

Energy-Spreading-Transform Based MIMO Systems: Iterative Equalization, Evolution Analysis, and Precoder Optimization

Xiaojun Yuan, *Member, IEEE*, Junjie Ma, and Li Ping, *Fellow, IEEE*

Abstract—In this paper, we develop a novel iterative equalization algorithm for energy-spreading-transform (EST) based multiple-input multiple-output (MIMO) systems. We show that the proposed scheme significantly outperforms the existing non-linear MIMO equalizers in various system setups. We further investigate the precoder design based on the signal-to-interference-plus-noise-ratio (SINR) variance evolution technique, so as to exploit the available channel state information at the transmitter (CSIT). We derive the optimal precoding directions, and show that the precoder optimization then boils down to a simple power allocation problem that is solvable using convex programming. Numerical results demonstrate that the optimized precoder can achieve a significant power gain, as compared with the non-optimized scheme.

Index Terms—Energy spreading transform, MIMO, iterative equalization, SINR-variance transfer chart, precoder design.

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) techniques [1], [2] have been extensively studied in wireless communication in the past decade. A critical issue for MIMO communication is to mitigate cross-antenna interference at the receiver. Various equalization techniques, e.g., in [3]–[11], have been developed for this purpose. Particularly, turbo equalization [10], [11], utilizing the decision feedback of a forward-error-control (FEC) decoder for efficient interference cancellation, proves remarkably successful. However, this requires iterative processing of the equalizer and the decoder, which considerably increases computational complexity and complicates the hardware design. As such, state-of-the-art industrial standards for MIMO techniques [12], [13] suggest sequential processing of the equalization and decoding operations, instead of turbo

Manuscript received September 9, 2013; revised February 22, 2014 and July 1, 2014; accepted July 4, 2014. Date of publication July 16, 2014; date of current version September 8, 2014. The work of J. Ma and Li Ping were supported by the Research Grant Council, Hong Kong, under Project CityU 118013. This paper was presented in part at the IEEE Wireless Communications and Networking Conference Shanghai, China, April 2013, and at the IEEE International Conference on Communications Budapest, Hungary, June 2013. The associate editor coordinating the review of this paper and approving it for publication was S. Bhashyam.

X. Yuan is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 200031, China (e-mail: yuanxj@shanghaitech.edu.cn).

J. Ma and L. Ping are with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong (e-mail: junjiema2-c@my.cityu.edu.hk; eeliping@cityu.edu.hk).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TWC.2014.2339839

equalization. It is then of great practical and theoretical interest to design equalizers without the help of the decoder, which is the focus of this paper.

Conventional approaches for MIMO equalization include maximum likelihood (ML) estimation, linear minimum mean-square error (LMMSE) estimation, and zero-forcing (ZF), etc. [14]. The optimal ML estimation is usually not desirable due to its high complexity. ZF and LMMSE estimators can reduce the complexity to a polynomial order, but usually suffer from significant performance degradation, as incurred by noise amplification and residual interference.

Various non-linear equalizers have been proposed in the literature. Among them, a promising approach is decision feedback equalization (DFE) [3], [4], [15], [16], in which the residual interference of a linear estimator is suppressed without significantly increasing computational complexity. Particularly, linear MMSE filtering based iterative soft-decision interference cancellation (MMSE-ISDIC) [17], which is equivalent to the probability data association (PDA) algorithm [18], [19], has been widely used for its outstanding performance and affordable complexity.

Linear precoding techniques, by exploiting the available channel state information at the transmitter (CSIT), can be used together with DFE to improve system performance. For example, the geometric-mean-decomposition (GMD) based linear precoder [5] is shown to be optimal in the sense of minimizing bit-error rate (BER) for ZF-DFE; and the uniform-channel-decomposition (UCD) based linear precoder [6] was shown to be optimal for LMMSE-DFE. The key idea in these approaches is to linearly transform the physical channel matrix into an upper-triangular matrix with equal channel gains in the diagonal. Then, the information streams are detected layer-by-layer using successive interference cancellation (SIC). However, it is known that SIC suffers from error propagation (as the errors in the upper layers propagate to the lower layers). To overcome this deficiency, the authors in [7] have recently proposed a root-mean-square decomposition (RMSD) based *energy-spreading-transform* (EST) approach for MIMO communications. The RMSD-EST scheme in [7] employs dispersive transmission at the transmitter, i.e., each data symbol is spread over space and time by an EST precoding matrix, so as to harvest the diversity provided by the MIMO channel; a low-complexity MMSE-ISDIC algorithm is applied at the receiver for efficient equalization. It was shown in [20] that the proposed EST scheme can achieve full diversity and multiplexing gains, and considerably outperforms the GMD and UCD schemes.

vector \mathbf{x} of length $J = NL$ to the receiver. Each entry of \mathbf{x} is equi-probably taken over a finite constellation $\mathcal{S} = \{s_1, s_2, \dots, s_{|\mathcal{S}|}\}$, where $|\mathcal{S}|$ represents the cardinality of \mathcal{S} . Without loss of generality, we assume that \mathcal{S} satisfies the following property:

$$\sum_{s_i \in \mathcal{S}} s_i = 0 \text{ and } \frac{1}{|\mathcal{S}|} \sum_{s_i \in \mathcal{S}} |s_i|^2 = 1. \quad (3)$$

Vector \mathbf{x} passes through a discrete Fourier transform (DFT) operator to yield

$$\mathbf{y} = \mathbf{F}\mathbf{x} \quad (4)$$

where \mathbf{F} is a J -by- J normalized DFT matrix, with the (m, n) th entry given by

$$F_{m,n} = \frac{1}{\sqrt{J}} \exp\left(-i \frac{2\pi mn}{J}\right), \text{ for } m, n = 0, \dots, J-1. \quad (5)$$

Let Π be a randomly chosen interleaver. The vector \mathbf{y} passes through Π , yielding the channel input vector

$$\mathbf{z} = \Pi\mathbf{y} = \mathbf{E}\mathbf{x} \quad (6a)$$

where

$$\mathbf{E} \equiv \Pi\mathbf{F} \quad (6b)$$

is the energy-spreading transform proposed in [7] and [20], and Π the J -by- J permutation matrix specified by Π .

The complexity of the transmitter is dominated by (4), which can be efficiently implemented using the renowned fast Fourier transform (FFT) algorithm with complexity $\mathcal{O}(J \log J)$.

III. PROPOSED ITERATIVE EQUALIZATION ALGORITHM

In what follows, we focus on the receiver design. We provide a new interpretation of the decision feedback equalization from the perspective of BP. Based on that, we develop a novel iterative equalizer that can significantly outperform the existing approaches in the literature.

A. Proposed Iterative Equalizer

The equalizer aims to detect \mathbf{x} based on the channel observation \mathbf{r} . The diagram of the proposed iterative equalizer is illustrated in the lower half of Fig. 1. For convenience, we divide the overall iterative equalizer into two virtual modules, namely the *estimation module* and the *demodulation module*. The estimation module consists of the flowchart starting from $\vec{\mathbf{z}}$ and ending at $\vec{\mathbf{x}}$, including the estimator and the two de-interleaver and IFFT blocks; the demodulation module consists of the flowchart starting from $\vec{\mathbf{x}}$ and ending at $\vec{\mathbf{z}}$, including the demodulator and the two FFT and interleaver blocks.

For the estimation module, the estimator estimates \mathbf{z} based on the channel observation \mathbf{r} and the input vector $\vec{\mathbf{z}}$, yielding an output vector $\vec{\mathbf{z}}$. Then, $\vec{\mathbf{z}}$ is de-interleaved by Π^{-1} to form the vector $\vec{\mathbf{y}}$. The output $\vec{\mathbf{x}}$ of the estimation module is obtained by calculating the extrinsic distributions, using the IFFT of $\vec{\mathbf{y}}$ as the *a posteriori* and the IFFT of the de-interleaved version of $\vec{\mathbf{z}}$ as the *a priori*.

For the demodulation module, the demodulator makes a soft decision on \mathbf{x} , denoted by $\vec{\mathbf{x}}$, based on the input $\vec{\mathbf{x}}$ and the fact that each entry of \mathbf{x} is constrained on the discrete constellation \mathcal{S} . Then, the output of the demodulation module, denoted by $\vec{\mathbf{z}}$, is obtained by calculating the extrinsic distributions (in which the interleaved FFT of $\vec{\mathbf{x}}$ is used as the *a posteriori* and the interleaved FFT of $\vec{\mathbf{x}}$ as the *a priori*).

The detailed iterative algorithm is described as follows. Let T_{max} be the number of iterations.

Algorithm 1

Initialization: Set $\vec{\mathbf{z}} = \mathbf{0}$ and $v = 1$ and the counter $t = 0$.

Step 1: With $\vec{\mathbf{z}}$ and v , calculate $\vec{\mathbf{x}}$ and u as follows.

a) Noting $\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}$, calculate the MMSE estimator and the MMSE matrix respectively as

$$\vec{\mathbf{z}} = \vec{\mathbf{z}} + v\tilde{\mathbf{H}}^H(v\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \sigma^2\mathbf{I})^{-1}(\mathbf{r} - \tilde{\mathbf{H}}\vec{\mathbf{z}})$$

$$\tilde{\mathbf{V}} = (\sigma^{-2}\tilde{\mathbf{H}}^H\tilde{\mathbf{H}} + v^{-1}\mathbf{I})^{-1}$$

b) Calculate the average variance as $w_x = J^{-1}\text{tr}\{\tilde{\mathbf{V}}\}$;

c) Calculate $\vec{\mathbf{x}}$ and u as

$$u^{-1} = w_x^{-1} - v^{-1}$$

$$\frac{\vec{x}_i}{u} = \frac{(\mathbf{E}^H\vec{\mathbf{z}})_i}{w_x} - \frac{(\mathbf{E}^H\vec{\mathbf{z}})_i}{v}, \text{ for } i = 1, \dots, J.$$

Step 2: With $\vec{\mathbf{x}}$ and u given in Step 1, calculate the following to update $\vec{\mathbf{z}}$ and v .

a) Calculate the means and variances respectively as

$$\vec{x}_i = \mathbb{E}[x_i | \vec{x}_i, x_i \in \mathcal{S}] = \frac{\sum_{s_j \in \mathcal{S}} s_j p(\vec{x}_i | x_i = s_j)}{\sum_{s_j \in \mathcal{S}} p(\vec{x}_i | x_i = s_j)},$$

$$t_i = \mathbb{E}\left[|x_i - \vec{x}_i|^2 | \vec{x}_i, x_i \in \mathcal{S}\right] \\ = \frac{\sum_{s_j \in \mathcal{S}} |s_j - \vec{x}_i|^2 p(\vec{x}_i | x_i = s_j)}{\sum_{s_j \in \mathcal{S}} p(\vec{x}_i | x_i = s_j)}$$

where

$$p(\vec{x}_i | x_i) = \frac{1}{\pi u} \exp\left(-\frac{|x_i - \vec{x}_i|^2}{u}\right).$$

b) Calculate the average variance as $w_y = J^{-1} \sum_{i=1}^J t_i$;

c) Calculate $\vec{\mathbf{z}}$ and v as

$$v^{-1} = w_y^{-1} - u^{-1}$$

$$\vec{z}_i = \frac{(\mathbf{E}\vec{\mathbf{x}})_i}{w_y} - \frac{(\mathbf{E}\vec{\mathbf{x}})_i}{u}, \text{ for } i = 1, \dots, J.$$

Step 3: Set $t = t + 1$. If $t = T_{max}$, make a hard decision of x_i using \vec{x}_i for each i and stop; otherwise, go to Step 1.

B. Estimation Operations

In the remainder of this section, we explain the motivations behind Algorithm 1. We start with the estimation operation (i.e., Step 1) in which the estimator estimates \mathbf{z} given $\vec{\mathbf{z}}$ and \mathbf{r} .¹ A fundamental assumption taken by the estimator is given below.

Assumption 1: The input vector $\vec{\mathbf{z}}$ is modeled as

$$\vec{\mathbf{z}} = \mathbf{z} + \mathbf{n} \quad (7)$$

where \mathbf{n} is a J -by-1 noise vector with the entries independently drawn from $\mathcal{CN}(0, v)$.

At the beginning of the iterative process, we initialize $\vec{\mathbf{z}}$ and v to $\mathbf{0}$ and 1, respectively, implying that there is no information available from the demodulation module. With (7), this initialization is equivalent to say that \mathbf{z} given $\vec{\mathbf{z}}$ is distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{I})$, which is in agreement with the following facts: first, from (3) and (6a), the initial mean and covariance of \mathbf{z} is indeed $\mathbf{0}$ and \mathbf{I} ; second, from (6a) and the central limit theorem, the entries of \mathbf{z} are Gaussian distributed for a sufficiently large J , no matter what signal constellation is used for generating \mathbf{x} . In the iterative process, $\vec{\mathbf{z}}$ and v are obtained from the output of the demodulation module. We defer the detailed justification of Assumption 1 in the iterative process to Subsection C.

Under Assumption 1, $\vec{\mathbf{z}} \rightarrow \mathbf{z} \rightarrow \mathbf{r}$ forms a Markov chain, i.e., the conditional probability density function (PDF) of $(\vec{\mathbf{z}}, \mathbf{r})$ given \mathbf{z} can be factorized as

$$p(\vec{\mathbf{z}}, \mathbf{r} | \mathbf{z}) = p(\vec{\mathbf{z}} | \mathbf{z}) p(\mathbf{r} | \mathbf{z}), \quad (8)$$

where the conditional PDF $p(\vec{\mathbf{z}} | \mathbf{z})$ is determined by (7), and $p(\mathbf{r} | \mathbf{z})$ is determined by (1). From (7), the likelihood function $p(\vec{\mathbf{z}} | \mathbf{z})$ specifies a CSCN distribution with mean $\vec{\mathbf{z}}$ and covariance $v\mathbf{I}$. This implies that estimating \mathbf{z} given $\vec{\mathbf{z}}$ and \mathbf{r} (without any other knowledge of \mathbf{z}) is equivalent to estimating \mathbf{z} given \mathbf{r} with \mathbf{z} *a priori* distributed as $\mathcal{CN}(\vec{\mathbf{z}}, v\mathbf{I})$.² The latter, by noting the linear system model in (1), is a standard LMMSE estimation problem [14]. The *a posteriori* distribution of \mathbf{z} is still CSCN, with the *a posteriori* mean given by (c.f., Theorem 12.1 of [14])

$$\hat{\mathbf{z}} = \vec{\mathbf{z}} + v \tilde{\mathbf{H}}^H (v \tilde{\mathbf{H}} \tilde{\mathbf{H}}^H + \sigma^2 \mathbf{I})^{-1} (\mathbf{r} - \tilde{\mathbf{H}} \vec{\mathbf{z}}) \quad (9)$$

and the *a posteriori* covariance matrix given by

$$\tilde{\mathbf{V}} = \mathbb{E}[(\mathbf{z} - \hat{\mathbf{z}})(\mathbf{z} - \hat{\mathbf{z}})^H] = (\sigma^{-2} \tilde{\mathbf{H}}^H \tilde{\mathbf{H}} + v^{-1} \mathbf{I})^{-1} \quad (10)$$

where $\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}$. In the above, the expectation is taken over the joint distribution of \mathbf{z} and \mathbf{r} given by $p(\mathbf{z})p(\mathbf{r} | \mathbf{z})$, with $p(\mathbf{z}) = \mathcal{CN}(\vec{\mathbf{z}}, v\mathbf{I})$ and $p(\mathbf{r} | \mathbf{z})$ determined by (2). It is also worth mentioning that the above LMMSE estimation approach has been widely used in turbo equalization [11], [27].

¹Here no other information on \mathbf{z} is available to the estimator. That is, for the estimator, it is only known that the entries of \mathbf{z} are arbitrary complex numbers.

²We give a heuristic explanation as follows. Suppose that \mathbf{z} is to be estimated based on the observed $\vec{\mathbf{z}}$ and no other prior knowledge of \mathbf{z} is available. From (7) we have

$$p(\vec{\mathbf{z}} | \mathbf{z}) = \frac{1}{\pi v} e^{-\frac{|\vec{\mathbf{z}} - \mathbf{z}|^2}{v}} = \mathcal{CN}(\vec{\mathbf{z}}, v\mathbf{I}).$$

Then, as no other knowledge of \mathbf{z} is available, we say that the information of \mathbf{z} provided by observing $\vec{\mathbf{z}}$ is $\mathbf{z} \sim \mathcal{CN}(\vec{\mathbf{z}}, v\mathbf{I})$. Similar conversions between observations and distributions will be frequently used in this paper. For example, in Assumption 2, (15) is equivalent to say that the information of \mathbf{x} provided by observing $\hat{\mathbf{x}} \sim \mathcal{CN}(\hat{\mathbf{x}}, w_x \mathbf{I})$.

It is required in the BP principle [21], [22] that extrinsic distributions, instead of *a posteriori* distributions, are delivered in message passing. In what follows, we describe how to calculate the extrinsic distributions.

As shown in Fig. 1, $\vec{\mathbf{z}}$ is de-interleaved by Π^{-1} to form $\overleftarrow{\mathbf{z}}$. From (10), $\tilde{\mathbf{V}}$ is block-diagonal by noting $\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}$. This implies that any two entries of $\vec{\mathbf{z}}$ from two different channel uses are independent of each other. That is, the entries of $\vec{\mathbf{z}}$ are only locally correlated. Then, after randomly interleaving $\vec{\mathbf{z}}$ of a sufficiently large length, such local correlations can be ignored in the subsequent processing, or in other words, the entries of given $\overleftarrow{\mathbf{y}}$ can be approximated as independent of each other. Moreover, it is well known that the distortion of an LMMSE estimator can be approximated as Gaussian distributed [11], [14], [23]. Thus, the conditional distribution of $\mathbf{y}(= \Pi^{-1} \mathbf{z})$ given $\overleftarrow{\mathbf{y}}$ is given by $\mathcal{CN}(\overleftarrow{\mathbf{y}}, \Pi^{-1} \mathbf{V}_{\text{diag}} \Pi)$, or equivalently, $\overleftarrow{\mathbf{y}}$ is modeled as

$$\overleftarrow{\mathbf{y}} = \mathbf{y} + \overleftarrow{\mathbf{w}} \quad (11)$$

where $\overleftarrow{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}, \Pi^{-1} \mathbf{V}_{\text{diag}} \Pi)$ is independent of \mathbf{y} , and \mathbf{V}_{diag} is the diagonal matrix obtained by setting the off-diagonal elements of $\tilde{\mathbf{V}}$ to zeros. The IFFT of $\overleftarrow{\mathbf{y}}$ is given by

$$\hat{\mathbf{x}} \equiv \mathbf{F}^H \overleftarrow{\mathbf{y}} = \mathbf{x} + \mathbf{F}^H \overleftarrow{\mathbf{w}}. \quad (12)$$

We have two observations on $\mathbf{F}^H \overleftarrow{\mathbf{w}}$: first, according to the central limit theorem, the entries of $\mathbf{F}^H \overleftarrow{\mathbf{w}}$ are approximately Gaussian distributed; second, the covariance of $\mathbf{F}^H \overleftarrow{\mathbf{w}}$ is given by

$$\mathbb{E}[\mathbf{F}^H \overleftarrow{\mathbf{w}} (\mathbf{F}^H \overleftarrow{\mathbf{w}})^H] = \mathbf{F}^H \Pi^{-1} \mathbf{V}_{\text{diag}} \Pi \mathbf{F} \rightarrow w_x \mathbf{I}, \text{ as } J \rightarrow \infty \quad (13)$$

where the last step follows from Theorem 3 in [20], and

$$w_x = J^{-1} \text{tr}\{\tilde{\mathbf{V}}\}. \quad (14)$$

With the above reasoning, we make the following assumption.

Assumption 2: The vector $\hat{\mathbf{x}}$ (c.f. (12)) is modeled as

$$\hat{\mathbf{x}} = \mathbf{x} + \hat{\mathbf{n}} \quad (15)$$

where $\hat{\mathbf{n}}$ is a J -by-1 vector with the entries independently drawn from $\mathcal{CN}(0, w_x)$.

We are now ready to calculate the extrinsic distributions. The extrinsic distribution of each x_i is obtained by excluding the contribution of the *a priori* distribution from the *a posteriori* distribution. From (15), the *a posteriori* distribution of each x_i is given by $\mathcal{CN}(\hat{x}_i, w_x)$; from (7), \mathbf{z} is *a priori* distributed as $\mathcal{CN}(\vec{\mathbf{z}}, v\mathbf{I})$. Thus, by noting $\mathbf{z} = \mathbf{E}\mathbf{x}$ (c.f., (6b)), the *a priori* distribution of each x_i is given by $\mathcal{CN}((\mathbf{E}^H \vec{\mathbf{z}})_i, v)$, where $(\mathbf{E}^H \vec{\mathbf{z}})_i$ represents the i th entry of $\mathbf{E}^H \vec{\mathbf{z}}$. Thus, both the *a priori* and *a posteriori* distributions of \mathbf{x} are CSCN. Based on the Gaussian-message combining rule (c.f., (54) and (55) in [28]), the extrinsic message is also CSCN, with the mean \hat{x}_i and variance u of each x_i respectively given by

$$u^{-1} = w_x^{-1} - v^{-1} \quad (16a)$$

$$\frac{\hat{x}_i}{u} = \frac{\hat{x}_i}{w_x} - \frac{(\mathbf{E}^H \vec{\mathbf{z}})_i}{v} = \frac{(\mathbf{E}^H \overleftarrow{\mathbf{z}})_i}{w_x} - \frac{(\mathbf{E}^H \overleftarrow{\mathbf{z}})_i}{v}. \quad (16b)$$

Then, the output of the estimation part is given by $\hat{\vec{x}} = [\hat{x}_1, \dots, \hat{x}_J]^T$.

We further model $\hat{\vec{x}}$ as follows. From Assumption 1, we represent the *a priori* mean vector as $\mathbf{E}^H \vec{z} = \mathbf{x} + \mathbf{F}^H \mathbf{\Pi}^{-1} \vec{\mathbf{n}}$, where $\mathbf{F}^H \mathbf{\Pi}^{-1} \vec{\mathbf{n}} \sim \mathcal{CN}(\mathbf{0}, v\mathbf{I})$. Thus, the entries of $\mathbf{E}^H \vec{z}$ given \mathbf{x} are conditionally independent and identically distributed (i.i.d.). Similarly, from Assumption 2, the entries of $\hat{\vec{x}}$ given \mathbf{x} are conditionally i.i.d. Therefore, from (16b), we conclude that the entries of $\hat{\vec{x}}$ given \mathbf{x} are conditionally i.i.d., and thus can be modeled as

$$\hat{\vec{x}} = \mathbf{x} + \vec{\mathbf{n}} \quad (17)$$

where the entries of $\vec{\mathbf{n}}$ are independently drawn from $\mathcal{CN}(0, u)$.

C. Demodulation Operations

We now consider the operations in the demodulation module illustrated in Fig. 1 (which corresponds to Step 2 in Algorithm 1). The demodulator estimates \mathbf{x} given $\hat{\vec{x}}$, together with the knowledge that the entries of \mathbf{x} are equi-probably taken over the constellation \mathcal{S} . Given $\hat{\vec{x}}$ in (17), the entries of \mathbf{x} are conditionally independent of each other, with the conditional mean and variance of each x_i respectively given by

$$\bar{x}_i = \mathbb{E}[x_i | \hat{x}_i, x_i \in \mathcal{S}] = \frac{\sum_{s_j \in \mathcal{S}} s_j p(\hat{x}_i | x_i = s_j)}{\sum_{s_j \in \mathcal{S}} p(\hat{x}_i | x_i = s_j)}, \quad (18a)$$

$$t_i = \mathbb{E} [|x_i - \bar{x}_i|^2 | \hat{x}_i, x_i \in \mathcal{S}] = \frac{\sum_{s_j \in \mathcal{S}} |s_j - \bar{x}_i|^2 p(\hat{x}_i | x_i = s_j)}{\sum_{s_j \in \mathcal{S}} p(\hat{x}_i | x_i = s_j)} \quad (18b)$$

where $p(\hat{x}_i | x_i)$ is obtained from (17) as

$$p(\hat{x}_i | x_i) = \frac{1}{\pi u} \exp\left(-\frac{|x_i - \hat{x}_i|^2}{u}\right)$$

Denote $\vec{\mathbf{x}} = [\vec{x}_1, \vec{x}_2, \dots, \vec{x}_J]^T$. We further model $\vec{\mathbf{x}}$ as

$$\vec{\mathbf{x}} = \mathbf{x} + \vec{\mathbf{w}} \quad (19)$$

where \mathbf{x} and $\vec{\mathbf{w}}$ are independent, and the i th entry of $\vec{\mathbf{w}}$ are independently drawn from a certain distribution $p(\vec{w}_i)$ with zero mean and variance t_i . Note that here we are not interested in the exact expression of $p(\vec{w}_i)$, though they are uniquely determined by (18a). With (19), we can express the *a posteriori* mean of the demodulation part as

$$\vec{\mathbf{y}} \equiv \mathbf{F} \vec{\mathbf{x}} = \mathbf{y} + \mathbf{F} \vec{\mathbf{w}}$$

where the covariance matrix of $\mathbf{F} \vec{\mathbf{w}}$ is given by

$$\mathbf{U} = \mathbf{F} \cdot \text{diag}\{t_1, t_2, \dots, t_J\} \cdot \mathbf{F}^H.$$

From Theorem 3 in [20], \mathbf{U} converges to $w_y \mathbf{I}$ as J tends to infinity, with w_y defined as

$$w_y \equiv \frac{1}{J} \sum_{k=1}^J t_k. \quad (20)$$

Further, from the central limit theorem, the entries of $\mathbf{F} \vec{\mathbf{w}}$ are approximately Gaussian distributed for a sufficiently large J . Therefore, we make the following assumption.

Assumption 3: The vector $\hat{\vec{z}} \equiv \mathbf{E} \hat{\vec{x}}$ is modeled as

$$\hat{\vec{z}} = \mathbf{z} + \hat{\mathbf{w}}$$

where $\hat{\mathbf{w}}$ is an J -by-1 vector with the entries independently drawn from $\mathcal{CN}(0, w_y)$.

The next step is to calculate the extrinsic messages at the output of the FFT modules. From Assumption 3, each z_i is *a posteriori* distributed as $\mathcal{CN}(\hat{z}_i, w_y)$; by noting that $\mathbf{z} = \mathbf{E} \mathbf{x}$ and that \mathbf{x} is *a priori* distributed as $\mathcal{CN}(\vec{\mathbf{x}}, u\mathbf{I})$, each z_i is *a priori* distributed as $\mathcal{CN}((\mathbf{E} \vec{\mathbf{x}})_i, u)$. Thus, similar to (16), the extrinsic mean and variance of each z_i , denoted by \vec{z}_i and v , respectively, satisfy

$$v^{-1} = w_y^{-1} - u^{-1} \quad (21a)$$

$$\frac{\vec{z}_i}{v} = \frac{\hat{z}_i}{w_y} - \frac{(\mathbf{E} \vec{\mathbf{x}})_i}{u} = \frac{(\mathbf{E} \vec{\mathbf{x}})_i}{w_y} - \frac{(\mathbf{E} \vec{\mathbf{x}})_i}{u}. \quad (21b)$$

The output of the demodulation module is given by $\vec{\mathbf{z}} = [\vec{z}_1, \dots, \vec{z}_J]^T$. $\vec{\mathbf{z}}$ is then treated as the input of the estimator in the next round of iteration.

We now justify Assumption 1 (in the second iterative step or after) as follows. From (17), we see that $\mathbf{E} \hat{\vec{x}}$ given \mathbf{z} is conditionally i.i.d.; from Assumption 3, the entries of $\hat{\vec{z}}$ given \mathbf{z} are also conditionally i.i.d. Therefore, from (21b), we see that the entries of $\vec{\mathbf{z}}$ given \mathbf{z} are conditionally i.i.d., with the distribution given by $\mathcal{CN}(\mathbf{0}, v\mathbf{I})$. We then arrive at the signal model (7) in Assumption 1.

D. Further Discussions

The complexity of Algorithm 1 is briefly discussed as follows. The complexity of the estimation module is dominated by the LMMSE estimation in (9). With $\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}$, we obtain

$$(v\tilde{\mathbf{H}}^H \tilde{\mathbf{H}} + \sigma^2 \mathbf{I})^{-1} = \mathbf{I} \otimes (v\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{I})^{-1}.$$

Thus, we only need to calculate the inverse of the M -by- M matrix $v\mathbf{H}^H \mathbf{H} + \sigma^2 \mathbf{I}$. The involved complexity is $\mathcal{O}(M^3)$ per block. The matrix-vector multiplication complexity of (9) is $\mathcal{O}(MJ)$ per block. The remaining operations can be efficiently implemented using FFT with complexity $\mathcal{O}(J \log_2 J)$. Therefore, the complexity for each round of iteration is $\mathcal{O}(M^3 + J \log_2 J + JM)$. Moreover, the convergence of Algorithm 1 is difficult to analyze, just like many other BP-based iterative algorithms. Empirically, we observe that the proposed algorithm converges quite fast, as will be demonstrated in the next section.

Before leaving this section, we compare the proposed iterative equalizer with alternative approaches. Specifically, the existing decision feedback algorithms, such as MMSE-ISDIC [17] and PDA [18], can be directly applied to the EST system in consideration. The involved complexity, dominated by the LMMSE estimation, is in general $\mathcal{O}(J^3)$, where J is the DFT length; see Appendix A. The authors in [20] mainly followed the MMSE-ISDIC principle in the equalizer design, and

reduced the computational complexity by making approximations in computation.

In this paper, we reinterpret the LMMSE based decision feedback equalization from the novel perspective of BP. We prove in Appendix A that the BP-based algorithm is equivalent to MMSE-ISDIC, if the calculation of extrinsic messages in (21) is omitted and no equal-variance assumption is taken. Similar to [20], approximations are introduced in our algorithm to reduce the involved complexity. *We emphasize that (21) is a critical step that distinguishes our algorithm from the existing approaches in the literature.* This is made possible by looking at the problem from the new perspective of BP. We will show that this new perspective further opens up the possibility to accurately characterize the behavior of the iterative equalizer. We will also demonstrate by numerical results that the new algorithm can significantly outperform its counterparts in various system setups.

IV. EVOLUTION ANALYSIS

In this section, we analyze the performance of the proposed iterative equalizer based on the SINR-variance evolution technique first proposed in [25] for turbo equalization.

A. Characterization of the Estimation Module

We first consider the estimation module which starts from \vec{z} and ends at \vec{x} in Fig. 1. We use the input variance v to characterize the reliability of the input \vec{z} , and the reciprocal of the output variance, denoted by $\rho = 1/u$ (with u given in (16a)), to characterize the reliability of the outputs. Note that ρ , referred to as the signal-to-interference-plus-noise-ratio (SINR) of x in \vec{x} , is actually the SNR of the equivalent AWGN channel in (17).

We now derive an input-output transfer function of the estimator. Define

$$\phi_1(v) = \left(N^{-1} \text{tr} \left\{ (\sigma^{-2} \mathbf{H}^H \mathbf{H} + v^{-1} \mathbf{I})^{-1} \right\} \right)^{-1} - v^{-1} \quad (22)$$

where $\text{tr}(\cdot)$ denotes the trace operation and v is the input variance of the estimator. We have the following proposition.

Proposition 1: The output SINR ρ of the estimation part is given by $\rho = \phi_1(v)$. Moreover, $\phi_1(v)$ satisfies

$$\lim_{v \rightarrow 0} \phi_1(v) = \frac{1}{N\sigma^2} \text{tr} \{ \mathbf{H}^H \mathbf{H} \}. \quad (23)$$

Proof: Noting $\tilde{\mathbf{H}} = \mathbf{I} \otimes \mathbf{H}$, we can rewrite (10) as

$$\tilde{\mathbf{V}} = \mathbf{I} \otimes (\sigma^{-2} \mathbf{H}^H \mathbf{H} + v^{-1} \mathbf{I})^{-1}.$$

Then, w_x in (14) can be written as

$$\begin{aligned} w_x &= J^{-1} \text{tr} \left[\mathbf{I} \otimes (\sigma^{-2} \mathbf{H}^H \mathbf{H} + v^{-1} \mathbf{I})^{-1} \right] \\ &= N^{-1} \text{tr} \left[(\sigma^{-2} \mathbf{H}^H \mathbf{H} + v^{-1} \mathbf{I})^{-1} \right]. \end{aligned}$$

Substituting the above w_x into (16a) and by noting $\rho = 1/u$, we conclude the proof of the first half of Proposition 1.

For the second half of the proposition, rewrite in (22) as

$$\begin{aligned} \phi_1(v) &= \frac{1}{N^{-1} \sum_{i=1}^N \frac{1}{\sigma^{-2} \lambda_i + v^{-1}}} - \frac{1}{v} \\ &= \frac{N^{-1} \sum_{i=1}^N \frac{v \sigma^{-2} \lambda_i}{v \sigma^{-2} \lambda_i + 1}}{N^{-1} \sum_{i=1}^N \frac{v}{v \sigma^{-2} \lambda_i + 1}} = \frac{\sum_{i=1}^N \frac{\sigma^{-2} \lambda_i}{v \sigma^{-2} \lambda_i + 1}}{\sum_{i=1}^N \frac{1}{v \sigma^{-2} \lambda_i + 1}} \end{aligned}$$

where λ_i is the i th eigenvalue of $\mathbf{H}^H \mathbf{H}$. By noting $\text{tr} \{ \mathbf{H}^H \mathbf{H} \} = \sum_i \lambda_i$, we readily obtain $\lim_{v \rightarrow 0} \phi_1(v) = (1/N\sigma^2) \text{tr} \{ \mathbf{H}^H \mathbf{H} \}$. ■

The right-hand-side of (23) represents the SNR of a transmit symbol as if there is no interference from the other symbols. Consider transmitting x over an AWGN channel $r = x + n$, where x is independent and uniformly drawn from the constellation satisfying (3), and the additive noise n is an independent symmetric complex Gaussian random variable with zero mean and variance $N\sigma^2(\text{tr}(\mathbf{H}^H \mathbf{H}))^{-1}$. The ML performance achieved over such an AWGN channel is referred to as the matched filter bound (MFB), and will be used as a benchmark for performance evaluation.

B. Characterization of the Demodulation Module

Now consider the demodulation module which starts from \vec{x} and ends at \vec{z} in Fig. 1. For a sufficiently large J , the average *a posteriori* variance in (20) can be calculated by

$$w_y = \frac{1}{J} \sum_{k=1}^J t_k = \gamma(\rho), \quad (24a)$$

with

$$\gamma(\rho) \equiv \text{E} \left[|x - \text{E}[x|y]|^2 \right] \quad (24b)$$

where x is uniformly taken over \mathcal{S} , and $n \sim \mathcal{CN}(0, 1/\rho)$ is independent of x ; the expectation is taken over the joint distribution of x and n . An explicit expression of $\gamma(\rho)$ is given by

$$\begin{aligned} \gamma(\rho) &= 1 - \text{E} \left[|\text{E}[x|y]|^2 \right] \\ &= 1 - \int \frac{\rho \left| \sum_{s \in \mathcal{S}} s \cdot \exp(-\rho|y-s|^2) \right|^2}{\pi |\mathcal{S}| \sum_{s \in \mathcal{S}} \exp(-\rho|y-s|^2)} dy \end{aligned}$$

where the integral above is taken over the whole complex domain. Then, we have the following result.

Proposition 2: The output variance v of the demodulation part is

$$v = (\gamma(\rho)^{-1} - \rho)^{-1} \equiv \psi(\rho). \quad (25)$$

Proof: The proof is straightforward by noting (21a) and (24). ■

C. SINR-Variance Transfer Chart

Now consider the overall iterative process. Based on the above discussions, the behavior of the iterative equalizer can be characterized by $\rho = \phi_1(v)$ and $v = \psi(\rho)$, i.e., the iterative process of the estimation and demodulation modules can be tracked by a recursion of ρ and v .

Specifically, let q be the iteration number. Then, we have

$$\rho^{(q)} = \phi_1(v^{(q-1)}) \text{ and } v^{(q)} = \psi(\rho^{(q)}), \quad q = 1, 2, \dots$$

The recursion continues and converges to a point v^* satisfying

$$\phi_1(v^*) = \psi^{-1}(v^*) \text{ and } \phi_1(v) > \psi^{-1}(v), \text{ for } v \in (v^*, 1],$$

where $\psi^{-1}(\cdot)$ is the inverse of $\psi(\cdot)$, which exists since $\psi(\cdot)$ is continuous and monotonic. The output performance (e.g., bit error rate (BER) or frame error rate (FER)) is uniquely determined by the convergence point v^* . Note that the functions of $\text{BER}(v^*)$ and $\text{FER}(v^*)$ can be obtained by pre-simulating the demodulation module. Therefore, the above evolution technique can be used to predict the performance of the proposed iterative equalization algorithm.

We emphasize that the above evolution analysis is based on Assumptions 1–3 stated in Section III. These assumptions are asymptotically true as J tends to infinity, and thus the proposed evolution technique works well for a sufficiently large J .

D. Numerical Example

Here we present numerical results to demonstrate the performance of the proposed iterative equalization algorithm, and also to verify the evolution technique developed above.

We start with a fixed 4×4 MIMO channel given by

$$\begin{bmatrix} -0.23+0.41i & 0.04-0.18i & -0.06-0.09i & -0.05+0.05i \\ 0.73+0.13i & 0.04-0.36i & -0.49-0.12i & -0.01+0.27i \\ -0.80-0.28i & 0.59-0.47i & -0.40-0.68i & -0.12-0.38i \\ 0.28+0.14i & 0.14-0.08i & -0.00-0.40i & -0.02+0.09i \end{bmatrix} \quad (26)$$

QPSK modulation is employed. Each frame consists of $J = 32768$ symbols, so that the DFT length J is sufficiently large to ensure that the evolution analysis agrees well with the simulation. The channel SNR is defined as $\text{SNR} \equiv 1/\sigma^2$.

Fig. 2(a) illustrates the SINR-variance transfer chart for the above channel at $\text{SNR} = 12$ dB. The transfer functions of the estimation and demodulation modules are respectively given by (22) and (25). The measured evolution trajectories are obtained by averaging over 500 frames. From Fig. 2(a), the measured trajectories agree well with the transfer curves.

As aforementioned, a key difference between our proposed algorithm and the EST-based algorithms in [7] and [20] is that the former calculates the extrinsic distributions at both the outputs of the estimation and demodulation modules, while the latter only calculates extrinsic distributions at the estimator output. It is interesting to compare our proposed equalizer with the one without calculating (21) (i.e., the *a posteriori* mean vector \vec{y} is directly passed to the estimation module and therefore the transfer function $\psi(\rho)$ of the demodulation module should be replaced by $\psi(\rho) = \gamma(\rho)$). We establish a similar SINR-

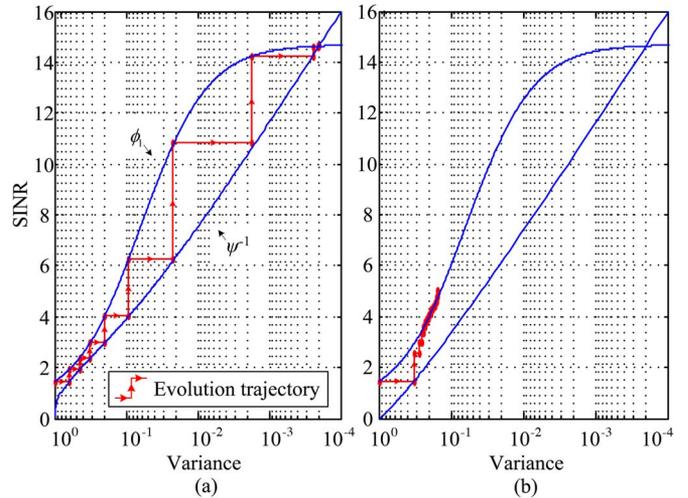


Fig. 2. The SINR-variance transfer charts for the proposed iterative equalizer in panel (a) and the equalizer without calculating (21) in (b). $\text{SNR} = 12$ dB.

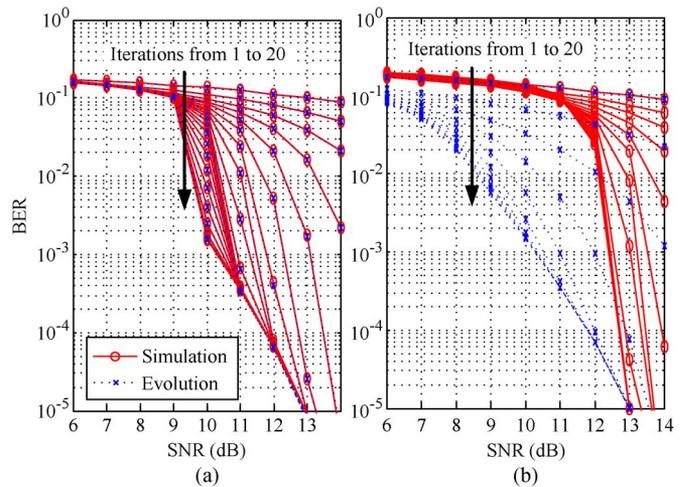


Fig. 3. The BER performances for the proposed iterative equalizer in panel (a) and the equalizer without calculating (21) in panel (b). $J = 32768$.

variance transfer chart to characterize the evolution process, as illustrated in Fig. 2(b) for $\text{SNR} = 12$ dB. We observe a clear mismatch between the measured evolution trajectories and the transfer curves. This mismatch is caused by delivering correlated messages in iterative equalization.

Fig. 3 compares the BER curves of the proposed iterative equalizer and the equalizer without calculating (21) for different numbers of iterations. The solid curves represent the simulated performance and the dotted curves represent the predicted performance using the evolution analysis. We see that the proposed equalizer outperforms the other one significantly. Moreover, the simulation and evolution match well for the proposed equalizer. For the equalizer without calculating (21), however, considerable discrepancies are observed between simulation and evolution.

Fig. 4 illustrates the BER performance of the proposed iterative algorithm and the equalizer without calculating (21), with the number of symbols J varying from 512 to 32768. Again, we see that the proposed equalizer significantly outperforms the other equalizer. Further, the performance of the proposed algorithm improves as J increases; particularly, for

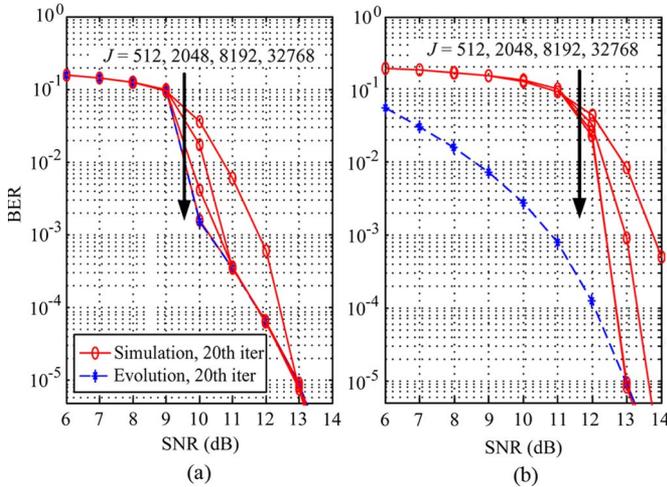


Fig. 4. The BER performances of the proposed equalizer [in panel (a)] and the equalizer without calculating (21) [in panel (b)] for different J . Iteration number = 20.

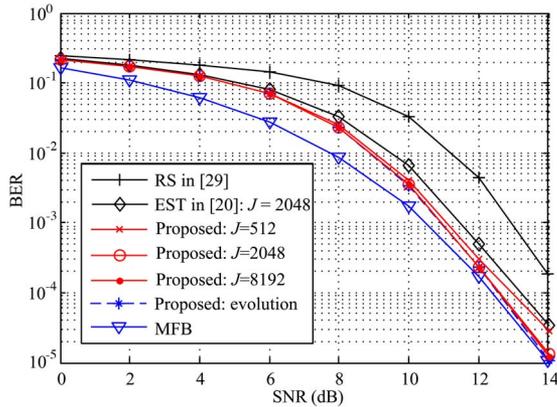


Fig. 5. The BER performances of the various schemes without utilizing CSIT in the quasi-static Rayleigh-fading 4×4 MIMO channel.

$J = 32768$, the simulated performance matches well with the evolution analysis. This is a verification of the fact that the three assumptions in Section III are justifiable for a sufficiently large J and hence the evolution analysis is established based on these assumptions becomes accurate for a large J . Empirically, for most channel realizations, a much smaller value of J is sufficient to ensure a good match of simulation and evolution analysis; for example, it suffices to set $J = 2048$ for the system considered in Fig. 5.

We next consider quasi-static Rayleigh fading MIMO channels, with the entries of \mathbf{H} independently distributed as $\mathcal{CN}(0, 1/N)$. For the iterative equalizer, the number of iterations is always set to 10 except otherwise specified. Fig. 5 compares the BER performance of the proposed scheme with the existing non-linear equalizers, including the EST-based scheme in [20] and the repetition-and-superposition (RS) approach [29]. We assume $M = N = 4$ and QPSK modulation with Gray mapping. For the RS scheme, 16 layers of signals are superimposed; in each layer, the data symbols are QPSK modulated, repeated by 16 times, and then randomly scrambled. The system throughput is 8 bits per channel use. At the receiver side, a message-passing algorithm is used for signal detection.

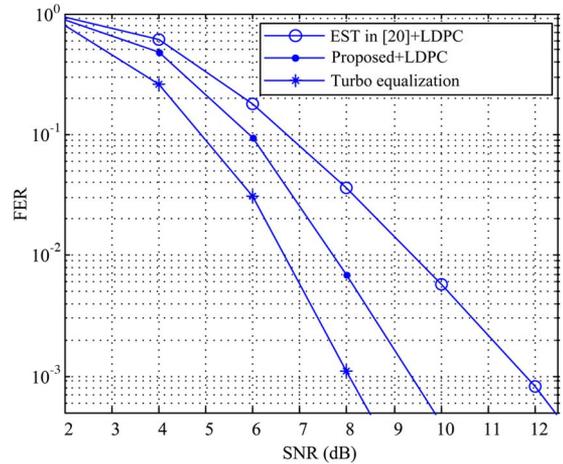


Fig. 6. Performance comparison of the proposed iterative equalization algorithm and the well-established channel-coding based turbo equalization schemes over quasi-static Rayleigh-fading MIMO channels with $M = N = 4$. Each frame consists of $J = 4096$ symbols.

From Fig. 5, we see that the BER performance of the scheme with the proposed iterative algorithm improves with increasing J . For this system setting, $J = 2048$ is sufficient to achieve satisfactory performance and further increasing J to 8192 does not provide much performance gain. We also see that our proposed scheme outperforms the EST scheme in [20] by about 0.7 dB at $\text{BER} = 10^{-4}$. This performance gain is attributed to the calculation of extrinsic distributions at the demodulation output. Also, our proposed scheme outperforms the RS scheme in [29] by about 2 dB at $\text{BER} = 10^{-3}$. This performance gain is because, in estimating each RS layer, the signals from the other layers are treated as interference, which lowers the effective SINR seen by each layer.

Fig. 5 also includes the MFB (with the definition given below the proof of Proposition 1) which can be viewed as the performance of a genie aided system in which the inter-symbol interference can be pre-cancelled at the receiver. From Fig. 5, the performance gap between the proposed scheme and the MFB is not significant throughout the whole SNR range of interest; this gap vanishes at high SNR.

We next study the performance of the proposed scheme serially concatenated with practical channel coding. The system setups are as follows. The transmitter employs a regular rate-1/2 binary (3, 6) LDPC code with code length = 8192. The coded bits are QPSK modulated with Gray mapping, and then EST-precoded and transmitted over a quasi-static Rayleigh fading 4×4 MIMO channel. The receiver processes the received data using the proposed iterative equalizer or the EST algorithm in [20], followed by a standard message-passing LDPC decoder. The simulated performances are illustrated in Fig. 6, in which our proposed algorithm significantly outperforms the EST-based scheme in [20] by 2.3 dB at $\text{FER} = 10^{-3}$.

Fig. 6 also includes the performance curves of the well-established iterative receivers following the principle of turbo equalization [11], [27]. In turbo equalization, the FEC decoder is included in the iterative process, which leads to better performance at the cost of considerably increased decoding complexity. From Fig. 6, we see that our proposed scheme

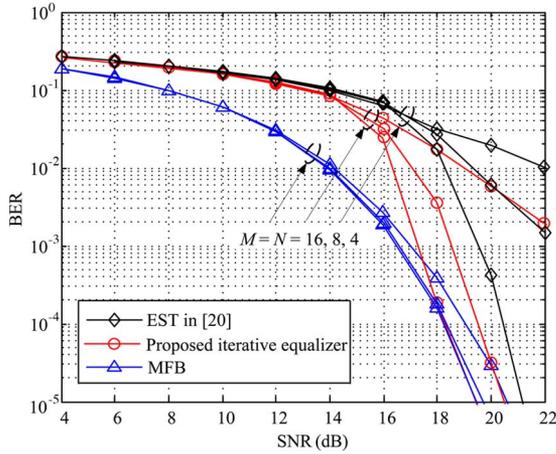


Fig. 7. The BER performances of the proposed scheme and the EST scheme in [20] in the quasi-static Rayleigh-fading channels with varying number of antennas. $J = 8192$. Gray-mapped 16-QAM is employed.

performs within 1.5 dB of the performance curve of turbo equalization throughout the SNR range of interest. This demonstrates that our proposed scheme offers an attractive tradeoff between performance and complexity.

In Fig. 7, we compare the BER performance of the proposed scheme with that of the EST scheme in [20] under various antenna setups, i.e., $M = N = 4, 8, 16$, respectively. Standard 16-QAM with Gray-mapping is employed. From Fig. 7, we see that the proposed scheme significantly outperforms the scheme in [20] for all cases, e.g., by over 2 dB at $\text{BER} = 10^{-4}$ for $M = N = 16$. The MFBs for the considered antenna setups have also been included for comparison. We see that the proposed scheme closely approaches the MFB for $M = N = 16$. Note that the power of each element of \mathbf{H} is normalized to $1/N$, and hence the MFBs for different antenna setups are close to each other.

V. PRECODER DESIGN WITH CSIT

In this section, we optimize the linear precoder using the evolution technique, by exploiting CSIT. We show that the system performance can be significantly improved with the optimal precoder design.

A. Channel and Transceiver Model

We now consider a fast fading MIMO Gaussian channel. The received signal vector at the i th channel use is modeled as

$$\mathbf{r}_i = \mathbf{H}_i \mathbf{z}_i + \boldsymbol{\eta}_i \quad (27)$$

where \mathbf{H}_i is an M -by- N channel transfer matrix, and the other variables remain the same as those in (1). Partial CSI at the transmitter and perfect CSI at the receiver are assumed. For partial CSIT, we assume the mean-feedback model [30], i.e., each \mathbf{H}_i is a realization of the random matrix

$$\mathbf{H} = \bar{\mathbf{H}} + \boldsymbol{\Delta} \quad (28)$$

where the entries of $\bar{\mathbf{H}}$ are independently drawn from $\mathcal{CN}(0, \alpha)$ with $\alpha \in [0, 1]$, and those of $\boldsymbol{\Delta}$ independently drawn from

$\mathcal{CN}(0, 1 - \alpha)$. Note that $\bar{\mathbf{H}}$ and $\boldsymbol{\Delta}$, respectively, represent the known and unknown part of the channel at the transmitter. The factor α represents the reliability of CSIT, with $\alpha = 1$ for the case of perfect CSIT and $\alpha = 0$ for the case of no CSIT.

The transceiver structure is depicted in Fig. 8. Each transmission frame consists of L channel uses. Thus, the overall channel matrix is given by $\text{diag}\{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_L\}$. A precoder \mathbf{P} is inserted between the EST module and the physical channel, so as to exploit the potential advantage provided by the knowledge of the channel mean $\bar{\mathbf{H}}$. For simplicity, we assume that $\bar{\mathbf{H}}$ remains unchanged in every transmission frame. The extension to the case of a varying channel mean is straightforward by considering a different \mathbf{P} for every different channel mean.

B. SINR-Variance Evolution Analysis

We begin with the SINR-variance evolution analysis. Since the transfer function of the demodulation module remains the same, we only need to consider the estimation module. Following the derivation of (22), we can show that the transfer function for the estimation module is written as

$$\rho = \left(\frac{1}{NL} \sum_{i=1}^L \text{tr} \left\{ (\sigma^{-2} \mathbf{P}^H \mathbf{H}_i^H \mathbf{H}_i \mathbf{P} + v^{-1} \mathbf{I})^{-1} \right\} \right)^{-1} - v^{-1}.$$

In general, the above is a random function of the channel $\{\mathbf{H}_i\}$. We assume that L is sufficiently large so that the above function becomes deterministic. Then

Proposition 3: As L tends to infinity, the transfer function of the estimation part is a deterministic function given by

$$\begin{aligned} \rho &\equiv \phi_2(v) \\ &= \left(\frac{1}{N} \mathbb{E} \left[\text{tr} \left\{ (\sigma^{-2} \mathbf{P}^H \mathbf{H}^H \mathbf{H} \mathbf{P} + v^{-1} \mathbf{I})^{-1} \right\} \right] \right)^{-1} - v^{-1} \end{aligned} \quad (29)$$

where the expectation is taken over \mathbf{H} .

C. Precoder Optimization

Our objective is to find the most power-efficient precoder at a target BER. Specifically, the optimization objective and the constraint are summarized as follows.

- We aim to minimize the transmission power $\text{tr}\{\mathbf{P}\mathbf{P}^H\}$ measured at the input of the physical channel (c.f., Fig. 8).
- At the same time, we ensure that the iterative equalizer converges to a designated output variance v^* , where v^* is determined by the target BER of the system.

This problem is formulated as follows:

$$\min_{\mathbf{P}} \text{tr}\{\mathbf{P}\mathbf{P}^H\} \quad (30a)$$

$$\text{s.t. } \phi_2(v) \geq \psi^{-1}(v), \text{ for } 1 \geq v \geq v^*, \quad (30b)$$

where v^* is determined by the target BER performance.

Now consider the optimization of the precoding matrix \mathbf{P} . The objective function is linear in $\mathbf{P}\mathbf{P}^H$. Also, it can be verified that $\phi_2(\cdot)$ in (29) is concave in $\mathbf{P}\mathbf{P}^H$. Thus, (30) can be solved numerically using convex programming. However, directly optimizing \mathbf{P} is computationally demanding, especially for a large

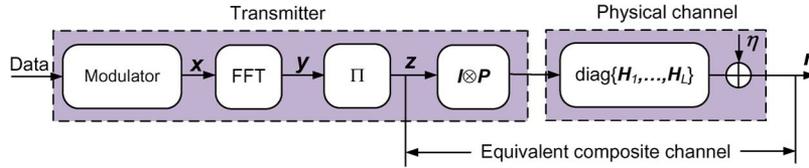


Fig. 8. An illustration of the linear precoder \mathbf{P} .

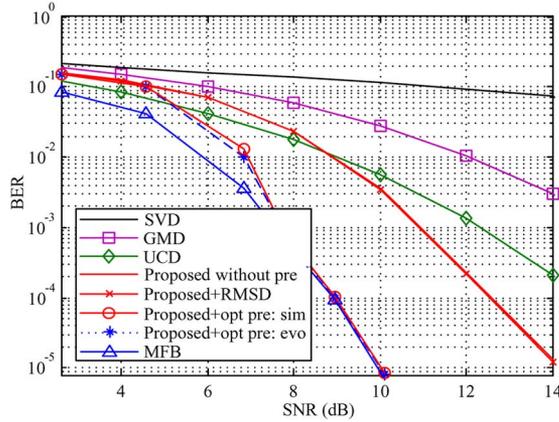


Fig. 9. The BER performance of the various schemes with optimized precoding based on full CSIT (i.e., $\alpha = 1$) in the quasi-static Rayleigh-fading 4×4 MIMO channel. System throughput = 8 bits per channel use.

MIMO size. We next discuss how to simplify the problem. Let the singular value decomposition (SVD) of \mathbf{P} be $\mathbf{P} = \mathbf{U}_P \mathbf{D}_P \mathbf{V}_P^H$. Similarly, let the SVD of the channel mean $\bar{\mathbf{H}}$ be

$$\bar{\mathbf{H}} = \mathbf{U}_{\bar{\mathbf{H}}} \Sigma_{\bar{\mathbf{H}}} \mathbf{V}_{\bar{\mathbf{H}}}^H. \quad (31)$$

We have the following result.

Proposition 4: Given $\bar{\mathbf{H}}$ and v^* , the optimal \mathbf{U}_P to the problem (30) is $\mathbf{U}_P = \mathbf{V}_{\bar{\mathbf{H}}}$, and the optimal \mathbf{V}_P is arbitrary.

The proof of Proposition 4 can be found in Appendix B. With Proposition 4, the optimization problem in (30) reduces to a simple power allocation problem with respect to \mathbf{D}_P . Though an explicit solution is not available, the optimal \mathbf{D}_P can be numerically found using standard convex programming. The details are omitted here for simplicity.

D. Numerical Results

We now present the numerical results to demonstrate the performance improvement. We first consider the special case of full CSIT (i.e., $\alpha = 1$). In this case, the precoder optimization in (30) directly applies to a quasi-static channel, by replacing $\phi_2(\cdot)$ by $\phi_1(\cdot)$. We now consider a quasi-static Rayleigh fading 4×4 MIMO channel with QPSK modulation. Each frame consists of $J = 32768$ symbols. The target BERs in precoder optimization are chosen as 1.5×10^{-1} , 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , and 10^{-5} . The well-known water-filling approach is included for comparison. Specifically, the MIMO channel is converted into parallel channels using SVD and then water-filling power allocation is applied to those parallel channels and QPSK modulated symbols are transmitted. We see that the performance of this water-filling approach is rather poor. This justifies the necessity of precoding across parallel channels. From Fig. 9, we

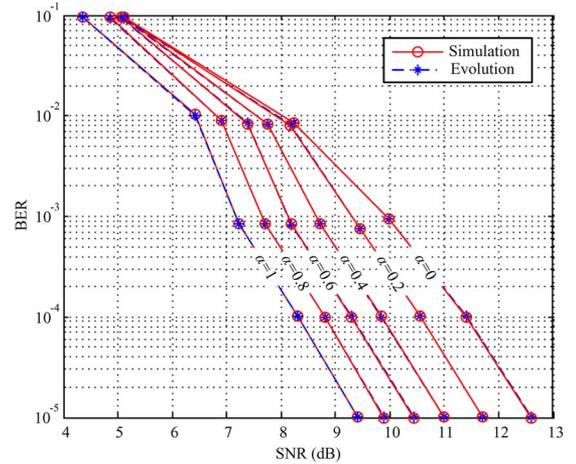


Fig. 10. The BER performance of the proposed scheme in a fast Rayleigh fading 4×4 MIMO channel with a varying quality of available CSIT.

see that the EST scheme with the proposed equalizer without exploiting CSIT (i.e., with non-optimized precoder $\mathbf{P} = \mathbf{I}$) can outperform the existing non-linear equalization schemes, including GMD [5] and UCD [6], when $\text{SNR} > 9$ dB. The MFB for the precoded system in Fig. 9 is given by (22) with \mathbf{H} replaced by $\mathbf{H}\mathbf{P}$. With precoder optimization, our proposed scheme can achieve an extra power gain of over 3 dB at $\text{BER} = 10^{-4}$. This demonstrates the advantages of our proposed scheme. In Fig. 9, we also include the performance of the proposed iterative equalizer with the RMSD precoder [7] at the transmitter. We see that this scheme does not bring any performance gain, as compared with the proposed equalizer without RMSD precoding. This implies that, for the proposed equalizer, RMSD precoding is not a good approach to exploit the potential benefit provided by the available CSIT.

We now consider the general case of partial CSIT. We assume a fast Rayleigh fading 4×4 MIMO channel. Each frame consists of $J = 32768$ symbols. In each frame, 32 samples of $\bar{\mathbf{H}}$ are independently generated, and for each sample of $\bar{\mathbf{H}}$, 256 samples of Δ are generated. In optimization, the target BERs are set to be 10^{-1} , 10^{-2} , 10^{-3} , 10^{-4} , and 10^{-5} . The other system settings remain the same. The BER performance of the above scheme is illustrated in Fig. 10. The solid lines represent the actual simulated system performance, while the dashed lines represent the performance predicted by the evolution analysis. In Fig. 10, the value of α varies from 0 to 1, with $\alpha = 0$ implying no CSIT and $\alpha = 1$ implying full CSIT. We see that full CSIT improves the system performance by over 3 dB at $\text{BER} = 10^{-4}$. Moreover, when α increases, the BER performance improves correspondingly. This implies that our scheme efficiently exploits the potential gain provided by the available CSIT.

VI. CONCLUSION

In this paper, we studied the EST-based linear precoding and iterative equalization scheme. We developed an iterative equalization algorithm based on the BP principle. We further use the SINR-variance technique to accurately analyze the performance of the proposed algorithm. We showed that, with full or partial CSIT, the linear precoder can be further optimized based on the newly established evolution technique. Optimal precoding directions were derived. We further showed that the remaining power allocation problem can be efficiently solved using convex programming. Numerical results demonstrate that the proposed scheme significantly outperforms the existing non-linear equalization algorithms, and provides an attractive tradeoff between performance and complexity, as compared with the well-established turbo equalization.

APPENDIX A

In this appendix, we establish an explicit connection between the MMSE-BP equalizer used in this paper and the MMSE-ISDIC equalizer in [17], [19].

A. Generic Linear System Model

Consider the estimation of a vector \mathbf{x} given a vector \mathbf{r} modeled as the output of a generic linear system:

$$\mathbf{r} = \mathbf{G}\mathbf{x} + \boldsymbol{\eta} \quad (32)$$

where \mathbf{G} is the transfer matrix, and the additive noise $\boldsymbol{\eta}$. Each entry of \mathbf{x} is taken over the constellation \mathcal{S} satisfying (3). The *a priori* mean and the covariance of \mathbf{x} are respectively given by $\bar{\mathbf{x}}$ and $\mathbf{V} = \text{diag}\{v_1, v_2, \dots, v_J\}$.

B. MMSE-BP

We first describe the MMSE-BP approach in which the extrinsic messages are calculated based on the principle of belief propagation. The linear MMSE estimator of \mathbf{x} given \mathbf{r} is given by [14]

$$\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{V}\mathbf{G}^H(\mathbf{G}\mathbf{V}\mathbf{G}^H + \sigma^2\mathbf{I})^{-1}(\mathbf{r} - \mathbf{G}\bar{\mathbf{x}}) \quad (33a)$$

and the corresponding MMSE matrix is given by

$$\hat{\mathbf{V}} = \left(\frac{1}{\sigma^2}\mathbf{G}^H\mathbf{G} + \mathbf{V}^{-1} \right)^{-1} = \mathbf{V} - \mathbf{V}\mathbf{G}^H\mathbf{R}^{-1}\mathbf{G}\mathbf{V} \quad (33b)$$

where

$$\mathbf{R} = \mathbf{G}\mathbf{V}\mathbf{G}^H + \sigma^2\mathbf{I}. \quad (33c)$$

Following (16), we obtain the extrinsic mean \bar{x}_i and variance \bar{v}_i of each x_i as

$$\bar{v}_i^{-1} = \hat{v}_i^{-1} - v_i^{-1}, \quad (34a)$$

$$\frac{\bar{x}_i}{\bar{v}_i} = \frac{\hat{x}_i}{\hat{v}_i} - \frac{\bar{x}_i}{v_i}, \quad (34b)$$

where \hat{v}_i is the i th diagonal element of $\hat{\mathbf{V}}$. Then, we model

$$\bar{x}_i = x_i + \bar{n}_i \quad (35)$$

where \bar{n}_i is independently drawn from $\mathcal{CN}(0, \bar{v}_i)$. Recall that each x_i is equi-probably taken over the constellation \mathcal{S} . Thus, given \bar{x}_i in (35), the conditional mean and variance of each x_i are respectively given by

$$m_i = \frac{\sum_{s_j \in \mathcal{S}} s_j p(\bar{x}_i | x_i = s_j)}{\sum_{s_j \in \mathcal{S}} p(\bar{x}_i | x_i = s_j)} \quad (36a)$$

$$t_i = \frac{\sum_{s_j \in \mathcal{S}} |s_j - m_i|^2 p(\bar{x}_i | x_i = s_j)}{\sum_{s_j \in \mathcal{S}} p(\bar{x}_i | x_i = s_j)} \quad (36b)$$

where $p(\bar{x}_i | x_i)$ is determined by (35) as

$$p(\bar{x}_i | x_i) = \frac{1}{\pi \bar{v}_i} \exp\left(-\frac{|x_i - \bar{x}_i|^2}{\bar{v}_i}\right). \quad (37)$$

C. MMSE-ISDIC

We now present the MMSE-ISDIC algorithm in [17], following the descriptions in [19]. We first calculate the soft interference-cancelled signal vector for detecting x_i as

$$\mathbf{r}_{\text{IC},i} = \mathbf{r} - \sum_{j=1, j \neq i}^J \mathbf{g}_j \bar{x}_j \quad (38)$$

and the corresponding linear MMSE filtering vector as

$$\mathbf{f}_i^M = \mathbf{g}_i^H (\mathbf{R}_i + \mathbf{g}_i \mathbf{g}_i^H)^{-1} \quad (39)$$

where \mathbf{g}_j is the j th column of \mathbf{G} , and

$$\mathbf{R}_i = \sum_{j=1, j \neq i}^J v_j \mathbf{g}_j \mathbf{g}_j^H + \sigma^2 \mathbf{I} = \mathbf{G}\mathbf{V}\mathbf{G}^H + \sigma^2 \mathbf{I} - v_i \mathbf{g}_i^H \mathbf{g}_i. \quad (40)$$

Then, we calculate the mean and variance of each x_i as

$$\tilde{m}_i = \frac{\sum_{s_j \in \mathcal{S}} s_j \gamma_{i,j}^M}{\sum_{s_j \in \mathcal{S}} \gamma_{i,j}^M} \quad \text{and} \quad \tilde{t}_i = \frac{\sum_{s_j \in \mathcal{S}} |s_j - \tilde{m}_i|^2 \gamma_{i,j}^M}{\sum_{s_j \in \mathcal{S}} \gamma_{i,j}^M} \quad (41a)$$

where

$$\begin{aligned} \gamma_{i,j}^M &= \exp\left(-\frac{|\mathbf{f}_i^M \mathbf{r}_{\text{IC},i} - \mu_i^M s_j|^2}{\mu_i^M (1 - \mu_i^M)}\right) \\ &= \exp\left(-\frac{\left|\frac{\mathbf{f}_i^M \mathbf{r}_{\text{IC},i}}{\mu_i^M} - s_j\right|^2}{\frac{1 - \mu_i^M}{\mu_i^M}}\right) \end{aligned} \quad (41b)$$

$$\mu_i^M = \mathbf{f}_i^M \mathbf{g}_i = \mathbf{g}_i^H (\mathbf{R}_i + \mathbf{g}_i \mathbf{g}_i^H)^{-1} \mathbf{g}_i. \quad (41c)$$

D. Equivalence Between MMSE-BP and MMSE-ISDIC

Here we establish the equivalence between the MMSE-BP approach and the MMSE-ISDIC approach. With (36) and (41), it suffices to show that

$$\hat{v}_i = \frac{1 - \mu_i^M}{\mu_i^M} \text{ and } \hat{x}_i = \frac{\mathbf{f}_i^M \mathbf{r}_{IC,i}}{\mu_i^M}. \quad (42)$$

We focus on proving $\hat{v}_i = (\mu_i^M)^{-1} - 1$. We first establish a useful relation. From the matrix inversion lemma, we obtain

$$\mathbf{g}_i^T (\mathbf{R}_i + \alpha \mathbf{g}_i \mathbf{g}_i^H)^{-1} = \mathbf{g}_i^H \mathbf{R}_i^{-1} - \frac{\alpha \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i \mathbf{g}_i^H \mathbf{R}_i^{-1}}{1 + \alpha \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i} \quad (43a)$$

$$= \frac{\mathbf{g}_i^H \mathbf{R}_i^{-1}}{1 + \alpha \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i} \quad (43b)$$

where α is an arbitrary constant. Then

$$\mathbf{g}_i^H \mathbf{R}_i^{-1} = \frac{\mathbf{g}_i^H \mathbf{R}_i^{-1} / (1 + \alpha \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i)}{1 - \alpha \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i / (1 + \alpha \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i)} \quad (44a)$$

$$= \frac{\mathbf{g}_i^H (\mathbf{R}_i + \alpha \mathbf{g}_i \mathbf{g}_i^H)^{-1}}{1 - \alpha \mathbf{g}_i^H (\mathbf{R}_i + \alpha \mathbf{g}_i \mathbf{g}_i^H)^{-1} \mathbf{g}_i}. \quad (44b)$$

By setting $\alpha = v_i$ and $\alpha = 1$ respectively, we obtain

$$\frac{\mathbf{g}_i^H \mathbf{R}_i^{-1}}{1 - v_i \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i} = \frac{\mathbf{g}_i^H (\mathbf{R}_i + \mathbf{g}_i \mathbf{g}_i^H)^{-1}}{1 - \mathbf{g}_i^H (\mathbf{R}_i + \mathbf{g}_i \mathbf{g}_i^H)^{-1} \mathbf{g}_i}. \quad (45)$$

Then

$$\frac{\mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i}{1 - v_i \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i} = \frac{\mu_i^M}{1 - \mu_i^M}. \quad (46)$$

We are now ready to prove $\hat{v}_i = (\mu_i^M)^{-1} - 1$. From (33b), we have $\hat{v}_i = v_i - v_i^2 \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i$. Plugging into (34a), we obtain

$$\hat{v}_i = (\hat{v}_i^{-1} - v_i^{-1})^{-1} = \left(\frac{1}{v_i - v_i^2 \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i} - \frac{1}{v_i} \right)^{-1} \quad (47)$$

$$= \frac{1 - v_i \mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i}{\mathbf{g}_i^H \mathbf{R}_i^{-1} \mathbf{g}_i} = \frac{1 - \mu_i^M}{\mu_i^M}, \quad (48)$$

where the last step utilizes (46). The other equality in (42) can be proven in a similar way. Therefore, we conclude that MMSE-BP and MMSE-ISDIC are equivalent.

E. Connections to Iterative Equalization

We now discuss the connections of the above two algorithms to the iterative equalization algorithm proposed in this paper, as well as to the algorithm proposed in [20]. The system considered in this paper is just a special case of the generic linear system in (32). To see this, we substitute (6) into (2) and compare the result with (32). Then, we see that the generic transfer matrix is chosen as $\mathbf{G} = (\mathbf{I} \otimes \mathbf{H}) \mathbf{\Pi} \mathbf{F}$ for the system considered in this paper. Clearly, we can follow (33)–(36),

as an alternative to the equations (9)–(18a), in the proposed algorithm. Then, the involved complexity is in general $\mathcal{O}(J^3)$, which is dominant by the inverse of the covariance matrix $\mathbf{G} \mathbf{V} \mathbf{G}^H + \sigma^2 \mathbf{I}$ in (33a). In the proposed algorithm, we simplify the calculation of this inverse as follows. We approximate the *a priori* covariance matrix \mathbf{V} as $v \mathbf{I}$, where v is the average variance. Then

$$\begin{aligned} (\mathbf{G} \mathbf{V} \mathbf{G}^H + \sigma^2 \mathbf{I})^{-1} &= (v \mathbf{G} \mathbf{G}^H + \sigma^2 \mathbf{I})^{-1} \\ &= \mathbf{I} \otimes (v \mathbf{H} \mathbf{H}^H + \sigma^2 \mathbf{I})^{-1}. \end{aligned}$$

The above only involves the inverse of an M -by- M matrix, and thus the complexity is reduced to $\mathcal{O}(M^3)$, where M is the number of receive antennas. Further, in Section III, we justify that this equal-variance approximation becomes exact as the block length J tends to infinity, based on the properties of the DFT. The system considered in [20] is another special case of (32) with $\mathbf{G} = (\mathbf{I} \otimes \mathbf{H}) \mathbf{\Pi} (\mathbf{I} \otimes \mathbf{F})$ (by noting that the modulated symbols in [20] are partitioned into several blocks and DFT is applied to each block, and that $\mathbf{\Pi}$ is a permutation matrix specified by the space-time encoder used in [20]). A similar equal-variance approximation is also taken in [20] to simplify the calculation. It is worth mentioning that the algorithm design in [20] basically follows the principle of MMSE-ISDIC described in Subsection C. We emphasize that, though MMSE-ISDIC is the same as MMSE-BP (as proven in Subsection D), it is difficult to justify the extra step (21) in our proposed algorithm for the calculation of the extrinsic messages at the demodulation output, if we stick to the principle of MMSE-ISDIC in the algorithm design. It is one of our major contributions to look at the design of a non-linear equalizer based on belief propagation, which gives a novel perspective different from the approach in [20]. With this new perspective, we are able to design iterative equalizers with analyzable behavior and with better performance.

APPENDIX B

A. Proof of Proposition 4

To start with, we have (49a), as shown at the bottom of the next page, where

$$C \equiv v(N - M). \quad (49b)$$

In the above, step (a) holds by noting that $\mathbf{A} \mathbf{A}^H$ and $\mathbf{A}^H \mathbf{A}$ share the same non-zero eigenvalues, (b) follows by substituting (28) and letting $\mathbf{Q} = \mathbf{V}_H^H \mathbf{P} \mathbf{P}^H \mathbf{V}_H$, (c) by substituting (31), and (d) from the fact that $\mathbf{U}_H \mathbf{\Delta} \mathbf{V}_H$ has the same distribution as $\mathbf{\Delta}$. With (29) and (49), the problem (30) can be rewritten as

$$\min_{\mathbf{Q}} \text{tr}\{\mathbf{Q}\} \quad (50a)$$

$$\text{s.t. } f(v, \mathbf{Q}) + C \leq \frac{1}{\psi^{-1}(v) + \frac{1}{v}}, \text{ for } 1 \geq v \geq v^*. \quad (50b)$$

The above optimization problem is independent of \mathbf{V}_P . Thus, the optimal \mathbf{V}_P is arbitrary.

What remains is to prove that the optimal \mathbf{U}_P is $\mathbf{U}_P = \mathbf{V}_{\bar{H}}$. From $\mathbf{Q} = \mathbf{V}_{\bar{H}}^H \mathbf{P} \mathbf{P}^H \mathbf{V}_{\bar{H}}$ and $\mathbf{P} = \mathbf{U}_P \mathbf{D}_P \mathbf{V}_P^H$, it suffices to show that the optimal \mathbf{Q} to (50) is diagonal. We prove by contradiction. Suppose that $\mathbf{Q} = \mathbf{Q}_0$ is an optimal solution to (50) and that \mathbf{Q}_0 is not a diagonal matrix. We need the following lemma, with the proof given in Subsection B.

Lemma 1: For any $v > 0$, $f(v, \mathbf{Q}) \geq f(v, \mathbf{Q}_{\text{diag}})$, where the equality holds if and only if \mathbf{Q} is diagonal.

From Lemma 1, we obtain

$$f(v, (\mathbf{Q}_0)_{\text{diag}}) < f(v, \mathbf{Q}_0) \leq (\psi^{-1}(v) + v^{-1})^{-1} - C. \quad (51)$$

Together with $\text{tr}\{\mathbf{Q}_0\} = \text{tr}\{(\mathbf{Q}_0)_{\text{diag}}\}$, $(\mathbf{Q}_0)_{\text{diag}}$ is also optimal to (50). However, from (51) and the continuity of the f -function, there must exist another precoding matrix \mathbf{Q}_1 (slightly perturbed away from $(\mathbf{Q}_0)_{\text{diag}}$ satisfying

$$\text{tr}\{\mathbf{Q}_1\} < \text{tr}\{(\mathbf{Q}_0)_{\text{diag}}\}$$

and

$$f(v, \mathbf{Q}_1) < (\psi^{-1}(v) + v^{-1})^{-1} - C.$$

Therefore, \mathbf{Q}_0 is not the optimal solution to (50), which leads to absurdity. This completes the proof.

B. Proof of Lemma 1

We cite without proof the following fact from [31]. Let J be an M -by- M diagonal matrix with the diagonal elements being ± 1 . There are $L = 2M$ different such matrices indexed from $k = 1$ to T . Let \mathbf{A} be an arbitrary M -by- M matrix. Then

$$(\mathbf{A})_{\text{diag}} = \frac{1}{T} \sum_{k=1}^T \mathbf{J}_k \mathbf{A} \mathbf{J}_k. \quad (52)$$

We obtain (53), as shown at the bottom of the page, where step (a) follows from the definition, and (b) from the fact that $\mathbf{J}_k \Delta \mathbf{J}_k$ has the same distribution as Δ . It can be verified that the f -function is a convex function of \mathbf{Q} . Thus

$$\begin{aligned} f(v, \mathbf{Q}) &\stackrel{(a)}{=} \frac{1}{T} \sum_{k=1}^T f(v, \mathbf{J}_k \mathbf{Q} \mathbf{J}_k) \\ &\stackrel{(b)}{\geq} f\left(v, \frac{1}{T} \sum_{k=1}^T \mathbf{J}_k \mathbf{Q} \mathbf{J}_k\right) \\ &\stackrel{(c)}{=} f(v, \mathbf{Q}_{\text{diag}}) \end{aligned} \quad (54)$$

where step (a) follows from (53), (b) from the Jensen's inequality, and (c) from (52). Noting that the equality in step (b) above holds if and only if \mathbf{Q} is diagonal, we complete the proof of Lemma 1.

$$\begin{aligned} &\text{tr} \left\{ \mathbb{E} \left[\left(\frac{1}{v} \mathbf{I} + \frac{1}{\sigma^2} \mathbf{P}^H \mathbf{H}^H \mathbf{H} \mathbf{P} \right)^{-1} \right] \right\} \\ &\stackrel{(a)}{=} \text{tr} \left\{ \mathbb{E} \left[\left(\frac{1}{v} \mathbf{I} + \frac{1}{\sigma^2} \mathbf{H} \mathbf{P} \mathbf{P}^H \mathbf{H}^H \right)^{-1} \right] \right\} + C \\ &\stackrel{(b)}{=} \text{tr} \left\{ \mathbb{E} \left[\left(\frac{1}{v} \mathbf{I} + \frac{1}{\sigma^2} (\bar{\mathbf{H}} + \Delta) \mathbf{V}_{\bar{H}} \mathbf{Q} \mathbf{V}_{\bar{H}}^H (\bar{\mathbf{H}} + \Delta)^H \right)^{-1} \right] \right\} + C \\ &\stackrel{(c)}{=} \text{tr} \left\{ \mathbb{E} \left[\left(\frac{1}{v} \mathbf{I} + \frac{1}{\sigma^2} (\Sigma_{\bar{H}} + \mathbf{U}_{\bar{H}}^H \Delta \mathbf{V}_{\bar{H}}) \mathbf{Q} (\Sigma_{\bar{H}} + \mathbf{U}_{\bar{H}}^H \Delta \mathbf{V}_{\bar{H}})^H \right)^{-1} \right] \right\} + C \\ &\stackrel{(d)}{=} \underbrace{\text{tr} \left\{ \mathbb{E} \left[\left(\frac{1}{v} \mathbf{I} + \frac{1}{\sigma^2} (\Sigma_{\bar{H}} + \Delta) \mathbf{Q} (\Sigma_{\bar{H}} + \Delta)^H \right)^{-1} \right] \right\}}_{f(v, \mathbf{Q})} + C \end{aligned} \quad (49a)$$

$$\begin{aligned} &f(v, \mathbf{Q}) \\ &\stackrel{(a)}{=} \text{tr} \left\{ \mathbb{E} \left[\left(\frac{1}{v} \mathbf{I} + \frac{1}{\sigma^2} (\Sigma_{\bar{H}} + \Delta) \mathbf{Q} (\Sigma_{\bar{H}} + \Delta)^H \right)^{-1} \right] \right\} \\ &= \text{tr} \left\{ \mathbb{E} \left[\left(\frac{1}{v} \mathbf{I} + \frac{1}{\sigma^2} (\Sigma_{\bar{H}} + \mathbf{J}_k \Delta \mathbf{J}_k) \mathbf{Q} (\Sigma_{\bar{H}} + \mathbf{J}_k \Delta \mathbf{J}_k)^H \right)^{-1} \right] \right\} \\ &\stackrel{(b)}{=} \text{tr} \left\{ \mathbb{E} \left[\left(\frac{1}{v} \mathbf{I} + \frac{1}{\sigma^2} (\Sigma_{\bar{H}} + \Delta) \mathbf{J}_k \mathbf{Q} \mathbf{J}_k (\Sigma_{\bar{H}} + \Delta)^H \right)^{-1} \right] \right\} \\ &= f(v, \mathbf{J}_k \mathbf{Q} \mathbf{J}_k) \end{aligned} \quad (53)$$

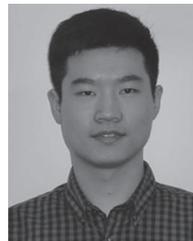
REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.
- [2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41–59, 1996.
- [3] P. Wonianski, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over rich scattering wireless channels," in *Proc. ISSE*, 1998, vol. 98, pp. 295–300.
- [4] B. Hassibi, "An efficient square-root algorithm for blast," in *Proc. IEEE ICASSP*, 2000, pp. 737–740.
- [5] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for mimo communications using geometric mean decomposition," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3791–3803, Oct. 2005.
- [6] Y. Jiang, J. Li, and W. W. Hager, "Uniform channel decomposition for MIMO communications," *IEEE Trans. Signal Process.*, vol. 53, no. 11, pp. 4283–4294, Nov. 2005.
- [7] T. Hwang and Y. Kwon, "Root mean square decomposition for EST-based spatial multiplexing systems," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1295–1306, Mar. 2012.
- [8] M. Sellathurai and S. Haykin, "Turbo-BLAST for wireless communications: Theory and experiments," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2538–2546, Oct. 2002.
- [9] T. Abe and T. Matsumoto, "Space-time turbo equalization in frequency-selective MIMO channels," *IEEE Trans. Veh. Technol.*, vol. 52, no. 3, pp. 469–475, May 2003.
- [10] C. Douillard *et al.*, "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecommun.*, vol. 6, no. 5, pp. 507–511, Sep./Oct. 1995.
- [11] 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.
- [12] *Group Radio Access Network, Evolved Universal Terrestrial Radio Access, Multiplexing and Channel Coding (Release 11)*, 3GPP Tech. Spec. Std. TS 25.212 v11.2.0, Jun. 2012.
- [13] *IEEE P802.11n/D1.0 Draft Amendment to Standard for Information Technology Telecommunications and Information Exchange Between Systems, Local and Metropolitan Networks, Specific Requirements*, IEEE 802.11 Working Group Std., Mar. 2006.
- [14] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [15] G. K. Kaleh, "Channel equalization for block transmission systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 1, pp. 110–121, Jan. 1995.
- [16] N. Benvenuto and S. Tomasin, "Iterative design and detection of a DFE in the frequency domain," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1867–1875, Nov. 2005.
- [17] J. F. Roßler and J. B. Huber, "Iterative soft decision interference cancellation receivers for DS-SS-CDMA downlink employing 4QAM and 16QAM," in *Conf. Rec. 36th Asilomar Conf. Signals, Syst. Comput.*, 2002, vol. 2, pp. 1488–1494.
- [18] J. Luo, K. R. Pattipati, P. K. Willett, and F. Hasegawa, "Near-optimal multiuser detection in synchronous CDMA using probabilistic data association," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 361–363, Sep. 2001.
- [19] F. Cao, J. Li, and J. Yang, "On the relation between PDA and MMSE-ISDIC," *IEEE Signal Process. Lett.*, vol. 14, no. 9, pp. 597–600, Sep. 2007.
- [20] T. Hwang, Y. Kim, and H. Park, "Energy spreading transform approach to achieve full diversity and full rate for MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6547–6560, Dec. 2012.
- [21] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA, USA: Morgan Kaufmann, 1988.
- [22] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [23] X. Yuan, L. Ping, and A. Kavcic, "Achievable rates of MIMO-ISI systems with linear precoding and iterative LMMSE detection," in *Proc. IEEE ISIT*, 2011, pp. 2909–2913.
- [24] X. Yuan, C. Xu, L. Ping, and X. Lin, "Precoder design for multiuser MIMO ISI channels based on iterative LMMSE detection," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 6, pp. 1118–1128, Dec. 2009.
- [25] X. Yuan, Q. Guo, X. Wang, and L. Ping, "Evolution analysis of lowcost iterative equalization in coded linear systems with cyclic prefixes," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 301–310, Feb. 2008.
- [26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
- [27] M. Tuchler, A. C. Singer, and R. Koetter, "Minimum mean squared error equalization using a priori information," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 673–683, Mar. 2002.
- [28] H.-A. Loeliger *et al.*, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, Jun. 2007.
- [29] L. Ping, K. Wu, L. Liu, and W. Leung, "A simple, unified approach to nearly optimal multiuser detection and space-time coding," in *Proc. ITW*, 2002, pp. 53–56.
- [30] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.
- [31] J. Wang and D. P. Palomar, "Worst-case robust MIMO transmission with imperfect channel knowledge," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3086–3100, Aug. 2009.



Xiaojun Yuan (S'04–M'09) received the B.S. degree in electronic and information systems from Shanghai Jiaotong University, Shanghai, China, the M.S. degree in circuit and systems from Fudan University, Shanghai, China, and the Ph.D. degree in electrical engineering from the City University of Hong Kong, Hong Kong, in 2008. From 2009 to 2011, he was a Research Fellow at the Department of Electronic Engineering, the City University of Hong Kong. He was also a Visiting Scholar at the Department of Electrical Engineering, the University

of Hawaii at Manoa in spring and summer 2009, as well as in the same period of 2010. From 2011 to 2014, he was a Research Assistant Professor with the Institute of Network Coding, The Chinese University of Hong Kong. He is now an Assistant Professor with the School of Information Science and Technology, ShanghaiTech University. His research interests cover a broad range of wireless communications, signal processing, and information theory including multiple-input multiple-output (MIMO) communications, rateless coding design, non-linear equalization, physical-layer network coding, cooperative communications, etc. He has published over 60 peer-reviewed research papers in the leading international journals and conferences, and has served on a number of technical programs for international conferences. Dr. Yuan was a recipient of the Best Paper Award of IEEE International Conference on Communications, Sydney, Australia, 2014.



Junjie Ma received the B.E. degree from Xidian University, Xi'an, China, in 2010. He is currently working towards the Ph.D. degree at City University of Hong Kong, Hong Kong. His research interests include wireless communications, iterative detection and decoding, and statistical signal processing.



Li Ping (S'87–M'91–SM'06–F'10) received the Ph.D. degree at Glasgow University in 1990. He was a Lecturer at Department of Electronic Engineering, Melbourne University, from 1990 to 1992. He worked as a Research Staff at Telecom Australia Research Laboratories from 1993 to 1995. Since January 1996, he has been with the Department of Electronic Engineering, City University of Hong Kong, where he is now a Chair Professor of information engineering. Dr. Li Ping received the IEE J J Thomson premium in 1993, the Croucher

Foundation Award in 2005 and a British Royal Academy of Engineering Distinguished Visiting Fellowship in 2010. He served as a member of the Board of Governors for IEEE Information Theory Society from 2010 to 2012.