Radio-communications Architectures
Antoine Diet, Martine Villegas, Geneviève Baudoin, Fabien Robert

To cite this version:

HAL Id: hal-00553229
https://hal.archives-ouvertes.fr/hal-00553229
Submitted on 7 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Radio-communications architectures

Antoine Diet*, Martine Villegas**, Geneviève Baudoin** and Fabien Robert**

*Université Paris-Sud 11, DRÉ, UMR 8506
**Université Paris-Est, ESYCOM, ESIEE Paris
France

1. Introduction

Wireless communications, i.e. radio-communications, are widely used for our different daily needs. Examples are numerous and standard names like BLUETOOTH, WiFi, WiMAX, UMTS, GSM and, more recently, LTE are well-known [Baudoin et al. 2007]. General applications in the RFID or UWB contexts are the subject of many papers. This chapter presents radio-frequency (RF) communication systems architecture for mobile, wireless local area networks (WLAN) and connectivity terminals. An important aspect of today’s applications is the data rate increase, especially in connectivity standards like WiFi and WiMAX, because the user demands high Quality of Service (QoS). To increase the data rate we tend to use wideband or multi-standard architecture. The concept of software radio includes a self-reconfigurable radio link and is described here on its RF aspects. The term multi-radio is preferred. This chapter focuses on the transmitter, yet some considerations about the receiver are given. An important aspect of the architecture is that a transceiver is built with respect to the radio-communications signals. We classify them in section 2 by differentiating Continuous Wave (CW) and Impulse Radio (IR) systems. Section 3 is the technical background one has to consider for actual applications. Section 4 summarizes state-of-the-art high data rate architectures and the latest research in multi-radio systems. In section 5, IR architectures for Ultra Wide Band (UWB) systems complete this overview; we will also underline the coexistence and compatibility challenges between CW and IR systems.

2. Transceiver aspects for radio-communications

2.1 Radio communications signals

Radio-communications applications deal with communicating and non-communicating links with their different parameters. People expect high quality from their different services (QoS) whatever the telecommunications system used. For example, voice (low data rate) or visio-phone and multimedia download (high data rate) are assumed to be present on the new generation mobile phone. This reveals the co-existence and interaction goals between mobile communication systems (GSM, GSM EDGE, UMTS…) and connectivity standards (BLUETOOTH, WiFi, WiMAX…) [Baudoin et al., 2007]. Thanks to Impulse Radio
Ultra Wide Band (IR-UWB), other fields of interest such as Radio Frequency Identification (RFID) and localization systems are examples of where radio-communications transceivers are currently being designed. Each kind of application can be classified by the resulting radio signals emitted/received. Determining factors are (i) the use of power efficient or spectrum efficient modulation schemes, (ii) the frequency and type of carrier signal used: Continuous Wave (CW) or Impulse Radio (IR) based signals and (iii) the data rate needed (defining a major subdivision of CW based systems). Depending on the choice of the factors involved, the design faced by the RF architect can be varied and challenging. We differentiate three types of cases in this chapter, as illustrated in Fig. 1.

1) NarrowBand CW (NB-CW) systems, like GSM (GMSK), EDGE (D-QPSK), BLUETOOTH, RFID tags, etc... These systems are often power efficient modulation schemes (EDGE is an exception) because these applications use a low data rate transfer. Major problems of the NB-CW architecture involve with the coexistence and the signal protection against interferers or blocking signals. Some spread-spectrum techniques are often added to improve the communication range (frequency-hopping in the case of BLUETOOTH). The NB-CW case is also considered to be the classic radio communications link and reference system because it corresponds to the popular AM/FM radio broadcasts. NB-CW systems are often considered as using constant envelope signals. This could be true only if FSK modulation schemes are performed, but it is also possible to use x-QAM low symbol rate modulations (EDGE). Additionally, the shaping filter (root raised cosine filter) implies amplitude variations (non-constant envelope) even on FSK modulated signals (GSM).
2) WideBand CW (WB-CW) systems, like: UMTS (W-CDMA), WiMAX (OFDM enhanced), LTE (OFDM based), WiFI (OFDM), UWB (OFDM version)... These systems correspond to high data rate transfer, often for multimedia applications. We can see an increasing need in this field of interest due to the large number of new standards which can be found. Due to the bandwidth limitations for each standard, the modulation scheme is often spectrum efficient (x-QAM) and the use of multi-carrier transmission is usually performed, e.g., OFDM or MC-CDMA. Since there are high amplitude variations of such signals, the transceiver architecture is designed in function with the unavoidable non-linear effects (NL) caused by the power amplification block. Emitting a high PAPR (Peak to Average Power Ratio) has always been a well-known transmitter challenge. The linearization of such a transmitter is mandatory and will be discussed in part 3. WB-CW transceivers also need a wideband design for all of its key elements: antennas, LNA (Low Noise amplifier), HPA (High Power Amplifier) and mixers. This often results in lower performance of the above-mentioned blocks than for NB-CW systems.

3) IR-UWB systems such as UWB localization systems, RFID-UWB... These systems are special because they are based on (one of the) spread spectrum techniques in order to protect the information. The idea is to spread the information in frequency while lowering the emitted power. The use of UWB (3.1-10.6 GHz in the USA) was highly discussed in order to evaluate its co-existence with NB-CW and WB-CW. What is important here is how the power amplification is processed differently for this type of communication. An average power is defined in function of the Pulse Repetition Frequency (PRF), given by the specifications of the IR-UWB standard. For a fixed emitted power: the shorter the pulse, the higher the instantaneous power and bandwidth. Transceivers for these signals are based on impulse generators and energy detectors or correlation receivers, see section 5.

![Fig. 2. Telecommunications spectrum sharing and power limitations up to 6 GHz](image)

The three types described can represent every kind of radio-communications signal. The characteristics which have an impact on the architecture are mainly the centre frequency (choice of the technology) and bandwidth (circuits' topologies and performance limitations), and also the PAPR for CW signals. Fig. 2 qualitatively summarizes the power limitations and frequency specifications for some telecommunications standards up to 6 GHz: cellular, WiMAX, WiFi, Bluetooth, UWB and DVB-H. The goal of RF architecture is to emit and receive such signals with no alteration of the information (no constellation distortion for CW). Designing the architecture implies other considerations such as noise, linearity, efficiency and
systems co-existence and immunity. The receiver part of a transceiver has to correctly identify information without adding too much noise, even when high power unwanted signals are close (in the frequency domain). The transmitter part has to linearly amplify the signal in order to emit as far as possible, respecting the standard power limitations (spectrum mask) for co-existence. Architectures for NB-CW, WB-CW and IR-UWB use some basic elements (blocks) that will be described in the next sub-section. Multi-radio is interpreted as a possible reconfigurable architecture for most of the signals presented (mainly NB-CW and WB-CW). This helps drive improvements on classic structures.

2.2 Basic elements and their imperfections

Transceiver architectures for radio-communications signals is defined from the Digital to Analog Converter (DACs, in the baseband section) to the transmitter (Tx) antenna and, respectively, from the receiving (Rx) antenna to the Analog to Digital Converter (ADCs). Each of the Tx and Rx sections deal with unavoidable tradeoffs such as linearity/efficiency (Tx), noise/gain (Rx). Other transceivers in the spectrum vicinity are supposed to correctly receive (co-existence) and/or emit their signals (immunity) without lowering QoS. The basic functions (blocks) in radio-communications are conversion (digital/analog), high frequency transposition and modulation, filtering, power or low noise amplification and radiation/sensing (Tx/Rx antenna). Here, we are describing a system and how it relates to these blocks in radio-communications architecture. We will focus on their imperfections and their impact on the system performance (noise, spectrum distortion...). Sometimes, for CW standards, the influence of noise or spectral re-growths is quantified by certain criteria such as the EVM and/or the ACPR as defined in Fig. 3 [Baudoin et al., 2007][Villegas et al., 2007].

Fig. 3. EVM and ACPR definitions

- DACs and ADCs will not be presented in details because their performance in terms of bandwidth (up to 100 MHz) and resolution (up to 16 bits) is almost sufficient for today’s radio-communications signals. The main limitations are the current consumption of fast converters and the difficulty in designing near-GHz Sigma Delta (ΣΔ) encoders, often in the context of polar transmitters as is presented in section 4. It should be noticed that for WB-CW signals, the bandwidth limitation is the criterion of choice and a source of important spectrum degradation (approximated qualitatively by a windowing effect).
Radio-communications architectures

- The IQ modulator is the block providing the transposition of the information at high frequency (up-converter) or at baseband/intermediate frequency (down-converter). We describe the case of the up-converter for simplicity. It needs the baseband information components I and Q (I and Q channels) and a carrier frequency signal, provided by a frequency synthesizer. Key components are the non-linear multipliers provided thanks to passive (PIN diodes…) or active circuits (Gilbert cell…). Every component should be carefully designed with regards to the synthesizer frequency value. Moreover, the noise added by the multipliers will be impossible to filter. The main difficulty for the modulator is the perfect matching between I and Q paths. As is reported in Fig. 4 (at one fixed frequency), gain and phase imbalances (a and \( \theta \)) create unwanted images of the information. Also the presence of an offset \( V_{DC} \) can create the emission of the synthesized carrier frequency (called Local Oscillator, LO). All of these imperfections result in the distortion of the information and are quantified by the IRR and LOR power ratio (the actual performance is in the range of -50 dB). These imperfections are not seen on the spectrum because their equivalent added noise is inside the main lobe. The biggest challenge concerns the modulation of wideband signals because it is hard for the modulator to perform and to be frequency independent (over the entire bandwidth). An example of this can be seen in the conversion gain (ratio between the RF output power and the baseband input power). A variation of this gain produces an unwanted AM on the emitted signal and can distort the information. Another problem is the image frequency. This effect is due to the multiplication of signals in a modulator. A multiplication of harmonic signals results in two signals whose frequencies are, respectively, the sum and the difference of the input frequencies. This duality is mathematically illustrated and discussed for the case of the receiver in Fig. 16, section 3. At the outgoing emission, an image is created at RF and
increases the EVM. At the reception (only for heterodyne architectures), an image frequency
different from RF can be sensed and interpreted as unwanted information which increases
the EVM, too. This is illustrated in section 3.3 with the presentation of the Hartley and
Weaver image rejection receiver architectures for NB-CW. To conclude, the effects of the IQ
modulator imperfections are: non-linearities, noise and the possibility of image frequencies.
Selective filtering and stable reference signals are needed to improve the system (easier for
NB-CW than for WB-CW).

- Frequency synthesizers are blocks which produce a stable reference signal for
transposition. RF architectures need flexibility and stability in the choice of their reference
frequency. It is usually not possible to provide this simultaneously with a simple oscillator
circuit such as Quartz (stable), SAWR... The stabilization and synthesis by a Phase Lock
Loop (PLL) is usually necessary. The PLL is a looped system whose different designs will
not be discussed here, only its system’s characteristics, which are reported in Fig. 5. The
worst imperfection of the PLL is its phase noise. An example of the resulting phase noise
profile of a synthesizer is illustrated in Fig. 5 (N=1). The different transfer functions of each
noise of the sub-blocks are reported in the same figure. The great challenge of the PLL is to
keep the stability performance of the synthesized signals for different values of N. It is also
possible to modulate the signal in frequency by directly adding the baseband information
on the Voltage Controlled Oscillator (VCO) input. This is called an “over the loop”
modulation. For CW systems, synthesizers are used to produce the different Local
Oscillators (LOs) needed for transposition(s) and for channel selection (fine tuning of the LO
value). PLL can be used to transpose the information, but only for angular modulations
schemes: PM, PSK, FSK or FM. This is not usually used for constant envelope WB-CW
signals for noise and stability reasons.

- RF filters are essential in communications chains for information selectivity (interferences,
noise, image-filtering for example). They are used at emission (limiting spectral re-growths),
reception (rejecting unwanted signals) and for channel selection (high selectivity for
discrimination and for image rejection). Their system characteristics are well-known:
attenuation/rejection, selectivity, ripple, group delay... Different technologies are used
depending on the frequency, implying different sources of imperfections.

![PLL functional blocks and frequency synthesizer phase noise](image-url)
At GHz frequencies, LC filtering is preferred for its low noise property but the components’ sensitivity (especially for integrated technology) implies low order filters. For higher selectivity, active or high speed digital filters can be used, but often need a frequency transposition section due to the circuit bandwidth and the sampling rate limitations. Moreover, these filters consume power and add much noise (circuit or quantification).

Whatever the technology used, the wider the bandwidth the higher the ripple (oscillation of the transfer function). This ripple problem introduces an unwanted amplitude variation. In order to reduce this, the order is increased as much as possible and some prototype functions like Butterworth or Cauer are chosen for the design. Attenuation in the rejection band is worst in these cases. Additionally, the group delay of the filter is mandatory for non-distortion of the information. For NB-CW, it is usually not a problem but the phase response has to be linear for WB-CW and IR-UWB systems. In the case of a multi-band system, being linear in each sub-band is sufficient. These remarks point out that filters for WB-CW are more difficult to design than filters for NB-CW due to the performance limitations over the bandwidth.

To conclude, the RF filters imperfections are modeled by their noise factors (noise added, distortions and non constant group delays), insertion losses (due to the ripples and mismatches) and selectivity (finite attenuation of unwanted frequencies).

- Power Amplifiers (PA) in RF architectures are designed to linearly amplify radio-communication signals with the highest efficiency possible and with the lowest spectrum re-growths or added noise (see spectrum mask/ACPR and EVM criteria to respect). Since the active components of the amplifier are operating at maximum power, non-linear (NL) compression/conversion and memory effects are unavoidable (Fig. 6). The PA design, identified as “class of operation”, impacts the performances [Villegas et al., 2007][Diet et al., 2004-2007]. Efficiency and linearity of the PA are mandatory for NB-CW and WB-CW architecture because the signal is transmitting information continuously. Low efficiency reduces battery lifetime and increases the dissipated power and the temperature of the circuit. Low linearity affects the quality of the signal. For IR-UWB systems, only peak performances in time are needed (see section 5). We are focusing upon the PA impact on the architecture in the case of the CW system. The most difficult case is for WB-CW due to the bandwidth, and the usual high PAPR of chosen modulation schemes (high data rate). A PA class of operation is determined by the hypothesis of transistor saturation (current source or switch). There are two families of PA classes: the switched mode (SW) and the continuous wave (CW) or biased mode. The different load-lines are illustrated in Fig. 6 (right).

Fig. 6. AM/AM, AM/PM effects (without memory effects) and PA class load-lines
Due to the need of polarization, CW classes (namely A, B, AB and C) present lower efficiency than SW classes (D, S, E and F). SW classes need a switching of the transistor and cannot reproduce an amplitude modulation for that reason. Their linearity is also worse than in CW classes because of this switching operation. Moreover, it is possible to restore the amplitude information by adjusting the voltage supply. This is theoretically linear in the case of SW classes and there is a tracking effect for CW classes. For AM signals the average efficiency will rely on the statistical properties of the signal itself. Additionally, it is important to consider that efficiency is given as a peak value for CW classes. 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. WB-CW signals for high data rate applications present such a high PAPR that the amplification by a CW class PA requires a power back off (very low efficiency) or a linearization technique to reduce the NL effects of compression (AM/AM) and conversion (AM/PM), see Fig. 6. Techniques which are interesting for the designer in the case of wideband and high PAPR signal are those providing the highest efficiency of the entire architecture including the PA (see section 3). Polynomial modeling of the AM/AM can be done by the PA response to 1 or 2 frequency signals, called 1 or 2 tons. These indicators are the 1 dB compression point (P1dB) and the 3rd order Interception Point (IP3), as defined in Fig. 7.

These are considered to illustrate the main PA imperfections. AM/AM and AM/PM measurements, when it is possible to take them, can be the best way of characterizing the effects of the PA on the architecture: EVM and noise factor increase and spectral re-growths (ACPR, spectrum mask). In CW systems, the PA has to amplify modulated signals (a sum of several tons). For NB-CW signals, the PA behavior is well-characterized by the P1dB and the input power is usually set at this value for the linearity/efficiency trade-off (if the PAPR is not too high). In the case of WB-CW, the IP3 is a good representation of NL effects on the
spectrum if the frequency separation of the two tones is coherent with the modulated bandwidth (symbol rate frequency). Moreover, the conversion effects are, by far, more complex to analyze. It is almost impossible to determine an equivalent of the P1dB or the IP3 for AM/PM. This relies upon the influence of transistor technology. For wideband and high PAPR signals, the conversion effect increases the EVM significantly and often destroys the information.

SW classes are based on the hypothesis that the transistor switches perfectly (no power dissipated). The filtering is mandatory in RF applications, except if the SW PA is dedicated to the amplification of a square signal (class D). The load-line of every SW PA class tends to be that of an ideal switch, which is impossible in practice due to the physical realities of the transistor (resistive and capacitive output effects). Although the switching cannot be perfect, SW classes present higher efficiency than CW classes, see the 100% efficiency class E PA [Sokal & Sokal, 1975][Raab et al., 2003][Diet et al., 2008]. Moreover, they are used for AM signals only with a recombination process: supply modulation or amplification of an envelope coded signal (PWM, ΣΔ) [Robert et al., 2009][Suarez et al., 2008]. Using supply modulation creates NL effects of compression and conversion on the output signal. These effects are named Vdd/AM and Vdd/PM [Diet et al., 2004]. For new RF architectures, high efficiency (SW) classes are preferred due to their high efficiency, but the challenge then shifts to the linearity of the amplified signals. Whatever the CW system is, NL effects of the PA are unavoidable and are sources of important imperfections (i.e., loss of linearity). In the case of high PAPR signals, a correction or a modification of the architecture is needed, as is presented in the following section.

Fig. 8. Antenna system characteristics in radio-communications

- Antenna characteristics from a “system” point of view are reciprocal interfaces between electrical and radiated signals. Depending on the application, antennas have to be adapted to their environment (omni-directivity of their radiation pattern, polarization…). Fig. 8 summarizes the different system parameters of an antenna that can influence the
performance of the radio-communications link: spatial and frequency variations of the radiation pattern, bandwidth limitations and phase distortion. Some of the imperfections are equivalent to those of the filter, but they depend on the signal direction and the channel. For wideband systems, i.e. WB-CW and IR-UWB, the use of a UWB antenna is mandatory for keeping the phase information unaltered [Schantz, 2005]. For NB-CW and separated multi-band signals using a wide bandwidth (see Fig. 8), the use of a multi-band antenna is sufficient. The latter can present phase linearity only for used sub-bands [Diet et al., 2006].

To conclude, the communications channel composed of the emitting and receiving antennas and the propagation channel is the source of amplitude and phase distortion, with a statistical dependence on time and space. Noise is added and the techniques of signal protection and antenna diversity (e.g., MIMO) are exploited as much as possible to improve the system’s range.

3. Architecture basics

RF architectures are adapted to radio-communications signals and provide emission and reception. Linearity improvements are needed when power amplification NLs can destroy the information. This section first focuses on the transmitter section where different major modifications are described. We especially focus on CW systems because of their important challenges about efficiency and linearity.

3.1 Classic architectures for RF transmitters

Radio-communications architecture is composed of the above-mentioned blocks in section 2 and provides the emission and reception of the signal. There exist three kinds of basic architectures, the classic ones, designed for NB-CW systems (the oldest application of which is radio broadcasting): homodyne, heterodyne and PLL modulated.

- The homodyne architecture means that the I and Q signals (the information after DACs) is transposed directly from baseband to RF see Fig. 9. If the modulation scheme is not x-QAM, there is only one frequency transposition in the transmitter. This type of architecture is the simplest combination of function blocks and theoretically requires the minimum number of components.

If the signal is transposed directly to RF, the frequency synthesizer output is at the same frequency value, and so is the HPA and the antenna. In a compact system (mobile phones for example), the spatial proximity is the cause of unavoidable coupling between the synthesizer, the HPA and the antenna. This problem of electromagnetic coupling (EMC, EM Compatibility) is that it highly distorts the signal (EVM increase). EMC effects can be reduced by circuit spatial optimization, when possible, and more efficiently with a shielding of the considered block (LO, HPA). Additionally, a ground plane acting as a reflector can be added between the antenna and these elements (HPA and/or LO synthesizer) to reduce the amount of radiated waves toward the circuit. Choosing the homodyne structure results from multiple trade-offs concerning the EVM, the simplicity and the size of the system.

- The heterodyne architecture, as represented in Fig. 9, means that the frequency transposition is achieved in two or more steps. In fact, heterodyne means that the frequency synthesizers are not at RF values. Any combination is possible, but for a minimum use of components only two transpositions are usually performed. As the frequencies are different, the coupling effect is highly reduced in this architecture. In heterodyne structures,
there are more components than in homodyne ones (with at least one filter and one additional mixer) and imperfection sources are added. For example, the phase noise is a function of the number of synthesizers. An additional mixer also means the possibility of an image frequency that can distort the emitted information if the intermediate frequency (IF) filter is not selective enough. This IF filter is traditionally an external SAW one (because of its selectivity). This increases the cost and complexity of building this system. An advantage of the heterodyne architecture is that the need of frequency flexibility for channel selection can be more easily achieved with two synthesizers than with one (homodyne case).

- Architectures using modulated PLL can benefit from PLL advantages. This corresponds to the direct modulation of a synthesizer (with N=1). The signal of a PLL is stable and its noise profile depends on the loop filter bandwidth. It is possible to modulate the PLL by introducing small variations of the reference frequency which will be stabilised by the loop reaction. This is called a modulation “in the loop” and is used for narrow-bandwidth modulations (low symbol rate). On the contrary, the input voltage of the VCO can be directly modulated to produce a wider bandwidth modulation (modulation “over the loop”) but this could affect the stability of the PLL. As is understood here, only angular modulations are possible with modulated PLL architectures. If the signal to transmit has AM information, this latter should be reintroduced after the PLL. To conclude, architectures using modulated PLL are very interesting with regards to their noise property (less noise...
than an IQ modulator) and their design for non-constant envelope signals implies some important modifications of the architecture.

The classic architectures presented are widely used for NB-CW signals. While the efficiency and linearity of the transmitter is not significantly affected, these well-known structures are preferred for their simplicity. In the case of WB-CW systems, the bandwidth and the probable increase of the PAPR (due to high data rate) lower the performance of such architectures. For power amplification in particular, high PAPR values of OFDM and other multi-carrier signals cause such compression and distortion/conversion effects that the information cannot be interpreted at reception. Additionally, standard limitations are, by far, not respected. The first choice is to perform a PA back-off, but this drives it to unacceptably low values for the architecture efficiency. To achieve a linear transmitter, linearization techniques are provided. The next sub-section is dedicated to their descriptions.

3.2 Linearization techniques for the transmitter

Wireless communications require highly efficient and compact transceivers, whatever the signal characteristics are. Transmitter architecture, at worst, must meet design constraints of: providing high efficiency and linearity for a wideband and high PAPR signal (or high dynamic). Power amplification of WB-CW multi-carrier signals (WiFi, WiMAX, LTE…) introduces crippling Non-Lineairities (NLs) in amplitude and in phase in the communication system. The linearization of such this kind of transmitter is mandatory. Identifying a type of architecture for such signals requires a careful study of linearization techniques and their performance. A linearization technique is beneficial only if it provides linearity with the maximum efficiency possible. 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 RF designers [Villegas et al., 2007]. Many criteria characterize the technique used: static/dynamic processing, adaptability, frequency (digital, baseband, IF or RF), correction of memory effects, complexity, stability, resulting efficiency, size increase… Herein, we basically classify these techniques in three types: (i) correction techniques, (ii) anticipation of NLs and (iii) those based on a decomposition and recombination of the signal, often dedicated to wideband signals.

Examples of correction techniques are (A) feed-back, (B) feed-forward and the anticipated technique of (C) pre-distortion, illustrated in Fig. 10. Their point in common is to modify the modulated signal as close as possible to the PA (before or after). The architecture considerations here, do not include the modulator nor the baseband signal processing. To linearize, we need a carefully selected NL model of the PA (Volterra series, Wiener or Saleh model…). Adaptability to the signal amplitude can be introduced in order to compensate for the model’s lack of accuracy and the PA memory effects (a temperature influence can be considered) [Baudoin et al., 2007]. Each structure contains a major defect. (A) Feed-back 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 performed 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 signals. (B) Feed-forward 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 costly in terms of consumption and size.
The advantages are stability and the possibility to process wideband signals. The most interesting of the three techniques is (C) pre-distortion because of its flexibility: the anticipation can be done in the digital part and, by doing so, can provide adaptability of the technique if using a feedback loop (with an additional DAC). The digital pre-distortion represents an additional and non-negligible consumption of a Digital Signal Processor (DSP) and often requires a look-up table [Jardin & Baudoin, 2007]. The signal is widened in frequency because of the non-linear law of the pre-distorter (as for NL effects of compression on the spectrum), requiring baseband and RF parts to be wideband designed. Interesting improvements of pre-distortion have been made with OFDM WB-CW signals in [Baudoin et al., 2007] [Jardin & Baudoin, 2007].

Fig. 10. Correction (A and B) and anticipation (C) linearization techniques

Other techniques presented are based on a vectorial decomposition of the signal. The goal is 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 used when NL effects are so great that feedback or pre-distortion cannot sufficiently improve the linearity. We can consider the problem of linearization in the communication chain from the digital part to the antenna (front end). This drives one to completely modify the architecture and its elements' specifications in baseband, IF/RF and power RF. After the amplification of constant envelope parts of the signals, the challenge is to reintroduce the variable envelope information with lower NLs than in a direct amplification case, keeping high efficiency of the architecture. Basic examples of these techniques are the LINC (Linearization with Non-linear Components) and the EER (Envelope Elimination and Restoration) methods (and their recent evolutions) [Cox, 1974] [Kahn, 1952] [Baudoin et al., 2003-2007] [Diet et al., 2004] [Suarez et al., 2008].
Radio Communications

Fig. 11. LINC decomposition and recombination at RF power amplification

The LINC principle relies on a decomposition of the modulated signal into two constant envelope signals as is shown in Fig. 11. The decomposition can be computed by a Digital Signal Processor (DSP) or by combining two VCOs in quadrature PLL configuration: CALLUM. This latter configuration is an interesting architecture but presents the possibility of instability and additional manufacturing costs. The amplification of the two constant envelope signals implies the design of two identical HPAs at RF frequency, and this often causes signal distortion due to imbalance mismatches. Also the HPA must be wideband because the signal decomposition is a non-linear process (widening of the spectrum), and the phase modulation index is increased. Whatever the decomposition technique is (LINC/CALLUM), the problem is that the efficiency is directly determined by the recombination step: a sum of the powers. It is very difficult to avoid losses at RF while designing a RF power combiner.

Fig. 12. Principle of the EER technique [Kahn, 1952]

Another decomposition technique was proposed by Kahn in 1952 and this is basically an amplitude and phase separation technique (polar): Envelope Elimination and Restoration (EER). This method was first proposed for AM signals as represented in Fig. 12. The advantage of EER is that it drives the RF PA with a constant envelope modulated signal (carrying the phase information), enabling the use of a SW high efficiency amplifier [Raab et al. 2003] [Sokal & Sokal, 1975] [Diet et al., 2005-2008]. The difficulty is to reintroduce the amplitude information linearly using the variations of the PA voltage supply. This implies a
Radio-communications architectures

power amplification of the envelope signal at a frequency equal to the symbol rate (lower than RF). The recombination can be done with a SW class PA because the output voltage is linearly dependant on the voltage supply for this PA mode. Two difficulties are to be considered in such a linearization technique: (i) synchronization between the phase and the amplitude information and (ii) linear and efficient amplification of the amplitude before the recombination (directly impacting the overall efficiency), as reported in [Diet, 2003-2005].

Recently, a lot of work has been done on the EER based architectures, often named as “polar” ones [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 digitally thanks to the power of DSPs, as is shown in Fig. 13. As was previously discussed in [Diet et al., 2003], the bandwidths of the envelope and phase signals are widened due to NL processing and make it necessary to design the circuit for three to four times the symbol rate (as for LINC or any other NL decomposition). Fortunately, a clipping in frequency and on the envelope is possible, increasing the EVM and ACPR under acceptable levels. These polar architectures are suited for new high data rate standards where efficiency of the emitter and linearization are mandatory. Also, the multi-standard and multi-radio concepts have helped polar architectures to evolve in multiple ways. For example, recombination on the PA’s input signal is possible because the amplitude information can modulate the phase signal (RF) and can be 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 PWM or ΣΔ envelope coding are the source of useless and crippling spectral re-growths. The efficiency is also penalized by the power amplification of such frequency components, but this is counter-balanced by the advantages of high flexibility of this architecture [Robert et al. 2009]. The digitally controlled PA and the mixed-mode digital to RF converter are key parameters in the evolution of these polar architectures [Suarez et al., 2008] [Robert et al., 2009] [Diet et al., 2008].

Fig. 13. Recent improvements of EER/polar architectures for wideband OFDM signals

We have presented the main linearization techniques. Actual needs, in terms of transmitter linearity and efficiency for high data rate applications, have caused the RF designer to consider RF architectures based on combined techniques. For example, digital pre-distortion is an improvement on polar based architectures, as is shown in section 4. Another example
is described in [Diet et al., 2008] where a combination of cascaded EER and LINC techniques can theoretically provide an architecture which cancels the PAPR influence, see Fig. 14.

This architecture needs some new components (e.g., a balanced switched PA with frequency transposition). This also represents an increase in complexity, but proposes new ways to improve WB-CW signals transmitters. In EER-LINC, the envelope information is converted to two angular modulated signals. The RF balanced PA (also called differential PA) is in SW class and is supply-modulated with the two above-mentioned signals. The fact that the PA is differentially supplied allows for the combining operation and a transposition at the same time (due to the multiplication). As the antenna is supposed to be differential, there is no balun after the PA.

This review of linearization techniques reveals that the different parts of an efficient and linear transmitter cannot be designed separately: baseband, frequency transposition and power amplification (and antenna for wideband systems). The modification of the architecture for global performance improvements must be done, considering each block’s impact, digital or analog (and their imperfections). To conclude this theoretical sub-section, linearized architectures are mandatory for the major part of actual and future CW systems, which are WB-CW (high data rate). The highest performance can be reached if a combination of different techniques is exploited: pre-distortion and EER seem to be the most popular.

3.3 Receiver architectures

The challenges of the receiver architectures are the noise level, the presence of other channels and blocking/interference signals. This is summarized by the immunity and the coexistence of the standards, illustrated in Fig. 15.

The information received is at such a low level that other RF emissions can mask it. It is easily possible to saturate the receiver if a signal is too strong in the vicinity of the spectrum. Other characteristics of the receiver are its sensitivity and selectivity. The sensitivity is the lowest level of power that can be received and demodulated correctly (providing a Bit Error Rate sufficient for interpretation). This latter, is altered by the total amount of noise added by the receiver itself. The noise factor (F) expresses an equivalent of a white Gaussian noise

![Fig. 14. EER-LINC method for high PAPR signals](image-url)
addition, for each block. It is possible to compute the global noise factor for the receiver thanks to Friis formula, see Fig. 15. This formula shows that amplifying the signal as close as possible to the antenna, at the front-end, is better for minimizing the total added noise. This is the goal of the Low Noise Amplifier (LNA) whose design is optimized for noise and not for gain or output power performance. Unfortunately, the presence of high power signals (blocking) often implies starting the reception chain with a selective band-pass filter, as is reported for the Rx architectures in Fig. 15. At this point, the signal bandwidth is another limiting factor for the design of the receiver because it directly impacts the reception filter design and the LNA. The selectivity and sensitivity performance are also affected. At reception, the decrease in the signal quality is expressed by the EVM and the BER. This degradation results from the noise in the system (noise factor) but other imperfections are due to the architecture design: block interactions, CEM, image frequency and so on. In digital radio-communications, the channel coding (block and convolution types) improves the robustness of the system against the noise, but this point will not be developed here.

\[ F = F_i + \sum_{j=2}^{N} (F_j - 1)/\prod_{j=1}^{j-1} G_j \]

Friis Formula:

Pros: Simple, noise
Cons: Auto-coupling ➔ DC offset

Fig. 15. Considerations for receiver architectures

The topic of this subsection is to present the main types of receiver architectures. As there is no PA NL effects (Tx architecture), there is no need for linearization techniques. If the receiver is saturated, it suffices to reduce the gain of the LNA (if possible) or to attenuate the signal (back-off). The basic receiver structures are similar to that of the classic transmitter structures, that is to say homodyne and heterodyne types as represented in Fig. 15.

- Homodyne receivers are composed of a direct IQ demodulator which comes after the LNA. RF filters, before and after the LNA, tend to avoid a masking effect by other signals received. The filter technology for RF is subject to a trade-off between the selectivity and the losses, which increase the noise factor of the filter. Swapping the LNA and the RF filter in the receiver chain creates a limitation of the received band. This is set in function with disturbing signals and the resulting noise factor (sensitiveness). This architecture is simple, as it requires few components, and theoretically limits the total amount of noise added. The
transposition needs a Local Oscillator (LO) signal that is stronger than the received signal. The coupling effects between the Rx antenna and the frequency synthesizer is very important and a shielding is mandatory to reduce CEM effects. Moreover, if a part of the LO signal is sensed by the antenna, it creates an offset (DC) by auto-mixing of the RF signal. This DC component may saturate the ADCs and is added to the baseband signal. The sensitivity of the homodyne receiver is not as low as expected due to this coupling effect whose impact is greater than that for Tx architecture.

To summarize, homodyne architecture is very attractive for its simplicity if the integration of the synthesizer is not causing crippling CEM effects. The addition of RF filtering can reduce the offset but reduces the receiver sensitivity (noise factor increased).

- Heterodyne receivers need two (or more) transpositions of the signal. The number of components is important and represents an additional current consumption compared to the homodyne structure (simpler). The coupling effect due to the strong LO\(_1\) and LO\(_2\) signals, see Fig. 15, is minimized due to the different frequency values. This architecture needs more filters than the homodyne one but this can be useful for improving the receiver selectivity: Intermediate Frequency (IF) filters can reach higher selectivity than in RF. Moreover, additional filters inevitably decrease the signal-to-noise ratio, as well as the receiver performance. Sensitivity and selectivity are the sources of the challenges faced by the receiver designer. Additionally, the mixers’ operation frequency and the IF filters enable different manufacturing technologies. A great challenge of the heterodyne structure is the possibility of receiving unwanted image information. An image is a signal producing an IF signal after the first mixer (IF = LO\(_2\) = LO\(_1\) - RF, as reported in Fig. 15). A signal whose frequency is “IM = 2LO\(_1\) - RF” produces, after mixing, an output signal whose frequency is “LO\(_1\) - RF = IF”. The receiver should reject this IM component in order to protect the receiver from this perturbation. Without selectivity, the receiver cannot attenuate IM components. If the addition of a filter after the antenna is not sufficient to do this, an improvement is possible with “image-rejection” architectures. The principle here is to eliminate the image signal by an addition of 180\(^\circ\) phase shifted copy. The considered improvements were proposed by Hartley and Weaver. This is possible if the total power level of received signals (information at RF and image at IM) are not saturating the receiver.

An RF filter is mandatory even if these architectures are used. Hartley and Weaver improvements are called “image-rejection” (heterodyne) receiver architectures. The different structures are summarized in Fig. 16.

As is reported, these structures are supposed to cancel the received signal whose frequency value is IM. The Hartley structure uses an all-pass filter with a phase shift of 90\(^\circ\). For NB-CW systems, this filter can be realized by R-C structures. For WB-CW systems, a filter with such frequency independent characteristics is called a Hilbert filter. The realization of this filter is a source of imperfections which directly impact the image rejection property. In order to avoid filter design difficulties, Weaver proposed a structure in which the 90\(^\circ\) phase shift is achieved by a second frequency transposition. This heterodyne structure can directly output the baseband signal (information), as expressed in Fig. 16. The cost of this improvement is the number of added components, which is almost twice as much when compared to a classic heterodyne structure. If the signal is IQ modulated, it is possible to create quadrature image rejection architecture by using two more mixers, as is illustrated by the IQ Weaver schematic in Fig. 16. The increase in size, complexity and consumption is balanced by the image rejection property, if IQ mismatches are low enough. Hartley,
Weaver and IQ Weaver were first designed for NB-CW systems. Their use for WB-CW can be discussed if IQ modulator performance and filters enables it. The bandwidth of the filter implies the choice of LO values. Some receivers, whose required bandwidths are too wide, are not possible to design. “Image-rejection” architectures are an interesting alternative to reduce selectivity constraints in the receiver design, at the cost of additional components and additional noise (lower sensitivity).

Another interesting improvement of heterodyne architecture is a group of Low-Intermediate Frequency (low-IF) receivers whose goal is to use ADCs at IF and not in baseband. There are two main possibilities for signal processing at IF frequencies: (i) using a poly-phase filter in a modified IQ Weaver architecture or (ii) using fast ADCs with digital signal processing on the baseband information. These ADCs imply a large increase in current consumption in function with the IF value. Nowadays, solution (ii) is more popular due to the state-of-the-art ADC performance (resolution versus sampling rate) and the high improvement possibilities provided by digital algorithms. Low-IF architectures with band-pass ADCs represent a sub-group of heterodyne architectures and are popular for its potential flexibility.

To conclude this section in the case of CW systems, there are two approaches for receiver architectures: (i) homodyne (also called “zero-IF”) with limitations of LO coupling effects and (ii) low-IF heterodyne structures, with a growing interest in IF ADCs.
4. High data rate state-of-the-art transmitters

4.1 Front-end considerations

Many new standards for high data rate communications have appeared recently or are under development whether in the frequency range below 6 GHz or in the millimeter wave range (60 GHz radio in particular). In the first case, the data rates are in the range of several tens to several hundreds of Mbps and in the second case they can be in the range of several Gbps. There are many challenges for the design of high data rate RF Transceivers. Among the most critical are:

- Designing power transmitters with a good linearity and efficiency for wideband, high PAPR signals with a large range of necessary transmit power control.

- Designing mobile transceivers using OFDM and multi-antenna MIMO and smart antenna techniques in order to achieve very high performance (throughput and BER) in mobile channels with large delay spreads while maintaining low power consumption and reduced size and cost.

- Designing flexible and scalable devices able to accommodate for many standards, frequency bands and modes

Multi-carrier (OFDM and OFDMA for multiple user access) and multiple-antenna (MIMO) techniques have emerged as enabling technologies for beyond 3G and 4G high data rate communication systems. Many new standards use OFDM and MIMO approaches. Examples include: IEEE 802.11n for wireless local area networks, IEEE 802.16d and IEEE 802.16e fixed and mobile WIMAX or IEEE 802.20 for wireless metropolitan area networks, IEEE 802.22 for wireless regional networks, 3GPP LTE (Long Term Evolution) for beyond 3G cellular networks. OFDM and MIMO have also generated new challenges in term of transmitter architectures with good efficiency and linearity, and in term of integration of MIMO transceiver and antennas in mobile user terminals. One of the main advantages of OFDM in mobile wireless channels is the simplification of the channel equalization in comparison with single carrier modulation. Indeed, for an OFDM modulation with \( N \) carriers, the original high data rate M-QAM data stream with a rate \( R \) is split into \( N \) parallel lower speed streams with a rate \( R/N \) that modulate one of the \( N \) carriers with a M-QAM modulation. In practice, the \( N \) parallel single carrier modulation is achieved thanks to a baseband IFFT. For a given data rate \( R \), the bandwidth of the M-QAM/OFDM signal is similar to that of a single carrier M-QAM signal. An OFDM symbol comprises \( N \) QAM original symbols. A guard interval (GI) is introduced between successive OFDM symbols in order to prevent inter-symbol interference. The duration of the GI is of the order of the delay spread of the impulse response of the channel. Therefore, the larger the number of carriers, the more efficient the system. Unfortunately, the PAPR value of the modulated signal also depends on the number of carriers. The larger the number of carriers, the larger the PAPR value, and the more difficult it is to design a power transmitter with a high efficiency and a high linearity. In order to accommodate for multiple users, a set of carriers can be allocated to each user resulting in the OFDMA principle. The LTE standard proposes an uplink (mobile emitter) variant called single-carrier FDMA (SC-FDMA) which has a smaller PAPR value. In a SC-FDMA emitter, the OFDM modulator is preceded by a DFT, which increases the baseband complexity.

MIMO technology uses multiple antennas at the transmitter and the receiver. The obtained diversity and spatial multiplexing allow for better BER performance and increased data rate or link range without increasing the bandwidth or the transmitted power. In comparison
with a SISO system (Single antenna at the emitter and at the receiver) the maximal achievable increase of the data rate depends on the minimum number of antennas in the transmitter and the receiver. For example, for 2 Tx and 2 RX antennas, the data rate can be multiplied by 2 at the maximum. There is a compromise between the diversity gain and the multiplexing gain. In the WiMAX standard, different modes are defined depending on whether one wants to increase the diversity (Space time block code STBC approach referred to as MIMO matrix A) or the spatial multiplexing (referred to as MIMO matrix B). It is possible to use MISO approach (multiple Tx and single Rx antennas) and use transmit diversity with space time coding (called MIMO matrix A). WiMAX also includes possibility for uplink collaborative MIMO technique. This is intended to allow for 2 separate user devices, each with a single transmit antenna, to communicate on the same frequency with a base station using 2 antennas. The MIMO approach can be associated with beam-forming to control the direction and shape of the radiation pattern.

MIMO techniques allow for tremendous improvements in throughput performances. But it is a challenge to integrate many antennas and transceivers in a small mobile user device. With a multiple antenna base station and a single antenna user device it is possible to achieve some improvement but a multiple antenna user device is necessary to really take advantage of MIMO technique. In the “WiMAX Wave 2” device certification, 2 receive antenna systems are mandatory. Another kind of challenge for high data rate front-end transmitters is the case of millimeter wave (mmWave) mobile communications. At 60 GHz, the available unlicensed bandwidth is very large and the mmWave technology is a good candidate for indoor very high data rate (several Gbits/sec) communication systems. The high data path loss allows for high frequency re-use. The small wave-length allows for the use of very small antennas and integrating multiple antennas for MIMO or beam-forming approaches can be easily achieved. The standard IEEE 802.15.3c is intended to provide Gbits data rate at distance of the order of a few meters. For fixed equipments, the situation at mmWave is quite similar to the preceding one (at frequency below 6 GHz) and OFDM technique is widely used. But the design of mobile devices with low power consumption constraint is still very challenging. The OFDM approach with high speed DAC/ADC and baseband processors has generally crippling power consumption. Among the challenges are: design of PAs with a sufficient output power, efficiency and linearity; power consumption of DAC and ADC with Gigahertz sampling frequencies; design of a physical layer with a low complexity and sufficient performance. Two approaches are commonly proposed: single carrier QPSK and UWB techniques.

4.2 Polar architectures
One of the main challenges for OFDM based mobile transmitters is to design power transmitters with a good linearity and efficiency for wideband, high PAPR signals with a large range of transmit power control. For example, for the WiMAX mobile standard:
- The signal bandwidth is scalable up to 10 MHz.
- 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.
- The Peak transmit power is typically 23 dBm for subscriber terminals.
- The power control of the transmitted signal (TPC) that compensates for variations in signal strength (e.g., distance variation) 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.
For common PAs operating in CW classes (A, AB, B or C), the power efficiency is better for high power output values (close to the P1dB) than for low output values while the linearity is usually better for small power values. In order to transmit a modulated signal with a good linearity, the maximum instantaneous power of the signal must be kept smaller than the P1dB. Therefore, for a given PAPR, the average input power is PAPR dB below P1dB. The amplifier is operated with a back-off depending on the value of the PAPR and for large PAPR values; the efficiency can be very small. The common efficiency obtained with class AB PA for WiMAX signal is typically smaller than 20% while it would be at least twice bigger for constant envelope signals such as GSM signals.

Another parameter to take into account is the average output power. For the same PAPR, the power amplifier efficiency also depends on the average output power. The power supply of the PA should be adjusted in order to take into account the desired average output power and keep a constant efficiency on a large range of average output power.

Different kinds of polar architectures have been proposed as candidate solutions for high PAPR management or for adjustment of power supply to the average power. The terminology is not very stable. But, one can distinguish:

- Polar lite architectures.
- Polar feedback loop.
- Dynamic power supply and amplifier gain control with dynamic biasing, envelope or power tracking for linear PA, drain (collector) modulation for non-linear PA.
- EER and Sampled-EER architectures.

Whatever the modulation, the complex envelope $z$ of the modulated signals can be expressed by its cartesian coordinates: real and imaginary parts usually called quadrature or $I$ and $Q$ components or by its polar coordinates: amplitude $\rho$ and phase $\phi$, see (1). As the amplitude $\rho$ and phase $\phi$ are obtained by very non-linear operation from $I$ and $Q$, their bandwidth is much higher than that of $I$ and $Q$, as discussed in section 3.

$$z(t) = I(t) + jQ(t) = \rho(t)\exp(j\phi(t)), \quad \rho(t) = \sqrt{I(t)^2 + Q(t)^2}, \quad \tan(\phi(t)) = \frac{Q(t)}{I(t)}.$$  \hspace{1cm} (1)

When the amplitude $\rho(t)$ is constant, the polar decomposition is interesting since only one signal ($\phi(t)$, or its time derivative) has to be digital to analog converted. GSM Transmitters can take advantage of this characteristic. As seen in section 3.1, the modulation of the frequency synthesis loop is a common GSM architecture that benefits of the good noise floor of the VCO and suppresses the need for external filtering. Since the envelope of the modulated signal is constant, the PA can operate in saturated CW or SW class (high efficiency). An average power controller must be added. This architecture is also called translational (or tracking offset) loop. The modulation of the PLL becomes very difficult for large bandwidth signals such as WCDMA. Therefore this approach is mostly used for GSM/EDGE signals. The name “polar transmitter” is sometimes limited to architectures in which the phase/frequency modulation is applied directly to the RF carrier by a modified PLL, using different techniques such as a “2-point modulation” [Durdodt 2001]. But we will use the adjective “polar” whenever the signal is decomposed in polar coordinates.

The translational loop technique was extended for EDGE-GSM signals using the so-called polar lite architecture. The 8-PSK EDGE-GSM modulated signal is decomposed in polar coordinates. The idea is to keep the benefits of the translational loop architecture but to introduce an AM capability. In such architecture, the signal is decomposed in polar coordinates. The phase modulates the loop and the amplitude multiplies the output of the
modulated loop by controlling the gain of a high dynamic range variable-gain amplifier (VGA). Since the signal at the input of the PA is envelope-varying, this architecture does not support a saturated PA. It uses a linear PA and has no particular efficiency advantage. It is sometimes associated with some dynamic power supply technique in order to increase the efficiency. An example is given in [Staszewski et al 2005].

The **polar feedback loop architecture** is a derivation of the polar lite architecture with the advantage of using a saturated PA. In the polar feedback loop, the PA is fed with a constant envelope signal and is modulated in amplitude. A feedback path takes a portion of the PA output signal (with a coupler) and an error signal is calculated between the ideal and actual output. The phase of the error drives the input of the PA and the error magnitude is used for the amplitude modulation of the PA. The feedback loops provides some linearization to the transmitter. But as in any feedback loop, the stability constraint limits the possible bandwidth. An example of a polar feedback loop architecture for GSM and EDGE is given in [Sowlati et al, 2004]. In that example, the PA efficiency is 54% in GSM mode at 33dBm output power and 37% in EDGE mode at 27dBm output power.

Many techniques are possible for dynamic power supply: dynamic biasing (at gate or drain), envelope or power tracking for linear PA, drain (collector) modulation for non-linear PA. Dynamic power supply using DC-DC switching regulators are interesting in terms of efficiency but they are still limited in term of signal bandwidth to a few tens of MHz. For envelope tracking (ET), the DC-DC is fed with the amplitude of the modulated signal (Fig. 17). ET was not presented in section 3 because it is not strictly a decomposition technique of linearization, but an improvement of the PA output power (and consequently the efficiency). This optimisation of PA biasing can increase of more than 40% the efficiency of a class A PA. In some standards the power-control dynamic range must be very wide (i.e., 80 dB for WCDMA). The amplitude modulation (envelope restoration) can be applied on a VGA or directly on the PA which leads to a better efficiency.

![Envelope tracking diagram](image)

**Fig. 17.** Envelope tracking

The principle and interest of **EER architectures** and their variants under different names such as polar architectures were explained in section 3.2. The critical points for this architecture are: (i) the time mismatch between the envelope and the RF phase signals, (ii) the NL of the envelope restoration, (iii) the distortion caused by difficulty of biasing the PA when the amplitude of the envelope signal is very small, (iv) the leakage of the RF PA input to the output, (v) the influence of the necessary limitation of the frequency bandwidth of the envelope and phase signals. It can be shown [Baudoin et al., 2003] that the signal to noise ratio due to a time mismatch \( \tau \) between these two paths is in a first approximation inversely
proportional to $\tau$ and to the bandwidth of the envelope signal. Therefore, the wider the signal bandwidth, the smaller should be the time mismatch. For an OFDM signal, such as in WiFi, the time mismatch $\tau$ must be kept smaller than typically 2 ns (which is a small percentage of the QAM symbol duration) in order to fulfil the specifications of the standard. The time mismatch can be corrected by adaptive techniques and the NL of the envelope restoration can be compensated by adaptive pre-distortion techniques. But, as already stated for dynamic power supply, it is difficult to design DC-DC converters with a bandwidth superior to a few tens of MHz. This is all the more critical as the envelope signal bandwidth is wider than the original modulated signal bandwidth. For example, an OFDM modulated WiFi signal has a frequency bandwidth close to 20 MHz. But the bandwidths of its envelope and phase signals cannot be filtered to a bandwidth smaller than respectively 40 MHz and 100 MHz if one wants to meet the specifications of the standard. The polar EER approach can be illustrated by Tropian’s Timestar (TM) RF IC supporting GSM/GPRS, EDGE and WCDMA signals [Wendell et al., 2003] [McCune et al., 2005] or Sequoia Communications Inc. SEQ7400 chip that supports HSDPA, GSM and EDGE and WCDMA [Groe et al., 2007-2008].

In polar architectures, the recombination of amplitude and phase signals can be done whether by PA amplitude modulation (EER architecture) or before the PA input. We will now consider the second approach, and we will focus on architectures using sampled signals and switched RF amplifiers (typically class D, E, F and their variants) [Jeong et al., 2007] [Hibon et al., 2005] [Berland et al., 2006] [Nielsen et al., 2007]. We will call these architectures “polar sampled architectures”. The major motivations for using sampled signals and switched amplifiers are the ease of integration of digital circuits and the very high theoretical efficiency of switched PA. In polar sampled architectures, the envelope signal is “sampled” (converted/coded) by a pulse width modulation (PWM) or a 1-bit Sigma-Delta ($\Sigma\Delta$) modulator with bipolar output $\pm A$ [Murmann et al., 2007]. This two level signal multiplies the RF phase modulated signal before the switched PA. The result is a constant envelope signal. The switched RF PA is fed with this constant envelope signal that controls the switching of the PA. The output of the PA must be filtered in order to suppress the PWM or $\Sigma\Delta$ noise and to recover the modulated signal. Fig. 18 illustrates this principle.

**Fig. 18.** Polar sampled architecture with envelope-phase recombination before the PA

In comparison with direct $\Sigma\Delta$ architectures [Rode et al 2003] in which the modulated RF signal is directly covered by a 1-bit $\Sigma\Delta$ coder before the switched amplifier, the coding of the envelope signal is interesting because it allows using smaller clock rates for the $\Sigma\Delta$ coder. One difference with EER architecture is the position and type of the filter used to eliminate
the noise of the ΣΔ or PWM modulator. As illustrated in Fig. 13, EER architectures use a low-pass filter at the output of the envelope BF amplifier to recover the envelope signal. But when the recombination of envelope and phase signals is done before the RF PA, the noise must be eliminated after the RF PA with a band-pass RF filter (Fig. 18). For a given standard, this RF band-pass filter should be unique and correspond to the full uplink bandwidth. It should not be specific to a given channel. Therefore the over-sampling ratio of the ΣΔ modulator must be calculated in reference to the full uplink bandwidth and not to channel bandwidth (bandwidth of the modulated signal). In [Andia et al., 2008] the possibility of using BW filters for WiMAX standard has been studied.

![Diagram](image)

**Fig. 19.** (a) Polar Sampled Architecture, PSA, with 2 DACs. (b) PSA with a single DAC and an analog mixer. (c) PSA with a single DAC and with digital mixing.

Different structures have been proposed for polar sampled architectures [Suarez et al., 2008], see Fig. 19. In the structure (a) of this figure, ΣΔ modulator output is digital, as well as \( I(t) \) and \( Q(t) \) (\( \rho \cos(\phi) \) and \( \rho \sin(\phi) \)). Therefore, two DACs are necessary before the IQ modulator. DACs sampling frequency is chosen according to the ΣΔ frequency. It has to be high enough to avoid ΣΔ noise overlapping. Targeted communication standards require high Sigma-Delta frequencies and therefore significant sampling frequency for DACs. In the structure (b), envelope and phase signals (\( \rho(t) \) and \( \phi(t) \)) are calculated and processed independently. The output of the low-pass sigma-delta modulator is analog. The digital phase signal is converted by a DAC and then modulated to the carrier frequency (\( f_c \)). Finally constant envelope and phase signals are recombined.
The advantage of this approach compared to the first one is that it only requires one DAC. Furthermore, DAC frequency requirements are mitigated. In the first architecture, modulation to the carrier frequency is performed by the IQ modulator bloc with help of an analog mixer. Similarly, this second architecture makes use of an analog multiplier to recombine envelope and phase. The third architecture, structure (c), uses a digital carrier and replaces the analog mixer by a digital one, like an AND gate for example. In this case, an All-Digital Phase Locked Loop (ADPLL) whose input is the phase signal is used. The digital mixing is advantageous because it offers all the typical advantages of a digital signal treatment and digital IC integration but it introduces harmonics of the carrier frequency in the output spectrum with replication of the spectrum every $3f_c$ (Fig. 20). If digital mixing is implemented in a polar $\Sigma\Delta$ architecture, frequency ratio between carrier frequency and sigma-delta frequency has to be carefully chosen to avoid $\Sigma\Delta$ noise overlapping.

Fig. 20. Power spectral density of the signal at the PA input for a PSA with digital mixing (right), WiMAX case. Zoom on the power spectral density before and after the RF filter (left)

Fig. 20 corresponds to the third architecture, (c) in Fig. 19, and gives the power spectral density (PSD) of the signal at the input of the PA in the case of a WiMAX signal with a 3.7 GHz carrier and for a second order $\Sigma\Delta$ modulator, using a clock equal to 3.7 GHz. Fig. 20 also shows a zoom on the PSD before and after the RF filter. In that case the considered filter bandwidth is 100 MHz (while the channel bandwidth is 10 MHz) and the over-sampling ratio is equal to $(3.7 \text{ GHz})/(200 \text{ MHz})=18$.

4.3 Digital architectures
The performance of ADCs and processors, and research activities on RF-DACs (e.g., Sigma Delta) argue for integrating more and more functions in the digital part of a transmitter. Based on the software radio concept, the actual idea is to “digitalize” the design of some basic blocks such as the modulator, DACs RF and PA (D-PAs). Brand new cellular devices lead to the use of two or three different standards in the same mobile platform. As explained in a previous part, it requires constringent spectrum performance under very high speed data treatments. Each element of the transmitter must be designed with regards with several parameters: (high) frequency, dynamic power control and consumption (leading to high efficiency). Due to the use of Sigma Delta modulators, WB-CW standards such as WiMAX 3.5 and 5.8 GHz need an over-sampling frequency in the range of 15 GHz (more than four times the carrier frequency).
Using CMOS 90 nm technology is a current solution to address this challenge, while reducing the number of analog blocks. In this sub part we present two digital architectures, representing the digitalization trend.

The first architecture presented is based on a classic direct conversion architecture using only one frequency transposition. The “Direct Digital to RF Modulator” (DDRM) architecture [Eloranta et al] was developed as a basis toward further architecture digitization. In this architecture the system is digitalized the closest possible to the amplifier, which is still an analog part of the transmitter, see Fig. 21.

Fig. 21. DDRM architecture (left) and DRFC (right) [Eloranta et al]

The advantage of being digital is that it limits the imperfections due to variations in the process. In analog architectures, the multiple filtering blocks see their basic characteristics varying, and so the architecture performance varies as well. This leads to the use of a calibration loop. Digitization provides size optimization and good stability of the circuit. As the first stage is an over-sampling stage, there is no baseband signal and no need to filter with high selectivity before the DACs. In this architecture, mixing and D/A conversion are performed by a single block: “Digital to RF converter” (DRFC), a kind of RF-DAC. Weighted Gilbert cells are parallelized. No baseband signal in the architecture implies a very low LO leakage (DC offset). The principle is that the data is over-sampled at the carrier frequency by switches. The signal amplitude is coded into a digital word (MSB to LSB). At the output of Gilbert cells, the current is proportional to the code word. The linearity performance and the signal resolution increase with the number of parallel cells. Due to this parallelization, IQ imbalance is limited and can only result from the average cell imbalances. Power control can be achieved by a bias current variation, which reduces the output current from each unit cell. As it uses no filter, the choice of the converter frequency is paramount. Indeed, the only filtering applied (SINC) is the zero-order hold. The frequency must be chosen so that baseband harmonics are cancelled thanks to the zeros of the filter response.

Thanks to CMOS evolutions, this architecture was optimized to address spectrum cohabitation issues [Pozsgay et al] under the name of “Sigma Delta – RFDAC” architecture, illustrated in Fig. 22. In this architecture, the gain control can take place throughout the transmitter. The first power control appears after the first up-sampling filter. It is then a low speed dynamic control, resulting from multiplying IQ signals with a binary word. This word size depends on the control resolution. A second power control stage is achieved, by deleting successive LSBs of the signal (6 dB steps). Several Delta-Sigma modulators in MASH structures are then used and I and Q channels are duplicated (I and I', Q and Q').
and Q signals go through a modulator, and signals I' and Q' are first delayed before a second modulator. This will result in the recombination of mixed signals around the carrier, and the creation of notches in desired frequency bands (certainly used by other standards). An additional 6 dB power control can be done by by-passing I' and Q', and the corresponding RF-DAC, losing the advantage of notches as a result.

![Digital Quadrature Modulator](image)

Fig. 22. Sigma Delta – RF-DAC architecture [Pozsgay et al]

The second architecture presented is based on an EER-like architecture. As will be explained, there is no complete separation between phase and amplitude in the transmitted signal processing. The architecture was developed by [Parikh et al], see Fig. 23. IQ signals are over-sampled and filtered in order to place the spectral re-growths far enough (zero order hold) and setting notches at multiples of the over-sampling frequency.

![Digital Quadrature Modulator](image)

Fig. 23. Digital Quadrature Modulator [Parikh et al]

Two signals are created: “abs(I + Q)” and “abs(I-Q)”. Moreover, we determine phase control signals (2 bits) with the signs of I and Q (complex quadrant). Amplitude signals are alternately sent to a digital differential amplifier (DDPA) at a rate of twice the carrier frequency. Thus we get a higher resolution of the variation of the original signal envelope. This can be seen as an over sampling by four of the IQ symbols leading to a better estimation of the amplitude, see Fig. 23. In Fig. 24, we can look at what would happen if I and Q take the value 4 and 3 respectively. This coded amplitude signal feeds a Digital Power Amplifier (DPA). This block will mix the signal around a carrier depending on the instantaneous sign of both I and Q. Knowing I and Q signs, gives us information on the phase (linked to the quadrant position). A part of the phase information can be taken from “abs(I+Q)” and “abs(I-Q)” signals.
This amplitude-coded signal can then be modulated by an RF carrier with an estimated phase. As seen in Fig. 23 and Fig. 24, four carrier signals with four different phases (0°, 90°, 180°, and 270°) feed the DDPA depending on I and Q signs (2-bit control signals). As we obtain a NZR signal, the filtering condition after the DDPA will be less stringent than when using I and Q directly. One of the carrier frequency signals and the coded amplitude are fed to the DDPA. The carrier frequency signal is logical ("0" or "1" values) and activates one or another of the two differential pairs in the DDPA, in function with the sign signals. If the carrier is set to “0”, “1”, respectively, the output is alternately “-abs(I-Q)” and “-abs(I+Q)”, or “abs(I+Q)” and “abs(I-Q)” (see Fig. 24). The output network of the DDPA is a band-pass filter around the carrier frequency, to reconstruct the original signal. Looking at the spectrum, performance is well suited to multi-standard applications. This is not due to flexibility in changing the operating frequency. In [Parikh et al] an example is given at 5.8 GHz for a 64QAM modulation scheme and 10 MHz bandwidth WiMAX signal, which is very stringent. The resulting spectrum re-growths are less than -50 dBc/Hz.

![Signal amplitude and phase estimation](image)

Fig. 24. Signal amplitude and phase estimation (left), DDPA (right). [Parikh et al]

To conclude, digital RF architectures for WB-CW are providing several advantages such as flexibility and size optimization. Their design is subject to the improvement in power consumption of RF blocks. This research topic is currently popular for mobile and connectivity high data rate standards.

5. IR-UWB Architectures

This section is dedicated to IR based and all the UWB systems. Why to consider the UWB systems in addition to already existing technologies? Currently, interfaces have different characteristics in terms of throughput, coverage, access efficiency, quality of service (QoS) and energy consumption. Some of these interfaces offer a QoS for multimedia applications, particularly through a given performance of transmission in a period of time. Other technologies offer specific services such as distance measuring (ranging) like the some UWB systems. UWB technologies are not made for metropolitan networks. The potential of UWB lies in the use of interconnections of a wireless pico-network.
5.1 UWB communications, goals and aspects
There are systems that transmit and receive waves whose relative bandwidth (BW) is greater than or equal to 0.25, see the definition in Fig. 25. The first definition has been amended and replaced by a new one proposed by the FCC. Under this new definition, a UWB signal is a signal whose “-10dB bandwidth” exceed 500 MHz and 20% of center frequency. The main UWB frequency band is between 3.1 and 10.6 GHz. This bandwidth of about 7 GHz could be divided into 14 sub-bands of 500 MHz. A system using the full bandwidth or a set of sub-bands will be considered a UWB system. Today, we can classify UWB into two main categories of applications: UWB Low Data Rate (UWB LDR) and UWB High Data Rate (UWB HDR). They are attached to the IEEE 802.15.4a and 802.15.3a.
Low data rate systems are generally characterized by data rates lower than to 2 Mbps, by ranges of up to 300 meters and finally by a low consumption. They may allow positioning and location functionalities. The high data rate systems are characterized by data rates exceeding 100 Mbps with short ranges (up to a few tens of meters). Their fields of application are the computer data transmission and multimedia systems. Fig. 25 shows the frequency masks used in the USA.

\[
BW = \frac{f_h - f_i}{f_c}
\]

\[
f_c = \frac{f_h + f_i}{2}
\]

\[
P_{\text{pic}}^{\text{lim}}(RBW) = 20 \log_{10} \left( \frac{RBW}{50} \right)
\]

Fig. 25. Frequency masks, relative bandwidth and peak radiated power (RBW in MHz).

The -41.3dBm/MHz limit correspond to a measurement of electromagnetic field equal to 500mVm⁻¹, in any sub-band of 1 MHz, at a distance of 3 meters from the antenna. This level is also named the “Part 15 limit”. In the FCC report, the peak power is also limited. It is measured around the frequency for which the radiation is at its maximum, and is defined in Fig. 25. For RBW = 50 MHz, the peak power must not exceed 0 dBm (1 mW). Several transmission techniques approaches have been studied and can be divided into:
- UWB mono-band: impulse approach (IR-UWB, or “classic UWB”)
- UWB Multi-bands: MB-OFDM approach and MB-OOK approach. These systems present high similarities with WB-CW ones.

5.2 IR-UWB transceiver architectures
The communication is based on short pulses transmission (a few hundred picoseconds), occupying all or part of the UWB spectrum, and repeated with a period of a few
nanoseconds. The signal is transmitted at the baseband frequency. It is interesting even in presence of multi-paths channel. By using a pulsed mode, it allows the measurement of the propagation delay time between the transmitter and the receiver and, consequently, the location. Informations are encoded in the shape of the impulse (amplitude or phase), by the position in time of the pulse or by random sequences of pulses.

Possibilities for modes transmission of the UWB pulses are: coherent, differentially coherent, and non coherent. In the case of coherent transmission, the modulations used are PSK, PAM, and PPM. The receiver will be of a correlator type, e.g. Rake. In the case of differentially coherent transmission, the modulation used is DPSK. Moreover, synchronization becomes un-necessary. For the last case of non coherent transmission, modulations used are OOK or PPM. The receiver will be based on energy detection. The transceiver architectures for IR-UWB are simpler than those for the MB-OFDM or any other WB-CW. The functional blocks of the architecture are given in Fig. 26.

![Functional blocks of IR-UWB transceivers and receiver windowing.](image)

Demodulation should be made on windows slots containing a pulse, as illustrated in Fig. 26. Architectures for impulse radio are divided into two types: Coherent Reception (Co-Rx), requiring knowledge of the phase of the received signal, and Non-Coherent Reception (NCo-Rx), based on detection of energy when pulses are transmitted. In the case of Co-Rx, the receiver uses the recognition of the phase signal. It is generally based on a sliding correlation scheme which produces almost the optimal performance. Therefore the complexity leads to a limitation on the implementation in low consumption equipment. Care should be taken to the synchronization at the acquisition that requires a very fine time to get the precision needed for coherent demodulation.

![Examples of Co-Rx transmitter and receiver](image)
The two diagrams in Fig. 27 show the blocks of the transmitter and receiver of a coherent system. Two configurations are possible, the first is mixed analog/digital circuits, the second one is all-digital. Key points of the architecture are: the co-design antenna/LNA, ADC converter performance and the design of the correlator.

For the NCo-Rx, several architectures are possible but the principle is still based on energy detection. This technique is less complex than the previous one but its performance is not as good. The problem is related to the difficulty in designing a receiver able to detect the pulse, when the signal to noise ratio is very poor.

In the case of a non-coherent system, architectures are based on energy detection whose principles are given in Fig. 28. Different types of pulses are used to perform the IR-UWB link: the Gaussian monocycle and its derivatives, the Hermite pulse (Hermite polynomials) or a sinusoid signal windowed by a Gaussian shape. Only the Gaussian pulse or the sinusoid signal windowed by a Gaussian shape are interesting because they can be implanted easily in practice [Marchaland et al., 2007].

5.3 MB-OFDM UWB transceiver architectures

As for CW systems, the MB-OFDM is characterized by a continuous transmission. It uses frequency hopping on at least three bands, 128 sub-carrier band of 528 GHz (in the case of 3.1 - 10.6 GHz band), and the QPSK modulation scheme. It allows a great flexibility in shaping the spectrum. This parallel multi-carrier operation minimizes inter-symbol interference, and the recovery of the energy available can be optimized. The MB-OFDM approach is based on an IFFT at emission and FFT at reception as shown in Fig. 29.

The overall topology of the receiver is very complex. The MB-OFDM technique is well suited to high speeds data transfer in indoor environments. It has good resistance to multi-paths channel, and, allows flexible shaping of the spectrum. Modulation is complex to implement and requires circuitry to perform an FFT in real time, so the digital part of the architecture is quite complex. It needs synchronization. This technique requires front-end with high linearity and low noise (RF elements). Additionally, the consumption of the analog area may be important.
6. Conclusion

The goal of this chapter is to demonstrate the absolute need of matching the architecture design and the signal carrying the information.

In section 1 and 2, considerations about current users’ needs helped to identify three types of carrier signals in the context of radio-communications: NB-CW, WB-CW and IR. A separation between CW and IR signals is unavoidable because it drives us to a different technological design. Basic blocks of transmitter architecture are optimized in function with the targeted performance over the bandwidth, as was discussed in section 3. Conclusions were given about actual trends in this research topic. The goal of section 4 was to illustrate the high degree of complexity for actual WB-CW systems, which represents high data rate applications. Section 5 was an overview of IR based architectures.

Whatever the system is, the design of the transceiver architecture has to fulfill challenges such as co-existence (for the transmitter part) and immunity (for the receiver part). This implies a careful reduction of the spectral emission (spectral re-growth) for a transmitter. It also implies a high selectivity for the receiver and, of course, with the lowest sensitivity possible. RF architectures are becoming more and more complex. Especially, in the cases of RF transmitters, high efficient power amplification often results in a combination of different linearization techniques. The ever-increasing performance of the digital part gives us the opportunity to provide more flexibility in the architecture when the dynamic control and the power delivered to the load are satisfying standard requirements. This can be done thanks to the digitalization of some functions such as constant amplitude envelope coding, RF converters and/or PA (D-PA).

The concept of multi-radio points out that the future of RF standards lies in favoring cooperation and flexibility in the managing of the system resources (bandwidths, power and time partitioning). Actual state-of-the-art results are that architectures are merging to reconfigurable RF blocks and when possible RF-digital blocks.
7. References


Pozsgay, A. et al. A Fully Digital 65nm CMOS Transmitter for the 2.4-to-2.7GHz WiFi/WiMAX Bands using 5.4GHz RF DACs. *IEEE ISSCC 2008*, pp 360-619.


