Data-Aided Channel Estimation in Large Antenna Systems

Junjie Ma and Li Ping, *Fellow, IEEE*

***Abstract*—This paper is concerned with a uplink scheme for multicell large antenna systems. We study a channel estimation technique where partially decoded data is used to estimate the channel. We show that there are two types of interference components in this scheme that do not vanish even when the number of antennas grows to infinity. The first type, referred to as cross-contamination, is due to the correlation among the data signals from different users. The second type, referred to as self-contamination, is due to the dependency between the channel estimate and the estimation error. Cross contamination is in principle similar to pilot contamination in a conventional pilot-based channel estimation scheme, while self-contamination is unique for the data-aided scheme. For efficient use of the channel, the data part in a signaling frame is typically much longer than the pilot part for a practical system. Consequently, compared with pilot signals, data signals naturally have lower cross correlation. This fact reduces the cross-contamination effect in the data-aided scheme. Furthermore, self-contamination can be effectively suppressed by iterative processing. These results are confirmed by both analyses and simulations.**

***Index Terms*—Data-aided channel estimation, iterative channel estimation and signal detection, large antenna system, massive MIMO system, pilot contamination.**

I. Introduction

ASSUME that all other system parameters, such as the number of users, the average transmission power per user and rate per user, are fixed in a cellular system. Then, provided that perfect channel state information (CSI) at the base station (BS) is available, cross user interference can be completely suppressed when $M \to \infty$, where M is the number of antennas at the BS. This fact motivates the research activities on large (or massive) antenna systems [1]–[12].

Channel estimation is a challenging problem for a large antenna system [1]. It is shown in [1] that, if CSIR is to be estimated using non-orthogonal pilot signals, the effect of interference does not vanish even when $M \to \infty$. This phenomenon is referred to as "pilot contamination" [1]–[11]. The name comes from the fact that the underlying cause is the cross correlation among the pilot signals of different users. Pilot contamination may result in serious performance deterioration in a large antenna system.

The treatments for pilot contamination have been extensively studied. It is shown in [2] that pilot contamination can be reduced by asynchronous transmissions among neighboring cells in a time duplex division (TDD) system. The scheme in [2] may result in strong interference when two closely located mobile terminals in neighboring cells are transmitting and receiving at the same time [13]. Pilot contamination can also be mitigated by techniques based on coordinated channel estimation [3] or by multi-cell joint processing [4]. These techniques involve cooperation and information exchange among multiple users in neighboring cells. Moreover, the technique of [3] relies on specific conditions on channel covariance matrix and the claim of [4] is only valid for the asymptotic case of infinite antennas. The singular value decomposition (SVD)-based blind channel estimation scheme proposed in [5], [6] is shown to be effective in mitigating pilot contamination. The disadvantage of this scheme is that SVD is a costly operation for a large antenna system.

In this paper, we study a scheme in which partially decoded data are used to aid channel estimation [14], [15]. We derived an analytical approach that can predict the performance of the data-aided channel estimation scheme with reasonable accuracy. (Strictly speaking, the approach is semi-analytical if the decoder function is obtained via simulation.) Our focus is on the distortion at the receiver output in this scheme. We show that there are residual interference terms that remain bounded away from zero when $M \to \infty$. These residual interference terms, broadly referred to as "contamination", can be divided into two types. The first type, referred to as cross-contamination, is due to the correlation among the data signals from different users. This is in principle similar to the correlation among the pilots (or data) in a conventional scheme [1], [9, Remark 5]. The second type, referred to as self-contamination, results from the dependency between channel estimate and estimation error. (Recall that these two are independent in a classical minimum mean-square-error (MMSE) estimation for a linear Gaussian model.)

With analysis and simulation, we show that the two types of contamination may cause considerable error if not treated properly. We also show that cross-contamination can be reduced by using a long data frame (as long data sequences naturally have low cross correlation), and that self-contamination can be alleviated by iterative processing.

Compared with other alternatives [2]–[6], the scheme studied in this paper is noncooperative and does not rely on specific

Manuscript received January 12, 2014; revised April 13, 2014; accepted April 14, 2014. Date of publication April 30, 2014; date of current version May 19, 2014. This work was supported by a grant from the Research Grant Council of the Hong Kong SAR, China [Project No. CityU 118014]. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Xiqi Gao.

The authors are with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, SAR, China (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/TSP.2014.2321120

channel models. Therefore it is simpler and more flexible in practice.

This paper is organized as follows. Section II introduces the data-aided channel estimation scheme. The data detection performance is then studied in Section III. Simulation results of an iterative channel estimation and signal detection system are given in Section IV and conclusions are drawn in Section V.

Notations

Boldface symbols denote matrices or vectors. $\mathbf{0}$ denotes a matrix of all-zero entries. $\boldsymbol{I}_N$ denotes the N-by-N identity matrix. $\mathrm{E}[\cdot]$ represents the expectation operation. $\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{C})$ represents circular symmetric complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{C}$. We use $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ to denote complex conjugate, transpose and conjugate transpose of a matrix, respectively. $\|\cdot\|$ represents the 2-norm of a vector. $\otimes$ denotes the Kronecker product. $\xrightarrow{P}$ represents convergence in probability.

II. Data-Aided Channel Estimation

In this section, we study a data-aided channel estimation scheme based on certain available *a priori* data information. This information could be obtained from the output of a soft-input soft-output channel decoder. For simplicity, we assume that only the *a priori* data information is utilized in channel estimation in this section. The impact of pilots are briefly discussed in Section II-E.

A. System Model

Consider an L-cell system. We adopt the following settings. Each cell contains only one user. (We will briefly discuss multi-user situations in Section IV-F.) Each base station (BS) is equipped with M antennas while each mobile unit is equipped with only one antenna. We will focus on the uplink transmission of cell 1. Our task is to estimate the channel during a frame of J consecutive symbols in time. A frame of received signals over different antennas at the BS of cell 1 are denoted as

$$y_{1m}(j) = h_{1m}x_1(j) + \sum_{i\neq 1}^{L} h_{im}x_i(j) + n_{1m}(j), \quad m = 1,\ldots,M,\ j = 1,\ldots,J \quad (1)$$

where h_{im} is the equivalent channel gain from the user in cell i to the mth antenna of BS 1, $x_i(j) \sim \mathcal{CN}(0,1)$ the jth symbol transmitted by the user in cell i, $n_{1m}(j)$ the additive white Gaussian noise sample with mean zero and variance N_0. The equivalent channel h_{im} is represented as $h_{im} = \beta_i^{1/2} \cdot g_{im}$, where β_i includes the large scale fading factor and the transmit power, and $g_{im} \sim \mathcal{CN}(0,1)$ is a Rayleigh fading factor. We assume that $\{\beta_i\}$ are known at the receiver. We will focus on a quasi-static channel, in which the channel remains constant in a frame and change independently from frame to frame. Denote $\boldsymbol{y}_1(j) \equiv [y_{11}(j), y_{12}(j), \ldots, y_{1M}(j)]^{\mathrm{T}}$, $\boldsymbol{h}_i \equiv [h_{i1}, h_{i2}, \ldots, h_{iM}]^{\mathrm{T}}$ and $\boldsymbol{n}_1(j) \equiv [n_{11}(j), n_{12}(j), \ldots, n_{1M}(j)]^{\mathrm{T}}$, we can now rewrite (1) as

$$\boldsymbol{y}_1(j) = \boldsymbol{h}_1 x_1(j) + \sum_{i\neq 1}^{L} \boldsymbol{h}_i x_i(j) + \boldsymbol{n}_1(j),\ j = 1,\ldots,J. \quad (2)$$

B. A Priori Data Information

Let $\{s(j), \forall j\}$ be the side information (or *a priori* information) of $\{x_1(j), \forall j\}$. Later in Section IV we will see that $\{s(j), \forall j\}$ can be generated using decoding feedbacks. Denote

$$\hat{x}_1(j) \equiv \mathrm{E}\left[x_1(j) | s(j)\right]. \quad (3)$$

Here, $\hat{x}_1(j)$ is the conditional mean of $x_1(j)$, i.e., the MMSE estimator of $x_1(j)$ based on $s(j)$ [16]. Note that the side information $s(j)$ usually contains random errors and is thus random. From (3), $\hat{x}_1(j)$ depends on $s(j)$ and therefore is also a random variable. For notational convenience, we also define an estimation error:

$$\Delta x_1(j) \equiv x_1(j) - \hat{x}_1(j). \quad (4)$$

We make the following assumptions:

1) $s(j)$ is given in the form
$$s(j) = x_1(j) + n(j) \quad (5)$$
where $x_1(j) \sim \mathcal{CN}(0,1)$ and $n(j) \sim \mathcal{CN}(0, \sigma_n^2)$ is independent of $x_1(j)$; and
2) both $\{x_1(j), \forall j\}$ and $\{n(j), \forall j\}$ contain independent identically distributed (i.i.d.) samples.

Under the linear Gaussian model in Assumption 1), the MMSE estimation $\hat{x}_1(j)$ and the estimation error $\Delta x_1(j)$ are mutually independent zero mean Gaussian random variables, following the well-known MMSE property [16]. Let $v_{\Delta x}$ be the variance of $\Delta x_1(j)$. As average power of $x(j)$ is normalized to 1, the variance of $\hat{x}_1(j)$ is $1 - v_{\Delta x}$. In particular, the *a priori* information is perfect when $v_{\Delta x} = 0$, and very unreliable when $v_{\Delta x} \approx 1$.

Under Assumption 2), both $\{\hat{x}_1(j), \forall j\}$ and $\{\Delta x_1(j), \forall j\}$ contain i.i.d. samples.

Assumptions 1) and 2) are introduced for simplicity of performance analysis in Section III. Similar Gaussian assumptions have been widely used in the literature on iterative detection [17], [18]. Assumptions 1) and 2) can be justified if superposition coded modulation (SCM) [19], [20] is involved. However, as will be shown in the numerical results in Section III-G (see Fig. 4), the results derived based on Assumptions 1) and 2) are reasonably accurate for other modulation methods such as Gray signaling.

C. Channel Estimation

We first combine the J time samples $\{\boldsymbol{y}_1(j),\ j = 1, 2, \ldots, J\}$ as follows,

$$\boldsymbol{z} = \sum_{j=1}^{J} \hat{x}_1^*(j)\boldsymbol{y}_1(j). \quad (6)$$

Substituting (2) into (6), we have

$$\boldsymbol{z} = \left(\sum_{j=1}^{J} \hat{x}_1^*(j)x_1(j)\right) \cdot \boldsymbol{h}_1 + \sum_{i\neq 1}^{L} \left(\sum_{j=1}^{J} \hat{x}_1^*(j)x_i(j)\right) \cdot \boldsymbol{h}_i + \sum_{i=1}^{J} \hat{x}_1^*(j)\boldsymbol{n}_1(j). \quad (7)$$

For notational brevity, denote $\hat{\boldsymbol{x}}_1 \equiv [\hat{x}_1(1), \ldots, \hat{x}_1(J)]^{\mathrm{T}}$, $\Delta\boldsymbol{x}_1 \equiv [\Delta x_1(1), \ldots, \Delta x_1(J)]^{\mathrm{T}}$ and $\boldsymbol{x}_i \equiv [x_i(1) \ldots, x_i(J)]^{\mathrm{T}}$. We can then rewrite (7) as

$$\begin{aligned}\boldsymbol{z} &= (\hat{\boldsymbol{x}}_1^{\mathrm{H}}\boldsymbol{x}_1)\cdot\boldsymbol{h}_1 + \sum_{i\neq 1}^{L}(\hat{\boldsymbol{x}}_1^{\mathrm{H}}\boldsymbol{x}_i)\cdot\boldsymbol{h}_i + \sum_{j=1}^{J}\hat{x}_1^*(j)\boldsymbol{n}_1(j)\\ &= \|\hat{\boldsymbol{x}}_1\|^2\cdot\boldsymbol{h}_1 + \tilde{\boldsymbol{n}},\\ \tilde{\boldsymbol{n}} &\equiv (\hat{\boldsymbol{x}}_1^{\mathrm{H}}\Delta\boldsymbol{x}_1)\boldsymbol{h}_1 + \sum_{i\neq 1}^{L}(\hat{\boldsymbol{x}}_1^{\mathrm{H}}\boldsymbol{x}_i)\boldsymbol{h}_i + \sum_{j=1}^{J}\hat{x}_1^*(j)\boldsymbol{n}_1(j). \quad (8)\end{aligned}$$

In the above, $\|\hat{\boldsymbol{x}}_1\|^2$ is known at the receiver and $(\hat{\boldsymbol{x}}_1^{\mathrm{H}}\Delta\boldsymbol{x}_1)\cdot\boldsymbol{h}_1$ is treated as uncorrelated noise. The linear MMSE (LMMSE) estimator [16] of $\boldsymbol{h}_1$ based on (8) is given by

$$\hat{\boldsymbol{h}}_1 = \theta\cdot\boldsymbol{z} \quad (9a)$$

where [16]

$$\begin{aligned}\theta &\equiv \frac{\|\hat{\boldsymbol{x}}_1\|^2\cdot\beta_1}{\|\hat{\boldsymbol{x}}_1\|^4\cdot\beta_1 + \|\hat{\boldsymbol{x}}_1\|^2\cdot\left(v_{\Delta x}\cdot\beta_1 + \sum_{i\neq 1}^{L}\beta_i + N_0\right)}\\ &= \frac{\beta_1}{\|\hat{\boldsymbol{x}}_1\|^2\cdot\beta_1 + v_{\Delta x}\cdot\beta_1 + \sum_{i\neq 1}^{L}\beta_i + N_0}. \quad (9b)\end{aligned}$$

The above procedure is optimal under the linear MMSE principle, as explained in Appendix A.

Note that $\tilde{\boldsymbol{n}}$ and $\boldsymbol{h}_1$ in (8) are not mutually independent, and also $\tilde{\boldsymbol{n}}$ is not Gaussian (as $(\hat{\boldsymbol{x}}_1^{\mathrm{H}}\Delta\boldsymbol{x}_1)\boldsymbol{h}_1$ and $(\hat{\boldsymbol{x}}_1^{\mathrm{H}}\boldsymbol{x}_i)\boldsymbol{h}_i$ are not Gaussian). The channel estimate $\hat{\boldsymbol{h}}_1$ in (9a) and the estimation error $\boldsymbol{h}_1 - \hat{\boldsymbol{h}}_1$ are also not independent, as opposed to the case of LMMSE estimation for a linear Gaussian model. This fact results in the so-called self-contamination, as will be discussed in Section III-D.

D. Impact of Data Length on Estimation Quality

The estimation quality in (9) can be measured by the following signal-to-interference-plus-noise ratio (SINR) contained in $\hat{\boldsymbol{h}}_1$ (or $\boldsymbol{z}$),

$$\begin{aligned}SINR_h &= \frac{\|\hat{\boldsymbol{x}}_1\|^4\cdot\beta_1}{\|\hat{\boldsymbol{x}}_1\|^2\cdot\left(v_{\Delta x}\cdot\beta_1 + \sum_{i\neq 1}^{L}\beta_i + N_0\right)}\\ &= \frac{\|\hat{\boldsymbol{x}}_1\|^2\cdot\beta_1}{v_{\Delta x}\cdot\beta_1 + \sum_{i\neq 1}^{L}\beta_i + N_0}. \quad (10)\end{aligned}$$

From the discussions in Section II-B, $\hat{\boldsymbol{x}}_1$ is also a random vector. The elements of $\hat{\boldsymbol{x}}_1$ are i.i.d. with mean zero and variance $(1 - v_{\Delta x})$. Taking average of $SINR_h$ over the distribution of $\hat{\boldsymbol{x}}_1$, we have

$$SINR_h = \frac{J\cdot(1 - v_{\Delta x})\cdot\beta_1}{v_{\Delta x}\cdot\beta_1 + \sum_{i\neq 1}^{L}\beta_i + N_0}. \quad (11)$$

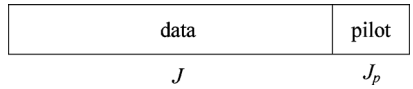


Fig. 1. Transmission frame structure of a communication system involving channel estimation.

From (11), we can see that $SINR_h$ grows linearly with J. This is expected, since $\hat{\boldsymbol{h}}_1$ is obtained using J independent observations $\{\boldsymbol{y}_1(j)\}$ in (6).

From (11), we can also see that $SINR_h$ depends on $v_{\Delta x}$, i.e., the accuracy of the *a priori* data information. In Section IV, we will see that $v_{\Delta x}$ can be reduced gradually using an iterative process.

E. Discussions

Consider the frame structure in Fig. 1. A transmission frame is divided into two phases: pilot phase and data phase. The total number of symbols in a frame, including pilots and data, is limited by the channel coherence time in practice. For example, one millisecond coherence interval corresponds to roughly 200 symbols in a typical LTE (Long Term Evolution) communication scenario [1], [11].

Conventionally, channel is estimated only based on the J_p pilots. In principle, we can improve the estimation quality by increasing J_p. However, since $J + J_p$ is limited by the channel coherence time, increasing J_p will decrease J and thus reduce the effective data rate.

For the data-aided scheme, the data frame length J is only limited by the channel coherence time and typically $J \gg J_p$. Therefore, the data-aided channel estimation scheme is more attractive than the pilot-only one.

In practice, pilots can also be used for a data-aided scheme, mainly for initialization purpose. The details will be discussed in Section IV.

III. Data Detection

In this section, we study the impact of channel estimation error on the performance of data detection. For ease of analysis, we assume that the receiver is based on a simple matched filter (MF) detector.

A. Data Detection

The received signal in (2) can be expressed in a signal-plus-distortion form,

$$\boldsymbol{y}_1(j) = \hat{\boldsymbol{h}}_1 x_1(j) + \boldsymbol{\eta}(j) \quad (12a)$$

where

$$\boldsymbol{\eta}(j) \equiv (\boldsymbol{h}_1 - \hat{\boldsymbol{h}}_1)x_1(j) + \sum_{i\neq 1}^{L}\boldsymbol{h}_i x_i(j) + \boldsymbol{n}_1(j). \quad (12b)$$

To maintain low receiver complexity, a simple matched filter (MF) detector is employed:

$$\begin{aligned}\hat{\boldsymbol{h}}_1^{\mathrm{H}}\boldsymbol{y}_1(j) = &\underbrace{\|\hat{\boldsymbol{h}}_1\|^2\cdot x_1(j)}_{\text{signal } S} + \underbrace{\hat{\boldsymbol{h}}_1^{\mathrm{H}}(\boldsymbol{h}_1 - \hat{\boldsymbol{h}}_1)\cdot x_1(j)}_{\text{self-interference } I_1}\\ &+ \sum_{i\neq 1}^{L}\underbrace{\hat{\boldsymbol{h}}_1^{\mathrm{H}}\boldsymbol{h}_i\cdot x_i(j)}_{\text{cross-interference } I_i} + \underbrace{\hat{\boldsymbol{h}}_1^{\mathrm{H}}\boldsymbol{n}_1(j)}_{\text{noise } N}. \quad (13)\end{aligned}$$

In (13), $\|\hat{\boldsymbol{h}}_1\|^2 x_1(j)$ is the desired signal, $\sum_{i\neq 1}^{L} \hat{\boldsymbol{h}}_1^{\mathrm{H}} \boldsymbol{h}_i \cdot x_i(j)$ represents the cross-interference from other cells. Since the detection is based on $\hat{\boldsymbol{h}}_1$, we treat $\hat{\boldsymbol{h}}_1^{\mathrm{H}}(\boldsymbol{h}_1 - \hat{\boldsymbol{h}}_1)\cdot x_1(j)$ as interference although it contains a part of the desired signal $x_1(j)$. We will refer to $\hat{\boldsymbol{h}}_1^{\mathrm{H}}(\boldsymbol{h}_1 - \hat{\boldsymbol{h}}_1)\cdot x_1(j)$ as self-interference.

B. Signal and Noise Power

The average signal and noise power in (13) are respectively given by

$$\mathrm{E}\left[|S|^2\right] = \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^4\right], \tag{14}$$

$$\mathrm{E}\left[|N|^2\right] = \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^2\right]\cdot N_0. \tag{15}$$

We show in Appendices B-B and B-C (see (53d) and (59c)) that when both M and J are large

$$\mathrm{E}\left[|S|^2\right] \approx \beta_1^2 \cdot M^2, \tag{16}$$

$$\mathrm{E}\left[|N|^2\right] \approx \beta_1 \cdot M \cdot N_0. \tag{17}$$

From (16) and (17), we see that the ratio of signal power to noise power is of order $O(M)$. This array gain comes from the fact that signals are combined "coherently" while noise is combined randomly.

C. Cross-Interference Power

Let us first discuss the following interference term inside the summation in (13)

$$I_i \equiv \hat{\boldsymbol{h}}_1^{\mathrm{H}} \boldsymbol{h}_i \cdot x_i(j), \quad i \neq 1. \tag{18}$$

When both M and J are large, we show that

$$\mathrm{E}\left[|I_i|^2\right] \approx \frac{\beta_i^2}{1-v_{\Delta x}} \cdot \frac{M^2}{J} + \beta_i \cdot \beta_1 \cdot M, \quad i \neq 1. \tag{19}$$

The detailed derivations of (19) are provided in Appendix B-D.

Comparing (16) and (19), we can see that the signal power and cross-interference power are both in the order of $O(M^2)$. If J is fixed, cross-interference imposes a limit on the SINR even when $M \to \infty$. This is similar to the pilot-contamination effect in a pilot based channel estimation scheme [1], except that the problem is caused by the correlation among data signals instead of pilot signals. A similar effect has also been discussed in [9, Remark 5]. In this paper, we will call this effect "cross-contamination".

Now note that the first term in (19) (which contains M^2) is attenuated by a factor of J. In a practical system, J can be quite large (e.g., around 200 for a typical scenario in LTE [1], [11]), resulting in effective suppression of cross-contamination. This is a main advantage of the data-aided scheme. A potential problem arises in (19) when $v_{\Delta x} \approx 1$, as then $1/(1-v_{\Delta x})$ become large. This corresponds to the situation that the *a priori* information on $\boldsymbol{x}_1$ is very unreliable (see the discussion below Assumption 2) in Section II-B and so the performance of the data-aided scheme is poor. Later in Section IV we will outline an iterative process that can overcome this problem.

D. Self-Interference Power

Let us next consider the self-interference term

$$I_1 \equiv \hat{\boldsymbol{h}}_1^{\mathrm{H}} \left(\boldsymbol{h}_1 - \hat{\boldsymbol{h}}_1\right)\cdot x_1(j). \tag{20}$$

For a large J and M, we derive in Appendix B-E the following approximations for $\mathrm{E}[|I_1|^2]$,

$$\mathrm{E}\left[|I_1|^2\right] \approx \frac{\beta_1^2 \cdot v_{\Delta x}}{1-v_{\Delta x}} \cdot \frac{M^2}{J}. \tag{21}$$

Similar to I_i in (19), the power of I_1 is also of the order $O(M^2)$. We refer to I_1 as "self-contamination". As discussed in Section II-C, $\hat{\boldsymbol{h}}_1$ and $\boldsymbol{h}_1 - \hat{\boldsymbol{h}}_1$ are not independent. This dependency essentially causes the self-contamination effect. This effect does not exist in a conventional pilot-based scheme [7]–[9] where pilots are assumed known at the receiver.

We can observe from (21) that, similar to cross-contamination, self-contamination is also attenuated by a factor of J. Furthermore, from (21), self-contamination will vanish when $v_{\Delta x} = 0$, i.e., when the *a priori* information on $\boldsymbol{x}_1$ is perfect. This is different from cross-contamination. The latter does not vanish even when $v_{\Delta x} = 0$, as can be seen from (19).

E. Correlation of Signal and Self-Interference

If self-interference is uncorrelated with signal, it can be treated as additive Gaussian noise, based on, e.g., the worst case Gaussian noise argument [21]. However, since the channel estimate and the estimation error are not independent (see the discussions in Section II-C), signal and self-interference in (13) are correlated. This correlation problem may complicate the performance analysis of the data detector.

We quantify the correlation and define the following correlation coefficient [22]

$$\alpha \equiv \frac{\left|\mathrm{E}[S^* \cdot I_1]\right|^2}{\mathrm{E}[|S|^2]\cdot \mathrm{E}[|I_1|^2]}. \tag{22}$$

In Appendix B-F, we show that, when J and M are large,

$$\alpha \approx \frac{v_{\Delta x}}{1-v_{\Delta x}} \cdot \frac{1}{J} \ll 1. \tag{23}$$

For this reason, we will ignore this correlation problem.

F. Numerical Results

Consider a 7-cell cellular system with normalized cell radius. The users are assumed to be uniformly randomly located. Assume a fourth-power path-loss attenuation law. Let us focus on a link from a user in cell i to the BS in cell j. The path-loss of this link is given by $\gamma \cdot d_{i\to j}^{-4}$, where γ is a constant and $d_{i\to j}$ the distance of this link. For simplicity, log-normal shadowing is not considered.

We adopt a power control policy [23] such that the receive power of a user to its own BS is P_{target}. The transmit power of this user is then $P_{target}/(\gamma \cdot d_{i\to i}^{-4})$. This implies that β_i is given by

$$\beta_i = d_{i\to 1}^{-4} \cdot \left(\frac{P_{target}}{d_{i\to i}^{-4}}\right).$$

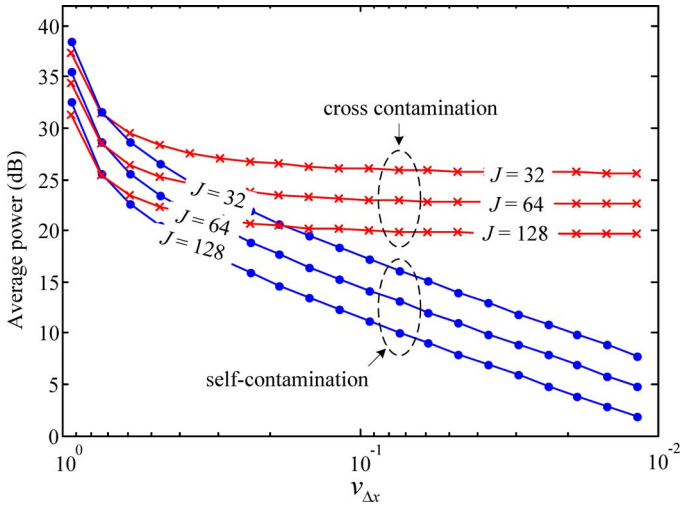


Fig. 2. Average power of cross-contamination and self-contamination for $M = 128$ and different J, $\mathrm{SNR}_{\mathrm{target}} = 0$ dB.

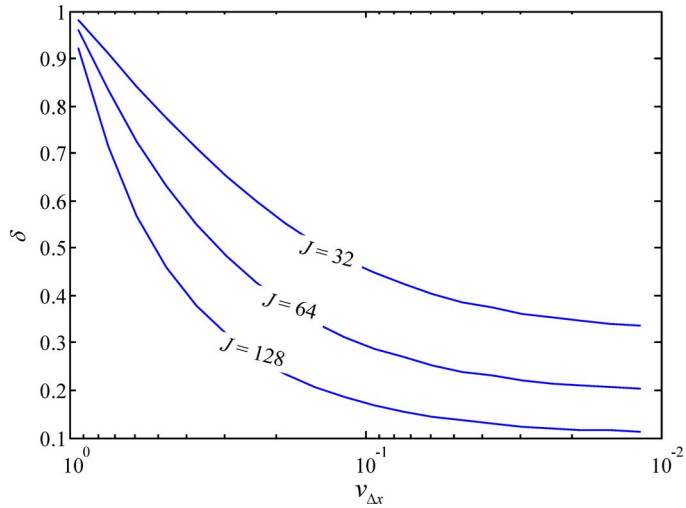


Fig. 3. The power ratio of contamination for $M = 128$ and different J. $\mathrm{SNR}_{\mathrm{target}} = 0$ dB. $\{\beta_i\}$ are generated in the same way as in Fig. 2.

(Recall that β_i includes both path-loss and transmit power.) With the above setting, the receive SNR is given by $\mathrm{SNR}_{\mathrm{target}} = P_{target}/N_0$.

Following the discussions in Sections III-C and III-D, and for convenience, we define the "cross-contamination" and "self-contamination" terms as follows (see the discussions related to (19) and (21)):

$$\text{cross-contamination} : \frac{\sum_{i\neq 1}^{L}\beta_i^2}{1-v_{\Delta x}} \cdot \frac{M^2}{J}, \tag{24a}$$

$$\text{self-contamination} : \frac{\beta_1^2 \cdot v_{\Delta x}}{1-v_{\Delta x}} \cdot \frac{M^2}{J}. \tag{24b}$$

In Fig. 2, the average of cross-contamination and self-contamination over the distributions of $\{\beta_i\}$ against $v_{\Delta x}$ are plotted for different J values. The number of antennas is fixed to be $M = 128$. From Fig. 2, we have the following observations:

- When $v_{\Delta x}$ is relatively large, both cross-contamination and self-contamination are serious.
- For a fixed J, both cross-contamination and self-contamination decrease as $v_{\Delta x}$ becomes smaller. When $v_{\Delta x} \to 0$, cross-contamination converges to a constant while self-contamination vanishes.
- Both cross-contamination and self-contamination reduce as J becomes larger.

To quantify the contamination effect in a data-aided scheme, we define the following ratio:

$$\delta = \frac{\text{power (self-contamination + cross-contamination)}}{\text{power (self-interference + cross-interference + noise)}}. \tag{25}$$

The plots of δ against $v_{\Delta x}$ are given in Fig. 3 for different J values, where the curves are obtained based on (17), (19), (21) and (24). We can see that the contamination effect becomes marginal when J is sufficiently large and $v_{\Delta x}$ is small.

G. SINR Performance

From (13), the SINR contained in the output of the MF detector, denoted as $SINR_x$, is defined in (27), at the bottom of the page. Furthermore, using (16), (17), (19) and (21), we have

$$SINR_x = \frac{\mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^4\right]}{\mathrm{E}\left[|\hat{\boldsymbol{h}}_1^{\mathrm{H}}(\boldsymbol{h}_1-\hat{\boldsymbol{h}}_1)|^2\right] + \sum_{i\neq 1}^{L}\mathrm{E}\left[|\hat{\boldsymbol{h}}_1^{\mathrm{H}}\boldsymbol{h}_i|^2\right] + \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^2\right]\cdot N_0}. \tag{27}$$

$$SINR_x^{app} = \frac{\beta_1^2 \cdot M^2}{\frac{\beta_1^2\cdot v_{\Delta x}}{1-v_{\Delta x}}\cdot\frac{M^2}{J} + \sum_{i\neq 1}^{L}\left(\frac{\beta_i^2}{1-v_{\Delta x}}\cdot\frac{M^2}{J} + \beta_i\cdot\beta_1\cdot M\right) + \beta_1\cdot N_0\cdot M} \tag{28}$$

$$= \frac{\beta_1^2}{\underbrace{\left(\frac{\beta_1^2\cdot v_{\Delta x}}{1-v_{\Delta x}} + \sum_{i\neq 1}^{L}\frac{\beta_i^2}{1-v_{\Delta x}}\right)\cdot\frac{1}{J}}_{\text{contamination}} + \underbrace{\left(\sum_{i\neq 1}^{L}\beta_i\cdot\beta_1 + \beta_1\cdot N_0\right)\cdot\frac{1}{M}}_{\text{conventional}}}. \tag{29}$$

the approximate SINR expression $SINR_x^{app}$ when M and J are large in (28)-(29) at the bottom of the previous page. The approximate SINR is more insightful in characterizing the performance of the data-aided channel estimation scheme. From (29), we have the following observations:

- The contamination terms are attenuated by a factor of J, as discussed in Sections III-C and III-D. The conventional interference terms, including inter-cell interference and noise, are attenuated by a factor of M.
- When J is fixed and $M \to \infty$, $SINR_x^{app}$ is bounded. The limiting value of $SINR_x^{app}$ is

$$SINR_x^{app} \xrightarrow{M\to\infty} \frac{\beta_1^2}{\left(\frac{\beta_1^2 \cdot v_{\Delta x}}{1-v_{\Delta x}} + \sum_{i\neq 1}^{L} \frac{\beta_i^2}{1-v_{\Delta x}}\right) \cdot \frac{1}{J}}. \quad (26)$$

  In this scenario, contamination is dominant and conventional interference terms disappear. Clearly, $SINR_x^{app}$ is proportional to J.
- When both M and J go to infinity, $SINR_x^{app}$ can be unbounded.

In Fig. 4, $SINR_x$ and $SINR_x^{app}$ are plotted for various values of M and J. The curve with an "approximation" legend is obtained from averaging (28) over $\{\beta_i\}$. For the two curves with a "simulation" legend, we first generate $\{\boldsymbol{x}_i\}$ and $\hat{\boldsymbol{x}}_1$, then estimate the channel using $\hat{\boldsymbol{x}}_1$, and finally measure $SINR_x$ according to (27) using Monte Carlo simulations. The difference between the two simulated curves is the generation of $\{\boldsymbol{x}_i\}$ and $\hat{\boldsymbol{x}}_1$. For the solid curve, $\{\boldsymbol{x}_i\}$ and $\hat{\boldsymbol{x}}_1$ are generated based on the Gaussian assumptions detailed in Sections II-A and II-B. For the dashed curve, $\{\boldsymbol{x}_i\}$ are generated from Gray-mapped 16QAM signal constellations. $\hat{\boldsymbol{x}}_1$ is generated following a standard approach in EXIT [17] analysis for bit-interleaved coded modulation (BICM) systems: we first produce the *a priori* log-likelihood ratios (LLRs) related to $\boldsymbol{x}_1$ and then estimate $\hat{\boldsymbol{x}}_1$ (and also $v_{\Delta x}$) based on the LLRs. (For details of estimating $\hat{\boldsymbol{x}}_1$ and $v_{\Delta x}$, refer to [19].) In this approach, the data variances are different for different symbols. In the channel estimator, we simply approximate the variances by the sample average, which is a common practice in the literature [24].

From Fig. 4, we can see that the two simulated curves are almost indistinguishable. This validates the usefulness of the Gaussian assumptions made in Sections II-A and II-B. Moreover, $SINR_x^{app}$ (28) (obtained based on the Gaussian assumptions) is a good approximation of $SINR_x$, except when $v_{\Delta x} \approx 1$. The inaccuracy is due to the approximation $\beta_1 \cdot J \cdot (1-v_{\Delta x}) \gg v_{\Delta x} \cdot \beta_1 + \sum_{k\neq 1}^{L} \beta_k + N_0$ in deriving (16), (17), (19) and (21) in Appendix B. This approximation is loose when $v_{\Delta x} \approx 1$ and J is relatively small.

We can observe from Fig. 4 that SINR grows as J and M increase. Also, $SINR_x^{app}$ is a function of $v_{\Delta x}$. In the next section, a practical iterative processing will be introduced where $v_{\Delta x}$ can be gradually reduced.

IV. Iterative Channel Estimation and Signal Detection

In Sections II and III, we studied channel estimation and data detection separately. In this section, we discuss these two functions in an iterative joint channel estimation and data detection process [14], [15].

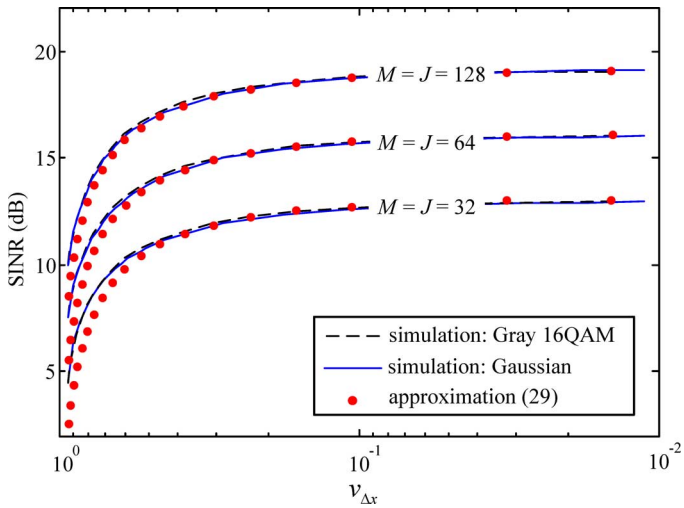


Fig. 4. Average SINR of the data aided channel estimation scheme for different M and J. $\mathrm{SNR}_{\mathrm{target}} = 0$ dB. $\{\beta_i\}$ are generated in the same way as in Fig. 2.

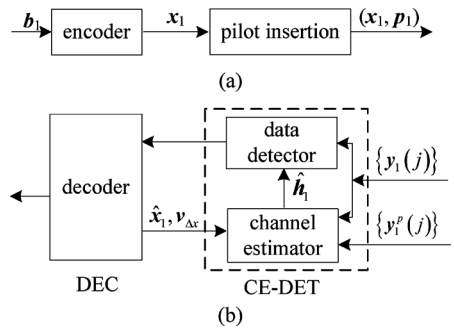


Fig. 5. Transceiver structure for user 1. (a) transmitter. (b) receiver.

A. Transmitter Structure

The transmitter structure for the user in cell 1 is illustrated in Fig. 5(a). We assume that one transmitted codeword spans several coherence blocks. These coherence blocks may be transmitted consecutively in time, or concurrently over different OFDM sub-carriers. For simplicity, we will assume that the channel conditions for different blocks are independent.

In each coherence block, the first J symbols are data symbols and the other J_p symbols are pilots. At the transmitter side, the input binary information sequence $\boldsymbol{b}_1$ is first processed by the encoder (which includes FEC coding, random interleaving and mapping) to get the data symbol $\boldsymbol{x}_1$, which are then multiplexed with the random pilot symbols $\boldsymbol{p}_1$ and transmitted through the antennas. The received signals corresponding to data and pilot transmissions are represented by $\boldsymbol{y}_1(j), j = 1, \ldots, J$ and $\boldsymbol{y}_1^p(j)$, $j = J+1, \ldots, J+J_p$ respectively. We set $J_p = 1$ for our simulations.

B. Iterative Channel Estimation and Signal Detection

The receiver structure is shown in Fig. 5(b), where iterative channel estimation and signal detection is adopted. The channel estimator and the signal detector have been discussed in Sections II and III respectively. The decoder module takes the output of the MF data detector as input and generates $\hat{\boldsymbol{x}}_1$ and $v_{\Delta x}$ as outputs. The decoder module can be further divided into

three parts: de-mapper, binary decoder, and mapper [19]. These operations are standard and will not be discussed here. For more details, refer to [19].

Pilot based channel estimation is performed in the first iteration. In subsequent iterations, data-aided channel estimation detailed in Section II will be employed with the help of the decoder feedback. The processions of the channel estimator, data detector and decoder are executed iteratively until convergence.

C. Some Details

In Section II, we assumed that only data is used for channel estimation. In practice, data and pilot can be combined to estimate the channel. The related discussions can be found in [14], [31].

There is another subtle point. From (6) and (9), the channel estimator $\hat{\boldsymbol{h}}_1$ is expressed as

$$\hat{\boldsymbol{h}}_1 = \frac{\beta_1 \cdot \sum_{j'=1}^{J} \boldsymbol{y}_1(j') \cdot \hat{x}_1^*(j')}{\sum_{j'=1}^{J} |\hat{x}_1(j')|^2 \cdot \beta_1 + v_{\Delta x} \cdot \beta_1 + \sum_{i \neq 1}^{L} \beta_i + N_0}. \quad (30)$$

It is well known that extrinsic message should be used in a turbo-type iterative receiver [25]. To meet the "extrinsic" requirement, the channel estimate used to detect the jth data symbol is modified as follows:

$$\hat{\boldsymbol{h}}_1^{ext}(j) = \frac{\beta_1 \cdot \sum_{j' \neq j}^{J} \boldsymbol{y}_1(j') \cdot \hat{x}_1^*(j')}{\sum_{j' \neq j}^{J} |\hat{x}_1(j')|^2 \cdot \beta_1 + v_{\Delta x} \cdot \beta_1 + \sum_{i \neq 1}^{L} \beta_i + N_0}, \quad (31)$$

where $\hat{x}_1(j)$ and $\boldsymbol{y}_1(j)$ are excluded in the summations. We will call $\hat{\boldsymbol{h}}_1^{ext}(j)$ in (31) *extrinsic channel estimation*. In simulation, we observed that the performance difference between (30) and (31) is marginal.

D. Complexity Analysis

Denote the complexities (per J data symbols) of the channel estimator, signal detector and decoder as C_{CE}, C_{DET} and C_{DEC} respectively. For the conventional pilot based scheme, only one channel estimation is required for J data symbols and each antenna can be estimated independently, therefore C_{CE} is $O(M)$. The complexity of the MF combining is $O(M)$ per data symbol, so we have $C_{\mathrm{DET}} = O(M \cdot J)$. The decoding complexity C_{DEC} depends on the FEC code used (but is independent of M).

We now consider the complexity of the data-aided scheme per iteration. The complexities of the signal detector and the decoder are the same as those of the pilot based one. For the channel estimator, since $O(J)$ combining operations are required to estimate each channel coefficient (see (6)), we have $C_{\mathrm{CE}} = O(M \cdot J)$, which is of the same order as C_{DET}.

From the above analysis, the complexity of the data-aided scheme per iteration is roughly the same as the pilot based one. The extra complexity is introduced by iteration. However, if the pilot based scheme already involves iterative processing, e.g., in BICM with iterative decoding (BICM-ID) [26] or turbo equalization [24] systems, the extra complexity of data-aided channel estimation is not a problem.

E. Evolution Analysis

The EXIT chart [17] technique is a useful tool for the analysis of an iterative receiver involving two constituent modules, in which each module is characterized by a transfer function and the overall performance is determined by the fixed point. To apply this technique, we divide the receiver in Fig. 5(b) into two modules: the decoder module and the CE-DET module, with the latter consisting of the channel estimator and the data detector. Instead of mutual information, we will characterize the two modules using their SINR versus variance relationships [27]. This is because our discussions in Section III readily provide such relationship for CE-DET. The followings are some details.

We characterize the decoder module by the transfer function $v_{\Delta x} = \psi(SINR_x)$, which can be generated using standard Monte Carlo simulations [27] and will not be discussed here. We characterize the CE-DET module using a transfer function $SINR = \phi(v_{\Delta x})$. Here, $SINR_x$ is defined in (27). We can obtained $\phi(\cdot)$ using two approaches. The first approach is simply to adopt the approximation in (28), which is fast but less accurate. We will denote the resultant function as $\phi_1(\cdot)$. The second approach is through numerical simulation. We denote the resultant function as $\phi_2(\cdot)$. The generation of $\phi_2(\cdot)$ is similar to that in Fig. 4, except now both pilots and data are used in the channel estimator, as mentioned in Section IV-C. When $v_{\Delta x} = 1$, no *a priori* data information is available and pilot only channel estimation is adopted.

Note that $\phi_1(\cdot)$ is formulated under the assumption that there is no pilot used (see (28)). This assumption results in certain discrepancy between $\phi_1(\cdot)$ and $\phi_2(\cdot)$. In particular, $\phi_1(1)$ is zero ($-\infty$ in dB) while $\phi_2(1)$ takes a non-zero value provided by pilots.

An example of the two $\phi(\cdot)$ functions are given in Fig. 6, together with a $\psi(\cdot)$ function for a rate-1/2 convolution code with Gray-mapped 256-QAM modulation. We can see from Fig. 6 that $\phi_1(\cdot)$ and $\phi_2(\cdot)$ are reasonably close except when $v_{\Delta x} \approx 1$. Using the principles developed in [17], [27], the performance of the underlying system can be predicted from the transfer functions.

F. Simulation Results

The BER performances of the above iterative scheme, the conventional pilot based one and the SVD blind estimation scheme [6] are demonstrated in Fig. 7. Define ρ to be the power of pilot symbol to the power of data symbol.

From Fig. 7, we can see that the BER performances for both $\rho = 1$ and $\rho = 100$ are very poor for the conventional pilot-only scheme. As the problem is caused by the correlation among pilots, increasing pilot power alone (even to an extremely large value of $\rho = 100$) cannot solve the problem.

On the other hand, the data-aided channel estimation technique can improve the BER performance drastically. After only 4 iterations, the performance is reasonably close to the benchmark scheme with perfect CSI. Clearly, the iterative scheme is very effective in treating contamination. The predicted BER performance using the evolution analysis in Section IV-E is also included. We see that it is a good approximation of the simulated one.

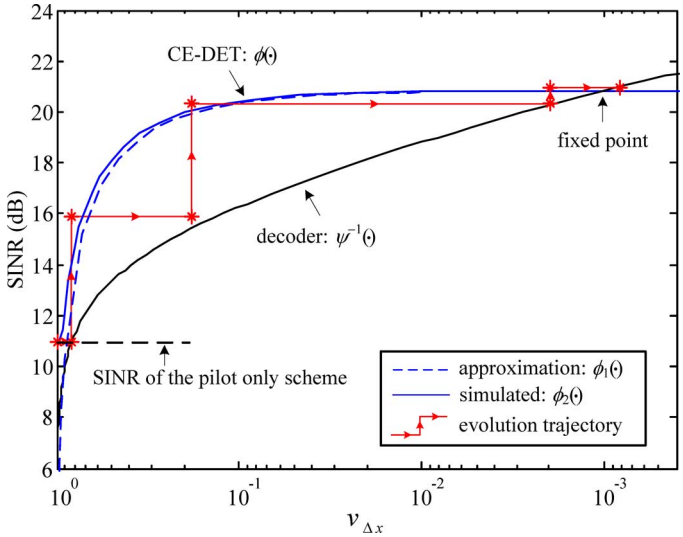


Fig. 6. An example of evolution trajectory. $L = 7$, $M = 128$, For simplicity, we set $\beta_1 = 1$ and $\beta_i = 0.1$ for $i \neq 1$. $J = 127$, $J_p = 1$. A code word spans 4096 symbols. The rate-1/2 $(23, 35)_8$ convolutional code with Gray-mapped 256QAM is employed. Channel SNR $= 4$ dB.

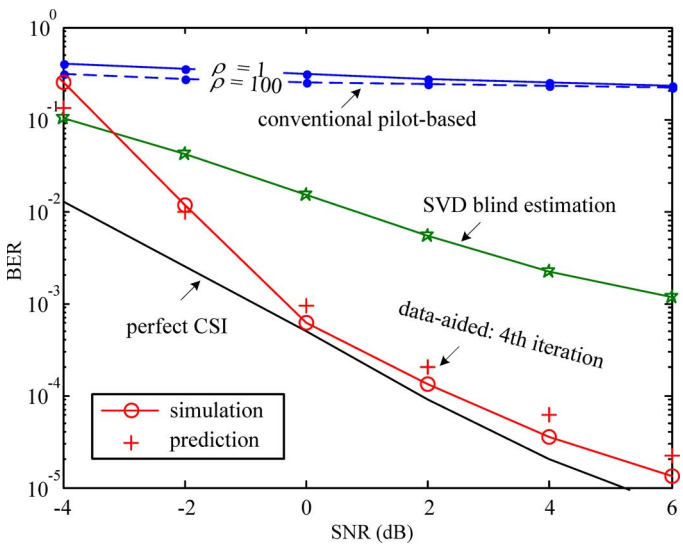


Fig. 7. BER performances of the pilot based channel estimation scheme, the data-aided scheme and the SVD blind scheme. $L = 7$ and $M = 128$. $J = 127$. $\beta_1 = 1$, $\beta_i = 0.1$ for $i \neq 1$. $J_p = 1$. A codeword spans 64 coherence blocks. The rate-1/2 $(23, 35)_8$ convolutional code is employed with Gray-mapped 256-QAM modulation. MAP de-mapping is used. SNR is defined as the ratio of data symbol power to noise power.

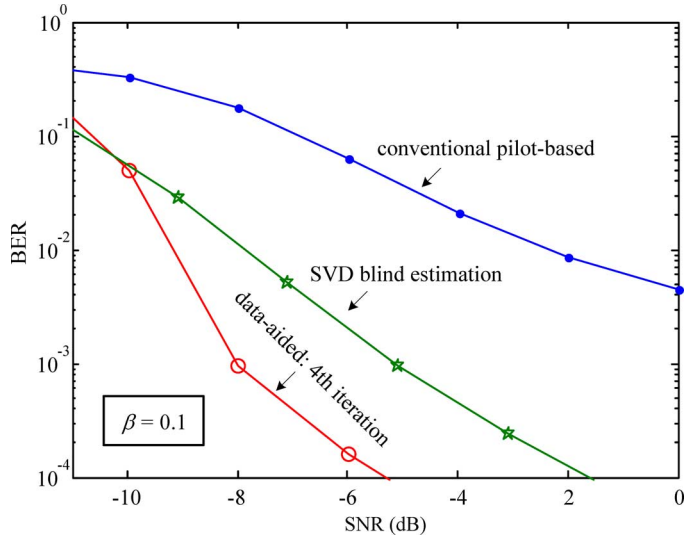


Fig. 8. BER performances in a multiple-user system. The number of users is $K = 4$. $\beta_1 = 1$, and $\beta_i = \beta$ for $i \neq 1$. 64-QAM modulation is employed. Other parameters are the same as Fig. 7.

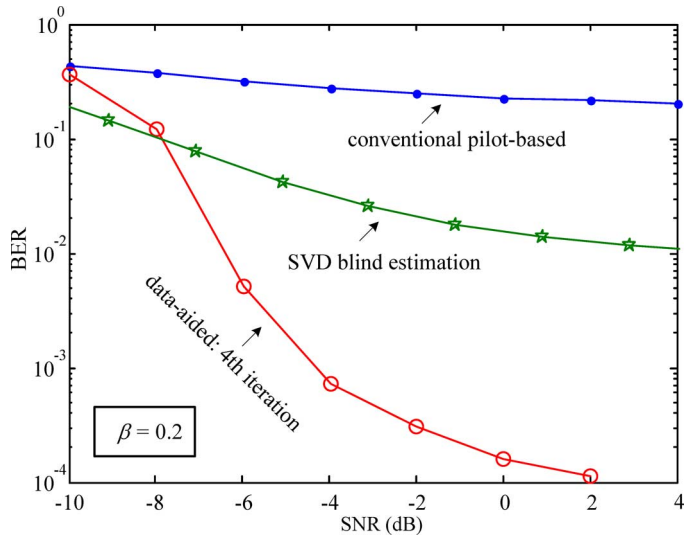


Fig. 9. $\beta_i = \beta = 0.2$ for $i \neq 1$. Other parameters are the same as Fig. 8.

It can also be observed from Fig. 7 that the performance of the blind SVD scheme is much better than that of the conventional pilot only scheme, but still worse that data-aided one in the high SNR region.

In the system model discussed in Section II-A, we assumed that each cell contains only one user. The discussions on data-aided channel estimation can also be extended to a system with multiple users in a cell if the interferences from the inter-cell and intra-cell users are simply treated in the same way. In Figs. 8 and 9, we provide simulation results for such a multi-user system. The number of users per cell is assumed to be $K = 4$. In this system, orthogonal pilots are adopted for the intra-cell users and the same orthogonal pilots are reused for different cells [1]. We consider two values of β: $\beta = 0.1$ and $\beta = 0.2$. We observe from Figs. 8 and 9 that the data-aided scheme performs the best among the three schemes and it is more robust against strong inter-cell interference.

We now briefly discuss the complexities of the three schemes shown in Figs. 8 and 9. The per user complexities of the pilot based scheme and the data-aided scheme have been analyzed in Section IV-D. For the blind SVD scheme, the complexity mainly comes from extracting the K left singular vectors of a (large) M-by-J matrix. For this purpose, the straightforward approach is to perform full SVD and truncate the interested singular vectors, which has complexity $O(M \cdot J \cdot \min(M, J))$. As pointed out in [6], more efficient algorithms exist, e.g., the implicitly restarted Lanczos/Arnoldi algorithm in [28].

In Fig. 10, we demonstrate the BER performances of the pilot-based, and the data-aided schemes as the number of antennas M varies. The channel SNR is fixed to be -6 dB. Other

parameters are the same as Fig. 8. We can see that for the relatively high inter-cell interference scenarios of $\beta = 0.2$ and $\beta = 0.3$, the pilot-only scheme suffer from high error floor even when $M = 256$, while the data-aided scheme performs significantly better.

V. Conclusion

We analyzed the performance of the data-aided channel estimation scheme in a multi-cell large antenna system. We showed that there are two types of contamination in this scheme: cross-contamination and self-contamination. Both analysis and simulation show that the data-aided scheme can effectively suppress the contamination effect and achieve improved performance in large antenna systems.

The analysis procedure developed in this paper also provides useful guidelines for system design. For example, from Fig. 6, the variance $v_{\Delta x}$ at the fixed point can be reduced if the transfer function of the decoder can be carefully shaped. It implies that the FEC code should be designed by taking into account the contamination effect. This is an interesting direction for future research.

Possible future work also includes extensions to multi-path channels. For this purpose, combining the data-aided channel estimation scheme and the time frequency training OFDM (TFT-OFDM) scheme proposed in [12] is an interesting topic.

Appendix A
Derivations of the LMMSE Channel Estimator

We now rigorously prove that the estimator in (9) is the optimal LMMSE estimator. For convenience, denote $\boldsymbol{y}_1 \equiv \left[\boldsymbol{y}_1(1)^{\mathrm{T}}, \ldots, \boldsymbol{y}_1(J)^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{n}_1 \equiv \left[\boldsymbol{n}_1(1)^{\mathrm{T}}, \ldots, \boldsymbol{n}_1(J)^{\mathrm{T}}\right]^{\mathrm{T}}$. We can rewrite (2) as

$$\boldsymbol{y}_1 = (\boldsymbol{x}_1 \otimes \boldsymbol{I}_M)\boldsymbol{h}_1 + \sum_{i\neq 1}^{L}(\boldsymbol{x}_i \otimes \boldsymbol{I}_M)\boldsymbol{h}_i + \boldsymbol{n}_1 \tag{32a}$$

$$= (\hat{\boldsymbol{x}}_1 \otimes \boldsymbol{I}_M)\boldsymbol{h}_1 + \tilde{\boldsymbol{n}}_1, \tag{32b}$$

where $\otimes$ denotes the Kronecker product and

$$\tilde{\boldsymbol{n}}_1 = (\Delta\boldsymbol{x}_1 \otimes \boldsymbol{I}_M)\boldsymbol{h}_1 + \sum_{i\neq 1}^{L}(\boldsymbol{x}_i \otimes \boldsymbol{I}_M)\boldsymbol{h}_i + \boldsymbol{n}_1. \tag{32c}$$

The LMMSE estimate [16] of $\boldsymbol{h}_1$ based on (32) is given by

$$\hat{\boldsymbol{h}}_1 = \mathrm{E}\left[\boldsymbol{h}_1\boldsymbol{y}_1^{\mathrm{H}}\right]\left(\mathrm{E}\left[\boldsymbol{y}_1\boldsymbol{y}_1^{\mathrm{H}}\right]\right)^{-1}\boldsymbol{y}_1 \tag{33}$$

where the expectation is conditioned on $\hat{\boldsymbol{x}}_1$. In (33), $\boldsymbol{h}_1$ and $\tilde{\boldsymbol{n}}_1$ are uncorrelated (note that $\Delta\boldsymbol{x}_1$ and $\boldsymbol{h}_1$ are mutually independent). Therefore

$$\mathrm{E}\left[\boldsymbol{h}_1\boldsymbol{y}_1^{\mathrm{H}}\right] = \mathrm{E}\left[\boldsymbol{h}_1\boldsymbol{h}_1^{\mathrm{H}}\right](\hat{\boldsymbol{x}}_1 \otimes \boldsymbol{I}_M)^{\mathrm{H}} \tag{34a}$$

$$= (\beta_1 \cdot \hat{\boldsymbol{x}}_1^{\mathrm{H}}) \otimes \boldsymbol{I}_M, \tag{34b}$$

$$\mathrm{E}\left[\boldsymbol{y}_1\boldsymbol{y}_1^{\mathrm{H}}\right] = (\hat{\boldsymbol{x}}_1 \otimes \boldsymbol{I}_M)\mathrm{E}\left[\boldsymbol{h}_1\boldsymbol{h}_1^{\mathrm{H}}\right](\hat{\boldsymbol{x}}_1 \otimes \boldsymbol{I}_M)^{\mathrm{H}} + \mathrm{E}[\tilde{\boldsymbol{n}}_1\tilde{\boldsymbol{n}}_1^{\mathrm{H}}] \tag{34c}$$

$$= \boldsymbol{R} \otimes \boldsymbol{I}_M \tag{34d}$$

where

$$\boldsymbol{R} \equiv \beta_1\hat{\boldsymbol{x}}_1\hat{\boldsymbol{x}}_1^{\mathrm{H}} + \left(\beta_1 v_{\Delta x} + \sum_{i\neq 1}^{L}\beta_i + N_0\right)\boldsymbol{I}_J. \tag{34e}$$

Steps (34b) and (34d) follow some basic properties of the Kronecker product [29, pp. 243]. Substituting (34) into (33), we get

$$\hat{\boldsymbol{h}}_1 = \left(\left(\beta_1 \cdot \hat{\boldsymbol{x}}_1^{\mathrm{H}}\boldsymbol{R}^{-1}\right) \otimes \boldsymbol{I}_M\right)\boldsymbol{y}_1 \tag{35a}$$

$$= \left(\left(\theta \cdot \hat{\boldsymbol{x}}_1^{\mathrm{H}}\right) \otimes \boldsymbol{I}_M\right)\boldsymbol{y}_1 \tag{35b}$$

$$= \theta \cdot \sum_{j=1}^{J}\boldsymbol{y}_1(j) \cdot \hat{x}_1^*(j), \tag{35c}$$

where

$$\theta = \frac{\beta_1}{\beta_1 \cdot \|\hat{\boldsymbol{x}}_1\|^2 + \beta_1 \cdot v_{\Delta x} + \sum_{i\neq 1}^{L}\beta_i + N_0}. \tag{35d}$$

In the above, (35a) is due to the following two properties of Kronecker product [29, pp. 244]: (1) $(\boldsymbol{A}\otimes\boldsymbol{B})(\boldsymbol{C}\otimes\boldsymbol{D}) = \boldsymbol{AC}\otimes\boldsymbol{BD}$, (2) $(\boldsymbol{P}\otimes\boldsymbol{Q})^{-1} = \boldsymbol{P}^{-1}\otimes\boldsymbol{Q}^{-1}$ for any invertible matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$; (35b) follows from the matrix inversion lemma, and (35c) follows directly form the definition of the Kronecker product and that of $\boldsymbol{y}_1$.

It can be verified that (35) is equivalent to the estimator (9) in Section II-C.

Appendix B
Derivations of Average Signal, Noise and Interference Power

For all derivations in this appendix, we will start from the conditional expectations over fixed $\{\boldsymbol{x}_i, i \neq 1\}$, $\Delta\boldsymbol{x}_1$ and $\hat{\boldsymbol{x}}_1$, and gradually remove these conditions.[1]

A. Linear Gaussian Model Related to $\hat{\boldsymbol{h}}_1$

Before the detailed derivations, we will first discuss a linear Gaussian model related to $\hat{\boldsymbol{h}}_1$, which is useful for following derivations. In this subsection, we assume that $\{\boldsymbol{x}_i, i \neq 1\}$, $\Delta\boldsymbol{x}_1$ and $\hat{\boldsymbol{x}}_1$ are fixed. (Note that when $\Delta\boldsymbol{x}_1$ and $\hat{\boldsymbol{x}}_1$ are fixed, $\boldsymbol{x}_1 = \Delta\boldsymbol{x}_1 + \hat{\boldsymbol{x}}_1$ is also fixed.)

For notational brevity, we define

$$\kappa_k \equiv \hat{\boldsymbol{x}}_1^{\mathrm{H}}\boldsymbol{x}_k, \quad \forall k. \tag{36}$$

Combining (8) and (9a), the channel estimate $\hat{\boldsymbol{h}}_1$ can be rewritten as

$$\hat{\boldsymbol{h}}_1 = \theta \cdot \kappa_i \cdot \boldsymbol{h}_i + \boldsymbol{\eta}_i, \quad \forall i, \tag{37a}$$

where

$$\boldsymbol{\eta}_i \equiv \sum_{k\neq i}\theta \cdot \kappa_k \cdot \boldsymbol{h}_k + \sum_{j=1}^{J}\theta \cdot \hat{x}_1^*(j)\boldsymbol{n}_1(j) \tag{37b}$$

and θ is defined in (9b). Since we assumed $\{\boldsymbol{x}_i, i \neq 1\}$, $\Delta\boldsymbol{x}_1$ and $\hat{\boldsymbol{x}}_1$ are fixed, $\theta \cdot \kappa_i$ are thus fixed and $\boldsymbol{\eta}_i$ is Gaussian. Under this

[1]To keep notational brevity, we will always use $\mathrm{E}[\cdot]$ for conditional expectation without explicitly indicating the involved conditions, which should be clear from the context.

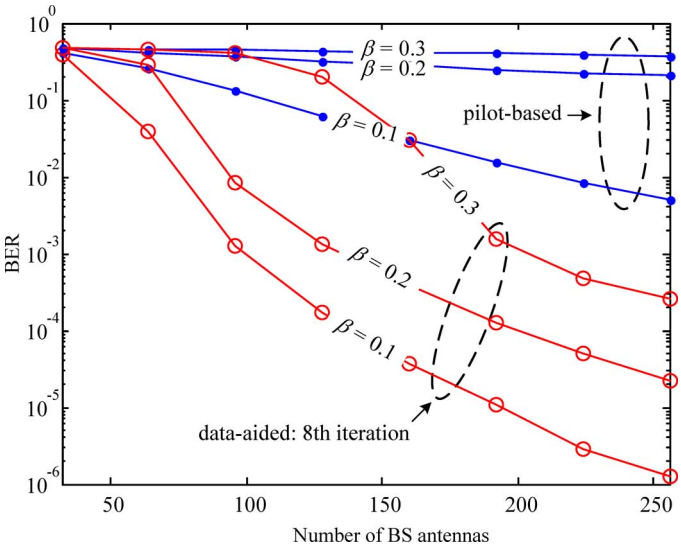


Fig. 10. BER performances of the pilot based scheme, the data-aided scheme for different M and β. $\beta_1 = 1$, $\beta_i = \beta$ for $i \neq 1$. SNR $= -6$ dB. Other parameters are the same as Fig. 8.

assumption, (37) becomes a standard linear Gaussian model. We rewrite $\boldsymbol{h}_i$ as

$$\boldsymbol{h}_i = \varphi_i \cdot \hat{\boldsymbol{h}}_1 + \boldsymbol{\zeta}_i, \quad \forall i, \tag{38}$$

where the linear combining coefficient φ_i is given by [16],

$$\varphi_i = \frac{\theta \cdot \kappa_i^* \cdot \beta_i}{v_{\hat{h}_1}}, \quad \forall i. \tag{39}$$

In the above

$$v_{\hat{h}_1} = \theta^2 \cdot \left(\sum_{k=1}^{L} |\kappa_k|^2 \cdot \beta_k + \|\hat{\boldsymbol{x}}_1\|^2 \cdot N_0 \right). \tag{40}$$

By the property of MMSE estimation of a linear Gaussian model, $\hat{\boldsymbol{h}}_i$ and $\boldsymbol{\zeta}_i \equiv \boldsymbol{h}_i - \varphi_i \cdot \hat{\boldsymbol{h}}_1$ are statistically independent [16]. It can also be shown that $\boldsymbol{\zeta}_i$ is Gaussian with zero mean and variance

$$v_{\zeta_i} = \beta_i - \frac{\theta^2 \cdot |\kappa_i|^2 \cdot \beta_i^2}{v_{\hat{h}_i}}, \quad \forall i. \tag{41}$$

B. Signal

The signal power in (13) is given by

$$\mathrm{E}\left[|S|^2\right] = \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^4\right]. \tag{42}$$

We now evaluate the expectation in (42). We will start from the conditional expectation over fixed $\{\boldsymbol{x}_i, i \neq 1\}$, $\Delta\boldsymbol{x}_1$ and $\hat{\boldsymbol{x}}_1$, and gradually remove these conditions.

When $\{\boldsymbol{x}_i, i \neq 1\}$, $\Delta\boldsymbol{x}_1$ and $\hat{\boldsymbol{x}}_1$ are fixed, $\hat{\boldsymbol{h}}_1$ is Gaussian with zero mean and variance $v_{\hat{h}_1}$ given in (40). Therefore, $\|\hat{\boldsymbol{h}}_1\|^2 \sim v_{\hat{h}_1}/2 \cdot \chi^2(2M)$, where $\chi^2(2M)$ denotes chi-square distribution with $2M$ degrees of freedom, and so

$$\mathrm{E}\left[|S|^2\right] = \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^4\right] = v_{\hat{h}_1}^2 \cdot (M^2 + M). \tag{43}$$

Next, we assume that only $\hat{\boldsymbol{x}}_1$ is fixed. We will derive the expectation of (43) over the distribution of $\{\boldsymbol{x}_i, i \neq 1\}$ and $\Delta\boldsymbol{x}_1$. To this end, it suffices to derive $\mathrm{E}[v_{\hat{h}_1}^2]$ condition on a fixed $\hat{\boldsymbol{x}}_1$. Recall the Gaussian assumptions on $\{\boldsymbol{x}_i, i \neq 1\}$ and $\Delta\boldsymbol{x}_1$ in Sections II-A and II-B. Under these Gaussian assumptions, $\{\kappa_k, \forall k\}$ defined in (36) are also Gaussian, i.e.,

$$\kappa_1 \equiv \hat{\boldsymbol{x}}_1^{\mathrm{H}} \boldsymbol{x}_1 \sim \mathcal{CN}\left(\|\hat{\boldsymbol{x}}_1\|^2, \|\hat{\boldsymbol{x}}_1\|^2 \cdot v_{\Delta x}\right), \tag{44a}$$

$$\kappa_i \equiv \hat{\boldsymbol{x}}_1^{\mathrm{H}} \boldsymbol{x}_i \sim \mathcal{CN}\left(0, \|\hat{\boldsymbol{x}}_1\|^2\right), \quad i \neq 1. \tag{44b}$$

Taking average over $\{\boldsymbol{x}_i, i \neq 1\}$ and $\Delta\boldsymbol{x}_1$ and using the properties of Gaussian distribution, we have [22]

$$\begin{aligned}
\mathrm{E}\left[|\kappa_1|^2\right] &= \|\hat{\boldsymbol{x}}_1\|^4 + \|\hat{\boldsymbol{x}}_1\|^2 \cdot v_{\Delta x}, \\
\mathrm{E}\left[|\kappa_i|^2\right] &= \|\hat{\boldsymbol{x}}_1\|^2, \quad i \neq 1, \\
\mathrm{E}\left[\kappa_1^* \cdot |\kappa_1|^2\right] &= \|\hat{\boldsymbol{x}}_1\|^6 + 2 \cdot \|\hat{\boldsymbol{x}}_1\|^4 \cdot v_{\Delta x}, \\
\mathrm{E}\left[\kappa_1^* \cdot |\kappa_i|^2\right] &= \|\hat{\boldsymbol{x}}_1\|^4, \quad i \neq 1, \\
\mathrm{E}\left[|\kappa_1|^4\right] &= \|\hat{\boldsymbol{x}}_1\|^4 \cdot (\|\hat{\boldsymbol{x}}_1\|^4 + 4 \cdot \|\hat{\boldsymbol{x}}_1\|^2 v_{\Delta x} + 2 v_{\Delta x}^2), \\
\mathrm{E}\left[|\kappa_i|^4\right] &= 2 \cdot \|\hat{\boldsymbol{x}}_1\|^4, \\
\mathrm{E}\left[|\kappa_1|^2 \cdot |\kappa_i|^2\right] &= \|\hat{\boldsymbol{x}}_1\|^6 + \cdot \|\hat{\boldsymbol{x}}_1\|^4 \cdot v_{\Delta x}, \quad i \neq 1, \\
\mathrm{E}\left[|\kappa_i|^2 \cdot |\kappa_k|^2\right] &= \|\hat{\boldsymbol{x}}_1\|^4, \quad i \neq 1,\ k \neq 1,\ i \neq k.
\end{aligned} \tag{45}$$

Applying (45) to (41) and after some straightforward manipulations, the expectation of $v_{\hat{h}_1}^2$ conditioned on $\hat{\boldsymbol{x}}_1$ is shown to be

$$\begin{aligned}
\mathrm{E}\left[v_{\hat{h}_1}^2\right] &= \theta^4 \cdot \|\hat{\boldsymbol{x}}_1\|^4 \cdot \left(\beta_1 \|\hat{\boldsymbol{x}}_1\|^2 + \beta_1 v_{\Delta x} + \sum_{i \neq 1}^{L} \beta_i + N_0 \right)^2 \\
&\quad + \theta^4 \cdot \|\hat{\boldsymbol{x}}_1\|^4 \cdot \left(\left(2 \cdot \|\hat{\boldsymbol{x}}_1\|^2 \cdot v_{\Delta x} + v_{\Delta x}^2\right) \cdot \beta_1^2 + \sum_{i \neq 1}^{L} \beta_i^2 \right),
\end{aligned} \tag{46}$$

where θ is given by (see (9b))

$$\theta = \frac{\beta_1}{\|\hat{\boldsymbol{x}}_1\|^2 \cdot \beta_1 + v_{\Delta x} \cdot \beta_1 + \sum_{i \neq 1}^{L} \beta_i + N_0}. \tag{47}$$

Finally, we will derive the expectation of (46) over $\hat{\boldsymbol{x}}_1$. The exact result is very complicated for finite J. Instead, we resort to approximations for large J.

Our approximation is based on the following theorem [30, Theorem 3.9].

Theorem 1: Suppose $a(1), a(2), \ldots, a(J)$ are i.i.d. observations with a finite fourth moment. The mean and variance of $a(j)$ are given by $\mathrm{E}[a]$ and $\mathrm{Var}[a]$. Let $g(\cdot)$ be a scalar function with four uniformly bounded derivatives. Denote by $\bar{a}$ the sample average, i.e., $\bar{a} = \sum_{j=1}^{J} a(j)/J$. Then

$$\mathrm{E}\left[g(\bar{a})\right] = g(\mathrm{E}[a]) + \frac{g^{(2)}(\mathrm{E}[a]) \cdot \mathrm{Var}[a]}{2J} + O\left(\frac{1}{J^2}\right), \tag{48}$$

where $g^{(2)}(\cdot)$ represents the second order derivative of $g(\cdot)$.

We now return to our problem. From assumptions 1) and 2) in Section II-B, $\{\hat{x}_1(1), \ldots, \hat{x}_1(J)\}$ are i.i.d. Gaussian random variables with zero mean and variance $(1 - v_{\Delta x})$. Therefore, $\{|\hat{x}_1(1)|^2, \ldots, |\hat{x}_1(J)|^2\}$ are i.i.d. samples and $\|\hat{\boldsymbol{x}}_1\|^2/J$ is the sample average. Also, the mean and variance of $\hat{x}_1(j)$ are respectively given by

$$\mu \equiv \mathrm{E}\left[|\hat{x}_1(j)|^2\right] = 1 - v_{\Delta x}, \tag{49a}$$

$$\sigma^2 \equiv \mathrm{Var}\left[|\hat{x}_1(j)|^2\right] = (1 - v_{\Delta x})^2. \tag{49b}$$

Note that (46) is a function of $\|\hat{\boldsymbol{x}}_1\|^2/J$, denoted by $f(\|\hat{\boldsymbol{x}}_1\|^2/J)$. Our aim is to approximate $\mathrm{E}[f\left(\|\hat{\boldsymbol{x}}_1\|^2/J\right)]$ when J is large, where the expectation is with respect to $\|\hat{\boldsymbol{x}}_1\|^2$. We verified that the four uniformly bounded derivatives condition of Theorem 1 is satisfied here. Applying Theorem 1, we have

$$\mathrm{E}\left[f(\|\hat{\boldsymbol{x}}_1\|^2/J)\right] = f(\mu) + \frac{f^{(2)}(\mu)\sigma^2}{2J} + O\left(\frac{1}{J^2}\right). \tag{50}$$

where μ and σ^2 are given in (49). $f(\mu)$ is obtained by substituting $\|\hat{\boldsymbol{x}}_1\|^2 = J \cdot \mu = J \cdot (1 - v_{\Delta x})$ into (46). $f^{(2)}(\mu)$ can be obtained by first deriving the second order derivative $f^{(2)}(\cdot)$ and then substituting in μ. Applying Taylor series expansion to $f(\mu)$ and $f^{(2)}(\mu)$, we finally have

$$f(\mu) = \beta_1^2 + O\left(\frac{1}{J}\right). \tag{51a}$$

$$f^{(2)}(\mu) = O\left(\frac{1}{J^2}\right). \tag{51b}$$

The detail of (51) are omitted here. Combining (50) and (51), we can write

$$\mathrm{E}\left[f(\|\hat{\boldsymbol{x}}_1\|^2/J)\right] = \beta_1^2 + O\left(\frac{1}{J}\right). \tag{52}$$

We now return to the signal power in (43). From (52) and (43), we have

$$\mathrm{E}\left[|S|^2\right] = \mathrm{E}\left[f(\|\hat{\boldsymbol{x}}_1\|^2/J)\right] \cdot (M^2 + M) \tag{53a}$$

$$= \left(\beta_1^2 + O\left(\frac{1}{J}\right)\right) \cdot (M^2 + M). \tag{53b}$$

$$= \beta_1^2 \cdot M^2 + O\left(\frac{M^2}{J}\right) \tag{53c}$$

$$\approx \beta_1^2 \cdot M^2. \tag{53d}$$

where we assumed that M is also large in (53c).

C. Noise

We next consider the noise power, given by (see (13))

$$\mathrm{E}\left[|N|^2\right] = \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^2\right]. \tag{54}$$

Similar to (43), when $\{\boldsymbol{x}_i, i \neq 1\}$, Δx_1 and $\hat{\boldsymbol{x}}_1$ are fixed, we have $\|\hat{\boldsymbol{h}}_1\|^2 \sim v_{\hat{h}_1}/2 \cdot \chi^2(2M)$ and

$$\mathrm{E}\left[|N|^2\right] = \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^2\right] = v_{\hat{h}_1} \cdot M, \tag{55}$$

where $v_{\hat{h}_1}$ was given in (40):

$$v_{\hat{h}_1} = \theta^2 \cdot \left(\sum_{k=1}^{L} |\kappa_k|^2 \cdot \beta_k + \|\hat{\boldsymbol{x}}_1\|^2 \cdot N_0\right). \tag{56}$$

We now assume that only $\hat{\boldsymbol{x}}_1$ is fixed and take average of $v_{\hat{h}_1}$ with respect to $\{\boldsymbol{x}_i, i \neq 1\}$ and Δx_1. Note that the averages of $\{|\kappa_k|^2, \forall k\}$ are given in (45). Combining this with (56), we can verify that

$$\mathrm{E}\left[v_{\hat{h}_1}\right] = \frac{\|\hat{\boldsymbol{x}}_1\|^2 \cdot \beta_1^2}{\|\hat{\boldsymbol{x}}_1\|^2 \cdot \beta_1 + v_{\Delta x} \cdot \beta_1 + \sum_{i \neq 1}^{L} \beta_i + N_0}. \tag{57}$$

Applying Theorem 1 and using similar steps as (49)–(52) in Subsection B, we further take the average of (57) with respect to $\hat{\boldsymbol{x}}_1$ and obtain

$$\mathrm{E}\left[v_{\hat{h}_1}\right] = \beta_1 + O\left(\frac{1}{J}\right). \tag{58}$$

Combining (55) and (58) we have

$$\mathrm{E}\left[|N|^2\right] = N_0 \cdot M \cdot \left(\beta_1 + O\left(\frac{1}{J}\right)\right). \tag{59a}$$

$$= N_0 \cdot M \cdot \beta_1 + O\left(\frac{M}{J}\right) \tag{59b}$$

$$\approx N_0 \cdot M \cdot \beta_1 \tag{59c}$$

when both M and J are large.

D. Cross-Interference

We now consider the average power of cross-interference in (13)

$$\mathrm{E}\left[|I_i|^2\right] = \mathrm{E}\left[\left|\hat{\boldsymbol{h}}_1^{\mathrm{H}} \boldsymbol{h}_i\right|^2\right], \quad i \neq 1. \tag{60}$$

Again, we first assume that $\{\boldsymbol{x}_i, i \neq 1\}$, $\Delta \boldsymbol{x}_1$ and $\hat{\boldsymbol{x}}_1$ are fixed. Under this assumption, we have shown in (38) that $\boldsymbol{h}_i$ can be decomposed as

$$\boldsymbol{h}_i = \varphi_i \cdot \hat{\boldsymbol{h}}_1 + \boldsymbol{\zeta}_i. \tag{61}$$

Substituting (61) into (60), we have

$$\mathrm{E}\left[|I_i|^2\right] = \mathrm{E}\left[\left|\hat{\boldsymbol{h}}_1^{\mathrm{H}} \left(\varphi_i \cdot \hat{\boldsymbol{h}}_1 + \boldsymbol{\zeta}_i\right)\right|^2\right] \tag{62a}$$

$$= \mathrm{E}\left[|\varphi_i \cdot \|\hat{\boldsymbol{h}}_1\|^2 + \hat{\boldsymbol{h}}_1^{\mathrm{H}} \boldsymbol{\zeta}_i|^2\right] \tag{62b}$$

$$= |\varphi_i|^2 \cdot \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^4\right] + v_{\zeta_i} \cdot \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^2\right] \tag{62c}$$

$$= |\varphi_i|^2 \cdot v_{\hat{h}_1}^2 \cdot \left(M^2 + M\right) + v_{\zeta_i} \cdot v_{\hat{h}_1} \cdot M, \tag{62d}$$

where φ_i, v_{ζ_i} and $v_{\hat{h}_1}$ are given in (39)–(40). The expectations in (62) are with respect to $\hat{\boldsymbol{h}}_1$ and $\boldsymbol{\zeta}_i$, conditioned on $\{\boldsymbol{x}_i, i \neq 1\}$, $\Delta \boldsymbol{x}_1$ and $\hat{\boldsymbol{x}}_1$. The equality in (62c) is due to the independency

between $\hat{\boldsymbol{h}}_1$ and $\boldsymbol{\zeta}_i$. In (62d) we have used (see also (43) and (55))

$$\mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^2\right] = v_{\hat{h}_1} \cdot M; \tag{63a}$$

$$\mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^4\right] = v_{\hat{h}}^2 \cdot \left(M^2 + M\right). \tag{63b}$$

We next average $|\varphi_i|^2 \cdot v_{\hat{h}_1}^2$ and $v_{\zeta_i} \cdot v_{\hat{h}_1}$ in (62d) over the distribution $\{\boldsymbol{x}_i, i \neq 1\}$ and $\Delta\boldsymbol{x}_1$. Using (45), $\mathrm{E}\left[|\varphi_i|^2 \cdot v_{\hat{h}_1}^2\right]$ and $\mathrm{E}\left[v_{\zeta_i} \cdot v_{\hat{h}_1}\right]$ are given in (65) at the bottom of the page. Next, we take the average of (65) with respect to $\hat{\boldsymbol{x}}_1$. Similar to (49)–(52) in Subsection B, for large J, we have

$$\mathrm{E}\left[|\varphi_i|^2 \cdot v_{\hat{h}_1}^2\right] = \frac{\beta_i^2}{1 - v_{\Delta x}} \cdot \frac{1}{J} + O\left(\frac{1}{J^2}\right), \tag{64a}$$

$$\mathrm{E}\left[v_{\zeta_i} \cdot v_{\hat{h}_1}\right] = \beta_i \cdot \beta_1 + O\left(\frac{1}{J}\right). \tag{64b}$$

Substituting (64) into (62d), the cross interference power $\mathrm{E}[|I_i|^2]$ can be approximated as follows

$$\mathrm{E}\left[|I_i|^2\right] = \mathrm{E}\left[|\varphi_i|^2 v_{\hat{h}_1}^2\right]\left(M^2+M\right) + \mathrm{E}\left[v_{\zeta_i} v_{\hat{h}_1}\right] M \tag{68a}$$

$$= \left(\frac{\beta_i^2}{1 - v_{\Delta x}} \cdot \frac{1}{J} + O\left(\frac{1}{J^2}\right)\right) \cdot \left(M^2 + M\right) + \left(\beta_i \cdot \beta_1 + O\left(\frac{1}{J}\right)\right) \cdot M \tag{68b}$$

$$= \frac{\beta_i^2}{1 - v_{\Delta x}} \cdot \frac{M^2}{J} + \beta_i \cdot \beta_1 \cdot M + O\left(\frac{M^2}{J^2}\right) + O\left(\frac{M}{J}\right), \tag{68c}$$

where we assumed that M is large in (68c). Since when J is large, $O(M^2/J^2)$ is in a lower order (in terms of J) than $(\beta_i^2/(1 - v_{\Delta x})) \cdot (M^2/J)$ and $O(M/J)$ in a lower order than $\beta_i \cdot \beta_1 \cdot M$, we thus omit the two error terms in (68c) and obtain the following approximation

$$\mathrm{E}\left[|I_i|^2\right] \approx \frac{\beta_i^2}{1 - v_{\Delta x}} \cdot \frac{M^2}{J} + \beta_i \cdot \beta_1 \cdot M, \quad i \neq 1. \tag{69}$$

E. Self-Interference

Form (13), the average self-interference power is given by

$$\mathrm{E}\left[|I_1|^2\right] = \mathrm{E}\left[\left|\hat{\boldsymbol{h}}_1^{\mathrm{H}}\left(\boldsymbol{h}_1 - \hat{\boldsymbol{h}}_1\right)\right|^2\right]. \tag{70}$$

The derivations of (70) will be similar to those of $\mathrm{E}[|I_i|^2], i \neq 1$, in Subsection D.

Firstly, we assume that $\{\boldsymbol{x}_i, i \neq 1\}$, Δx_1 and $\hat{\boldsymbol{x}}_1$ are fixed. We decompose $\boldsymbol{h}_1$ into two mutually independent parts as

$$\boldsymbol{h}_1 = \varphi_1 \cdot \hat{\boldsymbol{h}}_1 + \boldsymbol{\zeta}_1. \tag{71}$$

Substituting (71) into (70) and similar to (62), we can write

$$\mathrm{E}\left[|I_1|^2\right] = \mathrm{E}\left[\left|\hat{\boldsymbol{h}}_1^{\mathrm{H}}\left(\varphi_1 \cdot \hat{\boldsymbol{h}}_1 + \boldsymbol{\zeta}_1 - \hat{\boldsymbol{h}}_1\right)\right|^2\right] \tag{72a}$$

$$= |\varphi_1 - 1|^2 \cdot \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^4\right] + v_{\zeta_1} \cdot \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^2\right] \tag{72b}$$

$$= |\varphi_1 - 1|^2 v_{\hat{h}_1}^2 \cdot \left(M^2 + M\right) + v_{\zeta_1} v_{\hat{h}_1} \cdot M, \tag{72c}$$

Next, we average $|\varphi_1 - 1|^2 \cdot v_{\hat{h}_1}^2$ and $v_{\zeta_1} \cdot v_{\hat{h}_1}$ in (72c) over $\{\boldsymbol{x}_i, i \neq 1\}$ and $\Delta\boldsymbol{x}_1$, conditioned on $\hat{\boldsymbol{x}}_1$. Using (45), we obtain $\mathrm{E}\left[|\varphi_1 - 1|^2 \cdot v_{\hat{h}_1}^2\right]$ and $\mathrm{E}\left[v_{\zeta_1} \cdot v_{\hat{h}_1}\right]$ in (66), as shown at the top of the next page. When J is large, following (49)–(52) in Subsection B, we finally have

$$\mathrm{E}\left[|\varphi_1 - 1|^2 \cdot v_{\hat{h}_1}^2\right] = \frac{\beta_1^2 \cdot v_{\Delta x}}{(1 - v_{\Delta x})} \cdot \frac{1}{J} + O\left(\frac{1}{J^2}\right); \tag{73a}$$

$$\mathrm{E}\left[v_{\zeta_1} \cdot v_{\hat{h}_1}\right] = O\left(\frac{1}{J}\right). \tag{73b}$$

We now combine (73b) and (72c) and approximate self-interference in the large M and J regime as follows

$$\mathrm{E}\left[|I_1|^2\right] = \mathrm{E}\left[|\varphi_1 - 1|^2 \cdot v_{\hat{h}_1}^2\right] \cdot \left(M^2 + M\right) + \mathrm{E}\left[v_{\zeta_1} \cdot v_{\hat{h}_1}\right] \cdot M \tag{74a}$$

$$= \left(\frac{\beta_1^2 \cdot v_{\Delta x}}{1 - v_{\Delta x}} \cdot \frac{1}{J} + O\left(\frac{1}{J^2}\right)\right) \cdot \left(M^2 + M\right) + O\left(\frac{1}{J}\right) \cdot M \tag{74b}$$

$$= \frac{\beta_1^2 \cdot v_{\Delta x}}{1 - v_{\Delta x}} \cdot \frac{M^2}{J} + O\left(\frac{M^2}{J^2}\right) + O\left(\frac{M}{J}\right) \tag{74c}$$

$$\approx \frac{\beta_1^2 \cdot v_{\Delta x}}{1 - v_{\Delta x}} \cdot \frac{M^2}{J}. \tag{74d}$$

$$\mathrm{E}\left[|\varphi_i|^2 \cdot v_{\hat{h}_1}^2\right] = \frac{\|\hat{\boldsymbol{x}}_1\|^2 \cdot \beta_i^2 \cdot \beta_1^2}{\left(\|\hat{\boldsymbol{x}}_1\|^2 \cdot \beta_1 + v_{\Delta x} \cdot \beta_1 + \sum_{k \neq 1}^{L} \beta_k + N_0\right)^2}, \tag{65a}$$

$$\mathrm{E}\left[v_{\zeta_i} \cdot v_{\hat{h}_1}\right] = \frac{\left(\|\hat{\boldsymbol{x}}_1\|^2 \cdot \beta_1 + v_{\Delta x} \cdot \beta_1 + \sum_{k \neq 1, k \neq i}^{L} \beta_k + N_0\right) \cdot \|\hat{\boldsymbol{x}}_1\|^2 \cdot \beta_i \cdot \beta_1^2}{\left(\|\hat{\boldsymbol{x}}_1\|^2 \cdot \beta_1 + v_{\Delta x} \cdot \beta_1 + \sum_{k \neq 1}^{L} \beta_k + N_0\right)^2}. \tag{65b}$$

$$\mathrm{E}\left[|\varphi_1-1|^2 v_{\hat{h}_1}^2\right] = \theta^4 \|\hat{\boldsymbol{x}}_1\|^2 \left(v_{\Delta x}\left(\beta_1 v_{\Delta x} + \sum_{i\neq 1}^{L}\beta_i + N_0\right)^2 + \beta_1^2 \|\hat{\boldsymbol{x}}_1\|^4 v_{\Delta x} + \beta_1^2 \|\hat{\boldsymbol{x}}_1\|^2 v_{\Delta_x}^2 + \sum_{i\neq 1}^{L}\|\hat{\boldsymbol{x}}_1\|^2 \beta_i^2 \right); \quad (66a)$$

$$\mathrm{E}[v_{\zeta_1}\cdot v_{\hat{h}_1}] = \beta_1\cdot\theta^2\cdot\|\hat{\boldsymbol{x}}_1\|^2\cdot\left(\sum_{i\neq 1}^{L}\beta_i + N_0\right). \quad (66b)$$

$$\mathrm{E}\left[(\varphi_1-1)\cdot v_{\hat{h}}^2\right] = \beta_1^2\cdot\theta^3\cdot\|\hat{\boldsymbol{x}}_1\|^4\cdot v_{\Delta x} - \theta^4\cdot\|\hat{\boldsymbol{x}}_1\|^4\cdot\left(\left(2\cdot\|\hat{\boldsymbol{x}}_1\|^2\cdot v_{\Delta x} + v_{\Delta x_1}^2\right)\cdot\beta_1^2 + \sum_{i\neq 1}^{L}\beta_i^2\right) \quad (67)$$

F. Correlation of Signal and Self-Interference

The correlation coefficient between signal and self-interference is defined in (22) as

$$\alpha \equiv \frac{\left|\mathrm{E}[S^*\cdot I_1]\right|^2}{\mathrm{E}[|S|^2]\cdot\mathrm{E}[|I_1|^2]}. \quad (75)$$

The average signal power $\mathrm{E}[|S|^2]$ and self-interference power $\mathrm{E}[|I_1|^2]$ in the denominator of (75) has been derived in (53) and (74), respectively. We now consider $E[S^*\cdot I_1]$ in the numerator of (75). Following the same procedure adopted in deriving self-interference power, we have

$$\mathrm{E}\left[S^*\cdot I_1\right] = \mathrm{E}\left[\|\hat{\boldsymbol{h}}_1\|^2\cdot\hat{\boldsymbol{h}}_1^{\mathrm{H}}\left(\boldsymbol{h}_1-\hat{\boldsymbol{h}}_1\right)\right], \quad (76a)$$

$$= (\varphi_1-1)\cdot v_{\hat{h}_1}^2\cdot\left(M^2+M\right). \quad (76b)$$

Again, using the formulas in (45), we obtain $\mathrm{E}[(\varphi_1-1)\cdot v_{\hat{h}_1}^2]$ in (67) at the top of the page. Applying similar steps as (49)–(52) to (67), we obtain the following result when J is large,

$$\mathrm{E}\left[(\varphi_1-1)\cdot v_{\hat{h}}^2\right] = \frac{-\beta_1^2\cdot v_{\Delta x}}{1-v_{\Delta x}}\cdot\frac{1}{J} + O\left(\frac{1}{J^2}\right). \quad (77)$$

From (76) and (77), when both M and J are large,

$$\left|\mathrm{E}\left[S^*\cdot I_1\right]\right|^2 = \left|\mathrm{E}\left[(\varphi_1-1)\cdot v_{\hat{h}_1}^2\right]\cdot\left(M^2+M\right)\right|^2 = \frac{\beta_1^4\cdot v_{\Delta x}^2}{\left(1-v_{\Delta x}\right)^2}\cdot\frac{M^4}{J^2} + O\left(\frac{M^4}{J^4}\right) + O\left(\frac{M^2}{J^2}\right). \quad (78)$$

Furthermore, from (53c) and (74b) we have

$$\begin{aligned}&\mathrm{E}\left[|S|^2\right]\cdot\mathrm{E}\left[|I_1|^2\right]\\ &= \left(\beta_1^2\cdot M^2 + O\left(\frac{M^2}{J}\right)\right)\\ &\quad\cdot\left(\frac{v_{\Delta x}\cdot\beta_1^2}{1-v_{\Delta x}}\cdot\frac{M^2}{J} + O\left(\frac{M^2}{J^2}\right) + O\left(\frac{M}{J}\right)\right)\\ &= \frac{\beta_1^4\cdot v_{\Delta x}}{1-v_{\Delta x}}\cdot\frac{M^4}{J} + O\left(\frac{M^4}{J^2}\right) + O\left(\frac{M^3}{J}\right). \quad (79)\end{aligned}$$

Now substituting (78)–(79) into the definition of α in (75), we have the following approximation when M and J are both large,

$$\begin{aligned}\alpha &= \frac{\left|\mathrm{E}\left[S^*\cdot I_1\right]\right|^2}{\mathrm{E}\left[|S|^2\right]\cdot\mathrm{E}\left[|I_1|^2\right]}\\ &= \frac{\frac{\beta_1^4\cdot v_{\Delta x}^2}{(1-v_{\Delta x})^2}\cdot\frac{M^4}{J^2} + O\left(\frac{M^4}{J^4}\right) + O\left(\frac{M^2}{J^2}\right)}{\frac{\beta_1^4\cdot v_{\Delta x}}{1-v_{\Delta x}}\cdot\frac{M^4}{J} + O\left(\frac{M^4}{J^2}\right) + O\left(\frac{M^3}{J}\right)}\\ &\approx \frac{v_{\Delta x}}{1-v_{\Delta x}}\cdot\frac{1}{J} \ll 1. \quad (80)\end{aligned}$$

From (80), the correlation between the signal and self-interference is negligible when both M and J are large.

ACKNOWLEDGMENT

The authors would like to thank Dr. J. Chen from Alcatel-Lucent Shanghai Bell Company Ltd. for enlightening discussions.

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, 2010.

[2] F. Fernandes, A. Ashikhmin, and T. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, 2013.

[3] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.

[4] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Int. Symp. Inf. Theory Proc. (ISIT)*, 2012, pp. 1137–1141.

[5] H. Q. Ngo and E. Larsson, "Evd-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2012, pp. 3249–3252.

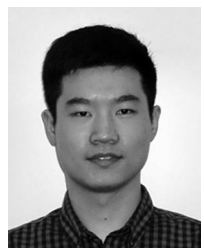
[6] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Top. Signal Process.*, 2013, preprint, doi: 10.1109/JSTSP.2014.2310053 , to be published.

[7] J. Jose, A. Ashikhmin, T. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, 2011.

[8] J. Hoydis, S. ten Brink, and M. Debbah, "Massive mimo in the ul/dl of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[9] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, 2013.

[10] H. Ngo, E. Larsson, and T. Marzetta, "The multicell multiuser mimo uplink with very large antenna arrays and a finite-dimensional channel," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2350–2361, 2013.
[11] E. Larsson, F. Tufvesson, O. Edfors, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, 2014.
[12] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, 2013.
[13] H. Holma, S. Heikkinen, O.-A. Lehtinen, and A. Toskala, "Interference considerations for the time division duplex mode of the UMTS terrestrial radio access," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 8, pp. 1386–1393, 2000.
[14] H. Schoeneich and P. A. Hoeher, "Iterative pilot-layer aided channel estimation with emphasis on interleave-division multiple access systems," *EURASIP J. Appl. Signal Process.*, vol. 2006, 2006 [Online]. Available: http://asp.eurasipjournals.com/content/2006/1/081729
[15] M. Zhao, Z. Shi, and M. Reed, "Iterative turbo channel estimation for OFDM system over rapid dispersive fading channel," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3174–3184, 2008.
[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1993.
[17] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *IEEE Trans. Commun.*, vol. 49, no. 10, pp. 1727–1737, 2001.
[18] S.-Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 657–670, 2001.
[19] L. Ping, J. Tong, X. Yuan, and Q. Guo, "Superposition coded modulation and iterative linear MMSE detection," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 6, pp. 995–1004, 2009.
[20] P. Hoeher and T. Wo, "Superposition modulation: Myths and facts," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 110–116, 2011.
[21] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, 2003.
[22] S. Ross, *A First Course in Probability*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2005.
[23] K. Gilhousen, I. Jacobs, R. Padovani, A. Viterbi, L. Weaver Jr., and C. Wheatley, III, "On the capacity of a cellular CDMA system," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 303–312, 1991.
[24] M. Tuchler, R. Koetter, and A. Singer, "Turbo equalization: Principles and new results," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 754–767, 2002.
[25] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, 1996.
[26] X. Li and J. Ritcey, "Bit-interleaved coded modulation with iterative decoding," *IEEE Commun. Lett.*, vol. 1, no. 6, pp. 169–171, 1997.
[27] X. Yuan, Q. Guo, X. Wang, and L. Ping, "Evolution analysis of low-cost iterative equalization in coded linear systems with cyclic prefixes," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 301–310, 2008.
[28] R. R. B. Lehoucq, D. D. C. Sorensen, and C.-C. Yang, *Arpack User's Guide: Solution of Large-Scale Eigenvalue Problems With Implicityly Restorted Arnoldi Methods*. Philadelphia, PA, USA: SIAM, 1998, vol. 6.
[29] R. A. Horn, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[30] A. DasGupta, *Asymptotic Theory of Statistics and Probability*. New York, NY, USA: Springer, 2008.
[31] T. Zemen, M. Loncar, J. Wehinger, C. Mecklenbrauker, and R. Muller, "Improved channel estimation for iterative receivers," in *Proc. IEEE Global Telecommun. Conf. (IEEE GLOBECOM)*, 2003, vol. 1, pp. 257–261.



Junjie Ma received the B.E. degree from Xidian University, Xi'an China, in 2010. He is currently working towards his Ph.D. degree at City University of 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 his Ph.D. degree at Glasgow University in 1990. He lectured at Department of Electronic Engineering, Melbourne University, from 1990 to 1992, and 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. Prof. 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 and is a fellow of IEEE.