A real-time MIMO-OFDM mobile WiMAX receiver: Architecture, design and FPGA implementation

O. Font-Bach,⇑, N. Bartzoudis, A. Pascual-Iserte, D. López Bueno

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Parc Mediterrani de la Tecnologia (PMT), Av. Carl Friedrich Gauss 7, 08860 Castelldefels, Barcelona, Spain

Article info

Article history:
Available online 30 March 2011

Keywords:
MIMO
Testbeds
IEEE 802.16e
Real-time systems
FPGAs
DSP

Abstract

The IEEE 802.16e-2005 standard, also denoted as mobile WiMAX, was introduced as one of the first real efforts towards the deployment of fourth generation communication systems providing fixed and mobile broadband wireless access. Mobile WiMAX supports multiple input multiple output (MIMO) antenna techniques which are considered a key technology in wireless communication systems for increasing both data rates and system performance. This paper presents a real-time 2×2 MIMO mobile WiMAX receiver with a detailed description of the architecture, design and implementation steps. The complexity of the real-time baseband signal processing has been scaled-up due to the high channel bandwidth that was adopted. Numerous equipment and instrumentation comprising our high performance experimental MIMO testbed were used to validate the operation of the mobile WiMAX receiver. The paper includes a subset of results that demonstrate the system-performance using standard 2×2 MIMO mobile channels.

1. Introduction

The multiple input multiple output (MIMO) technology using multiple antennas at both the transmitter and receiver sides is widely proposed as one of the key techniques to enhance the link quality and/or improve the spectrum efficiency of cellular systems. However, increasing the performance of multi-antenna mobile terminal devices implies the use of processing-intensive algorithms at baseband. The selected signal processing solutions should satisfy a trade-off between performance, numerical stability and hardware efficiency. This obviously results in tremendous challenges for real-time hardware implementations at baseband. The field programmable gate arrays (FPGAs) provide the necessary technology to deploy bit-intensive systems as long as the proposed algorithms are realistic for real-time implementations.

The ability to verify the benefits of new MIMO techniques or the performance of new communication standards based on orthogonal frequency-division multiplexing (OFDM)-MIMO schemes is an emergent goal of both academic and industrial research initiatives. Testbeds allow the validation of such developments in realistic environments accounting for hardware limitations and software or coding constraints. In this paper we present GEDOMIS® (GENeric hardware DemoNstrator for MIMO Systems) a multi-antenna wireless communication testbed that enables the prototyping and evaluation of MIMO physical layer (PHY) algorithms. GEDOMIS® is currently hosting a real-time implementation of the IEEE 802.16e standard (i.e., mobile WiMAX) on a 2×2 MIMO configuration.

⇑ Corresponding author.
E-mail addresses: ofont@cttc.cat (O. Font-Bach), nbartzoudis@cttc.cat (N. Bartzoudis), antonio.pascual@upc.edu (A. Pascual-Iserte), dlopez@cttc.cat (D.L. Bueno).
featuring a bandwidth of 20 MHz and using matrix-A encoding based on Alamouti’s space–time block code (STBC) [11] in a per carrier basis. The physical layer algorithms of the receiver were modeled in Matlab, designed in VHDL and implemented using a real-time FPGA platform.

The aim of our work is twofold; a principle objective is to present the baseband algorithms that are necessary to efficiently design the architecture of a real-time MIMO mobile WiMAX receiver; another important goal is the analysis of the design, implementation and debugging issues that have to be considered when embarking in the challenging task of building a real-life wireless broadband communication system.

2. Review of the state-of-the-art

The first wave of MIMO-OFDM technology is found in testbeds featuring the IEEE 802.11 family of standards, commonly referred as wireless local area network (WLAN) [2,3]. The 3rd Generation Partnership Project (3GPP) Long Term Evolution (LTE) has also embraced the MIMO-OFDM technology. A real-time 2 × 2 pre-LTE MIMO software-radio testbed implemented using FPGA and DSP technology is described in [4]. In [5] a real-time 12 × 12 MIMO-OFDM LTE testbed (with 20 MHz bandwidth), is implemented using cell-processors; however, its description does not reveal the underlying immature microprocessor technology in terms of available pre-verified functions, design software and coding peculiarities. Another LTE-based real-time 2 × 2 MIMO-OFDM system with 20 MHz bandwidth is presented in [6]. This work can be considered as a prime reference in the field of experimental research using MIMO testbeds, demonstrating the functionality of a base station (BS) and several mobile subscribers in an urban environment using a wideband channel emulator.

There are various other MIMO testbeds in the literature that make certain assumptions or simplifications to scale-down the processing complexity at baseband. For instance, [7] presents a FPGA implementation of a low bandwidth, real-time 2 × 2 MIMO system, using a software-generated indoor channel. Another indicative example is presented in [8] involving a low-bandwidth 2 × 2 MIMO testbed implementing offline baseband processing.

A well established wireless open-access research platform, namely WARP [9] has been adopted by various researchers. An indicative example is presented in [10] where the WARP testbed realizes a real-time OFDM-based cooperative system, using a distributed version of Alamouti’s block code, to analyze its capacity versus a 2 × 1 multiple input single output (MISO) system. Another initiative is the Vienna MIMO testbed [11] which appears in numerous recent publications like [12], where the PHY throughput of a 2 × 2 MIMO fixed WiMAX system featuring Alamouti coding is measured in an urban and an alpine scenario.

Several other fixed WiMAX implementations were encountered focusing on the characterization of certain channels with in-the-field test-campaigns. A 2 × 2 MIMO testbed featuring offline baseband processing is presented in [13]. The proposed system is used to evaluate the performance of differential STBC (DSTBC) transmissions over indoor channels. In [14] a real-time, low bandwidth 2 × 2 MIMO testbed, build with FPGAs, is used to carry-out channel capacity measurements. In [15] path loss measurements are conducted in a rural environment using a WiMAX-based network.

Although mobile WiMAX is currently considered a mature technology, there are still scarce sources in the literature, mainly presenting single antenna implementations [16] with offline baseband processing [17]. Real-time mobile WiMAX testbeds featuring MIMO are even more uncommon in the literature. We have only encountered partial FPGA-based implementations presenting ambiguous performance or results. For instance, in [18] a real-time 2 × 2 MIMO fixed sphere decoder (FSD) is implemented in a FPGA, while the rest of the system is developed in Matlab. In [19] a 2 × 2 MIMO system with 10 MHz bandwidth using a single cell processor, presents the IEEE 802.16e orthogonal frequency-division multiple access (OFDMA) PHY for a BS transceiver. Nevertheless, we could not efficiently evaluate this work due to the absence of mobility in the channels under test. Finally, in [20] the acclaimed FPGA-based transceiver implementation of a 2 × 2 MIMO system presents limited simulation and emulation results of a partially described system-architecture.

2.1. Contribution

The contribution of our real-time broadband MIMO-OFDM implementation compared to other mobile WiMAX physical layer implementations is explained in the following lines (comparison with the WLAN implementations is omitted, since their scope and operating requirements are different). Although our implementation is lacking the full scalability defined by the IEEE 802.16e-2005 standard, it is not merely an isolated hardware acceleration of specific PHY-layer algorithms (i.e., the mobile WiMAX implementations quoted previously); our solution is a prototype of the complete PHY-layer processing-chain of a MIMO-OFDM mobile WiMAX receiver. On top of this, our system was designed with the required modularity, which enables and facilitates the aggregation of scalable OFDM features (the limitations of our proposed solution in respect to the implementation overheads of the mobile WiMAX OFDM scalability are discussed in the following sections).

Furthermore, the system described herein is a real-time FPGA implementation capable of carrying out challenging baseband signal processing on-the-fly, while offering the required design modularity and structure to host a variety of advanced MIMO schemes (e.g., an open-loop MIMO configuration was implemented and presented herein). This can be considered an important operating difference when compared to the already presented non real-time implementations, which are easy, fast and cheap to deploy but feature as well significant downsides (e.g., processing of long data frames becomes unaffordable, closed-loop schemes cannot be directly evaluated).

A key difference of our work compared to that of other authors is the fact that their implementations are applying certain simplifications that facilitate their rapid
configuration parameters of the OFDM frame used in our subset of this flexible configuration of the PHY layer. The channel model).

It is also worth mentioning that both the current WiMAX IPs targeting mobile terminals and the encountered experimental MIMO mobile WiMAX testbeds, operate at 10 MHz. The system described in this paper is using a 20 MHz channel bandwidth, a state-of-the-art feature that exceeds the WiMAX Forum Radio Conformance Tests. It has to be underlined that this bandwidth forms part of mandatory implementation-specifications that will be introduced in the IEEE 802.16m version of the standard (i.e., in 2012).

3. Mobile WiMAX: outline of the physical layer specifications

The IEEE 802.16e-2005 standard, on top of various enhancements of its previous version regarding stationary operations, supports mobile subscriber stations at vehicular speeds and thus specifies a system for combined fixed and mobile broadband wireless access. The PHY layer of the mobile WiMAX supports scalable OFDMA architectures. The scalability is achieved by modifying the fast Fourier transform (FFT) size, a feature that facilitates the support of various channel bandwidths (i.e., from 1.25 to 20 MHz). Mobile WiMAX also supports adaptive modulation and coding (AMC), various subchannelization permutation techniques and MIMO-aided transmit/receive diversity. OFDM is used for both downLink (DL) and upLink (UL) transmissions.

In order to create the OFDM symbol in the frequency domain, the modulated symbols are mapped onto the subchannels that have been allocated for the transmission of the data block. A subchannel, as defined by IEEE 802.16e-2005, is a logical collection of subcarriers. The number and distribution of the subcarriers that comprise a subchannel depends on the permutation mode. The number of subchannels allocated for transmitting a data block depends on various parameters, such as the size of the data block, the modulation format, and the coding rate. There are numerous variations of permutation schemes defined in the standard. The OFDMA frame may include multiple zones, hosting different permutation schemes with partial usage of subchannels (PUSC) being the one utilized in our system. The 802.16e makes use of different MIMO techniques, such as STBC, beamforming and spatial multiplexing (SM).

Our system as indicated in Fig. 1b adopts only a fixed subset of this flexible configuration of the PHY layer. The configuration parameters of the OFDM frame used in our testbed (DL), are defining a single burst with a fixed predefined format (i.e., FCH and DL-MAP are not decoded) as seen in Fig. 1a. The selected encoding scheme is Alamouti's STBC (defined as matrix A in the WiMAX standard). The scalability of the OFDM-based mobile WiMAX standard is adds a top-up complexity in the control plane (i.e. different permutation schemes, variable length of the cyclic prefix (CP), different modulation scheme, etc.). However, it is important to underline that the bulk of the data plane processing challenges have been met by using the high bandwidth of 20 MHz. The latter dramatically increased the design and implementation considerations at baseband, since it implied additional processing complexity and memory requirements.

The first processing block of a WiMAX transmitter as defined by the standard is a randomizer that pseudo-randomly scrambles input data. The modulator maps data into constellation points and the subcarrier mapper allocates the symbols to the corresponding subcarriers. A preamble generator produces a symbol that precedes the burst, facilitating the timing synchronization. The pilot subcarrier mapper inserts pilot subcarriers into each data burst. The inverse FFT transforms the frequency-domain signal into a time-domain signal and a CP is inserted to obtain the complete OFDM signal. The standard does not define the exact structure of the receiver, which in all respects follows the reverse signal processing sequence from the one described in the signal transmitting stage.

3.1. MIMO technology

One of the most common problems faced by designers of wireless communication systems is the phenomenon of fading that arises due to the spatio-temporal variations of the wireless channel. This is inevitable in wave-reflecting and scattering environments that are subject to changes over time. The multiple received versions caused by reflections are referred to as multipath and can eventually produce a deep fade in the signal.

One of the main proposed techniques to tackle this effect is found in MIMO systems, which comprise multiple antennas at the transmitter and at the receiver sides. MIMO systems use diversity techniques to mitigate the effects of fading by providing multiple copies of the same signal. The use of multiple antennas dramatically reduces the probability of simultaneous deep-fades in all the receive antennas. MIMO technology may also be exploited to implement SM that significantly increases the spectral efficiency, and hence the capacity of a wireless communication system. SM realizes high data rates by transmitting independent information streams in parallel over different transmit antennas. Hence, MIMO technology features a trade-off between quality of service provided in diversity schemes and high data rates provided by SM [21].

There are different ways and transmission strategies to capitalize the benefits of diversity, which primarily depend

In addition to PUSC, the AMC permutation scheme has been implemented and validated only for the SISO mobile WiMAX receiver. Thus, the results are not included in this paper.
on the degree of knowledge of the channel response, i.e., the channel state information (CSI). In order to get such CSI at the transmitter, when channel reciprocity does not apply, a feedback channel from the receiver to the transmitter can be implemented.

An indicative MIMO transmission technique that does not require channel knowledge at the transmitter is space–time coding (STC), which utilizes both Block [22] (STBC) and Trellis [23] (STTC) codes. Alamouti’s block code, which benefits from its inherent orthogonality and uses 2 antennas at the transmit side, has become increasingly popular among other codes because of its optimal and low-complexity decoding stage at the receiver.

4. Receiver architecture and design

4.1. Received signal model

The WiMAX signal comprises frames which encapsulate user data and silence periods which are inserted between these frames. During the silence periods the receiver is continuously monitoring the incoming signal in order to detect the beginning of the following frame using a synchronization algorithm. In a real-world MIMO-OFDM testbed, the received radio frequency (RF) signal, on top of the noise, is impaired due to the performance characteristics of the equipment used (e.g., channel emulator, RF front-end, baseband signal processing boards, etc.). If the received subcarriers lose their orthogonality due to analog and RF impairments, the performance of the MIMO-OFDM system degrades dramatically. Thus, such signal-impairments have to be determined and removed before making the symbol decisions.

The specifications and performance of the equipment composing the GEDOMIS® testbed (as detailed in Section 5) allows our signal model to safely ignore certain negligible signal-impairments such as: the in-phase and quadrature (I/Q) gain and phase imbalances, the inaccuracy between the sampling clocks of the transmitter and receiver in respect to the ideal sampling frequency, the local oscillator (LO) drifts and finally, and the random phase noise due to LO instability. The resulting received signal model at the output of the RF down-converters at the ith receive antenna can be expressed as:

\[
c_t(t) = \Re\{x_t(t) \cdot e^{j2\pi f_{IF}t + \Delta f_t} + A_i + B_i \cdot \cos(2\pi f_{IF}t + \Delta f_t) + \varphi_i + w_i(t),
\]

where \(x_t(t)\) represents the useful part of the received baseband signal, \(f_{IF}\) is the intermediate frequency (IF), \(\Delta f_t\) is the carrier frequency offset (CFO), \(A_i\) is the direct current (DC) level introduced by the baseband board chassis, \(B_i \cdot \cos(2\pi f_{IF}t + \Delta f_t)\) represents the unwanted residual carrier, located at the center of the useful signal-spectrum (i.e., introduced by the LO coupling at the transmitter) and finally, \(w_i(t)\) is the Gaussian noise. The useful part of the received baseband signal at the ith receive antenna can be expressed as follows:

\[
x_t(t) = \sum_{j=1}^{n_i} \tilde{x}_j(t) \star H_{ij}(t),
\]

where \(\tilde{x}_j(t)\) is the equivalent baseband signal transmitted from the jth transmit antenna, with \(n_i\) being the number of assumed transmit antennas, and \(H_{ij}(t)\) is the equivalent baseband of the time impulse response of the MIMO channel between the jth transmit antenna and the ith receive antenna.

4.2. Design of the processing blocks at baseband

An indicative representation of the receiver processing blocks at baseband is shown in Fig. 2. The algorithmic foundation and functionality of each block consisting the mobile WiMAX physical layer will be explained in the following subsections. The entire system including the transmitter, channel and receiver was initially modeled in Matlab. Due to the bit-intensive nature of the physical layer algorithms utilizing the MIMO technology at the receiver and the real-time system-constraints, a suitable processing platform with FPGA devices was selected to implement the receiver. The selected Matlab-based algorithms were implemented by writing custom VHDL code following a register transfer level (RTL) design approach. The Matlab model and the VHDL implementation were co-simulated to verify the system precision and overall performance.

4.2.1. RF front-end and analog to digital conversion (ADC)

The tasks performed at the receiver’s RF front-end is the low-noise amplification, the downconversion from RF to IF
(i.e., in our case centered at $f_{\text{RF}} = 156.8 \text{ MHz}$) and the suppression of out-of-band unwanted signals such as noise and spurs.

Each active component in the receiver chain has a limited dynamic range. Thus, signals exceeding this range are subject to saturation or clipping. Since saturation is a detrimental factor of the system-performance, countermeasures should be taken to prevent it. Moreover, the active RF or baseband processing components are subject to thermal or other types of noise. The dynamic range in the baseband part of the receiver may potentially be affected by the presence of a DC offset. Static DC offsets occur due to bias mismatches in the baseband boards, but they could be generated as well by LO coupling at the RF transmitter or self-mixing of the LO signals at the RF receiver.

Sampling an analog signal at IF results in replicas of the signal’s spectrum which are repeated at uniform intervals. The choice of the sampling rate of such signals is dependent on the signal’s bandwidth and the IF center frequency. The chosen bandpass sampling architecture requires only one analog to digital converter (ADC) for the final IF to baseband conversion to occur in the digital domain. The ADC is performed using under-sampling and taking as ADC sampling rate $f_s = 89.6 \text{ MHz}$. Therefore, after the ADC, one of the aliases of the discrete signal will be located at 22.4 MHz, which is the baseband sampling frequency of the receiver as described by the WiMAX standard. The digital spectrum after the sampling is depicted in Fig. 3. The delta at baseband represents the DC coupling of the baseband hardware which has to be taken into account in the design of the baseband signal processing stages. The delta at the center of the signal spectrum represents the coupling of the analog LOs at both the up and down-converters.

4.2.2. Automatic gain control

The automatic gain control (AGC) is an analog–digital hybrid processing block providing an interface between the FPGAs and the RF front-end. The programmable gain amplifier PGA is a digitally-controlled analog circuit with a discrete set of possible gain values, while the algorithm that decides the new gain value of the PGA is implemented in the digital domain.

The correct operation of the AGC is a decisive factor for the overall performance of a mobile receiver. The AGC adjusts in a timely manner the power-level of the input IF signal to utilize the full dynamic range of the ADC and overcome the variations caused by the mobile channel fading. Frame-based OFDM systems are specially prone to high peak-to-average power ratio (PAPR); the inclusion of back-off margin that prevents signal clipping is therefore a prerequisite. The ADC device is indicating its saturation with a state signal. When saturation occurs the AGC does not forward data until the signal is attenuated at an optimal dynamic range.

The heart of the AGC algorithm is a signal peak-detector, operating in a per-frame basis (i.e., fixed gain for an entire frame), which provides a baseline trade-off between
implementation complexity and efficiency. In a frame-based communication system like mobile WiMAX, where the channel varies rapidly in high mobility conditions, the AGC algorithm has a very limited timing budget to operate. This is because the AGC must calculate the gain of the next frame and apply it to the PGA registers during the inter-frame silence period (i.e., taking into account the peak value of the previous data frame).

The PGA used in our receiver has 16 gain steps with 1.5 dB of separation (i.e., resolution of the gain corrections). Thus, starting from the gain value applied during the previous frame, the optimal adjustment of the IF input power level during the following frame, $\Delta G$, is calculated as follows:

$$\Delta G = 10 \cdot \log_{10} \left( \frac{v_{DS}}{v_{PK}} \right) \text{ dB} = 10 \cdot \log_{10} g \text{ dB},$$

where $v_{DS}$ is the digital full scale of the quantizer in the ADC, $v_{PK}$ accounts for the back-off safety margin, and $v_{PK}^2 = \max|c[n]|^2$, with $c[n]$ representing the samples at the output of the ADC during the previous frame. In order to minimize the processing complexity and the implementation latency, we have calculated the contents of a look-up table (LUT) that correlates all the possible values of $g$ in relation to the applicable gain-corrections (i.e., number of steps, $\Delta G$).

4.2.3. Digital down converter

The digital down converter (DDC) implements three functions: channel frequency translation, I/Q components extraction and signal decimation. The direct digital synthesizer (DDS) component of the DDC translates any frequency band within the analog bandwidth of the ADCs down to zero frequency (i.e., baseband), while a complex finite impulse response (FIR) low-pass filter is responsible for eliminating out-of-band components. Finally, an output decimator and formatter, which keeps one out of every four samples, delivers the complex representation of the digitalized signal whose spectrum is shown in Fig. 3.

The digital filtering stage of the DDC is important because it prevents aliasing during the sub-sampling process. Hence, it is critical to account for the system-wide signal impairments when designing the digital filter. For this reason, the bandpass and reject frequencies should be carefully selected keeping the useful signal spectrum intact while at the same time eliminating the effects of the DC level which is an inherent feature of the baseband processing boards (e.g., the DC is transformed to a synchronization-altering sinusoid which is eventually filtered). The designed low-pass filter has 103 coefficients and is implemented jointly with the decimation stage as a polyphase decimator filter. Fig. 4 shows the frequency domain representation of the operations performed within the DDC.

The output frequency of the DDS, $f_{DDS}$, is controlled by the phase increment, $\Delta \theta$, which is related to $f_{DDS}$ by:

$$f_{DDS} = \frac{f_s \cdot \Delta \theta}{2 \cdot 10^{6}} \text{ Hz},$$

where $f_s$ is the ADC sampling rate, as described in Section 4.2.1, and $B_{DS}$ is the resolution in bits of the internal accumulator used in the DDS (32 bits in our case). On power-up the $f_{DDS}$ is tuned to 22.4 MHz and then is constantly updated in real-time to compensate the effects of the CFO:

$$f_{DDS} = 22.4 + \Delta f \text{ MHz},$$

$\Delta f$ represents the CFO that is defined in terms of the separation between adjacent subcarriers:

$$\Delta f = \frac{22.4 \cdot 10^6}{2^{10} - 2^{6}} \text{ MHz},$$

where $\alpha$ is the CFO normalized with respect to the intercarrier separation (i.e., in practice the CFO will not be higher than one half the intercarrier separation, $\alpha \in [-0.5, 0.5]$). 22.4 MHz is the sampling frequency and 2048 is the FFT size. Combining (4)–(6) the phase increment is given by:

$$\Delta \theta = \left( 22.4 + \alpha \right) \frac{2^{12}}{89.6} = 2^{10} + \alpha \cdot 2^{19}$$

4.2.4. Synchronization, CFO estimation and correction

The WiMAX frame comprises several OFDM symbols; in our system-configuration 46 symbols are used for the data each one having 2560 samples. Symbol detection is required to properly locate the FFT window of the samples corresponding at each OFDM symbol. This is feasible with the inclusion of a CP at the beginning of each OFDM symbol. Considering channels with a maximum delay spread of 2510 ns, only 455 out of the 512 samples in the CP can be used for the timing synchronization (i.e., the remaining 57 samples are discarded to avoid unreliable operation of the FFT window-locator). The implemented synchronization technique is based on a sliding window of 2048 + 455 samples, which allows us to calculate the cross-correlation of two groups of 455 samples (having a separation of 2048 samples). The expression corresponding to the square of the correlation when the sliding window starts at the nth sample is given by:

$$|r[n]|^2 = \frac{\sum_{l=-\infty}^{\infty} \sum_{l \neq 0}^{454} s[n+l] \cdot s[n+l+2048]^2}{\sum_{l=-\infty}^{\infty} \sum_{l \neq 0}^{454} |s[n+l]|^2 \cdot (\sum_{l=-\infty}^{\infty} \sum_{l \neq 0}^{454} |s[n+l+2048]|^2)}.$$
where \( s_i[n] \) is the equivalent complex baseband signal at the output of the DDC, sampled at 22.4 MHz, at the \( i \)th receive antenna processing branch, and \( n_R \) denotes the number of receive antennas (i.e., \( n_R = 2 \) in our case). Due to the extremely resource-demanding implementation imposed by (8) the following simplification in terms of complexity was applied:

\[
|r_s[n]|^2 = \frac{|dn[n]|^2}{ds0[n] \cdot ds1[n]},
\]

where:

\[
dn[n + 1] = \begin{cases} 
    dn[n] + \sum_{i=1}^{n_R} s_i'[n + 455] \cdot s_i[n + 2048 + 455] & \text{if } n \leq 455, \\
    dn[n] - \sum_{i=1}^{n_R} s_i'[n] \cdot s_i[n + 2048] + \sum_{i=1}^{n_R} s_i'[n + 455] \cdot s_i[n + 2048 + 455] & \text{if } n > 455,
\end{cases}
\]

with \( dn[0] = 0 \). It should be noted that \( ds0[n], ds1[n] \) are calculated in a similar manner. With this optimization only four samples need to be introduced to the already calculated correlation. A peak in \( |r_s[n]|^2 \), indicates the detection of the symbol and thus the sample where the CP starts, i.e., \( pos_{cp} = \arg \max_n |r_s[n]|^2 \). Additionally the phase of the correlation (i.e., the numerator of \( |r_s[n]|^2 \)), at \( pos_{cp} \) can be used to estimate the phase shift of the received signal in the presence of CFO. Using the notation given in (6), the phase shift between two signal samples delayed by 2048 positions is equal to \( e^{j2\pi f_t \cdot \frac{1}{22.4M} \cdot \frac{1}{2048}} = e^{j\alpha P} \). Therefore, the estimated CFO or, equivalently, the constant \( \alpha \) can be defined as:

\[
\alpha = \frac{1}{2\pi} \angle (r_s[pos_{cp}]) = \frac{1}{2\pi} \angle (dn[pos_{cp}]),
\]

where \( \alpha \) can be calculated using a coordinate rotation digital computer (CORDIC) algorithm. As already mentioned, the estimated CFO is used to fine tune the DDS.

The architecture of the proposed synchronization scheme is shown in Fig. 5. Due to the stringent real-time constraints of our system, a pipelined structure has been deployed requiring a latency of more than one clock cycle. First in first out (FIFO) memories provide a latency-leveler temporary storage for the incoming signal. A custom design for such memories allows the retrieval of the four specific samples employed to calculate the cross-correlation at each clock cycle. Additionally, once the location of the first sample of each OFDM symbol is determined by the data-forwarding control logic, the window of subcarriers com-

---

**Fig. 4.** Frequency representation of the operations performed in the DDC.

**Fig. 5.** Architecture of the synchronization block.
posing this symbol is synchronously forwarded from the FIFOs.

The detection of the correlation peak is a critical part of the synchronization algorithm because it is error-prone under the presence of spurious signals. In our system, we experience a parasitic sinusoid because of the presence of CFO drifts and due to the DC-level created by the digital mixing of the unwanted residual carrier with a digitally generated sinusoid at the DDS stage. The presence of this sinusoid during the silence period can result in erroneous performance of the aforementioned correlator, which may indicate the presence of peaks and consequently misplace the window of samples forwarded to the FFT processing block. Failure to prevent an erroneous symbol detection may render the system unusable. Thus, the peak detection algorithm, which is usually based on a triggering threshold and the selection of the maximum value in a window, must be optimized to recognize the legitimate peaks. When the trigger issues a correlation value above the threshold, the shape of the correlation curve determines whether the located peak indicates the beginning of an OFDM symbol or a silence period. The correlation curve tends to have high-values and nearly no variations during the silence periods (Fig. 6); on the other hand it presents high values only during the processing of an OFDM symbol while processing the CP (Fig. 7).

4.2.5. Pilot extraction and channel estimation

The channel estimation in our receiver is based on the pilot subcarriers that are being transmitted in each OFDM symbol. Let \( S_i[k] \) be the \( k \)th subcarrier in the OFDM symbol received by the \( i \)th receive antenna after the FFT (i.e., \( S_i[k] = \text{FFT}(s_i[n]) \)), with \( k \in [0 \cdots n_U - 1] \), where \( n_U \) represents the number of subcarriers used to transmit user data and pilot tones. The IEEE 802.16e standard defines the value, \( p_v = \frac{4}{3} \) and location (i.e., frequency), \( p_{k,j} \), of such special subcarriers for each transmit antenna \( j \), i.e., \( p_{k,j} \in [0 \cdots n_U - 1] \). The number and distribution of pilot subcarriers depends on the subcarrier permutation scheme. When the PUSC permutation scheme is used, clusters of 14 contiguous subcarriers are defined, with two of them used to transmit pilot tones. Additionally, out of the 2048 subcarriers available at each OFDM symbol, 1440 will be used for data transmission and 240 for pilot tones transmission, i.e., \( n_U = 1679 \), while the rest are being utilized for the guard-bands and DC carrier. When \( n_T = 2 \) the PUSC permutation scheme distributes the pilot tones for each antenna in two consecutive OFDM symbols; this imposes an implementation constraint of storing two complete OFDM symbols per receive antenna enabling in this way the pilot-based channel estimation. This also implies that the channel estimation will be applied in pairs of consecutive OFDM symbols (i.e., the estimated channel frequency response will be the same for both). Fig. 8 shows the detailed cluster structure, which is cyclically repeated each four OFDM symbols, where \( O_n \in [0 \cdots 23] \) is the index of the transmitted OFDM symbol pair within the frame and \( T_x \) represents the \( i \)th transmit antenna. Therefore, the cluster structure defines \( p_{k,j} \) for each OFDM symbol and each transmit antenna \( j \). Note that when \( S_1[k] \) is used as a pilot tone then no transmission occurs for \( S_2[k] \) (i.e., null subcarrier) and vice versa, to avoid interferences in the pilot positions.

Each processing branch of the MIMO-enabled receiver has to estimate the corresponding channels from all
transmit antennas. First, the channel frequency response at the pilot tones, $H_{ij}[p_{kj}]$, is estimated as follows:

$$H_{ij}[p_{kj}] = \frac{S[i][p_{kj}]}{S_{ij}},$$

where $S[i][p_{kj}]$ represents the $k$th pilot tone from the $j$th transmit antenna after the FFT in the $i$th receive antenna processing chain, with $j \in [1,2]$ in our case. Thus, $H_{ij}[p_{kj}]$ is a discrete function calculating the channel frequency response at the pilot tones between the $i$th receive antenna and the $j$th transmit antenna. An interpolation of the pilot positions is then required to estimate the channel at the frequencies where data subcarriers were transmitted for each transmit-receive antenna pair. After comparing different algorithms, we have selected a second order polynomial interpolation, which provides the best trade-off between accuracy and implementation complexity (accounting for the number of pilot tones in each OFDM symbol and the channel specifications). The channel frequency response for the data subcarriers is calculated as follows:

$$H_{ij}[k] = H_{ij}[p_{cj,i}] + \frac{H_{ij}[p_{cj,i}] - H_{ij}[p_{cj,j}]}{p_{cj,j} - p_{cj,i}} \cdot (k - p_{cj,i})$$

$$+ \frac{H_{ij}[p_{cj,j}] - H_{ij}[p_{cj,j}]}{p_{cj,j} - p_{cj,j}} \cdot (k - p_{cj,i}) \cdot (k - p_{cj,j}),$$

where $p_{cr,j}$ represents the location of one of the three closest pilot tones, originated from transmit antenna $j$, to $S[i][k]$, respectively for each $r \in [1 \ldots 3]$.

The general processing architecture of the channel estimation is depicted in Fig. 9. The subcarriers are stored in two separated memories, considering that the channel estimation only commences when all the pilot carriers of two OFDM symbols are received. First, the channel frequency response at the pilot tones is estimated. A carefully designed memory system is required to store the calculated coefficients, until the whole $H_{ij}[p_{kj}]$ is computed. The calculation of the estimated channel frequency response coefficients is implemented using groups of three neighboring pilots (i.e., $H_{ij}[p_{kj}]$); their locations within the memory will be indicated by their associated indexes (i.e., FFT output index of the pilot tone), and replicated in three memory blocks, avoiding in this way extra latencies and computational complexity in the memory-management plane. Once $H_{ij}[p_{kj}]$ is available the computation of the interpolation is performed. It must be noted that incoming subcarriers have lost their sequential order for each data subcarrier after the FFT calculation, therefore an algorithm is required to calculate the indexes of the three closest pilot subcarriers, $p_{cr,j}$, accounting for the position of the actual carrier at the output of the FFT. A FIFO memory component is used to compensate the latency introduced by the calculations and provide an aligned output.

![Fig. 8. Distribution of the pilot subcarriers in the PUSC permutation.](image)

![Fig. 9. Architecture of the channel estimation block.](image)
4.2.6. Matrix A space–time decoding

The Alamouti’s STC (matrix A) transmission scheme encodes data symbols, $d_k$, in pairs and distributes them in groups of two OFDM symbols. In order to decode and estimate the transmitted data symbols the following operations are applied:

$$d_k[2O_n] = \frac{1}{c} \sum_{i=1}^{n} H_1[k] \cdot S_i[k, 2O_n] + H_2[k] \cdot S_i[k, 2O_n + 1]$$

$$d_{k+1}[2O_{n+1}] = \frac{1}{c} \sum_{i=1}^{n} H_1[k] \cdot S_i[k, 2O_n] - H_2[k] \cdot S_i[k, 2O_n + 1]$$

where $O_n$ and $O_n + 1$ represent the indexes of the two consecutive OFDM symbols. Note that in (14) and (15) it is assumed that the gain applied by the AGC to the incoming sample-streams is equal for both receive antennas.

Taking into account the implemented channel estimation for the data subcarriers it is required to store one OFDM symbol per receive antenna before applying on-the-fly the space–time decoding during the reception of the second OFDM symbol of the pair. In other words, four data symbols will be estimated at each operation of the block, with an initial latency equal to the length of an OFDM symbol. It is worth mentioning that there is no need to store the channel coefficients for the first OFDM symbol of the pair, because they are the same as the posterior one (i.e., the pilot tones are distributed in two OFDM symbols).

The processing stages that follow the space–time decoding calculations are related to the PUSC permutation, clustering, channelization and mapping of the data symbols, which enable the recovery of the originally transmitted bit sequence.

5. GEDOMIS® testbed description

A graphic-overview of the GEDOMIS® testbed setup for a point-to-point $2 \times 2$ MIMO system is shown in Fig. 10. The baseband part of the transmitter was designed in MATLAB. The separate I/Q baseband outputs of this model are written to data-files, which are fed to two instances of Agilent’s Signal Studio Toolkit. The data-files are then downloaded to two ESG4438C instruments configuring in this way the operating parameters of the transmitter. Since the two ESG4438C need to be time and phase aligned for the MIMO signal generation, several adjustments were applied (i.e., master–slave connection of the instruments, 2)

The testbed uses a $2 \times 2$ MIMO scheme applying an Alamouti’s STBC (matrix A). Increasing the number of antennas would only imply a linear increase of the complexity of the space–time block decoding (with respect to the number of antennas), as long as an orthogonal STBC is still applied at the transmitter. Moreover, our testbed has certain hardware constraints that prevent the implementation of higher order MIMO schemes (i.e., FPGA capacity, maximum number of simultaneous channels that can be emulated in our channel emulator and absence of the required instrumentation for generating a MIMO signal having more than two antennas).

Implementing more permutation schemes would merely involve the design of additional blocks with control logic whose complexity in terms of implementation and cost is much lower than the remaining of the system detailed in this paper.

2 The testbed uses a $2 \times 2$ MIMO scheme applying an Alamouti’s STBC (matrix A). Increasing the number of antennas would only imply a linear increase of the complexity of the space–time block decoding (with respect to the number of antennas), as long as an orthogonal STBC is still applied at the transmitter. Moreover, our testbed has certain hardware constraints that prevent the implementation of higher order MIMO schemes (i.e., FPGA capacity, maximum number of simultaneous channels that can be emulated in our channel emulator and absence of the required instrumentation for generating a MIMO signal having more than two antennas).

3 Implementing more permutation schemes would merely involve the design of additional blocks with control logic whose complexity in terms of implementation and cost is much lower than the remaining of the system detailed in this paper.
time alignment of the two signals, etc.). Finally, the two ESG4438C are utilizing their embedded arbitrary waveform generator to playback in real-time the baseband I/Q waveforms, up-convert the signal and finally provide the RF output centered at 2.595 GHz.

The Elektrobit Propsim C8 radio channel emulator, allows accurate multi-channel emulation of custom or standardized models in the laboratory. This is feasible by adding complex and time-varying effects of multipath and Doppler-shifts in the digital domain (e.g., adjusting the tap amplitude, delay spread, operation frequency and mobile speed). For the 2 × 2 MIMO scenario considered in this paper, four uncorrelated multipath fading channels were created (with different distribution seeds), using either the ITU Vehicular A or the ITU Pedestrian B channel model [24]. The channel emulator has to be tuned to provide optimal performance in terms of noise floor, dynamic range and error vector magnitude (EVM), allowing sufficient safety margin for its operation. This is necessary to avoid signal-distortion, degradation of the received signal to noise ratio (SNR), saturation of the ADCs and DACs (the signal PAPR has to be accounted for) and quantization errors.

At the receiver side, the Mercury Computer Systems Echotek Series RF 3000 Tuners comprised by 1 DDS module and 4 receiver modules, apply phase-coherent down-conversion from an RF signal of 2.6 GHz to an IF frequency of 156.8 MHz featuring high spectral purity and dynamic range. A prototyped SAW filter board is used to optimally match the signal bandwidth at IF and thus reduce the out-of-band noise level and eliminate spurious effects introduced by the channel emulator. Finally, two broadband RF noise generators provide extremely flat white noise at IF and facilitate the measurement campaign and the assessment of the system’s performance under variable SNR conditions (e.g., the two noise sources were calibrated and balanced for every 2 dB attenuation step).

The signal is then delivered to the signal acquisition and processing development platform equipped with Lyrtech’s ADC and FPGA/DSP boards. The VHS-ADC board includes 8 phase synchronous channels with 14-bit ADCs and digitally-controlled PGAs for each channel. The board also includes a Xilinx Virtex-4 FPGA that hosts the AGC, the DDC and synchronization processing blocks of the receiver. The baseband signal (20 MHz, 22.4 Msps) is then fed to the SignalMaster Quad board that features two clusters of a Xilinx Virtex-4 FPGA and two TSM320C6416 DSPs capable hosting the rest of the MIMO mobile WiMAX demodulation blocks.

6. Measurement and results

As already mentioned, the MIMO-enabled mobile WiMAX receiver was fitted in two Virtex-4 LX160 devices, having the following resource utilization: 81% of slices, 93% of RAMB16s and 100% of DSP48 for the first FPGA and 49% of slices, 71% of RAMB16s and 57% of DSP48 for the second one (i.e., using the Xilinx ISE 9.2). The real-time debugging of the receiver and data visualization was made using the Xilinx ChipScope Pro.

The results presented in this section were acquired after an extensive measurement campaign and a post-processing of the captured data. Matlab scripts and a parser were developed for this reason. The system-performance has been evaluated in terms of the EVM and the raw bit error rate (BER). Obviously, the inclusion of a channel coding technique would significantly improve the system performance by lowering the final BER values in respect to the ones presented herein.\footnote{Channel coding is a standard processing block that has similar implementation issues with other broadband wireless communication standards. Taking into account the previous argument, the focus of the contribution of this paper and the limitations of our hardware platform, we have omitted the inclusion of channel coding in our system.} The real-time channel emulator was configured with two MIMO channel models that form part of the radio conformance tests of WiMAX forum for mobility scenarios (i.e., ITU-T vehicular A (60 km/h) and pedestrian B (3 km/h) channels).

The obtained results allow the quantification of the performance losses caused by the hardware deployment of...
the signal processing algorithms. The co-simulation results (i.e., Matlab versus testbed data) revealed implementation losses of approximately 3 dBs (e.g., quantization, finite bit representation, etc.), a figure that can be considered acceptable taking into account that the entire hardware setup is also contributing to these losses. In Fig. 11a and b a single static realization for both random channels (i.e., no mobility is emulated) has been used.

Finally, in Fig. 12a and b the same measurements are repeated for 100 different realizations of the random channel (i.e., using a different channel seed), applying 6 attenuation steps of the additive white Gaussian noise (AWGN) generators per channel for each of the vehicular A and pedestrian B models. The curves are produced by averaging the 100 data captures for each of the 6 attenuation steps. This method of obtaining the results allows an accurate analysis of the receiver performance under different mobility scenarios.

As could be expected, the performance of the system using the pedestrian B channel model is higher than the one presented in the vehicular A case. This is due to the fact that in the first case the mobile speed is significantly lower, allowing an improved tracking of the channel variations by the AGC algorithm (i.e., the power level of the input signal is better accommodated to the dynamic range of the ADCs). As far as the diversity is concerned, the theoretical order is expected to be 4 however, the average BER curves obtained in the real implementation tend to have a slope of 2 (i.e., diversity order equal to 2). This reduction in the diversity order with respect to the theoretical calculated case is based mainly on two reasons. First, the theoretical results generally assume that the receiver tracks the channel perfectly, which is not realistic in a real-life implementation. Second, the floating point logic dominating the proposed MIMO algorithms is subject to quantization errors when it is transformed to fixed point logic; truncation and finite bit representation was indeed a constraint in our FPGA-based implementation. Nonetheless, the obtained results allowed us to demonstrate, validate and measure the actual performance of a real MIMO system implemented in hardware. Indeed the performance profile is within the expectable margins and constitutes the prelude of a testbed that can host real 4G systems.

6.1. Power consumption

The FPGA-based design of the MIMO mobile WiMAX receiver presented in this paper did not follow an energy-efficiency design-path (i.e., this is not part of the scope of this paper). However, it is useful to include preliminary power-consumption metrics to enable a relative assessment of our prototype’s power-consumption footprint (power optimizations will be part of our future work). It is important to take into account that the target FPGA device does not belong to a power-efficient family of Xilinx, while at the same time the power-reduction margin that could be achieved by the Xilinx ISE 9.2 tool is very limited. The Xpower software tool of Xilinx has been used to estimate the PHY-layer power consumption of the prototyped MIMO mobile WiMAX receiver.

The conducted analysis accounts for both quiescent and dynamic power dissipation.² Obviously, the design decision to divide the PHY-layer implementation of the mobile WiMAX receiver in two FPGA devices, plays a major role in the presented power-consumption metrics. The first part,

<table>
<thead>
<tr>
<th>Table 1</th>
<th>FPGA-deployed MIMO baseband power consumption estimation.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption</td>
<td>Quiescent (W)</td>
</tr>
<tr>
<td>FPGA-1</td>
<td>1.22</td>
</tr>
<tr>
<td>FPGA-2</td>
<td>1.13</td>
</tr>
</tbody>
</table>

The power consumed when no signal-switching occurs is defined as quiescent, while the dynamic power consumption represents the accumulated power dissipation of the operating components comprising the FPGA design.
nearly the digital front-end (from the AGC up to the synchronization processing block) is always operating, while the second part (from the CP-removal up to the de-mapping processing block) is not operating during the silence period between data frames, which can be considered as the system’s idle state. As may be observed in Table 1, when the system is in idle state, the second part of the design (noted as FPGA-2) will only present quiescent power consumption, achieving a 14% power consumption savings. The presence of multiple clock regions in FPGA-1 increases the overall power consumption. As expected, the power consumption in idle state is quite high due to the inherent high power-leakage of the FPGA Virtex-4 technology. Mapping our design to an application-specific integrated circuit (ASIC) implementation could dramatically reduce this metric.

7. Conclusions and future work

A designer of real-time MIMO systems is confronted by various conventional and other less predictable software and hardware issues during modeling, implementation and debugging time. This paper presented the deployment of the complete PHY-layer of a real-time MIMO mobile WiMAX receiver. The overall system validation under realistic channel mobility emulation was facilitated by using the GEDOMIS™ testbed. The presented work covers an in-depth analysis of the receiver’s architecture, design and FPGA implementation, especially focusing on the computational and deployment complexity of this undertaking. The system-design conceptualization followed a modular approach that allows for more OFDM or MIMO schemes to be implemented in the future. The contribution of our work is mainly found in building and validating a realistic MIMO system based on the mobile WiMAX standard; the 20 MHz channel bandwidth and the real-time implementation have scaled the implementation complexity, facilitating at the same time the experimentation of advanced research concepts.

The benefits of different multi-antenna schemes when compared to single antenna implementations could be investigated in the future by applying minor changes to the currently available implementation (i.e., SIMO and MISO). More results could be obtained in the future with the inclusion of channel coding processing blocks. Finally, a real-time closed-loop MIMO mobile WiMAX system could be compared with our existing open-loop implementation, by deploying an FPGA-based MIMO transmitter.

References


O. Font-Bach (Terrassa, 1981) received his M.Sc. degree in Computer Engineering from Universitat Autònoma de Barcelona (UAB, 2004). He also obtained a Master in Design of Integrated Circuits from UAB (2006), as part of a postgraduate internship in the Institut de Microelecònica de Barcelona (IMB-CNM CSIC) where he worked in the design automation of AMBA-based NoC systems for FPGA-based implementations. He is currently a Research Engineer at the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) working in the implementation of real-time baseband signal processing algorithms for high-bandwidth MIMO-OFDM(A) systems. He is also pursuing his Ph.D thesis at the Department of Signal Theory and Communications of Universitat Politècnica de Catalunya (UPC). His research interests focus on baseband digital signal processing for multi-carrier communication systems.

N. Bartzoudis (Alexandroupoli, Greece 1976) obtained his B.Sc. in Electronic Engineering at the Technical Educational Institute of Thessaloniki (Greece, 2000). He then pursued postgraduate studies and received his M.Sc. degree (in Digital Communication Systems) and Ph.D. degree (in dependable embedded systems) from Loughborough University (UK) in 2001 and 2006 respectively. He has also worked as a post-doctoral researcher at the University of Essex (UK). Nikolaos is currently holding a Research Associate position at CTTC and he is in charge of the GEDOMIS® testbed. His core working area focuses on baseband digital signal processing implementations of wideband multi-carrier communication systems. Nikolaos has served as TPC member in various international conferences and he is a regular reviewer for numerous scientific journals.

A. Pascual-Iserte (Barcelona, 1977) received the Electrical Engineering degree and the Ph.D. degree from the Technical University of Catalonia (UPC), Barcelona, in September 2000 and February 2005, respectively. He was awarded with the “First National Prize of 2000/2001 University Education” by the Spanish Ministry of Education and Culture, and with the “Best 2004/2005 PhD Thesis Prize” by UPC. In September 2003 he became Assistant Professor at the Escola d’Enginyeria de Telecomunicació i Aeroespacial de Castelldefels of the UPC, where he teaches currently undergraduate courses in Linear Systems and Digital Signal Processing. He also teaches post-graduate courses in Advanced Signal Processing in the Dept. of Signal Theory and Communications. Since April 2008 he is Associate Professor at UPC. His current research interests include: array processing, robust designs, OFDM, MIMO channels, multiuser access and optimization theory.

D. López Bueno (Granollers, 1978) received the Electrical Engineering degree from Universitat Politècnica de Catalunya (UPC) in 2005. From 2003 to 2006 he worked as RF designer, first in Mier Comunicaciones and later in Thales Alenia Space España, participating in several Space projects like Amazonas 1, TRMO, Galileo and ICO G1. By the end of 2006, he joined CTTC as Research Engineer where he works in the RF hardware development and test and measurement operations of GEDOMIS testbed (MIMO wireless system demonstrator). He is also working in interference risk assessment between satellite and terrestrial communication systems, and currently pursuing a PhD in Electronics Engineering at UPC. His main interests include the architecture evaluation and design of high-end spectrum sensing and agile RF Transceivers to tackle next generation wireless applications.