Challenges and Some New Directions in Channel Coding

Erdal Arıkan, Najeeb ul Hassan, Michael Lentmaier, Guido Montorsi, and Jossy Sayir

Abstract: **Three areas of ongoing research in channel coding are surveyed, and recent developments are presented in each area: Spatially coupled low-density parity-check (LDPC) codes, non-binary LDPC codes, and polar coding.**

Index Terms: **Channel polarization, low-density parity-check (LDPC) codes, non-binary codes, polar codes, spatial coupling.**

I. INTRODUCTION

THE history of channel coding began hand in hand with Shannon's information theory [1]. Following on the pioneering work of Golay [2] and Hamming [3], the majority of linear codes developed in the early ages of coding theory were "error correction" codes in the sense that their aim was to correct errors made by the channel. The channel was universally assumed to be a binary symmetric channel (BSC). The study of error correction codes culminated with the invention of Reed-Solomon codes [4] in 1960, which are maximum distance separable (MDS) over non-binary fields and hence are guaranteed to correct or detect the largest number of errors possible for a given code length and dimension.

In parallel to the evolution of linear block codes, the invention of convolutional codes by Elias in 1955 [5] lead to a different approach and to the invention of trellis-based decoding methods such as the Viterbi algorithm [6], [7] and the Bahl, Cocke, Jelinek, and Raviv (BCJR) algorithm [8]. Both of these algorithms can be easily adapted to any channel and hence generalise the concept of error correction to general channels for which the channel output alphabet is not necessarily the same as the input alphabet. For such channels, it does not make much sense to talk about "channel errors". We now speak of "channel coding" rather than "error correction coding". Further progress in channel coding was made by Ungerboeck [9] by linking coding to modulation for convolutional codes.

In 1993, Berrou and co-authors shocked the coding research community in [10] by designing a coding system known as "turbo codes" that achieved a quantum leap in the performance of codes over general channels. They obtained very good error performance within a small margin of the channel capacity, something that had been thought impossible with practical systems and moderate complexity by most coding theorists. Yet Berrou's approach achieved this in an eminently implementable system and with linear decoding complexity in terms of code block length. In the subsequent scramble to explain the theory behind this puzzling performance, a method originally developed by Gallager in his Ph.D. dissertation [11], known as low-density parity-check (LDPC) coding was rediscovered in [12] and shown to have comparable properties. Both these methods have become the workhorses of modern communication standards, with arguments about the technical advantages of one over the other mostly obscured by business and standardization interests of who is arguing. What is clear and undisputed is that LDPC codes are easier to explain and analyse and hence should probably take precedence over turbo codes in teaching. It is nowadays well known that both LDPC codes and turbo codes can be viewed as sparse codes on graphs. As a consequence they share a lot of properties, and any construction or analysis method that can be applied to one of them can usually be replicated for the other.

We could conclude this history of coding here and bury the topic into dusty textbooks, sending it the same way as classical Newtonian mechanics[1] and other topics made obsolete by quantum leaps in research. Many coding researchers nowadays are confronted with the recurrent "coding is dead" motto [13] of experts claiming that, now that capacity is achieved, there is nothing further to be researched in the field. In fact, as this paper will contribute to showing, coding is still an ongoing and very active topic of research with advances and innovations to address important and practical unsolved problems.

Current hurdles in the applicability of modern coding techniques can be classified in two categories:

Complexity While turbo and LDPC codes have brought capacity-approaching performance within reach of implementable systems, implementable does not necessarily mean practical. The complexity of codes that perform well under practical constraints such as limited decoding delay and high spectral efficiency is still a major hurdle for low power implementations in integrated circuits. There is a serious need for new methods that simplify code design, construction, storage, and decoder implementation.

New applications Turbo and LDPC codes can be seen to "solve" the capacity problem for elementary point-to-point channels. Recent years have seen advances in information the-

Manuscript received December 15, 2014.

This work was supported by the FP7 Network of Excellence NEWCOM# under grant agreement 318306.

E. Arıkan is with Bilkent University, Ankara, Turkey, email: arikan@ee.bilkent.edu.tr.

N. Hassan is with TU Dresden, Germany, email: Najeeb_ul.Hassan@tu-dresden.de.

M. Lentmaier is with Lund University, Sweden, email: michael.lentmaier@eit.lth.se.

G. Montorsi is with Politecnico di Torino, Italy, email: guido.montorsi@polito.it.

J. Sayir is with the University of Cambridge, UK, email: jossy.sayir@eng.cam.ac.uk.

Digital object identifier 10.1109/JCN.2015.000063

[1] Apologies to mechanics researchers for the seemingly disparaging remark. In fact, we are aware that classical mechanics is an ongoing and modern research topic as evidenced by many journals and conferences, just as coding theory is.

ory for many multi-user channels such as the multiple access, broadcast, relay and interference channels. As communication standards become more ambitious in exploiting the available physical resources such as spectrum and geographical reach, there is a push to switch from interference limited parallel point-to-point protocols to true multi-user processing with joint encoding and/or decoding. There is a need for coding methods that can do this efficiently for all of the scenarios described. Furthermore, theory has gone further than pure communications by expanding to distributed compression and joint source/channel coding, distributed storage, network coding, and quantum channels and protocols. All of these new theories come with their own requirements and constraints for coding, and hence coding research is far from dead when it comes to these new applications.

The paper will present three areas of ongoing research in coding, all of which have some degree of relevance to the two challenges described.

In Section II, we will address spatially coupled LDPC codes, which have a structure akin to convolutional codes. For spatially coupled codes the asymptotic performance of an iterative decoder is improved to that of an optimal decoder, which opens the way for new degrees of freedom in the code design. For example, it is possible to achieve capacity universally for a large class of channels with simple regular spatially coupled-LDPC (SC-LDPC) codes where irregular LDPC codes would require careful individual optimizations of their degree profiles. We will discuss the design of SC-LDPC codes for flexible rates, efficient window decoding techniques for reduced complexity and latency, and the robustness of their decoding for mobile radio channels. In Section III, we will address non-binary LDPC and related codes. These are codes over higher order alphabets that can, for example, be mapped directly onto a modulation alphabet, making them interesting for applications requiring high spectral efficiency. While these have been known for a while, the complexity of decoding has made them unsuited for most practical applications. In this section, we will discuss research advances in low-complexity decoding and also present a class of LDPC codes with an associated novel decoding algorithm known as analog digital belief propagation (ADBP) whose complexity does not increase with alphabet size and hence constitutes a promising development for very high-spectral efficiency communications. Finally, in Section IV, we will introduce polar coding, a new technique introduced in [14] based on a phenomenon known as channel polarization, that has the flexibility and versatility to be an interesting contender for many novel application scenarios.

II. SC-LDPC CODES

The roots of LDPC codes [11] trace back to the concept of random coding. It can be shown that a randomly generated code decoded with an optimal decoder exhibits very good performance with high probability. However, such a decoder is infeasible in practice because the complexity will increase exponentially with the code length. The groundbreaking idea of Gallager was to slightly change the random ensemble in such a way that the codes can be decoded efficiently by an iterative algorithm, now known as belief propagation (BP) decoding. His LDPC codes were defined by sparse parity-check matrices $\mathbf{H}$ that contained a fixed number of K and J non-zero values in every row and column, respectively, known as *regular* LDPC codes. Gallager was able to show that the minimum distance of typical codes of the ensemble grows linearly with the block length, which guarantees that very strong codes can be constructed if large blocks are allowed. The complexity per decoded bit, on the other hand, is independent of the length if the number of decoding iterations is fixed.

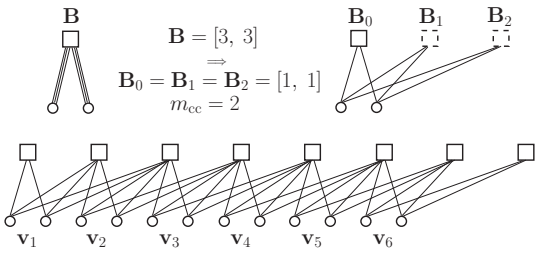

Fig. 1. Illustration of edge spreading: the protograph of a (3,6)-regular block code represented by a base matrix $\mathbf{B}$ is repeated $L = 6$ times and the edges are spread over time according to the component base matrices $\mathbf{B_0}$, $\mathbf{B_1}$, and $\mathbf{B_2}$, resulting in a terminated LDPC code.

The asymptotic performance of an iterative decoder can be analyzed by tracking the probability distributions of messages that are exchanged between nodes in the Tanner graph (*density evolution*) [15]. The worst channel parameter for which the decoding error probability converges to zero is called the *BP threshold*. The BP thresholds of turbo codes are actually better than those of the original regular LDPC codes of Gallager (see Fig. 2 of [15]). A better BP threshold is obtained by allowing the nodes in the Tanner graph to have different degrees [15]. By optimizing the degrees of the resulting *irregular* LDPC code ensembles it is possible to push the BP thresholds towards capacity. However, this requires a large fraction of low-degree variable nodes, which leads to higher error floors at large signal-to-noise ratios (SNRs). As a consequence of the degree optimization, the capacity-achieving sequences of irregular LDPC codes no longer show a linear growth of the minimum distance.

LDPC convolutional codes were invented by Jiménez Felström and Zigangirov in [16]. Like LDPC block codes, they are defined by sparse parity-check matrices, which can be infinite but have a band-diagonal structure like the generator matrices of classical convolutional codes. When the parity-check matrix is composed of individual permutation matrices, the structure of an LDPC code ensemble can be described by a *protograph* [17] (a prototype graph) and its corresponding base matrix $\mathbf{B}$. The graph of an LDPC convolutional code can be obtained by starting from a sequence of L independent protographs of an LDPC block code, which are then interconnected by spreading the edges over blocks of different time instants [18]. The maximum width of this *edge spreading* determines the memory, m_{cc}, of the resulting chain of length L that defines the LDPC convolutional code ensemble. Since the blocks of the original protograph codes are coupled together by this procedure, LDPC convolutional codes are also called SC-LDPC. Fig. 1 shows an illustration of the edge spreading procedure.

A BP threshold analysis of LDPC convolutional codes shows that the performance of the iterative decoder is improved significantly by spatial coupling. In fact, the results in [19] show that asymptotically, as L tends to infinity, the BP threshold is boosted to that of the optimal maximum a posteriori (MAP) decoder. Stimulated by these findings, Kudekar, Richardson, and Urbanke developed an analytical proof of this *threshold saturation* phenomenon [20], [21]. More recently, potential functions have been identified as a powerful tool for characterizing the connection between MAP thresholds and BP thresholds [22]. All these approaches make use of the *area theorem* [23] in order to derive bounds on the MAP threshold and prove threshold saturation for spatially coupled codes. Since the MAP thresholds of regular LDPC ensembles with increasing node degrees are known to converge to capacity, it follows that spatial coupling provides a new way of provably achieving capacity with low-complexity iterative BP decoding - not only for the BEC but also for the AWGN channel. Furthermore, the spatially coupled code ensembles inherit from the uncoupled counterparts, the linearly increasing minimum distance property [24]. This combination of capacity achieving thresholds with low complexity decoding and linearly increasing distance is quite unique and has attracted a lot of interest in the research community.

The capacity achieving property of regular SC-LDPC codes raises the question whether irregularity is still needed at all. In principle, it is possible for any arbitrary rational rate to construct regular codes that guarantee a vanishing gap to capacity with BP decoding. On the other hand, for some specific code rates, the required node degrees and hence the decoding complexity increase drastically. But even if we neglect the complexity, there exists another problem of practical significance that so far has not received much attention in the literature: For large node degrees J and K, the threshold saturation effect will only occur for larger values of the coupling parameter m_{cc}, as illustrated for the BEC in Fig. 2 [25]. We can see that for a given coupling width $w = m_{\mathrm{cc}} + 1$, the gap to capacity becomes small only for certain code rates R, and it turns out that these rates correspond to the ensembles for which the variable node degree J is small.

Motivated by this observation, in [25] some nearly-regular SC-LDPC code ensembles were introduced, which are built upon the mixture of two favorable regular codes of same variable node degree. The key is to allow for a slight irregularity in the code graph to add a degree of freedom that can be used for supporting arbitrary rational rates as accurately as needed while keeping the check and variable degrees as low as possible. These codes exhibit performance close to the Shannon limit for all rates in the rate interval considered, while having a decoder complexity as low as for the best regular codes. The exclusion of variable nodes of degree two in the construction ensures that the minimum distance of the proposed ensembles increases linearly with the block length, i.e., the codes are asymptotically good.

A. Efficient Decoding of Spatially Coupled Codes

In order to achieve the MAP threshold, the number L of coupled code blocks should be sufficiently large for reducing the rate loss due to termination of the chain. But running the BP decoder over the complete chain of length L would then result in a large latency and decoding complexity and hence is not feasible

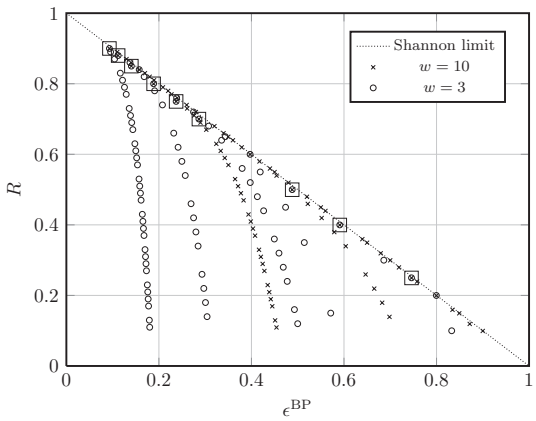

Fig. 2. Density evolution thresholds ϵ^{BP} for (J, K)-regular SC-LDPC ensembles in comparison with the Shannon limit ϵ^{Sh}. The coupling width w is equal to $m_{\mathrm{cc}} + 1$. For a given rate $R = 1 - J/K$, the smallest pair of values J and K are chosen under the condition that $J \geq 3$. The ensembles with minimum variable node degree $J = 3$ are highlighted with squares.

in practical scenarios. However, thanks to the limited width of the non-zero region around the diagonal, SC-LDPC codes can be decoded in a continuous fashion using a sliding window decoder [26] of size W ($W \ll L$). As a result, decoding latency and decoding complexity become independent of L. Moreover, the storage requirements for the decoder are reduced by a factor of L/W compared to a non-windowed decoder. An example of the window decoder of size $W = 4$ is given in Fig. 3.

It has been shown in [27] that for equal structural latency, SC-LDPC codes under window decoding outperform LDPC codes for short to long latency values and outperform convolutional codes from medium to long latency values. For applications requiring very short latency, Viterbi decoded convolutional codes were still found to be the optimal choice [27]–[29]. Note that only structural latency was considered in all these comparisons which is defined as the number of bits required before decoding can start. It therefore can be concluded that for low transmission rate applications (in the range of bit/seconds), convolutional codes with moderate constraint length are favorable since the delay in filling the decoder buffer dominates the overall latency. Whereas, for applications with transmission rates in excess of several Gb/s., e.g., short range communication, medium to large structural latency is tolerable and strong codes such as SC-LDPC codes provide gain in performance compared to the conventional convolutional codes. Another advantage of using a window decoder is the flexibility in terms of decoding latency at the decoder. Since the window size W is a decoder parameter, it can be varied without changing the code, providing a flexible trade-off between performance and latency [27].

In BP decoding, messages are passed between the check and variable nodes according to a parallel (flooding) or serial (on-demand) rule [30]. In both schedules, all the nodes in the graph are typically updated at every decoding iteration (*uniform schedules*). For both LDPC and SC-LDPC, a uniform serial decoding schedule results in a factor of two in complexity reduction when applied over the complete length of the code [30]. However, this gain in complexity reduction reduces to only 20% when uniform serial schedules are applied within a decoding window [31],

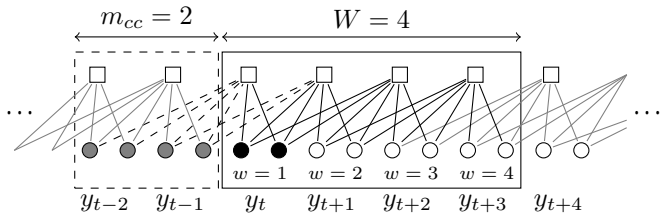

Fig. 3. Window decoder of size $W = 4$ at time t. The gray variable nodes represent decoded blocks and the black variable nodes (y_t) are the target block within the current window. The dashed lines represent the read access to the m_{cc} previously decoded blocks.

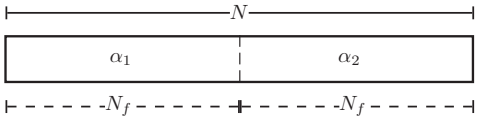

Fig. 4. Illustration of block-fading channel for a codeword of length N and $F = 2$.

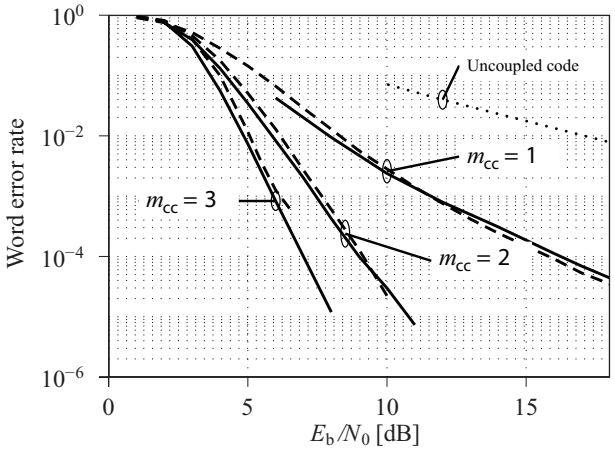

Fig. 5. Density evolution outage for SC-LDPC codes with memory 0,1,2, and 3. The bold lines represent the DEO and dashed lines represent the simulation results when a code with $N = 200$, $L = 100$, is decoded using a window decoder, $F = 2$.

[32]. In order to reduce the decoding complexity for window decoding, non-uniform window decoding schedules has been introduced in [31], [32], which result in 50% reduction in complexity compared to uniform decoding schedules. The reduction in decoding complexity can be achieved by avoiding unnecessary updates of nodes not directly connected to the first position in the window. Only nodes that show improvement based on their bit error rate (BER) compared to the previous iteration are updated in the next iteration.

B. Performance over Mobile Radio Channels

One of the most remarkable features of spatially coupled codes is their universality property, which means that a single code construction performs well for a large variety of channel conditions. For discrete-input memoryless symmetric channels the universality of SC-LDPC codes has been shown in [21]. In this section we consider the block-fading channel and demonstrate that SC-LDPC codes show a remarkable performance on this class of channels.

The block-fading channel was introduced in [33] to model the mobile-radio environment. This model is useful because the channel coherence time in many cases is much longer than one symbol duration and several symbols are affected by the same fading coefficient. The coded information is transmitted over a finite number of fading blocks to provide diversity. An example where a codeword of length N spreads across $F = 2$ fading realizations is shown in Fig. 4. In general, when dealing with block-fading channels, two strategies can be adopted: Coding with block interleaving or coding with memory [34]. Spatially-coupled codes, with their convolutional structure among LDPC codes, are expected to be a nice example of the second strategy.

The block-fading channel is characterized by an *outage probability*, which serves as a lower bound on the word error probability for any code decoded using a maximum likelihood decoder. In terms of density evolution, the density evolution outage (DEO) is the event when the bit error probability does not converge to zero for a fixed value of SNR after a finite or an infinite number of decoding iterations are performed [35]. The probability of density evolution outage, for a fixed value of SNR, can then be calculated using a Monte Carlo method considering significant number of fading coefficients.

Since the memory of the code plays an important role to exploit code diversity, we consider SC-LDPC codes with increasing memory from 0 to 3. The diversity of the code, which is defined as the slope of the WER curve, is calculated numerically from the DEO curves presented in Fig. 5. For uncoupled LDPC codes, the diversity is limited to $d = 1.3$ (see dotted line in Fig. 5). This case can be interpreted as an SC-LDPC code with $m_{\mathrm{cc}} = 0$. If we now increase the coupling parameter to 1, 2, and 3, then the diversity of SC-LDPC codes increases to 3, 6, and 10, respectively [36]. The figure also shows the simulation results (dashed lines) for finite length codes when the length of each individual coupled code block is $N = 200$. The simulation results match closely with the calculated DEO bounds.

An alternative approach to codes with memory is taken by the *root-LDPC* codes [35] with a special check node structure called *rootcheck*. Full diversity ($d = F = 1/R$) is provided to the systematic information bits only by connecting only one information bit to every rootcheck. However, designing root-LDPC codes with diversity order greater than 2 requires codes with rate less than $R = 1/2$. The special structure of the codes makes it a complicated task to generate good root-LDPC codes with high diversity (and thus low rate).

Another key feature of SC-LDPC codes is its robustness against the variation in the channel. In case of root-LDPC codes, the parity-check matrix has to be designed for the specific channel parameter F to provide a diversity of $d = F$ to the information bits. However for SC-LDPC codes, it can be shown that the code design for a specific value of F is not required whereas the diversity order strongly depends on the memory of the code. This feature makes them very suitable for a wireless mobile environment.

III. NON-BINARY CODES AND HIGH SPECTRAL-EFFICIENCY CODES

As already mentioned, LDPC codes were originally proposed by Gallager [11] and re-discovered by MacKay & al. [12] in the years after the invention of turbo codes [10]. LDPC codes have been adopted in several current standards, e.g., IEEE 802.11n wi-fi standard, DVB-S2, T2, and C2 digital video broadcasting standards for satellite, terrestrial, and cable channels, 10 GBase-T Ethernet over twisted pairs, G.hn/G.9960 home networking over power lines. Together with turbo codes, they are the modern coding technique of choice when it comes to designing communication systems that approach the theoretical limits of physical transmission media in terms of data rate, transmission power, geographical reach and reliability.

All LDPC codes in current standards are binary codes. LDPC codes over non-binary alphabets were mentioned in [11] and fully described in [37]. They offer two practical advantages and one major disadvantage with respect to binary codes:

- Advantage 1: Encoding directly over the q-ary alphabet corresponding to the signal constellation used for modulation saves the mapping and de-mapping operations needed to transfer between binary coding alphabet and non-binary modulation signal space. Furthermore, the de-mapping operation is costly in terms of complexity and introduces a loss of sufficient statistic and a resulting performance loss that can only be partially countered by proper choice of the mapping, or fully recovered by costly iterations over the de-mapper and the decoder. With non-binary codes, there is no mapping and no loss of efficiency through de-mapping as the input messages to the decoder are a sufficient statistic for the transmitted symbols, making non-binary LDPC codes a tempting proposition for high spectral-efficiency coding over higher order constellations.
- Advantage 2: Non-binary LDPC codes tend to exhibit less of a performance loss when the block length is shortened to accommodate delay constraints, as compared to binary codes.
- Disadvantage: The decoding complexity of LDPC codes increases with the alphabet size.

The complexity issue has been addressed in a number of refinements of the non-binary LDPC iterative decoding algorithm. The plain description of the decoder requires convolutions of q-ary distribution-valued messages in every constraint node of the associated factor graph. A first and appealing improvement [37] is obtained by switching to the frequency domain where convolutions become multiplications. While the savings by the frequency-domain approach are significant, the resulting complexity is still much higher than that of the equivalent binary decoder. The currently least complex methods known for decoding non-binary LDPC codes are various realizations of the extended min-sum (EMS) [38] algorithm. In this method, convolutions are evaluated directly in the time domain but messages are first truncated to their most significant components, and convolutions are evaluated on the truncated alphabets, resulting in a significant complexity reduction with respect to the q^2 operations needed for a full convolution. Many technical improvements of the EMS can be achieved by hardware-aware implementation of the convolution operations, e.g., [39], [40].

In this section, we discuss two current research areas related to non-binary codes. First, we will look at frequency-domain methods that operate on truncated messages. The aim here is to achieve a fairer comparison of complexity between the EMS and frequency-domain methods, since much of the gain of the EMS is achieved through message truncation, but in complexity comparisons it is evaluated alongside frequency domain decoders operating on full message sets. In the second part of this section, we will look at a novel non-binary code construction operating over rings rather than fields, with a decoding algorithm known as analog digital belief propagation (APBP) [41]. This promising new approach has the merit that its complexity does not increase with the alphabet size, in contrast to regular belief propagation for LDPC codes over q-ary fields, making it an appealing proposition for communications at very high spectral efficiencies.

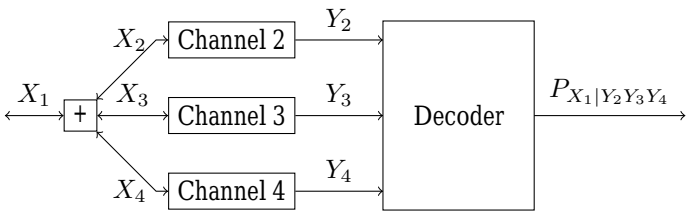

Fig. 6. Conceptual scenario for a degree 4 constraint node decoder

A. Frequency Domain Decoding with Truncated Messages

The ideal constraint node operation of an LDPC decoder operating on a field $\mathcal{F}$ implements a Bayesian estimator for the conceptual scenario illustrated in Fig. 6.

The estimator provides the a-posteriori probability distribution of code symbol X_1 given the observations Y_2, Y_3, and Y_4 of the code symbols X_2, X_3, and X_4, respectively, where the sum of X_1, X_2, X_3, and X_4 is zero over $\mathcal{F}$. Assuming that the input to the decoder is provided in terms of a-posteriori probability distributions $P_{X_2|Y_2}$, $P_{X_3|Y_3}$, and $P_{X_4|Y_4}$, i.e., as distribution-valued messages, it follows that the distribution $P_{X_1|Y_2Y_3Y_4}$ to be computed is a type of convolution of the input distributions. For example, if $\mathcal{F} = \mathrm{GF}(3)$, i.e., the field of numbers $\{0, 1, 2\}$ using arithmetic modulo 3, then the output probability that X_1 be zero given Y_2, Y_3, and Y_4 is the sum of the probabilities all configurations of X_2, X_3, and X_4 that sum to zero, i.e., 0,0,0 or 0,1,2 or 0,2,1 or 1,0,2 or 1,1,1 or 1,2,0 or 2,0,1 or 2,1,0 or 2,2,2. This case results in a cyclic convolution of the three distribution-valued input messages. Over the more commonly used binary extension fields $\mathrm{GF}(2^m)$, where the sum is defined as a bitwise sum, the corresponding operation is a componentwise cyclic convolution in multi-dimensional binary space.

Convolution can be efficiently operated in the frequency domain. For a pure cyclic convolution such as the one illustrated over $\mathrm{GF}(3)$, the transform required is the discrete Fourier transform (DFT). The convolution of vectors in the time domain is equivalent to the componentwise product of the corresponding vectors in the transform domain. This process is illustrated in Fig. 7. For the more practically relevant binary extension fields $\mathrm{GF}(2^m)$, the same process applies but the transform required is the Walsh-Hadamard transform (WHT).

Both the DFT and the WHT can be operated efficiently using

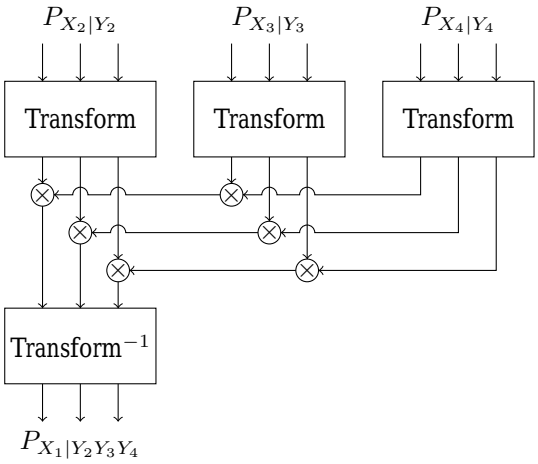

Fig. 7. Frequency domain convolution

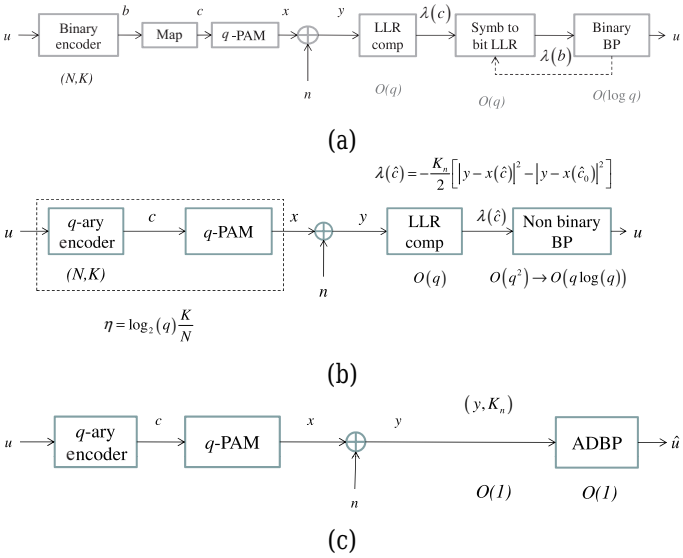

Fig. 8. High spectrally efficient systems using binary codes and: (a) Pragmatic receiver, (b) non-binary codes and non-binary BP, and (c) ADBP.

a fast butterfly structure as the fast Fourier transform (FFT) or the fast Hadamard transform (FHT), requiring $q \log q$ operations where q is the alphabet size of the code. In a typical non-binary LDPC decoder realization, these transforms despite their efficient implementation still use up over 90% of the computing resources and hence constitute the main hurdle for the practical implementability of non-binary LDPC when compared to binary LDPC codes. The approach of the EMS is to revert to time-domain convolutions but operate them on reduced alphabet sizes $q' \ll q$ by truncating each incoming distribution-valued message to its largest components. The resulting algorithm is more difficult to operate than may at first appear, because in such partial convolutions one needs to retain which output values emerge from the mappings of the differing truncated alphabets of each input message, so the implementation needs to perform operations in $\mathcal{F}$ in parallel to the convolution operations over the probabilities. The complexity comparison becomes a comparison between q'^2 and $q \log q$. For example, when operating in $\mathrm{GF}(64)$, the complexity of the frequency-domain based decoder is on the order of $6 \times 64 = 384$ operations per constraint node per iteration, whereas the EMS with messages truncated to $q' = 8$ is in the order of $8 \times 8 = 64$ operations per constraint node per iteration. An added benefit of performing convolutions in the time domain is that one can operate in the logarithmic domain, replacing products by max operations using the well established approach that also underpins the min-sum method for decoding binary LDPC codes.

The comparison described above is not completely fair because it fails to take into account that message truncation may also be of benefit when operating in the frequency domain. Specifically, evaluating a FHT for truncated messages can be made more efficient if we neutralise all operations that apply to the constant message tail corresponding to the truncated portion of the message. In [42], the expected number of operations in a FHT on truncated messages was evaluated both exactly and using an approximation approach that makes it easier to compute for large alphabet sizes. The resulting comparison is promising and shows that much can be gained in operating in the frequency domain on truncated messages. The study however is limited to the direct transform and stops short of treating the more difficult question of how to efficiently evaluate the inverse transform when one is only interested in its q' most significant output values.

B. LDPC Codes over Rings and ADBP

Consider the problem of designing a high-spectral-efficiency transmission system making use of an encoder of rate r_c and a high order q-pulse amplitude modulation (PAM) constellation, yielding a spectral efficiency $\eta = r_c \log_2(q)$ [bits/dimension].

The current state-of the art solution for such systems, adopted in most standards, is the *pragmatic* approach of Fig. 8.(A). A *binary* encoder is paired to a q-PAM modulation using an interleaver and a proper mapping that produces a sequence of constellation points. At the receiver a detector computes binary log-likelihood ratios (LLRs) from symbol LLRs and passes them to the binary iterative decoder through a suitably designed interleaver. The complexity of the LLR computation is linear with q and consequently exponential with the spectral efficiency η for a general mapping without any structure that can be exploited for low-complexity.

The feed-forward receiver scheme described above is associated to a "pragmatic" capacity that is smaller than that of the modulation set and can be maximized using Gray mapping. The feedback structure (dashed line in Fig. 8(a)) can recover this capacity loss if coupled with a proper binary code design. However, iterating between detector and decoder increases the receiver complexity as the conversion from bit to symbol LLRs and vice versa is included in the loop, so that its complexity is multiplied by the number of detector iterations.

A straightforward extension of an (N, K) binary encoder is obtained by substituting the binary quantities at the input of the encoder with q-ary symbols. Parity-check symbols are obtained by performing $\mathrm{mod}\, q$ sums instead of $\mathrm{mod}\, 2$ sums in the encoding procedure. The set of codewords is then defined as follows

$$\mathcal{C} = \{\mathbf{c} \in \mathbb{Z}_q^N : \mathbf{Hc} = \mathbf{0}\},$$

where the matrix elements are constrained to take only value in $\{0, \pm 1\}$. The asymptotic properties of this class of codes were studied in [43] and [44], where they were named "modulo-q" or quantized coset (MQC) codes. Both papers showed that they achieve the random coding exponent and thus are capable of achieving capacity.

The q-ary output symbols c from the encoder can then be directly mapped to q-PAM constellations. At the receiver (Fig. 8(b)) the use of the regular non-binary BP iterative decoding algorithm requires computing the log-likelihood ratios of the transmitted symbols in the form of $(q-1)$-ary vectors. For AWGN the LLRs take the following form

$$\lambda(\hat{c}) = -\frac{K_n}{2}\left[|y - x(\hat{c})|^2 - |y - x(c_0)|^2\right], \quad \forall c \neq c_0$$

where $K_n = 1/\sigma_n^2$ is the *concentration* of the noise.

A straightforward implementation of non-binary BP results in memory and complexity requirements of the order of $O(q)$ and $O(q^2)$, respectively. In order to reduce the complexity of non-binary decoding, several decoding schemes have been proposed in recent years. These were discussed in the previous section and we summarize them again here.

The first straightforward simplification is obtained at check nodes by replacing the discrete convolution of messages, having complexity $O(q^2)$, with the product of the message Fourier transforms. The use of FFT brings down the complexity to $O(q \log q)$. In [45], the authors introduce a log-domain version of this approach that has advantages in terms of numerical stability.

As mentioned before, some further simplifications have been proposed in [38] with the EMS algorithm, where message vectors are reduced in size by keeping only those elements in the alphabet with higher reliability. In [39] and [46], the same authors propose a hardware implementation of the EMS decoding algorithm for non-binary LDPC codes.

In [47], the min-max algorithm is introduced with a reduced complexity architecture called selective implementation, which can reduce by a factor 4 the operations required at the check nodes; however, complexity is still in the order of $O(q^2)$.

Several studies on VLSI implementation of non-binary decoders based on the previous algorithms have been presented in literature [48]–[54]. The results of such studies confirm that all non-binary decoders require complexity growing with the size of the alphabet.

The ADBP algorithm proposed in [41] represents a breakthrough in the reduction of the complexity and memory requirements with respect to previous proposed algorithms, as for ADBP both complexity and memory requirements are *independent* of the size q of the alphabet. The main simplification of ADBP is due to the fact that messages are not stored as vector of size q containing the likelihood of the discrete variables (or equivalently their LLRs) but rather as the two moments, or related quantities, of some suitable predefined class of Gaussian-like distributions. ADBP can be cast into the general class of expectation-propagation algorithms described by Minka [55]. The main contribution in [41] is the definition of the suitable class of distributions for the messages relative to wrapped and discretized variables and the derivation of the updating equations for the message parameters at the sum and repetition operations of the Tanner graph.

A receiver system using the ADBP (Fig. 8(c)), takes then as input messages directly the pair (K, y) of the concentration of the noise and the received samples. This pair identifies a member of the predefined class of Gaussian-like likelihoods and ADBP performs the BP updating by constraining the messages in the graph to stay in the same distribution class.

The exact ADBP updating equations however are not suitable for a straightforward implementation due to the presence of complex non linear operations. Some simplifications to the updating equations have been presented in [56]. In [57], the practical feasibility of the ADBP decoding is established by providing post-synthesis results of a hardware implementation of the required ADBP processing functions.

The ADBP decoder cannot be applied to all types of linear codes over $GF(q)$ as multiplication by field elements different from ± 1 is not allowed in the graph. This constraint has not been taken into consideration previously at the code design stage and requires the construction of new and efficient codes. Although [43] and [44] show that asymptotically this class of codes can achieve capacity, in literature there are no examples of good code constructions with finite size.

The exceptional complexity reduction achieved by using the ADBP, together with the asymptotic results motivates further research effort in the design of good LDPC encoders within this class.

IV. POLAR CODES

Since its inception, the major challenge in coding theory has been to find methods that would achieve Shannon limits using low-complexity methods for code construction, encoding, and decoding. A solution to this problem has been proposed in [14] through a method called "channel polarization". Rather than attacking the coding problem directly, the polarization approach follows a purely information-theoretic route whereby N independent identical copies of a given binary-input channel W are manipulated by certain combining and splitting operations to "manufacture" a second set of binary-input channels $\{W^{(i)}\}_{i=1}^{N}$ that have capacities either near 0 or near 1, except for a fraction that vanishes as N becomes large. Once such polarized channels are obtained, "polar coding" consists of transmitting information at full rate over $W^{(i)}$ that are near perfect and fixing the inputs of the remaining channels, say, to zero. In [14], it was shown that polar codes contructed in this manner could achieve capacity with encoding and decoding methods of complexity $O(N \log N)$. In subsequent work [58], it was shown that the probability of frame error for polar codes goes to zero roughly as $e^{-\sqrt{N}}$ for any fixed rate below capacity; this result was later refined by [59] who determined the explicit form of the dependence of the exponent on the code rate.

The basic binary polar code is a linear code defined for any block length $N = 2^n$ in terms of a generator matrix

$$\mathbf{G_N} = \mathbf{F}^{\otimes \mathbf{n}}, \quad \mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{1}$$

where $\mathbf{F}^{\otimes \mathbf{m}}$ denotes the nth Kronecker power of F. In polar coding one encodes a data word $\mathbf{u} = (\mathbf{u_1}, \cdots, \mathbf{u_N})$ into a codeword $\mathbf{x} = (\mathbf{x_1}, \cdots, \mathbf{x_N})$ through the transformation $\mathbf{x} = \mathbf{uG_N}$. For a rate K/N polar code, one fixes $N - K$ of the coordinates of $\mathbf{u}$ to zero, effectively reducing $\mathbf{G_N}$ to a $K \times N$ matrix. For example, for a $(N, K) = (8, 4)$ polar code, one may fix u_1, u_2, u_3, u_5 to zero and obtain from

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

the 4×8 generator matrix

$$\mathbf{G_{4,8}} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The polar code design problem consists in determining which set of $(N - K)$ coordinates to freeze so as to achieve the best possible performance under successive-cancellation (SC) decoding on a given channel. It turns out that the solution to this problem depends on the channel at hand, so in general there is no universal set of coordinates that are guaranteed to work well for all channels of a given capacity. In [14], only a heuristic method was given for the polar code design problem. The papers [60]–[62] provided a full solution with complexity $O(N)$. With this development, polar codes became the first provably capacity-achieving class of codes with polynomial-time algorithms for code construction, encoding, and decoding.

Other important early theoretical contributions came in rapid succession from [63]–[67]. Polar coding was extended to non-binary alphabets in [68]–[71]. Polar code designs by using alternative generator matrices with the goal of improving the code performance were studied in [72]–[76].

As stated above, polar coding is a channel dependent design. The performance of a polar code under "channel mismatch" (i.e., using a polar code optimized for one channel on a different one) has been studied by [77], who showed that there would be a rate loss. As shown in [78], the non-universality of polar codes is a property of the suboptimal low-complexity successive cancellation decoding algorithm; under ML decoding, polar codes are universal. More precisely, [78] shows that a polar code optimized for a BSC achieves the capacity of any other binary-input channel of the same capacity under ML decoding. This result is very interesting theoretically since it gives a constructive universal code for all binary-input channels; however, it does this at the expense of giving up the $O(N \log N)$ decoding algorithm. In more recent work [79] and [80], universal polar coding schemes have been described, which come at the expense of lengthening the regular polar code construction.

It was recognized from the beginning that the finite length performance of polar codes was not competitive with the state-of-the-art. This was in part due to the suboptimal nature of the standard SC decoding algorithm, and in part due to the relatively weak minimum distance properties of these codes. Another negative point was that the SC decoder made its decisions sequentially, which meant that the decoder latency would grow at least linearly with the code length, which resulted in a throughput bottleneck. Despite these shortcomings, interest in polar codes for potential applications continued. The reason for this continued interest may be attributed to several factors. First, polar codes are firmly rooted in sound well-understood theoretical principles. Second, while the performance of the basic polar code is not competitive with the state-of-the-art at short to practical block length, they are still good enough to maintain hope that with enhancements they can become a viable alternative. This is not surprising given that polar codes are close cousins of Reed-Muller (RM) codes, which are still an important family of codes [81] in many respects, including performance. Third, polar codes have the unique property that their code rate can be adjusted from 0 to 1 without changing the encoder and decoder. Fourth, polar codes have a recursive structure, based on Plotkin's $|u|u + v|$ construction [82], which makes them highly suitable for implementation in hardware. For these and other reasons, there have been a great number of proposals in the last few years to improve the performance of polar codes while retaining their attractive properties. The proposed methods may be classified essentially into two categories as encoder-side and decoder-side techniques.

Among the encoder-side techniques, one may count the non-binary polar codes and binary polar codes starting with a larger base matrix (kernel); however, these techniques have not yet attracted much attention from a practical viewpoint due to their complexity. Other encoder side techniques that have been tried include the usual concatenation schemes with Reed-Solomon codes [83], and other concatenation schemes [84]–[86].

Two decoder-side techniques that have been tried early on to improve polar code performance are BP decoding [87] and trellis-based ML decoding [88]. The BP decoder did not improve the SC decoder performance by any significant amount; however, it continues to be of interest since the BP decoder has the potential to achieve higher throughputs compared to SC decoding [89].

The most notable improvement in polar coding performance came by using a list decoder [90] with cyclic redundancy check (CRC), which achieved near ML performance with complexity roughly $O(LN \log N)$ for a list size L and code length N. The CRC helps in two ways. First, it increases the code minimum distance at relatively small cost in terms of coding efficiency, thus improving code performance especially at high SNR. Second, the CRC helps select the correct codeword from the set of candidate codewords offered by the list decoder. It should be mentioned that the above list decoding algorithm for polar codes was an adaptation of an earlier similar algorithm given in [91] in the context of RM codes. The vast literature on RM codes continues to be a rich source of ideas in terms of design of efficient decoding techniques for polar codes. A survey of RM codes from the perspective of decoders for polar codes has been given in [92].

We end this survey by giving a performance result for polar codes. Fig. 9 compares the performance of a $(2048, 1008)$ polar

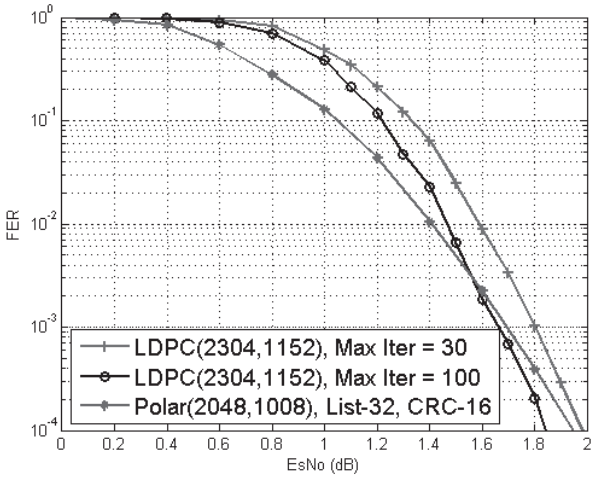

Fig. 9. Performance comparison of polar and LDPC codes.

code with the WiMAX (2304,1152) LDPC code. The polar code is obtained from a $(2048, 1024)$ code by inserting a 16-bit CRC into the data and is decoded by a list-of-32 decoder. The LDPC code results are from the database provided by [93]; decoding is by belief propagation with the maximum number of iterations limited to 30 and 100 in the results presented. The realization that polar coding performance can rival the state-of-the-art has spurred intense research for practical implementations of these codes. We omit from this survey the implementation-oriented papers since that is already a very large topic by itself. Whether polar codes will ever appear as part of the portfolio of solutions in future systems remains uncertain. The state-of-the-art in error correction coding is mature, with a firm footprint by turbo and LDPC codes. Whether polar codes offer significant advantages to make room for themselves in practical applications depends in large part on further innovation on the subject.

V. CONCLUSION

We have presented three areas of active research in coding theory. We discussed spatially coupled LDPC codes for which the asymptotic performance of the iterative decoder is improved to that of the optimal decoder. We have discussed non-binary LDPC codes and have introduced a new decoding algorithm, ADBP, whose complexity does not increase with the alphabet size. Finally, we have described polar coding, a novel code construction based on a phenomenon coined channel polarization, which can be proved theoretically to achieve channel capacity. We have stated a number of open problems, among them:

- When decoding non-binary LDPC codes in the frequency domain, can we design a reduced complexity inverse transform if we are only interested in the larger components of the resulting distribution-valued message?
- How do we design LDPC codes over rings of integers to optimize the performance of the ADBP decoder?
- While the potential of polar codes is already well-established, how can we improve the performance of these codes either by encoder or decoder side techniques while retaining their low-complexity properties? Are there decoding techniques that approximate ML performance with complexity significantly lower than the existing list-decoding techniques?

We hope to have shown in this paper that coding theory is an active area of research with many challenges remaining and a number of promising innovations on their way to maturing into technological advances in the coming years.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948.
[2] M. J. E. Golay, "Notes on digital coding," *Proc. IRE*, vol. 37, p. 657, 1949.
[3] R. W. Hamming, "Error detecing and error correcting codes," *Bell Syst. Tech. J.*, vol. 29, pp. 147–160, 1950.
[4] I. Reed and G. Solomon, "Polynomial codes over certain finite fields," *J. Soc. Indust. Appl. Math.*, vol. 8, June 1960.
[5] P. Elias, "Coding for noisy channels," *IRE Convention Record, Part 4*, pp. 37–46, 1955.
[6] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Trans. Inf. Theory*, vol. 13, pp. 260–269, Apr. 1960.
[7] G. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 258–278, Mar. 1973.
[8] L. Bahl *et al.*, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, pp. 284–287, Mar. 1974.
[9] G. Ungerboeck, "Channel coding with multilevel/phase signals," *IEEE Trans. Inf. Theory*, pp. 55–67, Jan. 1982.
[10] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limimt error correcting coding and decoding: Turbo-codes," in *Proc. IEEE Int. Conf. Commun. (ICC)*, (Geneva, Switzerland), vol.2, May 1993, pp. 1064–1070.
[11] R. G. Gallager, *Low Density Parity Check Codes*. Ph.D. thesis, Massachussetts Institute of Technology (MIT), Cambridge, Mass., 1963.
[12] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inf. Theory*, vol. 45, pp. 399–431, Mar. 1999.
[13] J. L. Massey, "Coding and modulation in digital communications," in *Proc. Int. Zurich Seminar on Communications (IZS)*, (Zurich, Switzerland), 1974.
[14] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, pp. 3051–3073, July 2009.
[15] T. Richardson, M. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, pp. 619–637, Feb. 2001.
[16] A. Jimenez Felstrom and K. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inf. Theory*, vol. 45, pp. 2181 –2191, Sept. 1999.
[17] J. Thorpe, "Low-density parity-check (LDPC) codes constructed from protographs," IPN Progress Rep. 42-154, JPL, Aug. 2003.
[18] M. Lentmaier *et al.*, "Approaching capacity with asymptotically regular LDPC codes," in *Proc. Inf. Theory and Applications Workshop*, (San Diego, CA, USA), Feb. 2009, pp. 173 –177.
[19] M. Lentmaier *et al.*, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Trans. Inf. Theory*, vol. 56, pp. 5274 –5289, Oct. 2010.
[20] S. Kudekar, T. Richardson, and R. Urbanke, "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC," *IEEE Trans. Inf. Theory*, vol. 57, pp. 803–834, Feb. 2011.
[21] S. Kudekar, T. Richardson, and R. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Cambridge, MA, USA), July 2012, pp. 453–457.
[22] S. Kumar *et al.*, "A proof of threshold saturation for spatially-coupled LDPC codes on BMS channels," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Oct. 2012, pp. 176–184.
[23] C. Measson and R. Urbanke, "Maximum a posteriori decoding and turbo codes for general memoryless channels," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Adelaide, Australia), Sept. 2005, pp. 1241–1245.
[24] D. Mitchell, M. Lentmaier, and D. Costello, "New families of LDPC block codes formed by terminating irregular protograph-based LDPC convolutional codes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Austin, Texas, USA), June 2010, pp. 824–828.
[25] W. Nitzold, G. Fettweis, and M. Lentmaier, "Spatially-coupled nearly-regular LDPC code ensembles for rate-flexible code design," in *Proc. IEEE Int. Conf. Commun. (ICC)*, (Sydney, Australia), June 2014, pp. 2027–2032.

[26] A. R. Iyengar *et al.*, "Windowed decoding of protograph-based LDPC convolutional codes over erasure channels," *IEEE Trans. Inf. Theory*, vol. 58, pp. 2303–2320, Apr. 2012.

[27] N. Ul Hassan, M. Lentmaier, and G. Fettweis, "Comparison of LDPC block and LDPC convolutional codes based on their decoding latency," in *Proc. Int. Symp. on Turbo Codes & Iterative Inf. Proc.*, (Gothenburg, Sweden), Aug. 2012.

[28] T. Hehn and J. Huber, "LDPC codes and convolutional codes with equal structural delay: a comparison," *IEEE Trans. Commun.*, vol. 57, pp. 1683–1692, June 2009.

[29] S. Maiya *et al.*, "Coding with a latency constraint: The benefits of sequential decoding," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Oct. 2010, pp. 201–207.

[30] E. Sharon, N. Presman, and S. Litsyn, "Convergence analysis of generalized serial message-passing schedules," *IEEE J. Sel. Areas Commun.*, vol. 27, pp. 1013–1024, Aug. 2009.

[31] N. Ul Hassan *et al.*, "Reduced complexity window decoding schedules for coupled LDPC codes," in *Proc. IEEE Inform. Theory Workshop (ITW)*, (Lausanne, Switzerland), Sept. 2012, pp. 20–24.

[32] N. Ul Hassan *et al.*, "Non-uniform windowed decoding schedules for spatially coupled codes," in *Proc. IEEE Glob. Comm. Conf. (GLOBECOM)*, (Atlanta, GA, USA), Dec. 2013, pp. 1862–1867.

[33] L. Ozarow, S. Shamai (Shitz), and A. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359–378, 1994.

[34] E. Biglieri, G. Caire, and G. Taricco, "Coding for the fading channel: a survey," *Signal Processing (EURASIP)*, vol. 80, no. 7, pp. 1135–1148, 2000.

[35] J. Boutros *et al.*, "Low-density parity-check codes for nonergodic block-fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4286–4300, 2010.

[36] N. Ul Hassan *et al.*, "Improving code diversity on block-fading channels by spatial coupling," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Honolulu, Hawaii, USA), pp. 2311–2315, June 2014.

[37] M. C. Davey and D. J. C. MacKay, "Low density parity check codes over GF(q)," in *Proc. IEEE Inform. Theory Workshop (ITW)*, (Killarney, Ireland), June 1998, pp. 70–71.

[38] D. Declercq and M. P. Fossorier, "Decoding algorithms for nonbinary LDPC codes over GF(q)," *IEEE Trans. Commun.*, vol. 55, pp. 633–643, Apr. 2007.

[39] A. Voicila *et al.*, "Low-complexity decoding for non-binary LDPC codes in high order fields," *IEEE Trans. Commun.*, pp. 1365–1375, May 2010.

[40] E. Boutillon and L. Conde-Canencia, "Bubble check: a simplified algorithm for elementary check node processing in extended min-sum non-binary LDPC decoders," *IET Electronic Lett.*, vol. 46, pp. 633–634, Apr. 2010.

[41] G. Montorsi, "Analog digital belief propagation," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1106–1109, 2012.

[42] J. Sayir, "Non-binary LDPC decoding using truncated messages in the walsh-hadamard domain," in *Proc. Int. Symp. Inf. Theory and Its App. (ISITA)*, (Melbourne, Australia), Oct. 2014.

[43] A. Bennatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 50, pp. 417–438, Mar. 2004.

[44] U. Erez and G. Miller, "The ML decoding performance of LDPC ensembles over Z_q," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1871–1879, 2005.

[45] H. Wymeersch, H. Steendam, and M. Moeneclaey, "Log-domain decoding of LDPC codes over GF(q)," in *Proc. IEEE Int. Conf. Commun. (ICC)*, vol. 2, (Paris, France), vol. 2, 2004, pp. 772–776.

[46] A. Voicila *et al.*, "Architecture of a low-complexity non-binary LDPC decoder for high order fields," in *Communications and Information Technologies, 2007. ISCIT '07. International Symposium on*, 2007, pp. 1201–1206.

[47] V. Savin, "Min-Max decoding for non binary LDPC codes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Toronto, Canada), 2008, pp. 960–964.

[48] C. Spagnol, E. Popovici, and W. Marnane, "Hardware Implementation of GF(q) LDPC Decoders," *IEEE Trans. Circuits Syst. I*, vol. 56, no. 12, pp. 2609–2620, 2009.

[49] J. Lin *et al.*, "An Efficient VLSI architecture for nonbinary LDPC decoders," *IEEE Trans. Circuits Syst. II*, vol. 57, no. 1, pp. 51–55, 2010.

[50] C. Zhang and K. Parhi, "A network-efficient nonbinary QC-LDPC decoder architecture," *IEEE Trans. Circuits Syst. I*, vol. 59, no. 6, pp. 1359–1371, 2012.

[51] X. Chen, S. Lin, and V. Akella, "Efficient configurable decoder architecture for nonbinary quasi-cyclic LDPC codes," *IEEE Trans. Circuits Syst. I*, vol. 59, no. 1, pp. 188–197, 2012.

[52] X. Chen and C.-L. Wang, "High-throughput efficient non-binary LDPC decoder based on the simplified min-sum algorithm," *IEEE Trans. Circuits Syst. I*, vol. 59, no. 11, pp. 2784–2794, 2012.

[53] Y.-L. Ueng *et al.*, "An efficient layered decoding architecture for nonbinary QC-LDPC codes," *IEEE Trans. Circuits Syst. I*, vol. 59, no. 2, pp. 385–398, 2012.

[54] Y.-L. Ueng *et al.*, "A high-throughput trellis-based layered decoding architecture for non-binary LDPC codes using Max-Log-QSPA," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2940–2951, 2013.

[55] T. P. Minka, "Expectation propagation for approximate bayesian inference," in *Proc. AUAI*, Morgan Kaufmann Publishers Inc., 2001, pp. 362–369.

[56] G. Montorsi, "Analog digital belief propagation: from theory to practice," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2012, pp. 2591–2595.

[57] M. Awais Aslam *et al.*, "VLSI implementation of a non-binary decoder based on the analog digital belief propagation," *IEEE Trans. Signal Process.*, pp. 3965–3975, July 2014.

[58] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Seoul, South Korea), June 28–July 3, 2009, pp. 1493–1495.

[59] S. H. Hassani *et al.*, "Rate-dependent analysis of the asymptotic behavior of channel polarization," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2267–2276, 2013.

[60] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Commun. Lett.*, vol. 13, pp. 519–521, July 2009.

[61] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, pp. 6562–6582, Oct. 2013.

[62] R. Pedarsani *et al.*, "On the construction of polar codes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (St. Petersburg, Russia), 2011, pp. 11–15.

[63] N. Hussami, S. B. Korada, and R. Urbanke, "Performance of polar codes for channel and source coding," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Seoul, South Korea), 2009, pp. 1488–1492.

[64] S. B. Korada and R. Urbanke, "Polar codes for slepian-wolf, wyner-ziv, and gelfand-pinsker," in *Proc. IEEE Inform. Theory Workshop (ITW)*, (Cairo, Egypt), 2010.

[65] E. Sasoglu, E. Telatar, and E. Yeh, "Polar codes for the two-user binary-input multiple-access channel," in *Proc. IEEE Inform. Theory Workshop (ITW)*, (Cairo, Egypt), 2010, pp. 1–5.

[66] S. B. Korada *et al.*, "An empirical scaling law for polar codes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Austin, Texas, USA), 2010, pp. 884–888.

[67] E. Abbe and I. Telatar, "Mac polar codes and matroids," in *Proc. Inf. Theory and Applications Workshop*, (San Diego, CA, USA), 2010, pp. 1–8.

[68] E. Sasoglu, I. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Proc. IEEE Inform. Theory Workshop (ITW)*, (Taormina, Italy), 2009, pp. 144–148.

[69] M. Karzand and I. Telatar, "Polar codes for q-ary source coding," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Austin, TX, USA), 2010, pp. 909–912.

[70] W. Park and A. Barg, "Polar codes for q-ary channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 955–969, 2013.

[71] A. G. Sahebi and S. S. Pradhan, "Multilevel polarization of polar codes over arbitrary discrete memoryless channels," in *Proc. Allerton Conf. on Communications, Control, and Computing*, 2011, pp. 1718–1725.

[72] S. B. Korada, E. Sasoglu, and R. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6253–6264, 2010.

[73] R. Mori and T. Tanaka, "Channel polarization on q-ary discrete memoryless channels by arbitrary kernels," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Austin, TX, USA), 2010, pp. 894–898.

[74] R. Mori and T. Tanaka, "Source and channel polarization over finite fields and reed-solomon matrix," *arXiv preprint arXiv:1211.5264*, 2012.

[75] N. Presman, O. Shapira, and S. Litsyn, "Binary polar code kernels from code decompositions," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (St. Petersburg, Russia), 2011, pp. 179–183.

[76] N. Presman, O. Shapira, and S. Litsyn, "Polar codes with mixed kernels," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (St. Petersburg, Russia), pp. 6–10, IEEE, 2011.

[77] S. H. Hassani, S. B. Korada, and R. Urbanke, "The compound capacity of polar codes," *arXiv preprint arXiv:0907.3291*, 2009.

[78] E. Sasoglu, *Polar coding theorems for discrete systems*. Ph.D. thesis, EPFL, 2011.

[79] E. Sasoglu and L. Wang, "Universal polarization," *arXiv preprint arXiv:1307.7495*, 2013.

[80] S. H. Hassani and R. Urbanke, "Universal polar codes," *arXiv preprint arXiv:1307.7223*, 2013.

[81] D. J. Costello and G. D. Forney Jr, "Channel coding: The road to channel capacity," *Proc. IEEE*, vol. 95, no. 6, pp. 1150–1177, 2007.
[82] M. Plotkin, "Binary codes with specified minimum distance," *IRE Trans. Inform. Theory*, vol. 6, pp. 445–450, Sept. 1960.
[83] M. Bakshi, S. Jaggi, and M. Effros, "Concatenated polar codes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (Austin, TX, USA), 2010, pp. 918–922.
[84] E. Arıkan and G. Markarian, "Two-Dimensional Polar Coding," in *Proc. Int. Symp. on Commun. Theory and App. (ISCTA)*, (Ambleside, U.K.), July 2009.
[85] H. Mahdavifar *et al.*, "Performance limits and practical decoding of interleaved reed-solomon polar concatenated codes," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1406–1417, 2014.
[86] P. Trifonov and P. Semenov, "Generalized concatenated codes based on polar codes," in *Proc. Int. Symp. on Wireless Comm. Systems (ISWCS)*, (Aachen, Germany), 2011, pp. 442–446.
[87] E. Arıkan, "A performance comparison of polar codes and Reed-Muller codes," *IEEE Commun. Lett.*, vol. 12, pp. 447–449, June 2008.
[88] E. Arıkan *et al.*, "Performance of short polar codes under ML decoding," in *Proc. ICT MobileSummit*, (Santander, Spain), June 2009.
[89] Y. S. Park *et al.*, "A 4.68gb/s belief propagation polar decoder with bit-splitting register file," in *VLSI Circuits Digest of Technical Papers, 2014 Symposium on*, June 2014, pp. 1–2.
[90] I. Tal and A. Vardy, "List decoding of polar codes," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, (St. Petersburg, Russia), 2011, pp. 1–5.
[91] I. Dumer and K. Shabunov, "Soft-decision decoding of reed-muller codes: recursive lists," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1260–1266, 2006.
[92] E. Arıkan, "A survey of reed-muller codes from polar coding perspective," in *Proc. IEEE Inform. Theory Workshop (ITW)*, (Cairo, Egypt), Jan 2010, pp. 1–5.
[93] *The Coded Modulation Library* [Online]. Available: http://www.iterativesolutions.com/Matlab.htm

Erdal Arıkan was born in Ankara, Turkey in 1958. He received the B.S. degree from the California Institute of Technology, Pasadena in 1981 and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, MA in 1982 and 1985, respectively, all in electrical engineering. Since 1987, he has been with the Electrical-Electronics Engineering Department of Bilkent University, Ankara, Turkey, where he is presently a Professor. His research Interests include coding techniques, coding theory, and Shannon theory. He has received many awards including hte IEEE information theory society paper award (2010) and the IEEE W.R.G. Baker award (2013) for his paper "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels".

Najeeb ul Hassan received the B.S. degree from NUCES Lahore, Pakistan in 2007 and M.Sc. in Digital Communications from Christian-Albrechts-University of Kiel, Germany in 2010. For his Master's studies, he received an award for a best Master's Thesis in the Faculty and also a DAAD-Prize for the outstanding achievement as a foreign student. Currently, he is a Ph.D. student in Vodafone Chair for Mobile Communications at the Technical University of Dresden Germany since 2011. His main research interests include forward error correction codes with the focus on design and analysis of the iterative decoding algorithms.

Michael Lentmaier received the Dipl.-Ing. degree in Electrical Engineering from University of Ulm, Germany in 1998, and the Ph.D. degree in Telecommunication Theory from Lund University, Sweden in 2003. He then worked as a Post-Doctoral Research Associate at University of Notre Dame, Indiana and at University of Ulm. From 2005 to 2007 he was with the Institute of Communications and Navigation of the German Aerospace Center (DLR) in Oberpfaffenhofen, where he worked on signal processing techniques in satellite navigation receivers. From 2008 to 2012, he was a Senior Researcher and Lecturer at the Vodafone Chair Mobile Communications Systems at TU Dresden, where he was heading the Algorithms and Coding research group. Since January 2013 he is an Associate Professor at the Department of Electrical and Information Technology at Lund University. His research interests include design and analysis of coding systems, graph based iterative algorithms and Bayesian methods applied to decoding, detection and estimation in communication systems. He is a Senior Member of the IEEE and served as an Editor for IEEE Communications Letters from 2010 to 2013 and IEEE Transactions on Communications since 2014. He was awarded the Communications Society & Information Theory Society joint paper award (2012) for his paper "Iterative decoding threshold analysis for LDPC convolutional codes".

Guido Montorsi is a Full Professor at the Politecnico di Torino, Italy, where he also received the Ph.D. degree in Telecommunications from the Dipartimento di Elettronica. His master thesis concerned the study and design of coding schemes for HDTV, developed at the RAI Research Center, Turin. Professor Montorsi has been at the Department of Electrical Engineering at the Rensselaer Polytechnic Institute, Troy, NY, USA, and later at Sequoia Communications, CA, USA, working with the innovative design and implementation of a 3rd generation WCDMA receiver. His interests are in the area of channel coding and wireless communications, particularly on the analysis and design of concatenated coding schemes and study of iterative decoding strategies. He is author of more than one hundred papers published in international journal and conference proceedings.

Jossy Sayir received his doctorate (Dr. Sc. Techn.) from ETH Zurich in 1999. From 1991 to 1993, he worked as a Development Engineer for Motorola Communications in Tel Aviv, Israel. From 1993 to 1999, he worked as a Research Assistant under the supervision of Prof. James L. Massey while writing his dissertation "On coding by probability transformation". From 2000 until 2009, he was with the Telecommunications Research Centre (ftw.) in Vienna, Austria, as a Senior Researcher and Project Manager in charge of part of the centre's strategic research activities. Since 2009, he has been with the Department of Engineering of the University of Cambridge, first as a Intra-European Marie Curie fellow and then as a fixed-term and affiliated lecturer. He is a Fellow of Robinson College and Director of Studies in engineering. He has served on the organisation committees of several international conferences, most recently as TPC Co-Chair of ISIT 2013 and will serve as General Co-Chair of ITW 2016. His research interests focus around low complexity iterative decoding algorithms.