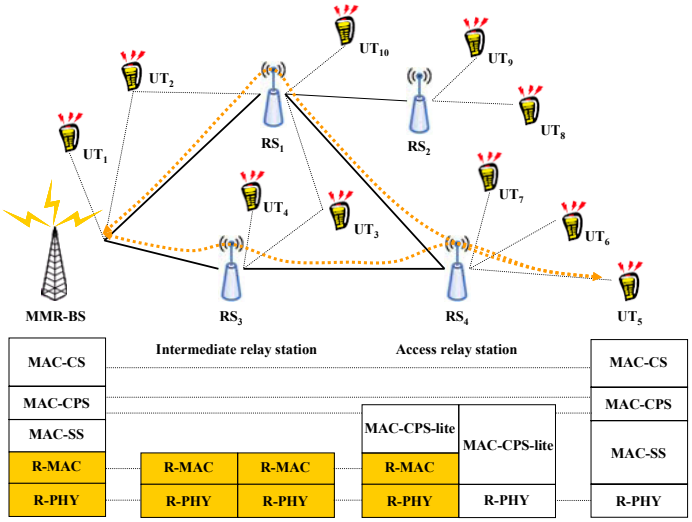



Fig. 4. Example of network topology and MMR data protocol stack for simple relay stations (Soldani & Dixit, 2008) Reproduced by permission of © IEEE 2008

Main functions of *MAC-Common Part Sublayer* (MAC-CPS) are: connection establishment and maintenance of the connection between the two peer entities; relay path management and routing; network entry and initialization; UL/DL hybrid ARQ (HARQ); handover procedures; multicast and broadcast services (MBS); relay station neighbourhood discovery; scheduling of bandwidth allocation and QoS management. *MAC-Security Sublayer* (MAC-SS) is mainly responsible for authentication, secure key ex-change, and encryption and integrity control procedures. *Relay-MAC* (R-MAC) takes care of concatenation and fragmentation of forwarded MAC PDU and control functions, such as scheduling, routing, flow control, etc. It also supports the use of the generic MAC header. *Relay-Physical* (R-PHY) implements all physical layer functionalities and procedures, such as sub-channelisation, modulation and coding for links between MMR-BS and RS and between RSs.

### 2.5 Frame Structure

The standard defines two different modes of operation for relaying: *transparent* and *non-transparent*; and the frame structure configuration depends on the type of relaying.

A *transparent relay station* (T-RS) does not transmit the preamble, frame control header (FCH) and MAP at the beginning of the frame. Instead, it receives the preamble, FCH and MAP, and an optional R-MAP transmission from the MMR-BS, as explained later. In transparent relaying, RSs do not forward framing information, and hence do not increase outer MMR-BS coverage; consequently, the main use case for transparent mode relays is to increase capacity within the MMR-BS coverage area. This type of RS is of lower complexity, and only operates in a centralized scheduling mode and for topology up to two hops. A simulation study on the gain in throughput that is possible to achieve in transparent mode relay-based 802.16j systems can be found in (Genc et al., 2008).

In a *non-transparent relay* (NT-RS), the contents of the FCH, DL-MAP and UL-MAP in the relay frame may be different from those in the MMR-BS frame. The frame header contains

essential scheduling information that nodes use to determine when they can transmit and receive data. Related to the modes of relaying are two different options for scheduling: *centralized* and *distributed*. In the former, all scheduling for all nodes in the MMR-Cell takes place in the MMR-BS; in the latter, the RSs have some autonomy and can make scheduling decisions for the nodes which they communicate with, as explained in detail later. In non-transparent relaying, RSs generate their own framing information or forward those provided by the MMR-BS depending on the scheduling approach (i.e., distributed or centralized). They can provide a wider coverage area and, therefore, are mainly used for increasing coverage. On the other hand, the transmission of the framing information can result in high interference between neighbouring RSs; hence, the capacity enhancement that can be achieved using these relays is rather limited. Also, they can operate in topologies larger than two hops, leading to different levels of complexity at the RS (Genc et al., 2008).

As transparent and non-transparent relays have different advantages, there can be some scenarios in which it makes sense to associate both with a single MMR-BS (Genc et al., 2008).

The *Frame Structure* required to support relay network architectures, similarly to the legacy 802.16 frame, is divided into two subframes (DL/UL). These subframes are further divided into different *zones* for MMR-BS-RS and RS-UT connectivity. The *access zones* are defined for supporting MMR-BS/NT-RS communications with the UT/T-RS. In transparent mode, a *transparent zone* is defined for T-RS communications with the UT; in non-transparent mode, *relay zones* are defined for MMR-BS/NT-RS communications with NT-RS.

As illustrated in Figure 5, for transparent mode (with two-hop topologies) it is only necessary to have a single access zone and one transparent zone in both DL and UL. The RSs switch mode (transmit ↔ receive) when the system switches zone; therefore, it is necessary to have a receive/transmit transition gap (R/TTG) between the two zones. The scheduling is performed via MAPs that are transmitted by the MMR-BS. Data can be transmitted from the MMR-BS to the RS in one frame and forwarded from the RS to UT in the subsequent frame.

An example of non-transparent mode frame structure (two-hop case) is shown in Figure 6. The MMR-BS and RS must be synchronized to transmit the frame preamble at the same time, and both the DL and UL subframes must be synchronized in both the MMR-BS and RS. The DL subframe includes at least one DL access zone and, optionally, one or more relay zones. The UL subframe may include one or more UL access zones and one or more relay zones. The system behaviour depends on the scheduling mode of operation; though, in both modes (centralized or distributed), the MMR-BS transmits two MAPs for scheduling transmissions to and from UTs to which connects directly, and one R-MAP for scheduling MMR-BS-RS communications. In centralized mode, the MMR-BS generates the MAP (or R-MAP, in topologies with more than two hops) for the RSs to schedule transmissions to/from their subordinates. The RSs transmit the received MAP messages at the start of its DL access zone. In distributed scheduling the RSs can perform their own scheduling; hence, they do not need these messages. Furthermore, non-transparent mode provides support for both single- and dual-radio RSs. For single-radio RSs, is necessary to introduce transition gaps between zones within the subframes, which are not necessary in the latter case. In the dual radio case, the RS communicates with its parent via one radio, and the other radio is used for communicating with its subordinates; typically, these radios would operate using different channels.

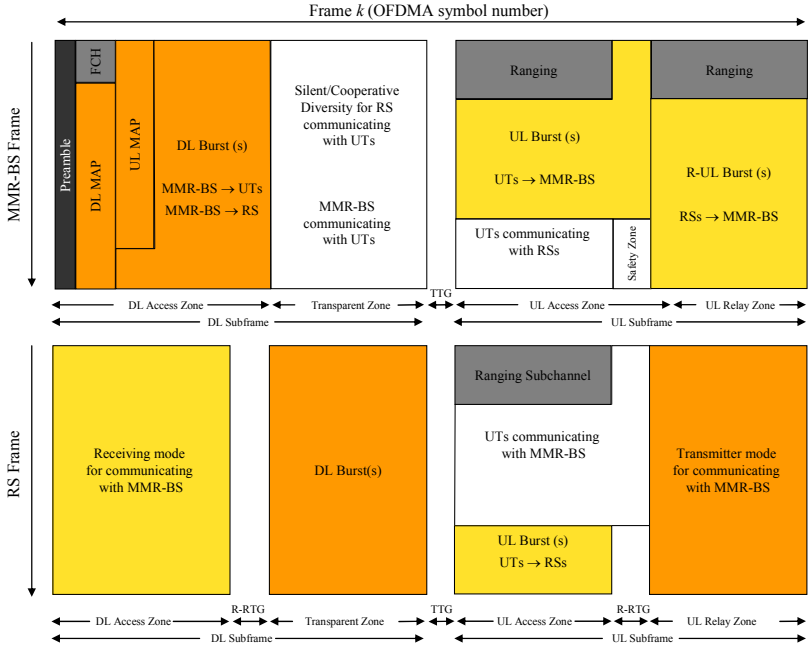



Fig. 5. Transparent mode frame structure as viewed at MMR-BS and at RS

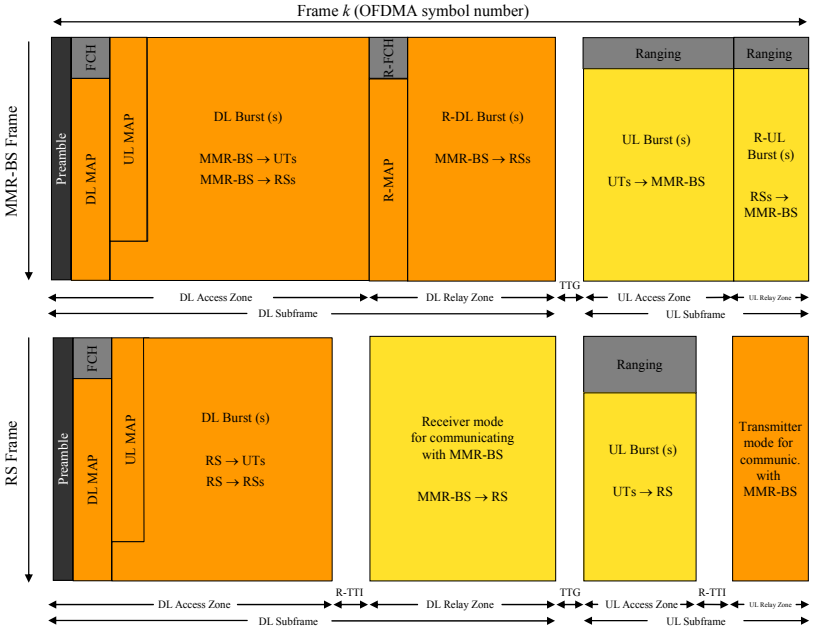



Fig. 6. Non-transparent mode frame structure (two-hop) as viewed at MMR-BS and at RS

### 2.6 Forwarding Schemes

Two different forwarding schemes are defined: the *tunnel-based* scheme and the *connection ID (CID)-based* scheme. As showed in Figure 7, both schemes are intended to maximize system efficiency by aggregating traffic wherever possible and requiring less signalling information, as groups of flows can be handled together. These schemes can be differentiated in terms of QoS management functionality, error handling, and overhead.

The tunnel-based scheme provides support for *explicit management tunnel connections* (bi-directional MT-CID) and *transport tunnel connections* (unidirectional T-CID), which are characterized by a unique CID, two specific endpoints, and QoS requirements. The tunnels aggregate traffic from disparate UTs on the MMR-BS-RS connection for either management or transport connections with similar QoS requirements. As depicted in Figure 7, the access station at the ingress of the tunnel adds a so-called *relay MAC header* to a packet or group of packets indicating the CID of the tunnel that the packet(s) should traverse. This header is removed when the packet arrives at the destination RS. It is worthwhile noting that *tunnel-in-tunnel* is also possible, which is essentially a recursive tunnel encapsulation, where a tunnel is constructed using a mix of individual connections and tunnels. In both centralized and distributed scheduling modes, the use of the tunnel requires some intelligence at the RSs along with the distribution of different service flows parameters to the RSs for *traffic prioritization*. In an MR network, where tunnels are established for traffic of the same service type, e.g. voice, 3 bits can be taken from the reserved bit field in the relay MAC header to implement the traffic prioritization function, especially when the scheduling is performed in a distributed manner at each RS. The tunnelling mechanism improves the efficiency of relay links application and can significantly simplify routing, QoS management and RS handover at the intermediate RSs along the relay path (Chen & De Marca, 2008).

The CID-based scheme has no such tunnels and supports only the legacy management and transport connections, as defined in the 802.16e-2005 standard. The packets are forwarded based on the CID of the destination station, which is contained in the MAC-PDU header. In centralized scheduling, the MMR-BS sends a message to the RSs describing the relay link channel characteristics, including an extra field specifying the delay associated with each packet in either the DL or UL. Thus, the RS knows in which frame each packet should be transmitted. This is necessary in order to meet the QoS requirements of each connection. In the distributed case, the RS has knowledge of the QoS requirements of each connection and can therefore make its own scheduling decisions (Genc et al., 2008).

The CID-based approach must be used in transparent relaying, whereas both schemes can be used in non-transparent mode, depending on the system configuration. The difference in terms of MAC efficiency provided by these schemes is accentuated as the number of hops increases in the system. Thus, the tunnel mode eventually leads to higher system efficiency as the number of hops in the system increases (Chen & De Marca, 2008).

### 2.7 Routing and Path Management

Routing and path management functions handle multi-hop paths between the MMR-BS and UT. As explained in Section 2.3, *routing* is tree-based, and algorithms for deciding which RS a particular UT should be associated with (routing/path selection) are left to vendors. The decisions are made at the MMR-BS based on information provided by the RSs. They may be based on metrics such as radio resource availability, radio link quality and traffic load at the RSs. *Path management* relates to path establishment, maintenance, and release.

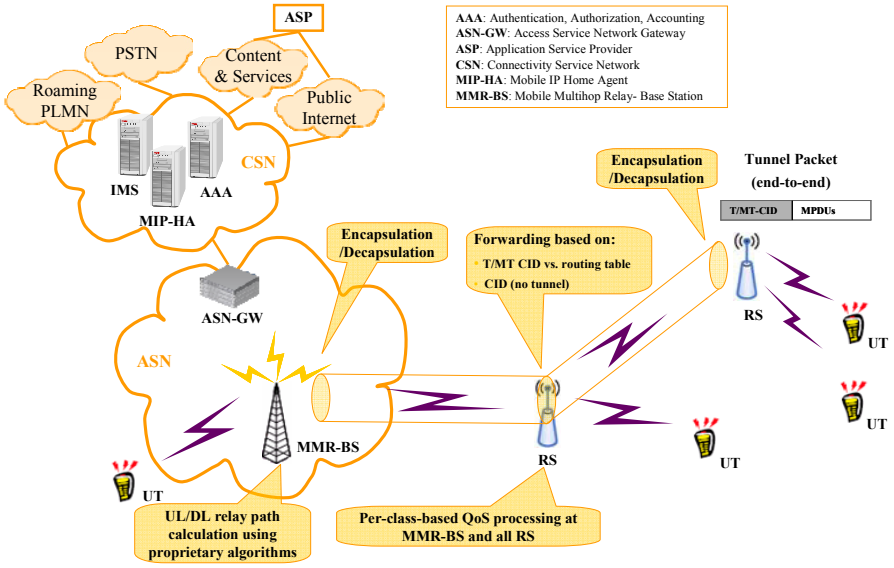



Fig. 7. WiMAX network architecture and an example of relay path management and routing (Soldani & Dixit, 2008) Reproduced by permission of © IEEE 2008

The standard defines two path management mechanisms for creating the most appropriate path: *embedded* and *explicit*. The main difference between the approaches is how signalling information to manage the path is distributed in the system (IEEE 802.16j, 2009).

In the *embedded path management* approach, a hierarchical CID allocation scheme is used in the MMR-Cell. The MMR-BS allocates CIDs to its subordinate stations such that the CIDs allocated to all subordinate RSs of any given station are a subset of the allocated CIDs for that station. In this way, there is no need of a specific routing table in each RS and very little signalling is required to update the path information.

The *explicit path management* mode uses an end-to-end signalling mechanism to distribute the routing table along the path. When a path is created, removed, or updated, the MMR-BS sends the necessary information to the RSs involved in the path. Each path is identified by a path ID, which the CIDs are bound to. This leads to smaller routing tables at the RSs and a reduction of the overhead required to update these tables. In distributed scheduling mode, to allow the RSs to make an independent decision regarding how to schedule the packet, the MMR-BS may optionally send (include) the QoS requirements associated with each CID.

The key difference of the two path management mechanisms stands in signalling overhead. The embedded scheme simplifies operation at the RS with no routing table, allows fast data forwarding, and generates less overhead and delay in route update than the explicit scheme. Thus, the embedded scheme is very suitable for managing paths in centralized scheduling. The explicit scheme should be used in distributed scheduling, as it allows the distribution of the routing table and service flow parameters to the RSs along the path.

An example of distributed path selection method for NT-RSs can be found in (Ann et al., 2008). The proposed method allows the NT-RS to select the optimal path using a path cost function with the PHY-layer loss rate, link bandwidth and hop count, as input performance metrics.

### 2.8 Initial Ranging and Network Entry

The network entry procedure for the UT remains unchanged. Some additional functions are introduced in the MMR-BS and RSs, in order to determine which node should be the access node for the UT. The initial ranging process depends on the scheduling and relay modes. In transparent relaying, the RSs monitor the ranging channel in the UL access zone (see Figure 5) and forward the ranging codes they receive to the MMR-BS. The MMR-BS determines the most appropriate path (i.e., direct or via an RS) from the messages with the same ranging code received from other RSs and sends a response directly, or indirectly via the RS, to the UT, depending on the path chosen. In non-transparent relaying, the UT selects the MMR-BS or NT-RS with the strongest preamble. The RS communicates with the MMR-BS only to ensure that the UT is "permitted network entry". The information exchanged with the MMR-BS depends on the scheduling mode.

The initial ranging process for RSs is similar to that for UTs in non-transparent mode relay. The MMR-BS (or NT-RS) can determine whether a node performing ranging is an RS or a UT based on the ranging code it receives, since specific set of codes are reserved for RSs. (In this way, transparent mode RSs can easily ignore ranging performed by other RSs, and RSs performing ranging can be also prioritized with respect to UTs.) In the RS network entry procedure, after the initial ranging, authentication, and registration processes, the MMR-BS may request the RS to provide the MMR-BS the signal strength of each of its neighbouring RSs. The MMR-BS can then determine the most suitable access station with which to associate the RS based, e.g., on traffic load, signal strength, and other relevant parameters. The final stage of network entry is then the configuration of the RS parameters, including its operation (e.g., transparent or non transparent) and scheduling modes (Genc et al., 2008).

### 2.9 Radio Resource Management

As already pointed out in the previous sections, bandwidth requests and allocations can be either distributed or centralized.

In *distributed scheduling*, see Figure 8 a), each MMR-BS and RS individually determines the bandwidth allocations on the links they control (i.e., downlinks to and uplinks from its immediate downstream stations) and create their own MAPs reflecting these decisions. Along a relay path, the MMR-BS and RSs must guarantee the performance of the following service flows: unsolicited grant services (UGS), real-time polling services (rtPS) and extended real-time variable rate (ERT-VR) services. Scheduling comprises bandwidth requests (including contention based CDMA bandwidth requests) from the immediate downstream stations, and grants and polling from the corresponding upstream stations.

In *centralized resource partitioning*, see Figure 8 b), the MMR-BS determines the bandwidth allocations for all links (access & relay) in its MMR-cell. Hence, before a station can transmit a packet to the MMR-BS, the bandwidth request from that station must first reach the MMR-BS, which in turn creates bandwidth allocations on the links along the path from the RS to the MMR-BS. The centralized scheduling encompasses the following procedures: contention-based CDMA bandwidth requests for relaying, continuous bandwidth allocation mechanism, dedicated uplink channel allocation, dedicated channel between MMR-BS and RS, and service flow based dedicated resource update.

An example of resource scheduling with directional antennas for multi-hop relay networks in a Manhattan-like environment can be found in (Chen & De Marca, 2008). Both MMR-BS and RSs were equipped with directional antennas. By taking advantage of the effect of high

degree of shadowing, the proposed methods could increase the system throughput up to 12 times, as compared to the system with omni-directional antennas.
The choice of a centralized or distributed system has profound ramifications throughout the entire network. For larger systems, in which there are more than two hops to the MMR-BSs, the benefits of distributed systems are most probably larger. However, many deployments are likely to be smaller topologies in which most communications will be via one or two hops; in these cases the centralized mode of operation seems preferable as it results in lower-complexity RSs, ultimately leading to lower overall costs (Genc et al., 2008).
Network coverage and cell throughput for transparent relaying system and non-transparent relaying system with centralized and distributed scheduling mode were compared in (Zeng & Zhu, 2008). The basic IEEE 802.16e system (with only BS, i.e. 1 hop system) was used as a comparison base line. Different RS configurations (1 or 3 RSs per sector) were considered. Performance results showed that the largest network coverage and highest system capacity were both achieved by the non-transparent relay system with 3 RSs per sector with distributed scheduling.
In IEEE 802.16j, the MMR-BS controls the *handover* (HO) process – centralized multi-hop relay network (MR network) control – and an RS relays HO associated messages between a UT and the MMR-BS, independent of the scheduling mode. After HO, the routing and QoS information along the old and new path is updated as described in Section 2.7.
The RS HO process hands off all the UTs attached to itself, along with the RS, to a target MMR-BS (see Figure 3, in the M-RS use case). It supports the same procedures as described for an UT HO process in the 802.16e with the following additional functionalities:

- *Access station selection*, which enables the target MMR-BS to indicate access station reselection.
- *RS operational parameters configuration*, which enables the target MMR-BS to reconfigure RS operational parameters.
- *Tunnel connection re-establishment*, which enables the target MMR-BS to re-establish tunnel connections for an RS.
- *UT CID mapping* (only for CID-based forwarding). This stage is used for the target MMR-BS to inform the RS, which is performing the handover, about the mapping of new CIDs and old CIDs of its subordinate UTs. The RS uses this information to swap the new CIDs with the old CIDs before forwarding the MAC PDUs to the subordinate UTs after the handover.

Quality measurements for each active UT that can be used form making an HO are: *received signal strength indicator* (RSSI), *carrier interference noise ratio* (CINR) and *timing adjust* (TA). The HO is executed after the signal strength from neighbour cell exceeds the signal strength of the current cell for a certain period of time (IEEE 802.16j, 2009).
A simulation study on the performance of HO techniques for IEEE 802.16j MR networks can be found in (Sultan et al., 2008). Simulation results, at the considered UT speeds, showed that Macro Diversity HO (MDHO) performed better than Fast Best Station Switching (FBSS) and Hard Handover (HHO). Also, the MDHO gain decreased as UT speed increased. FBSS performed slightly better than HHO. However, the overall average DL CINR decreased for the three handover techniques, as the UT speed increased.

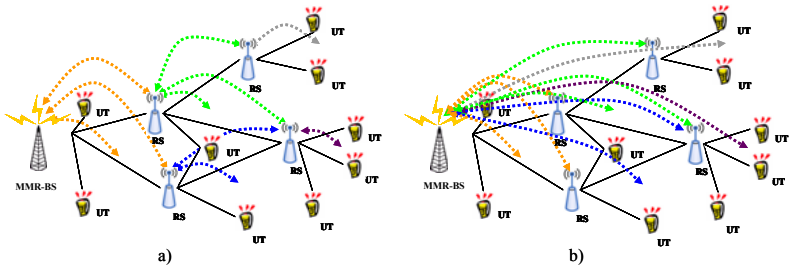



Fig. 8. Bandwidth request /allocation: a) distributed scheduling; b) centralized scheduling

MR networks support a *power control* algorithm for the uplink channels from RSs and UTs with both initial calibration and periodic adjustment procedure without loss of data. Power control of the RS downlink channels is supported as well. In the case of RSs operating in centralized scheduling mode, the UL power control algorithm is located in the MMR-BS and the MMR-BS controls the transmission power of all uplink channels served by the MMR-BS and its subordinate RSs. In distributed scheduling, an UL power control algorithm is located in both the MMR-BS and RSs to control the uplink channels they serve (see Figure 8).
For more information on the procedures described above the reader is referred to the Multi-hop Relay Specification (IEEE 802.16j, 2009).
Admission control (AC), caching of subscriber profiles and encryption keys, AAA client functionality, establishment and management of mobility tunnel with base stations, QoS and policy enforcement, mobile IP foreign agent (FA) functionality for mobile IP (MIP), and routing to the selected connectivity service network (CSN) are a part of the functionalities of the access service network gateway (ASN-GW). Load Sharing or balancing can be seen as an integral part of the proprietary algorithms for relay path management and handover control (Navaie et al., 2006).

### 2.10 Cooperative Relaying

In an MMR-Cell, cooperative relaying can be achieved using either an MMR-BS and one or more RSs or multiple RSs transmitting in cooperation. These RSs are either transparent RSs or non-transparent RSs transmitting the same frame start preamble, FCH and MAPs, or non-transparent using a common permutation zone during the transmission. Diversity gains and pilot collision mitigation are achieved by sending appropriately coded signals (correlated signals) to subordinate stations (e.g., S1, S2 and S3, in Figure 1) across different MMR-BS and RS transmit antennas.
In 802.16j, three modes of operation for cooperative relaying are possible: *cooperative source diversity*, *cooperative transmit diversity*, and *cooperative hybrid diversity*, which is a combination of the previous cooperative relaying schemes (IEEE 802.16j, 2009).
In *cooperative source diversity* (CSD), transmitting antennas simultaneously transmit the same signal using the same time-frequency resource. In practice, the transmission timing differences from multiple signal sources need to be within a cycling prefix (CP) period. An example of CSD is illustrated in Figure 9. In the simulation, the signals (2hop SISO and CSD) arriving at the UT were of the same power, and the signal noise ratio (SNR) on the relay link (MMR-BS – RS) was 30 dB. The channel model used in the simulation was SUI-4 model and the signal was transmitted using QPSK modulation with 1/2 convolution code.

In *cooperative transmit diversity* (CTD), space-time block coded (STBC) signals are transmitted across the transmitting antennas using the same time-frequency resource. The transmit structure is identical to the cooperative source diversity. However, in this diversity scheme, the main idea is that both the source and the relays transmit a part of the codeword symbols. Diversity is achieved based on the fact that a portion of the code word symbol arrives at the destination through a different and independent fading channel. An example of CTD is depicted in Figure 10. (The simulation assumptions were as for Figure 9.)
*Cooperative hybrid diversity* (CHD) is identical to cooperative transmit diversity except that at least one value for virtual antenna assignment is assigned to multiple physical antennas. An example of CHD is showed in Figure 11. (The simulation assumptions were as for Figure 9.)

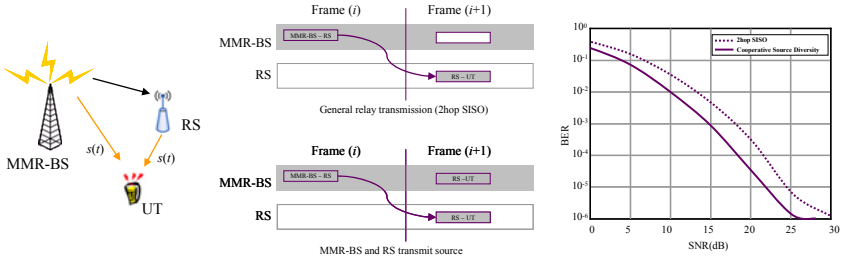



Fig. 9. Example of Cooperative Source Diversity (CSD): the received signals from different sources (MMR-BS and RS) is exactly the same

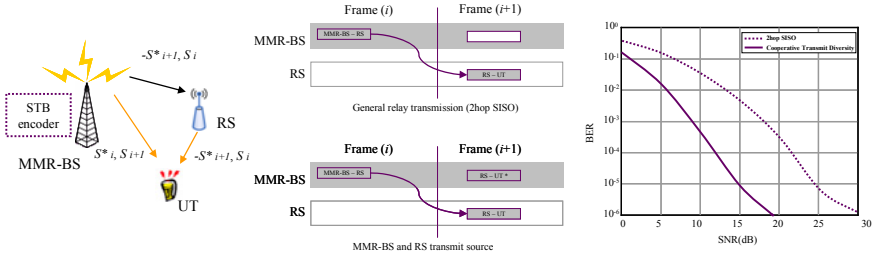



Fig. 10. Example of Cooperative Transmit Diversity (CTD) with full encoding in the MMR-BS (STB encoder): the received signal from different sources (MMR-BS and RS) is different

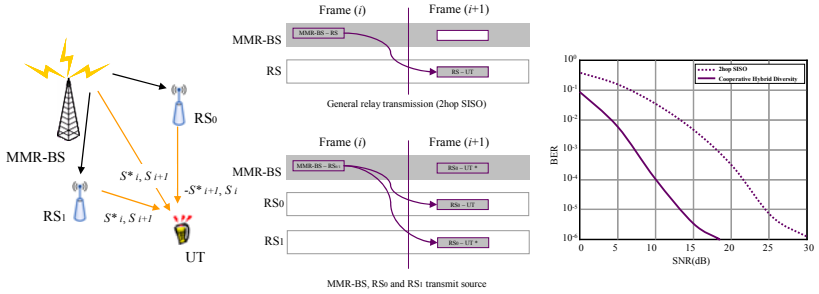



Fig. 11. Example of Cooperative Hybrid Diversity (CHD): multiple signal sources (MMR-BS and $RS_1$) can transmit the same STBC encoded signal to implement an Mx1 STBC scheme

## 3. Impact on Network Planning

The network planning process consists of the following phases: *preplanning*, *planning*, *detailed planning*, *acceptance*, and *optimization*.
In the preplanning phase, criteria for deploying wireless relays are agreed with customers. This encompasses the definition of traffic, services, network topology and deployment scenarios; coverage, throughput and system capacity requirements, relay type and usage model are discussed and agreed as well.
In the dimensioning phase, coverage range and an estimate of the number of MMR-cells based on the average SINR distribution in the cell is calculated. The noise limited maximum allowed path loss is a combination of link budgets calculated separately for RS to MMR-BS link, RS to UT link and MMR-BS to UT link, e.g., link 1, 2 and 3 in Figure 1. An example of link-budget for DL is illustrated in Figure 12. This assumes a sector antenna gain of 15 dBi and a maximum transmission power of 42.3 dBm at the MMR-BS antenna port, which yields an effective isotropic radiated power (EIRP) of 57.3 dBm.
The detailed network planning makes use of a tool for combined coverage and capacity computation. The tool usually supports traffic models, propagation and RRM models, and all necessary radio parameters for the MMR Cells.
In the network configuration and optimization phases, the additional MMR parameters are set and fine tuned based on service and network performance measurements. The latter activities depend heavily on the self-organizing capabilities of the network.
The following sections present some key indicators for link and system level performance assessment and an example of deployment cost analysis in different traffic scenarios and propagation conditions. For more information on evaluation methodologies and models, dimensioning and capacity evaluation for multi-hop WiMAX networks, and methods for cost analysis, the reader may consult the following relevant references (Puthenkulam et al., 2006; Hart et al., 2007; Hoymann et al., 2007; and Moberg et al., 2007).

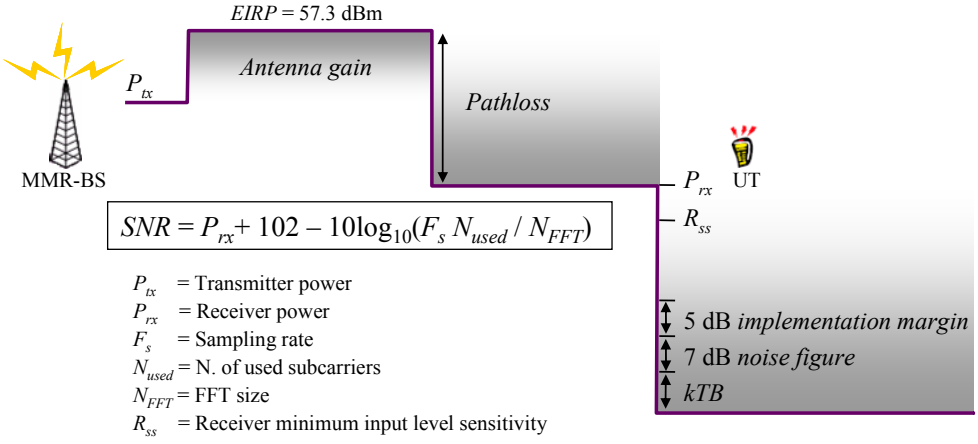



Fig. 12. Example of DL link budget

### 3.1 Key performance metrics

In order to obtain a viable metric for performance assessment, *coverage* (allowable outage probability) must be coupled with *capacity* (# of admissible UTs at a given min bit-rate).
This is due to the fact that, although a user may be covered for a certain percentage area (e.g. 99%) for a given service, the resources (time, frequency, power) are to be shared among UTs in a sector/MMR-BS. It can be expected that the UT average throughput may be reduced by a factor of $N$, when there are $N$ active users compared to a single user rate (assuming no multi-user diversity gain and resources equally shared). On the other hand, the total cell throughput could be adversely affected if users having poor channel quality were provided more resources for better QoE. Thus, there is a trade-off between coverage and capacity, and any measurement of capacity should be provided with the associated coverage and quality of end user experience.
The combined index for *coverage & capacity* (CC) assessment is given by the number $N$ of simultaneous UTs per cell (e.g. MMR-cell or legacy cell) that can be supported achieving a target information throughput $R_{min}$ with specified coverage reliability.
The *packet call throughput* (PCT) is given by

$$PCT = \frac{1}{K}\sum_{k=1}^{K}\frac{BitsPacketCall(k)}{t_{end}(k) - t_{arrival}(k)} \tag{1}$$

where $K$ is the total number of packet calls.
The *system data throughput* of a MMR-BS is defined as the number of information bits per second that a site can successfully deliver or receive using scheduling algorithms.
The corresponding *effective system spectral efficiency* (ESSE) is calculated as:

$$ESSE = \frac{SystemDataThroughput}{TotalSiteBWallocted} \tag{2}$$

In an MR network, a key performance indicator for quantifying the end-to-end quality of an $M$–hop routing path in terms of throughput is given by:

$$EtoET = \left(\sum_{m=1}^{M}\frac{1}{LinkThroughputHop(m)}\right)^{-1} \tag{3}$$

where $LinkThroughputHop(m)$ is the throughput per link over the $m$–hop.
Another end-to-end PHY abstraction metric is given by the harmonic mean of the capacities ($C$) over individual wireless links, as a function of the SINR:

$$C = \left(\sum_{m=1}^{M}\frac{1}{C_m}\right)^{-1} = \left(\sum_{m=1}^{M}\frac{1}{\log_2(1+SINR_{eff,m})}\right)^{-1} \tag{4}$$

where

$$SINR_{eff} = 2^{C} - 1 \tag{5}$$

is the effective SINR for the multi-hop route; $C$ the effective end-to-end capacity for the multi-hope route; $C_m$ the effective capacity over hop $m$, $m$ = 1, …, $M$; $SINR_{eff,m}$ the effective SINR over hop $m$; and $M$ the number of hops over the established route between MMR-BS and UT. (The above formula is valid in the case of orthogonal channels (e.g. slots) for inter-relay communication. This relay capacity model applies only for small $M$ (1-3). For large $M$, the same resource (e.g. slot) can be reused in the relays farther a part, and, hence, this needs to be accounted for in the capacity calculation.)

An example of multi-hop wireless networks capacity is illustrated in Figure 13. The capacity is related to a one-dimensional network, where an MMR-BS and UT communicates through multiple intermediate relay stations located equidistantly, as depicted in the figure.

In the simulations, the channel model included path loss and lognormal shadowing. No spatial reuse, no interference, no synchronization error were considered. Outage was defined as the event in which the achieved end-to-end data rate felled below the target data rate. In Figure 13, the Spectral Efficiency, i.e. $C_m$, denotes the maximum achievable rate per Hz on hop–$m$, and $d$ is the inter-relay stations distance.

As shown in Figure 13, the deployment of relay stations improves the spectral efficiency. Also, the simulation results demonstrate that *a MR network with maximum 2-3 hops provides the best network performance*. More hops in the MR network would not improve the situation.

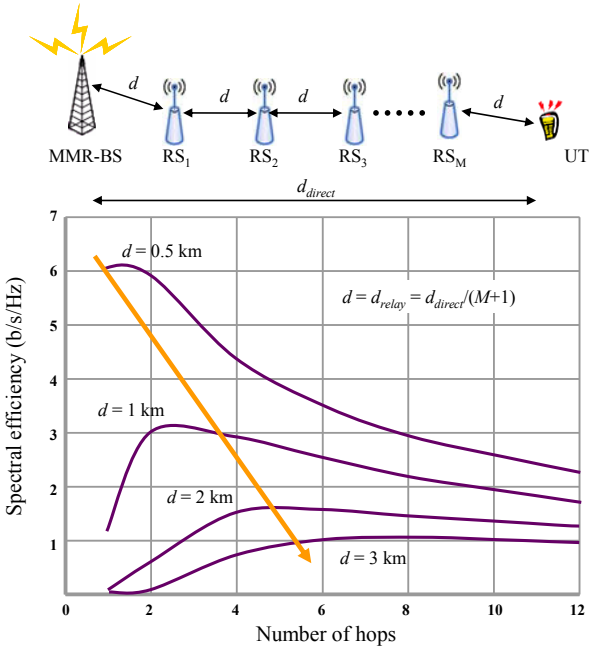



Fig. 13. Example of example of capacity of multi-hop wireless networks

### 3.2 Example of Deployment Cost Analysis

This section discusses the relative CAPEX and OPEX (total cost of ownership) of an MMR approach versus a conventional WiMAX deployment at 3.5 GHz, to meet the same coverage and capacity requirements. This is studied for the urban environment with heavy traffic, and for the urban, suburban and rural environments with light traffic.

The cell structures are dimensioned for a minimal SINR of 3.7dB at the edge. The cell split for conventional WiMAX is based on capacity demand, whereas the MMR system is dimensioned for heavy load. The channel bandwidth is 30, 20 and 10 MHz for conventional WiMAX, MMR-BS and RS, respectively. The spectral efficiency is 5 b/s/Hz for conventional WiMAX and MMR-BS, and 2 b/s/Hz for the RS.

In the analyzed deployment scenarios, the MMR-BS to RS ratio is 1:56, 1:33, and 1:12. CAPEX consists of site acquisition and construction costs per cell, wired backhaul costs and station costs (e.g., hardware, software).

Backhauling and station costs for a MMR-BS are assumed to be higher than for a conventional BS. Civil work expenditures are supposed to be the same for base stations and much lower for deploying a RS, which is also considered much cheaper than any BS.

OPEX comprises all administrative costs for backhaul, access points, and network. This expenditure is considered to be the same for the base stations and much lower for a RS.

A sample of the analyzed networks and the resulting deployment costs normalized and relative to the MMR CAPEX value with RS to MMR-BS ratio 56 are showed in Figure 14.

In the conventional WiMAX deployment, CAPEX is a significant cost with respect to OPEX. In the MMR approach, CAPEX decreases if the MMR-BS to RS ratio increases and it is considerably less than OPEX in the capacity limited scenario (heavy traffic).

Further, the total costs of the MMR approach are always less than those for the conventional WiMAX, and savings in expenditure from capacity improvement in heavy traffic scenarios, e.g., in urban environment, is significantly higher than those from range extension.

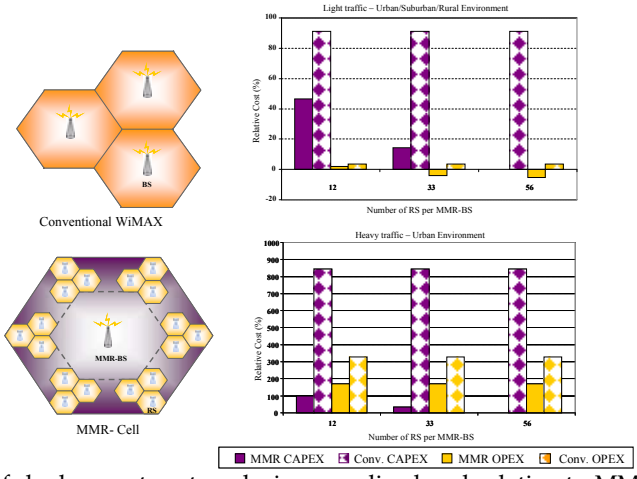



Fig. 14. Results of deployment cost analysis normalized and relative to MMR CAPEX with RS to MMR-BS ratio 56, for urban environment with heavy traffic and for urban, suburban, and rural environment with light-traffic density (Soldani & Dixit, 2008) Reproduced by permission of © IEEE 2008

## 4. Conclusions and Future Work

Relay technology to extend coverage and range has been receiving a lot of attention due to its simplicity, flexibility, speed of deployment, and cost effectiveness. This is particularly so in scenarios where first responders need to communicate in the disaster and emergency situations. Relaying also offers a cost-effective way to deliver broadband data to the rural communities where the distances may be large and population density sparse.

Some key advantages of relays are: (a) they do not require backhauling resulting in lower CAPEX and OPEX, (b) flexibility in locating relay stations, (c) when located in a cell, relays can enlarge the coverage area and/or increase the capacity at cell border, (d) decrease transmit power and interference, and (e) mobile relays enable fast network rollout, indoor-outdoor service, and macro diversity by way of cooperative relaying.

However, relaying is not without drawbacks, namely increased use of radio resources in in-band relaying (time domain) and need for multiple transceivers in out-of-band relaying (frequency domain). Relays also introduce additional delays.

Overall, the substantial amount of choice, coupled with a general lack of understanding of the impact of the different design decisions, makes the system design difficult, and much research remains to be carried out, in order to understand how 802.16j systems perform under different configurations and at what cost compared to 802.16e systems.

As a matter of facts, the MR network architecture is currently a relatively new design and introduces many complexities within the already challenging environment of radio access networks with mobility support. Many of the issues remain still unsolved, and more work is necessary to really understand the cost/benefit trade-offs that arise in IEEE 802.16j systems. Also, resource allocation in MR networks requires the design of novel scheduling algorithms with QoS differentiation for improving QoE, e.g., in terms of reliability, fairness, and latency. In this respect, there are many aspects that require further investigation; these include the approaches to realize distributed systems, ways to maximize spatial reuse, and dynamic mechanisms to control the amount of resources allocated to each of the zones in both the transparent and non-transparent relaying modes.

Fast-forwarding into the future, the relay stations will not be confined to just decode and forward, but will also support additional capabilities, such as being able to connect to more than one RS both in the downstream and upstream direction, support routing, multicasting, and dynamic meshing. (These are a part of the advanced relay station (ARS) characteristics defined in IEEE 802.16m (IEEE 802.16m, 2008). The ARS supports procedures to maintain relay paths, mechanisms for self configuration and self optimization and multi-carrier capabilities.) When such evolution will have occurred, the relay network beyond the MMR-BS will mimic a mesh topology and the MMR-BS will simply function as a gateway to the Internet core while connecting to the nearest relay nodes in the downstream direction. Mesh and self organizing capabilities will enable connection reliability, traffic load balancing, and proactive topology management.

Ultimately, it remains to be seen how wireless relays will compete against other important solutions, such as femto base stations, and conventional broadband networks that will use lower carrier frequencies and optimized backhauling, for example, using digital subscriber lines (xDSLs), passive optical networks (xPONs), and broadband meshed microwave links.

Overall, wireless relays offer great advantages and will continue to receive a lot of attention both in the research and business communities.

## 5. References


Andrews, J. G.; Ghosh, A. & Muhamed, R. (2007). *Fundamental of WiMAX – Understanding Broadband Wireless Networking*, Prentice Hall, ISBN: 0132225522, USA

Ann, S.; Lee, G. K. & Kim, S. H. (2008). A Path Selection Method in IEEE 802.16j Mobile Multi-hop Relay Networks, *Proceedings of the 2nd International Conference on Sensor Technologies and Applications*, pp. 808-812, ISBN: 978-0-7695-3330-8, Cap Esterel, Aug. 2008, IEEE

Chen, K. C. & De Marca J. R. B. (2008). *Mobile WiMAX*, Wiley & Sons and IEEE, ISBN: 978-0-470-51941-7, UK.

Genc, V.; Murphy, S. & Murphy, J. (2008). Performance analysis of transparent relays in 802.16j MMR networks, *Proceedings of the 6th international Symposium on Modeling and Optimization in Mobile*, Ad Hoc, and Wireless Networks, pp. 273-281, ISBN: 978-963-9799-18-9, Berlin, Apr. 2008, IEEE

Genc, V.; Murphy, S.; Yang, Y. & Murphy, J. (2008). IEEE 802.16j relay-based wireless access networks: an overview, *IEEE Wireless Communications Magazine*, Vol., N. 15, (October 2008), pp. 56-63

Hart, M. et al. (2007). Multi-hop Relay System Evaluation Methodology (Channel Model and Performance Metric), *Contribution to the IEEE 802.16 Broadband Wireless Access Working Group*, IEEE 80216j-06/013r3, http://www.ieee802.org/16/relay/

Hoymann, C.; Dittrich, M. & Goebbels, S. (2007). Dimensioning and capacity evaluation of cellular multihop WiMAX networks, *Proceedings of the Mobile WiMAX Symposium*, pp.150-157, ISBN: 1-4244-0957-8, Orlando, Mar. 2007, IEEE

IEEE 802.16j Draft Standard P802.16j/D9 (delta), *Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems - Multihop Relay Specification*, Feb. 2009, IEEE, http://www.ieee802.org/16/published.html

IEEE 802.16m Draft Standard 802.16m-08/003r6, *IEEE 802.16m System Description Document*, Dec. 2008, IEEE, http://www.ieee802.org/16/published.html

Moberg, P.; Skillermark, P.; Johansson, N. & Furuskar A. (2007). Performance and cost evaluation of fixed relay nodes in future wide area cellular networks, *Proceedings of the 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1-5, ISBN: 978-1-4244-1144-3, Athens, Sept. 2007, IEEE

Navaie, K.; Liu, Y.; Abaii, M.; Florea, A.; Yanikomeroglu, H. & Tafazolli, R. (2006). Routing mechanisms for multi-hop cellular communications in the WINNER air interface, *Proceedings of the 64th Vehicular Technology Conference*, pp. 1-4, Montreal, Sept. 2006, IEEE

Pabst, R.; Walke, B. H.; Schultz, D. C.; Herhold, P.; Yanikomeroglu, H.; Mukherjee, S.; Viswanathan, H.; Lott, M.; Zirwas, W.; Dohler, M.; Aghvami, H.; Falconer, D. D. & Fettweis, G. P. (2004). Relay-based deployment concepts for wireless and mobile broadband radio, *IEEE Communications Magazine*, Vol., N. 42, (Sept 2004), pp. 80-89

Puthenkulam, J. et al. (2006). Tutorial on 802.16 Mobile Multihop Relay, *Contribution to the IEEE 802.16 Broadband Wireless Access Working Group*, 802 Plenary, Mar. 2006, IEEE 802.16mmr-06/006, http://www.ieee802.org/16/sg/mmr/

Soldani, D. & Dixit, S. (2008). Wireless relays for broadband access, *IEEE Communications Magazine*, Vol. 46, March 2008, pp. 58-68, ISSN: 0163-6804

Sultan, J.; Ismail, M. & Misran, N. (2008). Downlink performance of handover techniques for IEEE 802.16j multi-hop relay networks, *Proceedings of the 4th IEEE/IFIP International Conference on Internet*, pp. 1-4, ISBN: 978-1-4244-2282-1, Tashkent, Sept. 2008, IEEE/IFIP

Van Der Meulen, E. C. (1971), Three-terminal communication channels, *Advances in Applied Probability, Applied Probability Trust*, Vol. 3, N. 1, spring 1971, pp. 120-154

WINNER and WINNER+, http://projects.celtic-initiative.org/winner+/index.html

Zeng, H. & Zhu C. (2008). System-level modelling and performance evaluation of multi-hop 802.16j systems, *Proceedings of the International Wireless Communications and Mobile Computing Conference*, pp. 354-359, ISBN: 978-1-4244-2201-2, Crete Island, Aug. 2008, IEEE

22

# Broadband communication in the high mobility scenario: the WiMAX opportunity

M. Aguado, E. Jacob, M.V. Higuero and P. Saiz
*University of the Basque Country (UPV/EHU)*
*Spain*

Marion Berbineau
*INRETS -Institute National de Recherche sur les Transports et leur Sécurité*
*France*

## 1. Introduction

Nowadays, the emerging broadband wireless access technologies face the long term challenge to properly address the air link channel limitations with the growing demand on services, fast mobility and wide coverage. One of the most demanding and challenging scenarios is the high mobility scenario; scenario that matches the railway domain.

The International Telecommunication Union (ITU) – radio division (ITU-R), in its standardization global role, has recently identified the IMT-Advanced family as those mobile communication systems offering technical support for such high mobility usage scenarios. In October 2007, ITU-R decided to include the WiMAX technology in the IMT-2000 family of standards; and, in the near future, the next IEEE802.16 specification, the IEEE802.16m project, will cover the mobility classes and scenarios supported by the IMT-Advanced, including the high speed vehicular one.

This chapter covers the WiMAX opportunity in this low dense, full mobility and high demanding railway scenario. In order to do so, this chapter is structured as follows:

Section 2 presents and characterizes one of the most typical high speed vehicular scenario: the railway scenario. Subsection 2.1 describes the existent data communication networks in the railway domain. Subsection 2.2 introduces the current trends in the railway domain regarding IT services .

Section 3 describes the current telecom context. Section 4 provides an analysis on the open WiMAX network specification as a valid player for matching the railway requirements previously specified in section 2. The currently set of technical, regulatory, and market aspects that contribute to identify the mobile WiMAX access technology as a competitive solution in the railway context are shown.

## 2. The Railway Context

Traditionally, railway transport is one of the industry sectors with a greatest demand on telecom services due to the intrinsic mobile nature of the resources involved.
However, the railway domain introduces quite specific and challenging requirements to a general wireless communication architecture, system or technology, such as: high mobility, high handover rate, compatibility with legacy or non-conventional applications, stringent quality of service (QoS) indicators and reliability. These legacy applications are related to signalling and train control and command systems. Such signalling systems highly demand communication availability; if there is any communication loss, the signalling system is disrupted and trains stop. The embedded information within these systems is related to control train movement and is based on very strict safety rules. Moreover, railway environment is also a really harsh environment from the electromagnetic point of view; high vibration, thermal noisy, high number of different radio systems everywhere, cohabitation between high power (traction) and low power systems (electronic)…
On the other hand, and once exposed the challenges, it is also fair to outline some facts that may turn the railway domain in a favourable scenario from the telecom point of view. In normal conditions, not in busy yards, the railway network is not a heavy loaded telecom network as it can be considered a traditional one. Secondly, from the operational railway point of view, the supported services are pretty well defined. Not only is the mobile node's mobility pattern predictable but also its data traffic profile. Being that way, it is possible to identify and predict the most complex and challenging use cases to be supported by the network architecture.

### 2.1 Railway Communication services

From a general point of view, three main types of communications flows exist in the guided transport context (railway and underground):

- Train to ground communications (vehicle to infrastructure communications).
- Train to Train communications (Vehicle to vehicle communications).
- In train communications (Intra vehicle communication).

The requirements for train to ground communications in the guided transport field are generally divided in two main families related to **safety** and **non safety** applications. The first family is quite demanding in terms of robustness and availability, but the amount of information exchanged is generally low. In order to attend these safety applications, railway communication dedicated architectures have traditionally been deployed. On the contrary, non safety applications require high data rate. They use dedicated communication architectures, shared communication architectures or, even sometimes, they rely on public and commercial communication systems. The number of these non-safety applications keeps growing.
Following the traditional UIC (Union Internationale des Chemins de Fer or International Union of Railways) classification, it is possible to classify the fundamental Train to ground communication needs in the following application fields.
Safety applications

- Voice and data communications between CCC (Command Control Centre) and drivers. This application consists in providing voice and data communications in order to control, ensure and increase the safe movement of trains.

- Data communication for Automatic Train Control (ATC) systems.
- Data communications for remote control applications such as: remote control of engine for shunting, remote control of trains at line opening and closure, remote control of customers information systems, remote control of interlocking, remote control of electrical substations, remote control of lighting, electrical stairs, lifts, emergency ventilation installations, etc…
- Voice communications for broadcast emergency calls, for shunting in depot areas and for workers during track maintenance activities.

Non safety applications

- Voice and data communications in depot, maintenance and yard areas.
- Voice and data communications from and towards a train for staff, customer's services, diagnosis and maintenance message. These information exchanges aim to increase operation efficiency.
- Voice and data transmission for crew members
- Voice and data transmission for security applications
  These applications consist of: the supervision with discreet voice listening inside trains from a central control room to the surface (Centralized Control room, Security Control room); supervision of trains with discreet digital video record for trains from a central control centre on the surface; digital video broadcast in the drivers' cabin of the platform supervision at stations
- Voice and data communications for passenger services
  Passengers on public transport (underground, train or plane) or private transport (car) expect the information they usually receive in day-to-day life, whether professional or private, to be available to them during their journeys. These demands will increase significantly with the growing market of mobile telecommunications. The main needs identified in general are listed here: public phone, fax, passenger call service, connection to external networks and computers, entertainment videos, live radio channels, live TV channels, video-on-demand, tourist, multimodal and traffic information, information panels at the platforms and inside the units, database queries for passengers or staff, E-mail, Internet browsing, other Internet services, VPN secure connection to company's Intranet, Audio and video streaming, Video-conference..

### 2.2 Railway Trends regarding IT services

The increasing complexity of railways systems, the new European directive regarding the separation between track owner and train operators and future deregulation regarding maintenance, push the development of a huge variety of information systems. In addition, the following current trends can be pointed out:

A. Suppress cables and discontinuous data communication equipment installed between the tracks in order to avoid vandalism and to decrease maintenance costs.
B. Use of open technology and IP equipment interoperability, avoiding protocols and proprietary solutions.
C. Utilization of telecommunication technologies that have been proven and validated in other industries (Component Off the Shelf –COTS). Essentially, well proven and cost-effective solutions are the main goal.

D. Minimize obsolescence. Due to the high cost of a telecommunication system deployment along a railway, all equipments and systems installed along the railway net are expected to have a working life of around 30 years. Currently this requirement is being slightly loosened.
E. Migrate from a dedicated network infrastructure towards an infrastructure supporting critical and complimentary services with prioritization.
F. Increase data acquisition from the train and from wayside equipment involving high capacity broadband networks (Fibre, Gigabit backbone networks) and then enhance safety through complimentary services.

Having into account these trends regarding IT railway services, a set of general requirements can be identified for the communication technologies in the railway domain.

*1. Broadband Wireless Digital Radio Access Support*

Railway technologies shall be based on wireless digital communication technology, minimizing cable deployments and this way lowering maintenance cost and contributing to higher availability indicators.

2. *Support for Full Mobility and High Speed Vehicular Scenario*

Railway communication technologies shall support the high speed vehicular profile (up to 500km/h), solving the mobility management and re-attachment problem, and providing low latency and seamless handover between cells without data loss.

*3. High Data Rate Support*

Railway communication technologies shall provide broadband communication in both uplink and downlink communication. It shall provide higher capacity (traffic volume/number of users) than second and third generation of mobile communication technology. This way the architecture shall provide support for the previously identified trend related to increase the high quantity of data acquisition from train and wayside equipment and high capacity network utilization.

*4. Low Latency*

Railway communication technologies shall cater for low end-to-end latency able to support high demanding real time applications in full mobility.

*5. End-to-end Quality of Support*

Railway communication technologies when making use of packet or connection oriented based technologies shall provide end-to-end QoS support. This means that, it shall be possible to provide support for critical applications prioritization. Emergency support and priority access is one of the important requirements for critical railway services. The radio access technology should be able to provide differentiated levels of QoS – coarse grained (per user) and/or fine-grained (per service flow per user). It will be able to implement admission control and bandwidth management.

*6. Advanced Security Scheme*

Railway communication technologies shall support a security scheme with mutual authentication, able to cope with the critical services messages vitality, integrity and authenticity. The mobility scheme chosen should support different levels of security requirements, such as user authentication, while limiting the traffic and time of security process, i.e., key exchange.

*7. Scalability, Extensibility, Coverage*

Railway communication technologies shall support incremental infrastructure deployment. The railway communication architecture may accommodate a variety of

backhaul links, both wireless and wire line and be able to be integrated in a fibre deployment.

*8. Operate at Licensed and Licensed exempt frequency bands*

The railway communication technology shall work at licensed and licensed exempt frequency bands. This requirement is aligned with another demand that is commonly manifested by railway operators. As seen before, due to the safety and critical nature of the train control communication service, railway operators have typically eschewed shared public and commercial network solutions and have been responsible for designing and maintaining their own telecom network. Railway operators normally demand the possibility of totally controlling the communication architecture due to the inherent responsibilities that failures, malfunctioning or low performance indicators in this architecture, may represent on railway operators´ own safety and performance.

*9. Cost-effective Deployment Based on Open and Standard Based Technology*

The railway communication technologies shall facilitate a cost effective deployment. In order to do so, these technologies will follow the international standardization framework, which further enhances the economic viability of the solution proposed. The architecture shall provide support for IP equipment interoperability.

There are some other important features such as maturity and mesh support that have to be taken into account when choosing the railway access technology. Mesh support is related to the demanded "direct mode" communication; in this case, every connection is not necessarily performed via the network.

The standards that define the new wireless digital communication technologies cover only the PHY and MAC layers. And just specifying these layers is not sufficient to build an interoperable broadband wireless network for railway critical services. Rather, it is necessary to propose an interoperable network architecture framework capable to deal with the end-to-end service aspects such as QoS and mobility management. A full railway communication architecture that may serve as a valid alternative to the existing GSM-R deployments shall be a full stack end-to-end architecture. It shall also provide robustness and redundancy, this way increasing availability. Mechanisms such as support for hot standby configuration and redundant coverage deployments shall be implemented.

Additionally, the architecture shall support a broad set of mobility, deployment and use case scenarios and co-existence of fixed, nomadic, portable and mobile (and full mobile) usage models. Last, but not least, and as a general good telecom practice, the communication architecture shall allow a functional decomposition and support management schemes based on open broadly deployable industry standards.

## 3. Telecom Context

In the last few years, traffic profile in Wireless Mobile Networks has changed abruptly. Figure 1 shows the data services as the key service driving the bandwidth demands in Wireless Mobile Networks, together with the migration from a circuit switching traditional approach towards a packet switching strategy where packets are routed between nodes over data links shared with other traffic. In each network node, packets are queued or buffered, resulting in variable delay.

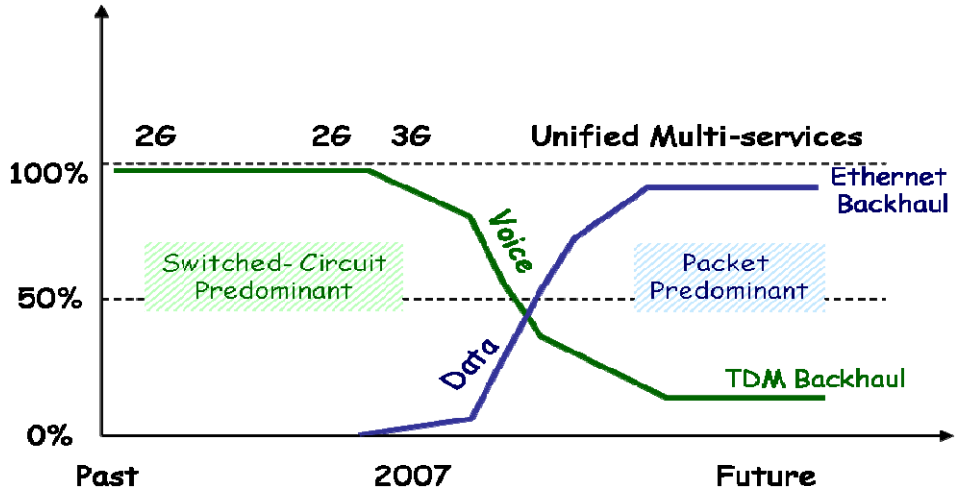



Fig. 1. Voice and data trends in mobile networks (Source International Wireless Packaging Consortium IWPC Milan 2008)

It is foreseen that the development of IMT-2000, the ITU global standard for third generation wireless communication, will reach a limit of around 30 Mbps. In the vision of the ITU [ITU-R M.2072], there may be a need for new wireless access technologies capable of supporting even higher data rates.

The ITU-R has recently proposed the International Mobile Telecommunications – Advanced (IMT-Advanced) technical requirements; one of the most demanding and challenging scenarios covered by the IMT-Advanced is the high speed scenario. The new capabilities of these IMT-Advanced systems are envisaged to handle a wide range of supported data rates according to economic and service demands in multi-user environments. Target peak data rates are up to approximately 100Mbit/s for high mobility, such as mobile access, and up to approximately 1 Gbit/s for low mobility such as nomadic/local wireless access. However, it is necessary to take into account that IMT-Advanced is a long term endeavour. The specification of IMT-Advanced technologies will probably not be completed until at least 2010.

Until recently, there was a technological gap regarding access techniques which could offer high transmission data rates and high interactivity (low latency) able to support real time applications in high mobility environments. However, research community efforts are underway to develop new generation wireless mobile networks that provide broadband data communication in this high speed vehicular scenario and new technologies capable of fulfilling the aforementioned technology gap have been developed, Figure 2. Currently, there are a number of initiatives that aim to provide ubiquitous connectivity at different mobility profiles.

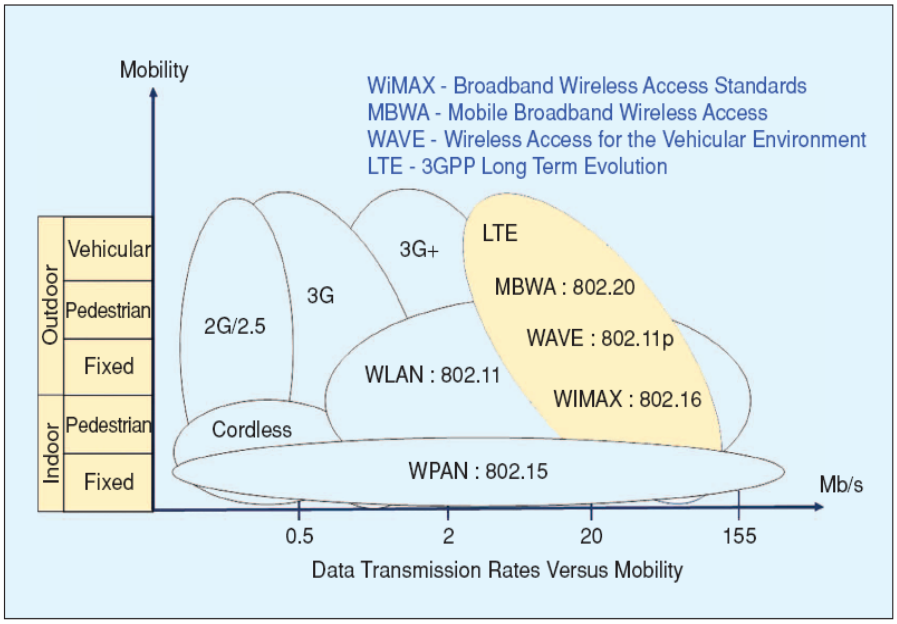



Fig. 2. Radio access technologies scenario: mobility versus data rate.

The *standard based* broadband wireless technologies able to support the vehicular mobility profile while offering a high transmission data rate are:

- IEEE802.11p or Wireless Access for the Vehicular Environment (WAVE),
- IEEE802.20 or Mobile Broadband Wireless Access (MBWA),
- IEEE802.16,
- Third Generation Partnership Project (3GPP) Long Term Evolution (LTE)

These emerging broadband and mobile access wireless technologies have some common features such as QoS support, low latency and advanced security mechanisms. They are also designed to support QoS and real-time applications such as voice-over-Internet protocol (VoIP), video, etc. They also may offer deployment bandwidth on the order of 40 to 100Mbps per base station.

OFDM and higher order MIMO antenna configurations are the core enabler for scaling throughput of these wireless mobile technologies. IEEE802.16, 3GPP and 3GPP2 standards bodies are all adopting OFDM & MIMO for 4G (WiMAX Forum, 2008). Figure 3 shows how all the three 4G candidates are based on OFDM and MIMO, consequently their major features are similar.

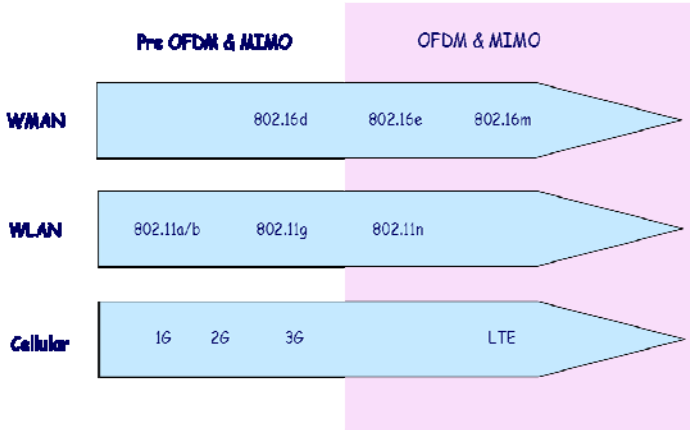



Fig. 3. All roads lead to OFDM and MIMO (WiMAX Forum, 2008)..

The IEEE802.16.m specification and the Third Generation Partnership Project (3GPP) long term evolution (LTE) specification are currently the only two candidates to cover the IMT-Advanced requirements.

## 4. The WiMAX opportunity in the railway domain

The scope of the IEEE 802.16 family is to develop the air interface technology for Wireless Metropolitan Area Networks (WirelessMAN), by specifying the medium access control layer (MAC) and the physical layer (PHY), of combined fixed and mobile *broadband wireless access* system providing *multiple* services. (IEEE 80216e, 2005)
The IEEE 802.16 standard has evolved from a fixed scenario IEEE802.16d towards a mobile typical vehicular (up to 120km/h) with IEEE802.16e. The group is now working on the revision of IEEE std 802.16-2004 (IEEE P802.16Rev2). WiBRO, the WMAN South Korean initiative, began to align itself with the WiMAX Forum implementation in 2004. Currently both approaches are harmonized in current WiMAX profile.
The IEEE 802.16 is focused on filling the existent gap between very high data rate wireless local area networks and very high mobility cellular systems. The next standardization effort in which the IEEE 802 is involved, *the IEEE 802.16m project*, follows this line. The IEEE 802.16m amends the IEEE 802.16 WirelessMAN-OFDMA specification providing an advanced air interface that meets the requirements of next generation mobile networks targeted by the cellular layer of IMT-Advanced. The purpose of this standard is to provide performance improvements necessary to support future advanced services and applications, such as those described by the ITU in Report ITU-R M.2072 [ITU-R M.2072].
IEEE 802.16m specification supports the mobility classes and scenarios supported by the cellular systems IMT-Advanced, including high speed vehicular scenario (up to 350km or even up to 500km/h). (P80216MPR2007)
Regarding the economic viability of IEEE 802.16, the proposed amendment is done within the framework of international standardization, which will further enhance the economic viability of the standard. Because IMT-Advanced is intended to be a globally deployed

system, it is expected that cost effective performance can be achieved through large economies of scale. (P80216M2006). IEEE 802.16e was first released in December 2005. Currently, the standard and its certification programs have clearly demonstrated its maturity.

IEEE802.16 specification provides mobile to mobile infrastructure-less communication. IEEE802.16j specification (to be included in next IEEE802.16 release) expands IEEE802.16 current deployment alternatives PTP, PTMP and mesh topologies by supporting a relay topology, this way enhancing the IEEE802.16 support for backhaul and last mille deployments. It specifies multi-hop relay capabilities and functionalities of interoperable relay stations and base stations.

IEEE802.16 standard supports license exempt profile operation. Additionally, it is currently working in the P802.16H/D8a proposal. This amendment provides measures to increase the efficiency and robustness of license exempt operation, specifies improved mechanisms as policies and medium access control enhancements and facilitate the coexistence of such systems with primary users.

The IEEE802.16 supports public safety first responders, military and emergency services such as call-prioritization, pre-emption and push-to-talk. (P802.16M2007). The involved tools to obtain this feature are the contention-based and the allocation-based radio resources.

Currently, WiMAX Forum plots a mobile profile in the 700MHz band, the recently identified by the ITU as the digital dividend spectrum; this spectrum is starting to be abandoned by TV broadcasters moving to all digital delivery, all around the world. WiMAX-700 specifications are already concluded (WiMax Forum, 2008B). The activities related to the operation of TDD and FDD in the 700MHz band are included in the development of the Release 1.5 profile. The certification of Release 1.5 based products, IEEE802.16REV2 compliant, is currently projected to begin in Q4 of 2009.

There are several interesting technical considerations from the railway low density context point of view regarding the 700MHz UHF frequency band:

- Lower path loss (26.5 dB lower than in the 2.5GHz or 3.5GHz bands)
- Lower Doppler Shift
- Better building penetration or indoor propagation to be taken into account inside the train composition and building stations.
- The enhanced signal processing reach (up to 65km) leads to the fact that the number of base stations necessary to cover the same area is about 10% of those at 3.5GHz profile. This represents an enhanced cost effective deployment when compared with 3.5GHz deployments.

A 700MHz PHY profile can compliment 2.5GHz and 3.5GHz networks. In dense environments such as the busy junctions situations, with higher subscriber density and high capacity demand small cells are necessary. In this case the optimal recommended deployment proposed is to deploy 700MHz umbrella cells for coverage and compliment then with 2.5GHz and 3.5GHz macro or micro cells to meet capacity requirements.

Most of the current implementations of IEEE802.16 in the access network are pre-WiMAX versions, such as the internet access provided along the 96km London-Brighton route with T-Mobile and Nomad Digital initiatives (Conti JP, 2005). There are also other research initiatives considering IEEE802.16 technology in the access network such as the European Research projects BOSS and also early studies like the one found in (Ritesh Kumar, 2008).

## 5. Conclusions

The access network between the train and the fixed network is definitely the most challenging one. Table 1 matches the railway requirements identified in Section 2 for each of the access technologies under study.

Nowadays, the most commonly used technologies are 2nd generation cellular-based (trackside) and satellite solutions. 2nd generation cellular systems, PMR included, and satellite solutions cannot be considered as an appropriate solution because either they are quite limited in bandwidth or suffer from an unacceptable delay and high cost.

The IEEE802.11 proposal for high speed vehicular scenario, IEEE802.11p, has a too limited coverage. Apart from that it is no mature enough. RoF and LCX solutions demand quite of an extended wired deployment. IEEE802.22 is not intended to provide support to mobile end users. HAPs do not support handover capability. The three technologies to be considered for the access network in the railway domain are HSPA, LTE and IEEE802.16.

4G technologies, LTE and IEEE802.16, are a step forward from HSPA technology when comparing data rate and latency features (Krapichler,C., 2007 ) , (Arthur D Little, 2007). Then, when considering LTE and IEEE802.16, mainly IEEE802.16m, both PHY layers are similar, OFDM and MIMO support and consequently there is no doubt about their capacity to cover the data rate and low latency needs of all emerging and future communication applications for the railway domain. Major LTE constraints are related to LTE maturity, cost and LTE operators' interest in low density, and consequently low return of investment, deployments.

There are some other identified features in the IEEE802.16 specification (mobile to mobile support, license exempt operation, public emergency services support and 700MHz profile), that currently contribute to consider the IEEE802.16 access technology as the best competitive access technology for the railway domain.

| | Satellite | GSM-R | HSPA | LTE | PMR | DVB-H | IEEE 802.11p | IEEE 802.16 | HAPs | IEEE 802.22 | IEEE 802.20 | RoF | LCX |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| High Speed Veh. Support | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | □ | ■ | ■ | ■ |
| HO capability | □ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | □ | □ | ■ | NA | NA |
| High Data Rate Support | □ | □ | ■■ | ■■■ | □ | ■ | ■■ | ■■■ | varies | ■■■ | ■■■ | ■■■ | ■ |
| Low Latency | □ | ■■ | ■■ | ■■■ | ■■ | ■ | ■■■ | ■■■ | ■ | ■ | ■■■ | ■■■ | ■■ |
| Wide Radio Coverage | ■■■ | ■ | ■ | ■■ | ■ | ■■■ | □ | ■■ | ■■■ | ■■■ | ■■ | NA | NA |
| Advanced Security Scheme | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | □ | □ |
| End2End QoS Support (MAC layer) | NA | NA | ■ | ■ | NA | NA | ■ | ■ | varies | ■ | ■ | varies | varies |
| Maturity | ■■■ | ■■■ | ■■ | □ | ■■■ | ■■■ | □ | ■■ | ■ | □ | □ | ■■ | ■■■ |
| Cost | □ | ■ | ■ | ■ | ■ | ■■ | ■■■ | ■■ | ■ | ■■ | ■■ | ■ | ■■ |
| Mesh Support | □ | □ | □ | □ | □ | □ | ■ | ■ | □ | ■ | □ | NA | NA |
| Licensed exempt operation | □ | □ | □ | □ | □ | □ | ■ | ■ | □ | ■ | □ | □ | □ |
| Wireless | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | ■ | □ | □ |
| Full stack E2E Arch. (all IP) | □ | □ | □ | ■ | □ | □ | □ | ■ | □ | □ | □ | □ | □ |
| Data Rate | Varies from few Kbps up to few Mbps | 9.6Kbps | Useful: 10.8/4.3 Mbps | Useful: 75/37.5Mbps | 7.2kbps | Downlink 30Mbps | 10-20Mbps | .16e Useful: 42/14Mbps .16m twice .16e | varies | 18Mbps | 16Mbps | few Gbps | 720 Kbps |

## 6. References

80216E2005, IEEE Standard for Local and metropolitan area networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment 2: Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1.

Arthur D Little, 2007. HSPA and Mobile WiMAX for Mobile Broadband Wireless Access. An Independent Report Prepared for the GSM Association.

Asenjo A, E González 2006 .Voz y datos digitales en el bus. *Comunicaciones world* , 58. 2006.

Berbineau,M, et all. 2006, High data rate transmission for high speed trains. Dream or reality?

Boch, 2006 . A Flexible Communication System for High-Speed Broadband Wireless Access from High Altitude Platforms: a CAPANINA Candidate.

Conti JP, 2005 Hot spots on rails. *Communications Engineer* 3[5], 18-21. 2005.

Gavrilovich Jr,CD, G C Ware, L L P Freiderich, C A , 2001 Broadband communication on the highways of tomorrow. *IEEE Communications Magazine* 39[4], 146-154. 2001.

InteGRail, 2008 InteGRrail: Intelligent Integration of Railway Systems. http://www.integrail.info/ 2008.

Ishizu,K, M Kuroda, H Harada, 2007 Bullet-train Network Architecture for Broadband and Real-time Access. Computers and Communications, 2007. *Proceedings of ISCC 2007.IEEE Symposium on , 241-248. 2007.*

ITU-R- M2072, 2006 World Mobile Telecommunication Market Forecast.

Krapichler,C., 2007 LTE, HSPA and Mobile WiMAX a comparison of technical performance. Hot Topics Forum: LTE vs WiMAX and Next Generation Internet, 2007 Institution of Engineering and Technology , 1-31. 2007.

Kuun,E, W Richard, 2004 Open Standards for CCTC and CCTV Radio-based Communication. *Alcatel telecommunications review* , 243-252. 2004.

Lannoo,B, D Colle, M Pickavet, P Demeester. 2003. Radio over fiber technique for multimedia train environment.

Lannoo,B, D Colle, M Pickavet, P Demeester., 2007. Radio-over-fiber-based solution to provide broadband internet access to train passengers. *IEEE Communications Magazine* 45[2], 56. 2007.

Lardennois,R., 2003 Wireless Communication for Signalling in Mass Transit. http://www.minatec.com .

p802.16m. 2006.IEEE 802.16m PAR.

p802.16m. 2007. Priority Access for IEEE 802.16m IEEE C802.16m-07/253.

p802.16m PR. 2007. IEEE 802.16m Performance Requirements, IEEE C802.16m-07/039, LG Electronics, Feb 2007.

Ritesh Kumar,K, P Angolkar, D Das, R Ramalingam, 2008. SWiFT: A Novel Architecture for Seamless Wireless internet for Fast Trains. Vehicular Technology Conference, 2008.*VTC-2008 Singapure IEEE 67th* . 2008.

Ruesche,SF, J Steuer, K Jobmann, 2008 The European Switch. *IEEE Vehicular Technology Magazine* 3[3], 37-46. 2008.

Vieira, P.MRS Logística switches to CBTC. *Railway Gazette* . 2009.

WiMAX Forum., 2007 WiMAX and IMT-2000.

WiMAX Forum, 2008, Deployment of Mobile WiMAX™Networks by Operators with Existing 2G & 3G Networks.

WiMAX Forum, 2008, B WiMAX Forum® Position Paper for WiMAX™ Technology in the 700 MHz Band.

Zhang,W, L Gui, W Ma, B Liu, J Xiong. Zhang, 2008 The television broadcasting network of Chinese High Speed Railway. *Proceedings of 2008 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* , 1-4. 2008.