Compressed Coding, AMP Based Decoding and Analog Spatial Coupling

Shansuo Liang, Chulong Liang, Junjie Ma, and Li Ping, *Fellow, IEEE*

Abstract

This paper considers a compressed-coding scheme that combines compressed sensing with forward error control coding. Approximate message passing (AMP) is used to decode the message. Based on the state evolution analysis of AMP, we derive the performance limit of compressed-coding. We show that compressed-coding can approach Gaussian capacity at a very low compression ratio. Further, the results are extended to systems involving non-linear effects such as clipping. We show that the capacity approaching property can still be maintained when generalized AMP is used to decode the message.

To approach the capacity, a low-rate underlying code should be designed according to the curve matching principle, which is complicated in practice. Instead, analog spatial-coupling is used to avoid sophisticated low-rate code design. In the end, we study the coupled scheme in a multiuser environment, where analog spatial-coupling can be realized in a distributive way. The overall block length can be shared by many users, which reduces block length per-user.

Index Terms

Compressed sensing, forward error control coding, approximate message passing, state evolution, area theorem and analog spatial-coupling.

S. Liang, C. Liang and L. Ping are with Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China (e-mail: ssliang3-c@my.cityu.edu.hk, liangchulong@qq.com, eeliping@cityu.edu.hk). J. Ma is with the John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA (e-mail: junjiema@seas.harvard.edu).

I. INTRODUCTION

This paper is concerned with the following compressed-coding scheme:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{c}, \tag{1}$$

where $\boldsymbol{c} \in \mathbb{R}^{N \times 1}$ is an forward error control (FEC) coded and modulated sequence and $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ refers to a compression matrix [1], [3]. We denote $\delta \triangleq M/N$ as the compression ratio of $\boldsymbol{A}$. For simplicity we consider the real case for (1) and the results can be directly extended to complex cases [4], [5].

Sparse regression codes, introduced in [6]–[8], can also be represented by (1). We will discuss this in Section II-B in detail using the equivalence between position modulation (PM) and Hadamard coding. Approximate message passing (AMP), originally developed for compressed sensing, has been applied to decode sparse regression codes [9]–[11]. The performance of AMP can be analyzed using a state evolution (SE) technique [12], [13]. It has been shown that detection algorithms based AMP and the related orthogonal AMP (OAMP) can potentially outperform the conventional Turbo-type detection algorithms in coded linear systems as (1) [4], [5], [14].

Spatial-coupling offers improved performance for Turbo and LDPC type codes [15], [16]. Most works on spatial-coupling are based on binary additions [15]–[18]. Analog spatial-coupling over real or complex fields have been investigated for applications involving code-division multiple-access (CDMA) [19], data-coupling [20] and compressed sensing [21]–[23]. It is shown that data-coupling can approach Gaussian capacity at a asymptotically high signal-to-noise-ratio (SNR) [20]. This is theoretically interesting, but the SNR range is outside the scope of most practical systems.

It has been shown that spatially-coupled sparse regression code (SC-SRC) is asymptotically capacity achieving [11], [24], [25]. However, SC-SRC requires a very small δ for good performance at low-to-medium SNRs. For $\boldsymbol{A}$ with a fixed M, N grows as δ decreases, which incurs an increase of memory for storing and decoding complexity. For this reason, most available simulation results of SC-SRC are for high SNR scenarios [11], [25], [26].

In this paper we study the scheme in (1) involving general FEC codes. We will show that combining some powerful techniques from signal processing and communications, namely compressed sensing, concatenated FEC coding, AMP based decoding and spatial-coupling, can offer significant performance gains. Our main findings are as follows.

- We derive the performance limit of compressed-coding with AMP based decoding. Our basic assumption is that the SE for AMP remains accurate in the presence of an FEC decoder. Based on the area property of extrinsic information transfer charts, we show that compressed-coding can approach Gaussian capacity, even though the underlying coded sequence is non-Gaussian. This is consistent with the Gaussian distribution of the signals after compression.

- We further extend the results to systems involving non-linear effects such as clipping and quantization. We show that the near-optimal performance can be maintained when generalized AMP (GAMP) is used for decoding. This alleviates the problem of high peak-to-average-power ratio (PAPR) related to Gaussian signaling. Incidentally, sparse regression codes suffer from the same problem [27].

- The above capacity approaching property requires careful code optimization using the curve matching technique, which is rate specific and lacks flexibility [28]–[30]. We will introduce a spatially-coupled compressed-coding (SC-CC) scheme to circumvent this difficulty. Compared with SC-SRC, SC-CC can offer good performance over a wider SNR range. We will provide a graphic illustration that clearly explains this advantage of SC-CC over SC-SRC.

- A code is said to be universal if it remains good after random puncturing. Such codes are useful in, e.g., type-II automatic repeat request (ARQ) applications [31]–[33]. The existing high rate coded modulation methods typically do not work well after heavy puncturing. We show that SC-CC is inherently universal and potentially capacity approaching at both low and high SNRs.

- We study SC-CC in a multiuser environment. A traditional view is that spatial-coupling will increase overall block length, which causes difficulty in applications with stringent latency requirements. Interestingly, in a multi-user system, analog spatial-coupling can be realized in a distributive way

and the overall length can be shared by many users, which effectively reduces block length per-user. This offers an interesting new solution for future multi-user wireless communication systems.

In summary, the proposed SC-CC scheme offers practical solutions to some open challenges in coding techniques: (i) a simple method to approach the ultimate capacity of Gaussian signaling (beyond that of discrete signaling), (ii) a simple treatment of non-linear effects during transmission, (iii) a low-cost universal coding and decoding strategy and (iv) a multi-user scheme with short per-user block length and good performance. These claims are supported by extensive theoretical and numerical results.

II. COMPRESSED-CODING SCHEME

A. Compressed-Coding

Fig. 1(a) illustrates the system model for compressed-coding. A binary information sequence $\boldsymbol{d} \in \mathbb{B}^{J \times 1}$ is encoded into $\boldsymbol{c} \in \mathbb{R}^{N \times 1}$ based on an FEC code $\mathcal{C}$ and a constellation $\mathcal{S}$. We assume that the entries of $\boldsymbol{c}$ are drawn from $\mathcal{S} \equiv \{s_j\}$ with equal probabilities and $\mathrm{E}[s] = 0$, $\mathrm{E}[|s|^2] = 1$. Consider transmitting $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{c}$ in an additive while Gaussian noise (AWGN) channel. The received signal $\boldsymbol{y} \in \mathbb{R}^{M \times 1}$ is given by

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ contains AWGN samples. For theoretical analysis, we assume that the entries of $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ are independent and identically distributed (i.i.d.) as $A_{m,n} \sim \mathcal{N}(0, 1/N)$[1]. The information rates of $\boldsymbol{c}$ and $\boldsymbol{A}\boldsymbol{c}$ are defined as $R_{\mathrm{C}} \triangleq J/N$ and $R_{\mathrm{AC}} \triangleq J/M$ bits per channel use (bpcu), respectively. Recall that the compression ratio of $\boldsymbol{A}$ is $\delta = M/N$, thus we have $R_{\mathrm{AC}} = R_{\mathrm{C}}/\delta$. When R_{C} is fixed, R_{AC} can be adjusted by choosing different δ. The task at the receiver is to recover $\boldsymbol{d}$ from $\boldsymbol{y}$ with known $\boldsymbol{A}$.

B. Connection to Sparse-Regression Codes

Sparse regression code can be represented by (1) with $\boldsymbol{c}$ segmented as $\boldsymbol{c}^T = [\boldsymbol{c}_1^T, ..., \boldsymbol{c}_l^T, ..., \boldsymbol{c}_L^T]$, where each $\boldsymbol{c}_l \in \mathbb{R}^{B \times 1}$ contains exact one entry of "1" and $B - 1$ "0". Such PM is equivalent to Hadamard

---

[1]Note that $\boldsymbol{A}$ is row-normalized to unit in this paper, while $\boldsymbol{A}$ is column-normalized to unit in the original AMP algorithm in [9]. We adopt row-normalization here to ensure that, at the fixed symbol power of $\boldsymbol{c}$, the symbol power of $\boldsymbol{x}$ does not change with δ. This makes the capacity expression in (10)–(12) relatively simpler.
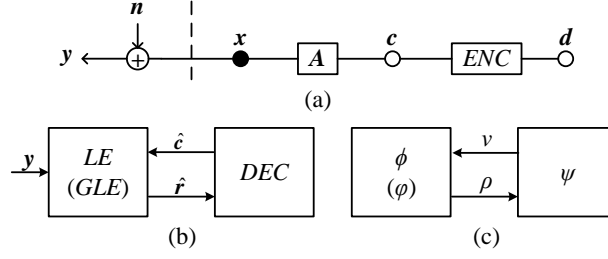
Fig. 1. Graphic illustrations for (a) system model, (b) the receiver structures and (c) state evolution. ENC and DEC denote the encoder and decoder, respectively.

coding. Too see this, let $H \in \mathbb{R}^{B \times B}$ be a Hadamard matrix over $\{-1, +1\}$ [34]. The columns of H form a Hadamard code. For c_l defined above, $\tilde{c}_l = H c_l$ is a Hadamard codeword. Define a block diagonal matrix $\widetilde{H} \triangleq \mathrm{diag}[H, H, ..., H] \in \mathbb{R}^{LB \times LB}$. Then $\tilde{c} \equiv \widetilde{H} c$ can be segmented as $\tilde{c}^T = [\tilde{c}_1^T, ..., \tilde{c}_l^T, ..., \tilde{c}_L^T]$, with each $\tilde{c}_l$ being a Hadamard codeword. Due to the orthogonality of Hadamard matrices ($\widetilde{H}^T \widetilde{H} = B \cdot I$), we have

$$x = Ac = B^{-1} A \widetilde{H}^T \widetilde{H} c = \widetilde{A} \tilde{c}, \tag{3}$$

where $\widetilde{A} \triangleq B^{-1} A \widetilde{H}^T$. Therefore, x can be generated using either PM ($x = Ac$) or Hadamard coding ($x = \widetilde{A} \tilde{c}$). Assume that the entries of A are i.i.d Gaussian. Due to the orthogonality of $\widetilde{H}$, the entries of $\widetilde{A}$ are also i.i.d Gaussian. Statistically, A and $\widetilde{A}$ are equivalent. This clearly shows the equivalence between sparse regression code and compressed-coding using a Hadamard code. Numerical results for such equivalence will be given in Fig. 7.

C. AMP-based Decoding

Initializing from $c^1 = 0$ and $r_{\mathrm{Onsager}}^1 = 0$, the AMP algorithm alternates between a linear estimator (LE) and an nonlinear estimator (NLE) as [2] [9]–[11]

$$\mathrm{LE}: \quad r^t = \hat{c}^t + \delta A^T (y - A \hat{c}^t) + r_{\mathrm{Onsager}}^t, \tag{4a}$$

$$\mathrm{NLE}: \quad \hat{c}^{t+1} = \eta(r^t), \tag{4b}$$

[2]In fact, AMP can be extended to a more general case, where $A_{m,n} \sim \mathcal{N}(0, \sigma_a^2/M)$. In this case, we can rewrite the system to $y' = \sigma_a^{-1} y = A' c + n'$, where $A'_{m,n} \sim \mathcal{N}(0, 1/M)$ and $n' \sim \mathcal{N}(0, \sigma_a^{-2} \sigma^2 I)$. Then, the original AMP algorithm and SE in [9] can be applied by replacing y, A, σ^2 with y', A', $\sigma_a^{-2} \sigma^2$, respectively. For example, for the column normalization considered in this paper, we set $\sigma_a^2 = \delta$ to make the results of this paper valid.

5

where η is a denoising function of $\boldsymbol{r}^t$. In (4a), $\boldsymbol{r}^t_{\mathrm{Onsager}}$ is an "Onsager" term defined as

$$\boldsymbol{r}^t_{\mathrm{Onsager}} \triangleq \frac{1}{\delta} \langle \eta'(\boldsymbol{r}^{t-1}) \rangle \left(\boldsymbol{r}^{t-1} - \boldsymbol{c}^{t-1} \right), \tag{5}$$

where $\langle \cdot \rangle$ denotes the average of the inputs. The final estimate is given by the hard decision based on $\hat{\boldsymbol{c}}^{T+1}$, where T is the maximum number of iterations. LE is used to handle the linear observation constraint $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{n}$ while NLE is used to explore the prior information of $\boldsymbol{c}$. The Onsager term is used to regulate correlation among messages during iterative processing.

Fig. 1(b) illustrates an iterative receiver for (2) based on AMP. The denoising function η in (4b) is given by an *a posteriori* probability (APP) decoder for the underlying FEC code $\mathcal{C}$. Its input $\boldsymbol{r}^t$ generated in (4a) is treated as a noisy observation of $\boldsymbol{c}$ using the following model [10], [11]

$$\boldsymbol{r}^t = \boldsymbol{c} + (\rho^t)^{-1/2} \cdot \boldsymbol{w}, \tag{6}$$

where $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \mathbf{I})$ is independent of $\boldsymbol{c}$ and ρ^t is the equivalent channel SNR. The output of η is the APP mean of $\boldsymbol{c}$ based on $\boldsymbol{r}^t$ in (6) as

$$\hat{\boldsymbol{c}}^t = \eta(\boldsymbol{r}^t) \triangleq \mathrm{E}[\boldsymbol{c}|\boldsymbol{r}^t, \mathcal{S}, \mathcal{C}]. \tag{7}$$

In practice, we compute (7) as follows. We feed $\boldsymbol{r}^t$ into a soft-output decoder that generates the APP log-likelihood ratio (LLR) for each coded bit. Such APP LLRs are then used to generate the APP mean $\hat{\boldsymbol{c}}^t$ (see [35, Section IV-A]). The corresponding mean squared-error (MSE) is denoted as

$$\psi(\rho^t) = \frac{1}{N} \cdot \mathrm{E}\left[\|\boldsymbol{c} - \mathrm{E}(\boldsymbol{c}|\boldsymbol{r}^t, \mathcal{S}, \mathcal{C})\|^2 \right], \tag{8}$$

where the expectation is over the distribution of $\boldsymbol{w}$ in (6) and the underlying code $\mathcal{C}$ (including modulation over $\mathcal{S}$).

D. Evolution Analysis

Define the large system limit as $M, N \to \infty$ with a fixed $\delta \in (0, \infty)$. The MSE performance of AMP in the large system limit can be tracked by a scalar SE recursion [9], [12]. For the t-th iteration, let v^t

be the *a priori* variance at the input of LE and ρ^t the *a priori* SNR at the input of the FEC decoder. Initializing with $v^1 = 1$, the SE recursion at the t-th iteration is given by [12]

$$\rho^t = \phi(v^t) \quad \text{and} \quad v^{t+1} = \psi(\rho^t), \tag{9}$$

where $\phi(v^t)$ gives the SNR at LE output and $\psi(\rho^t)$ gives the variance at decoder output. Fig. 1(c) illustrates the SE recursion.

The accuracy of SE was proved for AMP and GAMP under the assumption that η is "separable" [12], [13]. In general, an FEC decoder cannot be regarded as "separable". The discussions in this paper are based on Assumption 1 below. Simulation results will be provided to support this assumption.

Assumption 1: SE is accurate for both AMP and GAMP algorithms involving an FEC decoder in the large system limit including a fixed δ that is arbitrarily close to zero.

Fig. 2(a) illustrated the SE recursion for a general ψ. Starting from $v^1 = 1$, the zigzag curve between ϕ and ψ illustrates the iterative recovery trajectory as $v^\infty = \psi(\phi(\cdots\psi(\phi(v^1 = 1))))$. The fixed point of SE is given by the first intersection of ϕ and ψ shown as (ρ_1, v_1). The final MSE performance of the AMP algorithm is given by the variance of the fixed point, i.e., v_1.

Property 1: Error-free decoding is achieved when $v_1 \to 0$, which requires $\psi(\rho) \leq \phi^{-1}(\rho)$ for $\rho \in [0, \infty)$ with $v = \phi^{-1}(\rho)$ being the inverse function of $\rho = \phi(v)$.

E. Area Property

According to [9], $\phi(v)$ in (9) is given by

$$\rho = \phi(v) = \frac{\delta}{v + \sigma^2}, \quad 0 \leq v \leq 1. \tag{10}$$

The capacity of a real-valued AWGN channel with Gaussian signaling is given by

$$C_{\mathrm{G}} = 0.5 \log(1 + 1/\sigma^2). \tag{11}$$

From (10)–(11), we have the following lemma:

Lemma 1:

$$\frac{1}{2} \int_0^1 \phi(v)\, \mathrm{d}v = \delta \cdot C_{\mathrm{G}}. \tag{12}$$
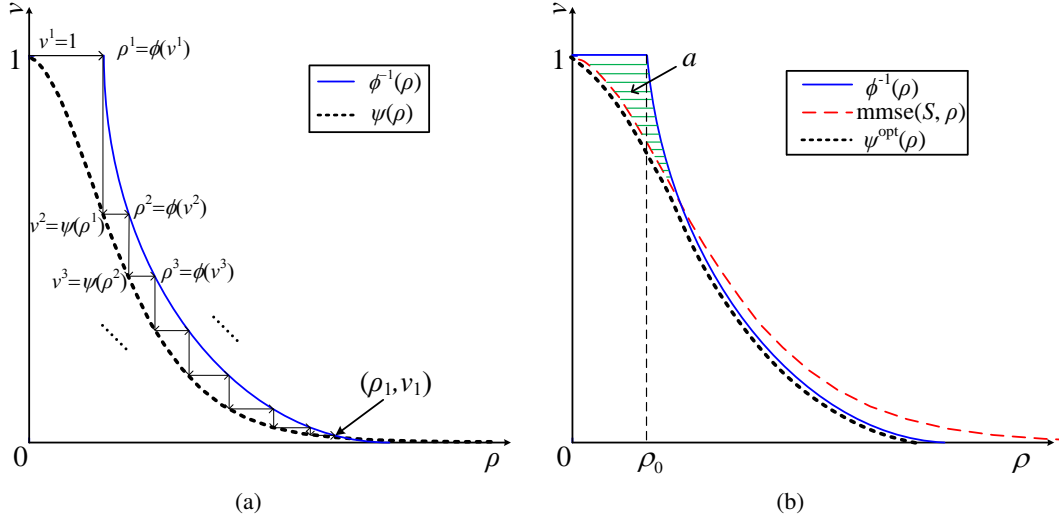
Fig. 2. Graphic illustrations for (a) an SE recursion with a general decoding function ψ and (b) the optimal ψ^{opt} satisfying the matching condition. The functions ϕ in two figures are the same.

The left hand side of (12) can be interpreted as the area under ϕ scaled by 0.5 (for a real-valued channel). Lemma 1 bridges the area under ϕ and the AWGN channel capacity.

Assume that η for decoding $\boldsymbol{c}$ in (7) is Bayes-optimal. Then ψ in (8) gives the minimum MSE for decoding $\boldsymbol{c}$. The lemma below follows [36, Corollary 1]:

Lemma 2:

$$\frac{1}{2}\int_0^\infty \psi(\rho)\,\mathrm{d}\rho = R_{\mathrm{C}}. \tag{13}$$

F. Achievable Rate

Define the inverse function of ϕ in (10) as

$$v = \phi^{-1}(\rho) = \begin{cases} 1, & 0 \le \rho < \rho_0, \\[2mm] \frac{\delta}{\rho} - \sigma^2, & \rho \ge \rho_0, \end{cases} \tag{14}$$

where $\rho_0 \triangleq \frac{\delta}{1+\sigma^2}$ is shown in Fig. 2(b).

Let $\mathrm{mmse}(\mathcal{S}, \rho)$ be the MMSE for the symbol-by-symbol detection (without the coding constraint $\mathcal{C}$):

$$\mathrm{mmse}(\mathcal{S}, \rho) = \mathrm{E}\left[|c - \mathrm{E}(c|c + \rho^{-1/2}w, \mathcal{S})|^2\right], \tag{15}$$

8

where $w \sim \mathcal{N}(0, 1)$ is independent of c and the expectation is taken over $\mathcal{S}$. The symbol-by-symbol estimation in (15) cannot outperform the one in (8) since the latter considers the coding constraint. Thus, we have $\psi(\rho) \leq \text{mmse}(\mathcal{S}, \rho)$ for $\rho \in [0, \infty)$. Combining with Property 1 yields

$$\psi(\rho) \leq \min\{\phi^{-1}(\rho), \text{mmse}(\mathcal{S}, \rho)\}, \quad \rho \in [0, \infty). \tag{16}$$

According to Lemma 2, the achievable rate of c can be maximized by maximizing the area under ψ. From (16), the maximal area is achieved when

$$\psi^{opt}(\rho) = \min\{\phi^{-1}(\rho), \text{mmse}(\mathcal{S}, \rho)\}, \quad \rho \in [0, \infty). \tag{17}$$

Based on (17), we define

$$a \triangleq \int_0^\infty \left(\phi^{-1}(\rho) - \psi^{opt}(\rho)\right) \mathrm{d}\rho. \tag{18}$$

Fig. 2(b) shows examples of $\psi^{opt}(\rho)$ and a that are illustrated by the dot line and the shadowed area, respectively. According to Lemmas 1–2, we have

$$a = \int_0^1 \phi(v)\mathrm{d}v - \int_0^\infty \psi^{opt}(\rho)\mathrm{d}\rho = 2\delta C_\mathrm{G} - 2R_\mathrm{C}. \tag{19}$$

The overall rate after the compression matrix is given by

$$R_\mathrm{AC} = R_\mathrm{C}/\delta = C_\mathrm{G} - a/2\delta. \tag{20}$$

Eq. (20) shows that the rate of compressed-coding is away from the AWGN capacity by a gap of $a/2\delta$. In the next subsection, we show that this gap vanishes as $\delta, R_\mathrm{C} \to 0$ with $R_\mathrm{AC} = R_\mathrm{C}/\delta$ fixed.

G. Approaching Capacity

We first consider BPSK for $\mathcal{S}$. Afterwards, we extend the results to more general cases.

For BPSK, the entries of c are drawn from $\mathcal{B} = \{+1, -1\}$ with equal probabilities. From (17), we have

$$\psi_\mathcal{B}^{opt}(\rho) = \min\{\phi^{-1}(\rho), \text{mmse}(\mathcal{B}, \rho)\}, \quad \rho \in [0, \infty), \tag{21}$$

where $\text{mmse}(\mathcal{B}, \rho)$ is the MMSE for detecting BPSK from an AWGN channel [37]:

$$\text{mmse}(\mathcal{B}, \rho) = 1 - \int_{-\infty}^{+\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \tanh\left(\rho - \sqrt{\rho}x\right) \mathrm{d}x. \tag{22}$$

9

Following (18), we define

$$a_\mathcal{B} \triangleq \int_0^\infty \left(\phi^{-1}(\rho) - \psi_\mathcal{B}^{opt}(\rho) \right) \mathrm{d}\rho. \tag{23}$$

The corresponding rate is given by

$$R_{\mathrm{AC}} = C_{\mathrm{G}} - a_\mathcal{B}/2\delta. \tag{24}$$

Theorem 1: Assume that the matching condition (21) holds for BPSK. Then, $R_{\mathrm{AC}} \to C_{\mathrm{G}}$ when $\delta, R_{\mathrm{C}} \to 0$ with R_{C}/δ fixed.

Proof: See Appendix A. ∎

Next, we consider a commonly used symmetric constellation $\mathcal{S}_C$ such that if $s \in \mathcal{S}_C$ then $-s \in \mathcal{S}_C$. Such a $\mathcal{S}_C$ includes quadrature phase shift keying (QPSK) and quadrature amplitude modulation (QAM) as special examples. For such $\mathcal{S}_C$, similar to (24), we can show that

$$R_{\mathrm{AC}} = C_{\mathrm{G}} - a_{\mathcal{S}_C}/2\delta, \tag{25}$$

where $a_{\mathcal{S}_C} = \int_0^\infty \left(\phi^{-1}(\rho) - \psi_{\mathcal{S}_C}^{opt}(\rho) \right) \mathrm{d}\rho$ with the following matching condition

$$\psi_{\mathcal{S}_C}^{opt}(\rho) = \min\{\phi^{-1}(\rho), \mathrm{mmse}(\mathcal{S}_C, \rho)\}, \quad \rho \in [0, \infty). \tag{26}$$

Theorem 2: Assume that the matching condition (26) holds for $\mathcal{S}_C$. Then, $R_{\mathrm{AC}} \to C_{\mathrm{G}}$ when $\delta, R_{\mathrm{C}} \to 0$ with R_{C}/δ fixed.

Proof: Due to the symmetry, we can treat $\mathcal{S}_C$ as the sum of multiple BPSK constellations multiplied by proper scalings. Recall the assumption that constellation points are drawn from $\mathcal{S}_C$ with equal probability. According to [37, Proposition 14], we have $\mathrm{mmse}(\mathcal{S}_C, \rho) \geqslant \mathrm{mmse}(\mathcal{B}, \rho)$ for $\rho \in [0, \infty)$.

Comparing (21) and (26) yields $\psi_{\mathcal{S}_C}^{opt}(\rho) \geqslant \psi_\mathcal{B}^{opt}(\rho)$ for $\rho \in [0, \infty)$. Furthermore, we have $a_{\mathcal{S}_C} \leq a_\mathcal{B}$ according to a defined in (18). Recall Theorem 1 that $a_\mathcal{B}/2\delta \to 0$ when $\delta, R_{\mathrm{C}} \to 0$ with R_{C}/δ fixed. Since $a_{\mathcal{S}_C} \leq a_\mathcal{B}$, $a_{\mathcal{S}_C}/2\delta \to 0$ also holds. From (25), we have $R_{\mathrm{AC}} \to C_{\mathrm{G}}$ and complete the proof. ∎

Remark 1: Theorem 2 shows that R_{AC} of compressed-coding with $\mathcal{S}_C$ can approach C_G under two sufficient conditions: (a) the matching condition (26) holds and (b) both R_C and $\delta \to 0$. These conditions require an underlying low-rate FEC code $\mathcal{C}$ that meets the matching condition (26). In practice, it is a highly complicated task to design such a low-rate code (see [28]–[30] for details). There is another difficulty. Due to the Gaussian distribution of $\boldsymbol{x}$, the compressed-coding scheme suffers from a high PAPR problem.

In what follows, we will address these two difficulties in the next two sections separately.

III. COMPRESSED-CODING WITH CLIPPING

The aforementioned results are for the linear system $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{c} + \boldsymbol{n}$. In this section, we extend the results to a more general system modeled below:

$$\boldsymbol{y} = f(\boldsymbol{x}) + \boldsymbol{n} = f(\boldsymbol{A}\boldsymbol{c}) + \boldsymbol{n}, \tag{27}$$

where f is a symbol-by-symbol function. This generalized scheme arises in various practical applications. An example is the clipping function for alleviating the high PAPR problem mentioned at the end of Section II that is given by

$$f(x) \triangleq \begin{cases} Z, & x > Z, \\ x, & -Z \leq x \leq Z, \\ -Z, & x < -Z, \end{cases} \tag{28}$$

where $Z > 0$ is the clipping threshold.

Alternatively, consider a slightly different system model:

$$\boldsymbol{y} = f(\boldsymbol{A}\boldsymbol{c} + \boldsymbol{n}). \tag{29}$$

As an example, f in (29) may represent the quantization effect of low-resolution analog-to-digital conversion on the received signal. In this section, we first focus on (27). The treatment for (29) is discussed in Appendix C-C.

11

A. GAMP and State Evolution

At the receiver side, GAMP can be used to recover $\boldsymbol{c}$ from $\boldsymbol{y}$ in (27) involving nonlinearity. Similar to AMP, the MSE performance of GAMP can be characterized by a SE recursion [13], [38]. We now briefly outline the GAMP algorithm and the corresponding SE recursion. Based on the SE, we analyze the achievable rate of GAMP for (27) following the procedure in Section II.

When $M, N \to \infty$ with $\delta = M/N$ fixed, we have $\|\boldsymbol{A}\|_F^2/M \to 1$ and $\|\boldsymbol{A}\|_F^2/N \to \delta$ since $A_{m,n} \sim \mathcal{N}(0, 1/N)$ (see (2)). Initializing $\hat{\boldsymbol{c}}^1 = \boldsymbol{0}$, $\boldsymbol{s}^0 = \boldsymbol{0}$ and $v^1 = 1$, the GAMP algorithm in [38] can be summarized by the following iteration between a generalized LE (GLE) and a NLE:

$$\text{GLE}: \quad \hat{\boldsymbol{p}}^t = \boldsymbol{A}\hat{\boldsymbol{c}}^t - v^t \boldsymbol{s}^{t-1}, \tag{30a}$$

$$\boldsymbol{s}^t = \frac{g(\hat{\boldsymbol{p}}^t, \boldsymbol{y}) - \hat{\boldsymbol{p}}^t}{v^t}, \tag{30b}$$

$$\boldsymbol{r}^t = \hat{\boldsymbol{c}}^t + \frac{v^t/\delta}{1 - \langle \partial g(\hat{\boldsymbol{p}}^t, \boldsymbol{y})/\partial\hat{\boldsymbol{p}}^t \rangle} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{s}^t, \tag{30c}$$

$$\text{NLE}: \quad \hat{\boldsymbol{c}}^{t+1} = \eta(\boldsymbol{r}^t). \tag{30d}$$

For GLE, $g(\hat{\boldsymbol{p}}^t, \boldsymbol{y})$ in (30b) performs the MMSE estimation of $\boldsymbol{x}$ based on its prior $\hat{\boldsymbol{p}}^t$ and the observation $\boldsymbol{y}$ in (27) as

$$g(\hat{p}_m^t, y_m) = \mathrm{E}[x_m | \hat{p}_m^t, y_m], \quad m = 1, 2, ..., M, \tag{31}$$

where $p(\boldsymbol{x}|\hat{\boldsymbol{p}}^t) = \mathcal{N}(\hat{\boldsymbol{p}}^t, v^t\mathbf{I})$ at the t-th iteration. Similar to (6), $\boldsymbol{r}^t$ in (30c) can be treated as an AWGN observation of $\boldsymbol{c}$. For NLE, η is given by an FEC decoder as illustrated in Fig. 1(b).

The MSE performance of GAMP in (30) can be tracked by an SE recursion [13], [38]. With abuse of notation, let v^t be the *a priori* variance at the input of GLE and ρ^t the *a priori* SNR at the input of the decoder as shown in Fig. 1(c). Initializing with $v^1 = 1$, the SE recursion for (30) is given by [12], [38]

$$\rho^t = \varphi(v^t) \quad \text{and} \quad v^{t+1} = \psi(\rho^t), \tag{32}$$

where ψ is the same as (8), and φ is given as follows.

Let $p(x, \hat{p})$ be a joint Gaussian distribution as

$$[x, \hat{p}]^T \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{33}$$

where

$$\Sigma = \begin{bmatrix} 1 & 1-v \\ 1-v & 1-v \end{bmatrix}. \tag{34}$$

Denote $p(y|x)$ as the likelihood of (27). According to [38], φ in (32) is given by

$$\rho = \varphi(v) = \frac{\delta}{v}\left(1 - \frac{\mathrm{mmse}(x|\hat{p}, y)}{v}\right), \tag{35}$$

with

$$\mathrm{mmse}(x|\hat{p}, y) \equiv \mathrm{E}[|x - \mathrm{E}(x|\hat{p}, y)|^2], \tag{36}$$

where the expectation is over $p(\hat{p}, x, y)$. The following result follows [38, Claim 1]:

Property 2: In (33)–(36), $\hat{p} \to x \to y$ forms a Markov chain. Thus, $p(\hat{p}, x, y) = p(y|x)p(x, \hat{p})$, where $p(y|x)$ and $p(x, \hat{p})$ are given by (27) and (33), respectively.

B. Area Property

The area properties derived below will be useful for achievable rate analyses in the next subsection.

Since ψ is still given by (8), Lemma 2 applies to ψ in (32). Theorem 3 below is presented first to underpin the area property of φ in Theorem 4.

Consider three random variables S_1, S_2, S_3 that form a Markov chain $S_1 \to S_2 \to S_3$. Assume that S_1 and S_2 are jointly distributed as

$$[S_1, S_2]^T \sim \mathcal{N}(\mathbf{0}, \Sigma), \tag{37}$$

where Σ is given by (34). $S_2 \to S_3$ is characterized by a likelihood as $p(S_3|S_2)$. Denote the mutual information between S_1 and S_3 as $I(S_1; S_3)$ and define

$$\mathrm{mmse}(S_2|S_1, S_3) \triangleq \mathrm{E}[|S_2 - \mathrm{E}(S_2|S_1, S_3)|^2], \tag{38}$$

where the expectation is over $p(S_1, S_2, S_3)$.

Theorem 3:

$$-\frac{\partial}{\partial v}I(S_1; S_3) = \frac{1}{2v}\left(1 - \frac{\mathrm{mmse}(S_2|S_1, S_3)}{v}\right). \tag{39}$$

Proof: See Appendix B. ∎

13

Denote $I(y; x)$ as the mutual information between y and x in (27), where $x \sim \mathcal{N}(0, 1)$. Theorem 4 below establishes an area property for φ.

Theorem 4:

$$\frac{1}{2} \int_0^1 \varphi(v) dv = \delta \cdot I(y; x). \tag{40}$$

Proof: According to Property 2, applying Theorem 3 to the Markov chain $\hat{p} \to x \to y$ yields

$$-\frac{\partial}{\partial v} I(\hat{p}; y) = \frac{1}{2v} \left(1 - \frac{\text{mmse}(x|\hat{p}, y)}{v} \right). \tag{41}$$

Taking integrations of the above equation yields

$$\frac{1}{2} \int_0^1 \frac{1}{v} \left(1 - \frac{\text{mmse}(x|\hat{p}, y)}{v} \right) dv = \int_0^1 -\frac{\partial}{\partial v} I(y; \hat{p}) dv \tag{42a}$$

$$= [I(y; \hat{p})]_{v=0} - [I(y; \hat{p})]_{v=1} \tag{42b}$$

$$\overset{(a)}{=} I(y; \hat{p} = x) - I(y; \hat{p} = 0) \tag{42c}$$

$$= I(y; x), \tag{42d}$$

where (a) is due to $p(\hat{p}, x)$ in (33)–(34). Comparing (42) and (35), it is clear to get (40). ∎

Remark 2: Beyond (27), Theorem 4 holds for general systems that can be modeled by $p(\boldsymbol{y}|\boldsymbol{x}) = \prod_{m=1}^M p(y_m|x_m)$, since the underlying proof of Theorem 3 does not specify $p(S_3|S_2)$. This can be seen from the proof of Theorem 4, where $p(S_3|S_2)$ characterizes the relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ in (27).

C. Achievable Rate

With the area properties of ψ and φ obtained in Lemma 2 and Theorem 4, respectively, we now evaluate the achievable rate of the generalized compressed-coding scheme according to the curve matching principle. The property below follows [24, Lemma 4.2].

Property 3: For $v \in [0, 1]$, $\varphi(v)$ is positive and decreasing with v.

From Property 3, $\rho = \varphi(v)$ is a one-to-one mapping from $v \in [0, 1]$. Following the definition of ϕ^{-1} in (14), we define the inverse function of $\rho = \varphi(v)$ as $v \equiv \varphi^{-1}(\rho)$ that is shown by the solid line in Fig. 3.
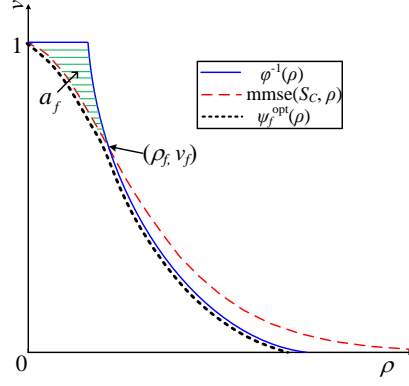
14

Fig. 3. Examples of transfer functions for GAMP.

Assume a symmetrical constellation $\mathcal{S}_C$ for c. Similar to (16), ψ in GAMP is upper bounded by

$$\psi_f(\rho) \leq \min\{\varphi^{-1}(\rho), \text{mmse}(\mathcal{S}_C, \rho)\}, \quad \rho \in [0, \infty). \tag{43}$$

According to Lemma 2, the achievable rate of ψ can be maximized when the equality holds in (43), i.e.,

$$\psi_f^{opt}(\rho) = \min\{\varphi^{-1}(\rho), \text{mmse}(\mathcal{S}_C, \rho)\}, \quad \rho \in [0, \infty). \tag{44}$$

Based on (44), define

$$a_f \triangleq \int_0^\infty \left(\varphi^{-1}(\rho) - \psi_f^{opt}(\rho) \right) \mathrm{d}\rho. \tag{45}$$

Fig. 3 shows an example of a_f that is illustrated by the shadowed area. The maximal rate of ψ_f^{opt} in (44) after compression is given by

$$R_{\text{AC}} = \frac{1}{2\delta} \int_0^\infty \psi_f^{opt}(\rho) d\rho. \tag{46}$$

Substituting (40) and (46) into (45) yields

$$a_f = 2\delta \cdot I(y; x) - 2\delta \cdot R_{\text{AC}} \Rightarrow R_{\text{AC}} = I(y; x) - a_f/2\delta. \tag{47}$$

The theorem below shows that the gap $a_f/2\delta$ vanishes when $\delta, R_{\text{C}} \to 0$ with R_{C}/δ fixed.

Theorem 5: Assume that the matching condition (44) holds for a symmetrical constellation $\mathcal{S}_C$. Then, $R_{\text{AC}} \to I(y; x)$ when $\delta, R_{\text{C}} \to 0$ with R_{C}/δ fixed.

Proof: See Appendix C-B. ∎

Theorem 5 shows that R_{AC} of the generalized compressed-coding scheme with a practical $\mathcal{S}_C$ can approach the mutual information between x and y in (27). To approach $I(y; x)$ in practice, a code should

15

be designed to meet the matching condition (44) and meanwhile the code rate should be kept as low as possible. However, it is complicated to design a low-rate code to ensure (44) using curve matching [28]–[30]. The same obstacle happens to Theorem 2 as discussed in Remark 1. Indeed, Theorem 2 is a special case of Theorem 5 since $I(y; x) = C_G$ when f in (27) is removed. In the next section, we will treat this issue together.

IV. ANALOG SPATIAL COUPLING

In this section, analog spatial-coupling is introduced to avoid the difficulty in curve matching, which provides a simple method to approach capacity without complicated code optimizations [28], [29]. Incidentally, we will see in Section V-C that SC-CC reveals a new multi-user scheme with short per-user block length and good performance.
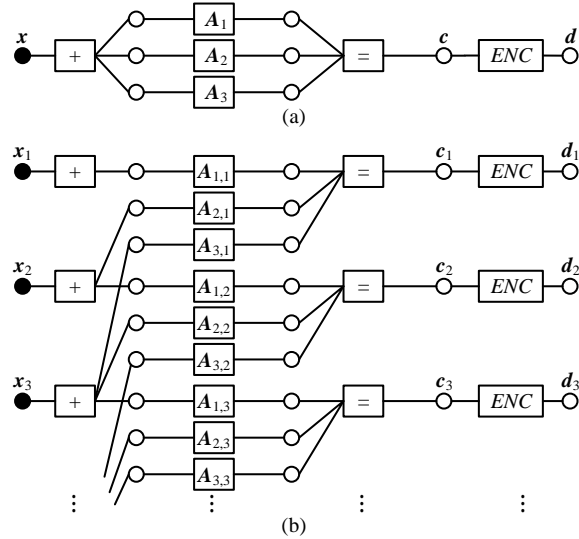
A. Spatial Coupling Principle

Fig. 4. (a) A modified version of Fig. 1(a) with $W = 3$. (b) An SC-CC system based on (a) with $W = 3$.

Fig. 4(a) shows a slightly modified form of Fig. 1(a), in which c is repeated for W times ($W = 3$ in Fig. 4(a)). Each replica is multiplied by a compression matrix A_i. The transmitted signal is

$$x = (A_1 + A_2 + \cdots + A_W)c. \tag{48}$$

When the entries of A and $\{A_i\}$ are i.i.d. Gaussian distributed, Figs. 1(a) and 4(a) are equivalent by setting $A = A_1 + A_2 + \cdots + A_W$ after proper normalization.

16

Applying the spatial-coupling principle [15], [16], [39] to K copies of Fig. 4(a), we obtain an spatially-coupled compressed-coding (SC-CC) scheme in Fig. 4(b) [1], [3], [27]. The coupling is "analog" in that the summation is on the real or complex field. Specifically, for the j-th $(1 \leq j \leq K + W - 1)$ copy, the transmitted signal is given by

$$x_j = \frac{1}{W} \sum_{k=1}^{K} A_{j+1-k,k} c_k, \tag{49}$$

where $\{A_{j+1-k,k}\}$ are assumed to be independent and the entries of each $A_{j+1-k,k}$ have the same distribution as A in (2). We assume the termination as $A_{j+1-k,k} = 0$ for $j + 1 - k < 1$ or $j + 1 - k > W$. When K is large, we approximately have $\mathrm{E}[||x|^2] = 1$ if we ignore the boundary effects for $j < W$ and $j > K$. Consider transmitting x_j over an AWGN channel as

$$y_j = x_j + n_j = \frac{1}{W} \sum_{k=1}^{K} A_{j+1-k,k} c_k + n_j, \tag{50}$$

where $j = 1, 2, ..., K + W - 1$ and n_j contains i.i.d. Gaussian noise. The AMP algorithm can be directly applied to the system in (50), of which the details can be found in [11], [24], [40].

B. State Evolution for the SC-CC System

Compared with the SE in (9) for compressed-coding, SE for the SC-CC system is a recursion between two vectors denoted as $\{v_j, j = 1, ..., K + W - 1\}$ and $\{\rho_k, k = 1, ..., K\}$. Initializing with $\{v_j^1 = 1, \forall j\}$, the vector SE given by [24, Definition 4.9] is equivalent to the following recursion

$$\rho_k^t = \frac{1}{W} \sum_{w=1}^{W} \phi\left(v_{k-1+w}^t\right), \quad \forall k, \tag{51a}$$

$$v_j^{t+1} = \frac{1}{W} \sum_{w=1}^{W} \psi\left(\rho_{j-W+w}^t\right), \quad \forall j, \tag{51b}$$

where $\rho_k^t = 0$ for $k < 1$ or $k > K$ due to the assumed termination. Functions $\phi(v)$ and $\psi(\rho)$ are the same as those in (9).

C. Potential Function Analysis

Following [41], we define the uncoupled potential function as[3]

$$U(v) = \int_{\phi(1)}^{\phi(v)} \left[\psi(\rho) - \phi^{-1}(\rho) \right] d\rho, \text{ for } v \in [0,1]. \tag{52}$$

Lemma 3: [41, Theorem 1] For the coupled recursion in (51), the variance of the fixed point is upper bounded by the minimizer of $U(v)$ when $K, W \to \infty$.

Here, we assume that $U(v)$ has a unique minimum. From Lemma 3, a sufficient condition for error-free decoding is that the minimizer of $U(v)$ tends to zero.
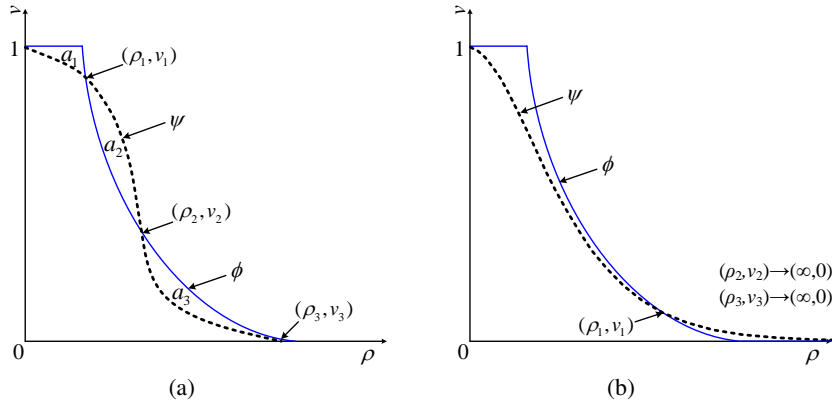

Fig. 5. Examples of ϕ and ψ in typical scenarios.

We consider typical scenarios of ϕ and ψ in Fig. 5 to illustrate Lemma 3. In Fig. 5(a), ϕ and ψ have three intersections marked as (ρ_i, v_i), $i = 1, 2, 3$. In particular, we assume that $v_3 \approx 0$, which implies error free decoding approximately. Fig. 5(b) can be seen as an special case of Fig. 5(a), where $(\rho_2, v_2) \to (\infty, 0)$ and $(\rho_3, v_3) \to (\infty, 0)$. The areas a_1, a_2 and a_3 in Fig. 5(a) are, respectively, calculated as

$$a_1 \equiv \int_{v_1}^{1} \left[\phi(v) - \psi^{-1}(v) \right] dv, \quad a_2 \equiv \int_{v_2}^{v_1} \left[\psi^{-1}(v) - \phi(v) \right] dv, \quad a_3 \equiv \int_{v_3}^{v_2} \left[\phi(v) - \psi^{-1}(v) \right] dv. \tag{53}$$

[3]Function $U(v)$ is defined based on the underlying uncoupled system in Section II. The definition here differs from [41, Equation (4)] by an additive constant. Such difference does not affect the minimizer of $U(v)$.

It can be verified that the minimizer of $U(v)$ is either v_1 or v_3. The critical point is $U(v_1) = U(v_3)$, or equivalently, $a_2 = a_3$. According to Lemma 2 and (53), we have from Fig. 5(a)

$$R_C = \frac{1}{2} \int_0^\infty \psi(\rho)\, \mathrm{d}\rho = \frac{1}{2} \left\{ \int_0^1 \phi(v)\, \mathrm{d}v - a_1 + a_2 - a_3 \right\}. \tag{54}$$

When the critical condition ($a_2 = a_3$) is reached, the achievable rate R_{AC} is given by

$$R_{AC} = \frac{R_C}{\delta} = \frac{1}{2\delta} \left\{ \int_0^1 \phi(v)\, \mathrm{d}v - a_1 \right\} = C_G - a_1/2\delta, \tag{55}$$

where the termination effects are ignored. For the special case in Fig. 5(b), according to (53), the critical point is given by $a_2 \to 0$ and $a_3 \to 0$.
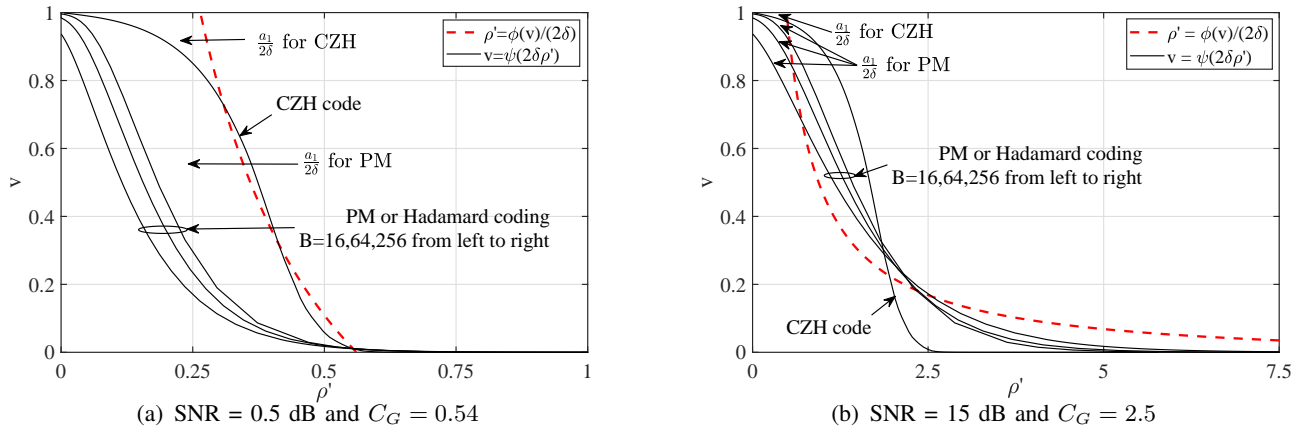
D. Achievable Rate of SC-CC

Fig. 6. Transfer functions ϕ and ψ for different underlying codes. The CZH code (non-systematic and un-punctured) has two component codes with BPSK modulation and the Hadamard code length is 64 [34]. (a) δ = 1.1977, 0.3628, 0.1065 for PM with B = 16, 64, 256, respectively. δ = 0.0916 for CZH. (b) δ = 0.1193, 0.0429, 0.0138 for PM with B = 16, 64, 256, respectively. δ = 0.0205 for CZH.

The critical condition $a_2 = a_3$ for (55) can be equivalently stated as

$$\int_0^{v_1} \phi(v)dv = \int_0^{v_1} \psi^{-1}(v)dv, \tag{56}$$

where $v_1 \in (0, 1)$ is the root of $\phi(v) = \psi^{-1}(v)$. Recall that $\phi(v) = \frac{\delta}{v+\sigma^2}$, implying that v_1 is a function of δ. Hence, the critical condition in (56) determines a unique value for δ, which in turn determines the achievable rate after compression, i.e., $R_{AC} = \frac{R_C}{\delta} = \frac{1}{2\delta} \int_0^\infty \psi(\rho)d\rho$.

Fig. 6 illustrates the curves $\rho' = \phi(v)/2\delta$ and $v = \psi(2\delta\rho')$ for PM with different B and a concatenated zigzag Hadamard (CZH) code [34]. The curves for PM are identical to those for Hadamard codes, as analyzed in Section II-B. More details on the CZH code will be given in Section V-A1. Fig. 6 involves scaling $\rho' = \frac{\phi(v)}{2\delta} = \frac{0.5}{v+\sigma^2}$ to make ρ' be independent of δ. The values of δ for all $\psi(2\delta\rho')$ curves are numerically computed such that (56) is satisfied. According to (12), (13) and (55), we have

$$\int_0^1 \phi(v)/(2\delta)dv = C_G \quad \text{and} \quad \int_0^\infty \psi(2\delta\rho')d\rho' = R_{AC} = C_G - a_1/(2\delta), \tag{57}$$

where the areas indicated by $\frac{a_1}{2\delta}$ in Fig. 6 give the gaps to the capacity.

For the relatively low SNR in Fig. 6(a), the achievable rate of SC-SRC is quite poor. This is because the slow roll-off rates of the PM curves, which lead to large values of δ from (56) and hence small values of $R_{AC} = R_C/\delta$. SC-SRC performs better at a high SNR in Fig. 6(b). The reason is that the flatter curves of $\rho' = \frac{0.5}{v+\sigma^2}$ lead to smaller values of δ and hence larger values of R_{AC}. We can see from Fig. 6 that the CZH code performs better as its curves have high roll-off rates in both low and high SNRs, resulting in quite small values of δ. Incidentally, the roll-off rate of PM increases when B increases. This is consistent of the claim that SC-SRC is asymptotically capacity approaching when $B \to \infty$. However, decoding cost can be a concern when B is large.

The above analyses provide useful insights into the rationales of SC-SRC and SC-CC. As a final note, Theorems 2 and 5 are under the assumption that $\psi(\rho)$ in the vicinity of $\rho = 0$ is given by symbol-by-symbol estimation. A classic non-concatenated code, such as PM or a Hadamard code, generally does not meet this assumption. A concatenated code fits this assumption better. The latter usually has performance close to symbol-by-symbol estimation before a certain SNR threshold and a sharp water-fall behavior afterward [16].

V. SIMULATION RESULTS

In this section, we provide simulation results to show the advantages offered by SC-CC in approaching Gaussian capacity, universal coding and short block length coding in multi-user systems.
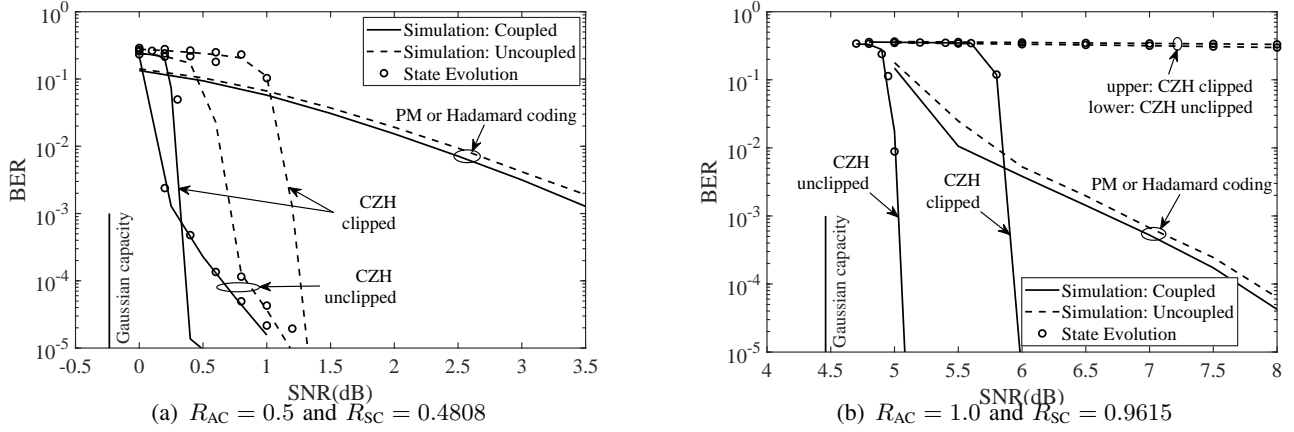
Fig. 7. Simulation results for compressed-coding. PM and Hadamard coding refer to sparse regression coding and compressed-coding, respectively. $L = 4096$ and $B = 128$ for PM. $K = 50$ and $W = 3$ for spatial-coupling. CR = 1 dB for clipping.

A. Compressed-Coding with a Low Rate Underlying Code

1) Low-Rate CZH Coding: A CZH code (non-systematic and un-punctured) with BPSK modulation is used for $\boldsymbol{c}$ with two component codes [34]. The Hadamard code length is 64, the number of information bits (per copy) is 24576 and $R_C = 3/64$. For complexity considerations, the compression matrix is generated using randomly selected rows from an $N \times N$ Hadamard matrix. We observed that the difference between Hadamard and i.i.d. Gaussian sensing matrices is typically very small for large N. Fast Hadamard transform (FHT) is used with complexity $\log_2(N)$ per bit. Soft output FHT is used for decoding the CZH code [34]. Iteration proceeds until convergence.

Theoretically, AMP requires asymptotically large sensing matrices, which incurs high cost even with FHT. Simulations show that the performance of the above scheme remains almost unchanged for $N > 65536$. Therefore, to reduce cost, each copy (coded length $= 524288$) is partitioned into 8 parts, each of length 65536, that are individually compressed.

2) Spatial-Coupling and Clipping: Recall that the theoretical analysis for spatial-coupling in Section IV-C requires $K, W \to \infty$. For practical K and W, termination incurs rate loss in spatial-coupling and the actual rate realized by SC-CC is given by $R_{\mathrm{SC}} \equiv R_{\mathrm{AC}} \cdot K/(K + W - 1)$ [15]. Fig. 7 shows the bit error rate (BER) performance of compressed-coding in various settings. For SC-CC with clipping, the

received signal is given by

$$\boldsymbol{y}_j = \alpha \cdot \text{clip}(\boldsymbol{x}_j) + \boldsymbol{n}_j, \quad j = 1, 2, ..., K + W - 1, \tag{58}$$

where $\text{clip}(\cdot)$ and $\{\boldsymbol{x}_j, \forall j\}$ are given by (28) and (49), respectively. The coefficient $\alpha \triangleq \sqrt{1/\text{E}[|\text{clip}(x)|^2]}$ normalizes the transmit power to unit. Define the clipping ratio (CR) as $\text{CR} \triangleq 10 \log_{10}(Z^2/\text{E}[x^2])$.

From Fig. 7, we have the following observations:

- The SE predictions are reasonably accurate for AMP or GAMP based decoding. Note that SE assumes i.i.d. Gaussian matrices while simulations use the Hadamard matrices due to complexity concerns.

- Spatial-coupling offers significant performance gains for compressed-coding. Such gains are marginal for sparse-regression code at low-to-medium SNRs. For the latter, the gain is more prominent at high SNRs [11], [25], [26].

- SC-SRC and SC-CC based on Hadamard codes have exactly the same BER performance. Similar results are reported in [3]. SC-CC based on a CZH code offers excellent performance. This is consistent with the analysis related to Fig. 6.

- Clipping causes performance loss in the uncoupled cases. However, very interestingly, clipping may improve performance in the coupled case. This phenomenon was first reported in [27], where an intuitive explanation is given based on the area property. The improvement is significant at $R_{SC} = 0.48$: SC-CC with clipping approaches about 0.7 dB away from the Gaussian capacity.
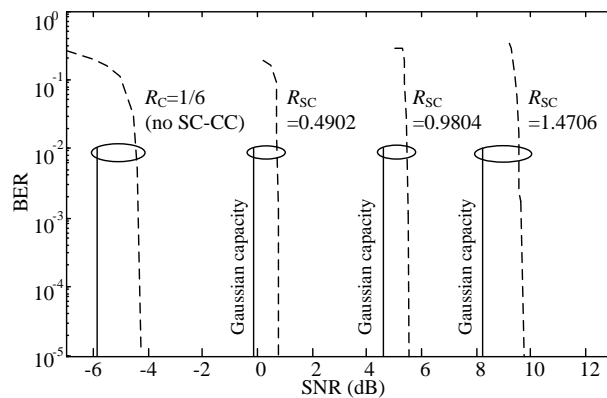
B. Universal Coding Scheme

Fig. 8. Performance of the modified SC-CC system.

22

We consider a modified version of Fig. 4(b), in which each input $\boldsymbol{d}_k$ is, after independent interleaving, encoded for W times. Graph illustration can be found in [1, Fig. 5]. We observed that the extra interleaving leads to performance improvement. However, so far, we are unable to provide analytical explanations. Fig. 8 shows the BER performance of the modified SC-CC system using a rate-$1/2$ non-systematic and un-punctured ZH code [34] with Hadamard code length $= 4$ and number of information bits per copy $= 16384$. The rate of the CZH code $= 1/6$. Each copy is partitioned into 3 parts that are individually compressed. The sensing matrices are based on a size $N_H = 32768$ Hadamard matrix. For spatial-coupling, $W = 3$ and $K = 100$.

We call this SC-CC scheme as a master code. We can puncture this master code to obtain different rates. Fig. 8 shows the simulation results for $R_{\text{SC}} = 0.4902$, 0.9804 and 1.4706, where random puncturing is used. The non-puncturing rate is $1/6$. From Fig. 8, we can see that this master code is universal since it can be randomly punctured without affecting the relative performance measured by the gaps toward capacity. At $R_{\text{SC}} = 0.4902$, nearly error-free performance is achieved at only 0.7 dB away from the theoretical limit for Gaussian signaling. Such universal codes have been widely discussed for various applications, such as type-II ARQ [32], [33] and distributed caching systems [42].

C. A Multiuser SC-CC System with Short Block Length per User

We now consider the application of SC-CC in a multi-user system of K users. The overall system structure is the same as that in Fig. 4(b) except that $\{\boldsymbol{d}_k\}$ are generated separately by K users. The transmitted signals from K users are encoded and transmitted in a decentralized way without information sharing except for proper transmission time scheduling. Specifically, we divide the time span into slots, each of which corresponds to a copy. In slot k, user k generates the signals based on its data $\boldsymbol{d}_k$ and transmits them over W slots starting from k. The signals from different users are separated by user-specific interleaving, following the interleave division multiple access (IDMA) principle [43]. At the receiver, the signals from K users are combined, which has the same effect as a linear summation in Fig. 4(b) [20].

In general, to achieve improved performance, a spatially-coupled LDPC code requires much longer

block length than the underlying LDPC code before coupling [16], [39]. This is because the former involves multiple copies of the latter. In a multiuser SC-CC system, the increased overall block length is shared by multiple users. This achieves the benefit of spatial-coupling without increasing the codeword length of each user.

Incidentally, the block length problem is usually due to the latency constraint. It cannot be solved by, e.g., increasing processing speed. It is a source problem, and in many real-time applications, the source can only generate a limited number of information bits within a fixed duration. The scheme below is to fill this fixed duration with the signals from multiple users, which effectively increases block length.
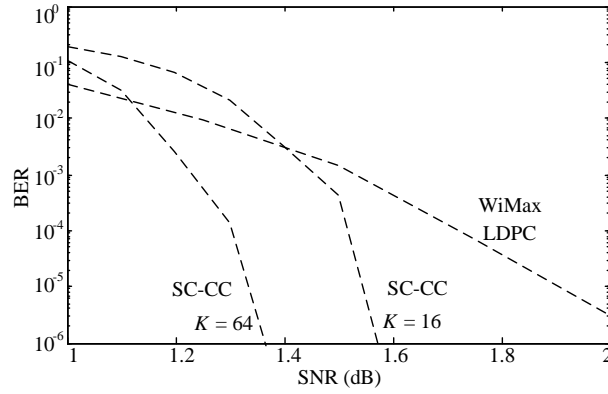

Fig. 9. Multiuser performance of the modified SC-CC scheme and the WiMax LDPC code [44]. All schemes have a coding rate 0.5.

Fig. 9 shows the performance of the SC-CC-IDMA scheme for both $K = 16$ and 64 using a rate-$1/4$ non-systematic and un-punctured ZH code with Hadamard code length = 16. The rate of the CZH code = 1/16. Number of information bits (per copy) = 1024. Each copy is partitioned into 4 parts that are individually compressed. To achieve a rate exactly $R_{AC} = 0.5$, for $K = 16$, each $\boldsymbol{A}_{j,w}$ is formed by randomly selecting 1725 rows in an Hadamard matrix with $N_H = 4096$ so $\delta = 0.4211$; for $K = 64$, each $\boldsymbol{A}_{j,w}$ is formed by randomly selecting 1956 rows in an Hadamard matrix with $N_H = 4096$ so $\delta = 0.4775$. For spatial-coupling, $W = 4$. The SC-CC performance improves with K. This is a common property of spatial-coupling [16], since the overhead due to termination reduces when K increases.

The performance of the rate-$1/2$ LDPC code for the WiMax standard [44] with 1152 information bits is compared in Fig. 9, for which users are separated by time division multiple access (TDMA). We can see that SC-CC outperforms the conventional LDPC coded TDMA scheme noticeably. Intuitively, the coding

rate of each user in SC-CC is $1/8$ (after compression) since it occupies $W = 4$ slots. This is much lower than the rate ($= 1/2$) of the LDPC code. Therefore, SC-CC might provide better coding gain if cross user interference can be ignored. However, interference does exist in SC-CC. It appears that SC-CC provides an efficient way for multiuser interference cancelation. This phenomenon was first noted in [45] for LDPC codes with SC but without compression.

CDMA with conventional successive interference cancelation (SIC) cannot help in this case. If a short code is used by each user in CDMA, decoding loss will accumulate in SIC, resulting in large overall loss. Superposition coding [46] also suffers from the same problem. Clearly, SC-CC offers an attractive solution to the latency problem by sharing a long block length by multiple users in multi-user systems.

VI. CONCLUSIONS

In this paper, we proposed a compressed-coding scheme that combines compressive sensing with FEC coding, where AMP is used for decoding. We derived the performance limit of compressed-coding and showed that compressed-coding can asymptotically approach Gaussian capacity. This capacity approaching property can be maintained in systems with non-linear effects such as clipping and quantization. We also studied an SC-CC scheme to circumvent the difficulty in optimizing low-rate codes for approaching capacity in compressed-coding. We showed that SC-CC can maintain universally good performance under random puncturing. By sharing the overall code length among multiple users, SC-CC also relieves the requirement on per-user block length to achieve good performance in multi-user environments. The above claims are supported by extensive theoretical and numerical results.

APPENDIX A

PROOF OF THEOREM 1

It can be verified that $\mathrm{mmse}(\mathcal{B}, \rho) = \phi^{-1}(\rho)$ has exactly one solution for $\rho \in [0, \infty)$, which is denoted as $\rho_{\mathcal{B}}$. Recalling $\rho = \phi(v) = \frac{\delta}{v + \sigma^2}$, we have $\rho \to 0$ as $\delta \to 0$. Apply the first-order Taylor expansion to $\mathrm{mmse}(\mathcal{B}, \rho)$ in (22) at $\rho = 0$ as [37]

$$v = \mathrm{mmse}\left(\mathcal{B}, \rho\right) = 1 - \rho + o\left(\rho^2\right). \tag{59}$$

Substituting (59) into $\phi(v)$ yields

$$\rho_{\mathcal{B}} = \frac{\delta}{1 - \rho_{\mathcal{B}} + \sigma^2} \iff \rho_{\mathcal{B}}^2 - (1 + \sigma^2)\rho_{\mathcal{B}} = -\delta \tag{60a}$$

$$\iff \left(\rho_{\mathcal{B}} - \frac{1 + \sigma^2}{2}\right)^2 = \left(\frac{1 + \sigma^2}{2}\right)^2 - \delta \tag{60b}$$

$$\overset{(a)}{\iff} \rho_{\mathcal{B}} = \frac{1 + \sigma^2}{2} - \sqrt{\left(\frac{1 + \sigma^2}{2}\right)^2 - \delta}, \tag{60c}$$

where (a) is due to that we consider the solution at the vicinity of $\rho = 0$ and the other solution (a larger

value) is abandoned.

According to (21), we have

$$\psi_{\mathcal{B}}^{opt}(\rho) = \begin{cases} \mathrm{mmse}(\mathcal{B}, \rho), & 0 \leq x \leq \rho_{\mathcal{B}}, \\ \phi^{-1}(\rho), & \rho_{\mathcal{B}} < x. \end{cases} \tag{61}$$

For $a_{\mathcal{B}}$ in (23), it can be shown from (61) that

$$a_{\mathcal{B}} = \int_0^{\rho_{\mathcal{B}}} \left(\phi^{-1}(\rho) - \mathrm{mmse}(\mathcal{B}, \rho)\right) \mathrm{d}\rho \leq \int_0^{\rho_{\mathcal{B}}} \left(1 - \mathrm{mmse}(\mathcal{B}, \rho)\right) \mathrm{d}\rho. \tag{62}$$

Combining (62) and (59) yields

$$a_{\mathcal{B}} \leq \int_0^{\rho_{\mathcal{B}}} \left(\rho + o(\rho^2)\right) \mathrm{d}\rho = 0.5\rho_{\mathcal{B}}^2 + \mathrm{o}(\rho_{\mathcal{B}}^3). \tag{63}$$

In (60c), we treat $\rho_{\mathcal{B}}$ as a function of δ and apply Taylor expansion at $\delta = 0$ as

$$\rho_{\mathcal{B}} = \frac{\delta}{1 + \sigma^2} + o(\delta). \tag{64}$$

Substituting (64) into (63), we have

$$\frac{a_{\mathcal{B}}}{2\delta} \leq 0.25 \frac{\delta}{(1 + \sigma^2)^2} + o(\delta). \tag{65}$$

As $\delta \to 0$, we can see from (65) that $a_{\mathcal{B}}/2\delta \to 0$, and therefore complete the proof.

APPENDIX B

PROOF OF THEOREM 3

Start with evaluating $I(S_1; S_3)$ as

$$I(S_1; S_2, S_3) = I(S_1; S_3, S_2) \overset{(a)}{\iff} I(S_1; S_3) + I(S_1; S_2|S_3) = I(S_1; S_2) + I(S_1; S_3|S_2) \tag{66a}$$

$$\overset{(b)}{\iff} I(S_1; S_3) = I(S_1; S_2) - I(S_1; S_2|S_3). \tag{66b}$$

In (66a), we simply rearrange the variables. In (a), we apply the chain rule of the conditional mutual information. In (b), we apply the property of Markov chain, i.e., $I(S_1; S_3|S_2) = 0$. From (66b), we evaluate $I(S_1; S_2)$ and $I(S_1; S_2|S_3)$ separately as follows.

Based on interpretation in [37], S_1 and S_3 can be viewed as two independent observations for S_2. In particular, S_1 is an AWGN observation of S_2 according to (37) and (34). From [37], we have $I(S_1; S_2) = 0.5\log(1 + \rho)$ with ρ being the effective SNR of the AWGN channel. Following the mutual information and MMSE identity for AWGN channel with side information, we have [37]

$$I(S_1; S_2|S_3 = s_3) = \frac{1}{2} \int_0^\rho \mathrm{mmse}(\rho, S_2|S_3 = s_3)d\rho, \tag{67}$$

where s_3 is a realization of S_3 and $\mathrm{mmse}(\rho, S_2|S_3 = s_3)$ represents the minimum MSE of estimating S_2 from the AWGN observation S_1 with side information $S_3 = s_3$. Taking expectation of $I(S_1; S_2|S_3 = s_3)$ with respect to (w.r.t.) S_3 yields

$$I(S_1; S_2|S_3) = \mathrm{E}[I(S_1; S_2|S_3 = s_3)] = \frac{1}{2} \int_0^\rho \mathrm{E}[\mathrm{mmse}(\rho, S_2|S_3 = s_3)]d\rho = \frac{1}{2} \int_0^\rho \mathrm{mmse}(\rho, S_2|S_3)d\rho. \tag{68}$$

For (66b), we substitute $I(S_1; S_2)$ and $I(S_1; S_2|S_3)$ and yield

$$I(S_3; S_1) = \frac{1}{2}\log(1 + \rho) - \frac{1}{2} \int_0^\rho \mathrm{mmse}(\rho, S_2|S_3)d\rho. \tag{69}$$

Note that $\mathrm{mmse}(\rho, S_2|S_3)$ in (69) is a function of the effective SNR ρ. Taking partial derivative of $I(S_3; S_1)$ w.r.t. ρ in (69) yields

$$\frac{\partial}{\partial \rho}I(S_3; S_1) = \frac{\partial}{\partial \rho}\left[\frac{1}{2}\log(1 + \rho) - \frac{1}{2} \int_0^\rho \mathrm{mmse}(\rho, S_2|S_3)d\rho\right] = \frac{1}{2(1 + \rho)} - \frac{1}{2}\mathrm{mmse}(\rho, S_2|S_3). \tag{70}$$

Next, the effective SNR ρ should be identified for the effective AWGN channel. According to (37) and (34), $p(S_1, S_2)$ is given by

$$p(S_2, S_1) \propto \exp\left(-\frac{1}{2}[S_2, S_1]\mathbf{\Sigma}^{-1}[S_2, S_1]^T\right) \propto \exp\left[-\frac{1}{2}\left(\frac{S_1^2}{v(1 - v)} - \frac{2}{v}S_2S_1 + \frac{S_2^2}{v}\right)\right]. \tag{71}$$

27

According to Bayes' rule, $p(S_1|S_2)$ is given by

$$p(S_1|S_2) = \frac{p(S_2, S_1)}{p(S_2)} \propto \frac{\exp\left[-\frac{1}{2}\left(\frac{1}{v(1-v)}S_1^2 - \frac{2}{v}S_2S_1 + \frac{1}{v}S_2^2\right)\right]}{\exp\left[-\frac{S_2^2}{2}\right]} \tag{72a}$$

$$\propto \exp\left[-\frac{1}{2}\left(\frac{S_1^2}{v(1-v)} - \frac{2}{v}S_2S_1 + \frac{1-v}{v}S_2^2\right)\right] \tag{72b}$$

$$\propto \exp\left[-\frac{S_1^2 - 2(1-v)S_2S_1 + (1-v)^2 S_2^2}{2v(1-v)}\right] \tag{72c}$$

$$\propto \exp\left[-\frac{(S_1 - (1-v)S_2)^2}{2v(1-v)}\right]. \tag{72d}$$

The conditional distribution of $p(S_1|S_2)$ is given by

$$p(S_1|S_2) = \mathcal{N}\left(S_1; (1-v)S_2, v(1-v)\right). \tag{73}$$

Eq. (73) is equivalent to the following AWGN channel

$$S_1 = (1-v) \cdot S_2 + \mathcal{N}\left(0, v(1-v)\right), \tag{74}$$

where the noise is independent of S_2. The effective SNR ρ is given by

$$\rho = \frac{(1-v)^2}{v(1-v)} \Leftrightarrow v = \frac{1}{1+\rho}. \tag{75}$$

Substituting $v = \frac{1}{1+\rho}$ into (70) yields

$$-\frac{\partial}{\partial v}I(S_3; S_1) = \left[-\frac{\partial}{\partial \rho}I(S_3; S_1)\right]_{\rho=\frac{1-v}{v}} \cdot \frac{d}{dv}\rho \tag{76a}$$

$$= \left(\frac{1}{2(1+\rho)} - \frac{\text{mmse}(\rho, S_2|S_3)}{2}\right)_{\rho=\frac{1-v}{v}} \cdot \left(\frac{1}{v^2}\right) \tag{76b}$$

$$= \frac{1}{2}[v - \text{mmse}(S_2|S_1, S_3)] \cdot \frac{1}{v^2} \tag{76c}$$

$$= \frac{1}{2v}\left(1 - \frac{\text{mmse}(S_2|S_1, S_3)}{v}\right). \tag{76d}$$

Therefore,

$$-\frac{\partial}{\partial v}I(S_3; S_1) = \frac{1}{2v}\left(1 - \frac{\text{mmse}(S_2|S_1, S_3)}{v}\right), \tag{77}$$

which completes the proof of Theorem 3.

A. An Upper Bound of $\varphi(v)$

For $\mathbf{y} = f(\mathbf{x}) + \mathbf{n}$, we assume that f does not change the average power of $\mathbf{x}$, i.e., $\mathrm{E}[|f(x)|^2] = \mathrm{E}[|x|^2]$. The following proposition gives an upper bound of φ.

Proposition 1: For $v \in [0, 1]$, $\rho = \varphi(v)$ in (35) is upper bounded by $\delta \cdot \left(\frac{1+\sigma^2}{4\sigma^2} \right)$.

Proof: Define $\mathbf{z} = f(\mathbf{x})$ and mmse$(z|y)$ as the minimum MSE of estimating $\mathbf{z}$ from $\mathbf{y} = \mathbf{z} + \mathbf{n}$. Note that $\mathrm{E}[|f(x)|^2] = 1$. According to [37], mmse$(z|y) \geq \frac{1}{1+1/\sigma^2}$. According to the data processing inequality, it is clear that mmse$(x|y) \geq$ mmse$(z|y)$.

Rewrite $\varphi(v)$ in (35) as

$$\rho = \varphi(v) = \delta \left(v + \frac{1}{\frac{1}{\text{mmse}(x|\hat{p},y)} - \frac{1}{v}} \right)^{-1}. \tag{78}$$

Combining mmse$(x|y) \geq \frac{1}{1+1/\sigma^2}$ with (78) yields

$$\rho \leq \delta \left(v + \frac{1}{1 + 1/\sigma^2 - \frac{1}{v}} \right)^{-1} = \delta \left(\frac{1 + 1/\sigma^2 - \frac{1}{v}}{v(1 + 1/\sigma^2)} \right) \tag{79a}$$

$$= \delta \left(-\frac{\sigma^2}{1 + \sigma^2} \cdot \frac{1}{v^2} + \frac{1}{v} \right) \tag{79b}$$

$$= \delta \left(-\frac{\sigma^2}{1 + \sigma^2} \left(\frac{1}{v} - \frac{1 + \sigma^2}{2\sigma^2} \right)^2 + \frac{1 + \sigma^2}{4\sigma^2} \right) \tag{79c}$$

$$\leq \delta \left(\frac{1 + \sigma^2}{4\sigma^2} \right), \tag{79d}$$

which completes the proof. ∎

B. Proof of Theorem 5

First, we prove that $R_{\text{AC}} \to I(y; x)$ as $\delta, R_{\text{C}} \to 0$ with R_{C}/δ fixed for the BPSK case as follows.

Recall $\rho = \varphi(v)$ in (35) as

$$\rho = \varphi(v) = \frac{\delta}{v} \left(1 - \frac{\text{mmse}(x|\hat{p}, y)}{v} \right) \leq \frac{\delta}{v}. \tag{80}$$

From Proposition 1, we can see that $\rho \to 0$ as $\delta \to 0$. Then, we apply Taylor expansion to mmse$(\mathcal{B}, \rho)$ at $\rho = 0$ as [37]

$$v = \text{mmse}(\mathcal{B}, \rho) = 1 - \rho + o\left(\rho^2\right). \tag{81}$$

29

Denote ρ_f as solutions of $\varphi^{-1}(\rho) = \mathrm{mmse}\,(\mathcal{B}, \rho)$. Combining (80) and (81), we have ρ_f at the vicinity of $\rho = 0$

$$\rho_f \leq \frac{\delta}{1 - \rho_f} \overset{(a)}{\Longrightarrow} \left(\rho_f - \frac{1}{2}\right)^2 \leq \frac{1}{4} - \delta \overset{(b)}{\Longrightarrow} 0 \leq \rho_f \leq \frac{1}{2} - \sqrt{\frac{1}{4} - \delta}, \tag{82}$$

where (a) and (b) are due to that $\delta \to 0$ and $\rho \to 0$. Define $\rho_f^{max} \triangleq \frac{1}{2} - \sqrt{\frac{1}{4} - \delta}$. From (45), we have

$$a_f \leq \int_0^{\rho_f} \left(1 - \mathrm{mmse}(\mathcal{B}, \rho)\right) \mathrm{d}\rho \leq \int_0^{\rho_f^{max}} \left(1 - \mathrm{mmse}(\mathcal{B}, \rho)\right) \mathrm{d}\rho \triangleq a_f^{max}. \tag{83}$$

Following (60c) and (63)–(65), we can show $\frac{(a_f^{max})^2}{2\delta} \to 0$ as $\delta \to 0$. Since $a_f \leq a_f^{max}$, we have $a_f/2\delta \to 0$ as $\delta \to 0$, which completes the proof for the BPSK case.

Following the same method in the proof of Theorem 2, the above result can be extended to a general symmetrical constellation $\mathcal{S}_C$.

C. Extension of Theorem 5

The proof of Theorem 5 relies on Proposition 1 to show that ρ is in the vicinity of zero when $\delta \to 0$. In fact, Theorem 5 holds for an arbitrary signal model as long as Proposition 1 holds. Another provable example is $\boldsymbol{y} = f(\boldsymbol{x} + \boldsymbol{n})$. When f is a one-to-one mapping function, it can be shown that $\mathrm{mmse}(x|\hat{p}, y) = \frac{v\sigma^2}{v + \sigma^2}$, which is the MMSE for $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$. When f is a many-to-one mapping function, we have $\mathrm{mmse}(x|\hat{p}, y) > \frac{v\sigma^2}{v + \sigma^2}$ due to the ambiguity from $\boldsymbol{y}$ to $\boldsymbol{x} + \boldsymbol{n}$. Thus, we have $\mathrm{mmse}(x|\hat{p}, y) \geq \frac{v\sigma^2}{v + \sigma^2}$. Combining with (78) yields

$$\rho = \varphi(v) \leq \delta \left(v + \frac{1}{\frac{v + \sigma^2}{v\sigma^2} - \frac{1}{v}}\right)^{-1} \leq \frac{\delta}{v + \sigma^2} \leq \frac{\delta}{\sigma^2}. \tag{84}$$

It is clear that $\rho \to 0$ as $\delta \to 0$.

REFERENCES

[1] C. Liang, J. Ma, and L. Ping, "Towards Gaussian capacity, universality and short block length," *Proc. 9th Int. Symp. Turbo Codes (ISTC)*, pp. 412–416, Sep 2016.

[2] S. Liang, C. Liang, J. Ma, and L. Ping, "Compressed coding and analog spatial coupling using AMP based decoding," *submitted to 2020 IEEE Global Communications Conference (GLOBECOM)*, 2020.

[3] C. Liang, J. Ma, and L. Ping, "Compressed FEC codes with spatial-coupling," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 987–990, May 2017.

[4] J. Ma, L. Liu, X. Yuan, and L. Ping, "On orthogonal AMP in coded linear vector systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5658–5672, Dec 2019.

[5] L. Liu, C. Liang, J. Ma, and L. Ping, "Capacity optimality of AMP in coded systems," *arXiv:1901.09559*, 2019.

[6] A. R. Barron and A. Joseph, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 2541–2557, Feb. 2012.

[7] A. Joseph and A. R. Barron, "Fast sparse superposition codes have near exponential error probability for $R < C$," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 919–942, Feb. 2014.

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

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

[10] 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.

[11] 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.

[12] 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.

[13] A. Javanmard and A. Montanari, "State evolution for general approximate message passing algorithms, with applications to spatial coupling," *Inf. Inference*, vol. 2, no. 2, pp. 115–144, 2013.

[14] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 1227–1231.

[15] 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.

[16] D. G. M. Mitchell, M. Lentmaier, and D. J. Costello, Jr., "Spatially coupled LDPC codes constructed from protographs," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, pp. 4866–4889, Sep. 2015.

[17] X. Ma, C. Liang, K. Huang, and Q. Zhuang, "Block Markov superposition transmission: Construction of big convolutional codes from short codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 6, pp. 3150–3163, Jun. 2015.

[18] W. Hou, S. Lu, and J. Cheng, "Spatially coupled repeater-combiner-convolutional codes," *IEEE Commun. Lett.*, vol. 20, no. 1, pp. 21–24, Jan. 2016.

[19] K. Takeuchi, T. Tanaka, and T. Kawabata, "Improvement of BP-based CDMA multiuser detection by spatial coupling," in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russian, Jul 2011, pp. 1489–1493.

[20] D. Truhachev and C. Schlegel, "Coupling data transmission for multiple-access communications," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4550–4574, Jul 2019.

[21] S. Kudekar and H. D. Pfister, "The effect of spatial coupling on compressive sensing," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2010, pp. 347–353.

[22] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Statistical physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, p. 021005, 2012.

[23] 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.

[24] J. Barbier, M. Dia, and N. Macris, "Universal sparse superposition codes with spatial coupling and GAMP decoding," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5618–5642, Sep. 2019.

[25] C. Rush, K. Hsieh, and R. Venkataramanan, "Capacity-achieving spatially coupled sparse superposition codes with AMP decoding," *arXiv:2002.07844*, 2020.

[26] K. Hsieh and R. Venkataramanan, "Modulated sparse superposition codes for the complex AWGN channel," *arXiv:2004.09549*, 2020.

[27] S. Liang, J. Ma, and L. Ping, "Clipping can improve the performance of spatially coupled sparse superposition codes," *IEEE Commun. Lett.*, vol. 21, no. 12, pp. 2578–2581, Dec 2017.

[28] G. Yue, L. Ping, and X. Wang, "Generalized low-density parity-check codes based on Hadamard constraints," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 1058–1079, Mar 2007.

[29] S. Abu-Surra, D. Divsalar, and W. E. Ryan, "Enumerators for protograph-based ensembles of LDPC and generalized LDPC codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 858–886, Feb 2011.

[30] Li Ping, W. K. Leung, and K. Y. Wu, "Low-rate turbo-Hadamard codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3213–3224, Dec 2003.

[31] S. Lin and D. J. Costello, *Error control coding*. Pearson Education India, 2004.

[32] Q. Huang, S. Chan, L. Ping, and M. Zukerman, "Improving wireless TCP throughput by a novel TCM-Based hybrid ARQ," *IEEE Trans. Wirel. Commun.*, vol. 6, no. 7, pp. 2476–2485, Jul 2007.

[33] R. Zhang and L. Hanzo, "Superposition-coding-aided multiplexed hybrid ARQ scheme for improved end-to-end transmission efficiency," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 8, pp. 4681–4686, Oct 2009.

[34] W. K. R. Leung, G. Yue, L. Ping, and X. Wang, "Concatenated zigzag Hadamard codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1711–1723, Apr. 2006.

[35] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear MMSE detection," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 6, pp. 995–1004, 2009.

[36] K. Bhattad and K. Narayanan, "An MSE-based transfer chart for analyzing iterative decoding schemes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 22–38, Jan. 2007.

[37] D. Guo, Y. Wu, S. Shamai, and S. Verdu, "Estimation in Gaussian noise: properties of the minimum mean-square error," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 2371–2385, Apr 2011.

[38] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE International Symposium on Information Theory (ISIT)*, Jul 2011, pp. 2168–2172.

[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, Aug 2015.

[40] E. Biyik, J. Barbier, and M. Dia, "Generalized approximate message-passing decoder for universal sparse superposition codes," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Jun 2017, pp. 1593–1597.

[41] 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, Nov 2014.

[42] S. Borst, V. Gupta, and A. Walid, "Distributed caching algorithms for content distribution networks," in *2010 Proceedings IEEE INFOCOM*, March 2010, pp. 1–9.

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

[44] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[45] A. Yedla, P. S. Nguyen, H. D. Pfister, and K. R. Narayanan, "Universal codes for the Gaussian MAC via spatial coupling," in *Proc. Allerton Conf. Commun., Contr. & Comput.*, Monticello, IL, USA, Sep 2011, pp. 1801–1808.

[46] X. Ma and L. Ping, "Coded modulation using superimposed binary codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3331–3343, Dec. 2004.