Proceedings of
SDR-WInnComm- Europe 2013
Wireless Innovation European Conference on Wireless Communications Technologies and Software Defined Radio
11-13 June 2013, Munich, Germany

Editors: Lee Pucker, Kuan Collins, Stephanie Hamill

SDR-WInnComm-Europe 2013 Organization

Kuan Collins, *SAIC* (Program Chair)

Thank you to our Technical Program Committee:

Marc Adrat, *Fraunhofer FKIE / KOM*
Anwer Al-Dulaimi, *Brunel University*
Onur Altintas, *Toyota InfoTechnology Center*
Masayuki Ariyoshi, *NEC*
Claudio Armani, *SELEX ES*
Gerd Ascheid, *RWTH Aachen University*
Sylvain Azarian, *Supélec*
Merouane Debbah, *Supélec*
Prof. Romano Fantacci, *University of Florence*
Joseph Jacob, *Objective Interface Systems, Inc.*
Wolfgang Koenig, *Alcatel-Lucent Deutschland AG*
Dr. Christophe Le Matret, *Thales*
Fa-Long Luo, *Element CXI*
Dr. Dania Marabissi, *University of Florence*
Dominique Noguet, *CEA LETI*
David Renaudeau, *Thales*
Isabelle Siaud, *Orange Labs, Reseach and Development, Access Networks*
Dr. Bart Scheers, *Royal Military Academy, Belgium*
Ljiljana Simic, *RWTH Aachen University*
Sarvpreet Singh, *Fraunhofer FKIE*
Olga Zlydareva, *University College Dublin*

Table of Contents

PHYSEC CONCEPTS FOR WIRELESS PUBLIC NETWORKS – INTRODUCTION, STATE OF THE ART AND PERSPECTIVES

Cong Ling
(Imperial College London; London, United Kingdom; c.ling@imperial.ac.uk)
François Delaveau, Eric Garrido
(Thales Communications & Security; Gennevilliers, France; francois.delaveau@thalesgroup.com; eric.garrido@thalesgroup.com)
Jean Claude Belfiore, Alain Sibille
(Institut Mines Telecom Paris Tech; Paris, France; belfiore@enst.fr; alain.sibille@telecom-paristech.fr)

ABSTRACT

This paper aims at providing elements about advances in physical security (physec) and about relevant application perspectives in public wireless networks. After a short introduction of existing protection of communications signals, we will introduce several notions relevant to information theory and point out the main physec concepts. Then, we discuss their theoretic advantages and the current knowledge about secrecy codes. Finally, the paper will highlight practical implantation perspectives of physec in existing and future public radio-networks, as stand-alone added modules operating at the physical player, or as addedalgorithm combined with classical solutions in order to upgrade and/or to simplify existing security procedures.
This work is supported by the PHYLAWS project (EU FP7-ICT 317562, www.phylaws-ict.org), starting Nov. 2012.

1. INTRODUCTION

Given the growing prevalence of wireless radio-communication technologies, security, privacy and reliability of the exchanged information becomes a major societal challenge for both personal and professional sphere. Moreover, the growing importance of sensing and cognitive procedures in future radio access technologies (white spectrum, cognitive networks) will occur numerous downloading and uploading procedures for geo-referenced sensing spectrum allocations, whose integrity and privacy are major industrial challenges for both operators and administrations. Secure air interface within wireless networks are thus crucial for various applications such as broadband internet, e-commerce, radio-terminal payments, bank services, machine to machine, health/hospital distant services. Most of citizens, professionals, stakeholders, services providers and economical actors are thus concerned by confidentiality lacks and by privacy improvements of the physical layer of wireless networks.

1.1. Existing protections within wireless networks

Several classical solutions already exist in order to protect privacy of radio transmissions.
. Wave forms can be designed in order to achieve Low Probability of Interception (LPI) and Low Probability of Detection (LPD), by using furtive frequency/time hopped or spread spectrum signals. LPD and LPI signal achieve transmission security (transec) at frame and at symbol level.
. At the signaling level, protocols can be designed in order to achieve low probability of decoding and low intrusion capabilities of the signaling messages by non-legitimate users, thanks to subscribers' and nodes' authentication procedures, thanks to advanced scrambling, interleaving, coding and ciphering techniques, etc. Such protections achieve network signaling security (netsec). They apply either at signal frame, at symbol level and at bit level. In addition, netsec may be re-enforced with early identification as Identification of Friend and Foe procedures.
. At the communication level, encryption algorithm and message integrity control schemes are used in order to avoid non legitimate interpretation and/or intrusion attempts of the users' messages. Such protections achieve communications security (comsec) by applying mainly at message/bit level.

1.2. Limits and drawbacks of existing protections

Nevertheless, all the classical protections above require a priori knowledge or exchanges of keys, thus improving the complexity of the network management and/or reducing the set of users of highest protected modes. In addition, these protections often require shared time references; thus induce added vulnerabilities or failure risks (ex: time reference is public itself - GPS for instance).
Moreover, all existing protections use added data, and thus decrease the spectrum efficiency, especially when facing short packet services. Finally, all their constraints trend to dramatically reduce the effective privacy of any wireless standards that targets a worldwide mass market.

1.3. New perspectives offered by physec.

Physical Layer Security (physec) is a radically novel concept that exploits the properties of the local radio-environments, especially when complex, dispersive and non-stationary.
. Since its introduction by Wyner [18], the fundamental model of wiretap channel (figure 1) has led to the definition of secrecy capacity for several propagation models (see [18-26] and to the design and/or to the re-use of advanced coding schemes in order to approach it (LDPC, lattice and polar, codes).
. A native and tremendous advantage of physec is the absence of keys: Security over radio-channel is achieved through channel coding in the same processing as signal transmission, thanks to "secrecy codes" that optimize information recovery by legitimate receiver and that mitigate information leakage about the legitimate link at any eavesdropper location. No external information is required nor exchanged.
. Because of its information-theoretic foundation, physec is intrinsically robust to any computer attack (unlimited computing power), even to quantum attack.

1.4. Practical expectations of physec for wireless networks

Physec appears as a potential "front end" solution for warranting privacy and security within wireless public networks. Physec mainly operates at the radio interface an uses software means only: Low imbrication should occur with upper layers of the transmission protocol and with network management.
Thus, many practical advantages may be expected from physec-native or physec-derivate security solutions:
- Reduced impact on terminal and on network architectures.
- Easy and low cost integration.
- Compatibility with existing encryption solutions.
- Compatibility with existing radio access technologies.
- Negligible impact on spectrum efficiency.

Therefore, physec solutions or security modules including physec concepts should address a wide class of wireless applications in the close future:
- Within wireless radio-cells: GSM and UMTS evolution, LTE and LTE–A.
- Within upgraded or new Local Loop standards: WiFi, extension of 802.11a/b/g/n, 802.11i/w, 802.11ac, WiGi.
- Within broadcast and point to point services supporting mobile broad band internet, machine to machine and internet of machines.
- Within cognitive networks: data base downloading, geo-referenced sensing and geo-referenced access procedures.
- Within private transmission systems (PMR).
- Within short range communications devices: Bluetooth Zigbee etc., even RFIDs

1.5. Structure of the paper

The following will first introduce several notions relevant to information theory and the main principle that are relevant to Physical Layer Security. Then, from a state of the current researches, we will highlight several security solutions that should take benefit of native physec concepts when facing passive eavesdroppers, such as adaptive modulation and coding schemes, cooperative jamming and space time diversity exploitation within MIMO RATs. We will then introduce physec perspectives to counter active threats such as radio-hacker intrusion attempts of signaling messages.
Possible drawbacks of physec will be discussed too, that may be relevant
- to secrecy codes determination,
- to implementation of coding and decoding schemes into standard wave forms, handsets and base stations,
- to embedded computing complexity (versus the expected performances of embedded computers),

The paper will conclude on practical implantation perspectives of physec in existing and future radio-networks, as stand-alone added modules operating at the physical player, or as added algorithm combined with classical solutions in order to upgrade and/or to simplify existing transec, netsec and comsec protections.

2. EXISTING PROTECTIONS OF RADIO SIGNALS

2.1. Native causes of security lacks in public wireless

Intrinsic causes deteriorate privacy of the radio-interface of public networks that are roughly summarized hereafter:
- The worldwide mass market nature of modern digital standards induces a native weakness of the broadcast signaling channel and of the early steps of radio access (that should be understood everywhere by any terminal)
- Roaming and handoff procedures of mobile handsets cause regular signaling exchanges of subscriber and/or terminal IDs for updating their registration and location. The relevant protections are often not sufficient.
- Cipher procedures within digital standards often remain limited. In addition, they may suffer of unexpected publication ([16]). In practice, confidentiality of algorithm used for authentication, key computation, ciphering within public standards cannot be warranted over years.
- Multiple standard handset and ascending compatibility of radio communication protocols are others weaknesses. In many cases, active attacks re-enforced by selective jamming can force terminals to the use of the weakest RATs.
- Other weaknesses come from sub-optimal operators practices, from subscribers' misunderstood (bad parameterization of secret key, no regular change of passwords, etc.), from legal restrictions, etc.

In the following, Alice and Bod form legitimate transmitter and receiver link and Eve is the Eavesdropper or the Radio Hacking System (RHS). Several principles of passive and active attacks are described into [31] and into the associated references. Figure 1 briefly introduces the model of the threat and the main relevant information-theoretic notions developed § 3.

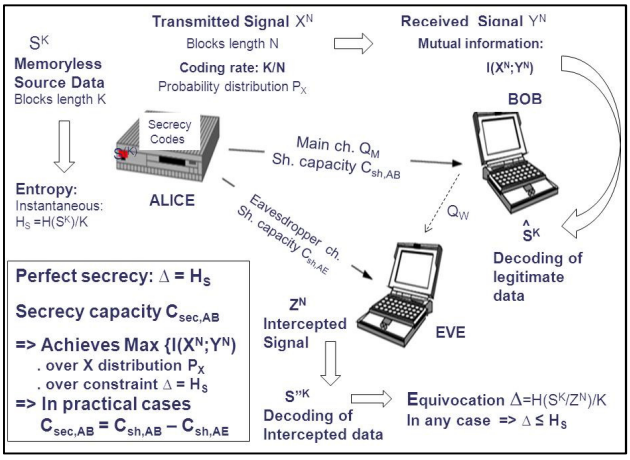


Figure 1: Model of wiretap channel - relevant information-theoretic notions (memoryless stationary source).

2.2 Transec with furtive LPI and LPD signals.

2.1.1 Principle of LPD LPI signals

"Interception" is the management (by Eve) of the concomitance of its own instantaneous carrier and bandwidth, with Alice's signal carrier and bandwidth.

. For fixed frequency signals, interception is achieved when Eve instantaneous bandwidth "meets" the signal carrier.

. For slotted signals and for hopped signals, interception is achieved when Eve's bandwidth "meets" at least one signal slot/burst during the acquisition duration.

«Detection» is a probabilistic estimation of the presence of an intercept signal, followed with a decision mechanism that maximizes the likelihood over two hypotheses H0: "signal non present" and H1: "signal present".

The main processing for detection use

- radiometer filters based on a signal power criteria, when no a priori information is available to signals
- Matched filter (inter-correlation processing) when a priori information is available to signals: this is usually the most efficient case when facing wireless standards (exceptions occur nevertheless for CDMA UL senses)

Basis of signal processing for spectrum monitoring on communication signals (including standards) can be found in [7-9]. Deeper considerations can be found in [10-11].

Signals of Low Interception Probability (LPI) avoid most of classical interception mechanisms (such as frequency scanning of low bandwidth receivers).

Frequency Hopped (FH) signal over wide frequency intervals and long periods are usual in military networks because they have good LPI characteristics - see fig 2. They are often merged with TDMA RATs (many VHF and UHF tactical radios) and with CDMA RATs (MIDS)

Signals used for opportunistic RATs within Cognitive radios (CR) or digital dividend of white space (DDWS) may have good LPI properties too thanks to their versatile spectrum access protocol (interaction with sensing capabilities and local spectrum usage) and their adaptive modulation.

Signals of Low Detection Probability (LPD) avoid most of classical detection mechanisms (such as radiometer and matched filter). DSSS signals [12] have good LPD properties when transmitted Spectrum Density Powers (SDP) remains low (thus countering radiometer), and when spreading and scrambling codes remain unknown (thus countering matched filters). Frequency hopped and Time hopped signal may have LPD properties when power remains weak (Short Range System, Ultra Wide Band RATs) and when carrier/slot allocation is random over wide periods. More generally, any highly non-stationary RAT (CR, DDWS) induce native LPD capabilities because it reduce integration duration, processing gain and association capabilities within adverse receivers.

2.2.2 Frequency Hopped (FH) signal in public wireless

GSM [3], Bluetooth and some other TDMA wireless use FH signals over a few carriers for their traffic channels. Similar procedures exist for Time Hopped Signals

The main motivation for FH in such networks is usually the use of frequency diversity (that averages effects of fading) and the flexibility of spectrum allocation within dense (urban) environments. Nevertheless, when taking place in high density network, when dealing with numerous carriers, when using unknown Frequency Hopped Sequences (FSH), TDMA/FH RATs may provide some LPI and LPD capabilities face to low bandwidth passive threats (fig 2).

Unfortunately, in many practical cases, cell frequency plans have a limited number of carriers and many of the FHS parameters remain stationary (even when ciphered - such as in GSM [3]).

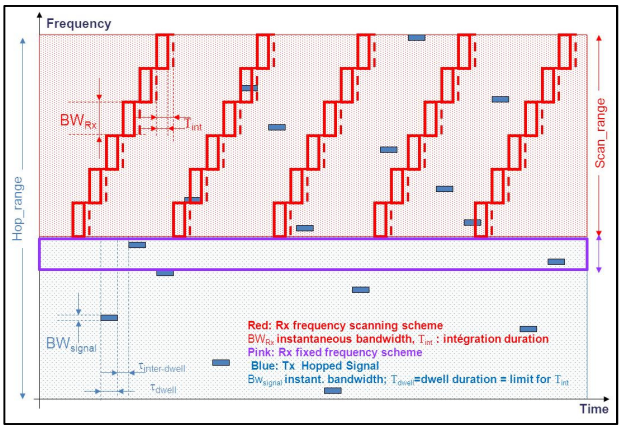


Figure 2: Low bandwidth receiver: fixed frequency or frequency scanning facing LPI frequency hopped signals

2.2.3. DSSS signals in public wireless

3GPP/UMTS, 3GPP2/CDMA2000 and several other systems such as GPS and Globalstar satellite constellations use DSSS signals (fig 3) and CDMA codes for achieving the best use of time and space diversity (Rake processing, Soft handoff) and the best flexibility of random radio access (numerous codes with flexible spreading factors).

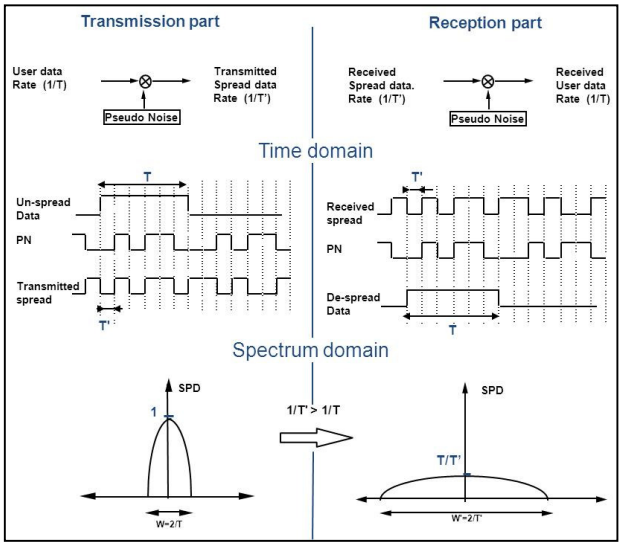


Figure 3: Principle of DSSS signals

Thanks to the fast control power, to the high combinatory of long scrambling codes, to accurate time resolution DSSS/CDMA could provide high LPI and LPD capabilities especially for uplink (weaker SPD), when facing passive and active threats. Unfortunately, in many practical cases, additional determinism is introduced into the wave form in order to facilitate the synchronization and the quality control of legitimate link, which largely deteriorates the transec:

3GPP/UMTS examples [1]: low combinatory synchronization codes (P/S-SCH) are included in the DL Frames. The combinatory of Pilot codes is then largely reduced (from 8192 to 8 hypothesis), which allows an easy and accurate recovering of Frame synchronization and DL codes (common control, paging and of traffic channels) over large distances ([27]),

Pilots symbols are included in UL traffic control channel. Their low combinatory allows an efficient external recovery of the UL slot and frame synchronization and then of the UL scrambling code for further dispreading ([28]).

3GPP2/CDMA2000 examples [2]: such networks are clocked by GPS system time, that facilitate frame cock recovery at first, and then long range detection and de-spreading of DL signaling paging and traffic channels. Then for public services, symbol scrambling and punctuation of power control bits are achieved with a sequence that is very similar to the UL spreading sequence (depending on a Long code Mask (LCM) linked to the Electronic Serial Number of the terminal). In the UL sense, the same LCM manages the long spreading code and the 64-Hadamard modulation scheme induces significant redundancy over time that facilitates the recovery of the LCM. Finally combining both DL UL analyses may provide all spreading and scrambling parameters to Eve [2][28][29].

2.3 Netsec aspects : signaling and access negotiation protocol, authentication and identification

2.3.1 Current netsec within public wireless

First of all, broadcast signaling and early access signals are usually transmitted with significant power and without transec because no power control is active at the early steps of the Radio access protocol and no transec secret could been exchanged. This facilitates long range detection and decoding of both signaling and access messages.

Moreover existing netsec protections of current wireless public network are usually very poor, providing thus much information about network engineering to any man in the middle (mitm) attacker or to external watchers of the network:

- Nowadays, no real protection of broadcast signaling applies in wireless public network neither to on-going access attempts. Thus, after decoding, their content is intelligible
- Subscriber initial registration often requires the complete IDs of terminals and/or subscribers.
- Authentication during access attempts is usually based on random parameter exchanges and on secret algorithms that are shared by terminals and by network nodes, but the integrity control of the relevant messages is often poor or inexistent.
- On-going identification and roaming procedure are usually managed with Temporary IDs (TMSI). In several standards, Ids are transmitted in clear text. Even when cipher mechanism is activated prior to TMSI reallocation (such as in GSM [3]), old TMSI are transmitted in clear text. Finally TMSI reallocation procedures in general induce a severe privacy weakness.

The examples above are detailed in [31]. They show the necessity for strong netsec upgrades of wireless public networks (especially crucial for CR). Several ways for this are introduced in the following.

2.3.2 Use of low power tag signals

Electronic marker of radio-electric source can be achieved by using DSSS weak SPD heterogeneous signals [30]. These tag signals are transmitted at the same time, at the same frame/slots and at same carrier than the user signal. Power, interference level and spreading factor of the tag signal are adjusted such a s in figure 4 in order to achieve

- Transec protection of the tag signal at first, thanks to the native interference (face to non-authorized receiver that do not know the tag spreading code),
- Easy detection and recognition of the tag signal by a suitable matched filter in any authorized receiver (thanks to the spreading factor exceeding the interference ratio).

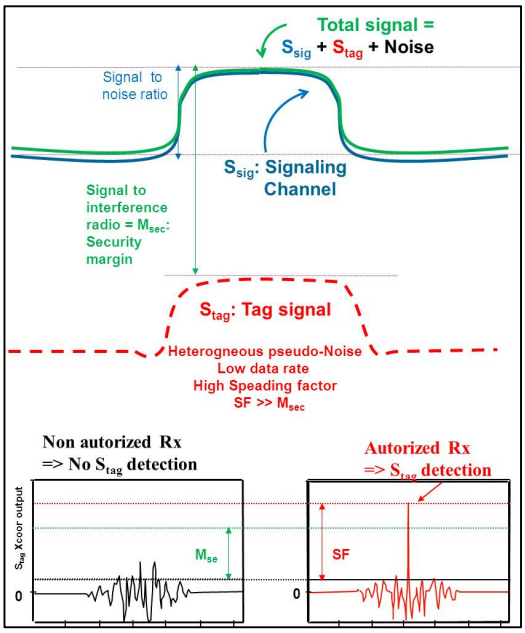


Figure 4: Principle of radio-electric tag signals

2.3.3 IFF-derived techniques

Netsec may be re-enforced with early identification such as Identification of Friend and Foe procedures. Most common Identification Friend and Foe (IFF) are fixed frequency interrogation (1 030 MHz) and response (1 090 MHz) protocols that use low duty-cycle signals for many applications such as airborne Traffic Control (both civilian – modes A C S - and military – modes 1 to 5-), battlefield surveillance, fratricide avoidance, etc. Various secure mechanisms (including time hopping, cryptography, DSSS) are developed in these systems. One challenge is to define similar procedures for public wireless.

From the current knowledge of IFF standards, several extensions could be imagined into wireless public frequency plans and into standard RATs in order to re-enforce security of access attempts:

. Electronic tag of networks nodes (§ 2.3.2) would achieve a preliminary recognition of networks by terminals and contribute thus to avoid most of active threats.

. Electronic tag of terminals would achieve a preliminary recognition of terminals before network download.

. Dedicated "early identification channels" would achieve preliminary interrogation and response before tuning into intelligible signaling paging and access messages.

2.4 Comsec aspects : ciphering of user's data

2.4.1 Providing comsec with classical cryptography

In the context of comsec and netsec protection of radio-communication with classical cryptography, the transmitter (Alice) and the receiver (Bob) share a common symmetric key that is used in authentication, integrity and confidentiality symmetric mechanisms.

Confidentiality is performed in such a way that it avoids demodulation/decoding error propagation after decipherment. The solution is based on a synchronized encryption mechanism, using

- a block cipher in counter mode
- or a Key Stream Generator that produces an pseudo random sequence to be xored to the plaintext.

When both confidentiality and integrity/authentication are required, an Authenticated Encryption (AE) scheme such as Galois Counter Mode (GCM) [14] is used to encrypt the plaintext and to compute an additional Message Authentication Code (MAC).

All these mechanisms use an Initial vector (IV). This IV may be random or a determinist nonce built with an ad-hoc frame counter and/or time reference, signalization information, addresses, physical information, etc. that are shared by Alice and Bob.

In traditional cryptography, the IV and MAC data are managed separately. Two main features must be outlined.

a) Additional bandwidth is required for security

Especially when security is applied on short durations (the frames of the radio link), the part of the bandwidth purely dedicated to security management (to transport MAC and IV materials) is often significant with respect to the part dedicated to the data itself. As examples

- The size of MAC is at least n = 64 bits for a standard security level and must be 128 bits for a upper security level. Each frame must include a MAC for integrity control.
- The length of a random IV is typically more than 48 bits to insure that the probability of IV collisions is small. When the IV is a determinist nonce based on an ad-hoc frame counter or a time reference, it should not be transmitted in each frame (depends on Alice and Bob synchronization), anyway, a synchronization pattern must be regularly transmitted to maintain the correct synchronization between Alice and Bob crypto.

b) The resilience in case of IV misuse (ex: IV repeated)

When using an Authenticated Encryption (AE) scheme like GCM mode, the security is dramatically lowered when the same IV value is used with the same key. A major point for security is therefore the real robustness of IV generation process (possibly based on the time reference, on some

frame counter, or on a random generator) to guarantee the uniqueness of the IV for each frame.

2.4.2. New cryptography concepts

In the future, some new enhanced cryptographic mechanisms may be adopted in order to provide an enhanced security at frame level (especially regarding the added data consumption and IV misuse), by managing both encryption and integrity control into a unique processing.

Good candidates may be taken from Determinist Authenticated Encryption Scheme such as the Synthetic Initial Vector mode (SIV) (see [13] and [15]).

In this kind of solution, a MAC=SIV is first computed on a message that includes the useful frame content and that can include an added context header, whose content is built for example with shared and non-transmitted signaling data. The computation itself uses a deterministic mechanism.

In a second step, the resulting MAC=SIV is also used as the IV for the encryption of the plaintext.

In this setting, the traditional couple of security data (IV, MAC) to be included in each frame is reduced to the single SIV pattern that plays both IV and MAC roles.

In [15], the specific good resilience feature of the Deterministic Authentication Encryption (DAE) scheme is outlined as following:

a) It is an IV-based AE scheme that is secure when its IV is an arbitrary nonce, not just when it is a random value.
b) In case of IV misuse (if the IV does get repeated) then
   - authenticity remains;
   - privacy is compromised only to the extent that some minimal amount of information is revealed: the almost information that may be revealed is the fact that the plaintext of a frame is equal to the plaintext of a prior frame. In addition, it is revealed only when the plaintext and the header content are equal over two frames.

3. OVERVIEW OF PHYSEC CONCEPTS

3.1. Notation and statement of the problem.

3.1.1. Entropy, Information and secrecy.

Figure 1 introduced briefly the kind of threat to be countered and the main relevant theoretic notions. Alice wants to send her data to Bob through the main channel, while Eve is an eavesdropper who overhears the signal. To maintain data confidentiality, Alice uses a secrecy code of rate K/N, namely, the input data block S^K is of length K, and the codeword length is N. Alice sends the codeword X^N. Bob and Eve receive corrupted versions of the codeword Y^N and Z^N due to channel noise or fading.

The Shannon entropy of a random signal X (of discrete values $x_1 \ldots, x_M$) is $H(X) = E_{PX}[-\log_2(Px_m)]$, where P_X ($P_X(X=x_m) \triangleq P_{Xm}$) is its probability distribution. H represents the degree of uncertainty of X. Its maximum $\log_2(M)$) occurs when X is uniformly distributed.

When considering a discrete stationary source S, and K outputs $S_1, \ldots S_K$ of S, the source entropy is then $H_S \triangleq \lim_{K\to\infty}[H(S_1, \ldots, S_K)/K]$. For more general cases, see [32].

When two random discrete signals are considered, such as X transmitted by Alice and Y received by Bob, the conditional Shannon entropy is $H(X/Y) = E_{PX,Y}[-\log_2(P_{Xm/Ym'})]$, that involves the conditional probability distribution $P_{X/Y}(X=x_m/Y=y_{m'})=P_{X,Y}(X=x_m,Y=y_{m'})/P_Y(Y=y_{m'}) \triangleq P_{Xm/Ym'}$.

The mutual information is defined by $I(X;Y)=H(X)-H(X/Y)$. The "propagation" channel being defined by the probability $P_{Y/X}$, I(X;Y) represents the amount of information about X that could be inferred after observing Y. "Perfect" channel $p_{Y/X} \equiv 1$ implies $H(X,Y)=H(X)$, $H(X/Y)=0$ and $I(X;Y)=H(X)$, while "full noisy" channel $p_{Y/X} \equiv p_Y$ implies $H(X,Y)=H(X)+H(Y)$, $H(X/Y)=H(X)$, $I(X;Y)=0$).

All these concepts lead to the definition of instantaneous (or stationary) Shannon capacity of channel $P_{Y/X}$, that is the superior bound of the mutual information I(X;Y) over any distribution P_X of the transmitted signal X:

$$C_{sh,AB} \triangleq \mathrm{Sup}\{I(X;Y)\ ;\ P_X\}.$$

For more general cases see [32].

To measure the secrecy against Eve, the equivocation is defined as the conditional Shannon entropy $H(S^K|Z^N)$, which is the remaining uncertainly of Eve about the source block after receiving Z^N. Initially, Shannon defined perfect secrecy as [17]:

$$H(S^K|Z^N) = H(S^K)$$

Which means that the signal Eve receives does not contain any information about the source data. Alternatively, perfect secrecy can be measured by information leakage, i.e., the mutual information is zero: $I(S^K;Z^N) = 0$. However, it is impractical to achieve perfect secrecy, since it essentially requires one-time pad.

3.1.2 Weak secrecy

To make secrecy coding practical, one may consider the limit of the ratio [20]:

$$\lim I(S^K;Z^N)/K = 0 \text{ as } K \to \infty.$$

It means that the average information leakage per symbol tends to zero. However, a major weakness of this notion is that the absolute information leakage $I(S^K;Z^N)$ can still tend to infinity, for example, on the order of the square root of K. This is not considered secure enough. For this reason, it is usually referred to as weak secrecy.

3.1.3 Strong secrecy

Strong secrecy overcomes this weakness by considering the un-normalized limit:

$$\lim I(S^K;Z^N) = 0 \text{ as } K \to \infty.$$

This notion of secrecy is now widely accepted in the community, and is regarded as secure enough. It can be shown that it is closely related to the standard notion of semantic security widely used in the crypto community. It is important to note that although the information leakage is

not absolutely zero, it can be made arbitrarily small by increasing the block length K. As long as the information leakage vanishes fast (for example, exponentially), the system with sufficiently large K is secure in practice.

3.1.4 Secrecy Capacity

The Secrecy capacity $C_{sec,AB}$ is defined as the maximum transmission rate of the legitimated link Alice-Bob under the constraint that secrecy is achieved with respect to Eve. It is known [18-20]:

. that the secrecy capacity is defined under the condition $C_{sh,AE} \leq C_{sh,AB}$ and verifies: $C_{sh,AB} - C_{sh,AE} \leq C_{sec,AB} \leq C_{sh,AB}$

. that $C_{sec,AB}$ is equal for weak or strong secrecy

. that in most practical cases where the channel satisfies certain symmetry, the following equality holds:

$$C_{sec,AB} = C_{sh,AB} - C_{sh,AE}$$

3.1.5 Intentional cooperative jamming

Another important direction to providing physical-layer security is the use of jamming/artificial noise. The limitation of the basic wiretap channel is that a positive secrecy capacity can be achieved only if the legitimate receiver has a better channel than the eavesdropper (see above). However, this assumption is not always true. When the receiver's channel is worse than the eavesdropper's channel, a promising technique is the use of interference or artificial noise to confuse the eavesdropper. In this scenario the transmitter has a helping interferer, which has a more detrimental effect on the eavesdropper than on the legitimate receiver [20]. Therefore, the secrecy capacity can be positive under the help of the relay.

3.1.6 Secret key generation

If Alice and Bob insist on using conventional crypto, they can use the noisy channel to generate a secret key. From a practical perspective, the design of such a scheme might be less challenging than the construction of secrecy codes. This is because the wiretap code needs to simultaneously guarantee reliability and security, while secret key generation from noisy channels makes it possible to handle them separately. It is still possible to ensure information-theoretic security if the key is used as a one-time pad. Of course, this requires the rate of the key is quite high.

3.2. Secrecy capacity for ideal channels.

3.2.1 Binary symmetric SISO channel

Binary symmetric channel (bsc) is defined for a binary source X=S (="0" or "1", with probability p_{x0} and $1-p_{x0}$), and crossover probabilities when X is transmitted:

. for the legitimate P(Y≠X)=p and P(Y=X)=1-p.

. for the eavesdropper P(Z≠X)=q and P(Z=X)=1-q.

When $p < q \leq 1/2$, the relevant Bernoulli law entropies verify $H(p)<H(q)$ and the secrecy capacity of the bsc wiretap channel is:

$$C_{sec,AB} = C_{sh,AB} - C_{sh,AE} = [1-H(p)]-[1-H(q)] = H(q)-H(p)$$

3.2.2 Gaussian SISO Channel

The Gaussian channel is defined by $Y(k) = \alpha.X(k)+n(k)$,

. x being a memoryless stationary signal of power $\pi_x=E_{PX}[X(k).X^*(k)]$, submitted to propagation attenuation α

. n being a X-independent centered Gaussian noise of variance σ^2, thus $P(Y=y/X=x)=P(n=y-x)=\exp[-(y-x)^2/2/\sigma^2]$.

. The signal-to-noise ratio (SNR) is $\rho_{SNR} = |\alpha|^2.\pi_x/\sigma^2$

Considering now the main channel defined by α_m and σ_m^2, the eavesdropper channel defined by α_e, and σ_e^2, the secrecy capacity of the Gaussian wiretap channel is given by:

$$C_{sec,AB} = C_{sh,AB} - C_{sh,AE} = \log_2(1+\rho_{SNRm}) - \log_2(1+\rho_{SNRe})$$

Where $\rho_{SNR,AB}$ and $\rho_{SNR,AB}$ ($\rho_{SNRm} > \rho_{SNRe}$) are the SNRs of the main and eavesdropping channel, respectively.

3.2.3 Secrecy capacity of Rayleigh Channels

The Rayleigh channel is defined by $y(k) = H(k).x(k) + n(k)$, H being a fading coefficient following Gaussian centered law (attenuation $|H|^2$ follows an exponential probability distribution over $[0 +\infty[$). In the following, secrecy capacity is defined for a power strategy of Alice who is supposed to perfectly know the fading coefficients H_m and H_e and the noise variance σ_m^2 and σ_e^2 of the main and eavesdropping channels, respectively (full channel state information). Alice adapts its transmitted power $\pi_x(H_m,H_e)$ over the max power constraint Π, to the instantaneous signal to noise ratios at Bob's and Eve's part $\pi_x(H_m,H_e)|H_m|^2/\sigma_m^2$ and $\pi_x(H_m,H_e)|H_e|^2/\sigma_e^2$, respectively. The relevant secrecy capacity is thus given by:

$$C_{sec,AB} = \max_{\pi_x(H_m,H_e)\leq\Pi} \left\{ \log_2\left(1+\frac{\pi_x(H_m,H_e)|H_m|^2}{\sigma_m^2}\right) - \log_2\left(1+\frac{\pi_x(H_m,H_e)|H_e|^2}{\sigma_e^2}\right) \right\}$$

3.2.4 Secrecy capacity for realistic SISO and SIMO Channel

In realistic cases, it is essential to investigate the degree of security generated from the channel, which strongly depends on the nature of the multi-link channel between Alice, Bob and Eve. Several strategies can be developed in order to maximize the security, e.g. by exploiting as much as possible the degrees of freedom of the channel and thus avoiding as much as possible channel knowledge leakage to Eve. All dimensions providing at least partially uncorrelated channel gains can be used, in the frequency (for wide band), time (for time variant) and spatial (for multi-antenna) domains.

3.2.5 Extension of secrecy capacity to MIMO Channels

The MIMO channel is defined by $Y(k) = \mathbf{H}(k).X(k) + N(k)$, X being a vector signal over an transmitter antenna (size $N_{Tx}.1$), Y being a vector signal over a receiver antenna (size

N_{Rx}.1), N being a noise vector (size N_{Rx}.1) of covariance matrix $\sigma^2.I_{NRx}$ (size $N_{Rx}.N_{Rx}$), **H** being a propagation matrix (size $N_{Rx}.N_{Tx}$).
Let $\mathbf{H}_m$ and $\mathbf{H}_e$ denote the channel matrices of the main and eavesdropping channel, respectively. The secrecy capacity of the Gaussian MIMO wiretap channel is

$$C_{\text{sec},AB} = \max_{tr(Kx)\le\Pi} \left\{ \log_2 \left| \mathbf{I} + \frac{1}{\sigma_m^2} \mathbf{H}_m \mathbf{K}_X \mathbf{H}_m^+ \right| - \log \left| \mathbf{I} + \frac{1}{\sigma_e^2} \mathbf{H}_e \mathbf{K}_X \mathbf{H}_e^+ \right| \right\}$$

Where the maximization is carried over all positive semi-definite matrices $\mathbf{K}_X$ such that the power $tr(\mathbf{K}_X) \le \Pi$ is satisfied.

3.3. Coding schemes that approach secrecy capacity.

The vast majority of work on physec is based on non-constructive random-coding arguments to establish the theoretic results. Such results demonstrate the existence of codes that achieve the secrecy capacity, but are of little practical usefulness. In recent years, significant progress has been made on the construction of practical codes for physec, to a more or less extent. The design methodology can be traced back to Wyner's work on coset coding [18].

3.3.1 Low Density Parity-Check Codes (LDPC)
LDPC codes have been used to build wiretap codes, with limited success. When the main channel is noiseless and the wiretap channel is the binary erasure channel (BEC), LDPC codes for the BEC, was presented in [21,22] and proved to achieve secrecy capacity.

3.3.2 Polar Codes (PC)
In the meantime, polar coding seems to offer a more powerful approach to design wiretap codes. In [23], it was shown that, with a minor modification of the original design, polar codes achieve strong secrecy (and also semantic security). However, they could not guarantee reliability of the main channel when it is noisy.

3.3.3 Lattice Codes (LC)
The aforementioned designs based on LDPC and polar codes only tackled discrete channels, yet the physical channels are continuous. For Gaussian wiretap channels, lattice coding is emerging as a prominent approach to implement information-theoretic security. In [24], the weak-secrecy rate for lattice coding over Gaussian wiretap channel was derived. The notion of secrecy gain was introduced in [25], which has great practical significance as a criterion to design wiretap lattice codes. It has also extended to fading channels later. In a more recent paper [26], semantically secure lattice codes were proposed.

4. THEORETIC ADVANTAGES AND PRATICAL EXPECTATIONS OF INFORMATION THEORIC SECURITY CONCEPTS

4.1. Theoretic advantages of secrecy coding

Unlike conventional cryptography, secrecy coding simultaneously provides capacity and security without resorting to computational hardness assumptions (which are often unproven in practice). Even if the eavesdropper has unlimited computation power, it is impossible to break the code, because physec comes from the Shannon capacity difference of the channels. Therefore, physec is resilient to the would-be forthcoming quantum computation attacks.

4.2. Determination of secrecy codes, even sub-optimal, that approach secrecy for real field radio-environments

However, there is still a long way to go in the direction of physec. The state of the art suffers a number of significant shortcomings. In particular, LDPC and polar codes are limited to some special channel models, while explicit design of wiretap lattice codes is still lacking.

4.3. Complexity and embedding constraints

In principle, the complexity of secrecy coding is the same as that of conventional channel coding. Thus, it may be seamlessly integrated into an existing communication system. However, the state of the art does not offer such a code for real radio environments yet. In addition, when code lengths are great, practical applications are reduced for burst signals or for short messages services.

4.4. Practical perspectives of physec: towards a merging of secrecy codes with existing protections.

Further, physec can be complementary to existing transec netsec and comsec protocols. At the very least, it offers another layer of protection to vulnerable wireless communications. Therefore, there is a strong potential that physec can be merged with existing protections.
There are many open problems in this direction. In addition to the design of explicit wiretap codes, the issue of attacks warrants more attention. So far, only passive eavesdropping is assumed in most of the literature, and often, an implicit hypothesis is done that legitimate links is established, and channel propagation measured for applying secrecy codes. Finally, it seems that the threat of active attacks has not been considered carefully neither the early steps of radio access protocols.

4.5 Secrecy Coding for fading channels

A layered broadcast approach may be used when the channel is varying. The basic idea is to employ multi-layered coding to encode information into a number of layers and use stochastic encoding for each layer in order to keep the corresponding information secret from an eavesdropper. The advantage of this approach is that the transmitter does not need to know the channel states to the legitimate receiver and the eavesdropper, but can still securely transmit certain layers of information to the legitimate receiver. The layers that can be securely transmitted are determined by the channel states to the legitimate receiver and the eavesdropper. So, in practice, the data that have to be transmitted are ranked in such a way that the bits which have to be the most secure will be encoded in the layers corresponding to the most critical channel (finest granularity), and so on. This approach guarantees the best security for data as a function of Eve's instantaneous Signal-to-Noise Ratio.

5. PERSPECTIVES OF PHYSEC INSIDE WIRELESS NETWORKS

Many radio measurements that are needed in nodes and terminals for achieving communications: equalization of SISO SIMO MIMO RATs, RAKE processing of CDMA RATs, control of Quality of Service (QoS), sensing procedures and adaptive modulation/coding schemes of cognitive radios, etc. The relevant information may provide added protections based on the relevant physical randomness during access phases and during established calls. We list below some perspectives for privacy upgrades.

5.1. Re-enforce transec with adaptive resource allocation

Existing transec protections, such as selection of FHS inside TDMA RATs and selection of scrambling/spreading codes inside CDMA RATs (§ 2.2) should be highly improved by adding physical randomness into the resource allocation of legitimate links. Adaptive resource allocation for establishing radio-links usually induce high disturbance at the eavesdropper part, especially in dense networks. Therefore, combining signal mixtures (full duplex RATs, MISO, MIMO, artificial jamming) and physical-dependent allocation process should be an efficient alternative: the resource allocation would thus depend on both propagation channel and interference level at Alice and Bob parts. Moreover, sensing outputs of cognitive and opportunistic radios would induce more versatility.

5.2. Upgrade netsec with “tag channels”

Privacy of early negotiation protocols should be highly improved by taking advantage of existing (high power) signaling channels that are broadcasted by network nodes, with added heterogeneous tag signals sharing the same carrier and slots. The existing broadcast channels, being initially protected by dedicated codes or encryption schemes (thus not intelligible), would play the role of cooperative jammers. DL and UL tag signals of DSSS type, of low data rate, of low SPD and of high spreading factor would be transmitted “under” the broadcast channels by following the principles of §2.2.2 and § 2.2.3. These tag signals would support early radio exchanges such as the following:

- Terminal’s and node’s preliminary identification based on the spreading codes and on low data rate spread messages
- Channel measurements based on the spreading codes (such as in rake receiver techniques)
- Computation of secrecy codes at terminal and node part (supported by channel measurements)
- Exchange of acknowledgement messages.

Then after successful acknowledgement,

- Broadcast signaling would commutate into intelligible text
- Radio access would continue such as specified in the standard by adding “tag channels” under each “main channel”. Tag channels would apply computed secrecy codes that would be adapted to radio environment during the process; even normal channels could apply secrecy codes. Moreover, information could be shared among main rate channel and (low rate) tag channels. Finally
    - each of the main channel would plays the role of a cooperative jammer for the associated tag channel,
    - the integrity control of each of the main channels would be achieved thanks to its associated tag signal,
    - the most private data (that remain low rate), such as subscribers IDs, encryption characteristics for future traffic messages etc., would be transmitted by the (more protected) tag channels.

The whole procedure should highly disturb:

- Any passive eavesdropper thanks to the native jamming of tag channel and to thank the added secrecy coding,
- Any active eavesdropper because of the heterogeneous nature of tag signals and because of native advantages of DSSS signals (spreading factor and time resolution). Note that even if the computed secrecy codes remains suboptimal for capacity of the legitimate link, this has no great importance for the tag channels that remain low data rate.

5.3. Merge physec and advanced comsec schemes

Within established traffic messages, combining SIV computation of advanced AE (see§ 2.4.2) and propagation-dependent random issued from receiver processing (equalization, rake etc.) appear as a promising way. In such a privacy improvement, outputs from receiver processing and from radio measurements (QoS estimates etc.) would be taken into account in order to build part of the context header of each frame.

6. CONCLUSION

In this paper, we introduced several concepts based on physical layer properties that may compensate security lacks occurring in civilian wireless networks when facing passive eavesdropper and active radio hacker systems.

By focusing on practical perspectives of secrecy codes in realistic radio-environments, we pointed out that the best perspectives for significant privacy upgrades for wireless networks rely in merging of traditional techniques, of cooperative jamming and of advanced channel codes (involving secrecy coding concepts) in order to build:

- Protected signaling channels
- Confidential negotiation schemes in the early stages of the radio access protocol
- Enhanced ciphering schemes that involve added propagation dependent random sources.

By this, it should be possible to hardly penalize any basic eavesdropper and radio-hacking systems that more or less re-use existing radio-components and protocol stacks.

Moreover, we conjecture that even facing advanced threats, the protection principles above should be efficient, relevant to privacy aspects, when they exploit radio-environment advantages that can be catch and/or generated locally by legitimate base stations, communication nodes and terminals for their proper communication services. The core idea is to convert in secrecy benefits

- radio-interferences and strong signals (such as signaling channels) that are present in the radio spectrum.
- information got by sensing, by equalization processing and by QoS management at each termination point of the radio link,

Based on this information, the final achievement is to embed adaptive modulation and coding schemes:

- that approach mean channel capacity in realistic radio-environment
- that maximize confusion at any threat location,
- that keep implantation complexity compatible with the performances of the future embedded computers.

We are confident that current national and European research programs will discover and proof feasibility of such adaptive modulation and coding schemes thus preparing standardization and industrial development of trustworthy and full-secure public RATs.

7. REFERENCES

[1] www.3GPP.org

[2] www.3GPP2.org

[3] Lagrange (X.), Godlewski (P.) and Tabbane (S.). Réseaux GSM. 5e édition, Éditions Hermès (2000).

[4] Bluetooth Vulnerability Assessment Tech. Publication ITSPSR-17A, Communications Security Establishment Canada, 06/2008.

[5] http://fr.wikipedia.org/wiki/Identification_friend_or_foe

[6] ITU-R SM 1600 « Technical identification of digital signals »

[7] QoSMOS project "Radio Context Acquisition algorithms » Deliverable D3.3. FP7-ICT-2009-4/248454

[8] F. Delaveau, D. Depierre, F. Sirven «Oriented processing of Communication signals for Sensing and Disseminated Spectrum Monitoring», SDR Winncomm Forum 2011 Brussels.

[9] F. Delaveau Y Livran "Radiosurveillance du spectre - TE 6890 Rôles et tendances, TE 6891 Interception Réception et Détection, TE 6893 Analyse et identification des transmissions» in Techniques de l'Ingénieur. 2012.

[10] J.G. Proakis, M. Salehi Prentice Hall Int. Ed2 2001., "Communication System Enginnering".

[11] H.L. Van Trees, "Detection, estimation and modulation Theory". Ed. J. Wiley (1968).

[12] Pickholtz (R.L.) and al. – Theory of Spread Spectrum Comm - A Tutorial. IEEE Trans. Com., vol. com 30-5,1982.

[13] D. Harkins "Synthetic Initialization Vector (SIV) Authenticated Encryption Using the Advanced Encryption Standard (AES)", 2008.

[14] McGrew, D. and J. Viega, "The Galois/Counter Mode of Operation (GCM)".

[15] Rogaway, P. and T. Shrimpton, "Deterministic Authenticated Encryption, A Provable-Security Treatment of the Key-Wrap Problem", Advances in Cryptology -- EUROCRYPT '06 St. Petersburg, Russia, 2006.

[16] B. Shneier "Applied Cryptography, Protocols, Algorithms, and Source Code in C". Wiley.

[17] C. E. Shannon, "Communication theory of secrecy systems," Bell System Technical Journal, vol. 28,pp. 656–715, 1948.

[18] A. D. Wyner, "The wire-tap channel," Bell System Technical Journal, vol. 54, no. 8, pp. 1355–1367, October 1975.

[19] U. Maurer, S. Wolf, « Information-theory Key afgreement From Weak to Stong secrecy for Free. EUROCRYPT 2000, International Conference on the *Theory* and Application of Crypto. Techniques

[20] M. Bloch and J. Barros, Physical-Layer Security: From Information Theory to Security Engineering. Cambridge University Press, 2011.

[21] A. Thangaraj, S.Dihidar, A.R.Calderbank, S.W.McLaughlin, and J.Merolla, "Applications of LDPC codes to the wiretap channel," *IEEE Trans. Inform. Theory,* vol. 53, no. 8, pp. 2933-2945, 08/2007.

[22] A. Suresh, A. Subramanian, A. Thangaraj, M. Bloch, and S.W.McLaughlin, "Strong secrecy for erasure wiretap channels," *Proc. IEEE Inf. Theory Workshop*, Dublin, Ireland, September 2010.

[23] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," IEEE Trans. Inform. Theory, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.

[24] L.-C. Choo, C. Ling, and K.-K. Wong, "Achievable rates for lattice coding over the Gaussian wiretap channel," in *ICC 2011 Physical Layer Security Workshop*, 2011.

[25] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," Mar. 2011.

[26] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, "Semantically secure lattice codes for the Gaussian wiretap channel," 2012.

[27] « Procédé d'optimisation de la planification dans un système de communications de type CDMA » Patent Thales FR 04.01475

[28] « Procédé ''protocole orienté'' de traitement des signaux stationnaires, partiellement stationnaires, ou cyclo-stationnaires, patent Thales FR 10.05017, PCT/EP2011/073420 WO2012.084956

[29] R. Gautier, G. Burel, J. Letessier and O. Berder « Blind estimation of Scrambler offset using encoder redundancy », in IEEE 2002

[30] « Procédé de taggage radio-électrique des signaux de brouilleurs et d'autres émetteurs» Patent Thales FR 12.03071

[31] F. Delaveau, A. Evestti ; A. Kotelba; R. Savola, N. Shapira; "active and passive eavesdropper threats within public and private civilian networks-existing and potential future countermeasures – an overview". Winncomm2013, Munich

[32] T.M Cover, J.A Thomas « Elements of Information Theory», Wiley 1991

ACTIVE AND PASSIVE EAVESDROPPER THREATS WITHIN PUBLIC AND PRIVATE CIVILIAN WIRELESS NETWORKS - EXISTING AND POTENTIAL FUTURE COUNTERMEASURES – AN OVERVIEW.

François Delaveau
(Thales Communications & Security; Gennevilliers, France; francois.delaveau@thalesgroup.com);
Antti Evesti ; Jani Suomalainen; Reijo Savola
(VTT Technical Research Centre; Oulu, Finland; Antti.Evesti@vtt.fi; Jani Suomalainen@vtt.fi; Reijo.Savola@vtt.fi);
Nir Shapira
(Celeno Comnunications Ltd; Ra'anana, Israël; Nir.Shapira@celeno.com).

ABSTRACT

This paper aims at providing an overview of threats that may deteriorate security level and trust in public wireless networks, because of eavesdropper and hacking technologies that operate at the radio interface, and aims at providing an introduction to relevant counter-measures that deal with "physical based" security in a large sense (Physec). We highlight selected promising Physec technologies that are expected in the future years by mixing classical protections and advanced issues of information-theoretic security, secrecy coding and cooperative jamming. These particular items are studied and developed in the PHYLAWS project (EU FP7-ICT 317562, www.phylaws-ict.org), starting Nov. 2012, which supports this work.

1. INTRODUCTION

Given the growing prevalence of wireless radio-communication technologies, the sufficient confidentiality, integrity, availability and reliability a person or an organization can have of the exchanged information is a major societal challenge for both personal and professional sphere. Moreover, the growing importance of sensing procedures and of pilot channels in future radio access technologies (white spectrum, cognitive networks), will result in numerous radio-transmissions of geo-referenced spectrum allocations and of radio engineering data, whose integrity and confidentiality are major industrial challenges for both operators and administrations. Security of radio-interface within wireless networks appears now as crucial for many applications such as broadband internet, e-commerce, radio-terminal payments, bank services, machine to machine, health/hospital distant services. Most of citizens, professionals, stakeholders, services providers and economical actors are thus concerned by confidentiality lacks and by privacy improvements of the physical layer of wireless networks

This paper first introduces the current radio access technologies and describes briefly the main security protocols that are used at the physical layer of wireless public networks, such as subscriber authentication, control of message integrity, and cyphering procedures of messages' content.

From several known examples inside radio-cells (GSM, UMTS, LTE), Wireless Local Area Networks (WiFi), Short Range Communications (Bluetooth, ZigBee), etc., we will focus on the main failures that may occur in existing security procedures and discuss their multiple causes. By considering these weaknesses, we will then describe possible threats during the initial access attempts, during negotiation protocols and during established calls. We will take into account both passive (radio-eavesdropper) and active (radio-hacker) attacks. Nevertheless, in order to avoid any paranoiac or angelic caricature, we will also consider the (severe) practical radio attack limitations that are caused by complex radio environments in many real field situations.

Then, existing countermeasures for improving security, thrust and privacy within wireless networks will be introduced, by distinguishing radio-signals (transmission security), signaling message content (network security), and content of users' messages (communication security). Advantages and drawbacks of these procedures regarding public worldwide use will be discussed. Additional elements about secure architectures for radio terminals and about risk-driven security metrics will be given too

Finally, we will introduce new protection concepts for radio-communications that exploit the physical properties of radio-environments. Especially when complex dispersive and non-stationary, radio propagation has to be measured by infrastructures and handsets: equalization, RAKE processing, MISO/MIMO coding schemes, sensing procedures of cognitive radios (CRs), etc. The relevant physical information provides significant opportunities in order to enhance security algorithms and protocols during access phases and during established calls.

2. OVERVIEW OF EXISTING PUBLIC RADIO ACCESS TECHNOLOGIES

2.1. Main class of public radio networks and relevant radio characteristics

Figure 1 illustrates the large variety of signals to be taken into account nowadays for privacy considerations in Ultra/Special High Frequency (300 MHz - 3 GHz – 6 GHz)

System	Uplink frequency plan (MHz)	Downlink Frequency plan (MHz)	Channel spacing	Modulation UL - DL	Radio Access Technlogy	Access mode	Range of Terminal Power	Typical propag. Range	ref standard
GSM 900	890 - 915	935 - 960	200 kHz	GMSK + variants	TDMA/FDMA	Aloha	2 W	100 m to 3 km	ETSI
DCS 1800	1710 - 1785	1805 - 1880							
PCS 1900	1850 - 1890	1930 - 1970							
UMTS	890 - 915	935 - 960	5 MHz	(OC) QPSK	DSSS/CDMA FDD and TDD, MISO	Aloha	0,25 W	10 m to 3 km	3GPP
	1920 - 1980	2110 - 2170							
LTE	890 - 915	935 - 960	1,4 - 5 MHz	OFDMA and SC-FDMA	FDD and TDD, MIMO	Aloha	0,25 W	10 m to 3 km	3GPP
	2500 - 2570	2620 - 2690							
IS-95 A/B	824-844 1850 - 1890 other	869-889 1930 - 1970 other	1,25 MHz	OQPSK - QPSK	DSSS/CDMA	Aloha	2 W	100 m to 3 km	3GPP2
CDMA2000 SR1/3GPP2			5 MHz						
CDMA2000 SR3/3GPP2			5 MHz						
WIMAX	2402 - 2480		10 MHz	OFDM and QPSK/CDMA	TDD, SC-OFDMA, MIMO	CSMA/CA	0,25 W	1 to 15 km	IEEE 802.16xxx
	3400 - 3600								
	5150 - 5850								
WIFI L band	2402 - 2480		20 MHz	OFDM and QPSK/MC-CDMA	TDD, MIMO	CSMA/CA	0,1 W	indoor	IEEE 802.11xxx
WIFI C band	5150 - 5850		20 - 80 MHz						
Bluetooth	2402 - 2480		157 kHz	0,5 BT GFSK	TDMA/TDD	CSMA/CA	0,01 W	indoor	IEEE 802.15.1
Zigbee	868 - 868.6		2 et 5 MHz	ASK, BPSK, O-QPSK, MSK	CDMA/TDMA	CSMA/CA	0,01 W	indoor outdoor < 50 m	IEEE 802.15.4
	902 - 928								
	2400-2483.5								
DVB-T		470-862	8 MHz	COFDM	FDD, MISO		20 - 200 kW	>> 10 km	ETSI

Figure 1: Public network to be improved relevant to privacy.

2.2. Main class of RATs – Early signaling exchanges

Roughly mains RATs can be shared in four classes (fig. 2):

FDMA: (Frequency Division Multiple Access)

- Signal repartition over frequency
- Exemples are 1G standards: NMT, AMPS, etc.
- Propagation equalization is required in receivers
- Hopped/opportunistic frequency variants: Military, ALE (HF)

TDMA: (Time Division Multiple Access)

- Signal repartition over time slot
- TFDMA/FDMA variant with hopped frequency
- Propagation equalization is required in receivers
- Examples are 2G public standards (GSM, D-AMPS), WLAN 802.11b, short range (Bluetooth, DECT), and most of tactical VHF Military ad-hoc networks

CDMA: (Code Division Multiple Access)

- Signal Repartition over spreading codes
- Receiver Rake processing
- CDMA/FDMA/TDMA variants with hopped frequency / slots
- Examples are 3G public standards ([1], [2]), and several UHF and SHF Military ad hoc networks (ex: MIDS).

OFDM: (Orthogonal frequency Division Multiplex)

- signal multiplexing over frequency
- simplified equalization within receivers
- numerous examples: DVBT/H, DRM, LTE, Wifi, Wimax
- advanced planning capabilities: Single Frequency Network;, MISO and MIMO
- derived RATs: COFDM, O-FDMA, SC-FDMA, SC-FDE

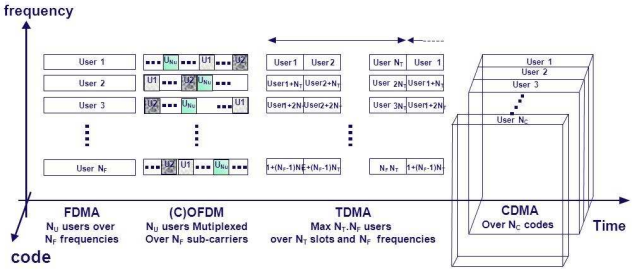


Figure 2: main radio access schemes

Note that in any case, establishing radio links involves early signaling exchanges among infrastructures, nodes and terminals, which are summarized on fig. 3 hereafter.

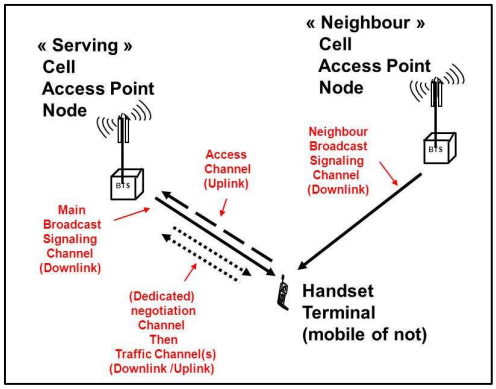


Figure 3: signaling exchanges before and during radio-access

3. SECURITY LACKS OF PUBLIC NETWORKS

3.1. Native privacy weaknesses of worldwide standards

The worldwide nature of modern digital standards induces intrinsic privacy lacks of the early negotiation protocol. Because it must remain simple and generic everywhere and for any subscriber, it is often achieved with unprotected broadcast signaling and access channels (through beacon frequencies, pilots codes, etc.) that provide local system time, easy decoded network parameters, frequency planning. Exchanges of subscribers' identifiers are required for registration, and they are recurrent for roaming and handoff of mobile (radio-cells, PMR, DVB-H, etc.). The frequent use of temporary identifier (such as TMSI in GSM) appears as a poor privacy improvement in many practical cases, even when ciphered (see below).

FDMA, TDMA and ODFM based RAT signals, especially when including synchronization midamble or words, are easy to detect and to demodulate in both DL and UL sense. Within CDMA, synchronization codes or pilots symbols, especially when clocked by GPS system time, allow easy detection and de-spreading of DL signaling and of UL access channels. Pilots symbols included in traffic facilitate both DL and UL synchronization recovery and de-spreading.

All these facilities fasten terminals computations, but they make passive and active attacks easier for recovering of synchronization, for decoding of broadcast and negotiation channels and for demodulation of traffic signals ([10],[14]).
In the following, we will consider the notations and geometry of fig. 4: legitimate link is Alice to Bob, Eve being the eavesdropper or the radio-hacking system.

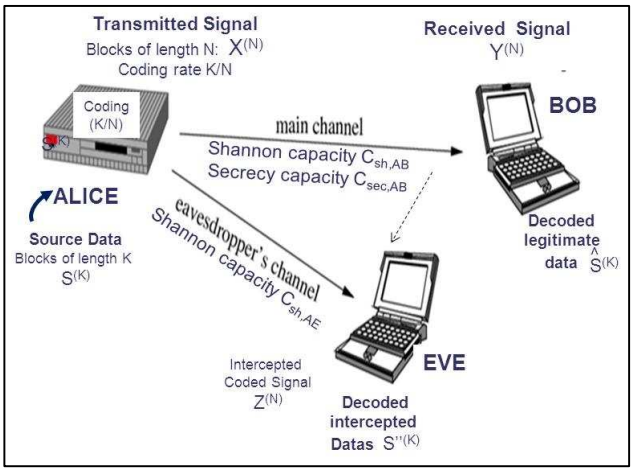


Figure 4: legitimate and eavesdropper geometry - Notations

3.2. (Non-exhaustive) examples of privacy weaknesses in modern public standards

Radio-cells authentication and location attack: When authentication is single sense only or weak, mobiles may be spoofed by using a virtual base station through forced roaming or paging procedures (that succeed when transmitted power is enough to overhead the propagation losses and the cell re-allocation criteria, when carrier and content of broadcast messages is convenient, etc.). These basic active attacks apply to 2G networks such as GSM [12], to many practical cases of 3GPP3/GPP2 and of 4G networks when operators apply weak Auth procedures. As a result ([22]), paging messages from RHS towards idle mobile stations can initiate connections and track location of mobile stations. These messages are selectively sent to the tracking area that the terminal is known to locate in unprotected format: paging messages identify the target terminal by using permanent Ids (IMSI) or temporary IDs (TMSI) which change only when user changes of location update area. A variant consist to initiate call requests (with MSIDSN) to victim terminals and then passively monitor paging messages whether the user locates in the monitored area. Disconnecting paging and call request is even possible before victim's terminal alerts, thus keeping attacks stealthy.

Subscriber's or terminal's dependent resource allocation: In some CDMA standards, close relationships occur between the allocated code for traffic and identifiers of the subscriber or the terminal. Moreover strong dependence of DL and UL links may exit, such as in 3GPP2 public mode (see [2] [14]): long scrambling codes masks (LCM) manage both UL spreading and DL scrambling among users, with and injective dependence on the Electronic Serial Number of terminals; this determinism may highly facilitate users' selective interceptions by both passive and active threats.

Pilot symbols within traffic CDMA channel: many CDMA signals, such as UMTS traffic channels [1], include low combinatory pilots symbols that highly facilitate exhaustive tests for slot and frame synchronization, for recovery of scrambling codes, in both DL and UL senses [10].

Sub-optimal module order in the transmission chain: In the transmission chain of several standards ([1][2]), coding occur before ciphering and modulation redundancies occur before scrambling. This usually facilitates key attacks.

Unexpected publications of cipher algorithm – GSM example: in this case successful cypher attacks may be facilitated. This occurred in the late 90s' for GSM A5/1-2 cipher algorithm (max key length is 64 bits), and for A3/A8 algorithm that compute authentication results (RES), temporary identifier (TMSI), and cipher keys (Kc) from internal Keys (Ki coded inside SIM card) and from random parameters (NRAND) that are transmitted over the air [16] [17]. In practice, as a result of economic competition and of hacker activities, full secrecy of wireless standards cipher algorithm can never be warranted over numerous years.

4G Networks: A number of potential attack vectors have been identified in [18]: Firstly, efficient jamming attacks can target OFDM pilot tones [19] [20], which are used to correct channel effects and to equalize transmission. Secondly, uplink channel quality information, which is adaptively used in base station to select modulation and coding for the downlink, may be targeted in Denial of Service (DoS) attacks. Thirdly, large amount of 'virtual terminals' (i.e. 'Sybil' identities) can be used to affect base stations resource usage and, in coordinated attacks, to achieve more resources for attacker's terminal. Primary user emulation attacks [21] can also be used in spectral herding to guide a victim into the wanted channel, which is chosen in order to facilitate man-in-the-middle (mitm) attacks. LTE Evolved Packet System–Authentication and Key Agreement solution (EPS-AKA) has been considered to be vulnerable for mitm attacks as it discloses the permanent identifier (IMSI) by sending it in clear text during the first connection [25].
Location privacy attacks such as identified in GSM in [22] are also applicable in LTE [23]. As cell size can in LTE be small the users' locations can thus be resolved in high detail.
The authentication and confidentiality of LTE is based on permanent security associations i.e. long term symmetric keys shared by terminal and network. Consequently, as noted in [24] the key derivation procedure in EPS-AKA

does not provide perfect forward secrecy. If a symmetric key is revealed, all derived session keys and content protected with these keys can be compromised. Moreover, EPS-AKA is backwards compatible with older authentication mechanisms and, therefore, an attacker may gain an access to the LTE network by utilizing security weaknesses found from the GSM or UMTS security algorithms [26].

Bluetooth and ZigBee Short range communications:
Bluetooth is intended to establish wireless ad-hoc networks by means of short range radios. Bluetooth is widely used to connect peripherals to computers and mobile devices. The complexity of the Bluetooth specification causes challenges for security [28]. Moreover, National Security Agency (NSA) lists following threats related to Bluetooth: identity detection, location tracking, DoS, unintended control and access of communication channel and unauthorized device control and data access. Furthermore, National Institute of Standards and Technology (NIST) lists the following Bluetooth specific attacks in its Bluetooth security guide [29]: Bluesnarfing makes it possible to gain access to data stored in a device, Bluejacking makes it possible to send messages for a Bluetooth device, Bleubugging offers access to data and device commands, Car Whisperer makes it possible to send audio to car's audio system and eavesdrop via car's microphone, Fuzzing attacks to send malformed data to device and observe device's behavior in order to reveal possible vulnerabilities in the Bluetooth stack, Pairing Eavesdropping to determine secret keys for data decryption, and Secure Simple Pairing Attacks to cause mitm attacks.
In [28], NSA states that Bluetooth should offer adequate security for situations where unclassified data is handled. In other words, Bluetooth is not applicable for classified information. Default or inappropriate passkeys are one important issue that has enabled attacks towards Bluetooth devices. However, from the physical point of view these relate to upper layers. Nevertheless, Bluetooth is intended for short range communication but the directed high gain antenna may offer signal reception over kilometers [30]. Thus, appropriate encryption is needed in order to avoid eavesdropping and also means to mitigate threats related to traffic analysis. Bluetooth utilizes a Frequency Hopping transmission mode, which actually does not offer much security in the physical layer: Frequency Hopping Sequence is delivered in a clear form during the link establishment, and thus, near devices are able to capture this information [30]. Lastly, it is known that random number generation in the Bluetooth is weak [30]. From the physic viewpoint, this issue can be improved by utilizing any random features of the communication channel.

ZigBee is intended to establish ad-hoc networks, where a low data rate and long battery life are perquisites. In ZigBee, security of the whole network depends on a master key. Thus, achieving the master key threatens the whole network. ZigBee security is investigated from the protocol and implementation viewpoints alike – where the protocol refers to security capabilities of the IEEE 802.15.4 and implementation for manufacturers' implementations. Most of the security risks are due to the implementation made by equipment manufacturers. Three main categories of attacks against ZigBee are physical attacks, key attacks, and replay and injection attacks [31]. From these categories, physical attacks are not performed via network, i.e. attack requires physical access to the programming interfaces of a device.
In addition, the minimal session checking of ZigBee makes it possible to mimic legitimate nodes.
Key attack is another well-known failure in ZigBee [31] [32]. It uses commercial traffic capturing device in order to collect wireless transmissions and analyses the collected data by means of KillerBee [33]. Based on the traffic analysis, such an attack is able to get network key.
Lastly, even if ZigBee is intended to support a long battery lifetime, jamming attacks are able to drain batteries faster than initially assumed. For instance, [32] presents an attack to abuse poll requests in a ZigBee system that prevents the utilization of the sleep mode, which in turn may cause power failures in ZigBee nodes/actuators.

WLAN: the direct use of subscriber identifiers or MAC address in WiFi registration procedure occur intrinsic vulnerability regarding user's privacy. WiFi encryption is applied on frame's payload only and not on MAC header which is present in all frames, thus user's privacy and identity is inherently compromised. This is particularly useful for Eve to classify traffic according to source-destination pairs. Until very recently, WiFi protocol management frames were not encrypted at all, thus exposing the network to various sorts of active attacks and DoS. In 2009 IEEE has standardized 802.11w, which defines encryption of management frames. Still, some management frames are excluded from 802.11w, amongst them all CSI feedback related frames, making them an easy target for interception and for both passive and active attacks. Moreover, strong failures of the initial Wifi WEP keys were highlighted in the early 2000s, and weaknesses are pointed out relevant to new WAP and WAP2 ciphering keys [15]).
Another physical layer vulnerability resulting from network security lapse is the unnecessary exposure of both the AP's and the terminal's capabilities. The capability exchange, which transpires during the association procedure before authentication and establishment of a secure link, includes many of the physical layer's attributes (supported modulations and error correction codes, beam-forming capabilities, etc.) that can be utilized in smart passive or attacks. Here, physical security could be greatly enhanced by simple protocol upgrades, i.e. exchanges of capabilities after authentication procedure, over a secure link.

Channel negotiation in MIMO RATs: advanced close loop MIMO RATs include early propagation channel estimation procedures. In particular the 802.11n/ac based WLAN protocol defines a closed loop sounding procedure wherein the terminal returns Channel State Information (CSI) to the Access Point (AP) for performing single or multi user beam-forming transmissions. The channel state UL feedback message (included in a Management frame and being not encrypted) is easy to intercept, thus it compromise security and facilitate passive and active attacks. The closed loop sounding procedure can also be easily attacked either on the DL (sounding frame) or the UL (feedback CSI frame), by a protocol aware jammer ([19] [20]).

Geo-location services: When un-protected (through most of SMS transmissions for example), geo-location services that use GPS location propagation delay measurements or decoding of signaling, induce serious privacy lacks. Protection of geo location messages becomes now crucial for subscribers, operators and administrations when considering the massive signaling procedures that are studied for future 4G and cognitive networks ([3] [4] [5]), such as:

- Downloading of network data in order to improve RAT through geo-referenced allocations of radio-resource.
- Geo-referenced uploading of sensing report by mobiles.

Multi-RATs handset: Multi-RAT handsets are now very usual. Unfortunately, the vulnerabilities of each RAT may be cumulated, especially when facing active threats: brief jamming procedures of the most protected mode is often enough to force commutation on the worst one.

Personal L–Band satellite communications (L-PCS): Most of L-PCS phones include dual ground-satellite modes and many of the usual satellite RATs are very weak regarding privacy (ex: public services of Iridium, Thuraya, etc.): terminal have high output power and low antenna directivity, waveform are easy to demodulate, un-ciphered transmissions of subscriber's location and ID are usual at early stages of access attempts (this facilitates roaming and billing), etc.

Sub-optimal radio-engineering practices regarding privacy. Examples are fixed frequency planning in GSM networks, low-random code allocation in CDMA, poor time recurrence for changing Temporary IDs and ciphering keys, single authentication sense (instead of dual sense), low power threshold values for cell re-allocation criteria, un-ciphered transmissions of IDs at borders zones, etc.

Users' misunderstood of security aspects (parameterization of secret key, regular change of personal passwords, etc.), and policies restrictions that may occur too, such as ciphering forbidden, Temporary identifier forbidden, etc.

3.3. About passive eavesdropper

3.3.1. Principle of passive attack in public networks:

Passive eavesdropper usually follows the usual RAT:

- Decoding of the broadcast signaling at first,
- Search and decoding of access and paging messages,
- Following of the complete access protocol such as the terminal and nodes do, demodulation of negotiation messages (including subscribers and/or terminals IDs, GPS locations, radio measurements, etc.).
- Recovering and demodulation of traffic channels
- Attempts to decipher negotiation and traffic messages

Several variants are described in [9]. In some cases (i.e. when the key is not found in real time), passive eavesdroppers conduct off line from massive signal records. Passive attacks take advantage of geometric propagation (close range), of high output powers, of easily detected and demodulated signals (FDMA, TDMA, OFDM). They may be disturbed by fast power control, by weak and complex signals (CDMA), by dense spectrum occupancy, by interferences and signal mixtures (MIMO, full duplex [6]). Relevant to access protocol, passive eavesdroppers directly take advantage of any privacy default: especially a priori knowledge or un-protected information relevant to subscribers or terminals IDs or relevant to network engineering allow strong reduction processing complexity.

3.3.2. Passive processing techniques:

Data-aided Processing (DAP) is very usual and efficient for signal processing when facing digital civilian radio-communication standards mentioned fig 1. Usually more sensitive and more accurate than all other techniques, DAP can process medium to strong interference when merged into smart antennas (see [35]). When based on matched filter (inter-correlation of synchronization words, of midambles, of pilot codes, etc.) DAP can achieve early recognition of signals, efficient synchronization and equalization (fig 5) and decoding of messages (fig 6 and 7).

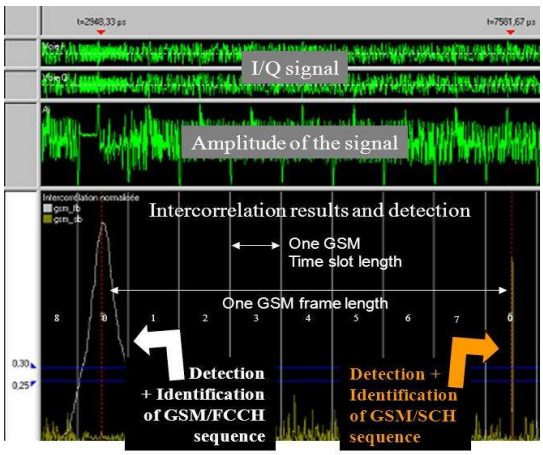


Figure 5: processing into passive eavesdroppers. Example of GSM slot+ frame synchronization (data-aided SISO)

Usually, data-aided techniques apply to any digital standard except CDMA UL traffic signals. In this later case and without extra knowledge, combinatory is prohibitive, and passive attacks take better advantages of symbol modulation characteristics and of pilot symbols into the frame for achieving synchronization and de-spreading ([10]).

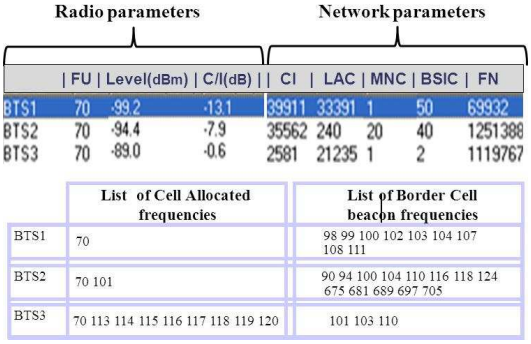


Figure 6: GSM beacon decoding (data-aided SIMO)

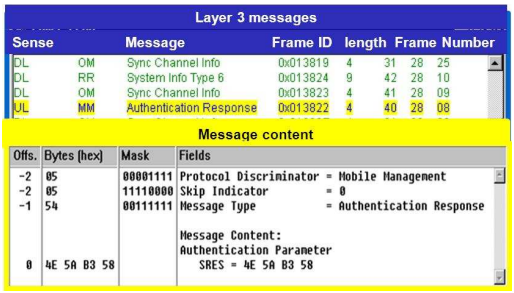


Figure 7: GSM SDCCH decoding (data-aided SISO)

3.3.3. Practical limitations of passive attacks in real field.

Communities of crypto-analysis and of physical layer security usually consider "maximal" attack risk and ideal attack situation: complete a-priori knowledge of the legitimate link, negligible demodulation errors and infinite message lengths. Nevertheless, when facing realistic radio networks and real field propagation, passive attacks are fully dependent of radio conditions (over the complete access protocol and during the data transmission) as they cannot influence communication protocol at nodes neither at terminals. In addition, they are limited by signal structures.

. The power control being relevant to the legitimate link only, it often induced high non-stationaries and low signal to noise ratios (SNRs) at Eve's part (often less that Bob's SNRs). Similarly, interference situations are often stronger for Eve than for Bob. Here, different radio-environments and geometries such as close range indoor, dense outdoor, pedestrian, vehicular, etc. may induce significant differences regarding operational efficiency of passive attacks, whatever are Eve's radio-performances.

. Time, space and Doppler coherence of the propagation channel is finite (and often limited in complex environment); length and redundancy of reference signals and of messages are usually limited by slots duration and by frame structure. All these constraints decrease Eve's integration capabilities

. Eves processing may be (unintentionally) hardened by the standardized procedures themselves (for example fast power control in DS/CDMA systems, soft handover procedures, MIMO/MISO and full duplex RATs [6] increase and the apparent randomness).

In practice numerous passive eavesdropper are thus highly disturbed when unexpected randomness occur into legitimate links and when radio-environment is complex.

3.4. About active radio-hacking systems (RHS)

3.4.1. Principle of active attacks

Radio hacking systems usually exploit weaknesses relevant to authentication or to integrity control in order to fool the victim terminal node or infrastructure, and to influence radio-access procedure in the weakest privacy modes.

3.4.2. Active processing techniques:

Active catching: One basic principle to control a victim terminal is to substitute the local communication node with a virtual node in order to force registration or roaming procedure of terminals (fig. 8). This simplest catching mode requires no synchronization with the real network but only achievement of a (cell re-selection) power criteria and a suitable cloning of beacon channels (see fig. 9). Then RHS controls the caught victim terminal and it can force its registration roaming and identification procedures, it can page it on IMSI IMEI or TMSI, it can intercept calls initiated by the victim terminal, call the victim terminal and force max power, etc. These procedures can be achieved with a protocol tester and a test mobile that access to the network and relay the message of the victim mobiles ([12]).

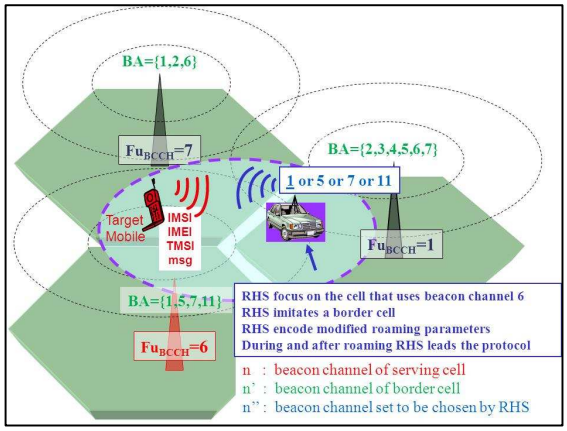


Figure 8: example of active catching and control of radio-cells.

More advanced RHS concepts ([13]) achieve prior synchronization with the real network, follow complete paging and access procedures and perform full duplex synchronized relay of the messages exchanged between the caught terminal and the real network.

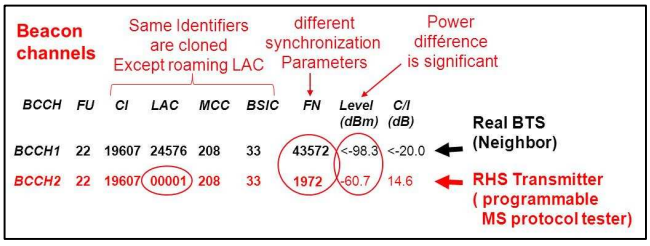


Figure 9: real field example of cloning a GSM beacon channel.

Active catchers practical efficiency fully depend on the network radio-engineering characteristics at the earliest access phases, and thus of relevant propagation conditions. Then, having the full control of the caught terminal, they are less disturbed by radio conditions.

Semi-active spoofing: Semi active variants consist to exploit large acknowledgement latencies and poor integrity control in order to repeat and change the content of dedicated UL or DL messages inside the real negotiation phase, before the acknowledgement of several protocol steps. For example,

- Suitable repetition of "subscriber identity" messages by including modified indications can force transmission of IMSI (or even IMEI) instead of TMSI,
- Suitable repetition and modification of "terminal cipher capability" messages can force clear text transmission.

Semi active variants thus require at least frame synchronization with the exchanged messages and repetition + propagation delay times that lower than protocol tempos.

Selective jamming attacks: Both Active catching and semi active spoofing may be sustained by selective jamming in order to influence the cell selection processing and to forbid more protected access protocols that may be used by multi-RATs terminals, etc. Several classical active attacks variant are described in [9]. Some authors have shown that "aware jamming attacks" of propagation negotiation protocol may be significant threats for MIMO RATs [19][20].

3.4.3. Practical limitations of (semi) active RHS in outdoor.

The main drawbacks of active attacks rely in the power constraints and in the real time mode of use. The received power at the victim terminal or node that has to be sufficient
. to forbid at first any unexpected handoff initiative of the terminal. In dense radio-environments, this may require additional selective jamming (and it thus it induces energy dispersion, it requires more accurate network recovery for suitable parameterization of jamming + beacon signals, etc.)
. to better influence the cell re-selection criteria.

Thus, when facing dense urban networks with BS at building roofs, outdoor pedestrian and vehicle embedded RHS usually have poor effective range (typically less than a few hundred meters, often a few ten meters only).

In addition, the intrinsic real time use mode of RHS reduces processing capabilities and practical efficiency when facing adverse radio conditions, dense environments, etc.

Moreover, as active non synchronized attacks are very intrusive within the real network (fig 9), they often disturb many mobiles in the neighborhood. This leads to saturation risk of the RHS which has to manage multiple roaming/detach procedures of non-targeted terminals.

Semi active variants are more discrete and they usually require less transmitted energy at the RHS part. Nevertheless, semi active RHS highly depend on the network engineering over the full protocol duration (and not only of early negotiation phases): frequency planning, multiplexing schemes, time reference of victim node/terminal, latencies of the negotiation protocol. They thus require an accurate recovery of frame/slot/symbol synchronization (more difficult when facing wideband CDMA), they need fast real time reactions and they have to take into account propagation delays. All these constraints dramatically limit their practical range of use: typically a few hundred meters, often to a few ten meters only.

3.5. Special threats relevant to SDR and CR.

Within CR networks (see [3-5]), access attempts should be sustained with numerous procedures such as the following:

. Geo-referenced database downloading will inform terminals about the available radio-networks and the relevant radio-access parameters.

. Terminal will perform sensing and report to nodes and cognitive managers about the local radio spectrum.

. Terminals should perform geo-referenced access attempts that would involve systematic transmission of subscribers' locations in the early stages of the negotiation protocols

. Dedicated "beacon" signals such as DL/UL-CPC (Down Link and Up Link Cognitive Pilot Channel) should be broadcasted in order to support both downloading and sensing + channel sounding procedures within terminal and nodes. Network downloading and terminal embedded sensing should be based on a DL-CPC. Sensing information reporting and BS/node sensing should be based on the use of a UL-CPC signal. For simplification, both DL-CPC and UL-CPC should be designed for fast recognition, accurate measurements and easy decoding. Thus, they should be very weak regarding both passive and active threats.

4. EXISTING COUNTERMEASURES PRINCIPLES

4.1. Transec countermeasures.

Transmission Security (transec) is relevant to the protection of the wave form face to interception/direction finding of the transmitted radio signal, face to jamming of the user receiver, and face to intrusion attempts into the radio access protocol. Some transec technics are described hereafter:

Randomization of transmitted signals: this can be achieved
. with frequency hopping and time hopping of burst signals (RATs involving TDMA and TDD mechanism),
. with pseudo noise spreading (such as CDMA RATs),
. with random jitter and scrambling of midamble, of synchronization words and or pilots symbols (any RAT),
. by using long term pseudo-random schemes for modulation, coding, scrambling and interleaving of signaling and access messages, for allocation of traffic radio resource.

Use of furtive signals for supporting access protocol and traffic: for improving privacy of access of negotiation phase and of traffic allocation, transec often uses short duration messages that are randomly jittered within long protocol frames, progressive protected and acknowledged DL and UL exchanges for early device's identification, such as applied in the domain of Identification Friend and Foe (IFF).
Transec applies mainly at the radio interface and usually re-enforce advantages of propagation diversity. Moreover, with convenient adaptations, several characteristics of public RATs could provide some transec capabilities:
. TDMA standards could provide some transec if frequency hopped modes (FHM) were used over numerous channels. Unfortunately, FHM in public TDMA RATs are usually dedicated to frequency diversity and not to privacy.

. CDMA standards provide native protection in the UL sense when using a-synchronous long pseudo noise codes. Unfortunately, it dramatically decreases when low combinatory pilots symbols are present (3GPP), when system time reference is GPS (3GPP2) and when code allocation is terminal dependent (3GPP2 public mode).

. Complex data multiplex schemes within OFDM signals and associated RATs would provide transec protections if scattered pilots and signaling were more difficult to synchronize and to decode (DVB-T/H, LTE).

. Any adaptive MISO and MIMO RAT full duplex access schemes induce native space time diversity and randomness that could re-enforce propagation non-stationary effects and generate signal mixtures that provide intrinsic interferences at eavesdropper part.

4.2. Netsec countermeasures.
Network Transmission Security (netsec) is relevant to the protection of the signaling of the network (including the subcriber's part). Netsec applies either at the radio interface and at the medium access protocol layer, with request to upper protocol layers. Netsec techniques involve mainly transmitter authentication protocols, integrity control and ciphering of signaling and of negotiation messages. Severe netsec weaknesses exist in most of public wireless networks:
. Only few public wireless networks are able to protect and to control the integrity of the signaling messages that are broadcasted by BSs or nodes, same lacks apply to first paging messages sent by node and to first access messages sent by terminals.
. Many networks apply no authentication or single sense authentication only, because of standards lacks or because of sub-optimal operators' engineering.
.Ciphering procedures are usually initiated in the later part of the RAT procedure, thus after authentication and identification steps, whose data remain un-ciphered.
. Netsec failures are often re-enforced into multi-RATs terminals, especially when involving L-PCS RATs.

4.3. Comsec countermeasures.
Communication Security (comsec) is relevant to the protection of the content of the user messages (voice, data). Comsec applies at the radio interface and at upper layers. Comsec techniques involve ciphering and integrity control of users messages at several protocol layers and even at several interfaces when transmission relay occur (examples are point to point ciphering of each user data flux before multiplexing, ciphering of IP packets, of artery, etc.).
Native comsec capabilities of public wireless standards are controlled by legal authorities (key lengths are limited).
Native comsec failures exist in many standards and un-expected failures were pointed out exist in several cases. Examples above pointed out lacks of integrity control in many standards, security conception errors (Wifi, Bluetooth), un-expected publication of (initially secret) cipher algorithm (GSM), etc.

4.4. Elements about secure terminal architectures.
Information security modules (infosec) are dedicated to the generation of random data.
In most of wireless public standards, infosec modules exploit shared keys (that are present into SIM cards, into terminal -electronic serial numbers-, and/or into operators' databases), and they generate random parameters that are transmitted over the radio interface. These parameters initiate or acknowledge computations of keys at both node and terminal part. Usually these procedures start during the later stages of the negotiation protocol and the earlier data exchanges remain thus un-protected.
There exist too "source ciphered" handsets [7] [8] that improve comsec. Nevertheless, such handsets induce heavy constraints for operational use (limited set of subscribers) and they usually remain non-operant for transec and netsec.
In more secure terminal and node architectures, infosec is based on an initial shared secret and avoids dedicated exchanges. It provides several independent pseudo-random sequences to be followed by RATs, to be xorred with the data stream, to be added inside messages for integrity control, etc. Therefore, infosec apply very early in the negotiation protocol and it is placed at the core of the terminal architecture so that it is called for any security procedure and it cannot be shortcut.

4.5. Elements about secure military radios - relevant opportunities for privacy of public standards.

Military communications usually use advanced transec netsec and comsec protections that are managed with transverse infosec modules, secret system time reference and a priori secret information that are shared over terminals and nodes. For example data bases are pre-computed and implanted into terminals and nodes that are relevant to frequency plans, to pseudo random sequences, to transec netsec and comsec keys, etc. Nevertheless, massive sharing of a priori secret information remains a major difficulty: even with medium number of terminal and nodes, this induces strong system constraints, mission preparation, etc.

Thus, in order to transpose some of privacy concepts from military communication skills to public worldwide mass market standards, there are strong needs for shared and private random sources. Physec solutions introduced in § 5 are expected to provide suitable alternatives for that.

4.6. Risk analysis and risk-driven security metrics

Systematic methods such as security metrics development and management are needed to be able to develop sufficient security and privacy countermeasures for physic (see §5). A high-quality risk analysis is the starting point of all security work and its results set the reference level for security metrics. The reference requirements used in security and privacy are either based on (i) security risk, or (ii) best practices and regulations. The former category assumes direct availability of risk analysis results, while the latter does not. To some extent, security can be managed with the help of best practices, but the lack of risk knowledge could result in costly and incorrect security countermeasures.

Security metrics can be used to reason about the effectiveness of countermeasures, to support configuration management and to show compliance to security and privacy regulations and legislation. Security metrics should be based on the prioritized collection of security risks, making them risk-driven. A security metrics development approach based on hierarchical decomposition of security objectives was introduced in [27].

5. PERSPECTIVES OFFERED BY PHYSICAL LAYER SECURITY (PHYSEC)

Physec concepts take advantage of the physical characteristics radio-environments, especially when complex, dispersive and non-stationary, and try to take the benefit of radio propagation parameters that have to be measured by infrastructures and handsets for the purpose of their proper communication services [34][36]. Nowadays, the relevant information is used for equalization in FDMA and TDMA RATs, for RAKE processing in CDMA RATs, for adaptive modulation/coding schemes in MIMO RATs, for interference mitigation in full duplex RATs [6]. Sensing procedures and opportunistic spectrum access within cognitive radios are other opportunities for privacy improvements. Finally, any intrinsic physical randomness, especially when measured by legitimate links during access attempts and established calls, should contribute to security the air interface, with low impact at upper layer and no constraints at other network interfaces (Abis, A).

. Dedicated coding schemes (secrecy codes) were proven to provide intrinsic secrecy of legitimate radio link facing passive eavesdropper (fig 4), when better radio quality is achieved for the legitimate [34]. Roughly, secrecy codes mitigate the information about the legitimate link at any radio-eavesdropper location, up to a given "secrecy capacity" ($C_{sec,AB}$). In general, secrecy capacity is (of course) less than the legitimate Shannon capacity and greater than the Shannon capacity difference of the legitimate and of the eavesdropper link: $0 \leq C_{sh,AB} - C_{sh,AE} \leq C_{sec,AB} \leq C_{sh,AB}$.

. Secrecy codes could be merged with advanced RATs that generate signal mixtures in order to disturb Eve: MISO and MIMO, artificial jamming [34], full duplex techniques [6].

. Diverse non-stationary artificial random source facilitate transec (versatile allocation of traffic resource, adaptive changes of modulation and coding etc.), thus making interception, eavesdropping and spoofing more complex.

. Propagation-dependent random sources are added value for generation of secret keys, of control pattern, etc. especially when combined with existing comsec schemes.

. In order to enhance security of access phases, early identification procedures should be designed by using weak/furtive low data rates signal that would be mixed with strong signals that are already broadcasted [11], before terminal's dedicated (and protected) signaling is made intelligible for further access attempts.

6. CONCLUSION

In this paper, we illustrated several security lacks of civilian wireless networks and relevant passive and active attacks at the radio interface that may dramatically deteriorate both subscribers' privacy and security and operators' confidence, especially when focusing on signaling, on first radio access attempts and on negotiation phases.

By considering realistic radio environments, existing counter-measures and secure terminal architectures, we pointed out that large perspectives exist for significant privacy upgrades by merging traditional privacy techniques, secrecy coding and other physec concepts.

The core idea is to combine high combinatory channel codes, advanced modulation schemes and traditional countermeasures in order:

- To trend toward secrecy capacity
- To take the maximal benefit of adverse radio environments that are often encountered by eavesdropper and by RHS.

These concepts largely exploit new randomness sources and propagation advantages that are measured and/or generated locally by legitimate communications nodes and terminals for their proper communication services. Complements about principles, theoretic advantages and practical expectations for wireless networks privacy can be found in [35] and deeper explanations are given in [34] .

We conjecture that introducing physec-privacy concepts into wireless public standards should be particularly efficient:

. For access attempts and negotiation phases in general

. For downloading/uploading procedures within CRs

. During established call: upgrade of current cipher schemes.

We are confident that current national and European research programs will establish convincing feasibility proofs in the future years, thus preparing standardization and industrial development of trustworthy and full-secure RATs.

7. REFERENCES

[1] www.3GPP.org

[2] www.3GPP2.org

[3] http://standards.ieee.org/findstds/standard/1900.6-2011.html

[4] https://standards.ieee.org/findstds/standard/1900.4a-2011.html

[5] J. Mittola: Conference "Secure Geospatial Dynamic Spectrum Access". GDR ISiS Telecom Paris tech 9 Mai 2011.

[6] "Full Duplex Radios for Local Access" EU EU FP7-ICT 316369. URL: http://www.fp7-duplo.eu

[7] URL: http://www.cryptophone.de/en/products/mobile/cp400/

[8] URL: http://www.defense.gouv.fr/dga/actualite/la-dga-livre-les-premiers-telephones-teorem

[9] Y. S Shiu and al. Physical Layer Security: a tutorial in IEEE Wireless Communications • April 2011

[10] Procédé « protocole orienté » de traitement des signaux stationnaires, partiellement stationnaires, ou cyclo-stationnaires, patent Thales FR 10.05017, PCT/EP2011/073420 WO2012.084956

[11] Procédé de taggage radio-électrique des signaux de brouilleurs et d'autres émetteurs. Patent Thales FR 12.03071

[12] Patent Rohde and Schwarz EP 1051 053 B1

[13] Method of controlling and analysing communications in a telephone network. Patents Thales FR 04.04043, PCT WO 2005/112497 A1

[14] R. Gautier, G. Burel, J. Letessier and O. Berder « Blind estimation of Scrambler offset using encoder redundancy », in IEEE 2002

[15] URL: http://fr.wikipedia.org/wiki/Wi-Fi_Protected_Access

[16] http://web.archive.org/web/20090318143444/http://www.scard.org/gsm/a3a8.txt http://cryptome.org/gsm-a512.htm (Originally on www.scard.org),.

[17] E. Biham, O. Dunkelman, Cryptanalysis of the A5/1 GSM Stream Cipher, Progress in Cryptology, proceedings of Indocrypt'00, Lecture Notes in Computer Science 1977, Springer-Verlag, pp. 43-51, 2000

[18] Shahriar, C., Sodagari, S. & Clancy, T.C. Physical-layer security challenges of DSA-enabled TD-LTE. Proceedings of the 4th Int. Conf. on CRadio and Advanced SM. Barcelona, Spain, New York, NY, USA: ACM, 2011. CogART '11. pp. 40:1.

[19] R. Miller, W. Trappe ; "On the Vulnerabilities of CSI in MIMO Wireless Communication Systems » IEEE Tran. on Mobile Computing, Vol. 11, N°. 8, August 2012

[20] Sodagari, S. & Clancy, T.C. Efficient jamming attacks on MIMO channels. IEEE (ICC) 2012. p. 852.

[21] Ruiliang Chen, Jung-Min Park & Reed, J.H. Defense against Primary User Emulation Attacks in Cognitive Radio Networks. Selected Areas in Com., IEEE Journal 2008, Vol. 26-1, pp. 25-37.

[22] Kune, D.F., Koelndorfer, J., Hopper, N. & Kim, Y. Location leaks on the GSM air interface. Network & Distributed System Security Symposium (NDSS). 2012.

[23] Ta, T. & Baras, J.S. Enhancing Privacy in LTE Paging System Using Physical Layer Identification. Teoksessa: Pietro, R., Herranz, J., Damiani, E. & State, R. (eds.).September 13-14, 2012. Vol. 7731. Springer Berlin Heidelberg, 2013. Lecture Notes in Computer Science. p. 15.

[24] Vintila, C., Patriciu, V. & Bica, I. Security Analysis of LTE Access Network. 10th Int. Conf. on Networks (ICN 2011). 2011. p. 29.

[25] Hyeran Mun, Kyusuk Han & Kwangjo Kim 3G-WLAN interworking: security analysis and new authentication and key agreement based on EAP-AKA. WTS 2009. p. 1.

[26] Bikos, A.N. & Sklavos, N. LTE/SAE Security Issues on 4G Wireless Networks. Security & Privacy, IEEE 2013, Vol. 11, No. 2, pp. 55-62.

[27] Savola, R. and Abie, H., Development of measurable security for a distributed messaging system, *Int. Journal on Advances in Security*, Vol. 2, No. 4, 2009, pp. 358–380.

[28] NSA National Security Agency, Bluetooth Security, URL: http://www.nsa.gov/ia/_files/factsheets/I732-016R-07.pdf

[29] Padgette J., Scarfone K., Guide to Bluetooth Sec., NIST, 2011.

[30] Bluetooth Vulnerability Assessment Technical Publication ITSPSR-17A, Communications Security Establishment Canada, 06/2008.

[31] Bowers B., ZigBee Wireless Sec. 2012: A New Age Penetration Tester's Toolkit. URL: http://www.ciscopress.com/articles/article.asp?p=1823368,.

[32] Vidgre and al. Security Threats in ZigBee-Enabled Systems: Vulnerability Evaluation, Practical Experiments, Countermeasures, and Lesson Learned.: 46th IEEE Hawaii Int. Conf. on System Sciences, 2012, pp. 5130-5136.

[33] KillerBee, URL: http://code.google.com/p/killerbee/

[34] M. Bloch and J. Barros, Physical-Layer Security: From Information Theory to Security Engineering. Cambridge University Press, 2011

[35] Procédé d'optimisation de la planification dans un système de communications de type CDMA Patent Thales FR 04.01475.

[36] J.C. Belfiore, A. Sibille, C. Ling, F. Delaveau, E Garrido ; "Physec Concepts for Wireless Public Networks, introduction, state of the art, perspectives" Winncomm 2013, Munich.

An Efficient Incremental Redundancy Implementation for 2.75G Evolved EDGE

Benjamin Weber, Harald Kröll, Christian Benkeser, Qiuting Huang

Integrated Systems Laboratory
ETH Zurich, 8092 Zurich, Switzerland
e-mail: {weberbe,kroell,benkeser,huang}@iis.ee.ethz.ch

***Abstract*—Incremental Redundancy (IR) in GSM was introduced with EDGE and later adopted in Evolved EDGE in order to increase the throughput. The distribution of IR controlling and processing over various parts of the PHY and RLC/MAC layer demands high loads to processing units and imposes difficulties with respect to system design. In particular, large amounts of data have to be moved between memories and processing units. In this work IR is implemented as part of baseband signal processing in order to unburden RLC/MAC layer processing units from IR. The open source baseband framework MatPHY is extended for packet switched operation to facilitate the development of efficient IR hardware architectures. With the design parameters obtained from performance evaluations, computed with MatPHY, we implemented an architecture in 130 nm CMOS technology to prove the suitability of the approach.**

I. Introduction

After its first release in 1990, GSM was fast becoming one of the most successful and widely used cellular standards worldwide. Although initially supporting circuit switched connections only, there was a need for packet switched operation. Correspondingly, GPRS and later EDGE (also referred to as EGPRS) services were introduced by the 3GPP standardization organization. Together with EDGE, additional modulation and coding schemes were introduced to maximize throughput. Irrespective of 3rd generation and LTE deployments, the GSM standard and its enhancements will be in use for many years to come. Bearing this in mind, the 3GPP introduced an enhancement to EDGE called Evolved EDGE (also referred to as EGPRS2) with up to 32QAM modulation and turbo codes. In order to keep the throughput at an acceptable level even under severe radio conditions, Incremental Redundancy (IR) operations were first introduced with EDGE and inherited by Evolved EDGE. IR is also referred to as Type II Hybrid ARQ. More detailed, if decoding of a data block proves unsuccessful, the encoded soft values (log-likelihood ratios) of the data block in question are kept at the receiver and a retransmission is requested. The retransmitted data block contains additional redundancy information. The receiver combines the retransmission with the previous transmission before a decoding attempt is started. Correspondingly, the coding rate decreases with each retransmission whereas the decoding success rate increases. Thus, IR aids to achieve high average throughput. Minimum average throughput requirements from the 3GPP specifications ask a Mobile Station (MS) to store up to 1024 Radio Link Control (RLC) blocks [1], [2]. Therefore, memory and processor challenges are two major concerns when implementing IR at MS side.

A hardware architecture of a MS consists of several processing components attached to a RF transceiver. The system processor for higher layer operations such as RLC/MAC is connected to memories and other peripherals. For digital baseband processing in the PHY a DSP with a number of accelerators is employed. Alternatively, there exist approaches (e.g. [3]) which make do without a DSP. Instead, the digital baseband processing is implemented in dedicated hardware only. According to literature (e.g. [4]), IR implementation is traditionally distributed over various processing components in a MS. Erroneous RLC blocks are stored in an external RAM. IR control processes run on the system processor. The decoding of the RLC blocks takes place in the digital baseband on a DSP's accelerator.

In such a conventional setup the IR mechanism claims high demands for the system processor, which is not desired. And, RLC/MAC software developers need to incorporate IR operations. Furthermore, the data bus between external memory, system processor and digital baseband can experience high loads due to IR processing. For emerging applications that require ultra low-cost and low-power devices, such as Internet of Things (IoT) or machine to machine (M2M) communication, which may use the GSM/EDGE standard due to its ubiquitous coverage, high computational loads are prohibitive. An IR processing block incorporated into the digital baseband comprising all IR related operations can remove IR complexity and load from the system processor and external components.

Contribution: In this paper IR implementation challenges are investigated and an efficient dedicated hardware architecture, which hides the IR complexity from the system processor and higher protocol layers, is proposed. In order to gain insight into the IR mechanisms, an Evolved EDGE capable version of the open source MatPHY GSM framework [5] was equipped with IR functionality. The minimum IR memory size in order to meet the 3GPP throughput requirements is determined based on the receiver performance of the extended framework

and a model RF front-end. The implemented IR hardware architecture using a dedicated on-chip memory for the storage of RLC blocks comprises depuncturing, memory management, and a soft information combiner unit. A corresponding VHDL model is implemented and synthesized.

Outline: The rest of this paper is organized as follows: Section II explicates IR operations and GSM/EDGE architectures in more detail. Subsequently, in Section III the MatPHY extensions are outlined followed by IR performance evaluations in Section IV. Afterwards, a dedicated hardware architecture is proposed followed by concluding remarks.

II. IR in GSM/EDGE MS Architectures

Packet switched connections use so called radio blocks consisting of a RLC/MAC header and a number of RLC blocks. IR operations between MS and network are performed on a RLC block basis. IR controlling (IR memory management) forms part of RLC/MAC whereas IR processing (combination of redundancy versions and subsequent decoding) is a pure PHY operation. A MS PHY receives a radio block and makes an attempt to decode the RLC/MAC header and RLC blocks. Depending on the radio conditions decoding may fail. In RLC acknowledged mode, the receiving RLC/MAC layer stores the encoded RLC block's soft values and automatically requests a retransmission by means of an ACK/NACK message sent to the network. The network side RLC/MAC layer makes sure that the network PHY uses a different puncturing pattern for the retransmission. The MS RLC/MAC layer knows when the retransmitted RLC block is received by analyzing the RLC/MAC header of each received radio block. It assures that the stored RLC block gets combined with the retransmitted version before decoding. In this manner, more redundancy bits are available at the MS after each retransmission. This procedure can be repeated various times until decoding succeeds [2].

A GSM/EDGE packet switched connection is referred to as Temporary Block Flow (TBF). Each TBF corresponds to a unique Temporary Flow Identity (TFI). RLC blocks within a TBF are numbered using the so called Block Sequence Number (BSN). Both TFI and BSN form part of the RLC/MAC header. A flow control mechanism is used in terms of a transmit window on a RLC block granularity.

A. IR in Classical GSM/EDGE Architectures

A GSM/EDGE MS architecture is typically based on two processing units and a RF transceiver. In Fig. 1 the architecture of [4] is given, comprising a DSP for baseband processing and a system processor for protocol layer handling.

The system processor has external components attached to it such as RAM and peripherals. The DSP uses accelerator blocks for computational intensive operations such as channel equalization or decoding. Alternatively, it is possible to employ dedicated hardware only for the PHY such as in [3].

IR operations are logically spread over the RLC/MAC and PHY as RLC/MAC is responsible for controlling and the PHY for rate adaption and decoding. Physically, they are spread over a DSP, an instruction set processor (system processor),

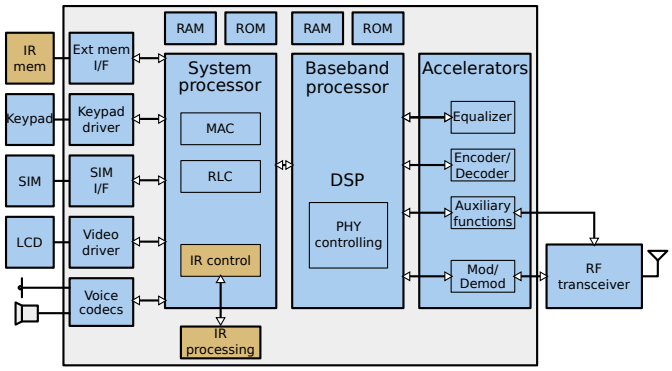


Fig. 1. Block diagram with signal processing components of a MS. IR processing as in [4] is colored in yellow.

and external RAM for storing soft value information. A large amount of data moves between PHY (DSP), RLC/MAC (instruction set processor), and IR memory (external RAM) is the result.

B. Accelerating IR Operations by Dedicated HW Architectures

In [4] an attempt was made to minimize aforementioned data moves by relieving the RLC/MAC layer from IR controlling operations. All IR controlling and processing, including decoding of RLC/MAC header and RLC blocks, were moved to a separate IR controlling and a IR processing module, respectively. They run partly on the CPU as well as in dedicated hardware as depicted in Fig. 1 in yellow. The RLC/MAC layer still needs to know whether decoding of a RLC block failed in order to produce corresponding ACK/NACK messages for the network. However, awareness of IR operations on a RLC/MAC layer level is no longer required as long as the IR controlling module is informed whether the TBF is operating in acknowledged mode. Still, a lot of data moves between IR memory (external RAM), IR controlling and processing modules (CPU and dedicated hardware), and the PHY (where equalized soft values for IR come from) remain.

Baseband processing in dedicated hardware allows further reduction of IR processing overhead, as shown in this work. We focus on PHYs as presented in [3] which allow minimization of data moves between various parts of a GSM/EDGE receiver with respect to IR. Moreover the IR complexity is completely invisible for the system processor.

However, moving IR controlling and processing into the baseband processing inherently requires some RLC/MAC layer related information to be present in the PHY. A list of TFIs corresponding to TBFs in acknowledged mode and their corresponding transmit window sizes are required. Everything else, such as BSN and puncturing version can be extracted from the decoded header bits. The decoded header is present in the PHY anyway provided that the header is actually decoded in the PHY instead of in a IR processing module attached to the system processor. Naturally, the transmit window size could simply be replaced with a hard coded value corresponding to the maximum transmit window size for

the multislot class (maximal timeslot configuration) supported (see [6]). Technically, status information for each BSN within the current transmit window per TBF is required for IR controlling to avoid decoding of already successfully received RLC blocks. In fact, it should be sufficient that IR controlling simply tracks the transmit window and clears IR memory entries corresponding to RLC blocks which drop out of the transmit window.

All things considered, it is possible and feasible to implement IR controlling and processing in the PHY by duplicating only a minimal subset of RLC/MAC layer related information. If, in addition, dedicated on-chip memory for storing soft value information is incorporated data moves with respect to IR can be reduced to

- configuration data from the RLC/MAC layer towards the PHY containing TFI information and
- notifications of decoding results from the PHY towards the RLC/MAC layer.

All other data moves, such as feeding soft values into the IR processing unit or storing soft value information in the IR memory, are inherently inevitable but can be implemented as local moves (within the PHY) only.

Variant	As in [4]	Dedicated HW (PHY)
Across layers	s_p	0
Off-chip memory read	$s_\mathrm{p} n_\mathrm{v}(s_\mathrm{d})$	0
Off-chip memory write	$s_\mathrm{p}(s_\mathrm{d})$	0
On-chip memory read	0	$s_\mathrm{p} n_\mathrm{v}(s_\mathrm{d})$
On-chip memory write	0	$s_\mathrm{p}(s_\mathrm{d})$

TABLE I
DATA MOVES FOR IR OPERATIONS.

Data moves of soft values across layers as well as corresponding IR memory accesses are summarized in Table I. The combination of a retransmitted RLC block with s_p punctured soft values with n_v redundancy versions stored in IR memory is evaluated. The depunctured RLC block consists of s_d soft values. It is assumed that only punctured redundancy versions are stored. Alternatively, whenever $s_\mathrm{p} n_\mathrm{v}$ is larger than s_d, it is wise to store depunctured RLC blocks. In that case there are s_d read and write memory accesses, irrespective of the implementation variant. Moves of control data, be it across layers or inside a layer, being negligibly small, are not considered. As Table I shows, using a dedicated hardware inside the PHY avoids cross-layer data moves. Memory accesses, however, cannot be avoided. Nevertheless, accessing an on-chip memory is more efficient as no external components are required. Keeping this savings and simplifications in mind a dedicated IR implementation inside the PHY is developed starting from an IR extension for the MatPHY framework.

III. MATPHY EXTENSIONS TOWARDS EVOLVED EDGE

The open source project MatPHY presented in [5] uses OsmocomBB (an open source GSM protocol stack [7]) and a custom PHY called *phydev* written in Matlab (see Fig. 2). The OsmocomBB distinct L1CTL protocol is used for data exchange between PHY and higher layers. MatPHY models digital baseband operations and has been structured into PHY controller, auxiliaries and three large signal processing blocks which there are Digital Front-End (DFE), Detector (DET), and Decoder (DEC). The signal processing blocks consist of controllers and various signal processing primitives.

Following the EDGE evolution MatPHY has been expanded into three directions:

- Support for packet switched data (GPRS, EDGE, Evolved EDGE Level A).
- Enhancement for the OsmocomBB L1CTL protocol.
- Configurable data source for on-the-fly IQ samples generation.

Each expansion is briefly described in the following.

A. Packet Switched Data Support

Packet switched data support in MatPHY is a two-fold endeavor. First, as OsmocomBB software does not support packet switched communication, a minimalistic RLC/MAC test-software is required. The latter can be used to test MatPHY packet switched operations and supports as much as receiving packet switched traffic channels. The minimalistic RLC/MAC software can communicate with the configurable data source in order to force retransmissions of RLC blocks.

Second, the phydev part of MatPHY has been enhanced for packet switched data. This includes additional controller functionality inside the various controllers including handling of packed switched L1CTL enhancements. In addition, the existing primitives for DFE, DET, and DEC were enhanced accordingly. Most notable is the required support for additional modulation schemes in DET and the incorporation of a turbo decoder for Evolved EDGE operation in DEC similar to [3]. Naturally, so far unused primitives in DEC such as depuncturing and straight forward IR with infinite memory needed to be added. Packet switched operation requires blind detection of the modulation order [2]. However, this task was omitted and a priori knowledge of the modulation order in the receiver was assumed.

B. L1CTL Enhancement

Even though the OsmocomBB project does not (yet) support packet switched data the idea of configuring the PHY by means of simple L1CTL messages seems worthwhile. Bearing this in mind, we propose an extension of the L1CTL messages for packet switched operation. In fact, only two additional L1CTL request messages are required:

`TBF_REQ`: This message holds a list of TFIs corresponding to TBFs for which the MS is currently configured. This is necessary as resource sharing is possible on the downlink [8]. The MS may distinguish payloads intended for decoding by analyzing the TFI field in the decoded header. In addition, this message indicates whether a TBF corresponding to a certain TFI is in RLC acknowledged mode or, in other words, whether IR operations are required. As described in Section II-B, the transmit window size is not indispensable for IR operations. However, if required, the configured transmit window size for

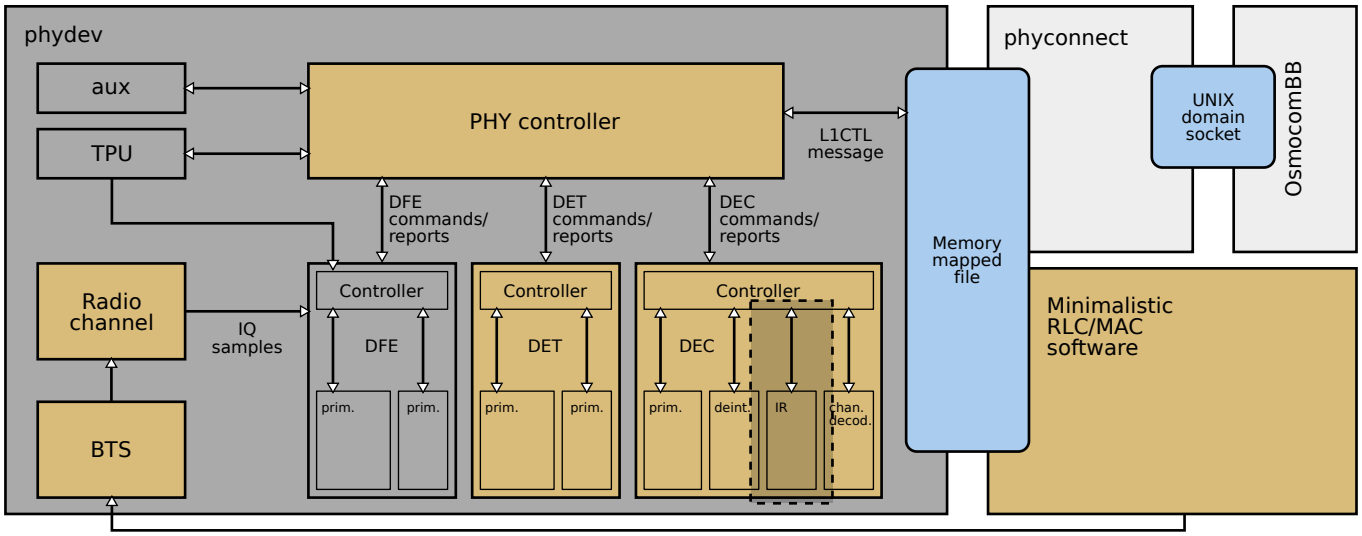


Fig. 2. MatPHY Evolved EDGE extensions: Modified or added parts as opposed to the original MatPHY are depicted in yellow. A zoom of the shaded selection in DEC can be seen in Fig. 5.

each active TBF in RLC acknowledge mode may form part of the message.

`DATA_CONF_REQ`: This message holds an information structure for each timeslot containing the Training Sequence Code (TSC) and the GSM mode (GPRS, EDGE or Evolved EDGE). In addition, it indicates whether in a certain timeslot transmission or reception is required. Furthermore, a starting frame number (FN) indicating the first valid frame of the new configuration forms part of this message. Naturally, this message can be enhanced with additional information as the need arises.

In the other direction (from PHY towards RLC/MAC) two additional L1CTL messages were defined:

`TBF_CONF`: This is a simple confirmation of the corresponding request.

`P_TRAFFIC_IND`: This message is used to indicate and report data on packet switched channels. It can be used to report RLC/MAC headers, RLC blocks and other information individually, as soon as it is decoded or, alternatively, complete radio blocks.

C. Configurable Data Source

The data source consists of a Base Transceiver Station (BTS) and a radio channel for on-the-fly IQ samples generation. The BTS can produce burst wise IQ samples for decoding. The radio channel block, on the other hand, is able to modify the transmit burst according to 3GPP channel profiles. The data source is closely connected to the minimalistic MS RLC/MAC software for e.g. the management of RLC block retransmission.

IV. Performance Evaluation

IR performance was assessed using the MatPHY extension from the previous section. As mentioned previously, soft value information of RLC blocks which could not be decoded successfully needs to be stored in the receiver. The required memory capacity is not given in the 3GPP specifications. However, a long-term throughput between the RLC/MAC and the layer on top per timeslot needs to be achieved [1].

	EDGE	Evolved EDGE
Required throughput	20 kbit/s/timeslot	33 kbit/ps/timeslot
Propagation conditions	Static, input level -97 dBm	Static, input level -94 dBm
Modulation and Coding Scheme	MCS9	DAS12
Acknowledgments polling period	32 RLC data blocks	32 RLC data blocks
Roundtrip time	120 ms	120 ms
Number of timeslots	Maximum capability of the MS	Maximum capability of the MS
Transmit window size	Maximum for the MS capability	Maximum for the MS capability

TABLE II
IR test case description [1].

Table II summarizes the test conditions under which the overall IR performance needs to be evaluated. Values such as acknowledgment polling period and transmit window size and their impact on IR test cases are described in [2]. For these simulations, the overall receiver needs to be taken into account. This includes the Noise Figure (NF) of the RF part, as well. Soft values with finite width w bits have been used in the DET and DEC blocks.

IR performance simulations have been split into two parts, one that uses parallelization and a sequential one. The extended MatPHY framework from the previous section was used in a single timeslot mode where retransmissions of RLC blocks occur immediately after decoding failed at the receiver. Correspondingly, the amount of necessary retransmissions for the IR test cases from Table II under the assumption of infinite IR memory and 0 ms roundtrip time can be recorded.

Unfortunately, such a simulation is very time consuming. Therefore, many simulations of this kind can be run in parallel in order to get results much faster and, in addition, get more accurate retransmission count information due to averaging over a larger amount of RLC blocks.

As the IR test scenario uses a static channel model, it is feasible to use the retransmission counts from the simulation to compute various decoding success probabilities of RLC blocks. There exist not more than three redundancy versions per RLC block. Therefore, the decoding success probabilities p_i for $i = 0, 1, 2, 3$ denoting the amount of redundancy versions already stored in the IR memory were computed.

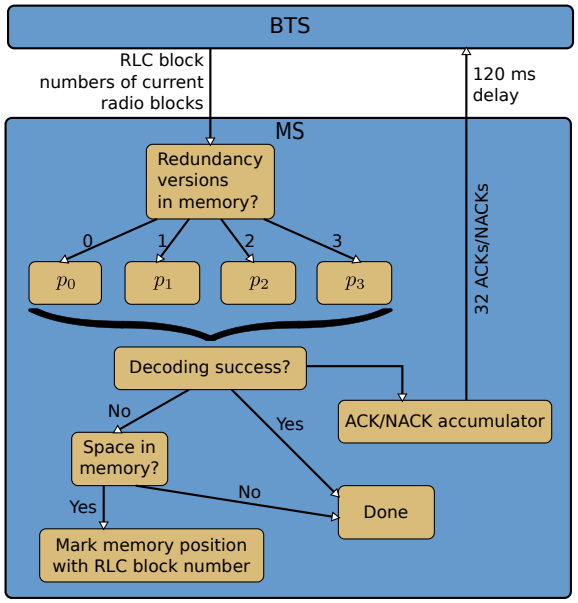


Fig. 3. IR flow setup used in the simulation.

A separate RLC block level simulation flow implementing a BTS and a MS was set up, see Fig. 3. It comprises the various requirements from Table II. In addition, the MS can be configured to various IR memory sizes in terms of punctured RLC blocks. The BTS decides using RLC/MAC procedures which RLC blocks need to be transmitted. The MS evaluates whether the decoding was successful using the decoding success probabilities described above, depending on the amount of redundancy versions already stored in IR memory. For various IR memory sizes the average throughput per timeslot was computed. Six downlink timeslots were used according to multislot class 45 as specified in [6]. As no IQ data is equalized or decoded in the RLC block level simulation, computation time is less critical and parallelization is not required.

Fig. 4 shows simulation results for MCS9 and DAS12, respectively. As required by the 3GPP IR performance requirements the throughput in kbit/s/timeslot is plotted on the vertical axis against the IR memory capacity in terms of punctured RLC blocks. Various receiver performances have been simulated. In fact, the RF performance model in terms of NF was altered whereas the digital baseband performs the same. The target throughput for MCS9 with a NF of 6 dB can be met with almost no IR memory. However, if a NF of 8 dB is used roughly 40 punctured RLC blocks need to stored. A similar increase of IR memory capacity can be observed using the same two NFs and the Evolved EDGE DAS12 test case. This clearly shows that in order to evaluate IR performance and to find a suitable IR memory size the entire PHY processing chain needs to be taken into account, including the RF transceiver. The overall receiver performance has a huge impact on IR performance.

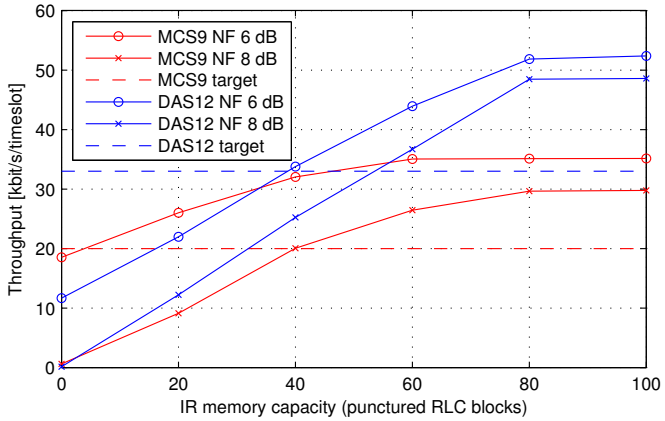


Fig. 4. Simulation results in terms of throughput (kbit/s/timeslot) against punctured RLC block memory capacity.

V. Dedicated Hardware Architecture

In order to check the feasibility of having IR operations in PHY alone a dedicated hardware architecture was developed.

A. Architecture

An enlarged version of the shaded section of Fig. 2 is depicted in Fig. 5. It consists of an IR control unit as part of the DEC controller and an IR processing unit, the latter a signal processing primitive between deinterleaver and channel decoder.

The decoded header bits, which contain among other things the BSNs of the payload, are present in the DEC controller. Consequently, it makes sense to have an IR control unit inside the DEC controller accessing the decoded header bits. All IR operations become thus completely transparent to RLC/MAC processes. The IR control unit has a small memory containing all necessary information regarding the stored RLC blocks in the IR memory such as puncturing scheme, BSN, TFI, position in IR memory, and the current transmit window position. Whenever a RLC block is ready to be deinterleaved, the IR control unit checks whether there already exists a version with the same BSN and transfers the required information to the IR processing unit. In addition, if a newly received RLC block moves the transmit window such that older RLC blocks drop out, the corresponding IR memory can be recycled. Once the decoding terminates, the IR control unit updates the control memory according to the result from the channel decoder.

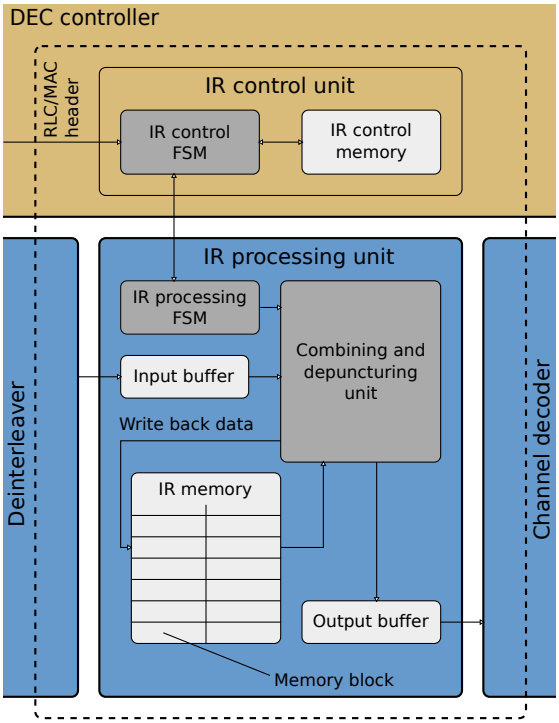


Fig. 5. Block diagram of the proposed IR architecture with dedicated memories. This corresponds to the shaded area inside DEC of Fig. 2.

The IR processing unit is responsible for depuncturing, combining previous transmissions, and storing soft values in the IR memory. Furthermore, it contains an input buffer for the newly received and punctured RLC block and an output buffer for the combined and depunctured RLC block ready to be decoded. The input buffer is used in order to write back the newly received RLC block to the IR memory in case decoding should fail. However, newly received RLC blocks are written back to the IR memory without waiting for the result of the channel decoder or, in other words, as soon as the combined version is in the output buffer. Correspondingly, the IR units can perform write back operations and process the next RLC block while the channel decoder is processing the previous RLC block. In case the same punctured version is already present in the IR memory the two transmissions are combined before write back in order to save IR memory. In order to ease memory access and management, IR memory is fragmented into memory blocks. A punctured RLC block occupies 1 or more such memory blocks depending on its size.

B. Synthesis and Clocking

The architecture as described above was implemented in VHDL and subsequently synthesized using Synopsys Design Compiler software with a 130 nm technology.

The input buffer must meet the largest punctured RLC block which consists of $s_{\mathrm{p,max}} = 1248$ soft values whereas the output buffer must be able to store the largest depunctured RLC block which consists of $s_{\mathrm{d,max}} = 2022$ soft values [9]. An example soft value width of $w = 5$ bits and an example memory capacity of $2^{15} = 32768$ soft values was used. The latter corresponds to 46 punctured or 16 depunctured DAS12 RLC blocks and in terms of MCS9 this corresponds to 53 punctured or 17 depunctured RLC blocks [9]. As can be seen in the performance simulations from the previous section MCS9 always requires less memory than DAS12 in terms of punctured RLC blocks. The memory sizes of input buffer, output buffer, IR memory, and control memory are listed in Table III. The circuit can be clocked with a maximum clock frequency of $f_{\mathrm{c,max}} = 187$ MHz which is sufficient for the fastest modes of Evolved EDGE. The circuit (without memory) corresponds to 50k Gate Equivalents (GE) at $f_{\mathrm{c,max}}$.

Memory	Soft values	Size (kbit)
Control memory	N/A	2.438
Input buffer	1248	6.24
Output buffer	2022	10.11
IR memory	32768	163.84
Total memory	36038 (without control)	182.628

TABLE III
MEMORIES AND THEIR SIZES FOR THE IMPLEMENTED ARCHITECTURE.

VI. Conclusions

The open source MatPHY framework has been extended with packet switched operation including Evolved EDGE and L1CTL extensions to connect with RLC/MAC software. IR performance with a number of receiver configurations has been studied. It has been proposed that EDGE and Evolved EDGE IR operations can be implemented efficiently as part of the PHY without burdening the system processor. RLC/MAC engineers no longer need to be aware of IR operations which simplifies protocol design. Such an implementation imposes less traffic between processors in a classical GSM/EDGE architecture. In fact, IR operations have been completely obfuscated from higher layers.

References

[1] *3GPP TS 45.005: Radio transmission and reception*, TS 45.005, Rev. 11.0.0, Jun. 2012. [Online]. Available: http://www.3gpp.org/
[2] E. Seurre, P. Savelli, and P. Pietri, *EDGE for Mobile Internet*. Artech House Publishers, 2003.
[3] C. Benkeser, A. Bubenhofer, and Q. Huang, "A 4.5 mW Digital Baseband Receiver for Level-A Evolved EDGE," in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*. IEEE, 2010, pp. 276–277.
[4] L. Chang and Y. Wang, "EDGE Incremental Redundancy Memory Structure and Memory Management," Jul. 23 2009, uS Patent App. 12/507,835.
[5] H. Kröll, C. Benkeser, S. Zwicky, B. Weber, and Q. Huang, "Baseband Signal Processing Framework for the OsmocomBB GSM Protocol Stack," in *Proceedings of the SDR'12 WinnComm Europe, 2012 Wireless Innovation Forum*. Wireless Innovation Forum, 2012, pp. 1–6. [Online]. Available: http://code.google.com/p/matphy/
[6] *3GPP TS 44.060: General Packet Radio Service (GPRS); Mobile Station (MS) - Base Station System (BSS) interface; Radio Link Control / Medium Access Control (RLC/MAC) protocol*, TS 44.060, Rev. 11.0.0, Mar. 2012. [Online]. Available: http://www.3gpp.org/
[7] "OsmocomBB," Jan. 2013. [Online]. Available: bb.osmocom.org
[8] *3GPP TS 43.064 General Packet Radio Service (GPRS); Overall description of the GPRS radio interface; Stage 2*, TS 43.064, Rev. 11.0.0, Sep. 2012. [Online]. Available: http://www.3gpp.org/
[9] *3GPP TS 45.003 Channel coding*, TS 45.003, Rev. 11.0.0, Sep. 2012. [Online]. Available: http://www.3gpp.org/

REAL-TIME VALIDATION OF A SDR IMPLEMENTATION OF TDD WIMAX STANDARD

Ángel Carro-Lagoa (University of A Coruña, A Coruña, Spain; acarro@udc.es);
Pedro Suárez-Casal (University of A Coruña, A Coruña, Spain; pedro.scasal@udc.es);
Paula Fraga-Lamas (University of A Coruña, A Coruña, Spain; paula.fraga@udc.es);
José A. García-Naya (University of A Coruña, A Coruña, Spain; jagarcia@udc.es);
Luis Castedo (University of A Coruña, A Coruña, Spain; luis@udc.es);
Antonio Morales-Méndez (Indra Sistemas S.A., Aranjuez, Spain; ammendez@indra.es)

ABSTRACT

This paper focuses on the validation of an innovative software-defined radio architecture for a WiMAX system based on commercially available field-programmable gate array and digital signal processor modules. We provide a real-time implementation of a standard-compliant time-division duplex physical layer including a mobile and a base station as well as downlink and uplink communications, thus obtaining a full-featured physical layer. Additionally, a set of different configurations are supported as described in the standard and in the WiMAX Forum. The main contribution of the paper consists in a reproducible and repeatable validation of the implementation in representative scenarios. At the same time, a characterization of the performance exhibited by the system is provided based on bit error rate measurements carried out using a custom-made, real-time channel emulator.

1. INTRODUCTION

Worldwide Interoperability for Microwave Access (WiMAX) was initially conceived for wireless broadband access, but evolved with time until becoming a candidate technology for the so-called 4G mobile networks. In late 90s, IEEE started a working group to create a Point-to-Multi-Point (PMP) air interface that was proposed as an alternative to cable and digital subscriber line. The name was next coined by the WiMAX Forum, which was constituted to promote the interoperability of the standard. The original standard was modified as the IEEE 802.16d for fixed applications using Orthogonal Frequency-Division Multiplexing (OFDM) as the transmission scheme. In 2005, mobility support was incorporated based on Scalable Orthogonal Frequency-Division Multiple Access (SOFDMA), resulting in the amendment 802.16e also known as Mobile WiMAX. Four years later, the standard IEEE 802.16-2009 was released to support both fixed and mobile wireless communications. A complete survey of the historical evolution of the standard up to 2010 can be found in [1]. Recently, in 2011, WiMAX evolved to 802.16m [2], which focuses on providing an advanced air interface to fulfill the requirements of IMT-Advanced while maintaining backward compatibility with existing specifications. In August 2012, the IEEE 802.16-2012 [3] was released, consolidating material from IEEE 802.16j-2009 for relay-based networks and the amendment 802.16h-2010, which implements coexistence enhancement for license-exempt operation. Such a standard also incorporated the IEEE 802.16m-2011, but excluding the WirelessMAN Advanced Air Interface, which is now specified in the IEEE 802.16.1-2012 [4]. Finally, improvements focused on machine-to-machine applications are examined in amendments 802.16p-2012 [5] and 802.16.1b-2012 [6].

The physical layer is in charge of multiplexing user and system data together with control signaling in order to ensure a proper utilization of the radio resources. The design of the physical layer specifies how to map and how to allocate those resources as either reference signals or to form various physical channels. WiMAX supports several physical layer modes. Among them, OFDMA is the most appealing given its flexibility and ability to support multiple users at the same time.

WiMAX specifies both Frequency-Division Duplex (FDD) as well as Time-Division Duplex (TDD) operating modes. The election of TDD was based on its capabilities for dealing with the asymmetrical uplink/downlink data flow. In this way, one of the incentives of the present work is the scarcity of complete, real-time, OFDMA-TDD, mobile physical layer implementations available in the literature. Existing works focus on performance analysis such as path-loss measurements using Fixed WiMAX commercial equipment in rural environments [7], tests in outdoor scenarios employing commercial Mobile WiMAX equipment [8], or evaluations of the IEEE 802.16e OFDMA downlink in vehicular environments (ITU-R M.1225) [9]. In this context, simulations [10], non-real-time deployments [11, 12], and Fixed WiMAX implementations [13] are found. However, none of the aforementioned approaches accounts for software constraints or hardware limitations.

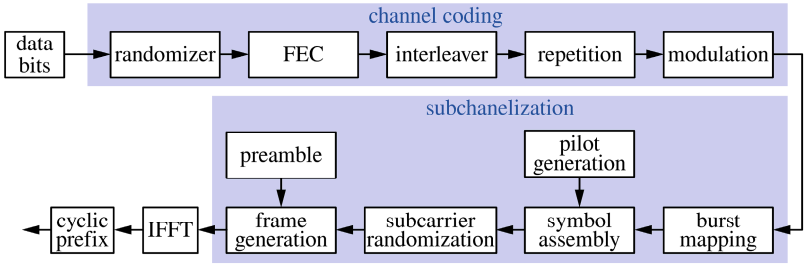


Figure 1: IEEE 802.16 transmitter block diagram.

The experimental validation of prototype baseband systems employing real-time hardware demonstrators confronts several challenges to be taken into account. Among them, formidable design complexity, long development time, high costs and manpower, or dealing with unsurmountable hardware issues are highlighted. To the knowledge of the authors, although several papers can be found in the literature describing real-time implementation and validation of downlink [14] or uplink channels [15], no one has been found dealing with both at the same time. It is worth mentioning that Mobile WiMAX duplex communications in a system-on-chip platform can be found in [16].

The main contribution of this paper is to describe the performance evaluation of the OFDMA-TDD physical layer implementation by means of repeatable and reproducible performance measurements of the Mobile WiMAX system under realistic conditions, validating the system as well as showing the versatility of the design through a wide variety of configurations.

The remainder of this article is organized as follows. Section 2 provides a brief description of the Mobile WiMAX physical layer, followed by the definition of the proposed hardware architecture in Section 3. Section 4 details the hardware components and the mapping of the real-time tasks to the hardware resources. The amount of resources consumed by the implementation is also detailed. Section 5 addresses the validation of the system, which is carried out by means of performance measurements over ITU-R channel models generated by a custom-made channel emulator. Finally, Section 6 is devoted to the conclusions and future research directions.

2. MOBILE WIMAX PHYSICAL LAYER

Mobile WiMAX is based on the OFDMA physical layer defined in the IEEE 802.16e standard. It supports both TDD and FDD operation modes whilst allowing for variable bandwidth and a scalable number of subcarriers ranging from 128 to 2048. Furthermore, WiMAX Forum specifies five profiles for interoperability as shown in Table 1. Such profiles combine different FFT sizes, bandwidths, and sampling frequencies.

The block diagram of the transmitter defined in the IEEE 802.16e standard is shown in Fig. 1. In TDD, a frame

Table 1: Mobile WiMAX profiles.

WiMAX profile	channel bandwidth [MHz]	sampling frequency [MHz]	FFT size
1	3.5	4	512
2	5	5.6	512
3	7	8	1024
4	8.75	10	1024
5	10	11.2	1024

is divided into downlink and uplink subframes. The downlink subframe consists of a preamble followed by a Frame Control Header (FCH), DL-MAP, and UL-MAP messages. If an UL-MAP message is sent to describe the uplink structure, it must be included in the first burst defined in the DL-MAP.

Mapping of bursts on subframes can be done using different permutation schemes such as non-adjacent groupings with Partial Usage of Subcarriers (PUSC) or Full Usage of Subcarriers (FUSC). The smallest data allocation unit is the slot, which is used to specify the time-frequency regions for data in the bursts. The slot definition varies according to the subcarrier grouping scheme.

Uplink resources are shared among mobile stations and their management is centralized on the Base Station (BS), which decides the amount of slots to be assigned to each mobile station based on the Quality-of-Service (QoS) parameters and bandwidth requirements for each connection. Additionally, the so-called ranging regions can be defined in the uplink to allow subscriber stations to perform network entry or to improve uplink synchronization parameters.

Data and pilot subcarriers are scrambled before the Inverse Fast Fourier Transform (IFFT) followed by Cyclic Prefix (CP) insertion are applied (see Fig. 1). Notice that the length of the cyclic prefix can take one the following values: 1/4, 1/8, 1/16, and 1/32.

In the IEEE 802.16e-2005, the channel coding stage consists of the following steps: data randomization, channel coding, bit-interleaving, repetition coding, and symbol mapping. Data randomization is performed in both uplink and downlink employing the output of a maximum-length shift-register sequence initialized at the beginning of every FEC block. Such a FEC block consists of an intenger num-

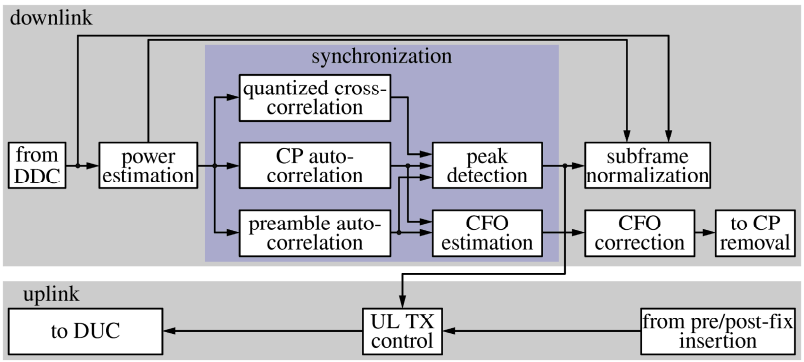


Figure 2: Downlink synchronization subsystem.

ber of subchannels. Channel coding is performed on a per-FEC-block basis employing one of the schemes defined in the standard, namely, Tail-Biting Convolutional Codes (TBCC), Block Turbo Codes (BTC), Convolutional Turbo Codes (CTC), and Low-Density Parity Check Codes (LDPC). Additionally, variable coding rate and modulation are supported, thus enabling for Adaptive Modulation and Coding (AMC) capabilities. Furthermore, Repetition coding with factors of 2, 4, or 6 are employed to increase the resilience of important control data. Finally, the modulation stage maps the coded bits into QPSK, 16-QAM, or 64-QAM constellations.

3. PROPOSED SYSTEM ARCHITECTURE

The real-time implementation described in this section focuses on the mandatory parts of the Mobile WiMAX physical layer for base and subscriber stations. It employs the OFDMA-TDD frame structure, PUSC permutation scheme both in the downlink and the uplink subframes, ranging, and channel coding based on TBCC.

3.1 Downlink Synchronization

Downlink synchronization consists of frame and symbol detection in the mobile station. Such tasks take advantage of the correlation properties exhibited by the preamble defined in the standard as well as those found in OFDM symbols when the cyclic prefix is included. The subsystem in charge of the downlink synchronization is shown in Fig. 2.

Preambles in Mobile WiMAX have a fixed structure with two guard subcarriers inserted between each pilot subcarrier and whose values are selected from a predefined set depending on the segment and base-station cell identifier. Frame detection is based on the Repetition Property Based (RPB) autocorrelation metric [17]. Next, two frequency offset estimations are obtained computing the angle of the previously declared autocorrelation values. The first one is extracted directly from the preamble, allowing for a wide frequency offset acquisition range. The second one is gathered from the cyclic prefixes, thus tracking the frequency offset on a per-OFDM-symbol basis.

3.2 Uplink Synchronization

Physical layers based on the OFDMA scheme require that uplink frames arrive at the base station at the same time and with a significant accuracy. This can only be achieved if all users are synchronized with the base station before the communication takes place. WiMAX standard states that the Round-Trip Delay (RTD) between the mobile station and the base station must be known beforehand by the mobile station, which is possible thanks to the so-called ranging.

For that purpose, mobile stations generate Pseudo Noise (PN) sequences from a shift register and they can be transmitted in specific regions of the uplink subframe. Such regions have to be reserved by the base station in a contention-based policy. At the receiver side (in the uplink), the base station detects the arrival of a ranging code, and then, it estimates the synchronization parameters. Finally, the mobile station adjusts its synchronization parameters from the base-station estimates sent back to the mobile station in a Medium Access Control (MAC) management message.

The standard defines different operation types depending on the current connection state: initial ranging on network entry, periodic ranging to update variations, bandwidth requests, and handover. During the initial ranging, OFDM symbols containing ranging codes are transmitted by the mobile station in pairs, the first one with a cyclic prefix, while the second one also adds a cyclic postfix, thus allowing for a wider time-synchronization margin.

Code detection and time-offset estimation at the base station takes place at the frequency domain over each received OFDM symbol. Firstly, a power threshold (see Fig. 2) followed by cross-correlation between the arriving subcarriers and all possible ranging codes is performed [18]. Once the ranging code is known, the received PN sequence is mapped back to the OFDM symbol. Since the initial rang-

ing forces mobile stations to transmit the same ranging code twice in two consecutive symbols, this property is exploited to extract the frequency offset through a correlation.

3.3 Subchannelization and Channel Equalization

Subchannelization involves interleaving and randomizing subcarriers followed by a permutation scheme and a pilot pattern. The base station specifies this structure for each frame using the dedicated DL-MAP and UL-MAP messages. Therefore, DL-MAP messages become critical since most of the processing of the downlink subframe at the receiver cannot start until this message has been completely decoded. On the other hand, subcarrier randomization in the uplink cannot be applied to the ranging bursts, thus this process depends entirely on the uplink burst scheme defined by the base station. We assign this task to the Digital Signal Processor (DSP) to provide the maximum flexibility with respect to the different sizes of the FFT, burst mapping, and eventual support of other permutation schemes (see Fig. 3).

Channel estimation and equalization is performed by inverting the pilot subcarriers and linearly interpolating the computed values for the remainder subcarriers.

3.4 Channel Coding

The proposed design supports a variable-rate TBCC coding scheme with constellation sizes varying from QPSK to 64-QAM both in downlink and uplink.

There are several techniques to design TBCC using standard convolutional encoders and Viterbi decoders [19]. The chosen technique offers a good trade-off between computational complexity and performance. The encoder is implemented adding a cyclic prefix to each FEC block with a size equal to the constraint length (in the case of WiMAX, such a value is set to 7). On the other hand, the decoder concatenates the first bits of the block at the end and vice versa, thus removing the additional bits from the decoder output.

The size of the chunks added at the beginning and at the end of the blocks is equal to the traceback length of the Viterbi decoder. If a block is shorter than the traceback length, it is just sent three times to the decoder and only the output corresponding to the second repetition is considered.

Additionally, the decoder computes a Carrier-to-Interference Noise Ratio (CINR) metric employing a soft decisor to estimate the transmitted symbols. It was verified that the algorithm provides accurate values of the CINR as long as decision errors are low. Otherwise, the CINR is overestimated.

Channel coding is mainly implemented in a Virtex II FPGA (see Fig. 3), although the optional repetition coding step performed over the constellation-mapped data and the processing control are both carried out in the DSP, using the FPGA as a coprocessor.

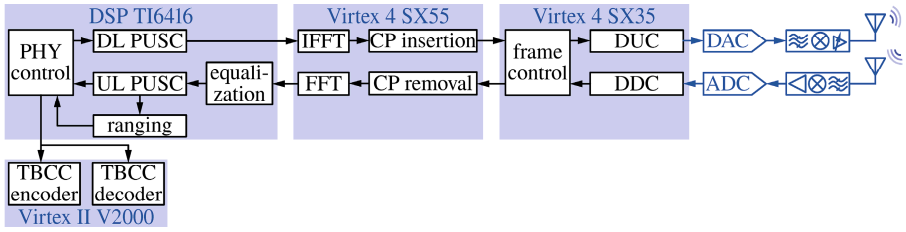
3.5 Physical Layer Control

Separation between MAC-level and physical-layer-level processing is obtained using the so-called OFDMA physical-layer Service Access Point (SAP) specification defined by Intel for its base stations [20]. Such a SAP provides the description of subframes, it sends and acquires data bursts, and it transmits and detects ranging codes.

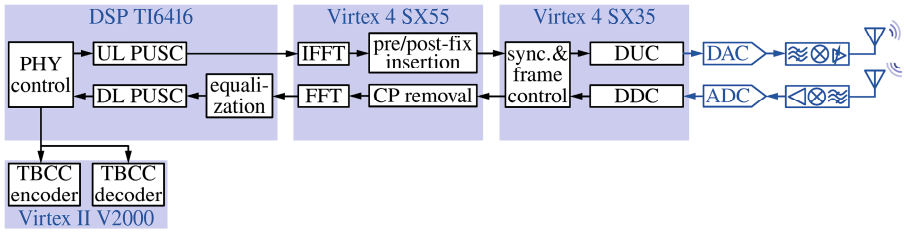
The subframe structure is transferred to the mobile station through MAC management messages (DL-MAP, UL-MAP, Downlink Channel Descriptor (DCD), and Uplink Channel Descriptor (UCD)). Mobile WiMAX requires data bursts to be rectangular-shaped while spanning a multiple of two symbols in time and a multiple of a subchannel in frequency (the so-called slot unit). Although the standard allows for more than a single burst per mobile station, the corresponding DL-MAP overhead is increased. Moreover, the standard also allows for more than a single connection packed into a burst. Finally, the base station has to distribute the available resources between users, guaranteeing their QoS requirements.

Multiple burst-mapping proposals for Mobile WiMAX are presented in [21]. An approach suitable for the downlink is the so-called Ohseki algorithm [22], a reference algorithm considering complexity, requested bandwidth, and the shape of the downlink burst.

Resource management in the uplink is more flexible since it is only necessary to indicate the number of slots allocated to each station. The size of such allocations is decided by the MAC layer considering the QoS restrictions.



(a) Base station.



(b) Mobile station.

Figure 3: Hardware components and real-time software tasks with their allocation for the base station (a) and the mobile station (b).

4. HARDWARE DESCRIPTION

In the proposed design for a bi-directional TDD WiMAX system, mobile and base stations were implemented utilizing the same type of hardware elements based on Commercial Off-The-Shelf (COTS) components. An overview of the architecture defined for the base station and the mobile station is shown in Fig. 3. Each station consists of three different FPGAs and a DSP module placed on a Peripheral Component Interconnect (PCI) carrier board.

The first module contains a Texas Instruments TMS320C6416 DSP together with a Xilinx Virtex-II XC2V2000 FPGA. The second module has a Xilinx Virtex-4 XC4VSX55 FPGA, while the third one is equipped with a Virtex-4 XC4VSX35 FPGA together with an analog add-on module containing a dual Digital-to-Analog Converter (DAC) and a pair of Analog-to-Digital Converters (ADCs). Note that both Virtex-4 FPGAs are equipped with a large number of embedded multipliers, thus enabling intensive signal processing operations.

Data exchange between hardware modules is achieve using proprietary buses that can reach up to 400 MB/s, together with control-data buses limited to 20 MB/s. A PCI bus links the carrier board containing the hardware modules to a host computer.

In order to validate the real-time implementation as well as to assess the performance of the system, a channel emulator was implemented on a Xilinx Xtreme DSP Development Kit consisting of a Virtex-4 FPGA plus a couple of DACs and ADCs (see Fig. 4).

4.1 Digital Up/Down Conversion

Digital Up-Conversion (DUC) and Digital Down-Conversion (DDC) adapt the signals to the sampling rate of ADCs and DACs (see Fig. 3). In the case of the DUC, the following tasks are done: upsampling, pulse shaping, and I/Q modulation to a configurable intermediate frequency. The DDC performs the complementary operations in the reverse order. The chosen profiles selected by WiMAX Forum are supported providing five different bit-streams for the FPGAs. Each bit-stream has a different up/down-sampling factor. Since the DACs and the ADCs are configured with a sampling frequency of 80 MHz, the factors for each profile are respectively 20, 100/7, 10, 8, and 50/7 for profiles 1 to 5 (see Table 1). Each FPGA bit-stream also has a different optimized combination of interpolation/decimation filters in order to efficiently implement these sampling-rate conversions [23].

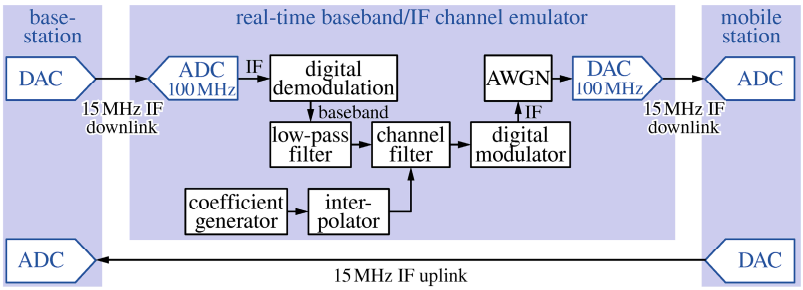


Figure 4: Real-time custom-made channel emulator utilized for validating the implementation.

Table 2: Resource utilization of the base station FPGAs.

base station	Virtex-II	Virtex-4 SX55	Virtex-4 SX35
Slice	10131/10752 (94%)	13785/24576 (56%)	6580/15360 (41%)
LUT	13509/21504 (62%)	18356/49152 (37%)	8261/30720 (26%)
RAMB16	52/56 (92%)	113/320 (35%)	45/192 (23%)
Multi-pliers	2/56 (3%)	116/512 (22%)	24/192 (12%)

Table 3: Resource utilization of the mobile station FPGAs.

mobile station	Virtex-II	Virtex-4 SX55	Virtex-4 SX35
Slice	10131/10752 (94%)	14692/24576 (59%)	15358/15360 (99%)
LUT	13509/21504 (62%)	19951/49152 (40%)	22625/30720 (73%)
RAMB16	52/56 (92%)	114/320 (35%)	52/192 (27%)
Multi-pliers	2/56 (3%)	117/512 (22%)	70/192 (36%)

4.2 Resource Utilization

FPGA designs were implemented using the Xilinx System Generator and built with Xilinx ISE 10.1. The resource utilization of the FPGAs is shown considering Slices, LUTs, RAMB16s, and multipliers for both base station (see Table 2) and mobile station (see Table 3).

The relatively high consumption of FPGA resources is explained because all FPGA tasks were implemented using the System Generator high-level development tool. Therefore, the implementation process becomes easier and faster, but the price to be paid is a non-optimum code in terms of resource utilization.

On the other hand, the higher resource utilization at the mobile station is because first, the synchronization block – which requires 68% of the slices available at the Virtex-4 SX35–, and second, the concatenation of cyclic postfixes at the IFFT output.

Table 4: Specification of the ITU-R M.1225 channel models employed in the performance evaluation.

	Pedestrian A	Pedestrian B	Vehicular A
Number of paths	4	6	6
Power of each path [dB]	0, –9.7, –19.2, –22.8	0, –0.9, –4.9, –8.0, –7.8, –23.9	0, –1.0, –9.0, –10.0, –15.0, –20.0
Path delay [ns]	0, 110, 190, 410	0, 200, 800, 1200, 2300, 3700	0, 310, 710, 1090, 1730, 2510
Speed [km/h]	3	3	60, 120

The utilization of a large FPGA as the Virtex-4 SX55 allows for a pipelined architecture dedicated to the FFT implementation. In case of the Virtex-II, an optimized FEC design was required due to the limited resources available.

Finally, apart from subchannelization, channel estimation and equalization, subcarrier mapping, and ranging (only at the base station); the SAP protocol, the minimum functionalities from the MAC required for the bi-directional operation, and data exchange with the host are all carried out by the DSP, thus demanding for a significant amount of resources from the DSP.

5. PERFORMANCE EVALUATION

This section describes the performance evaluation of the developed bi-directional TDD WiMAX physical layer. A diagram of the employed setup is shown in Fig. 4. For evaluating the downlink in a repeatable and in a reproducible fashion, the corresponding downlink subframes are transmitted across the channel emulator, whilst the uplink is connected with a cable. The reverse configuration is considere for evaluating the uplink.

The custom-made, real-time channel emulator was configured following the WiMAX Forum recommendations. Therefore, ITU-R M.1225 Pedestrian A at 3 km/h, Pedestrian B at 3 km/h, Vehicular A at 60 and at 120 km/h were considered assuming the tapped delay line characteristics summarized in Table 4. Notice that frequency selectivity

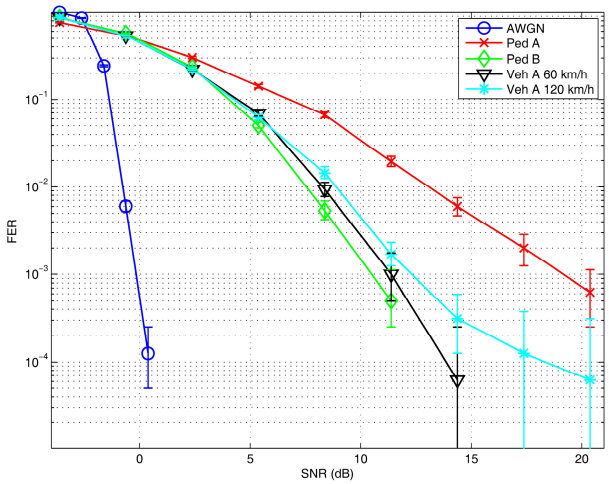


Figure 5: Downlink subframe detection in AWGN and ITU-R channels expressed in terms of FER with respect to the SNR.

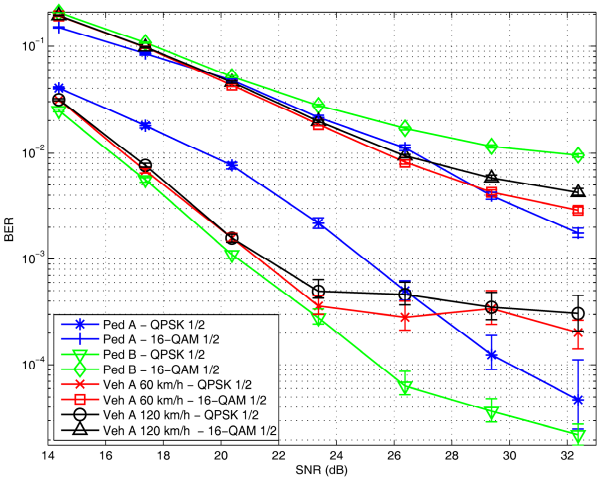


Figure 7: BER over ITU-R channels using the 8.5 MHz downlink profile.

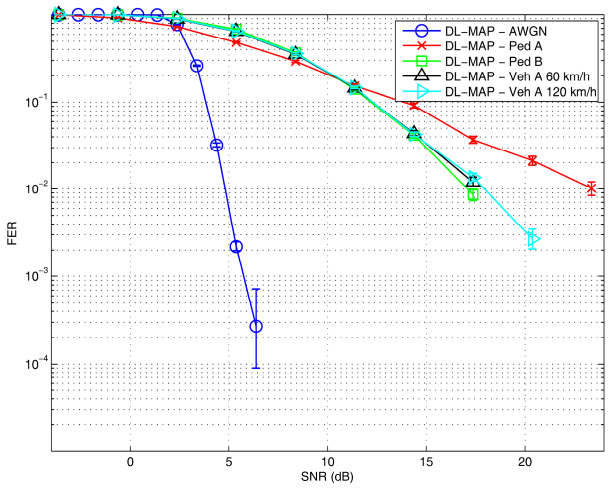


Figure 6: DL-MAP FER in AWGN and ITU-R channels.

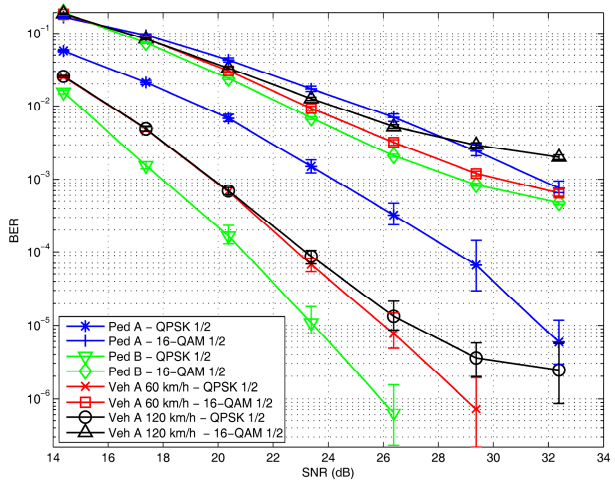


Figure 8: BER over ITU-R channels using the 8.5 MHz uplink profile.

exhibited by Pedestrian A is very low, whereas Pedestrian B and Vehicular A models provide richer multipath diversity and higher path delay spread than Pedestrian A.

Doppler spread computation is performed based on the Jakes Doppler power spectrum density assuming transmissions at 2.4 GHz. Note also that the maximum channel delay (3700 ns) does not exceed in any case the default 1/8 cyclic prefix length (11.429 ns). Consequently, the system is immune to Inter-Symbol Interference (ISI).

In order to plot the figure of merit with respect to the SNR, the AWGN generator included in the channel emulator is calibrated to match the SNR estimation obtained during the synchronization process.

First, downlink frame detection is evaluated transmitting more than 10^4 frames and counting the number of them correctly detected at the mobile station. The performance over AWGN and ITU-R channel models is shown in Fig. 5. In order to provide a measure of the uncertainty of the results, all performance results include 90% confidence intervals for the mean computed using bootstrapping.

Figure 5 shows that the results for the AWGN channel clearly outperform those obtained for the ITU-R channel models, which is mainly due to the channel fading. Pedestrian B and Vehicular A present a similar behavior, achieving FER values of 10^{-3} with a SNR value of 12 dB. Pedestrian A exhibits the worst performance caused by its lower

multipath diversity, thus making the synchronization more difficult.

According to the WiMAX standard, downlink synchronization is acquired if DL-MAP messages are correctly decoded. Following this criterion, FER is obtained distinguishing between undetected frames and frames in which the FDH or the DL-MAP are not correctly decoded (see Fig. 6). DL-MAP messages were sent using QPSK, convolutional coding at a rate of 1/2 with no repetitions, and a size of 28 bytes including the header and the Cyclic Redundancy Check (CRC).

Comparing Figs. 5 and 6 we can conclude that the frame detection has no negative impact on the system performance since the SNR for incorrectly detecting a frame is 5 dB lower than that for misdetecting the DL-MAP.

Uplink time-offset synchronization module is evaluated by the mobile station sending ranging codes while the base station stores the computed estimations. Following this procedure, it can be shown that the time offset estimation in the uplink is not affected by the particularities of the channel models, thus being suitable for all of them.

BER measurements are carried out considering a known structure for the subframes in order to avoid the case in which FCH or the DL-MAP messages cannot be decoded correctly. Figure 7 presents the BER for the 8.75 MHz downlink profile when the ITU-R channel models are employed. The Pedestrian A presents consistent results with those obtained for frame detection (see Figs. 5 and 6). Once higher SNR levels are achieved, the Pedestrian A outperforms other channel models since its channel frequency response is easier to equalize.

BER measurements for the uplink transmission for the 8.75 MHz profile are shown in Fig. 8. An improvement with respect to the downlink transmission is observed. Although surprising in principle, this behavior is perfectly justified when comparing the pilot structure of the uplink and downlink subframes. The effect is especially outstanding in the case of the Pedestrian B channel model. Due to the limited coherence time of Vehicular A channel models, there is a channel estimation error causing the error floor observed in Figs. 7 and 8, corresponding to the downlink and the uplink, respectively.

6. CONCLUSION

In this paper we have proposed the design and implementation of a real-time, bi-directional TDD physical layer compliant with the Mobile WiMAX standard. We have detailed the utilized SDR hardware architecture consisting of commercial off-the-shelf modules based on FPGAs and DSPs for both mobile and base stations. We have described in detail most of the design decisions pursuing an efficient utilization of the FPGA resources.

The implementation is validated in a repeatable and in a reproducible way by means of performance measurements carried out with the help of a real-time, custom-made channel emulator. Such a channel emulator implements AWGN as well as Pedestrian A and B, and Vehicular A channel models recommended by the WiMAX Forum. First, we tested the suitability of our synchronization algorithms, ensuring that they do not limit the performance of the system. Secondly, BER measurements were carried out for both the uplink and the downlink for the 8.75 MHz profile. The measurement results confirm that the proposed implementation is suitable for the scenarios modeled with the aforementioned channel models.

Finally, future research will be devoted to adapt our designs to the recently proposed WirelessMAN-Advanced Air Interface included in the IEEE 802.16m.

ACKNOWLEDGEMENTS

This work has been partially supported by Indra Sistemas S.A. and the Spanish Ministry of Defence with the technical direction of PEC/ITM under grant DN8644-COINCIDENTE. The authors wish to thank J. M. Camas-Albar from Indra Sistemas S.A. for his help.

This work has been additionally funded by Xunta de Galicia, Ministerio de Ciencia e Innovación of Spain, and FEDER funds of the European Union under grants with numbers 10TIC003CT, 09TIC008105PR, TEC2010-19545-C04-01, and CSD2008-00010.

REFERENCES

[1] D. Pareit, B. Lannoo, I. Moerman, P. Demeester, “The History of WiMAX: A Complete Survey of the Evolution in Certication and Standardization for IEEE 802.16 and WiMAX”, IEEE Communications Surveys Tutorials.

[2] IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 3: Advanced Air Interface, IEEE Std 802.16m-2011(Amendment to IEEE Std 802.16-2009).

[3] IEEE Standard for Air Interface for Broadband Wireless Access Systems, IEEE Std 802.16-2012 (Revision of IEEE Std 802.16-2009).

[4] IEEE Standard for WirelessMAN-Advanced Air Interface for Broadband Wireless Access Systems, IEEE Std 802.16.1-2012.

[5] IEEE Standard for Air Interface for Broadband Wireless Access Systems. Amendment 1: Enhancements to Support Machine-to-Machine Applications, IEEE Std 802.16p-2012 (Amendment to IEEE Std 802.16-2012).

[6] IEEE Standard for WirelessMAN-Advanced Air Interface for Broadband Wireless Access Systems Amendment 1: Enhancements to Support Machine-to-Machine Applications, IEEE Std 802.16.1b-2012 (Amendment to IEEE Std 802.16.1-2012).

[7] P. Imperatore, E. Salvadori, I. Chlamtac, “Path Loss Measurements at 3.5 GHz: A Trial Test WiMAX Based in Rural Environment”, 3rd International Conference on Testbeds and

Research Infrastructure for the Development of Networks and Communities, 2007. TridentCom 2007, 2007, pp. 1 - 8. doi:10.1109/TRIDENTCOM.2007.4444709.

[8] G. Zaggoulos, M. Tran, A. Nix, "Mobile WiMAX system performance - simulated versus experimental results", IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications, 2008. PIMRC 2008, 2008, pp. 1 -5. doi:10.1109/PIMRC.2008.4699670.

[9] R. Colda, T. Palade, E. Pucchita, I. Vermecan, A. Moldovan, "Mobile WiMAX: System performance on a vehicular multipath channel", Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP), 2010, 2010, pp. 1 - 5.

[10] Z. F. Mohamed M.A., M. R.H., "Simulation of WiMAX Physical Layer: IEEE 802.16e", IJCSNS International Journal of Computer Science and Network Security 10 (11).

[11] S. Hu, G. Wu, Y. L. Guan, C. L. Law, Y. Yan, S. Li, "Development and performance evaluation of mobile WiMAX testbed", IEEE Mobile WiMAX Symposium, 2007., 2007, pp. 104 -107. doi:10.1109/WIMAX.2007.348688.

[12] C. Mehlführer, S. Caban, M. Rupp, "Experimental Evaluation of Adaptive Modulation and Coding in MIMO WiMAX with Limited Feedback", EURASIP Journal on Advances in Signal Processing 2008, Article ID 837102.

[13] H. Lai, S. Boumaiza, "WiMAX baseband processor implementation and validation on a FPGA/DSP platform", Canadian Conference on Electrical and Computer Engineering, 2008. CCECE 2008, 2008, pp. 001449-001452. doi:10.1109/CCECE.2008.4564781.

[14] K.-C. Chang, J.-W. Lin, T.-D. Chiueh, "Design of a downlink baseband receiver for IEEE 802.16E OFDMA mode in high mobility", IEEE International SOC Conference, 2007, 2007, pp. 301-304. doi:10.1109/SOCC.2007.4545479.

[15] Y.-J. Wu, J.-M. Lin, H.-Y. Yu, H.-P. Ma, "A baseband testbed for uplink mobile MIMO WiMAX communications", IEEE International Symposium on Circuits and Systems, 2009. ISCAS 2009, 2009, pp. 794-797. doi: 10.1109/ISCAS.2009.5117874.

[16] G. Chuang, P.-A. Ting, J.-Y. Hsu, J.-Y. Lai, S.-C. Lo, Y.-C.Hsiao, T.-D. Chiueh, "A MIMO WiMAX SoC in 90nm CMOS for 300km/h mobility", IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011, 2011, pp. 134-136. doi:10.1109/ISSCC.2011.5746252.

[17] H. Kang, S.-B. Im, H.-J. Choi, D.-J. Rhee, "Robust OFDMA Frame Synchronization Algorithm on Inter-Cell Interference", Communications, 2006. APCC '06. Asia-Pacific Conference on, 2006, pp. 1-5.doi:10.1109/APCC.2006.255832.

[18] H. Mahmoud, H. Arslan, M. Ozdemir, "Initial Ranging for WiMAX (802.16e) OFDMA", IEEE Military Communications Conference, 2006. MILCOM 2006, 2006, pp. 1 -7. doi:10.1109/MILCOM.2006.302240.

[19] R. Cox, C. Sundberg, "An efficient adaptive circular Viterbi algorithm for decoding generalized tailbiting convolutional codes", IEEE Transactions on Vehicular Technology 43 (1) (1994) 57 -68. doi:10.1109/25.282266.

[20] OFDMA PHY SAP Interface Speccation for 802.16 Broadband Wireless Access Base Stations, Tech. rep., Intel Corporation (2007).

[21] C. So-In, R. Jain, A.-K. Tamimi, "Scheduling in IEEE 802.16e Mobile WiMAX Networks: Key Issues and a Survey", IEEE Journal on Selected Areas in Communications 27 (2) (2009) 156 -171. doi:10.1109/JSAC.2009.090207.

[22] T. Ohseki, M. Morita, T. Inoue, "Burst Construction and Packet Mapping Scheme for OFDMA Downlinks in IEEE 802.16 Systems", IEEE Global Telecommunications Conference, 2007. GLOBECOM '07, 2007, pp. 4307-4311. doi:10.1109/GLOCOM.2007.819.

[23] P. Suárez-Casal, A. Carro-Lagoa, J. García-Naya, L. Castedo, "A Multicore SDR Architecture for Reconfigurable WiMAX Downlink", 13th Euromicro Conference on Digital System Design: Architectures, Methods and Tools (DSD), 2010, 2010, pp. 801-804.doi:10.1109/DSD.2010.108.

A LTE Receiver Framework Implementation in GNU Radio

Johannes Demel, Sebastian Koslowski, Friedrich K. Jondral
Karlsruhe Institute of Technology (KIT)
{johannes.demel, sebastian.koslowski, friedrich.jondral}@kit.edu

***Abstract*—We present an open source LTE receiver framework. Using GNU Radio's block-based signal processing capabilities, various LTE baseband specific functionality has been implemented in dedicated easily reconfigurable blocks. These can be used to decode and analyze arbitrary channels in the LTE downlink signal. As an example we decode the Master Information Block (MIB) transmitted on the Broadcast Channel (BCH). Our work is focused on performance measurements in order to identify critical processing operations on a General Purpose Processor (GPP). By optimizing critical components, e.g. channel estimation, synchronization, we can improve the overall system performance and therefore the system's real-time capabilities.**

***Index Terms*—LTE, GPP, GNU Radio, Performance**

I. Introduction

Over the last decades digital radio systems evolved from GSM with Time Division Multiple Access (TDMA) to UMTS with Code Division Multiple Access (CDMA) to Long Term Evolution (LTE) which uses Orthogonal Frequency Division Multiple Access (OFDMA). Goals for LTE development include higher data rates and lower latencies. Along with higher bandwidth modes some new techniques were introduced to reach these goals. For example Multiple Input Multiple Output (MIMO) is employed to achieve higher diversity or higher data rates.

The Long Term Evolution (LTE) standard [1] defines a multi-mode air interface based on Orthogonal Frequency Division Multiplex (OFDM) within its downlink structure. LTE requires high performance hardware to compute all its signal processing. This need is either met by Field Programmable Gate Arrays (FPGAs), by Application-specific integrated Curcuits (ASICs) or by optimized software for GPPs.

Modern radio communication systems rely on the ability to adopt to new environments. One approach to achieve more flexiblity is to utilize Software Defined Radio (SDR) techniques [2]. GNU Radio is a flexible open source SDR framework for GPPs, which provides a flexible and extendable signal processing infrastructure to easily adopt and add new capabilities. It also includes a rich signal processing library which is split into blocks that can be reused and regrouped as needed. Signal processing functions are implemented using C++ to hit high performance needs. A GNU Radio application consists of a flowgraph made up of several blocks and its connections between them. The setup and construction of the flowgraph is done using Python to facilitate easy usability. The GNU Radio Companion (GRC) further eases the development of a flowgraph by providing a graphical interface to GNU Radio.

In the following we describe our LTE receiver framework in GNU Radio. The focus of our work is on the implementation and performance of the physical layer receiver components. As an example we show the extraction of the MIB which is always broadcasted by a LTE base station (BS) to convey its basic configuration.

The remainder of this paper is structured as follows: In Section II the LTE downlink air interface and an outline of the receiver algorithms is described. Its implementation in GNU Radio is presented in Section III. The results of our performance measurements and its outcome for future optimization is evaluated in Section IV.

II. LTE Downlink Air Interface

LTE transmissions from a base station to a mobile terminal are frame-based multi-carrier signals. The transmitted symbols are organized in a scheme referred to as the OFDM time-frequency grid. A Resource Element (RE) is one modulation symbol on one subcarrier. REs are grouped into Resource Blocks (RBs) which are composed of twelve consecutive subcarriers and six or seven OFDM symbols depending on the Cyclic Prefix (CP) mode, normal or extended, respectively. The system bandwidth is defined as a multiple of RBs in frequency domain and may range from six to one hundred RBs as indicated in Tab. I.

RBs	bandwidth	subcarriers
6	1.25 MHz	72
15	3 MHz	180
25	5 MHz	300
50	10 MHz	600
75	15 MHz	900
100	20 MHz	1200

TABLE I
LTE bandwidth configuration options

In the time domain a transmission is organized in frames, see Fig. 1. Each frame consists of ten subframes, which include two slots with each six or seven OFDM symbols (OFDM symbols) grouped together in one slot. The number of symbols per slot varies according to the CP mode (and therefore its length). Assuming normal CP-length (N_{CP}) there are seven OFDM symbols in one slot and 14 OFDM symbols in

one subframe. A frame hence consists of 140 OFDM symbols. The transmitted frames are numbered by the System Frame Number (SFN) which is a cyclic counter from zero to 1023 and thus is represented as a ten bit binary value.

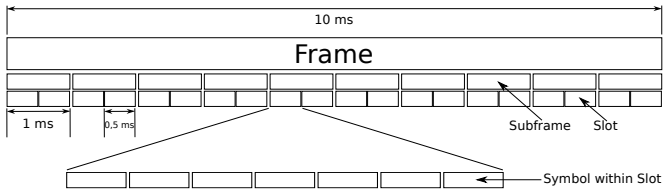


Fig. 1. LTE downlink frame structure

Together with the Reference symbols (RSs) for channel estimation and synchronization symbols, six channels are multiplexed on the time-frequency grid: the transport channels Physical Broadcast Channel (PBCH) which carries the essential system information, Physical Downlink Shared Channel (PDSCH) which carries the downlink data and Physical Multicast Channel (PMCH) for multicast purposes and the control information channels Physical Hybrid-ARQ Indicator Channel (PHICH) which provides an acknowledgement process and Physical Downlink Control Channel (PDCCH) and Physical Control Format Indicator Channel (PCFICH) which carry downlink control information.

A. Synchronization

For digital radio signal reception timing and frequency information of a transmitting station must be recovered in order to decode its signal. The synchronization used in our work was first described by [3]. Timing recovery is divided into several steps which include estimation of symbol timing along with Primary Synchronization Symbol (PSS) and Secondary Synchronization Symbol (SSS) detection. Whereas frequency recovery is divided into Integer Frequency Offset (IFO) and Fractional Frequency Offset (FFO) detection. PSS and SSS are not solely used for synchronization but also carry information on the Cell ID (N_{ID}). The N_{ID} is an identifier of the transmitting BS and has to be extracted to allow descrambling and the localization of RSs within each frame.

The first step in the synchronization process is symbol clock recovery exploiting the signal redundancy introduced by the CP. Symbol start detection is achieved by a sliding window correlation with a length of N_{CP} and a fixed lag of N_{FFT}:

$$\gamma(n) = \sum_{m=n}^{n+N_{\text{CP}}-1} r(m)\, r^*(m - N_{\text{FFT}}) \tag{1}$$

A subsequent magnitude peak detection within a search window of one OFDM symbol and averaging over the results yields a coarse estimate of the symbol timing.

In order to recover frame start there are two synchronization symbols embedded in the transmitted signal. First, the PSS which is sent every half frame is detected by correlation. The PSS is chosen amongst one of three Zadoff-Chu sequences with length-63 according to the Cell ID Number (N^2_{ID}) and enables half frame timing recovery. On the time-frequency grid it is located around the DC-carrier spanning over 6 RBs. Its 32nd value is set to zero because it is placed on the DC-carrier.

Besides half frame timing recovery using the PSS, IFO detection may be accomplished by a cross correlation with the received signal and all possible sent Zadoff-Chu sequences. FFO detection and compensation is accomplished afterwards by using the phase of the magnitude peak detected from (1).

Frame synchronization is finally achieved by a detection of the SSS. The SSS is also sent every half frame and consists of two interleaved m-sequences of length-31 according to the Cell Identity Group (N^1_{ID}). These sequences are interleaved differently depending on the SSS position within a frame and thus allow frame timing recovery.

After the synchronization is completed, the receiver has compensated frequency and timing offsets. Furthermore, the N_{ID} can be calculated from the separately extracted value N^1_{ID} and N^2_{ID}:

$$N_{\text{ID}} = 3 * N^1_{\text{ID}} + N^2_{\text{ID}} \tag{2}$$

B. Channel estimation and equalization

A transmission over the air is influenced by several types of distortion such as multipath scattering, fading and phase shifts. Especially a mobile communication system is therefore required to perform equalization on the received data which, in case of OFDM, can be achieved in the frequency domain.

For the purpose of correct data reception, the frequency response of the radio channel has to be estimated using the scattered pilots, or Reference symbols, embedded in each frame. Magnitude and phase distortion between received and known RSs are calculated and then a linear interpolation is performed to get channel estimates for the RE carrying data. These operations are based on the assumption that the maximum multipath delay is within the CP duration and channel coherence between RSs in time and frequency domain.

Equalization is performed in the frequency-domain using a one-tap equalizer per subcarrier. To simplify the channel estimation process a zero-forcing equalization was chosen for now.

C. Precoding

To improve diversity or data rate, LTE supports 2x2 and 4x4 MIMO configurations using Space Time Block Codes (STBCs) – or Alamouti Codes – [4]. In case of 2x2 MIMO the symbols $x(n)$ are coded and assigned to the transmit antennas t_n according to (3).

$$\begin{array}{c} \\ t_0 \\ t_1 \end{array} \begin{array}{c} \begin{array}{cc} \text{time}_0 & \text{time}_1 \end{array} \\ \begin{pmatrix} x(n) & x(n+1) \\ -x(n+1)* & x(n)* \end{pmatrix} \end{array} \tag{3}$$

As described above, the transmission channels introduce independent frequency flat channel distortions h_n to a receiver antenna signal $r_k(n)$.

$$\begin{array}{rcccl} r_0(n) & = & h_0\, x(n) & - & h_1\, x^*(n+1) \\ r_1(n) & = & h_0\, x(n+1) & + & h_1\, x^*(n) \end{array} \tag{4}$$

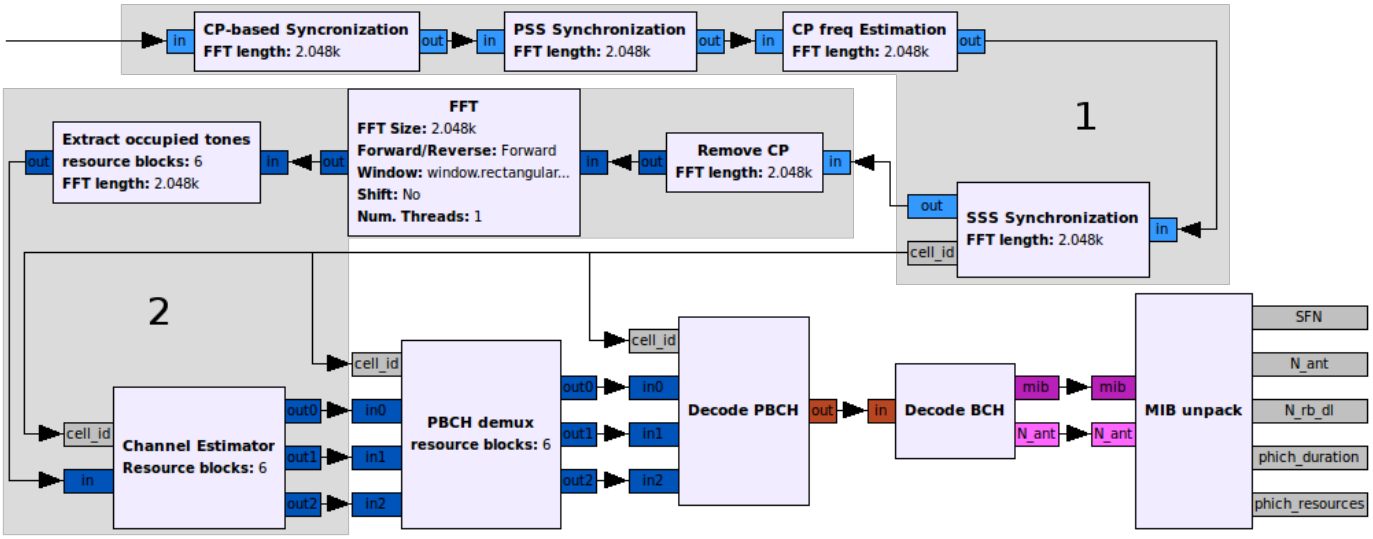


Fig. 2. GNU Radio flowgraph with Synchronization (1), OFDM receiver operation (2)

The transmitted symbols can be estimated by

$$\begin{aligned} \hat{x}(n) &= h_0^* r_0(n) + h_1 r_1^*(n) = 2x(n) \\ \hat{x}(n+1) &= h_0^* r_1(n) - h_1 r_0^*(n) = 2x(n+1) \end{aligned} \tag{5}$$

Assuming perfect equalization, the result indicates a 3 dB gain. For the LTE receiver presented here we are using one receive antenna for now.

D. Modulation and Coding

LTE supports several modulation schemes depending on the link quality. All channels can use a Quadrature Phase Shift Keying (QPSK) constellation for modulation. PDSCH and PMCH optionally allow the usage of 16QAM or 64QAM instead.

For transport channels tail biting convolutional coding and turbo coding are utilized. While PBCH uses tail biting convolutional coding the PDSCH uses turbo coding. For the control information channels a tail biting convolutional code is employed together with a block or repetition code.

III. LTE Receiver Implementation

In this section we present our GNU Radio implementation of the LTE receiver algorithms outlined in the previous section. All source code of the receiver implementation is available under the terms of GPL on Github [5].

The input can either be a recorded, generated or live data stream. It is expected to be a complex base band signal sampled at a multiple (power of two) of the LTE subcarrier spacing (15 kHz). Depending on the number of RBs to be extracted, the signal bandwidth has to be chosen according to Tab. I.

The overall flowgraph structure is presented in Fig. 2. Five main sections can be identified in this flowgraph: Synchronization (1), the OFDM receiver operation (2) and the PBCH-, BCH- and MIB-decoding.

A. OFDM receiver

In the first section synchronization is achieved together with the extraction of Cell ID (N_{ID}). The synchronization information is propagated downstream from block to block by using stream tags, which provide meta information on individual samples in the stream. The parameter N_{ID} is not associated with any sample in particular and is therefore published using GNU Radio's message passing infrastructure, which provides asynchronous data transfer capabilities between blocks.

The signal processing itself starts with the block *CP-based Synchronization* as described in Sec. II-A. Then, the PSS detection is performed in the block *PSS Synchronization*, which is a hierarchical block comprised of several blocks. After an initial acquisition phase all of these blocks are set into a tracking mode which reduces computing complexity by only correlating with the detected PSS at its expected position. Note, that due to the possibility of different lengths of the CP within one LTE slot fine frequency compensation based on CP correlation is only possible after *PSS Synchronization*. To complete the timing recovery processing, the *SSS Synchronization* block detects the SSS, computes the N_{ID} and, afterwards, also switches to tracking mode.

Once the synchronization is completed, the complex valued data stream is routed to section 2, which performs an inverse OFDM operation. This includes CP removal, Fourier transformation, extraction of subcarriers of interest and channel estimation. After those operations are completed, the received data is available as described in a time-frequency grid together with the corresponding channel estimates. All physical downlink channels can now be extracted from the time-frequency grid with their channel estimate values.

The next two sections of the flowgraph in Fig. 2 handle PBCH- and BCH-decoding. This is described in subsections III-B and III-C. The final step here is the *MIB unpack* block, which decodes the received MIB data and makes it available on

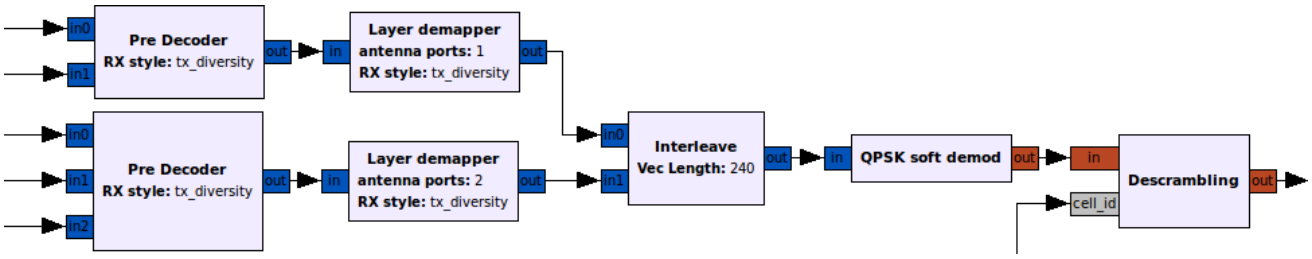


Fig. 3. Decode PBCH blocks

its message ports. It contains the number of RBs, the PHICH duration and resources and the SFN.

B. PBCH Decoding

The PBCH is transmitted in slot two of 20 using six RBs. Its payload is the BCH encoded using a repetition code, scrambled and spread over four OFDM frames. However, the whole BCH is transmitted in each OFDM frame using a different, initially unknown, phase of the scrambling sequence. In addition, the antenna configuration of the BS is unknown at this point. Therefore we process all of the different configurations in parallel *Pre Decoder* and *Layer Demapper* blocks. The resulting streams are interleaved as all further processing is the same. Once a MIB is decoded successfully, the antenna configuration can be deduced.

The block *Decode PBCH* in Fig. 2 combines the receiver side signal processing blocks of the PBCH shown in Fig. 3 into one hierarchical block [1]. First, the inverse Precoding operation is performed as indicated in Sec. II-C with channel estimate values calculated by the *Channel Estimator* block. The output of the *Pre Decoder* block is arranged according to the employed layer mapping which is described in [1] and the block *Layer demapper* does the inverse operation.

The block *QPSK soft demod* does QPSK soft demodulation and therefore marks the transition from complex to real values. In contrast to hard symbol decision soft demodulation was chosen to increase decoding performance.

The *Descrambling* block receives the N_{ID} via message passing which triggers an internal refresh of the descrambling sequence. The PBCH consists of 16 repetitions of the same data unit scrambled with a Pseudo Random Noise (PRN) sequence. For decoding, the data of one OFDM frame is repeated four times, descrambled and split into 16 data units according to the initial size of those data units. The position within the descrambled sequence of the successfully decoded data unit determines the Least Significant Bits (LSBs) of the SFN.

C. BCH Decoding

After the PBCH decoding operations the data units are passed to the hierarchical *Decode BCH* block which performs BCH-decoding [6] shown in Fig. 4.

The *Rate unmatch* block deinterleaves a data unit and by this prepares the correct sample order for the block *BCH Viterbi*. This block is a hierarchical block which parameterizes the GNU Radio Viterbi block and provides the correct data format. The output of the *BCH Viterbi* block are data units which carry the MIB information including a Cyclic Redundancy Check (CRC).

The block *CRC Calculator* checks if the received checksum is correct for different LTE antenna configurations assumed above. Depending on the antenna configuration of the BS a different scrambling scheme is chosen for the checksum and thus enables the receiver to determine this antenna configuration.

IV. Performance Analysis

In order to identify critical signal processing operations we use the profiling tool Valgrind [7]. For the performance analysis a computer with an Intel Core i3-2330M with 4 GiB of RAM running Xubuntu Linux 12.04 is used.

In Fig. 5 the different performance measurements are shown for the first running prototype and the currrent, optimized implementation. As the bar chart represents relative values, it can be seen that the sections *Decode PBCH*, *Decode BCH* and *MIB* consume far less CPU power. The performance improvements in these sections have been achieved by refactoring and optimization. GNU Radio's SIMD-optimized VOLK library has been employed where appropriate. The optimization of the synchronization section is yet to be completed and therefore shows no improvement over the initial version.

The receiver implementation allows to independently set the input sampling rate and the number of RBs processed after the inverse OFDM operation. The input sampling rate for the flowgraph directly corresponds to the Fast Fourier Transformation (FFT)-length. Thus increasing the FFT-length is expected to increase the CPU load of synchronization algorithms because these operate at the full input sampling rate. Fig. 6 shows the Instruction Fetch results of all custom GNU Radio LTE blocks. It can be observed that a smaller FFT-length results in a relative increase of CPU load for all sections but synchronization.

In Fig. 7 a comparison between different numbers of processed RBs is shown. The relative processing costs increase for

Fig. 4. Decode BCH blocks

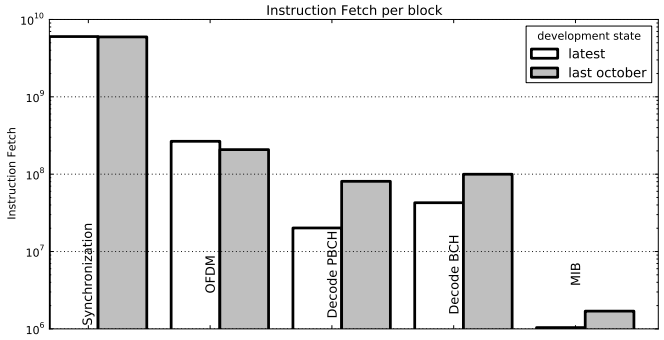


Fig. 5. Comparison: October 2012 vs. April 2013

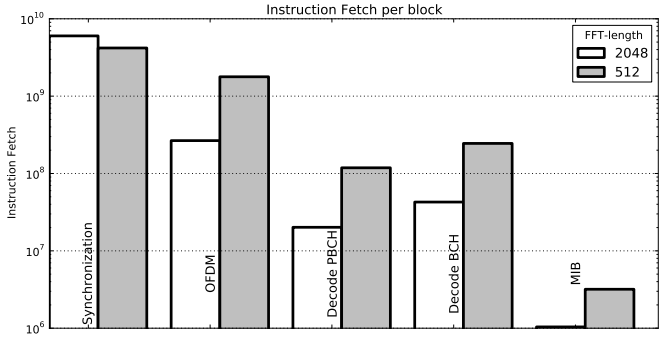


Fig. 6. Comparison: effect of different FFT-lengths

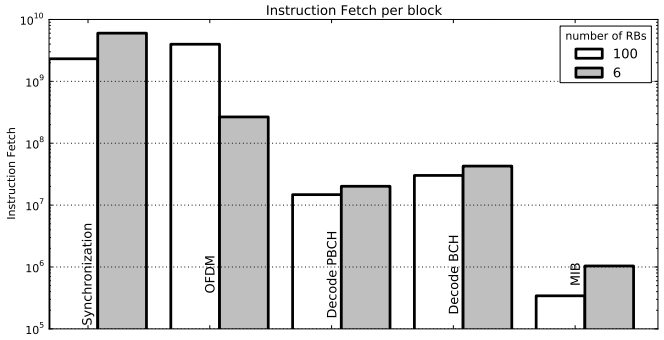


Fig. 7. Comparison: effect of different number of RBs

the OFDM section as it is the section which is configured by the number of RBs. For the synchronization section the relative CPU load increases with fewer RBs processed. However, the absolute CPU load remains the same. The processing costs shift towards the OFDM section for higher numbers of RBs processed. The main reason for that is the *Channel Estimator* block. Since we are only extracting the PBCH no additional CPU load is required in the other sections. If we extracted physical channels transported on a greater number of RBs the CPU load for their processing would increase.

V. Conclusion

In this paper we described a LTE receiver framework implementation in GNU Radio. We have outlined the algorithms necessary for synchronization, channel estimation and equalization as well as decoding arbitrary LTE physical channels. Our implementation uses advanced GNU Radio capabilities such as stream tags and message passing. Due to its modular design more functionality can easily be added by extending the flowgraph. Our future work will concentrate on further optimization, decoding of additional channels and the adoption of a partial event-based processing model which was recently introduced into GNU Radio.

References

[1] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation," 3rd Generation Partnership Project (3GPP), TS 36.211 v10.0.0, Jan. 2011.

[2] J. Mitola, "Software radios-survey, critical evaluation and future directions," *National Telesystems Conference*, May 1992.

[3] K. Manolakis, D. M. G. Estévez, V. Jungnickel, W. Xu, and C. Drewes, "A closed concept for synchronization and cell search in 3gpp lte systems," in *WCNC*, 2009, pp. 616–621.

[4] S. Alamouti, "A simple transmit diversity technique for wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1451 –1458, oct 1998.

[5] gr-lte Repository. (2013, Apr.) https://github.com/kit-cel/gr-lte. accessed 11th April 2013.

[6] 3GPP, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding," 3rd Generation Partnership Project (3GPP), TS 36.212 v10.0.0, Jan. 2011.

[7] Valgrind Website. (2013, Apr.) valgrind.org. accessed 10th April 2013.

CONTEXT-AWARE COGNITIVE RADIO[1]

James Neel (Cognitive Radio Technologies, LLC, Lynchburg, VA, USA; james.neel@crtwireless.com); Pete Cook (Peter G. Cook Consultancy); Ihsan Akbar (Harris Corporation); Daniel Devasirvatham (Wi-Plan Wireless Consulting); Charles Sheehe (NASA); Neal Mellen (Wireless Spectrum Management, LLC); and Bob Schutz (Artisan Wireless Solutions)

ABSTRACT

To make effective decisions, a cognitive radio needs to understand its operating context. Over the past several months, the Wireless Innovation Forum's Cognitive Radio Work Group has been exploring how to enable a cognitive radio to represent, understand, and share its context. Material to be covered in this paper includes the following:

- What exactly is meant by "context" in varying published existing context-aware applications
- The role of context in communications and information systems
- A survey of tools and software for developing context-aware applications
- A new model of the interactions of the real world, symbolic reasoning and representation, and acting on the reasoning
- Relating the components of a key context-aware tool to the new model
- Initial work coding a java-based context-aware application for a cognitive radio to reason and act on its context

This paper provides greater detail and context to a presentation given at WinnComm 2013 and reviews work performed on the subject since the earlier presentation.

1. INTRODUCTION

The CRWG's study into the development of context-aware cognitive radio (CR) came together from a few different threads of discussion. How to manage and reason over the vast amount of information that will be soon available to CR applications is a problem we call "Big RF", which we think is made more usable by better understanding of the context of collected data. How to ensure effective and efficient communications across a wireless link was the subject of an earlier study where we saw that an understanding of the originator's, recipient's and the message's context was critical. Finally, the emergence of CR applications that require information beyond what is available as metrics from an RF chipset implies that a CR be able to understand at least its operational and communications context. This section discusses those three motivating insights in more detail to facilitate a better understanding of the envisioned applications of this work.

1.1. Big Data and Big RF

In the IT world, IBM notes that [1]: "Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone."

This data is produced by a variety of sources - users posting to Facebook, records of banking transactions, cell phone GPS data for E-911, video and imagery. This data in theory provides the raw information needed to gain critical insights about customers but is complicated to extract because of the dimensions of:

Volume the amount of data, e.g., Intel's factories generate 5 Terabytes / hour [2] and 12 Terabytes of Tweets are created daily [3]
Velocity the rate of data acquisition, e.g., high frequency trading
Variety the range of data types and sources, e.g., text, HTML, pictures, sound, etc.
Veracity the accuracy of the data, e.g., weather forecasts are only valid for a certain period of time but also include an element of confidence

Because traditional relational database algorithms have proven insufficient for the task, the IT world is turning to "Big Data" to address this problem. Big Data is a loosely defined term that refers to both: a) a collection of emerging techniques and processes for rapidly acquiring, classifying, and synthesizing meaning from Terabytes or Petabytes of data and b) the data itself. Numerous tools have been developed for Big Data analysis with many being open

[1] This Document is Not Export Controlled. This Information is approved for publishing per the ITAR as "Fundamental Research" and the EAR as "Educational Information".

source, such as Hadoop for batch analysis, Storm for real-time analysis, Alteryx for Analytic, and Drill for interactive analysis.

Similarly, the wireless community is beginning to develop and deploy massive databases for collecting real-time data about the RF environment. This includes ambitious projects such as the nationwide Spectrum Dashboard [4] which hopes to maintain a view of spectrum usage for the entire nation, the TV White Space and other Dynamic Spectrum Access (DSA)-enabling databases, and the real-time data collection, monitoring, and management of commercial, public safety, and military wireless networks. The advent of cognitive radio (CR) introduces both potential consumers of this data (to gain better insights into how to manage radio resources in real-time) and potential sources of more data (from the spectrum sensing capabilities proposed for most CRs).

Because of the volume, velocity, variety, and veracity of the data in these envisioned RF databases, we believe that, like with Big Data, traditional relational database techniques will be insufficient for the purpose of gaining meaningful insights about the RF environment in a way that can be used by CRs, regulators, and network managers. For example a real-time nationwide Spectrum Dashboard could be faced with the following problems.

Velocity A single spectrum logger, such as D-TA's RFVision 2 logs data at the rate of19.2 Gbps [5]. To maintain a nationwide network of such sensors mounted at each cell tower in the US (260,000 towers) [6] would generate approximately 5 Petabits of data per second.

Volume To analyze trends over a single year without loss of data, this would then require 15.7 Zetabits of storage

Variety More realistically, spectrum measurements would come from a variety of sources, such as cell phones, base stations, and access points, which would have differing data formats and often duplicate observations of the same phenomena though with seemingly disparate measurements.

Veracity An important aspect of many communication links and networks is verification of the identity of the radio to determine the validity of the data being transmitted.

To refer to this problem space, the Cognitive Radio Work Group (CRWG) in the Wireless Innovation Forum (WinnF) has adopted the term "**Big RF"** to reflect the similarities between the IT and RF problem domains. Big RF conditions apply across all network domains from relatively small local networks to larger nationwide deployments. Big Data resources, techniques and concepts could be applied and extended to RF data analysis problems once we account for the unique properties of radios (e.g., data converter and sensor limitations, component nonlinearities) and the radio environment (e.g., propagation effects, lossy channels, interference, and bandwidth constraints).

In many Big Data problems, and likely with emerging Big RF problems, it is often necessary to remove the context from, or at a minimum anonymize, the data in order to provide for data security and privacy. It is envisioned that new tools will need to be developed in order to properly frame metadata the data to have useful information for both the Cognitive System and for the user of the information. In this paper, we address the special needs of Big RF and Cognitive Systems in the identification and application of context to data and provide additional thoughts on the tools and techniques that will be required to meet these challenges.

1.2. Context in Communications

In an earlier project [7], the CRWG explored how actionable communications occur and identified that shared context between the originator and recipient of a message was critical to both understanding the communication (effectiveness) and minimizing the bandwidth required for communications (efficiency). These concepts were encapsulated in the Information System Flow Model shown in Figure 1, which represents a single functional cycle of an information system.

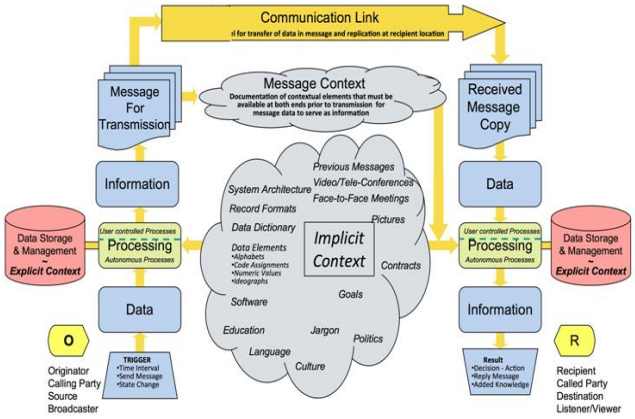


Figure 1. Information System Flow Model. From Figure 9 in [7].

The model has three basic parts: Origination (O), Communication, and Reception (R). An event at the Originator triggers the cycle and creates data, often called a transaction. After O-space processing, that data becomes useable information. A message for transmission is extracted from the information and delivered to the Recipient over a communication link. Upon receipt, R-space processing restores transmitted data to the state of useful information,

available for use by the recipient. Completion of a cycle often serves as a triggering event for subsequent cycle chains in the system. To understand the transmission, O and R must have a shared message context. That context can include formatting (syntax), enabling the received bit stream to serve as data, as well as a means to properly understand the meaning of the data (semantics, e.g., if "apple" refers to the company, the fruit, or the record label). The figure also shows the existence of implicit context, a variety of rarely considered but necessary conditions such as a common language and shared goals that enable cooperative action between O and R. [7] and [8] showed that this model can be applied to a wide variety of communications systems, including Air Traffic Control, Amazon transactions, TV broadcasts, and public safety dispatch.

[7] and [8] also presented security as an inverse function wherein denying a (presumably unwanted) recipient the necessary context hinders reception of the message. For instance, denying access to the communications format of the data (e.g., encryption) or hiding the meaning of the data (e.g., communicating in code words) both are means of obscuring the context of data transmitted in a message.

1.3. Context, Cognitive Radio, and Intelligent Decisions

While subsequent sections will consider a more formal definition of *context*, for a cognitive radio (CR), consider context to be all relevant information about the CR's operation. Thus a CR's context includes the metrics that measure the CR's communications performance – the so-called "meters" of a CR [9] – such as BER, jitter, end-to-end delay, and signal-to-noise ratio. But relevant information may go beyond waveform stack metrics, such as CR location, information from or about other radios in the area, mission objectives, user identity, and the specific application(s) being supported.

Thus for the purpose of this paper, a *context-aware cognitive radio* is a CR that is also aware of information about its operation beyond the metrics provided by the waveform stack. By virtue of being a cognitive ***radio***, this additional information is then used to control (adjust the knobs of) radio behavior. Such a scenario is illustrated in Figure 2, though it should be noted that the same concept could be extended to a cognitive network or cognitive system depending on what hardware is controlled by the intelligent software.

While most CR publications only make use of waveform chipset statistics (e.g., [9]), several publications have taken the broader view advanced in this paper of incorporating additional operational information into the CR decision process.

CR-One proposed in [10] incorporates information about the user's intentions to help guide the radio's adaptations. The WinnF Public Safety Special Interest Group's report on applying CR to a chemical plant disaster scenario [11] proposed automatic role-based reconfiguration wherein CRs would automatically recognize the role being fulfilled by the users and the current scenario being faced by the users to dynamically adjust device priorities, manage device profiles, and define talk-groups. [12] considers the role of CR in a disaster response wherein the extent of the damage is unknown, and replacement transmitters may unwittingly interfere with transmitters that survived the disaster while formal coordination of replacement through FCC processes is impractical. In this scenario, frequency allocations, power levels, operating waveforms, and device configurations are all dynamically managed based on information gathered about the type of scenario, specific information gleaned on the presence, location, and identity of systems from sensing and databases, and from interactions with human users (an example of the context concept known as *mediation*). Finally, the rules for US TV White Space operation [13], and geographically-constrained CRs in general, provide a salient example of context-aware CR wherein information beyond what is available to the chipset (specifically, location and emitter information managed in an online database) determines the frequencies and power levels available for communication.

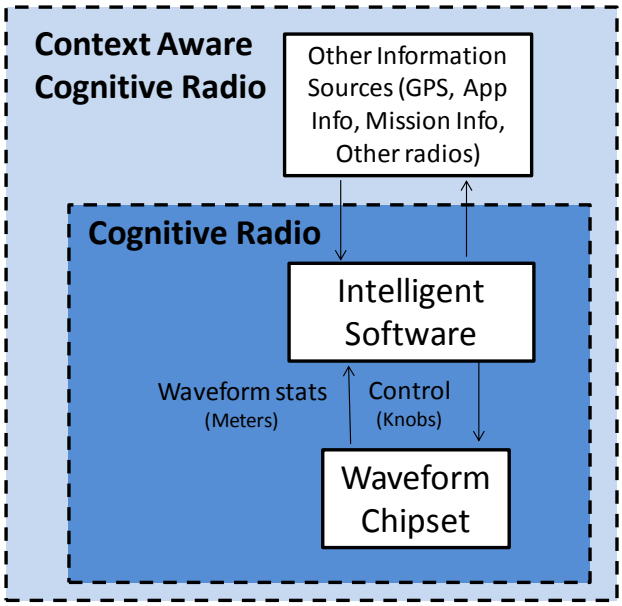


Figure 2: A Context-Aware Cognitive Radio incorporates sources of information beyond what is available to a baseband chipset.

Thus to allow CRs to better process the expected vast wealth and variety of data that will soon be available (Big RF), to make communications more effective and efficient (IPA Communications Cycle), and to enable a myriad of potential CR applications, CRs should be aware of and reason over their context. In the following sections, this

paper considers what is meant by context in communications and wireless applications, proposes a generalization of how an intelligent system interacts and reasons about the real-world, presents an initial context-aware CR application, and makes recommendations for developing context-aware CR applications.

2. CONTEXT IN RELATED LITERATURE

To better understand the role of context, the CRWG surveyed over 100 papers on developing context-aware applications, models, tools, and languages. This section presents a selection of these papers to highlight varying perspectives and subtle differences on the meaning of context and related terms, specific context-aware applications, and tools and languages developed to support these context-aware applications.

2.1 Context and Related Terms

Dictionary.com defines context as "the parts of [communication] that precede and follow a word and passage and contribute to its full meaning" and as "the conditions and circumstances that are relevant to an event, fact, etc." [14] defines context as "any information that can be used to characterize the situation of an entity." Other authors have different definitions, but all share the common thread of additional information about some "thing" that aids in the understanding of that "thing". Depending on what it is being applied to, the content of context can vary greatly.

For instance, message context considered in Section 1.2, provides additional information about a transmitted message and may be critical to understanding the message. As discussed in [7], this context can be explicitly conveyed as part of the message, as in the first definition from Dictionary.com, or may be understood to be shared implicitly between sender and receiver based on past communications or a shared understanding of the world (consistent with all three definitions). Context in the Big RF implies the use of meta-data (and meta-meta-data…) to better understand and process the data. The applications considered in Section 1.3 imply a need for operational context (e.g., location, what the user is trying to do, who is nearby) and communications context (e.g., the past, current, and expected future state of the link or network). Each of these types of context is important for a context-aware CR application.

Beyond the term context, the following are additional terms used in context-aware applications.

Ambiguous – information that cannot be resolved by a mediation process, information having two or more possible meanings. [15]

Confidence – degree to which the subject and object are valid. [16]

Mediation - the dialogue between the user and computer that resolves questions about how the user's input should be interpreted [15]

Situation - external semantic interpretations of context objects having properties and standing in relations to one another [16]

2.2 Context Aware Applications

The following is a small selection of context-aware applications that have been developed.

- A wearable device that automatically provides the user with information about the paper and presenter at a conference, e.g., if a recording is being made, based on the room the user is in, the time, and the conference schedule. [17]
- Nokia Situations, an app that adjusts the operation of your phone (ringer settings, launching other applications, changing voicemail, auto-replying to emails) based on location, and calendar information. [18]
- Word prediction and auto-completion for severely disabled users based on location, previous words, and other indicators [15]
- Context-aware light source from Disney that changes light based on the scene [19]

A much longer list will be provided in a context-survey paper being prepared by the CRWG.

While many different context-aware applications have been created that modify the behavior of an application, there are very few publications that consider using context information to modify the behavior of radio hardware. The one application of context-aware CR that is commonly discussed is policy controlled DSA radios, which is effectively a context-aware CR app that controls transmit characteristics based on geo-fencing (database approach) and direct measurements of the RF environment (sensing approach).

2.3 Context Models and Tools

The following is a short selection of available models and tools for developing context-aware applications.

- Qualcomm's Gimbal [20] uses location, time of day, web history, apps and app usage to infer information about the user and provides a series of libraries (e.g., geo-fencing) to build and deploy context-aware applications for iOS and Android platforms.
- Really Simple Context Middleware [21] is open source software for developing context-aware applications for Android platforms. It provides libraries for sensing battery levels and location and for reasoning about user activity (e.g., inferring if the user is walking).

- Context Modeling Language (CML) [22] is a graphical-based modeling language for analyzing and specifying requirements for a context-aware application. It provides varying context classes for context sources, can represent imperfect information, and can model dependencies, histories, and constraints for context fact types.
- Ontological models of context-aware applications that are principally driven by reasoning over spatial and geographic considerations. [23]
- Context Broker Architecture, an OWL-DL based middleware for modeling context-aware applications and facilitating communications between components. [24]
- A hybrid model that mixes both fact-based context capabilities with ontologies and reasoning to model, design, and implement context-aware applications. [25]

As tools already exist to develop context-aware applications for smartphones, the principal challenge for developing context-aware CR will be to translate this into context-aware applications that also control radio hardware.

2.4 Context ToolKit

Developed initially by researchers at Georgia Tech [26], the Context Toolkit is an open-source collection of software modules and processes for modeling and developing context-aware mobile applications in Java and XML. The code for the Context ToolKit can be downloaded from http://contexttoolkit.googlecode.com/ and documentation is available at www.contexttoolkit.org. The principle components of the Context Toolkit are the following:

Widgets are all purpose class that can store information, or context and the base class for other components

Sensors are widgets that turns real world data into contextual information where one or more attributes are normally accessible to other widgets

Enactors perform reasoning on inputs (attributes from other widgets) to create outputs

Generators update the sources of information used by the sensors and models real world processes when no real world component exists and is typically implemented as an enactor without an input

Services are the actuators of the model and are responsible for performing actions in the real world based on computations in the context reasoning model.

To support the reuse of contextual information and reasoning across applications, to better enable distributed applications, and to support dynamically constructed systems, the Context Toolkit provides a *Discovery System*. The Discovery System maintains a repository of running components, the addresses and ports used to communicate with components, descriptions of components, and mechanisms for subscribing to the computations of components. The Context Toolkit has also been extended to provide a means for mediation whereby the user or other context-aware applications can be queried for clarification when the context may be unclear. [14]

A particularly attractive feature of the Context Toolkit is built-in support for controlling hardware via its Services / Actuators components. Thus, constructing context-aware CR applications may only require extending the Context Toolkit to support Services that specifically control radio hardware. This is examined further in Section **Error! Reference source not found.**.

3. DEVELOPING A GENERALIZED MODEL OF CONTEXT-AWARE INFORMATION SYSTEMS

Building on the context survey and prior work in the IPA efforts, the CRWG has been exploring how to appropriately model context-aware information systems. This has led to development of a new model for context awareness, the Wireless Information System Descriptive Model (WISDM), which, though intended for wireless applications, is general enough to encompass all of the context-aware information system models surveyed by the CRWG to date. This section briefly overviews WISDM and how other context-aware reasoning models fit into the WISDM framework.

3.1 WISDM

People have communicated with each other for thousands of years, initially verbally, then with written symbols. But a dramatic disruptive change has occurred in the past one hundred years. Information, represented by binary digits, can be manipulated and stored using electronic computers, and communicated electronically at the speed of light, often without wires. Wireless information systems make information instantly available anywhere in the world, are accumulating vast amounts of stored data, and provide new insights into all aspects of society.

Recognizing these changes and building upon the insights and generalizations gleaned from the Information System Flow Model of Figure 1, the CRWG is developing a general model of how contextually aware information systems operate, which is called the Wireless Information System Descriptive Model (WISDM). WISDM, shown in Figure 3 is a high-level abstraction of information systems that has the objectives of describing the common characteristics that disparate systems have to facilitate interaction between them, designing better information systems, and suggesting new ways they can aid civilization. The concept of "systems" includes both physical systems and their information footprint in the digital world.

The WISDM space, as shown in Figure 3 is divided into four quadrants. The horizontal divider separates the real world (physical space embodied with atoms) from digital information space, i.e., the conceptual world, with binary digits (bits) as the fundamental unit. The vertical divider has

systems in operation on the left and development and enhancement of those systems on the right.

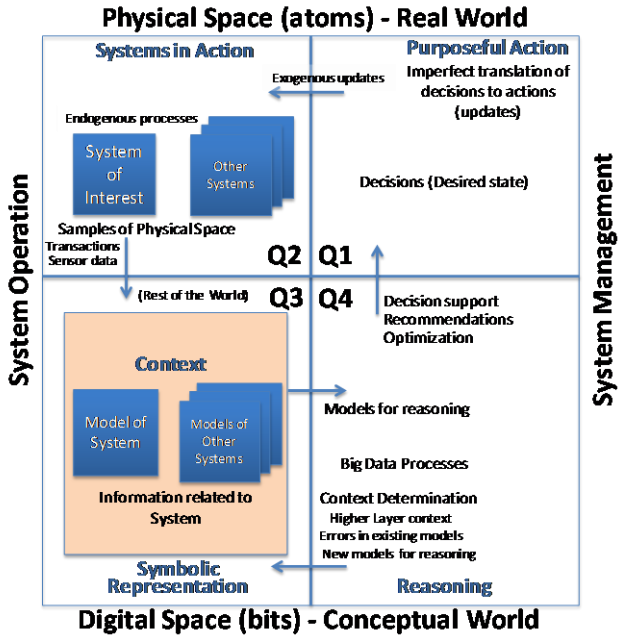


Figure 3: Wireless Information System Descriptive Model (WISDM)

Each of the quadrants has its own perspective. The upper left quadrant, Q2, is Systems in Action, the common real world in which we live. The lower left, Q3, is the system's information model of the real world, with all of the inherent limitations of models[2]. The lower right quadrant, Q4, represents the processes applied by the system to the models and accumulated data of Q3 to gain insights about the real world (context determination) and to alter operation of the system (decisions). The upper right, Q1, imperfectly translates these decisions into actions taken by the system in the real world, and represents the creation of new things, processes, and systems, and modification of how systems work.

The basic flow within WISDM is generally counter-clockwise. Q3 information gathered about the Q2 real world (context acquisition, e.g., sensor data) is used in Q4 to update the models and their context (relationships between models and additional information about the system). These models are then reasoned over, decisions are arrived at, and then imperfectly translated into real world actions, which may or may not lead to the desired result.

However, the flow is not always counter-clockwise, as Q4 reasoning about the models can lead to further insights about the models themselves and delivery of revisions directly to Q3. We call this resulting context about the context "higher-layer context" and it is used to discover additional relationships and improve decision support. This is conceptually similar to the learning modes supported by back-propagation analytics developed for neural-net architectures in the 1990s.

These concepts will be more fully fleshed out in a subsequent paper and report to be released by the CRWG.

3.2 Relationships Between Context Models

While working on WISDM and surveying existing models of context-aware systems, the CRWG had the insight that the models were generally isomorphic to one another. In other words, one context model could be used to represent or implement another. To illustrate this, consider Figure 4 and Figure 5, which show how WISDM represents the Context Toolkit components and the CR OODA (Observe-Orient-Decide-Act) loop model.

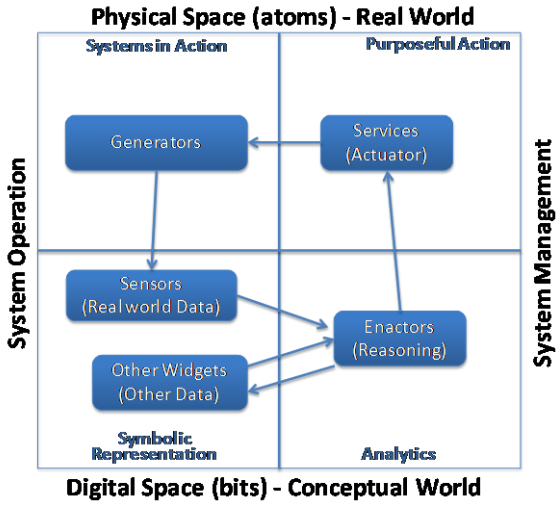


Figure 4: Context Toolkit components in WISDM

In Figure 4, the Context Toolkit assumes the existence of a "real world" (i.e., systems in action in Q1) whose processes of interest can be modeled with generators. The current states of systems in action are sampled by sensors which are stored in widgets (Q2) that provide attributes that can be used for contextual analysis (Q3). The Enactors (Q3) implement the reasoning processes which can then lead to the use of Services (Q4) that yield changes to the real world (e.g., increasing light levels). Alternately Enactors can derive new contextual information that updates widgets that store this new context information in Widgets that correspond to Q2.

In Figure 5, the Outside World of OODA corresponds directly to Q1 and the sampling of the Outside World leads to a low level awareness in Q2. The Orienting and Learning

[2] "All models are wrong, but some are useful," George E. P. Box, *Empirical Model-Building and Response Surfaces* (1987), co-authored with Norman R. Draper, p. 424.

processes of OODA span Q2 and Q3 of WISDM to reason to new contextual awareness and new models based on learned insights and detected errors. The results of Deciding and Planning of OODA in Q3 of WISDM lead to the Acting process in Q4, which is then realized in Q1 (Outside World).

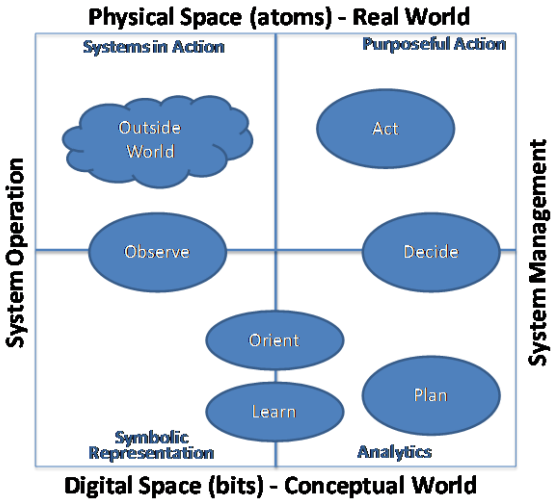


Figure 5: Mapping the CR OODA model [10] into WISDM

In both cases, the principle flow between components in the Context Toolkit components and the CR OODA loop match the counter-clockwise flow of WISDM. Further, it is relatively easy to see how human reasoning similarly maps into WISDM with mental (or physical) models of the system of interest substituting for WISDM's symbolic (digital) models. Because of the wide applicability of WISDM to context-aware systems and context-aware models and tools, WISDM appears to be a valuable tool for context-aware CR.

4. DEVELOPING A CONTEXT-AWARE CR APPLICATION

To better understand the utility of existing context tools and models for developing context-aware CR applications, to help refine the CRWG's study into context aware systems, and to identify potential gaps in existing capabilities for context-aware CR, the CRWG is developing a context-aware CR application using Context Toolkit components, illustrated in Figure 6. This example uses information commonly incorporated into other context-aware applications to control radio hardware – selectively enabling transmission and controlling the network to communicate over. This is a key differentiator between traditional context-aware applications and context-aware CR.

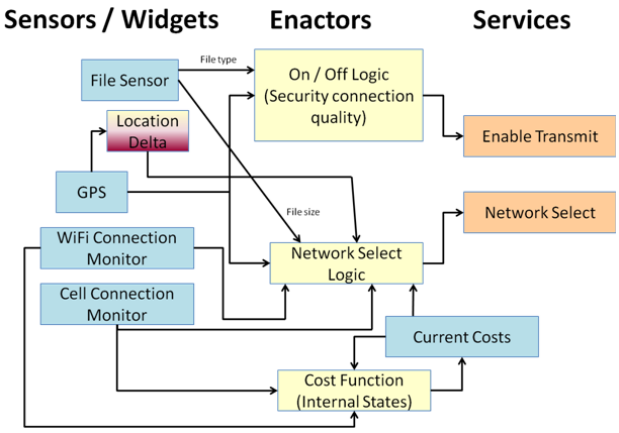


Figure 6: Block Diagram of Context-aware Cognitive Radio Application using Context Toolkit components. Blue blocks indicate simple sensors, yellow blocks enactors, orange blocks services (actuators), and the mixed color block an example of a widget providing a derived context attribute (which necessitates derivation from an enactor).

In this application the functions of primary concern are the policy considerations for utilizing the link, the physical availability of the network, the technical quality of the link, and lastly the economics of using the link. The application determines if the radio should be allowed to transmit or not and over which network (cellular or WiFi), the control of which is implemented in two different Services modules (Enable Transmit and Network Select). Enactors, which manage the reasoning logic in the Context Toolkit, are driven by the input from the following Sensors:

- the presence and achievable speed of a WiFi connection
- the presence and achievable speed of a cellular connection
- location via GPS
- movement rate (derived from GPS)
- file type and size to be transmitted.

We are currently implementing these objects in Java, whose widespread adoption can facilitate the incorporation of components originally intended for other uses. For example code from British Telecom Labs widget for interfacing with a GPS device was repurposed for use in this application with little if any modification. The flexibility of

the Context Toolkit also allows for implementation via XML descriptors via the WidgetXML function that allows for the creation of a Widget from an XML schema file. This potentially could be the basis for a Context Factory – a mechanism being explored in the CRWG for automatically managing the loading and unloading of context applications and the sharing of components and information across context-aware CR applications.

From a CR perspective, the weakness of the Context Toolkit is that it takes over where the communication link ends, i.e., it was not designed with radio applications in mind. For instance, the Context Toolkit puts the data provided by a communication link in a useable format, but it may not provide the necessary meta-data to establish a communication link. Establishing these links may be necessary for an application to operate, e.g., if it depends on a wirelessly connected sensor. However, it should be possible to create lower level Sensors, Widgets, Enactors and Services to fill this gap and to work largely within the existing framework.

5. CONCLUSIONS AND FUTURE WORK

As CR applications become more sophisticated and incorporate information from an expanding and more varied set of sources, being able to quickly process and extract meaning from the sea of data will be increasingly important. Because this is analogous to the Big Data problems faced in the IT world, we call this set of problems "Big RF". By reasoning to higher layers of context from the raw sensor data, we believe that CRs can better handle Big RF problems.

Understanding what context is and how it is shared between the originator and recipient of a message allows for more efficient communications by permitting significantly fewer bits to be transmitted to convey the same resulting information. Similarly, by concealing message context from undesired recipients, message security can be enhanced. These were concepts first examined in the previous CRWG IPA projects, and are serving as a springboard into the CRWG's further studies into enabling context-aware CR and information systems.

CR will benefit from understanding its context – message context of transmitted data, operational context of the system, and communications context of its managed link or network. While the role of context for CR has been briefly examined by previous authors, e.g., [10], [11], [12], we believe this is the first paper to systematically consider how context can be incorporated into CR processes.

As part of this process, the CRWG first conducted a broad survey of existing literature on developing context-aware applications, and subsequently focused on mobile applications. However, among the different authors, there were varying perspectives and subtle differences on the meaning of context and on which attributes were most critical to determining a device's context. We saw where tools and languages had been developed to support these context-aware applications, and we identified the Context Toolkit as a promising starting point for developing context-aware CR applications.

As part of this effort, the CRWG developed the Wireless Information System Descriptive Model (WISDM). WISDM models context-aware information systems and divides basic operations into four quadrants – (Q1) Systems in Action, (Q2) Symbolic Representation, (Q3) Analytics, and (Q4) Purposeful Action where Q1 and Q4 capture operations in the Physical Space (Real World) while Q2 and Q3 represent operations in Digital Space (Conceptual World) and where Q1 and Q2 describe system operation and Q3 and Q4 describe system management. Within and between the quadrants are various processes for updating models, e.g., from sensors, from reasoning, or from modeled endogenous processes. By comparing WISDM to the surveyed models of context-aware systems, we showed that WISDM applies to a wide variety of context-aware information processing systems, including humans!

To better understand the utility of existing context tools and models for developing context-aware CR applications, to help refine the CRWG's study into context-aware systems, and to identify potential gaps in existing capabilities for context-aware CR, the CRWG is developing a context-aware CR application using Context Toolkit components. The application determines if the radio should be allowed to transmit or not and over which network based on context sensed data from location, network availability, and intended file transmission type.

In the future, the CRWG will continue to explore and refine the concepts outlined in this paper. Planned activities include developing a Big RF tool to automatically manage changing context applications via a mechanism we tentatively call the Context Factory. In general, much of the work will focus on identifying what would constitute a comprehensive toolset for managing CR context with an emphasis on identifying which tools are still in need of further development and what kinds of applications could be deployed today.

6. REFERENCES

[1] http://www-01.ibm.com/software/data/bigdata/

[2] J. Bertolucci, "Intel Cuts Manufacturing Costs With Big Data",*Information* Week, March 18, 2013. Available online: http://www.informationweek.com/software/business-intelligence/intel-cuts-manufacturing-costs-with-big/240150978

[3] E. Naone, "What Twitter Learns from All Those Tweets," *MIT Technology Review*, September 28, 2010. Available online: http://www.technologyreview.com/view/420968/what-twitter-learns-from-all-those-tweets/

[4] FCC Spectrum Dashboard, http://reboot.fcc.gov/spectrumdashboard/
[5] D-TA, "Ultra-Wideband Scanning Transceiver with Signal Activity Detection, Real-Time Recording, IF Playback & Data Analysis Capabilities," White Paper, July 2012. Available online: http://www.d-ta.com/verification.php?site=Applications/ultra-wideband-scanning-transceiver-July.pdf
[6] http://wiki.answers.com/Q/How_many_cell_towers_are_in_the_US
[7] Wireless Innovation Forum, "IPA – Information Process Architecture Volume I," WINNF-09-P-0020-V1.0.0, Nov 01, 2010.
[8] Wireless Innovation Forum, "Information Process Architecture Volume 2: Survey of IPA Like Systems," WINNF-09-P-0021, Jun 27 2012.
[9] C. Bostian, J. Reed, "Understanding the Issues in Software Defined Cognitive Radio," Tutorial presented at *IEEE DySPAN*, Baltimore, MD, Nov. 2005.
[10] J. Mitola, "Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio," PhD Dissertation, Royal Institute of Technology (KTH), May, 2000.
[11] Wireless Innovation Forum, "Use Cases for Cognitive Applications in Public Safety Communications Systems Volume 2: Chemical Plant Explosion Scenario," WINNF-09-P-0015-V1.0.0, Feb 11, 2010.
[12] D. Devasirvatham, J. Neel, C. Tompsett and K. Link, "3 Layers of Disaster Recovery," *MissionCritical Communications*, March 2013, pp. 62-65.
[13] FCC, "Second Memorandum Opinion and Order In the Matter of Unlicensed Operation in the TV Broadcast Bands and Additional Spectrum for Unlicensed Devices Below 900 MHz and in the 3 GHz Band," FCC 10-174, Sep 23, 2010.
[14] A. Dey, J. Mankoff, "Designing Mediation for Context-Aware Applications," *ACM Transactions* 2005.
[15] A. Dey, J. Mankoff, G. Abowd, and S. Carter, "Distributed Mediation of Ambiguous Context in Aware Environments," *Proc. 15th Ann. Symp. User Interface Software and Technology (UIST '02),* pp. 121-130, Oct. 2002.
[16] J. Ye, S. McKeever, L. Coyle, S. Neely and S. Dobson "Resolving Uncertainty in Context Integration and Abstraction," *ICPS'08*, July 2008 pp. 131-140.
[17] A. Dey, M. Futakawa, D. Salber and G. Abowd, "The Conference Assistant: Combining Context-Awareness with Wearable Computing," *Proceedings of the 3rd International Symposium on Wearable Computers (ISWC '99)*, San Francisco, CA, Oct 1999. pp. 21-28.
[18] http://www.readwriteweb.com/archives/nokias_new_situations_app_makes_phones_self-aware.php
[19] O. Wang et al., "A Context-Aware Light Source," ICCP 2010. Available online: http://zurich.disneyresearch.com/~owang/pub/context.html
[20] https://developer.qualcomm.com/mobile-development/mobile-technologies/context-aware-gimbal
[21] http://code.google.com/p/rscm/
[22] K. Henricksen, J. Indulska, A. Rakotonirainy, "Modeling context information in pervasive computing systems," 1st International Conference on Pervasive Computing (Pervasive), vol. 2414 of Lecture Notes in Computer Science, Springer, 2002.
[23] A. Frank, "Tiers of ontology and consistency constraints in geographical information systems," *International Journal of Geographical Information Science,* 15 (7), 2001, pp. 667–678.
[24] H. Chen, T. Finin, A. Joshi, "Semantic Web in the Context Broker Architecture,"*Proceedings of the Second IEEE International Conference on Pervasive Computing and Communications*, 2004.
[25] K. Henricksen, S. Livingstone, J. Indulska, "Towards a Hybrid Approach to Context Modelling, Reasoning and Interoperation," *Proceedings of the First International Workshop on Advanced Context Modeling*, *Reasoning and Management*, 2004.
[26] A. Dey, "Providing Architectural Support for Building Context-Aware Applications," PhD thesis, Georgia Tech, Dec. 2000.

RECONFIGURABLE NATO IV RF FRONT-END FOR SDR TERMINALS

Javier Baltasar (Indra Sistemas, Madrid, Spain; jbaltasar@indra.es)
Antonio Morales (Indra Sistemas, Madrid, Spain; ammendez@indra.es)
José M. Camas (Indra Sistemas, Madrid, Spain; jmcamas@indra.es)
Carlos Rodriguez (GRADIANT, Vigo, Spain; cralemparte@gradiant.org)
Florian Palade (GRADIANT, Vigo, Spain; fpalade@gradiant.org)

ABSTRACT

The paper describes the development process of a state of the art 2x2 MIMO NATO IV (4.4 GHz – 5 GHz) RF Front End.

First, requirements over the design are introduced, where it is important to emphasize that a low phase noise, fast synthesizer tuning time, high spurious rejection and fast AGC solution is pursued in the whole frequency band to allow the compatibility of the Front End with the most innovative wideband waveforms.

Second, the paper will present the architecture selected, that is based on COTS SISO WiMAX chipsets but synchronized among them to achieve a MIMO solution. The output frequency of these chipset will be moved into the desired NATO IV RF band by external RF circuitry. Digital control of the architecture is based on reprogrammable devices (FPGA) what allows complete flexibility and reduces the impact for its integration with any digital base band from hardware as well as from software standpoint.

Last but not least, an overview of implementation key challenges is provided: critical component selection based on validation by simulation and mock-ups implementation, key PCB implementation issues and the different microwave technologies utilized.

1. INTRODUCTION

The work described in this paper gathers and details all the development stages of a military NATO IV enabled RF Front-End, from technical specification to implementation, integration and trials.

The RF Front-End presented is intended to be integrated in a military vehicle for being used in tactical land deployments. To achieve this, considerations as easing the integration with different digital base band units and adaptability to support different waveforms have been taken into account. Software and hardware architectural details of the RF Front-End are provided in section 3 of this paper. First integration trials have been done considering a WiMAX (IEEE 802.16e compliant) digital base band unit.

Requirements coming from legacy and new state of the art waveforms (narrow and wide band) have been considered for the technical specification of the RF Front-End. The most challenging technical requirements and associated design decisions are outlined in section 2 of this paper.

Targeted frequency band is 4.4 GHz to 5.0 GHz, so the integrated equipment will be military NATO IV enabled. There is clear intention of migrating tactical radio systems to this frequency band in the near future. Emerging SDR platforms are expected to facilitate the introduction of cost-effective system in this band, allowing the support and provision of advanced services in LOS/NLOS Point-to-Point (PtP), Point-to-Multipoint (PtMP) and access deployments.

The RF Front-End has been designed to be integrated in a ruggedized enclosure adapted for vehicular usage and considering the fulfillment of EMI/EMC and environmental constraints (MIL STD 461 and MIL-STD-810). Both aspects have had impact in the implementation of the RF Front-End, such as layout (size, mechanical, gerbers..), components and PCB technology selection. All these mechanical and implementation considerations, whose most relevant aspects are introduced in section 0 of this paper, ensure the readiness of the implementation for field deployment.

The work presented in this paper has been performed in the scope the EUREKA project Broadpro [1] and has been co-funded by Spanish National Authority CDTI (*Centro para el Desarrollo Tecnológico e Industrial*) and Indra Sistemas S.A, with the Technological Centre GRADIANT acting as Indra's subcontractor.

2. REQUIREMENTS AND DESIGN DECISIONS

As introduced in section 1, main drivers of technical specification are to ease the integration of the RF Front-End with any SDR digital subsystem and to provide support to legacy and also to the newest and most advanced waveforms.

From a hardware standpoint these goals are translated into a very broad and stringent RF specification and an entirely flexible RF Front-End control approach. This flexibility is obtained through a FPGA-based proxy that implements the control and monitoring functions of the RF Front-End and an interface with the Digital Subsystem based on a high density connector with all the pins routed to the FPGA, providing full flexibility-

Software architecture of this FPGA component is based on a microblaze soft-core, embedded in the FPGA, integrated with specific verilog cores for time-constrained functionalities, such as AGC control or carrier frequency switching.

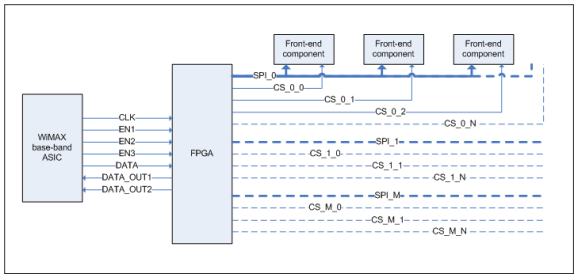


Figure 1. - RF-Front-End Control FPGA Proxy

More details of the hardware and software architecture of the RF Front-End are provided in section 3 of this paper.

As it has been **WiMAX** the first target waveform to be used and integrated with the RF Front-End, the proposed solution is based in the MAX2838 chipset [2]. This MAXIM transceiver is presenting a reconfigurable **RF channelization from 1.5 MHz to 28 MHz**, thus allowing the integration with other waveforms. The output frequency band of the MAXIM transceiver is from 3.3 GHz to 3.9 GHz so a hardware frequency converter module has been used to move the 600 MHz working bandwidth to **4.4 GHz to 5GHz frequency band.**

MIMO functionality is required. For this purpose, two transmission / reception branches, based on MAX2838 and working synchronously, are considered.

For testing the performances of the MAX2838, and confirm its selection, a specific MAX2838 mock-up has been implemented.

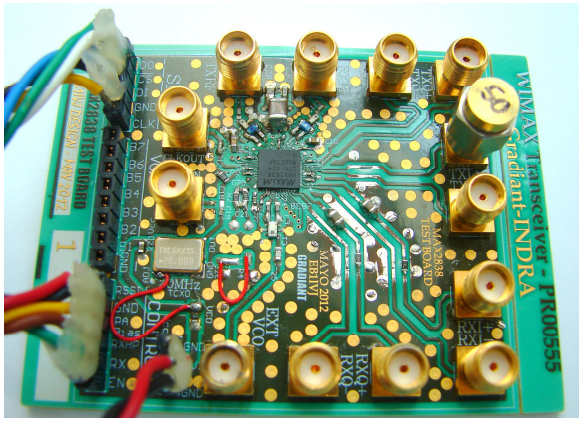


Figure 2. - MAX2838 mock-up

This MAX2838 test board has been used to analyze the compliancy with the requirements that rely on it, as frequency range, ripple, transmitted power, power control, harmonics and phase noise; and also for performing link tests, with modulated signals (DVB-T) , in transmission and reception.

A **transmission power from -20 dBm up to +24 dBm** is achievable in current design, with the possibility of **adjusting this power in 1 dB step**. This granularity, together with the RF control implementation based on FPGA, allows the integration of efficient waveform power control algorithms.

Transmitter output power is limited by the power consumption of the transmission power amplifier that is linked to the power provided to the RF Front-End (by the Digital Base Band). Current integration, with WiMAX baseband digital unit, limits the output power to +24 dBm but this figure could be incremented if more power is provided. Current integration with WiMAX standard is also considering a **back-off of 6dB** (suitable for multi-carrier modulations) that could be reduced in case of other kind of modulation is used, improving output power vs consumption performances. For testing the performances of the power amplifier selected (CREE CMPA2560025F, [3]), a specific mock-up (Figure 3) has been implemented.

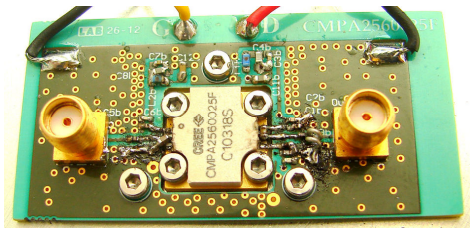


Figure 3. - Power Amplifier mock-up

Quite stringent linearity requirements, for transmission and reception, have been defined, specifying a **transmitter OIP3 of greater than + 35 dBm and a receiver IIP3 greater than + 25dBm**. Besides, high spectral purity is also required, with **spurious and harmonics rejection levels between 60 dB and 80 dB and a noise figure of less than 7 dB**. These requirements' fulfillment relies on the performance of the MAX2838 and the OL and mixer of the external hardware stage that makes the translation to NATO IV frequency band.

Down-conversion and Up-conversion mixers are based on Mini-Circuits Ultra Low Noise MMIC amplifier PMA-5453+ [4] and SIM-U712H mixers.

Specific mock-ups have been implemented for the up-conversion and down-conversion mixers (Figure 4), as well as for the local oscillator for assessing individually and in integration with the MAX2838 mock-up (Figure 2) the performances and confirming its selection.

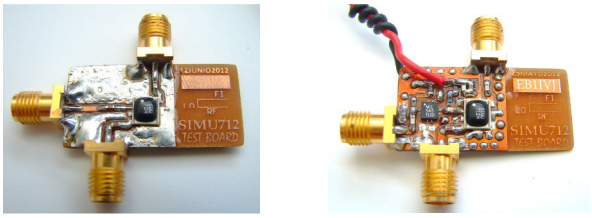


Figure 4. - Up-converter and down-converter mixers mock-ups

In order to support Fast Frequency Hopping waveforms, a carrier **frequency tuning time of less than 80 µs** is specified while keeping high performances in phase noise performance (-114 dBc @ 1 Mhz offset from carrier frequency), together with a **carrier stability equal or less than 1 ppm**. These requirements are affected by the MAX2838 and the external hardware to move to NATO IV frequency band.

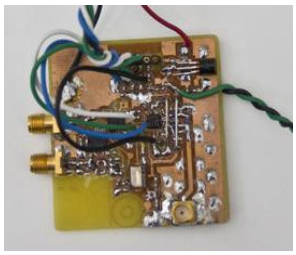


Figure 5. - Local Oscillator mock-up

The local oscillator is based on the Analog Device ADF4360 VCO [5] and a 20 MHz TCXO and a specific mock-up (Figure 5) has been implemented to assess its performances and confirm its selection. The TCXO selected ensures the fulfillment of the carrier stability requirement.

In order to allow the support to different waveforms (narrow and wide band waveform) a **broad reception dynamic range** has been specified (**94 dB**) that will be handled by a FPGA-based AGC.

The implementation of the AGC control unit in the FPGA embedded in the RF Front-End allows, on one hand, full reconfigurability of AGC dynamics for its adaptation to specific waveforms (amplitude modulation, phase modulations, multi-carrier schemes,...) and, on the other hand, the achievement of the **AGC attack time** specified (**< 36 µs**).

The stringent specification found for the AGC derives in a design decision of using logarithmic detectors that present, in front of RMS detectors, higher dynamic ranges and much better response time. Power detectors selected are Analog Device AD318. For AGC algorithm calculation, the MAX2838 RSSI detectors are also used.

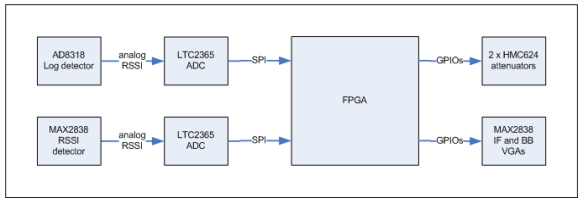


Figure 6. - AGC Components (for one branch)

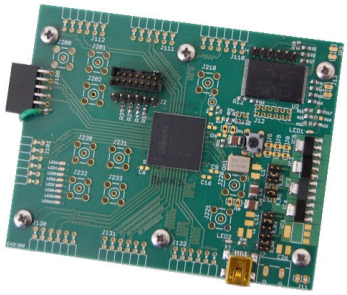


Figure 7. - FPGA mock-up

For initial validation of the firmware (microblaze and verliog cores components), a specific FPGA mock-up has been implemented that has been used in integration with the rest of the analog hardware mock-ups for architecture and design decision validation.

3. RF FRONT-END ARCHITECTURE

From previous section in which HW and SW requirements, and derived design decision, for the MIMO NATO Band IV RF Front End are detailed, present section is focused on the RF Front-End Architecture description.

Present section is structured in two main sub-sections: HW architecture of the RF Front-End and SW/FW Architecture as an enabler of the RF Front-End full reconfigurability.

A. HARDWARE ARCHITECTURE

In Figure 8 it is presented a high level diagram of the RF Architecture defined for the MIMO NATO Band IV RF Front-End.

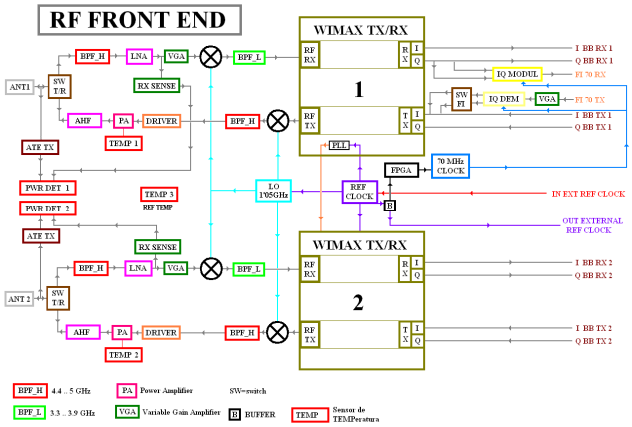


Figure 8. - HW Architecture

RF Architecture design is based on the usage of commercial WiMAX direct conversion transceivers typically used in WiMAX certified equipments. Specifically, RF architecture selected considers the usage of two MAX2838 transceiver chipsets working synchronously, due to the MIMO requested architecture, which work in the 3.3 to 3.9 GHz commercial band for WiMAX communications.

Frequency band requirement for the MIMO RF Front End is to operate in NATO Band IV (from 4.4. to 5 GHz), therefore is requested to add a frequency conversion stage to move the original 600 MHz chipsets bandwidth to the desired frequency band.

MIMO architecture requirement implies the necessity to duplicate TX and RX chains, besides a critical time control of both branches.

First, HW modules making up the two down-conversion branches of the RF Front End are described. RF Front End operating in RX performs a down-conversion from NATO IV band to 3.3 – 3.9 GHz frequency band to subsequently feed MAX2838 chipset. Hereinafter. HW modules that go though the received signal path are briefly described
:

- **Unidirectional High Band Pass Filter** for global selectivity of 600 MHz in NATO IV band (from 4.4 to 5.0 GHz). Due to the reduced space available for filters, the need to provide MIMO support, the conversion architecture selected and stringent RF performance, an ad-hoc manufacture of the filters of this RF Front End has been needed.
- **RF Switch** for selecting the correct branch (TX or RX).
- **Low Noise Amplifier** to obtain a relevant gain factor and keep controlled noise factor parameter.
- **Power Detector module**, which is in charge to detect the in band power signal. These estimations are used as inputs for correct operation of the Automatic Gain Control (AGC).
- **Variable Gain Amplifiers,** which are in charge of selecting the appropriate gain or attenuation in order to feed mixer with the proper signal level (important to achieve optimum inter-modulation figures and fulfill requirements). These amplifiers are controlled by the FW architecture.
- **Mixer**, which is in charge of the down-conversion process. 1050 MHz reference local oscillator, properly synchronized in both MIMO branches, is used for the mixing process.
- **Unidirectional Low Band Pass Filter**, which is in charge of selecting the 3.3 to 3.9 GHz native frequency band for the WiMAX transceiver chipset.

Regarding TX operation, TX branch is responsible for moving native frequency band of WiMAX Transceiver (from 3.3 GHz to 3.9 GHz) to NATO IV frequency band. Both up-conversion branches are made up of the following RF modules:

- **Mixer**, which is in charge to mix RF signal from WIMAX transceiver and 1050 MHz local oscillator to move operative frequency band to Nato Band IV. There is a mixer module in each TX branch.
- **Unidirectional High Band Pass Filter** applied to the mixer output previously to any amplification stage.
- **Pre-Amplification stage (driver amplification)**, which is in charge to apply the required gain factor to drive correctly the power amplifier.

- **Power Amplifier,** providing the amplification stage in charge of obtaining the maximum output level according requirements.
- **RF Switch** for selecting the TX or the RX branch.
- **Unidirectional Anti-harmonic Filter,** with low insertion losses
- **Power Detector module** feed by the transmission signal after the fixed required attenuation to not affect power detector and avoiding any kind of reflection problem.

Apart from TX and RX branches, RF architecture also contains the following modules:

- **Reference Clock circuit** which is in charge to manage the reference clock. This reference clock can be provided form an external source and it is distributed to all RF components using it (WiMAX transceiver chipsets, local oscillator and FPGA). In case no external reference is present at input, TXCO provides its internal reference clock, fulfilling stability requirements. External reference has always priority against internal reference.
- **Local Oscillator**, which is in charge to generate the 1.1. GHz tone input for the up-mixers and down-mixers. Therefore, output tone needs to be amplified and split for the 4 mixers.
- **Temperature sensors** included in the RF architecture to control operation temperature of the two power amplifiers and in specific and critical.

B. FIRWARE/SOFTWARE ARCHITECTURE

Apart from the Hardware Architecture, FW/SW architecture allows fully reconfigurability of the RF Front End, easing RF Front End integration with different SDR Platforms.

Firmware/Software architecture is based on the usage of a FPGA device providing a registers map accessible, using a standard SPI bus, by the SDR platform to configure, control and monitor all RF Front End parameters.

Figure 1 depicts a high level diagram of the HW/SW Architecture designed for the MIMO NATO IV RF Front-End.

Selected FPGA FW/SW architecture is based on a MicloBlaze soft-core CPU processor which is in charge to manage a SPI bus for communication between external SDR platform and MicroBlaze CPU.

Additionally, as can be seen in Figure 1, other SPI buses are managed by the MicroBlaze processor to support configuration and control of the RF Front-End:

- **RF Subsystem General Control** (with the digital base band): activation of the RF Front End, deactivation of specific branch (for MIMO or SISO operation mode), etc.
- **Local Oscillators** control.
- **Transmission power control.**
- **VGAs** control and monitoring supporting AGC capability.
- **ADCs** for power detection and analog signal provision to SDR platform.
- **AGC parameterization.**
- **Temperature sensors** control.
- **Clock Reference** Control.

SW/FW architecture is implemented on FPGA device XA Spartan-6 Family, which support and extended temperature range compared with typical FPGA from Spartan-6 family. This extended range feature has been needed to allow the system to fulfill MIL-STD standards.

4. IMPLEMENTATION DETAILS

RF Front-End has been designed for its integration on a ruggedized enclosure, together with the digital base band hardware, ready to be installed in a military vehicle

Figure 9 shows the mechanical design of this ruggedized enclosure and the space reserved for the RF Board and the Digital Baseband Board. .

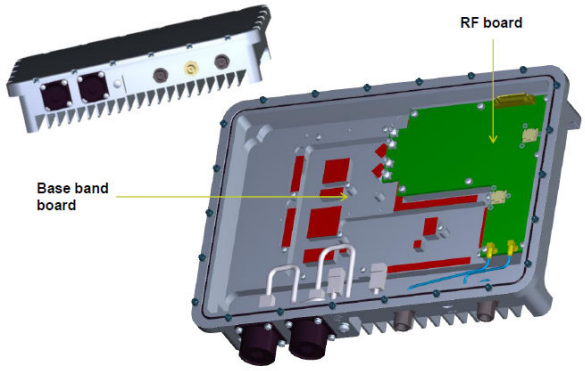


Figure 9. - Ruggedized Enclosure

The space limitations have impacted in components selection that has derived in the need of custom implementation of the high-band filters (4.4 GHz – 5 GHz) and in the decision of reusing components for transmission

and reception branches, feasible thanks to the half-duplex characteristics of the transceiver.

The system, integrated in the enclosure, must fulfill MIL-STD environmental and EMI/EMC requirements, presenting an operational temperature range from -40ºC to +65 ºC.

Specific power dissipation, vibration, shock and stress analysis have been performed over the system.

Figure 10 shows some details of the outcomes obtained in the thermal analysis performed.

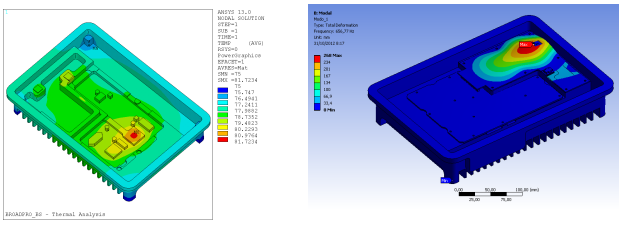


Figure 10. - Thermal and structural analysis (some results)

Compliancy with MIL-STD standards has been obtained constraining the placing of most critical components as, for example, FPGA and Power Amplifiers and implementing in the enclosure mechanics specific heat dissipation towers for the critical components.

RF Front-End FPGA has been selected with extended temperature range and in contact with a specific dissipation tower. This ensures the compliancy with MIL-STD standards, even in the case of control FW evolution for its adaptation to other waveforms.

Due to the hard restrictions in this project, a special 6-layer PCB was designed. It uses a symmetrical stack-up, using microwave low loss RO4350 cores at the top and bottom sides, and a standard central FR4 core, being this glued by intermediate prepegs to the RF cores. Only one core is used for RF devices, but symmetrical stack-up avoid PCB bending due to differential physical properties against temperature, as in a bimetallic thermostat.

Double side components placement is needed, being impossible the design if only one side is used.Via hole managed levels are 3:

- A full through via hole level, connecting the 6 layers, used to interconnect digital and analog parts and for general grounding and shielding.
- A 1-2 level, reserved for RF devices on the top layer
- A 3-6 level, used mainly for bottom layer digital devices, but too in striplines.

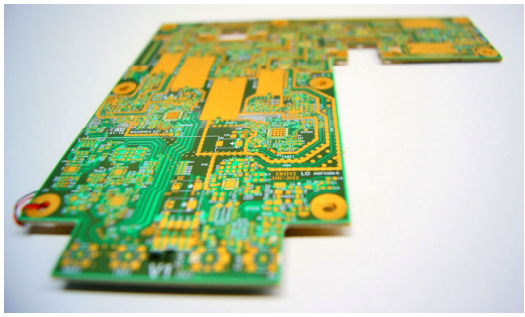


Figure 11. - NATO IV RF Front-End PCB

All digital noisy devices, as FPGA, charge pump, ADCs, high voltage switched power supply for the power amplifier, etc., are in the bottom side, while analog RF signals are mainly in the top side, with ground plane keeping away the noisy part from the analog parts.

Special care is needed when defining via holes levels, because each of this level requires a manufacturer re-metallization in the copper layer, for metallic deposition in the via-hole inner cylinder. Each of this re-metallization adds some tens of um in the copper thickness, and we need to take in account this when we are using 10 mils RF substrate cores. Copper is thicker now, and this modifies RF tracks impedance.

150um tracks are used to manage high density routing for FPGA pin-out. Via-hole re-metallization process limits available resolution, making a little difficult the routing near FPGA. Differential lines were used in the base band signal to avoid noise. Microstrip lines in the external cores are used as transmission lines, using stripline in the internal layers to provide RF crossings and in order to keep highly shielded some critical signals as the 4/3 RF carrier used for MAX2838 synchronization in MIMO operation. This signal is not a big problem in the MAX2838 native frequency band (3300-3900MHz) but go across all the target NATO IV frequency band (4400-5000MHz), so it can be a excellent auto-jammer if care is not taken. StripLine provides a big shielding, 90 dB or more, while microstrip at 5GHz microwave in our design can only warrants values around 60 to 65 dB for shielding between lines. Stripline higher performance is coming from the 2 ground plane around the active line, quite similar to a coaxial structure.

It can be summarized this design as a hybrid digital-analog prototype, 6 level multilayer PCB and with both sides components, being really a not trivial task requiring enough engineering skills, managing old and classic concepts and others quite new, mixing RF and FPGA world in a real prototype

Figure 12 shows final implantation of the NATO IV RF Front-End.

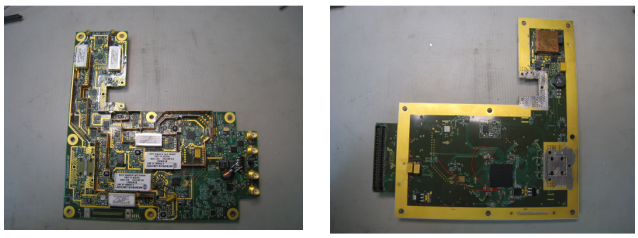


Figure 12. - NATO IV RF Front End (final implementation)

5. CONCLUSIONS

In this paper, the challenges found when developing a fully reconfigurable, multi-mode, SDR-ready and NATO IV capable RF Front End has been detailed.

Architectural design decisions, component selection process and implementation constraints has been described in front of the stringent set of requirements that have to be fulfilled. The presented RF Front-End is intended to be integrated with the most innovative state-of-the-art waveforms.

Feasibility has been demonstrated through its implementation and validation in a vehicular oriented form factor, integrated with a WiMAX standard based digital Base Band unit.

6. REFERENCES

[1] EUREKA E!5767 Broadpro (Broadband Technologies for Professional Applications) Project (http://www.eurekanetwork.org/project/-/id/5767)

[2] MAX2838 3.3 GHz – 3.9 GHz Wireless Broadband RF Transceiver (http://www.maximintegrated.com/datasheet/index.mvp/id/5611)

[3] CMPA2560025F, 25W 2500 – 6000 MHz, GaN MMIC Power Amplifier (http://www.cree.com/~/media/Files/Cree/RF/Data%20Sheets/CMPA2560025F.pdf)

[4] Ultra Low Noise MMIC Amplifier, PMA-5453+ (http://217.34.103.131/pages/npa/PMA-5453+_NPA.pdf)

[5] Integrated Synthesizer and VCO, ADF4360 (http://www.analog.com/static/imported-files/data_sheets/ADF4360-7.pdf)

OPTIMIZATION OF SQUELCH PARAMETERS FOR EFFICIENT RESOURCE ALLOCATION IN SOFTWARE DEFINED RADIOS

Nesrine Damak, Christoph Krall, Rainer Storn
(ROHDE & SCHWARZ GmbH & Co. KG
Radiocommunications Systems Division
81671 Muenchen, Germany, Email: rainer.storn@rohde-schwarz.com)

ABSTRACT

The squelch function is an important element in almost every radio, especially the airborne radio. It suppresses the audio output of the radio receiver when the desired signal does not have sufficient SNR and/or signal strength.
In legacy radios the squelch characteristic has usually evolved through many iterations stemming from customer feedback as well as extensive lab and field tests, so that the user finally experiences the most convenient squelch behaviour. If the computation of the squelch algorithm has to be changed, for example due to efficiency reasons, it must be ensured that the carefully obtained squelch characteristic remains the same, preferably without being forced to redo all iterations mentioned above. This paper describes an optimization-based approach which allows to meet these requirements.

1. INTRODUCTION

The squelch function of a radio connects the desired signal to the audio output of the radio receiver when the desired signal exhibits the required characteristics such as sufficient SNR and/or signal strength. It suppresses the audio output in case these characteristics are not met. Hence the proper function of the squelch is imperative for the communication between the pilot and the air traffic controller. Since the squelch controls the incoming signal, it is implemented in the receiver of a radio. The main components for digital signal processing in the receiver module of a radio are a Digital-Down-Converter (DDC), a Field-Programmable-Gate-Array (FPGA), the Digital Signal Processor (DSP), and a flash memory. We are mainly interested in the DSP where the absolute value of the in-band signal and its differential phase are used to compute the squelch criteria for muting the AF output as illustrated in Figure 1.1 and Figure 1.2.
Note that the receiver module includes two receiving blocks: a main receiver (the primary block for voice and data communication) and a guard receiver (for emergency calls), which operate concurrently. Both main and guard receivers

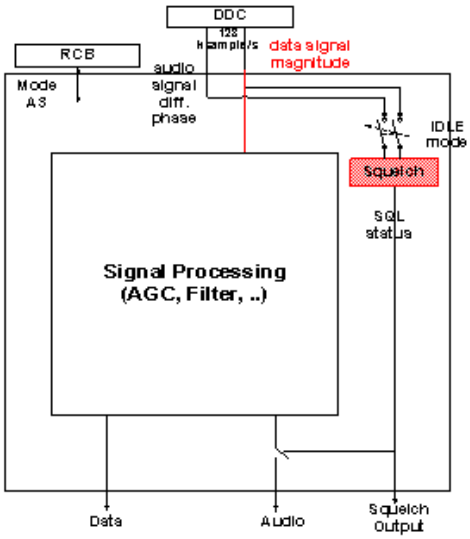


Figure 1.1: The squelch in the receiver module of the airborne radio (A3E Mode).

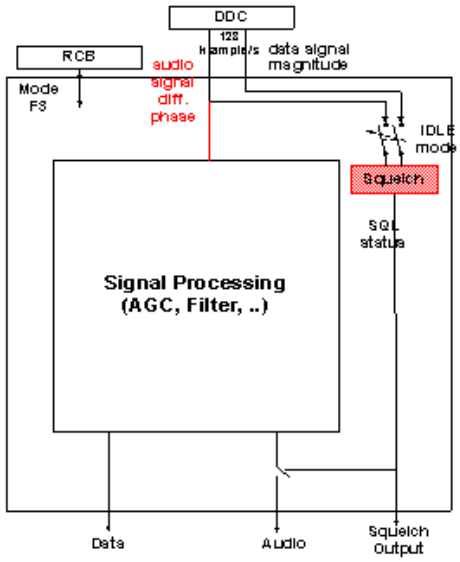


Figure 1.2: The squelch in the receiver module of the airborne radio (F3E Mode).

have a main block for signal processing and a separate block for squelch calculation. In this work, we only focus on the main receiver. The information-bearing input of the signal processing block is, in case of frequency modulation, the differential phase of the signal, and the absolute value of the in-band signal in case of amplitude modulation (compare Figure 1.1 and Figure 1.2). While the output of the squelch block controls the audio output of the receiver, the data output is not under squelch control.

2. A CLOSER LOOK AT THE SQUELCH

There are many state-of-the-art squelch techniques, and Figure 2.1 shows two of them: the S/N squelch and the carrier squelch. Both of them are illustrated for the F3E-mode [1]. We will investigate the S/N squelch more deeply because the carrier squelch has a much simpler implementation and is less challenging in terms of optimization.

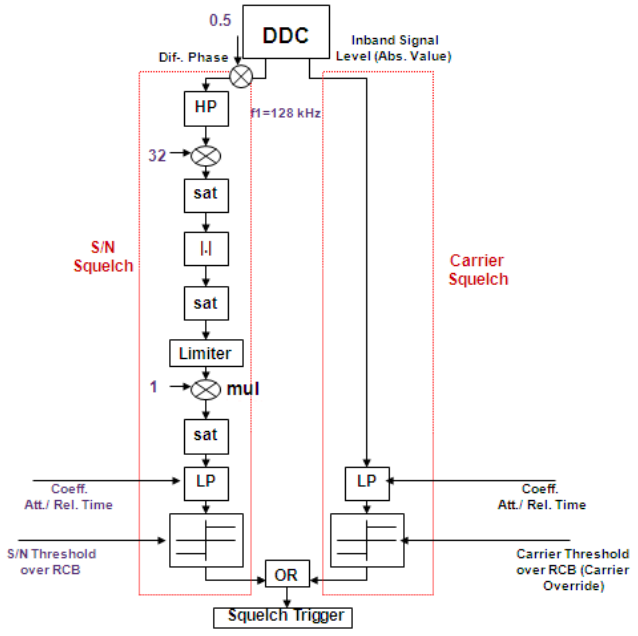


Figure 2.1: Fixed point implementation of the squelch in F3E Mode.

The S/N squelch detects valid radio signals which have a certain signal-to-noise ratio. Before that, the received RF signal is sampled and demodulated in the receiver. The demodulated signal is then fed into the squelch. In this work, the squelch receives the signal from the digital down converter (DDC), which converts the signal to a complex basebanded signal centered at zero frequency and represented in polar coordinates (absolute value and differential phase). The squelch gating of received signals, represented in polar coordinates, is a method that was introduced and patented in [2]. In this approach the noise is estimated and compared to thresholds which are determined by the characterization of the front-end to get meaningful signal-to-noise ratio estimates.

To this end the differential phase from the DDC is considered as a basis of the S/N squelch decision where it is assumed that an increase in noise power leads to an increase in the differential phase. This can be shown as follows:

Suppose we receive a 1 kHz noiseless tonal sound signal, which can be written as

$$y(t) = e^{j\omega t} \qquad (2.1)$$

The amplitude of the signal is constant, while the phase is a linear function $\Phi = \omega t$. Hence, the differential phase is a constant $d\Phi/dt = \omega$. Let's suppose that a white Gaussian noise is added to the tonal sound. This random process also propagates into the differential phase of the signal which henceforth exhibits increased phase noise. The phase is no longer linear, and consequently the differential phase is not a constant any more. With increased noise level, also the variance of the phase noise is increased and therefore the energy of the differential phase. Therefore, the high-pass filtered differential phase can be considered as a measure for the noise level. In [3], the monotonic relation between noise in the complex baseband and in polar coordinates was shown.

3. MODES OF THE S/N SQUELCH

An airborne radio operates in different modes: fixed frequency modes, TRANSEC modes or other special modes. Each mode has a different payload routing for the signal transmission and reception. In this work, we are only interested in the signal reception for the fixed frequency modes. They can be classified into two categories: amplitude-modulated AM and frequency-modulated FM modes, depending on the modulation technique used to convey the signal.

The signal processing chain of the squelch is similar for all fixed frequency modes (cf. Fig. 2.1): the differential phase goes through a high-pass, so that the remaining noise can be estimated afterwards. However, the channel filter differs from mode to mode, allowing either narrow-band or wide-band signals to pass through. For instance, the amplitude-modulated AM A3E mode has a channel filter with bandwidth equal to 25000 Hz while the AM A1D mode has 50000 Hz , allowing both audio and data signals to be processed.

For each mode, the S/N squelch has a specified threshold function: the squelch estimates the signal-to-noise ratio of the received signal. If the measured S/N value is above the desired S/N ratio, the received signal is treated as a

meaningful signal and is applied to the pilot's headphones or data modem.

In order to be able to perform an efficient development of the squelch algorithm a simulation in MATLAB® is available. When feeding a simulated differential phase into both the actual DSP implementation of the target radio MR6000A [4] and the simulation we obtain the result shown in Figures 3.1 and 3.2 verifying the validity of the simulation.

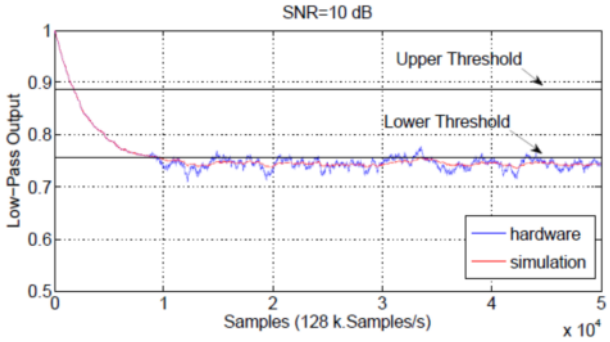


Figure 3.1: SNR=10dB in AM mode A3E.

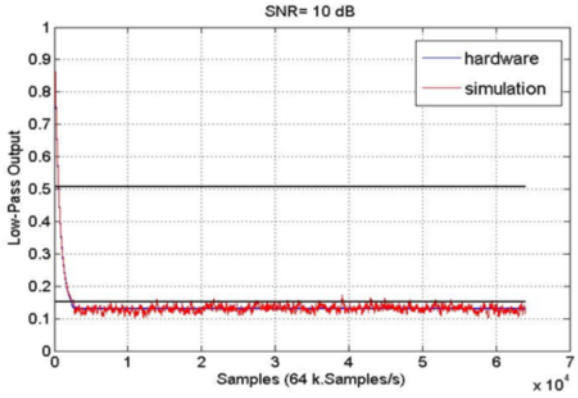


Figure 3.2: SNR=10dB in FM mode F3E.

4. SQUELCH ALGORITHM OPTIMIZATION

In this chapter, we focus on the task of preserving the S/N-squelch behavior while reducing the processing load that the S/N-squelch computation imposes upon the DSP. The latter may be necessary in order to free up DSP cycles for further functionality that needs to be added to the radio. The former, i.e. the preservation of the squelch algorithm behavior, is important since it has been optimized by means of exhaustive measurements and customer feedback. For instance, we consider the following situation: the squelch operates in A3E mode on signals with a sample rate of fa = 128 kHz. Consequently, all IIR filters (HP,LP) also have a clock rate of 128 kHz. In order to reduce the computational complexity, we want to perform down sampling on the LP filter. After a down sampling of factor 4, the low-pass filter will have a clock rate of 32 kHz, i.e., only every fourth sample will be considered by the LP filter. Hence, the output of the LP filter will deviate from the original one. With the appropriate optimization algorithm, we optimize the release- and attack- time coefficients of the low-pass filter so that its output approaches the one with the initial sampling rate (fa = 128 kHz).

The most critical function of the S/N-squelch is to estimate the SNR of the received signal. This is done by estimating the noise power and subsequently inverting it so that the SNR in the complex baseband can be computed [3]. Since the LP output of the squelch is a measure for the noise it is strictly decreasing with increasing SNR. Figure 4.1 confirms the decreasing behavior of the average squelch output as a function of the SNR, and also shows that the output changes considerably for the different lowpass sampling frequencies (initial frequency 128kHz and modified frequency 32kHz).

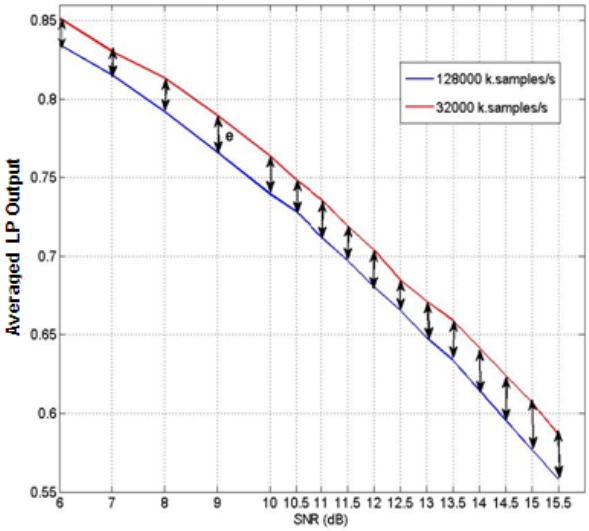


Figure 4.1: Averaged LP output of the squelch before and after downsampling.

The goal is now to alter the LP coefficients for a sampling frequency of 32kHz such that the noise estimation behaves in the same way as for the sampling frequency of 128kHz.

The optimization problem consists of minimizing the squared error between the two curves shown in Figure 4.1, i.e.

$$\min_{a_{lp},\,b_{lp}} (f) = \sum_{i=1}^{16} (e_i)^2 \qquad (4.1)$$

where e_i represents the difference between the outputs at SNR_i, and a_{lp} and b_{lp} are the optimized coefficients of the IIR low-pass filter.

It has to be noted that the LP is not a simple fixed coefficient LP but has two sets of coefficients instead. One set determines the attack time of the squelch, the other set determines the release time. For this reason we have to split the minimization task of Eq. 4.1 into two which is expressed in the two equations Eq. 4.2 und Eq. 4.2. Hence we first minimize

$$\min_{a_{lp,att},b_{lp,att}} (f_{att}) = \sum_{i=1}^{16} (e_i)^2 \quad (4.2)$$

and then we minimize

$$\min_{a_{lp,rel},b_{lp,rel}} (f_{rel}) = \sum_{i=1}^{16} (e_i)^2 \quad (4.3)$$

where $a_{lp,att}$ and $b_{lp,att}$ are the attack time coefficients, i.e., the coefficients of the low-pass which control how quickly the filter modifies the open state of the squelch, and $a_{lp,rel}$ and $b_{lp,rel}$ are the release time coefficients, i.e., those coefficients of the low-pass which determine how quickly the modification comes to an end.
A longer attack time means that a longer period of time is used to lower the noise output, and a longer release time means that a longer time is taken for the noise to increase again in case of an abrupt decrease of the SNR. This allows separate control of the smoothing of the signal depending upon whether it is increasing or decreasing in amplitude.

In the following the Differential Evolution (DE) algorithm is described which is used to solve the minimization problems stated above.

5. DIFFERENTIAL EVOLUTION (DE)

The error functions introduced in Eq. (4.2) and Eq. (4.3) exhibit the following difficulties:

1) The error functions have regions of non-differentiability, because the filter coefficients are quantized.
2) The error functions are potentially multimodal, i.e. there may be more than one minimum, so there is a chance that the optimization gets stuck in a local minimum if the optimization method does not have mechanisms to escape local minima.
3) The optimization has to deal with constraints enforced by the stability criterion required for IIR-filters.

DE belongs to the direct search methods, is very simple to implement and use, possesses global optimization capability, and is able to deal with nonlinear as well as mixed-integer optimization problems [5], [6]. Hence it has been chosen for the minimization of the error functions stated in Eq. (4.2) and Eq. (4.3).
The optimization problem in chapter 4 aims to minimize the squared difference of the LP output before and the LP output after the sample rate reduction.
We already mentioned that the optimizers are the coefficients of the LP. For the first order IIR low-pass filter, these coefficients have the form $a_{lp} = [1;-(1 - c_0); 0]$ and $b_{lp} = [c_0, 0, 0]$, which means that we can optimize over one variable c_0 ex $[0,1[$. The optimization problem has two constraints: First, the IIR LP filter should be stable, i.e., the absolute value of the pole $1 - c_0$ should be in
$[0,1]$. Secondly, the filter parameters are represented as internal program variables, i.e. the values need to be from the interval $[0, 1]$.

Initialization of the population ($g = 0$): $c_{0i,0} = rand(1) * (c_{0U} - c_{0L}) + c_{0L}$, $i = 1, .., N_p$
Initialization of the optimum $c_{0opt} = c_{01,0}$
while f(modified rate, c_{0opt}) $> \epsilon$ **do**
  for every element i of the population **do**
    $r_0 = rand(Np)$
    $r_1 = rand(Np)$
    $r_2 = rand(Np)$
    //Mutation
    $\mathbf{v}_{i,g+1} = c_{0r_0,g} + F(c_{0r_1,g} - c_{0r_2,g})$ $(F = 0.85)$
    //Crossover
    $u_{i,g+1} = \begin{cases} v_{i,g+1}, & \text{if } rand(1) \leq CR \\ c_0(i,g), & \text{otherwise.} \end{cases}$
    //Bounce Back
    if $u_{i,g+1} > c_{0U}$ **then**
      $u_{i,g+1} = c_{0r_1,g} + rand(1)(c_{0U} - c_{0r_1,g})$
    end if
    if $u_{i,g+1} < c_{0L}$ **then**
      $u_{i,g+1} = c_{0r_1,g} + rand(1)(c_{0L} - c_{0r_1,g})$
    end if
    //Selection
    if f(modified rate, v) $< f$(modified rate, $c_0(i)$) **then**
      $c_0(i) = v$
    end if
  end for
  choose the best c_{0opt} from the new generation
end while

Figure 5.1: Pseudo code for DE applied to the squelch optimization problem.

From the two constraints, we conclude that
$0 < c_0 < 1$. Fig. 5.1 shows the pseudo-code of the DE-algorithm applied to the problem stated in this chapter.
The function to be minimized in Figure 5.1 is f(modified_rate, c_0) which is regarded in the SNR-interval [6dB,…,15.5dB].

6. RESULTS FOR AMPLITUDE MODULATION

As discussed in chapter 4, we optimize over attack coefficients and release coefficients separately, i.e., we

minimize the function f(modified_rate, $c_{0attack}$) and f(modified_rate, $c_{0release}$), then we combine the results to obtain the optimal c_0 for both attack and release time. If the DE algorithm of Fig. 5.1 is applied to the A3E and the A1D modes the optimization yields the results depicted in Fig. 6.1 and Fig. 6.2.

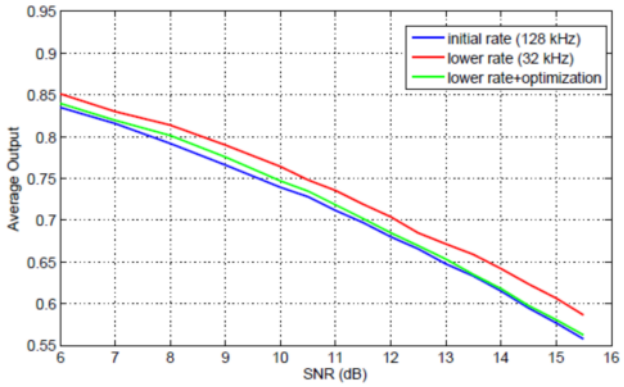


Figure 6.1: Average output of the squelch after 5 iterations for A3E and A1D using a population size Np = 5.

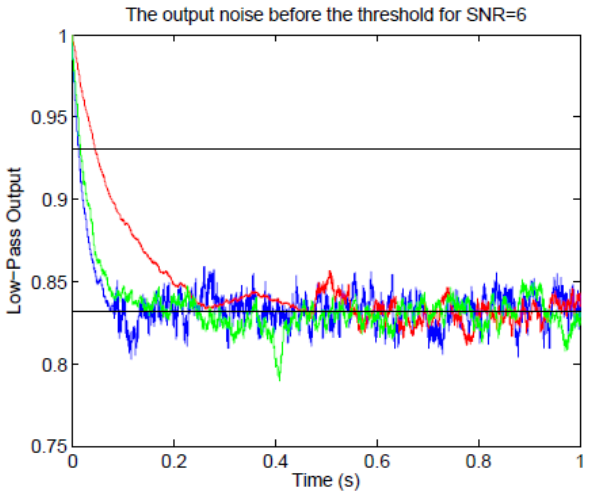


Figure 6.2: Lowpass output for SNR=6dB after 5 iterations of DE. Color coding is the same as in Fig. 6.1.

Obviously DE works very well for the optimization in the amplitude modulated modes. Only five iterations are needed to achieve good results.

7. RESULTS FOR FREQUENCY MODULATION

For the F3E Mode DE fails to find an optimal solution with optimizing only parameter c_0: Figure 7.1 shows the results after 50 iterations. It can be seen that the blue curve to be fitted is not approximated sufficiently.

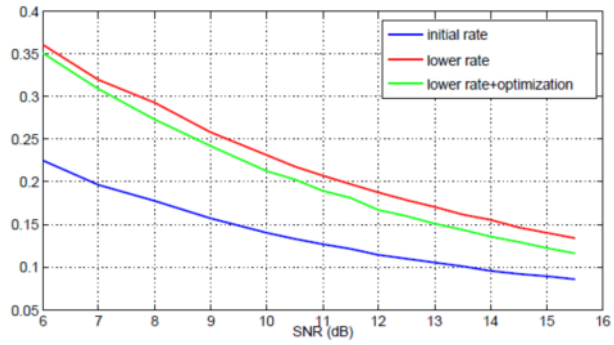


Figure 7.1: Optimization result after 50 iterations in case of one parameter $c_{0attack}$ (attack time coefficient) in F3E mode after minimizing f(rate, $c_{0attack}$).

In order to improve the optimization we increase the number of parameters for the frequency modulated modes: we now optimize with respect to c_0, the limiter bounds lim_{min} and lim_{max} and the multiplier mul (shown in Fig. 2.1). For this optimization we are faced with a mixed integer optimization problem, because the multiplier coefficients are powers of 2, i.e., the multiplier only performs right or left shifting of the input.

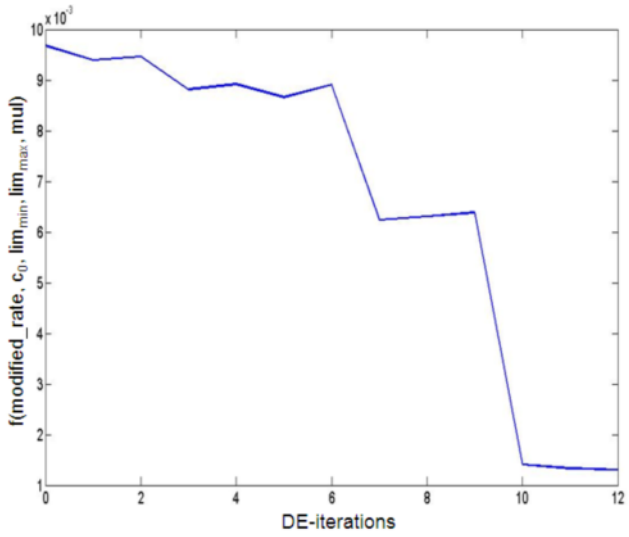


Figure 7.2: Convergence graph of the error function f(modified_rate, c_0, lim_{min}, lim_{max}, mul) while being minimized by DE. DE employed a population size Np = 20.

Since DE does not employ any gradient-based techniques the extension of the optimization to accommodate an error

function f(modified_rate, c_0, lim_{min}, lim_{max}, mul) with four parameters is straightforward, and the results are depicted in Fig. 7.2, Fig. 7.3 and Fig. 7.4. It is evident that the optimization has been much more effective than in the previous attempt where only one parameter c_0 was adjusted.

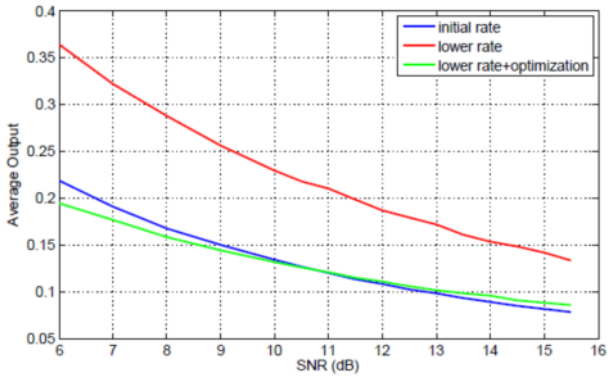


Figure 7.3: Optimization result in F3E mode after minimizing f(modified_rate, c_0, lim_{min}, lim_{max}, mul). The optimization took only 12 iterations of DE using a population size of Np = 20.

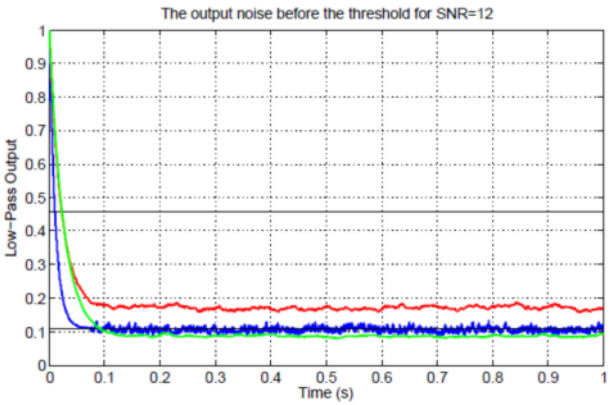


Figure 7.4: Lowpass output for SNR=12dB for F3E after minimizing f(modified_rate, c_0, lim_{min}, lim_{max}, mul). Color coding is the same as in Fig. 7.2.

8. CONCLUSION

The squelch algorithm of a radio is a component of the auditory user interface and therefore its behavior has direct influence on the acceptance of the radio by the user. If, for example for efficiency reasons, the sequelch algorithm needs to be changed it is very important that the behavior perceivable by the user remains the same. In this contribution it has been demonstrated that it is feasible to alter the squelch algorithm of an established implementation with very little effort while still maintaining its behavior. This was possible using an optimization procedure which is able to cope with nonlinear and mixed integer minimization problems which potentially exhibit multiple minima. The optimization algorithm of choice was Differential Evolution (DE).

8. ACKNOWLEDGEMENT

The help of and discussions with R. Arnaoudov are greatfully acknowledged.

9. REFERENCES

[1] *Determination of Necessary Bandwidths Including Examples for their Calculation*, Recommendation ITU-R SM.1138, Geneva, 1995

[2] L. Brueckner and R. Hausdorf. *Method and Apparatus for Squelch Gating a Received Signal*, European Patent EP1843467 2007.

[3] Thomas Gebauer. *Simulation and Implementation of Climax Squelch Functions for ATC Receivers*. Master's thesis, Technical University of Munich, ROHDE & SCHWARZ GmbH & Co. KG, October 2009.

[4] Dolfen, M, Lipp, F., and Storn, R., *R&S®MR6000A: the first choice – and not just for the Airbus A400M transport aircraft*, Rohde & Schwarz MILnews, vol. 13, 2012.

[5] K. V. Price, R. Storn, J. Lampinen, *Differential Evolution – A Practical Approach to Global Optimization*, Springer, Berlin, Heidelberg, New York, 2005.

[6] Storn, R., "Optimization of Wireless Communication Applications using Differential Evolution", *SDR Technical Conference SDR'07*, Denver, Colorado, Nov. 5-9, 2007.

SDR-Based Channel Emulator using a Semi-Stochastic Radio Propagation Model

Johannes Schmitz, Xiang Xu, Florian Schröder, Milan Zivkovic, Rudolf Mathar
Institute for Theorectical Information Technology
RWTH Aachen University
D-52062 Aachen, Germany
Email: {schmitz, xu, schroeder, mathar}@ti.rwth-aachen.de

Index Terms—**SDR, Channel Emulator, GNU Radio**

***Abstract*— We present the design of a low cost Software Defined Radio Based Channel Emulator. The proposed system can be used to emulate channels for link level testing of any of type radio equipment. It comprises of one or multiple RF frontends at the input, respectively output side. Different channel models can be easily exchanged due to the software defined nature of the system. However, in this work we focus on a semi-stochastic, ray-tracing based urban channel model. The model combines advantages of a site specific physical model with a geometry based stochastic model. Being optimized for signal strength prediction and network planning, the ray-tracer is not able to simulate small scale effects like antenna diversity or doppler spread, yet the gap can be closed by the stochastic part of the model.**

I. Introduction

During the last years several freely available Software Defined Radio (SDR) frameworks have enabled researchers to develop a multitude of radio transmission systems and waveforms. However, the same technology can be applied for slightly different tasks or use cases. In this presentation we propose an SDR-based channel emulator using the GNU Radio framework [1]. This device can be used to emulate an RF channel in order to test a wireless transmission link. The system under test (SUT) could be any arbitrary radio equipment. During development a second SDR-based radio system serves as an exemplary SUT. The OFDM transceiver system has been used for a couple of cognitive radio demonstrations [2], [3] and has recently been extended to a 2x2 MIMO transmission mode.

The flexibility of the emulator is characterized by the reconfigurable number of parallel in- and output channels, while the different channel models can be easily emulated due to the software defined nature of the system. A 2x2 configuration of the channel emulator system is shown in Figure 1. Universal Software Radio Peripheral (USRP) devices are used to interface with the SUT. To allow for higher MIMO order channel emulation, the number of USRPs can be increased. The channel model used in the emulator system is based on a ray-tracing simulation of the radio propagation [4] combined with a stochastic model [5]. When running the system with the targeted semi-stochastic channel model, the online signal processing part is a relatively simple filter, which has to be updated with the channel impulse response generated in a preliminary offline phase.

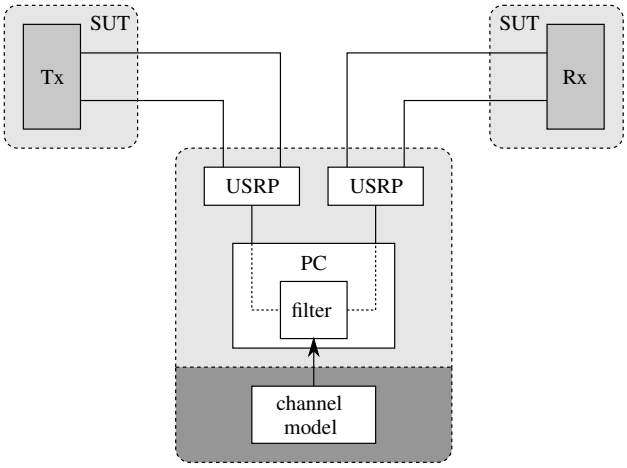


Fig. 1. System Architecture

A challenge lies in the conversion of the ray-tracer output into a suitable input format for the channel emulator. The ray-tracing simulation does not have a discrete sampling grid but rather the different paths arrive with arbitrary distances in the time domain. In order to be able to use the simulation output for the digital signal processing, it has to be remapped to a uniform sampling rate using an appropriate method. As described in [5], to generate the channel matrices $\mathbf{H}(t)$, the stochastic part of the model has to be applied to the simulation output.

References

[1] "GNU Radio," `http://gnuradio.org/`.
[2] Milan Zivkovic, Dominik Auras, and Rudolf Mathar, "OFDM-based Dynamic Spectrum Access," in *IEEE DySPAN 2010*, Singapore, Apr. 2010.
[3] Milan Zivkovic, Johannes Schmitz, and Rudolf Mathar, "Acquisition and Identification of OFDM Signals using Cyclostationary Signatures," in *ACM WiNTECH 2011*, Las Vegas, USA, Sept. 2011, pp. 91–92.
[4] Florian Schröder, Michael Reyer, and Rudolf Mathar, "Field Strength Prediction for Environment Aware MIMO Channel Models," in *6th European Conference on Antennas and Propagation (EUCAP)*, Prague, Czech, Mar. 2012, pp. 1–4.
[5] Xiang Xu, Michael Reyer, Florian Schröder, Alexander Engels, and Rudolf Mathar, "A Semi-Stochastic Radio Propagation Model for Wireless MIMO Channels," in *International Symposium on Wireless Communication 2011*, Aachen, Deutschland, Nov. 2011, pp. 619–623.

A FRAMEWORK AND ARCHITECTURE FOR A COGNITIVE MANAGER BASED ON A COMPUTATIONAL MODEL OF HUMAN EMOTIONAL LEARNING

Urban Bilstrup and Mahboobeh Parsapoor
(Halmstad University, Halmstad, Sweden)
Urban.Bilstrup@hh.se; Mahboobeh.Parsapoor@hh.se

ABSTRACT

In this paper we propose an architecture for a cognitive engine that is based on the emotional learning cycle instead of the traditional cognitive cycle. The cognitive cycle that traditionally has been used as reference for cognitive radio is on the basis of the Unified Theories of Cognition (UTC) to model rational decision making in humans. UTC represents a rational goal-oriented decision-action made by an intelligent agent. However, the emotional cycle represents an emotional reaction-oriented cycle instead. These two models differ in function and structure of learning, decision making and optimization. In this work the structure of these two learning cycles are compared and a computational model for artificial emotional learning based engine is suggested.

1. INTRODUCTION

Today, the context of a military operation can span from small special unit operations to large multinational military endowers. To support this large mission spectrum it requires that a force has "tactical agility" [1], i.e. is able to quickly comprehend unfamiliar situations, creatively apply doctrine, and make timely decisions. These requirements demand for a communication infrastructure that is highly mobile and adaptive to the context of operation. Furthermore, the continuously increasing requirement of situation awareness requires more communication bandwidth at the tactical edge. Considering the heterogeneous set of systems of systems that is interacting in forming such infrastructure of networks of networks, the configuration, management and optimization is not a trivial problem. Traditional network management systems often rely on centralized and human-controlled management decisions propagated to network elements, which are clearly not adequate as a centralized entity can never be expected to have and analyze the network state information that is necessary to make informed network management decisions in such a dynamic and agile system that the communication infrastructure of a military operation must be today. The fulfillment of present requirements can only be achieved by autonomous network management. In the work reported in this paper such a cognitive radio network management system architecture is proposed in the context of next generation military tactical communication system. The emphasis is on the proposed cognitive engine, especially considering features like: capabilities for local processing of goals, monitoring of the local environment, reaction to contextual events by self-configuration and information propagation for interactions with and between components and network elements, objective functions and their interrelations to measuring and control parameters.

The rest of this paper is organized as follows: Section 2, describes a brief review of related works in autonomous network management. In Section 3, cognitive manager architecture is presented. Section 4 describes emotion-based reasoning and learning module. Section 5 presents the preliminary results of the suggested emotion-based engine. In Section 6, we conclude with some final notes.

2. BACKGROUND

The autonomous management of complex systems is a hard challenge it include different traditional research domains: system theory, artificial intelligence, communication systems signal processing, self-organizing systems and control theory etc. The starting point for creating an autonomous system must be to understand what it should handle and how it is handled manually and in our case these domains are spectrum management and network management of radio communication systems. Traditional spectrum management of military operations follows standardized processes and policies [2], it is handled by the use of spectrum management tools and centralized spectrum databases [3]. In the context of a military operation this coordination is referred to as battle space spectrum management [2]. Important to notice is that the use of electromagnetic spectrum is not limited to communication, many other aspect of a military operation relay on the use of electromagnetic spectrum, some examples are [4]: target acquisition, weapons control and guidance, navigation and terminal control, etc. Experiences from recent operations have

indicated that "current operational and tactical radio frequency (RF) spectrum planning and management practice do not keep pace with operations tempo" [5], indicating the lack of automation of managing frequency assignments. Important to notice is that these traditional spectrum management methods lack the ability to dynamically adapt in real-time to dynamic changes in the battle space. Network management refers to cooperative interaction between application processes in managing and managed systems for the management of telecommunications resources [6]. A more concrete interpretation of the concept can be made through the standardized decomposition of network management into five distinct functional areas:[1]

Fault management: is a set of functions that enable the detection, isolation and correction of abnormal operation of the network and its environment.

Configuration management: provides functions to exercise control over, identify, collect data from and provide data to network elements.

Accounting management: enables the measurement of the use of network services and the determination costs to the service provider and charges to the customer for such use

Performance management: provides functions to evaluate and report upon the behaviour and effectiveness of the network or network element, and through monitoring and control actions correct the behaviour and effectiveness of the network.

Security management: simply consists of the functions that are needed for maintain the secure operation of a communications network.

The goal of a joint spectrum/network management system is to achieve coherent and optimal system level behavior through interactions between operators and a large set of low-level control interfaces on network and radio elements. A set of new management challenges, [6]-[10], arise from the increased interconnection of heterogeneous networks, the increasing number of multimodal applications, increasing quality requirements, and the increasing complexity of many mobile networking environments. One such challenge stems from the increased variability reducing the ability of a human actor to perform any form of low-level decision. Next generation tactical radio network includes three main entities: platform, waveform and network. The platform is considered as a hardware that can load and run a set of different waveforms[2] (and application processes), figure 1. The waveform is the lowest entity representing a software defined transceiver (including relevant protocols) executing on a platform. A network is defined as an autonomous system (AS) of several platforms executing the same waveform forming a joint communication link or a network of communication links to its peer/peers. Each platform also includes routing capability that enabling forwarding between different waveforms, radio networks, and wired physical interfaces, i.e. Ethernet interface(s). A platform can also instantiate specific application processes except the actual waveforms, e.g. network management agents, cognitive management agents, security management agents platform management agent etc.

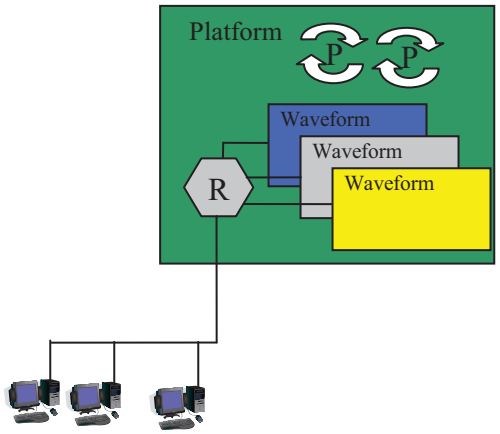


Figure 1. Platform with several waveforms.

2.1 Cognitive Radio and Networks

A cognitive architecture describes the necessary infrastructure for an intelligent system. Such a system should be generic and support different domains and knowledge bases. Cognition is a mental process that includes attention, memory, learning, reasoning, and decision making. A central aspect of a cognitive architecture is the underlying feedback loop for learning in which past interaction guide current and future interaction [11]. The cognitive loop is sometimes referred to as: observe - orient – decide - act (OODA) loop originally derived by John Boyd [12] for fighter pilots to understand the thought process of their adversaries, figure 2.

[1] The descriptions of the functional areas are adopted from ITU-T M.3400, TMN management functions.

[2] Important to notice is that can be several instances of one waveform or several different waveforms.

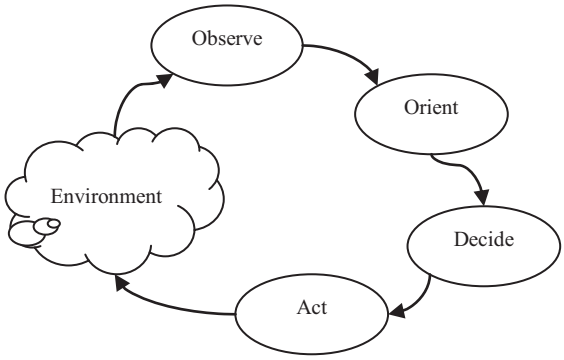


Figure 2. The OODA cycle [12].

The OODA loop has been adopted in subject areas spanning from business management to artificial Intelligence. The cognitive architecture proposed by Joe Mitola [13] includes a cognitive cycle that is an adoption of the OODA cycle that has ability to perform self-configuration tasks, learn from environment and its pervious action to react on new situation and reconfigure.

3. SYSTEM ARCHITECHTURE

To fulfill the previously mentioned agility requirement of next generation military communication systems, the management system should be based on a multi-tier structure, where individual tiers can operate autonomous without overlaying tiers. These tiers reflects the horizontal structure of a management system, in a networks of networks, including everything from individual parameters of a waveform executing on a platform to higher level policy respiratory for spectrum management and databases for operation plans and geographical information.

The cognitive manager architecture consists of two main parts: low level control/optimization functions and high level reasoning. The low level functions, the blue box in figure 1, is representing all algorithms that control and optimize the operation of the wave form in short term.

The main idea for the high level reasoning module of a cognitive manager, the purple box in figure 1, is to conduct the long term reasoning, learning and decision making, which is the focus of the rest of this paper.

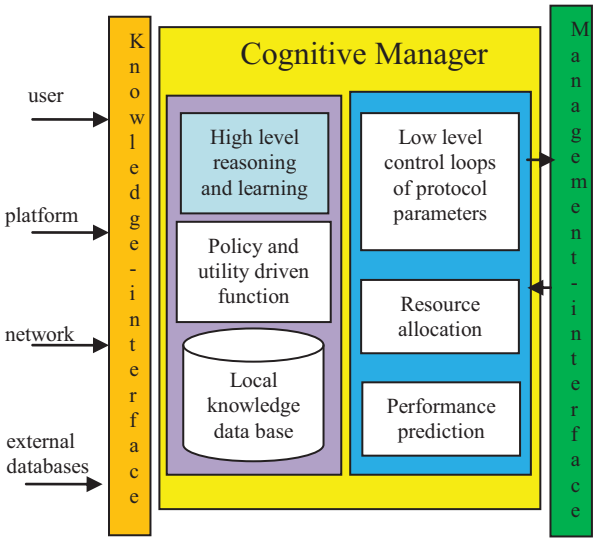


Figure 3. Abstraction of functional modules and interfaces.

4. EMOTION-BASED REASONING AND LEARNING MODUL

As was mentioned earlier, the reasoning and learning module of the cognitive manager is on the basis of brain emotional learning. We illustrate this module using an emotion-based engine that is inspired by human emotional system.

Functionally, the emotion-based engine imitates the function of the emotional system in learning and controlling basic instincts in particular fear; structurally, the suggested engine mimics the internal interaction of those regions of brain that are responsible for emotional processing.

4.1. Emotion and Anatomical Aspect of Emotion

For a long time, emotion was not assumed to be related to intelligence in human beings [14]. Hence, the emotional aspect of human behavior has so far received somewhat limited attention in the artificial intelligence research fields. In 1988 emotion was first proposed to be a principle part in human reaction [15]. Neuroscientists and psychologists have made a lot of efforts to analyze emotional behavior and describe emotion on basis of different hypotheses, e.g., psychological, neurobiological, philosophy and learning hypothesis. Studies of the emotional system have not only led to the explanation of emotional reactions by the application of different theories, e.g., central theory and cognitive theory [1]; they have also contributed to the development of computational models of emotional learning that is the basic of emotion-based artificial intelligence (AI) tools, intelligent controller [14], [16], [17] and data driven prediction methodologies [16], [18], [19], [20], [21], [22]. A good example of a computer based model that is relaying on the central theory [16] (i.e., which explains how a primary evaluation of emotional stimuli forms emotional

experiences) is called a computational model of emotional learning and imitates the associative learning aspect of emotional processing [16], which is based on fear conditioning [23], [24].

MacLean defined a group of the brain regions as the limbic system to describe the anatomical structure of brain emotional learning [25]. The limbic system consists of thalamous, sensory cortex, amygdala, hippocampus and hypothalamus, etc., The roles of the main regions of the limbic system with regard to fear conditioning can be summarized as follows:

1) ***Thalamus*** is the entrance gate of emotional stimuli. It determines the effective values of stimuli [26]-[32] to be pass to the amygdala and the sensory cortex [32].

2) ***Hypothalamus*** consists of several small nuclei and performs a variety of functions. Its main function is to connect the limbic system to the nervous system.

3) ***Sensory cortex*** is a part of the sensory area of the brain and is responsible for analysis and process of received signals [33]-[25], [31].

4) ***Amygdala*** is the central part of the limbic system of mammals and has a principle role in fear conditioning [18]-[26]. The amygdala consists of several parts with different functional roles (see Fig. 1) and it connects through them to other regions of the limbic system (e.g., the insular cortex, orbital cortex and frontal lobe). It has connections to the thalamus, the orbitofrontal cortex and the hypothalamus [34], [35]. During emotional learning, the amygdala participates in many tasks such as: reacting to emotional stimuli, storing emotional responses [36], evaluating positive and negative reinforcement and [37], learning the association between unconditioned and conditioned stimuli [23], [24] and [38], predicting the association between stimuli and future reinforcement [38], forming an association between neutral stimuli and emotionally charged stimuli [37]. The two main parts of the amygdala are the basolateral part (the largest portion of the amygdala) and the centeromedial part. The basaloteral part has a main role in mediating of memory consolidation [39], in providing the primary response. The centeromedial part, which is another important part of the basaloteral [35], is also divided into several regions[33], [24], [34], [35]. It is responsible for the hormonal aspects of emotional reactions [34] or mediating the expression of the emotional responses [34], [35].

5) **Hippocampous** is the main part of the mammalians' brain and performs function in the consolidation of information from short-term memory to long-term memory.

6) **Orbitofrontal cortex** is located close to the amygdala and has a bidirectional connection to the amygdala. This part is also involved in processing the stimuli, learning the stimulus–reinforcement association. It also evaluates the reinforcement signal to prevent the amygdala from providing an inappropriate response [39].

4.2. Emotion-based methods

We categorize emotion-based methods into three groups: emotion–based decision making model, emotion-based controller, and emotion-based machine learning approach.

1) *Emotion–based decision making model*: Some artificial intelligence (AI) emotional agents such as EMAI (emotionally motivated artificial intelligence) and DARE (Emotion-based Robotic Agent Development) have been developed. EMAI was applied for simulating artificial soccer playing [40] and its results were fairly good. While, DARE was based on the Somatic Marker theory, was examined in modeling social and emotional behavior [43].

2) *Emotion-based Controller*: BELBIC (Brain Emotional Learning-based Intelligent Controller) [7] is an emotion-based controller. The basic of BELBIC was the computational model that was presented by Moren et al [7],[27],[44]. This model has a simple structure (see Figure 4) that has been inherited from the anatomical structure of limbic system, e.g., the amygdala, the thalamus, the sensory cortex. It imitates the interaction between those parts of the limbic systems and formulates the emotional response using mathematical equations [27]. This model that is referred to as amygdala-orbitofrontal subsystem consists of two main subsystems: the amygdala and the orbitofrontal. Each subsystem has several linear neurons and receives a feedback (a reward). BELBIC have been tested for a number of applications: controlling heating and air conditioning [43] aerospace launch vehicles [44], intelligent washing machines [45] and trajectory tracking of stepper motor [46]. BELBIC has shown an excellent performance overcome uncertainty and complexity issues of control applications. Specifically, the BELBIC has been proven to outperform other in terms of simplicity, reliability and stability [16], [41]-[44].

3) *Emotion-based machine learning approach*: Several machine learning approaches have also been developed based on emotional processing of the brain. Some examples are hippocampus-neocortex model and amygdala hippocampus model [47], [48]. These methods combine associative neural network with emotional learning concepts. Moreover, emotion-based prediction models [47]-[48] have been developed to predict the future behavior of complex system. They have been applied for different applications, e.g., solar activity prediction [18]-[20]. Recently, a classification model based on amygdala-orbitofrontal system has also been developed; the obtained results from testing this model on two benchmark data sets has verified its good performance.

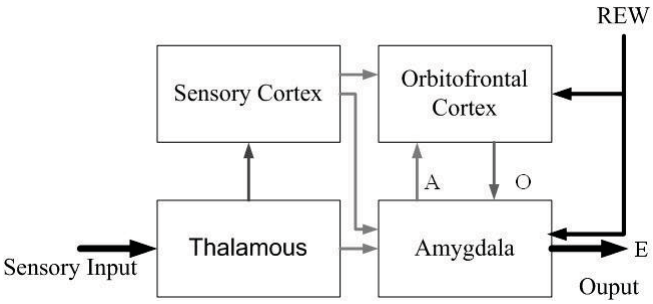


Figure 4. The graphical description of amygdala-orbitofrontal subsystem.

4.3. Emotion-based Engine

Joseph Mitola defined the term cognitive cycle (see Figure 5) to describe the intelligent behavior of radio nodes through five iterative steps: observe, orient, plan, decide and act. Mitola's cognitive cycle has been developed on the basis of the OODA loop that uses Unified Theories of Cognition (UTC) to model rational decision making in human [49]. As a matter of fact, the Mitola's cognitive model is on the basis of logical decision making processes in human considering the all environmental states are fully observable.

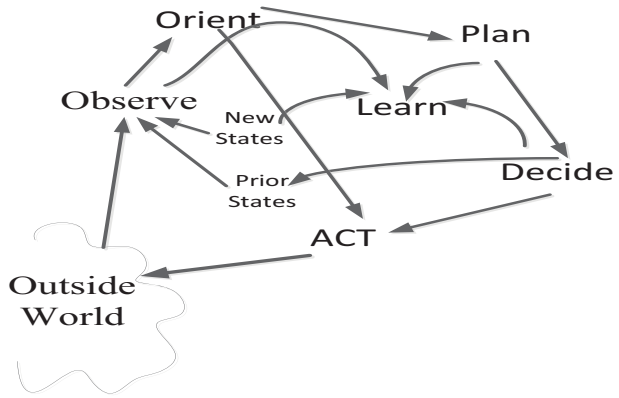


Figure 5.Cognitive radio cycle

4.3.1. Emotional Cycle

We develop an emotional cycle to represent the functions of the reasoning and learning module, assuming environmental changes elicits emotional stimuli that lead to emotional reactions. The emotional cycle imitates the brain's pathway from an emotionally charged stimulus (e.g. fear full stimulus) to an emotionally response (e.g. freezing) and it consists of three steps: sensing, learning and acting (see Figure 6).

In the cognitive radio network context, the emotional stimulus is a vector of environmental changes that can be triggered by different sources, e.g., users, the network policies, radio frequency channels, etc. The emotional response is a waveform configuration. In the following we explain the cognitive cycle assuming the emotional stimulus that is a vector of environmental information, e.g., channel environment, network interface and user interface is triggered. The sensing step recognizes and interprets the received information. The procedure is led to a low level learning that provides useful information for the following steps. The learning step deals with making an optimal decision. It combines prediction algorithms (regression or classification) with optimization algorithms. The reacting step forms the waveform configuration that is equivalent with the emotional response and sends it to the radio platform.

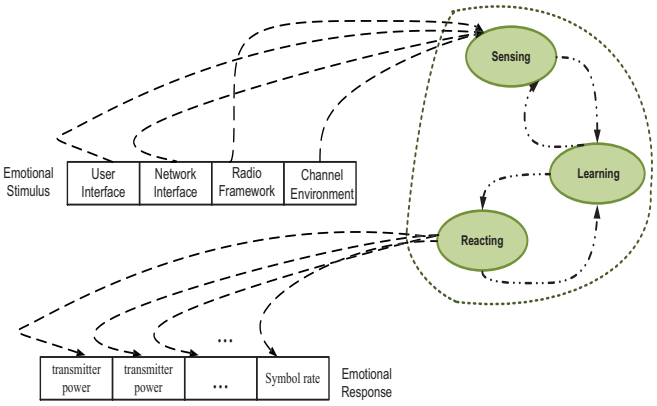


Figure 6. The emotional cycle.

4.3.2. Emotional Cycle and Limbic System

As was discussed earlier, the limbic system is the main responsible of emotional behavior e.g., fear conditioning in mammals. Figure 7 describes the pathways of fear conditioning and shows the connection between the regions of the limbic system. Each step of emotional cycle is equivalent to some regions of limbic system and mimics the functionality of those parts. The sensing step imitates the functionality of the thalamus and sensory cortex i.e., recognition, interpretation, quick perception, accurate representation. This step provides high level information copying the selective and processing procedure in the thalamus and sensory cortex. This information indicates which specific algorithm should be activated in the next step, learning step. The learning step mimics the roles and interactions of the amygdala, hypothalamus and orbitofrontal cortex using prediction, optimization, decision making algorithms. The reacting steps imitate the functions of hippocampus by constantly produce a response (here a waveform) and adjusts it.

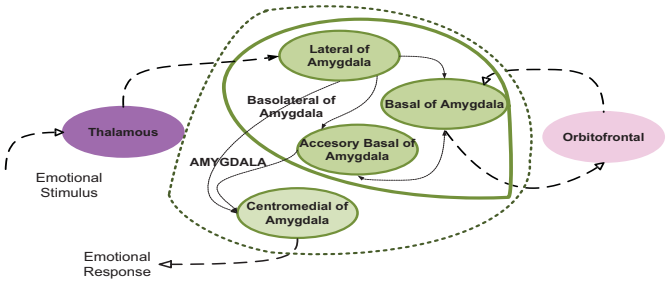


Figure 7. The components of brain emotional system and their connections according to fear conditioning.

4.3.3. Emotional Cycle and Mitola's Cognitive Cycle

As mentioned earlier, Mitola's cognitive cycle and UTC express how a rational goal-oriented decision can be made by an intelligent agent. However, the emotional cycle represents an emotion-oriented action. In addition, the emotional cycle and Mitola's cognitive cycle differs on implementing the learning, decision making and optimization algorithms. Figure 7 depicts Mitola's cognitive cycle and shows how it can be mapped to the emotional cycle. As Figure 8 indicates the sensing is corresponding with: observe, orient and act states of the cognitive cycle. The learning is correspondent with the plan, learn and decide. And the reacting is corresponding with decide and act parts.

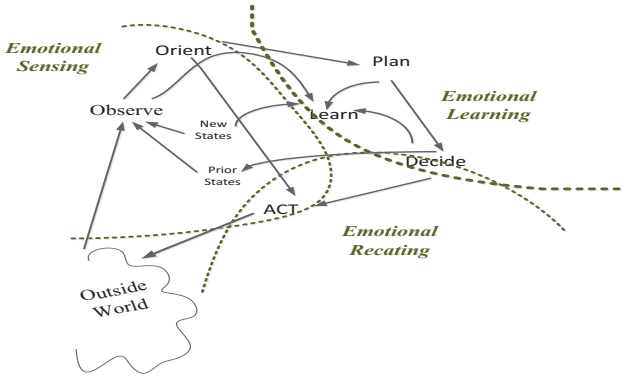


Figure 8. The emotional cycle and Mitola's Cognitive Cycle.

4.3.4. A Simple Structure of the Suggested Engine

A simple structure of the emotion-based engine has been developed and tested for prediction and classification application [18]-[20].

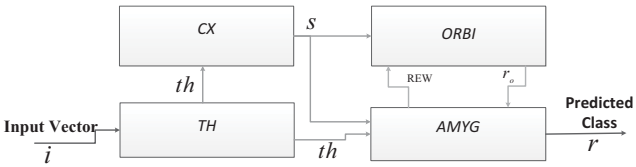


Figure 9. A simple emotion-based engine.

As Figure 9 shows, the structure consists of four parts: THalamous (TH), sensory CorteX (CX), AMYGdala (AMYG) and ORBItofrontal cortex (ORBI). Let us assume i as an input vector of stimulus; the functionality and the connection between the parts can be described as follows:

1) TH extracts high level information from the input stimulus, e.g., the maximum value of input vector, i. TH provides th and sends it to AMYG.

2) CX also receives th and provides, $\mathbf{s}$ that is sent to both AMYG and ORBI.

3) AMYG has connections with all other parts. It receives two inputs th and $\mathbf{s}$ which are originated from TH and CX, respectively. Using a data-driven method e.g. neural network or neuro-fuzzy method, the AMYG provides a primary response. In addition, a reinforcement signal, REW, is provided by AMYG for this response and is sent it to ORBI.

4) ORBI also adapts a learning algorithm and provides a secondary response r_o; however, the final response, r, is provided by AMYG.

5. PRELIMINARY RESULTS

Applying the suggested emotional learning based engine for predicting time series and classification have indicated excellent results. For example the brain emotional learning-based recurrent fuzzy system (BELRFS) [18] is based on the four previous mentioned modules: TH, CX, AMYG and ORBI. Figure 8 depicts the basic structure of the model and the connection between its sub-modules. BELRFS has been applied for predicting some chaotic time series, e.g. Lorentz time series, and the result was compared with another black box models, local linear neuro fuzzy model with local tree model algorithm (LoLiMoT). Some of the result indications for prediction ability are reproduced in table I below. For comparison of the results, the error index of normalized mean square error (NMSE) is considered as the error measure.

Table I. The NMSEs of BELRFS and LoLiMoT to predict multi-step ahead of Lorenz

Learning Model	NMSE index for multi-step ahead prediction		
	10 step ahead	*20 step ahead*	*30 step ahead*
BELRFS	1.463e-4	0.1250	0.6023
LoLiMoT	0.0012	0.1509	0.7086

BELRFS has also been applied benchmarked by predicting sunspot number time series. The obtained results indicate that BELFIS is a reliable, nonlinear predictor model for solar activity forecasting. So as conclusion one can state that it can be applied as a predictor model for the long-term and short-term prediction of chaotic and nonlinear systems which are one important feature for an cognitive manager.
Classification ability for a computational model of emotional learning has also been tested by using some benchmark data sets. Table II presents the results of classification of Wine data using the simple emotion-based engine. The Wine data set is a multiclass data set with 13 features and can be categorized into three classes [50]-[51].

The detail of the emotion-based engine that is referred to as Brain Emotional Learning Based Ensample Classifier (BeLBeC) has been explained at [50]. As the table verifies, the emotion-based engine has excellent performance as the classification method.

Table II. The classification accuracy of BELBEC for the wine data set with 60 samples as training data and 118 samples test data.

Classification Model	Specification of results		
	Structure	*The average per class accuracy*	*The number of training samples*
BELBEC without normalized[50]	16 neighbors	%99.02	60
McNN [51]	9 neurons	%98.49	27
PBL-McRBFN [51]	11 neurons	%98.69	29

6. CONCLUSION

This paper suggested architecture for a cognitive engine based on brain emotional learning, which differs from previously proposed architectures of cognitive engines. Previous cognitive engines are based on a cognitive cycle that is defined according to the rational reasoning system of the human brain. A rational goal-oriented decision is made by an intelligent agent while the emotional cycle represents an emotional reaction-oriented action. The conducted work indicates that the two systems differs in function and structure of learning, decision making and optimization algorithms. Mitola's cognitive cycle can be mapped to the emotional cycle. As indicated the sensing is corresponding with: observe, orient and act states of the cognitive cycle. The learning is correspondent with the plan, learn and decide. And the reacting is corresponding with decide and act parts. Our suggested engine is based on the emotional system. The emotional system is faster and has a more basic structure, from an evolutionary perspective, than the rational system. Preliminary result, for prediction and classification indicate very promising performance and previous work has shown that the computational models of brain emotional learning also can be used for intelligent control. Ongoing work is developing an entity for emotion-based decision making and an emotion-based optimization.

7. REFERENCES

[1] S. R. Atkinsson and J. Moffat, *The Agile Organization – From informal networks to complex effects and agility*, Command and Control Research Program (CCRP) publishing series, 2005.

[2] *GUIDE TO SPECTRUM MANAGEMENT IN MILITARY OPERATIONS*, ACP 190 (C), a Combined Communications-Electronics Board (CCEB) publication, September 2007.

[3] W. J. Morgan, *DoD Spectrum Management: A Critical Analysis*, Air Force Institute of Technology, June 2008.

[4] *POLICY FOR THE COORDINATION OF MILITARY ELECTROMAGNETIC SPECTRUM ALLOCATIONS AND*

ASSIGNMENTS BETWEEN COOPERATING NATIONS ACP 194, a Combined Communications-Electronics Board (CCEB) publication, June 2011.

[5] R. Poe, R. Shaw H. Zebrowitz W. Kline, W. Heisey, F. Loso, and Y. Levy, "Optimal Spectrum Planning and Management with Coalition Joint Spectrum Management Tool (CJSMPT)," in proceedings of *Military Communications Conference (MILCOM) 2008*, San Diego, U.S. 2008.

[6] Douglas E. Comer, *Automated Network Management Systems – Current and Future Capabilities*, Pearson Prentice Hall, 2007.

[7] S. Dobson et al, "A Survey of Autonomic Communications", *ACM Transactions on Autonomous and Adaptive Systems*, Vol. 1 No. 2, December 2006.

[8] R. Chadha and L. Kant, *Policy-Driven Mobile Ad Hoc Network Management*, Wiley Series in Telecommunications and Signal Processing, 2008.

[9] M. J. Ryan and M. R. Frater, *Tactical Communications for the Digitized Battlefield*, Artech House, 2002.

[10] Q. H. Mahmoud, *Cognitive Networks – Towards Self-Aware Networks*, Wiley, 2007.

[11] R. W. Thomas, L. A. DaSilva, and A. B. MacKenzie, "Cognitive Networks," in proceedings of *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks 2005*, DySPAN 2005, Baltimore Maryland, U.S. 2005.

[12] J. Boyd, A discourse on winning and losing: Patterns of conflict, 1986.

[13] J. Mitola, Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio, PhD thesis, Royal institute of Technology (KTH), Sweden 2000.

[14] L. Custodio, , R. Ventura, , C. Pinto-Ferreira, 'Artificial emotions and emotion-based control systems', *Proceedings of 7th IEEE International Conference on Emerging Technologies and Factory Automation,* pp. 1415-1420, 1999.

[15] J. M. Fellous, J. L. Armony, and J. E. Le Doux, "Emotional Circuits and Computational Neuroscience,'' in *The Handbook of Brain Theory and Neural Networks*, The MIT Press, Cambridge, MA, 2003.

[16] C. Lucas, , D. Shahmirzadi, , and N. Sheikholeslami, (2004) 'Introducing BELBIC: Brain Emotional Learning Based Intelligent Controller', *International Journal of IntelligentAutomation and Soft Computing (Autosoft),* Vol. 10, pp. 11-22.

[17] S. H. Zadeh, S. B. Shouraki, and R. Halavati, "Emotional behaviour: A resource management approach", *Adaptive Behaviour Journal,* Vol.14, pp. 357-380,2006.

[18] M. Parsapoor, M, U. Bilstrup, "Neuro-fuzzy models, BELRFS and LoLiMoT, for prediction of chaotic time series," in *Proc. IEEE Int. Conf. INISTA.,* pp.1-5, 2012.

[19] M. Parsapoor, C. Lucas and S. Setayeshi, "Reinforcement _recurrent fuzzy rule based system based on brain emotional learning structure to predict the complexity dynamic system," *in Proc. IEEE Int. Conf. ICDIM,* , pp.25-32, *2008*.

[20] M. Parsapoor, U. Bilstrup, "Brain Emotional Learning Based Fuzzy Inference System (BELFIS) for Solar Activity Forecasting," in *Proc. IEEE Int. Conf. ICTAI 2012*, 2012.

[21] C. Cavada., W. Schultz., "The Mysterious of Orbitofrontal Cortex. Foreword. Cereb Cortex,'', *J. Cerebr. Cortex.* vol. 10, no. 3, pp .205, 2000.

[22] C. A. Winstanley, D. E. H. Theobald, R. N. Cardinal, and T. W. Robbins, "Constracting Roles of Basolateral Amygdala and Orbitofrontal Cortex in Impulsive Choice," *J. Neurosci.*, vol. 24, no. 20, pp. 4718-4722, 2004.

[23] M. L. Kringelbach, and E. T. Rolls, "The functional neuroanatomy of the human orbitofrontal cortex: evidence from neuroimaging and neuropsychology," *J., Prog. Neurobiol,* vol. 72, pp. 341–372, 2004.

[24] Damas, B. D. and Custódio, L.: "Emotion-Based Decision and Learning Using Associative Memory and Statistical Estimation," *J. Informatica (Slovenia),* vol. 27, no. 2, pp. 145-156, 2004.

[25] J. D. Velásquez., "When Robots Weep: Emotional Memories and Decision-Making," in *Proc. Conf. on Artifitial Intelligence*, pp.70-75, 1997.

[26] S. H. Zadeh, S. B. Shouraki, and R. Halavati, R. "Emotional behaviour: A resource management approach," *J. Adaptive Behaviour, vol.* 14, pp. 357-380, 2006.

[27] M. Maçãs and L. Custódio, "Multiple Emotion-Based Agents Using an Extension of DARE Architecture," *J. Informatica (Slovenia),* vol. 27, no. 2, pp. 185-196, 2004.

[28] E. Daryabeigi, G. R. A. Markadeh, C. Lucas, "Emotional controller (BELBIC) for electric drives — A review," in *Proc. Annual Conference on IEEE Industrial Electronics,* pp.2901-2907, 2010.

[29] N. Sheikholeslami, D. Shahmirzadi, E. Semsar, C. Lucas., "Applying Brain Emotional Learning Algorithm for Multivariable Control of HVAC Systems,", *J. INTELL. FUZZY. SYST.*vol.16, pp. 1–12, 2005.

[30] *A. R. Mehrabian, C. Lucas, J. Roshanian,"Aerospace Launch Vehicle Control: An Intelligent Adaptive Approach", J. Aerosp. Sci. Technol*., vol.10, pp. 149–155, 2006.

[31] J. L. Armony, and J. E. LeDoux, "*How the Brain Processes Emotional Information*," *J. Ann. N. Y. Acad*., no. 821, pp. 259-270, 1997.

[32] J. M. Jenkins, K. Oatley, N. L. Stein, *Human emotions: A READER*, Blockwell publisher, U.K., 1998.

[33] D. H. Hubel, M. S. Livingstone, "Color and Contrast Sensitivity in the Lateral Geniculate Body and Primary Visual Cortex of the Macaque Monkey," *J., Neuroscience.* vol. 10, no.7, pp. 2223-2237, 1990.

[34] J. P. Kelly, "The Neural Basis of Perception and Movement, Principles of Neural Science," London: Prentice Hall. 1991.

[35] K. Amunts., O. Kedo., M. Kindler., P. Pieperhoff., H. Mohlberg., N. Shah., U. Habel., F. Schneider., K. Zilles., "Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex : intersubject variability and probability maps,'' *J. Anatomy and Embryology.,* vol. 21, no. 5-6, pp. 343-352, 2005.

[36] C. I. Hooker., L. T. Germine., R. T. Knight., M. D. Esposito., "Amygdala Response to Facial Expressions Reflects Emotional Learning,'' *Neuroscience. J.,* vol. 26, no.35, pp. 8915-8930, Aug. 2006.

[37] J. Moren, C. Balkenius,"A computational model of emotional learning in the amygdala,'', *in From Animals to Animats*, MIT, Cambridge, 2000.

[38] B. D. Damas, and L. Custódio, "Emotion-Based Decision and Learning Using Associative Memory and Statistical Estimation," *J. Informatica (Slovenia),* vol. 27, no. 2, pp. 145-156, 2004.

[39] J. D. Velásquez., "When Robots Weep: Emotional Memories and Decision-Making," in *Proc. Conf. on Artifitial Intelligence*, pp.70-75. 1997.

[40] S. H. Zadeh, S. B. Shouraki, and R. Halavati, R. "Emotional behaviour: A resource management approach," *J. Adaptive Behaviour, vol.* 14, pp. 357-380, 2006.

[41] M. Maçãs and L. Custódio, "Multiple Emotion-Based Agents Using an Extension of DARE Architecture," *J. Informatica (Slovenia),* vol. 27, no. 2, pp. 185-196, 2004.

[42] E. Daryabeigi, G. R. A. Markadeh, C. Lucas, "Emotional controller (BELBIC) for electric drives — A review," in *Proc. Annual Conference on IEEE Industrial Electronics,* pp.2901-2907, *2010.*

[43] N. Sheikholeslami, D. Shahmirzadi, E. Semsar, C. Lucas., "Applying Brain Emotional Learning Algorithm for Multivariable Control of HVAC Systems,", *J. INTELL. FUZZY. SYST.*vol.16, pp. 1–12, 2005.

[44] *A. R. Mehrabian, C. Lucas, J. Roshanian,"Aerospace Launch Vehicle Control: An Intelligent Adaptive Approach", J. Aerosp. Sci. Technol*.,vol.10,pp. 149–155, 2006

[45] *M. Milasi, C. Lucas, B. N. Araabi, "Intelligent Modeling and Control of Washing Machines Using LLNF Modeling and Modified BELBIC,"* in *Proc. Int. Conf. Control and Automation.,* pp.812-817, 2005.

[46] A. M. Yazdani1, S. Buyamin1, S. Mahmoudzadeh2, Z. Ibrahim1 and M. F. Rahmat1., "Brain emotional learning based intelligent controller for stepper motor trajectory tracking," *J. IJPS*., vol. 7, no. 15, pp. 2364-2386, 2012.

[47] T. Kuremoto, T. Ohta, K., Kobayashi, M., Obayashi, "A dynamic associative memory system by adopting amygdala model," *J. AROB, v*ol.13, pp.478–482, 2009.

[48] T. Kuremoto, T. Ohta, K. Kobayashi, K., *M.* Obayashi, "A functional model of limbic system of brain," in *Proc*. *Int. Conf. Brain informatics*, pp.135-146, 2009.

[49] A, Amanna and J.H., Reed, "Survey of cognitive radio architectures," Proceedings of the IEEE SoutheastCon 2010, pp.292-297, 18-21 March 2010.

[50] M. Parsapoor and U. Bilstrup, "Brain Emotional Learning-based Ensemble Classifier (BELBEC)," submitted to *8th International Symposium Advances in Artificial Intelligence and Applications, AAIA 2013*.

[51] G. S. Babu and S. Suresh, "Sequential Projection-Based Metacognitive Learning in a Radial Basis Function Network for Classification Problems," *Neural Networks and Learning Systems, IEEE Transactions on* , vol.24, no.2, pp.194,206, Feb. 2013

DESIGN AND IMPLEMENTATION OF A FFT PRUNING ENGINE FOR DSA-ENABLED COGNITIVE RADIOS

Manuele Cucchi , Deepak Revanna , Roberto Airoldi and Jari Nurmi
Tampere University of Technology
Department of Electronics and Communications Engineering
P.O. Box 553, FIN-33101, Tampere Finland
name.surname(at)tut.fi

ABSTRACT

This research work presents the design and implementation of an FFT pruning block, as an extension of an FFT core for OFDM demodulation, enabling run-time pruning of the FFT algorithm. The pruning engine allows any pruning pattern without any restrictions on the distribution pattern of the active/inactive sub-carriers, enabling the efficient implementation of dynamic spectrum access (DSA) cognitive radios. The system was prototyped on an ALTERA STRATIX V FPGA in order to evaluate the performance of the pruning engine. Synthesis and simulation results showed that the logic overhead introduced by the pruning block is limited to a 10% of the total resources utilization. Moreover, in presence of a medium-high scattering of the sub-carriers, the power consumption of the FFT core was reduced up to 40%.

1. INTRODUCTION

The continuous evolution of services available through the network has pushed mobile users to increasingly demand more and more bandwidth. Because of the increasing bandwidth demand, police makers and communication technologists had to seek new solutions for resolving the issues of limited spectrum availability. In fact, the spectrum is a limited resource and the spectrum utilization will soon saturate. However, a recent study of the Federal Communication Commission (FCC) has shown that the spectrum scarcity is a false problem, which arises from inefficient spectrum utilization [1]. In fact, from the FCC study it emerges that the spectrum is poorly utilized across frequency, space and time. As a consequence, researchers from industry and academia have developed new paradigms for mitigating the spectrum utilization inefficiency. One proposed paradigm for efficiently utilizing spectrum is dynamic spectrum access (DSA) [13].

To realize DSA communications, highly reconfigurable wireless platforms are needed in order to provide the required spectral agility. Flexible and efficient baseband processing platforms are available through the software-defined radio (SDR) technology and cognitive radio systems [12]. These systems can be utilized to enable DSA communications. Besides the architectural solutions, new agile communication techniques are required to more efficiently exploit the DSA concept: conventional wireless communication systems based for example on frequency division multiplexing (OFDM), might not possess an adequate level of spectral agility, required by DSA communications. As a result, variants of OFDM called non-contiguous OFDM (NC-OFDM) and Discontinuous-OFDM (D-OFDM) were introduced [4] [5].

NC-OFDM systems can transmit data across non-contiguous frequency blocks of sub-carriers turning off the remaining sub-carriers, which are located within the spectral vicinity of existing primary user communications, in order to avoid interference. One of the main advantages of OFDM and its variants is that its implementation of parallel modulated streams of subcarrier data can be efficiently realized using a Fast Fourier Transform (FFT), where each FFT point represents an OFDM sub-carrier. The utilization of NC-OFDM opens opportunities for efficient implementations of the demodulation/modulation blocks. In fact, sub-carriers that are turned off are interpreted by the FFT block as zero-value inputs. Therefore, FFT pruning algorithms could be utilized for lightening the computational load of the FFT block, leading to more power-efficient implementations.

This research work presents the design of an FFT pruning engine, as an add-on block for an existing FFT core. The FFT pruning block allows the elimination of redundant operations at run-time, enabling an efficient implementation of a demodulation block for NC-OFDM systems. The manuscript is organized as follows: Section 2 introduces the theory behind the FFT pruning; Section 3

analyses the state-of-the art in the design and implementation of FFT pruning engines; Section 4 describes in detail the proposed FFT pruning engine and its application to an FFT core; Section 5 and Section 6 present and discuss the achieved results; finally, Section 7 draws conclusions and points out future directions.

2. FFT PRUNING THEORY

FFT pruning was introduced by Markel in [1] as an effective way to reduce the computation complexity of FFT algorithms whenever in presence of zero-valued inputs. The idea behind FFT pruning is to reduce the computational complexity of the algorithm via the elimination of redundant operations, such as: multiplication or addition of neutral terms as well as multiplications by zero factors. In fact, the results of such operations are either a copy of one of the two operands (multiplication/addition by neutral elements) or zero (multiplication by a zero factor) and therefore the mathematical operation can be pruned without any consequences on the correctness of the algorithm. To better underline the simplifications that can be introduced by FFT pruning, Figure 1 presents an 8-point FFT data-flow and an example of input distribution. In the particular example, only 2 out of the 8 inputs are non-zero (x1 and x5). The dashed lines in Figure 1 represent the operations that can be pruned.

The mathematics behind FFT pruning and its advantages have been widely studied. As an example, the computation complexity reduction of DFT and FFT algorithms is discussed in detail in [2] and [3]. From an implementation point of view, the challenges introduced by FFT pruning reside in how to design and implement efficiently the pruning, without any over complications of the control plane. Many different implementations of the pruning algorithm have been proposed.

Alves et al. in [4] present a pruning algorithm based on a configuration matrix and an *if-then-else* statement. The configuration matrix size is *N x log2 N*, where N is the FFT size. Each column of the matrix represents an FFT stage, while the rows represent the input vector for each FFT stages. The matrix is a binary matrix: logical 0 corresponds to a zero-valued input, while logical 1 corresponds to a non-zero input. The computation of the algorithm is then based on a conditional execution of the operations on the basis of the valued stored in the configuration matrix. This approach is based on an extensive utilization of *if-then-else* statements, which potentially limits the advantages of the complexity reduction.

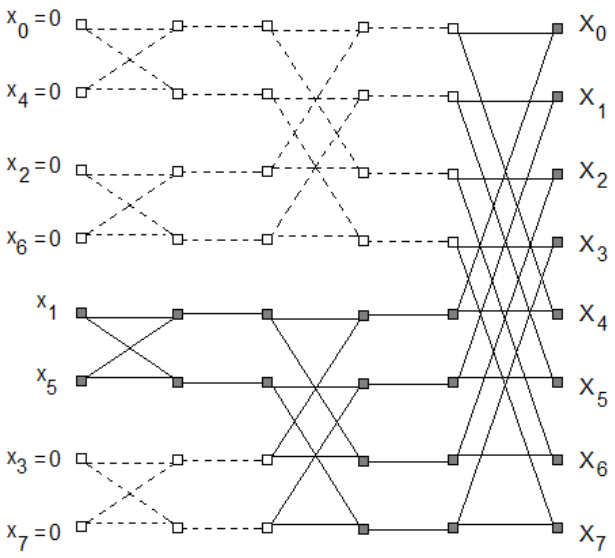


Figure 1: 8-point FFT data-flow and example of pruning simplification.

$$M_{[4]} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix} \quad M_{[5]} = \begin{bmatrix} 2 & 4 & 8 \\ 4 & 4 & 0 \\ 5 & 5 & 1 \\ & 6 & 2 \\ & 7 & 3 \\ & & 4 \\ & & 5 \\ & & 6 \\ & & 7 \end{bmatrix} \quad M_{[6]} = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 2 & 0 \\ & 3 & 1 \\ & & 2 \\ & & 3 \end{bmatrix}$$

Figure 2: Configuration matrix for the algorithms proposed in [4][5][6] according to the input distribution presented in Figure 1.

Rajbanshi et al. in [5] introduce a different configuration matrix: Rajbanshi's configuration matrix is composed of *log2 N* columns and *N/2+1* rows. The first row indicates how many non-zero inputs are present for each one of the FFT stages (columns), while the following elements of the column indicate the actual position of the non-zero inputs.

An improved version of Rajbanshi's configuration matrix algorithm is proposed in [6] by Airoldi et al.; the configuration matrix is reduced in size, leading to a more embedded-system friendly implementation. The reduction of the size of the configuration matrix was made possible by storing in the configuration matrix a butterfly identifier instead of the position of each element. However, this

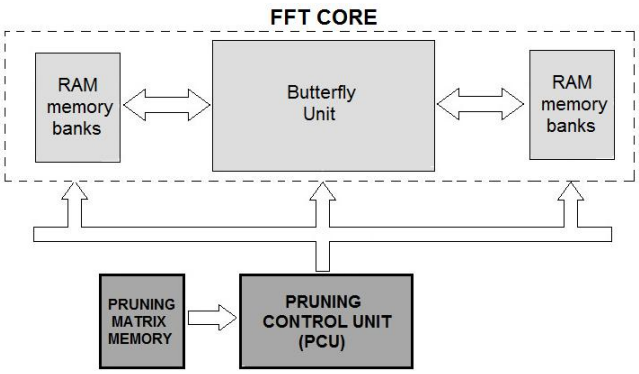


Figure 3: Simplified block diagram of the FFT core and Pruning engine integration.

introduced a restriction in the pruning, allowing the pruning of the butterflies in only the case where both inputs are equal to zero. The effectiveness of the pruning algorithm is then reduced but the actual implementation of the algorithm is simplified, leading to good performance figures. To highlight the differences between the presented algorithms, the configuration matrixes for [4], [5] and [6] are reported in Figure 2. The matrixes were constructed for an 8-point FFT and the example input distribution proposed in Figure 1. From an implementation point of view, the configuration matrixes proposed in [5] and [6] are more suitable for software implementations, while the configuration matrix proposed in [4] can be more efficiently ported to hardware implementations.

3. RELATED WORKS

Different software-hardware solutions have been proposed for the implementation of the FFT pruning algorithm. This section gives an overview of significant research works related to the efficient hardware implementation of FFT pruning for cognitive radios applications.

Venilla et al. in [7] present a 64-point FFT/IFFT with pruning for OFDM-based cognitive radios. The proposed architecture is implemented on a Field Programmable Gate Array (FPGA). To provide the required flexibility for the implementation of the pruning, the work is based on the dynamic partial reconfiguration provided by FPGA technology [10]. The architecture implements a radix-2 FFT algorithm. However, no details are given about the utilization of a configuration matrix for the pruning selection. The results achieved showed a power reduction of 68% when only 16 out of 64 inputs are non-zero. The proposed implementation allows the pruning of the input only for contiguous blocks of 16 inputs at the time, prohibiting the application of the approach to different types of scenario. Moreover, the proposed architecture is tightly bounded to a particular FPGA implementation.

Xu and Lim in [8] propose an FFT pruning design based on a split radix implementation of the FFT kernel. The pruning of the input is managed through an *N x log2 N* configuration matrix. The matrix is a binary matrix and it is similar to the one proposed in [4]. However, the elements of the pruning matrix identify a multiplication and not a single data input. The analysis of the reduction of the computational complexity is evaluated for a 1024-point FFT. Moreover, a hardware implementation of the algorithm (for a 64-point FFT) is presented, but no power consumption and resources utilization figures are given.

Chen et al. in [9] introduce an FFT processor for OFDMA communication systems (e.g. IEEE 802.16e/m and 3gpp-LTE). OFDMA in respect of OFDM allows multi-modal transmission utilizing different numbers of sub-carriers utilization. Therefore, more energy efficient implementation can be achieved through the use of FFT pruning. The proposed FFT architecture is able to perform N-point FFT with N ranging from 128 to 1024. The pruning algorithm introduced relies on OFDMA specifications and therefore is not suitable for cognitive radios applications, where the dynamics of the inputs could be more variegated.

Jang et al. in [15] describe a Synchronous Data Flow (SDF) architecture for the implementation of a 2048-point FFT algorithm. The architecture is described in Verilog HDL and synthetized using Samsung 130nm standard cell libraries. The analysis of power figures underlines a 22.3% reduction of power consumption when the zero input ratio is increased from 30% to 90%.

4. PROPOSED PRUNING ENGINE

The pruning engine proposed by this research work was designed as an add-on block for an existing FFT core [17]. Therefore, the internal control structure of the FFT core was left unaltered. Figure 3 presents a conceptual view of the integration of the FFT core and the pruning engine. Moreover, because the design of the pruning engine was not tailored to the particular FFT core utilized, the proposed solution can be ported to similar types of FFT cores. The design of the pruning engine is composed of two parts: 1) the pruning algorithm and 2) the pruning architecture.

The pruning algorithm defines how the pruning is performed and on which constraints, while the pruning

architecture takes care of the generation of the right control signals for the FFT core, in order to efficiently implement the pruning algorithm.

4.1 Pruning algorithm

The proposed pruning algorithm is based on a pruning matrix. The pruning matrix was derived from [4] and [6]. In
particular, given an algorithm of N-point FFT, the pruning matrix is a $n \times m$ matrix, where n = $log_2(N)$ columns while $m = N/2$. Each column of the matrix represents a stage of the
radix-2 FFT and each one of $N/2$ rows identifies one of the $N/2$ butterflies that compose a given stage (row). The matrix maps the precise location of each butterfly in the data-flow. Each butterfly is represented by one bit (one matrix element), which indicates if the butterfly operation has to be computed or not. To better underline the construction of the pruning matrix, Figure 4(a) presents the pruning matrix for the 8-point FFT example presented in Figure 1. The pruning matrix is stored into a RAM memory. The storage of the pruning matrix is shown in Figure 4(b): a single memory word stores the configuration for up to 32 consecutive butterflies. Finally, the configuration words steer a clock gating system for the activation of the butterfly unit and the memory address generations for the data fetch.

4.2 Pruning architecture

As shown by Figure 3, the FFT core is built around a computational block named Butterfly Unit. The Butterfly Unit can compute two butterfly operations in parallel [14]. A simplified schematic view of one of the two butterfly blocks is given in Figure 5. The figure also highlights the control signals of the pruning engine (*Bfy0_enableX* in the Figure 6). In fact, the pruning control unit (PCU) enables

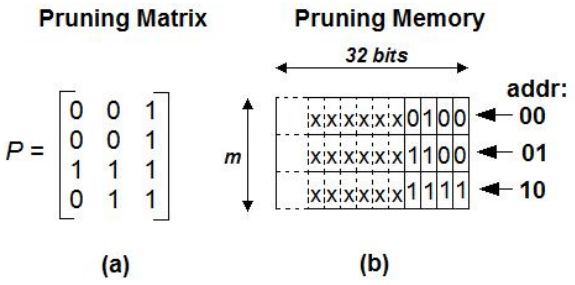


Figure 4: a) Example of pruning matrix in respect of example of Figure 1; b) Memory organization for the storage of the pruning matrix.

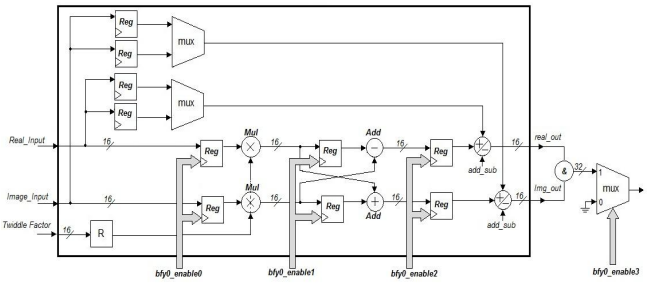


Figure 5: Simplified schematic view of one of the two a butterfly units of the FFT core.

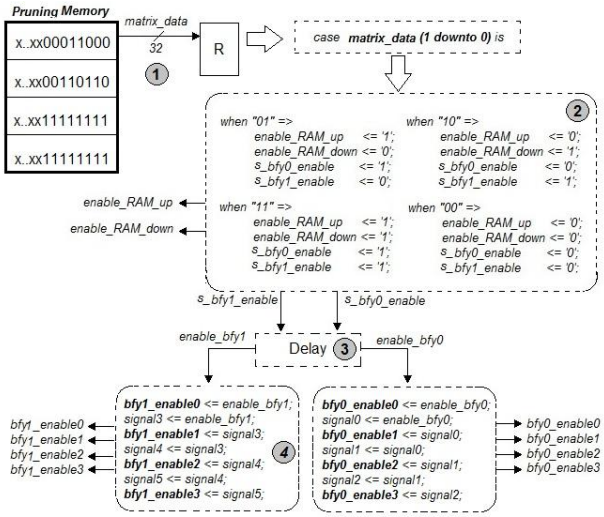


Figure 6: Conceptual view of the Pruning Control Unit design.

or disables the butterfly operations if both inputs are zero. In particular, the enable signals are latched only to the multipliers: the energy consumption of adders is not as significant as the energy consumed by the four multipliers. At each step of the algorithm, the PCU reads and decodes the configurations from the pruning memory and generates the enable signals accordingly. The whole process is timed by the clock of the FFT core in order to keep the two control units synchronized. Indeed, the generation of the enable signals has to respect the pipeline timing of the butterfly unit. Together with the butterfly enables, the PCU generates other two enable signals, which are connected to the two RAM memories. In fact, whenever a butterfly is pruned, the associated reading from the memories is disabled. To better emphasize the steps that characterize the pruning algorithm, Figure 6 provides a simplified view

of the control flow for the PCU. Initially, the PCU fetches from the pruning matrix memory the configurations for the required butterfly operations. The configuration word serves 32 following butterflies within a FFT plane. For larger FFT sizes (requiring more than 32-bit configuration words for each FFT plane) the configuration words are stored at sequential addresses. In such cases, multiple read phases from the memory are needed. In the second phase, the PCU decodes the configuration fetched and generates the enable signals for the data memory accesses. As an example, if the first pairs of bits are both zero, then the calculation of the first two butterflies is avoided. Thus, the enable signals for the memories are not generated. Finally, in the third and last phase, the PCU generates the enable signals for the butterfly units.

FPGA resources	*FFT core*	*Pruning Engine*
Logic Combinational Elements	1057	113
Register	1811	157
DSP blocks	4	0
Memory Size	128 Kbits	11 Kbits

Table 1: Synthesis results of the FFT core and pruning engine on an Altera STRATIX V FPGA.

4.3 Hardware implementation

The pruning engine together with the FFT core were prototyped on a Stratix V GS FPGA (device: 5SGSMD5K2F40C2N) in order to evaluate the overhead of resource utilization introduced by the pruning engine. The synthesis was carried out utilizing Quartus II (release v12.1). Beside area overheads, power measurements were carried out to evaluate the efficiency of the introduced pruning engine. Details about power consumption are given in Section 5. The FFT core is able to perform FFT of different sizes up to 2048-point. Table 1 summarizes the resource utilization for the FFT core and the pruning add-on block. As shown by the resource utilization figures, the pruning engine introduces a logic and memory overhead of about a 10%. In the Table 1 the memory size implemented in the architecture are reported.

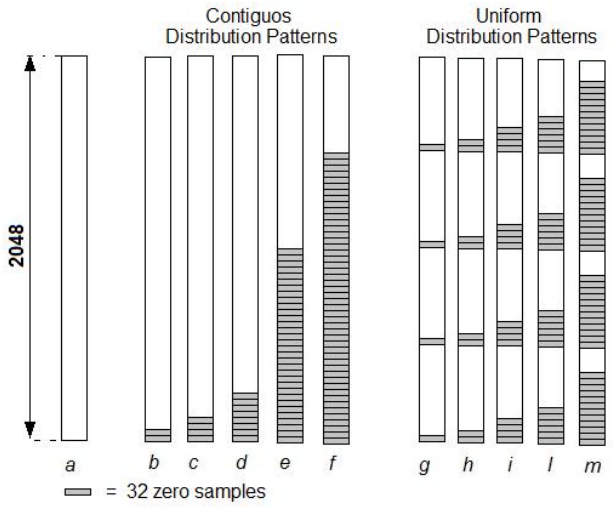


Figure 7: Analyzed input distribution patterns for the evaluation of the FFT pruning engine.

5. RESULTS

The joint system composed of FFT core and pruning engine was evaluated through RTL simulations. In particular, the computation of a 2048-point FFT was considered. Different levels of pruning and different pruning distributions were considered. Figure 7 presents the different input distributions analyzed. The first pattern (pattern *a* of Figure 7) is the reference case where the input sequence contains no zero-inputs. The remaining patterns are divided into two classes: the first considers a scenario where the transmitters delete different contiguous sequences of samples (patterns *b,c,d,e,f*), while the second sub-set shows a uniform distribution of zeros patterns (patterns *g,h,i,l,m*). These distribution topologies represent typical working scenarios in cognitive radio applications, as presented by IEEE 802.22 standard [16]. The estimations of power consumption were performed through the power analyzer tool provided by Altera Quartus II. To obtain more accurate estimations, switching activity files were generated for each one of the analyzed input distribution patterns. Table 1 collects the performance of the pruned FFT architecture in terms of power consumption. Moreover, the Table collects the relative saving obtained via the utilization of the pruning engine in relation to the reference pattern *a*. The reduction of the power consumption mainly depends on two factors: the number of zero samples, and the distribution topology used. For the patterns *b, c, d,* and *f*

increasing the length of the zero sequences leads to higher power savings. Instead, for the patterns belonging at the Uniform Distribution topology, the percentage of the power saved is lower than in the first scenario, because the pruning has a reduced impact on the radix-2 algorithm. In particular, in presence of low pruning levels (e.g. pattern *g*) no power saving is achieved: the advantage of the pruning is compensated by the power consumption of the pruning engine.

Pattern	*Total Power Consumption*	*Power Saved*
a	182 mW	-
b	146 mW	19.8 %
c	144 mW	20.9 %
d	137 mW	24.7 %
e	116 mW	36.3 %
f	106 mW	41.8 %
g	183 mW	-
h	177 mW	2.7 %
i	162 mW	11.0 %
l	138 mW	24.2 %
m	128 mW	29.7 %

Table 2: Estimation of the power consumption of the proposed system for the different operating conditions analyzed.

6. DISCUSSION OF ACHIEVED RESULTS

The results achieved indicate a significant power savings introduced by the pruning engine, leading to a more efficient implementation of the demodulation block in NC-OFDM application scenarios. Moreover, the proposed pruning engine is designed as an add-on block and therefore could be integrated into different FFT cores than the one utilized in this research work. The introduction of the pruning engine does not affect the computation time of the FFT core. On one hand, this design choice limits the power saving achieved (more efficient dynamic power management techniques could be applied), on the other hand it maintains the original specifications of the FFT block, without the need to redesign the interface between the FFT core and the host System on-Chip.

Compared to the power efficiency achieved in [7] the proposed pruning engine offers a reduced power saving. However, the work presented in [7] is strictly tided to FPGA implementations and therefore its application domain might be limited, while the pruning engine proposed in this article is not tided to any particular implementation technology.

In [15] the proposed FFT architecture is not tied to any particular technology either and was synthetized with standard cells libraries. The power consumption is reduced of 22.3% when moving from a 30% to 90% pruning ratio. No power figures are given for an un-pruned scenario. However, the relative power savings achieved by the pruning engine proposed in this article point out a more efficient implementation of the pruning with power saving factors up to 41.8%.

7. CONCLUSIONS

This research work presented an FFT pruning engine as an adds-on block for FFT core architectures. The pruning engine improves the power efficiency of the FFT core without significantly modifying its internal structure and leaving unaltered the FFT core interface. Synthesis and simulation results have shown that for a 10% increase in resource utilization the pruning engine is able to deliver power savings up to 40% in medium high pruning scenarios. Comparisons with related works have shown the competitiveness of the proposed pruning engine.

Future works will cover two directions: 1) ASIC implementation of the proposed pruning engine and 2) a merged implementation of the FFT core and pruning engine, in order to enable the utilization of dynamic power management techniques, such as dynamic frequency and voltage scaling to further improve the power efficiency of the system.

ACKNOLEDGEMENTS

The work was financially supported by the Graduate School in Electronics, Telecommunications and Automation (GETA) and Academy of Finland under contract #258506 (DEFT: Design of a Highly-parallel Heterogeneous MP-SoC Architecture for Future Wireless Technologies). Research grants were received from the Tuula ja Yrjö Neuvo Foundation, the Nokia Foundation, the Ulla Tuominen Foundation and the Tekniikan Edistämissätiö, which are all gratefully acknowledged.

10. REFERENCES

[1] J. Markel, “FFT pruning,” IEEE Transactions on Audio Electroacoustic, vol. 19, no. 4, pp. 305–311, 1971.

[2] H. V. Sorensen and C. S. Burrus, “Efficient computation of the DFT with only a subset of input or output points,” IEEE Transactions on Signal Processing, vol. 41, no. 3, pp. 1184–1200, 1993.

[3] D. Skinner, “Pruning the decimation in-time FFT algorithm,” IEEE Transactions on Acoustic Speech, Signal Processing, vol. 24, no. 2, pp. 193–194, 1976.

[4] R. G. Alves, P. L. Osorio, and M. N. S. Swamy, “General FFT pruning algorithm,” in Proceedings of the 43rd IEEE Midwest Symposium on Circuits and Systems, vol. 3, 2000, pp. 1192–1195.

[5] R. Rajbanshi, A. M. Wyglinski, and G. J. Minden, “An Efficient Implementation of NC-OFDM Transceivers for Cognitive Radios,” in Proceedings of the 1st International Cognitive Radio Oriented Wireless Networks and Communications Conference, 2006, pp. 1–5.

[6] R. Airoldi, F. Garzia and J. Nurmi, "Efficient FFT Pruning Algorithm for Non-Contiguous OFDM Systems," in Proceedings of the 2011 Conference on Design and Architectures for Signal and Image Processing (DASIP), 2011, pp.144-149.

[7] C. Vennila, C.T.K. Palaniappan, K.V. Krishna, G. Lakshminarayanan, Ko Seok-Bum, "Dynamic partial reconfigurable FFT/IFFT pruning for OFDM based Cognitive radio," in Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS), 2012, pp. 33-36.

[8] Yihu Xu; Myong-Seob Lim, "An efficient design of split-radix FFT pruning for OFDM based Cognitive Radio system," in Proceedings of the 2011 International SoC Design Conference (ISOCC), 2011, pp.368-372.

[9] Chao-Ming Chen, Chien-Chang Hung and Yuan-Hao Huang, "An Energy-Efficient Partial FFT Processor for the OFDMA Communication System," IEEE Transactions on Circuits and Systems II: Express Briefs, vol.57, no.2, pp.136-140, 2010.

[10] J. Becker, M. Hubner, G. Hettich, R. Constapel, J. Eisenmann, and J. Luka, "Dynamic and Partial FPGA Exploitation," in Proceedings of IEEE, vol.95, no.2, pp.438-452, 2007.

[11] Federal Communication Commission, “Spectrum policy task force report,” ET Docket No. 02135, Tech. Rep., 2002.

[12] J. Mitola, and J.G.Q. Maguire, “Cognitive radio: making software radios more personal,” IEEE Personal Communications, vol. 6, no. 4, pp. 13–18, 1999.

[13] Q. Zhao and B. M. Sadler, “A Survey of Dynamic Spectrum Access,” IEEE Signal Processing Magazine, vol. 24, no. 3, pp. 79–89, 2007.

[14] J. Takala and K. Punkka, “Butterfly Unit supporting Radix-4 and Radix-2 FFT,” in Proceedings of 2005 International TICPS Workshop on Spectral Methods and Multirare Signal Processing (SMMSP), 2005, pp. 47-54.

[15] In-Gul Jang, Zhe-Yan Piao, Ze-Hua Dong, Jin-Gyun Chung and Kang-Yoon Lee, "Low-power FFT design for NC-OFDM in cognitive radio systems," in Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS), 2011, pp.2449,24.

[16] C. Cordeiro, K. Challapali, D. Birru, and N. Sai Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in Proceedings of the First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2005, pp.328-337.

[17] D. Revanna, O. Anjum, M. Cucchi, R. Airoldi and J. Nurmi, “A Scalable FFT Processor Architecture for OFDM Based Communication Systems” to appear in Proceedings International Conference on Embedded Computer Systems (SAMOS), 2013.

Assessing Performance of Software Defined Radios on Multicore Hardware

Nick Green (University of Illinois at Chicago; Chicago, Illinois, USA ngreen21@uic.edu)
Ugo Buy (University of Illinois at Chicago; Chicago, Illinois, USA; buy@cs.uic.edu)
Redge Bartholomew (Rockwell Collins; Cedar Rapids, Iowa, USA; rgbartho@rockwellcollins.com)

Abstract—Multicore hardware is now ubiquitous in society. While Moore's Law predicts that the number of transistors on a chip will double every two years, its effect on the speeds of single core processors have levelled off. To work around this limitation, hardware designers are producing chips with an increasing number of cores. This allows vast increases of processing on a single chip without increases in clock speeds.

However, this extra processing power does not come for free. There are significant design issues that need to be taken into account, such as parallelisation and synchronisation issues. In some cases there can be bottlenecks that extra cores will not help overcome. Here we empirically assess the potential performance changes when running SDR applications on multicore platforms. We conclude that the performance of SDRs can in fact benefit from multicore hardware; however, various factors may have an adverse effect on potential performance gains.

1 Introduction

Computer systems have seen many transformations since their inception. A relatively recent, albeit significant, evolution has been with respect to the CPU. Traditionally, CPUs have been based on a serial pipeline where an application is executed sequentially. This has given a somewhat easy approach to coding applications. For a developer, it is straightforward to see that one instruction leads onto another, giving a predictable flow to the program.

Over the years, increases in system performance have come from the ever increasing speeds of the CPU. Hardware manufacturers have increased clock speeds by increasing the number of transistors on the CPU while also decreasing their overall size [1]. Moore's Law [2] has applied whereby performance has roughly doubled every two years. This has worked for decades, however, since the mid-2000s there has been a plateau in the performance gains that have been achieved by using this method due to power and heating constraints. As mentioned elsewhere [1], it is possible to increase performance using this method, but processors would need to double in size while offering marginal gains.

However, this does not necessarily mean that Moore's Law has become invalid. Since it was apparent that the traditional method of performance increases would not last forever, hardware manufacturers have been devising new ways of getting the most out of a CPU. The method that has been most successful over recent years has been the multicore CPU.

1.1 Multicore CPU

The multicore CPU is a change from the traditional single core processor in that rather than making a single pipeline faster, it enables multiple pipelines with slightly lower performance to be utilised at once. The optimal outcome, for instance, is that performance may double when doubling the number of cores on a given processor. While this is generally good news [3], it also means that the free lunch for software developers is clearly over. Developers have had the privilege of creating software with the assumption that hardware would increase in performance over the foreseeable future with no effort on their part. An application that runs slowly on a given hardware platform may run at an adequate speed when the very same application is executed on a faster CPU. This is no longer the state of affairs for developers today.

When extra cores are provided on a CPU, the application has to know how to utilise them. This is the biggest effect when programming single vs. multicore systems. In order for the application to take advantage of these hardware capabilities, it must know what parts of the system can be executed in parallel. This can be achieved with the concept of concurrent programming. A single application can make a choice at a particular time when it can divide processing between multiple units of execution, such as a thread. Thankfully, it is at this point that the hardware level of abstraction hides the handling of the cores. Therefore, software need only know concurrency at a software threading level in order to exploit the full performance of multicore hardware.

Existing software can also take advantage of multicore hardware. If an application makes use of multiple threads, which can be a desirable design decision in single core environments, then each of those threads may utilise the multicore feature of the CPU. An issue for existing software is how far it uses concurrency and asynchronous processing. As stated by Amdahl's Law [4], if an application has a particular portion that cannot be parallelised, the ill-effects of this bottleneck will persist no matter how many processors are allocated to the application.

Various factors may limit the performance gains that can be obtained by parallelising software for deployment on multicore hardware. For instance, some I/O operations are serial, meaning that each operation cannot be effectively speeded up, although it is sometimes possible to carry out multiple operations in parallel with each other (e.g., by reading a file while updating a user display). In addition, data dependencies can

prevent effective parallelisation. Data dependencies happen, for instance, when a unit produces data which is later consumed by another. In this case, the units affected cannot be executed in parallel, and so on.

1.2 Software Defined Radio

The software defined radio [5] is the application of choice in our research. SDRs are an incredibly flexible way of programming a radio for use under multiple frequencies, bands and codecs. SDRs take much of the custom hardware out of a radio and instead implements hardware functionality in software on a general purpose host, such as a standard desktop PC. This host receives a transmission from a hardware radio and processes it natively. The host is the flexible part of the equation as it is possible to vary its software (e.g., the operating system), hardware and CPU architecture, while still allowing a vast number of software radio components to be utilised at low cost.

1.3 Research Questions

Here, we detail our investigation into performance changes that occur in SDR systems when used on a multicore platform. Through this research we should be able to answer the following questions:

1) Can an existing SDR benefit from execution on multicore hardware?
2) Is it possible to extend SDR functionality to make use of multicore features?
3) Are there cases when multicore adversely affects performance? and ultimately
4) Is it worth using multicore hardware for SDR applications?

Our empirical investigation will evaluate SDR performance under different circumstances. In Section 2, we give an overview of software defined radios and the software used to drive them. Sections 3 and 4 detail the platforms that we used for our experiments with OSSIE and GNU Radio, respectively. In Sections 5 and 6, we present our empirical results with OSSIE and GNU Radio. Finally, in Section 7 we give our conclusions on these results and attempt to answers the research questions above.

2 Software Defined Radio Environment

There are many different options available to users who wish to use a software defined radio. However, there are two main components; the hardware radio and the software host. The first step for us in our research was to decide what environment we should use to carry out our experimentation. A popular choice for this is the versatile USRP [6] from Ettus Research [7]. The USRP is a modular system that can take a number of daughter-boards, each providing different features such as varying bands and receiving/transmitting capabilities. In our case, we wanted to perform reception within the FM frequency range; thus, we opted for the WBX daughter-board with the USB enabled USRP1, which houses the daughter-board.

There are also a few choices that can be made for the software component of the environment. We evaluated two systems in depth. First, we looked at the Joint Tactical Radio System (JTRS) [8] compliant OSSIE [9], which gives a highly modular SDR software solution based on a component graph architecture. Another software framework that we considered is GNU Radio, which like OSSIE is also a graph based system but is not JTRS compliant. It is apparent that the graph based architecture is a popular choice for developing streaming based applications, just like some non-SDR related frameworks such as some DirectX APIs [10]. Even though both OSSIE and GNU Radio are architecturally similar, we chose to focus our investigation to GNU Radio due to the maturity of the platform, also considering that OSSIE uses some components of GNU Radio in order to act as an interface to the USRP. Most of the empirical results that we report later on are based on GNU Radio, however, we will also report some of our preliminary results obtained with OSSIE.

3 OSSIE Experimental Setup

In our research, we empirically studied the effects of multicore on software defined radio usage. Our first set of experiments involved the OSSIE framework. Since our aim was to investigate in depth real-world applications with GNU Radio, we ran a smaller set of experiments at a more abstract level with OSSIE.

The OSSIE framework is built upon a graph architecture just like GNU Radio; there are many similarities between the two frameworks. One key difference is that OSSIE is JTRS compliant, which in turn makes it highly modular with a CORBA [11] interface between components. In terms of system resources, a significant difference is that OSSIE will then spawn each graph component in a different OS process, whereas GNU Radio spawns components in different threads (i.e., light-weight processes) within the same process.

We performed some preliminary experiments with OSSIE in order to get a feel for applications within the domain. The simple set of experiments that we performed was a pure simulation of a component graph. Using a dummy source component from the OSSIE toolbox (TX Demo), empty packets would be injected into the system at a predefined interval of 1 ms. The goal of our experiments was to determine with how long it would take for a packet to make the transition to a null sink component (RX Demo), also provided by OSSIE. By imposing a variable load on a set of simulation components, we could then see the effects of the workload relative to packet throughput. Furthermore, we investigated different graph configurations, in particular, serial and parallel graphs, where our workload is either performed in a sequential or a parallel pipeline. The hypothesis here is that since each component within an OSSIE graph is executed in their own process, splitting workload among components should yield an improvement on a multicore platform, where an improvement is denoted by a decrease in time for a packet to make its transition from the source node to its destination.

Since this set of experiments was only concerned with the OSSIE graph, we did not use any SDR hardware. For the

software host, we selected a dual-core PC (with an Intel Core i3 CPU) to perform the experiments with Ubuntu 10.04 (32-bit), which is the latest compatible version of Ubuntu that is supported by OSSIE 0.8.2.

4 GNU RADIO EXPERIMENTAL SETUP

In our experimentation with GNU Radio [12], we investigated in detail more real-world systems using the USRP antenna. The primary metric for our study is concerned with the all important concern of signal integrity. We created a test framework using GNU Radio where we could easily modify various parameters and system configurations in relation to the software host. By methodically varying such parameters, we identified properties that could affect overall system quality. For example, if introducing more threads into the system introduced gaps into the audio playback from the radio, we could deduce that the introduction of threads in this situation is not desirable for the application. This is very applicable to our work since there are several parameters that are related to multicore hardware, such as process affinity and the number of software threads within the application. On the one hand, process affinity defines the number of cores that can be assigned to a given application. For instance, by setting an application's process affinity to one, we can force an application to run on a single core. Software threads define units that can be executed in parallel within the context of the same OS process. Threads can be executed simultaneously on different cores, if the affinity of the application containing the cores is greater than one.

We ran our experiments with GNU Radio on a multicore processor, with the Intel Core i7 CPU, which includes four hardware cores with one hyper-thread per core. As we want to use GNU Radio, a reliable host operating system is a Linux-based OS; in our case we chose the 32-bit variant of Ubuntu 12.10.

4.1 Test Framework

The test framework is the basis of our experimentation. GNU Radio offers us a very modular graph-based architecture. The aim of a GNU Radio graph is to transform an input into an output via a series of transforms. An example of a very simple graph is one that contains two components. Figure 1 shows such a graph, where a source component generates a sine wave, which will then pass the signal to an output component which in turn renders it to the PC's speaker. A more complex graph could be made to receive, say, FM radio from a USRP, as can be seen in Figure 2. Such a graph would consist of an input node which interfaces with the USRP hardware, passes through numerous transform nodes by performing operations such as demodulation, with the output then sent to the speaker.

Our experiments used this kind of architecture to create our testing framework. This was done by measuring the amount of data that is output to the speaker. Since SDRs are real-time systems, there has to be a steady flow of data throughout the graph. If there were any blocks in the graph that temporarily stop data flow, samples would be dropped at or near that point

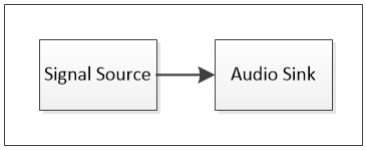


Fig. 1. GNU Radio Companion graph that plays back a test signal.

of time. In the case of the previously mentioned graphs, the source components would skip the generation of samples. This could be heard by the user as simple sound jitters. The greater the bottleneck, the more disrupted the output was.

However, to get a more accurate reading we did not use audible sound quality, but instead counted the samples that made it through the graph. This was done by counting the number of samples per second that passed through a custom component which is placed just before the output component. It is through this sample rate that we determined if degradation occurs, potentially due to a bottleneck in the pipeline. Through our experimentation we varied parameters and observed how these parameters correlated with this sample rate.

We were able to develop two different modes of operation with this framework:

- Manual mode—Here we used GNU Radio Companion [13] to present a user interface (Figure 3) to the investigator, which contains various sliders that vary the available parameters. We can specifically vary the following parameters: (1) the total number of software threads executing in our GNU Radio custom component; (2) The amount of simulated workload imposed on the various cores; and (3) The frequency at which we receive an FM signal. This is an ideal way of quickly gauging the effects of each parameter on system performance. Output can easily be heard via degradation of output.
- Automatic mode—For a deeper analysis, it is also possible to automate parameter modifications via a configuration file. Samples are taken at set intervals (in our setup, 10 samples of 2 second duration) which will indicate the throughput of the system. This throughput gives an insight into the overall system performance.

An important aspect of the framework is our custom experimentation component which not only acts as a hook in getting timing data out of the system, but also as a way of inducing some system-specific logic into the workflow. This is a useful way for a developer in the real-world to see how their code reacts to multicore parameter modifications. A use case of this methodology is to implement a feature within a component that performs a transform, for example, encryption of the signal or multiplexing of two signals. Each case could introduce their own parameters into the equation which could be exposed via our test framework, allowing a systems engineer to change parameters and see how performance would be affected in their application. In our case, we performed some general processing with variable workload in a so-called *Processor Block*.

The Processor Block performs the simple operation of

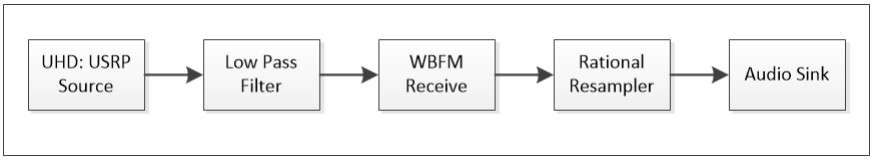


Fig. 2. GNU Radio Companion graph for FM receiving and playback.

batch in place floating point division, where the workload is correlated to the number of division performed. However, this operation is designed to utilise multiple cores via threading. When a packet is received from the radio, the workload is shared among a variable number of threads by creating a thread with the job of performing the workload function an assigned number of times (i.e., n floating point divisions).

The hypothesis here is that as we change the number of threads, active cores, and workload, we should see some changes in performance. If performance increases when we introduce more threads and cores into the component, we can conclude that multicore can improve performance in SDR applications. We conducted several experiments varying parameters and component graphs to see the effects that these have on overall system performance.

Serial1 (iterations)	Serial2 (iterations)	Mux (iterations)	Serial4 (iterations)	2 Cores (ms)	1 Core (ms)	Perf. Delta%
1	1	1	1	1.20	1.20	1.00
1	1	1	1000	109	110	0.99
1	1	1	5000	548	548	1.00
1	1	5000	5000	547	1095	0.50
1000	1000	5000	1000	655	876	0.75
1000	1000	1	1	110	220	.050
1000	1000	1	5000	548	781	0.71
5000	5000	5000	5000	1084	2153	0.50

TABLE 1
Timing results for packet transfer in OSSIE.

5 OSSIE Emperical Results

Table 1 shows the results that were collected from the graph in Figure 4. It details the number of iterations of workload that were applied in several cases, along with the amount of time it took for a packet to make its transition from the source component to the sink on a dual core system. Times for each component shown in Figure 4 are expressed in milliseconds.

When a light workload was applied to the system (Row 1), packets would make their transition from the source to the sink is roughly 1.20 ms. As workload is applied to any custom filter, the transition time increases as expected. Row 2 shows results when the custom filter has a slight workload of 1000 floating point divisions. This gives an expected slow down in the transition speed since the packet will be blocked waiting for the processing to complete. The benefits of multicore came into play when there was a high workload applied to multiple components. Row 4 shows two filters in serial having a workload of 5000 divisions applied to both. The time for a packet to go through the system takes twice as long when running on a single core. This will be due to the parallel processing of the workload in different components occurring in different processes, thus different cores. This hypothesis is backed up by the recurrence of this in other reported cases. We conclude that in our OSSIE simulations, we were able to achieve near-optimal or optimal (linear) speed-ups when using multiple cores.

6 GNU Radio Empirical Results

To further our investigation into multicore performance for SDRs applications, the next line of research was to switch to a more mature SDR framework, GNU Radio. In this section we detail our findings on an experiment by experiment basis, while showing the process that we followed to gain extra information for further clarification.

6.1 How do Multiple Threads Affect Performance

One of the key questions to answer here is how the addition of threads within an application affects overall system performance. To answer this question we looked at the performance difference between using a single thread for all our simulated workload and using up to four threads. Two experiments were executed—one using a test signal graph and another using an FM reception graph.

Figures 5 and 6 show the overall performance of the SDR application for the two different graphs. Performance can be measured based on signal quality, ranging from 0% (no signal) to 100% (full signal). The signal quality was calculated based on the number of dropped packets from the audio stream. If a stream is transmitting 96,000 samples per second, if only 90,000 are processed in one second, we deem the signal quality to be 90,000/96,000 $\simeq$ 84%. When there is a drop in quality we conclude that the workload applied to the graph at that point is greater than the host can handle. From here, all GNU Radio results will plot the trend of workload (x-axis) vs. signal quality (y-axis) based on a set of parameters, such a thread count and active cores.

Figure 6 shows how multiple threads affect performance of the host. Maximum throughput is achieved regardless of the parallelism preformed in the Processor Block whilst the workload is low. However, when the workload reaches about 10,000 iterations of the simulation function (a sequence of floating point divisions), performance is affected. There is a gradual decline in performance for the single threaded Processor Block; however, the effects of the workload are not seen

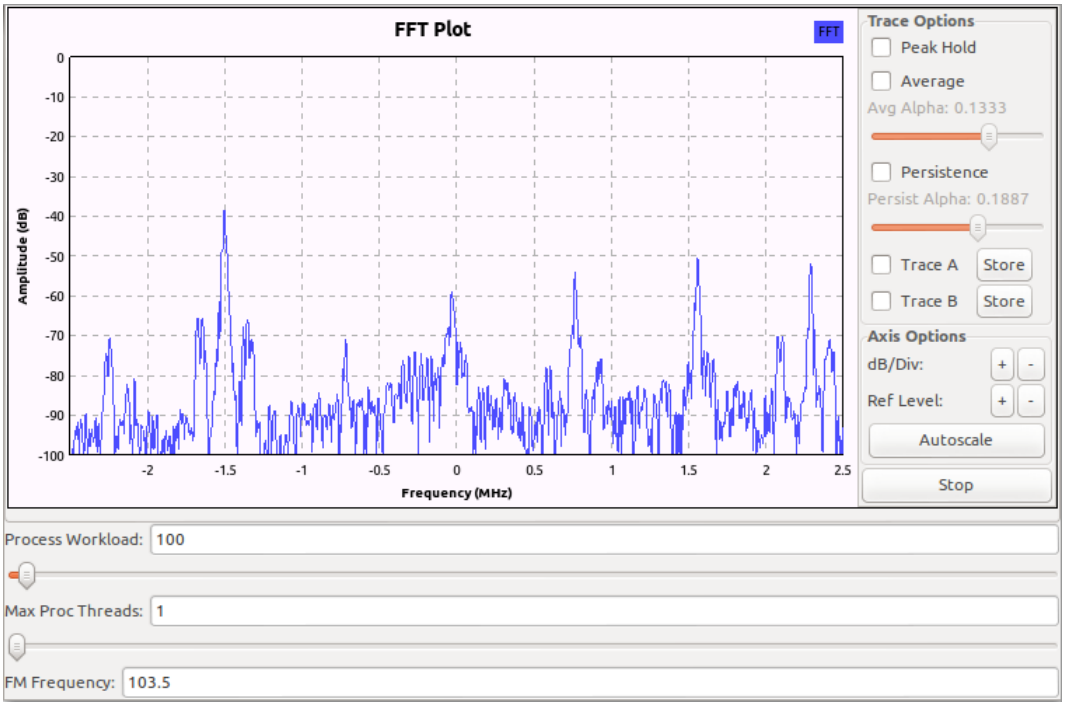


Fig. 3. Test Framework GUI in manual mode.

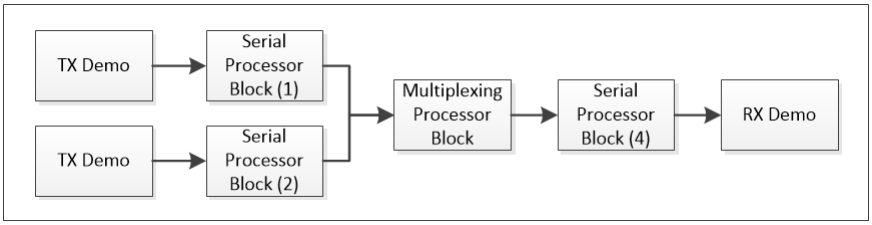


Fig. 4. Multiplexing OSSIE simulation graph.

until later for a multi-threaded block. This clearly shows that the introduction of threads gives an immediate performance improvement over a single threaded solution. This could be attributed to a couple of properties, such as the workload being distributed over different cores. Another possibility is that there could be a blocking call within the Processor Block. We discovered through additional studies that the cause was the former, that is, the way the workload is distributed among the various cores. We discuss this issue further later on when we report about experiments in which we restricted the use of cores.

Another point to mention from this experiment is to what extent performance improved when more threads were introduced. There is a notable improvement when using more threads but the degree of these improvements vary with the workload. For example, at the 40,000 division workload point, doubling the number of threads almost doubles performance (22% vs. 44% vs. 82% measure of quality depending on whether 1, 2, or 4 cores are used). This is near the maximum theoretical speed-up that can be achieved by increasing the number of resources. However, looking at another data point (e.g., 60,000 floating point divisions), the performance difference is still substantial, but not quite as dramatic with a performance of 16% vs. 32% vs. 52%. This could indicate that while introducing threads can improve performance to near optimal levels, there may be other limiting factors.

As for the test signal graph in Figure 5, we see a similar pattern with a different scale (since there is more work in processing an FM signal than generating a sine wave). When the number of utilised threads doubles, the performance also doubles. Even though this is not even using the SDR but a self contained software experiment, it is useful to see how the parallelisation is not strictly dependent on the source.

6.2 Threading on a Single Core

We have already shown that introducing threads into the SDR host application may dramatically improve performance. However, it could be the case that software threads are responsible for the improvement and not the introduction of multiple hardware cores (which may be the case if blocking calls are

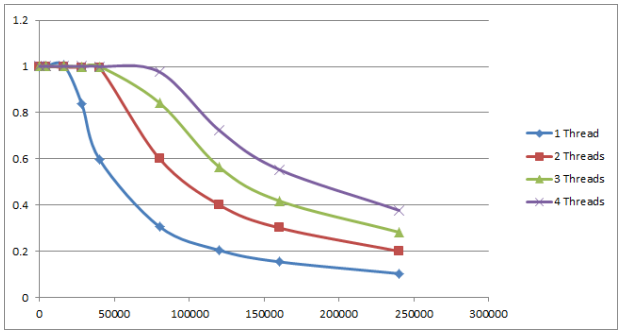


Fig. 5. Results showing throughput while playing back a test signal.

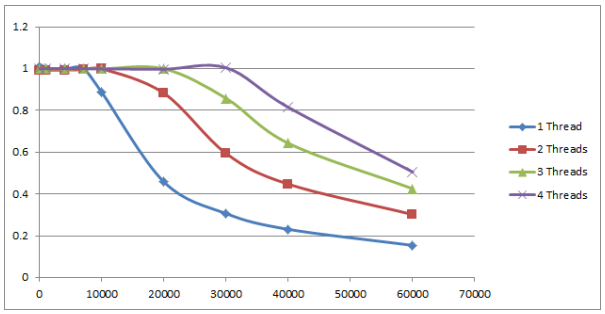


Fig. 6. Results showing throughput while playing back FM on multiple cores.

made). Our next set of experiments used the FM receiver on a single core. This was accomplished by setting the process affinity (locking a process to a subset of hardware cores) of GNU Radio Companion to run on only one of the available hardware cores.

Figure 7 shows the performance when only one core is active using the floating point division simulation function. This helps prove our earlier claim that the original performance gains were due to threads being allocated to different cores. Evidently, when there is only a single core running the SDR application, there are no performance increases when multiple software threads are introduced. This shows that in this system multicore is responsible for the vast performance increases and that software threads simply unlock this potential.

There are two more points of interest here. First, Figure 7 also shows that even if threads are introduced, performance will not be degraded. This is quite an important point since code that can be written using threads can have instant benefits when run on a multicore system. In such cases, there is the potential for only an insignificant penalty when run on a system with fewer cores.

While software threads can unlock the potential gains that can be accomplished by multicore hardware, the indiscriminate addition of software threads may also hurt performance. Threads do incur an overhead, such as context switching. If a significant number of additional threads were to be created and used, these overheads may accumulate into poor performance.

We also compared the data in Figure 7 to the multicore execution when using a single threaded Processor Block (Figure 6). Even though in both cases only one software thread is used in the Processor Block, there is higher performance when using multicore. This could be down to multi-threading elsewhere in GNU Radio. GNU Radio, by default, will run each of its nodes in a different thread. So even if the Processor Block is running in a single thread, the USRP Source block will run in a different thread, as well as each other node in the graph. This is an important finding as we inherit some implicit multi-threaded properties from the framework that can be used by the multicore CPU.

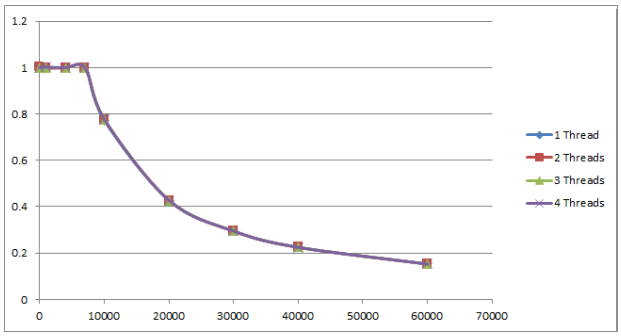


Fig. 7. Performance playing back FM on one core.

6.3 The Single Threaded Framework

It was already seen that performance gains can be instantly made in GNU Radio by utilising more cores due to the GNU Radio framework. One interesting option GNU Radio gives to the user is to select different component schedulers in the framework. In particular, a user can choose whether the framework should run using a multi-threaded scheduler (MTS) or single-threaded scheduler (STS), where the multi-threaded mode will execute each component in a different thread, while the single-threaded scheduler runs all node operations in a single thread. It is mentioned that a reason for using the single-threaded mode instead is when there is a need to use hardware, such as graphics cards via Cuda [14], to do some processing in a component.

When each components runs in a single thread, this does not affect the execution of other threads created within that component. Thus, for some insight into this framework property, we ran experiments with a single-threaded Processor Block using either a multi-threaded or single-threaded scheduler. Figure 9 shows the difference between using the two framework modes. This figure shows the performance under both framework modes as well as 1 vs. 4 active cores.

The results shown in the figure are quite interesting. The only case in which there is a significant drop in performance is when the single threaded scheduler is used on multiple cores. Performance actually improves when the scheduler is locked to a particular core. Looking at the Linux system monitor during execution, for the single-threaded scheduler on multiple cores there is quite a bit of variation over what cores the application is run on. This is in contrast to the case when there is only one

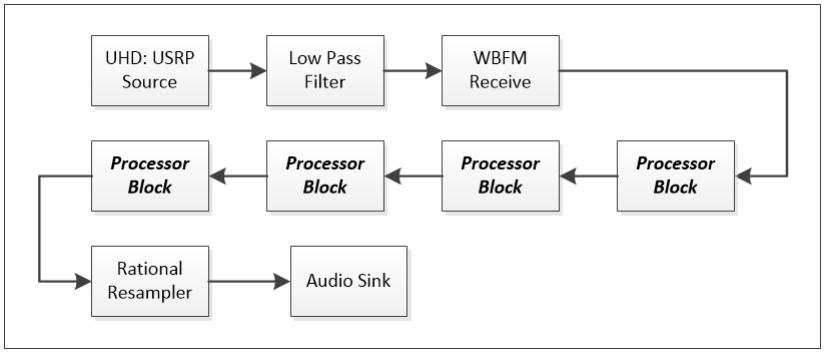


Fig. 8. GNU Radio Companion graph with multiple Processor Blocks.

core for it to use; there is a single flat line of usage on one core. Therefore, even though an application can get instant benefits moving to multi-core, if it is in nature a serial application, like GNU Radio when using the single-threaded scheduler, multicore may actually hurt performance. (See Figure 7).

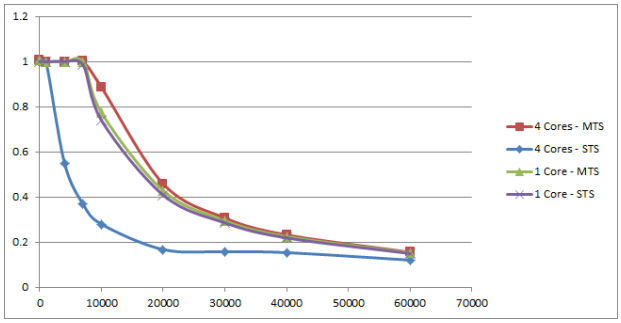


Fig. 9. Comparison of scheduler types

6.4 Inter-Component Parallelism

Another variation of the system configuration is to use the inherent multi-threaded framework given in GNU Radio to exploit multicore architectures. This works similarly to earlier experiments using OSSIE. Since GNU Radio runs each component in its own thread (in multi-threaded mode) it should be possible to share some workload between components, more specifically, components of the same type. Therefore, instead of running a single Processor Block with four threads, we should achieve similar performance by chaining together four Processor Blocks each using a single thread. Such a GNU Radio graph can be seen in Figure 8. Intuitively, this model has many benefits. The major benefit here is it mitigates concurrent development from the component developer to the framework. Since developing multi-threaded code is time consuming, this requires additional expertise on part of the developer, and can introduce concurrency issues. Thus, any mitigation is desirable. This extra encapsulation also reinforces good software engineering principles. Another benefit is reuse, as this component can then be used as many times as necessary via a graph modification instead of editing the code, introducing more threads within a component. One potential downside to this assignment of threads to components is the overhead that may be imposed by the framework, as there will be a cost in performing these extra transitions within the graph.

The results of splitting workload between blocks can be seen in Figure 10. There is the expected improvement when spreading workload over more blocks, similar to intra-component parallelism (i.e., parallelism obtained by creating threads within the Processor Block). However, it can also be seen that there is a notable decrease in performance when moving from the intra-component to inter-component parallelism. As we already mentioned, some performance degradation is expected, however, there are cases when there is a 20% difference in performance. Even though inter-component parallelism has yielded some improvements, there are cases where performance is significantly worse than their intra-component counterparts. We further investigate this issue below.

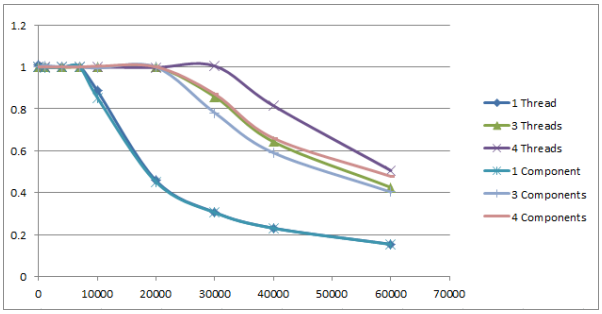


Fig. 10. Performance of intra vs. inter-component parallelism.

6.5 Isolating Inter Component Bottleneck

In the previous subsection, we reported a test case where inter-component performance differs from intra-component performance. We decided to investigate this discrepancy using performance profilers. Some popular profilers, such as gprof [15], can be fairly intrusive. They can require rebuilding the entire executable which will effectively change the code, hence changing program flow and performance. This may

not be desirable, especially when we are trying to identify bottlenecks in a large project.

A less intrusive option is OProfile [16], which has also already been used to successfully profile GNU Radio [17]. Instead of relying on instrumentation, OProfile polls the OS kernel to identify what applications, and functions within those applications, are being executed at any one time. The polling rate is fine enough to get a reliable idea of where the program is spending most of its time whilst having little effect on the overall runtime. The results of profiling our two different experiments, which can be seen in Table 2, were surprising.

Profiling has shown a couple of differences between the two parallelism methods. The first increase is in libpangoft2, which is a text rendering library [18]. This seems like an unlikely culprit in this case and may also be a side-effect of the additional text messages printed by GNU Radio Companion during a period of stress on the graph. The other increase is within the component itself. To be more specific, there was a slight increase in the amount of time the application spent within the simulation function of the Processor Block when aiming for inter-component parallelism. This may suggest that the floating point division is more efficient when executed in different threads within the same component rather than in separate components.

6.6 Differing Simulation Processing

Our findings on Processor Block performance introduced yet another variable into the pipeline; the type of simulation that is being executed in the Processor Block. So far we have found that there is more time spent within the simulated processing function when using inter-component threading. We were to discover what processing, if any, promotes this behaviour. Therefore, we created several different test cases that exercise different software properties to see what operations can affect performance. We experimented with various kinds of blocks differing in the kind of simulation operation that they perform. The kind of operations that we considered included the following:

- In place integer addition and subtraction
- Memory copy with floating point division
- Recursive function calling
- Large integer array reversal (using a temporary variable)
- Large integer array reversal (using XOR swap)
- In place integer addition and subtraction
- An empty nested loop
- Large dynamic memory allocation
- In place floating point addition and subtraction
- In place floating point division

In each of these cases we observed the difference between performance with respect to inter vs. intra-component parallelism. Interestingly, it was found that it is possible to achieve similar performance with either methodology if the Processor Block exhibited certain computational properties.

We consider our previous simulation block, in place floating point division, as our baseline (Figure 10). In this case it is clear there is a performance difference of up to 20% at times. However, when we perform an in place integer addition and subtraction in our experiments (Figure 12), the differences are eradicated with both inter-component and intra-component parallelism yielding similar results. This suggests that there are some inherent properties with the original Processor Block (performing floating-point divisions) that do not work as well when encapsulated in other components. This could be due to contention for shared hardware resources, such as arithmetic units.

We also observed that some operations showed a general degradation between components. Reversing an array of 1 Million integers (Figure 11) also showed a slight degradation of around 5%; however, the method of performing the swap (using a temporary variable or exclusive or) did not make a substantial difference.

There were further issues when experimenting with array reversal methods in intra-component mode. As can be seen in Figure 11, the previous optimal gains were achieved when the number of threads was doubled from 1 to 2. However, the biggest discrepancy seen here from previous intra-component experiments is with the marginal improvement from 3 to 4 threads. The improvement is less than 5%, which is far less than the gains seen in the addition/subtraction experiments. This could hint towards memory or cache related bottlenecks. In such a case, multicore does not offer the gains that we'd hope for due to contention for other hardware resources.

The most significant difference was with respect to dynamic memory allocation. Figure 13 shows a simulation function which allocates a 100MB block of RAM on a system (which has 6GB of RAM). Any kind of concurrent allocation has a significant performance slow-down of over 80%. This indicates that memory allocation in a concurrent environment may cause an exorbitant performance penalty. We have not looked further into what parameters affect performance; however, using a test framework like the one we developed may be of use to a systems developer. They could add various memory allocation parameters to the framework, such as allocation length, and see how this value affects their system. After all, there may be other limiting factors, such as the memory bus on the motherboard and the speed of RAM.

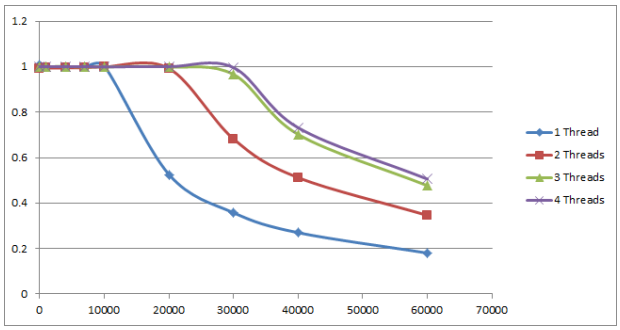


Fig. 11. Performance of integer array reversal between threads.

Profile for intra-component parallelism		Profile for inter-component parallelism	
% of Time in Symbol	**Symbol Name**	**% of Time in Symbol**	**Symbol Name**
88.9296	ProcessorBlock::InPlaceFpDivision(int)	89.6622	ProcessorBlock::InPlaceFpDivision(int)
4.3517	fcomplex_dotprod_sse	3.7241	fcomplex_dotprod_sse
1.4393	libpangoft2-1.0.so.0.3000.1	1.2072	no-vmlinux
1.0082	no-vmlinux	1.1898	libpangoft2-1.0.so.0.3000.1
0.8701	python2.7	0.8408	python2.7
0.7006	no-vmlinux	0.6902	no-vmlinux

TABLE 2
Output of O-Profile, showing system wide CPU utilisation.

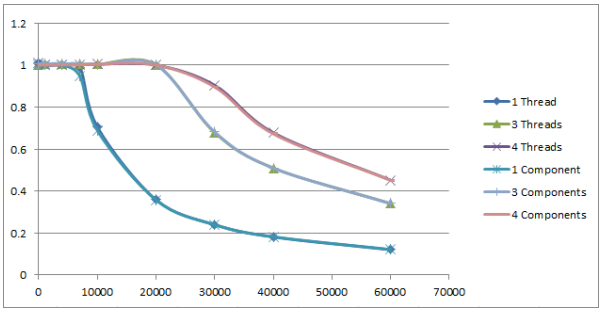


Fig. 12. Performance of inter/intra component parallelism with an addition/subtraction simulation function.

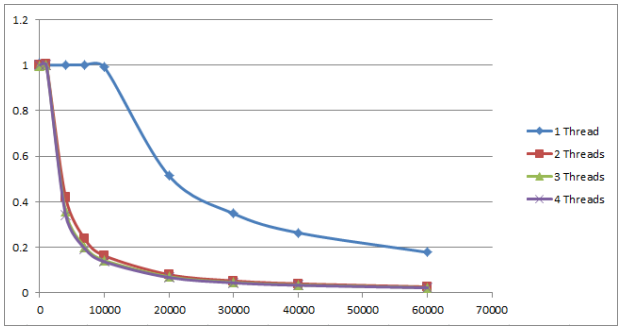


Fig. 13. Performance of concurrent dynamic memory allocation.

7 Conclusions and Future Work

Our empirical study leads to several conclusions. First, we have demonstrated a test framework that shows the conditions upon which multicore CPUs and software threading can be effective in an SDR environment. We feel that such a simple interface can be of use to any concurrency project and the ideas here can be domain independent and platform agnostic. The automated approaches here have been invaluable in our experimentation, but use of this approach could reach farther afield, for example, performance testing could be used during the normal automated software build and test cycles. It would be easy to use frameworks similar to ours in test cases; issues could be reported if metrics exceed threshold values. Even for the developer, our framework can be a useful tool in order to tune their use of concurrency. A simple example here is to avoid large concurrent memory allocations.

Secondly, it is clear from our results that going multicore can have instant improvements. In the case of GNU Radio, its use of multiple threads makes it an ideal candidate as switching from one core to four increased performance significantly. In general, parallelising over multiple cores gave good increases in performance. However, if the application is a serial one, going to multicore may actually hurt performance. Another setback for multicore is when there is the need to interact with other resources, such as RAM, which can significantly degrade performance.

Finally, even though splitting work between cores can work for a particular case, a system's architecture may come into play. We showed some evidence that splitting work between components, which should be promoted as good software design, may have an adverse effect on system performance.

Here, we have shown some results into our investigation on the effects that multicore CPUs have within the domain of software defined radios. Some of the results were conclusive that a vast, near maximal increase in performance can be achieved, but only in certain circumstances. A possible future line of research is to find out what operations can be parallelised, what cannot be parallelised, or maybe to find out what different operations can be performed in parallel in order to get maximum performance out of the system.

8 Acknowledgements

We would like to thank our sponsor, Rockwell Collins, for funding research into this project.

References

[1] D. Geer, "Chip makers turn to multicore processors," *Computer*, vol. 38, no. 5, p. 1113, 2005. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1430623

[2] R. R. Schaller, "Moore's law: past, present and future," *Spectrum, IEEE*, vol. 34, no. 6, p. 5259, 1997. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=591665

[3] H. Sutter, "The free lunch is over: A fundamental turn toward concurrency in software," *Dr. Dobbs Journal*, vol. 30, no. 3, p. 202210, 2005. [Online]. Available: http://www.info2.uqam.ca/ tremblay/INF5171/Liens/sutter.pdf

[4] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities," in *Proceedings of the April 18-20, 1967, spring joint computer conference*, 1967, p. 483485. [Online]. Available: http://dl.acm.org/citation.cfm?id=1465560

[5] F. K. Jondral, "Software-defined radio: basics and evolution to cognitive radio," *EURASIP Journal on Wireless Communications and Networking*, vol. 2005, no. 3, p. 275283, 2005. [Online]. Available: http://dl.acm.org/citation.cfm?id=1088725

[6] M. Ettus, "Universal software radio peripheral," *Ettus Research, Mountain View, CA*, 2009.

[7] ——, "Ettus research, LLC," *Online information on USRP board). http://www. ettus. com*, 2008.

[8] R. North, N. Browne, and L. Schiavone, "Joint tactical radio system-connecting the GIG to the tactical edge," in *Military Communications Conference, 2006. MILCOM 2006. IEEE*, 2006, p. 16. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4086476

[9] M. Robert, S. Sayed, C. Aguayo, R. Menon, K. Channak, C. Vander Valk, C. Neely, T. Tsou, J. Mandeville, and J. H. Reed, "OSSIE: open source SCA for researchers," in *SDR Forum Technical Conference*, vol. 47, 2004. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.138.7486&rep=rep1&type=pdf

[10] M. D. Pesce, *Programming Microsoft DirectShow for Digital Video, Television, and DVD*. Microsoft Press, 2002. [Online]. Available: http://dl.acm.org/citation.cfm?id=863598

[11] R. Otte, P. Patrick, and M. Roy, *Understanding CORBA (Common Object Request Broker Architecture)*. Prentice-Hall, Inc., 1995. [Online]. Available: http://dl.acm.org/citation.cfm?id=230247

[12] E. Blossom, "GNU radio: tools for exploring the radio frequency spectrum," *Linux journal*, vol. 2004, no. 122, p. 4, 2004. [Online]. Available: http://dl.acm.org/citation.cfm?id=993251

[13] J. Blum, *GNU Radio Companion*, 2011.

[14] C. Nvidia, *Programming guide*, 2008.

[15] S. L. Graham, P. B. Kessler, and M. K. Mckusick, "Gprof: A call graph execution profiler," *ACM Sigplan Notices*, vol. 17, no. 6, p. 120126, 1982. [Online]. Available: http://dl.acm.org/citation.cfm?id=806987

[16] J. Levon and P. Elie, *Oprofile: A system profiler for linux*, 2004.

[17] T. W. Rondeau, "Application of artificial intelligence to wireless communications," Ph.D. dissertation, Virginia Polytechnic Institute and State University, 2007. [Online]. Available: http://scholar.lib.vt.edu/theses/available/etd-10052007-081332/

[18] O. Taylor, *Pango: internationalized text handling*, 2001. [Online]. Available: https://old.lwn.net/2001/features/OLS/pdf/pdf/pango.pdf

DESIGN OF BASIC RECEIVING FUNCTIONS FOR AN SDR BASED QPSK BASE BAND DEMODULATOR FOR A RECONFIGURABLE DATA-LINK

Angelo Manco (a.manco@cira.it)
Ivan Iudice (i.iudice@cira.it)
Vittorio Ugo Castrillo (v.castrillo@cira.it)
Embedded Systems and Communications
CIRA – Italian Aerospace Research Centre, Capua, Caserta, Italy

ABSTRACT

An Unmanned Aerial Vehicle (UAV) needs "talk" to the ground control station (GCS) about mission and service information like the flight plan and telemetry. This paper focuses on the design and implementation of the base-band basic receiving functions, for a Quadrature Phase Shift-Keying demodulator case study, as independent modules of a complete Reconfigurable Data-Link (RDL) for UAVs communications. A model-based approach and Software Defined Radio (SDR) paradigm are used for the design. The implementation will be executed on Field-Programmable Gate Array (FPGA) based hardware.

1. INTRODUCTION

UAVs could operate in different scenarios that require multi-standard operations and environment link adaption, so flexibility and scalability are critical. These requirements can be addressed developing a Reconfigurable Data-Link (RDL), which is a system able to provide different communication functions without changing hardware. To design such a system, a model-based approach and SDR paradigm are investigated.
Model-based approach is helpful to obtain modular architectures with stand-alone blocks which can be used afterwards for other systems designs simply changing their interconnections. Therefore one of the main goals is the development of a library of blocks (each one with its own close hardware model) useful to implement a RDL.
Pushing the analog-to-digital conversion as close as possible to the antenna, SDR based systems allow to realize multiple radio features on the same hardware platform and to adapt the designed communications system to the radio context changes. In this panorama, Field-Programmable Gate Arrays (FPGAs) show good balance among computational power, configurability and design costs, so they are suitable to realize the advanced signal processing tasks necessary in the physical layer of a SDR based system.
On the basis of the above considerations, this paper focuses on the design of basic receiving functions as part of a complete RDL using a model-based approach. Since Phase Shifting-Keying modulations play an important role in the area of aeronautical communications, in particular in the satellite field, the receiving functions are investigated for a differential encoded QPSK demodulator case.
Digital signal processing can be executed at high or intermediate frequencies, but baseband processing is much more suitable by hardware performance and cost point of view; therefore in this work all operations are done on in-phase (I) and on quadrature (Q) components coming from a down conversion stage that is not described here. Synchronization is taken into account and feedback solutions are adopted. In particular, both timing and carrier recovery are implemented as described in sections 3 and 4.

2. BACKGROUND

A QPSK modulated signal is described by the following equation:

$$s_m(t) = \sqrt{\frac{2E_s}{T}} g(t) \cos\left(2\pi f_c t + \frac{\pi}{4}(2m-1)\right) = \sqrt{\frac{2E_s}{T}} g(t) \cos\left(\frac{\pi}{4}(2m-1)\right) \cos(2\pi f_c t) + \ldots - \sqrt{\frac{2E_s}{T}} g(t) \sin\left(\frac{\pi}{4}(2m-1)\right) \sin(2\pi f_c t) \quad (1)$$

where E_s is the symbol energy, T the symbol period, f_c the carrier frequency, $g(t)$ is the pulse shape and $m=1,2,3,4$ depends on the transmitted symbol.

An ideal SDR communication system has Analog-to-Digital Converter (ADC) pushed close to the antenna, so there is minimum or null signal conditioning prior to sampling. Today technologies allow us to have something very close to the ideal case only for not too high frequencies

and data-rates. However we consider the case where the down-conversion stage is in the analog domain and the output of this stage is the complex baseband signal represented through its I/Q components, neglecting the noise component:

$$I_m = \sqrt{E_s}\, g(t)\cos\left(\frac{\pi}{4}(2m-1)+\phi+\Delta f\right)$$
$$Q_m = \sqrt{E_s}\, g(t)\sin\left(\frac{\pi}{4}(2m-1)+\phi+\Delta f\right) \quad (2)$$

where ϕ, Δf are phase and frequency offset due to mismatch between carrier and local oscillator of the quadrature mixer in the down conversion stage. The I/Q signals are sampled by an ADC, with a sampling period T_s equal to T/N, where N is the number of samples for symbol, and then processed by the demodulator. A Raised Cosine $g(t)$ pulse shape filter mitigates the Inter Symbol Interference (ISI) over a band limited channel. A Decision Directed Phase Locked Loop (DD-PLL) is adopted to track the initial phase and the frequency drift. In order to resolve the phase ambiguity introduced by the DD-PLL, a differential encoding is adopted and so the data are mapped to the phase shifts of the modulated carrier.

thanks to a Decision-Directed Phase-Locked Loop. It is constituted by a Phase Error Detector (PED), a Loop Filter (LF), a NCO and a Rotational CORDIC, as depicted in Fig. 1.

The phase error θ_k, computed by the PED, is showed in the following equation

$$\theta_k = sign(I_k)\cdot Q_k - sign(Q_k)\cdot I_k \quad (3)$$

The PED works at the sampling rate $1/T_s$ and its output is accumulated after filtering operation. PI loop filter constants K_p and K_i can be calculated considering a certain dumping factor ζ and an equivalent noise bandwidth B_n expressed in terms of maximum frequency drift Δf_{max} loop can track, using the following formula [8]:

$$B_n \approx \frac{\Delta f_{max}}{2\pi\sqrt{2}\zeta} \quad (4)$$

Anyway tracking errors are proportional to the equivalent noise bandwidth, so the optimum choice for the right value of B_n has to be based on a trade-off between fast acquisition and good tracking.

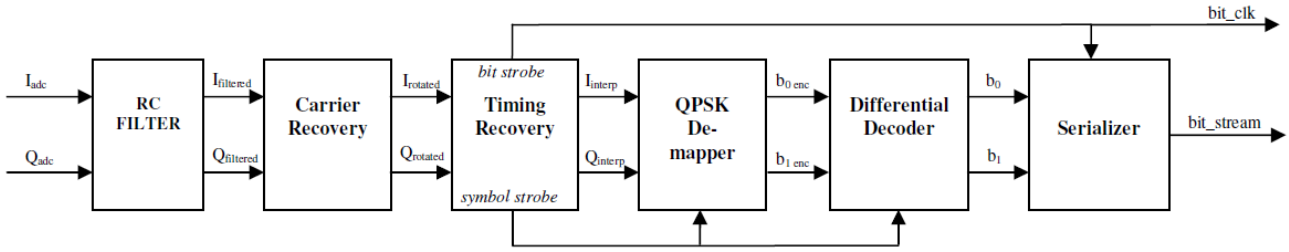


Fig. 1 – Functional architecture of the QPSK Demodulator

3. CARRIER RECOVERY

The phase and the frequency offset are acquired and tracked

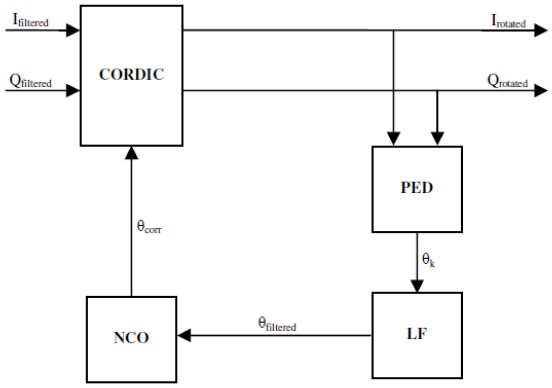


Fig. 2 – Carrier Recovery Unit

The filtered and accumulated error is used to adjust the phase and the frequency of the complex baseband signal by means of rotation of the vector with components I/Q. For this purpose, a CORDIC (COordinate Rotation DIgital Computer) algorithm in *rotational mode* is used [1]. It is an iterative procedure that only needs of adders, barrel shifters and a look-up table with off-line calculated values and, therefore, well suited for low cost FPGAs where dedicated multiplier resources are limited. Since the algorithm doesn't converge for input vectors having phase (in module) greater than 99°, a range extension is used [2].

4. BIT TIMING RECOVERY

The sampling period of analog-to-digital converter in the down-conversion stage is not aligned to the symbol period, i.e. it is asynchronous with symbols, and so timing errors affect demodulation performance in presence of noise. Real-world symbol pulse shapes have a peak in the centre of the symbol period. Sampling the symbol at this peak means to have the best signal-to-noise-ratio and it will mitigate interference from other symbols. There are possible several approaches to fix this issue depending on the characteristics of the receiver. We consider the case where the timing correction is all digital. This means that it is not possible to adjust the sampling frequency/phase of the ADC (it is fixed), passing from digital domain to analog one. A feedback scheme is adopted and in particular a Non-Data-Aided (NDA) timing recovery technique is used [3][5][6]. The structure is depicted in Fig. 3.

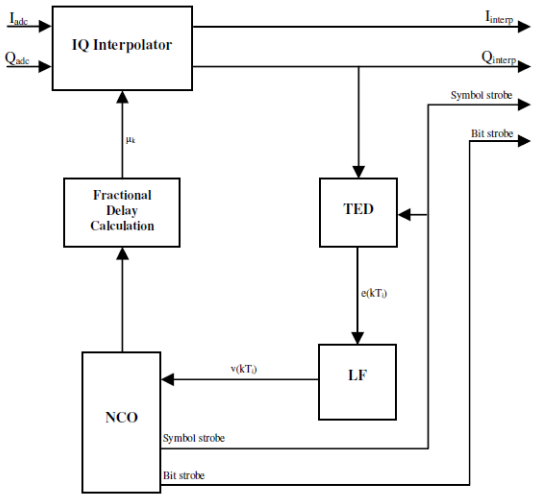


Fig. 3 – Bit Timing Recovery Unit

Let are {..., (k-1)T, kT, (k-1)T, ...} the desired interpolation instants and {..., (n-1)T_s, nT_s, (n-1)T_s, ...} the ADC sampling instants with n ≈ kN. The right sample for the k-th symbol is computed by means of an interpolation of ADC I/Q output samples on the basis of the fractional delay μ_k and on the last L samples available from ADC, where μ_k is the distance between the desired optimum sample (i.e. the sample at the time kT) and the closest ADC preceding sample (whose index, indicated with mk, is named base-point index) and L depends on the interpolation order. For example, the interpolated Q component sample is:

$$q(kT_i) = q((m_k + \mu_k)T_s) \qquad (5)$$

A Lagrange polynomial interpolation is considered. Since the Lagrange coefficients are expressed as a polynomial in μ, the interpolation is implemented as Farrow structure [8]. This approach allows the fractional delay to vary in a continuous way. A Timing Error Detector (TED) is used to estimate the error between the right sampling instant and the current one. In particular the Gardner algorithm is used. It is based on finding zero crossing between two consecutive symbols. It uses two samples per symbol and it generates the following error signal:

$$\hat{e}(kT_i) = \left(y(kT_i) - y((k-1)T_i)\right) \cdot y((k-\tfrac{1}{2})T_i) \qquad (6)$$

In this case the TED works at the sampling rate $1/T_s$ and so its output is decimated according to the base-point index m_k to extract the right error.

The TED output is filtered in order to have the control signal for the timing adjustment. A first order Proportional-plus-Integrator (PI) loop filter is used for this purpose. In particular it consists of two paths. The proportional path multiplies the error signal by the proportional gain K_p. It is able to track out a phase step error. An integral path multiplies the error signal by the integral gain K_i in order to track out a ramp phase error (i.e. a frequency error). Constants K_p and K_i can be calculated considering a certain dumping factor ζ and an equivalent noise bandwidth B_n using a procedure similar to the one of a PLL design as explained in [8]. Note that the gain of the linearized TED and the NCO gain must be known to make the above procedure.

The Interpolation Control provides the base-point index and the fractional delay on the basis of the filtered error signal $v(kT_i)$. It is performed by a Numerically Controlled Oscillator (NCO) [5]. In this case, the NCO is constituted by an accumulator, operating at the sampling rate $1/T_s$, that overflows every N samples. The overflow in the k-th period indicates the base-point m_k and a symbol strobe is generated. This signal gives the timing for the right decimation at the de-mapper and at the differential decoder that must work at symbol time. Another strobe signal at a rate of two times the symbol rate, that gives the timing for the serializer unit, is necessary. This signal, named bit strobe, must depend on the m_k and so it is derivate from it.

The loop filter output $v(kT_i)$ adjusts the amount by which the accumulator increments. The fractional delay is computed using the content η of the accumulator as showed in the following equation

$$\mu(m_k) \approx N \cdot \eta(m_k) \qquad (7)$$

The fractional delay is updated at symbol rate.

4. DEMODULATION

The demodulation is done by means of a de-mapper and of a differential decoder. Then a serialization process take place in order to have a serial output stream at a frequency of 2*symbol rate. The timing, as seen previously, is provided by the Interpolation Control Unit. De-mapper works at the symbol rate and it compares the input I/Q with thresholds (depending on the source code chosen, e.g. Gray code) and outputs two parallel bits. The DD-PLL introduces a phase ambiguity multiple of $\pi/2$ due to the rotational symmetry of the constellation. In order to resolve this issue, a differential encoding is adopted and the data are mapped to the phase shifts of the modulated carrier instead of the phase. Therefore a differential decoding, implemented through a look-up table, takes place. The last process is relative to the transition from a parallel data to serial one, whose timing is "slaved" to the previous units one.

6. DESIGN FLOW

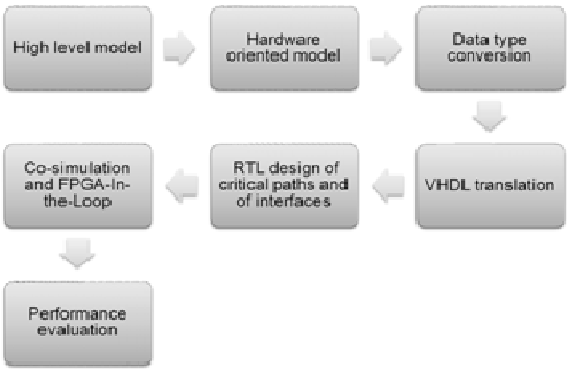


Fig. 4 – Design flow for QPSK Demodulator

QPSK base-band demodulator design is composed by the following steps. First of all the demodulator is high-level modelled and tested in MATLAB/SIMULINK environment. Required functions are implemented using standard maths functions (as the sine and cosine functions) and all signals are represented with a double data type. In the second step a close hardware model is designed. The general maths functions are calculated with algorithms whose hardware implementations are convenient (for example, vector rotation is performed by CORDIC algortihm). During the third step the data type is changed to fixed-point in order to obtain blocks models structure well suited to get a VHDL code. At this stage, comparing the simulation results of the double data type blocks models and the ones related to fixed-point models, it is possible to determinate the best choice for words and fractional parts lengths, choosing a good trade-off between accuracy and resources utilization. At the fourth step a VHDL code is designed for each block and for the overall system, taking into account the FPGA target. The fifth step is relative to simulation of demodulator VHDL code using SIMULINK and co-simulation tools; Next, an hardware test using signal generator and Bit Error Rate Tester (BERT). In a final step, performance tests will be executed to estimate the demodulator quality in presence of carrier frequency and timing errors, and in presence of noise tracing the BER vs E_b/N_0 curve.

7. SIMULATION

The overall system is simulated in MATLAB/SIMULINK environment. The aim of the simulation is to verify that the demodulator, in particular the synchronization functions, works properly. Timing and carrier errors are considered separately.

The coefficients of the loop filter for the timing recovery are designed considering an unitary damping factor and a single-sideband noise bandwidth Bn of 0.5% of the symbol rate. A PN11 bit sequence is generated through a Linear Feedback Shift Register and used, after modulation, as data input of the system. Whereas the sampling period of analog-to-digital converter in the down-conversion stage is not aligned to the symbol period, a step timing error is firstly considered. No carrier errors are introduced. Fig. 5 and Fig. 6 show the

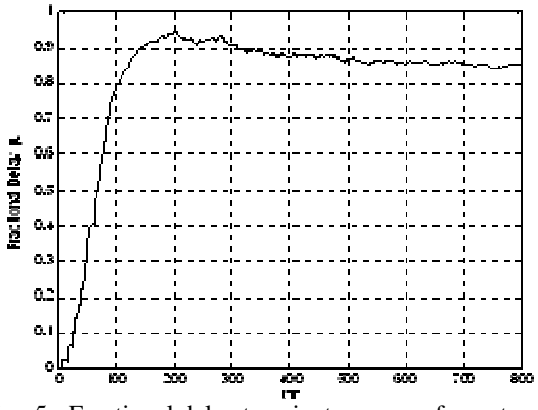


Fig. 5 - Fractional delay transient response for a step timing error

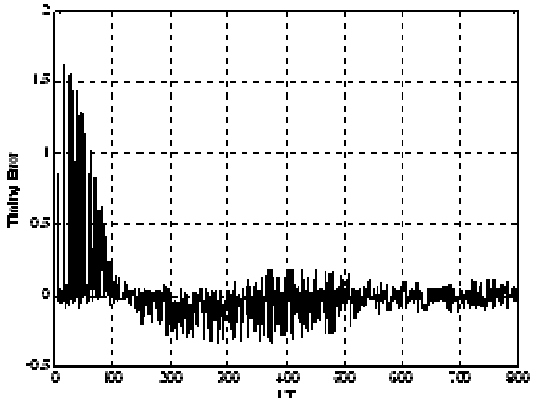


Fig. 6 – TED transient response for a step timing error

transient responses of the fractional delay μ_k and of the TED error signal. The fractional delay μ_k reaches a steady-state value of 0.85 (i.e. the target value) after about 500 symbols. Response to a timing error ramp of 1% of symbol period is shown in Fig. 7 and Fig. 8. The TED error signal goes to

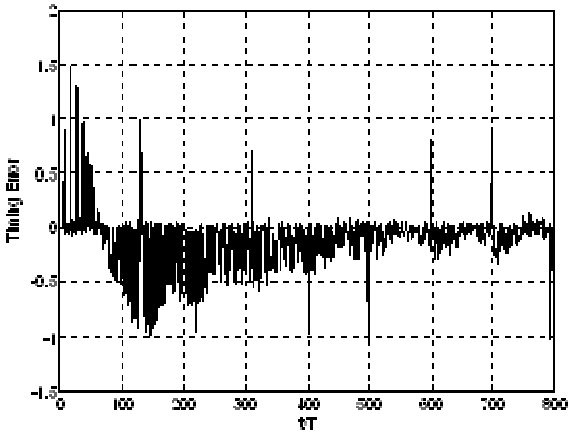


Fig. 7 - TED transient response for a ramp timing error

zero thanks to the loop filter that is capable to track out a frequency error. Because a residual timing error accumulates, the fractional delay μ_k decreases with time [8]. When the accumulated residual timing error exceeds a

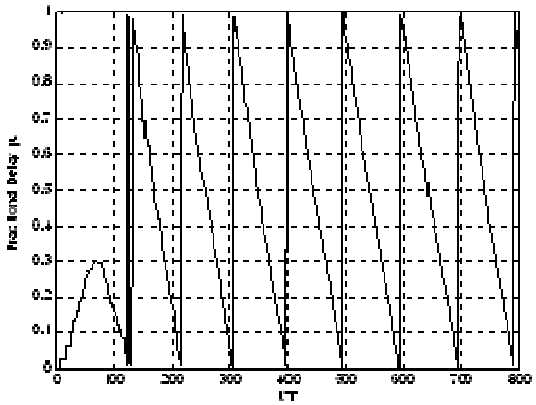


Fig. 8 - Fractional delay transient response for a step timing error

sample period, μ_k wraps around to 1. It happens every 100 symbols, accordingly to the introduced error.

The behaviour of the carrier recovery loop is also examined. The coefficients of the loop filter are designed considering a damping factor $\zeta = 1/\sqrt{2}$ and a single-sideband noise bandwidth B_n of 5% of the symbol rate. the phase error is $\pi/6$ and the frequency error is 0.1% of the symbol rate.

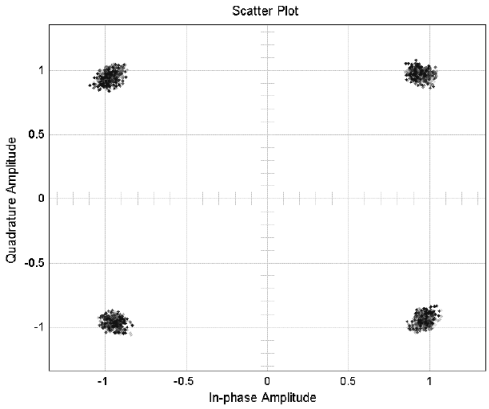


Fig. 9 – Corrected constellation

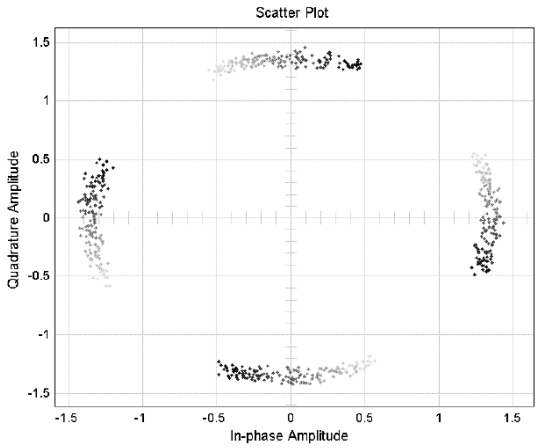


Fig. 10 – Frequency error implies constellation rotation

8. REFERENCES

In this paper, basic receiving functions of a communication system were investigated for a QPSK base-band demodulator case with the aim to create a library of modular blocks that can be used to implement a Reconfigurable Data-Link.

A model-based approach was used in the design flow. In such way the transition from a high level blocks models to a close hardware ones was simplified and the development time was considerably reduced. The obtained models were

tested with software simulation showing a proper operation. The VHDL code will be generated and implemented on low cost FPGA. The demodulator will be tested with laboratory instrumentation.

9. REFERENCES

[1] Volder Jack, "The CORDIC trigonometric computing technique", IRE Transactions on Electronic Computers, vol. 8, no. 3, September 1959, pp. 330-334

[2] Andraka Ray, "A survey of CORDIC algorithms for FPGA based computers", Proceedings of the ACM/SIGDA sixth international symposium on Field programmable gate arrays, Monterey, CA, February 22-24 1998, pp. 191-200

[3] Meyr Heinrich, Moeneclaey Marc, Fechtel Stefan A., "Digital Communication Receivers, Synchronization, Channel Estimation and Signal Processing", vol. 2, Wiley, NY, November 1997

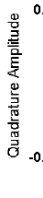
[4] Navjot Singh, "Design and implementation of optimum interpolation filter using Farrow structure", International Journal of Engineering Science and technology, vol. 3, no. 5, May 2011, pp. 4108-4113

[5] Gardner Floyd, "Interpolation in Digital Modems – Part I: Fundamentals, IEEE Transactions on Communications, vol. 41, no. 3, March 1993, pp. 501 – 507

[6] Erup Lars, Gardner Floyd, Harris Robert, "Interpolation in Digital Modems – Part II: Implementation and Performances", IEEE Transactions on Communications, vol. 41, no. 6, June 1993, pp. 998 – 1008

[7] Harris Frederic, Rice Micheal, "Multirate Digital Filters for symbol Timing Synchronization in Software Defined Radios", IEEE Journal on selected areas in Communications, vol. 19, no. 12, December 2001, pp. 2346 – 2357

[8] Rice Michael, "Digital Communications: A Discrete-Time Approach", Pearson Prentice-Hall, Upper Saddle River, NJ, 2009



FULLY RECONFIGURABLE FPGA-BASED COGNITIVE RADIO PLATFORM FOR RELIABLE COMMUNICATIONS

Félix Casado (fcasado@ikerlan.es); Raúl Torrego (rtorrego@ikerlan.es); Aitor Arriola (aarriola@ikerlan.es); Iñaki Val (ival@ikerlan.es);
IK4-Ikerlan, Arrasate-Mondragón, Guipúzcoa, Spain,

ABSTRACT

Cognitive radio and dynamic spectrum access are two of the technologies that enable nowadays the securing of reliable communications. In order to solve some of the problems that appear in wireless communications (e.g. wideband interferences or multipath fading) it is necessary to carry out a complete change of the transmission band and not only small in-band channel changes. It is necessary to ensure that the new chosen band is uncorrelated with respect the previous one, in terms of coherence bandwidth and RF interference. A complete change of band does not only imply the reconfiguration of some parameters in the digital domain of the cognitive-radio, but requires the reconfiguration of the analog part too. On this basis, this paper presents a fully reconfigurable FPGA-based cognitive-radio platform for reliable communications. The platform provides runtime reconfiguration at all levels: FPGA baseband processing, RF front-end and antenna. This enables both in-band channel change and dual-band (ISM 868 MHz and 2.45 GHz) reconfiguration.

1. INTRODUCTION

Software Defined Radio (SDR) [1] and Cognitive Radio (CR) [2] are the technological answer that nowadays fulfils with the requirements that robust wireless communications demand (such as reliability, interoperability, opportunistic spectrum sharing, advanced networking, etc). Dealing with reliable communications, industrial environments are one of the use cases where SDRs or CRs can show their potential due to the harsh conditions present in this type of environment (electromagnetic interferences, metallic objects...). Dynamic Spectrum Access (DSA) is one of the techniques that cognitive radios can use in order to achieve the previous objective. The change of transmission frequency looking for an unused channel has proved to be an effective way of avoiding interferences [3,4]. Unfortunately, the presence of wide band interferences, multipath fading or attenuation-related issues require a change not in the channel but in the transmission band. This change of the transmission band cannot be carried out by just updating some parameters in the digital domain (actually the domain of SDRs or CR) but also requires some changes in the analog domain (RF front-end and antenna). Consequently, this paper continues with the work presented in [3,4] and introduces a fully reconfigurable, FPGA-based cognitive-radio platform for reliable communications, where the use of a reconfigurable antenna allows to carry out the change of the transmission band.

Therefore, the proposed architecture provides runtime reconfiguration capabilities at all the levels of the system, namely: FPGA baseband processing, RF front-end and antenna. The main contributions with respect to the previous works are the design and implementation of a frequency reconfigurable antenna that can switch between the two ISM bands, and the implementation of a cognitive signal detection algorithm based on the spectral characteristics of the signal. Similarly to the previous developments, FPGA dynamic partial reconfiguration and rapid prototyping tools are used in the design and implementation of the digital part of the platform.

2. PLATFORM DESCRIPTION

The designed cognitive radio platform consists of a transmitter and a receiver node that implement a cognitive wireless communication at 1.6 Mbps. The system uses QPSK modulation and is able to change its transmission frequency depending on the availability of the channel. The platform aims to achieve the maximum possible flexibility and reconfigurability at the service of reliable communications. On this purpose, it is reconfigurable at the three main parts that make up each of the nodes: the Virtex 6 FPGA in charge of baseband processing, the RF front-end and the antenna. A block diagram and the implementation of the platform can be seen on Figure 1 and Figure 2 respectively.

2.1. Baseband Processing

Reconfiguration in the FPGA baseband processing is carried out via FPGA dynamic partial reconfiguration [5]. This

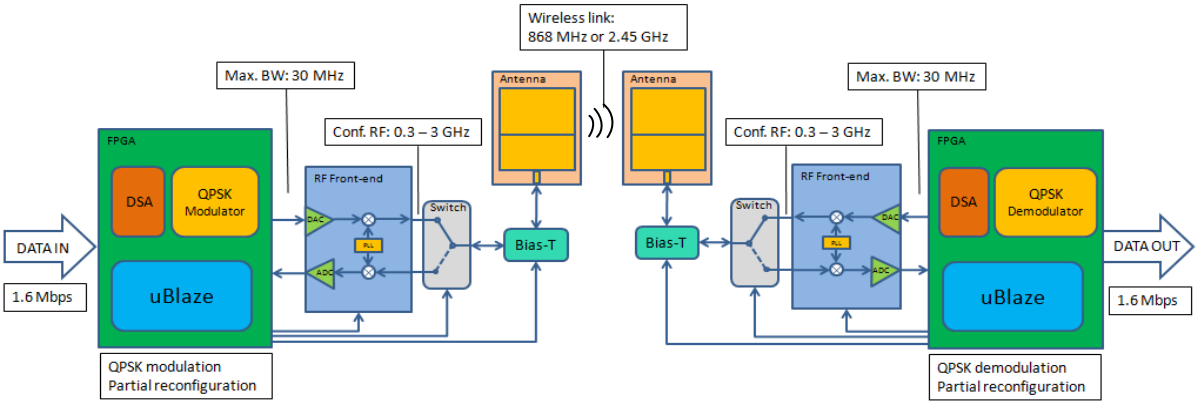


Figure 1: Block diagram of the fully reconfigurable cognitive radio platform

feature enables the change of functionality of part of the FPGA while the rest continues unaltered, which leads to a reduction in the required amount of FPGA resources and consequently in its size. The implemented baseband processing is mainly divided into two tasks, present both in the transmitter and in the receiver node: data modulation/demodulation and DSA (Dynamic Spectrum Access). The first task, in which FPGA dynamic partial reconfiguration is applied, performs the QPSK modulation of the input data in the transmitter and recovers it in the receiver. The DSA task, in turn, looks for channel availability in the transmitter prior to any transmission and performs the search of signal in the receiver. Taking into account that no information about the transmission frequency is provided to the receiver, it carries out an analysis of the spectral characteristics of the received signal in order to separate the valid one from the possible interferences. These signal processing tasks have been hardware-implemented in the FPGA using System Generator [6], Xilinx's rapid prototyping tool. Besides, both FPGAs in the transmitter and the receiver implement a MicroBlaze processor mainly in charge of the control of the different parts of the architecture: dynamic partial reconfiguration, programming of the front-end and control of the antenna. FPGA dynamic partial reconfiguration is used in this implementation in order to perform small in-band frequency changes (< 10 MHz) in the transmission frequency by reconfiguring the digital oscillator that generates the IF frequency; however, it could also be used in order to change any other parameter in the aforementioned baseband signal processing tasks (e.g. modulation, data rate, etc.).

2.2. RF Front-End

A Nutaq Radio420S module has been chosen as reconfigurable RF front-end for the transmitter and the receiver [7]. This module provides a carrier frequency

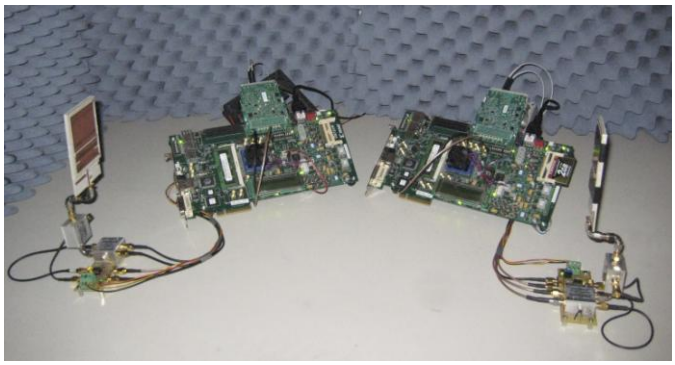


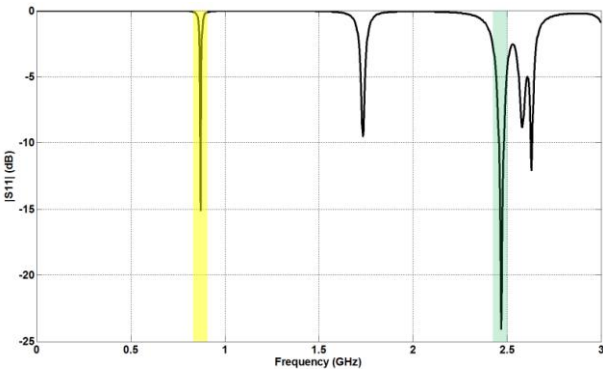
Figure 2: Implementation of the fully reconfigurable cognitive radio platform

between 300 MHz and 3 GHz with a bandwidth between 1.5 MHz and 28 MHz. Its transmitted power can be adjusted between -21.5 dBm and 10 dBm, and its sensitivity ranges between -90 dBm (0.3-1.5 GHz) and -103 dBm (1.5-3 GHz). It should be raised that this module has separate RF transmission and reception ports. Hence, in order to use a single antenna, a RF-switch (Mini-Circuits ZYSW-2-50DR) has been used. This switch has a wideband operation (DC-5 GHz) and is controlled with a TTL signal coming from the main FPGA.

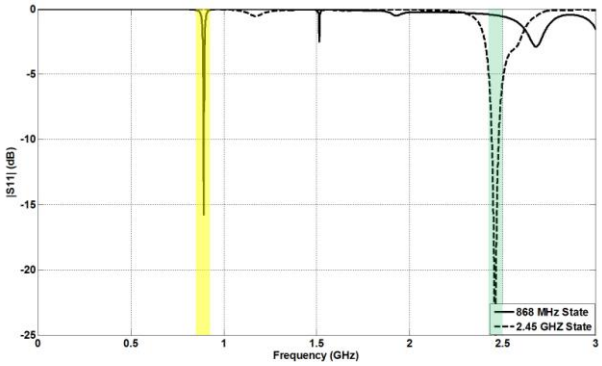
2.3. Antenna

In industrial environments, the antenna is likely to have metallic objects in its vicinity. As it has been studied in [8] a monopole-like antenna exhibits significant variations in its input impedance and in its radiation efficiency when it is placed near a metallic object, so its clearance distance (minimum distance between the antenna and the metallic object where the antenna still works properly) must be determined. Hence, to make the antenna suitable for harsh environments, a microstrip topology has been chosen which shows a great robustness against any object placed under the antenna.

The use of reconfigurable antennas provides extra filtering against undesired signals, compared to the multi-band antennas used in current communications systems (e.g. smart-phones). When using a multiband antenna both bands are covered at the same time, which means that if any interference is detected, even though the whole system changes its operating frequency the interference is received and the front-end has to be able to filter the undesired signal. In Figure 3 a comparison is shown between the return loss of a microstrip patch multiband antenna and a microstrip patch reconfigurable antenna. It can be observed that while the multiband antenna (Figure 3(a)) is receiving the power at both bands, the reconfigurable antenna (Figure 3(b)) receives only the power of the working band.



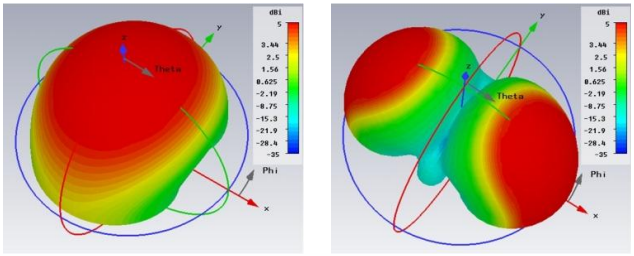
(a)



(b)

Figure 3: Simulated return loss: (a) Multiband antenna (b) Reconfigurable antenna

Moreover, by using microstrip patch reconfigurable antennas, it is possible to keep the radiation features of the different bands alike. It should be taken into account that if the radiation pattern is not kept constant it could happen that the antenna performs a broadside radiation pattern in one band and an endfire radiation pattern in the other. Hence, in the direction where one of the bands has its maximum the other will perform a null of radiation, resulting in a problem



f = 868 MHz f = 2.45 GHz

Figure 4: Simulated radiation pattern of a microstrip patch multiband antenna

to align properly the antennas in order to ensure the correct work of the two bands. In Figure 4 the radiation pattern of a microstrip patch multiband antenna is shown. As it can be observed, at the 868 MHz band it shows a broadside radiation pattern while it exhibits an endfire radiation pattern at 2.45 GHz. In that case, if the antennas are meant to be placed facing one to each other, at 2.45 GHz the antenna will exhibit a null of radiation in the direction where a maximum would be desirable, leading to a decrease in the range of the communication link. However, when using a microstrip patch reconfigurable antenna the radiation patterns between the two bands are kept alike as shown in Figure 5 (b) and Figure 6 (b).

Therefore, a frequency-reconfigurable antenna has been designed to work in two ISM bands (868 MHz and 2.45 GHz). In order to ensure the required bandwidth in both bands, a reconfigurable matching network (RMN) has been designed and added to the antenna. The RMN is a microstrip stub based network and the reconfiguration is handled through PIN Diodes such as in [9]. At the lower band, the antenna has a matching bandwidth of 6 MHz, while at the upper band the bandwidth is 100 MHz. The switching between the two bands is handled by a pair of PIN Diodes. These PIN Diodes are placed in the antenna and they are fed by DC lines. It should be noted that both the RF signal and the DC power are provided by the FPGA and combined with a Bias-T (Mini-Circuits ZFBT-4R2G-FT). Depending on the frequency of operation, the FPGA will provide a DC voltage to the PIN diodes that control the antenna and the RMN states. In Table 1 it is shown the PIN diode's state depending on the operation band.

Table 1: States of the PIN Diodes

		ISM 868 MHz	***ISM 2.45 GHz***
Antenna's PIN Diode	***#1***	***ON***	***OFF***
	#2	***ON***	***OFF***
RMN's PIN Diode	***#1***	***OFF***	***ON***
	#2	***OFF***	***ON***

3. MEASUREMENTS

In this section the measurements of the basic characteristics of the designed system are presented. Table 2 shows the FPGA resources consumed by the baseband digital part of both the transmitter and the receiver of the cognitive-radio platform.

Table 2: FPGA Resource Consumption

	SLICE	Flip-Flop	LUT	BRAM	DSP48
Transmitter	7180 (20%)	13943 (4%)	14410 (4%)	97 (23%)	630 (82%)
Receiver	8834 (23%)	23902 (6%)	22765 (15%)	98 (23%)	368 (47%)

Table 3: Frequency Reconfiguration Time

FPGA	Front-end	Antenna
94 μs	79 μs	53 μs

It can be observed the small form factor of the designed baseband processing algorithms occupying only a 20–23 % of the FPGA (assuming the number of SLICES as the most representative value of the design's size).

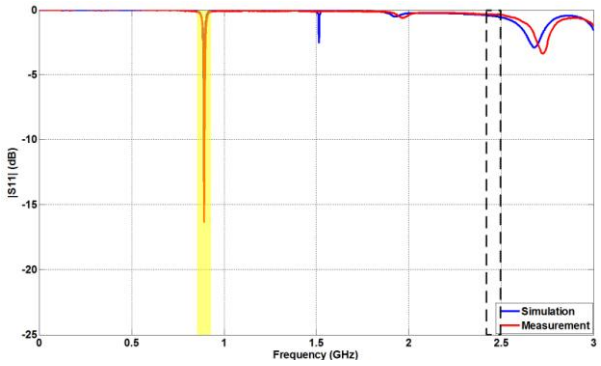
On the other hand, Table 3 shows the time that the different components in the system require for carrying out a frequency change.

Two different cases have to be distinguished though: the in-band channel change and the complete band change. The first case only requires the partial reconfiguration of the FPGA; hence the front-end and the antenna remain unaltered. On the other hand, in case a complete band change is needed, the front-end and the antenna have to be reconfigured and the FPGA remains unchanged in this specific application.

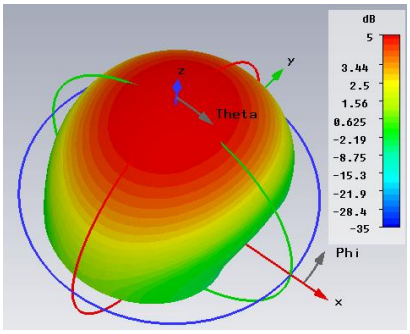
Eventually, Figure 5 and 6 show the return loss and the radiation pattern for the two frequency states (i.e. 868 MHz and 2.45 GHz). In the return-loss pictures, the operational band is colored and the non-operational band is marked with a dashed line. It can be seen that in the working band (e.g. at 868 MHz, Figure 2 (a)) the antenna shows a high degree of impedance matching (low return loss), which means that the vast majority of the received power is delivered to the antenna, while in the other band (e.g. 2.45 GHz at Figure 2 (a)) the impedance matching is very low, so almost all the power is reflected. Regarding to radiation features, it can be observed that the radiation patterns are quite alike in both bands, as it was expected.

4. CONCLUSION

This paper has presented a fully reconfigurable FPGA-based cognitive-radio platform for reliable communications. The proposed architecture provides runtime reconfiguration at all levels (FPGA baseband processing, RF front-end and antenna) enabling both in-band channel change via FPGA

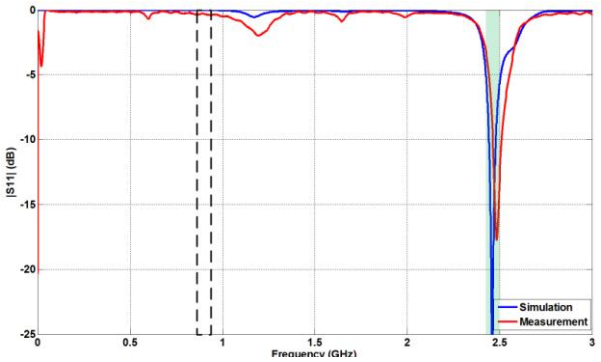


(a)

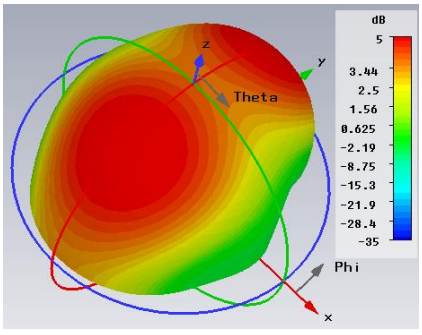


(b)

Figure 5: Reconfigurable antenna at ISM 868 MHz state: (a) Simulated and measured return loss; (b) Simulated radiation pattern at 868 MHz



(a)



(b)

Figura 6: Reconfigurable antenna at ISM 2.45 GHz state: (a) Simulated and measured return loss; (b) Simulated radiation pattern at 2.45 GHz

dynamic partial reconfiguration and dual-band (ISM 868 MHz and 2.45 GHz) reconfiguration. The system demonstrates the feasibility, necessity and benefits of joining reconfigurable components at different system levels in a single architecture in order to secure reliable communications.

5. REFERENCES

[1] Raut, R.D., and Kulat, K.D.: "SDR Design for Cognitive Radio". *4th International Conference on Modeling, Simulation and Applied Optimization (ICMSAO), 2011*

[2] The Software Defined Radio Forum Inc. "Quantifying the Benefits of Cognitive Radio" Version V1.0.0, 2 December 2010.

[3] Torrego, R., Val, I., and Muxika, E.: "OQPSK cognitive modulator fully FPGA-implemented via dynamic partial reconfiguration and rapid prototyping tools". *Proc. Wireless Innovation Forum European Conference on Communications Technologies and Software Defined Radio (SDR'11 – WInnComm – Europe), Brussels (Belgium), 2011*, pp. 142-147

[4] Torrego, R., Val, I., and Muxika, E.: "Small-form-factor cognitive radio, implemented via FPGA partial reconfiguration, replacing a wired video transmission system". *Proc. Wireless Innovation Forum European Conference on Communications Technologies and Software Defined Radio (SDR'12 – WInnComm), Washington, 2012*.

[5] XILINX: 'UG702 - Partial Reconfiguration User Guide', 2011, 13.3 edn., pp. 130.

[6] XILINX: 'UG640 - System Generator for DSP User Guide', 2012, 14.3 edn., pp. 1-424.

[7] Nutaq Radio420S Multimode SDR FMC RF transceiver product sheet, 2012.

[8] Casado, F., Arriola, A., Arruti, E., Parron, J., Ortego, I., and Sancho, J.I.: "2.45 GHz printed IFA on metallic environments: Clearance distance and retuning considerations" in *6th European Conference on Antennas and Propagation (EUCAP), 2012*, pp. 921-924.

[9] Casado, F., Arriola, A., Parron, J., Arruti, E., Ortego, I., and Sancho, J.I.: "Reconfigurable Matching Network for 2.45 GHz printed IFA on metallic environments" *Loughborough Antennas and Propagation Conference (LAPC), 2012*.

JOINT UTILIZATION OF TEMPORAL AND SPATIAL DIVERSITY FOR VEHICULAR SPECTRUM SENSING

Haris Kremo (Toyota InfoTechnology Center, Tokyo, Japan; hkremo@jp.toyota-itc.com);
Onur Altintas (Toyota InfoTechnology Center, Tokyo, Japan; onur@jp.toyota-itc.com)

ABSTRACT

Diversity combining is a well known technique which improves signal detection performance through use of quasi–independently faded signal replicas. The conventional wireless propagation modeling decouples large scale shadow fading from the "rapidly changing" small scale fading and treats them as independent random processes. Using different spatiotemporal scales of fading correlation we decouple time periods of individual and collaborative sensing for the vehicular cognitive networks. We evaluate the proposed sensing scheme by means of computer simulations in rural and urban propagation environments using energy detection for simplicity. On the small scale, scheduling of large number of sensing intervals results in a significant gain from small scale fading, even if collaborating nodes already provide more than a few independent measurements. On the large scale, the ratio between the large scale decorrelation distance and the distance at which the nodes must re–evaluate spectrum availability determines whether collaboration can be substituted by temporal diversity on each sensor.

1. INTRODUCTION

In the prior work [1] we argued that, although it introduces significant challenges, dynamic spectrum access in the vehicular environment is necessary to satisfy growing demand for wireless communications. We also argued that, for highly mobile cognitive nodes, primary spectrum user protection based on geolocation database lookup has significant limitations. Therefore, enhancement of spectrum sensing accuracy remains important area of our research interest.

To that goal, we evaluate application of diversity combining to improve performance of spectrum sensing in the vehicular environment based on the conventional propagation modeling paradigm, which separates dynamics of the radio channel into two spatiotemporal scales. The first one is small scale fading, which describes power fluctuations and delay spread in areas comparable in diameter to the carrier wavelength λ_c. The second one is the large scale fading, which describes combination of the distance related path loss and the slow signal strength undulation. This local mean undulation is frequently, although somewhat arbitrarily, referred to as the shadow fading.

We introduce a two–tier scheduling scheme which exploits diversity gain from both the small and the large scale fading. The most important parameter for design of scheduling is the distance at which the nodes are required to reassess spectrum usage. It is imposed by a regulating authority such as The Federal Communications Commission (FCC), which currently sets it to 100 m [2]. We refer to this distance as the *decision distance*. First, while the vehicles traverse the decision distance, we apply temporal diversity to the small scale fading by repeating sensing intervals with the period larger than its coherence time T_s. Second, depending on the imposed decision distance, the diversity gain from large scale fading can be exploited either through temporal or spatial diversity. When the decision distance is much larger than the decorrelation distance of large scale fading D_l, a single sensor can achieve the same detection performance as multiple sensors spread over the decision distance. However, collaboration must be used to maximize diversity gain if the decision distance is comparable to, or smaller than D_l.

It is well known that the small scale fading gain does not grow linearly with each additional diversity branch [3]. Sufficient separation of sensors on the large scale guarantees sufficient separation on the small scale as well, and the small scale fading gain diminishes as more cars/sensors are added. We simply overcome this by exponentially increasing the number of sensing intervals. This is practically impossible to do in space using multiple antennas. However, it is straightforward to use temporal diversity with cars as mobile sensor. The additional delay to collect samples is limited by the decision distance, which is typically by orders of magnitude larger than the coherence distance of the small scale fading D_s.

To the best of our knowledge, performance of multiple sensors under correlated composite (large and small scale) fading in the vehicular environment, and with regard to the regulatory domain requirements was not previously considered. The benefits of diversity combining in general are, for instance, evaluated in [4]. The tradeoff between temporal and spatial diversity on a large fading scale is

Table 1: Simulation settings and parameters

Environment	Rural	Urban
Shadow fading	mild	severe
standard dev. σ	3 dB	10 dB
decorrelation dist. D_l	100 m	10 m
local area size (m)	10 λ_c	5 λ_c
Small scale fading	LOS, GSM rural	NLOS, GSM urban
tap delays (μs)	0 0.2 0.4 0.6	0 0.2 0.6 1.6 2.4 5.0
relative powers (dB)	0 -2 -10 -20	-3 0 -2 -6 -8 -10
Rice K-factor	1	n/a
Doppler spectra	LOS: Jakes+δ(0.7f_{max}) all other taps: Jakes	all taps: Jakes
Sensor speed v	100 km/h	50 km/h
Carrier frequency f_c	700 MHz	
Sensing bandwidth	100 kHz	
Baseline sensing interval	0.1 ms (N = 10 samples)	1 ms (N = 100 samples)
Sensing period	40 ms	80 ms
Decision distance	100 m or 10 m	10 m or 107 m
SNR	-10 dB or -5 dB	
Sensor link budget	-5 dB	

analyzed in [5]. Performance of energy detection under small scale fading is modeled analytically in [6], and under correlated shadowing in [7]. Detection performance of a single energy detector under composite fading is evaluated in [8]. Our system model builds upon [9], which addresses performance of vehicular sensors under path loss and shadow fading.

In the following, we use the terms coherence distance D_s and coherence time T_s to quantify coherence of the small scale fading. The terms decorrelation distance D_l and the decorrelation time T_l are used to quantify correlation of the large scale fading.

In the next section we introduce the system model. In Section 3 we explain specific settings used in the simulations. Section 4 contains results of the performance evaluation. We formulate conclusions in Section 5.

2. SYSTEM MODEL

We consider two types of propagation environments to describe radio channel between the primary spectrum user and the sensors positioned on the vehicles. The *urban* and the *rural* environments have different large and small scale fading parameters, which are summarized in Table 1. The rationale behind selection of some of the parameters in Table 1 is explained in Section 3. With respect to propagation, we use the rural, open space model to describe fading of the primary signal on a freeway surrounded by flat terrain. The urban model describes fading in a downtown city area, which is characterized by much larger delay spread, absence of the line–of–sight (LOS), and shorter decorrelation distance of shadowing.

We focus on the TV white space and assume a channel centered at f_c = 700 MHz.

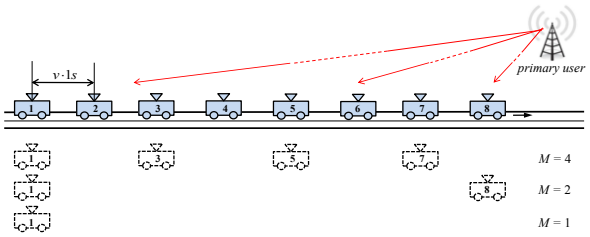


Fig. 1: Mobility model.

2.1. Mobility

We assume a model with either a single car, or a convoy of up to M = 8 vehicles (presented in Fig. 1) traveling straight in the same direction with constant speed v. The vehicles are separated by a fixed distance. This distance is the distance each vehicle passes in a second, and it corresponds to a dense traffic scenario. By expressing the distance in time units we take into account intuitive behavior of drivers to positively correlate the speed and the separation among vehicles. The reason for this is to maintain distance between the vehicles which is sufficient to accommodate the reaction time in case a decisive maneuver is needed.

It is important to emphasize that shortening the distance among vehicles has profound influence on sensing, since the diversity gain depends on the correlation between channel realizations.

2.2. Fading

2.2.1. Large Scale Fading

We assume that the distance between the sensors is much smaller than the distance between the primary user and the sensors. Therefore, we neglect change in primary power due to path loss and take into account only the lognormal shadowing.

The standard deviation of shadowing is provided in Table 1. Selected values are representative for corresponding environments, that is, larger in the urban environment and smaller in the rural environment. The correlation of shadowing is described by the empirical exponential model [10]. Correlation is realized by multiplying M uncorrelated normally distributed vectors with a Cholesky decomposition of the desired correlation matrix. The decorrelation distance D_l corresponds to the distance d between the sensors at which the correlation coefficient is equal to 0.5:

$$\rho = \exp\left(-\ln 2 \cdot \frac{d}{D_l}\right). \quad (1)$$

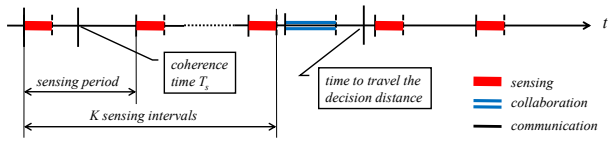


Fig. 2: Scheduling of sensing.

Once a realization of shadowing is created, the moving vehicles observe the same values at the corresponding positions.

2.2.2. Small Scale Fading

Small scale fading is modeled using the simplified GSM wideband propagation models, applicable for channels with bandwidth of up to 10 MHz [11]. Since small scale fading decorrelates rapidly, we generate independent initial tap coefficients at the start of every sensing interval. The model parameters are also given in Table 1.

2.2.3. Received Signal

Since the purpose of our work is to evaluate diversity combining on different spatiotemporal scales, we assume simple energy detection as the underlying local sensing algorithm. For simplicity, we assume that the signal to be detected is a pilot tone. This somewhat resembles detection of the pilot tone which is embedded into the Advanced Television Systems Committee (ATSC) digital TV signal by adding the constant 1.25 to the baseband signal [12].

The constant signal of amplitude A is passed through the time varying filter representing the small scale fading $h_s(t;\tau)$ and then scaled with the value which is zero mean normally distributed on the logarithmic scale

$$10\log_{10}(h_l) \sim \mathcal{N}(0,\sigma^2). \tag{2}$$

Since the value of shadowing represents local mean of the channel response, it is constant during a sensing interval. The complex baseband representation of the faded signal is then given as

$$y(t) = \left[A * h_s(t;\tau)\right] \cdot h_l \tag{3}$$

where * represents convolution.

After passing it through the random filter representing the channel between the primary transmitter and the sensors, we decimate the signal y to narrow the bandwidth to 100 kHz and insert additive white Gaussian noise (AWGN) $n(t)$ with power corresponding to that bandwidth

$$r(t) = y(t) + n(t). \tag{4}$$

We model the contribution of the receiver noise figure, cable losses, and antenna gain with a 5 dB increase in the noise floor.

2.3. Primary User Detection

To formulate the decision statistics we use a number of samples of the received signal r. Let us represent a vector of these complex values with $\boldsymbol{R}$. We collect these samples in two ways:

1. In the proposed scheduling scheme, presented in Fig. 2, vector $\boldsymbol{R}$ contains $N{\cdot}K$ samples. At each of K sensing intervals N samples are collected. The sampling intervals occur with the period much larger than the coherence time of the small scale fading T_s. On the other hand, the time to acquire N samples is shorter than T_s. In other words, N samples are collected while the channel is statistically time invariant. In this manner K uncorrelated sets of samples are acquired.
2. As the benchmark test, we repeat the same simulations with distinction that all $N{\cdot}K$ samples are collected consecutively. In this case, the total sensing time needed to acquire these samples is shorter or comparable to T_s.

For a given detection threshold η, we decide between the two hypotheses

$$\begin{cases} H_1 : \text{Primary user present} \\ H_0 : \text{Channel is free} \end{cases}. \tag{5}$$

2.3.1 Hard Decision Combining

Let index m denote one of M mobile sensors. The local hypothesis testing is

$$\boldsymbol{R}_m' \boldsymbol{R}_m \underset{H_0}{\overset{H_1}{\gtrless}} KN\eta \tag{6}$$

where $'$ represents conjugate transpose. These M local decisions can then be combined using different logical rules: AND, OR, or majority.

2.3.2. Soft Decision Combining

The local test statistics can be averaged across M vehicles

$$\sum_{m=1}^{M} \boldsymbol{R}_m' \boldsymbol{R}_m \underset{H_0}{\overset{H_1}{\gtrless}} MKN\eta \tag{7}$$

resembling the equal gain combining (EGC). Another approach is to put more weight to stronger signals, similar to the maximum ratio combining (MRC)

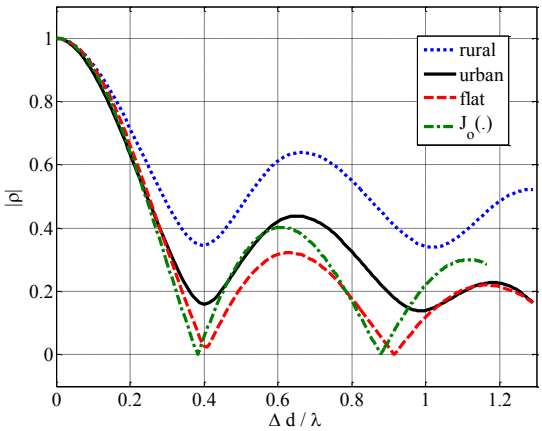


Fig. 3: Numerically generated normalized correlation coefficients of 100 kHz channels for different small scale fading models.

$$\sum_{m=1}^{M} a_m \boldsymbol{R}_m' \boldsymbol{R}_m \underset{H_0}{\overset{H_1}{\gtrless}} MKN\eta \tag{8}$$

with weights a_m provided from crude signal–to–noise ratio (SNR) estimates

$$a_m = \frac{\boldsymbol{R}_m' \boldsymbol{R}_m}{\sum_{m=1}^{M} \boldsymbol{R}_m' \boldsymbol{R}_m}. \tag{9}$$

3. NUMERICAL ANALYSIS IN DETAIL

In practice, a regulatory body imposes requirement that the decision about presence of a primary user is reached in regular intervals. For instance, the FCC [2] requires that spectrum sensing should be performed at least once every 60 s. This requirement is not intended for highly mobile white space devices. More adequate requirement is the one intended for devices which rely on geolocation database access to acquire spectrum occupancy. The database should be accessed each time a device moves by 100 m. It is reasonable to assume similar rule for vehicular networks. By relating primary user protection to traveled distance it can then be made independent of the speed, which varies widely with location and time of day.

We selected two values for the decision making distance. The first one—100 m—is following the current FCC rules. The second—10 m—is more stringent, with rationale that it should provide better protection of primary users. Both values are comparable to realistic values of the large scale fading decorrelation D_l.

We assume that the sensors move by 50 km/h in the urban environment and 100 km/h in the rural environment. The one–second separation between the cars then corresponds to 13.9 m and 27.8 m, respectively. The convoy of $M = 8$ vehicles is 97.2 m and 194.4 m long, respectively.

To decide on duration of the sensing interval we performed simple tests on used multipath models to determine the coherence distance D_c statistically from a number of realizations of $y(t)$

$$\rho(\Delta t) = \frac{\sum_i \left(y(t_i + \Delta t) - \mu_{i,\Delta t}\right)\overline{\left(y(t_i) - \mu_i\right)}}{\sqrt{\sum_i \left|y(t_i + \Delta t) - \mu_{i,\Delta t}\right|^2} \sqrt{\sum_i \left|y(t_i) - \mu_i\right|^2}} \tag{10}$$

where $\mu_{i,\Delta t}$ and μ_i represent the sample means calculated over samples $y(t_i+\Delta t)$ and $y(t_i)$, respectively. The bar denotes complex conjugate.

For a given constant speed v it is trivial to convert time lag into distance $\Delta d = v\Delta t$, and normalize the lag with respect to the carrier wavelength λ_c. In Fig. 3 we present these results for decimated (100 kHz) channels in a form independent of speed and carrier frequency. The Rice model selected to represent small scale fading in the rural environment decorrelates much slower than the Rayleigh urban model due to the deterministic portion of the LOS component. The Jake's Doppler spectra of the taps in the urban model result in correlation very similar to the flat (single tap) Rayleigh fading. As a sanity check we also provide theoretical flat fading correlation [3], given by the Bessel function of the first kind and zero order J_o.

When evaluating the sensing performance we vary the number of vehicles $M \in \{1,2,4,8\}$ (Fig. 1). For $M = 2$ we simply keep the first and the last car in the convoy $m \in \{1,8\}$. When $M = 4$, we consider only cars with odd indexes $m \in \{1,3,5,7\}$. To make fair comparison for different M we scale the sensing interval (number of samples N). For instance, when $M = 2$ we increase the number of samples four times. This way we keep the time–bandwidth product constant in all simulations.

When $M = 8$ in the urban environment we set the sensing interval to 1 ms ($N = 100$). Convenient setting of the period of sensing intervals to 40 ms for rural and 80 ms for urban environment results in $K = 100$ in the rural and $K = 10$ in the urban environment. To keep constant the product $MKN = 8000$ we set $N = 10$ in the rural environment. If, for instance, $M = 1$ the number of samples increases to $N = 80$ in the rural and $N = 800$ in the urban environment.

In each simulation run, as the vehicles traverse the decision making distance, we keep the shadow fading h_l constant in discrete steps representing the local areas. Following the channel modeling heuristics detailed in [13] and [10] we conveniently set the local area to be ten carrier

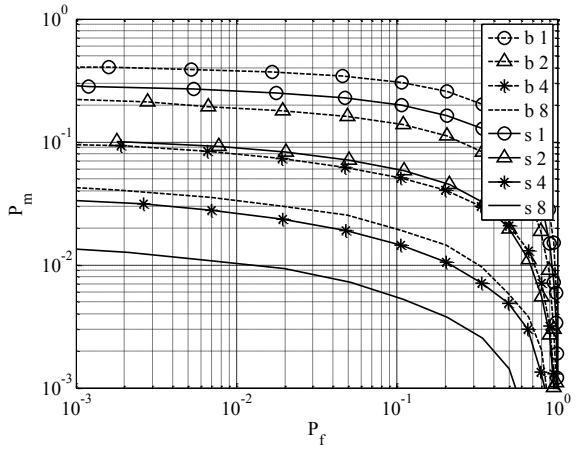


Fig. 4: Performance of EGC in urban environment for 1, 2, 4 and 8 collaborating cars, under -5 dB SNR (-22.8 dB SNR for 6 MHz channel), and with decision distance 10 m. Dashed line represents benchmark sensing in a single slot (denoted with 'b') and full line represents proposed slotted sensing (denoted with 's').

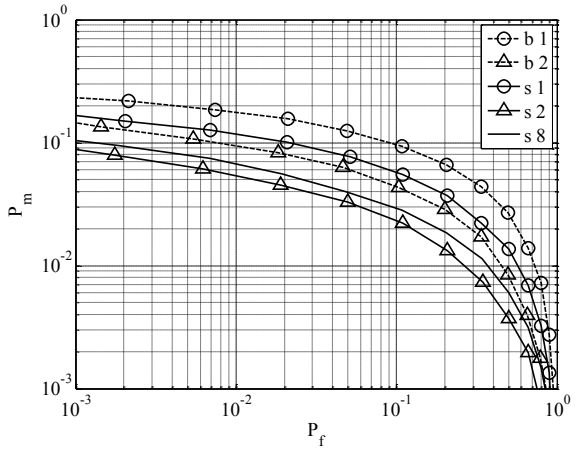


Fig. 5: Performance of EGC in rural environment for 1, 2 and 8 cars, under -5 dB SNR (-22.8 dB SNR for 6 MHz channel), for decision distance 100 m. Dashed line represents benchmark sensing in a single slot (denoted with 'b') and full line represents proposed slotted sensing (denoted with 's').

wavelengths λ_c in the rural environment and five carrier wavelengths in the urban environment.

The detection probability P_d—and its complement, the probability of missed detection P_m—are evaluated for different SNRs by creating a channel realization, adjusting the amplitude A, and calculating the test statistics. This procedure is repeated for a large number of channel realizations. The false alarm probability P_f is evaluated by setting $A = 0$. It is then possible to relate P_f and P_m through different values of threshold η and generate complementary Receiver Operating Curves (ROC).

4. RESULTS

We assume the SNR to be the ratio between the power of signal that must be detected A^2 and the noise floor of 100 kHz sensing bandwidth (without mentioned 5 dB budget penalty). We consider two such SNRs, -10 dB and -5 dB. These correspond to -27.8 dB and -22.8 dB in comparison to the noise floor of a 6 MHz TV channel, respectively. Of course, the performance of a detection system suffers from many implementation issues, but it can be further improved in two ways: 1) by increasing the sensing interval; and 2) by applying feature detection. Since energy detection is used as the underlying sensing algorithm for simplicity, we assume perfect estimation of the noise floor.

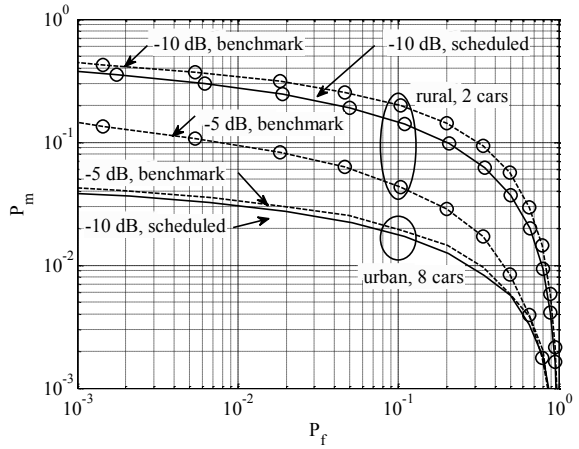


Fig. 6: Diversity gain achieved through the scheduling of sensing is between 1 and 2 dB in the rural environment, and around 5 dB in the urban environment.

4.1. Diversity Gain

In Figs. 4 and 5 we present detection performance of EGC in the urban and in the rural environment with -5 dB SNR, for different number of sensors M.

In the urban environment (Fig. 4), assuming 10 m decision distance, scheduling of sensing with $M = 2$ and 4 sensors shows practically the same performance as sensing in a single slot for the same time with $M = 4$ and 8 sensors, respectively.

In the rural environment (Fig. 5), the diversity gain is much smaller for two reasons. First, half of the power related to the small scale fading is deterministic. Second, the variance of shadowing is much smaller in comparison to the urban environment. With decorrelation distance of 100 m, even only $M = 2$ sensors, separated by almost 200 m, achieve maximum diversity. We attribute slightly worse performance of $M = 8$ sensors to their dense arrangement being exposed to the same shadowing as they move. For clarity, we omitted the results for benchmark sensing with eight cars, which is similar to benchmark sensing with two cars.

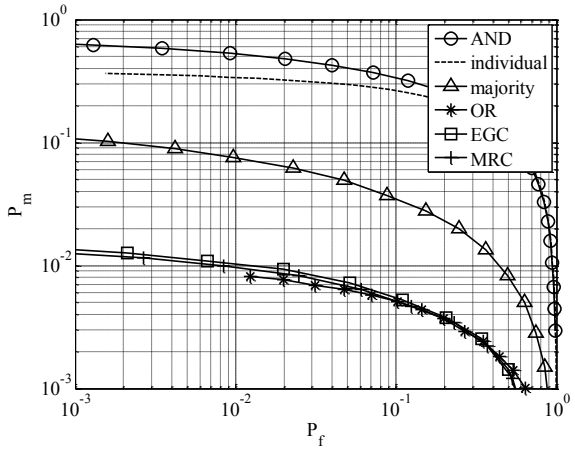


Fig. 7: Performance of different hard and soft fusion algorithms in urban environment for 10 m decision distance, and -5 dB SNR (-22.8 dB SNR for 6 MHz channel).

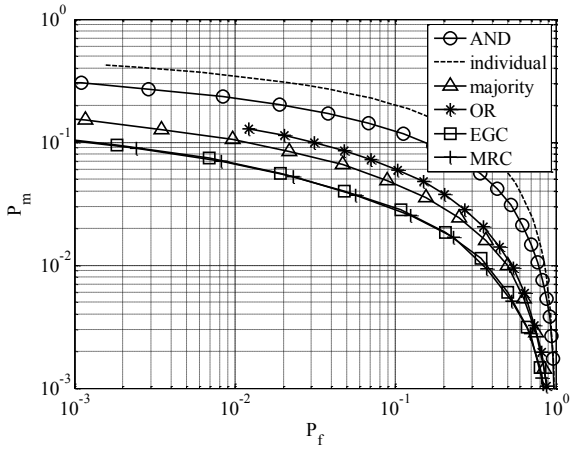


Fig. 8: Performance of different hard and soft fusion algorithms in rural environment for 100 m decision distance, and -5 dB SNR (-22.8 dB SNR for 6 MHz channel).

To illustrate the gain achieved by scheduling, we present in Fig. 6 results for the rural environment with two cars, together with results for the urban environment with eight cars. The gain in the rural environment is between 1 and 2 dB. The gain in the urban environment is approximately 5 dB.

4.2. Performance of Different Fusion Algorithms

Fig. 7 provides an overview of detection performance of our scheduling scheme across different fusion algorithms for SNR -5 dB in the urban environment. Here "individual" sensing represents the average of local decisions across all $M = 8$ vehicles and all realizations.

Among hard fusion algorithms, the OR rule shows the best performance. This is easy to justify having in mind that all sensors are considered equally reliable and, on the average, all experience the same fading statistics. Therefore, when a sensor correctly detects the primary user, it is most likely the sensor with the best propagation conditions.

The weighted sum performs only marginally better than the equal gain combining. It is well known that, when all diversity branches exercise the same average power, MRC performs only slightly better than EGC [3]. However, when path loss fading is not negligible across the sensors, the MRC should provide larger performance improvement.

To complement results in Fig. 7 we present in Fig. 8 the same results for the rural environment. In this case, with less variation in signal strength due to shadowing, soft methods perform better than all hard rules. These results qualitatively agree well with the results reported in [9] for a similar shadowing statistics.

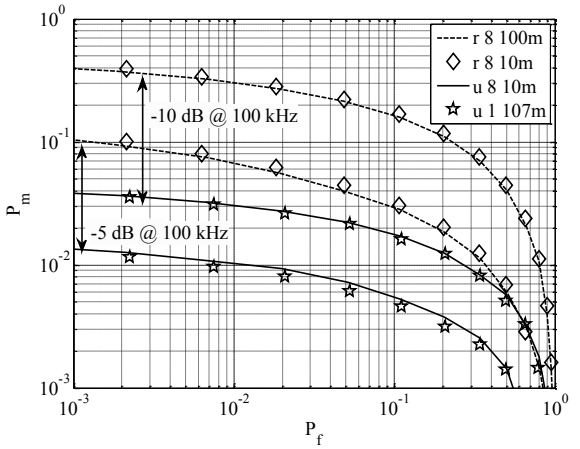


Fig. 9: Performance of EGC in rural and urban environments for different decision distances, and -10 dB and -5 dB SNR (-27.8 dB and -22.8 dB SNR for 6 MHz channel).

4.3. Decision Distance versus Spatiotemporal Tradeoff

The same argument we used to describe performance of two and eight sensors in Fig. 5 can be used to explain some of the results in Fig. 9, represented with dashed line and diamond marker. Here we look into the rural environment with eight sensors which are distributed over a distance twice the shadowing decorrelation distance. In such a scenario any decision distance smaller or equal to the decorrelation distance (e.g. 10 m or 100 m) does not help to decorrelate the shadow fading.

Complementary to that, if a single sensor can "wait enough" to experience the same diversity gain experienced by many collaborating nodes, it can achieve the same performance as they do. For instance, $M = 8$ nodes cover 107 m while each travelling for 10 m. If a single node travels the same distance before making decision, it achieves the same performance as eight nodes, indicated with star markers in Fig. 9.

The tradeoff between space and time has both pros and cons. Collaboration involves communication overhead to

exchange local sensing information and distribute the fusion outcome by means of unreliable wireless communication. In comparison to that, a single sensor utilizing temporal diversity suffers from additional delay needed to traverse the same distance.

5. CONCLUSION

We proposed and evaluated a method to improve performance of spectrum sensing in the vehicular environment by carefully scheduling sensing intervals to accommodate for different time scales of the primary user signal fading. Assuming energy detection, splitting of the sensing interval into many shorter intervals (with a period much larger than the coherence time of small scale fading) results in approximately 1 to 2 dB gain in the rural environment, and around 5 dB gain in the urban environment.

The soft combining techniques provide consistently good performance irrespective of the environment. This is not the case for hard combining. Under strong fading and equal average powers at the sensors, the OR rule is comparable to the soft methods. Under less severe fading simple majority outperforms all hard rules, but fails short of the accuracy achieved with the soft methods.

We also discussed influence of regulatory domain requirements, namely, the distance at which mobile sensors should re–evaluate spectrum occupancy, on the accuracy of detection. In severe fading, with detection distance much larger than the fading decorrelation distance, diversity gain can be utilized with a single sensor by trading spatial diversity for temporal diversity. This approach has the advantage of avoiding communication overhead (including failures) associated with cooperative sensing. Under mild fading conditions, for very short detection distance, the diversity must be exploited through collaboration of sensors.

6. REFERENCES

[1] H. Kremo, R. Vuyyuru and O. Altintas, "Spectrum Sensing in the Vehicular Environment: An Overview of the Requirements," in *Wireless Innovation Forum European Conference on Communications Technologies and Software Defined Radio*, Brussels, 2012.

[2] *Second Memorandum Opinion and Order In the Matter of Unlicensed Operation in the TV Broadcast Bands, Additional Spectrum for Unlicensed Devices Below 900 MHz and in the 3 GHz Band,* Federal Communication Commission, 2010.

[3] A. Goldsmith, Wireless Communications, New York: Cambridge University Press, 2009.

[4] S. M. Mishra, A. Sahai and R. W. Brodersen, "Cooperative Sensing Among Cognitive Radios," in *IEEE International Conference on Communications - ICC*, Istanbul, 2006.

[5] A. W. Min and K. G. Shin, "Impact of mobility on spectrum sensing in cognitive radio networks," in *ACM workshop on Cognitive radio networks (CoRoNet '09)*, 2009.

[6] S. P. Herath, N. Rajatheva and C. Tellambura, "Energy Detection of Unknown Signals in Fading and Diversity Reception," *IEEE Transactions on Communications,* vol. 59, no. 9, pp. 2443-2453, 2011.

[7] A. Ghasemi and E. S. Sousa, "Asymptotic Performance of Collaborative Spectrum Sensing under Correlated Log-Normal Shadowing," *IEEE Communication Letters,* vol. 11, no. 1, pp. 34-36, 2007.

[8] H. Rasheed and N. Rajatheva, "Spectrum Sensing for Cognitive Vehicular Networks over Composite Fading," *International Journal on Vehicular Technology,* vol. 2011, 2011.

[9] D. Borota, G. Ivkovic, R. Vuyyury, O. Altintas, I. Seskar and P. Spasojevic, "On the Delay to Reliably Detect Channel Availability in Cooperative Vehicular Environments," in *IEEE Vehicular Technology Conference*, Budapest, 2011.

[10] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *IET Electronic Letters,* vol. 27, no. 23, pp. 2145-2146, November 1991.

[11] European Telecommunications Standards Institute, "Global System for Mobile Commuincations 05.05: Radio Transmission and Reception," Sophia Antipolis, 1996.

[12] Advanced Television Systems Committee, "ATSC Digital Television Standard Part 1 – Digital Television System (A/53, Part 1:2007)," 2007. [Online]. Available: http://www.atsc.org/cms/standards/a53/a_53-Part-1-6-2007.pdf. [Accessed 12 4 2013].

[13] T. S. Rappaport, Wireless Communications: Principles and Practice, 2 ed., Prentice Hall, 2002.

Non-parametric Spectrum Sensing based on Censored Observations in Quasi-static Fading Channel for Cognitive radio

D. K. Patel
Department of Electrical Engineering
Institute of Technology,Nirma University
Ahmedabad, Gujarat 380015
Email: dkpatel@ieee.org

Y. N. Trivedi
Department of Electrical Engineering
Institute of Technology,Nirma University
Ahmedabad, Gujarat 380015
Email: yogesh.trivedi@nirmauni.ac.in

***Abstract*—The detection of primary user (PU) is a challenging task for a cognitive radio user to access the spectrum opportunistically. In this paper, a spectrum sensing method based on censored observations is proposed as Censored Anderson Darling sensing (CAD). We evaluate the performance of the CAD sensing method with Monte Carlo simulations. It is shown that the proposed method outperforms the energy detection at about 6 dB gain over quasi-static fading channel at lower signal to noise ratio. Also, it gives better detection performance compared to AD sensing and OS based sensing methods which have recently been proposed in the literature.**

***Keywords*—Spectrum sensing, goodness of fit test, quasi-static channel, censored data**

I. Introduction

Today's wireless networks are characterized by fixed spectrum assignment policy. With ever increasing demand for frequency spectrum and limited resource availability, the FCC has decided to make a paradigm shift by allowing more and more number of secondary users (SU) to transmit their signals in licensed bands for utilizing the available spectrum of primary users (PU) efficiently. This can be achieved using Cognitive radio (CR) [1]. One of the most important components in Cognitive radio is spectrum sensing. The main function in spectrum sensing is to detect the PU (or licensed user). This task is performed by SU (or unlicensed user) which can use the spectrum of PU such that they do not cause interference to PU. The spectrum sensing function is suffered by multipath fading, receiver's uncertainty, interference etc. Therefore, the design of a spectrum sensing algorithm for future wireless communications is a challenging problem in the research community.

In the last couple of years, many efforts are put by researches to provide spectrum access in an opportunistic way. There are different spectrum sensing techniques proposed under the category of parametric sensing in which some information about PU is available at SU. The different parametric sensing methods are Cyclo-stationary detection, matched filter, waveform-based sensing etc [2] [3]. In category of nonparametric sensing, Energy Detection (ED) [4] and Goodness of Fit (GoF) tests based sensing like Anderson Darling (AD) sensing [5], Kolmogorov-Smirnov (KS) sensing [6], Student t- sensing [7] and Order statistics based sensing [8] are proposed wherein no information about PU is required at SU.

The energy detection (ED) is the most common method for spectrum sensing due to its low complexity. However, the performance of the ED degrades at low signal-to-noise ratio (SNR) with uncertainty in noise power. Also, the detection performance is degraded at the less number of observations. In this scenario of low SNR and received observations, GoF test based sensing is preferred. GoF based sensing always means a statistical test for the presences of certain distribution [9]. More specifically, all received observations are independent and identically distributed (i.i.d) random variables with cumulative distribution function (CDF), denoted by F. In this kind of sensing, the test of the null hypothesis ($F = F_0$) against the alternative hypothesis ($F \neq F_0$) has been done, where F_0 is an available CDF. For performing any GoF test, the empirical CDF (ECDF) is determined from the received observations. This ECDF is compared with the known CDF (F_0) under the null hypothesis. The distance of the ECDF from the CDF decides whether PU is present or absent.

Based on this GoF testing method, AD sensing was proposed, wherein a special weight function has been used to give more emphasis to the tails of the CDF. Furthermore, in [10], AD test is used for the detection of PU under the condition of unknown noise power. The Student's t-distribution is used for the testing of null hypothesis instead of gaussian distribution. Recently in [11], the distance between the ECDF of the received observations and the known CDF, is measured using characteristic functions, instead of the CDFs, has been considered for the testing of null hypothesis. Also, GoF testing based on order statistics in [8], has been proposed for an AWGN channel.

All types of GoF based sensing methods which are proposed in literature so far, have used all observations to determine ECDF. However, the distance of the CDF and ECDF is higher especially at the right tail due to less number of observations. This incomplete information of CDF on the right tail introduces an error in determining statistics in AD sensing, especially at low SNR. To overcome this, we have used the concept of censored data which has already been used in survival analysis [12]. In view of this, we drop some observations in the right tail, which carry incomplete

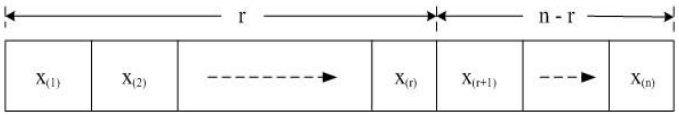


Fig. 1 Number of received (N) and censored $(N-R)$ observations

information for the CDF.

In this paper, we have proposed a nonparametric sensing method using GoF testing based on censored observations, also called Censored Anderson Darling (CAD) sensing. In this method, the observations from right tail are removed and we use modified statistic for the testing of null hypothesis as derived in [13]. This statistic has been obtained by modifying the upper limit of the integration. We have found that CAD sensing outperforms the AD sensing at lower values of SNR and less numbers of received observations under a quasi-static channel. Furthermore, the processing become simpler for the detection of PU. We have also compared CAD sensing with ED and OS based sensing methods.

The rest of the paper is organized as follows. The problem of spectrum sensing as GoF testing for censored observations is formulated as null hypothesis testing problem in Section II. In Section III, the detection performance of the CAD sensing algorithm is presented and compared with OS, AD and ED sensing methods. Finally, the paper is concluded in Section IV.

II. Goodness of fit testing for censored observations

Let $\mathbf{y} = [y_1, y_2, ..., y_N]^T$ be the received signal vector at the secondary user (SU), where N denotes total number of observations. We assume received observations are real valued and each y_i is represented as,

$$y_i = \sqrt{\rho} h s + w_i, \quad i = 1, 2, 3, \cdots\cdots N, \tag{1}$$

where $s \in \{0, 1\}$, ρ is the received SNR, h represents the fading factor, which is assumed to be random variable with the standard normal distribution. We also assume that the channel is quasi-static rayleigh fading. In (1), w_i, where $1 \leq i \leq N$, denotes noise samples. The CDF of w_i is denoted by $F_0(y)$. In (1), $s = 1$ and 0 denote presence and absence of PU respectively. Without loss of generality, we assume that all N observations are in ascending order. It means $y_1 \leq y_2 \leq \cdot\cdot, y_N$. Now, we retain first R observations and drop or censor the last $N-R$ observations as shown in Fig. 1. Hence, y_R is the highest valued observation. This method of censoring $N-R$ highest valued observations is known as right censoring with Type-2 [12].

In this scenario, the problem of spectrum sensing as null hypothesis testing problem as GoF testing is defined as [5],

$$\begin{aligned} H_0 &: F_Y(y) = F_0(y) \\ H_1 &: F_Y(y) \neq F_0(y) \end{aligned} \tag{2}$$

For CAD sensing, we use modified Cramer-von Mises GoF statistic to measure distance between $F_Y(y)$ and $F_0(y)$. Let $F_n(y)$ be the Empirical Cumulative Distribution Function (ECDF) of the received observations $\mathbf{y}$, which can be expressed as

$$F_n(y) = \frac{|\{i : y_i \leq y, 1 \leq i \leq N\}|}{N}, \tag{3}$$

where $|\cdots|$ indicates cardinality. In this case, based on the asymptotic distribution of censored observations, statistic can be expressed as [13],

$${}_{q,p}A_N^2 = N \int_q^p \frac{(F_n(y) - F_0(y))^2}{F_0(y)(1 - F_0(y))} dF_0(y), \quad 0 \leq q < p \leq 1,$$

where p denotes censoring ratio which can be expressed as

$$p = \lim_{n \to \infty} \frac{R}{N}.$$

Here, we take $q = 0$. In this case, statistic can be written as,

$${}_pA_N^2 = N \int_0^p \frac{(F_n(y) - F_0(y))^2}{F_0(y)(1 - F_0(y))} dF_0(y) \tag{4}$$

The above quadratic statistics ${}_pA_N^2$ can be solved using integration by parts and approximated as [13],

$$\begin{aligned} {}_pA_N^2 = &-\frac{1}{N} \sum_{i=1}^{R} (2i-1)(lnz_i - ln(1 - z_i)) - 2 \sum_{i=1}^{R} ln(1 - z_i) \\ &- \frac{1}{N}[(R-N)^2 ln(1 - z_R) - R^2 lnz_R + n^2 z_R], \end{aligned} \tag{5}$$

where $z_i = F_0(y_i)$. For sensing at secondary user, based on censored observations, H_0 is rejected when ${}_pA_N^2 > \lambda$, where λ is the value of threshold. The λ is selected such that the false alarm probability (P_f) under H_0 is at a desired level α,

$$\alpha = \mathbb{P}\{ {}_pA_N^2 > \lambda | H_0\} \tag{6}$$

To find λ, it is worth to mention that the distribution of ${}_pA_N^2$ under H_0 is independent of the $F_0(y)$. To observe this, apply probability integration transform (PIT) for available observations. Hence,

$${}_pA_N^2 = N \int_0^p \frac{(F_z(z) - z)^2}{z(1 - z)} dz, \tag{7}$$

where $z = F_0(y)$ and $F_z(z)$ denotes ECDF of z_i. Here, $z_i = F_0(y_i)$ for $1 \leq i \leq R$. All statistics of observations up to z_R are independent and uniformly distributed over $[0, p]$, where $p \in [0, 1]$. As shown in [5] for AD sensing, the distribution of A_c^2 is independent of the $F_0(y)$. The same is also true for the distribution of ${}_pA_N^2$. As given in [9], the value of λ is determined for a specific value of P_f and censoring ratio p. For example, when $P_f = 0.05$ and $p = 0.4$, the value of λ is 1.133.

Let us summarize, the above discussion in the following steps for CAD sensing algorithm:

Step:1 Find the threshold λ for a given probability of false alarm P_f using (6).

Step:2 Sorting all the observations in ascending order, we get

$$y_1 \leq y_2 \leq \cdot\cdot \leq y_R \leq y_{R+1} \leq \cdot\cdot \leq y_N,$$

where $y_{R+1} \leq y_{R+2} \cdot\cdot \leq y_N$ observations are censored.

Step:3 Calculate the required test statistic ${}_pA_N^2$ for the observations $y_1 \leq y_2 \leq \cdot\cdot \leq y_R$ as defined in (5).

Step:4 If ${}_pA_N^2 > \lambda$, then reject null hypothesis H_0.

Step:5 Compute performance metric as Probability of Detection (P_d) with a given value of P_f.

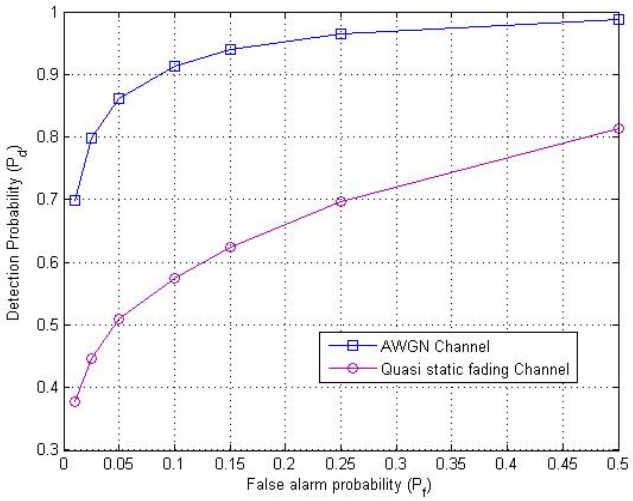


Fig. 2 ROC graph for Censored Anderson-Darling sensing at SNR = -2 dB

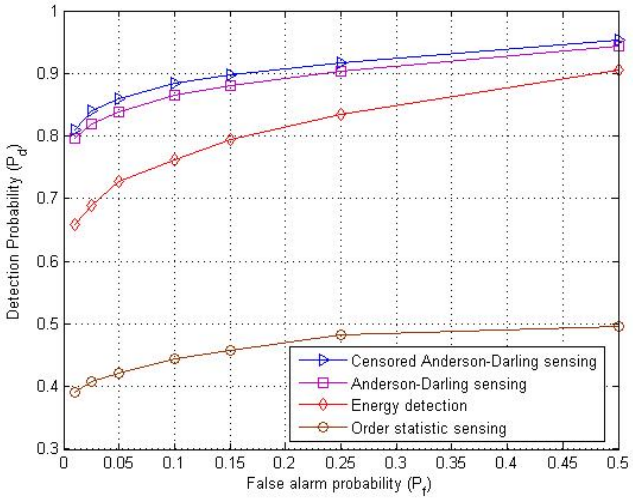


Fig. 3 ROC graphs for CAD, AD and ED Sensing at SNR = 6 dB under Quasi-static fading channel

III. Simulation Results

In this section, we have shown the performance of the CAD sensing algorithm with receiver operating characteristics (ROC) using monte carlo simulations. The ROC is a curve between Probability of detection (P_d) versus Probability of False alarm (P_f). These curves are obtained for different values of observations (N), censoring ratio (p) and SNR. We have also presented ROC for ED, AD and OS sensing algorithms and compared them with the proposed one.

Fig. 2 shows ROC for CAD sensing at SNR of -2 dB for $P_f = 0.05$ under AWGN and quasi-static rayleigh fading channel. We have taken R = 24 corresponding to the value of $N = 40$, so that p remains constant as 0.6. We can see that CAD gives higher detection performance under AWGN

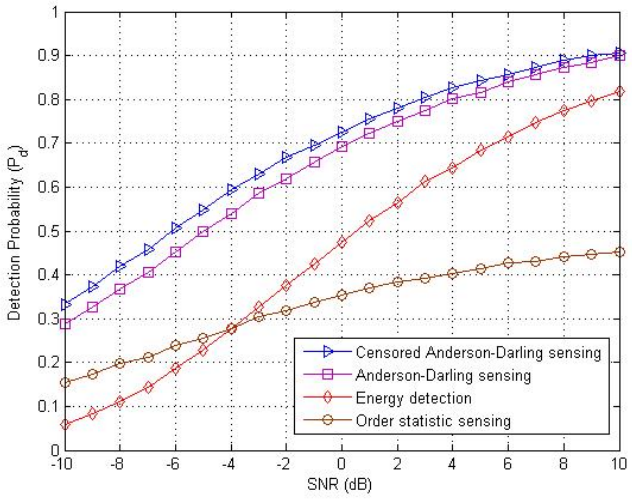


Fig. 4 Probability of detection for CAD, AD, OS and ED sensing at $P_f = 0.05$ under Quasi-static fading channel

compared to quasi-static rayleigh fading channel as expected.

Fig. 3 shows ROC for CAD sensing for $P_f = 0.05$ and an SNR of 6 dB for fixed value of N = 25. We have taken R = 24 corresponding to the value of N, so that p remains constant as 0.6. We have also presented ROC of ED and AD sensing algorithms for the same values of P_f, N and SNR. For specified P_f and N, it can be seen that the P_d = 0.8419 is obtained in AD sensing. For CAD sensing, N = 40 and R = 24 are taken. It means 24 observations are used, which are almost same as $N = 25$ in case of AD sensing. It can be seen that better $P_d = 0.8615$ for CAD sensing is achieved compared to AD sensing. In case of ED and OS sensing, $P_d = 0.7274$, $P_d = 0.4216$ are achieved respectively. Thus, the proposed CAD sensing outperforms OS, ED and AD sensing algorithms.

Furthermore, in fig. 4, we have shown P_d versus SNR for $P_f = 0.05$, $N = 40$ and $p = 0.6$ for CAD sensing. As SNR increases, P_d increases as per expectation. We have also presented performance of ED, AD and OS sensing methods in the same figure with $N = 25$ and same P_f. We can see that $P_d = 0.1106,\ 0.1967,\ 0.3655$ and 0.4225 for ED, OS, AD and CAD respectively at SNR of -8 dB.

It can be shown from fig. 4 that when the signal is transmitted on quasi-static rayleigh fading channel, CAD sensing has almost 6 dB gain over ED sensing i.e signal detected by ED sensing at 10 dB SNR with $Pd = 0.8183$, the similar detection performance is achieved at 4 dB in case of CAD sensing at the same value of P_f. Furthermore, gain of 1 dB is archived compared to AD sensing. We can see significant improvement in P_d compared to OS sensing. Thus, the CAD sensing outperforms at lower value of SNR compared to the remaining schemes for the whole range of SNR from -10 dB to 10 dB.

IV. Conclusion

In this paper, we have discussed the problem of spectrum sensing as null hypothesis testing problem for censored obser-

vations under quasi-static channel. The ROC is presented for the proposed CAD algorithm and compared with conventional OS, AD and ED sensing methods. The CAD sensing method gives significant improvement in detection of primary user compared to the ED sensing at about 6 dB gain at lower signal to noise ratio. The simulation and numerical results show that CAD sensing outperforms OS sensing as well as AD sensing with a gain of 1 dB.

REFERENCES

[1] S. Haykin, "Cognitive radio: Brain-Empowered Wireless Communications," *IEEE J. Sel. Areas in Commun.*, vol. 23, no. 2, pp. 201-220, Feb. 2005

[2] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys and Tutorials*, vol. 11, no. 1, pp. 116-130, Quat. 2009.

[3] E. Axell, G. Leus, E. G. Larsson and H. V. Poor, "Spectrum sensing for cognitive Radio : State-of-the-art and recent advances," *IEEE Signal Processing Mag.*, vol. 29, no. 3, pp. 101-116, May. 2012.

[4] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523-531, Apr. 1967.

[5] H. Wang, E. H. Yang, Z. Zhao and W. Zhang, "Spectrum sensing in cognitive radio using goodness of fit testing," *IEEE Trans. Wireless Commun.*, vol. 8, no. 11, pp. 5427-5430, Nov. 2009.

[6] G. Zhang, X. Wang, Y. C. Liang and J. Liu, "Fast and robust spectrum sensing via kolmogorov-smirnov test," *IEEE Trans. Commun.* vol. 58, no. 12, pp. 3410-3416, Dec. 2010.

[7] K. Arshad, K. Briggs and K. Moessner, "Robust spectrum sensing for cognitive radio based on statistical tests," *ACM 4th Int. Conf. Cognitive Radio and Advanced Spectrum Management*, New York, pp. 12:1-12:6, 2011.

[8] S. Rostami, K. Arshad and K. Moessner, "Order-Statistic Based Spectrum Sensing for Cognitive Radio", *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 592-595, May. 2012.

[9] R. D.Agostino and M. Stephens, Goodness of Fit Techniques, ser. Statistics: Textbooks and Monographs. M. Dekker, 1986. [Online]. Available: http://books.google.co.in/books?id=1BSEaGVBj5QC

[10] L. Shen, H. Wang, W. Zhang and Z. Zhao, "Blind spectrum sensing for cognitive radio channels with noise uncertainty," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1721 -1724, June. 2011.

[11] L. Shen, H. Wang, W. Zhang and Z. Zhao, "Multiple antennas assisted blind spectrum sensing in cognitive radio channels," *IEEE Commun. Lett.*,vol. 16, no. 1, pp. 92-94, Jan. 2012.

[12] J. Lawless, Statistical Models and Methods for Lifetime Data, ser. Wiley Series in Probability and Statistics. Wiley, 2002. [Online]. Available: http://books.google.co.in/books?id=YsFlQgAACAAJ

[13] A. N. Pettitt and M.A Stephens, "Modified Cramer-von-Mises statistics for censored data," *Biometrika*,vol. 63,pp. 291-298, Aug. 1976.

SPECTRUM SHARING AMONG MULTIPLE SECONDARY USERS USING CHANNEL ASSIGNMENT METHOD OF HIGH SPATIAL EFFICIENCY BASED ON MUTUAL INTERFERENCE

Takashi Kosugi (Advanced Wireless Communication research Center (AWCC),
The University of Electro-Communications, Choufu-shi, Tokyo, Japan;
t-kosugi@awcc.uec.ac.jp)
Kei Inage (Advanced Wireless Communication research Center (AWCC),
The University of Electro-Communications, Choufu-shi, Tokyo, Japan;
inagek@awcc.uec.ac.jp)
Takeo Fujii (Advanced Wireless Communication research Center (AWCC),
The University of Electro-Communications, Choufu-shi, Tokyo, Japan;
fujii@awcc.uec.ac.jp)

ABSTRACT

The most important target of a spectrum sharing is maximizing the performance of an unlicensed user (secondary user: SU) under the protection of licensed users (primary user: PU). A lot of researchers have been proposed spectrum sharing methods of maximizing the SU performance while protecting the PU. However, the spectrum sharing among multiple SUs with considering spatial efficiency has not been considered. When multiple SUs exist, channel assignment methods are required to maximize each SU performance under the limitation of the aggregated interference toward the PU. However, the minimization of mutual interference among SUs causes sparsely spectral utilization of spectrum in a point of spatial domain. In this paper, in order to realize high dense SUs with keeping interference constraint, we propose a novel channel assignment method targeting high spatial efficiency while keeping minimum required signal power to interference power plus noise power ratio (SINR) of each SU. In order to achieve high spatial efficiency, the central control server is prepared and it assigns the channel of white space (WS) to SUs based on relative distance among SUs and interference conditions. The central control server assigns the channels to shorten the distance among SUs under considering aggregated interference to be kept less than the allowable interference at cell boundaries of SUs. We confirm the effectiveness of the proposed method through computer simulations.

1. INTRODUCTION

In recent years, the demand of frequency resource increases with diversification of wireless communication systems. The current frequency resource allocation policy is exclusively allocating the dedicated spectrum to each wireless communication system. Therefore, it is difficult to allocate sufficient frequency resources to new application services and to enhancing existing systems. In a future wireless communication, it is necessary to bring a new paradigm shift of the spectrum resource allocation. According to a report of Federal Communications Commission (FCC), there is temporal and spatial vacant spectrum and the current utilization ratio of spectrum is less than 15 percent [1]. Therefore, spectrum is not always used fully and unused licensed bands called White Space (WS) exist.

In order to improve the spectral efficiency, a spectrum sharing using cognitive radio technology has been proposed. Cognitive radio is able to change communication parameters according to the surrounding wireless environment [2]. In the spectrum sharing, unlicensed user (secondary user: SU) can only access licensed band which is not used by licensed user (primary user: PU) if the interference toward the PU can be avoided. In other words, the SUs have to access the WS with lower priority than the PU. In order to meet the condition, the method that protects the PU by controlling the transmit power of the SU has also been proposed [3]. The constraint of the PU protection is also applied when multiple SUs exist. In this case, it is necessary to control aggregated interference toward the PU. Furthermore, in the case of multiple SUs use the same channel, it has to be noted that there is a possibility each SU communication quality decreases due to mutual interference among SUs. The performance of SUs has to be assured under protecting PU. In [4], the medium access control protocol that tries not to degrade the PU communication quality by limiting the

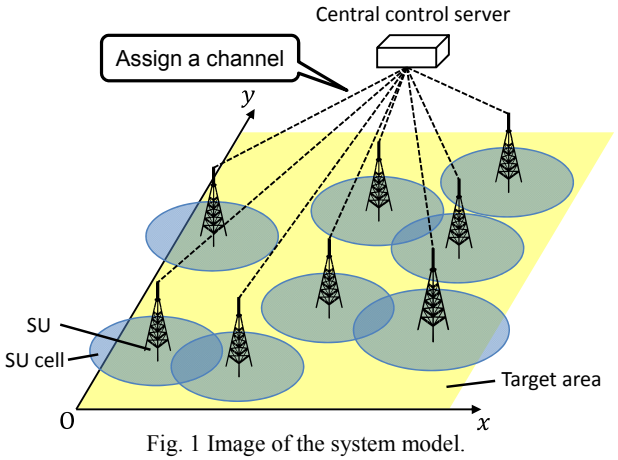


Fig. 1 Image of the system model.

aggregated interference at PU has been proposed and improves SU performance. In [5], [6], the PU is protected by controlling the transmit power of SUs. The medium access control protocol for avoiding collision of signals among SUs has been studied in [7]. In [8], the channel assignment method for maximizing the throughput of each SU with constraint of aggregated interference toward PU has been proposed. However, such the channel assignment causes a spatially sparse utilization of frequency resources if multiple SUs exist, because the relative distance between SUs on the same channel is maximized in order to minimize the mutual interference among SUs. As a result, communication opportunities of SUs which want to start communication are reduced. To solve this problem, it is better that SUs have to be packed densely as much as possible. Therefore, in this paper, we propose a novel channel assignment method targeting high spatial efficiency while keeping minimum required SINR of each SU. In the proposed method, the central control server assigns WS channels to SUs based on relative distance among SUs and interference conditions. The central control server makes channel allocation to shorten the distance among SUs under limitation of the aggregated interference less than allowable interference at the cell boundaries of SUs. We confirm the effectiveness of the proposed method through computer simulations. In simulation results, we confirm that it is possible to perform the channel assignment with high spatial efficiency while keeping every SU communication quality by using the proposed method.

2. SYSTEM MODEL

In this section, the detail of the system model is described. This paper focuses on the spectrum sharing among multiple SUs under multiple candidate channels. The coexistence environment model of multiple SUs is considered as shown in Fig. 1. In this model, each SU has a fixed cognitive base station like a base station of cellular communication system.

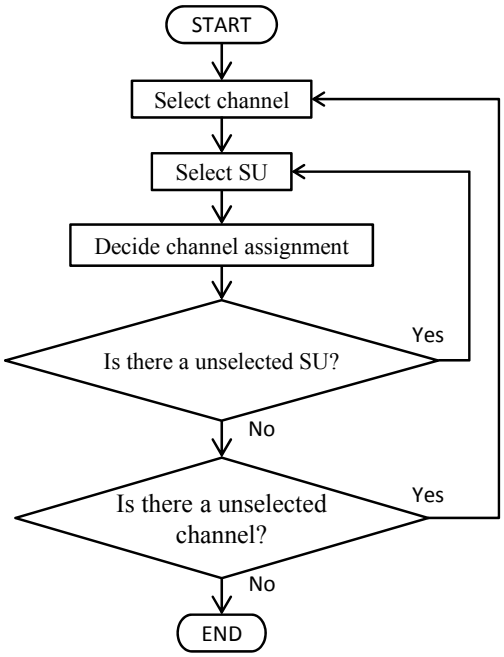


Fig. 2 Flow diagram the proposed method.

Furthermore, in our assumption, all SUs are registered to the central control server with parameter information as location information, transmit power, cell radius and minimum required SINR of the cell boundary. Communication link of each SU is established on the selected channel from multiple candidate channels. These cognitive base stations of the SUs can access only available channel assigned by the central control sever.

In this paper, the central control sever calculates channel conditions whether the WS can access the channel with satisfying the constraint or not, and only approved WS channels are listed as available channels for SU communication. Here, the mutual interference between the PU and each SU is ignored, because we consider the SU is located far enough away from the PU. Therefore, we just require considering spectrum sharing among SUs. The terminals of the SU are interfered by the cognitive base station of other SUs accessing the same frequency band. The central control server decides the target area for spectrum allocation based on the location information of all SU, and selects an anchor point as the origin O which is located one corner of the target area.

3. PROPOSED CHANNEL ASSIGNMENT METHOD

In this section, the detail of the proposed channel assignment method is described. In order to achieve high spatial efficiency of spectrum, it is necessary to minimize the relative distance between SUs accessing the same channel. Additionally, to maintain SU communication

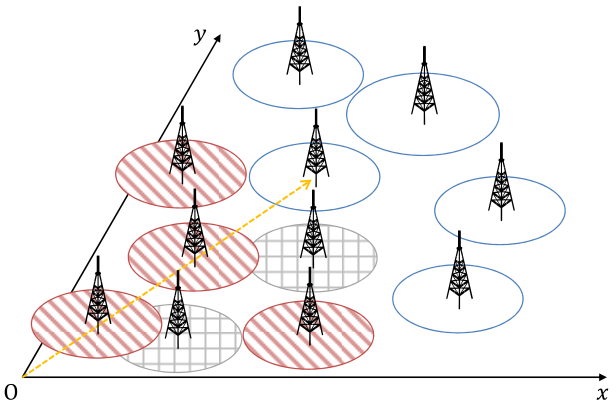


Fig. 3 Example of selecting the SU.

quality, the minimum required SINR of SUs has to be kept at the SU cell boundary. Figure 2 shows the flow diagram of the proposed channel assignment method. Three blocks drawn in Fig. 2, "Select channel", "Select SU" and "Decide channel assignment" are shown detail as follows.

- Select channel

The central control server selects the candidate channels from the order of a lower frequency channel in the target area. Let C_{sel} denote the channels which are selected by central control server.

- Select SU

The central control server selects the nearest SU from the origin O from the SUs that satisfy the condition. The condition is that the SU does not have any assigned channel and has not been selected in order to assign the channel C_{sel}.

Figure 3 shows the example of selecting SU. In Fig.3, SUs of cell with a diagonal pattern denote SUs that already have an assigned channel. SUs of cell with a grid pattern denote SUs that do not have any assigned channel C_{sel} yet due to not satisfying the condition of channel assignment. SUs of plain cell denote SUs do not have already assigned channel because those SUs do not have a chance of processing channel assignment yet. In the case of Fig. 3, the central control server selects the SU that is arrowed by the dashed line.

Let s denote the SU that is selected by the central control server for starting channel assignment can be stated as

$$s = \arg\min_{i \in \Pi} \left(\sqrt{{x_i}^2 + {y_i}^2} \right), \tag{1}$$

where i is arbitrary SU, x_i, y_i denote coordinates of i-th SU, Π is a set of unprocessed SUs. Here, i, j, k are different arbitrary SU.

- Decide channel assignment

The central control server decides whether the channel C_{sel} can be assigned to SU with satisfying the interference

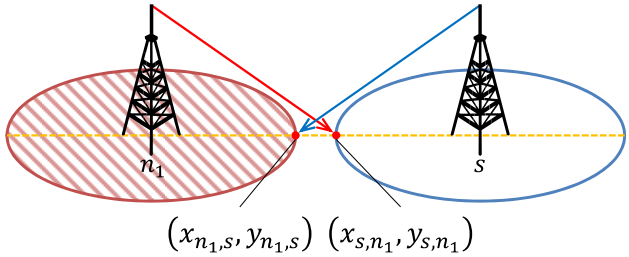


Fig. 4 Channel assignment decision ($n_A = 1$).

condition. Let s-th SU denote the SU which is selected by the central control server in "Select SU" process. The basic policy of the proposed method is that two adjacent SUs using the channel C_{sel} to s-th SU are chosen and the central control server checks the condition of channel assignment. The algorithm to check the condition for assigning the channel is different according to the number of SUs that already have the same channel C_{sel} in the field. Let n_A denote the number of SUs with the channel C_{sel}. The algorithms for channel assignment are classified into three cases: (i) $n_A = 0$, (ii) $n_A = 1$ and (iii) $n_A \geq 2$. The channel assignment algorithms of each case are described below.

(i) $n_A = 0$

In the case (i), there is no SU that already has the channel C_{sel}. In this case, the central control server unconditionally assigns the channel C_{sel} to the s-th SU.

(ii) $n_A = 1$

When the number of SUs that already have the channel C_{sel} is one, it is necessary to consider the mutual interference between the s-th SU and the SU has the channel C_{sel} in order to keep the SU communication quality. In this case, it is necessary to consider the SINR of both the s-th SU and the SU has the channel C_{sel}, because the channel assignment affects not only the s-th SU but also the SU has the channel C_{sel}. Thus, in the case (ii), the central control server makes the channel allocation decision to satisfy the interference constraint at two points of cell boundaries. In this paper, transmit signal is affected by only the propagation loss depended on the distance. The propagation loss is calculated by

$$L(d)\ [\mathrm{dB}] = -10 \log_{10} \left(\frac{\lambda}{4\pi d_0} \right)^2 + 10 n_{\mathrm{loss}} \log_{10} \left(\frac{d}{d_0} \right), \tag{2}$$

where λ is the wavelength of the carrier frequency, n_{loss} is propagation factor, d_0 denotes a reference distance, d is the distance from SU transmitter. Let $S_{\mathrm{BDRY},i}$ denote the signal power of i-th SU at cell boundary, $S_{\mathrm{BDRY},i}$ can be calculated as

$$S_{\mathrm{BDRY},i}\ [\mathrm{dBm}] = W_i\ [\mathrm{dBm}] - L\left(\sqrt{{h_i}^2 + {R_i}^2} \right) [\mathrm{dB}], \tag{3}$$

where R_i means the cell radius of i-th SU, W_i denotes the transmit power of i-th SU, h_i is the antenna height of i-th SU. Therefore, the allowable interference of i-th SU is given by

$$P_i = \frac{S_{\mathrm{BDRY},i}}{\gamma_i} - N, \tag{4}$$

where P_i denotes the allowable interference of i-th SU, γ_i means the minimum required SINR of i-th SU, N denotes the average noise power. In eq. (4), all parameters are a true value. In order to assign the channel under the SINR constraint, it is necessary to calculate the SINR of the SU at the closest cell boundary to other cognitive base station. Because, the closest cell boundary to other cognitive base station is the minimum SINR point within the cell coverage. Let n_1 denote the SU is the closest to s-th SU. Here, $(x_{i,j}, y_{i,j})$ denotes the closest cell boundary point of i-th SU to the cognitive base station of j-th SU. $(x_{i,j}, y_{i,j})$ is the intersection point of the cell boundary with the line connecting the centers of the two adjacent cells. Figure 4 shows the case of $i = s$ and $j = n_1$. In this case, (x_{s,n_1}, y_{s,n_1}) is the closest cell boundary point of s-th SU to the cognitive base station of n_1-th SU. Let $D_{i,j}$ denote the length of the line connecting the centers of the two cells. $(x_{i,j}, y_{i,j})$ can be written by

$$x_{i,j} = x_i + \frac{R_i}{D_{i,j}}(x_j - x_i), \tag{5}$$

$$y_{i,j} = y_i + \frac{R_i}{D_{i,j}}(y_j - y_i). \tag{6}$$

Since, the closest cell boundaries of n_1-th SU and s-th SU are expressed as (x_{s,n_1}, y_{s,n_1}), $(x_{n_1,s}, y_{n_1,s})$ illustrated in Fig. 4. In this paper, we consider the assignment conditions as the 95% interference limit based on SINR is the threshold for decision of channel assignment when the $n_A = 1$ for taking the SINR margin. The reason for this margin is that it is difficult to satisfy the condition of which the channel C_{sel} is assigned to s-th SU and there are possibility that the channel C_{sel} is not assigned to s-th SU in the case (iii) if the interference is nearly equal to allowable interference at cell boundary in the case (ii). From the above, if the interference from n_1-th SU is lower than $0.95P_s$ at (x_{s,n_1}, y_{s,n_1}) and the interference from s-th SU is lower than $0.95P_{n_1}$ at $(x_{n_1,s}, y_{n_1,s})$, the channel C_{sel} is assigned to the s-th SU by the central control server. Let $I_{i,j}^k$ denote the interference from the k-th SU to $(x_{i,j}, y_{i,j})$, it is given by

$$I_{i,j}^k[\mathrm{dBm}] = W_k[\mathrm{dBm}] - L\left(\sqrt{{h_k}^2 + (x_{i,j} - x_k)^2 + (y_{i,j} - y_k)^2}\right)[\mathrm{dB}]. \tag{7}$$

Therefore, the condition of which the channel C_{sel} is assigned s-th SU is expressed following equations,

$$\begin{cases} I_{s,n_1}^{n_1} \le 0.95P_s \\ I_{n_1,s}^{s} \le 0.95P_{n_1} \end{cases}. \tag{8}$$

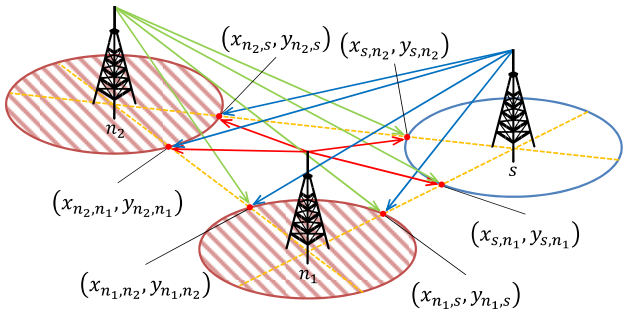


Fig. 5 Channel assignment decision ($n_A \ge 2$).

(iii) $n_A \ge 2$

In this case, two or more SUs already have the channel C_{sel} in the previous step. As well as the case (ii), it is necessary to consider the mutual interference among s-th SU and multiple SUs have the channel C_{sel} in order to keep SU communication quality. If the channel C_{sel} is assigned to the new SU, the mutual interference from the new SU affects all SUs that have the same channel C_{sel}. Therefore, in the case (iii), the central control server makes channel assignment decision whether the aggregated interference is less than the allowable interference at the affected cell boundaries. However, it is too complicate to calculate the interference situation caused by the effect of channel assignment results considering all SUs. Therefore, this paper proposes the simplified channel assignment evaluation for multiple SUs. In the proposed method, only two adjacent SUs using the same channel to s-th SU are chosen for the impact evaluation. Here, the SUs adjacent to s-th SU are decided by the closer distance SUs from the candidate SUs. In order to evaluate the interference, we have to consider six points at cell boundaries. Six points are six red dots that is intersection points of the line connecting the centers of the two SUs and each cell boundary, such as shown in Fig. 5. Let A is a set of the SUs that have the channel C_{sel}, n_1 is the closest SU to s-th SU, and n_2 denotes the second closest SU to s-th SU. Let n_1 and n_2 are given as follows.

$$n_1 = \arg\min_{i \in A}\left(\sqrt{(x_s - x_i)^2 + (y_s - y_i)^2}\right) \tag{9}$$

$$n_2 = \arg\min_{i \in A}\left(\sqrt{(x_s - x_i)^2 + (y_s - y_i)^2}\right)(i \ne n_1) \tag{10}$$

Namely, the six points are expressed as (x_{s,n_1}, y_{s,n_1}), $(x_{n_1,s}, y_{n_1,s})$, $(x_{n_1,n_2}, y_{n_1,n_2})$, $(x_{n_2,n_1}, y_{n_2,n_1})$, $(x_{n_2,s}, y_{n_2,s})$, (x_{s,n_2}, y_{s,n_2}) and illustrated in Fig. 5 and can be calculated by using eq. (5) and eq. (6). From the above, the condition of assignment channel to the SU can be written as following equations,

$$\begin{cases} I_{s,n_1}^{n_1} + I_{s,n_1}^{n_2} \le P_s \\ I_{n_1,s}^{s} + I_{n_1,s}^{n_2} \le P_{n_1} \\ I_{n_1,n_2}^{s} + I_{n_1,n_2}^{n_2} \le P_{n_1} \\ I_{n_2,n_1}^{s} + I_{n_2,n_1}^{n_1} \le P_{n_2} \\ I_{n_2,s}^{s} + I_{n_2,s}^{n_1} \le P_{n_2} \\ I_{s,n_2}^{n_1} + I_{s,n_2}^{n_2} \le P_s \end{cases}. \quad (11)$$

4. SIMULATION RESULTS

In this section, the performance of the proposed channel assignment method is evaluated through computer simulations. Here, we consider many SUs are uniformly distributed in the square area of 5×5 [km^2]. The same parameters are utilized among all SUs, such as transmit power, minimum required SINR and antenna height of a cognitive base station except cell radius. The cell radius of each SU is determined at random in the range of 100 ~ 150 [m]. Thus, in this environment, it is important to assign the channel considering the mutual interference in order to keep the required SINR under different coverage size. Five channels can be used for spectrum sharing among all SUs located in the same area. The center frequency of each channel can be selected from 500, 550, 600, 650 and 700 [MHz]. The other simulation parameters are shown in Table 1. The simulation results are obtained by averaging the thousand trials.

In simulations, we evaluate the ratio of the number of successful channels assigned to SUs to the total number of SUs in the area. In a channel assignment algorithm, it is important to keep the SINR constraint. In order to evaluate whether the proposed algorithm satisfies the minimum required SINR in the cell boundary of SUs that have assigned channel, the ratio of the satisfied minimum required SINR at the cell boundary is calculated by checking SINR per 1 degree angle at the cell boundary of the SU have assigned channel. Moreover, the spectral efficiency in a space domain is evaluated. In order to evaluate the spectral efficiency, spatial distribution of interference and noise power are derived in the area when the number of SUs is 130. For deriving the distribution of interference and noise power, the target area is divided into blocks of 5×5 [m^2] and the interference and the noise power at the center of each block is calculated in each channel.

In this simulation, random assignment method is chosen to compare with the proposed channel assignment method. In this random assignment method, the central control server does not exist. Each SU decides the access channel if the aggregate interference is lower allowable interference at cell boundary in all direction. In detail, as first step, the SU is selected randomly from all SU. Secondly, the selected SU select a channel at random. Thirdly, the selected SU decides the access channel based on whether condition is sufficient. If the condition is not sufficient, s-th SU randomly reselects other channel.

Table 1 Simulation parameters.

Transmit power of SU: W_i	20 [dBm]
Average noise power: N	-100 [dBm]
Minimum required SINR: γ_i	10 [dB]
Propagation factor: n_{loss}	3.5
Reference distance: d_0	10 [m]
Antenna height: h_i	10 [m]

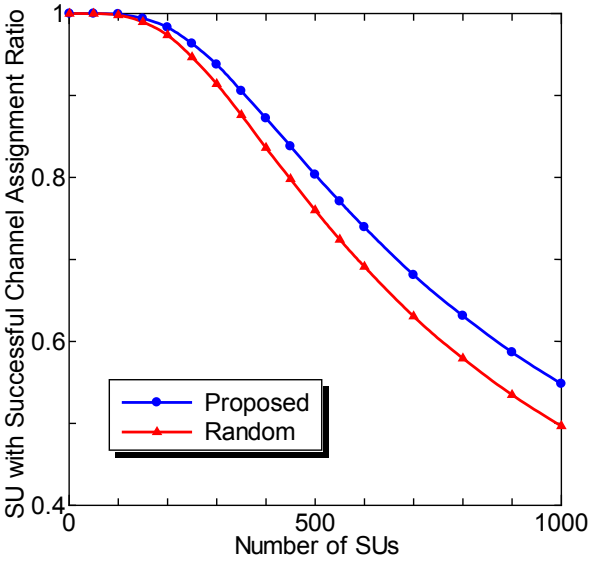


Fig. 6 SU with successful channel assignment ratio.

Figure 6 shows SU with successful channel assignment ratio. In Fig. 6, we can confirm that SU with successful channel assignment ratio of the proposed method is 5 percent higher than that of random channel assignment method when the number of SUs is 1000. Therefore, the proposed method can efficiently assign the channels to SU compared with random assignment method. It can be also seen that the improvement effect is increased as the number of SU increases.

Figure 7 shows the outage probability of the minimum required SINR at cell boundary 10 [dBm]. In Fig. 7, we can find the outage probability of SINR at cell boundary is less than 10 [dBm]. The proposed method achieves lower outage probability than that of random assignment method. Moreover, as the number of SUs increases, it is confirmed that the decrease of this ratio is small in the proposed method compared with random assignment method. Thus, the proposed method can improve spatial efficient channel assignment with satisfying the required SINR.

Figure 8 shows the distribution of the interference and the noise power in the case the number of SUs is 130. In Fig.8, there is one peak at around -85 [dBm] in the random

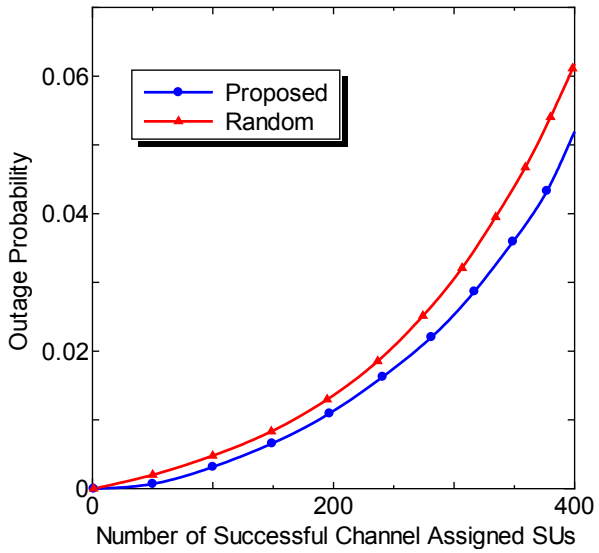


Fig. 7 Outage probability of the minimum required SINR at cell boundary.

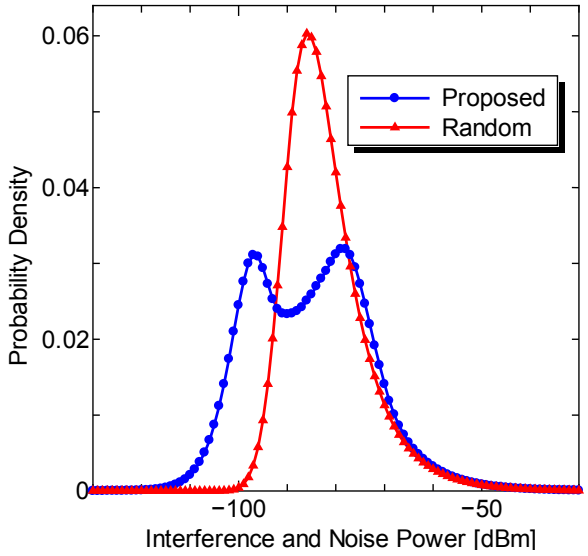


Fig. 8 Distribution of the interference and the noise power.

assignment method. On the other hand, in the proposed method, there are two peaks at approximately -80 and -100 [dBm]. In the proposed method, the central control server assigns low frequency channel preferentially. As a result, the channels assigned to SUs are packed densely in lower frequency channels and a peak appears at approximately -80 [dBm]. On the other hand, the number of SUs that with assigned channel is a few and a peak appears at approximately -100 [dBm]. This level is equivalent of the noise level. From these results, we can confirm that the proposed method enables the spatially dense utilization of spectrum and much room can be remained for the other wireless communication systems using the same spectrum as secondary systems.

5. CONCLUSION

In this paper, we propose a channel assignment method targeting for high spatial efficiency while keeping the minimum required SINR of each SU. In the proposed method, the central control server assigns WS channels to SUs based on relative distance among SUs and interference conditions. The central control server makes a channel allocation to shorten the distance among SUs with keeping aggregated interference less than the allowable interference at the cell boundaries of SUs. In our simulation results, we can confirm that it is possible to perform the channel assignment with high spatial efficiency while keeping every SU communication quality by using the proposed method.

6. REFERENCES

[1] Federal Communications Commission, “Spectrum policy task force report, FCC 02-155,” Nov. 2002.

[2] J. Mitola III and G. Q. Maguire, “Cognitive Radio: Making Software Radios More Personal,” *IEEE Pers. Commun.*, vol. 6, pp. 13-18, Aug. 1999.

[3] X. Kang, Y. Liang, A. Nallanathan, H. K. Garg and R. Zhang, “Optimal Power Allocation for Fading Channels in Cognitive Radio Networks: Ergodic Capacity and Outage Capacity,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 940-950, Feb. 2009.

[4] H. A. B. Salameh, M. M. Krunz and O. Younis, “MAC Protocol for Opportunistic Cognitive Radio Networks with Soft Guarantees,” *IEEE Trans Mobile Computing*, vol. 8, no. 10, pp. 1339-1352, Oct. 2009.

[5] Z. Chen, C. Wang, X. Hong, J. S. Thompson, S. A. Vorobyov, X. Ge, H. Xiao and F. Zhao, “Aggregate Interference Modeling in Cognitive Radio Networks with Power and Contention Control,” *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 456-468, Feb. 2012.

[6] S. Huang, X. Liu, and Z. Ding, “Decentralized Cognitive Radio Control Based on Inference from Primary Link Control Information,” *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 394-406, Feb. 2011.

[7] S. C. Jha, U. Phuyal, M. M. Rashid and V. K. Bhargava, “Design of OMC-MAC: An Opportunistic Multi-Channel MAC with QoS Provisioning for Distributed Cognitive Radio Networks,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 10, pp. 3414-3425, Oct. 2011.

[8] A. T. Hoang, Y. Liang and M. H. Islam, “Power Control and Channel Allocation in Cognitive Radio Networks with Primary Users' Cooperation,” *IEEE Trans. Mobile Computing*, vol. 9, no. 3, pp. 348-360, March 2010.

COGNITIVE SUPPRESSION OF MULTIPATH INTERFERENCE IN ANGULAR DOMAIN

Giulio Bartoli (Università degli Studi di Firenze, Florence, Italy; giulio.bartoli@unifi.it); Romano Fantacci (Università degli Studi di Firenze, Florence, Italy; romano.fantacci@unifi.it); Dania Marabissi (Università degli Studi di Firenze, Florence, Italy; dania.marabissi@unifi.it); Marco Pucci (Università degli Studi di Firenze, Florence, Italy; m.pucci@unifi.it); Claudio Armani (SELEX ES; claudio.armani@selex-es.com), Lorenzo Niccolai (TICom Consortium).

ABSTRACT

This paper deals with a cognitive approach used to reduce the interference generated by/toward another system operating on the same frequency band. A typical scenario is represented by an Underlay Cognitive Networks where a secondary system makes use of the same frequency band of the primary system but should transmit without affecting the primary system reception. However, this situation can arise also when an intentional jammer tries to destroy the primary system communications.

In this paper the interference is detected by using a spatial sensing approach instead of the classical spectrum sensing: multiple-antenna technology is considered to exploit the angle dimension as a new spectrum opportunity. In particular Multiple Signal Classification (MuSiC) algorithm is used to detect the Direction of Arrival (DoA) of the interference source. We assume actual propagation environments, with multipath components characterized by large angle spread, and actual antenna size with a limited number of elements. This leads to scenarios in which the number of resolvable directions is higher than the number of antennas that typically limits the DoA estimation capabilities.

DoA information is then used in transmission inserting nulls in the estimated directions thus protecting the useful information. The proposed scheme permits to increase the signal to interference ratio at the receiver side thus reducing the achieved error rate.

1. INTRODUCTION

The increasing development of wireless systems and the insufficient spectrum availability lead the need of an efficient frequency utilization. However, this can introduce interference issues that must be carefully addressed.

For example in next generation cellular systems inter-cell interference and interference arising from heterogeneous network deployment have become significant limiting factors [1], [2]. Similarly, in Cognitive Networks where the coexistence of two systems, primary and secondary, over the same frequency resources is allowed [3], the primary network communications should not be degraded by the secondary interference. Hence, secondary network should employ advanced communication techniques to exploit underutilized dimensions in the signal space. In particular, in Underlay Cognitive networks primary and secondary systems communicate simultaneously, but secondary system has cognitive capabilities in order to learn from the environment and adapt its transmission by means distribute power control algorithms and/or suitable resource allocation schemes [3] to not interfere with the primary system.

Interference issues concern also with intentional jammers that are used to destroy specific communications.

In all the cases specific countermeasures must be adopted in transmission and/or in reception in order to reduce the effects of interferences and intentional disturbs and increase the communication robustness.

With recent advances in multi-antenna technologies, space and angle dimensions can be exploited to reduce the interference both in transmission or in reception.

An useful approach is represented by the use of beamforming. With Cognitive Beamforming (CB) the secondary system adapts its transmission/reception in order to maximize the secondary system performance while the interference on the primary system receiver is minimized [4], [5]. Similarly schemes can be adopted to reduce the inter-cell interference [6] or jamming [7]. However, beamforming requires complex numerical solutions and the knowledge of all propagation channels that can be impracticable in some actual systems.

Another opportunity is represented by the estimation of the Direction of Arrival (DoA): the interference can be mitigated by steering spatial nulls in the direction of the interference sources (in reception) or the primary system (in transmission) [8], [9].

This approach presents many challenges in wireless communication systems, in particular in the DoA estimation when actual propagation conditions are considered.

This paper proposes a cognitive system able to reduce the interference generated by a secondary network on the primary network. The secondary terminal (ST) is equipped with a multiple antenna system and is able to estimate the DoA of different signal replicas received from the primary terminal (PT) during suitable sensing intervals. Then the ST transmits using a null steering algorithm to reduce the interference generated toward the PT. This scheme could be easily applied also to different scenarios, where the null steering is applied in reception (i.e., inter-cell interference or jamming). The analysis considers actual propagation conditions. In particular different propagation paths with different DoAs varying during time are taken into account. Hence, the effects of not ideal conditions due to a number of signal higher than the number of antennas and due to DoA variations are considered. The paper is organized as follows: Sec. II and Sec. III present the working hypothesis and the proposed cognitive system, respectively. Numerical results are provided in Sec. IV and, finally in Sec. V some conclusions are shown.

2. WORKING HYPOTHESIS

The intensive use of the spectrum leads to the coexistence of several wireless communication systems sharing the same radio resources. This is the case of Cognitive Networks where a secondary system operates on frequencies already assigned to a primary system adopting cognitive capabilities to listen the surrounding environment. In particular, in the *Interwave Cognitive* Networks the secondary system looks for resource availability, not used by the primary, in the time or frequency domain. While in *Underlay Cognitive* Networks the secondary system transmits simultaneously with the primary, but without affecting its communications. Hence, the secondary system must adapt its transmissions in order to generate a negligible interference toward the primary. This is the operating scenario considered in the paper. However it is important to underline that the approach here proposed could be used also at the receiving end by the interference victim.

In particular, in this paper we introduce the spatial domain as a new transmission opportunity.

The secondary system detects the presence of a primary system and its spatial information through a suitable sensing step. The sensing is periodically repeated taking into account the mobility of the terminals. We set the sensing period equal to 1 *ms*.

The ST is equipped with an antenna system consisting of omnidirectional elements, separated by $d = \lambda/2$, where λ is the wavelength. The number of antenna elements is a trade-off between performance and size of the device. Indeed, it is known in literature that increasing the number of antenna elements allows to improve the accuracy in the beams steering and in the detection of the directions of arrival.

The carrier frequency of the secondary system is a crucial key in the choice of the number of antenna elements. For a given spacing between the antenna elements and low operational frequencies, good performance can be achieved only with unfeasible dimensions of the device. For these reasons, we select actual conditions. The ST is equipped with an Uniform Linear Array (ULA) with $L = 4$ elements, and the operating frequency is equal to f_0=2GHz resulting in a linear dimension of the array of D=21cm.

The antenna elements are spatially correlated and therefore the received signal can be used to perform DoA estimation and digital beamforming. These techniques are of particular interest when the signal of the primary devices is received through a series of temporal replicas which must be characterized in the spatial domain. For actual mobile environments where devices are able to move, this modeling is necessary because different paths are originated from multiple scatterers and obstacles existing between primary and secondary and they can temporarily change their response. In general a specific spatial relationship between the arrival directions of the different replicas doesn't exist but this is highly dependent on the specific environment and thus the assumption that the paths are randomly distributed in $[0, 2\pi]$ is justified [10]. The number of resolvable paths is a function of the sampling frequency of the receiving system and the channel model taken into consideration. In particular, the impulse response of the channel is modeled according to the tapped-delay-line described in M.1225 ITU-R Recommendation [11].

The "A-Channel" model is characterized by M multiple propagation paths and **h** contains the channel coefficients

$$\mathbf{h} = [\alpha_1 e^{j\varphi_1}, \alpha_2 e^{j\varphi_2}, \dots, \alpha_M e^{j\varphi_M}]^T$$

where α_i and φ_i are modeled as independent random variables following Rayleigh and uniform. The i-th replica on the antennas forms the angle θ_i with the array perpendicular. The temporal delay of the i-th signal between the consecutive elements of antenna system is

$$\tau = \frac{d \sin(\theta_i)}{c} = \frac{\sin(\theta_i)}{2 f_0}$$

where f is the carrier frequency and c is the speed of light. Considering a narrowband signal, $\tau \gg T_S$, where T_S is the sampling period, the arriving signal phase is rotated by $2\pi f_0 \tau$. Hence, the paths are independent of each other and the n-th sample of the multiplex received on the l-th antenna is

$$r_l[n] = \sum_{i=1}^{M} x[n - t_i]\alpha_i e^{j[\phi_i + (l-1)\pi \sin(\theta_i)]} + v_l[n]$$

where t_i is the delay due to scattered propagation of the i-th replica, $x[n]$ is n-th sample of the signal transmitted by the

primary system and $v_l \sim \mathcal{N}(0, \sigma_v^2)$ is the AWGN noise which is independent among the antennas. Denoting $\mathbf{s}(\theta)$ as the steering vector, where $s_l(\theta) = e^{j\pi(l-1)\sin(\theta)}$ and $l = 1, 2, \dots, L$ then it can steer the antenna pattern on the direction θ. The matrix containing the steering vectors of the incoming signals is $\mathbf{S} = [\mathbf{s}(\theta_1), \mathbf{s}(\theta_2), \dots, \mathbf{s}(\theta_M)]$ and the arriving signals model can be represent in matrix form by

$$\mathbf{r} = \mathbf{S} \cdot \mathrm{diag}(\mathbf{h}) \cdot \mathbf{x} + \mathbf{v}$$

as $\mathbf{x} = [x[n - t_i]]^T$ and $\mathrm{diag}(\cdot)$ denotes the diagonal matrix. The number of replicas arriving to the receiver varies as a function of the sampling frequency of the secondary system, as shown in the Table 1

$f_S(MHz)$	2	4	7.5	15	30	40
Model paths	1	2	3	4	6	6

Table 1: working sampling frequencies and related numbers of paths

In our performance analysis, the outdoor model was selected that is suitable for many scenarios. Finally, the temporally related channel samples of $\mathbf{h}$ are filtered to obtain a relationship between the different directions of the arrival of the signal. First a set of K angles of arrival is generated through a random distribution in the $[0, \pi]$ interval and then it is interpolated by a k factor related to the devices speed [12]. In particular k can be derived as

$$k = \frac{f_S}{f_{Ch}}$$

where f_{Ch} can be calculated as a function of the wavelength, the speed of primary device and a suitable parameter A:

$$f_{Ch} = \frac{1}{T_C} = \frac{\lambda A}{v}$$

In this manner some representative values of k have been set, as shown in Table 2

	k	v [km/h]
No mobility	Inf	0
Low mobility	1500	5
Medium mobility	300	15
High mobility	150	25

Table 2: channel speed parameters

3. PROPOSED COGNITIVE SYSTEM

In the described context the secondary system must coexist with the primary system operating on the same area. The most critical scenario occurs when the two systems use the same carrier frequency and bandwidth. For this reason, this condition has been considered as a worst case scenario and it has been used to obtain numerical results. The purpose of the proposed system is to identify a technique that enables the secondary system to share resources without causing interference and other degradations in the performance of the devices already deployed. Initially, the secondary system must determine if the primary system is present or not in its coverage area through a sensing stage. The knowledge of the noise power in the absence of signal allows it to detect the presence of a primary system. If it is identified it is necessary to estimate the direction of arrival of the different signal replicas received from different propagation paths. This is done by using the antenna array with which the secondary system is equipped. The ST must periodically repeat this step to update the tracking relating to the mobility of detected devices. When the directions of arrival have been identified it is possible to provide this information to the Null Steering algorithm which will include some nulls in suitable positions of the radiation pattern.

3.1. DOA Estimation

In literature are proposed many algorithms to identify the direction of arrival of the incident signal on the antennas [13], [14]. The algorithms based on the autocorrelation matrix decomposition in its eigenvectors provide an efficient compromise between accuracy and resolution of performance and reduced computational complexity [15]. The autocorrelation matrix of the received signal is defined as

$$\begin{aligned}\mathbf{R_r} = E[\mathbf{r}\mathbf{r}^H] &= E[\mathbf{S}\,\mathrm{diag}(\mathbf{h})\mathbf{x}\mathbf{x}^H\,\mathrm{diag}(\mathbf{h})^H\,\mathbf{S}^H] \\ &= \mathbf{S}\mathbf{P}\mathbf{S}^H + \sigma_v^2\mathbf{Id}_L\end{aligned}$$

where $(\cdot)^H$ represents hermitian operator, $\mathbf{Id}_L$ is the identity matrix with dimension $L \times L$ and $\mathbf{P}$ is defined as

$$\mathbf{P} = E[\mathrm{diag}(\mathbf{h})\mathbf{x}\mathbf{x}^H\mathrm{diag}(\mathbf{h})^H]$$

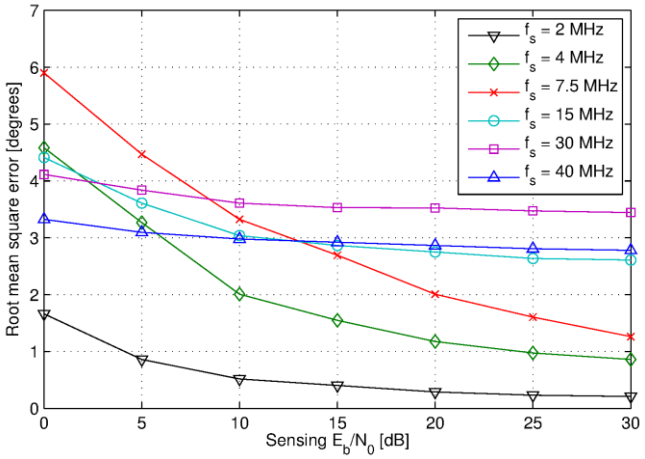


Figure 1: DOA estimation error for different sampling frequencies.

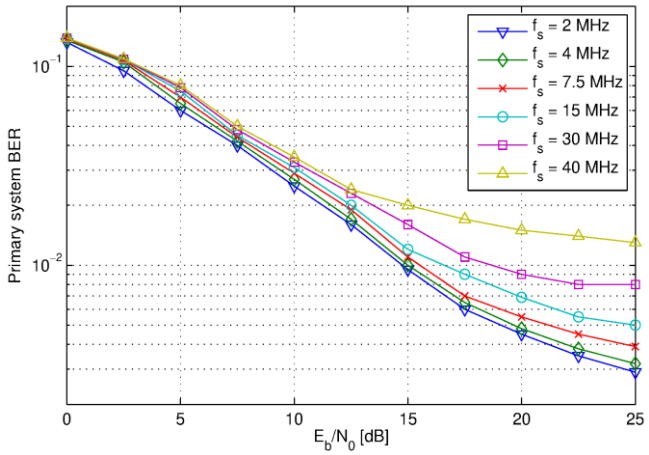


Figure 2:Primary system BER for different sampling frequencies.

The $\mathbf{R_r}$ decomposition in eigenvectors identifies two disjointed eigenspaces: the first is composed of the eigenvectors of the signal and is named *signal subspace* $\mathbf{U}_S$ and the latter is named *noise subspace* $\mathbf{U}_N$ and it is made up of the remaining eigenvectors. The eigenvectors are sorted according to the value of its eigenvalue and the M largest eigenvalues are associated with the M signal replicas. Therefore the two eigenspaces have dimension $L \times M$ and $L \times (L - M)$ and all possible steering vectors belong to one of the two subspaces, in particular the steering vectors related to signal DOAs belong to the signal subspace. Among all the steering vectors the MuSiC algorithm looks for ones that are orthogonal to the noise subspace, maximizing the function

$$P_{SM}(\theta) = \frac{1}{\|\mathbf{s}^H(\theta)\mathbf{U}_N\|}$$

where $\|\cdot\|$ is the norm of the vector. The function $P_{SM}(\theta)$ has a very wide codomain and the identification of the maxima can be hard. The introduction of a logarithm that compresses the shape and two derivatives which underline the curve concavities allow to achieve better performance resulting in a DOA estimation which is much more concentrated around the correct value. Hence we have

$$P'_{SM}(\theta) = \frac{d^2(\log_{10} P_{SM}(\theta))}{d\theta^2}$$

The identification of the maxima is based on the detection of local maxima greater than an suitable threshold. The MuSiC algorithm provides performance that depends on the noise overlapping on the signal and on the snapshot dimension. The snapshot represents the amount of signal samples received through which the autocorrelation matrix is estimated. Another key contribution is the number of antenna elements in the array: the maximum number of paths that are identifiable by Spectral MuSiC and by other algorithms of the same family is $L - 1$ because otherwise the decomposition in two disjointed subspaces is unfeasible. It is therefore evident that a greater number of antennas allows to identify a high number of DOA angles but it also increases the overall receiver dimensions and the computational complexity of the estimation. In an actual environment it can't be taken for granted, except in particular scenarios, that the several replicas come from neighboring directions and therefore they may be considered as separate signals. The Spectral MuSiC takes into account that the number of paths to be searched is $K = L - 1$. All angles of arrival can be detected as long as the number of paths M is smaller than $L - 1$, whereas the estimates of DOAs are limited to K when $M > L - 1$.

3.2. Null Steering

The information obtained by the algorithm of DOA estimation is used as input for subsequent cognitive phase. The secondary system needs to adjust its radiation pattern by placing some nulls to avoid interference in the directions just detected. This change is achieved by means of a pre-processing of the signal, in particular the symbols to be transmitted are multiplied by a weight vector $\mathbf{w}$ with dimension $L \times 1$.

The Null Steering algorithm is based on the same mathematical construct used by the classical beamformer. The weights are complex numbers that modify the amplitude and the phase of the output signal transmitted from each antenna. This allows algorithm to adjust the value of the radiated power in a certain direction and then it generates a set of K nulls. The power radiated after the pre-processing doesn't vary and therefore the secondary system maintains coverage performance similar to the previous case. This algorithm inherits the main limitation from the MuSiC and then it isn't able to generate more than $L - 1$ nulls [16]. In this scenario, we used the algorithm described in [13], that is

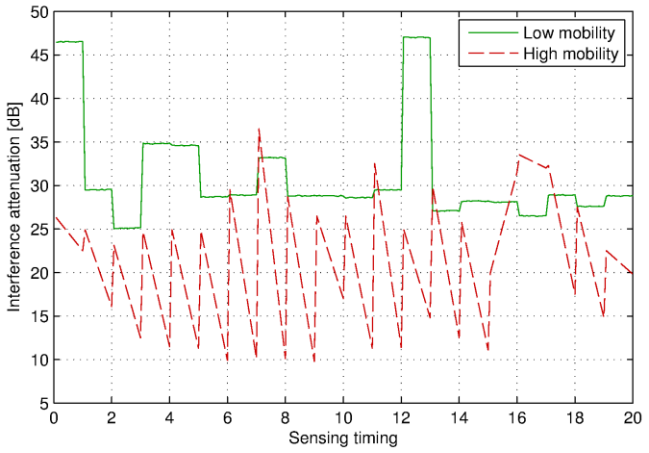


Figure 3: Interference attenuation on the primary user directions.

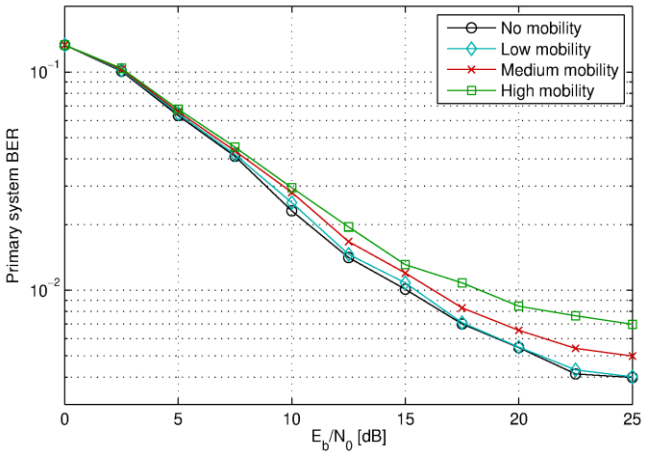


Figure 4: Primary system BER for different channel speeds.

$$\mathbf{w}^H = \mathbf{c}^H \mathbf{A}^H (\mathbf{A}\mathbf{A}^H)^{-1}$$

The weights $\mathbf{w}$ are obtained by imposing the steering vector $\mathbf{s}(\theta_0)$ equal to 1 and the steering vectors where null beams are required, $\mathbf{s}(\theta_1), \dots, \mathbf{s}(\theta_K)$, equal to 0. Let us denote the matrix containing the steering vectors of interest as $\mathbf{A} = [\mathbf{s}(\theta_0), \mathbf{s}(\theta_1), \mathbf{s}(\theta_2), \dots, \mathbf{s}(\theta_K)]$ and the vector $\mathbf{c} = [1, 0, \dots, 0]^T$ which contains $K + 1$ elements. The system can perform some nulls depending on the number of paths estimated by DOA algorithm.

The obtained pattern is very effective because nulls are very deep. Its shape is similar to a notch filter which excludes the transmission in the θ_i directions and the values of interference mitigation are about same decades. The limit due to the number of feasible nulls can be overcome by increasing the antenna elements which are installed on the secondary system, as explained in [13], [16] and [17]. This introduces the same problems addressed about the eigenstructure algorithms used for the DOA estimation. Anyway, when the number of paths is greater than $L - 1$, the nulls can be placed in the direction of the strongest sources of interference.

4. NUMERICAL RESULTS

In order to evaluate the performance of the proposed scheme, we resorted to a numerical approach by computer simulations. The results are provided in this section.

The center frequency considered here was set to 2 GHz, while different sampling frequency have been taken into account. Indeed, as explained in the Sec. 2, different sampling frequencies may lead to a different numbers of paths, according to [11].

We start our analysis by investigating the DOA estimation accuracy. Figure 1 shows the root mean square error of the MuSiC algorithm as a function of the sensing E_b/N_0 for different sampling frequencies. It is possible to see that the higher is the number of paths, the worse is the performance. Moreover, if the number of resolvable path is greater than the number of antennas, the accuracy does not improve significantly when the E_b/N_0 grows. However, a higher sampling frequency allows to collect a larger number of samples during the sensing interval, hence, if the number of path is the same (as for 30 and 40 MHz), it leads to better performance.

Figure 2 presents the performance of the primary link (primary transmitter and primary receiver) in terms of bit error rate (BER) as a function of the E_b/N_0 at the primary receiver for different sampling frequencies. By a joint analysis of Figure 1 and Figure 2, it is evident that the performance degradation of the primary link is strictly dependent on the DOA estimations of the secondary receiver.

Relevant interest relies on the sensing and mitigation algorithm capability to adapt to channel variations. Hence, the estimation has to be performed on a periodical base.

Figure 3 presents an example of interference attenuation on the primary user directions introduced by the proposed scheme during the time. Two different channel speed have been considered: when the speed is slow (i.e., 5 km/h), the algorithm is able to follow the DOA variations, in fact the SIR maintains constant over the time until the next estimation. On the other hand, when the channel speed is higher (25 km/h) the estimation accuracy has a general deterioration and gets worse until the following sensing period. However, in both the cases the proposed scheme allows a remarkable improvement of the SIR, and hence the secondary transmission has a small impact on the performance of the primary receiver. This is evident from Figure 4, where the BER of the primary system is shown for different channel speeds. It is possible to see that the primary performance degradation is limited to the case of high E_b/N_0 values at the primary receiver, and is modest

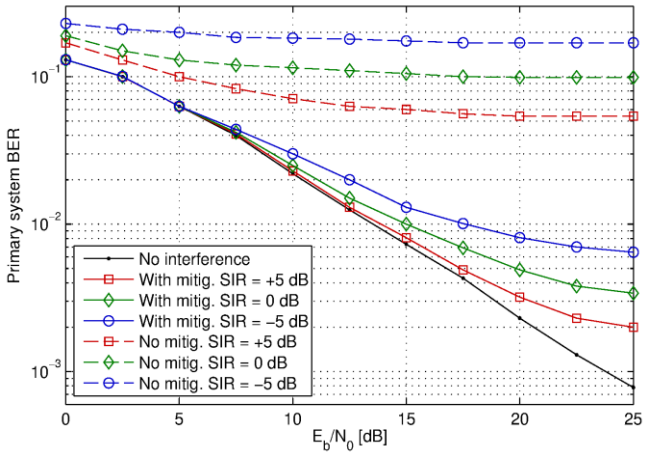


Figure 5: Primary system BER for different SIR.

even in that case. Finally, Figure 5 provides the comparison of the primary system performance with and without the proposed interference mitigation scheme for different value of mean signal-to-interference ratio (SIR, the power of the primary signal is assumed to be normalized). From this figure is evident the remarkable improvement due to the jointed use of MuSiC and null steering algorithm: without a suitable interference cancellation scheme, the performance quickly reaches a floor, and the BER does not decrease even if the E_b/N_0 grows. On the other hand the proposed scheme allows to approach the curve relative to the performance without interference.

5. CONCLUSIONS

This paper has dealt with a cognitive scheme used to reduce the interference generated toward another system. The interference detection is based on a spatial sensing approach: beamforming technology has been considered to exploit the angle dimension as a new spectrum opportunity. The paper has focused on the Multiple Signal Classification (MuSiC) algorithm, used to detect the Direction of Arrival (DOA) of the interference source, and on a null-steering technique to cancel interference. The analysis has considered an actual scenario: the propagation environment has been characterized by large angle spread multipath components, while actual antenna size with a limited number of elements has been considered at the receiver.
The proposed scheme has been evaluated by means of computer simulations, showing the capability to increase the signal to interference ratio at the receiver side, thus reducing the achieved error rate.

6. REFERENCES

[1] A. Hamza, S. Khalifa, H. Hamza, and K. Elsayed, "A Survey on Inter-Cell Interference Coordination Techniques in OFDMA-Based Cellular Networks", *IEEE Communications Surveys & Tutorials*, vol.PP, no.99, pp.1,29, 2013.

[2] A. Zyoud, M. H. Habaebi, J. Chebil, and M. R. Islam, "Femtocell interference mitigation", *2012 IEEE Control and System Graduate Research Colloquium (ICSGRC)*, vol., no., pp.94,99, 16-17 July 2012

[3] Ying-Chang Liang, Kwang-Cheng Chen, G. Y. Li, and P. Mahonen, "Cognitive Radio Networking and Communications: An Overview", *IEEE Transactions on Veh. Techol.*, vol.60, no.7, pp.3386,3407, Sept. 2011

[4] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process*., vol. 2, no. 1, pp. 88–102, Feb. 2008.

[5] S. Yiu, C.-B. Chae, K. Yang, and D. Calin, "Uncoordinated Beamforming for Cognitive Networks" *IEEE Trans. Commun*., vol. 60, no. 5, pp.1390–1397, May 2012.

[6] J. Uk, L. Kang-Yong, C. Kee-Seong, R. Won, "Transmit Beamforming Based Inter-Cell Interference Alignment and User Selection with CoMP," *2010 IEEE 72nd Vehicular Technology Conference Fall (VTC 2010-Fall)*, vol., no., pp.1,5, 6-9 Sept. 2010.

[7] W. Xin-Huai, S. Xiao-Wei, L. Ping, B. Yan-Fu, L. Bo, L. Rui, and L. Hao-Jia, "Smart antenna design for GPS/GLONASS anti-jamming using adaptive beamforming," *2010 International Conference on Microwave and Millimeter Wave Technology (ICMMT)*, vol., no., pp.1149,1152, 8-11 May 2010

[8] E. Yaacoub and Z. Dawy, "Enhancing the performance of OFDMA underlay cognitive radio networks via secondary pattern nulling and primary beam steering," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2011, pp. 1476–1481.

[9] J. Xie, Z. Fu, and H. Xian, "Spectrum sensing based on estimation of direction of arrival," *in International Conference on Computational Problem-Solving (ICCP),* Dec. 2010, pp. 39–42.

[10] Q. Spencer, B. Jeffs, M. Jensen, and A. Swindlehurst, "Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel," *IEEE J. Sel. Areas Commun.*, Vol. 18, No. 3, pp. 347–360, 2000.

[11] ITU-R, "Guidelines for evaluation of radio transmission technologies for IMT-2000," International Telecommunication Union, Recommendation M.1225, Feb. 1997.

[12] Y. S. Cho, J. Kim, W. Y. Yang and C. G. Kang, *MIMO-OFDM Wireless Communications with MATLAB*, John Wiley & Sons, Singapore, Nov. 2010

[13] L. Godara, "Application of antenna arrays to mobile communications. II. Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, Vol. 85, No. 8, pp. 1195–1245, Aug. 1997.

[14] T. Lavate, V. Kokate, and A. Sapkal, "Performance analysis of MUSIC and ESPRIT DOA estimation algorithms for adaptive array smart antenna in mobile communication," *in 2010 Second International Conference on Computer and Network Technology (ICCNT)*, Apr. 2010, pp. 308–311.

[15] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, Vol. 34, No. 3, pp. 276–280, Mar. 1986.

[16] B. Friedlander and B. Porat, "Performance analysis of a null-steering algorithm based on direction-of-arrival estimation,"

IEEE Trans. Acoust., Speech, Signal Process., Vol. 37, No. 4, pp. 461–466, Apr. 1989.

[17] R. Qamar and N. Khan, "Null steering, a comparative analysis," *in IEEE 13th International Multitopic Conference*, 2009. INMIC 2009., Dec. 2009, pp. 1–5.