

Capacity Optimality of AMP in Coded Systems

Lei Liu, *Member, IEEE*, Chulong Liang, Junjie Ma, and Li Ping, *Fellow, IEEE*

Abstract—This paper studies a large random matrix system (LRMS) model involving an arbitrary signal distribution and forward error control (FEC) coding. We establish an area property based on the approximate message passing (AMP) algorithm. Under the assumption that the state evolution for AMP is correct for the coded system, the achievable rate of AMP is analyzed. We prove that AMP achieves the constrained capacity of the LRMS with an arbitrary signal distribution provided that a matching condition is satisfied. As a byproduct, we provide an alternative derivation for the constraint capacity of an LRMS using a proved property of AMP. We discuss realization techniques for the matching principle of binary signaling using irregular low-density parity-check (LDPC) codes and provide related numerical results. We show that the optimized codes demonstrate significantly better performance over un-matched ones under AMP. For quadrature phase shift keying (QPSK) modulation, bit error rate (BER) performance within 1 dB from the constrained capacity limit is observed.

Index Terms—Approximate message passing (AMP), large random matrix system, arbitrary input distributions, channel capacity, channel coding.

I. INTRODUCTION

A. Large Random Matrix System (LRMS) and Approximate Message Passing (AMP)

Consider the problem of signal reconstruction for a large random matrix system (LRMS):

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (1)$$

where \mathbf{A} is an $M \times N$ matrix with independent and identically distributed (IID) entries and \mathbf{x} a length- N vector with IID entries. Furthermore, we assume that the entries of \mathbf{A} are Gaussian, but those of \mathbf{x} are not necessarily Gaussian. If \mathbf{x} is generated using a forward error control (FEC) code with rate R_C , the overall rate of this scheme (per received symbol) is NR_C/M .

In a special case when \mathbf{x} is un-coded, if \mathbf{x} is Gaussian, the optimal solution can be obtained using the linear minimum mean square error (MMSE) methods. Otherwise, the problem

is in general NP hard [2], [3]. Approximate message passing (AMP), derived from belief-propagation (BP) with Gaussian approximation and first order Taylor approximation, has attracted extensive research interest for this problem [4], [5]. A basic assumption of AMP is that \mathbf{A} has IID Gaussian (IIDG) entries. This assumption will hold throughout this paper.

AMP works by iterating between two local processors: namely, a linear detector (LD) and a non-linear detector (NLD). (An NLD is sometimes referred to as a de-noiser [6].) There is no matrix inversion involved, so its complexity is low. AMP has been studied for various signal processing [6]–[19] and communication applications [20], [21]. Recently, it has been observed that AMP and its variations such as expectation propagation (EP) [22], [23] and orthogonal AMP (OAMP) [24] outperform the conventional Turbo linear MMSE (Turbo-LMMSE) in coded linear systems involving FEC coding [25]–[27]. The applications of such systems include inter-symbol interference (ISI) channels [25], multi-user systems [26] and multiple-input multiple-output (MIMO) systems [27]. Most works on AMP in coded systems are simulation based [25]–[27]. There is still a lack of rigorous analysis on the information theoretical limits of AMP in coded systems.

B. Contributions of this Paper

In this paper, we discuss the LRMS in (1) with FEC coding. The receiver is a variation of AMP with NLD formed by an *a posteriori probability* (APP) decoder. For convenience of discussions, we define two classes of optimality for a receiver.

- A receiver is MMSE-optimal if it can achieve MMSE when \mathbf{x} is an IID sequence.
- A receiver is information theoretically optimal if it can achieve error free performance when \mathbf{x} is coded with a rate which equals to the mutual information $I(\mathbf{x}; \mathbf{y})$.

The state evolution (SE) technique of AMP was originally derived to track the mean square error (MSE) in AMP during iterative processing. SE involves a scalar recursion of the transfer functions of LD and NLD. It has been shown via SE analysis that AMP can achieve MMSE asymptotically in the un-coded case when the transfer functions of LD and NLD have only one fixed-point [28]–[30]. In this paper, we will show via SE analysis that AMP is information theoretically optimal, while the conventional methods, such as the well-known Turbo-MMSE algorithm [31], [32] (also referred to as the Wang and Poor algorithm [27], [33], [34]), are not.

Our discussions are based on the following background works: (i) the I-MMSE relationship between mutual information and MMSE [35], (ii) the area property of iterative decoding systems [36], (iii) the MMSE-optimality of AMP [28]–[30], and (iv) the capacity of an LRMS recently derived in [28], [29], [37]. Similarly to [31], [32], the performance of

Lei Liu was with the Department of Electronic Engineering, City University of Hong Kong (CityU), Hong Kong, SAR, China, and is currently with the School of Information Science, Japan Institute of Science and Technology (JAIST), Nomi 923-1292, Japan (e-mail: leiliu@jaist.ac.jp).

Chulong Liang was with the Department of Electrical Engineering, City University of Hong Kong, Hong Kong SAR, China. He is now with the Algorithm Department, ZTE Corporation, Shenzhen 518057, China, and also with the State Key Laboratory of Mobile Network and Mobile Multimedia Technology, ZTE Corporation, Shenzhen 518057, China (e-mail: liang.chulong@zte.com.cn).

Junjie Ma is with the Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, and also with University of Chinese Academy of Science, Beijing, China (e-mail: majunjie@lsec.cc.ac.cn).

Li Ping is with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, SAR, China (e-mail: eeliping@cityu.edu.hk).

AMP can be optimized by matching the transfer functions of LD and decoder. The achievable rate can be analyzed using an area property similar to that for low density parity check (LDPC) decoders [31], [32]. However, there is a key difference. The area property for LDPC decoders is based on the so-called extrinsic information, for which perfect matching is theoretically possible [31], [32]. We will see that perfect matching is not possible for AMP: there is an inherent gap between the two transfer functions. Interestingly, AMP is still information theoretically optimal despite this gap, in the sense that its achievable rate can approach the mutual information $I(\mathbf{x}; \mathbf{y})$.

The main contributions of this paper are summarized as follows.

- We show that the constrained capacity of a coded LRMS with an arbitrary input distribution (Gaussian or non-Gaussian) can be graphically interpreted as the area determined by the transfer functions of LD and MMSE NLD of an AMP. We establish an area property for AMP and derive its achievable rate under a matching condition. We prove that this achievable rate equals to the constrained capacity of an LRMS derived in [28], [29], [37], thereby showing the potential information theoretic optimality of AMP in coded linear systems.
- We develop a matching strategy for AMP. We show the existence of a capacity approaching superposition coded modulation (SCM) scheme for Gaussian signaling. We also provide numerical results to demonstrate the efficiency of the matching strategy for binary signaling. These findings provide a promising direction to significantly enhance the performance of coded linear systems.
- As a byproduct, we provide an alternative derivation for the capacity of an LRMS. This capacity has been recently derived in [28], [29], [37]. In our opinion, the approach in this paper is more concise, taking advantage of the available results of AMP.

C. Connection to Existing Works

In [28], the authors derived the constrained capacity and MMSE of an LRMS by establishing some properties of the finite-length MMSE and mutual information sequences, and then using these properties to uniquely characterize their limits. The authors of [29], [37] provided a rigorous proof for the replica formula of the constrained capacity (see Theorem 2) by using a Guerra-Toninelli type interpolation method to yield an upper bound for the capacity, and spatial-coupling and AMP to yield a lower bound. In addition, the MMSE optimality of AMP was rigorously proved in [29], [37]. In this paper, we give a different concise derivation of the capacity. Our derivations are built on existing results, namely, the I-MMSE theorem [35], [36], the MMSE optimality of AMP [28]–[30] and the decoupling property of AMP [38].

Sparse regression coding (SRC) [17], [18] is a special case of (1) in which \mathbf{x} is generated using position modulation. A position modulation scheme of block length B is a special form of FEC coding with coding rate $R_{\text{PM}} = \log_2(B)/B$, which is also equivalent to a length- B Hadamard code [19], [20].

The decoding technique for SRC [17], [18] can be regarded as a special form of AMP, with FEC decoding implemented by position demodulation. SRC is capacity approaching when $B \rightarrow \infty$ or, equivalently, $R_{\text{PM}} \rightarrow 0$. (Note: The overall rate $R_{\text{PM}} \times N/M$ of SRC can remain finite even though $R_{\text{PM}} \rightarrow 0$ if $N/M \rightarrow \infty$.) The cost of position demodulation grows with B . For affordable complexity (e.g. B up to thousands), there is a considerable gap between SRC performance and capacity. Detailed discussions on this issue can be found in [19].

The compressed coding scheme introduced in [19] is equivalent to the LRMS in (1). Arbitrary FEC coding is assumed in [19]. The achievable rate of compressed coding using AMP approaches Gaussian capacity in the limiting case when the underlying FEC rate $R_C \rightarrow 0$ [19], [20], which is similar to SRC [17], [18]. It is shown that compressed coding using a properly designed FEC code can outperform SRC under practical complexity constraint (i.e., a limited block length of position modulation in SRC) [19].

Recall that the overall rate of the LRMS in (1) is NR_C/M . It was proved in [17]–[19] that SRC and compressed coding are capacity approaching when $R_{\text{PM}} \rightarrow 0$ or $R_C \rightarrow 0$. In these cases, if the overall rate is finite, we need $N/M \rightarrow \infty$ in (1), which may incur excessively high receiver cost. In this paper, we remove this limitation. We will show that AMP is information theoretically optimal for any R_C under a matching condition.

D. Notations

Boldface lowercase letters represent vectors and boldface uppercase symbols denote matrices. $I(\mathbf{x}; \mathbf{y})$ for the mutual information between \mathbf{x} and \mathbf{y} , \mathbf{I} for the identity matrix with a proper size, $|\mathcal{S}|$ for the cardinality of set \mathcal{S} , \mathbf{a}^H for the conjugate transpose of \mathbf{a} , $\|\mathbf{a}\|$ for the ℓ_2 -norm of the vector \mathbf{a} , $\det(\mathbf{A})$ for the determinant of \mathbf{A} , $\text{Tr}(\mathbf{A})$ for the trace of \mathbf{A} , A_{ij} for the i th-row and j th-column element of \mathbf{A} , $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for the circularly-symmetric Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, $\mathbb{E}\{\cdot\}$ for the expectation operation over all random variables involved in the brackets, except when otherwise specified. $\mathbb{E}\{a|b\}$ for the expectation of a conditional on b , $\text{var}\{a\}$ for $\mathbb{E}\{(a - \mathbb{E}\{a\})^2\}$, $\text{mmse}\{a|b\}$ for $\mathbb{E}\{(a - \mathbb{E}\{a|b\})^2 | b\}$, $\langle \mathbf{x} \rangle = \sum_{i=1}^N x_i/N$, and $\eta'(r) = \frac{\partial}{\partial r} \eta(r)$.

Capacity is defined by default as the maximum mutual information over all possible choices of input distribution. A constrained capacity is defined as the mutual information under a fixed input distribution $x \sim P_X(x)$. For the systems considered in this paper, capacity is always achieved by Gaussian signaling. Hence sometimes we will call it ‘‘Gaussian capacity’’. For simplicity, we will call the constrained capacity given $P_X(x)$ ‘‘capacity’’ if it is clear by context.

E. Paper Outline

This paper is organized as follows. Section II gives the area property, the constrained capacity of an LRMS and the AMP algorithm. Section III proves the capacity optimality of AMP under Gaussian assumption. Numerical results are shown in Section IV.

II. PRELIMINARIES

In this section, we briefly outline some existing results that will be used in this paper, such as the area property for single-input-single-output additive white Gaussian noise (SISO-AWGN) channel, the capacity of a large random matrix system, and the AMP algorithm.

A. Area Property for SISO-AWGN Channel

A SISO-AWGN channel is defined as

$$y = \sqrt{\rho}x + z, \quad (2)$$

where $x \sim P_X(x)$, $z \sim \mathcal{CN}(0, 1)$, and ρ denotes the signal-to-noise-ratio (SNR). The MMSE of (2) is denoted as

$$\omega(\rho) \equiv \text{mmse}(x|\sqrt{\rho}x + z, x \sim P_X(x)). \quad (3)$$

The following lemma, proved in [35], establishes the connection between MMSE and the capacity given $P_X(x)$ for a SISO-AWGN channel.

Lemma 1 (Scalar I-MMSE): Let $\text{SNR} = \rho^*$. The capacity of a SISO-AWGN channel equals to the area under $\omega(\rho)$ from $\rho = 0$ to $\rho = \rho^*$, i.e.,

$$C_{\text{SISO}}(\rho^*) = I(x; \sqrt{\rho^*}x + z) = \int_0^{\rho^*} \omega(\rho) d\rho. \quad (4)$$

Fig. 1 gives a graphical illustration of Lemma 1. The following are some instances of the MMSE function $\omega(\rho)$.

- *Gaussian Signaling:* For $x \sim \mathcal{CN}(0, 1)$, $\omega(\rho)$ is given by

$$\omega_{\text{Gau}}(\rho) \equiv 1/(\rho + 1). \quad (5a)$$

The channel capacity with Gaussian signaling is given by

$$C_{\text{SISO}}^{\text{Gau}} = \int_0^{\rho^*} \omega_{\text{Gau}}(\rho) d\rho = \int_0^{\rho^*} \frac{1}{1+\rho} d\rho = \log(1+\rho^*). \quad (5b)$$

- *Discrete Signaling [39]:* For an arbitrary discrete constellation $\mathcal{S} = \{s_1, \dots, s_{|\mathcal{S}|}\}$ with equal probability $1/|\mathcal{S}|$, $\omega(\rho)$ is given by

$$\omega_{\mathcal{S}}(\rho) \equiv 1 - \frac{1}{\pi} \int \frac{\left| \sum_{l=1}^{|\mathcal{S}|} s_l e^{-|y - \sqrt{\rho} s_l|^2} \right|^2}{|\mathcal{S}| \sum_{l=1}^{|\mathcal{S}|} e^{-|y - \sqrt{\rho} s_l|^2}} dy. \quad (6a)$$

The constrained capacity for $x \in \mathcal{S}$ is given by $\int_0^{\rho^*} \omega_{\mathcal{S}}(\rho) d\rho$. In particular, if x is un-coded, the information rate is

$$R_{\mathcal{S}} = \int_0^{\rho^*} \omega_{\mathcal{S}}(\rho) d\rho = \log |\mathcal{S}|. \quad (6b)$$

- *QPSK Signaling [35]:* As a special case of (6a), for quadrature phase-shift keying (QPSK) signaling $x \in \{\frac{1}{\sqrt{2}}(\pm 1 \pm j)\}$, $\omega(\rho)$ is given by

$$\omega_{\text{QPSK}}(\rho) \equiv 1 - \int_{-\infty}^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} \tanh(\rho - \sqrt{\rho}y) dy. \quad (7)$$

- *Code-Rate-MMSE Lemma [36]:* Let the code length be N and code rate $R = K/N$. We treat the code-book $\mathbf{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_{2^K}\}$ as a uniformly distributed N -dimension constellation with 2^K discrete points. When $\text{SNR} \rightarrow \infty$,

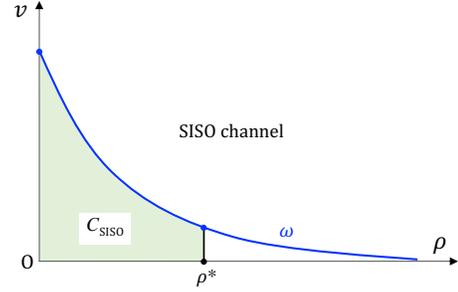


Fig. 1. Graphical illustration of the capacity of a SISO-AWGN channel. ρ^* denotes the channel SNR, and ω is the MMSE of the SISO-AWGN channel.

the capacity per length- N code block approaches to the entropy of \mathbf{C} , i.e., $\log(2^K) = K$. The entropy per dimension is K/N . Hence, we have

$$R_C = \int_0^{\rho^*} \omega_C(\rho) d\rho = K/N, \quad (8)$$

where $\omega_C(\rho) \equiv \frac{1}{N} \text{mmse}(x|\sqrt{\rho}x + z, x \in \mathbf{C})$ and R_C is the rate of \mathbf{C} .

B. LRMS Capacity

Return to the LRMS in (1): $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where $\mathbf{y} \in \mathbb{C}^{M \times 1}$ is a vector of observations, $\mathbf{A} \in \mathbb{C}^{M \times N}$ an IIDG matrix with $A_{ij} \sim \mathcal{CN}(0, 1/M)^1$, $\{x_i \sim P_X(x), \forall i\}$, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ a vector of Gaussian additive noise samples. Fig. 2(a) shows a modulated LRMS. In this paper, we consider a large system with $M, N \rightarrow \infty$ and a fixed $\beta = N/M$. The transmit SNR is defined as $\text{snr} = \mathbb{E}\{\|x_i\|^2\} / \mathbb{E}\{\|n_j\|^2\} = \sigma^{-2}$. We assume that \mathbf{A} is known at the receiver, but unknown at the transmitter². This assumption has been widely used in multiple-input-multiple-output (MIMO) and multi-user MIMO (MU-MIMO) systems [26], [27], [32].

The constrained capacity of an LRMS given $P_X(x)$ was proved in [28], [29], [37].

Lemma 2 (Capacity): Assume that the signal distribution $P_X(x)$ satisfies the single-crossing property, i.e., $\zeta = \beta \text{snr} \omega(1/[\beta(1+\zeta)])$ has exactly one positive fixed point ζ^* . Then, the capacity of the LRMS in (1) is given by

$$C = \beta^{-1} \left[\log(1+\zeta^*) - \zeta^*/(1+\zeta^*) \right] + C_{\text{SISO}}(\text{snr}/(1+\zeta^*)), \quad (9)$$

where $C_{\text{SISO}}(\cdot)$ is defined in (4).

The capacity in (9) is equivalent to that desired in [28]. For the details, see APPENDIX A.

¹In fact, it can be easily extended to a more general case $A_{ij} \sim \mathcal{CN}(0, \sigma_a^2/M)$, where σ_a^2 is finite. In this case, we can rewrite the system to $\mathbf{y}' = \sigma_a^{-1} \mathbf{y} = \mathbf{A}' \mathbf{x} + \mathbf{n}' = \sigma_a^{-1} \mathbf{A} \mathbf{x} + \sigma_a^{-1} \mathbf{n}$, where $A'_{ij} \sim \mathcal{CN}(0, 1/M)$ and $\mathbf{n}' \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \sigma_a^{-2} \mathbf{I})$. Then, all the results in this paper are still valid by replacing σ^2 with $\sigma^2 \sigma_a^{-2}$. For example, if $A_{ij} \sim \mathcal{CN}(0, 1/N)$, we replace σ^2 by $\beta \sigma^2$ to make the results of this paper be valid.

²If \mathbf{A} is also known at the transmitter, then the LRMS in (1) can be converted to a set of parallel SISO-AWGN channels using singular value decomposition (SVD). Then, the well-known water filling technique is capacity approaching.

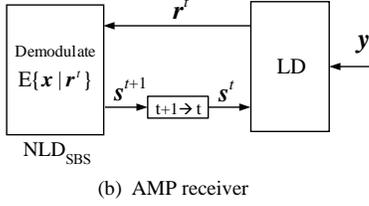
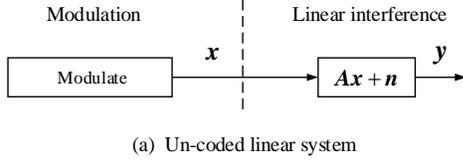


Fig. 2. Un-coded LRMS: transmitter and AMP receiver, where “Demodulate” and LD in (b) correspond to “Modulate” and “ $\mathbf{A}\mathbf{x} + \mathbf{n}$ ” in (a) respectively.

C. Overview of AMP

AMP [4] finds an approximate MMSE solution to the problem in (1) using the following iterative process (initialized with $t = 0$ and $\mathbf{s}^0 = \mathbf{r}_{\text{Onsager}}^0 = \mathbf{0}$):

$$\text{LD} : \mathbf{r}^t = f(\mathbf{s}^t) \equiv \mathbf{s}^t + \mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{s}^t) + \mathbf{r}_{\text{Onsager}}^t, \quad (10a)$$

$$\text{NLD}_{\text{SBS}} : \mathbf{s}^{t+1} = \eta(\mathbf{r}^t) \equiv \mathbf{E}\{\mathbf{x}|\mathbf{r}^t\}, \quad (10b)$$

where $\eta(\mathbf{r}^t)$ is a symbol-by-symbol (SBS) MMSE demodulate function, and $\mathbf{r}_{\text{Onsager}}^t$ is an “Onsager term” defined by $\mathbf{r}_{\text{Onsager}}^t = \beta \langle \eta'(\mathbf{r}^{t-1}) \rangle (\mathbf{r}^{t-1} - \mathbf{s}^{t-1})$ [4]. Fig. 2(b) is a graphical illustration of AMP, where the linear detector (LD) and non-linear detector (NLD) correspond to (10a) and (10b) respectively. We define the errors at the LD and NLD respectively as

$$\mathbf{h}^t \equiv \mathbf{r}^t - \mathbf{x} \quad \text{and} \quad \mathbf{q}^t \equiv \mathbf{s}^t - \mathbf{x}. \quad (11)$$

Let ρ^t be the signal-to-interference-plus-noise-ratio (SINR) for \mathbf{r}^t and v^t the MSE for \mathbf{s}^t :

$$\rho^t \equiv N[\mathbf{E}\{\|\mathbf{h}^t\|^2\}]^{-1}, \quad v^t \equiv \frac{1}{N}\mathbf{E}\{\|\mathbf{q}^t\|^2\}. \quad (12)$$

The following lemma summarizes the findings in [5].

Lemma 3: Let $M, N \rightarrow \infty$ with a fixed $\beta = N/M$. For AMP, \mathbf{h}^t defined in (11) can be modeled by a sequence of IIDG samples independent of \mathbf{x} . The LD and NLD of AMP can be characterized by the following transfer functions [5]

$$\text{LD} : \rho^t = \phi(v^t) = (\beta v^t + \sigma^2)^{-1}, \quad (13a)$$

$$\text{NLD}_{\text{SBS}} : v^{t+1} = \omega(\rho^t), \quad (13b)$$

where $\omega(\cdot)$ is the MMSE function given in II-A.

The iterative process in (10) can be written as (see Fig. 3)

$$\mathbf{r}^0 = f(\mathbf{s}^0), \mathbf{s}^1 = \eta(\mathbf{r}^0), \mathbf{r}^1 = f(\mathbf{s}^1), \mathbf{s}^2 = \eta(\mathbf{r}^1), \dots \quad (14a)$$

From (12) and Lemma 3, we can track the SINR and MSE in (14a) as

$$\rho^0 = \phi(v^0), v^1 = \omega(\rho^0), \rho^1 = \phi(v^1), v^2 = \omega(\rho^1) \dots \quad (14b)$$

Assumption 1: There is exactly one fixed point for $\omega(\rho) = \phi^{-1}(\rho)$ for $\rho > 0$, where $\phi^{-1}(\cdot)$ is the inverse of $\phi(\cdot)$.

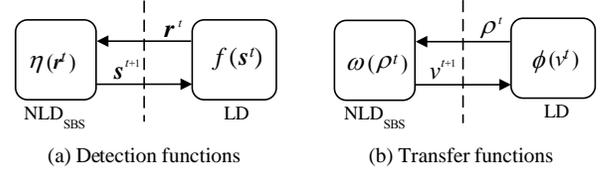


Fig. 3. Detection functions (a) and transfer functions (b).

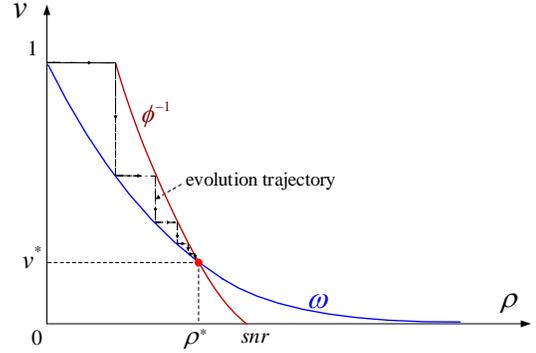


Fig. 4. Graphical illustration of an AMP, where ϕ^{-1} is the inverse of ϕ given in (13a) and ω is defined in (13b). The iterative process of AMP is illustrated by the evolution trajectory, and the fixed point (ρ^*, v^*) gives the MMSE. From (13a), we have $\phi(0) = snr$. (Figure parameters: $\omega(\rho) = \omega_{\text{QPSK}}(\rho)$, $\beta = 0.65$, $snr = 5$ dB, $\rho^* = 2.25$, $v^* = 0.20$.)

Fig. 4 provides a graphical illustration of Assumption 1. The evolution trajectory of AMP converges to a unique fixed point (ρ^*, v^*) with $v^* = \omega(\rho^*)$. The following theorem was first established in [30] via replica method, and then was rigorously proved in [28], [29].

Lemma 4 (MMSE Optimality): Let $\hat{\mathbf{x}}_{\text{MMSE}} = \mathbf{E}\{\mathbf{x}|\mathbf{y}, x_i \sim P_X(x), \forall i\}$ be the conditional mean of \mathbf{x} given \mathbf{y} and $\{x_i \sim P_X(x), \forall i\}$ and suppose that Assumption 1 holds. Then

$$v^* = \frac{1}{N}\mathbf{E}\{\|\mathbf{x} - \hat{\mathbf{x}}_{\text{MMSE}}\|^2\}, \quad (15)$$

i.e., AMP converges to the MMSE of the un-coded LRMS.

III. CAPACITY OPTIMALITY OF AMP UNDER GAUSSIAN ASSUMPTION

In this section, we investigate the achievable rate of the AMP receiver with FEC decoding.

A. Coded System Model and AMP

Fig. 5(a) shows an LRMS involving FEC coding. We write $\mathbf{x} \in \mathcal{C}$ for coded \mathbf{x} . The other conditions are the same as that in Fig. 2. We focus on the AMP receiver in Fig. 5(b) for a coded LRMS.

AMP: Initialized with $t=0$ and $\mathbf{s}^0 = \mathbf{r}_{\text{Onsager}}^0 = \mathbf{0}$,

$$\text{LD} : \mathbf{r}^t = f(\mathbf{s}^t) \equiv \mathbf{s}^t + \mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{s}^t) + \mathbf{r}_{\text{Onsager}}^t, \quad (16a)$$

$$\text{NLD}_{\text{DEC}} : \mathbf{s}^{t+1} = \eta_{\mathcal{C}}(\mathbf{r}^t) \equiv \mathbf{E}\{\mathbf{x}|\mathbf{r}^t, \mathbf{x} \in \mathcal{C}\}. \quad (16b)$$

Comparing (16) and (10), we can see that the symbol-wise NLD in AMP for un-coded \mathbf{x} is replaced by an *a-posteriori* probability (APP) decoder in AMP for coded \mathbf{x} .

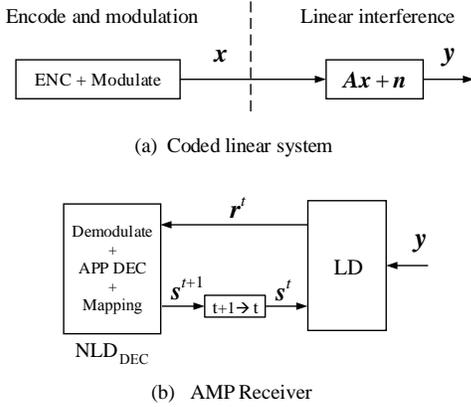


Fig. 5. Coded linear system: Transmitter and AMP receiver. “APP DEC” (*a-posteriori* probability decoding), “Demodulate” and LD in (b) correspond to “ENC” (encode), “Modulate” and “ $Ax + n$ ” in (a) respectively.

Lemma 3 gives the IIDG property for AMP for un-coded \mathbf{x} . The discussions in this paper are based on the following assumption for coded \mathbf{x} .

Assumption 2: Lemma 3 still holds for AMP for coded \mathbf{x} , i.e., \mathbf{h}^t is IIDG and independent of \mathbf{x} , and LD and NLD of AMP can be characterized by

$$\text{LD} : \rho = \phi(v), \quad (17a)$$

$$\text{NLD}_{\text{DEC}} : v = \omega_C(\rho) \equiv \frac{1}{N} \text{mmse}(\mathbf{x} | \sqrt{\rho}\mathbf{x} + \mathbf{z}, \mathbf{x} \in \mathcal{C}). \quad (17b)$$

The $\phi(v)$ in (17a) is the same as that in (13a), and $\omega_C(\rho)$ depends on the code constraint.

B. Area Property of LRMS and Capacity Optimality of AMP

In the un-coded case in (10), as shown above, AMP converges to a fixed (ρ^*, v^*) in Fig. 4. Detection is not error free as $v^* > 0$.

In the coded case, it is possible to achieve error-free detection using a properly designed $\omega_C(\rho)$. As illustrated in Fig. 6, the key is to create a detection tunnel that converges to $v = 0$, implying zero error rate. There should be no fixed point between $\omega_C(\rho)$ and $\phi^{-1}(\rho)$, since otherwise the tunnel will be closed at $v > 0$. This requires that

$$\omega_C(\rho) < \phi^{-1}(\rho), \quad \text{for } 0 \leq \rho \leq snr. \quad (18a)$$

Also, by definition, the MMSE NLD_{DEC} in (17b) should achieve an MSE lower than that of a symbol-by-symbol detector, i.e.,

$$\omega_C(\rho) < \omega_S(\rho), \quad \text{for } \rho \geq 0. \quad (18b)$$

Combining (18a) and (18b), we obtain a necessary and sufficient condition for AMP to achieve error-free detection:

$$\omega_C(\rho) < \omega_C^*(\rho), \quad \text{for } 0 \leq \rho \leq snr, \quad (19a)$$

where

$$\omega_C^*(\rho) = \min \{ \omega_S(\rho), \phi^{-1}(\rho) \}, \quad \text{for } 0 \leq \rho \leq snr. \quad (19b)$$

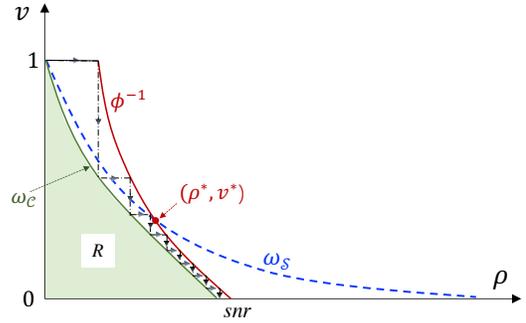


Fig. 6. Graphical illustration of AMP, where ω_S is a demodulation function (un-coded case) and ω_C is a transfer function of a decoder (coded case). The iterative process of AMP is illustrated by the evolution trajectory between ϕ^{-1} and ω_C .

Proposition 1: Suppose that Assumption 1 holds. Then the constrained capacity of an LRMS with the given \mathcal{S} is

$$C = A_{\omega_C^*}, \quad (20a)$$

where $A_{\omega_C^*}$ is the area covered by ω_C^* , i.e.,

$$A_{\omega_C^*} \equiv \int_0^{snr} \omega_C^*(\rho) d\rho \quad (20b)$$

$$= \beta^{-1} [\rho^*/snr - \log(\rho^*/snr) - 1] + \int_0^{\rho^*} \omega(\rho) d\rho. \quad (20c)$$

Proof: See APPENDIX A. ■

Combining (8), (19) and (20), we obtain the capacity optimality of AMP below.

Theorem 1 (Capacity Optimality): Assume that Assumptions 1 and 2 hold and AMP converges to $v = 0$. Then,

$$R_C \rightarrow C, \quad (21)$$

if $\omega_C(\rho) \rightarrow \omega_C^*(\rho)$ in $[0, snr]$.

Fig. 7 gives a graphical illustration of Theorem 1. The following are some notes on Theorem 1.

- Theorem 1 is based on a matching condition:

$$\omega_C(\rho) \rightarrow \omega_C^*(\rho). \quad (22)$$

A proof for the existence of a code achieving (22) can be found in Appendix C-B for Gaussian signaling. For other signaling, the existence of such a code is a conjecture only. We will discuss techniques to approximately achieve (22) for QPSK modulations in IV-A.

- The situations for multi-ary modulations are more complicated. Various techniques have been developed to match the extrinsic information transfer (EXIT) functions of two local processors in conventional Turbo receivers [40]–[42]. These methods can be borrowed to achieve (22) approximately. Detailed discussions on the multi-ary systems are beyond the scope of this paper.
- Interestingly, the discussions above provide an alternative proof for the constraint capacity of an LRMS. The key is to prove $A_{\omega_C^*} = C$ without prompting the result in [28], [29], [37]. This is indeed possible using the properties of AMP directly. The details can be found in Appendix B.

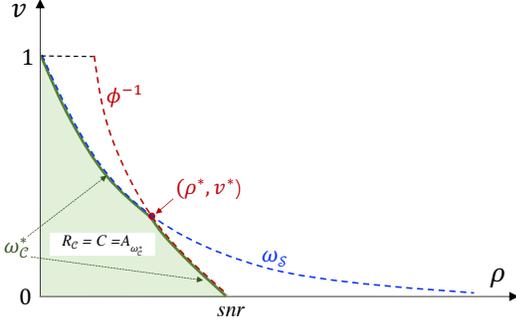


Fig. 7. Graphical illustration of the capacity, the maximum achievable rate of AMP and the optimal transfer function of decoder. The maximum achievable rate of AMP equals to the capacity, which is the area covered by ω_C^* .

Area Properties: Based on the above discussions, we can obtain some interesting area properties as illustrated in Fig. 8.

- (i) Area A_{AGPO} gives the entropy of the constellation, e.g. $\log |\mathcal{S}|$, as discussed in II-A, which is maximum achievable rate in the noiseless case.
- (ii) Area A_{AGQO} gives the capacity of SISO-AWGN channel as discussed in II-A (see Fig. 1).
- (iii) Area A_{GPQ} gives the rate loss after SISO coding to combat the channel noise. As SNR goes to infinity, point Q moves right to infinity (e.g. point P) and the SISO code rate approaches the maximum $\log |\mathcal{S}|$.
- (iv) Area A_{AFQO} gives the capacity C of an LRMS and also the achievable rate of AMP.
- (v) Area A_{AFHO} gives the achievable rate of a receiver with an AMP detector cascaded by a decoder [43], [44]. There is no iteration between the two. In this case, AMP achieves the MMSE by treating the codeword as an IID sequence. However, the overall algorithm is not capacity optimal (see III-D).
- (vi) Area A_{FQO} gives the rate loss for the cascading receiver in (v).
- (vii) Area A_{AEF} gives the shaping gain of Gaussian signaling. Curve AEP denotes the un-coded Gaussian NLD for Gaussian signaling.
- (viii) Area A_{FGQ} gives the capacity gap of LRMS and parallel SISO channels. When $\beta \rightarrow 0$, $\lim_{\beta \rightarrow 0} \mathbf{A}^H \mathbf{A} = \mathbf{I}$, indicating that the capacity of an LRMS will converge to that of a set of parallel SISO channels in the limiting case. In this case, the cross-symbol interference disappears, so $B \rightarrow S$ and areas A_{FGQ} and A_{FQO} become negligible. Then, the separate detection and decoding strategy becomes optimal.

C. Gaussian Signaling

We now study a special case of III-B when \mathbf{x} is Gaussian. We show that in this case both Assumptions 1 and 2 asymptotically hold. In addition, the results in III-B have simpler derivations as well as closed-form expressions.

For Gaussian signaling, $\omega(\rho) = \omega_{\text{Gau}}(\rho) = 1/(1 + \rho)$ [44]. Thus we can rewrite (19b) as

$$\omega_{C-\text{Gau}}^*(\rho) = \min\{1/(\rho + 1), \phi^{-1}(\rho)\}, \quad 0 \leq \rho \leq snr. \quad (23)$$

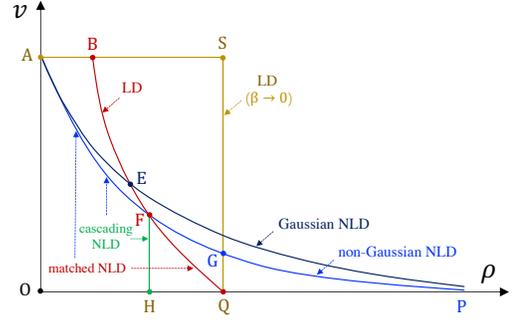


Fig. 8. Interpretations of SINR-variance transfer charts and areas of AMP. Curve BQ represents LD ϕ^{-1} . When $\beta \rightarrow 0$, it moves to the “interference-free” SQ. AEP: un-coded Gaussian NLD ω_{Gau} ; AFP: un-coded non-Gaussian NLD ω_S ; AFQ: matched decoder ω_C^* ; AFH: cascading decoder.

Then we have the following results for an LRMS with Gaussian signaling.

- *Unique Fixed Point:* Assumption 1 holds. That is, equation $\omega_{\text{Gau}}(\rho) = \phi^{-1}(\rho)$ has a unique positive solution:

$$\rho_{\text{Gau}}^* = \frac{(1-\beta)snr - 1 + \sqrt{[(1-\beta)snr - 1]^2 + 4snr}}{2}. \quad (24)$$

- *Area Property:* Let the area covered by $\omega_{C-\text{Gau}}^*$ be

$$A_{\omega_{C-\text{Gau}}^*} \equiv \int_0^{snr} \omega_{C-\text{Gau}}^*(\rho) d\rho \quad (25a)$$

$$= \beta^{-1} \log(1 + \beta snr v_{\text{Gau}}^*) - \log(v_{\text{Gau}}^*) + v_{\text{Gau}}^* - 1, \quad (25b)$$

where $v_{\text{Gau}}^* = \phi^{-1}(\rho_{\text{Gau}}^*)$. Then area $A_{\omega_{C-\text{Gau}}^*}$ equals to the Gaussian capacity of an LRMS, i.e.,

$$A_{\omega_{C-\text{Gau}}^*} = C_{\text{Gau}}. \quad (26)$$

Proof: See APPENDIX C-A. ■

- *Code Existence:* For Gaussian signaling, there exists an n -layer superposition coded modulation (SCM) code with rate R_n and transfer function $\{\omega_{C_n}(\rho) < \omega_{C-\text{Gau}}^*(\rho), \forall \rho \geq 0\}$, and as $n \rightarrow \infty$,

$$R_n \rightarrow A_{\omega_{C-\text{Gau}}^*}. \quad (27)$$

Proof: See Appendix C-B. ■

That is, for Gaussian signaling, there exists an SCM code that asymptotically matches with $\omega_{C-\text{Gau}}^*$ and its rate $\rightarrow A_{\omega_{C-\text{Gau}}^*}$.

- *Capacity Optimality:* Assume that Assumption 2 holds, AMP achieves the Gaussian capacity when $\omega_C = \omega_{C-\text{Gau}}^*$:

$$R_C = C_{\text{Gau}}. \quad (28)$$

D. Comparisons with Alternative Algorithms

1) *Comparison with Turbo-LMMSE:* It is proved in [31], [32] that Turbo-LMMSE is capacity achieving for Gaussian signaling. In the following, we show that Turbo-LMMSE is sub-optimal for non-Gaussian signaling.

The main difference between AMP and Turbo-LMMSE is as follows. To avoid the correlation problem in the iterative process, Turbo-LMMSE uses extrinsic local processors (e.g. an extrinsic LD and an extrinsic decoder), while AMP uses an “Onsager”-term.

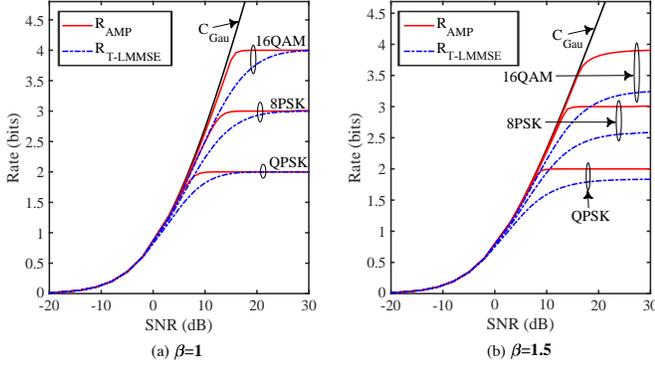


Fig. 9. Comparison between the capacity and the achievable rates of AMP and Turbo-LMMSE of an LRMS with $\beta = N/M = \{1, 1.5\}$, where C_{Gau} denotes the Gaussian capacity and also the achievable rates of AMP and Turbo-LMMSE with Gaussian signaling, R_{AMP} and $R_{\text{T-LMMSE}}$ respectively denote the achievable rates of AMP and Turbo-LMMSE with QPSK, 16QAM and 8PSK modulations.

Assume that the transfer functions of the detector and the decoder in Turbo-LMMSE are matched. The achievable rate of Turbo-LMMSE is given in [31]

$$R_{\text{LMMSE}} = \log |S| - \int_0^{+\infty} \omega_S(\rho + \phi(\omega_S(\rho))) d\rho. \quad (29)$$

Fig. 9 shows the capacity and the achievable rates of AMP and Turbo-LMMSE. The capacity for Gaussian signaling is achieved by both AMP and Turbo-LMMSE. For QPSK, 8PSK and 16QAM modulation, the achievable rate of AMP equals to capacity when Assumption 1 holds, while Turbo-LMMSE always has rate loss. Similar results can be obtained for other non-Gaussian signaling. In addition, the gap between AMP and Turbo-LMMSE increases with β . This gap $\rightarrow 0$ when $\beta \rightarrow 0$. The reason why Turbo-LMMSE has performance loss is that extrinsic update leads to performance loss for non-Gaussian signal processing, which was first pointed out in [27]. For more details, please refer to [27].

2) *Comparison with Cascading AMP and Decoding:* We define a cascading AMP and decoding (AMP-DEC) scheme [43], [44] as follows. We run AMP until it converges. The result is used by decoder. There is no iteration between AMP and the decoder. The area A_{AFHO} in Fig. 8 shows the achievable rate of AMP-DEC, i.e.,

$$R_{\text{AMP-DEC}} = A_{\text{AFHO}} = \int_0^{\rho^*} \omega_S(\rho) d\rho. \quad (30)$$

For Gaussian signaling, $\omega_{\text{Gau}}(\rho) = 1/1 + \rho$. Hence,

$$R_{\text{AMP-DEC}} = \log(1 + \rho_{\text{Gau}}^*), \quad (31)$$

where $\rho_{\text{Gau}}^* = 0.5 \left[(1-\beta)snr - 1 + \sqrt{[(1-\beta)snr - 1]^2 + 4snr} \right]$ (see (24)). If $\beta > 1$, when $snr \rightarrow \infty$, we have

$$\rho_{\text{Gau}}^* \rightarrow (\beta - 1)^{-1}, \quad (32)$$

and

$$R_{\text{AMP-DEC}} \rightarrow -\log(1 - \beta^{-1}). \quad (33)$$

That is, the achievable rate of AMP-DEC converges to a finite value, and it goes to zero as $\beta \rightarrow \infty$. This is very different

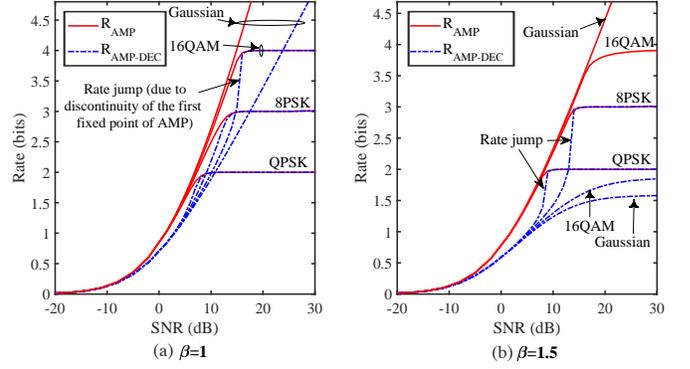


Fig. 10. Comparison between the capacity, the achievable rates of AMP, and separate optimal MMSE detection and ideal SISO decoding in [43], [44] with $\beta = N/M = \{1, 1.5\}$, where C_{Gau} denotes the Gaussian capacity and also the achievable rates of AMP with Gaussian signaling, R_{AMP} and $R_{\text{AMP-DEC}}$ respectively denote the achievable rates of AMP and “cascading AMP and decoding” scheme with QPSK, 16QAM and 8PSK modulations.

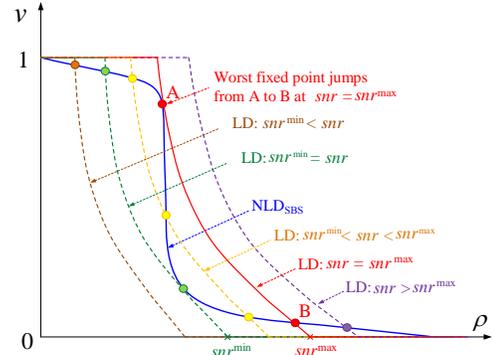


Fig. 11. Multiple fixed points illustration of AMP with QPSK modulation, $\beta = N/M = 2$, $snr^{\min} = 9.05$ dB, $snr^{\max} = 15.77$ dB. Assumption 1 holds when $snr < snr^{\min}$ or $snr > snr^{\max}$; Three fixed points for $snr^{\min} < snr < snr^{\max}$. Besides, the first fixed point jumps from Point A to Point B at $snr = snr^{\max}$. “semilogy”-plot is used to see the cross points clearly.

from the Gaussian system capacity that $C \rightarrow \infty$ as $snr \rightarrow \infty$.

Fig. 10 compares AMP and AMP-DEC. For QPSK, 8PSK and 16QAM modulations, the achievable rate of AMP-DEC is lower than that of AMP. This gap increases with β , but is negligible if β is small (e.g. $\beta < 0.5$ based on our experimental findings). Furthermore, different from the rate of AMP that always increases with the size of constellation, the rate of AMP-DEC decreases with the increasing of the constellation size for large β .

As shown Fig. 10, the achievable rate of AMP-DEC jumps at certain snr values. This happens when the number of fixed points changes. Fig. 11 illustrates this phenomenon for $\beta = 2$. For $snr = snr^{\max}$, the number of fixed points is 3 at the left vicinity and 1 at the right vicinity. While increasing snr , the worst fixed point jumps from Point A to Point B at $snr = snr^{\max}$, resulting in rate jump of AMP-DEC in Fig. 10.

3) *Comparison with AMP with Internal Iteration:* Fig. 12 illustrates an alternative for AMP. It involves the iteration of a decoder (DEC) module and an AMP module. In each global iteration, there are multiple internal iteration within the AMP module. The scheme in Fig. 12 may have the advantage of low cost if the overall complexity is dominated by that of DEC.

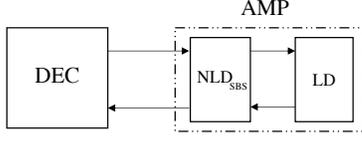


Fig. 12. AMP with internal iteration, in which “DEC” denotes decoder, and AMP is discussed in II-C with SBS-NLD being the demodulation for the constellation constraint S .

However, if the decoding complexity (e.g. sum-product LDPC decoding) is lower than that of AMP, the scheme in Fig. 12 may have higher complexity due to the internal iteration. Since there is no closed-form transfer function for the AMP module, it makes the achievable rate analysis and optimization more difficult. We conjecture that the scheme in Fig. 12 has the same overall performance of AMP without internal iteration. However, we do not have a proof of this conjecture.

IV. LDPC CODE DESIGN AND SIMULATION RESULTS

This section discusses matching techniques using optimized LDPC codes and QPSK modulation. Simulation results will be provided.

A. LDPC Code Optimization for AMP

According to the code design principle in Section III, the object of code optimization is to design an code with an *a-posterior* variance transfer function $v = \omega_C(\rho)$ satisfying the following matching condition:

$$v = \omega_C(\rho) = \omega_C^*(\rho), 0 \leq \rho \leq \text{snr}, \quad (34)$$

with a coding rate equal to the capacity.

The extrinsic information transfer (EXIT) chart matching techniques [45]–[47] can be used for this purpose by choosing optimized degree distributions. The discussions below follow [47, Appendix 5G] to design irregular LDPC codes. The difference is that mutual information is used instead of log likelihood ratio (LLR) in tracking the evolution process.

According to [46], the decoder characteristic for an LDPC code can be computed as

$$I_{E,V} = \sum_{i=1}^{d_{v,\max}} \lambda_i \cdot J \left(\sqrt{(i-1) [J^{-1}(I_{E,C})]^2 + 4\rho} \right), \quad (35a)$$

$$I_{E,C} = 1 - \sum_{j=1}^{d_{c,\max}} \eta_j \cdot J \left(\sqrt{j-1} \cdot J^{-1}(1 - I_{E,V}) \right), \quad (35b)$$

where

- $I_{E,V}$ (resp., $I_{E,C}$) is the extrinsic information from variable node (resp., check node) to check node (resp., variable node),
- $d_{v,\max}$ (resp., $d_{c,\max}$) is the maximum variable node (resp., check node) degree,
- λ_i (resp., η_i) is the fraction of edges in the bipartite graph of the LDPC code connected to variable nodes (resp., check node) with degree i ,

- the function $J(\cdot)$ is

$$J(\sigma_{ch}) = 1 - \int_{-\infty}^{\infty} e^{-\frac{(y - \sigma_{ch}^2/2)^2}{2\sigma_{ch}^2}} \frac{1}{\sqrt{2\pi\sigma_{ch}^2}} \log_2(1 + e^{-y}) dy, \quad (36)$$

and $J^{-1}(\cdot)$ is the inverse of $J(\cdot)$,

- ρ is the decoder input SNR.

Substituting (35b) into (35a), we can characterize the LDPC code by one single variable $I_{E,V}$ using

$$I_{E,V} = \sum_{i=1}^{d_{v,\max}} \lambda_i J \left(\sqrt{(i-1) [J^{-1}(I_{E,C}(I_{E,V}))]^2 + 4\rho} \right). \quad (37)$$

To satisfy (34), the converged extrinsic information $I_{E,V}$ of the LDPC decoder, denoted as $I_{E,V,\text{fin}}(\rho)$, should lead to an output *a posterior* variance $v = \omega_C(\rho)$, i.e., given ρ , $I_{E,V,\text{fin}}(\rho)$ should satisfy the following equation.

$$\sum_{i=1}^{d_{v,\max}} \Lambda_i \omega_{\text{QPSK}} \left(\frac{i [J^{-1}(I_{E,C,\text{fin}})]^2 + 4\rho}{4} \right) = \omega_C(\rho), \quad (38)$$

where

- Λ_i is the fraction of variable node of degree i and is computed as

$$\Lambda_i = \lambda_i / i \left/ \sum_{i=1}^{d_{v,\max}} \lambda_i / i \right., \quad (39a)$$

- $\omega_{\text{QPSK}}(\cdot)$ is the MMSE function of QPSK demodulation given in (7),
- $I_{E,C,\text{fin}}$ is the converged check node to variable node message and is computed as

$$I_{E,C,\text{fin}} = 1 - \sum_{j=1}^{d_{c,\max}} \eta_j J \left(\sqrt{j-1} J^{-1}(1 - I_{E,V,\text{fin}}) \right). \quad (39b)$$

The optimization problem can be formulated to maximize the code rate under the constraint (38):

$$\max_{\{\lambda_i\}} \sum_{i=1}^{d_{v,\max}} \lambda_i / i \quad \text{s.t.} \quad \sum_{i=1}^{d_{v,\max}} \lambda_i = 1, \quad \{\lambda_i\} \in \Xi. \quad (40)$$

where Ξ is defined as the constraint:

$$\sum_{i=1}^{d_{v,\max}} \lambda_i J \left(\sqrt{(i-1) [J^{-1}(I_{E,C}(I_{E,V}))]^2 + 4\rho} \right) > I_{E,V}, \quad (41a)$$

for $\forall 0 < \rho < \infty$ and $I_{E,V,\text{ini}}(\rho) \leq I_{E,V} \leq I_{E,V,\text{fin}}(\rho)$, and $I_{E,V,\text{ini}}(\rho)$ is the initial extrinsic information given by the channel:

$$I_{E,V,\text{ini}}(\rho) = \sum_{i=1}^{d_{v,\max}} \lambda_i J \left(\sqrt{(i-1) [J^{-1}(0)]^2 + 4\rho} \right) = J(2\sqrt{\rho}). \quad (41b)$$

The optimization problem in (40) is non-convex. However, given $\{\eta_i, i = 1, 2, \dots, d_{c,\max}\}$, the problem in (40) can be solved using a standard linear programming technique. In this paper, an iterative way to optimize $\{\lambda_i, i = 1, 2, \dots, d_{v,\max}\}$ with $\{\eta_i, i = 1, 2, \dots, d_{c,\max}\}$ fixed is used (see Algorithm 1). In Algorithm 1, degree distributions are denoted as

$$\lambda(x) = \sum_{i=1}^{d_{v,\max}} \lambda_i x^{i-1} \quad \text{and} \quad \eta(x) = \sum_{i=1}^{d_{c,\max}} \eta_i x^{i-1}. \quad (42)$$

We observed that a maximum trial $T = 5$ and a threshold $\epsilon = 10^{-3}$ are good choice for Algorithm 1. Also, we force the degree-1 fraction $\lambda_1 = 0$. For results given in Table I, Algorithm 1 is repeated by manually tuning check edge distribution $\eta(x)$ and maximum variable degree $d_{v,\max}$ until a matching code is found. For the LDPC code optimization for Turbo-LMMSE, we only need to change (38) into to

$$\sum_{i=1}^{d_{v,\max}} \Lambda_i \cdot \omega_{\text{QPSK}} \left(\frac{i \cdot [J^{-1}(I_{E,C,\text{fin}})]^2}{4} \right) = \omega_C^{\text{ext}}(\rho), \quad (43)$$

where $\omega_C^{\text{ext}}(\rho)$ is matched to the Turbo-LMMSE LD transfer function.

With the optimized degree distributions $\eta(x)$ and $\lambda^{(T)}(x)$, a parity-check matrix for simulation is generated as follows:

- Set a code length to 10^5 .
- Randomly generate a parity-check matrix of columns equal to the code length according to the optimized degree distributions $\eta(x)$ and $\lambda^{(T)}(x)$.
- Remove cycle-4 loops in the generated parity-check matrix by removing one edge in every founded cycle-4 loop in the parity-check matrix.

Removing cycle-4 loops using the above method will slightly change the degree distributions. However, for a code length of 10^5 , this change has negligible effect on overall performance.

The high error floors in Fig. 13 are due to the fixed points of $\phi^{-1}(\cdot)$ and $\omega_C(\rho)$. More research effort is still required to improve this problem.

Related Works: Recently, LDPC codes are optimized to support much higher sum spectral efficiency and user loads for linear systems [47]–[49]. Based on the EXIT analysis [45], [46], a LDPC code is constructed to obtain a near capacity performance [50], [51]. To support massive users, an irregular

Algorithm 1 Algorithm for LDPC Code Optimization

Input: Target decoder transfer $v = \omega_C(\rho)$, check edge distribution $\eta(x)$, maximum trial T , Threshold ϵ and maximum variable degree $d_{v,\max}$.

Output: The optimized variable edge distribution $\lambda^{(T)}(x)$.

- 1: Initialize $\lambda^{(0)}(x) = x$.
 - 2: **for** $t = 1$ to T **do**
 - 3: Solve (40) by linear programming to obtain $\lambda^{(t)}(x)$, where $I_{E,V,\text{fin}}(\rho)$ in (40) is obtained by solving (38) using $\lambda^{(t-1)}(x)$.
 - 4: **if** $1 - \frac{\sum_{i=1}^{d_{v,\max}} \lambda_i^{(t)} \lambda_i^{(t-1)}}{\sqrt{\left(\sum_{i=1}^{d_{v,\max}} (\lambda_i^{(t)})^2 \right) \left(\sum_{i=1}^{d_{v,\max}} (\lambda_i^{(t-1)})^2 \right)}} \leq \epsilon$ **then**
 - 5: $\lambda^{(T)}(x) = \lambda^{(t)}(x)$.
 - 6: **return** $\lambda^{(T)}(x)$.
 - 7: **end if**
 - 8: **end for**
 - 9: **return** $\lambda^{(T)}(x)$.
-

repeat-accumulate (IRA) code is optimized in [52], [53]. More recently, a Turbo-LMMSE receiver with an optimized IRA code approaches the capacity for various of system loads [54]. However, the results in [54] mainly focused on low-rate coding schemes (e.g. $R_C = 0.1$ or 0.2). In these cases, Turbo-LMMSE is near-optimal (see the region of $R_C \leq 0.5$ in Fig. 9). In this paper, we will show that AMP performs much better than Turbo-LMMSE in high transmission rate.

For higher order modulation, the design of a curve-matching code is more complicated. Irregular bit-interleaved coded modulation (Ir-BICM) was developed for high-order modulation [40]–[42]. These methods can be borrowed to design curve-matching codes for high-order modulations. Detailed discussions on the high-order modulations are beyond the scope of this paper. We leave it as our future work.

B. BER Comparison with AWGN Irregular LDPC Code and Regular LDPC Code

Fig. 13 provides the BER simulations for an LRMS, in which x is generated using optimized irregular LDPC codes [47], [48] with code length = 10^5 . The AMP (see Fig. 5) for an optimized LDPC coding (see IV-A) LRMS is denoted as “Opt-Irreg”. The APP decoder is implemented using a standard sum-product decoder. The channel loads are $\beta = \{0.1, 0.5, 1, 2\}$ with $(N, M) = (250, 2500), (250, 500), (500, 500)$ and $(500, 250)$, respectively. The corresponding optimized code parameters are given in Table I, which illustrates that these decoding thresholds are very close (about 0.1 dB~0.2 dB away) to the Shannon limits.

To verify the finite-length performance of the irregular LDPC codes with code rate ≈ 0.5 , we provide the BER performances of the optimized codes. QPSK modulation is used. The rate of each symbol is $R_C \approx 1$ bits/symbol, and the sum rate is $R_{\text{sum}} \approx N$ bits per channel use. The maximum iteration number is 200 ~ 700. Fig. 13 shows that for all β , gaps between the BER curves of the codes at 10^{-5} and the corresponding Shannon limits are within 0.7 ~ 1 dB.

To validate the advantage of matching principle, we provide AMP for a standard regular (3, 6) LDPC code (denoted as “(3, 6)”) [55], and a SISO irregular LDPC code [56] (denoted as “SISO-Irreg”), corresponding to $R_{\text{AMP-DEC}}$ discussed in Section III-D. The parameters of “SISO-Irreg” are $\lambda(x) = 0.170031x + 0.160460x^2 + 0.112837x^5 + 0.047489x^6 + 0.011481x^9 + 0.091537x^{10} + 0.152978x^{25} + 0.036131x^{26} + 0.217056x^{99}$ and $\eta(x) = 0.0625x^9 + 0.9375x^{10}$, whose rate is 0.50004 and decoding threshold is 0.0247 dB away from the binary input AWGN capacity.

As shown in Fig. 13, when the BER curves of three systems are at 10^{-5} , the optimized irregular LDPC codes have 0.8 ~ 2 dB performance gains over the un-optimized regular (3, 6) LDPC code for $\beta = \{0.1, 0.5, 1, 2\}$, and 0.5 ~ 6 dB performance gains over “SISO-Irreg” for $\beta = \{0.5, 1, 2\}$. For small β (e.g. $\beta = 0.1$), the “SISO-Irreg” is good enough, since the interference is negligible in this case (see Fig. 8). These results demonstrate that code optimization provides attractive performance improvement, especially for the large β .

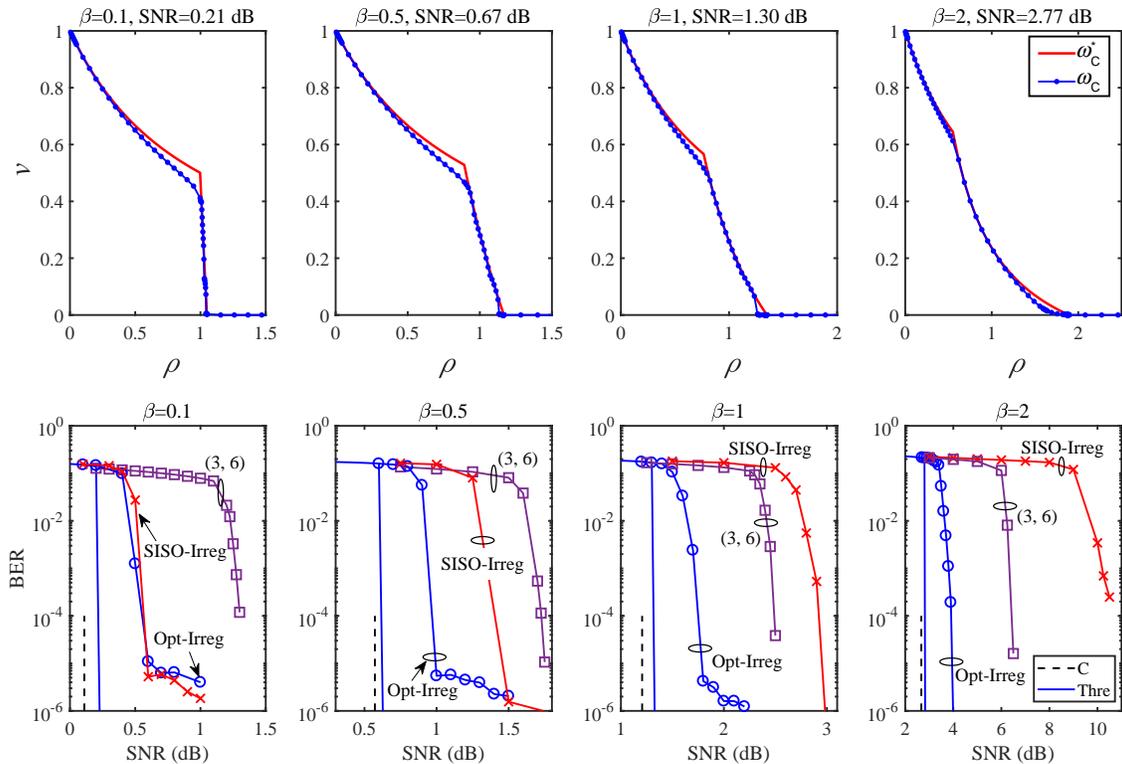
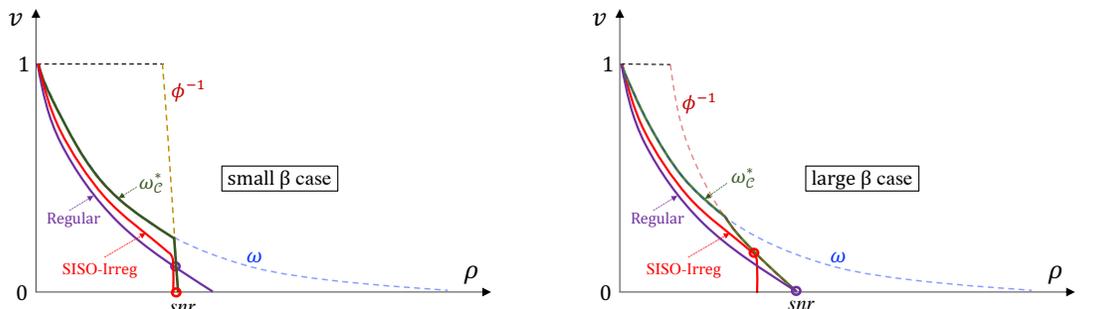


Fig. 13. Transfer function matching and BER performances of AMP, where “ ω_C^* ” denotes the fully matched transfer function of AMP (target), “ ω_C ” the optimized transfer function of decoder of AMP, C the capacity limit, “Thre” the BER threshold, “Opt-Irreg” the BER of AMP-optimized irregular LDPC codes, “SISO-Irreg” the BER with SISO-optimized irregular LDPC codes, “(3, 6)” the BER of AMP with regular (3, 6) LDPC code. Code length = 10^5 , code rate ≈ 0.5 , QPSK modulation, iterations = 200 ~ 700, and $\beta = N/M = \{0.1, 0.5, 1, 2\}$. For more details, refer to Table I.



(a) Small β case: “SISO-Irreg” matches better than “Regular”. Thus, “SISO-Irreg” converges to a lower fixed point.
 (b) Large β case: “Regular” matches better than “SISO-Irreg”. Thus, “Regular” converges to a lower fixed point.

Fig. 14. Graphical illustration of AMP with the regular (3, 6) LDPC code and the SISO-Irreg LDPC code.

Stability of Regular LDPC Code, SISO-Irregular LDPC Code and the Optimized Code:

- When β is small, the curve of “SISO-Irreg” matches the target curve ω_C^* better than the regular LDPC code. Therefore, for $\beta = \{0.1, 0.5\}$, “SISO-Irreg” outperforms the regular LDPC code. (See Fig. 14(a).)
- When β is moderately large, the (3, 6) regular LDPC code matches the target curve ω_C^* better than “SISO-Irreg”. Therefore, for $\beta = \{1, 2\}$, the regular LDPC code outperforms “SISO-Irreg”. (See Fig. 14(b).)
- The optimized code always has the best performance following the matching principle.

Number of Iterations: The number of iterations mainly depends on β . AMP has lower convergence speed when β is large since the AMP LD decreases more slowly. Consequently, it needs more iterations to converge. In our simulations, for $\beta = \{0.1, 0.5, 1, 1.5\}$, 200 iterations is sufficient, while more iterations (e.g. 700) for higher β (e.g. $\beta = 2$).

SE of AMP with LDPC Code: Fig. 15 compares the simulated and predicted BER performances of AMP with optimized irregular LDPC code. As we can see, the SE predictions are accurate when the number of iterations is small (e.g. Itc ≤ 30). The gap increases with the number of iterations. For Itc = 200, the simulated BER is about 0.5 dB away from the SE curve.

TABLE I
OPTIMIZED IRREGULAR LDPC CODES FOR AMP AND TURBO-LMMSE UNDER QPSK MODULATION

Methods	AMP					Turbo-LMMSE
β	0.1	0.5	1	1.5	2	1.5
N	250	250	500	500	500	500
M	2500	500	500	333	250	333
Code length	10^5					
Target Code rate	0.5			0.75	0.5	0.75
Capacity limit	0.110	0.572	1.206	5.384	2.669	7.994
Designed Code rate	0.5000	0.5013	0.5029	0.7370	0.5021	0.7369
R_C	1.0000	1.0026	1.0058	1.4741	1.0042	1.4738
R_{sum}	249.99	250.67	502.90	737.06	502.10	736.91
Iterations	200	200	200	200	700	200
Check edge distribution	$\eta_{10} = 1$	$\eta_9 = 1$	$\eta_8 = 1$	$\eta_8 = 0.5$ $\eta_{20} = 0.5$	$\eta_7 = 1$	$\eta_{12} = 0.8$ $\eta_{80} = 0.2$
Variable edge distribution	$\lambda_2 = 0.1922$ $\lambda_3 = 0.1694$ $\lambda_7 = 0.2201$ $\lambda_8 = 0.0511$ $\lambda_{26} = 0.0759$ $\lambda_{27} = 0.1315$ $\lambda_{80} = 0.0351$ $\lambda_{90} = 0.1247$	$\lambda_2 = 0.2254$ $\lambda_3 = 0.2066$ $\lambda_7 = 0.1101$ $\lambda_8 = 0.1377$ $\lambda_{27} = 0.1294$ $\lambda_{50} = 0.0969$ $\lambda_{60} = 0.0939$	$\lambda_2 = 0.2746$ $\lambda_3 = 0.2622$ $\lambda_{10} = 0.2098$ $\lambda_{40} = 0.1950$ $\lambda_{45} = 0.0223$ $\lambda_{90} = 0.0361$	$\lambda_2 = 0.5546$ $\lambda_3 = 0.1450$ $\lambda_{40} = 0.1750$ $\lambda_{45} = 0.1255$	$\lambda_2 = 0.4655$ $\lambda_3 = 0.1183$ $\lambda_{20} = 0.1020$ $\lambda_{21} = 0.1827$ $\lambda_{140} = 0.1315$	$\lambda_2 = 0.4882$ $\lambda_{19} = 0.3228$ $\lambda_{65} = 0.0002$ $\lambda_{67} = 0.0002$ $\lambda_{100} = 0.1201$ $\lambda_{110} = 0.0685$
snr_{dB}^*	0.3	0.69	1.33	5.62	2.77	8.5

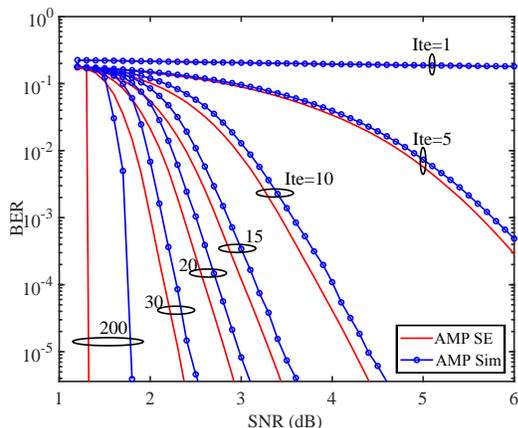


Fig. 15. Comparison of simulation and SE predictions for AMP with optimized irregular LDPC code under QPSK modulation. The curves from right to left correspond to iterations $ite = [1, 5, 10, 15, 20, 30, 200]$. Other parameters are the same as those of the case $\beta = 1$ in Fig. 13 and Table I.

C. BER Comparison with Optimized Turbo-LMMSE

We now compare AMP and Turbo-LMMSE [54]. We consider a 500×333 QPSK LRMS with $\beta = 1.5$. As shown in Fig. 9(b), the SNR limits of AMP and Turbo-LMMSE for the target rate $R_C = 1.48 \approx 1.5$ are 5.38 dB and 7.99 dB respectively. Fig. 16(a) shows the transfer functions of AMP and Turbo-LMMSE after optimization. The parameters of the optimized LDPC codes are listed in Table I. The output of the DEC transfer function is an *a-posteriori* variance, while that of Turbo is an *extrinsic* variance. Fig. 16(b) shows the BER performances of AMP and Turbo-LMMSE (with iterations = 200) using optimized LDPC codes. As we can see, the thresholds of AMP and Turbo-LMMSE are 5.62 dB and

8.50 dB respectively, 0.24 dB and 0.51 dB away from the corresponding achievable rate limits, and 0.6 dB and 1.2 dB away from their respective thresholds. We can see that, AMP has 3.5 dB improvement in BER over Turbo-LMMSE.

Complexity Comparison: The overall complexity of an iterative receiver including an LD and a DEC is $O((\Xi_{LD} + \Xi_{DEC})N_{ite})$, where N_{ite} is the number of iterations, Ξ_{LD} and Ξ_{DEC} are complexities of LD and DEC per iteration respectively. For an LDPC decoder, $\Xi_{DEC} \approx 4\bar{d}_v N_c$, where N_c is the code length and $\bar{d}_v = [\sum_i \lambda_i / i]^{-1}$ denotes the averaged variable-node degree. For AMP, $\Xi_{LD}^{\text{AMP}} = O(MN)$. For Turbo-LMMSE, $\Xi_{LD}^{\text{Turbo}} = O(MN^2)$ since it involves matrix inverse. Thus, AMP has much lower cost than Turbo-LMMSE.

V. CONCLUSION

This paper is on an AMP based scheme for a coded LRMS with arbitrary input distributions. We show that AMP is information theoretically optimal using a curve matching principle and the IIDG assumption. In addition, a code design principle is provided for AMP, and the irregular LDPC codes are considered for binary signaling as an example. The numerical results show that AMP is capacity-approaching (i.e. within 1dB away from the limit) based on optimized irregular LDPC codes, and significant performance improvements (0.8 dB \sim 4 dB) are observed over the system without code optimization. Apart from that, AMP has lower complexity and better performance than the well-known Turbo-LMMSE algorithm.

The proof of Assumption 2 is an interesting future work. A rigorous SE proof for a certain kind of non-separable (e.g. uniformly Lipschitz) functions was established in [59], which may be used to prove Assumption 2 in this paper.

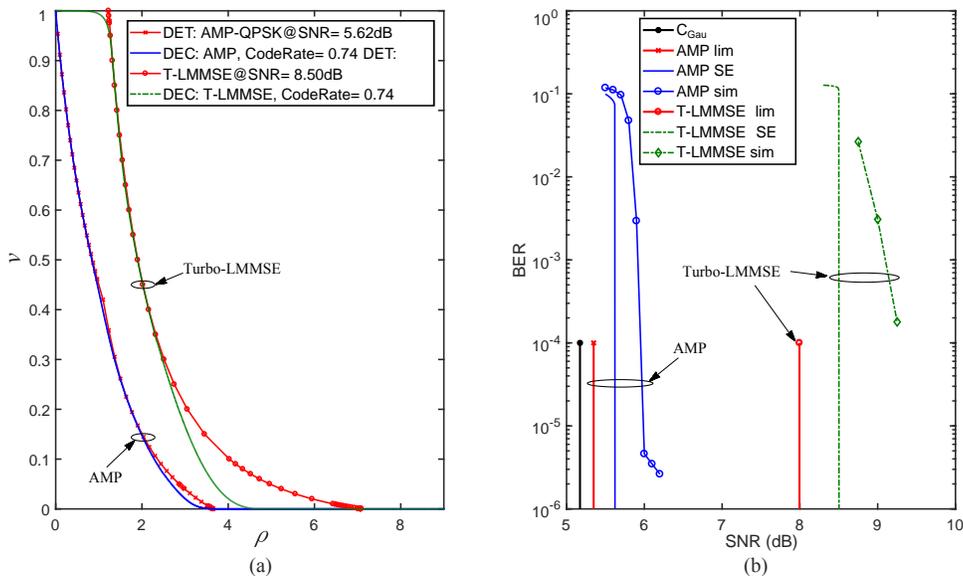


Fig. 16. Transfer function matching (left) and BER performances (right) of AMP and Turbo-LMMSE [32], [54] with optimized irregular LDPC codes, where “DET” denotes the LD transfer function of AMP/Turbo-LMMSE, “DEC” the optimized decoding transfer function of AMP/Turbo-LMMSE, C_{Gau} the Gaussian capacity, “SE” the state evolution, “lim” the QPSK achievable rate limits of AMP/Turbo-LMMSE, “sim” the simulated BERs. Code length = 10^5 , code rate ≈ 0.74 , QPSK modulation, iterations = 200, and $\beta = 1.5$ with $N = 500$ and $M = 333$, irregular LDPC codes are optimized for both AMP and Turbo-LMMSE. For more details, refer to Table I.

APPENDIX A PROOF OF PROPOSITION 1

In this appendix, we will first show the consistency between the capacity given in (9) and that derived in [28]. We then show $A_{\omega_C^*} = C$, where $A_{\omega_C^*}$ is given in (20) and C in (9).

A. Consistency Between (9) and the Results in [28]

Reference [28] is based on the following system model:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \quad (44)$$

where \mathbf{A} is an IIDG matrix with $A_{ij} \sim \mathcal{N}(0, 1/N)$, $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$, and $M, N \rightarrow \infty$ with a fixed $\delta = M/N$. Note that (44) is a real system, which is different from (1). The capacity of (44) was derived in [28] as follows.

Lemma 5: Assume that $\zeta = \text{mmse}_X(\delta/(1+\zeta))$ has one positive fixed point ζ^* . The capacity of the LRMS in (44) is given by

$$C = \frac{\delta}{2} \left[\log(1 + \zeta^*) - \zeta^*/(1 + \zeta^*) \right] + I(\delta/(1 + \zeta^*)), \quad (45)$$

where $I(s) = I(x; \sqrt{s}x + n)$, $\text{mmse}_X(s) = \text{mmse}(x | \sqrt{s}x + n)$.

Note that the following:

- (i) In (9), $1/2$ is removed as (1) is a complex LRMS, while (44) is a real one.
- (ii) Let v_x be the variance of x_i in (45), we have

$$\text{mmse}_X(s) = v_x \omega(v_x s), \quad (46a)$$

$$I(s) = I(x; \sqrt{s}x + z) = C_{\text{SISO}}(v_x s). \quad (46b)$$

- (iii) We have $v_x = \delta^{-1} \sigma^{-2} = \beta snr$ with $\beta = \delta^{-1} = N/M$, since (45) considers $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M)$ and $A_{ij} \sim$

$\mathcal{CN}(0, 1/N)$. Combining $v_x = \beta snr$ and (46), we have

$$\text{mmse}_X(\delta/(1+\zeta)) = \beta snr \omega(sn/(1+\zeta)), \quad (47a)$$

$$I(\delta/(1+\zeta^*)) = C_{\text{SISO}}(sn/(1+\zeta^*)). \quad (47b)$$

Following (i)-(iii), we can see the consistency between the capacities in (9) and (45).

B. Proof of Proposition 1

We now show $A_{\omega_C^*} = C$, where $A_{\omega_C^*}$ is given in (20) and C in (9).

Let $\rho^* = sn/(1+\zeta^*)$, i.e. $\zeta^* = sn/\rho^* - 1$. Then the fixed point function in (9) is rewritten to

$$sn/\rho^* - 1 = \beta snr \omega(\rho^*), \quad (48)$$

which is equivalent to the fixed point function $\omega(\rho) = \phi^{-1}(\rho)$. Substituting (4) and $\rho^* = sn/(1+\zeta^*)$ into (20), we have

$$A_{\omega_C^*} = \beta^{-1} \left[\rho^*/snr - \log(\rho^*/snr) - 1 \right] + \int_0^{\rho^*} \omega(\rho) d\rho \quad (49a)$$

$$= \beta^{-1} \left[\log(1 + \zeta^*) - \frac{\zeta^*}{1 + \zeta^*} \right] + C_{\text{SISO}} \left(\frac{snr}{1 + \zeta^*} \right). \quad (49b)$$

This is the same as the capacity C in (9). Hence, we complete the proof of Proposition 1.

APPENDIX B

AN ALTERNATIVE PROOF OF THE CAPACITY OF AN LRMS

In this appendix, we provide an alternative proof for the capacity of an LRMS using the properties of AMP. We call $\mathcal{M}_x(sn) \equiv \frac{1}{N} \mathbb{E} \{ \|\mathbf{x} - \hat{\mathbf{x}}_{\text{MMSE}}\|^2 \}$ the MMSE of an LRMS and $\mathcal{M}_{A_x}(sn) \equiv \frac{1}{N} \mathbb{E} \{ \|\mathbf{A}\mathbf{x} - \mathbf{A}\hat{\mathbf{x}}_{\text{MMSE}}\|^2 \}$ the measurement MMSE of the LRMS. The following lemma gives the capacity of an LRMS.

Lemma 6 (Measurement MMSE and Capacity): Assuming $\omega(\rho) = \phi^{-1}(\rho)$ has a unique positive solution ρ^* . The measurement MMSE of an LRMS is given by

$$\mathcal{M}_{Ax}(snr) = \rho^* \omega(\rho^*) / snr = \rho^* \mathcal{M}_x(snr) / snr, \quad (50a)$$

and the capacity is given by

$$C = A \omega_C^*, \quad (50b)$$

where $A \omega_C^*$ is defined in (20).

The next two subsections respectively give the proofs of (50a) and (50b) in Lemma 6.

- In B-A, we prove the measurement MMSE in (50a) of an LRMS using the MMSE optimality [28]–[30] and the decoupling property [38] of AMP.
- In B-B, we prove the capacity in (50b) of an LRMS based on the measurement MMSE in (50a) using the vector-I-MMSE theorem [35].

A. Proof of the Measurement MMSE in (50a)

As $t \rightarrow \infty$, the AMP in (10) converges to

$$\mathbf{s}^* = \mathbf{s}^\infty, \quad \mathbf{r}^* = \mathbf{r}^\infty, \quad \rho^* = \rho_\infty, \quad \omega^* = \omega(\rho^*) = v_\infty. \quad (51)$$

If Assumption 1 holds, from Lemma 4, we have $\mathbf{s}^* = \hat{\mathbf{x}}(\mathbf{y}; snr)$ and $\mathcal{M}_x(snr) = \omega(\rho^*)$. The following proposition is proved for AMP in [38] (see Theorem 1(A-a) in [38]).

Proposition 2: Note that $\langle \eta'(\mathbf{r}^*) \rangle = \rho^* \omega^*$. Define

$$\tilde{\mathbf{s}} = [\omega^{*-1} - \rho^*]^{-1} [\omega^{*-1} \mathbf{s}^* - \rho^* \mathbf{r}^*], \quad (52a)$$

$$\tilde{\mathbf{z}} = \tilde{\mathbf{s}} - \mathbf{x}, \quad (52b)$$

where the entries of $\tilde{\mathbf{z}}$ are IID with zero mean and variance $(\omega^{*-1} - \rho^*)^{-1}$. Then, $\tilde{\mathbf{z}}$ can be treated as a random variable that is asymptotically independent³ with \mathbf{n} and \mathbf{A} .

Using Proposition 2, we have the following lemma.

Lemma 7: The following equation asymptotically holds for the fixed point of AMP:

$$\frac{1}{N} \mathbb{E}\{(\mathbf{x} - \mathbf{s}^*)(\mathbf{x} - \mathbf{s}^*)^H\} = \frac{1}{N} [(\omega^{*-1} - \rho^*) \mathbf{I} + snr \mathbf{A}^H \mathbf{A}]^{-1}. \quad (53)$$

Proof: We rewrite (52) in Proposition 2 as

$$\mathbf{r}^* = \frac{\mathbf{s}^*}{\rho^* \omega^*} + \left(1 - \frac{1}{\rho^* \omega^*}\right) \tilde{\mathbf{s}}. \quad (54)$$

The output of AMP LD in (10) converges to

$$\mathbf{r}^* = \mathbf{s}^* + \mathbf{A}^H (\mathbf{y} - \mathbf{A} \mathbf{s}^*) + \beta \langle \eta'(\mathbf{r}^*) \rangle (\mathbf{r}^* - \mathbf{s}^*). \quad (55)$$

For MMSE function η [5],

$$\langle \eta'(\mathbf{r}^*) \rangle = \rho^* \omega^*. \quad (56)$$

With (56) and the fixed point function $snr = \rho^* [1 - \beta \rho^* \omega(\rho^*)]^{-1}$, (55) can be rewritten to

$$\mathbf{r}^* = [\mathbf{I} - snr \mathbf{A}^H \mathbf{A} / \rho^*] \mathbf{s}^* + snr \mathbf{A}^H \mathbf{y} / \rho^*. \quad (57)$$

From (54) and (57), we have

$$\mathbf{B} \mathbf{s}^* = snr \mathbf{A}^H \mathbf{y} + (\omega^{*-1} - \rho^*) \tilde{\mathbf{s}}, \quad (58)$$

³Let $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}$. In [38], it is proved that the entries of $\mathbf{b} = \mathbf{V} \tilde{\mathbf{z}}$ are IIDG and independent with \mathbf{n} and $\mathbf{U} \mathbf{\Lambda}$. Based on this, substituting $\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}$ and $\mathbf{b} = \mathbf{V} \tilde{\mathbf{z}}$ into (60b), we obtain (60c), which is the same as that $\tilde{\mathbf{z}}$ is independent with \mathbf{n} and \mathbf{A} .

where $\mathbf{B} = (\omega^{*-1} - \rho^*) \mathbf{I} + snr \mathbf{A}^H \mathbf{A}$. Thus,

$$\mathbf{s}^* = \mathbf{B}^{-1} [snr \mathbf{A}^H \mathbf{y} + (\omega^{*-1} - \rho^*) \tilde{\mathbf{s}}]. \quad (59)$$

Substituting (52) and $\mathbf{y} = \mathbf{A} \mathbf{x} + \mathbf{n}$ into (59), we have

$$\frac{1}{N} \mathbb{E}\{(\mathbf{x} - \mathbf{s}^*)(\mathbf{x} - \mathbf{s}^*)^H\} \quad (60a)$$

$$= \frac{1}{N} \mathbb{E}\left\{\left[\mathbf{B}^{-1} [snr \mathbf{A}^H \mathbf{n} + (\omega^{*-1} - \rho^*) \tilde{\mathbf{z}}]\right]^2\right\} \quad (60b)$$

$$= \frac{1}{N} [(\omega^{*-1} - \rho^*) \mathbf{I} + snr \mathbf{A}^H \mathbf{A}]^{-1}, \quad (60c)$$

where (60c) follows Proposition 2. Thus, we complete the proof of Lemma 7. ■

Based on Lemma 7, the MSE of AMP is given by

$$\mathcal{M}_x(snr) = \frac{1}{N} \mathbb{E}\{\|\mathbf{x} - \mathbf{s}^*\|^2\} = \omega(\rho^*) \quad (61a)$$

$$= \frac{1}{N} \text{Tr}\left\{[(\omega^{*-1} - \rho^*) \mathbf{I} + snr \mathbf{A}^H \mathbf{A}]^{-1}\right\} \quad (61b)$$

$$= \mathbb{E}_{\lambda_{\mathbf{A}^H \mathbf{A}}}\left\{[(\omega^{*-1} - \rho^*) + snr \lambda_{\mathbf{A}^H \mathbf{A}}]^{-1}\right\}, \quad (61c)$$

where $\lambda_{\mathbf{A}^H \mathbf{A}}$ is the eigenvalue of $\mathbf{A}^H \mathbf{A}$. Also, the measurement MMSE is derived as

$$\mathcal{M}_{Ax}(snr) \quad (62a)$$

$$= \frac{1}{N} \mathbb{E}\{\|\mathbf{A} \mathbf{x} - \mathbf{A} \mathbf{s}^*\|^2\} \quad (62b)$$

$$= \frac{1}{N} \text{Tr}\left\{\|\mathbf{A}(\mathbf{x} - \mathbf{s}^*)(\mathbf{x} - \mathbf{s}^*)^H \mathbf{A}^H\right\} \quad (62c)$$

$$= \frac{1}{N} \text{Tr}\left\{\mathbf{A}[(\omega^{*-1} - \rho^*) \mathbf{I} + snr \mathbf{A}^H \mathbf{A}]^{-1} \mathbf{A}^H\right\} \quad (62d)$$

$$= \mathbb{E}_{\lambda_{\mathbf{A}^H \mathbf{A}}}\left\{\lambda_{\mathbf{A}^H \mathbf{A}}[(\omega^{*-1} - \rho^*) + snr \lambda_{\mathbf{A}^H \mathbf{A}}]^{-1}\right\} \quad (62e)$$

$$= snr^{-1} \left[1 - \mathbb{E}_{\lambda_{\mathbf{A}^H \mathbf{A}}}\left\{[1 + snr(\omega^{*-1} - \rho^*)^{-1} \lambda_{\mathbf{A}^H \mathbf{A}}]^{-1}\right\}\right] \quad (62f)$$

$$= \rho^* \omega(\rho^*) / snr \quad (62g)$$

$$= \rho^* \mathcal{M}_x(snr) / snr, \quad (62h)$$

where (62d) follows (53), and (62g) and (62h) follows (61). Therefore, we obtain (50a).

B. Proof of the Capacity in (50b)

The connection between the measurement MMSE and the capacity of an LRMS is given by Lemma 8 proven in [35].

Lemma 8 (Vector I-MMSE): Consider a system $\mathbf{y} = \sqrt{snr} \mathbf{A} \mathbf{x} + \mathbf{z}$ where $\mathbf{x} \sim P_{\mathbf{x}}$ and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. Then, the capacity of this system is given by

$$C = \frac{1}{N} \mathbb{I}(\mathbf{x}; \sqrt{snr} \mathbf{A} \mathbf{x} + \mathbf{z}) = \int_0^{snr} \mathcal{M}_{Ax}(\rho) d\rho. \quad (63)$$

From (50a) and Lemma 8, we have

$$C = \int_0^{snr} \mathcal{M}_{Ax}(s) ds \quad (64a)$$

$$= \int_0^{snr} \rho \omega(\rho) / s ds \quad (64b)$$

$$= \int_0^{\rho^*} [1 - \beta \rho \omega(\rho)] \omega(\rho) d \frac{\rho}{1 - \beta \rho \omega(\rho)}, \quad (64c)$$

where (64a) follows (63), (64b) follows (50a), and (64c) follows the fixed point function. The following manipulations

show that $C = A\omega_c^*$.

$$C = \int_0^{\rho^*} \frac{\omega + \beta\rho^2\omega'}{1 - \beta\rho\omega} d\rho \quad (65a)$$

$$= \int_0^{\rho^*} \frac{\omega + \rho\omega' - \rho\omega'(1 - \beta\rho\omega)}{1 - \beta\rho\omega} d\rho \quad (65b)$$

$$= - \int_0^{\rho^*} \rho d\omega + \int_0^{\rho^*} \frac{\omega + \rho\omega'}{1 - \beta\rho\omega} d\rho \quad (65c)$$

$$= \int_0^{\rho^*} \rho^* d\omega + \beta^{-1} \int_0^{\rho^*} d \log(1 - \beta\rho\omega) \quad (65d)$$

$$= [-\rho\omega - \beta^{-1} \log(1 - \beta\rho\omega)]_{\rho=0}^{\rho=\rho^*} + \int_0^{\rho^*} \omega d\rho \quad (65e)$$

$$= \beta^{-1} [\rho^*/snr - \log(\rho^*/snr) - 1] + \int_0^{\rho^*} \omega(\rho) d\rho \quad (65f)$$

$$= A\omega_c^*, \quad (65g)$$

where ω' is the derivative of $\omega(\rho)$, (65e) follows the fixed point function, and (65g) follows (20). Thus, we obtain (50b).

APPENDIX C GAUSSIAN SIGNALING

We now study a special case of Gaussian \mathbf{x} . In C-A, we will show a simpler proof as well as a closed-form expression of the area property. In C-B, we will show the curve matching condition asymptotically holds, i.e., there exists an n -layer SCM code whose transfer function matches with the desired $\omega_{C-Gau}^*(\rho)$ as $n \rightarrow \infty$.

A. Gaussian Area Property

The following lemma gives an explicit form of the measurement MMSE for Gaussian signaling.

Lemma 9 (Gaussian Measurement MMSE): For Gaussian signaling, the measurement MMSE of an LRMS is given by

$$\mathcal{M}_{Ax}(snr) = snr^{-1} \rho_{Gau}^* / (1 + \rho_{Gau}^*), \quad (66)$$

where ρ_{Gau}^* is given in (24).

Proof: For $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, the following LMMSE detection is a global MMSE estimation of an uncoded LRMS.

$$\mathbf{s}^* = (\mathbf{A}^H \mathbf{A}^H + snr^{-1} \mathbf{I})^{-1} \mathbf{A}^H \mathbf{y}. \quad (67)$$

Its average MSE is the corresponding MMSE, i.e.,

$$\mathcal{M}_x(snrc) = \frac{1}{N} \text{Tr}\{(\mathbf{s}^* - \mathbf{x})(\mathbf{s}^* - \mathbf{x})^H\} \quad (68a)$$

$$= \frac{1}{N} \text{Tr}\{(snr \mathbf{A}^H \mathbf{A} + \mathbf{I})^{-1}\} \quad (68b)$$

$$= \text{E} \lambda_{\mathbf{A}^H \mathbf{A}} \left\{ (1 + snr \lambda_{\mathbf{A}^H \mathbf{A}})^{-1} \right\}. \quad (68c)$$

Then, we obtain the measurement MMSE as

$$\mathcal{M}_{Ax}(snrc) = \frac{1}{N} \text{Tr}\{\mathbf{A}(\mathbf{s}^* - \mathbf{x})(\mathbf{s}^* - \mathbf{x})^H \mathbf{A}^H\} \quad (69a)$$

$$= \frac{1}{N} \text{Tr}\{\mathbf{A}^H \mathbf{A} (snr \mathbf{A}^H \mathbf{A} + \mathbf{I})^{-1}\} \quad (69b)$$

$$= \text{E} \lambda_{\mathbf{A}^H \mathbf{A}} \left\{ \lambda_{\mathbf{A}^H \mathbf{A}} (1 + snr \lambda_{\mathbf{A}^H \mathbf{A}})^{-1} \right\} \quad (69c)$$

$$= snr^{-1} (1 - \mathcal{M}_x(snrc)). \quad (69d)$$

In addition, according to (15) and $\mathcal{M}_x(snrc) = \omega_{Gau}(\rho_{Gau}^*) = 1/(1 + \rho_{Gau}^*)$, we have

$$\mathcal{M}_{Ax}(snrc) = \rho_{Gau}^* \mathcal{M}_x(snrc) / snrc \quad (70a)$$

$$= snr^{-1} \rho_{Gau}^* / (1 + \rho_{Gau}^*). \quad (70b)$$

Thus, we obtain Lemma 9. \blacksquare

Now we are ready to prove (26). The Gaussian capacity of an LRMS is [57]

$$C_{Gau} = \frac{1}{N} I(\mathbf{x}; \mathbf{y}) = \frac{1}{N} \log \det(\mathbf{I} + snr \mathbf{A}^H \mathbf{A}), \quad (71a)$$

which is achieved by $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. For IIDG \mathbf{A} with $A_{ij} \sim \mathcal{CN}(0, 1/M)$, we have [58]

$$C_{Gau} \rightarrow \log[1 + snr - \mathcal{F}] + \beta^{-1} \log[1 + snr\beta - \mathcal{F}] - snr^{-1} \beta^{-1} \mathcal{F}, \quad (71b)$$

where $\mathcal{F} = 0.25 \left(\sqrt{snr(1 + \sqrt{\beta})^2 + 1} - \sqrt{snr(1 - \sqrt{\beta})^2 + 1} \right)^2$.

Next, we show that $A\omega_{C-Gau}^*$ is equal to the Gaussian capacity C_{Gau} . From (25),

$$A\omega_{C-Gau}^* = \beta^{-1} \log(1 + \beta snr v_{Gau}^*) - \log(v_{Gau}^*) + v_{Gau}^* - 1, \quad (72a)$$

with

$$v_{Gau}^* = \phi^{-1}(\rho_{Gau}^*) \quad (73a)$$

$$= \frac{\beta - 1 - snr^{-1} + \sqrt{(\beta - 1)^2 + 2(\beta + 1)snr^{-1} + snr^{-2}}}{2\beta} \quad (73b)$$

$$= 1 - snr^{-1} \mathcal{F} / \beta. \quad (73c)$$

Therefore,

$$A\omega_{C-Gau}^* = -\log\left(1 - \frac{\mathcal{F}}{\beta snr}\right) + \frac{1}{\beta} \log(1 + \beta snr - \mathcal{F}) - \frac{snr \mathcal{F}}{\beta} \quad (74a)$$

$$= \log(1 + snr - \mathcal{F}) + \frac{1}{\beta} \log(1 + \beta snr - \mathcal{F}) - \frac{\mathcal{F}}{\beta snr}, \quad (74b)$$

where the second equation follows from $(1 - snr^{-1} \mathcal{F} / \beta)(1 + snr - \mathcal{F}) = 1$. From (74) and (71), we have (26):

$$C_{Gau} = A\omega_{C-Gau}^*. \quad (75)$$

B. Code Existence for Gaussian Signaling

It was proved in [31] that there exists an infinite-layer SCM code whose extrinsic decoding transfer function asymptotically approaches a target monotonically decreasing curve. In this subsection, we extend this result to the APP decoding transfer function. We will show that there is an SCM code whose transfer function $\omega_{C_n}(\rho)$ asymptotically approaches ω_{C-Gau}^* . Then, we will then show that the rate of this SCM code asymptotically approaches to $A\omega_{C-Gau}^*$.

We first show that R_{AMP} can be approached with an infinite-layer SCM code. It is easy to verify that $\phi^{-1}(\rho)$ satisfies the following regularity conditions:

- (a) $\phi^{-1}(\rho) \geq 0$, for $\rho \in [0, snrc]$;
- (b) monotonically decreasing in $\rho \in [0, \infty)$;
- (c) equation $(p^{-1} + \rho)\phi^{-1}(\rho) = 1$ has only one positive solution ρ_p^* for any $p \in (0, 1]$;
- (d) continuous and differentiable in $[\rho_{Gau}^*, \infty)$ except for a countable set of values of ρ .

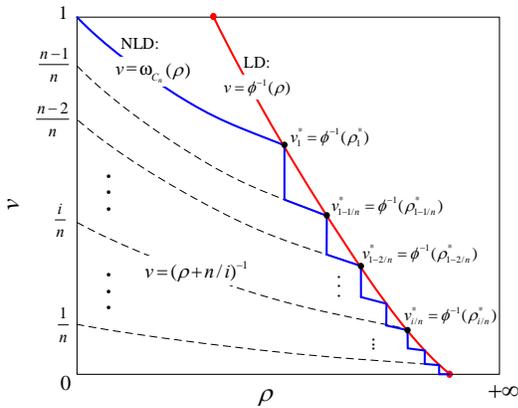


Fig. 17. An illustration of the transfer function of AMP LD $v = \phi^{-1}(\rho)$ and the corresponding transfer function $v = \omega_{C_n}(\rho)$ of the asymptotically matched n -layer SCM decoder.

Consider an n -layer SCM code $x = \sum_{i=0}^{n-1} x_i$ and the power of x_i is $p_{x_i} = 1/N$. In addition, x_i is encoded using an idea random code with code rate⁴

$$R_{n,i} = \log \left(1 + \frac{1/n}{\rho_{1-i/n}^{*-1} + (n-i-1)/n} \right), \quad (76)$$

where $i \in \{0, \dots, n-1\}$, and $\rho_{i/n}^*$ is the positive solution of $(\rho + n/i)\phi^{-1}(\rho) = 1$. Condition (c) ensures the existence of $\{\rho_{i/n}^*\}$, and condition (b) ensures $\rho_1^* < \dots < \rho_{2/n}^* < \rho_{1/n}^*$.

For any $i \in \{0, \dots, n-1\}$ and the decoder's input $\rho \in [\rho_{1-i/n}^*, \rho_{1-(i+1)/n}^*)$, the first $i+1$ layers $[x_0, \dots, x_i]$ can be successively decoded in the order from x_0 to x_i . Thus, under APP decoding, the transfer function of $x = \sum_{i=0}^{n-1} x_i$ for input $x + \rho^{-1/2}z$ with $z \sim \mathcal{CN}(0, 1)$ is

$$\omega_{C_n}(\rho) = \begin{cases} \frac{1}{\rho+1}, & 0 \leq \rho < \rho_1^* \\ \frac{1}{\rho+n/(n-i)}, & \rho_{1-i/n}^* \leq \rho < \rho_{1-(i+1)/n}^* \\ 0, & \rho_{1/n}^* < \rho < \infty \end{cases}, \quad (77)$$

where $i = 1, \dots, n-2$.

Fig. 17 shows the transfer functions of n -layer-SCM decoder ($v = \omega_{C_n}(\rho)$) and LD ($v = \phi^{-1}(\rho)$). Conditions (a)-(c) ensure that the decoder transfer function lies below that of LD, i.e.,

$$\omega_{C_n}(\rho) \leq \phi^{-1}(\rho), \quad \forall \rho \geq 0. \quad (78)$$

Define $v \equiv f(\rho) = [1/\phi^{-1}(\rho) - \rho]^{-1}$, and we have $\rho_v^* = f^{-1}(v)$, where $f^{-1}(\cdot)$ is the inverse function of $f(\cdot)$. Then, as

⁴Note that (76) considers Gaussian signaling for each x_i . For discrete signaling, from Lemma 1 in [35], we can complete the proof by replacing (76) with

$$R_{n,i} = \frac{1/n}{\rho_{1-i/n}^{*-1} + (n-i-1)/n} + o\left(\frac{1/n}{\rho_{1-i/n}^{*-1} + (n-i-1)/n}\right),$$

for all $i \in \{0, \dots, n-1\}$.

$n \rightarrow \infty$, the sum rate of the SCM code is given by

$$R_n = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} \log \left(1 + \frac{1/n}{\rho_{1-i/n}^{*-1} + (n-i-1)/n} \right) \quad (79a)$$

$$= \lim_{n \rightarrow \infty} \frac{1/n}{\rho_{1-i/n}^{*-1} + (n-i-1)/n} \quad (79b)$$

$$= \int_0^1 [\rho_v^{*-1} + v]^{-1} dv \quad (79c)$$

$$= \int_0^1 [[f^{-1}(v)]^{-1} + v]^{-1} dv \quad (79d)$$

$$= \left[[\rho^{-1} + f(\rho)]^{-1} f(\rho) \right]_{\rho=\rho_0^*}^{\rho=\rho_1^*} + \int_{\rho_0^*}^{\rho_1^*} f(\rho) d[\rho^{-1} + f(\rho)]^{-1} \quad (79e)$$

$$= [\rho \phi^{-1}(\rho)]_{\rho=\rho_0^*}^{\rho=\rho_1^*} + \int_{\rho_1^*}^{\rho_0^*} \frac{\phi^{-1}(\rho)}{1 - \rho \phi^{-1}(\rho)} d\rho (1 - \rho \phi^{-1}(\rho)) \quad (79f)$$

$$= [\log(1 - \rho \phi^{-1}(\rho))]_{\rho=\rho_0^*}^{\rho=\rho_1^*} + \int_{\rho_1^*}^{\rho_0^*} \phi^{-1}(\rho) d\rho \quad (79g)$$

$$= \log(1 + \rho_1^*) + \int_{\rho_1^*}^{\infty} \phi^{-1}(\rho) d\rho \quad (79h)$$

$$= A \omega_{C-\text{Gau}}^*, \quad (79i)$$

where (79e) follows the *inverse integral lemma* below

$$\int g(y, f^{-1}(y)) dy \quad (80a)$$

$$= g(f(x), x) f(x) - \int f(x) d g(f(x), x) + \text{Constant}, \quad (80b)$$

(79f) from $f(\rho) = [1/\phi^{-1}(\rho) - \rho]^{-1}$, (79i) follows (20) and $\omega_{\text{Gau}}(\rho) = 1/(1 + \rho)$, and (79h) utilizes the following facts:

- $1 - \rho_1^* \phi^{-1}(\rho_1^*) = 1/(1 + \rho_1^*)$;
- $\rho_0^* \phi^{-1}(\rho_0^*) = 0$ from $\phi^{-1}(\rho_0^*) = 0$ if ρ_0^* is finite;
- $\phi^{-1}(\rho) = 0$ for any $\rho > \rho_0^*$ if ρ_0^* is finite, since $\phi^{-1}(\rho_1^*) = 0$, and $\phi^{-1}(\rho)$ is positive and monotonically decreasing in $\rho \in [0, \infty)$.

Thus, we obtain the desired result: $R_n = A \omega_{C-\text{Gau}}^*$.

REFERENCES

- [1] L. Liu, C. Liang, J. Ma, and L. Ping, "Capacity optimality of AMP in coded systems," *IEEE ISIT*, 2021. (Accepted)
- [2] D. Micciancio, "The hardness of the closest vector problem with preprocessing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1212–1215, Mar. 2001.
- [3] S. Verdú, "Optimum multi-user signal detection," Ph.D. dissertation, Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, Aug. 1984.
- [4] 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.
- [5] 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.
- [6] P. Schniter, S. Rangan, A. K. Fletcher, and M. Borgerding, "Vector AMP and its connections to deep learning," *IEEE Info. Thy. Workshop*, Cambridge, UK, Sept. 2016.
- [7] U. Kamilov, S. Rangan, A. Fletcher, and M. Unser, "Approximate message passing with consistent parameter estimation and applications to sparse learning," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2969–2985, May 2014.
- [8] S. Som and P. Schniter, "Compressive imaging using approximate message passing and a markov-tree prior," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3439–3448, Jul. 2012.

- [9] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," in *Proc. Allerton Conf. Commun., Contr., Comput.*, Oct. 2012, pp. 815–822.
- [10] Y. Ma, J. Zhu and D. Baron, "Approximate message passing algorithm with universal denoising and Gaussian mixture learning," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5611–5622, Nov. 2016.
- [11] Y. Kabashima, "A CDMA multiuser detection algorithm on the basis of belief propagation," *Journal of Physics A: Mathematical and General*, vol. 36, no. 43, p. 11111, 2003.
- [12] J. P. Neirotti and D. Saad, "Improved message passing for inference in densely connected systems," *EPL (Europhysics Letters)*, vol. 71, no. 5, p. 866, 2005.
- [13] P. Schniter, "A message-passing receiver for BICM-OFDM over unknown clustered-sparse channels," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1462–1474, Dec 2011.
- [14] M. Nassar, P. Schniter, and B. Evans, "A factor graph approach to joint OFDM channel estimation and decoding in impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1576–1589, Mar. 2014.
- [15] S. Wu, L. Kuang, Z. Ni, J. Lu, D. Huang, and Q. Guo, "Low-complexity iterative detection for large-scale multiuser MIMO-OFDM systems using approximate message passing," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 902–915, Oct. 2014.
- [16] M. Nabae and F. Labeau, "Bayesian quantized network coding via generalized approximate message passing," in *Proc. Wireless Telecommun. Symp. (WTS)*, Apr. 2014, pp. 1–7.
- [17] 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.
- [18] 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, March 2017.
- [19] S. Liang, C. Liang, J. Ma, L. Ping, "Compressed coding, AMP based decoding and analog spatial coupling," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7362–7375, Dec. 2020.
- [20] C. Liang, J. Ma and L. Ping, "Towards Gaussian capacity, universality and short block length," *ISTC*, 2016, pp. 412–416.
- [21] 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.
- [22] T. P. Minka, "Expectation propagation for approximate bayesian inference," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, 2001, pp. 362–369.
- [23] B. Çakmak and M. Opper, "Expectation propagation for approximate inference: Free probability framework," *IEEE ISIT*, 2018, pp. 1276–1280, preprint *arXiv:1801.05411*, 2018.
- [24] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017, preprint *arXiv:1602.06509*, 2016.
- [25] I. Santos, J. J. Murillo-Fuentes, R. Boloix-Tortosa, E. Arias-de-Reyna and P. M. Olmos, "Expectation propagation as Turbo equalizer in ISI channels," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 360–370, Jan. 2017.
- [26] X. Meng, S. Wu, L. Kuang, Z. Ni and J. Lu, "Expectation propagation based iterative multi-user detection for MIMO-IDMA systems," *2014 IEEE 79th VTC (VTC Spring)*, Seoul, 2014, pp. 1–5.
- [27] J. Ma, L. Liu, X. Yuan and L. Ping, "On orthogonal AMP in coded linear vector systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 12, pp. 5658–5672, Dec. 2019.
- [28] 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, April 2019.
- [29] 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, July 2020.
- [30] A. M. Tulino, G. Caire, S. Verdú, and S. Shamai (Shitz), "Support recovery with sparsely sampled free random matrices," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4243–4271, Jul. 2013.
- [31] X. Yuan, L. Ping, C. Xu and A. Kavcic, "Achievable rates of MIMO systems with linear precoding and iterative LMMSE detector," *IEEE Trans. Inf. Theory*, vol. 60, no.11, pp. 7073–7089, Oct. 2014.
- [32] L. Liu, C. Yuen, Y. L. Guan, and Y. Li, "Capacity-achieving MIMO-NOMA: Iterative LMMSE detection," *IEEE Trans. Signal Process.*, vol. 67, no. 7, 1758–1773, April 2019.
- [33] X. Wang and H. V. Poor, "Iterative (Turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul 1999.
- [34] H. A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, June 2007.
- [35] 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, Apr. 2005.
- [36] K. Bhattad and K. R. 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] J. Barbier, M. Dia, N. Macris and F. Krzakala, "The mutual information in random linear estimation," *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Monticello, IL, 2016, pp. 625–632.
- [38] K. Takeuchi, "A unified framework of state evolution for message-passing algorithms," *IEEE ISIT*, 2019, pp. 151–155, preprint *arXiv:1901.03041*, 2019.
- [39] A. Lozano, A. M. Tulino and S. Verdú, "Optimum power allocation for parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.
- [40] R. Y. S. Tee, R. G. Maunder and L. Hanzo, "EXIT-chart aided near-capacity irregular bit-interleaved coded modulation design," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 32–37, Jan. 2009.
- [41] L. Kong, S. X. Ng, R. Y. S. Tee, R. G. Maunder and L. Hanzo, "Reduced-complexity near-capacity downlink iteratively decoded generalized multi-layer space-time coding using irregular convolutional codes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 684–695, Feb. 2010.
- [42] K. Wu, K. Anwar, T. Matsumoto, "BICM-ID-based IDMA: Convergence and rate region analyses," *IEICE Trans. Commun.*, vol. E97.B, no. 7, pp. 1483–1492, July 2014.
- [43] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.
- [44] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, Jun. 2005.
- [45] S. ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, Oct. 2001.
- [46] S. ten Brink, G. Kramer, and A. Ashikhmin, "Design of low-density parity-check codes for modulation and detection," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 670–678, April 2004.
- [47] X. Yuan, *Low-complexity iterative detection in coded linear systems*, PhD thesis, City University of Hong Kong, Hong Kong, China, 2008.
- [48] S.-Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," vol. 47, no. 2, pp. 657–670, Feb. 2001.
- [49] Y. Hu, C. Liang, L. Liu, C. Y. Y. Y, and L. Ping "Interleave-division multiple access in high rate applications," *IEEE Commun. Letter*, vol. 8, no. 2, pp. 476–479, April 2019.
- [50] X. Wang, S. Cammerer, and S. Brink, "Near-capacity detection and decoding: Code design for dynamic user loads in Gaussian multiple access channels," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7417–7430, Nov. 2019.
- [51] X. Wang, C. Liang, L. Ping and S. ten Brink, "Achievable rate region for iterative multi-user detection via low-cost Gaussian approximation," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3289–3303, May 2020.
- [52] G. Song, X. Wang and J. Cheng, "A low-complexity multiuser coding scheme with near-capacity performance," *IEEE Trans. on Vehi. Techn.*, vol. 66, no. 8, pp. 6775–6786, Aug. 2017.
- [53] G. Song, J. Cheng, and Y. Watanabe, "Maximum sum rate of repeat-accumulate interleave-division system by fixed-point analysis," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3011–3022, Oct. 2012.
- [54] Y. Chi, L. Liu, G. Song, C. Yuen, Y. L. Guan and Y. Li, "Practical MIMO-NOMA: Low complexity and capacity-approaching solution," *IEEE Trans. Wireless Commun.*, vol. 17, no. 9, pp. 6251–6264, Sept. 2018.
- [55] R. G. Gallager, "Low-density parity-check codes," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 21–28, Jan. 1962.
- [56] S.-Y. Chung, G. D. Forney, Jr., T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Commun. Lett.*, vol. 5, pp. 58–60, Feb. 2001.
- [57] Tse David and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [58] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Commun. and Inf. theory*, 2004.
- [59] R. Berthier, A. Montanari, P. M. Nguyen, "State evolution for approximate message passing with non-separable functions," *Information and Inference: A Journal of the IMA*, vol. 9, no. 1, pp. 33–79, Mar. 2020.