



## 216715 NEWCOM<sup>++</sup>

### DR.4.2

### Mid-Project Report

**Contractual Date of Delivery to the CEC:** T0+18

**Actual Date of Delivery to the CEC:** T0+18

**Editor(s):** J. Sayir (FTW), E. Arikan (Bilkent), C. N. Manchon (AAU), P. Tyczka (PUT)

**Participating institutions:** Bilkent/Kadir Has, Technion, NKUA, CNIT, CNRS, LNT, RWTH Aachen, UCL/UGent, FTW/VUT, PUT, Chalmers, AAU

**Contributors:** E. Arikan (Bilkent), J. Bhatti (UGent), P. Duhamel (CNRS), B.H. Fleury (AAU), A. Graell i Amat (CNRS), T. Gucluoglu (Bilkent/KH) G.E. Kirkelund (AAU), I-W. Lai (RWTH), C.N. Manchon (AAU), G. Masera (CNIT), M. Moeneclaey (UGent), G. Montorsi (CNIT), Z. Naja (CNRS), E. Panayirci (Bilkent/KH), S. Papaharalabos (NKUA/NOA), C. Poulliat (CNRS), P. Remlein (PUT), J. Sayir (FTW), S. Schwandter (VUT), M. Senst (RWTH), M. Sybis (PUT), P. Tyczka (PUT), D. Zhang (RWTH),

**Internal Reviewer(s):** S. Benedetto (ISMB)

**Workpackage number:** WPR.4 Iterative Receivers for Wireless Communications

**Nature:** R

**Total Effort Spent:** 32

**Dissemination Level:** Pu

**Version:** 1.0

#### Abstract:

This deliverable documents the status of WPR4 after 18 months, mixing past achievements with plans for ongoing and future collaborations. The work is concentrated along three main directions, namely polar codes, joint receivers, and reduced complexity receivers. The first direction addresses a recent coding technique developed by NEWCOM<sup>++</sup> partners that is receiving much attention internationally because it is possibly the first constructive coding method to achieve capacity for certain channels. The aim of the work within WPR4 is to determine whether this method has practical implications. The second direction addresses the application of iterative techniques to components in a receiver beyond the decoder. Parallels between iterative decoders and other iterative techniques are exploited, and tools are developed to optimize the performance of receivers that perform decoding, estimation, detection and synchronization iteratively. The third direction addresses specifically the design of iterative receivers under complexity constraints, where the aim is to obtain a good performance comparable to the optimal algorithms while using a structure that is more amenable to practical implementation. This direction is of prime importance for the adoption of iterative techniques in practical communication standards.

**Keyword list:** Iterative Processing, Turbo Codes, LDPC Codes, Convex Optimization, Information Geometry, Factor Graphs, Polar Codes, Non-Binary Codes, Bit-Interleaved Coded Modulation, EXIT Charts, Density Evolution, Synchronization, Phase Noise Estimation, Turbo Coded Modulation, Quantization, Clipping

**TABLE OF CONTENTS**

<b>1</b>	<b>Introduction</b>	<b>4</b>
1.1	Abbreviations . . . . .	5
<b>2</b>	<b>Polar Codes</b>	<b>7</b>
2.1	General Background on Polar Coding . . . . .	7
2.1.1	Overview of polar coding . . . . .	7
2.2	Brief literature survey on polar coding . . . . .	7
2.3	Polar coding research subjects under WP4 . . . . .	8
2.4	Polar coding for general classes of channels . . . . .	9
2.5	Polar codes over real and complex numbers . . . . .	10
2.5.1	A first method that fails . . . . .	11
2.5.2	Encoding with the mod 1 sum . . . . .	13
2.6	Polar coding and compressed sensing . . . . .	14
2.7	Non-binary Iterative decoding of Polar codes. . . . .	15
2.8	ML decoding performance of polar codes . . . . .	15
2.8.1	Trellis representation of polar codes . . . . .	16
2.8.2	Simulations . . . . .	16
2.8.3	Summary and conclusions . . . . .	18
2.9	Two-dimensional polar codes . . . . .	18
2.9.1	Polar code construction . . . . .	19
2.9.2	Product coding . . . . .	20
2.10	Are Reed-Muller codes capacity-achieving codes under ML decoding? . . . . .	22
2.11	Summary and Planned Collaboration . . . . .	23
<b>3</b>	<b>Joint Receivers</b>	<b>24</b>
3.1	Introduction . . . . .	24
3.2	Optimum receiver approach . . . . .	25
3.3	The Divergence Minimization Principle . . . . .	27
3.3.1	The DM Framework . . . . .	28
3.3.2	Application of the DM framework to channel estimation with overlapped pilots . . . . .	29
3.3.3	Application of the DM framework to channel estimation and multi-user detection . . . . .	30
3.4	Advanced Signal Processing Algorithms for synchronization of Wireless Communication systems . . . . .	31
3.4.1	Introduction . . . . .	31
3.4.2	Main knowledge gaps . . . . .	32
3.4.3	An example: A Monte-Carlo Implementation of the SAGE Algorithm for Joint Soft Multiuser and Channel Parameter Estimation . . . . .	33
3.4.4	General Conclusions . . . . .	35
3.5	Phase Noise Estimation . . . . .	35
3.5.1	Introduction . . . . .	35
3.5.2	Phase noise estimation using a basis expansion model . . . . .	36
3.5.3	Phase noise estimation using Monte-Carlo sampling . . . . .	36
3.5.4	Collaboration and future work . . . . .	36
3.6	Information Geometry . . . . .	37
3.6.1	Introduction . . . . .	37
3.6.2	Geometrical interpretation of iterative algorithms . . . . .	37
3.7	Collaboration and Future Plans . . . . .	38

<b>4</b>	<b>Reduced Complexity Decoding</b>	<b>39</b>
4.1	Reduced complexity algorithms for decoding turbo codes and turbo trellis-coded modulation . . . . .	39
4.1.1	New algorithms investigated . . . . .	39
4.1.2	Simulation results . . . . .	41
4.1.3	Complexity estimation . . . . .	44
4.1.4	Future work . . . . .	48
4.2	Space time coded systems with reduced complexity receivers over frequency selective fading channels . . . . .	48
4.2.1	Space time turbo codes . . . . .	48
4.2.2	Space time turbo coded systems with reduced complexity receivers over frequency selective fading channels . . . . .	49
4.3	Applications of the Role Model framework to reduced complexity decoder design . . . .	51
4.4	Log-likelihood processing . . . . .	52
4.4.1	Log-likelihood ratio clipping . . . . .	52
4.4.2	Selective and modified LLR updates . . . . .	52
4.5	Optimization of the redundancy in channel coding and application to layered channel coding	53
<b>5</b>	<b>Conclusion</b>	<b>54</b>

## 1 INTRODUCTION

The present deliverable is the second deliverable issued within the lifetime of workpackage WPR4 of NEWCOM++. The first deliverable concerned itself with establishing the state of the art in the area of iterative receivers to serve as a basis for further research within the network of excellence. In contrast, this second deliverable is a mid-term report concentrating on partial results and plans for further research. As such, it is neither quite all about past achievements, nor exclusively about future plans, but contains a mixture of both, in line with the status of the workpackage after its first 18 months of activities.

For a global perspective on the workpackage evolution, it is helpful to review the history of the workpackage, starting with the preparation work in the pre-project phase when the proposal was being written, and ending with the preparation of this deliverable. During the writing phase, this workpackage was positioned as an “interface” topic, researching techniques that are relevant for many of the other workpackages in the physical and networking layers, thereby attracting the interests of many partner institutions and many in related workpackages. Consequently, the workpackage description contains a large list of topics reflecting the broad interests of those involved, structured in 8 tasks that were grouped into 4 research directions. After the initial workpackage meeting in Munich in February 2008, it was decided to concentrate on 4 slightly modified general directions: theory of iterative systems; code design; low complexity and implementation issues; and synchronization. These 4 directions are reflected in the state-of-the-art document DR4.1. Several topics of central interest were divided across these 4 directions, e.g., polar coding was treated both under “code design” and under “low complexity and implementation”.

Collaborative work in some of these areas took off as intended and has produced a number of joint publications, in particular on phase noise estimation / synchronization, and on low complexity decoder architectures. In other fields, various circumstances made it more difficult for the collaboration to take off, like for example the departure of some project participants from their partner institutions or the complete novelty of a topic. This was recognized during a monitoring video meeting in the autumn of 2008 and it was decided to hold another plenary workpackage meeting in January 2009 in Istanbul. This meeting lasted 3 days and its aim was to offer an opportunity for new workpackage participants to meet and launch new collaborations. Following this, a consolidation of tasks was undertaken which is reflected in the present deliverable. The current 3 tasks of the workpackage, reflected in the 3 technical chapters of this deliverable, are:

**Polar Codes** This novel coding technique was devised by Erdal Arıkan from Bilkent and presented prior to publication at the initial WP meeting in Munich, where it raised great interest among WP participants. Meanwhile it has triggered an intense international response, with several papers on the subject submitted and published by groups inside and outside NEWCOM++, justifying our choice to dedicate a whole task to this technique within WPR4. Collaborations on this topic are still at an early stage due to the fact that the topic was not part of the initial proposal. Several researchers from CNIT/Polito, CNRS/ENSEA and FTW are interested in collaborating with the initiator from Bilkent and intense exchanges took place at our recent meeting in Istanbul, that is expected to bear fruits within the next reporting period. The information in this deliverable highlights the known results on polar coding upon which future work will elaborate, and reports the conclusions from the discussions held in Istanbul.

**Joint Receivers** This task reunites all activities relating to iterative receiver structures beyond decoding, where different parts of the receiver exchange messages in view of achieving a global efficiency as opposed to efficiency only at the level of the decoder. The consolidation of this task is in response to feedback received during the first year evaluation of NEWCOM++, where it was observed that activities in WPR4 targeting joint processing within the receiver were split across tasks and not sufficiently emphasized. The information in the corresponding chapter documents on one hand mature work that has been the subject of joint publications, e.g. collaboration by UCL/UGent and Bilkent/Kadir Has on phase noise estimation, and new initiatives and associated research plans that have been launched during the meeting in Istanbul.

**Reduced Complexity Decoding** This task is unchanged from the previous deliverable and gave rise to a fruitful collaboration between NKUA/NOA and PUT. In addition, ideas on quantization that were presented at the initial Munich meeting and in the previous deliverable gave rise to individual publications by FTW/VUT. Furthermore, a new collaboration between PUT and Bilkent/Kadir Has was initiated in Istanbul and the associated research plan is presented in the corresponding chapter.

### 1.1 Abbreviations

Throughout this report, abbreviations are defined in the text where they are used. For convenience, we also provide a list of abbreviations used:

**APP** A-Posteriori Probability

**AWGN** Additive White Gaussian Noise

**BCJR** Bahl Cocke Jelinek Raviv

**B-DMC** Binary-input Discrete Memoryless Channel

**BEC** Binary Erasure Channel

**BER** Bit Error Rate

**BICM** Bit-Interleaved Coded Modulation

**BPSK** Binary Phase Shift Keying

**BSC** Binary Symmetric Channel

**CDMA** Code Division Multiple Access

**DFE** Decision Feedback Equalization

**DM** Divergence Minimization

**DMC** Discrete Memoryless Channel

**EM** Expectation-Maximization

**EXIT** EXtrinsic Information Transfer

**FER** Frame Error Rate

**FS** Frequency Selective

**GF** Galois Field

**ISI** Inter Symbol Interference

**KL** Kullback Leibler

**LDPC** Low-Density Parity Check

**LLR** Log Likelihood Ratio

**LMMSE** Linear Minimum Mean Square Estimation

**Log-MAP** Logarithmic Maximum A Posteriori

**LUT** Look-Up Table

Reference DR.4.2

**Max-Log-MAP** Maximum Logarithm Maximum A Posteriori

**MAP** Maximum A Posteriori

**MIMO** Multiple Input Multiple Output

**MSE** Mean Squared Error

**OFDM** Orthogonal Frequency Division Multiplexing

**OFDMA** Orthogonal Frequency Division Multiple Access

**PSK** Phase Shift Keying

**QPSK** Quadrature Phase Shift Keying

**RM** Reed Muller

**SAGE** Space-Alternating Generalized Expectation-maximization

**SISO** Soft Input Soft Output

**SMC** Sequential Monte Carlo

**SNR** Signal to Noise Ratio

**SOVA** Soft-Output Viterbi Algorithm

**STC** Space Time Codes

**VB** Variational Bayesian

**VBEM** Variational Bayesian Expectation Maximization

## 2 POLAR CODES

The aim of this chapter is to report progress made on polar coding as part of NEWCOM++ WP4 activities in the period from T0+6 to T0+18. An initial research agenda for polar coding was defined in the Deliverable DR4.1, Sections 3.2 and 4.7. This agenda was revised and expanded in the WP4 Istanbul Meeting, 19-21 January 2009. A task group was formed on polar coding in the same meeting. This chapter will report the technical work carried out by this task group.

### 2.1 General Background on Polar Coding

A detailed description of polar coding was given in DR4.1 and will not be repeated here. Instead we will begin with an overview of the most important properties of polar codes and review the existing work on polar coding to provide a perspective on current and proposed work under N++.

#### 2.1.1 Overview of polar coding

Polar codes are a class of codes that achieve the symmetric capacity  $I(W)$  of any given binary-input discrete memoryless channel (B-DMC)  $W$  [Ari07]. The symmetric capacity  $I(W)$  is defined as the mutual information between the input and output terminals of  $W$  subject to using the channel input letters with equal frequency.  $I(W)$  equals the channel capacity for channels with certain symmetry properties, such as the Binary Symmetric Channels (BSC) and Binary Erasure Channel (BEC).

The encoding and decoding complexity for polar codes is given by  $O(N \log N)$  in terms of the code block-length  $N$ . Polar codes are the first class of codes that are *provably* capacity-achieving with such low-complexity encoding and decoding algorithms. The well-established LDPC and turbo codes are also capacity-achieving codes with low-complexity encoding and decoding algorithms, however, there is no rigorous proof of this fact for arbitrary B-DMCs.

It has been shown in [AT08] that the probability of block-error for polar codes is bounded by  $O(2^{-N^\beta})$  for any fixed rate  $R < I(W)$  and  $\beta < \frac{1}{2}$ . (We will refer to this fact by saying that polar codes have error exponent  $\frac{1}{2}$ .) This error bound combined with the low-complexity encoding and decoding algorithms for polar codes make them special in the following sense. For polar coding at rate  $R$  and block-length  $N$ , let  $P_b(N, R)$  and  $\chi_b(N, R)$  denote, respectively, the bit-error rate and the per-bit encoding and decoding complexity of achieving  $P_b(N, R)$ . Then, asymptotically, as  $N$  grows, these two parameters are related as

$$\chi_b(N, R) \approx \log \log \frac{1}{P_b(N, R)} \quad \text{for any fixed } R < I(W).$$

This offers the best known trade-off between complexity and performance among all known classes of capacity-achieving error-correction codes.

Despite these nice asymptotical properties on complexity and performance, it is not yet clear if polar coding will have an impact on the practice of error-correction coding, where the performance at non-asymptotic regime is important. One of the main motivations for studying polar coding as part of WP4 is to answer this question, as well as addressing interesting theoretical problems.

### 2.2 Brief literature survey on polar coding

Polar codes were defined and analyzed in [Ari07], [Ari08a]. The exponential error bound for polar codes was first proved in [AT08], and later a simplified proof was given in [AT09].

The polar codes in [Ari07] were based on a channel combining matrix  $F = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ , and it was conjectured that polar code constructions based on other matrices would be possible. Korada and Şaşıoğlu [KŞ08] gave a general characterization of combining matrices that polarize a binary-input channel. They showed that any matrix that is not upper-triangular is sufficient (and necessary) for polarization. In subsequent work, Korada, Şaşıoğlu, and Urbanke [KŞU09] constructed polar codes that improved the error exponent of the basic construction from  $\frac{1}{2}$  to any number arbitrarily close to 1. However, they also

showed that in order to improve the error exponent beyond  $\frac{1}{2}$  one needs to use a combining matrix of size 16-by-16. So, it is unclear if the improvement of the exponent will have an impact for practical purposes.

Mori and Tanaka [MT09] showed that if one assumes that one can do infinite precision density evolution (DE) at unit cost, the complexity of constructing polar codes is bounded as  $O(N)$ . The assumption of unit-cost DE is unrealistic, and it is of interest to study the actual complexity under a more realistic computational model.

Polar coding performance was compared with Reed-Muller codes in [Ari08b], where the main conclusion was that polar codes perform better than RM codes under belief propagation (BP) decoding. In [AKM<sup>+</sup>09], polar codes and RM codes were compared under ML decoding as well as BP decoding. The main conclusion was that RM codes perform better than polar codes at high SNR values due to their better minimum distance properties.

Hussami et al. [HKU09] considered variations of BP decoding of polar codes. They also considered ML decoding of polar codes for the case of a BEC. Their main conclusion was that polar codes are not record-breaking; existing LDPC and turbo codes outperform polar codes at practical block-lengths.

By now it is emerging that the idea of polar coding alone will not be sufficient to construct codes that can compete with state-of-the-art FEC schemes. In an effort to improve the performance of polar codes at small block-lengths, a two-dimensional polar coding scheme was considered in [AM09]. However, the results there are not record-breaking either.

On the more theoretical side, Korada and Urbanke [KU09] have shown that polar codes are optimal (achieving the Shannon limit) for a large set of source and channel coding problems, other than the simple error correction scheme for which they were introduced. Furthermore, they showed that the low-complexity nature of polar coding is preserved. In each problem they studied, they obtained the first provably-optimal low-complexity coding solution. Among the problems considered by Korada and Urbanke are lossy source coding, Wyner-Ziv coding, and Gelfand-Pinsker coding.

### 2.3 Polar coding research subjects under WP4

At present the research on polar coding in WP4 can be itemized as follows.

- *Polar coding for general channels.* There are two directions in which work is underway to generalize the basic polar coding scheme.
  - *Polar coding for arbitrary DMCs.* Polar codes were initially introduced for the class of binary-input DMCs. It is of interest to generalize polar coding to DMCs with non-binary input alphabets. This is discussed in Section 2.4.
  - *Polar coding for channels with continuous inputs.* It is of great theoretical and practical interest to generalize the notion of channel polarization and devise polar coding schemes for channels with continuous input alphabets, such as the AWGN channel. This is an area where we have identified convergence of interest under WP4. Initial research results on this subject is reported in Sections 2.4 and 2.5.
- *Polar coding for compressed sensing.* Compressed sensing [Don06] is a new paradigm that asserts that many real-life signals can be recovered from far fewer samples than dictated by the Nyquist rate. Motivated by the fact that polar coding is optimal for source coding [KU09], we initiated a research effort to apply polar coding to the CS problem. The initial ideas along this direction are reported in 2.6. Since CS is usually carried out over continuous vector spaces, this subject has a cross-section with polar coding for continuous channels.
- *Improvement of polar coding performance.* Although polar codes are capacity-achieving codes, initial trials indicate that they cannot compete with existing state-of-the-art FEC schemes. It is of interest to search for variants of polar coding in combination with other coding techniques to improve the basic polar coding performance. The work items under this category are as follows.

- *Variations on BP decoding for polar coding.* The initial presentation of polar codes were based on a successive cancellation (SC) decoding scheme. Although SC decoders are tractable, for practical applications it is desirable to use more powerful decoding algorithms, such as variations of belief propagation (BP) decoding.
- *ML decoding for short polar codes.* ML decoding of polar codes is feasible only at block-lengths less than a few hundred bits. Such short polar codes cannot achieve in standalone fashion the BER (bit error rate) performances required in many applications. However, short polar codes may be combined in various forms (such as product coding) to achieve the desired BER performance. A study on the performance of polar codes under ML decoding is presented in Section 2.8.
- *Two-dimensional polar coding.* One way of improving polar coding performance is to use it in a two-dimensional coding scheme in combination with other codes. One such scheme is described in Section 2.9.
- *Comparison with Reed-Muller codes.* An interesting research problem of a fundamental nature that we propose for study in the next reporting period is the performance of Reed-Muller codes under ML decoding. The challenge here is explained in Section 2.10.

## 2.4 Polar coding for general classes of channels

Consider an arbitrary DMC  $W : \mathcal{X} \rightarrow \mathcal{Y}$  where the input alphabet is of the form  $\mathcal{X} = \{0, 1, \dots, q-1\}$  for some  $q \geq 2$  and the output alphabet  $\mathcal{Y}$  is arbitrary. Let  $(X, Y)$  be an input-output ensemble for this channel with  $P_{X,Y}(x, y) = Q(x)W(y|x)$  where  $Q$  is some PMF on  $\mathcal{X}$ . Let  $I(X; Y)$  denote the mutual information between  $X$  and  $Y$ . The channel capacity  $C(W)$  is defined as

$$C(W) = \max_Q I(X; Y).$$

The symmetric channel capacity  $I(W)$  is defined as

$$I(W) = I(X; Y), \quad X \text{ is uniform on } \mathcal{X}.$$

In [Ari07], polar coding was introduced for binary-input channels ( $q = 2$ ) and it was shown that it achieves  $I(W)$  for such channels. The extension of polar coding to non-binary channels ( $q > 2$ ) and the extension of the method so that it achieves  $C(W)$  rather than  $I(W)$  was left for future study. In this part, we will outline a proposal to extend polar coding to arbitrary DMCs. Although the discussion is limited to DMCs, one of the channels we have in mind is the AWGN channel, which is important in applications. Such channels with continuous input and output alphabets can be studied by properly quantizing them into DMCs.

The main idea of the proposed approach is to reduce the general coding problem to the already-solved binary polar coding problem. In the proposed scheme, this reduction is achieved by attaching a front-end processor  $f : \{0, 1\}^m \rightarrow \mathcal{X}$  to the given channel  $W$ , where  $m$  is a design parameter. The mapping  $f$  may be thought of as a *modulator* that maps bit sequences to channel input symbols. The equivalent channel consisting of  $f$  and  $W$  is a DMC  $\tilde{W} : \{0, 1\}^m \rightarrow \mathcal{Y}$  with transition probabilities

$$\tilde{W}(y|u_1^m) = W(y|f(u_1^m)), \quad \text{for all } y \in \mathcal{Y} \text{ and } u_1^m \in \{0, 1\}^m.$$

Let  $(U_1^m, Y)$  denote an input-output ensemble for the channel  $\tilde{W}$  so that  $U_1^m$  is uniformly distributed on  $\{0, 1\}^m$  and  $P_{Y|U_1^m}(y|u_1^m) = \tilde{W}(y|u_1^m)$ . Clearly, the symmetric capacity for the channel  $\tilde{W}$  is given by  $I(\tilde{W}) = I(Y; U_1^m)$ . By the data processing theorem,  $I(\tilde{W}) \leq C(W)$ . On the other hand, for any given  $\epsilon > 0$ , it is possible to choose  $m$  large enough so that there exists an  $f$  for which  $I(\tilde{W}) > C(W) - \epsilon$ . Thus, any coding method that achieves  $I(\tilde{W})$  can be used to achieve  $C(W)$  with appropriate design. We will now describe a polar coding method for achieving  $I(\tilde{W})$ .

The method we use combines polar coding with multi-level coding, and we refer to the original work on the subject by Imai and Hirakawa [IH77]. The multi-level coding approach splits the channel  $\tilde{W}$  into  $m$  binary-input channels:

$$\tilde{W}^{(i)} : \{0, 1\} \rightarrow \mathcal{Y} \times \{0, 1\}^{i-1}, \quad i = 1, \dots, m,$$

with transition probabilities

$$\tilde{W}^{(i)}(y, u_1^{i-1} | u_i) = \sum_{u_{i+1}^m} \frac{1}{2^{m-i}} W(y | u_1^m).$$

We observe that

$$I(\tilde{W}^{(i)}) = I(U_i; Y, U_1^{i-1}) = I(U_i; Y | U_1^{i-1})$$

where the second equality follows by the independence of  $U_i$  and  $U_1^{i-1}$ . Thus, by the chain rule for the mutual information function, we have

$$I(\tilde{W}) = \sum_{i=1}^m I(\tilde{W}^{(i)}).$$

Now consider  $N = 2^n$  independent copies of the channel  $\tilde{W}$  for some  $n \geq 1$ . This gives  $N$  copies of each channel  $\tilde{W}^{(i)}$ ,  $1 \leq i \leq m$ . For each channel  $\tilde{W}^{(i)}$  we design a polar code of rate  $R_i < I(\tilde{W}^{(i)})$  and block-length  $N$ . In  $N$  uses of the channel  $\tilde{W}$ , we transmit one codeword over each of the channels  $\tilde{W}^{(i)}$ . At the receiver, first the polar code for channel  $\tilde{W}^{(1)}$  is decoded, regarding the codewords by the other channels as random noise. The decoder succeeds in recovering the correct codeword with probability  $O(2^{-N^\beta})$  (where  $\beta$  may be taken as  $\frac{1}{2} - \delta$  for some fixed  $\delta > 0$ ). The result of this first decoding step is fed to the second decoder which decodes the codeword for channel  $\tilde{W}^{(2)}$ . The decoding continues in this manner, by successive cancellation, until all  $m$  codewords are decoded. By a standard argument, it can be shown that the probability of overall error for the entire codeword is bounded by  $O(m2^{-N^\beta})$ . Thus, polar codes achieve  $I(\tilde{W})$ .

The above approach can be used to apply polar coding to the AWGN channel by suitably discretizing its input and output alphabets. Recall that for the AWGN channel, the channel output is given by  $Y = X + W$  where  $X$  is the channel input with a power constraint  $E[X^2] \leq P$  for some  $P$  and  $W \sim N(0, \sigma^2)$ . The channel-splitting approach for achieving the capacity of this channel has been considered by Duan et al. [DRU97] where the mapping  $f$  was chosen as an adder:

$$f(u_1^m) = \sqrt{\frac{P}{m}} \left( \sum_{i=1}^m (1 - 2u_i) \right).$$

Several such mappings were discussed by Ma and Ping [MP04]. Other methods for polar coding over the AWGN channel will be described in Section 2.5.

## 2.5 Polar codes over real and complex numbers

One approach to polar coding over channels with continuous input alphabets has been described at the end of Section 2.4. In this section, we describe alternative approaches. These alternatives address the polarization problem more directly, rather than trying to reduce the problem to one that has already been solved. The direct approaches promise to be more robust since they rely on a fewer number of design parameters.

In describing the methods studied, we will consider only the case of channels for which the input alphabet is the set of real or complex numbers.

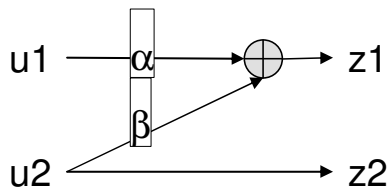


Figure 1: Basic cell of polar codes over reals

### 2.5.1 A first method that fails

The first approach we investigated substitutes in place of the modulo-2 adder in the basic butterfly of polar codes an adder over the real or complex number field.

The encoder procedure for a polar code over the real field is defined starting with a matrix similar to that defined for the binary field, but with the introduction of real positive scalars to keep bounded the variance of the generated random variables.

In general the  $2 \times 2$  matrix is defined as

$$F = \begin{pmatrix} 1 & 0 \\ \beta & \alpha \end{pmatrix}$$

where we fix  $\alpha^2 + \beta^2 = 1$ . A block diagram of the basic cell of the polar code (butterfly) is reported in Figure 1.

Input to the encoder are in general antipodal values  $u_i \in \{\pm 1, \dots, \pm M/2\}$ , properly normalized to have unitary energy, or 0 for frozen symbols. The encoding system is defined by taking the  $\log_2(N)$ th Kronecker power of  $F$  to generate a  $2^N$ -by- $2^N$  transform, and to fix a subset of frozen symbols  $I$  and its complementary set  $\bar{I}$  of transmitted symbols.

The rate of the encoder is then obtained as

$$R = \frac{|\bar{I}|}{2^N} \log_2 M \text{ bits/dim}$$

By construction, the encoding scheme guarantees that all the generated variables have zero mean and unitary variance. Furthermore, by exploiting the central limit theorem, the variables at the output of the encoder will a Gaussian distribution when  $N$  grows.<sup>1</sup>

### Message passing decoder

The main simplification resulting with the transmission of Gaussian variables on a Gaussian channel is that also the messages (likelihoods) at the receiver become Gaussian distributed and message passing decoder simplifies considerably. Gaussian messages for a given variable  $a$  are conveniently represented through their mean  $\mu_a$  and variance  $\sigma_a^2$ .

The derivation of the BP decoder for a graph with real values and Gaussian messages associated with them can be obtained following the general approach described in [LDH<sup>+</sup>07]; we report here the main results relevant to this framework:

- The message (sum and variance) at a sum node  $z = x + y$  is updated according to

$$\begin{aligned} \mu_z &= \mu_x + \mu_y \\ \sigma_z^2 &= \sigma_y^2 + \sigma_x^2 \end{aligned}$$

<sup>1</sup>We neglect here the border effect of those few variables that are only a sum of a small number of terms. The central limit theorem approximation is not needed for these terms since they are the best-protected ones in the SC decoding chain.

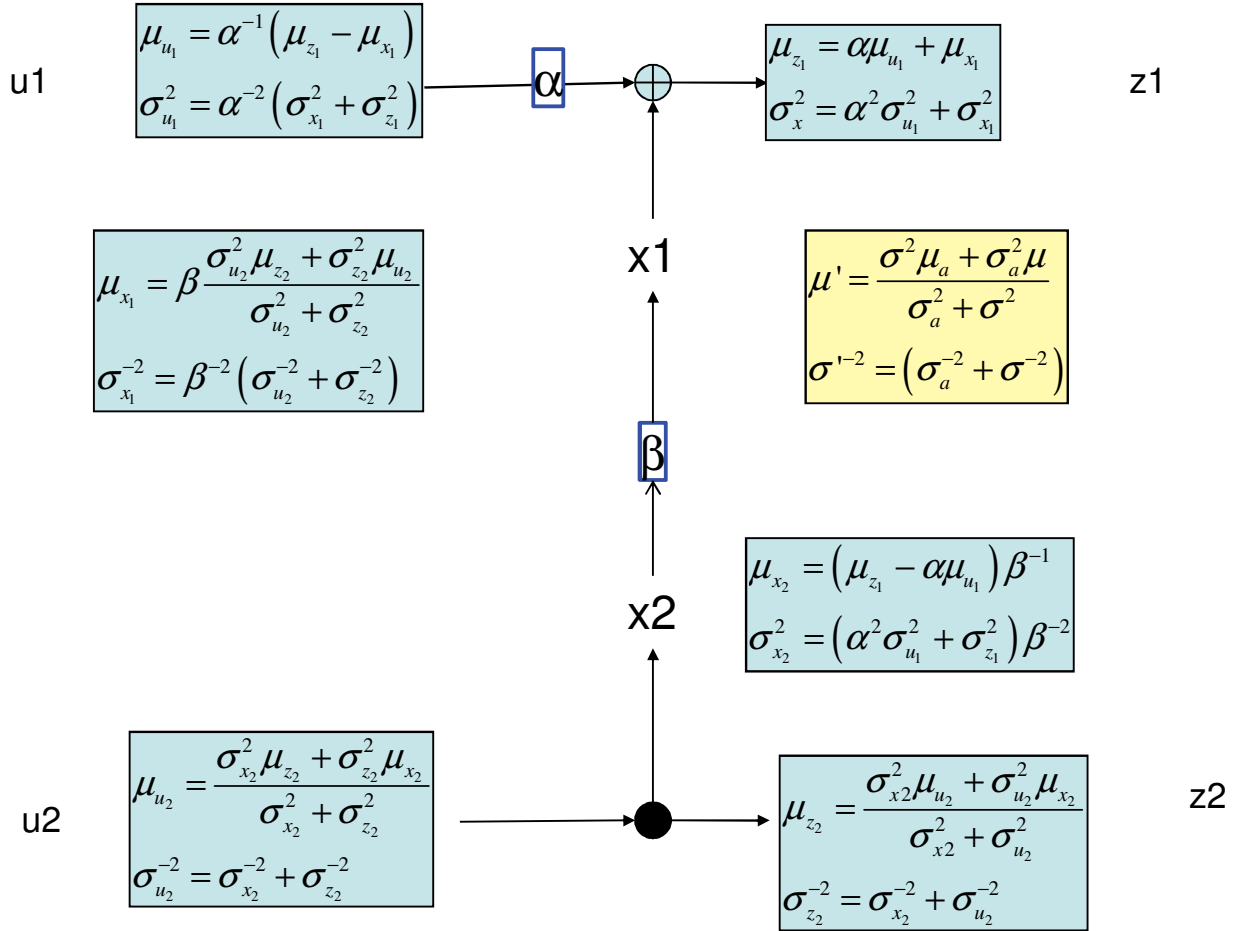


Figure 2: Message passing relationships (SISO decoder) for the basic cell in the polar codes over the reals field.

- The message at a repetition node  $z = x = y$  is updated according to

$$\sigma_z^{-2} \mu_z = \sigma_x^{-2} \mu_x + \sigma_y^{-2} \mu_y$$

$$\sigma_z^{-2} = \sigma_y^{-2} + \sigma_x^{-2}$$

- The message associated to the variable  $z = \alpha x$  is computed as

$$\sigma_z^2 = \alpha^2 \sigma_x^2$$

$$\mu_x = \alpha \mu_z$$

The SISO message passing algorithm, for Gaussian messages in the butterfly of polar code 1) is reported for in Figure 2.

### Failure of the method

The use of operations on the real field leads to an encoding system that is linear, and is then equivalent to a filtering operation. Unfortunately filtering in general modifies the covariance matrix of the transmitted Gaussian variables. This in turn may lead to a loss in capacity if the resulting covariance matrix is not diagonal.

In fact, comparing the information theory inequalities with those obtained for the binary case we can still prove, using the chain rule, that

$$I(Z_1, Z_2; Y_1, Y_2) = I(W') + I(W'') = I(U_1; Y_1 Y_2) + I(U_2; Y_1 Y_2 U_1),$$

where  $W'$  and  $W''$  are the component channels in channel polarization. However, due to the dependency of  $Z_1$  and  $Z_2$  at the output we have

$$I(Z_1, Z_2; Y_1, Y_2) < 2I(Z; Y);$$

so that channel polarization comes at the expense of a capacity loss.

In particular one can directly find closed-form expressions for the symmetric capacities  $I(W')$  and  $I(W'')$  as follows:

$$\begin{aligned} I(W') &= \frac{1}{2} \log \left( 1 + \frac{(1 - \beta^2)(1 + \sigma^2)}{\sigma^2(1 + \beta^2 + \sigma^2)} \right) \\ I(W'') &= \frac{1}{2} \log \left( 1 + \frac{1 + \beta^2}{\sigma^2} \right), \end{aligned}$$

and verify the inequality

$$I(W') + I(W'') = \frac{1}{2} \log \left( \frac{(1 + \sigma^2)^2 - \beta^2}{\sigma^4} \right) < 2I(Z; Y) = \frac{1}{2} \log \left( \frac{(1 + \sigma^2)^2}{\sigma^4} \right).$$

The only transformations that do not imply a loss of capacity are unitary transformations. These transformations however do not show any polarization effect and furthermore do not admit a loop free (tree) graph representation.

### 2.5.2 *Encoding with the mod 1 sum*

Given the above failed method based on encoding with real sum operations, we started investigating other kind of operations.

We observed that the loss of the capacity at the encoder is related to the fact that output variables have a correlation matrix that is not diagonal although inputs are independent.

In order to circumvent this problem, which is connected to the real sum in the upper part of the butterfly we considered alternative operations which preserve the independence of the output given the independence of the inputs. For example, we considered the substitution of the sum with the product of the signs of the incoming variables together with the propagation of the absolute value of the variable  $u_1$ . This operation is isomorphic to the binary sum and preserves independence of the outputs  $z_1$  and  $z_2$ .

Considering complex Gaussian variables, a natural extension of the previous operation would be the mod  $2\pi$  sum of the phases of the two complex Gaussian variables. For both operations however the amplitude of the output variable will be equal to the amplitude of the input variable  $u_1$ . The encoder is then equivalent to an encoder which accepts and provides random variables uniformly distributed in some range (say  $[0:1]$ ), and uses the values of these variables to generate the phase of a complex Gaussian variable, whose amplitude is modulated according to a Rayleigh distribution.

The encoding system which uses the mod 1 sum is actually the natural extension of non-binary encoder over rings, when the size of the ring grows to infinity.

The study of encoders with the use of mod 1 sum and related BP decoders brings up a series of questions, which are currently under investigation and can be listed as follows.

1. Channels suited to these encoders are channels with a bounded and continuous input. The choice of the channel determines the message type distribution at the input of the decoder and consequently the BP decoder equations. We are studying under which channel assumptions the polar code butterfly shows the polarization effect while preserving the channel capacity.
2. The use of a class of messages that can be described with few parameters is the key point for the simplification of the BP decoder. BP can be performed by deriving the updating equations of these parameters instead of the values of its samples. To have this simplification however it is

necessary to use message classes that are closed under the operations performed at the decoding graph nodes. Unfortunately Gaussian messages are the only distributions that are closed under both the sum and repetition operations at the encoder. We are currently considering Tychonov, wrapped normal messages and sampled wrapped normal messages, which are "almost" closed under sum and repetition, to derive BP equations and check the impact of the approximation.

3. As the inputs to the encoder are *discrete* uniform variables, the use of their a-priori probabilities in the BP equation has the effect of sampling the channel messages. The effect of this sampling on the behavior of the BP decoder with continuous messages should be studied.
4. The use of mod 1 operations open the possibility of designing encoding systems more general than Polar Codes. Once their usefulness has been shown, we are planning to consider also different encoding schemes, like for example LDPC.

## 2.6 Polar coding and compressed sensing

The celebrated Nyquist-Shannon sampling theorem, a fundamental result in the field of Information Theory, specifies that the number of samples required to reconstruct a signal without errors is dictated by its bandwidth. In particular, the signal must be sampled at a sampling rate at least twice the signal bandwidth. Conventional signal acquisition methods (for example, to reconstruct a digital image) follow this basic principle. Unfortunately, this approach presents two main drawbacks. First, the complexity might be too high: for several applications, such as medical imaging or high-speed analog-to-digital converters, increasing the sampling rate is very expensive. Second, the Nyquist rate is too high: it generates a huge amount of data, but only relatively little information about the interesting portions of the signal is provided. In practice, most of the sampled data can be removed with almost no perceptual loss (this is the principle that underlies most modern lossy source encoders, such as JPEG-2000). A fundamental question arises in this context: Is it possible to reconstruct signals accurately from a number of samples significantly smaller than the limit given by the Nyquist-Shannon theorem? In the past few years Compressed Sensing (CS), motivated by Donoho's seminal work [Don06], has emerged as a new paradigm that gives an affirmative answer to this question: signals can be recovered from far fewer measurements than are usually considered as necessary.

The main idea in CS is to exploit the fact that most interesting signals have an inherent structure or contain redundancy. In particular, many real-world signals are sparse or compressible, in the sense that they contain many coefficients close to or equal to zero, when represented in the proper basis. These signals can be recovered from significantly fewer samples compared to conventional sampling: it is not necessary to invest a lot of power into observing the entries of a sparse signal in all coordinates when most of them carry little or no information. Today, CS is a rapidly growing research field with interactions with other several fields, such as statistics, signal processing, computer science, and coding theory.

In [ZP08a, ZP08b] Zhang and Pfister showed that the basic CS problem can be regarded as a syndrome source coding problem, in which an  $n$ -dimensional signal  $\mathbf{x}$  is compressed to an  $m$ -dimensional ( $m < n$ ) observation  $\mathbf{y} = \mathbf{H}\mathbf{x}$  (the syndrome) by multiplying with the parity check matrix  $\mathbf{H}$  of a linear  $(n, k)$  block code with  $m = n - k$ . Thus,  $\mathbf{H}$  can be used as the measurement matrix in CS. Decoding can then be obtained for example via belief propagation on the bipartite graph associated with  $\mathbf{H}$ . However, since the effective code rate  $k/m$  is larger than one, in general additional a priori information about the source must be used to facilitate error-free decoding with high probability. In the case of CS this side information is given by the fact that  $\mathbf{x}$  is a sparse vector with a very small number of non-zero elements.

A main difference between CS and coding is that the latter deals in general with discrete alphabets while CS deals with signals over the real numbers. Therefore, codes must be properly generalized to the real numbers for the CS problem. In [ZP08a, ZP08b] Zhang and Pfister apply LDPC codes to CS by modifying the encoding/decoding algorithm to deal with real numbers.

Recently, in [HKU09] it has been shown that polar codes can be used for source coding. In particular, it is shown in [HKU09] that polar codes achieve the Shannon bound for lossless compression of a binary

memoryless source, and the optimal rate for zero distortion in lossy source coding. Polar codes are therefore appealing for CS. We intend to consider polar codes, suitably extended to the real numbers, for CS.

This research item has obvious close connections with the successful development of an error correction coding scheme based on polar codes for channels with inputs over real numbers. We foresee close collaboration between partners on these joint research issues.

## 2.7 Non-binary Iterative decoding of Polar codes.

It has been shown in [Ari07] that polar coding under successive cancellation (SC) decoding can achieve symmetric channel capacity. However, for practical codeword block lengths of interest, other kinds of decoding algorithms have to be considered to try to achieve better performance. To this end, using the normal graph representation of polar codes, it has been shown in [Ari07], [Ari08b] that Belief Propagation (BP) decoding algorithm can be efficiently applied to this class of codes: the authors have shown in particular that polar codes exhibit better performance than Reed-Muller codes. More recently, in [HKU09], the authors considered in detail the polar coding performance under iterative decoding. In particular, they proposed a particular scheduling scheme for BP decoding that has significantly outperformed the SC decoder. Further, they proposed to further improve performance under iterative decoding using an overcomplete representation of the codes based on different equivalent graph representations of these codes.

An on-going collaboration within NEWCOM++ WP4 aims to identify new trends in iterative decoding of polar codes to improve their performance. In particular, it has been previously identified that well-suited non-binary iterative decoding of binary codes can improve their performance, at the expense of increasing complexity. Therefore, it has been proposed to investigate the possible application of non binary iterative decoding in the context of polar codes. Towards this goal, some interesting issues arise. For example:

- What is the most suitable non-binary representation for efficient non binary BP decoding (i.e. order of the binary group extension and associated graph representation)?

Indeed, binary BP decoding is efficient since there exists a suitable representation of polar codes. However, good performance is achieved only if proper scheduling is used. When considering non-binary iterative decoding, one has to identify first what the most suitable representation is.

- Is the scheduling a critical issue for this representation?

This work in an on-going research task within WP4.

## 2.8 ML decoding performance of polar codes

In this part, we report research results on the performance of polar codes under ML decoding. Specifically, we consider trellis-based representations of polar codes and consider their BER and FER (frame error rate) performance under the Viterbi and BCJR algorithms. We use Reed-Muller (RM) codes as a benchmark in this study. As shown in [Ari08b], polar codes and RM codes share much in common in their construction and the well-established RM codes provide an excellent benchmark for both error rate and complexity comparisons. The material in this part is based on the paper [AKM<sup>+</sup>09].

Unlike BP and SC decoders, the complexity of ML decoding increases exponentially as the number of states in the associated trellis is increased for any fixed non-zero coding rate. So, ML decoding of polar codes becomes infeasible after a certain block length. It is of interest to determine the range of block-lengths where ML decoding is feasible for polar codes and also compare ML and BP decoder performance to find out how much loss is incurred by using the sub-optimal BP decoder. These questions are addressed here.

The study of polar coding at small to moderate block-lengths (64 to 512) may be of interest in applications where there is little tolerance for delay but the BER requirements are not very stringent. In such

cases, polar codes may be applied as stand-alone codes. Polar codes under ML decoding may also be employed as component codes in iterative coding schemes.

### 2.8.1 Trellis representation of polar codes

In this section, we study the performance and complexity of polar coding under trellis-based ML decoding. Throughout the study, RM codes are used as a benchmark. For terminology and details of trellis-based representation of block-codes, we refer to [LDJC04] and [HM97].

We list in Table 1, the trellis complexity of the two families of codes for various code dimensions  $(N, K)$ . Listed in the table are the code minimum distance  $d$ , the state complexity  $V$ , which equals the total number of vertices in the trellis representation, and the branch complexity  $E$ , which equals the total number of edges in the trellis representation. There are two types of trellis representations listed in the table: *bit-level* and *parallelized*. In the bit-level representation, the trellis consists of  $N$  sections. The parallelized trellis representation also consists of  $N$  sections; however, the trellis is split into a number of identical and disjoint sub-trellises using the method described in [LDJC04, Sect.9.7]. Although parallelization does not reduce the total state complexity, it simplifies the decoder implementation both in hardware and software. The complexity numbers in the table for the case of parallelized trellises refer to the complexity of one component sub-trellis.

Table 1: Trellis complexity of polar and RM codes.

		Polar					RM				
		bit-level			parallelized		bit-level			parallelized	
$N$	$K$	$d$	$V$	$E$	$V$	$E$	$d$	$V$	$E$	$V$	$E$
32	16	4	1222	1628	232	292	8	4798	6396	434	540
32	26	4	638	1180	390	700	4	638	1180	390	700
64	57	2	818	1556	818	1556	4	2638	5084	1630	3100
128	120	2	3338	6516	3338	6516	4	10734	21084	6670	13020

We note that the bit-level trellis complexity numbers in the table for the RM codes agree with those in Table II of [KDM<sup>+</sup>96]. The table shows that for a given  $(N, K)$ , polar codes tend to have significantly lower bit-level trellis complexity. The table also shows that RM codes benefit from parallelization more than polar codes do; however, polar codes still remain less complex.

To gain further insight into the trellis complexity issue, let us look at the RM(64,57) and P(64,57) codes more closely. The matrices  $G_P(64,57)$  and  $G_{RM}(64,57)$  happen to differ only in one row. While the RM generator matrix uses the row  $10^{15}10^{15}10^{15}10^{15}$  (here  $0^{15}$  denotes 15 consecutive 0s), in its place the polar generator matrix uses the row  $1^20^{62}$ . This choice reduces the minimum Hamming distance of the code from 4 to 2; however, it also reduces the trellis complexity since the second row has a shorter span of 1s. The preference of low-span rows by polar coding is also in evidence in the (8,5) code example whose generator matrix was given above. Note that  $G_P(32,26)$  and  $G_{RM}(32,26)$  are identical, which explains the identical figures in row 2 of the table.

### 2.8.2 Simulations

We now give simulation results for the BER and FER performance of polar codes and RM codes over a BPSK channel with additive Gaussian noise. Due to space limitations, we will present only a limited number of simulation results. The conclusions that will be drawn on the given examples are consistent with experimental results performed for other code lengths and rates.

Figure 3 shows the performance for (64,57) codes under Viterbi decoding. These codes are important as component codes in BTC (block turbo coding) schemes as we will see in the next section. From the figure, we observe that polar codes are more energy efficient at low  $E_b/N_0$  values (with BER values above

roughly  $10^{-4}$ ); at high SNRs, RM codes have a performance advantage which can be explained by their larger code minimum distance. The FER performance of RM codes appears better than that of polar codes at all  $E_bN_0$  values, which can also be attributed to the larger minimum distance of RM codes. Since polar codes have a significantly lower trellis complexity, the P(64,57) code may be seen as a viable alternative to the RM(64,57) code at least for certain applications where a target BER of  $10^{-4}$  is sufficient.

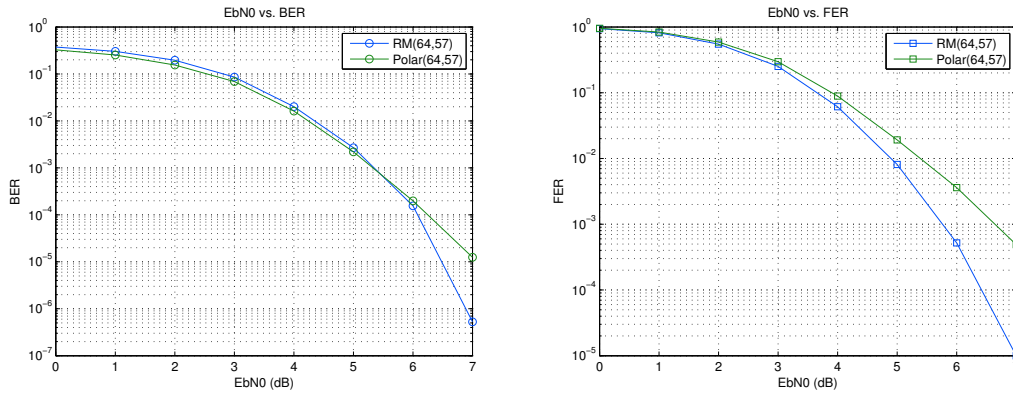


Figure 3: Error rates for (64,57) polar and RM codes on a BSPK channel.

In Fig. 4, we show the performance of P(64,57) code under three decoding algorithms, namely, Viterbi, BCJR, and BP decoders. We observe that the Viterbi and BCJR algorithms give roughly the same performance throughout the  $E_bN_0$  range. The BP algorithm on the other hand has a slightly worse BER performance than the other two, while its FER performance is not markedly different. Since the complexity of the BP algorithm is  $O(N \log N)$  in terms of code block-length, BP algorithm can be used at much higher block-lengths while the ML algorithms (Viterbi and BCJR) are restricted to small block-lengths.

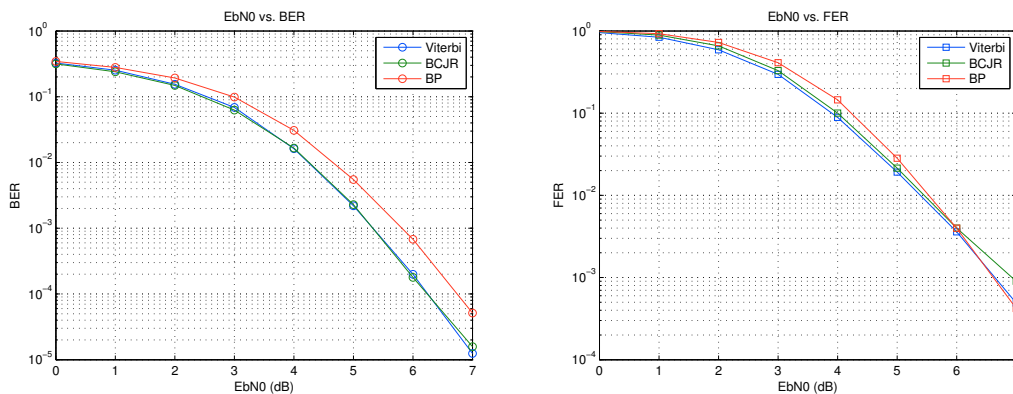


Figure 4: Error rates for the (64,57) polar code under various decoding algorithms.

We now illustrate that under BP decoding, a performance advantage emerges in favor of polar codes as one increases the block-length. In Fig. 5, we show results under BP decoding for the codes RM(64,57), P(64,57), RM(256,228), and P(256,228), all of which have the same rate. It is seen that while the high- $E_bN_0$  performance of the RM code is better than that of the polar code for a code size of (64,57), the situation is reversed when the code size is multiplied by four. This phenomenon had earlier been demonstrated in [Ari08b].

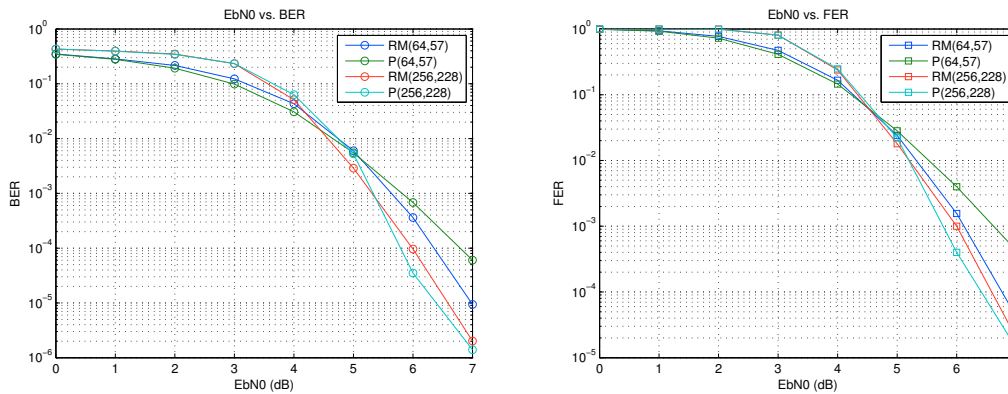


Figure 5: Error rates for polar and RM codes of sizes (64,57) and (256,228) under BP decoding on a BSPK channel.

### 2.8.3 Summary and conclusions

We have compared polar codes and RM codes under trellis-based ML decoding at short block-lengths. One observation has been that, when the two codes differ for a given code size  $(N, K)$ , the polar code tends to have a significantly lower trellis complexity, while the RM code has a larger minimum distance. The minimum-distance advantage of RM codes translates into better performance in their favor at high SNR, although polar codes seem to perform slightly better at low SNR. In conclusion, the results do not suggest that polar codes are a clear winner against RM codes at short block-lengths where ML decoding is feasible.

A second issue addressed in the paper has been the investigation of the performance of polar codes under BP decoding vs. ML decoding. BP decoders are suboptimal; however, they have complexity  $O(N \log N)$ , while ML decoding has exponential complexity in  $N$ . So, one may make up for the deficiency of BP decoding by using a polar code with a larger block-length. We demonstrate that, first of all, the performance disadvantage of BP decoding relative to ML decoding is not too large to begin with (the comparisons for P(64,57) code). Second, we demonstrate that, by increasing the block-length somewhat, BP decoding can outperform the ML decoder (the comparison of codes of sizes (256,228) vs. (64,57)). Finally, we demonstrated that at long block-lengths, where ML decoding is no longer feasible, polar codes appear to be better than RM codes under BP decoding; it appears that the minimum-distance advantage of RM codes over polar codes come into play at higher and higher  $E_bN_0$  values as the block-length is increased, which makes the minimum-distance a moot parameter in the comparison of the two codes, as far as BER values of practical significance are concerned.

This research work has laid the groundwork for developing iterative coding schemes such as BTC codes that use polar codes as component codes. Depending on the block-length of the polar code that is used as a component code, one may use ML decoding (at short block-lengths) or BP decoding (at long block-lengths). It is also possible to use an ML-decoded short RM code as one component and a BP-decoded long polar code as another component. These are subjects for further research.

## 2.9 Two-dimensional polar codes

The material in this section summarizes the paper [AM09], which serves as a basis for further investigations within WPR4.

Recall that polar coding uses a construction that recursively transforms  $N$  independent copies of a B-DMC  $W$  to obtain a second set of  $N$  binary-input channels  $\{W_N^{(i)} : 1 \leq i \leq N\}$  such that the symmetric capacity terms  $I(W_N^{(i)})$  are near 0 or 1 for all but a vanishingly small fraction as  $N$  tends to infinity. The construction is applicable for any block-length  $N = 2^n$  for  $n \geq 0$ . For practical lengths  $N$ , the polarization

effect is not perfect and many of the channels  $W_N^{(i)}$  have symmetric capacities  $I(W_N^{(i)})$  that are far from being polarized as shown in Fig. 6 for the case of a BEC with erasure probability  $\frac{1}{2}$ . The polar coding method as presented in [Ari07] has a step-wise rate allocation algorithm, which allocates a rate of 0 or 1 to each subchannel  $W_N^{(i)}$  according as  $I(W_N^{(i)})$  is above a certain threshold or not. It is conceivable that the code performance can be improved by a graded rate allocation so that each subchannel  $W_N^{(i)}$  is assigned a rate from a larger set of possible rates. One convenient method of graded rate allocation is to consider a two-dimensional (2D) array code, as we discuss in this part. For a self-contained treatment, we give a brief review of polar code construction.

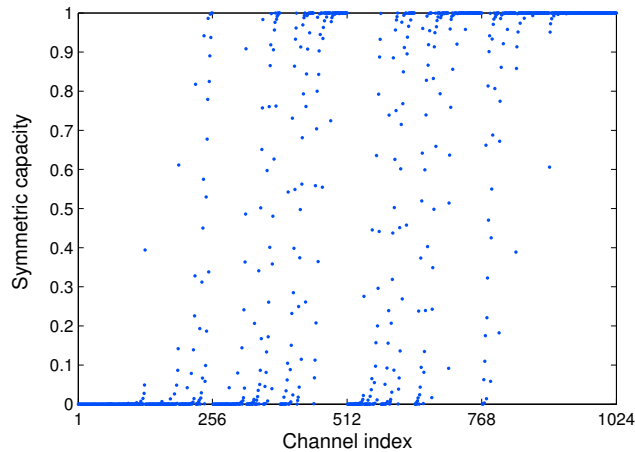


Figure 6: Plot of  $I(W_N^{(i)})$  vs.  $i = 1, \dots, N = 2^{10}$  for a BEC with erasure probability  $\frac{1}{2}$ .

### 2.9.1 Polar code construction

For any  $n \geq 0$ ,  $N = 2^n$ , and  $0 \leq K \leq N$ , there exists a polar code with block-length  $N$  and dimension  $K$ , denoted  $P(N, K)$ . A polar code  $P(N, K)$  is a linear code over  $\text{GF}(2)$  with a generator matrix  $G_P(N, K)$ . The matrix  $G_P(N, K)$  is selected as a submatrix of  $G_N = B_N F^{\otimes n}$ , where  $B_N$  is the *bit-reversal* operator and  $F^{\otimes n}$  denotes the  $n$ th Kronecker power of  $F$ , according to a selection rule defined in [Ari07]. Although the same  $G_N$  is used for all B-DMCs, the selection rule for  $G_P(N, K)$  is channel-specific. The selection rule for the BEC is particularly simple and we will restrict the discussion here to a BEC with erasure probability  $\frac{1}{2}$ .

As an example, let us consider the construction of a  $P(8, 4)$  code for this case. First, we form the generator matrix

$$G_8 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Then the polar code selection rule in [Ari07] selects the generator matrix of  $P(8,4)$  as

$$G_P(8,4) = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The code  $P(8,4)$  maps data words  $u_1^8 = (u_1, \dots, u_8)$  to codewords  $x_1^8 = (x_1, \dots, x_8)$  via  $x_1^8 = u_1^8 G_8$  where the data words  $u_1^8$  are restricted to vectors over  $GF(2)$  such that  $u_1 = u_2 = u_3 = u_5 = 0$ . This is equivalent to the usual viewpoint of encoding as the mapping  $x_1^8 = (u_4, u_6, u_7, u_8) G_P(8,4)$ . When considering the full vector  $u_1^8$ , we refer to the subvector  $(u_1, u_2, u_3, u_5)$  as the *frozen* vector and to  $(u_4, u_6, u_7, u_8)$  as the *information* vector. A vector  $u_1^8$  an *admissible* source vector for the code  $P(8,4)$  if its frozen part equals the zero vector.

### 2.9.2 Product coding

A 2D coding strategy with polar coding is to select an information array  $U = (u_{i,j})$  such that each row of  $U$  is admissible as an information vector for a particular  $P(N,K)$  code and each column is a codeword in a specific set of block codes. For example, consider the case where  $U$  is a 4-by-8 array with each row an admissible data vector for  $P(8,5)$  and the columns of  $U$  selected as codewords in the length-4 codes as indicated in Table 1. An entry  $(N,K,d)$  in the table designates a code with block-length  $N$ , dimension  $K$ , and minimum distance  $d$ . For example,  $(4,0,\infty)$  designates a trivial code consisting of only the all-zero codeword. The code  $(4,1,4)$  consists of two codewords 0000 and 1111. For example, an admissible

Table 2: Column codes for the 4-by-8 code example.

Column	Code type
1	$(4,0,\infty)$
2	$(4,0,\infty)$
3	$(4,0,\infty)$
4	$(4,1,4)$
5	$(4,0,\infty)$
6	$(4,3,2)$
7	$(4,3,2)$
8	$(4,4,1)$

source array for this code is

$$U = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

This array is converted to a codeword array  $X = UG_8$  and transmitted over 32 independent copies of  $\text{BEC}(\frac{1}{2})$ . Let  $Y$  denote the 4-by-8 received array at the channel output. Decoding is performed in interleaved row and column operations. The decoder knows that the first three columns of  $U$  are fixed to zero. So, it sets the estimates of these columns to zero:  $\hat{u}_{i,j} = 0$  for  $1 \leq i, j \leq 3$ . The actual decoding task begins with the fourth column of  $U$ , which is the first column containing information. First, independent row-by-row decoding operations are carried out to generate the estimates  $\tilde{u}_{i,4}$ ,  $1 \leq i \leq 4$ , for the elements of the fourth column; this decoding step ignores the constraints imposed by the  $(4,1,4)$  code checking on the 4th column, using only the constraints imposed by the polar code. A successive cancellation (SC) decoder is suitable for this task. Next,  $\tilde{u}_{i,4}$ ,  $1 \leq i \leq 4$  are used by an ML decoder designed for the  $(4,1,4)$

column code to generate the final decisions  $\hat{u}_{i,4}$ ,  $1 \leq i \leq 4$  for the fourth column. Next, the decoder moves on to the 5th column, which is a column frozen to zero. So, the decoder sets  $\hat{u}_{i,5} = 0$  for  $1 \leq i \leq 4$  and moves over to column 6. Column 6 is not a frozen column; so, a decoding operation similar to the one on column 4 is carried out. Similarly, columns 7 and 8 are decoded, and the decoding task is finished.

Table 3: Erasure probabilities for the 4-by-8 code

Column index $j$	Code type			
	(4,4,1)	(4,3,2)	(4,1,4)	(4,0, $\infty$ )
1	1.0E+000	1.0E+000	9.8E-001	0
2	8.8E-001	8.8E-001	6.0E-001	0
3	8.1E-001	8.0E-001	4.3E-001	0
4	3.2E-001	2.2E-001	1.0E-002	0
5	6.8E-001	6.6E-001	2.2E-001	0
6	1.9E-001	9.0E-002	1.3E-003	0
7	1.2E-001	3.9E-002	2.2E-004	0
8	3.9E-003	4.6E-005	2.3E-010	0

The performance of this code on  $\text{BEC}(\frac{1}{2})$  can be estimated using Table 2.9.2. This table lists the erasure probabilities for various bits of the array code. The entries under the column (4,4,1) relate to the case where effectively no column code is used; the entries in this column equal the channel parameters  $Z_8^{(j)}$  obtained by the recursion

$$Z_{2k}^{(2i-1)} = 2Z_k^{(i)} - (Z_k^{(i)})^2$$

$$Z_{2k}^{(2i)} = (Z_k^{(i)})^2$$

for  $1 \leq i \leq k$ , starting with  $Z_1^{(1)} = 1$ . The parameter  $Z_8^{(j)}$  can be interpreted as the erasure probability as seen by a SC decoder in decoding the  $j$ th bit of a polar code P(8,8) on a  $\text{BEC}(\frac{1}{2})$ , as explained in more detail in [Ari07].

In general we will denote an entry of this table by  $Z(j, \ell)$ , where  $j$  indexes columns of the 2D code-word and  $\ell$  indexes the type of candidate column codes that may be used. E.g.,  $Z(5, 1) = 6.8 \cdot 10^{-1}$  and  $Z(2, 3) = 6 \cdot 10^{-2}$ . The entry  $Z(j, \ell)$  is interpreted as the erasure probability at the decoder output for any designated bit of the  $\ell$ th code. For example, we have

$$Z(j, 2) = Z(j, 1)(1 - (1 - Z(j, 1))^3),$$

which is the probability that a designated bit of the (4,3,2) code is erased (when used on column  $j$ ) and at least one other bit is also erased, so that the decoder cannot recover the erasure in the designated bit.

If we denote the parameter of the actual column code selected for use in the  $j$ th column by  $(N_{\ell_j}, K_{\ell_j}, d_{\ell_j})$ , then the block-error rate (BLER)  $P_e$  for the overall code is bounded as

$$P_e \leq \sum_{j=1}^N K_j Z(j, \ell_j).$$

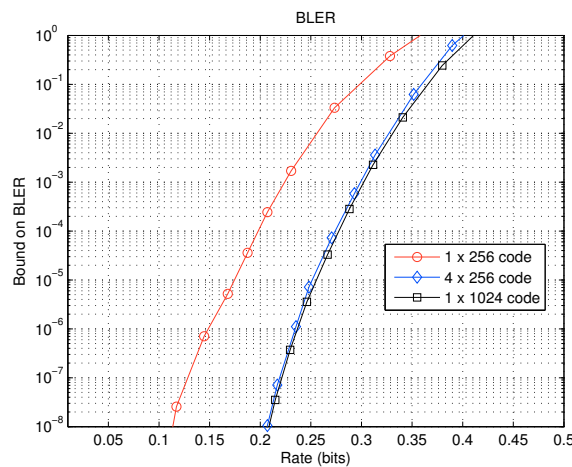
The overall coding rate is given by

$$R = \frac{1}{N} \sum_{j=1}^N \frac{K_{\ell_j}}{N_{\ell_j}}.$$

For example, consider the case where the column codes are assigned as in Table III. The rate of the array

Table 4: An assignment of column codes for the 4-by-8 case

$j$	$(N_{\ell_j}, K_{\ell_j}, d_{\ell_j})$	$K_{\ell_j}/N_{\ell_j}$	$Z(j, \ell_j)$
1	(4, 0, $\infty$ )	0	0
2	(4, 0, $\infty$ )	0	0
3	(4, 0, $\infty$ )	0	0
4	(4, 1, 4)	0.25	1.0E-2
5	(4, 0, $\infty$ )	0	0
6	(4, 3, 2)	0.75	9.0E-2
7	(4, 4, 1)	1	2.2E-4
8	(4, 4, 1)	1	2.3E-10

Figure 7: Block error rates for various array codes on a BEC( $\frac{1}{2}$ ).

code is then  $R = 1$  and the BLER is bounded roughly by  $1.0E-1$ .

Figure 1 shows simulation results for three array codes. We note that the 2D  $4 \times 256$  array code achieves significantly better performance than the 1D  $1 \times 256$  ordinary polar code. This proves that the effectiveness of the graded rate allocation.

It is natural to compare the performance of the  $4 \times 256$  code with that of the  $1 \times 1024$  polar code. This comparison shows that the two codes perform roughly the same and also have roughly the same complexity. Recall from [Ari07] that the complexity of SC decoding of a polar code of length  $N$  is  $O(N \log N)$ . The decoding complexities of the codes in Table 1 are bounded by a constant  $c$  independent of  $N$ . So, a 2D code with size  $4 \times N/4$  has decoding complexity approximately  $O(N \log N)$  as does the  $1 \times N$  polar code.

The performance may be improved by using more powerful column codes. Figure 2 gives simulation results for array codes for which the column codes available are selected from codes with parameters  $(8,8,1)$ ,  $(8,7,2)$ ,  $(8,4,4)$ ,  $(8,1,8)$ , and  $(8,0,\infty)$ .

In conclusion, we have shown that 2D array coding with a graded rate allocation is an effective method for improving the performance of polar coding.

## 2.10 Are Reed-Muller codes capacity-achieving codes under ML decoding?

Although Reed-Muller (RM) codes are one of the oldest codes in coding theory [Ree54], [Mul54], it is not known if these codes are capacity-achieving codes under ML decoding. This issue does not appear to have received much attention in the past. There is strong motivation to study this problem in connection

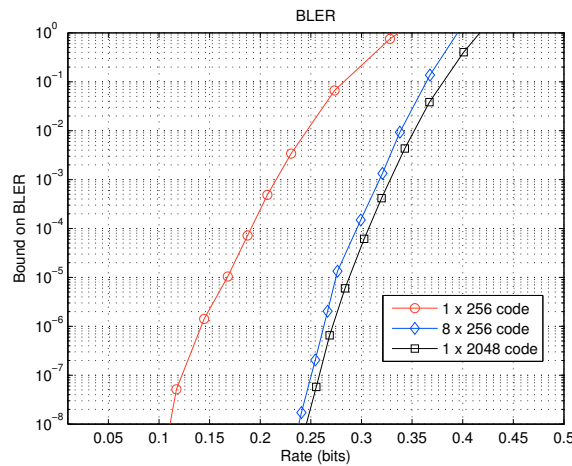


Figure 8: Block error rates for various array codes on a  $\text{BEC}(\frac{1}{2})$ .

with polar codes since RM coding and polar coding are closely related.

At present, we know that the successive-cancellation decoding algorithm that is capacity-achieving for polar codes fails to achieve capacity if applied to RM codes. However, there is not yet any strong evidence to believe that RM codes will not be able to achieve channel capacity under ML decoding. If RM codes were proven to be capacity-achieving, this would be an enormously significant result. So, we include this problem in our work agenda.

## 2.11 Summary and Planned Collaboration

We have identified above a number of highly novel research challenges for partners in the polar coding task. One of the central problems appears to be the generalization of channel polarization notions and methods to channels with real or complex input alphabets. On this challenge, three partners have already contributed material to this document and we foresee collaboration developing around this subject.

A second research area is the development of more effective code construction and decoding methods for polar codes over the binary field. Two partners have already contributed material to this report on this subject. We also see some joint research work emanating from the already proposed topics as we go forward.

Among other modes of collaboration, we may cite that a number of software simulation tools have already been shared among the partners through a web site on polar coding that was set up expressly for this purpose: <http://polar.ee.bilkent.edu.tr/portal/>. Similar tools will be made available through the NEWCOM++ portal to a larger group of users.

Since polar coding is a new research topic, at the moment we do not have any joint papers on the subject by WP4 partners. The main joint work until now has been open sharing of ideas and setting up a joint research agenda. We believe joint papers will be a natural outcome as we move forward.

### 3 JOINT RECEIVERS

#### 3.1 Introduction

In digital communication systems, the physical channel disturbs the transmitted signal in several ways. Channel coding is used to add redundancy to the information sequence in order to protect it against the distortion introduced by the channel. The receiver, however, needs to perform more tasks than channel decoding. Offsets in timing, carrier phase and frequency may occur between transmitter and receiver due to mismatches and instabilities of the respective oscillators. Before decoding, accurate synchronization is of particular importance as even small mismatches can severely degrade the system performance in terms of bit error rate (BER).

When considering wireless communication, both transmitter and receiver are surrounded by reflectors. Their presence creates multiple propagation paths that the transmitted signal is able to traverse. As a result, the receiver sees the superposition of multiple copies from different fading paths. This superposition can be constructive or destructive, since each signal copy may experience different attenuations, delays and phase offsets. Mathematically, the fading effect from the physical channel is usually modeled as a set of complex random variables, which are unknown to the receiver.

All these uncertainties lead to an extremely high complexity of optimal (in terms of e.g. bit or block error rate) receiver structures. Denoting the (coded or uncoded) transmitted data as  $\mathbf{x}$ , the received signal as  $\mathbf{y}$ , and the unknown parameters as  $\boldsymbol{\theta}$ , and assuming statistical independence of  $\mathbf{x}$  and  $\boldsymbol{\theta}$ , the optimal decision (in terms of block error rate) would be

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}) = \arg \max_{\mathbf{x}} p(\mathbf{x}) \int_{\Theta} p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (1)$$

Since the set of possible transmitted vectors is discrete, this is an NP hard problem, and the optimal receiver would have to perform a brute force search.

Due to the constraints on implementation complexity, conventional designs separate the receiver into two parts, namely inner and outer receiver. The task of inner receiver is to synchronize the receiver and to estimate channel parameters. The outer receiver then uses these estimated parameters as if they were the true value to optimally decode the information bits. In terms of (1) this means that the receiver approximates the true parameter distribution by  $p(\boldsymbol{\theta}) \approx \delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$ , which leads to the detection metric

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} p(\mathbf{x}) p(\mathbf{y}|\mathbf{x}, \hat{\boldsymbol{\theta}}). \quad (2)$$

This type of receiver algorithm is called *mismatched*. In most scenarios, it will result in a significant performance loss, because the uncertainty in the parameter estimate is neglected.

In order to perform synchronization and channel estimation in the inner receiver with reasonable complexity, pilot symbols which are known to the receiver must be included at the transmitter side. However, as the transmission of pilot symbols reduces the spectral efficiency of the system, it is desirable to keep their number as small as possible. This becomes more difficult in the context of multiple antenna transmission, since more unknown parameters have to be estimated than in single antenna transmission scenarios. More importantly, the performance of the receiver is compromised by separating the receiver into two parts, namely inner and outer receiver.

With the invention of turbo codes [BGT93], iterative receivers have attracted more and more research interests. As compared to non-iterative receiver structures, such systems are capable of achieving a certain BER at significantly reduced signal-to-noise ratios (SNR). The basic idea is that soft information is exchanged not only between constituent decoder components but also between other modules of the receiver in accordance with the so-called turbo principle. With the aid of these soft symbols, parameter estimates are generated and subsequently used to compensate distortions in the received signal prior to performing further decoding iterations. By doing so, the number of pilot symbols used for initialization can also be reduced.

The focus of this section is the iterative information exchange between channel estimator and data detector, which will be referred to as *iterative channel estimation*. In the following, a brief overview of

the channel estimation algorithms that have been proposed by the participants of this work package will be given. Detailed descriptions follow in the next sections.

The line of work [TB05, TC07, CRMT08b] by Taricco et al. considers a space-time coded MIMO transmission over a quasi-static, i.e., blockwise constant channel. Each packet consists of a preamble  $\mathbf{X}_p$  which is known to the receiver, followed by the payload data  $\mathbf{X}$ . The key point of this work is that instead of calculating explicit estimates  $\hat{\mathbf{H}}$  of the unknown channel matrix  $\mathbf{H}$ , data detection is based on the metric

$$\hat{\mathbf{X}} = \arg \max_{\mathbf{X}} p(\mathbf{X}|\mathbf{X}_p, \mathbf{Y}, \mathbf{Y}_p) = \arg \max_{\mathbf{X}} \mathbb{E}_{\mathbf{H}} [p(\mathbf{Y}|\mathbf{X}, \mathbf{H})p(\mathbf{Y}_p|\mathbf{X}_p, \mathbf{H})], \quad (3)$$

i.e., the channel estimation is implicitly contained in the detection metric. This corresponds to using the optimum receiver (1) instead of the mismatched receiver (2). However, the term ‘‘optimum’’ here only refers to the optimal usage of the pilot symbols—the structure of the outer channel code is not taken into account.

The focus of [MKF<sup>+</sup>09a] is also on pilot-aided channel estimation. It considers a multi-user system, where each transmitter is equipped with a single transmit antenna, while the receiver has several receiver antennas. This setup is known in the literature as *distributed MIMO* because from the receiver’s perspective, it is similar to a MIMO transmission with multiple spatial data streams. The key difference to a system with co-located transmit antennas is that these systems typically employ pilot symbol vectors that are orthogonal across the transmit antennas (for example, when one transmit antenna sends a pilot symbol, all other transmit antennas remain silent). This way, the subchannels from the different transmit antennas to each receive antenna can easily be separated. In the multi-user system considered in [MKF<sup>+</sup>09a], the pilot symbols from the different users are assumed to be non-orthogonal, which leads to co-channel interference. The paper proposes an iterative channel estimation based on the Variational Bayes (VB) framework, which can be interpreted as a generalization of the well-known Expectation-Maximization (EM) algorithm, the key difference being that the VB framework estimates posterior distributions of all unknown variables instead of calculating point estimates. The proposed algorithm iteratively first updates the channel estimate of the desired user, based on the current channel estimate of the interfering user, and then updates the interfering user’s channel estimate.

Another generalization of the EM algorithm, the Space Alternating Generalized EM (SAGE) algorithm, is used in [PKPR09] for joint multiuser detection and channel estimation in an asynchronous DS-CDMA system. In the SAGE framework, both the parameter vector  $\boldsymbol{\theta}$  and the vector  $\mathbf{x}$  of hidden variables are partitioned into smaller groups  $\{\boldsymbol{\theta}_k, \mathbf{x}_k\}$ . In each iteration, the estimates of only one group of parameters are updated, conditioned on the current estimates of the other parameters. In the system considered in [PKPR09], the hidden variables are the transmitted data symbols, while the unknown parameter vector consists of the complex channel gains and the time delay of each user. The authors propose to implement the (analytically intractable) expectation step by means of a Gibbs sampler, which belongs to the family of Markov Chain Monte Carlo (MCMC) techniques.

In [Sch08], Schmitt has developed a unified framework for systematically deriving complete and practically implementable iterative receiver structures. The key idea is based on the equivalence of the ML sequence detection problem to an unconstrained estimation problem for the log-probability ratios. Using this equivalence it can be shown that the complete structure of the receiver, performing the tasks of channel estimation, demapping and decoding, is the solution of a joint ML parameter estimation problem.

All the approaches mentioned above have been studied and debated within the work package. In the following subsections, more background information and results about these approaches will be given. Finally, some of the ongoing and future plans of the work package on this topic will be outlined.

### 3.2 Optimum receiver approach

A typical scenario for multiple-input multiple-output (MIMO) radio communication systems assumes a channel changing so slowly that an entire frame can be transmitted without any appreciable variation in channel characteristics. If this assumption (also known as block or quasistatic fading channel) is valid,

the system performance can be enhanced if the receiver is made aware of the so-called channel state information (CSI), i.e., the realization of the random fading gains affecting the transmission paths from each transmit to each receive antenna. To this purpose, a portion of the transmitted frame may consist of pilot symbols, whose composition is known to the receiver and is used by the latter to estimate the channel parameters. Due to noise and to the finite number of pilot symbols in a frame, the channel estimate is *not perfect*. The optimum receiver for flat fading MIMO channels proposed in [TB05] does not assume perfect CSI at the receiver. It "implicitly" estimates the channel with the primary goal to decode the transmitted code words. The basic idea is to reformulate the MAP decision metric as follows:

$$\begin{aligned}
\hat{\mathbf{X}} &= \arg \max_{\mathbf{X}} \Pr(\mathbf{X}|\mathbf{Y}, \mathbf{Y}_P, \mathbf{X}_P) \\
&= \arg \max_{\mathbf{X}} \frac{p(\mathbf{Y}, \mathbf{Y}_P, \mathbf{X}, \mathbf{X}_P)}{p(\mathbf{Y}, \mathbf{Y}_P, \mathbf{X}_P)} \\
&= \arg \max_{\mathbf{X}} \frac{\Pr(\mathbf{X}) \Pr(\mathbf{X}_P)}{p(\mathbf{Y}, \mathbf{Y}_P, \mathbf{X}_P)} p(\mathbf{Y}, \mathbf{Y}_P|\mathbf{X}_P, \mathbf{X}) \\
&= \arg \max_{\mathbf{X}} \mathbb{E}_{\mathbf{H}}[p(\mathbf{Y}, \mathbf{Y}_P|\mathbf{X}_P, \mathbf{X}, \mathbf{H})] \\
&= \arg \min_{\mathbf{X}} -\ln \mathbb{E}_{\mathbf{H}}[p(\mathbf{Y}|\mathbf{X}, \mathbf{H})p(\mathbf{Y}_P|\mathbf{X}_P, \mathbf{H})], \tag{4}
\end{aligned}$$

where  $\mathbf{X}$ ,  $\mathbf{Y}$ ,  $\mathbf{X}_P$ , and  $\mathbf{Y}_P$  denotes the transmitted (space-time) codeword matrix, the received codeword matrix, the transmitted matrix containing pilots, and the received matrix containing pilots, respectively. Note that the pilot and data symbols do not interfere due to the flat fading assumption. In equation (4), we (i) applied Bayesian rule, (ii) assumed equally likely transmitted codewords, and (iii) used the fact that  $\mathbf{Y}$  and  $\mathbf{Y}_P$  are independent conditioned on  $\mathbf{X}_P$ ,  $\mathbf{X}$ , and  $\mathbf{H}$ . Note that there are no approximations in equation (4) and thus it is equivalent to the MAP criterion.

The expectation value in (4) can be easily evaluated for Gaussian channels by solving a multidimensional Gaussian integral [TB05]. It should be noted that the resulting decision metric depends on the statistics of the propagation channel. Thus, the optimum receiver assumes channel *distribution* information (CDI). However, the statistics of the channel changes much slower than the channel realizations itself, so that CDI is much more realistic than CSI. CDI can be obtained by using sample averages during a preliminary training phase. During the training phase, the receiver operates as mismatched receiver. When a sufficient number of blocks is available to estimate the statistics of the channel, it switches to the optimum receiver.

The approach in [TB05] has been extended recently to more general channel models. In [TC07], the authors assume separately correlated Rician fading. In [CRMT08b], the optimum receiver has been extended to the wideband case assuming separately correlated Rician fading.

An important issue for the optimum receiver is complexity, in particular for more advanced channel models. Complexity reduction can be obtained as follows:

1. Rewriting the resulting expression in equation (4) as a sum of metric branches. This allows for a *trellis based implementation* for the optimum receiver [TB05, TC07, CRMT08b].
2. Projecting onto *dominant eigenspaces* when the eigenvalues of the correlation matrices of the channel decay sufficiently fast. This method turned out to be extremely efficient for the *wideband case* where the matrix which models correlation between the several carriers becomes very large [CRMT08b]. A detailed analysis of this method can be found in [CRMT08a].

Figure 9 and 10 show simulation results for the optimum wideband receiver obtained from [CRMT08b]. The blocks that compose the transmitter are compliant with the 802.11n draft 2.0 standard. The number of subcarriers is  $n_F = 117$ . The random bit stream is fed into a binary convolutional encoder with standard generator polynomials,  $g_0 = 133_8$  and  $g_1 = 171_8$ , of rate  $R = 1/2$ . Coded bits are modulated with QPSK and Gray mapping. Complex QPSK symbols are pairwise (since the presence of STBC) mapped on the data subcarriers. For every subcarrier  $k$  the standard STBC maps each pair of symbols  $(x_{k,1} \ x_{k,2})$  to

a  $2 \times 2$  matrix ( Alamouti code) to be transmitted by the  $n_T = 2$  transmit antennas. We choose to transmit the maximum amount of bits (208 of data plus 6 of tail and 2 of padding) contained in one OFDM symbol. Hence, each subcarrier matrix has length  $N = 2$ . The data symbol matrix is preceded by a unitary pilot symbol matrix of the same length.

Simulations are carried out for the mismatched and the optimum receivers; the genie receiver has been considered as an ideal reference. We consider a zero mean channel, with correlation between subcarriers, in absence/presence of spatial correlation at the transmitter and the receiver. Transmit and receive correlation matrices  $\mathbf{T}$  and  $\mathbf{R}$  are given according to the exponential model:  $\mathbf{T}_{ij} = \rho_T^{j-i}/n_T$  and  $\mathbf{R}_{ij} = \rho_R^{j-i}/n_R$ . The correlation between carriers is modeled as follows:

$$(\mathbf{F})_{ij} = \left[ \sqrt{1 + \left( |i-j| \frac{\Delta f}{\Delta f_c} \right)^2} \right]^{-1} \quad (5)$$

where  $\Delta f = 312.5$  kHz is the separation between carriers set by the standard and  $\Delta f_c = 10$  MHz is the coherence bandwidth of the channel.  $m_F$  denotes the number of eigenvalues taken into account for the matrix  $\mathbf{F}$ .

It can be noticed that taking into account the  $m_F = 4$  largest eigenvalues for the matrix  $\mathbf{F}$  is not enough, we have a penalty and diversity loss. The curves for  $m_F = 8$  and  $m_F = 16$  are overlapping, which implies that  $m_F = 8$  eigenvalues are enough. There is a 0.7 dB loss from the genie receiver, and more than 3 dB gain from the mismatched receiver. Moreover, it can be noticed that spatial correlation introduces 2.5 dB performance degradation.

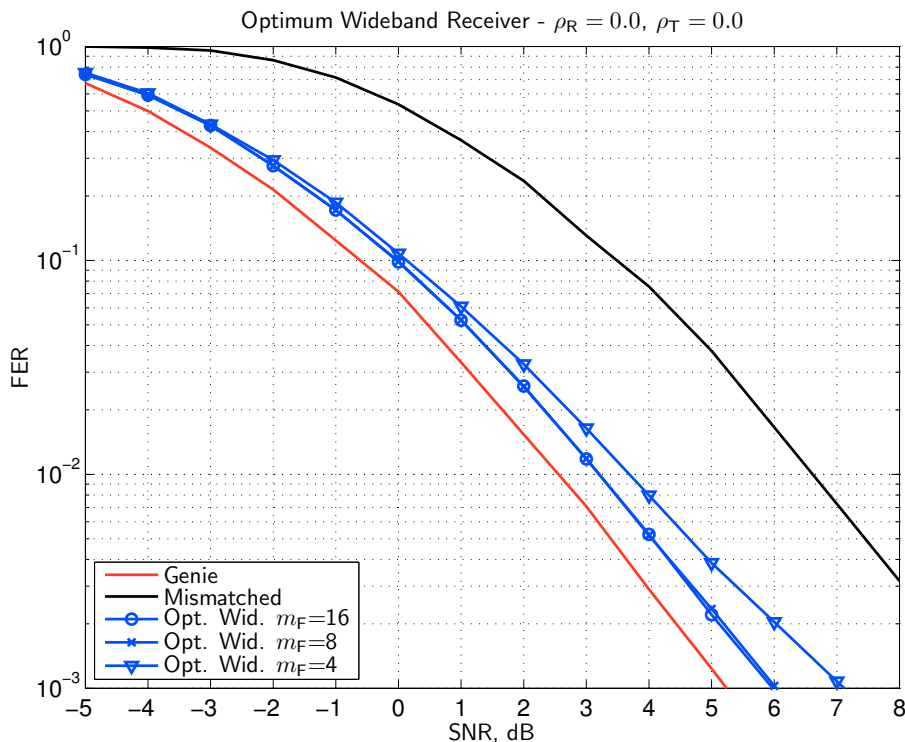


Figure 9: FER vs. SNR of the MIMO-OFDM system with  $\rho_R = \rho_T = 0.0$ . Genie, mismatched, and optimum ( $m_F = 4, 8, 16$ ) receivers.

### 3.3 The Divergence Minimization Principle

The Divergence Minimization (DM) principle was first proposed by Hu *et al.* in [HLR<sup>+</sup>08] as an approach for joint multiuser decoding for coded CDMA. The DM principle, which is closely related to

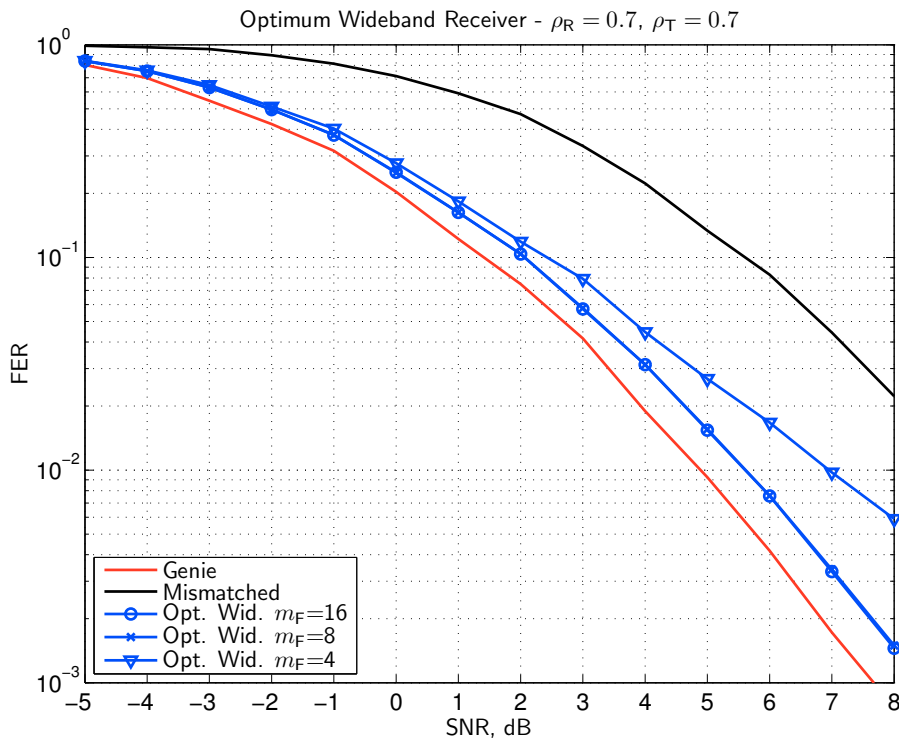


Figure 10: FER vs. SNR of the MIMO-OFDM system with  $\rho_R = \rho_T = 0.7$ . Genie, mismatched, and optimum ( $m_F = 4, 8, 16$ ) receivers.

the variational Bayesian expectation-maximization (VBEM) algorithm [Bea03], is a systematic, holistic framework for designing joint parameter estimation and decoding receivers. It is based on approximating the desired posterior probability distribution by an auxiliary model, which is iteratively updated by minimizing the Kullback-Leibler (KL) divergence (also called information divergence or cross entropy) of the auxiliary function with respect to the true posterior distribution. The auxiliary function is chosen to factorize with respect to the different parameters, so that each of the factors is updated independently. This process yields a sequence of estimates with non-increasing divergence, which typically convergence to a high quality approximation to the true posterior probability. As a practical output of the formal framework, an iterative algorithm taking into account all statistical parameters as well as code constraints is obtained.

In the following, we will first formally describe the general DM framework, and some specific application of it in practical scenarios will be illustrated later.

### 3.3.1 The DM Framework

Let  $\Phi$  be the vector of all the unknown parameters to be estimated at the receiver. At the receiver, an observation vector  $\mathbf{r}$  is available, and we define  $p(\Phi|\mathbf{r})$  as the joint posterior distribution of the parameters given the observation. A direct maximization of  $p(\Phi|\mathbf{r})$  is usually too complex to be feasible, so the goal is to find a feasible suboptimal solution.

To do so, an auxiliary distribution  $q(\Phi)$  of the unknown parameters is defined. We impose the condition that the auxiliary function factorizes as

$$q(\Phi) = \prod_i q_{\Phi_i}(\Phi_i), \quad (6)$$

where  $\Phi_i$  denotes each of the unknown parameters to be estimated. This factorization will lead to an iterative structure of the algorithm. In order to approximate the true posterior distribution,  $q(\Phi)$  is optimized

so the KL divergence [CT91]

$$D\left(q(\Phi) \parallel p(\Phi|\mathbf{r})\right) \triangleq \int d\Phi q(\Phi) \log \frac{q(\Phi)}{p(\Phi|\mathbf{r})} \quad (7)$$

is minimized. Specifically, the divergence is optimized alternatively with respect to each of the factors while the rest are kept fixed, which results in the iterative optimization algorithm.

As a consequence of the framework's formulation, some important properties can be highlighted. First, since the divergence is minimized in each step of the algorithm, it is guaranteed that the divergence will be non-increasing over the different iterations and, thus, the algorithm is guaranteed to converge in the divergence. Second, the factors  $q_{\Phi_i}(\Phi_i)$  of the auxiliary function  $q(\Phi)$  can be seen as approximations of the posterior distribution of each of the parameters, i.e.,  $q_{\Phi_i}(\Phi_i) \approx p(\Phi_i|\mathbf{r})$ . Finally, it is worth remarking that since the framework is based on the inference of probability distributions, rather than point estimates of the parameters. Since the covariance of the estimates are also updated in each updating step, the quality of the estimates is inherently taken into account in the algorithm, and this represents an advantage over most of the state-of-the art receivers.

In the following, and to illustrate the application of the DM framework to different scenarios, two specific applications are shown. These are part of ongoing research on applying the DM framework to MIMO-OFDM systems, have been discussed in the NEWCOM++ WP4 meetings, and have been submitted to international conferences in [MKF<sup>+</sup>09a, MKF<sup>+</sup>09b, KM<sup>+</sup>09].

### 3.3.2 Application of the DM framework to channel estimation with overlapped pilots

In the first application [MKF<sup>+</sup>09a], we apply the DM framework to the problem of MIMO pilot-based channel estimation with overlapped pilot patterns. In this scenario, two or more synchronized transmitters send their pilot signals in the same time-frequency resources, thus making it complicated for the receiver to estimate the different channels.

In order to apply the DM framework to this scenario, we define the set of parameters to be estimated as  $\Phi = \{\mathbf{h}_1, \dots, \mathbf{h}_M, \sigma_w^{-2}\}$ , where  $\mathbf{h}_m$  denotes the channel coefficients from the  $m^{\text{th}}$  transmitter, and  $\sigma_w^{-2}$  is the inverse of the noise-plus-interference variance. In this scenario, the factorization in (6) reads

$$q(\Phi) = q_{\sigma_w^{-2}}(\sigma_w^{-2}) \prod_{m=1}^M q_{h_m}(\mathbf{h}_m). \quad (8)$$

Such a factorization leads to a solution in which each of the channels is alternatively updated, as well as the update performed on the noise-plus-interference variance value. The updates of the channels consists on an interference cancellation step on the pilot signal, in which the contribution of other transmitters in the pilot positions is subtracted, and an MMSE filtering of the interference-free pilot signal. Furthermore, the continuous update of the noise-plus-interference variance allows for a faster convergence of the iterative algorithm.

The performance of the achieved channel estimator is evaluated in Figure 11 in terms of mean-squared error of the estimates. The performance of the pilot-based MMSE filter is also shown for comparison's sake. Two different channels are considered: the Indoor A channel, which has a low delay spread, and the Typical Urban channel, which has much higher frequency selectivity. As it can be observed, the performance of the DM channel estimator approaches that of the MMSE channel estimator with sufficient number of iterations. The more frequency-selective the channel is, the more iterations are required before achieving convergence. Furthermore, splitting the estimation into an interference cancellation step and a filtering step allows a significant complexity reduction of the estimation process as compared to the MMSE estimation, since the computationally complex matrix inversions required in the latter can be avoided.

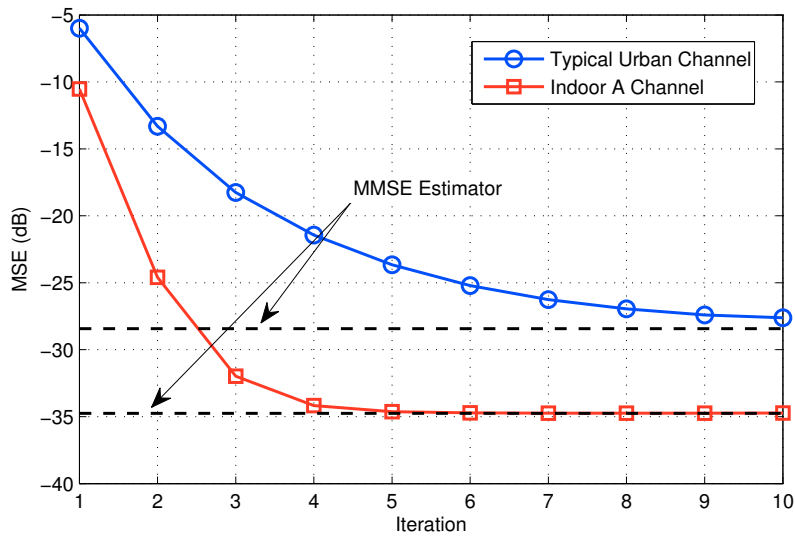


Figure 11: MSE of the estimator

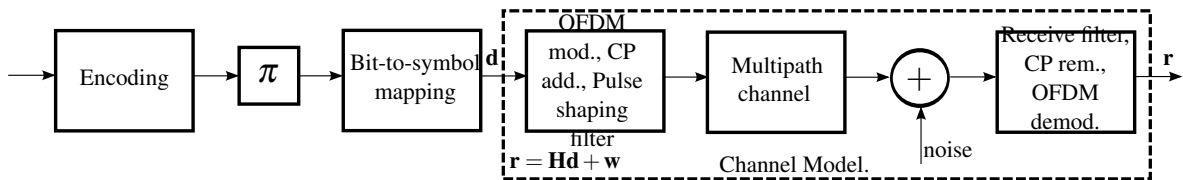


Figure 12: Considered MIMO-OFDM signal model, where  $\mathbf{d}$  consists of contributions from multiple antenna, which we assume independent. The assumed channel consists of OFDM modulation (which is a multiplication with the inverse Fourier matrix), cyclic prefix (CP) extension to avoid inter symbol interference, pulse shaping filter, and the receive filter, CP removal and OFDM demodulation. Also included in the signal model are the interleaver  $\pi$  and the encoder.

### 3.3.3 Application of the DM framework to channel estimation and multi-user detection

The contributions of this section are closely related to ongoing research, see [KM<sup>+</sup>09, MKF<sup>+</sup>09b]. We apply the DM framework to joint multi-user detection/decoding and channel estimation. The system considered is MIMO-OFDM system with a time- and frequency selective channel.

The parameters of interest are  $\Phi = \{\mathbf{H}, \mathbf{d}_1, \dots, \mathbf{d}_K, \Lambda\}$ , where  $\mathbf{H}$  is the complex MIMO channel matrix,  $\mathbf{d}_k$  is the codeword from transmit-antenna  $k$ ,  $K$  is the number of transmit antennas and  $\Lambda$  is the precision matrix for the estimates. The matrix  $\Lambda$  is given by the inverse of the noise and interference covariance matrix.

We consider the following signal model, see Fig. 12: the received vector  $\mathbf{r}$  contains contributions from all receive antennas

$$\mathbf{r} = \mathbf{H}\mathbf{d} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}\sigma^2), \quad (9)$$

where  $\mathbf{d} = [\mathbf{d}_1 \dots \mathbf{d}_K]$ . We apply the DM framework by imposing the following factorization of the auxiliary function, see (6):

$$q(\Phi) = q(\mathbf{H})q(\mathbf{d}_1) \dots q(\mathbf{d}_K)q(\Lambda). \quad (10)$$

By performing the minimization of the Kullback-Leibler divergence  $D(q(\Phi) \| p(\Phi|\mathbf{y}))$ , we obtain an iterative structure with sequential interference cancellation. Dictated by the framework a-posteriori distributions are passed between all blocks. See Fig. 13 for the structure of the obtained receiver.

The expectation of the estimate of the channel distribution  $q(\mathbf{H})$  have the structure of an MMSE channel estimator. Furthermore, in the calculation of the expectation  $q(\mathbf{H})$  a penalty term is included,

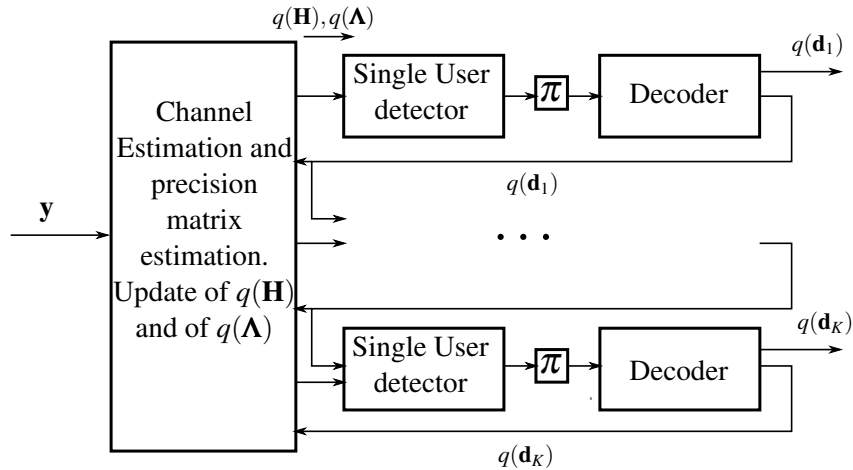


Figure 13: Obtained receiver structure for MIMO-OFDM performing joint channel estimation, precision matrix estimation and sequential multi-user decoding.

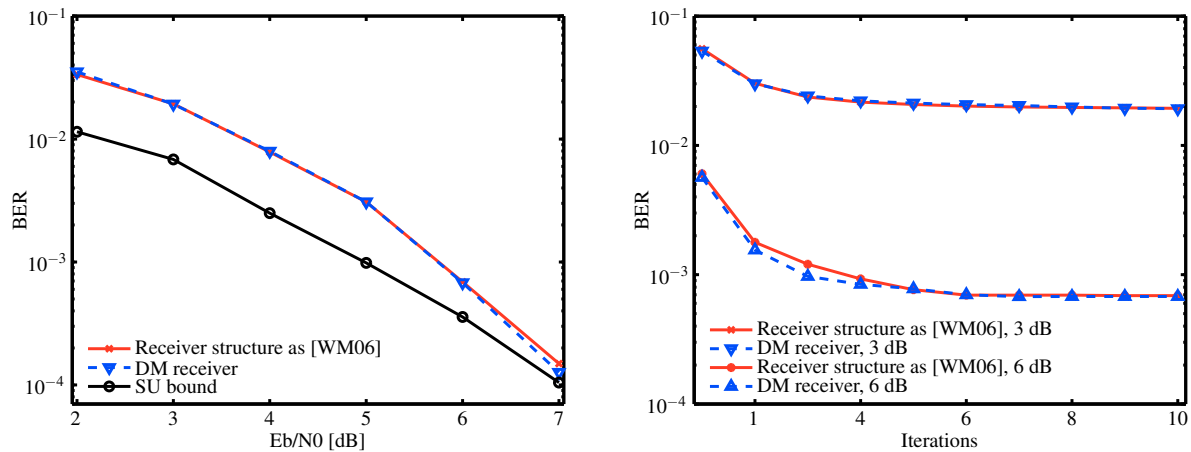


Figure 14: Typical urban channel profile, 10 channel tabs. 2x2 MIMO system. a) Bit error rate performance of the DM receiver and the LMMSE receiver after convergence (10 iterations). b) BER of the two receivers across 10 iterations, see also [KM<sup>+</sup>09]

which can be directly linked to the uncertainty on the data symbols after data detection and decoding.

The update of the expectation of each codewords distribution  $q(\mathbf{d}_k)$  has the structure of a matched filter with an additional penalty term due to uncertainties on the detected data from the other transmit antennas and the uncertainty on the channel estimates. The detector is followed by an interleaver and a single user decoder. The decoding can for certain codes be directly included in the DM framework.

Furthermore, the update of the precision term  $q(\Lambda)$  provides a way of estimating the noise and interference on a current estimate, and thus increase the performance of the next. It was shown in [HLR<sup>+</sup>08], that inclusion of estimation of this term increased the performance of the receiver.

Fig. 14 show Monte-Carlo bit-error-rate simulations of the performance in a typical urban channel profile.

### 3.4 Advanced Signal Processing Algorithms for synchronization of Wireless Communication systems

#### 3.4.1 Introduction

Traditional wireless technologies are confronted with new challenges in meeting the ubiquity and mobility requirements of cellular systems. Hostile channel characteristics and limited bandwidths in wireless applications provide key barriers that future generation systems must cope with. Advanced sig-

nal processing methods, such as the expectation-maximization(EM) algorithm[DLR77, NP96, PC07, DCP07, CPD06], the space alternating generalized expectation maximization(SAGE) algorithm [FH94, PS99, KF03, KLF07], the Kalman filters and their extensions[SCP04], Markov Chain Monte Carlo (MCMC)[DW05, BZS06b] and sequential Monte Carlo (SMC) techniques[PS99, PCMN06, LC98], stochastic approximation algorithms, sphere decoding and convex relaxation techniques (semidefinite relaxations) for detection, in collaboration with inexpensive and rapid computing power provide a promising avenue for overcoming the limitations of current technologies. Applications of advanced signal processing algorithms mentioned above include, but are not limited to, joint/blind/sequence detection, decoding, synchronization, equalization as well as channel estimation techniques employed in advanced wireless communication systems such as OFDM/OFDMA, Space-Time-Frequency Coding, MIMO, CDMA and with Multi User Detection, Time- and Frequency-Selective MIMO Channels [XLG07, X. 98, DP02, P. 04, DCP07, DCP08, I. 06, HW07]. Especially, the development of suitable algorithms for wireless multiple-access systems in non-stationary and interference-rich environments presents major challenges to us. While considerable previous work has addressed many aspects of this problem separately, e.g., single user-channel equalization, interference suppression for multiple access channels and tracking of time varying channels, the problem of jointly combatting these impairments in wireless channels has only recently become significant. On the other hand, the optimal solutions often present a prohibitively high computational complexity, impeding thus their implementation. The statistical tools offered by the advanced signal processing techniques above have provided a promising new route for the design of low complexity signal processing algorithms with performance approaching the theoretical optimum for fast and reliable communication in the highly severe and dynamic wireless environment.

Although over the past decade such methods have been successfully applied in a variety of communication contexts, many technical challenges remain in emerging applications, whose solutions will provide the bridge between the theoretical potential of such techniques and their practical utility.

### 3.4.2 Main knowledge gaps

Key knowledge gaps here concern:

- (i) Theoretical performance and convergence analysis of these algorithms;
- (ii) New efficient algorithms need to be worked out and developed for some of the problems mentioned above;
- (iii) Computational complexity problems of these algorithms when applied to on-line implementations of some algorithms running in the digital receivers must be handled;
- (iv) Implementation of these algorithms based on batch processing and sequential (adaptive) processing depending on how the data are processed and the inference is made has not been completely solved for some of the techniques mentioned above;
- (v) Some class of algorithms requires efficient generation of random samples from an arbitrary target probability distribution, known up to a normalizing constant. So far two basic types of algorithms, the Metropolis algorithm and the Gibbs sampler have been widely used in diverse fields. But it is known that they are substantially complex and difficult to apply for on-line applications. There are gaps for devising new types of more efficient algorithms that can be effectively employed in wireless applications.
- (vi) Although the research on the sequential Monte Carlo signal processing has recently started, many optimal signal processing problems found in wireless communications, such as mitigation of various types of radio-frequency interference, tracking of fading channels, resolving multipath channels dispersion, space-time processing by multiple transmitter and receiver antennas, exploiting coded

signal structures represent few problem waiting for to be solved under the powerful Monte Carlo signal processing framework.

### 3.4.3 An example: A Monte-Carlo Implementation of the SAGE Algorithm for Joint Soft Multiuser and Channel Parameter Estimation

In this section we summarize the work on the receiver design for asynchronous CDMA systems which perform joint channel coefficient and transmission delay estimation within the SAGE framework [FH94, KF03, KLF07]. The implication of the Monte-Carlo method in the SAGE framework [DW05, BZS06a, RC99] makes it possible to compute *soft*-data estimates for all users at polynomial computational complexity, as well. The resulting receiver architecture works in principal fully blind and is guaranteed to converge. A few pilot bits are inserted to resolve the phase ambiguity at the output of the Monte-Carlo SAGE receiver. This is achieved as follows: The Monte-Carlo SAGE provides soft-data symbols which are fed into a MMSE smoother (similar to the [WP99]) and a device that compares the phase of the pilot symbols with the phase of the estimated symbols. The main novelties of the work are that the theoretical framework for the joint transmission delays and channel estimation as well as the data detection algorithms can easily be extended to coded transmission. The resulting synchronizer can also manage to operate within plus minus 10 dB amplitude range.

**System model:** asynchronous CDMA received signal model for  $K$  users

$$r = S(\boldsymbol{\tau})\mathbf{A}\mathbf{d} + w. \quad (11)$$

Here,  $S(\boldsymbol{\tau})$  contains the signature sequences of all the users  $S(\boldsymbol{\tau}) = [S_1(\tau_1), S_2(\tau_2), \dots, S_K(\tau_K)]$  where  $\tau_k$ 's,  $k = 1, 2, \dots, K$  are the unknown transmission delays.  $A$ : The block diagonal channel matrix given by  $A = \text{diag}\{A_1, \dots, A_K\}$ .  $A_k$  is the channel matrix for user  $k$ , given by  $A_k = \text{diag}(a_k, \dots, a_k)$  where,  $a_k$ :  $k$ th user's channel coefficient  $\sim N(0, \sigma_k^2)$ .  $\tau_k$ , the  $k$ th user's transmission delay, is assumed to be uniformly distributed. Given the observation vector  $\mathbf{r}$ , the ultimate goal of the receiver is to estimate

- the channel coefficients  $\mathbf{a} = [a_1, a_2, \dots, a_K]^T$  and
- the propagation delays,  $\boldsymbol{\tau} = [\tau_1, \tau_2, \dots, \tau_K]^T$
- jointly, when the transmitted data vector  $\mathbf{d} = [d_1, d_2, \dots, d_K]^T$  is unknown.

**Main motivation:** The Maximum likelihood (ML) estimates of the unknown parameters  $(\mathbf{r}, \boldsymbol{\tau})$  are obtained by maximizing the likelihood function averaged over the data symbols  $\mathbf{d}_k$  for  $k = 1, 2, \dots, K$

$$(\hat{\mathbf{r}}, \hat{\boldsymbol{\tau}}) = \arg \max_{\mathbf{a}, \boldsymbol{\tau}} E_{\mathbf{d}} \{ \log p(\mathbf{r} | \mathbf{a}, \boldsymbol{\tau}) \}$$

The direct maximization of the averaged likelihood function can be quite complex or impossible analytically. However the lower complexity SAGE algorithm may be used to solve the complexity problem. It typically converges to maximum ML solution. Let us choose the parameter vector as  $\boldsymbol{\theta} = \{\mathbf{a}, \boldsymbol{\tau}\}$ . At iteration  $(i)$ , only the parameters of user  $k$ ,  $\boldsymbol{\theta}_k = (a_k, \tau_k)$  are updated, while the parameters of the other users  $\bar{\boldsymbol{\theta}}_k = \boldsymbol{\theta} \setminus \boldsymbol{\theta}_k$  are kept fixed. **Expectation (E)-Step** and the **Maximization (M)-Step** are expressed as follows:

- **E-Step** computes the objective function

$$Q(\boldsymbol{\theta}_k | \boldsymbol{\theta}^{(i)}) = E_{\mathbf{d}} \left\{ \log p(\mathbf{r} | \mathbf{d}, a_k, \tau_k, \bar{\mathbf{a}}_k^{(i)}, \bar{\boldsymbol{\tau}}_k^{(i)}) | \mathbf{r}, \mathbf{a}^{(i)}, \boldsymbol{\tau}^{(i)} \right\}.$$

- **M-Step** maximizes the objective function  $Q$  with respect to  $\boldsymbol{\theta}_k$  to obtain the update

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_k^{(i+1)} &= \max_{\boldsymbol{\theta}_k} Q(\boldsymbol{\theta}_k | \widehat{\boldsymbol{\theta}}^{(i)}) \\ \widehat{\boldsymbol{\theta}}_k^{(i+1)} &= \widehat{\boldsymbol{\theta}}_k^{(i)}\end{aligned}$$

After some algebra, the E-step of the SAGE algorithm is obtained as follows:

$$Q_k(\boldsymbol{\theta}_k | \boldsymbol{\theta}^{(i)}) = \frac{2}{N_0} \sum_{\ell=0}^{L-1} a_k^*(\ell) \Psi(\ell, \boldsymbol{\tau}_k) - \frac{L}{N_0} |a_k|^2 - \frac{1}{\sigma_k^2} |a_k^2|$$

where,

$$\Psi(\ell, \boldsymbol{\tau}_k) \triangleq \Re \left\{ S_k^\dagger(\ell, \boldsymbol{\tau}_k) \left( \widetilde{d}_k^{[i]}(\ell) \mathbf{r} - \mathcal{I}_k \right) \right\}$$

and the interference term

$$\mathcal{I}_k \triangleq \sum_{j \neq k} a_j^{(i)} \left( S_j(\ell+1, \boldsymbol{\tau}_j^{(i)}) \widetilde{d}_{k,j}^{[i]}(\ell, \ell+1) + S_j(\ell, \boldsymbol{\tau}_j^{(i)}) \widetilde{d}_{k,j}^{[i]}(\ell, \ell) + S_j(\ell-1, \boldsymbol{\tau}_j^{(i)}) \widetilde{d}_{k,j}^{[i]}(\ell, \ell-1) \right).$$

And,  $\widetilde{d}_k^{[i]}(\ell)$ ,  $\widetilde{d}_{k,j}^{[i]}(\ell, \ell+u)$  are the expectations taken over data  $\mathbf{d} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_K]^T$ ,  $\mathbf{d}_k = [d_k(0), d_k(1), \dots, d_k(L-1)]^T$

$$\widetilde{d}_k^{[i]}(\ell) \triangleq E \left\{ d_k(\ell) | \mathbf{r}, \boldsymbol{\tau}^{(i)}, \mathbf{a}^{(i)} \right\}$$

$$\widetilde{d}_{k,j}^{[i]}(\ell, \ell+u) \triangleq E \left\{ d_k(\ell) d_j(\ell+u) | \mathbf{r}, \boldsymbol{\tau}^{(i)}, \mathbf{a}^{(i)} \right\}, \text{ for } j \neq k$$

where  $d_k(\ell), d_k(\ell+u) \in \{-1, +1\}$  and  $u \in \{-1, 0, +1\}$ . We take derivative with respect to  $\boldsymbol{\tau}$  and  $\mathbf{a}$ , setting the results equal to zero, and solving: it follows that

$$\boldsymbol{\tau}_k^{(i+1)} = \arg \max_{\boldsymbol{\tau}_k} \left| \sum_{\ell=0}^{L-1} \Psi(\ell, \boldsymbol{\tau}) \right|;$$

$$a_k^{(i+1)} = \frac{1}{L + N_0 / \sigma_k^2} \sum_{\ell=0}^{L-1} S_k^\dagger(\ell, \boldsymbol{\tau}_k^{(i+1)}) \left( d_k^{(i)}[\ell] \mathbf{r} - \mathcal{I}_k \right).$$

**The main computational problem:** Direct computation of the expectations above has an exponential complexity. This very high complexity can be avoided by adopting the **Markov chain Monte Carlo (MCMC)** statistical method which provides computationally efficient way to calculate the expectations. The basic idea is first, generate a number of random samples  $\overline{\mathbf{D}}^{(t)}$ ,  $t = 1, 2, \dots, N_T$  from the joint conditional posterior distribution,  $P(\overline{\mathbf{D}} | \mathbf{r}, \boldsymbol{\tau}^{(i)}, \mathbf{a}^{(i)})$ . Then, approximate the expectation  $E \{ d_k(\ell) | \mathbf{r}, \boldsymbol{\tau}^{(i)}, \mathbf{a}^{(i)} \}$  based on the samples of  $\overline{\mathbf{D}}^{(t)}$ , by

$$E(d_k(\ell) | \mathbf{r}, \boldsymbol{\tau}^{(i)}, \mathbf{a}^{(i)}) \approx (1/N_s) \sum_{t=1}^{N_s} \xi^{(i)}(\overline{\mathbf{D}}^{(t)}).$$

Similarly, we can compute  $E \{ d_k(\ell) d_j(\ell+u) | \mathbf{r}, \boldsymbol{\tau}^{(i)}, \mathbf{a}^{(i)} \}$ , based on the samples of  $\overline{\mathbf{D}}^{(t)}$  by

$$E \left\{ d_k(\ell) d_j(\ell+u) | \mathbf{r}, \boldsymbol{\tau}^{(i)}, \mathbf{a}^{(i)} \right\} \approx \frac{1}{N_s} \sum_{t=1}^{N_s} \xi^{(i)}(\overline{\mathbf{D}}^{(t)}).$$

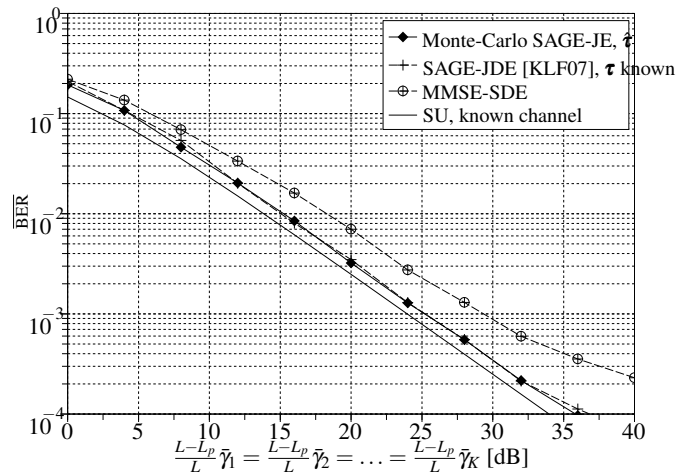


Figure 15: Performance in quasi-static Rayleigh fading:  $K = 8$ ,  $N_c = 8$ ,  $L_p = 4$ ,  $L = 80$ , 5 stages.

**Simulation Results:** the asynchronous DS-CDMA system with the following parameters is considered: Number of Users:  $K = 8$ , the processing gain:  $N_c = 8$  (i.e., system load  $\eta \triangleq K/N_c = 1$ ) is considered, the blocks of size:  $L = 80$  symbols are transmitted,  $L_p = 4$ , pilot symbols are embedded in each block to determine the initial estimates  $\mathbf{a}^{[0]}$  and  $\boldsymbol{\tau}^{[0]}$ . We refer to this method as MMSE-separate detection and estimation (MMSE-SDE). For comparison purpose, the SAGE-JDE scheme in for known transmission delays. We refer to these scheme as "SAGE-JDE,  $\boldsymbol{\tau}$  known". The average bit-error-rate ( $\overline{\text{BER}}$ ) of the proposed receiver is plotted versus the average *effective* SNR  $\frac{L-L_p}{L} \bar{\gamma}_k$ ,  $\bar{\gamma}_k \triangleq \sigma_k^2/N_0$ ,  $k = 1, \dots, K$ . The number of stages in the receiver is limited to 5. Based on the simulation results summarized in Figure 1, we conclude that (i) The initial MMSE-SDE scheme exhibits floor out for average effective SNR values larger than 30 dB, indicating that the MMSE-SDE is not robust to high correlations between the users' signature sequences. (ii) The proposed Monte-Carlo SAGE-JE scheme for unknown delay, "Monte-Carlo SAGE-JE,  $\hat{\boldsymbol{\tau}}$ " and "SAGE-JDE,  $\boldsymbol{\tau}$  known" perform roughly the same over the entire range of SNR. Both schemes have a multiuser efficiency of roughly 1 dB over the entire range of SNR values.

### 3.4.4 General Conclusions

- Applications of advanced signal processing techniques, in collaboration with inexpensive and rapid computing power, seem to very promising way to overcome the limitations of current technologies.
- The statistical tools offered by the advanced signal processing techniques have enabled us to design several low complexity signal processing algorithms with performance approaching the theoretical optimum for fast and reliable communication in the highly severe and dynamic wireless environment.
- However, more research are needed to come up with some novel solutions to narrow the knowledge gaps that I mentioned in the beginning of my presentation.

## 3.5 Phase Noise Estimation

### 3.5.1 Introduction

Oscillator imperfections at the transmitter and receiver in a communication system give rise to random perturbations in the carrier phase. The resulting phase of the received signal contains a time-varying term

which is called phase noise. Phase noise reduces the system noise margin and increases the probability of incorrect detection of data symbols, thereby considerably degrading the performance of digital communication systems. Several carrier synchronization methods have been detailed in the first deliverable DR.4.1. In this chapter we will consider two different approaches to the phase noise estimation problem. In a first (pilot-based) approach the basis expansion model (BEM) is used to approximate the phase noise [BM07a, BM07b, BM08, BM09]. The second (non-data-aided) approach, exploits the statistical properties of the phase noise process to obtain a phase estimate using Monte-Carlo (MC) sampling techniques [PCM05, SDC<sup>+</sup>09]. Section 3.5.2 presents the phase noise estimation algorithm using the basis expansion model. In section 3.5.3 the Monte-Carlo based estimator is discussed. The state of current collaboration and future research scope is detailed in section 3.5.4. In the following, transmission over an additive white Gaussian noise (AWGN) channel affected by Wiener phase noise is assumed.

### 3.5.2 *Phase noise estimation using a basis expansion model*

The phase noise process is represented using the basis expansion model [BM09]. In this model, a time-varying function (Wiener phase noise in our case) is expanded as a linear combination of a set of basis functions. Instead of estimating the phase directly, an estimate of the expansion coefficients is obtained from pilot symbols. The phase estimate is then computed from these coefficient estimates. Because of the lowpass nature of the phase noise, the higher-order basis functions are neglected in the expansion. Hence, only a small number of expansion coefficients need to be estimated. Considering only the lower-order coefficients introduces a phase noise modeling error which gives rise to an error floor in the BER performance. Increasing the number of coefficients in the expansion reduces the error floor, but increases the sensitivity to noise. In [BM09] the discrete cosine transform (DCT) basis functions are used, and an algorithm that estimates the corresponding DCT coefficients without requiring detailed knowledge about the phase noise statistics is devised. Analytical and simulation results have shown that the performance can be optimized by a suitable choice of the pilot symbol positions and the number of DCT coefficients estimated.

### 3.5.3 *Phase noise estimation using Monte-Carlo sampling*

The approach from [PCM05] attempts to obtain the optimal maximum a posteriori (MAP) estimate of the phase noise, without using any pilot symbols. When performing MAP estimation, the computation of the a posteriori distribution typically involves integration/averaging over a complicated distribution, which often defies an analytical expression for the estimate. The use of Monte-Carlo sampling to represent the complicated distribution is a convenient way to avoid this complex operation. In the Monte-Carlo framework a distribution is represented by means of samples (or particles) drawn from it. If a sufficient number of samples are generated from the distribution, further processing can be performed using these particles instead of the distribution and hence integration is no longer required. The use of Monte-Carlo methods is further motivated by the Markov-type behaviour of the considered Wiener phase noise process.

### 3.5.4 *Collaboration and future work*

- In [SDC<sup>+</sup>09], which is a result of collaborative work between UGent and Kadir Has University, several Monte-Carlo techniques were applied to AWGN channels affected by Wiener phase noise. It turned out that there are two feasible Monte-Carlo approaches to tackle the phase noise problem. The first one boils down to drawing samples from the a posteriori distribution of the symbols and updating them in a recursive manner. The carrier phase trajectory is hereby tracked analytically. This approach has previously been examined in [PCM05]. The other approach entails the sequential sampling of the a posteriori carrier phase distribution. Substantial performance improvement was found for both coded and uncoded transmission over an AWGN channel, as compared to phase noise estimators using the extended Kalman filter with symbol decisions or pilot symbols.

- Currently, collaborative research between UGent and Kadir Has University is focussing on the extension of both the BEM and the Monte-Carlo estimation algorithms from single-carrier transmission to multi-carrier transmission, more specifically to the orthogonal frequency division multiplexing (OFDM) scheme. Accurate phase noise estimation in OFDM systems is critical since a time-varying phase destroys the orthogonality of the subcarriers and hence nullifies the advantages of an OFDM system. The fact that OFDM is used in numerous international standards (e.g. standards for DVB, WiMax, WLAN . . .) further justifies research in this field.
- The BEM based estimate from [BM07a, BM07b, BM08, BM09] makes use of pilot symbols only. The estimate can be iteratively improved in a decision directed set-up. By doing so, the observations containing the unknown data symbols are also exploited. The phase estimates resulting from the BEM based algorithm using pilot symbols can then be used to initialize the iterative algorithm. This extension is particularly interesting for turbo coded transmission since the decoding process in this case is already iterative. Collaborative work between UGent and PUT will focus on developing a BEM based algorithm for joint iterative synchronization and decoding using turbo codes.

### 3.6 Information Geometry

#### 3.6.1 Introduction

Iterative algorithms can be used to solve a large variety of signal processing and communication problems. Especially iterative decoding and detection schemes received a lot of attention due to their excellent performance, which is in fact often close to theoretical limits. Although often simple to implement, the reason why iterative receivers feature excellent performance has not been well understood so far. Different approaches, like extrinsic information transfer (EXIT) charts [tB01] or density evolution [DDP01] have been considered but there exist no real breakthrough results.

Information Geometry [Csi75] is a mathematical theory that applies methods of differential geometry to problems of statistics and probability theory. We plan to use information geometrical methods to gain insight into iterative algorithms.

The underlying idea of information geometry is that probability distributions are considered as points in a manifold. This manifold does not have the properties of a Euclidean space, nevertheless it is possible to define distances in terms of Kullback-Leibler (KL) divergence. Another important concept of information geometry are projections, which find the point inside a submanifold closest to a given point in the manifold. To compute these information geometric projections, a simple iterative algorithm called alternating minimization [CT84] can be used.

In the following, we present examples for the application of information geometry to analyze iterative algorithms, discuss ongoing collaborations within the NEWCOM++ framework and future plans for cooperations.

#### 3.6.2 Geometrical interpretation of iterative algorithms

The Blahut-Arimoto algorithm [Bla72, Ari72] for computing the capacity of arbitrary discrete memoryless channels is an example of an iterative algorithm working with probability distributions. In [MD04, NAD09], we used basic tools of information geometry to find an interpretation of this iterative algorithm based on projections onto linear and exponential families of probability distributions. This understanding allows for a reformulation of the Blahut-Arimoto algorithm as a proximal point algorithm. It is shown that this version has an improved convergence rate compared to the initial algorithm.

Information geometry has been successfully applied to the analysis of Turbo and LDPC codes [ITA04] as well as to bit-interleaved coded modulation (BICM) systems [Muq01, MDdC02]. A proposal to accelerate turbo decoders has been made in [ITA04] but appears infeasible in practice. An implementable scheme to acceleration turbo-demodulation in BICM was made in [AND09] but yielded only moderate gains.

CNRS/L2S and VUT aim at reconciling the approach in [ITA04] and in [AND09] to arrive at accelerated versions of turbo decoding and turbo demodulation. Joint work will also be undertaken between CNRS/L2S and VUT aiming at a better understanding of the relation between information geometry and EXIT charts. More specifically, the information geometry interpretation of iterative algorithms concentrates on the convergence of the turbo process for a single channel realization and results in a rather deterministic interpretation with respect to the channel values and noise samples in terms of LLRs. In contrast, EXIT charts are based on an approximate statistical model of the LLRs, i.e., in some sense they build on approximate distributions of the quantities (LLRs) on which information geometry is based. The target of this work is to obtain an information geometry based interpretation of EXIT charts, as well as statistical (with respect to channel coefficients) analysis of the convergence characteristics obtained by information geometry.

### 3.7 Collaboration and Future Plans

The approaches previously mentioned have been presented and discussed inside the work package during the meeting in Istanbul in January, and follow-up discussions have taken place after this. As a consequence, the following topics for collaboration and future work have been identified:

- The optimum receiver structure is very attractive as it gives the best possible performance under the required conditions. However, integrating the channel estimation in the calculation of the decoding metric leads to a rather large computational complexity. In order to solve this problem, the channel estimation and detection could be decoupled from the decoding, and apply iterative processing to try to approach the same performance. The iterative processing could be defined by optimizing the receiver according to a certain metric, as for instance the KL divergence. This structures will be investigated, and the performance/complexity tradeoff of the optimum receiver with respect to the DM approach will be evaluated.
- One of the most novel aspects of the DM approach is that the message passing between the different modules is done in the form of probability distributions, and not point estimates. For instance, when estimating the channel, a measure of the channel estimate covariance is also obtained. Inspired by this method, the development of MIMO detectors which accept soft channel estimates as inputs will also be studied, for instance how to integrate these soft estimates in the sphere detector.
- The relation between the DM approach and other iterative processing frameworks as the EM or the SAGE algorithm are discussed in [HLR<sup>+</sup>08]. Specifically, it is stated that the SAGE algorithm can be seen as a special case of the DM algorithm under certain conditions on the auxiliary function (namely that the distribution of the parameters of interest factors and each factor is modeled as a Dirac delta function). It is interesting to investigate more in depth which are the advantages and drawbacks of both algorithms, and how the differences in performance depend on the scenarios considered, assumptions made, etc.
- The DM approach provides a framework for formal optimization of the receiver processing. It would be interesting to see how it relates to other framework, such as the sum-product algorithm in factor graphs [KFL01].

## 4 REDUCED COMPLEXITY DECODING

In this Section the summary of the ongoing and planned joint research on reduced complexity decoding issues is presented. Section 4.1 gives the overview of the collaborative research between NOA/NKUA, PUT and PoliTo/CNIT on reduced complexity decoding algorithms for turbo codes [BGT93, BG96] and turbo trellis-coded modulation (TTCM) [RW98]. The collaborative work concerns new decoding algorithms based on the simplification of the well-known optimal Log-MAP algorithm [RHV97]. The research is motivated by the constant interest in employing turbo codes and TTCM in various future digital communication systems due to their excellent performance. However, one of the challenges to take up remains reduced-complexity decoding with negligible performance degradation as compared to the optimal decoding. This issue is undertaken in the joint research of NOA/NKUA, PUT and PoliTo/CNIT. In Section 4.1 the brief presentation of the results achieved so far within this collaboration is given and plans for future research are indicated. The outcomes of this joint research in the first half of NEWCOM<sup>++</sup> lifetime were published in [STPM09, PST<sup>+</sup>09, PSTM09].

In Section 4.2 the background, motivation and plans for joint research of a recently established collaboration between Bilkent/Kadir Has University and PUT are described. The joint work focuses on space time turbo coded (STTC) systems [Jaf05, KSP<sup>+</sup>02] with low complexity equalization/decoding algorithms and antenna selection.

Further sections document planned collaborations between the University of Cambridge and CNRS/ENSEA, and between CNRS/L2S, PUT, VUT and Alcatel-Lucent.

### 4.1 Reduced complexity algorithms for decoding turbo codes and turbo trellis-coded modulation

#### 4.1.1 New algorithms investigated

##### A. Boyd approximations

In a recent publication [HKB08], the convex log-sum-exp (*lse*) function has been approximated using: (i) A certain number of piecewise linear (PWL) approximation terms, and (ii) The *max* operation. As shown in [HKB08], the exact number of PWL terms depends on the approximation error resulting from the original bivariate *lse* function. In particular, the approximation error reduces in the order of  $\frac{\sqrt{2}}{r^2}$ , where  $r$  is the number of PWL terms.

We observe that the *lse* function with two arguments is equivalent to the *max\** operator used to decode turbo codes in the so-called Log-MAP decoding [RHV97]. Consequently, the *max\** operator can be approximated using: (i) A small number of PWL terms; and (ii) The very simple *max* operator. Hence, novel reduced complexity decoding algorithms can be obtained with near-optimal Log-MAP performance.

As shown in [HKB08], the *max\** operator can be expressed as

$$\max^*(x_1, x_2) \approx \max(\kappa_1 x_1 + \lambda_1 x_2 + \mu_1, \dots, \kappa_i x_1 + \lambda_i x_2 + \mu_i) \quad (12)$$

where  $\kappa_i$ ,  $\lambda_i$  and  $\mu_i$  are real positive values and  $i \geq 2$ . The best PWL approximations of the *max\** operator with different number of terms are shown in Table 5. It is underlined that in case of turbo decoding, the  $r = 2$  approximation is identical to the Max-Log-MAP decoding.

##### B. The Two Largest (TL) values approach

The biggest drawback of the algorithm presented in Section 4.1.1 (A) (as it is also for Log-MAP) is that the computation of the expression  $\ln(\exp x_1 + \dots + \exp x_n)$ ,  $n > 2$ , which is necessary for both turbo codes and TTCM, still requires successive application *max\** approximation ( $n-1$ ) times. As it can be expected, if we modify this algorithm and apply *max\** approximation less than ( $n-1$ ) times we will decrease overall computational effort of the algorithm. Therefore, to further reduce the complexity of the modified Log-MAP algorithm, we propose to perform the simple approximation considering only the two largest values within the decoding step. In other words, first find the two largest values of  $x_i$  and

Table 5: Best approximations of the  $\max^*$  operation using a certain number of PWL terms and the  $\max$  operation [HKB08]

No. of terms ( $r$ )	Resulting $\max^*$ approximation	No. of max ops.	Approx. error
2	$\max(x_1, x_2)$	1	0.693
3	$\max(x_1, 0.5 \cdot x_1 + 0.5 \cdot x_2 + 0.693, x_2)$	2	0.223
4	$\max(x_1, 0.271 \cdot x_1 + 0.729 \cdot x_2 + 0.584, 0.729 \cdot x_1 + 0.271 \cdot x_2 + 0.584, x_2)$	3	0.109
5	$\max(x_1, 0.167 \cdot x_1 + 0.833 \cdot x_2 + 0.45, 0.5 \cdot x_1 + 0.5 \cdot x_2 + 0.693, 0.833 \cdot x_1 + 0.167 \cdot x_2 + 0.45, x_2)$	4	0.065

then, apply the single  $\max^*$  approximation only to these two values. We denote this approach as TL (*Two Largest*).

### C. $\max^*$ approximation with $n$ arguments (LM- $n$ algorithm)

As it is known, the conventional Log-MAP algorithm makes use of the so-called  $\max^*$  operator [RHV97]. To obtain the  $\max^*$  operator for more than two arguments, denoted with  $n$ , e.g.  $n = 3$ , the following recursive formula is applied:

$$\max^*(x_1, x_2, x_3) = \max^*(\max^*(x_1, x_2), x_3) \quad (13)$$

Since the above formula has to be applied recursively for  $n > 2$  arguments, it is intuitive to expect that a  $\max^*$  approximation with  $n$  arguments may offer complexity reductions as compared with the conventional Log-MAP decoding.

Motivated by the above, the following approximation for the  $\max^*$  operator with  $n$  arguments can be applied:

$$\max^*(x_1, x_2, \dots, x_n) = \max(x_i) + \ln\left(\frac{2}{n+1} \cdot [n + (n-1) \cdot \exp(-z)]\right) \quad (14)$$

where  $z$  is the absolute difference between the maximum value and the second maximum value among  $n$  arguments of  $\max^*$  operator. The first term on the right hand side of (14) is the simple  $\max$  operation. The second term on the right hand side of (14) can be practically implemented with a correction function  $f_c(\cdot)$ . For a given value  $n$ , the curve of the correction function can be analyzed and approximated with several values stored in small look-up table (LUT). For our computer-based simulations we have approximated the correction function with eight values, as it is the case of the conventional Log-MAP decoding for turbo codes. Hereafter, the Log-MAP algorithm using this novel approximation will be abbreviated as LM- $n$ .

## 4.1.2 Simulation results

### A. Turbo codes

A 16-states turbo code is considered with coding rate equal to 1/2 and generator polynomials (1, 33/23) in octal form representing the feedforward and backward polynomials, respectively. An information sequence (frame) of  $N=1000$  bits is assumed, whereas the total number of transmitted frames is  $10^5$ . For the turbo interleaver, a pseudo-random interleaver is considered and at the receiver a maximum of 10 decoding iterations are performed. The modulation type is binary phase-shift keying (BPSK) and the transmission channel is modelled to be the additive white Gaussian noise (AWGN) with the one-sided noise power spectral density  $N_0$ .

Typical turbo code performance evaluation results with the Boyd approximations depicted in Table 5 are illustrated in Figs. 16 and 17. As shown in Fig. 16, the  $r=5$  approximation achieves almost identical Log-MAP performance. The performance degradation of the  $r=4$  approximation with respect to the Log-MAP algorithm is less than 0.1 dB at BER of  $10^{-5}$ . The  $r=3$  approximation has performance degradation of approximately 0.1 dB at the same BER value. When scaling is used as in [VF00], it is expected that all algorithms improve the BER performance. In particular, the scaling factor, denoted with  $s$ , had the following values: (i) 0.65 for Max-Log-MAP; (ii) 0.9 for Log-MAP; and (iii) 0.85 for the rest of the algorithms. Note that for a fair comparison, scaling has been applied in the Log-MAP algorithm. As shown in Fig. 17, both the  $r=5$  and  $r=4$  approximations achieve essentially identical Log-MAP performance. The  $r=3$  approximation degrades less than 0.1 dB at BER of  $10^{-5}$ .

Typical turbo code performance evaluation results with the LM- $n$  approximation as expressed in Eq. (14) are illustrated in Figs. 18 and 19. In particular, two reduced complexity decoding algorithms have been investigated. The first algorithm, denoted with *LM- $n$  max*, assumes the  $\max$  operator for forward and backward, i.e.  $\alpha$  and  $\beta$ , computation. The second algorithm, denoted with *LM- $n$  max\**, assumes the  $\max^*$  operator for forward and backward computations. This is because for these computations there are always two arguments. For LLR, i.e. soft-output, computation both algorithms assume the simplified  $\max^*$  operation with  $n$  arguments from Eq. (14).

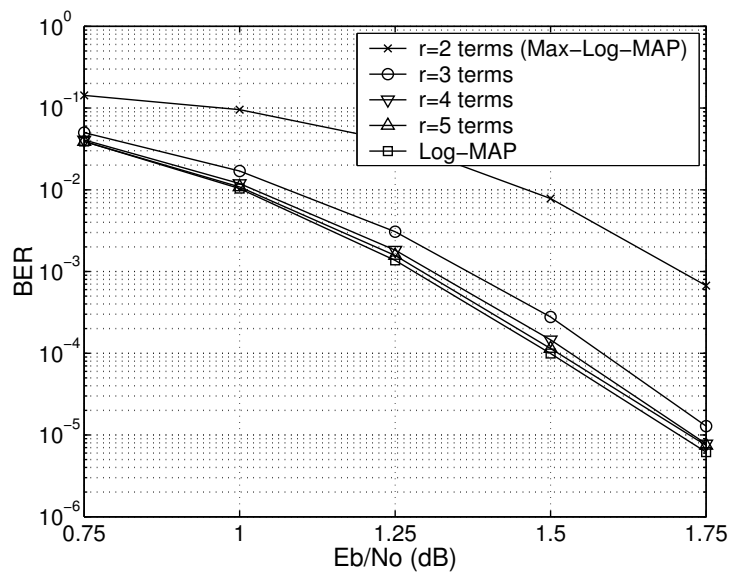


Figure 16: Turbo code BER performance with the best approximations of the max\* operator and different number of PWL terms, denoted with  $r$ .  $N=1000$  bits, BPSK modulation, AWGN channel, and 10 iterations.

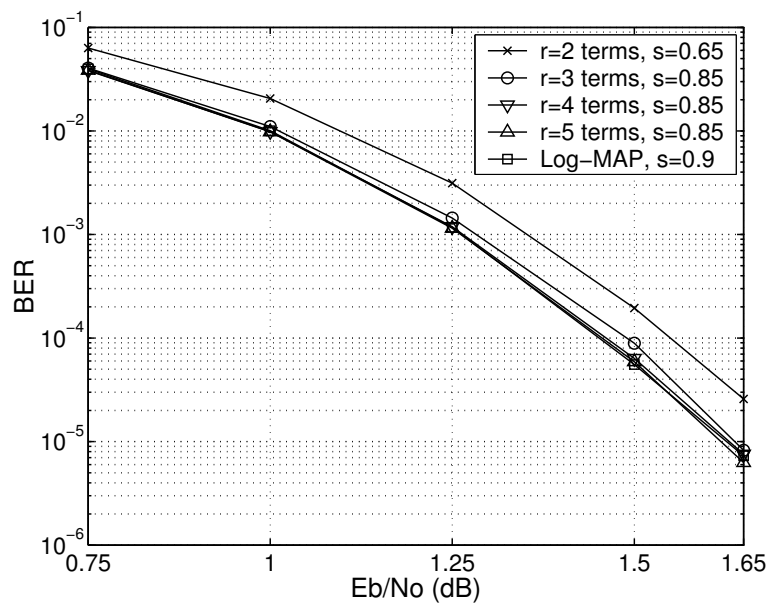


Figure 17: As in Fig. 16 but with the extrinsic information scaled by a factor of  $s$ .

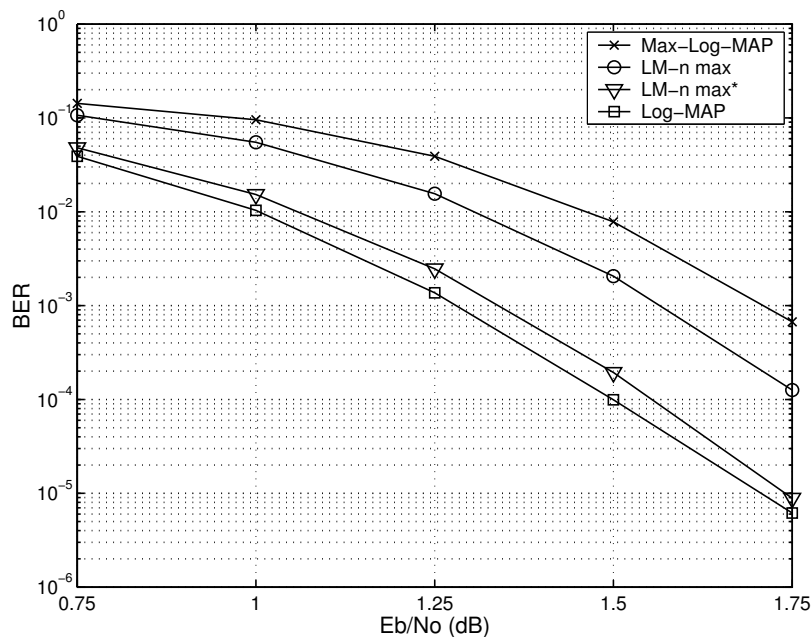


Figure 18: Turbo code BER performance with the LM-n, Log-MAP and Max-Log-MAP algorithms.  $N=1000$  bits, BPSK modulation, AWGN channel, and 10 iterations.

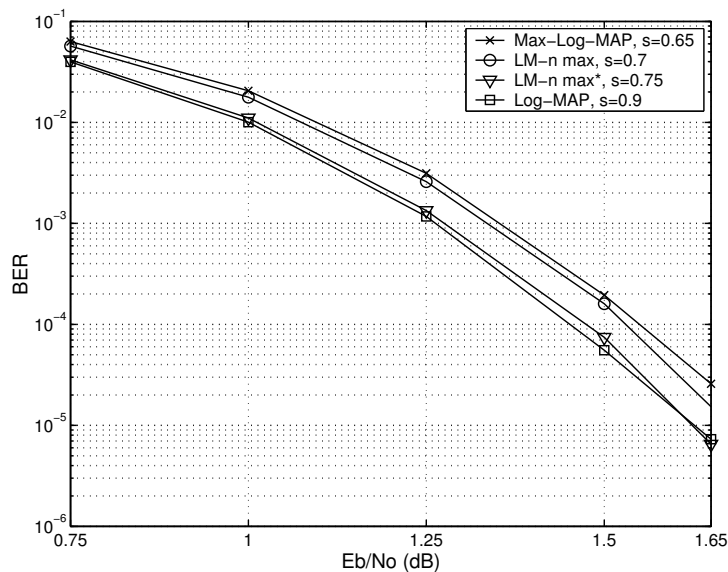


Figure 19: As in Fig. 18 but with the extrinsic information scaled by a factor of  $s$ .

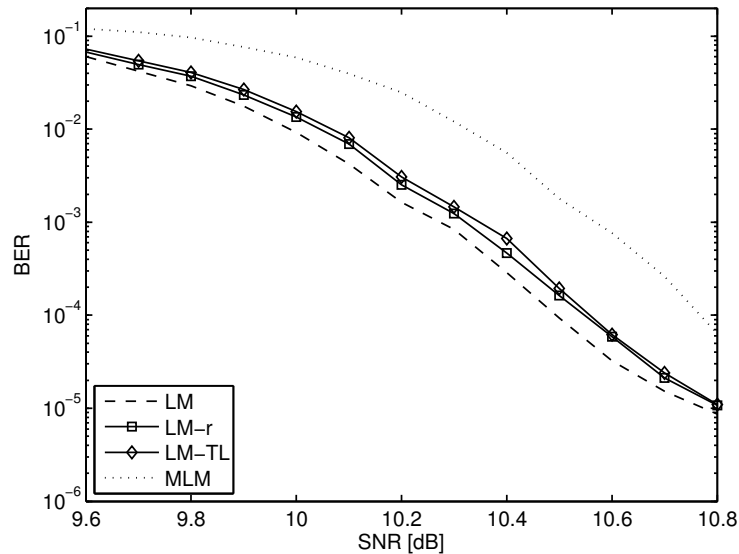


Figure 20: BER performance for the LM-r, LM-TL, Log-MAP and Max Log-MAP with  $N = 684$

As shown in Fig. 18, the  $LM-n \max^*$  algorithm is approximately 0.05 dB inferior to Log-MAP decoding at BER of  $10^{-4}$ . The  $LM-n \max$  algorithm is approximately 0.25 dB inferior to the Log-MAP decoding at the same BER value. The use of scaling [VF00] has also been investigated. The scaling factor had the following values: (i) 0.65 for Max-Log-MAP; (ii) 0.9 for Log-MAP; (iii) 0.75 for  $LM-n \max^*$ , and (iv) 0.7 for  $LM-n \max$ . As shown in Fig. 19, the  $LM-n \max^*$  algorithm achieves essentially identical Log-MAP performance. The  $LM-n \max$  algorithm degrades approximately 0.1 dB at BER of  $10^{-4}$ .

### B. Turbo TCM

We have investigated algorithms presented in Section 4.1.1 in the following scenario: parallel concatenated TCM scheme with two rate-3/4 8-state Ungerboeck's TCM encoders and M-ary quadrature amplitude modulation (i.e. 16 QAM) in the AWGN channel. Simulation results are given for two  $S$ -random interleavers with spreading factor  $S = 13$  and sizes 684 and 5000 symbols, respectively. For each algorithm the TCM decoder performs 8 iterations.

In the Fig. 20 and 21 we have BER performance results for LM-r and LM-TL algorithms for the smaller ( $N = 684$ ) and bigger ( $N = 5000$ ) interleaver sizes. To be precised, all the presented results for the proposed algorithms are for  $r = 3$  terms. For the comparison purposes we have also evaluated BER performance for the Log-MAP and Max Log-MAP algorithms. As shown in the figures the proposed algorithms have almost identical BER performance and are about 0.1 dB inferior to the optimal Log-MAP decoding algorithm at the BER level  $10^{-4}$ . Further improvement of the BER performance is possible after applying scaling of the extrinsic information [VF00]. More results (also with scaling and for  $r$  more than 3 terms) can be found in [STPM09, PST<sup>+</sup>09].

In the Fig. 22 and 23 we have BER performance results for the LM-n for both interleaver sizes  $N = 684$  and  $N = 5000$  respectively. Similarly to previous figures, curves performance for Log-MAP and Max-Log-MAP are also depicted for the comparison purposes. As shown in the figures the proposed algorithm are only about 0.05 dB inferior to the optimal Log-MAP decoding algorithm at the BER level  $10^{-4}$  for both interleaver sizes. This algorithm is also prone to scaling of the extrinsic information [VF00]. More results, also with scaling, can be found in [PSTM09].

### 4.1.3 Complexity estimation

#### A. Complexity estimation for turbo codes

In order to ease hardware implementation, the  $r = 3$  and  $r = 4$  PWL approximations introduced in

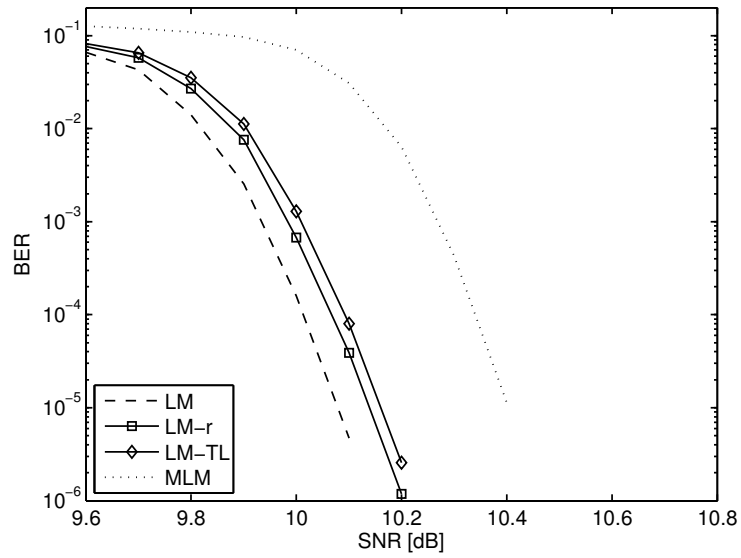


Figure 21: BER performance for the LM-r, LM-TL, Log-MAP and Max Log-MAP with N = 5000

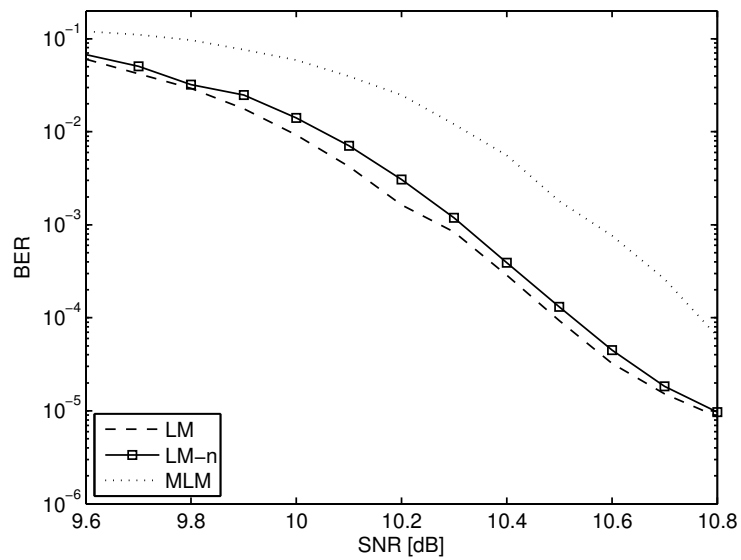


Figure 22: BER performance for the LM-n, Log-MAP and Max Log-MAP with N = 684

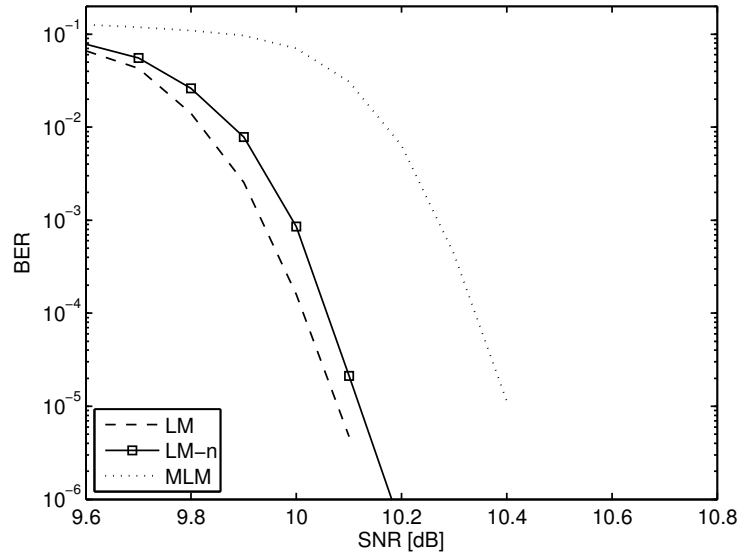


Figure 23: BER performance for the LM-n, Log-MAP and Max Log-MAP with N = 5000

Table 6: Occupied area (square  $\mu\text{m}$  on a 130 nm standard cell technology)

Metric represent.	Max-Log-MAP	Log-MAP	Proposed $r = 3$	Proposed $r = 4$	Constant Log-MAP [GG98]	Linear Log-MAP [CO00]	Average Log-MAP [CBD02]	Ref. [WYY06]	Ref. [TSSA07]
8 bits	250.133	1022.72	653.57	883.53	599.11	978.34	1069.1	891.6	1137.7
10 bits	312.666	1276.888	831.08	1148.8	764.52	1264.8	1363.6	1135.7	1456.4
12 bits	373.182	1613.76	1006.6	1426.2	931.95	1559.3	1555.3	1377.7	1758.9

Section 4.1.1 (A) have been modified, respectively as

$$\max^*(x_1, x_2) \approx \max[x_1, 0.5 \cdot (x_1 + x_2 + 1), x_2] \quad (15)$$

$$\max^*(x_1, x_2) \approx \max[x_1, 0.25 \cdot x_1 + 0.75 \cdot x_2 + 0.5, 0.75 \cdot x_1 + 0.35 \cdot x_2 + 0.5, x_2] \quad (16)$$

Synthesizable VHDL descriptions have been produced for the proposed  $\max^*$  approximations, as well as for Log-MAP (eight entry LUT), Max-Log-MAP algorithm and  $\max^*$  approximations compared in Section 4.1.1. In order to derive fair comparisons, the same area optimisation effort of the synthesis tool must be guaranteed for all cases. To this purpose, although all considered implementations of the  $\max^*$  are pure combinational architectures, registers have been placed at the architecture inputs and output: this allows setting a unique clock frequency constraint for all considered cases,  $f_{CK}=200$  MHz. Synthesis results obtained in terms of area occupied by the combinational part on a 130 nm standard cell CMOS technology are given in Table 6, for metrics represented with 8, 10 and 12 bits. These results show that the proposed approximations have a very low implementation cost and outperform most of the previously published methods, i.e. [CO00, CBD02, WYY06, TSSA07]. Particularly, the occupied area with  $r = 3$  is nearly 35% smaller than that required by the Log-MAP and it is only inferior to the Constant Log-MAP algorithm by 8%. On the other hand, the proposed  $r = 4$  approximation has comparable complexity with the method in [WYY06] and outperforms [CO00, CBD02, TSSA07]. Note that the  $r = 5$  approximation, which is not shown in Table 6, occupies an area greater than that required by the Log-MAP.

The implementation of the  $r = 3$  approximation is obtained in two steps. In the first step the sign of the difference between the two inputs is used to select the maximum between  $x_1$  and  $x_2$ . Concurrently, in the second step the term  $0.5 \cdot (x_1 + x_2 + 1)$  is computed and a further subtractor/multiplexer couple is used to obtain the final result (see Fig. 24). Similarly, the  $r = 4$  approximation is made of two steps, the first step is the same used in the architecture that implements the  $r = 3$  approximation. The second step

Table 7: Complexity comparison of decoding algorithms per decoding step in TTCM scheme

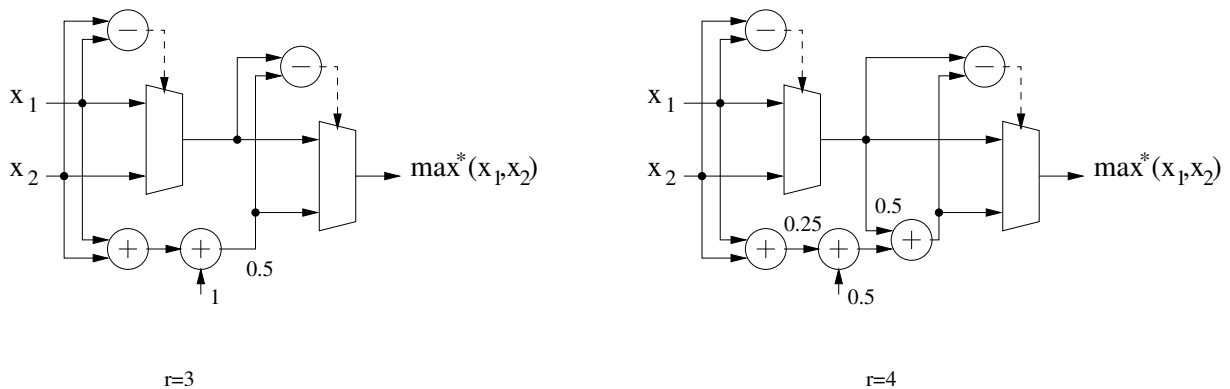
Algorithm	LM	LM-r	LM-TL	LM-n	MLM
Additions	680	680	392	392	344
Multiplication	0	0	0	24	0
Comparisons	357	357	357	357	189
Bit shifts	168	168	0	0	0
Conversion to integer	168	0	0	24	0
Assignment	233	281	269	269	173
Overall	1606	1486	1042	1066	706

is obtained observing that

$$0.25 \cdot x_1 + 0.75 \cdot x_2 + 0.5 = 0.25 \cdot (x_1 + x_2) + 0.5 + 0.5 \cdot x_2 \quad (17)$$

$$0.75 \cdot x_1 + 0.25 \cdot x_2 + 0.5 = 0.25 \cdot (x_1 + x_2) + 0.5 + 0.5 \cdot x_1 \quad (18)$$

As a consequence, the selection between these two expressions depends only on the two inputs. Furthermore, the two expressions share a common term that is computed only once by the means of two adders. The second term is the maximum between the two inputs, that is the result of the first step multiplied by 0.5. Thus it is added to common term by a further adder. Finally the maximum value is obtained with a subtracter/multiplexer couple (see Fig. 24). It is worth pointing out that all the multiplications are implemented as simpler hard-wired right shift operations.

Figure 24: Architecture of the  $r = 3$  and  $r = 4$  approximations

### B. Complexity estimation for turbo TCM

To evaluate the usefulness of all proposed algorithms: LM-r ( $r = 3$ ), LM-TL and LM-n in the context of their practical implementation, Table 7 depicts complexity comparison against the Log-MAP and Max-Log-MAP algorithms. The table contains a summary of the required number of operations (i.e. additions, comparisons, bit shifts, conversion to integer and assignment) per decoding step. In practice, to avoid the operations of multiplication in LM-r, LM-TL and LM-n algorithms, the multiplications by 0.5 in  $max^*$  approximation have been realized as bit shifts operations. It can be noticed that the reduction of the complexity of the LM-r, LM-TL and LM-n algorithms is about 7%, 35% and 34% as compared with the Log-MAP algorithm, respectively. The application of the scaling factor in the considered algorithms requires 8 additional operations (i.e. multiplications) per decoding step, which translates into only about 1% increase of complexity given in Table 7.

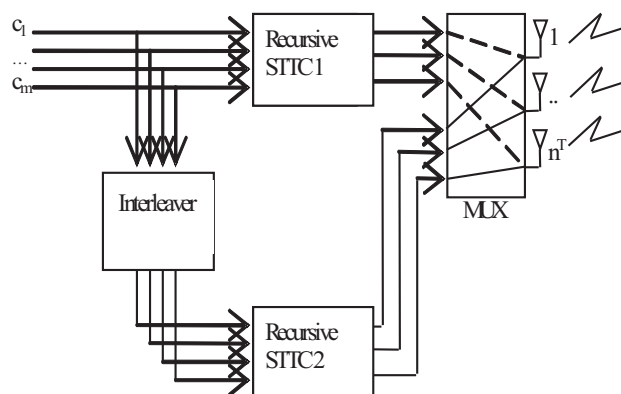


Figure 25: Space time turbo encoder.

#### 4.1.4 Future work

The collaborative research between NOA/NKUA, PUT and PoliTo/CNIT will address in the near future essentially two groups of issues:

- Modifications of the SOVA algorithm aiming at improving its error performance for turbo codes and TTCM without significant complexity increase. It is intended to consider both the bidirectional and classical SOVA algorithms. Complexity estimation of the new algorithms will also be evaluated for hardware and software implementations.
- Application of the algorithms from Section 4.1.1 and new modified SOVA algorithms to other iteratively decoded schemes (e.g. serial concatenation of TCM) and their performance analysis in fading channels.

Apart from the above mentioned joint research topics, it should be pointed out that discussions and interactions between collaborating partners during the NEWCOM++ project lifetime also foster new insights and research directions undertaken individually by the involved researchers.

## 4.2 Space time coded systems with reduced complexity receivers over frequency selective fading channels

### 4.2.1 Space time turbo codes

The insatiable demand for high data rates and high link quality caused a broad research in new transmit technologies. Nowadays the multiple input multiple output (MIMO) systems are commonly used in the wireless transmission [Ton03]. Thanks to the MIMO systems it is possible to increase the rate of transmission or improve the quality in comparison to one receive-transmit antenna systems. Recently the space time turbo codes (STTC) [Jaf05, KSP<sup>+</sup>02] have been proposed as an alternative that integrates space time code and turbo codes. Typically in such systems packet data transmission is used. We research system consisting of  $n_T$  transmit antennas and  $n_R$  receive antennas. Data are transmitted over the Rayleigh fading channel. As an encoder we use the binary space-time turbo trellis encoder or the space-time turbo trellis encoder over ring  $Z_M$ . Information bits input the recursive STTC encoder in the upper branch and the interleaver, which is connected to the STTC encoder in the lower branch. Each the STTC encoder operates on a block consisting of groups of  $m$  information bits. The block diagram of recursive STTC encoder for M-ary modulation is shown in Figure 25 [HYCV04].

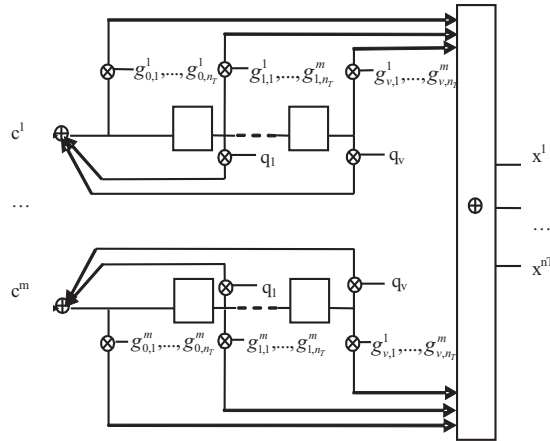


Figure 26: Recursive STTC encoder for M-ary modulation.

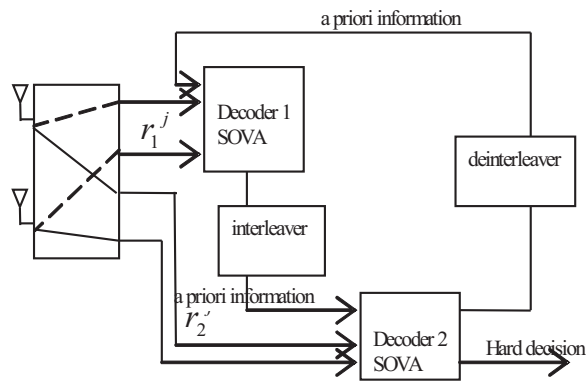


Figure 27: Space-Time Turbo Trellis Decoder.

Input sequence of binary vectors is transformed by the convolutional encoders shown in Figs. 25 and 26 into a coded sequence of the symbols which belong to the ring  $Z_M$ . The coefficients in the encoder structure (Fig. 26) are binary or are from to the same ring  $Z_M$ . The memory cells are capable of storing binary symbols. Multipliers and adders perform multiplication and addition respectively in the ring of integers modulo-M.

We investigate different receiver structures. One of them is the receiver which uses the iterative Soft Output Viterbi Algorithm (SOVA) [CXX99]. The structure of this receiver is shown in Fig. 27 [HYCV04].

In the receiver decoder 1 and decoder 2 (Fig. 27) use the iterative Soft Output Viterbi Algorithm (SOVA) [CXX99]. It selects maximum likelihood path on the trellis diagram with *a priori* received symbols probability taken into consideration. At the receiver side there are  $n_R$  antennas. The signal received by each antenna is demultiplexed into two vectors, denoted by and contributed by the Recursive STTC 1 (upper) and Recursive STTC 2 (lower) encoder, respectively.

#### 4.2.2 Space time turbo coded systems with reduced complexity receivers over frequency selective fading channels

We now present the background, collaboration and future plans about simplified approaches for space-time turbo coded transmissions over frequency-selective fading channels will be presented.

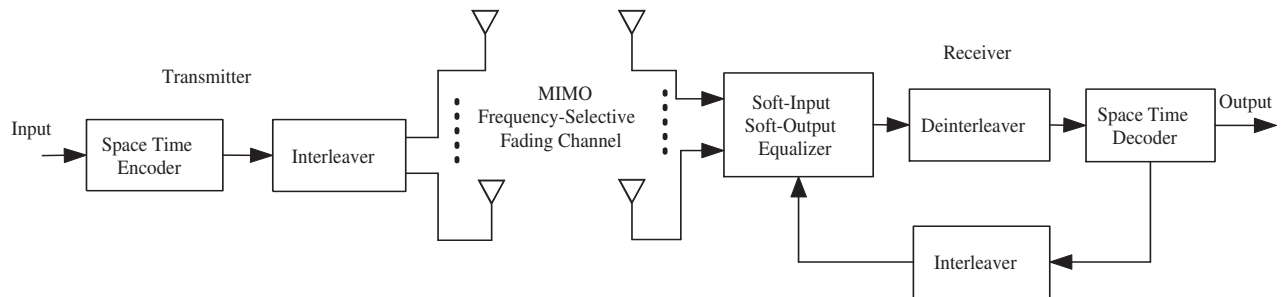


Figure 28: Block diagram of iterative equalization/decoding for space time coded systems.

In high rate wireless transmission systems, frequency selective (FS) channel model is more common than flat fading. When the transmission rates are increased by making the symbol periods smaller, the wireless channels are more likely to become frequency-selective which results in intersymbol interference (ISI) and hence increased error rates. ISI can be cancelled with the use of an equalizer [Pro00] which may provide multipath diversity in wireless channels. Next generation communication systems can have STC transmission over frequency selective fading channels [Ton03, VB03].

The major drawback of coded or uncoded transmissions over FS fading channels with single or multiple antennas is that the optimal receivers for them require prohibitively high computational complexity which depends exponentially on the number of ISI taps and the number of transmit antennas. With the motivation of reducing the hardware and software complexity of coded or uncoded single or multiple antenna systems, several simplified approaches have been studied in the literature.

In the literature, there are many low complexity soft-input soft output MIMO equalizers which can be used at the iterative receivers of space time coded systems (Figure 28).

For example, with motivation of reducing the complexity of equalization over single or multiple antenna frequency selective fading channels, a near-optimum soft-input soft-output stack algorithm is proposed in [GD05]. It is shown that the soft input soft-output stack equalization with modified metrics based on a previously designed soft-input soft-output stack detection algorithm [SM01] can be used at the receiver of the space time coded systems over multiple antenna frequency selective fading channels. The complexity of the receiver remains low while offering a bit error rate performance close to that of the optimal BCJR algorithm. Unlike the MAP equalizer, the complexity of the new equalizers does not depend on the memory of the channel, hence the soft-input soft-output stack equalizer are quite desirable for channels with large ISI spans. Moreover, decision feedback equalization (DFE) is also well known low complexity equalization technique based on minimizing the mean squared error (MSE), and it can be extended to MIMO systems [ADS00, TAS01]. With the motivation of using DFE in coded systems over ISI channels (or, frequency selective fading channels) as part of a turbo equalization scheme, an iterative receiver with MIMO DFE can be designed [Guc06] in which the MIMO DFE is modified to utilize the soft inputs from the decoder [MSPB01]. We would like to investigate the performance of MIMO iterative DFE receivers with space time turbo coding especially with M-ary signalling.

In order to reduce the hardware and software complexity of multiple antenna systems, antenna selection [Mol03] has become a popular technique which requires only a few RF chains switched to selected antennas. This can be highly effective in reducing the cost and the complexity of space time coded (STC) systems especially over frequency-flat [GD07, GD08] and frequency-selective [GDG04] fading channels. It has been shown that with antenna selection full diversity order can still be achieved when full rank space time codes are used. When the space time codes are rank-deficient then the achievable diversity order depends on the number of selected antennas. Even with channel estimation errors, antenna selection does not degrade the achievable diversity order [GP08]. In our joint research, the performance of space time turbo coding with transmit and/or receive antennas will be evaluated over frequency selective fading channels (Figure 29).

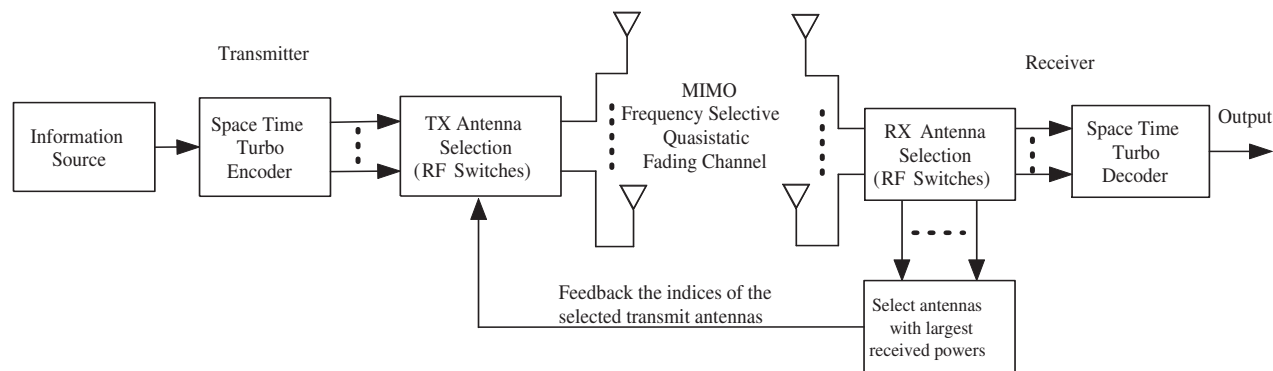


Figure 29: Space time turbo coded systems with transmit and receive antenna selection.

In conclusion, during the WPR4 Istanbul meeting (January 2009), Bilkent/Kadir Has University and PUT group have agreed to collaborate, and as a new project, we have decided to investigate space time turbo coded (STTC) systems over frequency selective fading channels with low complexity equalization/decoding algorithms and antenna selection. Furthermore, the effect of channel estimation errors, channel correlation, performance comparison with STTC-OFDM systems, incorporation of iterative estimation and EXIT charts can also be considered in STTC over FS fading channels scenario. We believe that our research will be helpful in the design of practical units for next generation high rate wireless systems.

### 4.3 Applications of the Role Model framework to reduced complexity decoder design

In [Say08], the “role model” framework was developed for the design of an estimator that seeks to imitate the output of an estimator with superior observations. This framework was developed specifically to facilitate the design of reduced complexity receiver components in an iterative framework.

An optimal receiver component is designed using Bayesian theory without setting any constraints on the complexity of its implementation. This often results in components that can be simulated but that are too complex for practical implementation. A reduced complexity receiver component is often designed heuristically by simplifying the mathematical function calculated by the optimal component. This leaves much room for tuning and improvement of the reduced complexity component, but there is little theory to guide this design phase.

The approach of the role model framework is to separate the component into a complexity reduction stage, where the component is designed heuristically without the help of theory, and a post-processing stage, where the output of the component is processed to optimize the performance. The post-processing stage can be designed optimally using the role model framework and is proved to converge to a Bayesian solution under the constraints posed by the choice of a reduced-complexity component.

In [Say08], no specific application to decoder design was tested and the method was illustrated through hypothetical examples using simple channels such as binary symmetric and binary erasure channels. In [Say09], the measurement of EXIT functions for reduced complexity components was discussed and the role model framework shown to work for the optimization of post-processing for binary LDPC decoders. However, the application to binary decoders is of limited interest since many heuristic solutions exist that are known to approach the optimal performance closely.

The real challenge lies in the application of the role model framework and associated EXIT chart measurement for reduced complexity non-binary decoders such as the Extended Min-Sum [DF07, VDV<sup>+</sup>07]. A collaboration between FTW and CNRS/ENSEA has been initiated to test this application. This collaboration also includes an external partner from University of Bretagne Sud and, while the contact was established through NEWCOM++, the collaboration is pursued in parallel within the FP7 STREP DA

VINCI. Furthermore, due to personnel changes the collaboration will be transferred from FTW to the University of Cambridge that is expected to join NEWCOM++ as a new partner in July 2009.

The targeted impact of this action is to optimize the performance of the reduced complexity EMS algorithm so as to bring it closer to the superior but complex sum-product algorithm. Since non-binary codes are known to exhibit superior performance in particular in the short block length regime, they are potentially of great relevance for applications and future standards. However, they have so far suffered from the far greater complexity of the decoder operations that require the implementation of fast Hadamard transforms. The reduced complexity EMS can go some way towards reducing the complexity of the decoder, but its performance and complexity both depend on a design parameter that sets the length of the reduced lists processed by the constraint nodes of the decoder. Choose this parameter too low and the complexity drops as intended, but the performance drops as well, thereby eradicating the advantage of non-binary codes with respect to their binary equivalents. Choose it too high and the performance is regained, but the complexity soon attains the same level as the optimal sum-product algorithm. The role model approach proposed seeks to optimize the performance, so that the performance loss of the low-complexity implementation with a small list size can be regained.

#### **4.4 Log-likelihood processing**

##### **4.4.1 Log-likelihood ratio clipping**

It can be desirable to restrict the log-likelihood ratios (LLRs) used in an iterative receiver to a certain dynamic range (“LLR clipping”). One reason is the restriction imposed by finite word length representation. Another motivation is that the complexity of practical demappers like the soft-output sphere decoder [SWBB06] can be strongly reduced by incorporating LLR clipping into the algorithm. The impact of LLR clipping on the capacity of a non-iterative MIMO-BICM receiver is studied in [SFNM09]. It is shown that rather small clipping thresholds do not have any impact on system performance. Furthermore, the authors give an interpretation of LLR clipping in terms of an information geometric projection onto a log-convex submanifold of probability distributions.

Discussions at the NEWCOM++ WPR4 meeting in Istanbul have led to the idea to combine the information geometric interpretations of BICM-ID (see above) and LLR clipping. CNRS/L2S and VUT will collaborate on this topic in view of studying the possibility of combining the information geometric interpretations of LLR clipping and of the BICM-ID receiver in order to get insights into BICM-ID systems where LLR clipping is applied.

##### **4.4.2 Selective and modified LLR updates**

LLRs convey reliability information about the corresponding bits. The results about LLR clipping suggest that only a limited LLR range has actual impact on the iterative decoding process. It is therefore feasible to stop updating LLRs whose absolute value exceeds a certain threshold and keep them constant for the remaining iterations. As shown in [NHM08] in the context of multiuser detection this results in significant reductions of receiver complexity.

In [AND09] the authors find a reformulation of the iterative BICM receiver as a hybrid minimum entropy algorithm, which improves performance and accelerates convergence of the traditional algorithm. This is achieved by modifying the LLR calculation in each receiver sub-block using a non-linearity.

Iterative BICM decoding can be regarded as a belief propagation algorithm on a factor graph. The selective update method corresponds to a specific scheduling of the messages that are exchanged between the graph nodes. The altered LLR calculation can be regarded as a modification of the messages.

The goals of CNRS/L2S and VUT in this area are to

- apply and analyze the selective LLR update method in the context of other iterative receivers (BICM-ID, turbo-decoding, etc.);
- combine selective LLR updates with modified LLR calculations to accelerate the turbo iterations;

- study the connection between information geometric interpretation and graphical model interpretation.

#### 4.5 Optimization of the redundancy in channel coding and application to layered channel coding

During some previous work by CNRS, which will be reported during ICC 2009 [SD09], some results were obtained about the transmission of scalable signals on Rayleigh channels using hierarchical modulation. During this work, it was observed that it was sometimes more efficient to increase the redundancy used in channel coding for correcting transmission errors by means of known ("pilot") bits mixed with the information bits, rather than by decreasing the code rate.

This observation is the basis of the proposed cooperative work between CNRS/L2S, PUT, and VUT, and with the cooperation of Alcatel-Lucent (affiliate N++ partner), which funded the first study.

The usefulness of these "pilot" bits was already observed in [CH93], but never really worked out, since this very preliminary paper was published before the turbo-codes were invented, and it may well be the case that this property is really useful only in this context. This question is part of the study.

The work can be decomposed in three steps:

1. Experimental study: checking the first results, and selecting the cases where this property holds. This part will answer the following questions, first:
  - What is the best redundancy allocation between pilot bits and the classical linear combinations ?
  - This question certainly does not have a unique answer, depending on the type of channel (AWGN, Rayleigh) and of code (convolutive, recursive systematic, etc.). In particular, we know that the pilot bits are not useful in LDPC's.
  - What is the best repartition of the pilot bits (equidistant, random locations ?)
2. Theoretical study:
  - Performance analysis. The "pilot bits" change the values of two essential parameters in channel coding: the minimal distance between two codewords, and the distance spectrum. It is still possible to evaluate these quantities, a fact which will explain (certainly) most of the observations obtained in step 1.
  - Complexity analysis. Since the proposed procedure amounts to suppress some choices in the trellises used for channel decoding, the complexity reduces compared to the classical way of working. A precise analysis of this reduction will be done, and means for turning them in actual acceleration of the programs will be investigated (the resulting graph should remain somewhat regular, otherwise it is likely that the programs will be too slow).
3. Finally, this process could lead to the proposal of "layered channel coding". The concept of "layered coding" has already been studied, but it implicitly assumes source coding. If steps 1 and 2 above are fruitful, there is a possibility that the whole process results in the proposal of layered channel codes, which would be well suited for use in conjunction with scalable multimedia streams, a situation of increasing practical importance.

## 5 CONCLUSION

This deliverable has presented the status of research within WPR4 after 18 months. The main research achievements so far within the workpackage have been:

- the discovery of polar codes (Section 2) and the start of joint efforts within the workpackage to investigate the practicality of this new technique, which the first constructive method to achieve capacity in coding history;
- the development of novel phase noise estimation techniques (Section 3.5) based on a basis expansion model and on Monte Carlo sampling;
- the invention of new reduced complexity algorithms for decoding turbo codes (Section 4.1).

Integration was achieved through two plenary meetings and several monitoring video conferences. A number of joint research activities are being pursued on polar coding, joint receivers, and reduced complexity implementation issues. Further events are planned to boost these activities and ensure that they will bear fruit by the scheduled end of NEWCOM++. During the autumn/winter 2009, it is foreseen to hold research retreats by topic where needed. These are not plenary events and offer an opportunity for those working on a specific topic to meet, exchange ideas, and collaborate for a period of 3 to 5 days. In the first half of 2010, we will seek to organize a joint workshop with related workpackages in order to encourage participants to formalize their results and bring them into presentable form. This will lead to the preparation of joint publication, and of a comprehensive overview of achievements and workpackage outputs that will constitute the material of our next and final deliverable DR4.3 to be delivered in December 2010.

**REFERENCES**

- [ADS00] N. Al-Dhahir and A. Sayed, “The finite-length multi-input multi-output MMSE-DFE,” *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2921–2936, Oct 2000.
- [AKM<sup>+</sup>09] E. Arıkan, H. Kim, G. Markarian, Ü. Özgür, and E. Poyraz, “Performance of short polar codes under ML decoding,” in *Proc. ICT MobileSummit*, Santander, Spain, Jun. 2009.
- [AM09] E. Arıkan and G. Markarian, “Two-dimensional polar coding,” in *Proc. Int. Symp. on Commun. Theory and App. (ISCTA)*, Ambleside, UK, Jul. 2009.
- [AND09] F. Alberge, Z. Naja, and P. Duhamel, “New criteria for iterative decoding,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 2493–2496.
- [Ari72] S. Arimoto, “An algorithm for computing the capacity of arbitrary discrete memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 18, pp. 14–20, 1972.
- [Ari07] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetry binary-input memoryless channels,” Oct. 2007, online: <http://arXiv:0807.3917v4>.
- [Ari08a] ———, “Channel polarization: A method for constructing capacity-achieving codes,” in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Toronto, Jul. 2008, pp. 1173–1177.
- [Ari08b] ———, “A performance comparison of polar codes and Reed-Muller codes,” *IEEE Commun. Lett.*, vol. 12, no. 6, pp. 447–449, Jun. 2008.
- [AT08] E. Arıkan and E. Telatar, “On the rate of channel polarization,” Jul. 2008, online: <http://arXiv:0807.3806v3>.
- [AT09] ———, “On the rate of channel polarization,” in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Seoul, Jun. 2009.
- [Bea03] M. Beal, “Variational algorithms for approximate inference,” Ph.D. dissertation, University of Cambridge, May 2003.
- [BG96] C. Berrou and A. Glavieux, “Near optimum error-correcting coding and decoding: Turbo-codes,” *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.
- [BGT93] C. Berrou, A. Glavieux, and P. Thitimajshima, “Near Shannon limit error correcting coding and decoding: Turbo codes,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Geneva, Switzerland, May 1993, pp. 1064–1070.
- [Bla72] R. E. Blahut, “Computation of channel capacity and rate-distortion functions,” *IEEE Trans. Inf. Theory*, vol. 18, pp. 460–473, 1972.
- [BM07a] J. Bhatti and M. Moeneclaey, “Data-aided phase noise estimation from a DCT basis expansion,” in *Proc. IEEE BENELUX/DSP Valley Signal Processing Symposium*, Antwerp, Belgium, Mar. 2007.
- [BM07b] ———, “Pilot-aided carrier synchronization using an approximate DCT-based phase noise model,” in *IEEE International Symposium on Signal Processing and Information Technology*, Cairo, Egypt, Dec. 2007.
- [BM08] ———, “Influence of pilot symbol configuration on data-aided phase noise estimation from a DCT basis expansion,” in *International Networking and Communications Conference*, Lahore, Pakistan, May 2008.

- [BM09] ———, “Feedforward data-aided phase noise estimation from a DCT basis expansion,” *EURASIP Journal on Wireless Communications and Networking*, 2009.
- [BZS06a] B. F. Boroujeny, H. Zhu, and Z. Shi, “Iterative multiuser receivers for CDMA channels: an em-based approach,” *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1896–1909, 2006.
- [BZS06b] ———, “Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems,” *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 1896–1909, 2006.
- [CBD02] B. Classon, K. Blankenship, and V. Desai, “Channel coding for 4G systems with adaptive modulation and coding,” *IEEE Wireless Commun. Mag.*, vol. 9, no. 2, pp. 8–13, Apr. 2002.
- [CH93] O. M. Collins and M. Hizlan, “Determinate state convolutional codes,” *IEEE Trans. Commun.*, vol. 41, no. 12, Dec. 1993.
- [CO00] J. F. Cheng and T. Ottosson, “Linearly approximated Log-MAP algorithms for turbo decoding,” in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Tokyo, Japan, May 2000, pp. 2252–2256.
- [CPD06] H. A. Cirpan, E. Panayirci, and H. Dogan, “Non-data-aided channel estimation for OFDM systems with space-frequency transmit diversity,” *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, Mar. 2006.
- [CRMT08a] G. Coluccia, E. Riegler, C. Mecklenbrauker, and G. Taricco, “Optimum MIMO-OFDM detection with pilot-aided channel state information,” *IEEE Journ. on Sel. Topics in Signal Proc.*, submitted 2008.
- [CRMT08b] ———, “Optimum mimo-ofdm receivers with imperfect channel state information,” in *Proc. IEEE Glob. Comm. Conf. (GLOBECOM)*, 2008, pp. 1–5.
- [Csi75] I. Csiszár, “I-divergence geometry of probability distributions and minimization problems,” *Ann. Prob.*, vol. 3, pp. 146–158, 1975.
- [CT84] I. Csiszár and G. Tusnády, “Information geometry and alternating minimization procedure,” *Statistics and Decisions*, pp. 205–237, 1984, supplemental Issue Nr. 1.
- [CT91] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, New York, 1991.
- [CXX99] L. Cong, W. Xiaofu, and Y. Xiaoxin, “On SOVA for nonbinary codes,” *IEEE Trans. Wireless Commun.*, vol. 3, pp. 335–337, Dec. 1999.
- [DCP07] H. Dogan, H. A. Cirpan, and E. Panayirci, “Iterative channel estimation and decoding of turbo coded SFBC-OFDM systems,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3090–3101, 2007.
- [DCP08] ———, “An efficient joint channel estimation and decoding algorithm for turbo-coded space-time orthogonal frequency division multiplexing receivers,” *IET Communications.*, vol. 2, no. 7, pp. 886–894, Aug. 2008.
- [DDP01] D. Divsalar, S. Dolinar, and F. Pollara, “Iterative turbo decoder analysis based on density evolution,” *IEEE J. Sel. Areas Commun.*, Jan. 2001.
- [DF07] D. Declercq and M. Fossorier, “Decoding algorithms for nonbinary LDPC codes over GF(q),” *IEEE Trans. Commun.*, vol. 55, no. 4, pp. 633–643, Apr. 2007. [Online]. Available: <http://publi-etis.ensea.fr/2007/DF07>
- [DLR77] A. Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977, series B.

- [Don06] D. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [DP02] H. Dai and H. V. Poor, “Iterative space-time processing for multiuser detection in multipath CDMA channels,” *IEEE Trans. Commun.*, vol. 50, no. 9, pp. 2116–2127, 2002.
- [DRU97] L. Duan, B. Rimoldi, and R. Urbanke, “Approaching the AWGN channel capacity without active shaping,” in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jun. 1997, p. 374.
- [DW05] A. Doucet and X. Wang, “Monte Carlo methods for signal processing,” *IEEE Signal Process. Mag.*, vol. 153, pp. 152–170, Nov. 2005.
- [FH94] J. Fessler and A. Hero, “Space-alternating generalized expectation-maximization algorithm,” *IEEE Trans. Signal Process.*, pp. 2664–2667, October 1994.
- [GD05] T. Gucluoglu and T. M. Duman, “Soft input soft output stack equalization for MIMO frequency selective fading channels,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 1, May 2005, pp. 510–514.
- [GD07] ———, “Performance of space-time coded systems with transmit antenna selection,” in *Proc. Conf. Inform. Sciences and Systems (CISS)*, March 2007, pp. 863–868.
- [GD08] ———, “Space-time coded systems with joint transmit and receive antenna selection,” *IEEE Trans. Wireless Commun.*, vol. 7, pp. 3056 – 3065, Aug. 2008.
- [GDG04] T. Gucluoglu, T. M. Duman, and A. Ghayeb, “Antenna selection for space time coding over frequency-selective fading channels,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, no. 10, May 2004, pp. iv–709 – iv–712.
- [GG98] W. J. Gross and P. G. Gulak, “Simplified MAP algorithm suitable for implementation of turbo decoders,” *IEE Electron. Lett.*, vol. 34, no. 16, pp. 1577–1578, Aug. 1998.
- [GP08] T. Gucluoglu and E. Panayirci, “Performance of transmit and receive antenna selection in the presence of channel estimation errors,” *IEEE Commun. Lett.*, vol. 12, pp. 371–373, May 2008.
- [Guc06] T. Gucluoglu, “Reduced complexity algorithms for space time coded systems over frequency selective fading channels,” Ph.D. dissertation, Arizona State University, Jun. 2006.
- [HKB08] K.-L. Hsiung, S.-J. Kim, and S. Boyd, “Tractable approximate robust geometric programming,” *Springer Optim. Eng. Journal*, vol. 9, no. 2, pp. 95–118, June 2008.
- [HKU09] N. Hussami, S. B. Korada, and R. Urbanke, “Polar codes for channel and source coding,” Jan. 2009, online: <http://arXiv:0901.2370v1>.
- [HLR<sup>+</sup>08] B. Hu, I. Land, L. Rasmussen, R. Piton, and B. Fleury, “A divergence minimization approach to joint multiuser decoding for coded CDMA,” *IEEE Journal on Selected Areas in Communications - Special Issue on Multiuser Detection*, vol. 26, no. 3, pp. 432–445, Apr. 2008.
- [HM97] B. Honary and G. Makarian, *Trellis decoding of block codes*. Kluwer, 1997.
- [HW07] X. Huang and H. C. Wu, “Robust and efficient intercarrier interference mitigation for OFDM systems in time-varying fading channels,” *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2517–2528, Sep. 2007.

- [HYCV04] Y. Hong, J. Yuan, Z. Chen, and B. Vucetic, "Space-time turbo trellis codes for two, three, and four transmit antennas," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 318–328, March 2004.
- [I. 06] I. Barhum, G. Leus, and M. Moonen, "Equalization for OFDM over doubly-selective channels," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1445–1458, Apr. 2006.
- [IH77] H. Imai and S. Hiraoka, "A new multilevel coding method using error correcting codes," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 3, pp. 371–377, May 1977.
- [ITA04] S. Ikeda, T. Tanaka, and S. Amari, "Information geometry of turbo and low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1097–1114, Jun. 2004.
- [Jaf05] H. Jafarkhani, *Space-Time Coding. Theory and Practice*. Cambridge University Press, 2005.
- [KDM<sup>+</sup>96] A. B. Kiely, S. J. Dolinar, R. J. McEliece, L. L. Ekroot, and W. Lin, "Trellis decoding complexity of linear block codes," *IEEE Trans. Inf. Theory*, vol. 42, pp. 1687–1697, Nov. 1996.
- [KF03] A. Kocian and B. Fleury, "EM-based joint data detection and channel estimation of DS-SS-CDMA signals," *IEEE Trans. Commun.*, vol. 51, no. 10, pp. 1709–1720, October 2003.
- [KFL01] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, 2001.
- [KLF07] A. Kocian, I. Land, and B. Fleury, "Joint channel estimation, partial successive interference cancellation, and data decoding for DS-SS-CDMA based on the SAGE algorithm." *IEEE Trans. Commun.*, vol. 55, no. 6, pp. 1231–1241, Jun. 2007.
- [KM<sup>+</sup>09] G. E. Kirek, C. N. Manchon, B. Fleury, B. Hu, and L. P. B. Christensen, "Relative entropy minimization in joint channel estimation and decoding for MIMO-OFDM," 2009, submitted to IEEE Global Communications Conference (GLOBECOM).
- [KŞ08] S. B. Korada and E. Şaşoğlu, "A class of transformations that polarize symmetric binary-input memoryless channels," Nov. 2008, online: <http://arXiv:0811.1770v1>.
- [KSP<sup>+</sup>02] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, pp. 1227–1239, Aug. 2002.
- [KŞU09] S. B. Korada, E. Şaşoğlu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," Jan. 2009, online: <http://arXiv:0901.0536v2>.
- [KU09] S. B. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," Mar. 2009, online: <http://arxiv.org/abs/0903.0307v1> [CS.IT].
- [LC98] S. Liu and R. Chen, "Sequential Monte Carlo method for dynamical systems," *Journal on the American Statistical Association*, vol. 93, pp. 1031–1041, 1998.
- [LDH<sup>+</sup>07] H. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, 2007.
- [LDJC04] S. Lin and J. Daniel J. Costello, *Error Control Coding*, 2nd ed. Pearson, New Jersey, 2004.
- [MD04] G. Matz and P. Duhamel, "Information geometric formulation and interpretation of accelerated blahut-arimoto-type algorithms," in *Proc. IEEE Inform. Theory Workshop (ITW)*, 2004.

- [MDdC02] B. Muquet, P. Duhamel, and M. de Courville, "A geometrical interpretation of iterative turbo decoding," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Lausanne, Switzerland, May 2002.
- [MKF<sup>+</sup>09a] C. N. Manchon, G. E. Kirkelund, B. Fleury, P. Mogensen, L. Deneire, T. B. Sørensen, and C. Rom, "Channel estimation based on divergence minimization for ofdm systems with co-channel interference," 2009, accepted for publication at IEEE International Conference on Communications (ICC).
- [MKF<sup>+</sup>09b] —, "Interference cancellation based on divergence minimization for mimo-ofdm receivers," 2009, submitted to IEEE Global Communications Conference (GLOBECOM).
- [Mol03] A. F. Molisch, "MIMO systems with antenna selection - An overview," in *Proc. Radio and Wireless Conference*, vol. 37, no. 20, Aug. 2003, pp. 167–170.
- [MP04] X. Ma and L. Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3331–3343, Dec. 2004.
- [MSPB01] M. Marandian, M. Salehi, J. G. Proakis, and F. Blackmon, "Iterative decision feedback equalizer for time dispersive channels," in *Proc. Conference on Information Sciences and Systems*, John Hopkins University, Baltimore, Maryland, March 2001.
- [MT09] R. Mori and T. Tanaka, "Performance and construction of polar codes on symmetric binary-input memoryless channels," Jan. 2009, online: <http://arXiv:0901.2207v1>.
- [Mul54] D. E. Muller, "Application of boolean algebra to switching circuit design and to error correction," *IRE Trans. Electronic Computers*, vol. EC-3, pp. 6–12, Sept 1954.
- [Muq01] B. Muquet, "Novel receiver and decoding schemes for wireless ofdm systems with cyclic prefix or zero padding," Ph.D. dissertation, ENST, Paris (France), Jun. 2001.
- [NAD09] Z. Naja, F. Alberge, and P. Duhamel, "Geometrical interpretation and improvements of the Blahut-Arimoto's algorithm," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2009.
- [NHM08] C. Novak, F. Hlawatsch, and G. Matz, "Low-complexity factor graph receivers for spectrally efficient MIMO-IDMA," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Apr. 2008, pp. 770–774.
- [NP96] L. Nelson and H. V. Poor, "Iterative multiuser receivers for CDMA channels: an EM-based approach," *IEEE Trans. Commun.*, vol. 44, no. 12, pp. 1700–1710, 1996.
- [P. 04] P. Schniter, "Low-complexity equalization of OFDM in doubly-selective channels," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1002–1011, Apr. 2004.
- [PC07] E. Panayirci and H. A. Cirpan, "EM-based MAP channel estimation and data detection for downlink MC-CDMA systems," in *IEEE Wireless Communications and Networking Conference 2007*, Honk Kong, Mar. 2007.
- [PCM05] E. Panayirci, H. Cirpan, and M. Moeneclaey, "A sequential Monte-Carlo method for blind phase noise estimation and data detection," in *Proc. European Sign. Proc. Conf. (EU-SIPCO)*, Antalya, Turkey, Sep. 2005.
- [PCMN06] E. Panayirci, H. Cirpan, M. Moeneclaey, and N. Noels, "Blind phase noise estimation in OFDM systems by sequential Monte Carlo method," *European Trans. on Telecommunications*, vol. 17, no. 6, pp. 685–693, Dec. 2006.

- [PKPR09] E. Panayirci, A. Kocian, H. V. Poor, and M. Ruggieri, “A Monte-Carlo implementation of the SAGE algorithm for joint soft multiuser and channel parameter estimation,” 2009, accepted for publication at Proc. Workshop on Signal Processing Advances in Wireless Communications (SPAWC).
- [Pro00] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill Inc., 2000.
- [PS99] M. Pitt and N. Shephard, “Filtering via simulation: auxiliary particle filter,” *Journal of the American Statistical Association*, vol. 94, pp. 590–599, 1999.
- [PST<sup>+</sup>09] S. Papaharalabos, M. Sybis, P. Tyczka, P. T. Mathiopoulos, G. Masera, and M. Martina, “Reduced complexity algorithms for near-optimal decoding of turbo and turbo TCM codes,” in *Proc. NEWCOM<sup>++</sup>-ACoRN Workshop*, Barcelona, Spain, Mar. 2009.
- [PSTM09] S. Papaharalabos, M. Sybis, P. Tyczka, and P. T. Mathiopoulos, “Modified Log-MAP algorithm for simplified decoding of turbo and turbo TCM codes,” in *Proc. IEEE Vehicular Tech. Conf. (VTC)*, Barcelona, Spain, April 2009.
- [RC99] C. Roberto and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.
- [Ree54] I. Reed, “A class of multiple-error-correcting codes and the decoding scheme,” *IRE Trans. Inform. Theory*, vol. 4, no. 4, pp. 39–44, Sept 1954.
- [RHV97] P. Robertson, P. Hoeher, and E. Villebrun, “Optimal and sub-optimal maximum a posteriori algorithms suitable for turbo decoding,” *European Trans. on Telecommunications*, vol. 8, pp. 119–125, Mar. 1997.
- [RW98] P. Robertson and T. Wörz, “Bandwidth-efficient turbo trellis-coded modulation using punctured component codes,” *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 206–218, Feb. 1998.
- [Say08] J. Sayir, “What makes a good role model,” Sep. 2008, preprint <http://arxiv.org/abs/0809.1300>.
- [Say09] —, “Measuring exit charts for low complexity decoders,” in *Proc. Int. Symp. on Commun. Theory and App. (ISCTA)*, Ambleside, UK, Jul. 2009.
- [Sch08] L. Schmitt, “On iterative receiver algorithms for concatenated codes,” Ph.D. dissertation, RWTH Aachen University, Mar. 2008.
- [SCP04] A. Sen, H. Cirpan, and E. Panayirci, “Blind multiuser detection: A subspace approach,” *Frequenz.*, vol. 58, no. 3, pp. 1–5, Dec. 2004.
- [SD09] S. M. S. Sadough and P. Duhamel, “On the interaction between channel coding and hierarchical modulation,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Dresden, Germany, Jun. 2009.
- [SDC<sup>+</sup>09] F. Simoens, D. Duyck, H. Cirpan, E. Panayirci, and M. Moeneclaey, “Monte Carlo solutions for blind phase noise estimation,” *EURASIP Journal on Wireless Communications and Networking*, 2009.
- [SFNM09] S. Schwandter, P. Fertl, C. Novak, and G. Matz, “Log-likelihood ratio clipping in MIMO-BICM systems: Information geometric analysis and impact on system capacity,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2009, pp. 2433–2436.
- [SM01] R. Sivasankaran and S. W. McLaughlin, “Twin-stack decoding of recursive systematic convolutional codes,” *IEEE Trans. Commun.*, vol. 49, pp. 1158–1167, Jul. 2001.

- [STPM09] M. Sybis, P. Tyczka, S. Papaharalabos, and P. T. Mathiopoulos, "Reduced-complexity algorithms for near-optimal decoding of turbo TCM codes," *IEE Electron. Lett.*, vol. 45, no. 5, pp. 278–279, February 2009.
- [SWBB06] C. Studer, M. Wenk, A. Burg, and H. Bölcskei, "Soft-output sphere decoding: Performance and implementation aspects," in *Proc. 40th Asilomar Conf. Signals, Systems, and Computers*, Nov. 2006, pp. 2071–2076.
- [TAS01] C. Tiestav, A. Ahlen, and M. Sternad, "Realizable MIMO decision feedback equalizers: structure and design," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 121–133, Jan 2001.
- [tB01] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, pp. 1727–1737, Oct 2001.
- [TB05] G. Taricco and E. Biglieri, "Space-time decoding with imperfect channel estimation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1874–1888, Jul. 2005.
- [TC07] G. Taricco and G. Coluccia, "Optimum receiver design for correlated Rician fading MIMO channels with pilot-aided detection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1311–1321, Sep. 2007.
- [Ton03] A. M. Tonello, "MIMO MAP equalization and turbo decoding in interleaved space-time coded systems," *IEEE Trans. Commun.*, vol. 51, pp. 155–160, Feb. 2003.
- [TSSA07] S. Talakoub, L. Sabeti, B. Shahrava, and M. Ahmadi, "An improved Max-Log-MAP algorithm for turbo decoding and turbo equalization," *IEEE Trans. Instrum. Meas.*, vol. 56, no. 3, pp. 1058–1063, June 2007.
- [VB03] R. Visoz and A. O. Berthet, "Iterative decoding and channel estimation for space-time BICM over MIMO block fading multipath AWGN channel," *IEEE Trans. Commun.*, vol. 51, pp. 1358–1367, Aug 2003.
- [VDV<sup>+</sup>07] A. Voicila, D. Declercq, F. Verdier, M. Fossorier, and P. Urard, "Low complexity, low memory ems algorithm for non-binary LDPC codes," in *IEEE ICC*, Glasgow, England, Jun. 2007. [Online]. Available: <http://publi-etis.ensea.fr/2007/VDVU07>
- [VF00] J. Vogt and A. Finger, "Improving the Max-Log-MAP turbo decoder," *IEE Electron. Lett.*, vol. 36, no. 23, pp. 1937–1939, Nov. 2000.
- [WM06] J. Wehinger and C. Mecklenbräuker, "Iterative CDMA multiuser receiver with soft decision-directed channel estimation," *IEEE Trans. Signal Process.*, vol. 54, pp. 3922–3934, October 2006.
- [WP99] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.
- [WYY06] H. Wang, H. Yang, and D. Yang, "Improved Log-MAP decoding algorithm for turbo-like codes," *IEEE Commun. Lett.*, vol. 10, no. 3, pp. 186–188, March 2006.
- [X. 98] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 667–690, 1998.
- [XLG07] Y. Xie, Q. Li, and C. N. Georghiades, "On some near optimal low complexity detectors for MIMO fading channels," *IEEE Trans. Commun.*, vol. 6, no. 4, pp. 1182–1186, 2007.
- [ZP08a] F. Zhang and H. D. Pfister, "Compressed sensing and linear codes over real numbers," in *Proc. Inf. Theory and applications Work.*, La Jolla, CA, Feb. 2008.

- [ZP08b] —, “On the iterative decoding of highrate LDPC codes with applications in compressed sensing,” in *Proc. Allerton Conf. on Communications, Control, and Computing*, Monticello, IL, 2008.