Coded Many-User Multiple Access via Approximate Message Passing

Xiaoqi Liu
University of Cambridge, UK
xl394@cam.ac.uk

Kuan Hsieh
Luminance Technologies Ltd, UK
kuanhsieh95@gmail.com

Ramji Venkataramanan
University of Cambridge, UK
rv285@cam.ac.uk

***Abstract*—We consider communication over the Gaussian multiple-access channel in the regime where the number of users grows linearly with the codelength. We investigate coded CDMA schemes where each user's information is encoded via a linear code before being modulated with a signature sequence. We propose an efficient approximate message passing (AMP) decoder that can be tailored to the structure of the linear code, and provide an exact asymptotic characterization of its performance. Based on this result, we consider a decoder that integrates AMP and belief propagation and characterize the tradeoff between spectral efficiency and signal-to-noise ratio, for a given target error rate. Simulation results are provided to demonstrate the benefits of the concatenated scheme at finite lengths.**

I. Introduction

We consider communication over an L-user Gaussian multiple access channel (GMAC), which has output of the form

$$\boldsymbol{y} = \sum_{\ell=1}^{L} \boldsymbol{c}_\ell + \boldsymbol{\varepsilon}\,, \tag{1}$$

over n channel uses. Here $\boldsymbol{c}_\ell \in \mathbb{R}^n$ is the codeword of the ℓ-th user and $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ is the channel noise. Motivated by modern applications in machine-type communications, a number of recent works have studied the GMAC in the *many-user* or *many-access* setting, where the number of users L grows with the block length n [1]–[4].

In this paper, we study the many-user regime where $L, n \to \infty$ with the user density $\mu := L/n$ converging to a constant. Each user transmits a fixed number of bits k (payload) under a constant energy-per-information-bit constraint $\|\boldsymbol{c}_\ell\|_2^2/k \le E_b$. Moreover, the spectral efficiency is the total user payload per channel use, denoted by $\mathsf{S} = (Lk)/n = \mu k$. In this regime, a key question is to understand the tradeoff between user density (or spectral efficiency), the signal-to-noise ratio E_b/N_0, and the probability of decoding error. Here $N_0 = 2\sigma^2$ is the noise spectral density. A popular measure of decoding performance is the per-user probability of error (PUPE), defined as

$$\texttt{PUPE} := \frac{1}{L} \sum_{\ell=1}^{L} \mathbb{P}(\boldsymbol{c}_\ell \neq \hat{\boldsymbol{c}}_\ell), \tag{2}$$

where $\hat{\boldsymbol{c}}_\ell$ is the decoded codeword for user ℓ.

Polyanskiy [2] and Zadik et al. [3] obtained converse and achievability bounds on the minimum E_b/N_0 required to achieve $\texttt{PUPE} \le \epsilon$ for a given $\epsilon > 0$, when the user density μ and user payload k are fixed. These bounds were extended to the multiple-access channels with Rayleigh fading in [5], [6]. The achievability bounds in these works are obtained using Gaussian random codebooks and joint maximum-likelihood decoding, which is computationally infeasible.

Coding schemes: Efficient coding schemes for the many-user GMAC, based on random linear models and spatial coupling with Approximate Message Passing (AMP) decoding, were proposed in [7]. Using similar ideas, Kowshik obtained improved achievability bounds in [8]. In the schemes proposed in [7], each user's codeword is produced by directly multiplying a matrix with the user's information sequence. The simplest such scheme is binary CDMA, where each user $\ell \in [L]$ transmits one bit of information by modulating a signature sequence $\boldsymbol{a}_\ell \in \mathbb{R}^n$, i.e., the codeword $\boldsymbol{c}_\ell = \boldsymbol{a}_\ell x_l$ where $x_\ell \in \{\pm\sqrt{E}\}$. Thus the decoding problem is to recover the vector of information symbols $\boldsymbol{x} = [x_1, \dots, x_L]^\mathsf{T}$ from the channel output vector

$$\boldsymbol{y} = \sum_{\ell=1}^{L} \boldsymbol{a}_\ell x_\ell + \boldsymbol{\varepsilon} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\varepsilon}\,, \tag{3}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_L] \in \mathbb{R}^{n \times L}$ is the matrix of signature sequences. If each user wishes to transmit $k > 1$ bits in n channel uses, the binary CDMA scheme requires k blocks of transmission, with each block (and each signature sequence) having length n/k.

The optimal spectral efficiency of CDMA in the large system limit (with random signature sequences) has been studied in a number of works, e.g. [9]–[13]. Assuming the signature sequences are i.i.d. sub-Gaussian, the best known technique for efficiently decoding $\boldsymbol{x}$ from $\boldsymbol{y}$ in (3) is AMP [14], [15], a family of iterative algorithms that has its origins in relaxations of belief propagation [16], [17]. An attractive feature of AMP decoding is that it allows an exact asymptotic characterization of its error performance through a deterministic recursion called 'state evolution'.

Main Contributions: The binary CDMA scheme described by (3) transmits *uncoded* user information. In this paper, we show how the performance can be significantly improved by using a concatenated coding scheme in which each user's information sequence is first encoded using a linear code before being multiplied with the signature sequence.

This work was supported in part by a Schlumberger Cambridge International Scholarship.

We propose a flexible AMP decoder that can be tailored to the structure of the linear code, and provide an exact asymptotic characterization of its error performance (Theorem 1). Specifically, we show how a decoder for the underlying code, such as a maximum-likelihood or a belief propagation (BP) decoder, can be incorporated within the AMP algorithm with rigorous asymptotic guarantees (Corollary 1). Simulation results validate the theory and demonstrate the benefits of the concatenated scheme at finite lengths. We focus on binary CDMA to highlight the gains in the simplest setting. The concatenated scheme as well as the AMP decoder and its analysis can be extended to the general random linear model based schemes studied in [7]. These will be described in the extended version of this paper.

We emphasize that our setting is distinct from unsourced random access over the GMAC [2], [18]–[20], where all the users share the same codebook and only a subset of them are active. In our case, each user has a distinct signature sequence and all of them are active. While the latter is particularly relevant in designing grant-free communication systems, coding schemes for this setting often rely on dividing a common codebook into sections for different users. Extending the ideas in this paper to unsourced random access is an interesting future direction.

Related Work: AMP algorithms were first proposed for compressed sensing [14], [15] and its variants [21], and have since been applied to a range of problems including estimation in generalized linear models and low-rank matrix estimation. We refer the interested reader to [22] for a survey. In the context of communication over AWGN channels, AMP has been used as a decoder for sparse regression codes (SPARCs) [23]–[25] and for compressed coding [26]. SPARC-based concatenated schemes with AMP decoding have been proposed for both single-user AWGN channels [27]–[29] and unsourced random access [19], [20].

In most of these concatenated schemes [19], [26]–[28] the AMP decoder for SPARCs does not explicitly use the structure of the outer code (which is decoded separately). Two key exceptions are the SPARC-LDPC concatenated schemes in [20], [29], which use an AMP decoding algorithm with an integrated BP denoiser. Drawing inspiration from these works, in Section IV-C we propose an AMP decoder with a BP denoiser for our concatenated scheme. Our scheme and its decoder differ from those in [20], [29] in a few important ways: i) we do not use the SPARC message structure, and ii) we treat each user's codeword as a row of a signal matrix and devise an AMP algorithm with matrix iterates, a notable deviation from prior schemes where AMP operates on vectors.

Notation: We write $[L]$ for the set $\{1, \dots, L\}$. We use bold uppercase letters for matrices, bold lowercase for vectors, and plain font for scalars. We write $\boldsymbol{a}_\ell$ for the ℓ-th row or column of $\boldsymbol{A}$ depending on the context, and $a_{\ell,i}$ for its ith component. A function $f : \mathbb{R}^d \to \mathbb{R}^d$ returns a column vector when applied to a column vector, and likewise for row vectors.

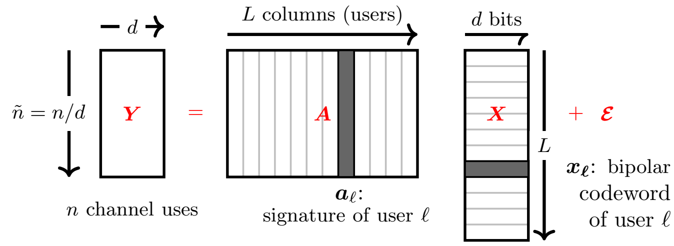

Figure 1. Concatenated coding scheme for GMAC

II. Concatenated coding scheme

The k-bit message of user ℓ, denoted by $\boldsymbol{u}_\ell \in \{0,1\}^k$, is mapped to a GMAC codeword $\boldsymbol{c}_\ell \in \mathbb{R}^n$ in two steps. First, a rate k/d linear code with generator matrix $\boldsymbol{G} \in \{0,1\}^{d\times k}$ is used to produce a d-bit binary codeword $\boldsymbol{G}\boldsymbol{u}_\ell \in \{0,1\}^d$. Each 0 code bit is then mapped to $\sqrt{E}$ and each 1 bit code bit to $-\sqrt{E}$ to produce $\boldsymbol{x}_\ell \in \{\pm\sqrt{E}\}^d$. The magnitude $\sqrt{E}$ of each BPSK symbol will be specified later in terms of the energy per bit constraint E_b. In the second step of encoding, for each user ℓ, we take the outer-product of $\boldsymbol{x}_\ell$ with a signature sequence $\boldsymbol{a}_\ell \in \mathbb{R}^{\tilde{n}}$, where $\tilde{n} := n/d$. This yields a matrix $\boldsymbol{C}_\ell = \boldsymbol{a}_\ell \boldsymbol{x}_\ell^\top \in \mathbb{R}^{\tilde{n}\times d}$. The final length-$n$ codeword transmitted by user ℓ is simply $\boldsymbol{c}_\ell = \text{vectorize}(\boldsymbol{C}_\ell) \in \mathbb{R}^n$.

Let $\boldsymbol{X} = \big[\boldsymbol{x}_1, \dots, \boldsymbol{x}_L\big]^\top \in \{\pm\sqrt{E}\}^{L\times d}$ be the signal matrix whose ℓth row $\boldsymbol{x}_\ell$ is the bipolar codeword of user ℓ. Let $\boldsymbol{A} = \big[\boldsymbol{a}_1, \dots, \boldsymbol{a}_L\big] \in \mathbb{R}^{\tilde{n}\times L}$ be the design matrix whose columns are the signature sequences. Then the channel output in (1) can be rewritten into matrix form:

$$\boldsymbol{Y} = \sum_{\ell=1}^{L} \boldsymbol{a}_\ell \boldsymbol{x}_\ell^\top + \boldsymbol{\mathcal{E}} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{\mathcal{E}} \in \mathbb{R}^{\tilde{n}\times d}. \tag{4}$$

See Fig. 1 for an illustration.

Assumptions: We consider i.i.d. Gaussian signature sequences. Specifically, we choose $A_{i\ell} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/\tilde{n})$, for $i \in [\tilde{n}]$, $\ell \in [L]$. We make the natural assumption that the information bits $\boldsymbol{u}_\ell \in \{0,1\}^k$ are uniformly random, for each user $\ell \in [L]$. We also assume that the noise variance σ^2 in (1) is known, a mild assumption since the i.i.d. Gaussian signature sequences allow σ^2 to be consistently estimated as $\hat{\sigma}^2 = \frac{\|\boldsymbol{Y}\|_F^2}{n} - Ed\mu$.

We consider the asymptotic limit where $L/n \to \mu$ as $n, L \to \infty$, for a user density $\mu > 0$ of constant order. We emphasize that d is fixed and does not scale as $n, L \to \infty$. Therefore $\tilde{n}/L = (n/d)/L \to 1/(d\mu)$ is also of constant order.

III. AMP decoder

The decoding task is to recover the signal matrix $\boldsymbol{X}$ from the channel observation $\boldsymbol{Y}$ in (4), given the design matrix $\boldsymbol{A}$ and the channel noise variance σ^2. A good decoder must take advantage of the prior distribution on $\boldsymbol{X}$: recall that each row of $\boldsymbol{X}$ is an independent codeword taking values in $\{\pm\sqrt{E}\}^d$, defined via the underlying rate k/d linear code. The prior distribution of each row of $\boldsymbol{X}$ induced by the linear code is denoted by $P_{\bar{\boldsymbol{x}}}$. Note that $P_{\bar{\boldsymbol{x}}}$ assigns equal probability to 2^k vectors in $\{\pm\sqrt{E}\}^d$.

The AMP decoder recursively produces estimates $\boldsymbol{X}^t \in \mathbb{R}^{L\times d}$ of $\boldsymbol{X}$ for iteration $t \geq 0$. This is done via a sequence of denoising functions η_t that can be tailored to the prior $P_{\bar{\boldsymbol{x}}}$. Starting from an initializer $\boldsymbol{X}^0 = \mathbf{0}_{L\times d}$, for $t \geq 0$ the AMP decoder computes:

$$\boldsymbol{Z}^t = \boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}^t + \frac{1}{n}\boldsymbol{Z}^{t-1}\left[\sum_{\ell=1}^{L} \eta'_{t-1}\left(\boldsymbol{s}_\ell^{t-1}\right)\right]^\top, \tag{5}$$

$$\boldsymbol{S}^t = \boldsymbol{A}^\top \boldsymbol{Z}^t + \boldsymbol{X}^t, \tag{6}$$

$$\boldsymbol{X}^{t+1} = \eta_t\left(\boldsymbol{S}^t\right), \tag{7}$$

where $\eta_t : \mathbb{R}^d \to \mathbb{R}^d$ applies row-wise to matrix inputs, and $\eta'_t(\boldsymbol{s}) = \frac{\mathrm{d}\eta_t(\boldsymbol{s})}{\mathrm{d}\boldsymbol{s}} \in \mathbb{R}^{d\times d}$ is the derivative (Jacobian) of η_t. Quantities with negative indices are set to all-zero matrices. When $d = 1$, (5)–(7) reduces to the classical AMP algorithm [15] for estimating a vector signal in a linear model.

State Evolution: In limit as $n, L \to \infty$ (with $L/n \to \mu$), the memory term $\frac{1}{n}\boldsymbol{Z}^{t-1}\left[\sum_{\ell=1}^{L} \eta'_{t-1}\left(\boldsymbol{s}_\ell^{t-1}\right)\right]^\top$ in (5) ensures that the row-wise empirical distribution of $\boldsymbol{Z}^t \in \mathbb{R}^{\tilde{n}\times d}$ converges to a Gaussian $\mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma}^t)$ for $t \geq 1$. Furthermore, the row-wise empirical distribution of $(\boldsymbol{S}^t - \boldsymbol{X}) \in \mathbb{R}^{L\times d}$ also converges to the same Gaussian $\mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma}^t)$. The covariance matrix $\boldsymbol{\Sigma}^t \in \mathbb{R}^{d\times d}$ is iteratively defined via the following state evolution recursion, for $t \geq 0$:

$$\boldsymbol{\Sigma}^{t+1} = \sigma^2 \boldsymbol{I} + \frac{1}{\delta}\mathbb{E}\left\{[\eta_t(\bar{\boldsymbol{x}} + \boldsymbol{g}^t) - \bar{\boldsymbol{x}}][\eta_t(\bar{\boldsymbol{x}} + \boldsymbol{g}^t) - \bar{\boldsymbol{x}}]^\top\right\}. \tag{8}$$

Here $\boldsymbol{I}$ is the $d \times d$ identity matrix, and $\boldsymbol{g}^t \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma}^t)$ is independent from $\bar{\boldsymbol{x}} \sim P_{\bar{\boldsymbol{x}}}$. The expectation in (8) is with respect to $\bar{\boldsymbol{x}}$ and $\boldsymbol{g}^t$, and the iteration is initialized with $\boldsymbol{\Sigma}^0 = (\sigma^2 + E/\delta)\boldsymbol{I}$.

The convergence of the row-wise empirical distribution of $\boldsymbol{S}^t$ to the law of $\bar{\boldsymbol{x}} + \boldsymbol{g}^t$ follows by applying standard results in AMP theory [22], [30]. This distributional characterization of $\boldsymbol{S}^t$ crucially informs the choice of the denoiser η_t. Specifically, for each row $\ell \in [L]$, the role of the denoiser η_t is to estimate the codeword $\boldsymbol{x}_\ell$ from an observation in zero-mean Gaussian noise with covariance matrix $\boldsymbol{\Sigma}^t$. In the next section, we discuss the Bayes-optimal denoiser and two other sub-optimal but computationally efficient denoisers. First, we provide a performance characterization of the AMP decoder with a generic Lipschitz-continuous denoiser in Theorem 1.

Decoding performance after t iterations of AMP decoding can be measured via either the user-error rate $\texttt{UER} = \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{1}\{\hat{\boldsymbol{x}}_\ell^{t+1} \neq \boldsymbol{x}_\ell\}$, or the bit-error rate $\texttt{BER} = \frac{1}{Ld}\sum_{\ell=1}^{L}\sum_{i=1}^{d} \mathbb{1}\left\{\hat{x}_{\ell,i}^{t+1} \neq x_{\ell,i}\right\}$. Here $\hat{\boldsymbol{x}}_\ell^{t+1} = h_t(\boldsymbol{s}_\ell^t)$ is a hard-decision estimate of the codeword $\boldsymbol{x}_\ell$, produced using a suitable function h_t, and $\hat{x}_{\ell,i}^{t+1}$ is the ith entry of $\hat{\boldsymbol{x}}_\ell^{t+1}$. For example, h_t may quantize each entry of $\boldsymbol{x}_\ell^{t+1} = \eta_t(\boldsymbol{s}_\ell^t)$ to a value in $\pm\sqrt{E}$. We note that the $\texttt{PUPE}$ defined in (2) is the expected value of the $\texttt{UER}$.

Theorem 1. *Consider the AMP decoding algorithm in* (5)–(7) *with Lipschitz continuous denoisers $\eta_t : \mathbb{R}^d \to \mathbb{R}^d$, for $t \geq 1$. Let $\hat{\boldsymbol{x}}_\ell^{t+1} = h_t(\boldsymbol{s}_\ell^t)$ be the hard-decision estimate in iteration t. The asymptotic* $\texttt{UER}$ *and* $\texttt{BER}$ *in iteration t satisfy the following almost surely, for $t \geq 1$:*

$$\lim_{L\to\infty} \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{1}\{\hat{\boldsymbol{x}}_\ell^{t+1} \neq \boldsymbol{x}_\ell\} = \mathbb{P}\left(h_t\left(\bar{\boldsymbol{x}} + \boldsymbol{g}^t\right) \neq \bar{\boldsymbol{x}}\right), \tag{9}$$

$$\lim_{L\to\infty} \frac{1}{Ld}\sum_{\ell=1}^{L}\sum_{i=1}^{d} \mathbb{1}\left\{\hat{x}_{\ell,i}^{t+1} \neq x_{\ell,i}\right\} = \frac{1}{d}\sum_{i=1}^{d} \mathbb{P}\left(\left[h_t\left(\bar{x} + g^t\right)\right]_i \neq \bar{x}_i\right). \tag{10}$$

Here $\bar{\boldsymbol{x}} \sim P_{\bar{\boldsymbol{x}}}$ and $\boldsymbol{g}^t \sim \mathcal{N}_d(\mathbf{0}, \boldsymbol{\Sigma}^t)$ are independent, with $\boldsymbol{\Sigma}_t$ defined by the state evolution recursion in (8).

The proof is given in the longer version of this paper.

IV. Choice of AMP denoiser η_t

A. Bayes-optimal denoiser

Since the row-wise distribution of $\boldsymbol{S}^t$ converges to the law of $\bar{\boldsymbol{x}} + \boldsymbol{g}^t$, the Bayes-optimal or minimum mean squared error (MMSE) denoiser η_t^{Bayes} estimates each row $\boldsymbol{X}$ as the following conditional expectation. For $\ell \in [L]$,

$$\begin{aligned}\boldsymbol{x}_\ell^{t+1} &= \eta_t^{\text{Bayes}}(\boldsymbol{s}_\ell^t) = \mathbb{E}\left[\bar{\boldsymbol{x}} \mid \bar{\boldsymbol{x}} + \boldsymbol{g}^t = \boldsymbol{s}_\ell^t\right] \\ &= \sum_{\boldsymbol{x}'\in\mathcal{X}} \boldsymbol{x}' \cdot \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x}' - 2\boldsymbol{s}_\ell^t)^\top(\boldsymbol{\Sigma}^t)^{-1}\boldsymbol{x}'\right)}{\sum_{\tilde{\boldsymbol{x}}'\in\mathcal{X}}\exp\left(-\frac{1}{2}(\tilde{\boldsymbol{x}}' - 2\boldsymbol{s}_\ell^t)^\top(\boldsymbol{\Sigma}^t)^{-1}\tilde{\boldsymbol{x}}'\right)}\end{aligned} \tag{11}$$

where $\mathcal{X} \subset \{\pm\sqrt{E}\}^d$ is the set of 2^k codewords. Since $|\mathcal{X}| = 2^k$, the cost of applying η_t^{Bayes} is $O(2^k d^3)$ which grows exponentially in k. In each iteration t, the decoder can produce a hard-decision maximum a posteriori (MAP) estimate $\hat{\boldsymbol{x}}_\ell^t$ from $\boldsymbol{s}_\ell^t$ via:

$$\hat{\boldsymbol{x}}_\ell^{t+1} = h_t(\boldsymbol{s}_\ell^t) = \arg\max_{\boldsymbol{x}'\in\mathcal{X}} \mathbb{P}\left(\bar{\boldsymbol{x}} = \boldsymbol{x}' \mid \bar{\boldsymbol{x}} + \boldsymbol{g}^t = \boldsymbol{s}_\ell^t\right). \tag{12}$$

In Section V (Fig. 2) we present numerical results illustrating the performance of AMP with denoiser η_t^{Bayes} for a Hamming code with $d = 7$ and $k = 4$. In practical scenarios where d is of the order of several hundreds or thousands, applying η_t^{Bayes} is not feasible, motivating the use of sub-optimal denoisers with lower computational cost.

B. Marginal-MMSE denoiser

A computationally efficient alternative to the Bayes-optimal denoiser is the marginal-MMSE denoiser [19], [27] which acts entry-wise on $\boldsymbol{s}_\ell^t$ and returns the entry-wise conditional expectation:

$$\boldsymbol{x}_\ell^{t+1} = \eta_t^{\text{marginal}}(\boldsymbol{s}_\ell^t) = \begin{bmatrix}\mathbb{E}[\bar{x}_1|\bar{x}_1 + g_1^t = s_{\ell,1}^t] \\ \vdots \\ \mathbb{E}[\bar{x}_d|\bar{x}_d + g_d^t = s_{\ell,d}^t]\end{bmatrix}, \text{ where}$$

$$\mathbb{E}[\bar{x}_i|\bar{x}_i + g_i^t = s_{\ell,i}^t] \overset{\text{(i)}}{=} \sqrt{E}\tanh\left(\sqrt{E}s_{\ell,i}^t/\Sigma_{i,i}^t\right). \tag{13}$$

The equality (i) follows from $g_i^t \sim \mathcal{N}(0, \Sigma_{i,i}^t)$ and $p(\bar{x}_i = \sqrt{E}) = p(\bar{x}_i = -\sqrt{E}) = \frac{1}{2}$ due to the linearity of the outer code. A hard decision estimate $\hat{\boldsymbol{x}}_\ell^{t+1}$ can be obtained by quantizing each entry of $\boldsymbol{x}_\ell^{t+1}$ to $\{\pm\sqrt{E}\}$.

This marginal denoiser has an $O(d)$ computational cost which is linear in d, but it ignores the parity structure of $\bar{\boldsymbol{x}}$, which is useful prior knowledge that can help reconstruction. One way to address this is by using the output of the AMP decoder as input to a channel decoder for the outer code, as in [19], [28]. In the next subsection, we show how to improve on this approach. Considering an outer LDPC code, we use an AMP denoiser that fully integrates BP decoding.

C. Belief Propagation (BP) denoiser

Assume that the binary linear code used to define the concatenated scheme is an LDPC code. We propose a BP denoiser η_t^{BP} which exploits the parity structure of the LDPC code in each AMP iteration by performing a few rounds of BP on the associated factor graph. Like the other denoisers above, η_t^{BP} acts row-wise on the effective observation $\boldsymbol{S}^t \in \mathbb{R}^{L\times d}$. For $\ell \in [L]$, it produces the updated AMP estimate $\boldsymbol{x}_\ell^{t+1}$ from $\boldsymbol{s}_\ell^t$ as follows, using R rounds of BP.

1) For each variable node $i \in [d]$ and check node $j \in [d-k]$, initialise variable-to-check messages (in log-likelihood ratio format) as:

$$L_{i\to j}^{(0)} = \ln\left[\frac{p(s_{\ell,i}^t \mid x_{\ell,i} = +\sqrt{E})}{p(s_{\ell,i}^t \mid x_{\ell,i} = -\sqrt{E})}\right] = \frac{2\sqrt{E}s_{\ell,i}^t}{\Sigma_{i,i}^t} =: L(s_{\ell,i}^t).$$

This initialization follows the distributional assumption $s_{\ell,i}^t \overset{\text{d}}{=} \bar{x}_i + g_i^t$, where $g_i^t \sim \mathcal{N}(0, \Sigma_{i,i}^t)$ and $p(\bar{x}_i = \sqrt{E}) = p(\bar{x}_i = -\sqrt{E}) = \frac{1}{2}$.

2) Let $N(i)$ denote the set of neighbouring nodes of node i. For rounds $1 \le \mathsf{r} \le \mathsf{R}$, compute the check-to-variable and variable-to-check messages, denoted by $L_{j\to i}^{(\mathsf{r})}$ and $L_{i\to j}^{(\mathsf{r})}$, as:

$$L_{j\to i}^{(\mathsf{r})} = 2\tanh^{-1}\left[\prod_{i'\in N(j)\setminus i} \tanh\left(\frac{1}{2}L_{i'\to j}^{(\mathsf{r}-1)}\right)\right], \tag{14}$$

$$L_{i\to j}^{(\mathsf{r})} = L(s_{\ell,i}^t) + \sum_{j'\in N(i)\setminus j} L_{j'\to i}^{(\mathsf{r})}. \tag{15}$$

3) Terminate BP after R rounds by computing the final log-likelihood ratio for each variable node $i \in [d]$:

$$L_i^{(\mathsf{R})} = L(s_{\ell,i}^t) + \sum_{j'\in N(i)} L_{j'\to i}^{(\mathsf{R})}. \tag{16}$$

Equations (14)-(16) are the standard BP updates for an LDPC code [31].

4) Compute the updated AMP estimate $\boldsymbol{x}_\ell^{t+1} = \eta_t^{\text{BP}}(\boldsymbol{s}_\ell^t)$, where

$$\left[\eta_t^{\text{BP}}(\boldsymbol{s}_\ell^t)\right]_i = \sqrt{E}\,\tanh\left(L_i^{(\mathsf{R})}/2\right), \quad i \in [d]. \tag{17}$$

The RHS above is obtained by converting the final log-likelihood ratio (16) to a conditional expectation, recalling that $x_{\ell,i}$ takes values in $\{\pm\sqrt{E}\}$. Following the standard interpretation of BP as approximating the bit-wise marginal posterior probabilities [31], the expression in (17) can be viewed as an approximation to

$$\mathbb{E}[\bar{x}_i \mid \bar{x}_i + g_i^t = s_{\ell,i}^t\,,\ \text{parities specified by } N(i) \text{ satisfied}].$$

We highlight the contrast between the conditional expectation above and the one in (13), which does not use the parity check constraints. As with the marginal-MMSE denoiser, a hard-decision estimate $\hat{\boldsymbol{x}}_\ell^{t+1}$ can be obtained by quantizing each entry of $\boldsymbol{x}_\ell^{t+1}$ to $\{\pm\sqrt{E}\}$. The computational cost of η^{BP} is $O(d\mathsf{R})$ which is linear in d.

While the derivative η_t' for the memory term can be easily calculated for η_t^{Bayes} and η_t^{marginal} via direct differentiation, the derivative for η_t^{BP} is less obvious because it involves R rounds of BP updates (14)-(16). Nevertheless, using the approach in [20], [29], the derivative can be derived in closed form provided the number of BP rounds R is less than the girth of the LDPC factor graph.

Lemma 1. *For $t \geq 1$, consider the AMP decoder with denoiser $\eta_t^{BP} : \mathbb{R}^d \to \mathbb{R}^d$, where BP is performed for fewer rounds than the girth of the LDPC factor graph. Let*

$$\boldsymbol{J} := \frac{d\eta_t^{BP}(\boldsymbol{s}_\ell^t)}{d\boldsymbol{s}_\ell^t} \in \mathbb{R}^{d\times d}, \quad \textit{for } \boldsymbol{s}_\ell^t \in \mathbb{R}^d. \tag{18}$$

Then for $i, h \in [d]$ and $i \neq h$,

$$J_{i,i} = \frac{1}{\Sigma_{i,i}^t}\left(E - \left[\eta_t^{BP}(\boldsymbol{s}_\ell^t)\right]_i^2\right), \quad J_{i,h} = 0. \tag{19}$$

The proof is given in the longer version of this paper. Lemma 1 implies that the characterization of the limiting `UER` and `BER` in Theorem 1 holds for AMP decoding with η^{BP}.

Corollary 1. *The asymptotic guarantees in Theorem 1 hold for the AMP decoder with any of the three denoisers: $\eta_t^{Bayes}, \eta_t^{marginal}$, and for η_t^{BP} assuming that the number of BP rounds is less than the girth of the LDPC factor graph.*

Proof. It can be verified by direct differentiation that the η_t^{Bayes} and η_t^{marginal} are bounded. For η^{BP}, we only need to show that $J_{i,i}$ in (19) is bounded for $i \in [d]$. This follows by observing that $\Sigma_{i,i}^t > \sigma^2$ (from (8)) and $[\eta^{\text{BP}}(\boldsymbol{s}_\ell^t)]_i^2 < E$ (from (17)). ☐

V. Numerical results

We numerically evaluate the tradeoffs achieved by concatenated coding scheme with different denoisers. For a target `BER` $= 10^{-4}$, we plot the maximum spectral efficiency $\mathsf{S} = Lk/n = (L/\tilde{n})(k/d)$ achievable as a function of signal-to-noise ratio $E_b/N_0 = (Ed/k)/(2\sigma^2)$. We use `BER` rather than `UER` since the `UER` of an uncoded scheme degrades approximately linearly with d. For each setting, we also plot the converse bounds from [3], and the achievability bounds from either [3] or [7] depending on which one yields the larger achievable region. These bounds can be adapted to obtain upper and lower bounds on the maximum spectral efficiency achievable for given values of E_b/N_0, per-user payload k, and `PUPE`. To adapt these bounds to target `BER` (rather than target

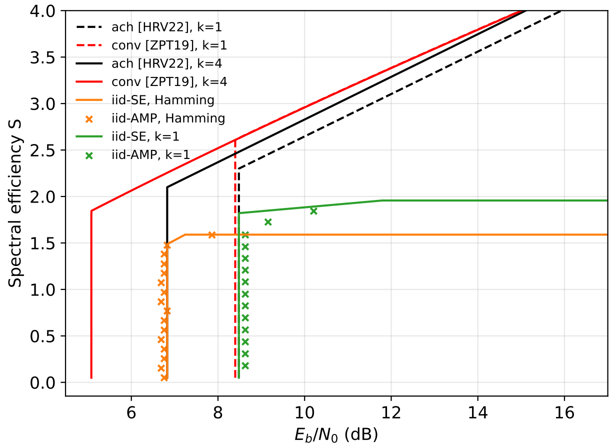

Figure 2. Comparison of the uncoded scheme and the concatenated scheme with $(7, 4)$ Hamming outer code and denoiser η_t^{Bayes}. $L = 20000$.

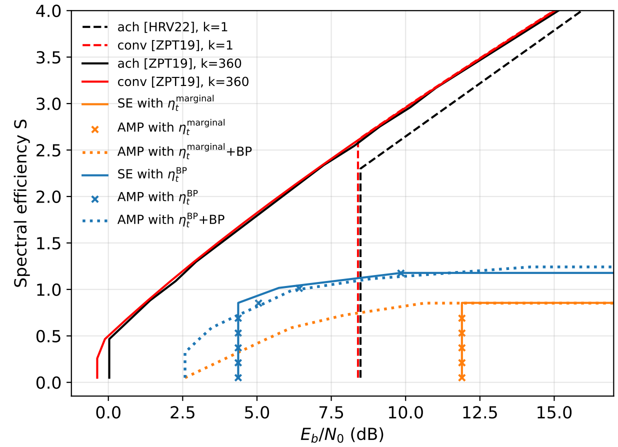

Figure 3. Comparison of marginal-MMSE denoiser η_t^{marginal} (orange) and BP denoiser η_t^{BP} (blue) for decoding LDPC outer code (with fixed rate 1/2). The dotted plots correspond to AMP decoding coupled with 200 rounds of BP after AMP has converged. $L = 2000, d = 720$.

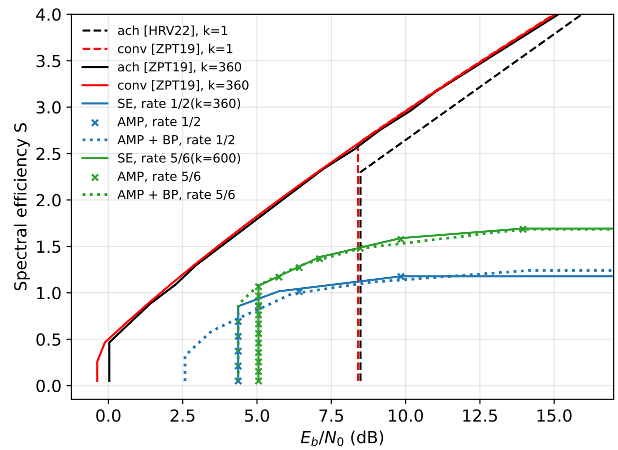

Figure 4. Comparison of LDPC outer code with rate 1/2 (blue) or 5/6 (green) under AMP decoding with BP denoiser η_t^{BP}. Dotted curves correspond to AMP decoding coupled with 200 rounds of BP after AMP has converged. $L = 2000, d = 720$.

PUPE), we using the random coding assumption that when a codeword is decoded incorrectly, approximately half of its bits are in error, i.e. $\mathbb{E}[\text{BER}] = \frac{1}{2}\text{PUPE}$.

In all the figures, 'SE' refers to curves obtained by using the state evolution result of Theorem 1, and 'AMP' (indicated by crosses) refers to points obtained via simulation. Fig. 2 compares the uncoded case (i.e. $d = 1$) with the concatenated scheme with a $(7, 4)$ Hamming code, decoded using AMP with Bayes-optimal denoiser η_t^{Bayes}. Even this simple code provides a savings of over 1dB in the minimum E_b/N_0 required to achieve positive spectral efficiency, compared to the uncoded scheme as well as the converse bound for $k = 1$.

Figs. 3 and 4 employ LDPC codes from the IEEE 802.11n standards as outer codes (with code length $d = 720$ bits). Fig. 3 compares the decoding performance of AMP with different denoisers: the marginal-MMSE η_t^{marginal} or the BP denoiser η_t^{BP} which executes 5 rounds of BP per AMP denoising step. The latter outperforms the former by around 7.5dB since η_t^{marginal} does not use the parity constraints of the code. The dotted orange curve in Fig. 3 shows that the performance of AMP with η_t^{marginal} is substantially improved by running a BP decoder (200 rounds) after AMP has converged. However, this additional BP decoding at the end also improves the performance of η_t^{BP} (blue dotted curve). We observe that the achievable spectral efficiency with $\eta_t^{\text{BP}}+$ BP is consistently about 40% higher than with $\eta_t^{\text{marginal}}+$ BP.

Fig. 4 compares the performance of the concatenated scheme with LDPC codes with different rates: 1/2 and 5/6. The AMP denoiser is η_t^{BP}, and the dotted curves show the effect of adding BP decoding (200 rounds) after AMP convergence. The code with the higher rate $5/6$ achieves higher spectral efficiency for large values of E_b/N_0, but the rate $1/2$ code achieves positive spectral efficiency for smaller E_b/N_0 values.

In all the figures, the asymptotic performance of AMP, predicted by state evolution, closely tracks its actual performance at large, finite L. Moreover, considering a metropolitan area with 10^6 to 10^7 devices and each device active a few times per hour, the user density μ is typically 10^{-4} to 10^{-3} [3, Remark 3]. For user densities in this range and per-user payload k on the order of 10^2 to 10^3, the spectral efficiency $\mathsf{S} = \mu k$ is less than 1. In all figures, the concatenated coding schemes exhibit the most substantial improvements for $\mathsf{S} < 1$.

Discussion: Though we assumed an i.i.d. Gaussian design matrix $\boldsymbol{A}$, using recent results on AMP universality [32], [33], the decoding algorithm and all the theoretical results remain valid for a much broader class of 'generalized white noise' matrices. This class includes i.i.d. sub-Gaussian matrices, so the results apply to the popular setting of random binary-valued signature sequences. Figs. 2–4 show that the gap between the spectral efficiency achieved by the concatenated scheme and the converse bounds grows with E_b/N_0. The spectral efficiency of our scheme can be substantially improved by using a spatially coupled design matrix [7]. This will be investigated in an extended version of this paper.

Acknowledgment: We thank Jossy Sayir for sharing his implementation of the belief propagation LDPC decoder.

References

[1] X. Chen, T.-Y. Chen, and D. Guo, "Capacity of Gaussian many-access channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3516–3539, 2017.

[2] Y. Polyanskiy, "A perspective on massive random-access," in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 2523–2527.

[3] I. Zadik, Y. Polyanskiy, and C. Thrampoulidis, "Improved bounds on Gaussian MAC and sparse regression via Gaussian inequalities," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 430–434.

[4] J. Ravi and T. Koch, "Scaling laws for Gaussian random many-access channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2429–2459, 2022.

[5] S. S. Kowshik, K. Andreev, A. Frolov, and Y. Polyanskiy, "Energy efficient random access for the quasi-static fading MAC," in *Proc. IEEE Int. Symp. Inf. Theory*, 2019, pp. 2768–2772.

[6] S. S. Kowshik and Y. Polyanskiy, "Fundamental limits of many-user MAC with finite payloads and fading," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 5853–5884, 2021.

[7] K. Hsieh, C. Rush, and R. Venkataramanan, "Near-optimal coding for many-user multiple access channels," *IEEE Journal on Selected Areas in Information Theory*, vol. 3, no. 1, pp. 21–36, 2022.

[8] S. S. Kowshik, "Improved bounds for the many-user MAC," in *Proc. IEEE Int. Symp. Inf. Theory*, 2022, pp. 2874–2879.

[9] S. Verdu and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, 1999.

[10] S. Shamai and S. Verdu, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, 2001.

[11] G. Caire, S. Guemghar, A. Roumy, and S. Verdu, "Maximizing the spectral efficiency of coded CDMA under successive decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 152–164, 2004.

[12] 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, 2002.

[13] D. Guo and S. Verdú, "Randomly spread CDMA: asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.

[14] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[15] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 764–785, 2011.

[16] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *J. Phys. A*, vol. 36, pp. 11 111–11 121, 2003.

[17] J. Boutros and G. Caire, "Iterative multiuser joint decoding: unified framework and asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1772–1793, 2002.

[18] V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "A coded compressed sensing scheme for unsourced multiple access," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6509–6533, 2020.

[19] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6894–6915, 2021.

[20] V. K. Amalladinne, A. K. Pradhan, C. Rush, J.-F. Chamberland, and K. R. Narayanan, "Unsourced random access with coded compressed sensing: Integrating AMP and belief propagation," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2384–2409, 2022.

[21] J. Ziniel and P. Schniter, "Efficient high-dimensional inference in the multiple measurement vector problem," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 340–354, 2013.

[22] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth, "A unifying tutorial on approximate message passing," *Foundations and Trends in Machine Learning*, vol. 15, no. 4, pp. 335–536, 2022.

[23] J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity achieving sparse superposition codes," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, pp. 4894–4927, Aug. 2017.

[24] C. Rush, A. Greig, and R. Venkataramanan, "Capacity-achieving sparse superposition codes via approximate message passing decoding," *IEEE Trans. Inf. Theory*, vol. 63, no. 3, pp. 1476–1500, Mar. 2017.

[25] R. Venkataramanan, S. Tatikonda, and A. Barron, "Sparse regression codes," *Foundations and Trends® in Communications and Information Theory*, vol. 15, no. 1-2, pp. 1–195, 2019.

[26] S. Liang, C. Liang, J. Ma, and L. Ping, "Compressed coding, AMP-based decoding, and analog spatial coupling," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7362–7375, 2020.

[27] A. Greig and R. Venkataramanan, "Techniques for improving the finite length performance of sparse superposition codes," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 905–917, 2018.

[28] H. Cao and P. O. Vontobel, "Using list decoding to improve the finite-length performance of sparse regression codes," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4282–4293, 2021.

[29] J. R. Ebert, J.-F. Chamberland, and K. R. Narayanan, "On sparse regression LDPC codes," in *Proc. IEEE Int. Symp. Inf. Theory*, 2023.

[30] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *Information and Inference: A Journal of the IMA*, vol. 2, no. 2, pp. 115–144, 2013.

[31] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[32] T. Wang, X. Zhong, and Z. Fan, "Universality of approximate message passing algorithms and tensor networks," 2022, arXiv:2206.13037.

[33] N. Tan, J. Scarlett, and R. Venkataramanan, "Approximate message passing with rigorous guarantees for pooled data and quantitative group testing," 2023, arXiv:2309.15507.