Area Theorems for Coded MIMO Systems with Message-Passing Decoding

Xiaojun Yuan and Li Ping
Department of Electronic Engineering, City University of Hong Kong, Hong Kong

Abstract—In this paper, we extend the area theorem of an iterative decoder on an additive-white-Gaussian-noise (AWGN) channel with binary input to the case of multiple-input multiple-output (MIMO) channels with arbitrary input. We also establish the area theorem for iterative detection involving suboptimal local processors.

I. INTRODUCTION

The optimal detection of a coded signal in a complicated channel environment may incur excessive complexity. Iterative detection provides a low-cost solution by decomposing the overall receiver into two or more local processors ([1]-[3] and the references therein). The analysis of an iterative detection process is an intriguing problem. The density evolution technique [2] shows that carefully designed low-density parity-check (LDPC) codes can achieve near-capacity performance in additive white Gaussian noise (AWGN) channels with iterative message-passing decoding algorithms. It has been shown in [4] that the achievable rate of an iterative scheme for an erasure channel can be measured by the area under the extrinsic information transfer (EXIT) curves and the channel capacity is approachable when the two local processors have matched EXIT curves. This area property is extended in [5] to AWGN channels based on the measure of minimum mean square error (MMSE), which establishes a sufficient condition to approach the capacity of a binary coded AWGN channel with iterative detection.

The work in [5] is limited to simple binary-coded single-input single-output AWGN channels. In this paper, we extend the area properties for more general multiple-input multiple-output (MIMO) transmission scenarios. We also extend the area property to iterative systems involving suboptimal local processors, such as those considered in [3] and [9]. We show by numerical results that the performance limit predicted by our analysis is actually approachable by proper coding design.

II. PRELIMINARIES

A. Message

Consider a random variable \( X \in \mathcal{S} \), where \( \mathcal{S} \) is called the constellation of \( X \). Let \( p_s(x) \) be the probability density function (PDF) of \( X \) defined on \( \mathcal{S} \). We refer to \( p_s(x) \) as the constellation information. Without loss of generality, we impose

\[
\int_{-\infty}^{\infty} x^2 p_s(x) = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} x^4 p_s(x) = 1 .
\]

Let \( Z \) be a random variable correlated with \( X \), and \( z \) be the observed value of \( Z \). A message \( \chi \) of \( X \) provided by \( Z \) is defined as the likelihood function \( p(z|x) \). By definition, \( \chi \) is a random function of \( Z \).

A likelihood function can be converted into a PDF by properly scaling. For example, \( p(z|x) \) can be converted into a PDF of \( X \), denoted by \( g(x) \), as

\[
g(x) = \frac{1}{\alpha} p(z|x) \quad \text{with} \quad \alpha = \int_{-\infty}^{\infty} p(z|x)dx .
\]

Based on (3a), we write

\[
g(x) \propto p(z|x)
\]

where "\( \propto \)" means equality up to a scaling factor invariant to \( x \).

B. Independent Messages

Now consider one more random variable \( U \). Assume that \( Z \) and \( U \) are conditionally independent given \( X \), i.e., \( Z \to X \to U \) forms a Markov chain. From the Bayesian rule, the overall message provided by \( Z \) and \( U \) can be factorized as

\[
p(z,u|x) = p(z|x)p(u|x) .
\]

We then say that the messages \( p(z|x) \) and \( p(u|x) \) are independent.

C. Constellation Information as a Message

We can express the constellation information as a message as follows. Let \( p_s(x) \) be the constellation information of \( X \) and \( z \) be a noisy observation of \( X \). The a posteriori PDF (APP) of \( X \) is given by

\[
p(x|z) = \frac{p(z|x)p_s(x)}{p(z)} .
\]

Since \( p(z) \) is not a function of \( x \), we can write

\[
p(x|z) \propto p(z|x)p_s(x) .
\]

Let \( U \) be a virtual noisy observation of \( X \) such that: (i) \( U = X + N \), where \( N \sim p_N(n) = p_N(-n) \) is independent of \( X \) and \( Z \); (ii) the observed value of \( U \) is \( u = 0 \). Using (2), we obtain a message

\[
p(z,u = 0|x) = p(z|x)p(u = 0|x) .
\]

By definition, \( p(u = 0|x) = p_N(-x) = p_N(x) \), and so

\[
p(z,u = 0|x) = p(z|x)p_s(x) .
\]

Comparing (4c) with (3), we can treat \( p_s(x) \) as a message generated by the virtual observation \( U \).

D. Factor Graph

Now consider the estimation of \( X \) based on two observations \( Z \) and \( Z' \). Assume that \( Z \) and \( Z' \) are conditionally independent given \( X \). From the Bayesian rule, the APP of \( X \) is

\[
p(x|z,z') \propto p(z|x)p(z'|x)p_s(x) \propto p(z|x)p(z'|x)p(u = 0|x) \propto p(z,z',u = 0|x)
\]
where $U = u = 0$ is the virtual observation representing the constellation information as defined earlier.

The factorizations in (5) can be characterized using a factor graph in Fig.1, where a white node represents a random variable to be estimated, a black node represents an observed random variable and a square node represents a constraint. The constellation information is generated by a virtual noisy observation $U$. In Fig.1, the leftward message on edge $e$ is the likelihood function of $X$ given $U$ and $Z'$, expressed as

$$p(z | u) = p(z | x)p_X(x).$$

More details on factor graphs can be found, e.g., in [6].

![Factor Graph](image)

**Fig. 1.** The factor graph representation of the factorization in (5). Circle nodes represent random variables and square nodes represent constraints. The black node $U$ is a virtual node that represents constellation information $p_X(x)$.

### E. Gaussian Messages

A Gaussian message of $X$ is a likelihood function that can be converted to a Gaussian distribution. It can be generated by an observation from an auxiliary AWGN channel modeled as

$$Z = X + N,$$

where $N \sim \mathcal{CN}(0, v)$ is independent of $X$, and $v$ is the noise variance. The Gaussian message given by $Z = z$ is computed as

$$p(z | x) = \frac{1}{\pi \sqrt{v}} \exp \left( -\frac{|x-z|^2}{v} \right) = \mathcal{CN}(z, v).$$

The above conversion of the observation over the auxiliary AWGN channel (7) to the Gaussian message in (8) plays a crucial role in establishing the area properties in Section III.

Consider two observations $Z = X + N$ and $Z' = X + N'$ with $N \sim \mathcal{CN}(0, v)$ and $N' \sim \mathcal{CN}(0, v')$. $Z$ and $Z'$ are conditionally independent given $X$. Thus

$$p(z, z' | x) = p(z | x)p(z' | x).$$

It can be shown that $p(z, z' | x)$ is also a Gaussian message with the mean $\bar{z}$ and variance $\bar{v}$, respectively given by [6]

$$\frac{\bar{z}}{\bar{v}} = \frac{z + z'}{v + v'}$$

Note that (10) is the combining rule of Gaussian messages.

### F. Multivariate Messages

Now consider messages for the multivariate case. Let $X$ be a length-$J$ vector of random variables. Similar to the treatment for a scalar, a message $\chi$ is a likelihood function that can be scaled to a distribution of $X$. Denote by $\chi(j)$ the marginal of $\chi$ with respect to $X(j)$. Let $Z(j) = z(j)$ be the random variable corresponding to $\chi(j)$ (i.e., $\chi(j) = p(z(j)|x(j))$). The following definitions are useful for our later discussions.

**Definition 1 (Uniform Message):** A message $\chi$ is said to be uniform if the conditional PDF of each $Z(j)$ given $X(j)$ is invariant to the index $j$, i.e.,

$$p_{\chi(j)}(z(j) | x(j)) = p_{\chi(j)}(z(j) | x), \forall j, k.$$  

**Definition 2 (Uncorrelated Message):** A message $\chi$ is said to be uncorrelated if $\chi$ is scalable to the product of $\{\chi(j)\}$.

We have some useful properties for a Gaussian message $\chi = \mathcal{CN}(\bar{z}, \Omega)$, where $\bar{z}$ is the mean vector and $\Omega$ is the covariance matrix. For example, if all diagonal entries of $\Omega$ are equal, $\chi$ is uniform. If $\Omega$ is a diagonal matrix, $\chi$ is uncorrelated.

![Factor Graph](image)

**Fig. 2.** The factor graph illustration of the messages defined in (13).

### III. APP PROCESSOR: MESSAGES AND AREA PROPERTIES

We now consider a processor that estimates $\bar{X}$ based on two observations $\bar{Z}_{\text{SIDE}} = z_{\text{SIDE}}$ and $\bar{Y} = \bar{y}$. We call $\bar{Z}_{\text{SIDE}}$ the side information, hence the superscript. Usually, $\bar{Y}$ represents the observation from the physical channel. Assume that

a) $\{\bar{Z}_{\text{SIDE}}(j)\}$ are conditionally independent given $\bar{X}$;

b) $\bar{Z}_{\text{SIDE}}$ and $\bar{Y}$ are conditionally independent given $\bar{X}$; and
c) $\bar{X}$ is i.i.d., i.e., the constellation information of $\bar{X}$ satisfies

$$p_{\bar{X}}(\bar{x}) = \prod_{j=0}^{J-1} p_{\bar{X}}(x(j)).$$

In this section, we establish the area properties of a processor based on the above assumptions. These assumptions will be examined in Section IV.

**A. Message Classification of an APP Processor**

We focus on a particular entry $X(j)$ (with the realization denoted by $x(j)$) in $\bar{X}$. Define

- **Total likelihood message:** $\chi^{\text{TL}}(j) = p(\bar{Y}, \bar{Z}_{\text{SIDE}} | x(j))$ (13a)
- **Side message:** $\chi^{\text{SIDE}}(j) = p(\bar{Z}_{\text{SIDE}}(j) | x(j))$ (13b)
- **Extrinsic message:** $\chi^{\text{EXT}}(j) = p(\bar{Y}, \bar{Z}_{\text{SIDE}} \setminus \bar{Z}_{\text{SIDE}}(j) | x(j))$ (13c)

1 This is equivalent to say that $\chi^{\text{SIDE}}$ in (13b) is an uncorrelated message.
We refer to a processor that calculates the messages in (13) as an APP processor. Fig. 2 illustrates the relations among the three types of messages. Noting that \( \{ \mathcal{Y}, \mathcal{Z}^{\text{SIDE}} \} \Rightarrow \mathcal{X} \Rightarrow \mathcal{Z}^{\text{SIDE}}(j) \) forms a Markov chain for each \( j \), we obtain
\[
\chi^{\text{SIDE}}(j) = \chi^{\text{EXT}}(j) \mathcal{X}^{\text{SIDE}}(j).
\] (14a)

From the Bayesian rule, the \textit{a posteriori} PDF of \( X(j) \) given \( \mathcal{Z}^{\text{SIDE}} \) and \( \mathcal{Y} \) is readily determined by the messages in (13) as
\[
p(x(j)|\mathcal{Z}^{\text{SIDE}},\mathcal{Y}) \propto \chi^{\text{EXT}}(j) \mathcal{X}^{\text{SIDE}}(j) p_x(x).
\] (14b)

\[\propto \chi^{\text{EXT}}(j) \mathcal{X}^{\text{SIDE}}(j) p_x(x).\] (14c)

\[B. \quad \text{Uniform and Gaussian (UG) Condition}\]

Assumption and Verifications

We assume that the messages defined in (13) are all uniform and Gaussian (UG). Techniques to partially relax this restriction are discussed in a separate paper. Under the UG condition, the messages in (13) can be expressed respectively as
\[
\chi^{\text{SIDE}}(j) \propto \mathcal{CN}(\mathcal{Z}^{\text{SIDE}}(j), \nu^{\text{SIDE}}(j)), \quad \chi^{\text{EXT}}(j) \propto \mathcal{CN}((\chi^{\text{SIDE}}(j), \nu^{\text{SIDE}}(j)), \nu^{\text{EXT}}(j)).
\] (15a)

(15b)

(15c)

For ease of discussion, define
\[
\rho^{\text{TL}} = 1/\nu^{\text{TL}}, \quad \rho^{\text{SIDE}} = 1/\nu^{\text{SIDE}}, \quad \rho^{\text{EXT}} = 1/\nu^{\text{EXT}}.
\] (16)

Recall that a Gaussian message can be generated by a random variable modeled as the observation of an auxiliary AWGN channel in (7). The variables defined in (16) are actually the signal to noise ratios (SNRs) of the auxiliary AWGN channels.

Eqn. (14a) says that \( \chi^{\text{TL}}(j) \) is obtained by combining \( \chi^{\text{SIDE}}(j) \) and \( \chi^{\text{EXT}}(j) \). From (10), (14a), (15), and (16), we have
\[
\rho^{\text{TL}} = \rho^{\text{EXT}} + \rho^{\text{SIDE}}\quad (17)
\]

This linearity of the SNRs is the key to the area properties discussed in the next subsection.

If the UG condition is met, we use an SNR transfer function
\[
\rho^{\text{EXT}} = \phi(\rho^{\text{SIDE}})\quad (18)
\]

to characterize the behavior of an APP detector. From (16) and (18), we have the following useful relationship,
\[
\rho^{\text{TL}} = \phi(\rho^{\text{SIDE}}) + \rho^{\text{SIDE}} = \rho^{\text{EXT}} + \phi^{-1}(\rho^{\text{EXT}})\quad (19)
\]

where \( \phi^{-1} \) represent the inverse of the \( \phi \)-function. Based on the SNR-incremental Gaussian channel in [12], it can be shown that \( \phi \) is monotonically increasing and thus its inverse exists.

\[C. \quad \text{Area Properties under the UG Condition}\]

From the estimation theory [18], the MMSE estimator of \( X(j) \) given \( \mathcal{Z}^{\text{SIDE}} \) and \( \mathcal{Y} \) is the \textit{a posteriori} mean \( E[X(j)|\mathcal{Z}^{\text{SIDE}},\mathcal{Y}] \). The related MMSE is given by
\[
\text{MMSE} = E\left[|X(j)|^2 - E[X(j)|\mathcal{Z}^{\text{SIDE}},\mathcal{Y}]|^2\right].
\] (20)

If the messages in (13) are UG, the MMSE in (20) is not a function of index \( j \), and so (14b) can be rewritten as
\[
p(x(j)|\mathcal{Z}^{\text{SIDE}},\mathcal{Y}) \propto \mathcal{CN}(\mathcal{Z}^{\text{SIDE}}(j), 1/\rho^{\text{SIDE}}(j)) p_x(x).
\] (21)

We can evaluate (20) based on the conditional PDF in (21). Let the result be written as a function of \( \rho^{\text{TL}} \):
\[
\text{MMSE} = \chi(\rho^{\text{TL}})\quad (22)
\]

where
\[
\chi(\rho) = E\left[|X - E[X|Z]|^2\right]\quad (23)
\]

with \( X \sim p_x(x) \) and \( Z \) given by (7).

Lemma 1 below is a straightforward extension of Lemmas 1 and 2 in [10] to arbitrary signaling constellations.

\[\text{Lemma 1: Under the UG condition in (15),}\]
\[
\int_{-\infty}^{\infty} \gamma(\rho^{\text{TL}}) d\rho^{\text{TL}} = J^{-1} H(\mathcal{X})\quad (24a)
\]

\[
\int_{-\infty}^{\infty} \gamma(\phi(\rho^{\text{SIDE}})) + \rho^{\text{SIDE}} d\rho^{\text{SIDE}} = J^{-1} H(\mathcal{X}|\mathcal{Y})\quad (24b)
\]

\[
\int_{-\infty}^{\infty} \gamma(\rho^{\text{EXT}} + \phi^{-1}(\rho^{\text{EXT}})) d\rho^{\text{EXT}} = J^{-1} I(\mathcal{X};\mathcal{Y})\quad (24c)
\]

The proof of Lemma 1 is similar to those of Lemmas 1 and 2 in [10]. We omit details due to lack of space. Lemma 1 provides the performance limit for a single processor. Next we consider iterative detection involving two concatenated processors.

\[\text{IV. ITERATIVE DETECTION AND AREA THEOREMS}\]

A generic complex-valued MIMO ISI channel is modeled by
\[
\mathcal{Y} = \mathbf{H}\mathcal{X} + \mathbf{N}
\] (25)

where \( \mathcal{Y} \) is the received signal, \( \mathbf{H} \) the known system transfer matrix, \( \mathcal{X} \) a coded signal vector generated by an forward-error-control encoder, and \( \mathbf{N} \sim \mathcal{CN}(0, \mathbf{\Sigma}) \).

A typical iterative receiver for the detection of \( \mathcal{X} \) in (25) consists of two concatenated processors: the elementary signal estimator (ESE) and the decoder (DEC). The ESE ignores the channel constraint only; the DEC ignores the channel constraint and handles the coding constraint only. Extrinsic messages are exchanged between the ESE and DEC. We use the subscripts "ESE" and "DEC" to represent the variables (or functions) related to the ESE and DEC, respectively. As illustrated in Fig. 3, the ESE output is used as the DEC input, i.e.
\[
\mathcal{X}^{\text{ESE}}(j) = \mathcal{X}^{\text{SIDE}}(j),\quad \text{for any } j.
\]

and similarly for the DEC output.

\[\text{A. Assumptions and Verifications}\]

We first examine the assumptions involved in Lemma 1 (since it plays a key role in our subsequent analysis).

The assumptions a)-c) stated at the beginning of Section III can be asymptotically ensured for both the ESE and DEC if the length of \( \mathcal{X} \) tends to infinity and random interleaving is applied.
in forming \( \bar{X} \). These assumptions have been widely used in analyzing turbo and LDPC codes [2][5], and is adopted as the fundamental assumptions in the sequel.

For the UG assumption in Section III.B, the Gaussian requirement can be approximately met using superposition coded modulation (SCM) [8], and the uniform requirement can be approximately met using linear precoding (LP) [9]. Due to space limitation, we omit the details on these issues.

In the sequel, we always assume that assumptions a)-c) and the UG assumption hold for both the ESE and DEC.

B. Evolution Analysis

The behavior of the ESE and DEC can be characterized by evolution analysis, as detailed below. From (18), we obtain

\[ \rho_{\text{ESE}}^{\text{SIDE}} = \phi_{\text{ESE}}(\rho_{\text{ESE}}^{\text{SIDE}}) \quad \text{and} \quad \rho_{\text{DEC}}^{\text{SIDE}} = \phi_{\text{DEC}}(\rho_{\text{DEC}}^{\text{SIDE}}). \]  

(26)

Denote by \( \rho_{\text{ESE}}^{\text{SIDE}}(1) \) the value of \( \rho_{\text{ESE}}^{\text{SIDE}} \) after the \( i \)th iteration. We start with an initial value of \( \rho_{\text{ESE}}^{\text{SIDE}}(0) = 0 \) (implying no information from the DEC). This process continues iteratively: after the first iteration, \( \rho_{\text{ESE}}^{\text{SIDE}}(1) = \phi_{\text{ESE}}(\rho_{\text{ESE}}^{\text{SIDE}}(0)) \); after the second iteration, \( \rho_{\text{ESE}}^{\text{SIDE}}(2) = \phi_{\text{ESE}}(\rho_{\text{ESE}}^{\text{SIDE}}(1)) \); and so on so forth.

The above iteration eventually converges to a fixed point \( \rho^* \geq 0 \) satisfying

\[ \rho^* = \phi_{\text{ESE}}(\rho^*) \quad \text{and} \quad \phi_{\text{DEC}}(\rho^*) \overset{t}{>}, \text{for} \ t \in [0, +\infty). \]  

(27a)

C. Curve Matching

Suppose that \( \rho_{\text{DEC}}^{\text{SIDE}} \) converges to infinity (i.e., \( \rho^* = +\infty \)) as the iterative process proceeds. We can then express (27a) as

\[ \phi_{\text{DEC}}(\rho(t)) \overset{t}{>}, \text{for} \ t \in [0, +\infty). \]  

(27b)

The uncertainty of \( X \) diminishes until \( X \) is perfectly determined, or in other words, the receiver performs error-free decoding.

Definition 4: An iterative receiver performs error-free decoding if

\[ \phi_{\text{ESE}}(\rho) > \phi_{\text{DEC}}(\rho), \text{ for any} \ \rho \in [0, +\infty). \]  

(28a)

Definition 5: The ESE and DEC are matched if

\[ \phi_{\text{ESE}}(\rho) = \phi_{\text{DEC}}(\rho), \text{ for any} \ \rho \in [0, +\infty). \]  

(28b)

D. Area Properties for Iterative APP Detection

The following analysis provides the performance limit for the iterative receiver in Fig.4. For the ESE, we obtain from (25c)

\[ \int_{0}^{\infty} \gamma(\rho_{\text{ESE}}^{\text{SIDE}}) + \phi^{-1}(\rho_{\text{ESE}}^{\text{SIDE}}) d\rho_{\text{ESE}}^{\text{SIDE}} = J^{\text{SIDE}} I(\bar{X} ; \bar{Y}). \]  

(29a)

Here the ESE does not take the coding constraint into consideration, and so \( I(\bar{X} ; \bar{Y}) \) is computed based on \( \bar{X} \sim p_{\text{ESE}}(\bar{X}) \) satisfying (12). For the DEC, as there is no channel observation available to the DEC, we have

\[ \int_{0}^{\infty} \gamma(\rho_{\text{DEC}}^{\text{SIDE}}) + \phi^{-1}(\rho_{\text{DEC}}^{\text{SIDE}}) d\rho_{\text{DEC}}^{\text{SIDE}} = J^{\text{SIDE}} H(\bar{X} \mid \bar{X} \in C) = R_{\text{C}}. \]  

(29b)

The last equality in (29b) is explained as follows. We combine the constellation information \( p_{\text{ESE}}(\bar{X}) \) with the coding constraints to form an overall signaling constellation over the \( J \)-dimensional space spanned by \( \bar{X} \) (with \( \bar{X} \) selected from the codebook \( C \)). In this case, the entropy \( H(\bar{X} \mid \bar{X} \in C) = J R_{\text{C}} \).

Note that the condition in (28b) can be equivalently written as

\[ \gamma(\rho) + \phi^{-1}(\rho) = \gamma(\phi_{\text{DEC}}(\rho) + \rho), \text{ for any} \ \rho \in [0, +\infty). \]  

(30)

Combining (29) and (30), we have the following result.

Theorem 1: If the ESE and DEC are matched APP processors with UG outputs, the information rate \( R_{\text{C}} \) (per entry of \( \bar{X} \)) of the iterative system in Fig.4 satisfies

\[ R_{\text{C}} = J^{\text{SIDE}} I(\bar{X} ; \bar{Y}). \]  

(31)

A special case of Theorem 1 was proven in [5, Theorem 2] for a binary \( \bar{X} \) over a memoryless AWGN channel. Here we extend Theorem 2 in [5] to any coded MIMO system in (26) with any signaling constellation.

E. Approximate APP Processing

The UG and APP assumptions in Theorem 1 are rather strong restrictions. From now on, we gradually relax these restrictions. We focus on the relaxation of the APP requirement for the ESE in this subsection. Recall from (13c) that an optimal processor calculates the extrinsic message as

\[ \tilde{X}^\gamma(j) = p(\tilde{Y}) x(j). \]  

(32)

The complexity involved in (32) may be excessively high when the channel in (26) is relatively complicated.

To reduce complexity, we consider a suboptimal processor that generates an approximate extrinsic message for \( X(j) \) as

\[ \tilde{X}^\gamma(j) = p(\tilde{Y}) x(j). \]  

(33a)

where \( U = \tilde{U} \) is a linear or non-linear transform of \( \tilde{Y} \) and \( Z^\text{SIDE} \backslash Z^\text{SIDE}(j) \) (e.g., \( U \) is a subset of \( \tilde{Y}, Z^\text{SIDE} \backslash Z^\text{SIDE}(j) \)). This transform may cause performance loss. We wish that (33a) can be easily calculated, while the related performance loss is acceptable. Similarly to (14a), we calculate the total message as

\[ \tilde{X}^\gamma(j) = \tilde{X}^\gamma(j) \tilde{X}^\gamma(j). \]  

(33b)

We henceforth refer to the processor calculating the messages based on (33) as an approximate APP processor.

Assume that the above suboptimal method still leads to UG messages. Similar to (18), we use an SNR transfer function

\[ \rho_{\text{ESE}}^{\text{SIDE}} = \phi(\rho_{\text{ESE}}^{\text{SIDE}}) \]  

(34)

to characterize an approximate APP processor. From the data processing inequality [10], we have

\[ \phi(\rho_{\text{ESE}}^{\text{SIDE}}) \leq \phi(\rho_{\text{SIDE}}^{\text{SIDE}}) \]  

(35)

where \( \phi(\cdot) \) is the transfer function for the corresponding APP processor defined in (18). Assume that \( U \) in (33a) is properly chosen such that \( \phi(\cdot) \) is monotonically increasing, and hence the inverse of \( \phi(\cdot) \) exists. Then, similar to (19), we still have the following relationship

\[ \rho_{\text{ESE}}^{\text{SIDE}} = \phi(\rho_{\text{SIDE}}^{\text{SIDE}}) + \rho_{\text{ESE}}^{\text{SIDE}} = \rho_{\text{SIDE}}^{\text{SIDE}} + \phi^{-1}(\rho_{\text{ESE}}^{\text{SIDE}}). \]  

(36)

The MSE related to the approximate method is defined as

\[ M_{SE} = \mathbb{E}\left[|X(j) - \bar{X}(j)|^2 + Z^\text{SIDE}(j), Z^\text{SIDE}(j)|^2\right]. \]  

(37)

Similarly to (23a), the MSE can also be expressed by

\[ M_{SE} = \gamma(\rho_{\text{ESE}}^{\text{SIDE}}). \]  

(38)

Substituting (36) into (38), we further have

2 An equivalent condition of (28a) is \( \phi_{\text{ESE}}(t) < \phi_{\text{DEC}}(t) \), for any \( t \in [0, +\infty) \), which shows the symmetry of the ESE and DEC.
MSE = \gamma \hat{\phi}(\rho_{\text{SIDE}}) + \rho_{\text{SIDE}} = \gamma (\rho_{\text{EXT}} + \hat{\phi}^{-1}(\rho_{\text{EXT}})). \quad (39)

Now consider both the ESE and DEC. The subscripts are added again to distinguish the variables related to the ESE and DEC. We extend the area property for an iterative receiver by assuming that the ESE is approximately APP. As an analogy to (30), we say that the ESE and DEC are matched if
\gamma (\rho + \hat{\phi}(\rho)) = \gamma (\hat{\phi}_{\text{ESE}}(\rho) + \rho), \quad \text{for any } \rho \in [0, +\infty). \quad (40)
Combining (29b) and (40), we have the following result.

**Theorem 2:** Let the ESE be an approximate APP processor, the DEC be a matched APP processor, both with UG outputs. Then, the information rate \( R_C \) of the system in Fig.4 satisfies
\[
R_C = \int_{0}^{\infty} (\gamma \rho_{\text{ESE}} + \hat{\phi}_{\text{ESE}}(\rho_{\text{ESE}})) \text{d}\rho_{\text{ESE}}.
\]
Using (29b), we can alternatively express \( R_C \) in (41) as
\[
R_C = J^{-1} I(\tilde{X};\tilde{Y}) - \Delta_{\text{ESE}} \quad (42a)
\]
where
\[
\Delta_{\text{ESE}} = \int_{0}^{\infty} (\gamma \rho_{\text{ESE}} + \hat{\phi}_{\text{ESE}}(\rho_{\text{ESE}})) - \gamma \rho_{\text{EXT}} + \hat{\phi}_{\text{ESE}}(\rho_{\text{ESE}}) \text{d}\rho_{\text{ESE}}. \quad (42b)
\]
Note that the DEC can also be an approximate APP processor, which results in a further reduction of the rate in (42a). The derivation is similar and thus is omitted here.

**V. NUMERICAL RESULTS**

Comparing (42a) with (31), we see that there is a performance loss due to the sub-optimality of the ESE. We now examine this gap using an example.

Consider a randomly generated 2x2 MIMO 3-tap ISI channel with the tap coefficients
\[
\begin{bmatrix}
0.5339 + j0.5395 & -0.4245 + j0.0648 \\
-0.3347 - j0.3727 & -0.4672 - j0.2420
\end{bmatrix} \quad (43a)
\[
\begin{bmatrix}
0.0582 - j0.2706 & 0.1525 - j0.7565 \\
-0.4968 - j0.1543 & -0.5243 - j0.5915
\end{bmatrix} \quad (43b)
\]
and
\[
\begin{bmatrix}
-0.5262 - j0.2654 & -0.3714 - j0.2865 \\
0.6721 - j0.1635 & 0.1607 - j0.2695
\end{bmatrix} \quad (43c)
\]
The linear precoding (LP) technique proposed in [9] is applied at the transmitter. The LP technique ensures that the UG condition is satisfied for the ESE. Another advantage of the LP technique is to achieve the water-filling gain by properly allocating power among the eigen-modes of the MIMO channel. Interested readers are referred to [9] for details. Iterative LMMSE detection (cf., [9]) is applied at the receiver side. The signaling constellation is the conventional 16-QAM that can be treated as two-QPSK-layer superposition coded modulation (SCM) with the first layer scaled by the coefficient 0.6325 and the second by 0.31625.

Fig.4 shows that the gap between the water-filling capacity and the achievable rate of the LP-LMMSE scheme is not significant in the interested rate region for such a 2x2 MIMO system (say, rate \leq 4 bits per channel use). Note that this gap is well expected, as predicted by Theorem 2.

We next show that the performance curve of LP-LMMSE in Fig.4 is approachable using SCM. Two irregular LDPC codes are adopted for the scheme. The details on code-design issues are beyond the scope of this paper and will be discussed in a separate paper in preparation. The first QPSK layer employs an LDPC code with the degree distributions \( \{ A(x) = 0.3495x + 0.2142x^2 + 0.1165x^4 + 0.0857x^6 + 0.3313x^8 + 0.1010x^{10} \} \) \( \rho(x) = x^1 \), and the other layer employs the other LDPC code with \( \{ A(x) = 0.3907x + 0.1854x^2 + 0.0968x^4 + 0.1877x^6 + 0.1394x^8 \} \) \( \rho(x) = x^2 \). The two LDPC-coded layers are properly scaled and superimposed to form a coded vector. Then, the LP technique in [9] is applied to yield the channel input \( \tilde{X} \). The system throughput is 4 bits per channel use.

From Fig.5, the design threshold is \( \text{SNR} = 6.7 \text{ dB} \), which is 1.1 dB away from the water-filling capacity, 0.5 dB away from the performance limit of the scheme. At code length = 10^5, the simulated system performance is 7.0 dB at BER = 10^{-4}, which is only 0.3 dB away from the design threshold. This demonstrates that the suboptimal LP-ILMMSE scheme can nearly achieve the optimal performance at a low-to-median rate.
VI. CONCLUSIONS

In this paper, we have established the area theorems for a coded MIMO system with message-passing decoding. Numerical results show that, with proper code design, a practical LDPC-coded MIMO scheme can closely approach the performance limit predicted by our analysis.

ACKNOWLEDGEMENT

This work has been performed in the framework of the ICT project ICT-248894 WHERE2, which is partly funded by the European Union.

REFERENCES


