

Coupling Data Transmission for Multiple-Access Communications

Dmitri Truhachev and Christian Schlegel

Abstract

We consider a signaling format where the information to be communicated from one or multiple transmitters to a receiver is modulated via a superposition of independent data streams. Each data stream is formed by error-correction encoding, constellation mapping, replication and permutation of symbols, and application of signature sequences. The relations between the data bits and modulation symbols transmitted over the channel can be represented by a sparse graph. In the case where the modulated data streams are transmitted with time offsets the receiver observes spatial coupling of the individual graphs into a graph chain enabling efficient demodulation/decoding. We prove that a two-stage demodulation/decoding method, in which iterative demodulation based on symbol estimation and interference cancellation is followed by parallel error correction decoding, achieves capacity on the additive white Gaussian noise (AWGN) channel asymptotically. We compare the performance of the two-stage receiver to the receiver which utilizes hard feedback between the error-correction encoders and the iterative demodulator.

I. INTRODUCTION

Recently, the technique of spatial graph coupling applied to iterative processing on graphs has attracted significant interest in broad areas of communications. The method was first introduced to construct low-density parity-check convolutional codes (LDPCCCs) [1] that exhibit the so-called *threshold saturation* behavior [2][3] which occurs when the limit (threshold) of suboptimal iterative decoding of LDPCCCs asymptotically achieves the optimum maximum a posteriori probability (MAP) decoding threshold [3] of underlying LDPC block codes with the same structure [20]. The idea of constructing graph structures from connected identical copies of a single graph has since been applied to compressed sensing [21][37], image recognition, quantum coding [26], and other fields.

A number of applications of spatial graph coupling to communications over multi-user channels have since been developed. Yedla et. al. studied threshold saturation for a two-user Gaussian multiple access channel (MAC) [7] while Kudekar and Kasai studied it for the binary erasure MAC [8]. In terms of multi-user signaling formats spatially-coupled partitioned-spreading code-division multiple-access (PS-CDMA) was studied in [22], [27]. It was proven that in a noiseless regime the coupled system can achieve arbitrary multi-user loads with the two-stage demodulation/decoding method contrary to the uncoupled PS-CDMA [11] or dense CDMA in general [30]. Sparse CDMA, extensively studied by Guo and Wang [9] in the uncoupled regime, was extended via spatial coupling in [25], [35] where significant threshold improvements over uncoupled CDMA were observed.

This work was partially supported by NSERC Discovery Grant and TELUS Corporation Canada and presented in part in ISIT 2013 [23].

In this work we focus on a generalized multi-user communication signaling format which lends itself to efficient spatial coupling and iterative demodulation. Each user's transmitter chain includes common elements such as error-correction encoding, higher-order constellation mapping, symbol repetition and permutation, and signature sequences. Its complexity and rate can be flexibly adjusted depending on the communication scenario. The system can be operated and studied in both coupled and uncoupled (block) regimes. For specific parameter settings some classes of coupled generalized modulation [10], PS-CDMA [22], or IDMA introduced by Ping et. al. [13], [14], [17] can be derived as particular cases. The iterative demodulation performed at the receiver is a sequence of symbol estimation and interference cancellation iterations and can be seen as a message-passing process on the system graph.

In this paper we investigate spectral efficiency achievable by iterative demodulation/decoding in block and coupled regimes and the resulting gap to the channel capacity. In particular we focus on the two-stage demodulation/decoding schedule where feedback between the error-correction decoders and the demodulator is absent. The two-stage schedule allows for a simplified system design and improved decoding latency compared to either sequential peeling schedules or turbo demodulation/decoding schedules based on soft information exchange between the demodulator and the error-correction decoders. The main contributions of our work can be summarized as follows.

A. Main Contributions

- We rigorously derive a coupled recursion characterizing the evolution of noise-and-interference powers throughout demodulation iterations for both block and coupled versions of the system. We show that for any given number of demodulation iterations there exists a system graph for which the iterative demodulation operates as a message passing on a set of trees. We then quantify the achievable data rate as a function of the signal-to-noise ratio (SNR), constellation symbol alphabet, and the sets of finite system parameters: the number of interfering data streams, repetition factor, and symbol constellation.
- For the case of binary symbols we derive an upper bound on the gap between the achievable rate and the channel capacity. The gap is a function of the SNR and the finite set of system parameters. We show that the asymptotic gap is inverse proportional to the capacity itself and vanishes as the system's SNR increases.
- We derive an expression for the asymptotic scaling of the gap between the achievable sum-rate and the channel capacity.
- We study an alternative demodulation/decoding schedule where hard decision feedback from the error-correction decoders to the demodulator is allowed. We derive a bound on the gap to capacity and compare the achievable rate with that of the two-stage schedule. Finally, we prove that in case of hard decision feedback the gap to capacity vanishes for any fixed SNR as the system load increases.

B. Related Work

A related multi-user communications problem is the problem of maximizing the number of users simultaneously communicating over the MAC channel or achieving high multi-user efficiency [42] measuring the impact of residual

post/detection multi-user interference on a user performance. A number of papers have been dedicated to study asymptotics of joint and individual multiuser detection for classic CDMA. Tse and Verdu investigated the asymptotic performance of joint MAP decoding of random CDMA, and proved that the multiuser efficiency approaches unity as the SNR goes to infinity [43]. Later Tanaka [30], Müller and Gerstacker [44], and Guo and Verdu [45] studied the performance of joint and individual multi-user detection of CDMA via the non-rigorous replica method, and derived the limits for the rates of coded random CDMA with joint and separate demodulation/decoding in terms of multi-user efficiency. In a paper of Takeuchi, Tanaka, and Kawabata [36], developed in parallel to ours [23], coupled sparse CDMA was studied in the asymptotic regime where the number of users, spreading gain and the system size were taken to infinity, while the multi-user load was kept constant. Based on a continuous approximation of the respective coupled recursion for the case of binary symbols the authors showed saturation of the iterative detection threshold of sparse CDMA to the MAP threshold of dense CDMA via spatial graph coupling.

The focus of our paper is on rigorously deriving and numerically demonstrating the rates achievable by the block and coupled sparse multi-user system we consider, where we quantify the resulting gap to capacity explicitly as a function of SNR rather than multi-user efficiency. We study a generalized system in terms of symbol alphabets and derive the achievable rate for finite parameter settings in terms of the number of data streams, repetition factor, symbol constellation alphabet etc. Our approach allows us to study the system performance for various explicitly defined decoding schedules such as the demodulation/decoding schedule with hard feedback, numbers of iterations, and receiver signal-processing options. Our framework leads to graphical interpretations of the achievable rates and the gap to capacity in a form similar to the threshold saturation theorem for LDPC in BEC [3].

It's worth mentioning that the capacity of the multiple-access channel can be achieved using peeling decoding, the fact that follows directly from the chain rule of the mutual information. Various types of the peeling or sequential cancellation decoders has been studied extensively for a variety of systems [5], [15], [16], [18]. The focus of this work, however, is on iterative demodulation using the low-latency two-stage demodulation/decoding schedule and, in particular on the role of spatial graph coupling in approaching channel capacity.

Interleaved-division modulation (IDM), a modulation format utilized in point-to-point transmission proposed by Hoeher et. al. [17], [19], has common elements with the multi-user format we consider here, such as superposition of data streams and bit interleaving, which is also a feature of bit-interleaved coded modulations in general [6]. Contrary to the reception techniques geared towards separation of symbol-synchronous data streams studied by Hoeher and Wo, we investigate cancellation-based low-complexity iterative interference removal which is agnostic to data stream synchronism and power fluctuations and is specially well-suited for coupling.

Another related signaling format are the sparse superposition codes studied by Baron and Joseph [50], later shown to achieve capacity in the block unequal power regime [51]. Sparse superposition codes are based on superimposing data streams based on Gaussian matrices, the technique which is closely related to compressed sensing reconstruction. A coupled version of sparse superposition codes was recently studied in [49] where it was shown that, given the coupled recursion describing the state evolution of the approximate message-passing (AMP)

decoder, the codes achieve channel capacity asymptotically. The AMP decoder of coupled superposition coding does not lend itself to a rigorous derivation of the coupled recursion, however.

C. Structure of the Paper

System model, transmission format and receiver processing are described in Section II. Section III presents our performance analysis, derivation of the main coupled recursions and formulation of the main results. Numerical results, discussion and relation to optimal MAP demodulation/decoding are given in Section IV-A. Finally Section V concludes the paper.

II. SYSTEM MODEL

We consider a communication scenario in which one or more transmitting terminals communicate to a single receiver. The signal observed at the receiver is formed by a superposition of L independently modulated data streams that originate at the terminals and then superimpose at the receiver.

The processes of generating data stream l , $l = 1, 2, \dots, L$ is depicted in Fig. 1. First a binary information sequence $\mathbf{u}_l = u_{1,l}, u_{2,l}, u_{3,l}, \dots, u_{K,l}$ is encoded by a binary error correction code of rate R and mapped to a sequence of signal constellation points denoted by $\mathbf{v}_l = v_{1,l}, v_{2,l}, v_{3,l}, \dots, v_{N,l}$, where $v_{n,l} \in \mathcal{A}_l$. By \mathcal{A}_l we denote a set of signal constellation symbols (which can be a PAM constellation, for example). At the next step, each symbol of the sequence \mathbf{v}_l is replicated M_1 times. After the replication the resulting symbols are permuted using a binary permutation matrix \mathbf{P}_l of size $NM_1 \times NM_1$ specific to the data stream l , $l = 1, 2, \dots, L$.

Finally each symbol $v'_{i,l}$, $i = 1, 2, \dots, NM_1$, $l = 1, 2, \dots, L$ of the permuted sequence is multiplied by a signature sequence $\mathbf{s}_{i,l}$ of length M_2 . We consider pseudo-random energy normalized binary signature sequences, however, Gaussian or other sequences can also be used. The symbols of the signature sequences satisfy

$$\mathbb{E}s_{n_1,l_1,m_1}s_{n_2,l_2,m_2} = \begin{cases} 1 & \text{if } (n_1, l_1, m_1) = (n_2, l_2, m_2) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The resulting sequence $\tilde{\mathbf{v}}_l = \tilde{v}_{1,l}, \tilde{v}_{2,l}, \dots, \tilde{v}_{NM,l}$ is multiplied by an amplitude a_l and transmitted over the channel. The overall symbol repetition factor is given by $M = M_1 M_2$.

Hence, generation of data streams is done in two steps. The first step encompasses individual error-correction encoding and constellation mapping. The second step is the preparation of the data for multi-user processing at the receiver and includes replication, permutation, and application of the signature sequences.

A. Block System

The *block system* scenario is a communication scenario in which all data streams are transmitted over the channel simultaneously. Assuming a real-valued Gaussian noise channel, the received signal is given by

$$\mathbf{y} = \sum_{l=1}^L a_l \tilde{\mathbf{v}}_l + \mathbf{n}$$

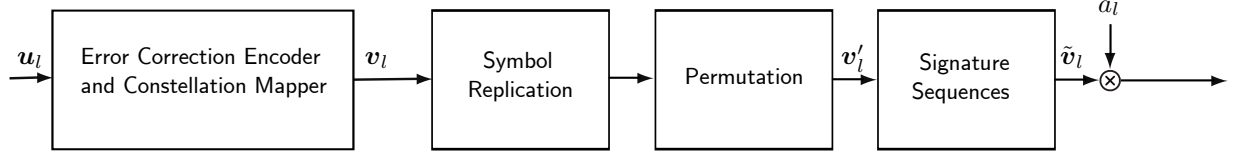


Fig. 1. Generation of the l th data stream at the transmitter side.

where \mathbf{n} is a vector of the iid Gaussian noise samples with zero mean and variance σ^2 . The block system can be represented by the graph shown on the left side of Fig. 2. The top part of the figure demonstrates three individual data stream graphs while the resulting graph observed at the receiver is shown at the bottom. Variable nodes representing symbols $v_{j,l}$ are depicted by circles and channel nodes representing received values y_t are depicted by hexagons.

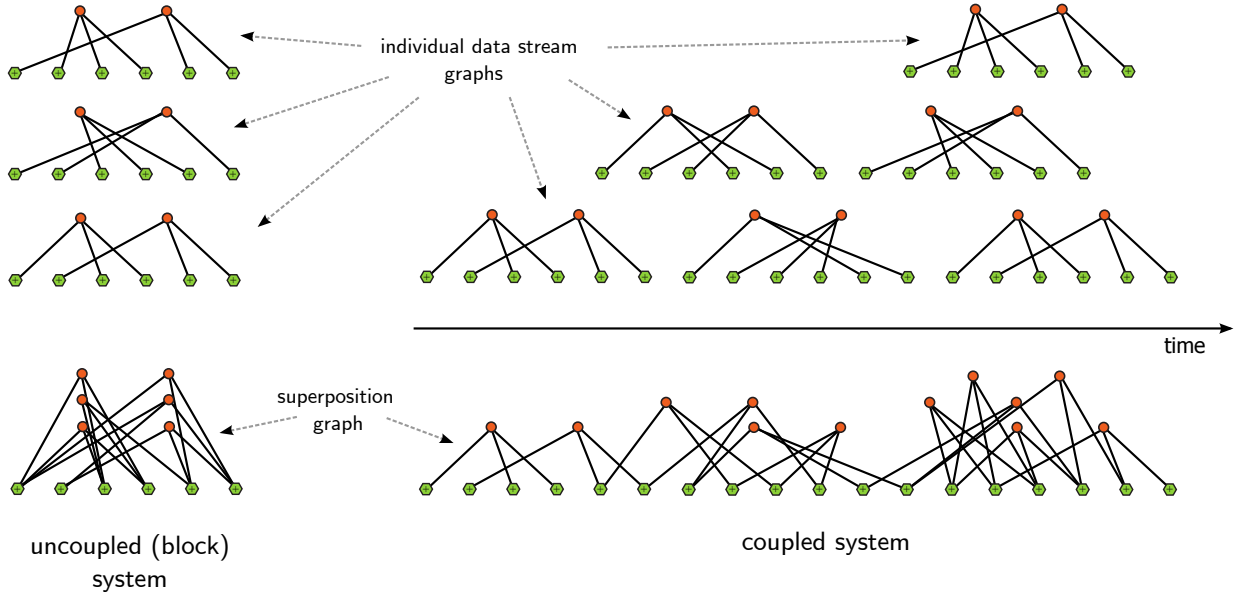


Fig. 2. Graph representation of the block and spatially coupled systems.

B. Coupled System

A spatially graph-coupled system for the presented modulation format arises naturally when the modulated data streams are transmitted over the channel with time offsets. This scenario is illustrated in the right side of Fig. 2. The graphs representing transmitted data streams couple through the channel nodes that correspond to the times at

which the data stream symbols are transmitted over the common channel. Contrary to the block system situation, where the data streams superimpose in groups of L and each group is represented by a graph at the receiver, the coupled system is represented by a single large graph also called the coupled graph chain. This draws parallels to coupled graph chains appearing in other applications such as SC-LDPC codes [1]. We will prove that iterative demodulation and decoding on this graph chain will give us substantially improved achievable rates.

The hexagonal channel nodes represent received symbols y_t , $t = 1, 2, 3, \dots$. The symbols $\tilde{v}_{j,l}$ transmitted at the same time t are connected to the same channel node. For the sake of graph representation we assume that the transmission is “symbol-synchronous”. It is important to note, however, that the receiver processing which is based on symbol estimation and interference cancellation is not based on this assumption nor is the iterative demodulation analysis.

We denote the time offset (delay) of data stream l by τ_l , $l = 1, 2, \dots, \hat{L}$ where \hat{L} is the total number of streams transmitted. Then the components of the received signal vector \mathbf{y} are given by

$$y_t = \sum_{l=1}^{\hat{L}} a_l \tilde{v}_{t-\tau_l, l} + n_t ; \quad t = 1, 2, 3, \dots$$

In order to study the performance of the coupled system we consider a regularized delay structure defined as follows. We introduce W , the coupling parameter and assume that each transmitted data stream consists of $2W + 1$ subsections of length $N_w = MN/(2W + 1)$. In addition the data streams are transmitted in groups of size $L_w = L/(2W + 1)$ streams. At time $t = 1$ the first group of data streams is transmitted, i.e., the first L_w time delays are given by $\tau_1 = \tau_2 = \dots = \tau_{L_w} = 0$. At time $N_w + 1$ the second group of L_w data streams is transmitted with the corresponding time delays given by $\tau_{L_w+1} = \tau_{L_w+2} = \dots = \tau_{2L_w} = N_w$ and so on. At every time $kN_w + 1$, $k = 0, 1, 2, \dots$ exactly L_w data streams are transmitted. In the remainder of the paper we use the terminology “coupled system” to address the system with regularized delays. The regularized delay structure is assumed for analysis purposes and to demonstrate the capacity-achieving properties of the coupled system and is not required for receiver signal processing. Poisson delays produce results close to those with regularized delays, however, the study of delay distributions and their impact is beyond the scope of this paper.

While the degrees of nodes in the coupled graph chain are predetermined, the connections between the nodes (edges) are not. We can define a graph ensemble describing the coupled system by considering random permutation matrices \mathbf{P}_l in the generation of the individual data streams. This representation is given in Appendix A along with a matrix representation of the coupled chain. In addition, signature sequences are also chosen randomly. The right side of Fig. 2 illustrates an example in which six data streams transmitted with delays $\tau_1 = 1$, $\tau_2 = 5$, $\tau_3 = 7$, $\tau_4 = 11$, $\tau_5 = 12$, $\tau_6 = 13$.

C. System Load and Rate

We define the load α of the block system by taking the ratio of the number of simultaneously transmitted data streams L to the repetition factor M , $\alpha = L/M$. Assume that all data streams have the same power normalized to $1/L$ and constellation alphabets \mathcal{A}_l , $l = 1, 2, 3, \dots$ that consists of 2^B signal points, where B is the modulation

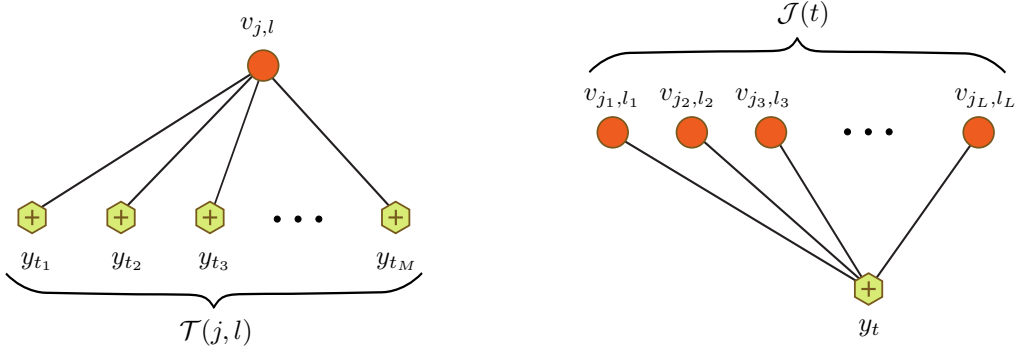


Fig. 3. The relation between the variable and channel nodes and the sets of indices involved.

index. The total system SNR measured as a ratio of the total transmit power over the noise power then equals $1/\sigma^2$ while the total transmit data rate equals $RBL/M = \alpha BR$ information bits per channel use. The signal-to-noise ratio per information bit equals $E_b/N_0 = 1/(2\alpha BR\sigma^2)$.

In the coupled case the system load in terms of the number of data streams per repetition factor M for the first NM time instances $t = 1, 2, \dots, NM$ is smaller than α . Assuming an infinite number of data streams in the system, $\hat{L} = \infty$ and constant system load α for $t > N$ we obtain the same rate and SNR as in the block case. The initial reduced load leads to an initial rate loss that vanishes as $\hat{L} \rightarrow \infty$. This is typical for all coupled systems including the one we consider.

D. Iterative Demodulation

The received signal vector $\mathbf{y} = (y_1, y_2, y_3, \dots)$ contains $M = M_1 M_2$ replicas of each transmitted symbol $v_{j,l}$, $j = 1, 2, \dots, l = 1, 2, \dots, L$. Let $\mathcal{T}(j, l) = \{t_1, t_2, \dots, t_M\}$ denote the set of time indices t such that the received vector component y_t , $t \in \mathcal{T}(j, l)$ contains $v_{j,l}$. Let's denote the corresponding symbols of the signature sequences that multiply $v_{j,l}$ during the modulation process by $s_{j,l}^{(t)}$, where $t \in \mathcal{T}(j, l)$. Moreover, by $\mathcal{J}(t)$ we denote a set of all index pairs (j, l) such that $v_{j,l}$ is included in y_t (see Fig. 3). The cardinality of the set $\mathcal{T}(j, l)$ is always equal to M . The cardinality of $\mathcal{J}(t)$ equals L for a block system and varies depending on t for the coupled system (see Section II-C).

Demodulation iteration i starts with the interference cancellation step computing signals $y_{t \rightarrow (j,l)}^{(i)}$ for $j = 1, 2, \dots, N$, $l = 1, 2, \dots, L$, $t \in \mathcal{T}(j, l)$

$$y_{t \rightarrow (j,l)}^{(i)} = y_t - \sum_{\substack{(j', l') \in \mathcal{J}(t) \\ \text{s.t. } (j', l') \neq (j, l)}} a_{l'} s_{j', l'}^{(t)} \hat{v}_{(j', l') \rightarrow t}^{(i-1)}, \quad (2)$$

$$= a_l s_{j,l}^{(t)} v_{j,l} + \xi_{j,l,t} + n_t, \quad (3)$$

where

$$\xi_{j,l,t} \stackrel{\text{def}}{=} \sum_{\substack{(j',l') \in \mathcal{J}(t) \\ \text{s.t. } (j',l') \neq (j,l)}} a_{l'} s_{j',l'}^{(t)} \left(v_{j',l'} - \hat{v}_{(j',l') \rightarrow t}^{(i-1)} \right) \quad (4)$$

is the residual interference component and the estimates $\hat{v}_{(j',l') \rightarrow t}^{(i-1)}$ have been generated for all data symbols $v_{j',l'}$, $j' = 1, 2, \dots$, $l' = 1, 2, \dots, L$ at the previous $(i-1)$ th iteration.¹

For each $\tau \in \mathcal{T}(j, l)$ we form a vector $\mathbf{y}_{j,l,\tau}^{(i)}$ out of estimates $\mathbf{y}_{(j,l) \rightarrow t}^{(i)}$ where $t \neq \tau$, $t \in \mathcal{T}(j, l)$. Since each signal $\mathbf{y}_{j,l,\tau}^{(i)}$ contains $v_{j,l}$ and some remaining interference we can write

$$\mathbf{y}_{j,l,\tau}^{(i)} = \mathbf{h}_\tau v_{j,l} + \boldsymbol{\xi}_{j,l,\tau} \quad \tau \in \mathcal{T}(j, l) \quad (5)$$

where the vector \mathbf{h}_τ consists of elements $a_l s^{(t)}$, $t \neq \tau$, $t \in \mathcal{T}(j, l)$ and $\boldsymbol{\xi}_{j,l,\tau}$ is the noise-and-interference vector with components given in (4). Let us denote the covariance matrix of the noise-and-interference vector $\boldsymbol{\xi}_{j,l,\tau}$ by $\mathbf{R}_{j,l,\tau}$. We now perform minimum mean-squared error (MMSE) filtering of $\mathbf{y}_{j,l,\tau}^{(i)}$ to form an SNR-optimal linear estimate of $v_{j,l}$, given by

$$z_{j,l,\tau} = \mathbf{w}_{j,l,\tau}^T \mathbf{y}_{j,l,\tau}^{(i)} \quad (6)$$

where

$$\mathbf{w}_{j,l,\tau}^T = (\mathbf{I} + \mathbf{h}_\tau^* \mathbf{R}_{j,l,\tau}^{-1} \mathbf{h}_\tau)^{-1} \mathbf{h}_\tau^* \mathbf{R}_{j,l,\tau}^{-1} \quad t \in \mathcal{T}(j, l)$$

minimizes $\|\mathbf{w}_{j,l,\tau}^T \mathbf{y}_{j,l,\tau}^{(i)} - v_{j,l}\|^2$.

Since we know that $v_{j,l}$ belongs to the symbol alphabet \mathcal{A}_l we can form a conditional expectation estimate

$$\hat{v}_{(j,l) \rightarrow \tau}^{(i)} = \mathbb{E}(v_{j,l} | z_{j,l,\tau}) \quad \tau \in \mathcal{T}(j, l) . \quad (7)$$

For the case of equiprobable binary symbols $v_{j,l} \in \{1, -1\}$ (2-PAM modulation) $\hat{v}_{(j,l) \rightarrow \tau}^{(i)}$ takes the form $\hat{v}_{(j,l) \rightarrow \tau}^{(i)} = \tanh(z_{j,l,\tau} \gamma_{j,l,\tau})$, where $\gamma_{j,l,\tau}$ is the SNR of the estimate $z_{j,l,\tau}$. We will compute this SNR in the next section.

This rules correspond to message passing on a system's graph. Each variable node (j, l) passes M estimates $\hat{v}_{(j,l) \rightarrow t}^{(i)}$ to the connected channel nodes $t \in \mathcal{T}(j, l)$ while each channel node passes $z_{j,l,t}^{(i)}$ where $(j', l') \in \mathcal{J}(t)$ to the connected variable nodes.

E. Demodulation and Decoding Schedules

The demodulation and error-correction decoding at the receiver can be scheduled in a number of ways. We consider two demodulation/decoding schedules which are of relatively low complexity. An alternative schedule of significantly higher complexity is the “turbo-schedule” with soft feedback between the error-correction decoders and the iterative demodulator.

Two-Stage Schedule

¹At the first iteration $i = 1$ we use $\hat{v}_{(j',l') \rightarrow t}^{(0)} = 0$ (no interference cancellation is attempted).

At the first stage I iterative demodulation iterations are performed as described above. The second stage comprises decoding of the forward error correction codes used to encode the information vectors \mathbf{u}_l , $l = 1, 2, \dots, \hat{L}$.

Hard Decoding Feedback and Cancellation Schedule

Demodulation iterations are performed until one or more data stream's SNR rise above the decoding threshold of the error-correction code. Once this happens the code is decoded and the decoded bits are given as hard feedback to the demodulator. The feedback bits are then utilized to modulate the respective data streams that are subtracted (cancelled) from the received signal. The demodulation process continues for the remaining undecoded data streams until one of these would have a sufficiently high SNR to be decoded, and so on.

III. PERFORMANCE ANALYSIS

At every iteration the demodulator attempts to reduce the amount of inter-stream interference. We take the approach to track the evolution of the signal-to-noise and interference ratio (SINR) throughout the decoding iterations [12], [14], [24]. The central part of the performance analysis is dedicated to the derivation of a noise-and-interference power evolution equation to track the evolution of the noise-and-interference power throughout the demodulation iterations.

We start with an assessment of the performance of the block system for which the evolution of the noise-and-interference power is independent of the time index t . After that we derive a characteristic equation that determines when the iterative demodulation process for the block system converges to a nearly interference-free state.

We then proceed with deriving the noise-and-interference power evolution for the coupled system and make a connection to the characteristic equation of the block system and its fixed points. Based on this connection we obtain the expressions for the achievable rates and the gap to capacity. Contrary to related work we make our derivation for finite parameter values of M, L, B and then derive the asymptotically achievable rates as particular cases.

A. Block System

We focus on the case of equal power data streams where the amplitudes are given by $a_l = a$, $l = 1, 2, \dots, L$. Without loss of generality we normalize data stream powers so that $a = \sqrt{1/L}$ and vary the noise power σ^2 . In this case the total system SNR per receive symbol equals $1/\sigma^2$ (see Section II-C). We then proceed with considering the evolution of the noise-and-interference power (4).

Let us denote the noise-and-interference power at iteration i by

$$x_i \stackrel{\text{def}}{=} \mathbb{E} \left(\sum_{\substack{(j',l') \in \mathcal{T}(t) \\ \text{s.t. } (j',l') \neq (j,l)}} \sqrt{\frac{1}{L}} s_{j',l'}^{(t)} (v_{j',l'} - \hat{v}_{(j',l') \rightarrow t}^{(i-1)}) + n_\tau \right)^2 = \sum_{\substack{(j',l') \in \mathcal{T}(t) \\ \text{s.t. } (j',l') \neq (j,l)}} \frac{1}{L} \mathbb{E} \left(v_{j',l'} - \hat{v}_{(j',l') \rightarrow t}^{(i-1)} \right)^2 + \sigma^2 \quad (8)$$

$$= \frac{1}{L} (L-1) \mu_{i-1} + \sigma^2 \quad (9)$$

where the mean-square error (MSE) of the symbols at iteration $i-1$ is given by μ_{i-1} , i.e.

$$\mu_{i-1} \stackrel{\text{def}}{=} \mathbb{E} \left(v_{j',l'} - \hat{v}_{(j',l') \rightarrow t}^{(i-1)} \right)^2 = \mathbb{E} (v_{j',l'} - \mathbb{E}(v_{j,l} | z_{j,l,t}))^2. \quad (10)$$

The expectation is taken over the respective channel node index t , variable node index (j',l') , and the system realizations. When breaking the variance of the noise-and-interference power in (4) into the sum of variances of the components $v_{j',l'} - \hat{v}_{(j',l') \rightarrow t}^{(i-1)}$ as in (8) we assume that these are independent. This can be guaranteed by making sure that the system graph has large girth, a property not required for practical implementation but important for an exact analysis.

Coming back to the graph representation of the system, if we connect the variable node $v_{j,l}$ to the channel nodes y_t where $t \in \mathcal{T}(j,l)$ and then connect each channel node y_t to $v_{j',l'}$ such that $v_{j,l} \in \mathcal{J}(t)$ we obtain a sub-graph describing one demodulation iteration for symbol $v_{j,l}$. If we expand it further by connecting the nodes $v_{j',l'}$ to their channel nodes $y_\tau \in \mathcal{T}(j,l)$ and so on we can obtain a sub-graph describing i demodulation iterations for symbol $v_{j,l}$. If the girth of the entire system's graph is at least $4i+1$ any demodulation sub-graph for any symbols $v_{j,l}$ will not contain repeated nodes and is, therefore, a tree. This description and conditions closely resemble those for LDPC block [34] and convolutional code analysis [40]. The following lemma elaborates on this connection and proves that a system graph with arbitrary large girth can be constructed for both block and coupled systems if the data stream length is chosen large enough.

Lemma 1. *For any number of iterations i , and parameters L, M, B, W , both coupled and block systems can be constructed such that the corresponding graph has girth larger than $4i$.*

Proof. See Appendix A. □

The next step is to express the MSE μ_{i-1} in terms of the noise-and-interference power x_{i-1} at iteration $i-1$. The Central Limit Theorem implies that the noise-and-interference vector $\boldsymbol{\xi}_{j,l,t}$ in (5) converges to a Gaussian random vector with independent zero-mean components and covariance matrix $\mathbf{R}_{j,l,t} = \text{diag}(\sigma_{t_1}^2, \sigma_{t_2}^2, \dots, \sigma_{t_M}^2)$ (t excluded) as L increases, where σ_τ^2 denotes the variance of the individual vector components $\xi_{j,l,t}$, $\tau \in \mathcal{T}(j,l)$, $\tau \neq t$. Typical values of L are around 100 or higher indicating that the Gaussian approximation is very accurate. The Lindeberg condition is satisfied for both coupled and block equal power systems. The resulting SNR of $z_{j,l,t}$ in (7) equals

$$\gamma_{j,l,t} = \mathbf{h}_t^* \mathbf{R}_{j,l,t}^{-1} \mathbf{h}_t = \frac{1}{L} \sum_{\tau \in \mathcal{T}(j,l)} \frac{1}{\sigma_\tau^2}. \quad (11)$$

The MSE expression (10) can be computed as MSE of symbol estimation in additive Gaussian noise and is a function of the symbol constellation and the SNR. We denote it by

$$\mu_{i-1} = g_{\text{mse}}(\gamma_{j,l}) . \quad (12)$$

The MSE for the binary symbol alphabet case takes the form

$$g_2(\gamma) = \mathbb{E} \left[(1 - \tanh(\gamma + \xi\sqrt{\gamma}))^2 \right] ; \quad \xi \sim \mathcal{N}(0, 1) . \quad (13)$$

A series expansion and a number of bounds and approximations for (13) are derived in [12].

In the case of the block system all noise-and-interference components (4) have the same power and, therefore, $\sigma_\tau^2 = x_{i-1}$, $\tau \in \mathcal{T}(j, l)$ for all k forming the diagonal covariance matrix $\mathbf{R}_{j,l} = \frac{1}{x_{i-1}} \mathbf{I}_M$ where \mathbf{I}_M is the $M \times M$ identity matrix. Hence (8), (13), (11) lead to the characteristic equation of the block system

$$x_i = \frac{L-1}{L} g_{\text{mse}} \left(\frac{M-1}{Lx_{i-1}} \right) + \sigma^2 . \quad (14)$$

Using a variable exchange we obtain a form of the characteristic equation which will be central to our analysis, viz.,

$$x'_i = g_{\text{mse}} \left(\frac{1}{\alpha' x'_{i-1}} \right) + \sigma'^2 \quad (15)$$

where

$$x'_i = x_i \frac{L}{L-1}, \quad \sigma'^2 = \sigma^2 \frac{L}{L-1}, \quad \alpha' = \frac{L-1}{M-1} . \quad (16)$$

This representation allows us to carry out a performance analysis of the system for finite values of L and M where the data stream size N is going to infinity for the purpose of allowing any number i of demodulation iterations to be considered as message passing on respective trees (relying on Lemma 1).

Equation (15) is initialized with

$$x'_0 = g_{\text{mse}}(0) + \sigma'^2$$

which constitutes the noise-and-interference power before the demodulation iterations start. Due to signal power normalization (see Section III-A) $g_{\text{mse}}(0) = 1$.

Depending on the value of σ'^2 (15) can have a different number of fixed points (see Section III-D) but it always converges to the largest fixed point that is less or equal than the initial variance $1 + \sigma'^2$.

B. Coupled System

We consider the regularized delay structure introduced in Section II-B where each data stream is divided into $2W + 1$ subsections of length N_w . Any two interfering data streams share a number of interfering subsections transmitted at the same starting times. A superposition of the transmitted data streams forms the received sequence that again consists of subsections of length N_w . We can, therefore, index the subsections of the received sequence by $t = 1, 2, 3, \dots$. For the purpose of system analysis we assume that time is given in terms of subsections. For the case of coupled systems the noise-and-interference power and MSE values are not just dependent on the demodulation iteration i but also on the time index t of the respective receive subsection.

In order to assess the noise-and-interference power x_i^t for subsection t at iteration i we look again at equation (8). All variable nodes connected to a channel node that belongs to receive subsection t are within subsections indexed by $\tau_1 = t - W, t - W + 1, \dots, t + W$. In addition, on average $(L - 1)/(2W + 1)$ variable nodes (j', l') connected to a single check node belong to each of the $2W + 1$ subsections. Hence

$$x_i^t \stackrel{\text{def}}{=} \sum_{\substack{(j', l') \in \mathcal{J}(t') \\ \text{s.t. } (j', l') \neq (j, l)}} \frac{1}{L} \mathbb{E} \left(v_{j', l'} - \hat{v}_{j', l', \tau}^{(i-1)} \right)^2 + \sigma^2 = \frac{L - 1}{L(2W + 1)} \sum_{\tau_1 = t - W}^{t + W} \mu_{i-1}^{\tau_1} + \sigma^2. \quad (17)$$

In turn, to compute the MSE $\mu_{i-1}^{\tau_1}$ for subsection τ_1 at iteration $i - 1$ we notice that equation (10) implies computation of the MSE for symbols of a data stream transmitted at time $\tau - W$ in terms of subsections. The $M - 1$ replicas of each data symbol in this data stream experience SNR values $1/(Lx_{i-1}^{\tau_2})$, $\tau_2 = \tau_1 - W, \dots, \tau_2 + W$. Again, on average each SNR is experienced by $(M - 1)/(2W + 1)$ symbols and we obtain

$$\mu_{i-1}^{\tau_1} = g_{\text{mse}} \left(\frac{M - 1}{2W + 1} \sum_{\tau_2 = \tau_1 - W}^{\tau_1 + W} \frac{1}{Lx_{i-1}^{\tau_2}} \right). \quad (18)$$

Combining (17) and (18) together, applying a variable substitution (16), we obtain the characteristic equation of the coupled system

$$x_i^t = \frac{1}{2W + 1} \sum_{\tau_1 = -W}^W g_{\text{mse}} \left(\frac{1}{\alpha'} \frac{1}{2W + 1} \sum_{\tau_2 = -W}^W \frac{1}{x_{i-1}^{t + \tau_1 + \tau_2}} \right) + \sigma'^2. \quad (19)$$

We assume that the transmission starts at time $t = 1$. At every time instant t , a new set consisting of L_w data streams is transmitted. As a result, the initial conditions for recursion (19) can be formulated as

$$x_0^t = 0 \quad t \leq 0, \quad (20)$$

$$x_0^t = \frac{t}{2W + 1} g_{\text{mse}}(0) + \sigma'^2 = \frac{t}{2W + 1} + \sigma'^2; \quad t \in [1, 2W + 1] \quad (21)$$

$$x_0^t = g_{\text{mse}}(0) + \sigma'^2 = 1 + \sigma'^2 \quad t > 2W + 1. \quad (22)$$

where (21) indicates the SINR of the data streams equals 0 before the demodulation starts.

C. Demodulation/Decoding Schedules

Two-Stage Schedule

For block systems the residual SINR per symbol after I demodulation iterations equals $\frac{1}{\alpha' x_i'}$. Assuming that each data stream is encoded with an identical error correction code with SNR threshold θ we obtain asymptotically error-free performance if

$$\frac{1}{\alpha' x_i'} \geq \theta.$$

For coupled system the SINR for subsection t after I iterations equals

$$\gamma_I^t = \frac{1}{\alpha'(2W + 1)} \sum_{j = -W}^W \frac{1}{x_I^{t+j}}$$

from (19). Hence the subsequent error correction decoding performed as the second stage of the two-stage receiver is successful if

$$\gamma_I^t \geq \theta \quad \text{for } t \geq 1. \quad (23)$$

Hard Decoding Feedback and Cancellation Schedule

For hard decision decoding and cancellation the receiver performs the decoding of any data stream for which the SINR exceeds the error-correction code threshold θ and cancels the impact of that stream from the received signal. This means that any data streams with SINR above θ will no longer contribute to the noise-and-interference power in (15) and (19). Let us define the MSE function including the error correction decoder as

$$g_{\text{mse,ecc}}(\gamma) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } \gamma \geq \theta \\ g_{\text{mse}}(\gamma), & \text{otherwise.} \end{cases}$$

As soon as the SINR reaches the error correction code threshold the function sets to 0 accounting for the fact that the respective data stream is eliminated in the cancellation process. The characteristic equations for the coupled and block systems are modified accordingly and we obtain

$$x'_i = g_{\text{mse,ecc}}\left(\frac{1}{\alpha' x'_{i-1}}\right) + \sigma'^2 \quad (24)$$

and

$$x_i^t = \frac{1}{2W+1} \sum_{\tau_1=-W}^W g_{\text{mse,ecc}}\left(\frac{1}{\alpha' \frac{1}{2W+1} \sum_{\tau_2=-W}^W \frac{1}{x_{i-1}^{t+\tau_1+\tau_2}}}\right) + \sigma'^2. \quad (25)$$

The systems converge iff x'_i (or x_i^t for every $t \geq 1$ respectively) converge to σ^2 after a number of iterations indicating that the residual of the received signal contains just noise while all data streams have been decoded.

D. Fixed Points of the Characteristic Equation

Since the MSE function $g_{\text{mse}}(\cdot)$ is differentiable and strictly decreasing [12], [46] the values x'_i in the recursive characteristic equation (15) are strictly decreasing with demodulation iterations i . Depending of the modulation symbol alphabet \mathcal{A} , σ'^2 , and α' (15) may have different numbers of fixed points. For each 2^B -PAM signal alphabet there exists a limiting threshold in terms of the noise power $\sigma_s^2(B)$ such that for $\sigma'^2 > \sigma_s^2(B)$ the characteristic equation (15) has a single fixed point x_s which depends on the system load α' and the noise power σ'^2 .

For $\sigma'^2 > \sigma_s^2(B)$ (15) can have one, two or three fixed points. We can define a system load value $\alpha_s(B, \sigma'^2)$ such that for $\alpha' < \alpha_s(B, \sigma'^2)$ (15) has only one fixed point, two for $\alpha' = \alpha_s(B, \sigma'^2)$, and three fixed points for $\alpha' > \alpha_s(B, \sigma'^2)$. Fig. 4 shows $\alpha_s(B, \sigma'^2)$ as a function of the SNR $1/\sigma'^2$ given in dB. The blue curve corresponds to the case $B = 1$ (2-PAM), the magenta curve is for $B = 2$ (4-PAM), and the black curve is for $B = 3$ (8-PAM) symbol alphabets. The threshold SNR levels $1/\sigma_s^2(B)$ in decibels are shown by the vertical lines.

Consider now the 2-PAM case and denote the three fixed points of the system by $x^{(1)} < x^{(2)} < x^{(3)}$. Lemma 2 states upper and lower bounds on the fixed points $x^{(1)}$ and $x^{(3)}$. The proposed bounds will be used in the proof

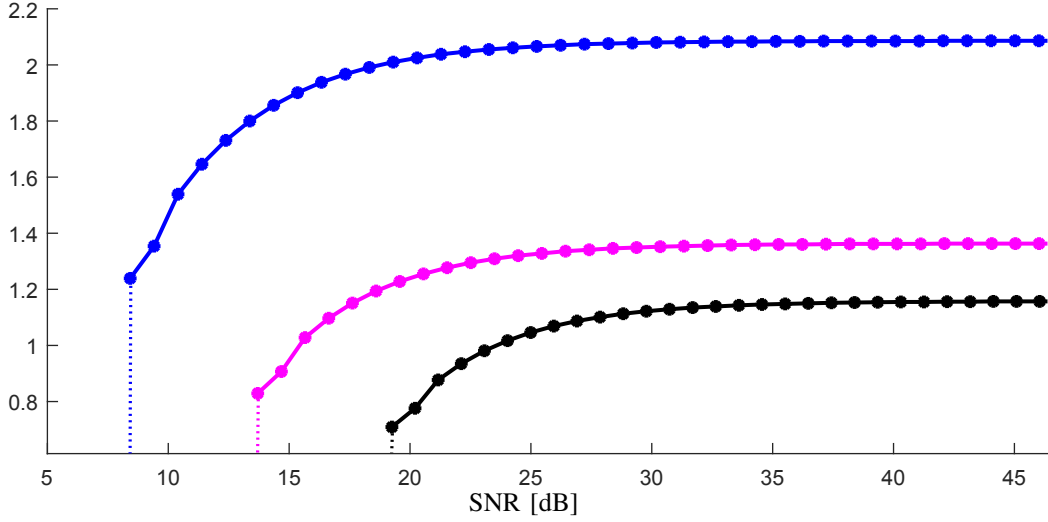


Fig. 4. System loads α_s for which the characteristic equation (15) has exactly two fixed points 2-PAM (blue), 4-PAM (magenta), 8-PAM (black) symbol alphabets are shown as a function of the total system SNR $1/\sigma^2$ in dB. Limiting SNRs $1/\sigma_s^2(B)$ for $B = 1, 2, 3$ are given by the blue, magenta, and black dotted vertical lines respectively.

of the main result stated in the next section. We note that in case (15) has a single fixed point x_s it satisfies Lemma 2 (a).

Lemma 2. Consider $\sigma'^2 \leq 1$. Then the following statements are satisfied:

(a) for $\alpha' \in [0, \mathcal{C}(\sigma'^2)]$

$$\sigma'^2 \leq x^{(1)} \leq (1 + e^{-1/\sigma'})\sigma'^2 \leq 2\sigma'^2, \quad (26)$$

(b) for $\alpha' \in [4, \mathcal{C}(\sigma'^2)]$

$$(1 + \sigma'^2)^2 - \frac{3}{\alpha'} \leq (x^{(3)})^2 \leq (1 + \sigma'^2)^2, \quad (27)$$

(c) for $\alpha' \in [4, \mathcal{C}(\sigma'^2)]$

$$\frac{1}{2}(1 + \sigma'^2) \leq 1 + \sigma'^2 - \frac{1}{\alpha'(1 + \sigma'^2)} - \frac{2}{\alpha'^2(1 + \sigma'^2)^3} \leq x^{(3)} \leq 1 + \sigma'^2 \quad (28)$$

where

$$\mathcal{C}(\sigma'^2) = \frac{1}{2} \log_2 \left(1 + \frac{1}{\sigma'^2} \right). \quad (29)$$

is the capacity of the AWGN channel with SNR $1/\sigma'^2$.

Proof. See Appendix B. □

In contrast to the block system (15), the fixed points \mathbf{x}^* of the coupled system (19) are vectors. Again we notice that due to the fact that the MSE function is decreasing the components x_i^t of the coupled recursion vectors are decreasing as well for each $t = 1, 2, 3, \dots$, and $\mathbf{x}_t^* = \lim_{i \rightarrow \infty} \mathbf{x}_i^t$, $t = 1, 2, 3, \dots$.

We show (see Appendix C) that the components of the vector fixed points x^* are concentrated around the fixed points $x^{(1)}$ and $x^{(3)}$ of the block system characteristic equation (15) depending on the parameters α', σ'^2 . The smallest fixed point $x^{(1)}$ is close to the value of the noise power σ^2 (Lemma 2 (a)). In the next section we will determine the rates achievable by block and coupled system and state the main result that the system's performance approaches channel capacity.

E. Achievable Rate and Capacity

For loads $\alpha' > \alpha'_s$ the block system's noise-and-interference power converges to the single fixed point of the characteristic equation (15) x_s and the corresponding SINR to $(\alpha' x_s)^{-1}$. In order to determine the maximum sum rate achieved by the entire system we assume that error-correction codes used to encode data streams are optimal with respect to the SINR $(\alpha' x_s)^{-1}$, the symbol constellation alphabet \mathcal{A} and transmission over the AWGN. The total communication rate (sum-rate) achievable by the block system in this case equals

$$\mathcal{R}_u(\alpha, \sigma^2) = \frac{L}{M} \mathcal{C}_{\mathcal{A}} \left(\frac{1}{\alpha' x_s} \right) = \alpha \mathcal{C}_{\mathcal{A}} \left(\frac{1}{\alpha' x_s} \right) \quad (30)$$

where $\mathcal{C}_{\mathcal{A}}(\gamma)$ denotes the capacity of the AWGN channel with input alphabet \mathcal{A} and SNR γ and x_s is the single fixed point of (15). The factor $\alpha = \frac{L}{M}$ accounts for the number of data streams L in the system and the repetition factor M . We recall that the system (19) operates with total signal power 1 and noise power σ^2 . Therefore, the corresponding capacity of the (real-valued) AWGN channel for these parameters equals $\mathcal{C}(\sigma^2)$ (where $\sigma^2 = \sigma'^2(L-1)/L$).

For the case when the characteristic equation has three fixed points, i.e., for $\sigma' < \sigma'_s$ the block system noise-and-interference power converges to the largest fixed point $x^{(3)}$ (15) and the achievable rate of the block system equals

$$\mathcal{R}_u(\alpha, \sigma^2) = \alpha \mathcal{C}_{\mathcal{A}} \left(\frac{1}{\alpha' x^{(3)}} \right), \quad (31)$$

The coupled system, however, converges to the smallest fixed point $x^{(1)}$ of the characteristic equation (15) in case the system load α' does not exceed a limiting load $\bar{\alpha} = \bar{\alpha}(\sigma'^2)$ determined via the system's potential function discussed in Appendix C. For higher loads $\alpha' > \bar{\alpha}$ the coupled system converges to the largest fixed point $x^{(3)}$ of (15) just like the block system. Hence the rates achievable by the coupled system are

$$\mathcal{R}_{\text{coup}}(\alpha, \sigma^2) = \begin{cases} \alpha \mathcal{C}_{\mathcal{A}} \left(\frac{1}{\alpha' x^{(1)}} \right), & \text{if } \alpha \leq \bar{\alpha}, W \geq \bar{W} \\ \alpha \mathcal{C}_{\mathcal{A}} \left(\frac{1}{\alpha' x^{(3)}} \right), & \text{otherwise} \end{cases} \quad (32)$$

for sufficiently large window size \bar{W} (see Appendix C). We are now ready to state the main result of the paper for the case of 2-PAM modulation ($B = 1$):

Theorem 1. *There exists a $\bar{W} > 0$ such that for any $W > \bar{W}$*

$$\mathcal{C}(\sigma^2) - \mathcal{R}(\bar{\alpha}, \sigma^2) \leq \mathcal{G}_{\text{asymptotic}}(\sigma^2) + \mathcal{G}_{\text{finite}}(L, M, \sigma^2) \quad (33)$$

where

$$\mathcal{G}_{\text{asymptotic}}(\sigma^2) = \frac{1.081}{\mathcal{C}(\sigma^2)} \quad \mathcal{G}_{\text{finite}}(L, M, \alpha, \sigma^2) = \frac{1}{2} \log_2 \left(\frac{\sigma^2 + 1}{\sigma^2 + 1 - \frac{1}{L}} \right) + \frac{\mathcal{C}(\sigma^2) - 1}{M - 1} \quad (34)$$

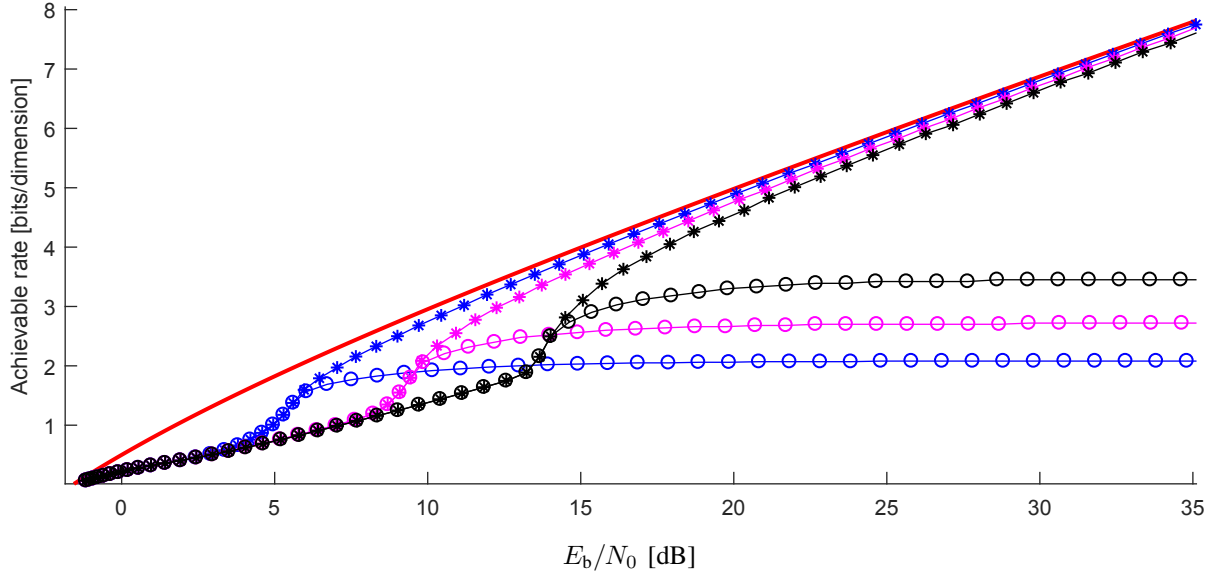


Fig. 5. Rates achievable by coupled (stars) and block (circles) systems in comparison to channel capacity (solid red curve). The performances of the systems with 2-PAM, 4-PAM, and 8-PAM symbol alphabets are shown by blue, magenta, and black curves respectively.

for $\mathcal{C}(\sigma^2) \geq 4.26$, and, therefore,

$$\lim_{L, M \rightarrow \infty} \mathcal{G}_{finite}(L, M, \alpha, \sigma^2) = 0, \quad (35)$$

$$\lim_{\sigma^2 \rightarrow 0} \mathcal{G}_{asymptotic}(\sigma^2) = 0. \quad (36)$$

The proof is given in Appendix C. The expression in Theorem 1² demonstrates the asymptotic behavior of the achievable rate and shows that the gap between the achievable rate and the AWGN channel capacity tends to zero as SNR $1/\sigma^2$ increases. The next theorem is dedicated to the asymptotic gap between the achievable rate and capacity.

Theorem 2.

$$\mathcal{C}(\sigma^2) - \max_{\alpha} \mathcal{R}_{coup}(\alpha, \sigma^2) = \frac{1}{4 \ln 2 \mathcal{C}(\sigma^2)} + \frac{1}{12 \ln 2 \mathcal{C}(\sigma^2)^2} + \mathcal{O}\left(\frac{1}{\mathcal{C}(\sigma^2)^3}\right) \quad (37)$$

as $\sigma^2 \rightarrow 0$ and $\mathcal{C}(\sigma^2) \rightarrow \infty$.

The proof is given in Appendix D. The gap, which is inverse proportional to the capacity itself, is plotted in Fig. 9 (squares) alongside with an upper bound given by Theorem 1 (circles), asymptotic scaling expression given by Theorem 2 (diamonds), and achievability (lower) bound (stars) discussed below.

Fig. 5 shows the maximum sum-rates achievable by the block (circles) and coupled (stars) systems for 2-PAM (blue), 4-PAM (magenta), and 8-PAM symbol alphabets as a function of E_b/N_0 . The solid red line shows AWGN

²A slightly tighter bound on the asymptotic gap to capacity can be found in the previous version of this paper posted on Arxiv. It requires a significantly lengthier derivation.

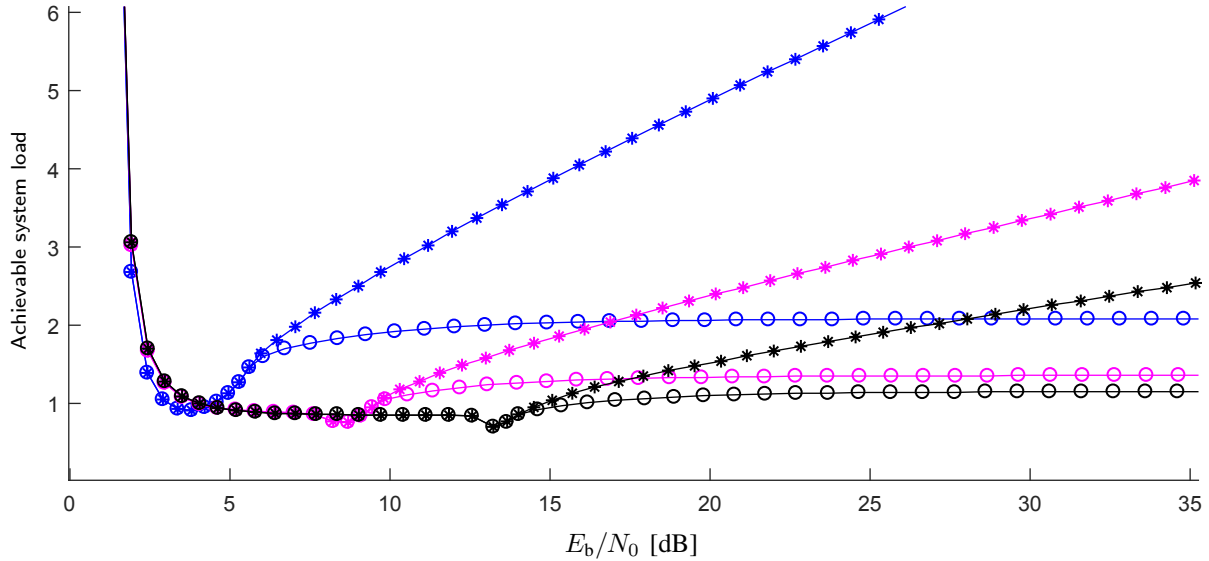


Fig. 6. System loads α maximizing the total achievable rate for coupled (stars) and block (circles) systems plotted as functions of E_b/N_0 . Loads for 2-PAM, 4-PAM, 8-PAM symbol alphabets are given by the blue, magenta, and black curves respectively.

channel capacity. We note that in the high SNR regime the rates achievable by the coupled systems approach the capacity curve. Conversely, the rates achievable by the block systems saturate with 8-PAM rate being the highest. This is because the coupled system can converge to the smallest (almost interference-free) fixed point of the characteristic equation while block system always converges to the largest fixed point. In low-SNR regime both block and coupled system rates coincide as they are determined by the same single fixed point of the characteristic equation.

Fig. 6 demonstrates the system loads maximizing the rates achievable by the block and coupled systems for 2-PAM, 4-PAM and 8-PAM constellation alphabets as a function of E_b/N_0 (that corresponds to the maximum achievable rate). For high SNRs the system loads maximizing the rate achievable by the block system coincide with the value $\alpha_s(B, \sigma'^2)$, the highest system load such that (15) has a single fixed. This is the case since the performance of the block system is driven by the largest fixed point of (15). The optimal loads the coupled system coincide with these of block systems for low SNRs. For high SNRs coupled systems can utilize higher loads since coupled systems are driven by the smallest fixed point of (15) at high SNRs the coupled system can get advantage of higher system loads. This behavior is consistent with the behavior of the individually optimal and jointly optimal MAP decoding thresholds studied in [44], [45] (using replica method) as discussed in Section IV-B.

F. Bounds on Achievable Rates

The proof of the asymptotic capacity-achieving property involves potential functions [48], [47] which can be used to characterize the conditions for convergence of coupled systems to their fixed points. While the proof in Appendix C

uses potential functions as in [48] here we use a potential function as discussed in [47] to visualize the capacity approaching behavior of the coupled systems for the two demodulation/decoding schedules from Section II-E. We start with the two-stage schedule. We consider the asymptotic scenario in this case where $L, M \rightarrow \infty$; $\alpha = L/M$ is constant. A variable exchange in the characteristic equation (15) leads to a characterization of fixed points u that satisfy³

$$\frac{1}{u} = \alpha g_{\text{mse}}(u) + \alpha \sigma^2. \quad (38)$$

Graphically the fixed points of (38) are illustrated in Fig. 7, which shows MSE vs SNR for a 2-PAM system with load $\alpha = 2.5$ and $\sigma^2 = 0.0254$ chosen such that the potential function of the coupled system equals 0. The left hand side of the (38) is shown by the blue solid curve while the right hand side of (38) is depicted by the red curve. The three black stars depict the three fixed points $(1/(\alpha x^{(3)}), \alpha x^{(3)})$, $(1/(\alpha x^{(2)}), \alpha x^{(2)})$, and $(1/(\alpha x^{(1)}), \alpha x^{(1)})$ of (38). According to [47] the parameter pair (α, σ^2) for which the two areas labeled P_1 and P_2 between the curves corresponding to the left and right hand sides of the characteristic equation (38) are equal correspond to the case when the potential function of the system equals 0. For fixed σ^2 this corresponds to the highest system load α such that the coupled system (19) converges to the smallest (nearly interference-free) fixed point $x^{(1)}$.

The area colored in light red (divided by $2 \ln 2$) can be computed by integration of the right hand side of (38) and equals the total rate achieved by the coupled system

$$\mathcal{R}_{\text{coup}}(\alpha, \sigma^2) = \int_0^{1/(\alpha x^{(1)})} \frac{1}{2 \ln 2} \alpha g_{\text{mse}}(u) du \quad (39)$$

while the area colored in blue (divided $2 \ln 2$) equals the AWGN channel capacity

$$\mathcal{C}(\sigma^2) = \frac{1}{2 \ln 2} \int_{1/(\alpha(1+\sigma^2))}^{1/(\alpha \sigma^2)} \frac{1}{u} du. \quad (40)$$

According to Lemma 2 the fixed points are contained between $(1/(\alpha(1+\sigma^2)), \alpha(1+\sigma^2))$ and $(1/(\alpha \sigma^2), \alpha \sigma^2)$. We also notice that the area R_1 belongs to the achievable rate only. The areas C_1 , C_2 , and C_3 (a tiny area between $1/(\alpha x^{(1)})$ and $1/(\alpha \sigma^2)$) belong exclusively to the AWGN capacity calculation. Both achievable rate and capacity share the main area M . A similar situation would happen in case of no-binary symbol alphabets and, as we will see later, in case of other demodulation/decoding schedules where usage of error-correction in the iterative demodulation loop impacts the MSE curve.

For the pictorial representation we can see that the difference between the achievable rate and capacity can be extracted from the areas R_1 , C_1 , C_2 , C_3 where R_1 and C_2 are the largest and play the main role. We can derive

³This equation appears in [47] where the system we study is discussed in an example. Here we add further steps relating it to the achievable rate and capacity.

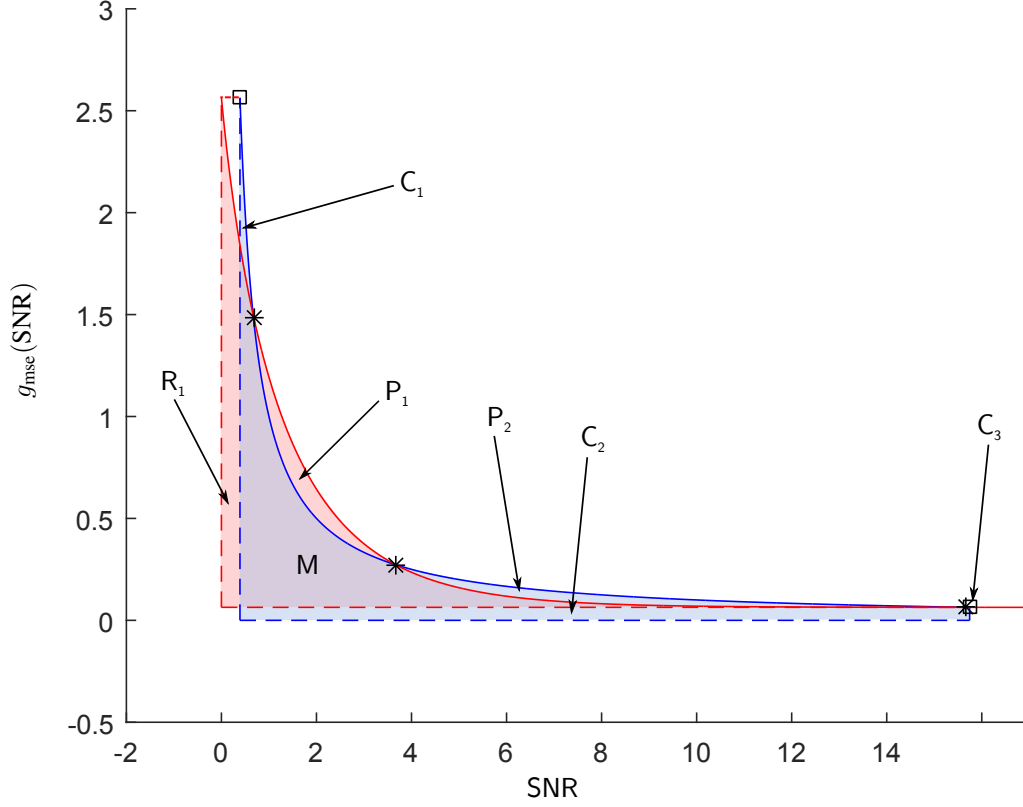


Fig. 7. Graphic representation of the MSE vs SNR, 2-stage demodulation/decoding.

a simple lower bound on the gap to capacity as

$$\begin{aligned}
 \mathcal{C}(\sigma^2) - \mathcal{R}_{\text{coup}}(\alpha, \sigma^2) &= \frac{1}{2 \ln 2} (C_1 + C_2 + C_3 - R_1) \geq \frac{1}{2 \ln 2} (C_2 - R_1) \\
 &\geq \frac{1}{2 \ln 2} \alpha \sigma^2 \left(\frac{1}{\alpha \sigma^2} - \frac{1}{\alpha(1 + \sigma^2)} \right) - \frac{1}{4 \ln 2} \frac{1}{\alpha(1 + \sigma^2)} \left(\alpha g_{\text{mse}}(0) + \sigma^2 + \alpha g_{\text{mse}} \left(\frac{1}{\alpha(1 + \sigma^2)} \right) + \sigma^2 \right) \\
 &= \frac{1}{2 \ln 2} \frac{1}{1 + \sigma^2} - \frac{1}{2 \ln 2} \frac{1}{1 + \sigma^2} \left(\frac{1}{2} + \frac{1}{2} g_{\text{mse}} \left(\frac{1}{\alpha(1 + \sigma^2)} \right) + \frac{\sigma^2}{\alpha} \right) \\
 &= \frac{1}{2 \ln 2} \frac{1}{1 + \sigma^2} \left(\frac{1}{2} - \frac{1}{2} g_{\text{mse}} \left(\frac{1}{\alpha(1 + \sigma^2)} \right) - \frac{\sigma^2}{\alpha} \right). \tag{41}
 \end{aligned}$$

For the case of 2-PAM signals we can specialize it further using the upper bound $(1 + s)^{-1} \geq g_2(s)$ derived in [12] and obtain

$$\begin{aligned}
 \mathcal{C}(\sigma^2) - \mathcal{R}_{\text{coup}}(\alpha, \sigma^2) &\geq \frac{1}{2 \ln 2(1 + \sigma^2)} \left(\frac{1}{2} - \frac{1}{2} \frac{\alpha(1 + \sigma^2)}{1 + \alpha(1 + \sigma^2)} - \frac{\sigma^2}{\alpha} \right) \\
 &= \frac{1}{2 \ln 2(1 + \sigma^2)} \left(\frac{1}{2(1 + \alpha(1 + \sigma^2))} - \frac{\sigma^2}{\alpha} \right). \tag{42}
 \end{aligned}$$

As $\sigma^2 \rightarrow 0$ the right hand side is approximately equal to $\frac{1}{2\alpha} \approx \frac{1}{2\mathcal{C}(\sigma^2)}$ which is consistent with the behavior showcased by Theorem 1. The upper bound on the gap derived in Theorem 1 involves estimation of all areas C_1 ,

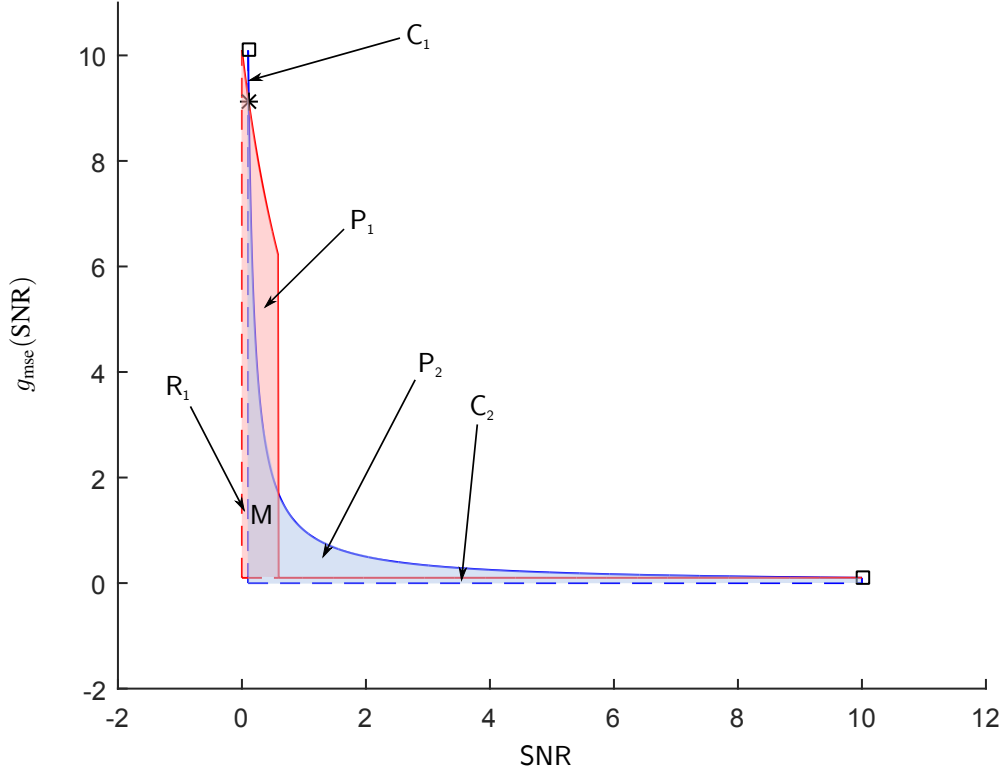


Fig. 8. Graphical representation of the MSE vs SNR, one time hard decoding feedback and cancellation schedule.

C_2 , C_3 and relates estimation of the largest fixed point $x^{(3)}$ of (15).

G. Rate Achievable for Hard Decoding Feedback and Cancellation Schedule

Consider now the hard decoding feedback and cancellation schedule. Once during the demodulation process a data stream reaches SNR equal or exceeding the threshold θ of the error correction code the data is passed to the error correction decoder. The decoder is then capable to perform error-free decoding and the effect of the data stream would be perfectly cancelled once the decoder gives its feedback to the demodulator.

Fixed points of the modified characteristic equation

$$\frac{1}{u} = \alpha g_{\text{mse}, \text{ecc}}(u) + \alpha \sigma^2 \quad (43)$$

as well as the areas between the curves characterizing the achievable rate and capacity are given in Fig. 8. The achievable rate expression is now given by

$$\mathcal{R}_{\text{coup}, \text{ecc}}(\alpha, \sigma^2) = \int_0^\gamma \frac{1}{2 \ln 2} \alpha g_{\text{mse}}(u) du \quad (44)$$

while the AWGN channel capacity is still computed as in (40). The gap to capacity in terms of the area is now given by

$$\begin{aligned}\mathcal{C}(\sigma^2) - \mathcal{R}_{\text{coup,ecc}}(\alpha, \sigma^2) &= \frac{1}{2 \ln 2} (\mathbf{C}_1 + \mathbf{C}_2 - \mathbf{R}_1) \geq \frac{1}{2 \ln 2} (\mathbf{C}_2 - \mathbf{R}_1) \\ &= \frac{1}{2 \ln 2(1 + \sigma^2)} \left(\frac{1}{2} - \frac{1}{2} g_{\text{mse}} \left(\frac{1}{\alpha(1 + \sigma^2)} \right) - \frac{\sigma^2}{\alpha} \right)\end{aligned}\quad (45)$$

which is the same lower bound as in (41). At the same time we can derive an upper bound

$$\mathcal{C}(\sigma^2) - \mathcal{R}_{\text{coup,ecc}}(\alpha, \sigma^2) = \frac{1}{2 \ln 2} (\mathbf{C}_1 + \mathbf{C}_2 - \mathbf{R}_1) \quad (46)$$

$$\leq \frac{1}{2 \ln 2} \frac{1}{2\alpha} \left(\frac{1}{x_s} - \frac{1}{1 + \sigma^2} \right) (\alpha(1 + \sigma^2 - x_s)) + \frac{1}{2 \ln 2(1 + \sigma^2)} \left(\frac{1}{2} - \frac{1}{2} g_{\text{mse}} \left(\frac{1}{\alpha(1 + \sigma^2)} \right) - \frac{\sigma^2}{\alpha} \right) \quad (47)$$

where x_s is the single fixed point of the characteristic equation (15). If we focus on the case of binary symbols we can now use Lemma 2 (c) and the lower bound $1 - y \leq g_2(y)$ [12] and based on (47) further obtain

$$\begin{aligned}\mathcal{C}(\sigma^2) - \mathcal{R}_{\text{coup,ecc}}(\alpha, \sigma^2) &= \frac{1}{2 \ln 2} \left[\frac{1}{2} \frac{(1 + \sigma^2 - x_s)^2}{x_s(1 + \sigma^2)} + \frac{1}{1 + \sigma^2} \left(\frac{1}{2} - \frac{1}{2} \left(1 - \frac{1}{\alpha(1 + \sigma^2)} \right) - \frac{\sigma^2}{\alpha} \right) \right] \\ &\leq \frac{1}{2 \ln 2} \left[\frac{1}{2} \frac{\left(\frac{1}{\alpha(1 + \sigma^2)} + \frac{1}{\alpha^2(1 + \sigma^2)^3} \right)^2}{\frac{1}{2}(1 + \sigma^2)^2} + \frac{1 - 2\sigma^2(1 + \sigma^2)}{2\alpha(1 + \sigma^2)^2} \right] \\ &\leq \frac{1}{2 \ln 2} \left[\frac{3}{\alpha^2(1 + \sigma^2)^4} + \frac{1 - 2\sigma^2(1 + \sigma^2)}{2\alpha(1 + \sigma^2)^2} \right] \leq \frac{1}{2 \ln 2\alpha} .\end{aligned}\quad (48)$$

For the case of the 2-PAM constellation the capacity-achieving property for hard decoding feedback and cancellation schedule has been proven in [27]. Here we look at the problem in more detail and prove that in case of a one time feedback and cancellation schedule convergence to channel capacity happens for any fixed σ^2 when we let $\alpha \rightarrow \infty$. For the two-stage decoding the system is only capacity achieving asymptotically as $\sigma^2 \rightarrow 0$. We can formulate this result as a theorem.

Theorem 3. *For any $\sigma^2 \leq 0.1$ (necessary for existence of at least one fixed point) and any $\alpha > 4$ there exists a $\bar{W} > 0$ such that for any $W > \bar{W}$ such that*

$$\mathcal{C}(\sigma^2) - \mathcal{R}_{\text{coup,ecc}}(\alpha, \sigma^2) \leq \frac{1}{2 \ln 2\alpha} \quad (49)$$

and, therefore, for any $\sigma^2 \leq 0.1$

$$\lim_{\alpha \rightarrow \infty} (\mathcal{C}(\sigma^2) - \mathcal{R}_{\text{coup,ecc}}(\alpha, \sigma^2)) = 0 .$$

IV. NUMERICAL RESULTS AND RELATION TO MAP DECODING

A. Numerical Results

We start by looking at the gap between the sum-rate achievable by coupled transmission and the AWGN channel capacity. Fig. 9 plots the difference between the capacity of the AWGN channel $\mathcal{C}(\sigma^2)$ and the rate $\mathcal{R}_{\text{coup}}(\alpha, \sigma^2)$ achievable by the coupled system with the two-stage schedule in the asymptotic regime $M, L \rightarrow \infty$ (black squares curve). The analytical upper bound $\mathcal{G}_{\text{asymptotic}}(\sigma^2)$ (see Theorem 1) is given by the blue circles curve. The asymptotic

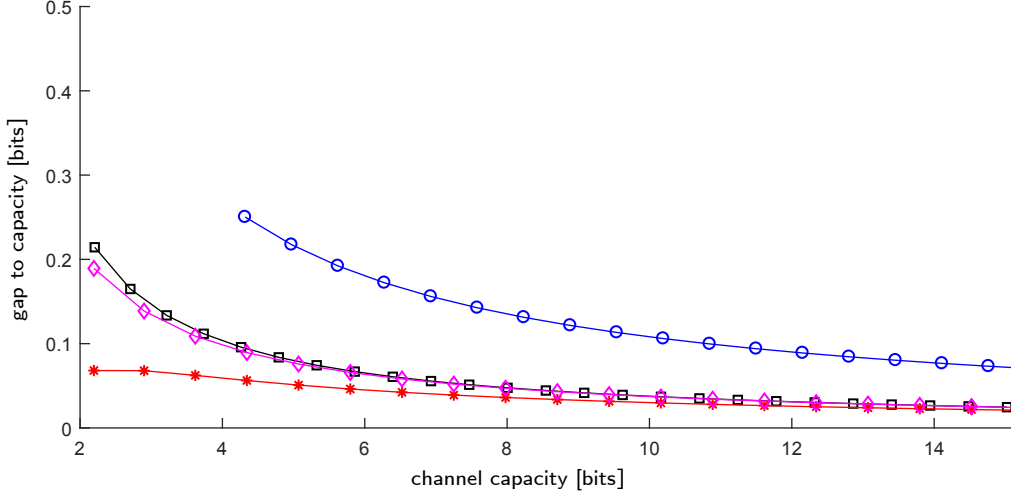


Fig. 9. The gap between the channel capacity and the total sum-rate achieved by the coupled system with two-stage demodulation/decoding (squares) with 2-PAM symbol constellation plotted as a function of the channel capacity. The upper bound $\mathcal{G}_{\text{asymptotic}}(\sigma^2)$ is given by blue circles, the asymptotic scaling expression by magenta diamonds, and the lower bound (42) by red stars.

scaling expression of Theorem 2 is given by magenta diamonds curve. Finally, the analytical lower bound (42) with $\alpha = \mathcal{C}(\sigma^2)$ is given by the red stars. We notice that all the there curves corresponding to the actual gap-to-capacity and the bounds on it decay as the channel capacity grows. The bounds get tighter as the capacity increases.

Simulation results of the spectral efficiency achievable by coupling data transmission are given in Fig. 10, which plots the sum-rates achievable by the coupled system with two-stage demodulation/decoding in bits per dimension as a function of the total SNR $1/\sigma^2$ in dB. The red curve corresponds to the capacity of the AWGN channel. The magenta curves correspond to capacities of the AWGN channels with constrained inputs where the lower magenta curve is for BPSK (2-PAM) modulation, and the other curves are for 4,8, and 16 PAM constellations (from bottom to top).

The black curve with circles plots simulated sum rate of the coupled data transmission with a block (data stream) length of $MN = 10,000$. The three simulated points correspond to loads $\alpha = 2, 3, 4$ with parameters $M = 250$ (and L chosen accordingly). The coupling window $W = 12$ was considered. The permutation matrices for each data stream have been chosen randomly with no attempt to eliminate short cycles. The SNR at the output of the demodulator is measured and the resulting sum-rate is computed assuming component codes achieve the capacity of the BIAWGN channel for this SNR. We note that the simulated performance curve is close to the theoretically predicted sum-rate performance (blue). Also for each $\alpha = 2, 3, 4$ the simulated performance is closer to the channel capacity than the performance of the corresponding 2,4,8, or 16-PAM capacity curves. Finally, we notice that the performance curve does not saturate at the modulation load of $\alpha = 2.07$ as it happens for the block system. The slight divergence at the higher load is arguably caused by the presence of short cycles in the system's graph which

becomes denser as the system load increases.

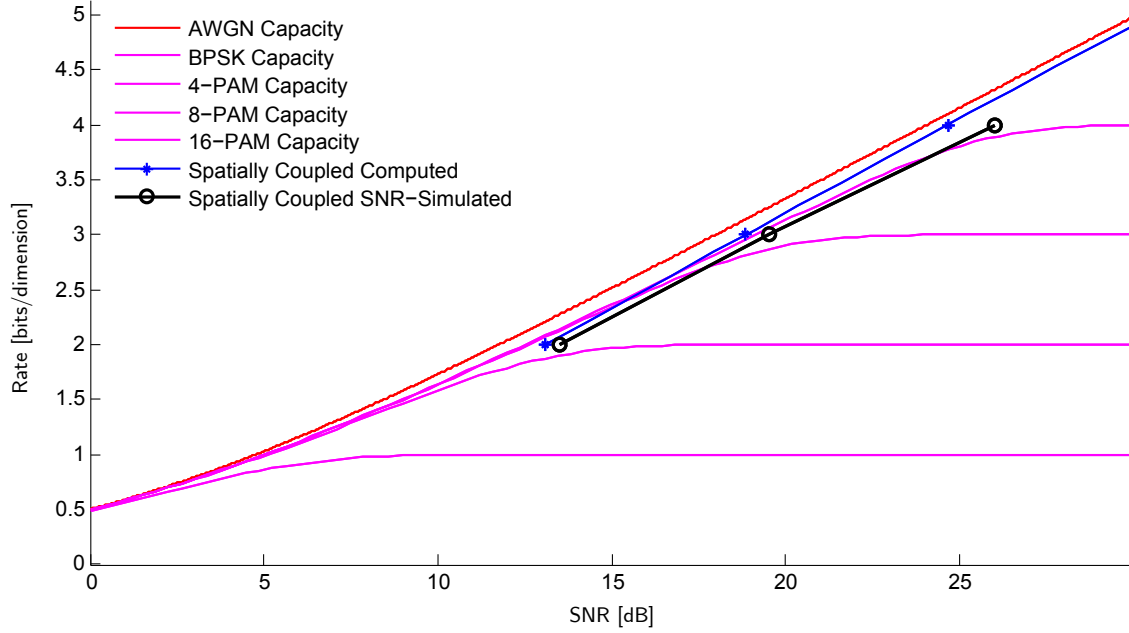


Fig. 10. Computed (blue) and simulated data rates (black) achievable by coupled data transmission as a function of the total system SNR. The red curve shows the capacity of the AWGN channel with power-constraint input while the magenta curves correspond to the capacities of the AWGN channel with the input constrained to 2,4,8,16-PAM symbol constellations.

B. Relation to MAP Decoding

Performance limits of joint and separate user decoding in context of multiple access communications have been considered in a number of prior papers. Tse and Verdu showed that in the high-SNR regime joint multi-user detection and decoding of random CDMA signals is capacity achieving [43]. Tanaka [30] applied the heuristic replica methods from statistical physics to study the spectral efficiencies achieved by iterative multiuser decoding of random CDMA systems in the large system limit. This results were later generalized by Guo and Verdu [45]. Müller and Gerstacker quantified the difference between the optimal rates achievable by joint and separate decoding of CDMA signals [44].

Before we compare the results achievable by our system under consideration with these achievable with optimal MAP detection and decoding of multi-user signals we would like to mention the two important differences in the system design. Our system model discussed in Section II includes repetition and permutation of signals at the modulation stage and then it includes coupling of signals at the receiver. The other important feature of our approach is the signal reception in which iterative detection is followed by individual ECC decoding. Our main contribution is a technique to approach the channel capacity using a low-complexity iterative detection followed by individual error correction decoding, without prohibitively complex joint detection/decoding or MAP demodulation

of multiuser signals. In the modulation format we consider each data bit is replicated M times and these replicas are spread among the entire block. This is different from random CDMA models considered prior works and is helpful in case of belief propagation decoding on the modulation graphs. Different bit replicas with different SNRs when combined for data estimation deliver a performance that approaches the performance of MAP demodulation due to the threshold saturation effect obtained via spatial graph coupling.

Often used for characterization of a multi-user detection performance the multi-user efficiency parameter defined in [42] is a ratio of the input and output SNR $\eta = \frac{\sigma^2}{\sigma_\infty^2}$. For optimum joint decoding (of CDMA signals) [30] η satisfies

$$\frac{1}{\eta} = 1 + \alpha \text{ snr } g_2(\eta \text{ snr}) = 1 + \frac{1}{\sigma^2} g_2\left(\frac{\eta}{\alpha \sigma^2}\right) \quad (50)$$

where $\text{snr} = 1/(\alpha \sigma^2)$ (see also [45] equation (34)). In case of multiple fixed points of the above equation it is shown in [45] that η resulting the smallest sum-rate

$$C_{\text{joint}}(\eta) = \frac{\eta - 1 - \log \eta}{\sigma^2} + \alpha C_{\text{BIAWGN}}\left(\frac{\eta}{\alpha \sigma^2}\right) \quad (51)$$

is the correct fixed point to choose. The equation (51) also appears in [44], [45] as a characterization of joint multiuser detection. The above characterization is, however, a conjecture derived via the informal replica method. The capacity of separate decoding and detection assuming optimum MAP multiuser decoding, again using the replica method analysis, is given by

$$C_{\text{sep}}(\eta) = \alpha C_{\text{BIAWGN}}(\eta \text{ snr}) = \alpha C_{\text{BIAWGN}}\left(\frac{\eta}{\alpha \sigma^2}\right) \quad (52)$$

as derived in [44]. The gap between joint and separate detection is also given by

$$\frac{\eta - 1 - \log \eta}{\sigma^2}. \quad (53)$$

The above equations have been generalized in [45] for the case of unequal user power distributions.

We notice that equation (50) coincides with (15), the characteristic equation for the block system. In this paper we have proven that the coupled system can achieve the smallest fixed point $x^{(1)}$ of (15) with iterative detection for loads α close to the capacity of the MAC channel. The resulting achievable spectral efficiency of our coupled system (32) formula coincides with (52). We also prove that our system design together with spatial coupling leads to an achievable $\eta = \sigma^2/x^{(1)}$ that rapidly approaches 1 as variance σ^2 of the noise diminishes (Lemma 2 a)). In such operational regime with $\eta \approx 1$ the gap between the performance of joint and separate decoding schedules (53) disappears and the capacity can be approached with a two-stage schedule.

V. CONCLUSION

We consider modulation of information in the form of a superposition of independent equal-power and equal-rate data streams. Each stream is formed by repetition and permutation of data and the streams are summed up with an offset initiating the effect of “stream coupling”. The convergence of block and coupled version of the system is studied for various symbol constellation alphabets. We have proven that the proposed system used with iterative

demodulation followed by external error control decoding achieves the capacity of the AWGN channel and the Gaussian multiple-access channel asymptotically and computed the gap to capacity as a function of the system's parameters. In addition, we show that for the hard feedback and cancellation schedule the gap to capacity is smaller than for two-stage schedule and can vanish at finite SNR.

REFERENCES

- [1] A. Jiménez Felström and K. Sh. Zigangirov, "Time-varying periodic convolutional codes with low-density parity-check matrices," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2181–2191, Sept. 1999.
- [2] M. Lentmaier, A. Sridharan, D. J. Costello, Jr., and K. Sh. Zigangirov, "Iterative decoding threshold analysis for LDPC convolutional codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5274–5289, Oct. 2010.
- [3] S. Kudekar, T. J. Richardson, and R. L. 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.
- [4] C. Schlegel and D. Truhachev, "Multiple Access Demodulation in the Lifted Signal Graph with Spatial Coupling," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2459–2469, Apr. 2013.
- [5] H. Imai and S. Hirakawa, "A new multilevel coding method using error correcting codes," *IEEE Transactions on Information Theory*, vol. 23, pp. 371–377, May 1977.
- [6] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.
- [7] A. Yedla, P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, "Universal codes for the Gaussian MAC via spatial coupling," in *Proc. Annual Allerton Conf. on Commun., Control, and Comp.*, Monticello, IL, Sept. 2011.
- [8] S. Kudekar and K. Kasai, "Spatially coupled codes over the multiple access channel," *IEEE International Symposium on Information Theory Proceedings (ISIT)*, pp. 2816–2820, St Petersburg, Russia, 2011.
- [9] D. Guo and C. C. Wang, "Multiuser Detection of Sparsely Spread CDMA," in *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 3, pp. 421–431, April 2008.
- [10] C. Schlegel, M. Burnashev, and D. Truhachev, "Generalized superposition modulation and iterative demodulation: A capacity investigation," *Hindawi Journal of Electr. and Comp. Eng.*, vol. 2010, Sep. 2010.
- [11] D. Truhachev, C. Schlegel, and L. Krzymien, "A two-stage capacity-achieving demodulation/decoding method for random matrix channels," *IEEE Tran. on Inform. Theory*, vol. 55, no. 1, pp. 136–146, Jan. 2009.
- [12] M. Burnashev, C. Schlegel, W. Krzymien, and Z. Shi, "Characteristics Analysis of Successive Interference Cancellation Methods", *Problemy Peredachi Informatsii*, vol. 40, no. 4, pp. 297–317, Dec. 2004.
- [13] L. Ping, L. Liu, K. Wu, and W. Leung, "Interleave division multiple-access," *IEEE Trans. Wireless Commun.*, vol. 5, no. 4, pp. 938–947, April 2006.
- [14] L. Liu, J. Tong, and L. Ping, "Analysis and Optimization of CDMA Systems with Chip-Level Interleavers," *IEEE Journal on Selected Areas in Communications*, pp. 141–150, Jan. 2006.
- [15] L. Duan, B. Rimoldi, R. Urbanke, "Approaching the AWGN channel capacity without active shaping," *Proc. IEEE Int. Symp. Inform. Theory*, p. 374, June–July 1997.
- [16] X. Ma, L. Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3331–3343, Dec. 2004.
- [17] P. A. Hoeher, H. Schoeneich, J. C. Fricke, "Multi-layer interleave-division multiple access: theory and practice," *European Transactions on Telecommunications*, vol. 19, no. 5, pp. 523–536, Aug. 2008.
- [18] H. S. Cronie, "Signal shaping for bit-interleaved coded modulation on the AWGN channel", *IEEE Trans. Commun.*, vol. 58, no. 12, pp. 3428–3435, Dec. 2010.

- [19] P. A. Hoeher and T. Wo, "Superposition modulation: myths and facts," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 110–116, Dec. 2011.
- [20] S. Kudekar and T. Richardson and R. Urbanke, "Spatially Coupled Ensembles Universally Achieve Capacity under Belief Propagation," submitted to *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec. 2013.
- [21] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *Proc. Allerton Conf. on Communications, Control, and Computing*, Monticello, IL, Sept. 2010.
- [22] C. Schlegel and D. Truhachev, "Multiple Access Demodulation in the Lifted Signal Graph with Spatial Coupling," in *Proc. IEEE Int. Symp. on Inf. Theory*, St. Petersburg, Russia, Aug. 2011.
- [23] D. Truhachev, "Universal Multiple Access via Spatially Coupling Data Transmission", *IEEE International Symposium on Information Theory 2013*, Istanbul, Turkey, July 2013.
- [24] L. Krzymien, D. Truhachev and C. Schlegel, "Coded Random CDMA with Partitioned Spreading", In *Proceedings of the 44th Annual Allerton Conference on Communications Control and Computing*, Monticello, USA, September 2006.
- [25] K. Takeuchi, T. Tanaka, and T. Kawabata, "Improvement of BP-based CDMA multiuser detection by spatial coupling," in *Proc. IEEE Int. Symp. on Inf. Theory*, St. Petersburg, Russia, Aug. 2011.
- [26] M. Hagiwara, K. Kasai, H. Imai, and K. Sakaniwa, "Spatially coupled quasi-cyclic quantum LDPC codes," in *Proc. IEEE Int. Symp. on Inf. Theory*, St. Petersburg, Russia, Aug. 2011.
- [27] D. Truhachev, "Achieving AWGN Multiple Access Channel Capacity with Spatial Graph Coupling," *IEEE Communications Letters*, vol. 16, no. 5, pp. 585–588, May 2012.
- [28] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.
- [29] S. Kumar, A. Young, N. Macris, and H. Pfister, "A Proof of Threshold Saturation for Irregular LDPC Codes on BMS Channels," in *Proc. Allerton Conf. on Communication, Control, and Computing*, Allerton, Illinois, USA, Sept. 2012.
- [30] T. Tanaka, "A statistical mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [31] M. S. Alencar, "A comparison of bounds on the capacity of a binary channel," in *Proc. Global Telecommun. Conf.*, Nov. 1996, vol. 2, pp. 1273–1275.
- [32] A. Yedla, Y.-Y. Jian, P. S. Nguyen and H. D. Pfister, "A Simple Proof of Threshold Saturation for Coupled Scalar Recursions," in *7th International Symposium on Turbo Codes and Iterative Information Processing*, Göthenburg, Sweden, August 2012.
- [33] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for "turbo-like" codes," in *Proc. 36th Allerton Conf. on Communication, Control, and Computing*, pp. 201–210, Allerton, Illinois, USA, Sept. 1998.
- [34] R. Gallager, *Information Theory and Reliable Communications*, Wiley & Sons, New York, 1968.
- [35] K. Takeuchi, T. Tanaka, and T. Kawabata, "A Phenomenological Study on Threshold Improvement via Spatial Coupling," *IEICE Transactions*, vol. 95-A, no. 5, pp. 974–977, 2012.
- [36] K. Takeuchi, T. Tanaka and T. Kawabata, "Performance Improvement of Iterative Multiuser Detection for Large Sparsely Spread CDMA Systems by Spatial Coupling," in *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1768–1794, Apr. 2015.
- [37] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing," *IEEE Trans. Inf. Theory*, Dec. 2013.
- [38] R. G. GALLAGER, "Low Density Parity Check Codes", M.I.T. Press, 1963.
- [39] M. Lentmaier, D. Truhachev, and K. S. Zigangirov, "On the theory of low density convolutional codes II," *Problems of Information Transmission*, vol. 37, pp. 288–306, Oct.-Dec. 2001.

- [40] D. Truhachev, M. Lentmaier, and K. S. Zigangirov, “Mathematical Analysis of Iterative Decoding of LDPC Convolutional Codes,” *Proc. IEEE Int. Symp. on Inf. Theory*, p. 191, Washington, USA, June 24–29, 2001.
- [41] G. A. Margulis, “Explicit Constructions of Graphs without Short Cycles and Low-Density Codes,” *Combinatorica*, vol. 2, no. 1, pp. 71–78., 1982.
- [42] S. Verdú, “Multiuser Detection.” Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [43] D. N. C. Tse and S. Verdú, “Optimum Asymptotic Multiuser Efficiency of Randomly Spread CDMA,” *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2718–2722, Nov. 2000.
- [44] R. R. Müller and W. H. Gerstacker, “On the Capacity Loss Due to Separation of Detection and Decoding,” *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1769–1778, Aug. 2004.
- [45] D. Guo and S. Verdú, “Randomly Spread CDMA: Asymptotics Via Statistical Physics,” *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 1983–2010, June. 2005.
- [46] D. Guo, Y. Wu, S. Shamai, and S. Verdú, “Estimation in Gaussian Noise: Properties of the Minimum Mean-Square Error,” *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr. 2011.
- [47] S. Kudekar, T. J. Richardson and R. L. Urbanke, “Wave-Like Solutions of General 1-D Spatially Coupled Systems,” in *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4117–4157, Aug. 2015.
- [48] A. Yedla, Y. Y. Jian, P. S. Nguyen and H. D. Pfister, “A Simple Proof of Maxwell Saturation for Coupled Scalar Recursions,” in *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6943–6965, Nov. 2014.
- [49] J. Barbier, M. Dia, N. Macris, “Proof of Threshold Saturation for Spatially Coupled Sparse Superposition Codes,” *Proc. IEEE Int. Symp. on Inf. Theory*, Barcelona, Spain, July 2016.
- [50] A. Barron and A. Joseph, “Toward fast reliable communication at rates near capacity with Gaussian noise,” in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, June 2010, pp. 3153–19.
- [51] C. Rush, A. Greig, and R. Venkataramanan, “Capacity-achieving Sparse Regression Codes via Approximate Message Passing Decoding,” in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, Hong Kong, July 2015.

APPENDIX

A. Proof of Lemma 1

We start by observing the fact that each modulated data stream \tilde{v}_l can be represented as a product of the encoded data sequence v_l and a binary matrix H_l which represents repetition and permutation operations,

$$\tilde{v}_l = H_l v_l ,$$

i.e., $H_l = P_l R$, where P_l is an $MN \times MN$ binary permutation matrix specific to data stream l , and $MN \times N$ matrix R is a binary repetition matrix, repeating each bit of v_l M times,

$$R = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & & & \\ & & & & & 1 & 1 & 1 & \cdots & 1 \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 & 1 & 1 & \cdots & 1 \end{pmatrix}^T \quad (54)$$

Without loss of generality we consider $M_1 = M$, $M_2 = 1$ and the regularized coupling structure of the data streams described in Section II-B. The resulting coupled modulated sequence equals

$$s = H v = C P \tilde{R} v,$$

where $\mathbf{v} = (v_1, v_2, \dots)$ is the composite vector of all data streams, and \mathbf{C} is the coupling matrix. The coupling matrix \mathbf{C} shown in Fig. 11 consists of identical sub-matrix blocks depicted by the dashed rectangles. Each block is filled with $(2W + 1) \times (2W + 1)$ identity matrices \mathbf{I} of size $N_w \times N_w$ as shown in Fig. 11. Each consequent block is shifted down by the size of one identity matrix. The permutation matrix \mathbf{P} is a block-diagonal matrix which

$$\mathbf{C} = \begin{bmatrix} \begin{array}{ccc} \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \dots & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & & \mathbf{I} \\ \vdots & \vdots & & \vdots \\ \mathbf{I} & \mathbf{I} & & \mathbf{I} \end{array} & \begin{array}{ccc} \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \dots & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & & \mathbf{I} \\ \vdots & \vdots & & \vdots \\ \mathbf{I} & \mathbf{I} & & \mathbf{I} \end{array} & \begin{array}{ccc} \mathbf{I} & \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & \dots & \mathbf{I} \\ \mathbf{I} & \mathbf{I} & & \mathbf{I} \\ \vdots & \vdots & & \vdots \\ \mathbf{I} & \mathbf{I} & & \mathbf{I} \end{array} & \dots \end{bmatrix}$$

Fig. 11. Coupling Matrix.

consists of permutation matrices of individual data streams $\mathbf{P} = \text{diag}(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \mathbf{P}_4, \dots)$ and the repetition matrix $\tilde{\mathbf{R}}$ is constructed as given in (54) but is infinitely long. Here we leave the signature sequence out of the consideration since they do not affect the structure of the system's graph.

The resulting matrix \mathbf{H} is a band-diagonal binary matrix with M ones in each row and L ones in each column. Such matrices belong to a class of syndrome formers (transposed infinite parity-check matrices) of LDPC convolutional codes [39]. An example of a syndrome former of a $(3, 6)$ -regular convolutional LDPC code constructed using \mathbf{P} that consists of smaller permutation matrices is given in Fig. 12.

Hence, it is interesting to note that if the addition of the binary modulated data streams was performed as a modulo 2 addition the resulting coupled sequence would be a codeword of an (L, M) -regular convolutional (spatially coupled) LDPC code. Such code can be graphically represented by an infinitely-long bipartite Tanner graph [1][39]. The same bipartite graph describes modulated data in our system (see Fig. 3 b)). For the case of block system several researchers have shown that it is possible to construct bipartite graphs of block LDPC codes without any cycles of length $2I$ for any fixed integer I [38](Appendix B), [41] if the block length of the code is chosen large enough. A counterpart of this result for convolutional LDPC codes has been shown in [40][39] where, it has been proven that regular LDPC convolutional codes with graphs of any given girth I can be constructed from block LDPC code graphs of large girths by an unwrapping procedure [40][39]. This directly implies that modulation graphs of any given girth I can be constructed if MN is chosen to be sufficiently large.

We note that the operation of symbol estimation (7) (see Fig. 3) and the operation of interference cancellation (2) both can be regarded as message message-passing operations on the system's graph. If the system's graph has no cycles of length smaller or equal to $2I$ demodulation with I iterations performed to estimate each bit $v_{j,l}$ is

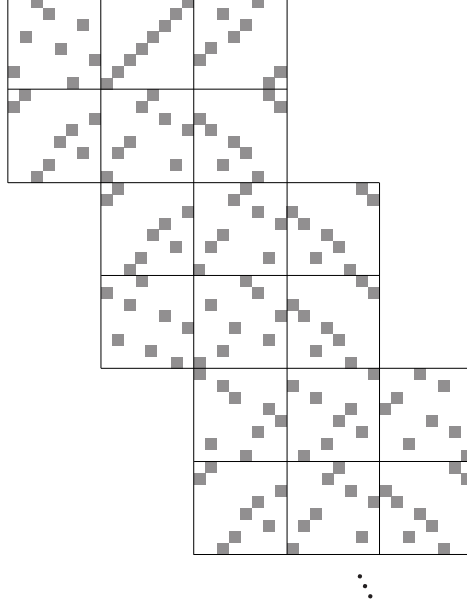


Fig. 12. Multiple convolutional permutor matrix.

operating on a tree. This proves the lemma. We note here that construction of modulation graphs with large girth is done for the purpose of analysis while in practice it is typically enough to expurgate cycles of length four and six.

B. Proof of Lemma 2

Let us define

$$h(x) \stackrel{\text{def}}{=} g_2\left(\frac{1}{\alpha'x}\right) + \sigma'^2 \geq \sigma'^2, \quad (55)$$

where the inequality follows from $g_2(y) > 0$, for $y \in (0, \infty)$. Similarly the fact that $g_2(\cdot)$ is non-negative implies the lower bound in Lemma 2 (a)

$$\sigma'^2 \leq g_2\left(\frac{1}{\alpha'x^{(1)}}\right) + \sigma'^2 = x^{(1)}. \quad (56)$$

We use the upper bound on $g_2(\cdot)$ derived in [12]

$$g_2(y) \leq \pi Q(\sqrt{y}) = \pi \int_{\sqrt{y}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (57)$$

and the condition $\alpha' \leq \mathcal{C}(\sigma'^2)$ of Lemma 2 (a) to upper bound

$$h(2\sigma'^2) \leq \pi Q \left(\frac{1}{\sqrt{\mathcal{C}(\sigma'^2)2\sigma'^2}} \right) + \sigma'^2 \leq \pi \frac{1}{\sqrt{2\pi}} \sqrt{\mathcal{C}(\sigma'^2)2\sigma'^2} \exp \left(-\frac{1}{\mathcal{C}(\sigma'^2)\sigma'^2} \right) + \sigma'^2 \quad (58)$$

$$= \exp \left(-\frac{1}{\mathcal{C}(\sigma'^2)\sigma'^2} + \frac{1}{2} \ln \frac{\pi}{2} + \frac{1}{2} \ln(2\sigma'^2 \mathcal{C}(\sigma'^2)) \right) + \sigma'^2 \quad (59)$$

$$= \exp \left(-\frac{1}{\sigma'^2 \frac{1}{2 \ln 2} \ln(1 + 1/\sigma'^2)} + \frac{1}{2} \ln \frac{\pi}{2} + \frac{1}{2} \ln \left(2\sigma'^2 \frac{1}{2 \ln 2} \ln(1 + 1/\sigma'^2) \right) \right) + \sigma'^2 \quad (60)$$

$$\leq \exp \left(-\frac{1}{\sigma'} + 2 \ln(\sigma') \right) + \sigma'^2 < 2\sigma'^2 . \quad (61)$$

The first inequality in (61) is valid for $\sigma'^2 \leq 1$ and is due to the fact that the term $-1/\sigma'^2$ dominates the exponent in (60) as $\sigma'^2 \rightarrow 0$. Bounds (55) and (58)–(61) imply the existence of a fixed point $\sigma'^2 \leq x^{(1)} \leq 2\sigma'^2$ for any $\alpha' \in [0, \mathcal{C}(\sigma'^2)]$ when $\sigma'^2 \leq 1$. The fixed point $x^{(1)}$ then satisfies the first inequality in (61) which is the upper bound given by (26). Lemma 2 (a) is proved. \square

The upper bound on $x^{(3)}$ in (27) is straightforward since $g_2(y) \leq 1$ for $y \in [0, \infty]$ and, therefore,

$$x^{(3)} = g_2 \left(\frac{1}{\alpha' x^{(3)}} \right) + \sigma'^2 \leq 1 + \sigma'^2 .$$

To prove the lower bound in (27) we use the lower bound derived in [12]

$$g_2(y) \geq 1 - y, \quad y \in [0, \infty] . \quad (62)$$

Inequality (62) implies that x_r , the largest fixed point of the equation

$$x = 1 - \frac{1}{\alpha' x} + \sigma'^2 , \quad (63)$$

is a lower bound on $x^{(3)}$. Hence

$$\left(x^{(3)} \right)^2 \geq (x_r)^2 = \left[\frac{1 + \sigma'^2}{2} \left(1 + \sqrt{1 - \frac{4}{\alpha'(1 + \sigma'^2)^2}} \right) \right]^2 \quad (64)$$

$$\begin{aligned} &\geq \frac{(1 + \sigma'^2)^2}{4} \left(2 - \frac{4}{\alpha'(1 + \sigma'^2)^2} + 2 \sqrt{1 - \frac{4}{\alpha'(1 + \sigma'^2)^2}} \right) \\ &\geq \frac{(1 + \sigma'^2)^2}{2} - \frac{1}{\alpha'} + \frac{(1 + \sigma'^2)^2}{2} \left(1 - \frac{4}{\alpha'(1 + \sigma'^2)^2} \right) = (1 + \sigma'^2)^2 - \frac{3}{\alpha'} . \end{aligned} \quad (65)$$

Since the value inside the square root in (65) needs to be non-negative the bound can be used for

$$\frac{4}{(1 + \sigma'^2)^2} \leq \alpha' ,$$

i.e. for $\alpha' \geq 4$ in particular. Lemma 2 (b) is proved.

To prove the lower bound in (28) (Lemma 2 (c)) we simply refine the bound (65) on x_r via a Taylor series expansion of $\sqrt{1 - \frac{4}{\alpha'(1 + \sigma'^2)^2}}$. Lemma 2 is proved. \square

C. Proof of Theorem 1

Proof. We start with deriving a convergence condition for the coupled system (19) using the method of potential functions [48]. The use of potential functions for coupled systems was suggested in [35], [36], where the authors showed that these functions are related to the Bethe free energy of continuous dynamical system describing a coupled model of a code-division multiple-access (CDMA) system and in [32], [47], where the technique was designed for more general types of coupled recursions. The potential function has also been used in [37] to study a coupled dynamical system describing reconstruction for compressed sensing.

The main part of the proof is dedicated to the derivation of the limiting load α^* for which the coupled system converges to the nearly interference-free state. To do so we use a relation between the mutual information and the MSE [28] to derive an analytically tractable lower bound on the minimum of the potential function. Finally, we use bounds on the fixed points $x^{(1)}$ and $x^{(3)}$ of the convergence equation, provided by Lemma 2, to compute the achievable communication rate $\mathcal{R}(\alpha^*, \sigma^2)$, prove that it is within a small gap from the channel capacity $\mathcal{C}(\sigma^2)$, and derive the asymptotic behavior of this gap.

We focus on the case of 2-PAM symbol constellation and for a fixed pair of parameters α', σ'^2 such that $\alpha' \in [4, \mathcal{C}(\sigma'^2)]$ consider the recursion

$$x_i = g_2 \left(\frac{1}{\alpha' (x_{i-1} + x^{(1)})} \right) + \sigma'^2 - x^{(1)} \quad (66)$$

equivalent to (15) in a sense that the fixed points of (66) are equal to these of (15) reduced by $x^{(1)} = x^{(1)}(\alpha', \sigma'^2)$ which is the smallest fixed point of (15). The fixed points of (66) are given by 0, $x^{(2)} - x^{(1)}$, and $x^{(3)} - x^{(1)}$. The coupled recursion corresponding to (66) can be written as

$$x_i^t = \frac{1}{2W+1} \sum_{\tau_1=-W}^W g_2 \left(\frac{1}{\alpha'} \frac{1}{2W+1} \sum_{\tau_2=-W}^W \frac{1}{x_{i-1}^{t+\tau_1+\tau_2} + x^{(1)}} \right) + \sigma'^2 - x^{(1)}, \quad t > 0, i > 0 \quad (67)$$

and is, in turn, equivalent to (19) from the convergence perspective. The resulting vector fixed points of (19) and (67) also differ by $x^{(1)}$. The recursion (66) can be stated in the form

$$x_i = f(g(x_{i-1}), \alpha') \quad (68)$$

where

$$f(x, \alpha) \stackrel{\text{def}}{=} g_2 \left(\frac{1}{\alpha'} \left(\frac{1}{x^{(1)}} - x \right) \right) + \sigma'^2 - x^{(1)}, \quad (69)$$

$$g(x) \stackrel{\text{def}}{=} \frac{1}{x^{(1)}} - \frac{1}{x^{(1)} + x}. \quad (70)$$

The potential function of the block system (68) is given [48] by

$$U(x, \alpha') = \ln \frac{x + x^{(1)}}{x^{(1)}} - \frac{\sigma'^2 x}{x^{(1)}(x + x^{(1)})} - \alpha' \int_{\frac{1}{\alpha'(x+x^{(1)})}}^{\frac{1}{\alpha'x^{(1)}}} g_2(y) dy. \quad (71)$$

and based on the fact that f, g are increasing and differentiable (admissibility conditions) we can utilize Theorem 1 [48] which states that the fixed points of (66) satisfy

$$\max_{t \geq 0} x_t^* \leq \max_x \left(\arg \min_x U(x, \alpha') \right).$$

for sufficiently large W . Hence, if for α', σ'^2 the potential function $U(x, \sigma'^2) = 0$ for $x = 0$ and $U(x, \sigma'^2) > 0$ for $x > 0$ the components of the fixed point vector x^* of the original coupled system (19) do not exceed $x^{(1)}$. Let us define

$$\alpha^*(\sigma'^2) = \sup \left\{ \alpha : \min_{x \in [0, 1+\sigma'^2]} U(x, \alpha') \geq 0 \right\} . \quad (72)$$

For any $\alpha' < \alpha^*$ (67) converges to $\mathbf{0}$ and, therefore, the original coupled system (19) converges to a fixed point vector with components not exceeding $x^{(1)}$. We will focus of estimating α^* and the respective load $\mathcal{R}_{\text{coup}}(\alpha^*, \sigma'^2)$ which we now know is achievable.

The function $U(x, \alpha')$ is plotted in Fig. 13 for $\sigma'^2 = 0.0129, 0.003347$, and 0.000855 and the corresponding $\alpha^* = 3, 4, 5$. Note that α^* is a decreasing function of σ'^2 .

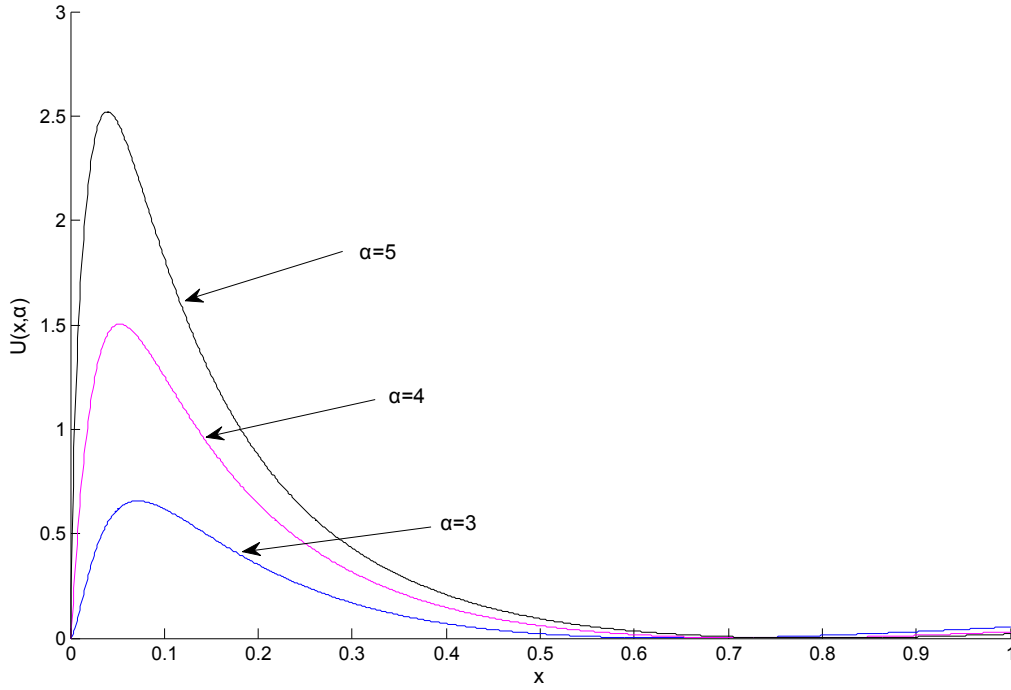


Fig. 13. Plot of $U(x, \alpha')$ for $\alpha' = 3$ and $\sigma'^2 = 0.0129$ (blue curve), $\alpha' = 4$ and $\sigma'^2 = 0.003347$ (magenta curve), and $\alpha' = 5$ and $\sigma'^2 = 0.000855$ (black curve).

To find the minimum of $U(x, \alpha')$ defined by (71) for a given α' and σ'^2 we compute its partial derivative with respect to x and equate it to 0

$$\frac{\partial U(x, \alpha')}{\partial x} = \frac{1}{x + x^{(1)}} - \frac{\sigma'^2}{(x + x^{(1)})^2} - g_2 \left(\frac{1}{\alpha'(x + x^{(1)})} \right) \frac{1}{(x + x^{(1)})^2} = 0 . \quad (73)$$

which is equivalent to

$$x + x^{(1)} = g_2 \left(\frac{1}{\alpha'(x + x^{(1)})} \right) + \sigma'^2 . \quad (74)$$

Since we consider $\alpha' \in (4, \mathcal{C}(\sigma'^2))$ and $4 > \alpha'_s$ the equation (74) has three fixed points. This implies that $U(x, \alpha')$ has two local minima achieved at $x = 0$ and $x_m = x^{(3)} - x^{(1)}$. Moreover, for $\alpha' = \alpha^*$ we have $U(x_m, \alpha^*) = 0$. Computing $U(x_m, \alpha')$ leads to

$$U(x_m, \alpha') = U(x^{(3)} - x^{(1)}, \alpha') = \ln \frac{x^{(3)}}{x^{(1)}} - \frac{x^{(3)} - x^{(1)}}{x^{(1)}x^{(3)}} \sigma'^2 - \alpha' \int_{\frac{1}{\alpha'x^{(3)}}}^{\frac{1}{\alpha'x^{(1)}}} g_2(y) dy \stackrel{\text{def}}{=} u(\alpha', \sigma'^2). \quad (75)$$

We will now focus on the function $u(\alpha', \sigma'^2)$, compute lower bounds for it and find out when these are positive in order to find a lower bound on $\alpha^*(\sigma'^2)$ and the achievable rate $\mathcal{R}(\alpha, \sigma^2)$.

The relationship between MSE (i.e. function $g_2(\cdot)$) and the mutual information for BIAWGN channel, derived in [28] implies

$$\int_{\frac{1}{\alpha'x^{(3)}}}^{\frac{1}{\alpha'x^{(1)}}} g_2(y) dy = 2 \ln 2 \left(\mathcal{C}_{\text{BIAWGN}} \left(\frac{1}{\alpha'x^{(1)}} \right) - \mathcal{C}_{\text{BIAWGN}} \left(\frac{1}{\alpha'x^{(3)}} \right) \right). \quad (76)$$

Applying (76) to (75) leads to

$$u(\alpha', \sigma'^2) = \ln \frac{x^{(3)}}{x^{(1)}} - \frac{1}{x^{(1)}} \sigma'^2 + \frac{1}{x^{(3)}} \sigma'^2 - 2 \ln 2 \alpha' \mathcal{C}_{\text{BIAWGN}} \left(\frac{1}{\alpha'x^{(1)}} \right) + 2 \ln 2 \alpha' \mathcal{C}_{\text{BIAWGN}} \left(\frac{1}{\alpha'x^{(3)}} \right) \quad (77)$$

For a given SNR (or equivalently σ'^2) in order to find a suitable α' such that $u(\alpha', \sigma'^2) > 0$ we first find a convenient lower bound on (77). We use the bounds on the capacity of the BIAWGN channel [34][31]

$$\frac{\gamma - \gamma^2}{2 \ln 2} \leq \mathcal{C}_{\text{BIAWGN}}(\gamma) \leq 1. \quad (78)$$

where the lower bound in (78) is valid for $\gamma < 1$. Lemma 2 (b) implies that the condition $1/(\alpha'x^{(3)}) < 1$ is satisfied for $\alpha' > 4$. Applying (78) to (77) leads to

$$u(\alpha', \sigma'^2) \geq \ln \frac{x^{(3)}}{x^{(1)}} - \frac{1}{x^{(1)}} \sigma'^2 + \frac{1}{x^{(3)}} \sigma'^2 - 2 \ln 2 \alpha' + \alpha' \left[\frac{1}{\alpha'x^{(3)}} - \frac{1}{(\alpha'x^{(3)})^2} \right] \quad (79)$$

We then extract the capacity term

$$u(\alpha', \sigma'^2) \geq 2 \ln 2 \left[\frac{1}{2} \log_2 \frac{1 + \sigma'^2}{\sigma'^2} \right] + \ln \frac{x^{(3)}}{1 + \sigma'^2} + \ln \frac{\sigma'^2}{x^{(1)}} \quad (80)$$

$$- \frac{1}{x^{(1)}} \sigma'^2 + \frac{1}{x^{(3)}} \sigma'^2 - 2 \ln 2 \alpha' + \alpha' \left[\frac{1}{\alpha'x^{(3)}} - \frac{1}{(\alpha'x^{(3)})^2} \right] \quad (81)$$

$$\geq 2 \ln 2 [\mathcal{C}(\sigma'^2) - \alpha'] + \ln \frac{x^{(3)}}{1 + \sigma'^2} + \ln \frac{\sigma'^2}{x^{(1)}} - \frac{1}{x^{(1)}} \sigma'^2 + \frac{1}{x^{(3)}} \sigma'^2 + \frac{1}{x^{(3)}} - \frac{1}{\alpha'(x^{(3)})^2} \quad (82)$$

At the next step we use Lemma 2 to bound the third and fourth terms in (82) which leads to

$$u(\alpha', \sigma'^2) \geq 2 \ln 2 [\mathcal{C}(\sigma'^2) - \alpha'] + \ln \frac{x^{(3)}}{1 + \sigma'^2} + \ln \frac{\sigma'^2}{\sigma'^2(1 + e^{-\frac{1}{\sigma}})} - 1 + \frac{1 + \sigma'^2}{x^{(3)}} - \frac{1}{\alpha'(x^{(3)})^2} \quad (83)$$

$$= 2 \ln 2 [\mathcal{C}(\sigma'^2) - \alpha'] + \ln \frac{x^{(3)}}{1 + \sigma'^2} - \ln(1 + e^{-\frac{1}{\sigma}}) - 1 + \frac{1 + \sigma'^2}{x^{(3)}} - \frac{1}{\alpha'(x^{(3)})^2} \quad (84)$$

Applying a simple lower bound on the second term on the right hand side of (84) implies

$$u(\alpha', \sigma'^2) \geq 2 \ln 2 [\mathcal{C}(\sigma'^2) - \alpha'] + 1 - \frac{1 + \sigma'^2}{x^{(3)}} - \ln(1 + e^{-\frac{1}{\sigma}}) - 1 + \frac{1 + \sigma'^2}{x^{(3)}} - \frac{1}{\alpha'(x^{(3)})^2} \quad (85)$$

$$= 2 \ln 2 [\mathcal{C}(\sigma'^2) - \alpha'] - \ln(1 + e^{-\frac{1}{\sigma}}) - \frac{1}{\alpha'(x^{(3)})^2}. \quad (86)$$

Now we notice that the fixed point $x^{(3)}$ of (15) is monotonically increasing both as a function of α' and σ'^2 and, therefore, $x^{(3)}(\alpha', \sigma'^2) \geq x^{(3)}(4, 0) = 0.7396$ for $\alpha' \geq 4$, $\sigma'^2 \geq 0$. Hence

$$u(\alpha', \sigma'^2) \geq 2 \ln 2 [\mathcal{C}(\sigma'^2) - \alpha'] - \ln(1 + e^{-\frac{1}{\sigma}}) - \frac{1}{\alpha'(x^{(3)}(4, 0))^2} \quad (87)$$

$$= 2 \ln 2 [\mathcal{C}(\sigma'^2) - \alpha'] - \ln(1 + e^{-\frac{1}{\sigma}}) - \frac{1.8281}{\alpha'}. \quad (88)$$

which simplifies to is a quadratic equation with respect to α' . A candidate solution

$$\alpha' = \alpha^{**} = \mathcal{C}(\sigma'^2) - \frac{1.08}{\mathcal{C}(\sigma'^2)} \quad (89)$$

guarantees $u(\alpha', \sigma'^2) > 0$ for σ'^2 such that $\alpha^{**} \geq 4$ ($\mathcal{C}(\sigma'^2) \geq 4.26$). We can now bound the gap to capacity based on parameters α^{**}, σ'^2

$$\mathcal{C}(\sigma'^2) - \alpha^{**} \mathcal{C}_{\text{BIAWGN}}\left(\frac{1}{\alpha^{**}x^{(1)}}\right) \leq \mathcal{C}(\sigma'^2) \left[1 - \mathcal{C}_{\text{BIAWGN}}\left(\frac{1}{\alpha^{**}x^{(1)}}\right)\right] + \frac{1.08}{\mathcal{C}(\sigma'^2)} \quad (90)$$

$$\leq 2 \ln 2 \mathcal{C}(\sigma'^2) \int_{\frac{1}{\alpha'x^{(1)}}}^{\infty} g_2(\gamma) d\gamma + \frac{1.08}{\mathcal{C}(\sigma'^2)} \leq 2 \ln 2 \mathcal{C}(\sigma'^2) \int_{\frac{1}{\alpha^{**}x^{(1)}}}^{\infty} \pi Q(\sqrt{\gamma}) d\gamma + \frac{1.08}{\mathcal{C}(\sigma'^2)} \quad (91)$$

$$\leq 2 \ln 2 \mathcal{C}(\sigma'^2) 2\pi \exp\left(-\frac{1}{2x^{(1)}}\right) + \frac{1.08}{\mathcal{C}(\sigma'^2)} \leq 2 \ln 2 \mathcal{C}(\sigma'^2) 2\pi \exp\left(-\frac{1}{4\sigma'^2}\right) + \frac{1.08}{\mathcal{C}(\sigma'^2)} \leq \frac{1.081}{\mathcal{C}(\sigma'^2)} \quad (92)$$

where we used an upper bound on the MSE function from [12], the MSE-capacity relationship [45], and then an upper bound $Q(\sqrt{\gamma}) \leq -\exp(\gamma/2)$.

We then bound the gap between the actual capacity

$$\mathcal{C}(\sigma^2) - \mathcal{C}(\sigma'^2) = \frac{1}{2} \log_2 \left(\frac{\sigma^2 + 1}{\sigma'^2 + 1 - \frac{1}{L}} \right)$$

and the achievable rate

$$\alpha^{**} \mathcal{C}_{\text{BIAWGN}}\left(\frac{1}{\alpha^{**}x^{(1)}}\right) - \mathcal{R}_{\text{coup}}(\alpha^{**}, \sigma^2) \leq \alpha^{**} \mathcal{C}_{\text{BIAWGN}}\left(\frac{1}{\alpha^{**}x^{(1)}}\right) - \alpha \mathcal{C}_{\text{BIAWGN}}\left(\frac{1}{\alpha^{**}x^{(1)}}\right) \quad (93)$$

$$\leq \alpha^{**} - \alpha = \frac{L-1}{M-1} - \frac{L}{M} \leq \frac{\alpha-1}{M-1} \leq \frac{\mathcal{C}(\sigma^2)-1}{M-1} \quad (94)$$

The Theorem is proved. \square

D. Proof of Theorem 2

Consider (α, σ^2) satisfying the conditions of Theorem 1 with σ^2 chosen such that $\mathcal{C}(\sigma^2) > 4.26$ and

$$\alpha \in [\mathcal{C}(\sigma^2) - 1.081/\mathcal{C}(\sigma^2), \mathcal{C}(\sigma^2)]. \quad (95)$$

We apply the low-SNR MSE $g_{\text{mse}}(\eta)$ series expansion [46] for $\eta \rightarrow 0$

$$g_{\text{mse}}(\eta) = 1 - \eta + \eta^2 + \mathcal{O}(\eta^3) \quad (96)$$

which is equivalent to the fact that there exists an $\eta_1 > 0$ and a constant $C_1 > 0$ such that for $\eta \in [0, \eta_1]$

$$g_{\text{mse}}(\eta) = 1 - \eta + \eta^2 + f(\eta) \quad \text{where} \quad |f(\eta)| \leq C_1 \eta^3. \quad (97)$$

Consider now the third root $x^{(3)}$ of the characteristic equation

$$x^{(3)} = g_{\text{mse}}\left(\frac{1}{\alpha x^{(3)}}\right) + \sigma^2 \quad (98)$$

that, according to Lemma 2 (c) satisfies

$$1 - \frac{1}{\alpha} - \frac{2}{\alpha^2} \leq 1 + \sigma^2 - \frac{1}{\alpha(1 + \sigma^2)} - \frac{2}{\alpha^2(1 + \sigma^2)^3} \leq x^{(3)} \leq 1 + \sigma^2. \quad (99)$$

The lower bound (99) implies that $1/(\alpha x^{(3)}) \rightarrow 0$ for $\sigma^2 \rightarrow 0$ and α satisfying (95). Therefore, the series expansion (97) can be applied to (98). There exists a σ_1 such that $\sigma < \sigma_1$ implies $1/(\alpha x^{(3)}) < \eta_1$ and, therefore,

$$x^{(3)} = 1 - \frac{1}{\alpha x^{(3)}} + \frac{1}{(\alpha x^{(3)})^2} + f\left(\frac{1}{\alpha x^{(3)}}\right) \quad \text{where} \quad \left|f\left(\frac{1}{\alpha x^{(3)}}\right)\right| \leq C_1 \frac{1}{(\alpha x^{(3)})^3}. \quad (100)$$

Proposition 1. *There exists σ_2 and a constant $C_2 > 0$ such that for $\sigma < \sigma_2$ and α satisfying (95)*

$$x^{(3)} = 1 - \frac{1}{\alpha} + f_2\left(\frac{1}{\alpha}\right) \quad \text{where} \quad \left|f_2\left(\frac{1}{\alpha}\right)\right| \leq C_2 \frac{1}{\alpha^3}. \quad (101)$$

or equivalently

$$x^{(3)} = 1 - \frac{1}{\alpha} + \mathcal{O}\left(\frac{1}{\alpha^3}\right) \quad (102)$$

as $\sigma \rightarrow 0$ and $\alpha \rightarrow \infty$ (α satisfies (95)).

Proof. First of all (99) directly implies that there exists a σ_2 such that for $\sigma < \sigma_2$ and α satisfying (95)

$$\left|x^{(3)} - \left(1 - \frac{1}{\alpha}\right)\right| \leq \frac{3}{\alpha^2}.$$

Consider now (100) and subtract the first two terms on the right hand side of (101) from it

$$\left|x^{(3)} - \left(1 - \frac{1}{\alpha}\right)\right| = \left|1 - \frac{1}{\alpha x^{(3)}} + \frac{1}{(\alpha x^{(3)})^2} + f\left(\frac{1}{\alpha x^{(3)}}\right) - 1 + \frac{1}{\alpha} + \frac{1}{\alpha^2} - \frac{1}{\alpha^2}\right| \quad (103)$$

$$= \left|\frac{1}{\alpha} \left(1 - \frac{1}{x^{(3)}} + \frac{1}{\alpha}\right) - \frac{1}{\alpha^2} \left(1 - \frac{1}{(x^{(3)})^2}\right) + f\left(\frac{1}{\alpha x^{(3)}}\right)\right| \quad (104)$$

$$\leq \frac{1}{\alpha^2 x^{(3)}} \left|\alpha x^{(3)} - \alpha + x^{(3)}\right| + \frac{1}{\alpha^2} \left|1 - \frac{1}{(x^{(3)})^2}\right| + \frac{C_1}{\alpha^3} \quad (105)$$

$$\leq \frac{2}{\alpha^2} \left|\alpha x^{(3)} - \alpha + x^{(3)}\right| + \frac{1}{\alpha^2} \left|x^{(3)} - 1\right| \frac{1 + x^{(3)}}{(x^{(3)})^2} + \frac{C_1}{\alpha^3} \quad (106)$$

$$\leq \frac{2}{\alpha^2} \left(\frac{1}{\alpha} + \frac{6}{\alpha^2}\right) + \frac{1}{\alpha^2} \left(\frac{1}{\alpha} + \frac{3}{\alpha^2}\right) 12 + \frac{C_1}{\alpha^3} \leq \frac{C_2}{\alpha^3} \quad (107)$$

where C_2 is a constant. Proposition is proved. \square

Proposition 1 implies

$$\ln(x^{(3)}) = -\frac{1}{\alpha} - \frac{1}{2} \frac{1}{\alpha^2} + \mathcal{O}\left(\frac{1}{\alpha^3}\right)$$

as $\sigma \rightarrow 0$ and $\alpha \rightarrow \infty$ (α satisfies (95)). Based on the relationship between the MSE (96) and the channel capacity [46] we have

$$\mathcal{C}_{\mathcal{A}}(\eta) = \frac{1}{2 \ln 2} \left(\eta - \frac{\eta^2}{2} + \frac{\eta^3}{3} \right) + \mathcal{O}(\eta^3) \quad (108)$$

for $\eta \rightarrow 0$ where the fraction $1/(2 \ln 2)$ is applied since the information we consider is measured in bits. Hence, using the above capacity expression and Proposition for $\eta = 1/(\alpha x^{(3)})$ we obtain

$$2 \ln 2 \alpha \mathcal{C}_{\mathcal{A}}(\eta) = \frac{1}{x^{(3)}} - \frac{1}{2 \alpha (x^{(3)})^2} + \frac{1}{3 \alpha^2 (x^{(3)})^3} + \mathcal{O}\left(\frac{1}{\alpha^3}\right) \quad (109)$$

$$= 1 + \frac{1}{\alpha} + \frac{1}{\alpha^2} - \frac{1}{2 \alpha} \left(1 + \frac{2}{\alpha} \right) + \frac{1}{3 \alpha^2} + \mathcal{O}\left(\frac{1}{\alpha^3}\right) = 1 + \frac{1}{2 \alpha} + \frac{1}{3 \alpha^2} + \mathcal{O}\left(\frac{1}{\alpha^3}\right) \quad (110)$$

as $\sigma \rightarrow 0$ and $\alpha \rightarrow \infty$ (α satisfies (95)). Coming back to the potential function (77) we compute the asymptotics of the potential function in terms of α

$$u(\alpha, \sigma^2) = \ln \frac{x^{(3)}}{x^{(1)}} - \frac{\sigma^2}{x^{(1)}} + \frac{\sigma^2}{x^{(3)}} - 2 \ln 2 \alpha \mathcal{C}_{\mathcal{A}}\left(\frac{1}{\alpha x^{(1)}}\right) + 2 \ln 2 \alpha \mathcal{C}_{\mathcal{A}}\left(\frac{1}{\alpha x^{(3)}}\right) \quad (111)$$

$$= \ln \sigma^{-2} + \ln x^{(3)} - 1 - 2 \ln 2 \alpha + 2 \ln 2 \alpha \mathcal{C}_{\mathcal{A}}\left(\frac{1}{\alpha x^{(3)}}\right) + \mathcal{O}\left(\frac{1}{\alpha^3}\right) \quad (112)$$

$$= \ln \left(1 + \frac{1}{\sigma^2} \right) - \frac{1}{\alpha} - \frac{1}{2 \alpha^2} - 1 - 2 \ln 2 \alpha + 1 + \frac{1}{2 \alpha} + \frac{1}{3 \alpha^2} + \mathcal{O}\left(\frac{1}{\alpha^3}\right) \quad (113)$$

$$= \ln \left(1 + \frac{1}{\sigma^2} \right) - \frac{1}{2 \alpha} - \frac{1}{6 \alpha^2} - 2 \ln 2 \alpha + \mathcal{O}\left(\frac{1}{\alpha^3}\right) \quad (114)$$

as $\sigma \rightarrow 0$ and $\alpha \rightarrow \infty$ (α satisfies (95)). By selecting

$$\alpha = \mathcal{C}(\sigma^2) - \frac{1}{4 \ln 2 \mathcal{C}(\sigma^2)} + \frac{1}{12 \ln 2 \mathcal{C}(\sigma^2)^2} + \mathcal{O}\left(\frac{1}{\mathcal{C}(\sigma^2)^3}\right)$$

we ensure that the potential function is positive and the iterative decoding converges to the smallest root $x^{(1)}$ of the characteristic equation. The resulting gap to capacity then scales as

$$\mathcal{C}(\sigma^2) - \alpha \mathcal{C}_{\mathcal{A}}\left(\frac{1}{\alpha x^{(1)}}\right) = \frac{1}{4 \ln 2 \mathcal{C}(\sigma^2)} + \frac{1}{12 \ln 2 \mathcal{C}(\sigma^2)^2} + \mathcal{O}\left(\frac{1}{\mathcal{C}(\sigma^2)^3}\right) \quad (115)$$

for $\sigma \rightarrow 0$. The Theorem is proved.