

# Near-Optimal Coding for Many-user Multiple Access Channels

Kuan Hsieh, Cynthia Rush, and Ramji Venkataramanan

**Abstract**—This paper considers the Gaussian multiple-access channel in the asymptotic regime where the number of users grows linearly with the code length. We propose efficient coding schemes based on random linear models with approximate message passing (AMP) decoding and derive the asymptotic error rate achieved for a given user density, user payload (in bits), and user energy. The tradeoff between energy-per-bit and achievable user density (for a fixed user payload and target error rate) is studied. It is demonstrated that in the large system limit, a spatially coupled coding scheme with AMP decoding achieves near-optimal tradeoffs for a wide range of user densities. Furthermore, in the regime where the user payload is large, we also study the tradeoff between energy-per-bit and spectral efficiency and discuss methods to reduce decoding complexity.

**Index Terms**—Multiple access, approximate message passing, spatial coupling, sparse superposition codes.

## I. INTRODUCTION

IN certain modes of modern communications, such as in massive machine-type communications, a large number of devices simultaneously transmit to a single receiver. Furthermore, the data payload of each user (or device) in such applications may be small, e.g., temperature readings from a wireless sensor network. These modes of communications do not lend themselves well to traditional large system analysis of multi-user systems, which often assume a fixed (small) number of users, and the size of the user payload scaling linearly with the code length (as information rate is the metric of interest), see for example [1, Chpt. 15] and [2, Chpt. 4]. This has motivated the study of multi-user channels in the *many-user* or *many-access* setting [3], where the number of users grows with the code length. This paper studies the asymptotic achievability of efficient coding schemes for the Gaussian multiple access channel (MAC) in the many-user setting.

The  $L$ -user Gaussian MAC produces its output  $\mathbf{y} \in \mathbb{R}^n$  according to

$$\mathbf{y} = \sum_{\ell=1}^L \mathbf{c}_\ell + \mathbf{w}, \quad (1)$$

where  $\mathbf{c}_\ell \in \mathbb{R}^n$  is the codeword of user  $\ell \in \{1, \dots, L\}$  and the noise vector  $\mathbf{w} \in \mathbb{R}^n$  has independent and identically distribution (i.i.d.) zero mean Gaussian entries with variance  $\sigma^2 > 0$ .

This work was supported in part by an Engineering and Physical Sciences Research Council Doctoral Training Partnership award. This paper was presented in part at the 2021 IEEE International Symposium on Information Theory. K. Hsieh and R. Venkataramanan are with Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK (e-mail: kh525@cam.ac.uk; rv285@cam.ac.uk). C. Rush is with the Department of Statistics, New York, NY 10027, Columbia University, USA (e-mail: cynthia.rush@columbia.edu).

This paper first studies the symmetric Gaussian MAC in the asymptotic regime proposed by Polyanskiy et al. [4], [5] where: (i) the number of users  $L$  grows linearly with the code length  $n$ , i.e.,  $L = \mu n$  for some fixed user density  $\mu \in (0, \infty)$ , (ii) the number of bits transmitted by each user (user payload), denoted by  $\log_2 M$ , is fixed and independent of  $n$ , and (iii) the energy-per-bit, denoted by  $E_b$ , is fixed and independent of  $n$ . The energy needed to transmit each user's payload is  $E = E_b \log_2 M$ . The decoding metric used is the *user error rate* (UER), which specifies the fraction of user messages decoded in error:

$$\text{UER} = \frac{1}{L} \sum_{\ell=1}^L \mathbb{1} \{ \hat{\mathbf{x}}_\ell \neq \mathbf{x}_\ell \}, \quad (2)$$

where  $\mathbf{x}_\ell$  denotes the message sent by user  $\ell$ , and  $\hat{\mathbf{x}}_\ell$  is the decoder's estimate of the message. The *per-user probability of error* (PUPE) error criterion used in [4], [5] is the expected value of the UER.

In this asymptotic regime, the works of Polyanskiy et al. [4], [5] obtained converse and achievability bounds on the minimum  $E_b/N_0$  required to achieve a decoding error of  $\text{PUPE} \leq \epsilon$  for a given  $\epsilon > 0$ , when the user density  $\mu$  and user payload  $\log_2 M$  are fixed. Here  $N_0 = 2\sigma^2$  is the noise spectral density. The achievability bound was based on the coding scheme where users encode their messages with i.i.d. Gaussian codebooks, and messages are decoded with (joint) maximum likelihood (ML) decoding. We note that it is infeasible to implement the ML decoder in the many-user setting as its complexity scales exponentially with the number of users.

## A. Main Contributions and Structure of the Paper

In this paper, we rigorously analyze the performance of coding schemes based on random linear models (which include i.i.d. Gaussian codebooks) and computationally efficient approximate message passing (AMP) decoding. Our results provide the exact achievable regions of these schemes in the asymptotic regime described above, and demonstrate numerically that the achievable region of a coding scheme based on *spatially coupled* Gaussian matrices and AMP decoding nearly matches the converse bound for a large range of user densities. An interesting feature of the spatially coupled scheme is that it can be interpreted as a block-wise time-division with overlap multiple-access scheme.

Section II introduces the random linear coding framework (based on either i.i.d. Gaussian or spatially coupled Gaussian matrices) and the associated AMP decoder. In Section III, we present Theorems 1 and 2 which give the asymptotic UER

achieved by these coding schemes for a fixed user payload. Then in Section IV we bound the asymptotic UER achieved by these coding schemes as the user payload grows large (Theorems 3 and 4). Our results show that in the limit of large user payload, reliable communication is not possible at any fixed user density, and the interesting asymptotic regime is when the number of users  $L$ , the code length  $n$ , and the user payload  $\log_2 M$  all tend to infinity with the spectral efficiency  $S = L \log_2 M / n$  (total bits/channel use) held constant (Remark 4.1). We also discuss how *modulation* can be used in the proposed coding framework to reduce decoding complexity at large user payloads.

For simplicity, we only analyze coding schemes for the real-valued Gaussian MAC, but the results in this paper can be extended to the complex setting as well (see Section IV-D).

### B. Related Works

The works of Chen et al. [3] and Ravi and Koch [6]–[8] study MACs under different user scaling regimes. Moreover, their main results pertain to the setting where the probability of error goes to zero with increasing code length. For the linear user scaling and finite error probability setting considered here, the recent papers [9], [10] study the fundamental tradeoffs in the quasi-static fading MAC.

The AMP algorithms used for decoding in this paper (and the analyses of their asymptotic performance) are similar to those used for random linear models and sparse superposition codes (SPARCs) [11]–[18]. The results in Section III also generalize early results in random code-division multiple access (CDMA) [19], [20]. Indeed, the random linear coding framework described in Section II includes random CDMA as a special case. The main novelty of our contribution is in applying AMP and spatial coupling techniques to design efficient coding schemes and prove that they achieve near-optimal tradeoffs for many-user MACs. Our AMP analysis for many-user MAC differs from previous AMP analyses for random linear models and SPARCs in some key technical aspects. In our setting the components of the signal/message vector can be drawn from a probability distribution over finite length vectors, which is more general than in previous work. Furthermore, previous works use the potential function method to analyze the mean-squared error of AMP, whereas our results are in terms of the UER defined in (2), which requires additional technical steps. We note that spatially coupled random CDMA and similar multiple-access schemes based on random spreading sequences have been studied under belief propagation decoding (with Gaussian approximations) in [21], [22]. However, these works do not analyze the many-user MAC setting considered in this paper.

Müller [23] recently proposed practical coding schemes for the many-user MAC setting considered here, based on Gaussian random codes with power optimization and iterative soft decoding. It was shown (non-rigorously) that the asymptotic spectral efficiency versus energy-per-bit tradeoff of the proposed scheme exceeds the achievability bound in [5] at low spectral efficiencies, and is close to the converse bound in [5] when the user payload is 100 bits. As mentioned

in [23, Sec. 3.2], power optimized random codebooks with iterative decoding can be rigorously analyzed by generalizing the AMP analysis of power allocated SPARCs for (single-user) AWGN channels [17]. We leave this analysis and an in-depth performance comparison of various many-user multiple-access schemes for future work.

Building on [10] and the first version of our paper, Kowshik [24] recently obtained a new achievability bound for many-user MACs based on spatially coupled random codebooks and a suboptimal AMP decoder. The advantage of this suboptimal decoder is that its analysis leads to a bound that can be easily computed for large user payloads (e.g., 100 bits). Our achievable region (computed using Theorem 2) is based on the optimal AMP denoiser, and though larger than the one in [24], cannot be computed for large payloads.

We emphasize that our problem setting is distinct from unsourced random access, where only a subset of users are active at any given time and all users use the same codebook (hence user messages are decoded only up to a permutation). In our setting of the Gaussian MAC, users have different codebooks and are all active during the transmission period. Several recent works have analysed the fundamental tradeoffs and the performance of practical coding schemes for unsourced random access in the many-user setting [4], [25], [26].

*Notation:* We use  $\log_2$  and  $\ln$  to denote the base 2 logarithm and natural logarithm, respectively. The Gaussian distribution with mean  $\mu$  and variance  $\sigma^2$  is denoted by  $\mathcal{N}(\mu, \sigma^2)$ . The indicator function of an event  $\mathcal{A}$  is denoted by  $\mathbb{1}\{\mathcal{A}\}$ . For a positive integer  $N$ , we use  $[N]$  to denote the set  $\{1, \dots, N\}$ . We use bold lower case letters or Greek symbols for vectors, and bold upper case for matrices and random vectors. We use plain font for scalars, and subscripts denote entries of a vector or matrix. For example,  $\mathbf{x}$  denotes a vector, with  $x_i$  being the  $i^{\text{th}}$  element of  $\mathbf{x}$ . Similarly, the  $(i, j)^{\text{th}}$  entry of matrix  $\mathbf{A}$  is denoted by  $A_{ij}$ . The transpose of matrix  $\mathbf{A}$  is denoted by  $\mathbf{A}^*$ .

## II. RANDOM LINEAR CODING AND AMP DECODING

We consider coding schemes where the codewords of user  $\ell \in [L]$  are constructed as  $\mathbf{c}_\ell = \mathbf{A}_\ell \mathbf{x}_\ell$ , where  $\mathbf{A}_\ell \in \mathbb{R}^{n \times B}$  is a random matrix and  $\mathbf{x}_\ell \in \mathbb{R}^B$  encodes the message of user  $\ell$ . In this coding framework, the channel model (1) can be written as

$$\mathbf{y} = \mathbf{A} \mathbf{x} + \mathbf{w}, \quad (3)$$

where the *design matrix*  $\mathbf{A} \in \mathbb{R}^{n \times LB}$  is the horizontal concatenation of matrices  $\mathbf{A}_1, \dots, \mathbf{A}_L$ , and the *message vector*  $\mathbf{x} \in \mathbb{R}^{LB}$  is the concatenation of vectors  $\mathbf{x}_1, \dots, \mathbf{x}_L$ . We will assume that the squared norm of each column of  $\mathbf{A}$  equals 1 in expectation. For example, an i.i.d. Gaussian design matrix  $\mathbf{A}$  has i.i.d.  $\mathcal{N}(0, \frac{1}{n})$  entries.

The  $L$  sections of  $\mathbf{x}$  (each of which corresponds to a user's message) are drawn i.i.d. from  $p_{\mathbf{X}_{\text{sec}}}$ , which is a probability mass function over a finite set of length  $B$  vectors. The per-user payload is therefore equal to the entropy  $H(\mathbf{X}_{\text{sec}})$ , where  $\mathbf{X}_{\text{sec}} \sim p_{\mathbf{X}_{\text{sec}}}$ . The codeword energy constraint is denoted by  $E = E_b \log_2 M$ , i.e., we require  $p_{\mathbf{X}_{\text{sec}}}$  to satisfy  $\mathbb{E} \|\mathbf{X}_{\text{sec}}\|^2 = E < \infty$ .

*Example 2.1 (Random Codebooks):* Let  $p_{\mathbf{X}_{\text{sec}}}$  be the distribution over length  $B$  vectors that chooses uniformly at random one of its entries to be non-zero, taking the value  $\sqrt{E}$ . This corresponds to a user payload of  $\log_2 M = \log_2 B$  bits. User  $\ell \in [L]$  selects (and scales by  $\sqrt{E}$ ) one-of- $B$  columns from random matrix  $\mathbf{A}_\ell$ , resulting in expected codeword energy  $\mathbb{E}\|\mathbf{c}_\ell\|^2 = E$ . In the rest of the paper, we denote the choice of  $p_{\mathbf{X}_{\text{sec}}}$  used in this example by  $p_1$ .

*Example 2.2 (Random Codebooks with Binary Modulation):* Let  $p_{\mathbf{X}_{\text{sec}}}$  be the distribution over length  $B$  vectors that chooses uniformly at random one of its  $B$  entries to be non-zero, taking values in  $\{\pm\sqrt{E}\}$  with equal probability. This corresponds to a user payload of  $\log_2 M = 1 + \log_2 B$  bits. User  $\ell \in [L]$  encodes  $\log_2 B$  bits by selecting one-of- $B$  columns from random matrix  $\mathbf{A}_\ell$ , and an additional 1 bit in whether to modulate that column with  $\sqrt{E}$  or  $-\sqrt{E}$ . When  $B = 1$ , this coding scheme corresponds to random CDMA with antipodal signalling.

One can generalize the distribution described in Example 2.2 to consider other modulation schemes such as pulse amplitude modulation, or even complex modulation schemes such as phase-shift keying when the channel is complex (see Section IV-D). Therefore, random linear coding with these choices of  $p_{\mathbf{X}_{\text{sec}}}$  can be viewed as a generalization of random CDMA, with each user encoding  $\log_2 B$  bits in the choice of one-of- $B$  random spreading sequences (in addition to the bits encoded in the choice of the modulation symbol).

#### A. Spatially Coupled Coding Schemes

A spatially coupled (Gaussian) design matrix  $\mathbf{A} \in \mathbb{R}^{n \times LB}$  is divided into  $R$ -by- $C$  equally sized *blocks*. The entries within each block are i.i.d. Gaussian with zero mean and variance specified by the corresponding entry of a base matrix  $\mathbf{W} \in \mathbb{R}_+^{R \times C}$ . The design matrix  $\mathbf{A}$  is constructed by replacing each entry of the base matrix  $W_{rc}$  by an  $\frac{n}{R} \times \frac{LB}{C}$  matrix with entries drawn i.i.d. from  $\mathcal{N}(0, \frac{W_{rc}}{n/R})$ , for  $r \in [R]$ ,  $c \in [C]$ . See Fig. 1a for an example. Hence, the design matrix  $\mathbf{A}$  has independent Gaussian entries

$$A_{ij} \sim \mathcal{N}\left(0, \frac{1}{n/R} W_{r(i) \mathbf{c}(j)}\right), \quad \text{for } i \in [n], j \in [LB]. \quad (4)$$

The operators  $r(\cdot) : [n] \rightarrow [R]$  and  $\mathbf{c}(\cdot) : [LB] \rightarrow [C]$  in (4) map a particular row or column index to its corresponding *row block* or *column block* index. We require  $C$  to divide  $L$ , resulting in  $L/C \geq 1$  users per column block. Recall that the design matrix  $\mathbf{A} \in \mathbb{R}^{n \times LB}$  is the horizontal concatenation of the random matrices  $\mathbf{A}_1, \dots, \mathbf{A}_L$  of all the users (see (3)). Therefore, the  $\ell$ th section of the design matrix (columns  $(\ell - 1)B + 1$  to  $\ell B$ ) is the random matrix of user  $\ell \in [L]$ .

The entries of the base matrix  $\mathbf{W}$  must satisfy  $\sum_{r=1}^R W_{rc} = 1$  for  $c \in [C]$  to ensure that the columns of the design matrix  $\mathbf{A}$  have unit norm in expectation. The trivial base matrix with  $R = C = 1$  (single entry equal to 1) corresponds to the design matrix with i.i.d.  $\mathcal{N}(0, \frac{1}{n})$  entries. In this paper we will consider a class of base matrices called  $(\omega, \Lambda, \rho)$  base matrices [18], [27].

*Definition 2.1:* An  $(\omega, \Lambda, \rho)$  base matrix  $\mathbf{W}$  is described by three parameters: coupling width  $\omega \geq 1$ , coupling length

$\Lambda \geq 2\omega - 1$ , and  $\rho \in [0, 1)$ . The matrix has  $R = \Lambda + \omega - 1$  rows and  $C = \Lambda$  columns, with each column having  $\omega$  identical non-zero entries in the band-diagonal. The  $(r, c)$ th entry of the base matrix, for  $r \in [R]$ ,  $c \in [C]$ , is given by

$$W_{rc} = \begin{cases} \frac{1-\rho}{\omega} & \text{if } c \leq r \leq c + \omega - 1, \\ \frac{\rho}{\Lambda-1} & \text{otherwise.} \end{cases} \quad (5)$$

When  $\rho = 0$ , as in Fig. 1a, the base matrix has non-zero entries only in the band-diagonal part. For example, the base matrix in Fig. 1a has parameters  $(\omega = 3, \Lambda = 7, \rho = 0)$ . For the simulations in Section III-D, we use  $\rho = 0$ , whereas for the theoretical results (Theorems 2 and 4) we choose  $\rho$  to be a small positive value. (The proofs of these results use the AMP concentration results in [18]; there are some technical difficulties in proving the concentration results for  $\rho = 0$ , which can be addressed by picking a suitable  $\rho > 0$ .)

Each entry of the base matrix corresponds to an  $\frac{n}{R} \times \frac{LB}{C}$  block of the design matrix  $\mathbf{A}$ , and each block can be viewed as an (uncoupled) i.i.d. Gaussian design matrix with  $\frac{L}{C}$  users, code length  $\frac{n}{R}$ , and user density

$$\mu_{\text{inner}} = \frac{L/C}{n/R} = \frac{R}{C} \mu = \left(1 + \frac{\omega - 1}{\Lambda}\right) \mu. \quad (6)$$

Since  $\omega > 1$  in spatially coupled systems, we have  $\mu < \mu_{\text{inner}}$ . This difference is often referred to as a “rate loss” in the literature of spatially coupled error correcting codes [27]–[30], and becomes negligible when  $\Lambda$  is much larger than  $\omega$ .

The spatially coupled coding scheme can be viewed as *block-wise time-division with overlap*. Consider the scenario depicted in Fig. 1b where there are  $L = 25$  users,  $n = 35$  channel uses, and a spatially coupled design matrix constructed using an  $(\omega = 3, \Lambda = 5, \rho = 0)$  base matrix  $\mathbf{W} \in \mathbb{R}_+^{7 \times 5}$ . Each block of the design matrix corresponds to 5 channel uses and 5 users. Fig. 1b shows how users communicate using this multiple access scheme, assuming that each channel use corresponds to one time instant. A red dot in the 2D grid represents a certain user being active (transmitting) during a certain time instant, and empty squares represent silence (no transmission). For example, users in the first column block (users 1 to 5) transmit during time instants 1 to 15 (corresponding to the first  $\omega = 3$  row blocks) but are silent afterwards; users 6 to 10 transmit during time instants 6 to 20 but are silent otherwise, and so on.

Users within the same column block transmit simultaneously over  $\omega$  row blocks of time (15 time instants), and users in neighbouring column blocks overlap in  $\omega - 1$  row blocks of time. At each time instant, up to  $\omega$  column blocks of users (15 users) simultaneously transmit, but the set of active users gradually shifts over time. When  $\omega = 1$ , there is no time overlap (no coupling) between neighbouring blocks of users and each block of users communicates using an i.i.d. Gaussian matrix. When  $\Lambda$  is large with respect to  $\omega$ , users are silent for most of the transmission period ( $n$  channel uses). This facilitates low-complexity encoding and decoding.

With a band-diagonal spatially coupled matrix as in Fig. 1a, the sections at the two ends of the message vector  $\mathbf{x}$  are more easily decoded than other sections. This is because the

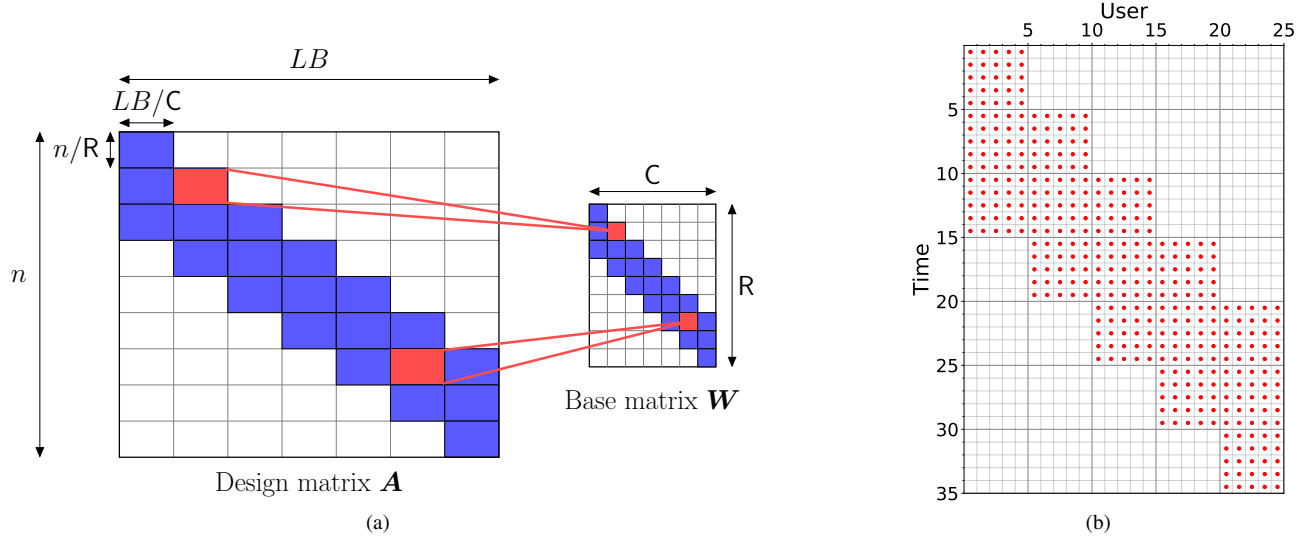


Fig. 1. (a) A spatially coupled design matrix  $\mathbf{A}$  defined using a base matrix  $\mathbf{W} \in \mathbb{R}_+^{9 \times 7}$ . Each square in  $\mathbf{W}$  represents a scalar entry that specifies the variance of the entries in a block of  $\mathbf{A}$ . The white parts of  $\mathbf{A}$  and  $\mathbf{W}$  correspond to zeros. (b) An example of how 25 users communicate over 35 uses of the channel using the multiple access scheme based on a spatially coupled design matrix constructed using an  $(\omega = 3, \Lambda = 5, \rho = 0)$  base matrix. A red dot in the 2D grid represents a certain user being active during a certain time instant and empty squares represent silence.

channel outputs containing information about the sections at the ends have less interference from other sections (e.g., the first and last row of the base matrix  $\mathbf{W}$  only have one non-zero entry). Once the sections at the ends have been decoded, their neighboring sections become easier to decode, and a decoding wave propagates towards the middle of the message vector. In contrast, for i.i.d. design matrices, all sections are equally hard to decode. For more intuitive explanations and figures that illustrate the decoding wave in spatially coupled systems, see [12], [13], [15], [18].

### B. AMP Decoding and State Evolution

We consider an efficient AMP decoder that aims to reconstruct the message vector  $\mathbf{x}$  from the channel output  $\mathbf{y}$ . The decoder knows the design matrix  $\mathbf{A}$ , the base matrix  $\mathbf{W}$ , the distribution  $p_{\mathbf{X}_{\text{sec}}}$ , and the channel noise variance  $\sigma^2$ . AMP algorithms are based on Gaussian approximations to loopy belief propagation on dense graphs [11] and have been proposed for estimation in the random linear model (3) with spatially coupled design matrices  $\mathbf{A}$  [12], [13], [15], [18].

The AMP decoder iteratively produces message vector estimates  $\mathbf{x}^t \in \mathbb{R}^{LB}$  for iterations  $t = 1, 2, \dots$  as follows. Initialize  $\mathbf{x}^0$  to the all-zero vector, and for  $t \geq 0$ , iteratively compute:

$$\begin{aligned} \mathbf{q}^t &= \mathbf{y} - \mathbf{A}\mathbf{x}^t + \tilde{\mathbf{v}}^t \odot \mathbf{q}^{t-1}, \\ \mathbf{x}^{t+1} &= \eta^t(\mathbf{x}^t + (\tilde{\mathbf{S}}^t \odot \mathbf{A})^* \mathbf{q}^t). \end{aligned} \quad (7)$$

Here  $\odot$  is the Hadamard (entry-wise) product and quantities with negative iteration indices are set to zero vectors. The vector  $\tilde{\mathbf{v}}^t \in \mathbb{R}^n$ , the matrix  $\tilde{\mathbf{S}} \in \mathbb{R}^{n \times LB}$ , and the denoising function  $\eta^t$  will be described in terms of the following state evolution parameters.

*State Evolution:* The performance of the AMP in the large system limit is succinctly captured by a deterministic recursion

called *state evolution* [31], [32]. State evolution iteratively defines vectors  $\gamma^t, \phi^t \in \mathbb{R}^R$  and  $\tau^t, \psi^t \in \mathbb{R}^C$  as follows. Initialize  $\psi_c^0 = E$  for  $c \in [C]$ , and for  $t \geq 0$ , iteratively compute:

$$\gamma_r^t = \sum_{c=1}^C W_{rc} \psi_c^t, \quad \phi_r^t = \sigma^2 + \mu_{\text{inner}} \gamma_r^t, \quad r \in [R], \quad (8)$$

$$\tau_c^t = \left[ \sum_{r=1}^R \frac{W_{rc}}{\phi_r^t} \right]^{-1}, \quad \psi_c^{t+1} = \text{mmse}(1/\tau_c^t), \quad c \in [C], \quad (9)$$

where  $\mu_{\text{inner}} = \frac{R}{C} \mu$  from (6), and

$$\begin{aligned} \text{mmse}(1/\tau) &= \mathbb{E} \left\| \mathbf{X}_{\text{sec}} - \mathbb{E}[\mathbf{X}_{\text{sec}} | \mathbf{X}_{\text{sec}} + \sqrt{\tau} \mathbf{Z}] \right\|^2 \\ &\stackrel{(i)}{=} E \left[ 1 - \mathbb{E} \left[ \frac{e^{\sqrt{\frac{E}{\tau}} Z_1}}{e^{\sqrt{\frac{E}{\tau}} Z_1} + e^{-E/\tau} \sum_{j=2}^B e^{\sqrt{\frac{E}{\tau}} Z_j}} \right] \right], \end{aligned} \quad (10)$$

where  $\mathbf{X}_{\text{sec}} \sim p_{\mathbf{X}_{\text{sec}}}$  and  $\mathbf{Z} = [Z_1, \dots, Z_B]$  is a standard Gaussian vector independent of  $\mathbf{X}_{\text{sec}}$ . The equality (i) holds when  $\mathbf{X}_{\text{sec}} \sim p_1$ , where  $p_1$  is described in Example 2.1.

The vector  $\tilde{\mathbf{v}}^t \in \mathbb{R}^n$  and the matrix  $\tilde{\mathbf{S}}^t \in \mathbb{R}^{n \times LB}$  in (7) both have a block-wise structure and are defined using state evolution parameters as follows. For  $i \in [n]$  and  $j \in [LB]$ ,

$$\tilde{v}_i^t = \frac{\mu_{\text{inner}} \gamma_{r(i)}^t}{\phi_{r(i)}^{t-1}}, \quad \tilde{S}_{ij}^t = \frac{\tau_{c(j)}^t}{\phi_{r(i)}^t}, \quad (11)$$

where we recall that  $r(i)$  and  $c(j)$  denote the row and column block index of the  $i$ th row entry and  $j$ th column entry, respectively. The vector  $\tilde{\mathbf{v}}^0$  is defined to be the all-zero vector.

In each iteration, the AMP decoder (7) produces an effective observation  $\mathbf{s}^t = \mathbf{x}^t + (\tilde{\mathbf{S}}^t \odot \mathbf{A})^* \mathbf{q}^t$ , which has the following approximate representation: for an index  $j$  in column block  $\mathbf{c}$  of the message vector  $\mathbf{x}$ , we have  $s_j^t \approx x_j + \sqrt{\tau_c^t} Z_j$ , where  $\{Z_j\} \sim_{\text{i.i.d.}} \mathcal{N}(0, 1)$ . The estimate  $\mathbf{x}^{t+1}$  in (7) is

then the minimum mean square error (MMSE) estimate of  $\mathbf{x}$  given  $\mathbf{s}^t$ , computed using the assumed distribution. This leads to the following definition of the denoising function  $\eta^t = (\eta_1^t, \dots, \eta_{LB}^t)$  in (7): for index  $j$  in section  $\ell \in [L]$ , which we denote by  $j \in \text{sec}(\ell)$ ,

$$\eta_j^t(\mathbf{s}) = \mathbb{E} \left[ (\mathbf{X}_{\text{sec}})_j \mid \mathbf{X}_{\text{sec}} + \sqrt{\tau_{\text{c}(j)}^t} \mathbf{Z} = \mathbf{s}_\ell \right] \quad (12)$$

$$\stackrel{(i)}{=} \sqrt{E} \cdot \frac{\exp \left( s_j \sqrt{E} / \tau_{\text{c}(j)}^t \right)}{\sum_{i \in \text{sec}(\ell)} \exp \left( s_i \sqrt{E} / \tau_{\text{c}(j)}^t \right)}, \quad (13)$$

where we recall that  $\mathbf{s}_\ell \in \mathbb{R}^B$  is the  $\ell$ th section of  $\mathbf{s} \in \mathbb{R}^{LB}$ . The equality (i) holds when  $\mathbf{X}_{\text{sec}} \sim p_1$ .

In addition, the decoder can also produce a *hard-decision* maximum a posteriori (MAP) estimate from  $\mathbf{s}^t$ , which we denote by  $\hat{\mathbf{x}}^{t+1}$ . For section  $\ell$  in column block  $\mathbf{c} \in [\mathbf{C}]$ , the  $\ell$ th section of this hard-decision estimate is given by

$$\hat{\mathbf{x}}_\ell^{t+1} = \arg \max_{\mathbf{x}' \in \mathcal{X}} \mathbb{P}(\mathbf{X}_{\text{sec}} = \mathbf{x}' \mid \mathbf{X}_{\text{sec}} + \sqrt{\tau_{\text{c}}^t} \mathbf{Z} = \mathbf{s}_\ell^t), \quad (14)$$

where  $\mathcal{X}$  is the support of  $p_{\mathbf{X}_{\text{sec}}}$ . When  $\mathbf{X}_{\text{sec}} \sim p_1$ , for  $j \in \text{sec}(\ell)$  this hard-decision estimate is given by

$$\hat{x}_j^{t+1} = \begin{cases} \sqrt{E} & \text{if } s_j^t > s_i^t \text{ for all } i \in \text{sec}(\ell) \setminus j, \\ 0 & \text{otherwise.} \end{cases} \quad (15)$$

Eqs. (13) and (15) give closed form expressions for the denoising function  $\eta^t$  and MAP estimator when  $\mathbf{X}_{\text{sec}} \sim p_1$ . Similar expressions can be easily obtained when  $\mathbf{X}_{\text{sec}}$  is drawn from the discrete distribution in Example 2.2 or its generalizations, but these are omitted for brevity.

**IID Gaussian A:** For the special case where the entries of the design matrix  $\mathbf{A}$  are i.i.d.  $\sim \mathcal{N}(0, \frac{1}{n})$ , the AMP decoder (7) and the state evolution (8)–(9) can be simplified. Setting  $\mathbf{x}^0$  equal to the all-zero vector, for  $t \geq 0$ , the AMP decoder computes:

$$\begin{aligned} \mathbf{q}^t &= \mathbf{y} - \mathbf{A} \mathbf{x}^t + \frac{\mu \psi^t}{\tau^{t-1}} \mathbf{q}^{t-1}, \\ \mathbf{x}^{t+1} &= \eta^t(\mathbf{x}^t + \mathbf{A}^* \mathbf{q}^t). \end{aligned} \quad (16)$$

At  $t = 0$ , the vector  $\frac{\mu \psi^0}{\tau^{-1}} \mathbf{q}^{-1}$  is set to the all-zero vector. The scalars  $\tau^t$  and  $\psi^t$  are computed via the state evolution recursion. Initializing with  $\psi^0 = E$ , for  $t \geq 0$  we have:

$$\begin{aligned} \tau^t &= \sigma^2 + \mu \psi^t, \\ \psi^{t+1} &= \text{mmse}(1/\tau^t), \end{aligned} \quad (17)$$

where the  $\text{mmse}$  function is defined in (10). Furthermore, in this case, the denoising function  $\eta^t$  in (12) and the hard-decision estimate  $\hat{\mathbf{x}}^{t+1}$  in (14) are defined using the state evolution parameter  $\tau^t$  as the Gaussian noise variance.

### III. ASYMPTOTIC UER ACHIEVED BY AMP DECODING

We now characterize the asymptotic UER (see (2)) achieved by coding schemes based on i.i.d. and spatially coupled Gaussian design matrices with AMP decoding. These results are stated in terms of a *potential function*.

#### A. Potential Function

Consider the single-section Gaussian channel with noise variance  $\tau$ :

$$\mathbf{S}_\tau = \mathbf{X}_{\text{sec}} + \sqrt{\tau} \mathbf{Z}, \quad (18)$$

where  $\mathbf{X}_{\text{sec}} \sim p_{\mathbf{X}_{\text{sec}}}$  and  $\mathbf{Z} \in \mathbb{R}^B$  is a standard Gaussian vector independent of  $\mathbf{X}_{\text{sec}}$ . The potential function for the random linear system (3) with user density  $\mu = \frac{L}{n}$  and channel noise variance  $\sigma^2$  is defined as

$$\mathcal{F}(\mu, \sigma^2, \psi) = I(\mathbf{X}_{\text{sec}}; \mathbf{S}_\tau) + \frac{1}{2\mu} \left[ \ln \left( \frac{\tau}{\sigma^2} \right) - \frac{\mu \psi}{\tau} \right], \quad (19)$$

where  $\psi \in [0, E]$ ,  $\tau = \sigma^2 + \mu \psi$ , and the mutual information  $I(\mathbf{X}_{\text{sec}}; \mathbf{S}_\tau)$  is computed using the channel (18).<sup>1</sup> If  $\mathbf{X}_{\text{sec}} \sim p_1$  and  $\mathbf{Z} = [Z_1, \dots, Z_B]$ , then

$$I(\mathbf{X}_{\text{sec}}; \mathbf{S}_\tau) = \frac{E}{\tau} + \ln B - \mathbb{E} \ln \left[ e^{\frac{E}{\tau} + \sqrt{\frac{E}{\tau}} Z_1} + \sum_{j=2}^B e^{\sqrt{\frac{E}{\tau}} Z_j} \right].$$

Define the set of potential function minimizers (w.r.t.  $\psi$ ) as:

$$\mathcal{M}(\mu, \sigma^2) = \arg \min_{\psi \in [0, E]} \mathcal{F}(\mu, \sigma^2, \psi). \quad (20)$$

Consider decoding  $\mathbf{X}_{\text{sec}}$  from  $\mathbf{S}_\tau$  produced by the Gaussian channel in (18). The MMSE decoder  $\hat{\mathbf{x}}_{\text{sec}}^{\text{MMSE}}(\mathbf{S}_\tau) = \mathbb{E}[\mathbf{X}_{\text{sec}} | \mathbf{S}_\tau]$  achieves the MMSE given by (10). The MAP decoder  $\hat{\mathbf{x}}_{\text{sec}}^{\text{MAP}}(\mathbf{S}_\tau) = \arg \max_{\mathbf{x}'} \mathbb{P}(\mathbf{X}_{\text{sec}} = \mathbf{x}' | \mathbf{S}_\tau)$  achieves the minimum probability of error given by

$$P_e(\tau) = \mathbb{P}(\hat{\mathbf{x}}_{\text{sec}}^{\text{MAP}}(\mathbf{S}_\tau) \neq \mathbf{X}_{\text{sec}}) \quad (21)$$

$$\stackrel{(i)}{=} 1 - \mathbb{E} \left[ \Phi \left( \sqrt{E/\tau} + Z \right)^{B-1} \right]. \quad (22)$$

where  $Z \sim \mathcal{N}(0, 1)$  and  $\Phi(\cdot)$  is the standard normal distribution function. The equality (i) holds when  $\mathbf{X}_{\text{sec}} \sim p_1$ .

Theorem 1 below shows that for i.i.d. Gaussian matrices, the fixed point of the state evolution equations (17) is characterized by a specific stationary point of the potential function (19), and that the asymptotic UER achieved by AMP decoding can be bounded using the fixed point. Theorem 2 gives an analogous result for spatially coupled Gaussian matrices, where the fixed point of the (spatially coupled) state evolution (8)–(9) is bounded using the global minimum of the potential function.

#### B. IID Gaussian Matrices

**Theorem 1 (IID Gaussian Matrices with AMP Decoding):** Consider the linear model (3), with the entries of the design matrix  $\mathbf{A}$  i.i.d.  $\sim \mathcal{N}(0, \frac{1}{n})$  and the  $L$  sections of the message vector  $\mathbf{x}$  i.i.d.  $\sim p_{\mathbf{X}_{\text{sec}}}$ . Let  $\hat{\mathbf{x}}^t$  be the AMP hard-decision estimate of  $\mathbf{x}$  after iteration  $t$  (see (14)), and recall that  $\tau^t, \psi^t$  are outputs of the state evolution (17).

<sup>1</sup>The potential function (19) has connections with the mutual information between the message vector and the channel output vector in the random linear estimation problem (3) [33], [34]. We do not use this connection here and only consider the relationship between the stationary points of the potential function and the fixed points of the state evolution recursion. The potential functions used in this paper can be derived using the method in [35], up to scaling factors and additive constants (which do not affect the desired properties).

1) The sequences  $\{\tau^t\}_{t \geq 0}$  and  $\{\psi^t\}_{t \geq 0}$  are non-increasing and converge to fixed points  $\tau^{\text{FP}}$ ,  $\psi^{\text{FP}}$ , where

$$\tau^{\text{FP}} := \sigma^2 + \mu\psi^{\text{FP}}, \quad (23)$$

$$\begin{aligned} \psi^{\text{FP}} &:= \max \left\{ \psi : \psi = \text{mmse} \left( \frac{1}{\sigma^2 + \mu\psi} \right) \right\} \\ &= \max \left\{ \psi : \frac{\partial \mathcal{F}(\mu, \sigma^2, \psi)}{\partial \psi} = 0 \right\}. \end{aligned} \quad (24)$$

The potential function  $\mathcal{F}(\mu, \sigma^2, \psi)$  is defined in (19).

2) Fix  $\delta > 0$ , and let  $T$  denote the first iteration for which  $\tau^t \leq \tau^{\text{FP}} + \delta$ . Then the UER of the AMP decoder after  $T + 1$  iterations satisfies the following almost surely:

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_\ell^{T+1} \neq \mathbf{x}_\ell\} \stackrel{\text{a.s.}}{=} P_e(\tau^T) \leq P_e(\tau^{\text{FP}} + \delta), \quad (25)$$

where the limit is taken with  $\frac{L}{n} = \mu$  held constant and  $P_e(\cdot)$  is defined in (21).

*Proof:* 1) Results similar to the first part of the theorem are known in the AMP literature and are sometimes used implicitly. We provide a proof here for completeness. We first prove that the sequence  $\{\psi^t\}_{t \geq 0}$  is non-increasing and converges to the fixed point  $\psi^{\text{FP}}$  defined in (24). Then the result for  $\{\tau^t\}_{t \geq 0}$  immediately follows since  $\tau^t = \sigma^2 + \mu\psi^t$ .

Writing the state evolution in (17) as a single recursion, we have:

$$\psi^{t+1} = \text{mmse}((\sigma^2 + \mu\psi^t)^{-1}). \quad (26)$$

Starting from  $\psi^0 = \mathbb{E}\|\mathbf{X}_{\text{sec}}\|^2 = E$ , we have that

$$\psi^1 = \text{mmse}((\sigma^2 + \mu E)^{-1}) \leq E = \psi^0, \quad (27)$$

where the inequality holds because the trivial all zero estimate of a random section  $\mathbf{X}_{\text{sec}}$  achieves an expected squared error of  $E$ . The mmse function defined in (10) is non-increasing in its argument [16], [36], and since its argument  $(\sigma^2 + \mu\psi^t)^{-1}$  is decreasing in  $\psi^t$ , the mmse function is non-decreasing in  $\psi^t$ . Therefore, together with (27), this shows that the sequence  $\{\psi^t\}_{t \geq 0}$  is non-increasing. Moreover, if  $\psi^t \geq \psi^{\text{FP}}$ , then  $\psi^{t+1} \geq \psi^{\text{FP}}$ . Indeed, for any  $\psi^t \geq \psi^{\text{FP}}$ ,

$$\begin{aligned} \psi^{t+1} &= \text{mmse}((\sigma^2 + \mu\psi^t)^{-1}) \\ &\geq \text{mmse}((\sigma^2 + \mu\psi^{\text{FP}})^{-1}) = \psi^{\text{FP}}. \end{aligned}$$

Since  $\{\psi^t\}_{t \geq 0}$  is a non-increasing sequence bounded below by  $\psi^{\text{FP}}$  (noting that  $\psi^0 \geq \psi^{\text{FP}}$ ), we conclude that it converges to  $\psi^{\text{FP}}$ .

To show that the fixed points of the state evolution correspond to the stationary points of the potential function  $\mathcal{F}(\mu, \sigma^2, \psi)$  defined in (19), we compute the derivative:

$$\frac{\partial \mathcal{F}(\mu, \sigma^2, \psi)}{\partial \psi} = \frac{\mu}{2(\sigma^2 + \mu\psi)^2} \left[ \psi - \text{mmse} \left( \frac{1}{\sigma^2 + \mu\psi} \right) \right],$$

where we have used  $\tau = \sigma^2 + \mu\psi$  and the vector I-MMSE relationship [37, Thm. 2]. Therefore, since  $\sigma^2 > 0$  and  $\mu > 0$ , we have that  $\partial \mathcal{F}(\mu, \sigma^2, \psi)/\partial \psi = 0$  corresponds to  $\psi = \text{mmse}((\sigma^2 + \mu\psi)^{-1})$ , which is the fixed point of the iteration (26).

2) We now prove (25). For  $\ell \in [L]$ , we denote by  $\mathbf{a}_\ell \in \mathbb{R}^B$  the  $\ell$ th section of a vector  $\mathbf{a} \in \mathbb{R}^{LB}$ . Consider the input to the

AMP hard-decision step in iteration  $t + 1$ , which we denote by  $\mathbf{s}^t = \mathbf{x}^t + \mathbf{A}^* \mathbf{q}^t \in \mathbb{R}^{LB}$  (see (14), (16)). The MAP estimator  $\hat{\mathbf{x}}_\ell^{t+1} = \hat{\mathbf{x}}_\ell^{t+1}(\mathbf{s}_\ell^t)$  in (14) partitions the space  $\mathbb{R}^B$  into decision regions. For each  $\mathbf{x}_\ell$  in the support of  $p_{\mathbf{X}_{\text{sec}}}$ , the decision region is

$$\mathcal{D}(\mathbf{x}_\ell) := \{\mathbf{s}_\ell^t : \hat{\mathbf{x}}_\ell^{t+1}(\mathbf{s}_\ell^t) = \mathbf{x}_\ell\}. \quad (28)$$

Note that  $\mathbb{1}\{\hat{\mathbf{x}}_\ell^{t+1}(\mathbf{s}_\ell^t) = \mathbf{x}_\ell\} = \mathbb{1}\{\mathbf{s}_\ell^t \in \mathcal{D}(\mathbf{x}_\ell)\}$ .

The distance between a vector  $\mathbf{v} \in \mathbb{R}^B$  and a set  $\mathcal{B} \subset \mathbb{R}^B$  is denoted by  $d(\mathbf{v}, \mathcal{B}) := \inf\{\|\mathbf{v} - \mathbf{u}\|_2 : \mathbf{u} \in \mathcal{B}\}$ . For any  $\epsilon > 0$ , define the functions  $\xi_{\epsilon,+}, \xi_{\epsilon,-} : \mathbb{R}^B \times \mathbb{R}^B \rightarrow \mathbb{R}$  as follows:

$$\xi_{\epsilon,+}(\mathbf{x}_\ell, \mathbf{s}_\ell^t) = \begin{cases} 1, & \text{if } \mathbf{s}_\ell^t \in \mathcal{D}(\mathbf{x}_\ell), \\ 0, & \text{if } d(\mathbf{s}_\ell^t, \mathcal{D}(\mathbf{x}_\ell)) > \epsilon, \\ 1 - d(\mathbf{s}_\ell^t, \mathcal{D}(\mathbf{x}_\ell))/\epsilon, & \text{otherwise,} \end{cases}$$

$$\xi_{\epsilon,-}(\mathbf{x}_\ell, \mathbf{s}_\ell^t) = \begin{cases} 1, & \text{if } d(\mathbf{s}_\ell^t, \mathcal{D}(\mathbf{x}_\ell)^c) > \epsilon, \\ 0, & \text{if } \mathbf{s}_\ell^t \in \mathcal{D}(\mathbf{x}_\ell)^c, \\ d(\mathbf{s}_\ell^t, \mathcal{D}(\mathbf{x}_\ell)^c)/\epsilon, & \text{otherwise.} \end{cases}$$

We note that  $\xi_{\epsilon,+}, \xi_{\epsilon,-}$  are Lipschitz-continuous (with Lipschitz constant  $1/\epsilon$ ), and

$$\xi_{\epsilon,-}(\mathbf{x}_\ell, \mathbf{s}_\ell^t) \leq \mathbb{1}\{\mathbf{s}_\ell^t \in \mathcal{D}(\mathbf{x}_\ell)\} \leq \xi_{\epsilon,+}(\mathbf{x}_\ell, \mathbf{s}_\ell^t),$$

and thus

$$\begin{aligned} \frac{1}{L} \sum_{\ell=1}^L \xi_{\epsilon,-}(\mathbf{x}_\ell, \mathbf{s}_\ell^t) &\leq \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_\ell^{t+1}(\mathbf{s}_\ell^t) = \mathbf{x}_\ell\} \\ &\leq \frac{1}{L} \sum_{\ell=1}^L \xi_{\epsilon,+}(\mathbf{x}_\ell, \mathbf{s}_\ell^t). \end{aligned} \quad (29)$$

A pseudo-Lipschitz function  $\xi : \mathbb{R}^m \rightarrow \mathbb{R}$  is one that satisfies the following for all  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ :

$$|\xi(\mathbf{u}) - \xi(\mathbf{v})| \leq C(1 + \|\mathbf{u}\|_2 + \|\mathbf{v}\|_2) \|\mathbf{u} - \mathbf{v}\|_2,$$

for some constant  $C > 0$ . The results in [31] and [17], [18], [38] imply that for any pseudo-Lipschitz function  $\xi : \mathbb{R}^B \times \mathbb{R}^B \rightarrow \mathbb{R}$ , the following holds almost surely:

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \xi(\mathbf{x}_\ell, \mathbf{s}_\ell^t) = \mathbb{E}\{\xi(\mathbf{X}_{\text{sec}}, \mathbf{S}_{\tau^t})\}, \quad (30)$$

where  $\mathbf{X}_{\text{sec}} \sim p_{\mathbf{X}_{\text{sec}}}$  and  $\mathbf{S}_{\tau^t}$  is given by (18). This result was proved in [31] for the  $B = 1$  case and extended in [17], [18] to the setting of SPARCs where the specific distribution  $p_{\mathbf{X}_{\text{sec}}}$  given in Example 2.1 (corresponding to random codebooks) is used. The proof for more general discrete distributions is essentially the same. In (30) and below,  $L/n = \mu$  as  $L \rightarrow \infty$ .

Applying (30) to the Lipschitz continuous functions  $\xi_{\epsilon,+}$  and  $\xi_{\epsilon,-}$ , we obtain almost surely:

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \xi_{\epsilon,+}(\mathbf{x}_\ell, \mathbf{s}_\ell^t) = \mathbb{E}\{\xi_{\epsilon,+}(\mathbf{X}_{\text{sec}}, \mathbf{S}_{\tau^t})\}, \quad (31)$$

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \xi_{\epsilon,-}(\mathbf{x}_\ell, \mathbf{s}_\ell^t) = \mathbb{E}\{\xi_{\epsilon,-}(\mathbf{X}_{\text{sec}}, \mathbf{S}_{\tau^t})\}.$$

Since  $\epsilon > 0$  is arbitrary, from (29) and (31), we almost surely have

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \mathbb{E}\{\xi_{\epsilon,-}(\mathbf{X}_{\text{sec}}, \mathbf{S}_{\tau^t})\} &\leq \liminf_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_{\ell}^{t+1}(\mathbf{s}_{\ell}^t) = \mathbf{x}_{\ell}\} \\ &\leq \limsup_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_{\ell}^{t+1}(\mathbf{s}_{\ell}^t) = \mathbf{x}_{\ell}\} \\ &\leq \lim_{\epsilon \rightarrow 0} \mathbb{E}\{\xi_{\epsilon,+}(\mathbf{X}_{\text{sec}}, \mathbf{S}_{\tau^t})\}. \end{aligned}$$

By the dominated convergence theorem, we have

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \mathbb{E}\{\xi_{\epsilon,-}(\mathbf{X}_{\text{sec}}, \mathbf{S}_{\tau^t})\} &= \mathbb{P}(\mathbf{S}_{\tau^t} \in \mathcal{D}(\mathbf{X}_{\text{sec}})) = 1 - P_e(\tau^t), \\ \lim_{\epsilon \rightarrow 0} \mathbb{E}\{\xi_{\epsilon,+}(\mathbf{X}_{\text{sec}}, \mathbf{S}_{\tau^t})\} &= \mathbb{P}(\mathbf{S}_{\tau^t} \in \mathcal{D}(\mathbf{X}_{\text{sec}})) = 1 - P_e(\tau^t). \end{aligned}$$

This completes the proof that, almost surely

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_{\ell}^{t+1}(\mathbf{s}_{\ell}^t) \neq \mathbf{x}_{\ell}\} = P_e(\tau^t). \quad (32)$$

*Remark 3.1:* Consider the setting of Theorem 1 and the section-by-section (SBS) MAP decoder for the linear model (3):

$$\hat{\mathbf{x}}_{\ell}^{\text{MAP}} = \arg \max_{\mathbf{x}' \in \mathcal{X}} \mathbb{P}(\mathbf{x}_{\ell} = \mathbf{x}' | \mathbf{y}, \mathbf{A}), \quad \text{for } \ell \in [L], \quad (33)$$

where  $\mathcal{X}$  is the support of the discrete prior for each section of the message vector. The SBS-MAP decoder in (33) minimizes the expected UER. Though computationally infeasible, the asymptotic error of the SBS-MAP decoder can be analyzed using the non-rigorous replica method. Using this technique, Tanaka [19] showed that for binary CDMA ( $B = 1$  and  $p_{\mathbf{X}_{\text{sec}}}$  uniform over  $\{1, -1\}$ ), the asymptotic UER of  $\hat{\mathbf{x}}^{\text{MAP}}$  can be characterized in terms of the probability of decoding error in the single-section Gaussian channel (18). Specifically, when  $\mathcal{M}(\mu, \sigma^2)$  in (20) is a singleton (i.e., when the global minimizer of the potential function is unique),

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_{\ell}^{\text{MAP}} \neq \mathbf{x}_{\ell}\} = P_e(\tau^*), \quad (34)$$

where the limit is taken with  $\frac{L}{n} = \mu$  held constant and

$$\tau^* = \sigma^2 + \mu \mathcal{M}(\mu, \sigma^2). \quad (35)$$

We expect that a similar result can be shown via replica analysis for  $B > 1$  and general discrete priors. Several arguments were put forward in [25, Sec. IV.A] to suggest that such an extension is possible.

*Remark 3.2:* The theoretical analysis in this paper (Theorems 1 to 4) is similar to the state evolution and potential function analyses of SPARCs and random linear estimation [15], [16], [18], [33], [34]. The key technical differences are: (i) our coding scheme is a more general random linear framework than SPARCs as we allow the sections of the message vector to be drawn from a general discrete distribution over length  $B$  vectors (standard SPARCs use the specific distribution in Example 2.1), and (ii) state evolution and potential function results are usually given in terms of the mean-squared error, whereas our results are given in terms

of the UER, which requires additional technical steps, e.g., the sandwiching argument used to prove the second part of Theorem 1.

### C. Spatially Coupled Gaussian Matrices

*Theorem 2 (Spatially Coupled Gaussian Matrices with AMP Decoding):* Consider the linear model (3) with a spatially coupled design matrix  $\mathbf{A}$  constructed using an  $(\omega, \Lambda, \rho)$  base matrix, and the  $L$  sections of the message vector  $\mathbf{x}$  i.i.d.  $\sim p_{\mathbf{X}_{\text{sec}}}$ . Let  $\hat{\mathbf{x}}^t$  be the AMP hard-decision estimate of  $\mathbf{x}$  after iteration  $t$  (see (14)), and recall that  $\tau^t \in \mathbb{R}^{\mathbf{C}}$  is an output of the state evolution recursion (8)–(9) with  $\mathbf{C} = \Lambda$  in this setting.

1) For any  $(\omega, \Lambda, \rho)$  base matrix, each entry of  $\tau^t \in \mathbb{R}^{\mathbf{C}}$  is non-increasing in  $t$ , and the  $\mathbf{c}$ th entry converges to a fixed point, denoted by  $\tau_{\mathbf{c}}^{\text{SC-FP}}$ , for  $\mathbf{c} \in [\mathbf{C}]$ .

2) For any  $\epsilon > 0$ , there are constants  $\omega_0 < \infty$ ,  $\Lambda_0 < \infty$  and  $\rho_0 > 0$  such that, for all  $\omega > \omega_0$ ,  $\Lambda > \Lambda_0$  and  $0 \leq \rho < \rho_0$ , the fixed points  $\{\tau_{\mathbf{c}}^{\text{SC-FP}}\}_{\mathbf{c} \in [\mathbf{C}]}$  satisfy

$$\max_{\mathbf{c} \in [\mathbf{C}]} \tau_{\mathbf{c}}^{\text{SC-FP}} \leq \bar{\tau}_{\vartheta} := \sigma^2 + \vartheta \mu (\max \mathcal{M}(\vartheta \mu, \sigma^2) + \epsilon), \quad (36)$$

where  $\vartheta = 1 + \frac{(\omega-1)}{\Lambda}$ , and the set of potential function minimizers  $\mathcal{M}(\vartheta \mu, \sigma^2)$  is defined in (20).

3) Fix base matrix parameters  $\omega > \omega_0$ ,  $\Lambda > \Lambda_0$ , and  $0 < \rho < \rho_0$ . Fix  $\delta > 0$ , and let  $T$  denote the first iteration for which  $\max_{\mathbf{c}} \tau_{\mathbf{c}}^t \leq \bar{\tau}_{\vartheta} + \delta$ . Then the UER of the AMP decoder after  $T + 1$  iterations satisfies the following almost surely:

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_{\ell}^{T+1} \neq \mathbf{x}_{\ell}\} \stackrel{\text{a.s.}}{=} \frac{1}{\mathbf{C}} \sum_{\mathbf{c}=1}^{\mathbf{C}} P_e(\tau_{\mathbf{c}}^T) \leq P_e(\bar{\tau}_{\vartheta} + \delta), \quad (37)$$

where the limit is taken with  $\frac{L}{n} = \mu$  held constant.

*Remark 3.3 (Threshold Saturation):* Theorem 2 shows that the asymptotic UER achievable with a suitable spatially coupled Gaussian matrix and AMP decoding is bounded by  $P_e(\bar{\tau}_{\vartheta} + \delta)$ . As  $\omega/\Lambda \rightarrow 0$ , we have  $\vartheta \rightarrow 1$ . Therefore, if  $\mathcal{M}(\mu, \sigma^2)$  defined in (20) is a singleton, (noting that  $\epsilon$  in (36) can be arbitrarily small) we have

$$\lim_{\omega \rightarrow \infty} \lim_{\Lambda \rightarrow \infty} \bar{\tau}_{\vartheta} \rightarrow \tau^*, \quad (38)$$

where  $\tau^*$  is defined in (35). Therefore, in the limit described in (38), the asymptotic UER of the spatially coupled scheme with AMP decoding is bounded by  $P_e(\tau^* + \delta)$  for any fixed  $\delta > 0$ . This matches the (predicted) asymptotic UER achieved by i.i.d. Gaussian matrices and SBS-MAP decoding (Remark 3.1). This phenomenon, where the performance of message passing decoding in a spatially coupled system matches the optimal decoding performance in the corresponding uncoupled system, has been shown in other applications and is known as *threshold saturation* [13], [16], [28], [29], [35], [39].

*Proof of Theorem 2:* 1) Consider the spatially coupled state evolution (8)–(9) as a single line recursion in the vector  $\gamma^t \in \mathbb{R}^{\mathbf{R}}$ : for  $\mathbf{r} \in [\mathbf{R}]$ ,

$$\gamma_{\mathbf{r}}^{t+1} = \sum_{\mathbf{c}=1}^{\mathbf{C}} W_{\mathbf{r}\mathbf{c}} \text{mmse} \left( \sum_{\mathbf{r}'=1}^{\mathbf{R}} W_{\mathbf{r}'\mathbf{c}} \frac{1}{\sigma^2 + \mu_{\text{inner}} \gamma_{\mathbf{r}'}^t} \right). \quad (39)$$

For any base matrix  $\mathbf{W}$  with non-negative entries, the result in [16, Cor. 4.3] shows that each entry of  $\gamma^t$  is non-increasing in  $t$  and converges to a fixed point. We denote these fixed points by  $\{\gamma_r^{\text{SC-FP}}\}_{r \in [R]}$ . The arguments used in [16, Cor. 4.3] are similar to those used in the proof of the first part of Theorem 1. The entries of the state evolution parameter  $\tau^t$  given by

$$\tau_c^t = \left[ \sum_{r=1}^R \frac{W_{rc}}{\sigma^2 + \mu_{\text{inner}} \gamma_r^t} \right]^{-1}, \quad \text{for } c \in [C],$$

are non-decreasing in  $\{\gamma_r^t\}_{r \in [R]}$ . Hence we conclude that each entry of  $\tau^t$  is also non-increasing in  $t$  and converges to a fixed point; these fixed points are denoted  $\{\tau_c^{\text{SC-FP}}\}_{c \in [C]}$ .

2) The result in (36) is obtained by using results from [35] that bound the fixed points of general coupled recursions. The uncoupled state evolution (17) can be written as a single line recursion:

$$\psi^{t+1} = \text{mmse} \left( \frac{1}{\sigma^2 + \mu \psi^t} \right). \quad (40)$$

The uncoupled recursion in (40) and the coupled recursion in (39) correspond exactly to [35, Eqs. (27)-(28)] when  $\mu$  of the uncoupled system is equal to  $\mu_{\text{inner}}$  of the spatially coupled system and  $\mathbf{W}$  is an  $(\omega, \Lambda, \rho = 0)$  base matrix. (We will discuss the implications of  $\rho$  being a small positive constant later.) Using the same arguments as in [35, Sec. VI.E] and the vector I-MMSE relationship [37, Thm. 2], we obtain the following result by applying [35, Theorems 1 and 2].

For  $\rho = 0$  and any  $\epsilon > 0$ , there is an  $\omega_0 < \infty$  and  $\Lambda_0 < \infty$  such that, for all  $\omega > \omega_0$  and  $\Lambda > \Lambda_0$ , the fixed point of (39) satisfies

$$\min \tilde{\mathcal{M}}(\mu_{\text{inner}}, \sigma^2) - \epsilon \leq \max_{r \in [R]} \gamma_r^{\text{SC-FP}} \leq \max \tilde{\mathcal{M}}(\mu_{\text{inner}}, \sigma^2) + \epsilon, \quad (41)$$

where

$$\begin{aligned} \tilde{\mathcal{M}}(\mu, \sigma^2) &= \arg \min_{\psi \in [0, E]} \tilde{\mathcal{F}}(\mu, \sigma^2, \psi), \\ \tilde{\mathcal{F}}(\mu, \sigma^2, \psi) &= 2 \left\{ I \left( \mathbf{X}_{\text{sec}}; \sqrt{\frac{1}{\sigma^2 + \mu \psi}} \mathbf{X}_{\text{sec}} + \mathbf{Z} \right) \right. \\ &\quad \left. - I \left( \mathbf{X}_{\text{sec}}; \sqrt{\frac{1}{\sigma^2}} \mathbf{X}_{\text{sec}} + \mathbf{Z} \right) \right. \\ &\quad \left. + \frac{1}{2\mu} \left[ \ln \left( 1 + \frac{\mu \psi}{\sigma^2} \right) - \frac{\mu \psi}{\sigma^2 + \mu \psi} \right] \right\}. \end{aligned}$$

Here  $\mathbf{X}_{\text{sec}} \sim p_{\mathbf{X}_{\text{sec}}}$  and  $\mathbf{Z} \in \mathbb{R}^B$  is a standard Gaussian vector independent of  $\mathbf{X}_{\text{sec}}$ . Since  $\tilde{\mathcal{F}}(\mu, \sigma^2, \psi)$  and the potential function  $\mathcal{F}(\mu, \sigma^2, \psi)$  defined in (19) are equivalent after removing constant scaling factors and terms that don't depend on  $\psi$ , their minimizers with respect to  $\psi$  are identical. Therefore, we can write (41) as

$$\min \mathcal{M}(\mu_{\text{inner}}, \sigma^2) - \epsilon \leq \max_{r \in [R]} \gamma_r^{\text{SC-FP}} \leq \max \mathcal{M}(\mu_{\text{inner}}, \sigma^2) + \epsilon, \quad (42)$$

where  $\mathcal{M}(\mu, \sigma^2)$  is the set of minimizers of  $\mathcal{F}(\mu, \sigma^2, \psi)$ .

Now we consider the effect of  $\rho$  being a small positive constant on the fixed point of the state evolution. We study this scenario as  $\rho$  needs to be lower bounded by a strictly positive constant for the AMP concentration result in (37) to

hold. First, the  $\text{mmse}(s)$  function defined in (10) is a smooth function of  $s$  on  $(0, \infty)$  [36, Prop. 7]. Therefore, the right-hand-side of (39) is a smooth function of the entries of  $\mathbf{W}$ . Hence, the fixed point of the state evolution recursion (39) is a smooth function of  $\rho$ . For  $\rho \geq 0$ , denoting this fixed point by  $\{\gamma_r^{\text{SC-FP}}(\rho)\}_{r \in [R]}$ , and letting

$$\Delta(\rho) := \max_{r \in [R]} |\gamma_r^{\text{SC-FP}}(\rho) - \gamma_r^{\text{SC-FP}}(0)|,$$

we have  $\Delta(\rho) \rightarrow 0$  as  $\rho \rightarrow 0$ . Consequently, the result for  $(\omega, \Lambda, \rho = 0)$  base matrices in (42) holds for  $(\omega, \Lambda, \rho > 0)$  base matrices with the deviation  $\epsilon$  replaced by the slightly larger value  $\epsilon + \Delta(\rho)$ . Equivalently, since  $\epsilon > 0$  is arbitrary and  $\Delta(\rho)$  is a smooth function with  $\Delta(0) = 0$ , there exists  $\rho_0 > 0$  such that for all  $\rho < \rho_0$ , the result (42) holds for  $(\omega, \Lambda, \rho > 0)$  base matrices.

We now obtain (36) using (42). For  $c \in [C]$ , we have

$$\begin{aligned} \tau_c^{\text{SC-FP}} &= \left[ \sum_{r=1}^R \frac{W_{rc}}{\sigma^2 + \mu_{\text{inner}} \gamma_r^{\text{SC-FP}}} \right]^{-1} \\ &\leq \left[ \frac{\sum_{r=1}^R W_{rc}}{\sigma^2 + \mu_{\text{inner}} \max_{r' \in [R]} \gamma_{r'}^{\text{SC-FP}}} \right]^{-1} \\ &\leq \sigma^2 + \mu_{\text{inner}} (\max \mathcal{M}(\mu_{\text{inner}}, \sigma^2) + \epsilon), \end{aligned}$$

where the last inequality is obtained using the  $\sum_{r=1}^R W_{rc} = 1$  constraint on base matrices, and the upper bound in (42). The result (36) follows by recalling from (6) that  $\mu_{\text{inner}} = \vartheta \mu$ , where  $\vartheta = 1 + (\omega - 1)/\Lambda$ .

3) We now prove (37). Consider the input to the AMP hard-decision step in iteration  $t + 1$ , which we denote by  $\mathbf{s}^t = \mathbf{x}^t + (\tilde{\mathbf{S}}^t \odot \mathbf{A})^* \mathbf{q}^t$  (see (7), (14)). For  $\ell \in [L]$ , we denote by  $\mathbf{a}_\ell \in \mathbb{R}^B$  the  $\ell$ th section of a vector  $\mathbf{a} \in \mathbb{R}^{LB}$ .

The results in [32] and [18] imply that for any pseudo-Lipschitz function  $\xi : \mathbb{R}^B \times \mathbb{R}^B \rightarrow \mathbb{R}$ , the following holds almost surely:

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \xi(\mathbf{x}_\ell, \mathbf{s}_\ell^t) = \frac{1}{C} \sum_{c=1}^C \mathbb{E} \{ \xi(\mathbf{X}_{\text{sec}}, \mathbf{S}_{\tau_c^t}) \}, \quad (43)$$

where the limit is taken with  $L/n = \mu$  held constant,  $\mathbf{X}_{\text{sec}} \sim p_{\mathbf{X}_{\text{sec}}}$  and  $\mathbf{S}_{\tau_c^t}$  is given by (18). This result was proved in [32] for the  $B = 1$  case, and extended in [18] to the setting of SPARCs where the specific distribution  $p_{\mathbf{X}_{\text{sec}}}$  given in Example 2.1 (corresponding to random codebooks) is used. The proof for more general discrete distributions is essentially the same. Then, following the same steps as (28)–(32) (using (43) instead of (30) in (31)) gives the desired result. ■

#### D. Numerical Results

Theorems 1 and 2 together with Remark 3.3 give us the asymptotic UER achieved by random linear coding and AMP decoding when i.i.d. and spatially coupled Gaussian design matrices are used. These results are given in terms of the largest stationary point and global minimum of the potential function defined in (19). In this section, we numerically evaluate these results to understand the achievable regions of random linear coding schemes with AMP decoding.



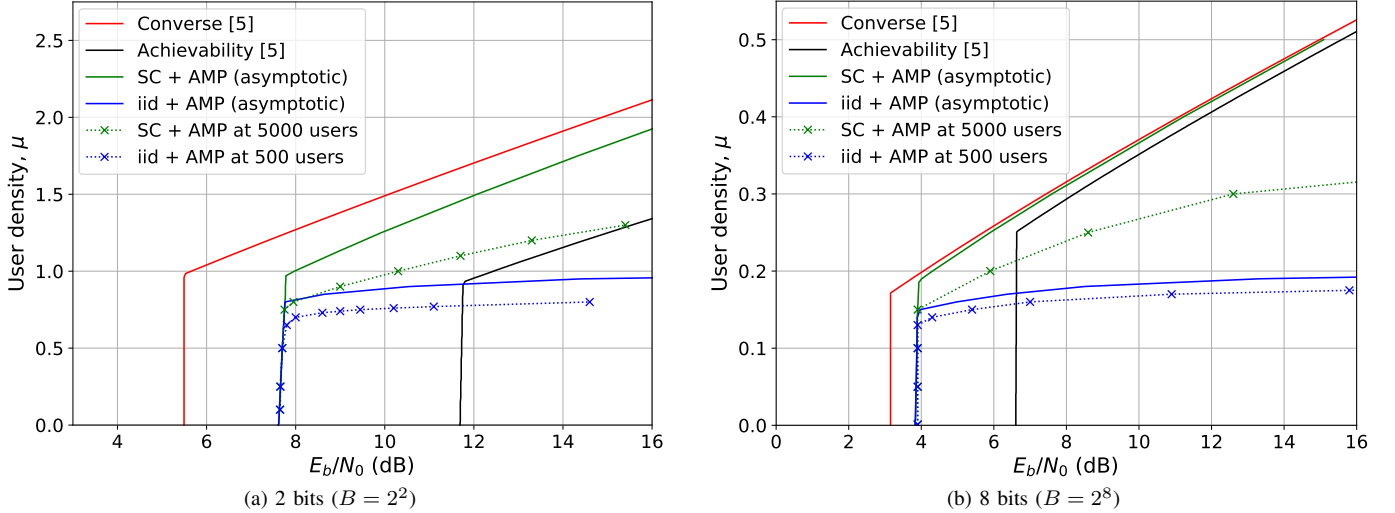


Fig. 2. Achievable regions of i.i.d. and spatially coupled Gaussian codebooks with AMP decoding, when the user payload is either 2 bits or 8 bits, and the the maximum tolerated expected (or average) UER is  $10^{-3}$ .

TABLE I  
OPTIMIZED COUPLING WIDTH VALUES USED IN FIGS. 2A AND 2B.

		Fig. 2a				
$\mu$		0.9	1.00	1.1	1.2	1.3
$\omega$		5	5	6	6	7
		Fig. 2b				
$\mu$		0.15	0.20	0.25	0.30	0.33
$\omega$		5	5	6	11	14

The solid blue and green curves in Figs. 2a and 2b plot the asymptotic achievable region of AMP decoding when the Gaussian design matrix  $\mathbf{A}$  is either i.i.d. or spatially coupled, and the sections of the message vector  $\mathbf{x}$  are drawn i.i.d. from  $p_1$  (see Example 2.1). Specifically, for a list of user densities  $\mu$ , we plot the minimum  $E_b/N_0$  required by the coding schemes to achieve a UER of less than  $10^{-3}$ , when the user payload is 2 or 8 bits. Recall that the asymptotic UER achieved by the two schemes are of the form  $P_e(\tau)$ , where  $P_e(\cdot)$  is defined in (22) and  $\tau$  is specified by (23) for the i.i.d. scheme and by (35) for the spatially coupled scheme. Furthermore, recall that  $E_b$  and  $N_0$  are related to  $E$  and  $\sigma^2$  using  $E = E_b \log_2 B$  and  $N_0 = 2\sigma^2$ .

In both Figs. 2a and 2b, the asymptotic achievable region of spatially coupled Gaussian codebooks with efficient AMP decoding (solid green) is strictly larger than the achievability bound in [5] (solid black), which is based on i.i.d. Gaussian codebooks and ML decoding. Moreover, in Fig. 2b where the user payload is 8 bits, it nearly matches the converse bound in [5] (solid red) for  $\mu \geq 0.2$ . The converse and achievability bounds in [5] are given in terms of the expected UER, which is also known as the per-user probability of error. See the end of this section for more details on the converse bound.

At low user densities ( $\mu \leq 0.80$  in Fig. 2a and  $\mu \leq 0.15$  in Fig. 2b), we observe that the minimum  $E_b/N_0$  required by the i.i.d. and spatially coupled coding schemes is the same; this is because the global minimum of the potential function coincides with its largest stationary point. However, the gap

between the achievable regions of the two schemes increases sharply for larger  $\mu$ . Furthermore, the shape of the solid blue curve for large  $E_b/N_0$  suggests that it might be impossible to achieve  $\text{UER} \leq 10^{-3}$  with i.i.d. Gaussian codebooks and AMP decoding above a certain user density ( $\mu \approx 1.0$  in Fig. 2a and  $\mu \approx 0.2$  in Fig. 2b).

The blue and green dotted lines with crosses in Figs. 2a and 2b show the simulated performance of the i.i.d. and spatially coupled coding schemes with 500 and 5000 users, respectively. For a list of user densities  $\mu$ , the crosses show the minimum  $E_b/N_0$  at which the coding scheme achieves an average UER of less than  $10^{-3}$  (averaged over many independent trials). Recall the construction details of spatially coupled matrices in Section II-A and the description of AMP decoding in Section II-B. Discrete Cosine Transform (DCT) based design matrices were used instead of Gaussian ones to reduce decoding complexity and memory usage—the error rates obtained by the two methods are similar for large matrix sizes. See [15], [17], [40, Sec. 2.5.1] for details on the DCT implementation.<sup>2</sup> The simulations for the spatially coupled coding scheme used  $(\omega, \Lambda, \rho)$  base matrices with  $\Lambda = 50, \rho = 0$ . For each user density  $\mu$ , we calculated the average UER obtained via simulations for a range of coupling widths  $\omega$  and picked the best one. Table I gives the optimized  $\omega$  for each  $\mu$ , with  $\Lambda = 50, \rho = 0$  fixed.

We observe that for both i.i.d. and spatially coupled coding schemes, the finite user and asymptotic curves match at low user densities (the near-vertical part of the curve). For the i.i.d. coding scheme (blue), although a gap between the two curves appears above a certain user density threshold ( $\mu = 0.65$  in Fig. 2a and  $\mu = 0.13$  in Fig. 2b), their overall shape remains similar. For the spatially coupled scheme (green), the gap between the asymptotic and finite user curves increases

<sup>2</sup>Although the cited works use Hadamard based design matrices and the Fast Walsh-Hadamard Transform in their simulations, our DCT construction is essentially the same.

with  $\mu$ . This gap is a finite length effect, due to the relatively small values of base matrix parameters.

Table I shows the values of the optimized coupling widths used in Figs. 2a and 2b for user densities above the near-vertical parts of each curve. We observe that the optimal coupling width increases with the user density. At lower user densities, a range of coupling widths (including the uncoupled case  $\omega = 1$ ) achieve similar average UERs.

The solid red curves in Figs. 2a and 2b show the following converse bound from [5] on the minimum  $E_b/N_0$  to achieve expected UER within a given  $\epsilon > 0$  (for fixed user density  $\mu$  and user payload  $\log_2 M$ ):

$$\frac{E_b}{N_0} \geq \max \left\{ \frac{[Q^{-1}(\frac{1}{M}) - Q^{-1}(1 - \epsilon)]^2}{2 \log_2 M}, \frac{2^{2\mu[\log_2 M - \epsilon \log_2(M-1) - H_b(\epsilon)]} - 1}{2\mu \log_2 M} \right\}, \quad (44)$$

where  $H_b(\cdot)$  is the binary entropy function.

#### IV. LARGE USER PAYLOADS

When coding with random codebooks, i.e., when the sections of  $\mathbf{x}$  are drawn i.i.d. from  $p_1$  (see Example 2.1), the size  $B$  of each section in  $\mathbf{x}$  increases exponentially with the user payload (which is  $\log_2 M = \log_2 B$  bits). For very large  $B$  it is computationally infeasible to evaluate the potential function (19). The potential function is needed to compute the asymptotic UER bounds in Theorems 1 and 2 (see (25) and (37)). In this section we bound the asymptotic UER achieved by i.i.d. and spatially coupled Gaussian codebooks with AMP decoding when the user payload is large. Furthermore, in Section IV-D we discuss simple ways to reduce the decoding complexity at larger user payloads. Both results in this section (Theorems 3 and 4) use the following lemma.

**Lemma 4.1 (Asymptotic UER Bound):** Consider the setting of either Theorem 1 or 2, and take the distribution  $p_{\mathbf{x}_{\text{sec}}}$  to be  $p_1$ . Let  $\hat{\mathbf{x}}^t$  be the AMP hard-decision estimate of  $\mathbf{x}$  after iteration  $t$ , and recall that  $\psi^t \in \mathbb{R}^C$  is an output of the state evolution (8)–(9). Then we have that the following limit exists almost surely and satisfies:

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_\ell^t \neq \mathbf{x}_\ell\} \leq \frac{4}{C} \sum_{c=1}^C \frac{\psi_c^t}{E}, \quad (45)$$

where the limit is taken with  $\frac{L}{n} = \mu$  held constant. Recall that  $C = 1$  when the design matrix  $\mathbf{A}$  has i.i.d. Gaussian entries  $\mathcal{N}(0, \frac{1}{n})$ .

The proof of Lemma 4.1 is given in Appendix A.

##### A. IID Gaussian Codebooks

**Theorem 3 (AMP Decoding of IID Gaussian Codebooks at Large User Payloads):** Consider the setting of Theorem 1 and take the distribution  $p_{\mathbf{x}_{\text{sec}}}$  to be  $p_1$ . The UER of the AMP decoder after its first iteration exhibits the following phase transition for sufficiently large payloads  $\log_2 B$ .

1) For any  $\delta \in (0, \frac{1}{2})$ , let  $f_{B,\delta} := \frac{B^{-k\delta^2}}{\delta\sqrt{\ln B}}$  where  $k$  is a positive constant. If

$$\mu \log_2 B < \frac{1}{2} \left( \frac{1}{(1 + \frac{\delta}{2}) \ln 2} - \frac{1}{E_b/N_0} \right), \quad (46)$$

then  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_\ell^1 \neq \mathbf{x}_\ell\} \leq 4f_{B,\delta}$ .

2) For any  $\tilde{\delta} \in (0, 1)$  and  $\delta_2 \in (0, \sqrt{2} - \sqrt{2 - \tilde{\delta}})$ , let  $g_{B,\tilde{\delta}} := B^{-k_1\tilde{\delta}^2}$  where  $k_1$  is a positive constant and  $h_{B,\delta_2} := \frac{B^{-\delta_2^2/2}}{\delta_2\sqrt{\ln B}} + B^{-\delta_2^2}$ . If

$$\mu \log_2 B > \frac{1}{2(1 - g_{B,\tilde{\delta}})} \left( \frac{1}{(1 - \frac{\tilde{\delta}}{2}) \ln 2} - \frac{1}{E_b/N_0} \right), \quad (47)$$

then  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_\ell^t \neq \mathbf{x}_\ell\} \geq 1 - h_{B,\delta_2}$  for all  $t \geq 1$ .

In both statements, the limits exist and are taken with  $\frac{L}{n} = \mu$  held constant.

**Remark 4.1:** From (47), we see that for any fixed values of  $\mu$  and  $\frac{E_b}{N_0}$ , the UER of AMP decoding is lower bounded by a value that approaches 1 with growing  $B$ . Therefore, the interesting regime for large user payloads is when the spectral efficiency

$$S := \mu \log_2 B = \frac{L \log_2 B}{n} \quad \text{bits/transmission}, \quad (48)$$

is of constant order. (The spectral efficiency is the total number of bits transmitted by all the users per channel use.) Theorem 3 can be extended (using analysis similar to [17], [18]) to this asymptotic regime where  $L, n, \log_2 B$  all tend to infinity with the spectral efficiency held constant. In this case, the UER of the AMP decoder exhibits the following phase transition in this large system limit almost surely:

$$\lim_{L, B, n \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_\ell^1 \neq \mathbf{x}_\ell\} = \begin{cases} 0 & \text{if } S < S_{\text{AMP}}, \\ 1 & \text{otherwise,} \end{cases} \quad (49)$$

where  $S_{\text{AMP}}$  is defined as

$$S_{\text{AMP}} := \frac{1}{2} \left( \frac{1}{\ln 2} - \frac{1}{E_b/N_0} \right). \quad (50)$$

From (49) and (50), we see that positive spectral efficiencies are achievable in this large system setting using i.i.d. Gaussian codebooks and AMP decoding if and only if  $\frac{E_b}{N_0} > \ln 2$ .

**Proof of Theorem 3:** Recalling the definition of the state evolution parameters  $\tau^t$  and  $\psi^t$  from (17), let

$$\nu^t := \frac{E}{\tau^t \ln B} = \frac{E}{(\sigma^2 + \mu \psi^t) \ln B}. \quad (51)$$

From [18, Lem. 4.1] we know that for sufficiently large  $B$  and any  $\delta \in (0, \frac{1}{2})$ ,  $\tilde{\delta} \in (0, 1)$ , we have

$$(1 - g_{B,\tilde{\delta}}) \mathbb{1}\{\nu^t < 2 - \tilde{\delta}\} < \frac{\psi^{t+1}}{E} \leq 1 - (1 - f_{B,\delta}) \mathbb{1}\{\nu^t > 2 + \delta\}, \quad (52)$$

where  $f_{B,\delta}$  and  $g_{B,\tilde{\delta}}$  are defined in the theorem statement. We now prove the two parts of the theorem in sequence.

1) Using (51) in the upper bound of (52) and recalling that  $E = E_b \log_2 B$ ,  $\sigma^2 = N_0/2$ , we obtain

$$\frac{\psi^{t+1}}{E} \leq f_{B,\delta} \quad \text{if} \quad \frac{\psi^t}{E} < \frac{\frac{2}{2+\delta} - \frac{\ln 2}{E_b/N_0}}{2\mu \ln B}. \quad (53)$$

By substituting the initial condition  $\psi^0 = E$  into (53), we have that  $\frac{\psi^1}{E} \leq f_{B,\delta}$  under the condition in (46). Furthermore, from Lemma 4.1 we know that the asymptotic UER after iteration  $t = 1$  satisfies

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{x}_\ell^1 \neq x_\ell\} \leq \frac{4\psi^1}{E} \leq 4f_{B,\delta}. \quad (54)$$

2) We prove the second statement of the theorem by first showing that under (47),

$$\frac{E}{\tau^t} < (2 - \tilde{\delta}) \ln B \quad \text{for all } t \geq 0. \quad (55)$$

For  $t = 0$ , we have  $\tau^0 = \sigma^2 + \mu E$ , and the condition in (47) ensures that  $\frac{E}{\tau^0} < (2 - \tilde{\delta}) \ln B$ .

Assume towards induction that  $\frac{E}{\tau^t} < (2 - \tilde{\delta}) \ln B$  for some  $t \geq 0$ . Then we have

$$\frac{E}{\tau^{t+1}} = \frac{E}{\sigma^2 + \mu\psi^{t+1}} < \frac{1}{(\sigma^2/E) + \mu(1 - g_{B,\tilde{\delta}})}, \quad (56)$$

where the inequality is obtained from (52) and noting that  $\nu^t = \frac{E}{\tau^t \ln B} < 2 - \tilde{\delta}$ . We further bound the right side of (56) by using the condition in (47) along with  $E = E_b \log_2 B$  and  $\sigma^2 = N_0/2$  to obtain:

$$\begin{aligned} \frac{E}{\tau^{t+1}} &< \frac{\ln B}{\frac{\ln 2}{2} \frac{1}{E_b/N_0} + \mu \log_2 B(1 - g_{B,\tilde{\delta}}) \ln 2} \\ &< \frac{\ln B}{\frac{\ln 2}{2} \frac{1}{E_b/N_0} + \frac{\ln 2}{2} \left( \frac{1}{(1 - \tilde{\delta}/2) \ln 2} - \frac{1}{E_b/N_0} \right)} \\ &= (2 - \tilde{\delta}) \ln B. \end{aligned} \quad (57)$$

This shows that (55) holds for all  $t \geq 0$ .

Using (55), we now prove that  $\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{x}_\ell^t \neq x_\ell\} \geq 1 - h_{B,\delta_2}$  for all  $t \geq 1$ . Using (25) and (22) (noting that the first equality in (25) holds for  $t \geq 0$ ), we have

$$\begin{aligned} \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{x}_\ell^{t+1} \neq x_\ell\} \\ = P_e(\tau^t) = 1 - \mathbb{E} \left[ \Phi \left( \sqrt{E/\tau^t} + Z \right)^{B-1} \right], \end{aligned} \quad (58)$$

where  $Z \sim \mathcal{N}(0,1)$  and  $\Phi(\cdot)$  is the distribution function of the standard normal. This holds specifically when  $\mathbf{X}_{\text{sec}} \sim p_1$ . Thus to show the main result we will prove

$$\mathbb{E} \left[ \Phi \left( \sqrt{E/\tau^t} + Z \right)^{B-1} \right] \leq h_{B,\delta_2}, \quad (59)$$

noting that  $\frac{E}{\tau^t} < (2 - \tilde{\delta}) \ln B$  as shown above. We will use the following bounds on the tail probability of the standard normal  $\Phi^c(u) := \mathbb{P}(Z > u)$  [41]. For all  $u > 0$ ,

$$\sqrt{\frac{2}{\pi}} \left( \frac{1}{u + \sqrt{u^2 + 4}} \right) e^{-u^2/2} < \Phi^c(u) < \frac{1}{\sqrt{2\pi}} \left( \frac{1}{u} \right) e^{-u^2/2}. \quad (60)$$

For arbitrary  $\delta_2 \in (0, \sqrt{2} - \sqrt{2 - \tilde{\delta}})$ , we have the bound

$$\begin{aligned} \mathbb{E} \left[ \Phi \left( \sqrt{E/\tau^t} + Z \right)^{B-1} \right] \\ \leq \mathbb{P} \left( |Z| > \delta_2 \sqrt{\ln B} \right) \\ + \mathbb{E} \left[ \Phi \left( \sqrt{E/\tau^t} + Z \right)^{B-1} \mid |Z| \leq \delta_2 \sqrt{\ln B} \right]. \end{aligned} \quad (61)$$

Label the two terms on the right side of (61) as  $T_1$  and  $T_2$ . Using (60), for the first term  $T_1$  we have:

$$\begin{aligned} T_1 &= 2\Phi^c \left( \delta_2 \sqrt{\ln B} \right) \leq \sqrt{\frac{2}{\pi}} \left( \frac{1}{\delta_2 \sqrt{\ln B}} \right) e^{-(\delta_2^2 \ln B)/2} \\ &\leq \frac{B^{-\delta_2^2/2}}{\delta_2 \sqrt{\ln B}}. \end{aligned} \quad (62)$$

To bound  $T_2$ , first notice that conditional on  $|Z| \leq \delta_2 \sqrt{\ln B}$  and  $\frac{E}{\tau^t} < (2 - \tilde{\delta}) \ln B$ , we have

$$0 \leq \left| \sqrt{E/\tau^t} + Z \right| \leq (\sqrt{2 - \tilde{\delta}} + \delta_2) \sqrt{\ln B}. \quad (63)$$

Using this and the substitution  $\delta_3 := \sqrt{2 - \tilde{\delta}} + \delta_2$ , we can bound  $T_2$  as follows:

$$\begin{aligned} T_2 &\stackrel{(i)}{\leq} \mathbb{E} \left[ \left[ 1 - \Phi^c \left( \delta_3 \sqrt{\ln B} \right) \right]^{B-1} \right] \\ &\stackrel{(ii)}{\leq} \mathbb{E} \left[ \exp \left\{ -(B-1) \Phi^c \left( \delta_3 \sqrt{\ln B} \right) \right\} \right] \\ &\stackrel{(iii)}{\leq} \exp \left\{ \frac{-\sqrt{2}(B-1)e^{-\delta_3^2 \ln B/2}}{\sqrt{\pi} \left( \delta_3 \sqrt{\ln B} + \sqrt{\delta_3^2 \ln B + 4} \right)} \right\} \\ &\stackrel{(iv)}{\leq} \exp \left\{ \frac{-\sqrt{2}B^{1-\delta_3^2/2}}{5\sqrt{\pi \ln B}} \right\}, \end{aligned} \quad (64)$$

where the labelled steps are obtained as follows: (i) using  $\Phi(u) \leq \Phi(|u|)$ , the second inequality in (63), and  $\Phi(u) = 1 - \Phi^c(u)$ ; (ii) using the bound  $(1-x) \leq e^{-x}$  when  $x > 0$ ; (iii) using (60); and in (iv) we use  $B-1 > (4/5)B$  when  $B > 5$ , and  $\delta_3 = \sqrt{2 - \tilde{\delta}} + \delta_2 \leq \sqrt{2}$  when  $\delta_2 \in (0, \sqrt{2} - \sqrt{2 - \tilde{\delta}})$ , so that for  $\ln B \geq 4$  we have  $\delta_3 \sqrt{\ln B} + \sqrt{\delta_3^2 \ln B + 4} < 4\sqrt{\ln B}$ .

Now we notice that (64) implies  $T_2 \leq B^{-\delta_2^2}$  since  $-\frac{\sqrt{2}B^{1-(\sqrt{2-\tilde{\delta}}+\delta_2)^2/2}}{5\sqrt{\pi \ln B}} \leq -\delta_2^2 \ln B$  for large enough  $B$  (because  $1 - (\sqrt{2 - \tilde{\delta}} + \delta_2)^2/2 > 0$ ). Finally, using (61) and combining the bounds on  $T_1$  and  $T_2$ , we conclude that  $\mathbb{E}[\Phi(\sqrt{E/\tau^t} + Z)^{B-1}] \leq h_{B,\delta_2}$ . ■

## B. Spatially Coupled Gaussian Codebooks

From Remark 4.1, we see that for large user payloads and spectral efficiencies less than  $S_{\text{AMP}}$ , one does not require spatial coupling for reliable AMP decoding. The following result shows that any spectral efficiency above  $S_{\text{AMP}}$  and below the converse can be achieved using spatially coupled Gaussian codebooks and AMP decoding. The converse is discussed in Remark 4.3.

**Theorem 4 (AMP Decoding of Spatially Coupled Gaussian Codebooks at Large User Payloads):** Consider the setting of Theorem 2 and take the distribution  $p_{\mathbf{X}_{\text{sec}}}$  to be  $p_1$ . Let  $\vartheta = 1 + \frac{\omega-1}{\Lambda}$ ,  $\mu = \frac{L}{n}$ ,  $\text{snr} = \frac{2E_b}{N_0} \mu \log_2 B$ , and define

$$\Delta := \frac{1}{2\vartheta} \ln(1 + \vartheta \text{snr}) - \mu \ln B, \quad (65)$$

$$\omega^* := \left( \frac{\vartheta \text{snr}^2}{1 + \vartheta \text{snr}} \right) \frac{1}{\Delta}, \quad (66)$$

$$\rho^* := \min \left\{ \frac{\Delta}{3 \text{snr}}, \frac{1}{2} \right\}. \quad (67)$$

Let  $\delta$  be an arbitrary constant in  $(0, \min\{\frac{\Delta}{2\mu \ln B}, \frac{1}{2}\})$  and  $S_{\text{opt}} > 0$  be the solution to

$$S_{\text{opt}} = \frac{1}{2} \log_2 \left( 1 + S_{\text{opt}} \frac{2E_b}{N_0} \right). \quad (68)$$

1) If the spectral efficiency satisfies

$$\frac{1}{\vartheta} S_{\text{AMP}} \leq \mu \log_2 B < \frac{1}{\vartheta} S_{\text{opt}}, \quad (69)$$

and the base matrix parameters satisfy  $\omega > \omega^*$  and  $0 < \rho \leq \rho^*$ , then, for  $t \geq 1$  and  $\mathbf{c} \leq \max\{\frac{\omega^t}{\omega^*}, \lceil \frac{\Lambda}{2} \rceil\}$ , we have

$$\psi_{\mathbf{c}}^t = \psi_{\Lambda - \mathbf{c} + 1}^t \leq E f_{B,\delta} \quad (70)$$

for sufficiently large  $B$ , where  $E = E_b \log_2 B$ ,  $f_{B,\delta} := \frac{B^{-k\delta^2}}{\delta \sqrt{\ln B}}$  and  $k$  is a positive constant.

2) Let  $T$  denote the first iteration for which  $\max_{\mathbf{c}} \psi_{\mathbf{c}}^t \leq E f_{B,\delta}$ . Then we have

$$T \leq \left\lceil \frac{\Lambda \omega^*}{2\omega} \right\rceil, \quad (71)$$

and the UER of the AMP decoder after  $T$  iterations satisfies the following almost surely:

$$\lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_{\ell}^T \neq \mathbf{x}_{\ell}\} \leq 4f_{B,\delta}, \quad (72)$$

where the limit is taken with  $\frac{L}{n} = \mu$  held constant.

*Proof:* The first part of Theorem 4 is a direct application of the state evolution analysis of spatially coupled SPARCs for channel coding over the (single-user) AWGN channel [18, Prop. 4.1]. The main change of variables required is that the signal-to-noise ratio in the AWGN channel is replaced by  $\text{snr} = \frac{L(E/n)}{\sigma^2} = \frac{2E_b}{N_0} \mu \log_2 B$ . Another change is that the AWGN rate  $R = \frac{L \ln B}{n}$  in [18] is replaced by  $\mu \ln B$ . The result in (72) is a direct application of Lemma 4.1. ■

*Remark 4.2:* A positive solution to (68) exists if and only if  $\frac{E_b}{N_0} > \ln 2$ .

*Remark 4.3 (Parameter Choice):* Consider spectral efficiency  $S = \mu \log_2 B$  bits/transmission. For any spectral efficiency  $S < S_{\text{opt}}$ , or equivalently any  $\frac{E_b}{N_0} > \frac{2^{2S}-1}{2S}$  (which matches the converse bound in [5] with  $B \rightarrow \infty$  and the target expected UER  $\epsilon \rightarrow 0$ ), we can choose design parameters as follows to guarantee that the AMP decoder achieves a small UER at large payloads.

1) If  $S < S_{\text{AMP}}$ , or equivalently  $\frac{E_b}{N_0} > (\frac{1}{\ln 2} - 2S)^{-1}$  for  $S < \frac{1}{2\ln 2}$ , we can use i.i.d. Gaussian codebooks. Theorem 3 then guarantees that the UER is bounded by a small constant at large payloads.

2) If  $S_{\text{AMP}} \leq S < S_{\text{opt}}$ , we can use spatially coupled codebooks with base matrix parameters  $\omega$  and  $\Lambda$  chosen as follows to satisfy the conditions of Theorem 4. Letting  $\vartheta_0 = S_{\text{opt}}/S$ , first choose  $\omega > \omega^*(\vartheta_0)$  (defined as in (66) with  $\vartheta$  replaced by  $\vartheta_0$ ). Then choose  $\Lambda$  large enough that  $\vartheta = 1 + \frac{\omega-1}{\Lambda} \leq \vartheta_0$ . This ensures that  $S < S_{\text{opt}}/\vartheta$  and  $\omega > \omega^*(\vartheta)$ . Theorem 4 then guarantees that the UER is bounded by a small constant at large payloads.

*Remark 4.4:* Theorem 4 can be extended to the setting where  $L, n, \log_2 B$  all tend to infinity with the spectral efficiency  $S =$

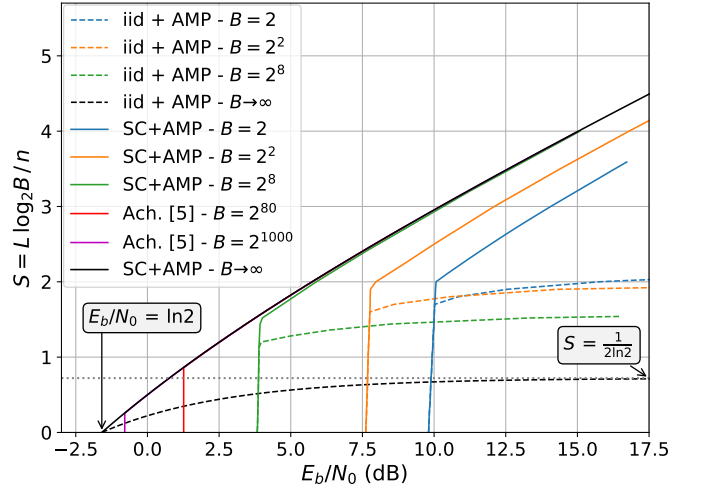


Fig. 3. Achievable regions of i.i.d. and spatially coupled Gaussian codebooks with AMP decoding at different user payloads ( $\log_2 B$  bits). The results at finite  $B$  show the minimum  $E_b/N_0$  required to achieve expected UER  $\leq 10^{-3}$ . The results for  $B = 2^{80}$  and  $2^{1000}$  use the achievability bound from [5].

$L \log_2 B/n$  held constant (see Remark 4.1). In this asymptotic regime, the result states that for any  $S_{\text{AMP}} \leq S < S_{\text{opt}}$ , the UER with AMP decoding converges almost surely to 0.

### C. Numerical Results

Fig. 3 shows the achievable regions of i.i.d. and spatially coupled Gaussian codebooks with AMP decoding, in the large system limit of  $L, n, \log_2 B$  all tending to infinity with the spectral efficiency  $S = L \log_2 B/n$  held constant (Remarks 4.1 and 4.4). Recall that the user payload is  $\log_2 M = \log_2 B$  bits. The dashed black line is the achievable region for i.i.d. Gaussian codebooks and the solid black line for spatially coupled codebooks. From (50), we see that i.i.d. Gaussian codebooks with AMP decoding cannot achieve spectral efficiencies  $S \geq \frac{1}{2\ln 2} \approx 0.7213$  (asymptote of the dashed black line). We note that the solid black line matches the converse bound in [5] with  $B \rightarrow \infty$  and the target expected UER  $\epsilon \rightarrow 0$ .

The solid and dashed black lines split Fig. 3 into three distinct regions, which are sometimes referred to as the ‘easy’, ‘hard’, and ‘impossible’ regions of inference in statistical physics [42]: 1) Below dashed black line: achievable with i.i.d. Gaussian codebooks and AMP decoding; 2) Between solid and dashed black lines: achievable with spatially coupled Gaussian codebooks and AMP decoding (or with i.i.d. Gaussian codebooks and symbol-by-symbol MAP decoding, see Remark 3.1); 3) Above solid black line: not achievable by any scheme.

In Fig. 3, we also plot the achievable regions of i.i.d. Gaussian codebooks (dashed lines) and spatially coupled Gaussian codebooks (solid lines) with AMP decoding at several finite payloads  $\log_2 B$  (with  $L, n \rightarrow \infty$  and the user density  $\mu = L/n$  held constant). For  $B = 2, 2^2, 2^8$ , we use the same setup as Fig. 2, and find the smallest  $\frac{E_b}{N_0}$  such that the coding scheme achieves UER  $\leq 10^{-3}$ . For  $B = 2^{80}, 2^{1000}$ , it is computationally infeasible to evaluate the potential function

(19), so we plot the achievability bound from [5] (red and purple curves).

For spatially coupled Gaussian codebooks with AMP decoding, the achievable region gets larger as the user payload increases, but at high spectral efficiencies (e.g.  $S > 1.5$ ), the improvement is insignificant after roughly  $\log_2 B = 8$  bits. Therefore, it is possible to communicate reliably at high spectral efficiencies with near-minimal  $\frac{E_b}{N_0}$  even when the user payload is finite. For i.i.d. Gaussian codebooks with AMP decoding, there is a tradeoff in the achievable region as the user payload increases: a lower  $E_b/N_0$  is required to communicate reliably at low spectral efficiencies, but the maximum achievable spectral efficiency decreases.

#### D. Implementation at Large User Payloads

When i.i.d. or spatially coupled codebooks are used and the user payload  $\log_2 M = \log_2 B$  bits is large, the computational complexity of the AMP decoder is too high for practical use even when DCT based design matrices are used instead of Gaussian ones. In this section we discuss the computational advantage of using smaller codebooks multiple times and the price paid in terms of a smaller achievable region. Furthermore, we discuss how introducing modulation to the encoding scheme can reduce the complexity and enlarge the achievable region.

A simple idea for reducing the complexity is to encode each user's message using several smaller codebooks instead of a single large codebook, e.g., by using a smaller codebook multiple times. For example, a user payload of 80 bits can be transmitted using a codebook of size  $B = 2^8$  ten times instead of a large codebook of size  $B = 2^{80}$  once. The smaller codebooks can be based on either i.i.d. or spatially coupled Gaussian (or DCT) matrices, and the messages can be decoded using an AMP decoder each time.<sup>3</sup>

The achievable region obtained using this method is closely related to the achievable region of the coding scheme where each user encodes a payload of 8 bits with a single codebook of size  $B = 2^8$  (the reference coding scheme). Compared to the reference coding scheme, the method described above effectively increases the code length by a factor of 10 due to repeated transmissions, and the achievable region obtained using this method in terms of the user density  $\mu$  versus  $E_b/N_0$  trade-off is the same as that obtained by the reference coding scheme, except the user density  $\mu$  is reduced by a factor of ten.<sup>4</sup>

This asymptotic achievable region of this method is shown in blue in Fig. 4, where we assumed that spatially coupled

<sup>3</sup>Another possible implementation involves the superposition of codewords from multiple (different) smaller codebooks. For example, user payloads of 80 bits can be transmitted with each user encoding their message with 10 codebooks of size  $B = 2^8$  each. The codewords from the 10 codebooks are summed together to form the final user codeword. This method is equivalent to each user encoding their message with a SPARC [14]. This implementation has the same asymptotic achievable region as using smaller codebooks multiple times but has higher decoding complexity.

<sup>4</sup>If instead we considered the spectral efficiency  $S = (\mu \times \text{user payload})$  versus  $E_b/N_0$  trade-off, then the achievable region of the described method would be exactly the same as that of the reference coding scheme. This is because the ten-fold increase in the user payload from 8 bits to 80 bits cancels out the factor of ten reduction in the user density  $\mu$ .

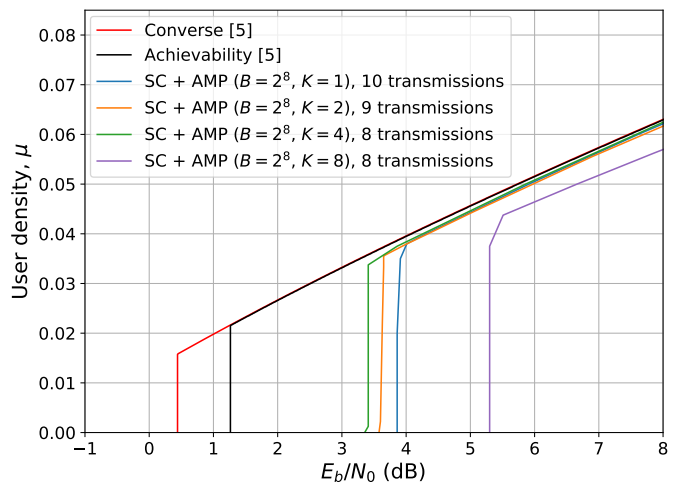


Fig. 4. Red and black: converse and achievability bounds from [5] for the many-user Gaussian MAC when the per-user payload is 80 bits and the maximum tolerated expected UER is  $10^{-3}$ . Blue, orange, green and purple: asymptotic achievable regions of coding methods based on spatially coupled complex Gaussian codebooks (of size  $B = 2^8$ ) with  $K$ -ary PSK modulation and AMP decoding for the same maximum tolerated expected UER.

complex Gaussian codebooks (of size  $B = 2^8$ ) and AMP decoding are used, and the maximum tolerated expected UER is  $10^{-3}$ . The red and black curves plot the converse and achievability bounds from [5] when the user payload is 80 bits and the maximum tolerated expected UER is also  $10^{-3}$ . We see that one can still achieve near-optimal  $\mu$  versus  $E_b/N_0$  trade-offs using these methods at user densities above  $\mu \approx 0.04$ . However, at lower user densities, there is a noticeable gap between the achievability of this method and the achievability of large ( $B = 2^{80}$ ) i.i.d. Gaussian codebooks with ML decoding (black).

We now discuss how to shrink this gap using modulation. We consider  $K$ -ary phase-shift keying (PSK) modulation in the context of complex Gaussian MACs. In the real-valued setting, this modulation technique is restricted to binary PSK.

*Introducing Modulation to Encoding:* Consider again the example where the user payload is 80 bits. If in addition to encoding 8 bits using a size  $B = 2^8$  codebook, we encode 1 extra bit in the sign of the chosen codeword (as in Example 2.2), then the final codeword would encode 9 bits and 80 bits can be sent in fewer than 9 transmissions. In comparison, to send 80 bits without modulation would require 10 transmissions of 8 bits each. For general  $K$ -ary PSK modulation, an additional  $\log_2 K$  bits are encoded the phase of the chosen codeword (using Gray coding). As the modulation factor  $K$  increases, the number of transmissions required (and the complexity) is reduced. The increase in AMP decoding complexity due to increased modulation is insignificant when  $K \ll \log(LB)$  [43, Sec. V]. In addition to reducing computational complexity, Fig. 4 shows that using binary PSK (orange) and 4-PSK modulation (green) increases the achievable region compared to the unmodulated case (blue); however, using 8-PSK (purple) significantly decreases the achievable region.

A user's message is decoded in error when either the chosen codeword from the user's codebook or the phase of the

codeword is estimated in error. We observe empirically that with 8-PSK modulation the phase of the codeword is much more likely to be estimated in error compared to 4-PSK or binary PSK, which results in a smaller achievable region.

*Complex setting:* To illustrate the effects of PSK modulation, Fig. 4 considered a complex Gaussian MAC and coding using complex Gaussian codebooks. With slight modifications to the AMP decoder, state evolution and potential function, Theorems 1 to 4 extend directly to the setting where the design matrix  $\mathbf{A}$ , message vector  $\mathbf{x}$  and channel noise are complex (see [40, Sec. 4.4] for more details). The main takeaway is that complex random linear coding with AMP decoding at user density  $\mu$  achieves the same asymptotic UER as real-valued random linear coding with AMP decoding at user density  $\frac{\mu}{2}$  when all other parameters are the same, e.g., user payload,  $E_b/N_0$  and  $p_{\mathbf{x}_{\text{sec}}}$ .

## V. CONCLUSION

In this paper we considered Gaussian MACs in the asymptotic setting where the number of users  $L$  and the code length  $n$  both tend to infinity with the user density  $\mu = L/n$  held constant, and where the user payload  $\log_2 M$  and energy-per-bit constraint are considered fixed and independent of  $n$ . We analyzed the asymptotic user error rate (the fraction of user messages decoded in error) achieved by coding schemes based on random linear models (which include i.i.d. Gaussian codebooks and random CDMA) and AMP decoding. We found that the asymptotic achievable region of a coding scheme based on spatially coupled Gaussian matrices and AMP decoding exceeds that obtained using the achievability bound in [5] and nearly matches the converse bound for a large range of user densities. The spatially coupled scheme can be interpreted as a block-wise time-division with overlap multiple-access scheme.

We then analysed the performance of these coding schemes as the user payload grows large and found that the interesting asymptotic regime is when  $L, \log_2 M, n$  all tend to infinity with the spectral efficiency  $S = L \log_2 M/n$  held constant. We also showed that using small random codebooks multiple times to transmit large user payloads can achieve near-optimal trade-offs at large user densities while having lower complexity, and adding modulation (e.g.,  $K$ -PSK modulation) to the encoding scheme can simultaneously increase the asymptotic achievable region and reduce complexity in such settings.

An interesting open question is how to close the gap between the converse bound and achievable region at low user densities in Fig. 2. Another exciting direction is to explore how ideas such as block-wise time-division with overlap (via spatially coupled codebooks) and the generalization of CDMA (by assigning a codebook of spreading sequences to each user) can be used to enhance the performance of practical techniques for multiple-access and unsourced random access. For example, in [25] the authors suggest spatial coupling as a way to improve the performance of their unsourced random access scheme based on SPARCs and AMP, and in [44], an outer code is used to induce coupling in the AMP-based random access scheme. Finally, it is also important to study the question of how spatial coupling can be applied in practical

random access settings where the number of users is random and unknown [45].

## APPENDIX A PROOF OF LEMMA 4.1

The existence of the limit in (45) is shown in Theorems 1 and 2. For  $\ell \in [L]$ , we denote by  $\mathbf{a}_\ell \in \mathbb{R}^B$  the  $\ell$ th section of a vector  $\mathbf{a} \in \mathbb{R}^{LB}$ . Let  $\mathbf{x}^{t+1}$  be the AMP estimate of  $\mathbf{x}$  after iteration  $t+1$  (defined in (13)). It was proved in [18] that the MSE of the AMP decoder after iteration  $t \geq 0$  converges almost surely to the following limit:

$$\lim_{L \rightarrow \infty} \frac{\|\mathbf{x}^{t+1} - \mathbf{x}\|^2}{L} = \frac{1}{C} \sum_{\mathbf{c}=1}^C \psi_{\mathbf{c}}^{t+1}, \quad (73)$$

where the state evolution parameters  $\psi_{\mathbf{c}}^t$  for  $\mathbf{c} \in [C]$  and  $t \geq 0$  are defined in (9)–(10).

To prove (45), we first notice that for the prior  $p_1$ , the hard-decision estimator  $\hat{\mathbf{x}}_\ell^{t+1}$  defined in (15) can equivalently be written as follows. For  $j \in \text{sec}(\ell)$ :

$$\hat{x}_j^{t+1} = \begin{cases} \sqrt{E} & \text{if } x_j^{t+1} > x_i^{t+1} \text{ for all } i \in \text{sec}(\ell) \setminus j, \\ 0 & \text{otherwise.} \end{cases} \quad (74)$$

Here  $\mathbf{x}^{t+1}$  is the AMP estimate computed according to (7) and (13).

Let  $j^* \in \text{sec}(\ell)$  denote the index of the unique non-zero entry of  $\mathbf{x}$  in section  $\ell \in [L]$ , i.e.,  $x_{j^*} = \sqrt{E}$ . From (13), we note that the sum of the entries in each section of  $\mathbf{x}^{t+1}$  equals  $\sqrt{E}$ . The decision rule (74) then implies that  $x_{j^*}^{t+1} \leq \sqrt{E}/2$  whenever  $\hat{\mathbf{x}}_\ell^{t+1} \neq \mathbf{x}_\ell$ . Therefore,  $\hat{\mathbf{x}}_\ell^{t+1} \neq \mathbf{x}_\ell$  implies that  $\|\mathbf{x}_\ell^{t+1} - \mathbf{x}_\ell\|^2 \geq E/4$ . Therefore,

$$\frac{1}{L} \sum_{\ell=1}^L \mathbb{1}\{\hat{\mathbf{x}}_\ell^{t+1} \neq \mathbf{x}_\ell\} \leq \frac{1}{L} \sum_{\ell=1}^L \frac{4\|\mathbf{x}_\ell^{t+1} - \mathbf{x}_\ell\|^2}{E}. \quad (75)$$

Combining (75) with (73) yields (45).

## ACKNOWLEDGMENT

We thank the reviewers and the Guest Editor for their comments which helped improve the paper.

## REFERENCES

- [1] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2006.
- [2] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.
- [3] X. Chen, T. Chen, and D. Guo, “Capacity of Gaussian many-access channels,” *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3516–3539, June 2017.
- [4] Y. Polyanskiy, “A perspective on massive random-access,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2017, pp. 2523–2527.
- [5] 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.
- [6] J. Ravi and T. Koch, “Capacity per unit-energy of Gaussian many-access channels,” in *Proc. IEEE Int. Symp. Inf. Theory*, 2019.
- [7] —, “On the per-user probability of error in Gaussian many-access channels,” in *Proc. Int. Zurich Seminar Inf. Commun.*, 2020, pp. 139–143.
- [8] —, “Scaling laws for Gaussian random many-access channels,” *arXiv:2012.10350*, 2021. [Online]. Available: <https://arxiv.org/abs/2012.10350>



- [9] 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.
- [10] 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.
- [11] 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. 18914–18919, 2009.
- [12] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Phys. Rev. X*, vol. 2, p. 021005, May 2012.
- [13] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7434–7464, Nov. 2013.
- [14] A. Joseph and A. R. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2541–2557, May 2012.
- [15] 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.
- [16] J. Barbier, M. Dia, and N. Macris, "Proof of threshold saturation for spatially coupled sparse superposition codes," in *Proc. IEEE Int. Symp. Inf. Theory*, July 2016, pp. 1173–1177.
- [17] 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.
- [18] C. Rush, K. Hsieh, and R. Venkataramanan, "Capacity-achieving spatially coupled sparse superposition codes with AMP decoding," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4446–4484, 2021.
- [19] 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.
- [20] D. Guo and S. Verdú, "Randomly spread CDMA: asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, 2005.
- [21] K. Takeuchi, T. Tanaka, and T. Kawabata, "Performance improvement of iterative multiuser detection for large sparsely spread CDMA systems by spatial coupling," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1768–1794, 2015.
- [22] C. Schlegel and D. Truhachev, "Multiple access demodulation in the lifted signal graph with spatial coupling," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2459–2470, 2013.
- [23] R. R. Müller, "Soft interference cancellation for random coding in massive Gaussian multiple-access," *Entropy*, vol. 23, no. 5, 2021.
- [24] S. S. Kowshik, "Improved bounds for the many-user MAC," *arXiv:2201.00866*, 2022. [Online]. Available: <https://arxiv.org/abs/2201.00866>
- [25] A. Fengler, P. Jung, and G. Caire, "SPARCs for unsourced random access," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, pp. 6894 – 6915, 2021.
- [26] V. K. Amalladinne, A. K. Pradhan, C. Rush, J.-F. Chamberland, and K. R. Narayanan, "On approximate message passing for unsourced access with coded compressed sensing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2020, pp. 2995–3000.
- [27] K. Hsieh, C. Rush, and R. Venkataramanan, "Spatially coupled sparse regression codes: Design and state evolution analysis," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2018, pp. 1016–1020.
- [28] 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.
- [29] S. Kudekar, T. Richardson, and R. L. Urbanke, "Spatially coupled ensembles universally achieve capacity under belief propagation," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7761–7813, Dec 2013.
- [30] D. J. Costello, L. Dolecek, T. E. Fuja, J. Kliewer, D. G. M. Mitchell, and R. Smarandache, "Spatially coupled sparse codes on graphs: theory and practice," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 168–176, July 2014.
- [31] 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, Feb. 2011.
- [32] 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.
- [33] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for random linear estimation with Gaussian matrices is exact," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2252–2283, 2019.
- [34] J. Barbier, N. Macris, M. Dia, and F. Krzakala, "Mutual information and optimality of approximate message-passing in random linear estimation," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4270–4303, 2020.
- [35] A. Yedla, Y.-Y. Jian, P. S. Nguyen, and H. D. Pfister, "A simple proof of Maxwell saturation for coupled scalar recursions," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6943–6965, 2014.
- [36] D. Guo, Y. Wu, S. S. Shitz, 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, 2011.
- [37] 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, 2005.
- [38] C. Rush and R. Venkataramanan, "The error probability of sparse superposition codes with approximate message passing decoding," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 3278–3303, May 2019.
- [39] S. Kudekar, T. J. Richardson, and R. L. Urbanke, "Wave-like solutions of general 1-D spatially coupled systems," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4117–4157, 2015.
- [40] K. Hsieh, "Spatially coupled sparse regression codes for single- and multi-user communications," Ph.D. dissertation, Dept. Eng., Cambridge Univ., Cambridge, UK, 2021. [Online]. Available: <https://doi.org/10.17863/CAM.70721>
- [41] P. Borjesson and C.-E. Sundberg, "Simple approximations of the error function Q(x) for communications applications," *IEEE Trans. Commun.*, vol. 27, no. 3, pp. 639–643, 1979.
- [42] L. Zdeborová and F. Krzakala, "Statistical physics of inference: thresholds and algorithms," *Advances in Physics*, vol. 65, no. 5, pp. 453–552, 2016.
- [43] K. Hsieh and R. Venkataramanan, "Modulated sparse superposition codes for the complex AWGN channel," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4385–4404, 2021.
- [44] J. R. Ebert, V. K. Amalladinne, J.-F. Chamberland, and K. R. Narayanan, "A hybrid approach to coded compressed sensing where coupling takes place via the outer code," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 4770–4774.
- [45] K.-H. Ngo, A. Lancho, G. Durisi, and A. G. i Amat, "Massive uncoordinated access with random user activity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2021, pp. 3014–3019.