Multiple Access Demodulation in the Lifted Signal Graph With Spatial Coupling

Christian Schlegel and Dmitri Truhachev

***Abstract*—Demodulation in a random multiple access channel is considered where the signals are chosen uniformly randomly with unit energy. It is shown that by lifting (replicating) the graph of this system and randomizing the graph connections, a simple iterative cancellation demodulator achieves the same performance as an optimal symbol-by-symbol detector of the original system. The iterative detector has a complexity that is linear in the number of users, while the direct optimal approach is known to be NP-hard. However, the maximal system load of this lifted graph is limited to $\alpha < 2.07$, even for large signal-to-noise ratios (SNRs)—the system is interference limited. Spatial coupling between subsequent lifted graphs is introduced, and anchoring the initial graphs, the achievable system load α can go to infinity as the SNR goes to infinity. Our results apply to several well-documented system proposals, such as interleave-division multiple access, partitioned spreading, and certain forms of multiple-input multiple-output communications.**

***Index Terms*—Iterative decoding, optimal joint detection, random signaling, spatial coupling.**

I. Introduction

ITERATIVE graph-based signal processing has enjoyed a tremendous rise in popularity with the introduction of turbo coding [2]. By breaking complex algorithms into local processes which communicate in iterative cycles, many highly complex problems can now be addressed with relative ease, and the iterative process reduces complexity in a way similar to how iterative solution methods break down the problem of solving large systems of equations into a repetition of operations with much lower complexity [26].

The analysis of the performance of iterative processors, on the other hand, has offered more resistance, and relatively little is known about the exact performance of these algorithms. The methods introduced by ten Brink [36], Divsalar *et al.* [6], and Richardson and Urbanke [23] are statistical methods to predict the average dynamical behavior of large iterative error control decoders. These analysis methods have much in common with Gallager's original probability propagation approach used to study low-density parity-check codes [9]. The same type of analysis has been utilized by others [1], [3], [30] to study the behavior of iterative decoders for multiple access systems, which were introduced shortly after the invention of the turbo decoding algorithm [1], [19], [21]. All of these methods can accurately predict the convergence thresholds of iterative processors for large-scale systems; however, they cannot predict exact performance in terms of error rates.

The question how close iterative processors, in particular decoders and demodulators, can approach the performance of an optimal decoder is also not well understood. While it appears that the turbo decoding algorithm has a performance very close to that of a maximum-likelihood processor, it is also known that it is in general not equivalent to maximum likelihood [18].

In this paper, we present a case where an augmented graph-based iterative processor achieves the optimal maximum-likelihood performance of the underlying original system in terms of error rates (and LLR statistics). The system in question is a multiple access system where the waveforms used for modulation are uniformly randomly selected from the $(N-1)$-dimensional unit(-energy) sphere in N-dimension. This model has practical counterparts in random code-division multiple access (CDMA) as explored in partitioned CDMA [27], [28], [38], in interleave-division multiple access (IDMA) [20], or in isotropic multiple antenna communications channel models.

The performance of a maximum-likelihood detector for random CDMA was calculated in [33] where a statistical mechanics approach was used. In this paper, we show that a proper augmentation of the factor graph of the random-code model leads to a system in which a basic message-passing algorithm can be applied which achieves the same performance as the optimal detector in the original system as computed in [33].

From a performance standpoint, the iterative system is ultimately interference limited. If we define the *system load* α as the number of transmitted signals K to the number of available signaling dimensions N, i.e., $\alpha = K/N$, then, in an equal received power situation, the maximum tolerated α satisfies $\alpha < 2.07$. We then show that this interference limitation can be overcome by a process called *spatial coupling*, whereby successive frames of the original system are interconnected, and the first few frames are initiated with known symbols which act as anchor or pilot symbols. We show that with appropriate such anchoring, the interference limitation disappears and the supportable load α can go to ∞, given that the signal-to-noise ratio (SNR) also goes to infinity.

Manuscript received October 19, 2010; revised March 16, 2011; accepted November 29, 2012. Date of publication December 11, 2012; date of current version March 13, 2013.This work was supported in part by iCORE Alberta and NSERC Canada. This paper was presented in part at the 2011 IEEE International Symposium on Information Theory.

C. Schlegel is with the Department of Electrical and Computer Engineering, Dalhousie University, Halifax, NS B3K 6R8 Canada (e-mail: christian.b.schlegel@gmail.com).

D. Truhachev is with the Department of Computing Science, University of Alberta, Edmonton, AB T6G 2R3 Canada (e-mail: dmitryt@ualberta.ca).

Communicated by P. O. Vontobel, Associate Editor for Coding Techniques.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2012.2232965

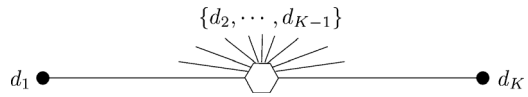


Fig. 1. Factor graph of the basic multiple-access system of (1).

II. System Model

We consider a general communication system with K data streams. Without essential loss of generality, the data consist of independent binary symbols $d_k \in \{-1, 1\}$, $k = 1, \ldots, K$, which come from the outputs of K parallel forward error control encoders. The binary symbols are then modulated onto random unit-energy waveforms $\mathbf{a}_k$ chosen uniformly from the (N–1)-dimensional unit sphere in N dimensions—see Appendix II for the signal description.

The composite signal after transmission at the receiver is given by

$$\mathbf{y} = \sum_{k=1}^{K} d_k \mathbf{a}_k + \sigma \mathbf{n} \tag{1}$$

where $\mathbf{n}$ is an N-dimensional vector of i.i.d. standard Gaussian noise samples and σ is the noise standard deviation. As shown in Appendix II, the average cross correlation of these waveforms is $E[\mathbf{a}_k^T \mathbf{a}_j] = 1/N, k \neq j$, which is all we need.[1]

In the sequel, we will use graph-based arguments and we, therefore, introduce the "factor graph" [11], [14] of (1), shown in Fig. 1, where the center node symbolizes the addition of K signals.

A decoder, or demodulator, can utilize this graph as an algorithmic blueprint. In this case, the graph is not very interesting, and consists of a single, cycle-free star graph. An (optimal) demodulator will have to compute

$$\hat{\mathbf{d}} = \arg\min_{\mathbf{d}} \left\| \mathbf{y} - \sum_{k=1}^{K} d_k \mathbf{a}_k \right\|^2 \tag{2}$$

which involves the computation of 2^K terms. In any case, it is known that (2) is NP-hard [41], and there is little hope that efficient algorithms for its exact solution exist unless $\mathrm{P} = \mathrm{NP}$, which most experts agree is not likely the case [8].

Even though the general problem is NP-hard, special instances of the multiple access problem (2) have efficient, low-complexity solution algorithms. For example, if $\mathbf{a}_k^T \mathbf{a}_l = 0, l \neq k$, which is usually done by design, see, e.g., [40], the demodulation complexity is of order $O(NK)$ only. If the cross correlations between the different signals take on only nonpositive values, optimal detection corresponds to a solution of a min-cut problem of the associated graph, whose complexity is $O(NK^3)$[25], [39]. Sequences $\mathbf{a}_k$ can also be constructed specifically such that lower complexity optimal detection is possible, even when the system is oversaturated, i.e., $\alpha = K/N > 1$ [16], [29]. In [29], the authors introduce controlled additional sequences such that the corresponding oversaturated system can be demodulated with a trellis decoder, whose complexity depends on α. All these approaches, of course, are only feasible if timing of the different signals can be tightly controlled and no transmission distortions occur.

Like the optimal sequence detector, the symbol-wise marginal-posterior-mode (MPM) detector, which maximizes the symbol-wise posteriori probability, is NP-hard in general. The probability of error for the MPM detector was computed in [33] using statistical mechanics tools. The ratio of signal to noise plus residual interference power, the signal-to-interference ratio γ^2 of the posterior probabilities of d_k, is computed as the solution to the implicit equation ([33, eqs. (45) and (43)])

$$\gamma^2 = \left[\sigma^2 + \alpha \mathrm{E} \left[1 - \tanh \left(\frac{F}{E} \left(\gamma^2 + \gamma \xi \right) \right) \right]^2 \right]^{-1} \tag{3}$$

where $\xi \sim \mathcal{N}(0, 1)$, and E and F are given in [33, eq. (43)], and are the mean and variance of the (individual) APP output signal. In general, $E > F$; however, in the case of binary random signaling as used in this paper $E = F$, which follows from the definitions of m, q, E, F in [33, eq. (27) ff.]. Result (3) holds for loads $\alpha < \alpha_{\mathrm{s}} \approx 1.49$, where α_{s} is the "spinodal value," where the statistical mechanics approach undergoes a phase transition and multiple solutions appear.

In this paper we show that by a process called graph lifting and randomization of connections, we can obtain an instance of the decoding problem where an iterative message-passing algorithm can achieve a posterior LLR signal-to-interference ratio that is identical to (3), and, therefore, achieves the same error performance as the MPM detector, but with a decoding complexity of $O(IN)$ per bit d_k, where I is the number of iterations. The number of iterations I does not depend on K, and in most cases, especially those of practical interest, $O(IN) \leq O(N2^K/K)$.

This result represents an extension of both [17] and [28]. In [17], the authors show that a belief propagation algorithm on the *dense* CDMA interference graph can generate the conditional expectation of the input symbols, that is, it can produce the minimum-mean square error (MMSE) output under certain assumptions, such as assuming that the input symbols have a Gaussian distribution. In [28], on the other hand, we showed that a sparse graph of the type discussed in this paper with binary inputs generates outputs whose posterior SNR is strictly lower-bounded by that of an MMSE processor for the corresponding dense system. In this paper, we show, in fact, that these outputs achieve the same performance as that of an optimal detector for the dense system.

Further related work is discussed in [12], where the authors substantiate the random CDMA capacity results with statistical mechanics methods by showing that those results coincide with an upper bound on the capacity of random CDMA, and further show that the upper bound is tight and exactly reproduces the capacity formulas for Gaussian-distributed input signals. In [13], the author shows that the computation of the free energy in random factor graphs can be carried out by a message-passing algorithm.

These recent advances linking statistical mechanics-based approaches to direct performance results, while not directly

[1]Note that the waveforms can, of course, be complex as is the case in typical wireless communications systems.

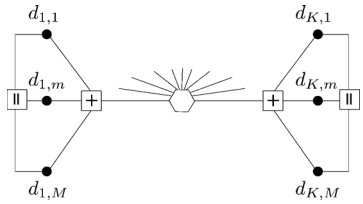


Fig. 2. Factor graph of the basic multiple-access system of (4).

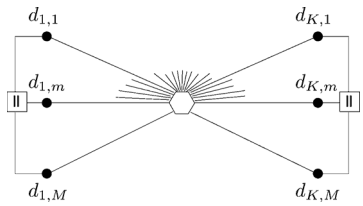


Fig. 3. Factor graph of the basic multiple-access system of (4).

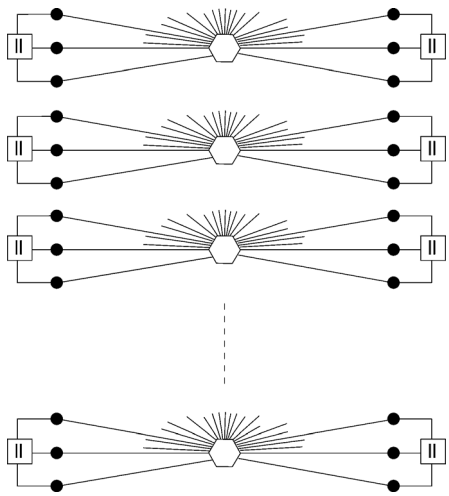


Fig. 4. L-fold lifting of the factor graph of the basic multiple-access system of (4).

addressing the situation we study in this paper, serve to lend credence to the MPM performance formulas of Tanaka [33] to which we relate some of our results.

III. Graph Lifting

We now consider an equivalent model for the system (1). Imagine that each bit is replicated M times as $d_{k,m} = d_k \forall m = 1, \ldots, M$, i.e., the transmission model is

$$\mathbf{y} = \sum_{k=1}^{K} \sum_{m=1}^{M} \frac{d_{k,m}}{M} \mathbf{a}_k + \sigma \mathbf{n}. \tag{4}$$

The factor graph of (4) is shown in Fig. 2.

However, since (4) is linear, we find the representation in Fig. 3 more useful. Note that the central multiple access node now has an M-fold larger incidence degree.

Fig. 4 shows an L-fold lifting of the graph of Fig. 3, which is basically an L-fold repetition. This can be thought of in practice as the transmission of a block of LM signals instead of the original M signals for each user.

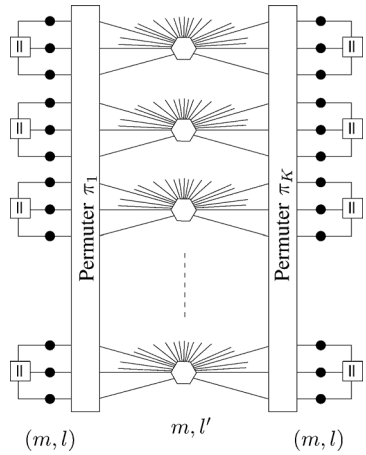


Fig. 5. Graphical structure created from the L-fold lifting of the original factor graph by adding individual permuters.

Of course, up to now nothing has changed in the system or the processing of the data. However, analogously to the construction of LDPC codes from protograph structures [5], [37], a concept initiated by Tanner [35], we now permute the edges among the different copies of the graph lifting using the permuters, or interleavers π_k in Fig. 5. This can be done in a number of convenient ways, such as to avoid contention by a memory-based receiver processor [32]. However, here we simply assume that the permutations are chosen randomly, and independently from each other, but once chosen they remain fixed and our results require no averaging of permutations. The resulting multiple access equation is given as

$$\mathbf{y}_{l'} = \sum_{k=1}^{K} \sum_{m',l'} \frac{d'_{k,m',l'}}{\sqrt{M}} \mathbf{a}_{k,l'} + \sigma \mathbf{n}_{l'} \tag{5}$$

where we have introduced the graph copy indices l and l', and (m', l') are simply the indices of the original location $Ml' + m'$ of an edge that is permuted to location $Ml + m$ in (5). We also denote the permuted symbols by d' for clarity. In order to keep the total energy of the system constant, we have to change the amplitudes of the waveforms to $1/\sqrt{M}$. This is because the individual portions of a given waveform do no longer add coherently as in (4). Fig. 5 shows the resulting graphical model.

The key observation is that instead of solving the NP-hard decoding problem for (1), we apply an iterative message-passing algorithm to the lifted graph in Fig. 5, and we show that this iterative algorithm achieves the same performance as the maximum-likelihood algorithm for the original problem. This is significant in that the exponential complexity per bit has been transformed into the complexity of the message-passing algorithm per graph section. The latter, as we will show, only grows linearly in K. It is important to point out that the iterative algorithm does not directly solve the original NP-hard problem, and is, therefore, not in conflict with the current view that P $\neq$ NP [8], but it produces a solution with the same bit error rate and output SNR performance.

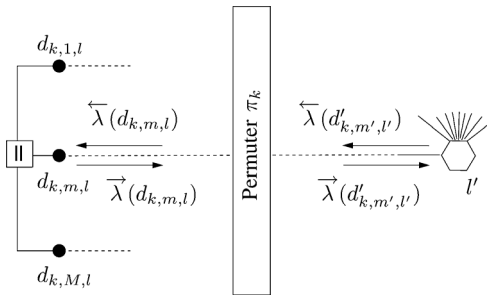


Fig. 6. Illustration of the message exchange mechanism and node functionalities in the message-passing algorithm.

IV. Belief Propagation Message Passing

Message passing in the lifted graph is carried out in the usual way. The equality nodes compute an extrinsic log-likelihood ratio of the bits $d_{k,m,l}$ as the sum of the incoming *a priori* log-likelihood ratios, i.e.,

$$\overrightarrow{\lambda}_i(d_{k,m,l}) = \sum_{n:n\neq m} \overleftarrow{\lambda}_{i-1}(d_{k,n,l}) \tag{6}$$

where $\lambda(d) = \log\left(\Pr(d=1)/\Pr(d=-1)\right)$ formally plays the role of a log-likelihood ratio. The index i indicates the detection iteration and is incremented at the equality nodes. The details of the message passing operation are illustrated in Fig. 6.

Message passing through the multiple access node is a little more involved, and an exact calculation of the extrinsic log-likelihood ratio would lead us back to the NP-hard problem we wanted to avoid in the first place. We, therefore, utilize a much simpler cancellation and subsequent unconstrained LLR computation. For each outgoing edge from the l'th multiple access node, the following canceled and filtered signal is computed and sent to the symbol nodes:

$$z'_{i,k,m',l'} = \mathbf{a}_{k,l'}^T \left(\mathbf{b}_{l'} - \sum_{\substack{\kappa,\mu \\ (\kappa,\mu)\neq(k,m')}} \tanh\left(\frac{\overrightarrow{\lambda}_i(d'_{\kappa,\mu,l'})}{2} \right) \frac{\mathbf{a}_{\kappa,l'}}{\sqrt{M}} \right)$$

$$\overleftarrow{\lambda}_i(d'_{k,m',l'}) = \frac{2}{\sqrt{M}\sigma_i^2} z'_{i,k,m',l'} \tag{7}$$

where σ_i^2 is the joint multiple access and noise interference of the (partially) canceled signal $z_{i,k,m',l'}$. $z_{i,k,m',l'}$ consists of the sum of a Gaussian noise term and a sum of bounded i.i.d. random variables, and becomes Gaussian with variance σ_i^2 and mean $d'_{k,m',l'}/\sqrt{M}$ as $K \to \infty$ by the central limit theorem. Evidently, its variance is independent of the lifting index l, the user index k, and the partition index m. The index i can be understood as an iteration index, since this variance changes with repeated iterations of the processing operations (6) and (7)—see also Fig. 6.

For $(k,l) \neq (\tilde{k},\tilde{l})$, the correlations $\mathrm{E}\left[\left(\mathbf{a}_{k,l}^T \mathbf{a}_{\tilde{k},\tilde{l}}\right)^2\right] = 1/N$ (see Appendix II). Noting that the LLRs in (7) are distributed with mean $2d'_{k,m',l'}/(M\sigma_i^2)$ and variance $4/(M\sigma_i^2)$, and accounting for (6), it follows that

$$\begin{aligned}\sigma_i^2 &= \frac{1}{NM} \sum_{\substack{\kappa,\mu \\ (\kappa,\mu)\neq(k,m')}} \mathrm{E}\left[d_{\kappa,\mu,l'} - \tanh\left(\frac{\overrightarrow{\lambda}_{i-1}(d_{\kappa,\mu,l'})}{2} \right) \right]^2 + \sigma^2 \\ &= \frac{1}{NM} \sum_{\substack{\kappa,\mu \\ (\kappa,\mu)\neq(k,m'))}} \mathrm{E}\left[1 - \tanh\left(\frac{M-1}{M\sigma_{i-1}^2} + \sqrt{\frac{M-1}{M\sigma_{i-1}^2}}\xi \right) \right]^2 + \sigma^2 \end{aligned} \tag{8}$$

where $\xi \sim \mathcal{N}(0,1)$ and the expectation is taken over the random interfering signals in $\overrightarrow{\lambda}_{i-1}(d'_{\kappa,\mu,l'})$. That is, assuming that (7) describes a Gaussian random variable—which is true for K large—(8) is an exact expression for all $i < I_{\mathrm{cycle}}/2$, where I_{cycle} is the length of the shortest cycle in the graph of Fig. 5. Equation (8) is, therefore, exact for as long as the system size L is large enough to ensure large minimum cycles. Typically, this means that L grows as $\propto e^{I_{\mathrm{cycle}}}$. Nonetheless, the analysis does not completely collapse for shorter L, but becomes increasingly approximative. We shall assume in what follows that $L \to \infty$, without deriving any practical suggestions from this statement.

In order to simplify the notation, we define

$$g\left(\frac{1}{x}\right) = \mathrm{E}\left[1 - \tanh\left(\frac{1}{x} + \sqrt{\frac{1}{x}}\xi \right) \right]^2 \tag{9}$$

and we note that $g\left(1/x\right)$ is a monotonically increasing function of x [4].

As a consequence, σ_i in (8) is monotonically nonincreasing with i, which can be seen by induction as follows. Consider the sequence $v_0^2, v_1^2, v_2^2, \ldots$, where $v_i^2 = g\left((M-1)/(M\sigma_{i-1}^2)\right)$, which we call *soft bit variance*, is the mean squared error of data bit estimates $\tanh\left(\overrightarrow{\lambda}_i(d_{\kappa,\mu,l})/2\right)$ at iteration i. Now, with $v_0^2 = 1$, and $v_1^2 \leq 1$, the sequence σ_i^2 in (8) is monotonically nonincreasing, i.e.,

$$v_i^2 = g\left(\frac{M-1}{\frac{KM-1}{N} v_{i-1}^2 + \sigma^2} \right) \leq g\left(\frac{M-1}{\frac{KM-1}{N} v_{i-2}^2 + \sigma^2} \right) = v_{i-1}^2. \tag{10}$$

Since $\sigma_i^2 \geq 0$, there must be a fixed point which is reached, and it is implicitly given by

$$\begin{aligned}\sigma_\infty^2 &= \frac{KM-1}{NM} \mathrm{E}\left[1 - \tanh\left(\frac{M-1}{M\sigma_\infty^2} + \sqrt{\frac{M-1}{M\sigma_\infty^2}}\xi \right) \right]^2 + \sigma^2 \\ &\to \alpha\, \mathrm{E}\left[1 - \tanh\left(\frac{M-1}{M\sigma_\infty^2} + \sqrt{\frac{M-1}{M\sigma_\infty^2}}\xi \right) \right]^2 + \sigma^2 \end{aligned} \tag{11}$$

where in the second row we have used $KM \gg 1$, which is the case for most situations of practical interest.

Note that the analysis up to this point does not depend on the specific binary sequence chosen in the K data streams, and that the second line of (8) is independent of the sign of $d_{\kappa,\mu,l'}$. Therefore, any fixed sequence can be used, for example, the sequence with $d_{k,m,l} = -1, \forall k,m,l$.

At each iteration, the equality nodes will collect the LLR messages reported from the network to them and form an ith soft symbol estimate (or *a posteriori* LLR value) to be reported to the external system. This LLR is given by

$$\lambda(d_{k,l}) = \sum_{n=1}^{M} \overleftarrow{\lambda}(d_{k,n,l}) \tag{12}$$

for each binary symbol $d_{k,l}$. Following (5), (11), and (6), $\lambda(d_{k,l})$ has the same distribution as $d_{k,l}$ after transmission through a binary-input AWGN channel with SNR $\gamma_i^2 = 1/\sigma_i^2$, which follows the iteration equation

$$\gamma_i^2 = \left(\alpha \mathrm{E}\left[1-\tanh\left(\frac{M-1}{M}\gamma_{i-1}^2 + \sqrt{\frac{M-1}{M}}\gamma_{i-1}\xi\right)\right]^2 + \sigma^2\right)^{-1}. \tag{13}$$

We note that for large partition numbers $M \to \infty$, the fixed point SNR of (13) and the implicit SNR equation for the optimal symbol detector (3) are identical. That is, the SNR γ_i^2 approaches the optimal symbol SNR as iterations approach the fixed point. This shows that simple message passing in the (large) lifted and randomized graph of the random signal multiple access system achieves "optimal" performance, in the sense that its error performance is identical to that of operating an NP-hard maximum marginal probability detector on the original system.

We summarize this result in the following.

Proposition 1: The performance of the NP-hard optimal bit-by-bit MPM detection in the (dense) multiple access system (1) is achieved by cancellation-based message passing in the corresponding sparse lifted multiple access system (5).

Comments: We call the aforementioned result a proposition since it relies on the correctness of (3), which is based on replica analysis, and is, therefore, only conjectured to be correct. In [12], progress has been made toward proving replica-based results for the capacity of the random CDMA channel by deriving a capacity upper bound which coincides with the corresponding replica formula. The replica approach also coincides with the capacity formula for Gaussian input signals. There exists, therefore, strong evidence for its correctness; however, a rigorous way of deriving (3) is still lacking.

V. Discussion

For $\sigma^2 > \sigma^2_{\mathrm{crit}} = 0.148$, (13) has only a single solution. However, for $\sigma^2 \leq \sigma^2_{\mathrm{crit}}$, one, two, or three solutions exist, depending both on σ^2 and α. The relevant solution, i.e., the fixed point approached by the iterative detector, is the one with the largest σ^2_∞. Only the case with a single solution is practically relevant, since the other solutions are high-noise fixed points where the decoder essentially fails, i.e., produces high bit error rates (see Fig. 7). For each $\sigma^2 \leq \sigma^2_{\mathrm{crit}}$, we have an associated maximum α such that a single solution exists. For $\sigma^2 \to 0$, this maximum α is the total system load that is supportable at the SNR $\to \infty$, given by $\alpha_{\max} = 2.07425$.

Fig. 7 shows how the appearance of additional solutions translates into a "turbo-cliff" behavior of the bit error rate. This abrupt phase change is, therefore, the characteristic of the optimal detector as well. Our results also suggest that the phase transitions observed in the statistical mechanics approach [33, Fig. 1] are artifacts of the analysis, and are not a phenomenon of the detector itself, whose error curves transition sharply as the fixed point jumps at the critical SNR.

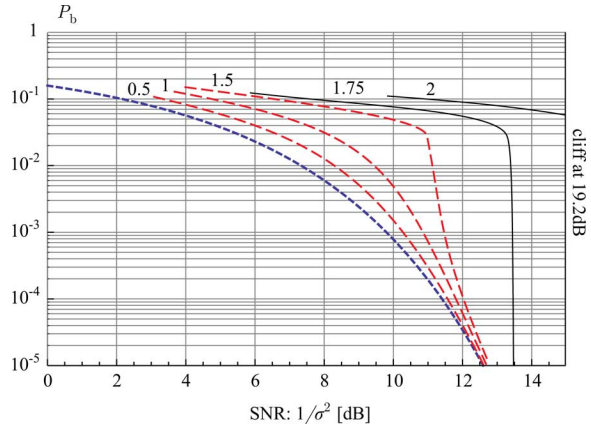


Fig. 7. Bit error rate of the MPM detector for various loads $\alpha = K/N$. The number on top of each curve indicates the load value α, for which the curve is plotted. The short-dashed line denotes the bit error rate of the BPSK signaling on an AWGN channel.

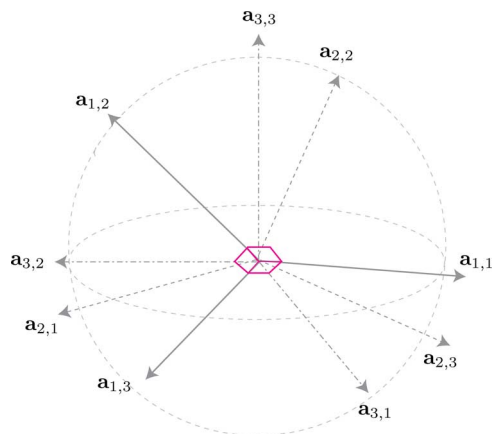


Fig. 8. Illustration of the redundant signal population using the lifted system.

The bit error rate was computed from the SNR γ^2_∞ as $P_{\mathrm{b}} = Q(\gamma_\infty)$, where $Q(\cdot)$ is defined in Appendix I. The short-dashed line denotes the bit error rate of antipodal signaling on an AWGN channel. We can deduce, for example, that at $\alpha = 1$ the cost of using random signaling instead of orthogonal signals is about 2 dB at $P_{\mathrm{b}} = 10^{-2}$, and shrinks to a fraction of a dB for lower error rates.

Just as with irregular low-density parity-check codes, where an optimization of the node degrees led to significant improvements in performance [22], the partition degree of our lifted graph does not have to be equal for all users, or all symbols for that matter. We are free to chose a partition degree profile $M_{1,1}, \ldots, M_{1,K}, M_{2,1}, \ldots, M_{L,K}$. The fact that gains in performance can be achieved can be seen through the following argument. Letting $M_{l,k} = M_k$, it is not hard to see that as long as each symbol is transmitted with constant power $1/M$, this amounts to assigning different powers to different users. However, it is already known that with exponentially distributed power assignments, or equivalent rate assignments, the multiple access channel capacity can be achieved with a cancellation system of this type [28], [38].

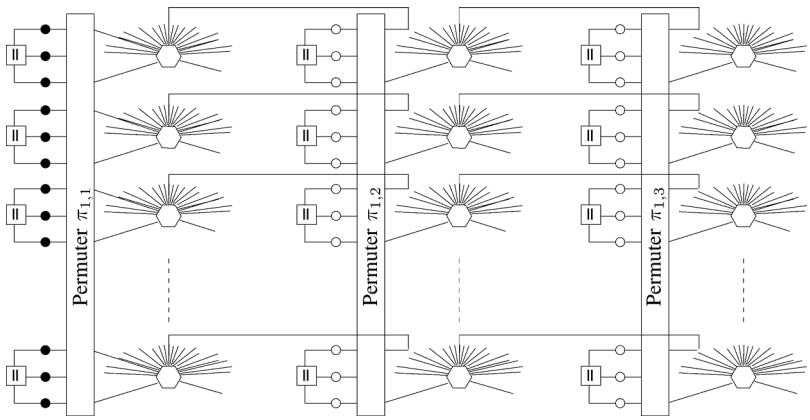


Fig. 9. Basic graphical coupling of different or successive lifted graphs. Couplings with wider span will be more effective.

Finally, we wish to comment on a subtle difference in the use of the signal space of the lifted system, compared to the original system. As mentioned in Section II (and Appendix II), the signals $\mathbf{a}$ are drawn uniformly and randomly from the N-dimensional signal space. With the lifting, each signal is split into M signals, also drawn uniformly randomly. Note that the total power and the total bandwidth utilization are equal in both systems. However, the lifted signal has a higher signal density, as M times as many signals crowd the original signal space. This is illustrated in Fig. 8.

Note that without loss of generality to the analysis, we may wish to enforce that the signals belonging to the same user k, and the same time interval l, are chosen to be mutually orthogonal, i.e., $\mathbf{a}_{k,l,m_1} \perp \mathbf{a}_{k,l,m_2}$ for all $m_1 \neq m_2$. This causes no significant design overhead, since instead of using M random signals, each user now uses a random M-dimensional basis to transmit its symbol fragments in time interval l. The analysis presented in this paper carries through without change. However, from an operational point of view, signal processing at the receiver is facilitated since no intrauser intersignal interference needs to be canceled.

VI. Spatial Coupling

In this section, we address the issue of the maximum system load of $\alpha_{\max}$ introduced in Section V. The question is in what way this is a limit to the message-passing iterative detector—in short, it is not. This will be demonstrated in this section.

In the development of LDPC convolutional coding [7], a phenomenon called "spatial coupling" has allowed these codes to be designed with decoding thresholds that are very close to the channel capacity. The effect of spatial coupling derives from anchoring initial symbols to known values, which causes a locally reduced information rate. This in turn allows the iterative decoder for the coupled system to converge to low bit error rates for SNRs where such convergence in the uncoupled system is not possible. Recently, it has been shown that spatial coupling can improve the convergence threshold of message-passing iterative decoding of low-density parity-check codes on binary-erasure channels all the way to the maximum-likelihood threshold [15].

In our system, we introduce spatial coupling by duplicating the graphical structures of Fig. 5, i.e., lifting the graph a second time. One can think of this as considering symbol transmission intervals in time. We then anchor the signals of the left-most graph to known values; these can be thought of as pilot signals, for example. The spatial coupling in its simplest form is achieved by connecting a fraction b of the nodes of the graph at replica (time) t, to the multiple access nodes of replica $t-1$. This is illustrated in Fig. 9, where the solid black nodes denote the anchored known symbols.

By this process, the first set of multiple access nodes effectively experiences a lower load, and will, therefore, converge faster. Astoundingly, this effect propagates through the semi-infinite lifted graph indefinitely for $L \to \infty$.

For the further discussion, we substitute $x_i^t = \sigma_{t,i}^2$. A density evolution analysis for large graphs then links the variances at the different replicas via the following equation:

$$x_i^1 = \alpha b g\left(\frac{1-b}{x_{i-1}^2} + \frac{b}{x_{i-1}^1}\right) + \sigma^2 \tag{14}$$

$$x_i^t = \alpha b g\left(\frac{1-b}{x_{i-1}^{t+1}} + \frac{b}{x_{i-1}^t}\right) + \alpha(1-b) g\left(\frac{1-b}{x_{i-1}^t} + \frac{b}{x_{i-1}^{t-1}}\right) + \sigma^2.$$

Here, we assume without essential loss of generality that M is large, so we can approximate $M - 1 \approx M$ in the density evolution equations.

The question now is if the iterative system of (14) allows convergence at higher values of $\alpha > \alpha_{\max}$. The answer is affirmative; however, the maximum load α_{coupling} now depends on the coupling connections, but $\alpha_{\text{coupling}} > \alpha_{\max}$ for any coupling. In particular, for the coupling of Fig. 9 with $b = 0.5$ a maximum coupled load $\alpha_{\text{coupling}} = 2.81 > \alpha_{\max}$ is achieved.

Fig. 10 shows that the variance values at the subgraphs at different indices t evolve completely homogeneously, with a constant iteration delay toward larger t. That is, the solutions to the balance (14) are linear shifts of each other.

Fig. 11 shows the convergence of the variance values at the subgraphs at an SNR of 10 dB. At that noise level, the system convergence is demonstrated for $\alpha = 1.95$. For comparison, without spatial coupling the system does not converge[2] at this SNR and load level; see Fig. 7.

[2]In these qualitative discussions, we use the term convergence in the loose sense of achieving convergence to a point of low noise, and not in the sense of reaching a fixed point, which always exists.

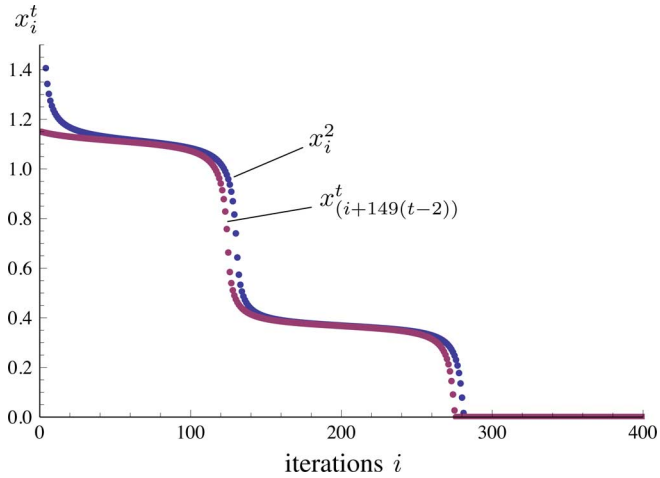


Fig. 10. Homogeneous convergence in the simply coupled graph of Fig. 9 with $b = 0.5$ and $\alpha_{\text{coupling}} = 2.81$ in the noiseless case. The index relating replica number t and iteration i above depend on b and are found experimentally.

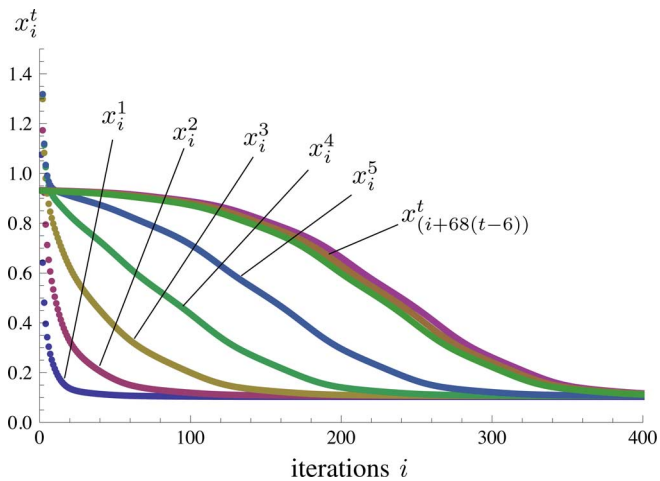


Fig. 11. Convergence in the simply coupled graph of Fig. 9 with $b = 0.5$, $\alpha = 1.95$, and $\text{SNR} = 10$ dB, hence the floor at $x_i^t = 0.1$.

We may generalize the coupling to include W previous and W future graphs, i.e., utilize a coupling window of $2W + 1$, instead of the basic example in (14). The first W subgraphs are now the anchor symbols and are set to fixed values. The variance update equations at replica t can now be computed as

$$x_i^t = \frac{\alpha}{2W+1} \sum_{j=-W}^{W} g\left(\sum_{l=-W}^{W} \frac{1}{2W+1} \frac{1}{x_{i-1}^{t+j+l}} \right) + \sigma^2 \quad (15)$$

for $i \geq 1$ and $t > 0$. Since we assume that all the messages transmitted at time prior to $t = 1$ are known (or anchored to 0 values), the initial condition can be formulated as

$$x_0^t = 0 \quad t \leq 0 \quad (16)$$

$$x_0^t = \frac{\alpha t}{2W+1} + \sigma^2; \quad t \in [1, 2W+1] \quad (17)$$

$$x_0^t = \alpha + \sigma^2; \quad t > 2W+1. \quad (18)$$

The system load of the coupled system described previously is time dependent. The average system load α_t at replica t equals

$$\alpha_t = \begin{cases} \alpha \frac{t+1}{2(2W+1)}; & t \in [1, 2W+1] \\ \alpha \frac{t-W}{t}; & t > 2W+1. \end{cases}$$

TABLE I
ACHIEVABLE LOADS AS A FUNCTION OF COUPLING WINDOW SIZE

W	0	1	2	3	4	5	10	20	50
α_{coupling}	2.07	3.17	3.6	3.9	4.1	4.3	4.9	5.5	6.2

Note that rate reduction induced by anchoring disappears as t grows, i.e., $\alpha_t \to \alpha$ as $t \to \infty$.

This allows us to achieve progressively larger maximum values α_{coupling} where convergence can be achieved. In fact, as shown in Theorem 1, an arbitrarily large system load can be achieved in the noise-free case. The argument of $g(\cdot)$ in the aforementioned equation, the signal-to-noise and interference value of the message received at replica $t+l-W$, is calculated assuming that $M \gg W$ and the M sections of each signal are uniformly spread among the $2W + 1$ neighboring sections according to (15).

Theorem 1: For $\sigma^2 = 0$, and with a sufficiently large W, i.e., $W = O\left(e^{\alpha \ln \alpha}\right)$, the iteration equations (15) converge to $x_i^t \to 0$ for all t and all values of α. In particular, we can let $\alpha \to \infty$.

Proof: See Appendix I.

Theorem 1 shows that the spatially coupled iterative demodulator is no longer "interference limited," in the sense that it can support any system load given the necessary SNR. This is in marked contrast to all known equal-power multiuser detectors, in particular also to the detector in Fig. 5. Some values of achievable loads are shown in Table I and illustrate the growth trend required for the window size W. The achieved values may be compared with the recent work of Takeuchi *et al.* [34] which have demonstrated that spatial coupling of "regular" CDMA systems allows system loads up to $\alpha = 2.95$ to be achieved. The addition of interleaving modules in Section III, i.e., the creation of a sparse processing graph, is therefore fundamental to achieving maximum performance.

VII. Spatial Coupling and the Multiple-Access Channel Capacity

The situation of a coupled system with $\sigma^2 \neq 0$ is substantially more complicated to treat. We begin our discussion with the following observations. Fig. 12 shows the convergence behavior of a system with $\alpha = 3$ and $W = 5$. We note, in particular, that two stable convergence levels establish themselves: an upper level convergence level and a lower convergence level. These are related to two stable (attracting) fixed points of (11). The upper fixed point is in fact given by σ_∞^2 in (11), while the lower fixed point cannot be reached in the original system.

The dynamics of the convergence phenomenon can be visualized with reference to Fig. 13, which shows the convergence equation (11), i.e.,

$$x_i = \alpha g\left(\frac{1}{x_{i-1}}\right) + \sigma^2 \quad (19)$$

as two parts, viz. the linear equation $x_i = \alpha w + \sigma^2$ and the expectation $w = g(1/x_i)$. While (19) holds for the uncoupled system, it is easy to see that both of its stable fixed points are also fixed points of the coupled system (15).

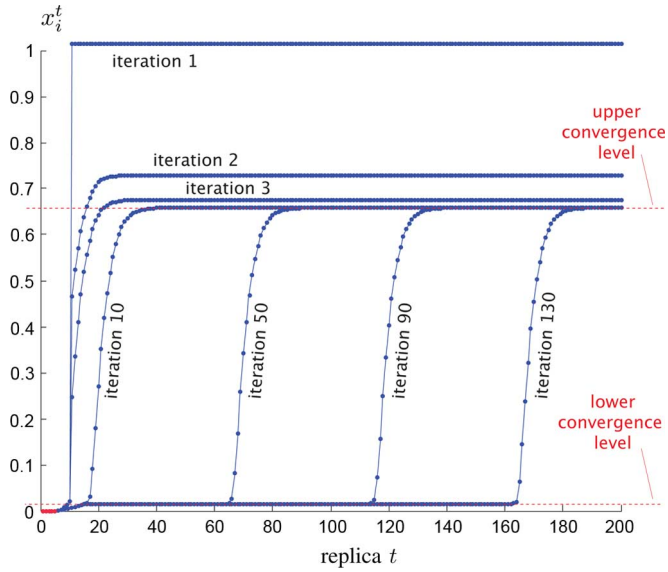


Fig. 12. Another convergence example illustrating a convergence "wave" propagating through the system.

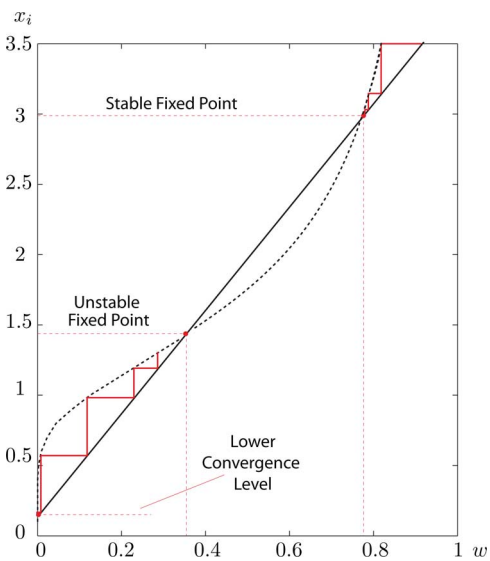


Fig. 13. Graphical illustration of the fixed points of (11).

While the uncoupled system converges to the first stable fixed point [of (11)], coupling enables the system to "tunnel" through the region of nonconvergence and achieve the low-noise stable fixed point, even at loads that are unachievable for the uncoupled system. This tunneling gives rise to the convergence "wave" seen in Fig. 12 that migrates away from the "anchors."

Let us denote the maximum load achievable by the uncoupled system at a given noise level σ^2 by $\alpha_{\max}(\sigma^2)$, and the load achievable by coupled system by $\alpha_{\text{coupling}}(\sigma^2)$. Then, we have the following.

Proposition 2: For $\alpha \in [\alpha_{\max}(\sigma^2), \alpha_{\text{coupling}}(\sigma^2)]$, the noise-and-interference variance x_i^t of the coupled system converges to $x_l(\sigma^2, \alpha)$, the smallest positive fixed point of (19) (out of three positive fixed points) as $i \to \infty$ for each replica t.

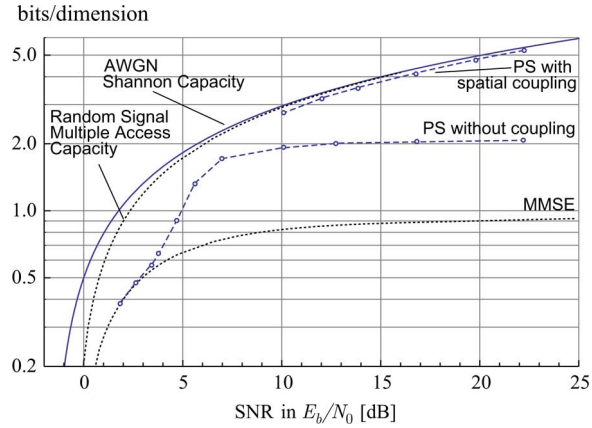


Fig. 14. Spectral efficiency values achievable with a coupled partitioned system under iterative decoding.

The resulting spectral efficiency in bits per dimension achievable by the coupled system is then calculated as

$$\mathcal{C}_{\text{eff}}(\sigma^2, \alpha_{\text{coupling}}(\sigma^2)) = \alpha_{\text{coupling}}(\sigma^2)\mathcal{C}_{\text{BIAWGN}}\left(\frac{\alpha_{\text{coupling}}(\sigma^2)}{x_l(\sigma^2, \alpha)}\right) \quad (20)$$

where $\mathcal{C}_{\text{BIAWGN}}(\gamma)$ is the capacity of the binary-input AWGN channel with SNR γ, and we assume that we are using binary input AWGN capacity achieving codes for each data stream.

Fig. 14 shows the numerical evaluations of (20) for large W and the maximum α_{coupling} achievable with the structure (15) (curve labeled "PS capacity with spatial coupling"), as well as those achievable with the uncoupled demodulator (curve labeled "PS capacity"), and a basic MMSE filter receiver. From these results, we conjecture that the coupled system (15) can achieve the multiple-access channel capacity for sufficiently large coupling windows W, and in the medium-to-high SNR regime. That is

$$\mathcal{C}_{\text{eff}}(\sigma^2, \alpha_{\text{coupling}}(\sigma^2)) \to \frac{1}{2}\log_2\left(1 + \frac{\alpha_{\text{coupling}}(\sigma^2)}{\sigma^2}\right)$$

as $W \to \infty$, since the random multiple-access channel capacity also approaches the AWGN channel capacity for large SNRs and load values α (see [10]).

A formal proof of this conjecture is currently not known and is the object of further investigations.

VIII. Conclusion

We have shown that simple cancellation-based message-passing demodulation in an properly "sparsified" system graph can achieve optimal maximum *a posteriori* performance of the underlying dense-graph system. In particular, we applied these concepts to the demodulation for random signal sets, as they occur, for example, in random CDMA. We then further introduced spatial coupling and "anchoring" of neighboring sparse graphs and showed that the spectral efficiency limitation of the original system could be eliminated allowing the system load to go to infinity as the SNR goes to infinity. We then conjectured, and numerically demonstrated, that spatial coupling of sufficient width can achieve the capacity of the random multiple access channel.

APPENDIX I
PROOF OF THEOREM 1

We restate Theorem 1 in a slightly more detailed way for the purpose of this proof.

Theorem 1: Consider an interference limited system with $\sigma^2 = 0$, $\alpha > 2$, and

$$W > \exp(15\alpha \ln \alpha + 20\alpha). \tag{21}$$

Then, for any t

$$\lim_{i\to\infty} x_i^t = 0$$

where after $i \gg 2t$ iterations x_i^t converges to 0 at least exponentially. (The specific numbers 15 and 20 in (21) are merely for convenience.)

Proof: The proof is divided into two parts. First, we show in Lemma 3 that for W satisfying (21) and after $2t$ iterations it holds that $x_{2t}^t < \alpha\epsilon$ where

$$\epsilon = \pi \exp\left(-\frac{1}{4\alpha\pi}(2W+1)^{\frac{1}{4\alpha}}\right). \tag{22}$$

In the next step (Lemma 4), we prove that

$$x_{i+1}^t < x_i^t/2$$

for $i > 2t + 2W$. The effect of spatial coupling initiated at position 0 requires $2t$ iterations to propagate to position t.

In the spatially coupled system of (15), let q_i^t denote the soft bit variance in the graph replica t. Then

$$x_i^t = \frac{\alpha}{2W+1}\sum_{j=-W}^{W} q_i^{t+j} \quad t>0, i>0 \tag{23}$$

and the soft bit variance

$$q_i^t = g\left(\frac{1}{2W+1}\sum_{l=-W}^{W}\frac{1}{x_{i-1}^{t+l}}\right) \quad t>0, i>0. \tag{24}$$

Combining these two equations, we obtain

$$\begin{aligned} x_i^t &= \frac{\alpha}{2W+1}\sum_{j=-W}^{W} q_i^{t+j} \\ &= \frac{\alpha}{2W+1}\sum_{j=-W}^{W} g\left(\frac{1}{2W+1}\sum_{l=-W}^{W}\frac{1}{x_{i-1}^{t+j+l}}\right) \end{aligned} \tag{25}$$

for $t > 0,\ i > 0$. We assume that transmission starts at time $t = 1$ with replica $t = 1$, i.e., the load of the system increases gradually. At every later replica t, the system load increases by $\alpha/(2W + 1)$ until it reaches α. The load then stays at level α for $t > 2W + 1$. The initial condition can be formulated as

$$q_0^t = 0 \quad t \le W \tag{26}$$

$$q_0^t = 1 \quad t \ge W+1 \tag{27}$$

which along with (23) implies (16)–(18).

Lemma 1: Consider an arbitrary replica t_0 and assume that for all t such that $t < t_0 + W$ the soft bit variance satisfies $q_{i-1}^t \le \epsilon$, and, consequently, for all t such that $t < t_0$ the noise and interference variance satisfies $x_{i-1}^t < \epsilon\alpha$, where ϵ is given by (22). Then

$$q_{i+2}^{t_0+W} < \epsilon$$

and as a result $x_{i+2}^{t_0} < \alpha\epsilon$.

Note that the initial conditions for spatially coupled system for $t_0 = 1$ satisfy the conditions of the lemma. Once the lemma is applied to graph replica t_0, it can be repeated again for graph replica $t_0 + 1$ and so on.

Proof of Lemma 1: It is easy to see that

$$q_{i-1}^t \le 1 \quad \text{for all } t \tag{28}$$

since $g(\cdot) \in [0, 1]$. According to (23), (28), and one of the conditions of Lemma 1 (namely $q_{i-1}^t \le \epsilon$ for $t < t_0 + W$), we obtain

$$x_{i-1}^{t_0+j} = \frac{\alpha}{2W+1}\sum_{l=-W}^{W} q_{i-1}^{t_0+j+l} \tag{29}$$

$$\le \frac{\alpha}{2W+1}((2W-j)\epsilon + j + 1)\ j \in [0, 2W]. \tag{30}$$

Let us proceed with iteration i. According to (24) and the aforementioned equation, we obtain

$$\begin{aligned} q_i^{t_0+W} &= g\left(\frac{1}{2W+1}\sum_{l=-W}^{W}\frac{1}{x_{i-1}^{t_0+W+l}}\right) \\ &= g\left(\frac{1}{2W+1}\sum_{j=0}^{2W}\frac{1}{x_{i-1}^{t_0+j}}\right) \\ &\le g\left(\frac{1}{\alpha}\left[\frac{1}{2W\epsilon+1}+\frac{1}{(2W-1)\epsilon+2}+\cdots+\right.\right. \\ &\qquad \left.\left.\frac{1}{2W+1}\right]\right) \end{aligned}$$

where the last inequality follows since $g(1/x)$ is an increasing function of x. Therefore,

$$q_i^{t_0+W} \le g\left(\frac{1}{\alpha}\frac{1}{2W\epsilon+1}\left[1+\frac{1}{2}+\cdots+\frac{1}{2W+1}\right]\right) \tag{31}$$

since for any $k > 0$ integer

$$\frac{2W\epsilon+1}{(2W-k)\epsilon+k} \ge \frac{1}{k}.$$

For W satisfying (21), and ϵ as in (22), it can be shown that

$$\epsilon < \frac{1}{2W} \tag{32}$$

and, therefore,

$$\begin{aligned} q_i^{t_0+W} &\overset{(a)}{\le} g\left(\frac{\ln(2W+1)}{2\alpha}\right) \\ &\overset{(b)}{\le} \pi Q\left(\sqrt{\frac{\ln(2W+1)}{2\alpha}}\right) \\ &\overset{(c)}{\le} \pi\exp\left(-\frac{\ln(2W+1)}{4\alpha}\right) \\ &= \pi(2W+1)^{-\frac{1}{4\alpha}} \end{aligned} \tag{33}$$

where

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-z^2/2}dz.$$

Inequality (a) in (33) follows from the logarithmic lower bound on the series in the argument of $g(\cdot)$ on the right-hand side of (31), the upper bound on ϵ(32), and the fact that $g(\cdot)$ is monotonically decreasing. Inequality (b) in (33) follows from an upper bound on function $g(\cdot)$ derived in [4]. Inequality (c) in (33) follows from a simple upper bound on $Q(\cdot)$.

Inequality (33) combined with (29) results in

$$x_i^{t_0} \leq \epsilon\alpha + \frac{\alpha}{2W+1}\pi(2W+1)^{-\frac{1}{4\alpha}}.$$

Likewise for W satisfying (21), and again using (22), we obtain

$$\epsilon < \pi(2W+1)^{-1-\frac{1}{4\alpha}} \tag{34}$$

which finally leads to

$$x_i^{t_0} \leq 2\alpha\pi(2W+1)^{-1-\frac{1}{4\alpha}}.$$

Now, we have

$$\begin{aligned} q_{i+1}^{t_0+W} &\leq g\left(\frac{1}{2W+1}\frac{1}{2\alpha\pi(2W+1)^{-1-\frac{1}{4\alpha}}} + \cdots\right) \\ &\leq g\left(\frac{1}{2\alpha\pi(2W+1)^{-\frac{1}{4\alpha}}}\right) \\ &\overset{(a)}{\leq} \pi\exp\left(-\frac{1}{2}\frac{(2W+1)^{\frac{1}{4\alpha}}}{2\alpha\pi}\right) = \epsilon, \end{aligned} \tag{35}$$

which implies

$$x_{i+1}^{t_0} \leq \epsilon(2W+1)\alpha/(2W+1) = \alpha\epsilon.$$

Inequality (a) in (35) is obtained analogously to (b) and (c) in (33). Lemma 1 is proven.

Let $\epsilon_{\rm crit}$ be the smallest positive root of

$$\pi\exp\left(-\frac{1}{2\alpha\epsilon}\right) = \epsilon/2. \tag{36}$$

Notice that for $\epsilon < \epsilon_{\rm crit}$

$$\pi\exp\left(-\frac{1}{2\alpha\epsilon}\right) < \epsilon/2. \tag{37}$$

Moreover, ϵ defined by (22) satisfies (37).

Lemma 2: Assume that for any t the inequalities $q_{i-1}^t \leq \epsilon$ and, consequently, $x_{i-1}^t < \epsilon\alpha$ are satisfied. If $\epsilon < \epsilon_{\rm crit}$ then

$$q_i^t < \epsilon/2$$

and as a result $x_i^t < \alpha\epsilon/2$ for any t as well.

Proof of Lemma 2:

$$\begin{aligned} q_i^t &\leq g\left(\frac{1}{2W+1}\left(\frac{1}{\alpha\epsilon} + \cdots + \frac{1}{\alpha\epsilon}\right)\right) \\ &\leq \pi\exp\left(-\frac{1}{2\alpha\epsilon}\right) < \epsilon/2 \end{aligned} \tag{38}$$

where the second inequality follows from upper bounding $g(\cdot)$ as in (33) (c). ∎

Since ϵ defined by (22) satisfies (37), it implies

$$x_{i+1}^t < x_i^t/2$$

for $i > 2t + 2W$. This means that for $T \gg t$, and after $i > T$ iterations x_i^t will decrease to zero exponentially.

Appendix II
Random Signal Waveforms

Consider a set of independent identically distributed random unit-length vectors in N-dimensional space. The endpoint of each vector is uniformly distributed in an $(N-1)$-dimensional sphere of unit radius centered at the origin.

Preliminaries:

Consider two unit-length vectors $\mathbf{a}$ and $\mathbf{b}$ in two dimensions. Then, the power of the interference imposed by one vector on the other is given by their cross product, i.e.,

$$\rho^2 = (\mathbf{a}\cdot\mathbf{b})^2 = |\mathbf{a}|^2|\mathbf{b}|^2\cos^2(\phi)$$

where ϕ is the angle between them.

Averaging this over $\phi \in [0, \pi]$ gives 1/2.

Lemma 3: The average squared cross correlation of two random unit-length vectors $\mathbf{a}$ and $\mathbf{b}$ in the sense above is given by

$$\overline{\rho^2} = \overline{(\mathbf{a}\cdot\mathbf{b})^2} = \frac{1}{N}. \tag{39}$$

Proof: Assume without loss of generality that the first basis vector of N-space coincides with $\mathbf{a}$. Now decompose $\mathbf{a}$ into two orthogonal parts: $\mathbf{a}_{\rm p} = \mathbf{a}\cos(\phi)$ and $\mathbf{a}_{\rm q} = \mathbf{a}\sin(\phi)$. Given that the endpoint of $\mathbf{a}$ is uniformly distributed over the $N-1$ bounding sphere, we first need to compute the ratio of such vectors $\mathbf{a}$ that form the angle ϕ with $\mathbf{a}$.

But $\mathbf{b}_{\rm q}$ is simply a vector of length $\sin(\phi)$ that lives in $(N-2)$-dimensional sphere of radius $\sin(\phi)$. Therefore, the probability density function of the pair forming an angle ϕ is the ratio of the volume of that $(N-2)$-dimensional sphere to the volume of the $(N-1)$-dimensional sphere that bounds the space of all vectors. Denoting the volume of an N-dimensional unit sphere by C_N, the volume of a sphere of radius R is given by $C_N R^N$, and its $N-1$ dimensional surface by $NC_N R^{N-1}$. Therefore, the ratio in question is

$$p(\phi) = \frac{C_{N-1}\sin^{N-2}(\phi)}{NC_N}.$$

Given that

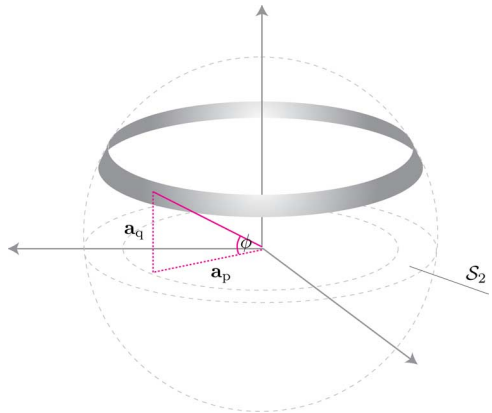
$$C_N = \frac{\pi^{N/2}}{\Gamma[N/2+1]}$$

it is a simple exercise to verify the lemma. q.e.d.

Now let us extend the lemma to the case of subspaces of dimension $K > 1$. We assume that K random vectors have been chosen, which span a K-dimensional subspace almost surely. We consider the case where a new vector is projected onto that K-dimensional subspace (or onto its orthogonal complement). We have the following lemma.

Lemma 4: The projection power of a random vector $\mathbf{a}$ onto the orthogonal complement of the space spanned by K random vectors $\mathbf{b}_1, \ldots, \mathbf{b}_K$ is given by

$$\rho_{K,N}^2 = \frac{N-K}{N}.$$



Proof: Again, consider that the vector $\mathbf{a}$ is decomposed into two orthogonal parts, $\mathbf{a}_\mathrm{p} = \mathbf{a}\cos(\phi)$ and $\mathbf{a}_\mathrm{q} = \mathbf{a}\sin(\phi)$, where $\mathbf{a}_\mathrm{p}$ lies in the space $\mathcal{S}_K$ spanned by $\mathbf{b}_1, \ldots, \mathbf{b}_K$, and $\mathbf{a}_\mathrm{q}$ lies in its complement. The angle ϕ is the angle $\mathbf{a}$ forms with $\mathcal{S}_K$. Conditioning on ϕ, we compute the area on the surface of C_N where the endpoints of $\mathbf{a}$ can lie. The ratio of this area to the area of C_N will give us the probability of ϕ, since $\mathbf{a}$ is uniformly, and *independently* distributed over the N-sphere.

The endpoints of $\mathbf{a}_\mathrm{p}$ lie in $\mathcal{S}_K$, in fact the surface of a K-dimensional sphere of radius $\cos(\phi)$. The differential volume of that sphere is $KC_K \cos^{K-1}(\phi)$. $\mathbf{a}_\mathrm{q}$ is orthogonal to $\mathcal{S}_K$ and lies in the $(N-K-1)$-dimensional sphere of radius $\sin(\phi)$ (one dimension is fixed because $\mathbf{a}$ has to lie on the surface of the N-sphere). The total differential area of possible endpoints for $\mathbf{a}$ is given by the following:

$$\text{Area of endpoints of } \mathbf{a} \text{ with } \mathbf{a}\cdot\mathcal{S}_K = \phi \text{ equals}$$
$$KC_K \cos^{K-1}(\phi)(N-K)C_{N-K}\sin^{N-K-1}(\phi).$$

The probability that $\mathbf{a}$ forms the angle ϕ with $\mathcal{S}_K$ is

$$\Pr(\mathbf{a}\cdot\mathcal{S}_K = \phi) = \frac{KC_K \cos^{K-1}(\phi)(N-K)C_{N-K}\sin^{N-K-1}(\phi)}{NC_N}.$$

This is illustrated in the figure below for $N = 3$ and $K = 2$.

The squared length of the projection $\mathbf{a}_\mathrm{p}$, i.e., the power of $\mathbf{a}$ in the subspace $\mathcal{S}_K$, is equal to $\cos^2(\phi)$, and therefore, the average projected power of $\mathbf{a}$ onto $\mathcal{S}_K$ is

$$\begin{aligned}\mathbf{E}\|\mathbf{a}_\mathrm{p}\|^2 &= \int_0^{\pi/2} \Pr(\mathbf{a}\cdot\mathcal{S}_k = \phi)\cos^2(\phi)d\phi \\ &= \int_0^{\pi/2} \frac{kC_k \cos^{K+1}(\phi)(N-K)C_{N-K}\sin^{N-K-1}(\phi)}{NC_N} d\phi \\ &= K/N. \end{aligned} \tag{40}$$

The last equation can be verified by any of the integration tools. Complementing $\|\mathbf{a}\|^2 = \|\mathbf{a}_\mathrm{p}\|^2 + \|\mathbf{a}_\mathrm{q}\|^2$, we obtain the lemma. ■

As observed by the editor, (39) and (40) can also be derived straightforwardly using symmetry and isotropy without reference to the geometry used in this appendix.

REFERENCES

[1] P. Alexander, A. Grant, and M. Reed, "Iterative detection in code-division multiple-access with error control coding," *Eur. Trans. Telecommun.*, vol. 9, pp. 419–426, Sep./Oct. 1998.

[2] 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.

[3] J. Boutros and G. Caire, "Iterative multi-user joint decoding: Unified framework and asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1772–1793, Jul. 2002.

[4] M. V. Burnashev, C. Schlegel, W. A. Krzymien, and Z. Shi, "Mathematical analysis of one successive cancellation scheme in iterative decoding," *Probl. Inf. Trans.*, vol. 40, no. 4, pp. 3–25, 2004.

[5] D. Divsalar, S. Dolinar, C. R. Jones, and K. Andrews, "Capacity-approaching protograph codes," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 6, pp. 876–888, Aug. 2009.

[6] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 5, pp. 891–907, May 2001.

[7] A. J. Felstrom and K. S. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrix," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 2181–2190, Sep. 1999.

[8] L. Fortnow, "The status of the P versus NP problem," *Commun. ACM*, vol. 52, no. 9, pp. 78–86, Sep. 2009.

[9] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[10] A. Grant and P. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2832–2836, Nov. 1998.

[11] J. Hu, H.-A. Loeliger, J. Dauwels, and F. R. Kschischang, "A general computation rule for lossy summaries/messages with examples from equalization," presented at the 44th Allerton Conf. Commun. Contr. Comp., Monticello, IL, Sep. 2006.

[12] S. B. Korada and N. Macris, "Tight bounds on the capacity of binary input random CDMA systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5590–5613, Nov. 2010.

[13] R. Mori, "Connection between annealed free energy and belief propagation on random factor graph ensembles," in *Proc. IEEE Int. Symp. Inf. Theory*, 2011, pp. 2016–2020.

[14] 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, Feb. 2001.

[15] 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, no. 2, pp. 803–834, Feb. 2011.

[16] R. Learned, A. Willsky, and D. Boroson, "Low complexity optimal joint detection for oversaturated multiple access communications," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 113–123, Jan. 1997.

[17] A. Montanari, B. Prabhakar, and D. Tse, Belief Propagation Based Multi-user Detection 2006, arXiv:cs/0510044v2.

[18] R. J. McEliece, E. R. Rodemich, and J.-F. Cheng, "The turbo decision algorithm," in *Proc. 33rd Allerton Conf. Commun. Contr. Comp.*, Monticello, IL, Oct. 1995, pp. 366–379.

[19] M. Moher, "An iterative multi-user decoder for near-capacity communications," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 870–880, Jul. 1998.

[20] L. Ping, L. Liu, K. Wu, and W. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, Apr. 2006.

[21] M. C. Reed, C. B. Schlegel, P. D. Alexander, and J. A. Asenstorfer, "Iterative multi-user detection for CDMA with FEC: Near-single-user performance," *IEEE Trans. Commun.*, vol. 46, no. 12, pp. 1693–1699, Dec. 1998.

[22] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, Feb. 2001.

[23] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, Feb. 2001.

[24] C. Schlegel and A. Grant, "Polynomial complexity optimal detection of certain multiple-access systems," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2246–2248, Sep. 2000.

[25] C. Sankaran and A. Ephremides, "Solving a class of optimum multiuser detection problems with polynomial complexity," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1958–1961, Sep. 1998.

[26] C. Schlegel, P. Hoeher, L. Perez, and O. Axelsson, "Iterative signal processing in communication," *J. Elect. Comput. Eng.*, vol. 2010, Sep. 2010.
[27] C. Schlegel, "CDMA with partitioned spreading," *IEEE Commun. Lett.*, vol. 11, no. 12, pp. 913–915, Dec. 2007.
[28] C. Schlegel, Z. Shi, and M. Burnashev, "Asymptotically optimal power allocation and code selection for iterative joint detection of coded random CDMA," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4286–4295, Sep. 2006.
[29] Z. Shi and C. Schlegel, "Spreading code construction for CDMA," *IEEE Comm. Lett.*, vol. 7, no. 1, pp. 4–6, Jan. 2003.
[30] Z. Shi and C. Schlegel, "Joint iterative decoding of serially concatenated error control coded CDMA," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1646–1653, Aug. 2001.
[31] Z. Shi and C. Schlegel, "Iterative multi-user detection and error control code decoding in random CDMA," *IEEE Trans. Signal Proc.*, vol. 54, no. 5, pp. 1886–1895, May 2006.
[32] O. Takeshita, "On maximum contention-free interleavers and permutation polynomials over integer rings," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1249–1253, Mar. 2006.
[33] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multi-user detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
[34] K. Takeuchi, T. Tanaka, and T. Kawabata, "Improvement of BP-based CDMA multi-user detection by spatial coupling," presented at the IEEE Int. Symp. Inform. Theory, Saint-Petersburg, Russia, Aug. 2011.
[35] R. Tanner, "On quasi-cyclic repeat-accumulate codes," in *Proc. 37th Allerton Conf. Commun. Contr. Comp.*, Monticello, IL, Sep. 1999, pp. 249–259.
[36] S. t. Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
[37] J. Thorpe, Low-density parity-check (LDPC) codes constructed from protographs, 2003, JPL IPN Progress Report 42-154.
[38] D. Truhachev, C. Schlegel, and L. Krzymien, "A two-stage capacity achieving demodulation/decoding method for random matrix channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 136–146, Jan. 2009.
[39] S. Ulukus and R. Yates, "Optimum multi-user detection is tractable for synchronous CDMA using m-sequences," *IEEE Commun. Lett.*, vol. 2, no. 4, pp. 89–91, Apr. 1998.
[40] S. Ulukus and R. Yates, "Iterative construction of optimum signature sequence sets in synchronous CDMA systems," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1989–1998, Jul. 2001.
[41] S. Verdú, "Computational complexity of optimal multi-user detection," *Algorithmica*, vol. 4, pp. 303–312, 1989.
[42] S. Verdú, "Minimum probability of error for asynchronous Gaussian multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 32, no. 1, pp. 85–96, Jan. 1986.
[43] X. Wang and H. V. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.

Christian Schlegel, biography not available at the time of publication.

Dmitri Truhachev, biography not available at the time of publication.