Downlink Multi-user Precoder Designs in Massive MIMO

Chia-Yi Yeh

***Abstract*—In this report, MU-MIMO downlink linear precoders such as matched beamforming and zero-forcing are studied. In addition, we discuss the incentive to increase the number of transmit antennas and examine the performance of such linear precoders when the number of transmit antennas approaches to infinity. Furthermore, we study the fundamental performance loss due to pilot contamination in multi-cell scenario.**

I. Introduction

In the class, we analyze the capacity of a single-user MIMO (SU-MIMO) system by doing singular value decomposition (SVD) to the channel matrix $\boldsymbol{H}$. We observe that the MIMO channel can be decomposed to several parallel channels, and the capacity is the sum of all the parallel channels. From this analysis, we also learn the capacity-achieving approach: the transmitter precodes the transmit signals with $\boldsymbol{V}$, and the receiver decode the receive signals with $\boldsymbol{U}^*$, where $*$ denotes conjugate transpose and $\boldsymbol{H} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{V}^*$ using SVD. $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices, and $\boldsymbol{\Lambda}$ is a diagonal matrix with only non-negative real numbers on the diagonal.

When considering multi-user MIMO (MU-MIMO), each receiver only knows its own channel information, and therefore the receivers can no longer apply $\boldsymbol{U}^*$ to their receive signals and achieve the sum rate achieved in the single-user scenario. In fact, the capacity region in multi-user MIMO is achieved with Dirty-Paper Coding (DPC) [1]. However, in practice, this capacity-achieving approach is rarely used due to its high complexity. Instead, a group of linear precoders and decoders are used.

MIMO brings multiplexing gain to the system, however, the gain depends on the channel realizations. One way to combat the variation in channel variation is to utilize a large antenna array. When the number of transmit antennas approaches to infinity, the channel vectors of different users become orthogonal due to the law of large numbers and is favorable for transmissions.

The favorable propagation conditions results from the large antenna array benefits both SU-MIOO and MU-MIMO. For single user MIMO, using a large antenna array for transmission ensures the singular values of the channel to be equal, which is the best possible distribution of singular values, and therefore achieves the higher bound of the capacity. For MU-MIMO, the orthogonality between channels for different users mitigates the inter-user interference. As a result, even linear precoders can achieve near optimal capacity.

Although using large antenna arrays helps achieve optimal capacity, the overhead of such large antenna system also needs to be taken into account considering a more realistic system. Since the array is large, the channel estimation process can be time consuming if we let each of the transmit antenna take turns estimating the downlink channel. Instead, such massive MIMO system operates in TDD mode and obtain the downlink channels from the uplink channels assuming channel reciprocity. That is, the uplink channels are estimated by letting users send uplink orthogonal pilots. And the downlink channel is the transpose of the uplink channel.

Due to the uplink estimation process, pilot contamination can happen when considering multi-cell scenario. Since there are only limited orthogonal pilots, the cells have to reuse those uplink pilots. As a result, when two users in different cells sending the same uplink pilot at the same time for the BS to estimate the channel, the channel estimation of one user will contains partial channel information of the other user. This will further lead to inter-cell interference for the downlink transmission. In addition, the effect of pilot contamination cannot be alleviated as the array size increases. Therefore, designs for the uplink pilot estimation process and precoding methods considering multi-cell is important.

In the following, I will first discuss the MU-MIMO system to introduce the capacity achieving dirty-paper coding, as well as linear precoders such as matched beamforming, zero-forcing beamforming, and regularized zero-forcing beamforming in Section II. In Section III, I will discuss the incentive to use a large antenna array for both single-user and multi-user scenario. Next, we study the performance of the linear precoders when assuming the BS knows the perfect or imperfect CSI in Section IV. In Section V, we extend the massive MIMO system to multi-cell, and discuss pilot contamination and the resulting inter-cell interference. Finally, I conclude in

Section VI.

II. Multi-user Downlink

A. System Model

The following system model is adapted from [2].

Assuming a general downlink MU-MIMO transmission with a total number of K users ($\mathbb{K} = \{1, \cdots, K\}$), n_t transmit antennas at the BS, $n_{r,q}$ receive antenna at mobile terminal q. When considering a simplified case in which all users have the same number of receive antennas, we drop the index q and denote the number of receive antennas as n_r.

The MIMO channel between the BS and user q is written as $\Lambda^{-1/2}\boldsymbol{H}_q$. Λ^{-1} refers to the large-scale fading. $\boldsymbol{H}_q \in \mathbb{C}^{n_{r,q} \times n_t}$ models the small scale time-varying fading process, and is normalized such that $\mathbb{E}[\|\boldsymbol{H}_q\|^2] = n_{r,q}n_t$. Here we also assume that the MIMO channel matrices $\boldsymbol{H}_q$ are full rank for every user q.

The received signal of a given user q can be written as

$$\boldsymbol{y}_q = \Lambda^{1/2}\boldsymbol{H}_q\boldsymbol{c}' + \boldsymbol{n}_q \tag{1}$$

where $\boldsymbol{y}_q \in \mathbb{C}^{n_{r,q}}$ and $\boldsymbol{n}_q$ is a complex Gaussian noise $CN(0, \sigma_{n,q}^2\boldsymbol{I}_{n_{r,q}})$. $\boldsymbol{c}'$ is the transmitted signal vector. Generally speaking, $\boldsymbol{c}'$ is written as the superposition of statistically independent signals $\boldsymbol{c}'_q$ encoded and destined to the co-scheduled users

$$\boldsymbol{c}' = \sum_{q=1}^{K} \boldsymbol{c}'_q$$

The transmit signals are subject to a transmit power constraint, $\mathrm{Tr}\{\boldsymbol{Q}\} \leq E_s$, where $\boldsymbol{Q}$ is the input covariance matrix $\boldsymbol{Q} = \mathbb{E}[\boldsymbol{c}'\boldsymbol{c}'^H]$. Based on the power constraint, we define the SNR of user q as $\eta_q = \frac{E_s\Lambda^{-1}}{\sigma_{n,q}^2}$.

We assume that the BS has perfect knowledge of the instantaneous channel state information (CSI). And the receivers have perfect knowledge of their own CSI. Also, we define the scheduled user set, denoted as $\boldsymbol{K} \subset \mathbb{K}$ as the set of users who are actually scheduled (with a non-zero transmit power) by the transmitter at the time instant of interest. The transmitter serves users belonging to $\boldsymbol{K}$ with total n_e data streams, and user $q \in \boldsymbol{K}$ is served with $n_{u,q}$ data streams, $n_e = \sum_{q \in \boldsymbol{K}} n_{u,q}$.

B. Capacity Achieving Dirty-Paper Coding

The capacity region of MU-MIMO downlink broadcast channel is an acute problem in information theory. Since the transmitter hosts multiple antennas,the users' channels cannot be ranked anymore. After Dirty-Paper Coding was introduced [3], people studied the achievable rate region using DPC [4] [5]. The achievable rate region of the MIMO broadcast channel using DPC was confirmed to be equal to the capacity region of the MIMO broadcast channel in [1].

The idea of DPC comes from the illustration of conveying information using a dirty paper. Imagined a paper partially covered with dirt that is indistinguishable from ink. DPC states that if the writer knows where the dirt is to start with, the writer can convey just as much information by writing on the paper as if it were clean, even though the reader does not know where the dirt is.

The remarkable result of DPC is provided by [3]. Assuming a system model

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{S} + \boldsymbol{Z}$$

where the state $\boldsymbol{S} \sim N(0, Q\boldsymbol{I})$ and noise $\boldsymbol{Z} \sim N(0, N\boldsymbol{I})$ are multivariate Gaussian random variables. The input $\boldsymbol{X} \in \mathbb{R}^n$ satisfies the power constraint $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \leq P$. If $\boldsymbol{S}$ is unknown to both transmitter and receiver then the capacity is $C = log(1 + \frac{P}{Q+N})$. However, if the state $\boldsymbol{S}$ is known to the encoder, the capacity is

$$C = log(1 + \frac{P}{N})$$

which is also the capacity of a standard Gaussian channel with SNR $\frac{P}{N}$. Therefore, the state $\boldsymbol{S}$ does not affect the capacity of the channel, even though $\boldsymbol{S}$ is unknown to the receiver. In other words, if the transmitter has full knowledge of the interference, the capacity of the dirty paper channel is equal to the capacity of the channel with the interference completely absent.

DPC works similar to successive interference cancellation (SIC). SIC decodes the interference first to removing it from the receive signal, so the subtracted signal appears as if there is no interference. In comparison, DPC does this in the reverse way. When encode a user Bob's signal, the transmitter uses the knowledge of previously decoded signals so that the previously decoded signals appears invisible to Bob.

Here is the process of encoding K users. The transmitter rst encodes user 1s signal into C_1. With full knowledge of C_1, the transmitter encodes user 2s signal into C_2. Now, C_2 appears like a Gaussian interference to user 1. However, C_1 appears invisible to user 2. Again the the transmitter encodes user 3s signal into C_3 with full knowledge of C_1 and C_2. C_3 appears like a Gaussian noise to user 1 and user 2, but C_1 and C_2 are invisible to user 3. The encoding process continues until all K users are encoded. Note that different encoding orders will result in different capacity region. To obtain the capacity region, all permutations of the user ordering need to be considered.

C. Linear Precoding

Although DPC can achieve the capacity region, implementation of DPC requires significant additional complexity at both transmitter and receiver, and the problem of finding practical dirty paper codes that approach the capacity limit is still unsolved [6]. Therefore, linear precoding methods are used in practical systems. In the following, we mathematically define linear precoding and introduce three common linear precoders.

As described previously, the transmitted signal vector $\boldsymbol{c}'$ is the combination of signals for all co-scheduled users. This transmitted signal is precoded so that each user can decode their own signals. Linear Precoding (LP) is a subclass of transmission schemes to serve multiple users simultaneously. Although LP is not the optimal approach, LP has the advantage of low complexity.

When using LP, the transmit signal vector $\boldsymbol{c}'$ for total n_e data streams can be written as

$$\begin{aligned}\boldsymbol{c}' &= \boldsymbol{P}\boldsymbol{c} = \boldsymbol{W}\boldsymbol{S}^{1/2}\boldsymbol{c} \\ &= \sum_{q\in\boldsymbol{K}} \boldsymbol{P}_q\boldsymbol{c}_q = \sum_{q\in\boldsymbol{K}} \boldsymbol{W}_q\boldsymbol{S}_q^{1/2}\boldsymbol{c}_q\end{aligned} \tag{2}$$

where $\boldsymbol{c}$ is the symbol vector made of n_e unit-energy independent symbols. $\boldsymbol{P} \in \mathbb{C}^{n_t\times n_e}$ is the precoder made of two matrices: a power control diagonal matrix denoted as $\boldsymbol{S} \in \mathbb{R}^{n_e\times n_e}$, and a transmit beamforming matrix $\boldsymbol{W} \in \mathbb{C}^{n_t\times n_e}$. The expression can be further decomposed to per-user level. $\boldsymbol{c}_q \in \mathbb{C}^{n_{u,q}}$ is the symbol vector only for user q. $\boldsymbol{P} \in \mathbb{C}^{n_t\times n_{u,q}}$, $\boldsymbol{S}_q \in \mathbb{R}^{n_{u,q}\times n_{u,q}}$, and $\boldsymbol{W}_q \in \mathbb{C}^{n_t\times n_{u,q}}$ are user q's sub-matrices of $\boldsymbol{P}$, $\boldsymbol{S}$, and $\boldsymbol{W}$.

Equation (2) shows that the symbol vector is multiplied by a linear precoder $\boldsymbol{P}$ for transmission through multiple antennas. Careful selection of the precoders can reduce (or eliminate) interference among different streams by taking advantage of spatial separation between users and thereby support multiple users simultaneously.

Combine Equation (1) and Equation (2), we get

$$\boldsymbol{y}_q = \Lambda^{1/2}\boldsymbol{H}_q\boldsymbol{W}_q\boldsymbol{S}_q^{1/2}\boldsymbol{c}_q + \sum_{p\in\boldsymbol{K},p\neq q} \Lambda^{1/2}\boldsymbol{H}_q\boldsymbol{W}_p\boldsymbol{S}_p^{1/2}\boldsymbol{c}_p + \boldsymbol{n}_q \tag{3}$$

We observe that the second term is the signal for other users, which is interference for user q.

1) Matched Beamforming (MBF): In single-user case, the precoder $\boldsymbol{w}$ of matched beamforming is

$$\boldsymbol{w} = \frac{\boldsymbol{h}^H}{\|\boldsymbol{h}\|} \tag{4}$$

where $\boldsymbol{h}$ is the channel vector to the user. Matched beamforming (MBF) maximizes the SNR at the user as it transmits along the direction of the matched channel. Since the beamforming weights is the conjugate transpose of the channel vector, it is also called Conjugate beamforming.

When extending this matched beamformer directly to multi-user scenario, matched beamforming simply maximizes the power to each of the user and ignores the interference. Matched beamforming is the simplest form of MU-MIMO precoder.

Assume a single receive antenna at each of the user for simplicity. When the channel vectors of user p and q are not orthogonal ($\boldsymbol{h}_p\boldsymbol{h}_q^H \neq 0$), user p and q will receive signals intended for the other, which is interference from their own perspective. The best case of matched beamforming happens when the channel vectors of user p and q are orthogonal, as p and q do not receive interference, and the power to p and q is maximized.

2) Zero-Forcing Beamforming (ZFBF): Zero-forcing beamforming [7] aims to eliminate the inter-user interference. Here we assume a single receive antenna at each of the user for simplicity, and therefore each user is served with 1 data stream. The channel matrix from a n_t-antenna transmitter to K single-antenna user is $\boldsymbol{H}[K \times n_t]$. To eliminate the inter-user interference, the beamforming metrix needs to satisfy the condition that $\boldsymbol{HW}$ is diagonal. When $K \leq n_t$ and $\boldsymbol{H}$ is full rank, the precoders can be chosen as the normalized columns of the right pseudo inverse of $\boldsymbol{H}$. Let the right pseudo inverse of $\boldsymbol{H}$ be $\boldsymbol{F}$.

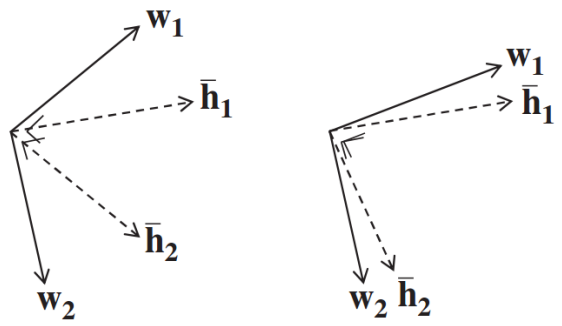
$$\boldsymbol{F} = \boldsymbol{H}^H\left(\boldsymbol{H}\boldsymbol{H}^H\right)^{-1} \tag{5}$$

The transmit precoder $\boldsymbol{w}_q$ for user q is the normalized columns of $\boldsymbol{F}$,

$$\boldsymbol{w}_q = \frac{\boldsymbol{F}(:,q)}{\|\boldsymbol{F}(:,q)\|} \tag{6}$$

Zero-forcing beamforming is illustrated by Fig. 1 for a two-user case. In Fig. 1, $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ are orthogonal to $\boldsymbol{h}_2$ and $\boldsymbol{h}_1$, respectively. Fig. 1a illustrates the case in which the channels of the two users are not orthogonal, and therefore the beamforming direction and the channel direction are different. As a result, the received power at the users decreases. Here we observe that zero-forcing beamforming trades power efficiency for interference-free transmissions, and ZFBF can be power inefficient when users' channels are not orthogonal. In comparison, when the channels of the two users are nearly orthogonal, the beamforming direction is close to the channel direction and less power is wasted on the other directions, as shown in Fig. 1b.

3) Regularized Zero-Forcing Beamforming (R-ZFBF): By forcing the interference to zero, a major problem of ZFBF is poor performance when the channels of the



(a) Non-orthogonal users (b) Quasi-orthogonal users

Fig. 1. Illustration of ZFBF precoding for a two-user scenario. Adapted from [2]

users are highly correlated, especially at low SNR. To tackle that issue, a regularized form of channel inversion, commonly denoted as Regularized Zero-Forcing Beamforming (R-ZFBF), is typically used. The regularized zero-forcing is modified as

$$\boldsymbol{F} = \boldsymbol{H}^H \left(\boldsymbol{H}\boldsymbol{H}^H + \alpha \boldsymbol{I}\right)^{-1} \tag{7}$$

where the regularization parameter α is design parameter, and can be optimized based on the design requirements. When $\alpha = 0$, R-ZFBF boils down to ZFBF and there is no multiuser interference. When $\alpha \to \infty$, R-ZFBF becomes MBF. The amount of interference increases with α, so one possible metric for choosing α is to maximize the SINR. When optimizing for SINR, the beamformer is also called MMSE beamforming.

III. Incentive for massive MIMO system

Large antenna array helps create favorable propagation condition and ensures the system to get the MIMO multiplexing gain, both in SU-MIMO and MU-MIMO. In the following, we discuss the limiting case when the number of transmit antennas approaches to infinity for SU-MIMO and MU-MIMO respectively, which is summarized from [8].

A. SU-MIMO

Consider a point-to-point MIMO link consists of a transmitter having an array of n_t antennas, a receiver having an array of n_r antennas. For each use of the channel we have

$$\boldsymbol{x} = \sqrt{\rho}\boldsymbol{G}\boldsymbol{s} + \boldsymbol{w}$$

where $\boldsymbol{s}$ is the n_t-component vector of transmitted signals, $\boldsymbol{x}$ is the n_r-component vector of received signals, $\boldsymbol{G}$ is the $n_r \times n_t$ propagation matrix of complex-valued channel coefficients, and $\boldsymbol{w}$ is the n_r-component vector of receiver noise, with each of the components being i.i.d. zero-mean and unit-variance circulary-symmetric complex-Gaussian random variables ($CN(0,1)$). The scalar ρ is a measure of the SNR of the link: it is proportional to the transmitted power divided by the noise-variance, and it also absorbs various normalizing constants. In what follows, we assume a normalization such that the expected total transmit power is unity,

$$\mathbb{E}\left[\|\boldsymbol{s}\|^2\right] = 1$$

With i.i.d. complex-Gaussian inputs, the mutual information between the input and the output of the point-to-point MIMO channel, under the assumption that the receiver has perfect knowledge of the channel matrix $\boldsymbol{G}$ is

$$C = \log\det\left(I_{n_r} + \frac{\rho}{n_t}\boldsymbol{G}\boldsymbol{G}^H\right) \tag{8}$$

where I_{n_r} denotes the $n_r \times n_r$ identity matrix, and the superscript H denotes the Hermitian transpose [9].

Another way to present Equation (8) is to express the achievable rate in terms of the singular values of the propagation matrix $\boldsymbol{G}$

$$\boldsymbol{G} = \boldsymbol{\Phi}\boldsymbol{D}_\nu\boldsymbol{\Psi}^H$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are unitary matrices of dimension $n_r \times n_r$ and $n_t \times n_t$ respectively, and $\boldsymbol{D}_\nu$ is a $n_r \times n_t$ diagonal matrix whose diagonal elements are the singular values, $\{\nu_1, \nu_2, \cdots \nu_{min(n_t,n_r)}\}$.

When expressing the achievable rate in Equation (8) in terms of the singular values, it is equivalent to the combined achievable rate of parallel links for which the lth link has an SNR of $\frac{\rho\nu_l^2}{n_t}$.

$$C = \sum_{l}^{min(n_t,n_r)} \log\left(1 + \frac{\rho\nu_l^2}{n_t}\right) \tag{9}$$

With respect to the achievable rate, we consider the best and the worst possible distribution of singular values, subject to the constraint obtained from [10]

$$\sum_{l}^{min(n_t,n_r)} \nu_l^2 = Tr\left(\boldsymbol{G}\boldsymbol{G}^H\right)$$

where Tr denotes trace. The worst case is when all but one of the singular values are equal to zero, and the best case is when all of the $min(n_t, n_r)$ singular values are equal, which is a direct consequence of the concavity of

the logarithm. The two cases bound the achievable rate as follows

$$\begin{aligned}\log\left(1+\frac{\rho Tr\left(\boldsymbol{G}\boldsymbol{G}^H\right)}{n_t}\right) &\leq C \\ &\leq min(n_t,n_r)\times\log\left(1+\frac{\rho Tr\left(\boldsymbol{G}\boldsymbol{G}^H\right)}{n_t min(n_t,n_r)}\right)\end{aligned} \tag{10}$$

If we assume that a normalization has been performed such that the magnitude of a propagation coefficient is typically equal to one, then $Tr(\boldsymbol{G}\boldsymbol{G}^H)\approx n_t n_r$, and the above bounds simplify as follows

$$\log\left(1+\rho n_r\right)\leq C\leq min(n_t,n_r)\log\left(1+\frac{\rho max(n_t,n_r)}{n_t}\right) \tag{11}$$

The worst case can potentially occurs for compact arrays under line-of-sight (LOS) propagation conditions such that the transmit array cannot resolve individual elements of the receive array. The equal singular value (best) case is approached when the entries of the propagation matrix are i.i.d. random variables.

Now we consider the limiting case when the number of transmit antennas grow large while keeping the number of receive antennas constant. Furthermore, we assume that the row-vectors of the propagation matrix are asymptotically orthogonal [11]

$$\left(\frac{\boldsymbol{G}\boldsymbol{G}^H}{n_t}\right)_{n_t\gg n_r}\approx\boldsymbol{I}_{n_t}$$

Together with capacity in Equation (8), the achievable rate becomes

$$\begin{aligned}C_{nt\gg n_r}&\approx\log\det\left(\boldsymbol{I}_{n_r}+\rho\boldsymbol{I}_{n_r}\right)\\&=n_r\log\left(1+\rho\right)\end{aligned} \tag{12}$$

which matches the upper bound in Equation (11).

From the above discussion, an excess number of transmit antennas, combined with asymptotic orthogonality of the propagation vectors, constitutes a highly desirable scenario, and give us the incentive to use large transmit antenna array.

B. MU-MIMO

After discussing SU-MIMO, now we study MU-MIMO. Consider a base station (BS) with an array of M antennas serves K single-antenna user terminals simultaneously. The propagation matrix $\boldsymbol{G}$ for downlink is a $K\times M$ matrix, which is composed of a $K\times M$ matrix $\boldsymbol{H}$ accounts for small-scale fading and a $K\times K$ diagonal matrix $\boldsymbol{D}_\beta^{1/2}$ accounts for large-scale fading. The diagonal elements of $\boldsymbol{D}_\beta$ constitute a $K\times 1$ vector $\boldsymbol{\beta}$ representing the large-scale fading coefficients to the K users.

$$\boldsymbol{G}=\boldsymbol{D}_\beta^{1/2}\boldsymbol{H}$$

The kth row-vector of $\boldsymbol{G}$ describes the channel vector between the M antennas at the BS and the kth terminal. We normalize the large-scale fading coefficients such that the small-scale fading coefficients typically have magnitudes of one.

The downlink transmission is again represented by

$$\boldsymbol{x}=\sqrt{\rho}\boldsymbol{G}\boldsymbol{s}+\boldsymbol{w}$$

where ρ is proportional to the ratio of power to noise-variance, and $\boldsymbol{w}$ is the $K\times 1$ vector of receiver noise whose components are independent and distributed as $CN(0,1)$. Just like before, the total transmit power is normalized such that the expected total transit power is unity and is independent of the number of antennas,

$$\mathbb{E}\left[\|\boldsymbol{s}\|^2\right]=1$$

The known capacity result for this channel assumes that the terminals as well as the base station know the channel [1] [12]. Let $\boldsymbol{D}_\gamma$ be a diagonal matrix whose diagonal elements constitute a $K\times 1$ vector $\boldsymbol{\gamma}$. To obtain the sum-capacity requires performing a constrained optimization,

$$\begin{aligned}C_{\text{sum}}&=\max_{\gamma_k}\log\det\left(\boldsymbol{I}_M+\rho\boldsymbol{G}^H\boldsymbol{D}_\gamma\boldsymbol{G}\right)\\&\text{subject to }\sum_{k=1}^{K}\rho_k=1\end{aligned} \tag{13}$$

Consider MU-MIMO with large transmit arrays, the number of antennas greatly exceeds the number of terminals. Under the most favorable propagation conditions, the row-vectors of the propagation matrix are asymptotically orthogonal

$$\left(\frac{\boldsymbol{G}\boldsymbol{G}^H}{M}\right)_{M\gg K}=\boldsymbol{D}_\beta^{1/2}\left(\frac{\boldsymbol{H}\boldsymbol{H}^H}{M}\right)_{M\gg K}\boldsymbol{D}_\beta^{1/2}\approx\boldsymbol{D}_\beta$$

The sum-capacity becomes

$$\begin{aligned}C_{\text{sum},M\gg K}&=\max_{\gamma_k}\log\det\left(\boldsymbol{I}_K+\rho\boldsymbol{D}_\gamma^{1/2}\boldsymbol{G}\boldsymbol{G}^H\boldsymbol{D}_\gamma^{1/2}\right)\\&\approx\max_{\gamma_k}\log\det\left(\boldsymbol{I}_K+M\rho\boldsymbol{D}_\gamma\boldsymbol{D}_\beta\right)\\&=\max_{\gamma_k}\sum_{k=1}^{K}\log\left(1+M\rho\gamma_k\beta_k\right)\end{aligned} \tag{14}$$

where $\boldsymbol{\gamma}$ is constrained as in Equation (13).

This result makes intuitive sense if the rows of the propagation matrix are nearly orthogonal, which occurs

asymptotically as the number of antennas grows. When the channels to different users are orthogonal, even the simplest conjugate beamforming achieves this sum rate. That is, adding extra number of transmit antennas at the BS makes the performance of linear precoders approaches to the optimal case.

IV. Performance of Downlink Linear Precoders in Multi-user MIMO

In this section, we derive the SNR or SINR expressions for matched beamforming and zero-forcing beamforming in the large system limit [8]. Specifically, we consider the number of transmit antenna M and the number of users K both approach to infinity, but with a fixed ratio $\alpha = \frac{M}{K}$. We consider both perfect CSI and imperfect CSI. The obtained expressions are summarized in Table I.

The downlink transmission is modeled as

$$\boldsymbol{x} = \boldsymbol{G}\boldsymbol{s} + \boldsymbol{w}$$

where $\boldsymbol{s}$ is a precoded version of the data symbols $\boldsymbol{q}$. Each component of $\boldsymbol{s}$ has average power $\frac{\rho}{M}$. Further, we assume that the channel matrix $\boldsymbol{G}$ has i.i.d. $CN(0,1)$ entries. User k receives the kth component.

Before diving into matched beamforming and ZF beamforming, we first discuss the performance of an intereference-free (IF) system that will subsequently serve as the upper bound of the system. IF system assumes no inter-user interference exists. The SNR of IF system is calculate by ignoring the inter-user interference. In the IF scenario, user k receives the signal x_k

$$x_k = \sqrt{\sum_{l=1}^{M} |g_{lk}|^2} q_k + w_k$$

Since $\frac{\sum_{l=1}^{M} |g_{lk}|^2}{M} \to 1$ as $M \to \infty$, and $\mathbb{E}\left[q_k q_k^H\right] = \frac{\rho}{K}$, the SNR for IF systems converges to $\rho\alpha$ as $M \to \infty$.

A. Perfect CSI

1) Zero-Forcing Beamforming: The zero-forcing precoder is the pseudoinverse of the channel. Intuitively, when the number of transmit antenna M grows, the channel matrix $\boldsymbol{G}$ tends to have nearly orthogonal rows as the terminals are not correlated due to their physical separation. This assures that the performance of ZF precoding will be close to that of the IF system. However, a disadvantage of ZF is that processing cannot be done distributedly at each antenna separately and an inversion of a large matrix is required when M is large. With ZF precoding, all data must instead be collected at a central node that handles the processing.

The ZF precoder sets

$$\boldsymbol{s} = \frac{1}{\sqrt{\gamma}} \boldsymbol{G}^H (\boldsymbol{G}\boldsymbol{G}^H)^{-1} \boldsymbol{q}$$

where γ normalizes the average power in $\boldsymbol{s}$ to ρ. A suiable choice for γ is $\gamma = \frac{Tr(\boldsymbol{G}^H\boldsymbol{G})^{-1}}{K}$, which averages fluctuations in transmit power due to $\boldsymbol{G}$. The receive signal x_k with ZF beamforming becomes

$$x_k = \frac{q_k}{\sqrt{\gamma}} + w_k$$

The received SNR at the terminal equals

$$\text{SNR} = \frac{\rho}{K\gamma} = \frac{\rho}{Tr(\boldsymbol{G}^H\boldsymbol{G})^{-1}}$$

When both the number of terminals K and the number of BS antennas M grow large, but with fixed ratio $\alpha = \frac{M}{K}$, $Tr(\boldsymbol{G}^H\boldsymbol{G})^{-1}$ converges to a fixed deterministic value [13]

$$Tr(\boldsymbol{G}^H\boldsymbol{G})^{-1} \to \frac{1}{\alpha - 1}, \text{ as } K, M \to \infty, \frac{M}{K} = \alpha$$

Therefore, the SNR approaches $\rho(\alpha - 1)$ as M and K approach to infinity, with a fixed ratio $\alpha = \frac{M}{K}$. We observe that ZF precoding achieves an SNR for an IF system with $M - K$ transmit antennas in the large antenna regime. Also, the SNR increases unbounded as the transmit power ρ increases. However, when $M = K$, SNR becomes 0.

2) Matched Beamforming: The Matched precoder sets

$$\boldsymbol{s} = \frac{1}{\sqrt{\gamma}} \boldsymbol{G}^H \boldsymbol{q}$$

with $\gamma = \frac{Tr(\boldsymbol{G}\boldsymbol{G}^H)}{K}$. A few simple manipulations lead to an asymptotic expression of the SINR $= \frac{\rho\alpha}{\rho+1}$. From the SINR expression, we observe that the SINR can achieve as high as desired by increasing the number of transmit antennas. However, the Matched precoder exhibits an error floor since no matter how large the transmit power is used, the SINR approaches α. That is, when $\rho \to \infty$, SINR $\to \alpha$.

TABLE I
SNR and SINR expressions for some standard precoding techniques as $K, M \to \infty$, $\frac{M}{K} = \alpha$

Precoding technique	Perfect CSI	Imperfect CSI
Benchmark: Interference Free	$\rho\alpha$	N/A
Zero-Forcing	$\rho(\alpha - 1)$	$\frac{\xi^2\rho(\alpha-1)}{(1-\xi^2)\rho+1}$
Matched Beamforming	$\frac{\rho\alpha}{\rho+1}$	$\frac{\xi^2\rho\alpha}{\rho+1}$

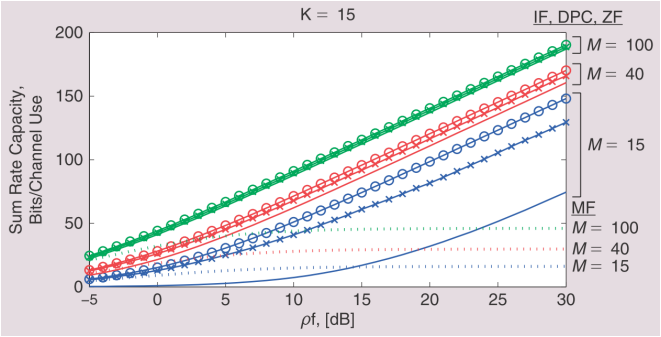


Fig. 2. Sum-rate capacities of single-cell MU-MIMO precoding techniques. The channel is i.i.d. $CN(0,1)$ and there are $K = 15$ terminals. Circles show the performance of interference free (IF) systems, DPC is denoted by "x", solid lines refer to ZFBF, and the dotted lines refer to matched beamforming.

3) Comparison of MBF and ZFBF: From the above discussion, we know the asymptotic SNR or SINR for Matched beamforming and ZF beamforming. For a fixed α, SNR for ZF beamforming can grow unrestricted as the transmit power increases, while Matched beamforming will reach an upper bound of α. Also, we observe that the SNR of ZF beamforming is lower than the SINR of Matched beamforming in the low power regime, specifically when $\rho < \frac{1}{\alpha-1}$. However, ZF beamforming has performance better than Matched beamforming once $\rho > \frac{1}{\alpha-1}$.

Figure. 2 shows the ergodic sum-rate capacities for Matched precoding, ZF precoding, and DPC. IF system is also shown as benchmark performance. In all cases, $K = 15$ users are served and we show results for $M = 15, 40, 100$.

First we observe that the gap between DPC and ZF beamforming becomes smaller as M increases. When $M = 100$, ZF precoding achieves performance nearly as good as DPC, even the IF system. This result suggests that the higher complexity of DPC cannot be justified by the performance gain.

We also observe that ZF precoding outperforms Matched precoding in high SNR region, while Matched precoding is better at low SNR, as suggested by the asymptotic SNR.

B. Imperfect CSI

Here we discuss the SINR of Matched precoding and ZF precoding when the BS has imperfect CSI. Let $\hat{\boldsymbol{G}}$ denote the minimum mean square error (MMSE) channel estimate of the downlink channel. The estimate can be expressed as

$$\hat{\boldsymbol{G}} = \xi \boldsymbol{G} + \sqrt{1-\xi^2}\boldsymbol{E}$$

where $< 0 \leq \xi \leq 1$ represents the reliability of the estimate, and $\boldsymbol{E}$ is an estimation error matrix with i.i.d. $CN(0,1)$ distributed entries.

For any reliability ξ, the SINR for zero-forcing beamforming is $\frac{\xi^2\rho(\alpha-1)}{(1-\xi^2)\rho+1}$. And the SINR for matched beamforming is $\frac{\xi^2\rho\alpha}{\rho+1}$, as summarized in Table I.

We observe that when the CSI is imperfect, the channel error reduce the receive signal strength for Matched beamforming. In comparison, ZF precoding encounter loss not only in signal strength but also the inter-user interference term. When the BS has perfect CSI, ZF beamforming prevent any inter-user interference. However, when the CSI is imperfect, ZF beamforming can no longer completely eliminate inter-user interference. As a result, we can conclude that ZF beamforming is more sensitive to channel error as the imperfect channel leads to both reduction in receive signal strength and increase in interference.

V. Multi-cell Scenario

In this section, we investigate a fundamental limit caused by pilot contamination in noncooperative cellular MU-MIMO systems. Although increasing number of transmit antennas M averages out the thermal noise and small-scale Rayleigh fading, the interference from other cells due to pilot contamination does not vanish with increasing number of BS antennas. The following discussion summarizes the results from [10] [8].

A. Pilot contamination

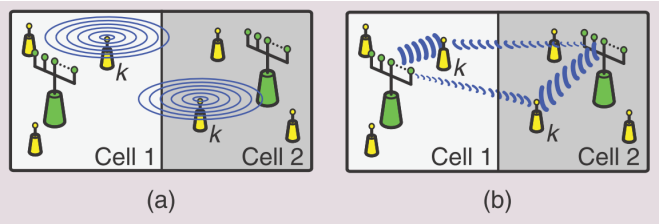


Fig. 3. Illustration of the pilot contamination concept. (a) during the training phase, the BS in cell 1 overhears the pilot transmission from other cells. (b) as a consequence, the transmitted vector from BS 1 will be partially beamformed to the terminals in cell 2.

In a large antenna system, to efficiently acquire the propagation matrix, the users send orthogonal uplink pilots for the BS to estimate the uplink channel. The downlink channel is obtained assume channel reciprocity. However, there are only limited number of orthogonal pilots, and the users in different cells have to reuse those pilots. Therefore, during the pilot estimation process, it is possible that users in different cells transmit the same

uplink pilot simultaneously and contaminate the channel estimation. As a result, the channel estimation contains channel for other users and cause inter-cell interference in the following downlink transmission. Figure 3 illustrates the process of pilot contamination.

B. Performance of Linear Precoders

Consider a multi-cell system where each cell has a BS with M antennas and serves K users, $K \leq M$. In the worst case, the pilot signals in all other cells are perfectly synchronized with the pilot signals in cell n. In this case, the channel estimate $\hat{\boldsymbol{G}}_{nn}$ gets contaminated from pilot signals in other cells

$$\hat{\boldsymbol{G}}_{nn} = \sqrt{\rho_p}\boldsymbol{G}_{nn} + \sqrt{\rho_p}\sum_{i\neq n}\boldsymbol{G}_{in} + \boldsymbol{V}_n$$

Where ρ_p is a measure of the SNR during of the pilot transmission phase, $\boldsymbol{G}_{in}$ represents the channel from the BS in cell n to the K users in cell i, and $\boldsymbol{V}_n$ is a matrix of receiver noise during the training phase, uncorrelated with all propagation matrices, and comprises i.i.d. $CN(0,1)$ distributed elements. Due to the geometry of the cells, $\boldsymbol{G}_{nn}$ is generally stronger than $\boldsymbol{G}_{in}, i \neq n$.

1) Matched Beamforming: When the BS use Matched beamforming for downlink, the precoder matrix is $\hat{\boldsymbol{G}}_{nn}^H$. The kth user in cell j receives signal the kth element in $\boldsymbol{x}_j$

$$\begin{aligned}\boldsymbol{x}_j &= \sqrt{\rho_f}\sum_n \boldsymbol{G}_{jn}\hat{\boldsymbol{G}}_{nn}^H\boldsymbol{q}_n + \boldsymbol{w}_j \\ &= \sqrt{\rho_f}\sum_n \boldsymbol{G}_{jn}\left[\sqrt{\rho_p}\sum_i \boldsymbol{G}_{in} + \boldsymbol{V}_n\right]^H \boldsymbol{q}_n + \boldsymbol{w}_j\end{aligned} \tag{15}$$

where ρ_f is a measure of the SNR during of the downlink data transmission phase, and $\boldsymbol{w}_j$ is the noise at the K receivers in cell j.

As M grows large, only terms where $j = i$ remain significant. Further, as M grows, the effect of small-scale Rayleigh fading vanishes. And the SIR of terminal l becomes

$$\text{SIR} = \frac{\beta_{jjl}^2}{\sum_{n\neq j}\beta_{jnl}^2} \tag{16}$$

where β_{kjl} represents the large-scale fading between terminal l in the kth cell and the BS in cell j.

Form Equation (16), we observe that that devoting more power to the training phase does not decrease the pilot contamination effect and leads to the same SIR. Note that this is a consequence of the worst-case-scenario assumption that the pilot transmissions in all cells overlap. If the pilot transmissions are staggered so that pilots in one cell collide with data in other cells, devoting more power to the training phase is indeed beneficial.

2) Zero-Forcing Beamforming: When the BS use ZF beamforming for downlink, the precoder matrix is the pseudoinverse of the channel estimate $\hat{\boldsymbol{G}}_{nn}^H(\hat{\boldsymbol{G}}_{nn}\hat{\boldsymbol{G}}_{nn}^H)^{-1}$. when M grows, only products of correlated terms remain significant. And the SIR of terminal l becomes

$$\text{SIR} = \frac{\beta_{jjl}^2 / \left(\sum_i \beta_{ijl} + \frac{1}{\rho_p}\right)^2}{\sum_{n\neq j}\beta_{jnl}^2 / \left(\sum_i \beta_{inl} + \frac{1}{\rho_p}\right)^2} \tag{17}$$

which is dependent of ρ_p, the SNR during the pilot training process.

3) Comparison of Matched and ZF beamforming With Numerical Results: Fig. 4 shows the SIR numerical results for Matched and ZF precoders as the number of BS antennas N increases. In the analysis, hexagonal cells are assumed and the large-scale fading factor is modeled by a log-normal distribution. The number of users in each cell $K = 10$. Also, large enough ρ_p and ρ_f are used so that the SNRs are large enough and the performance is limited by pilot contamination.

In Fig. 4, the two uppermost curves show the mean SIR as $M \to \infty$. The tow bottom curves show the mean SIR for Matched and ZF precoding for finite M. We observe that the limit of Matched beamforming is higher than ZF beamforming. However, ZF outperforms Matched beamforming for finite M. Furthermore, with finite M, ZF beamforming can achieve closer to the limit case. Although the limit of Matched beamforming is higher, Matched precoding requires much larger antenna array than ZF precoding does to reach a given mean SIR.

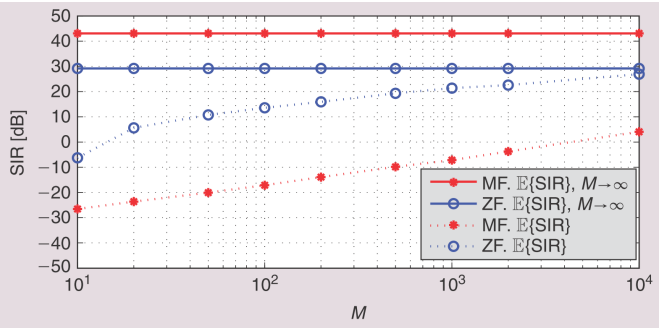


Fig. 4. SIRs for Matched and ZF precoders as a function of M. The two uppermost curves are asymptotic mean values of the SIR as $M \to \infty$. The bottom two curves show mean values of the SIR for finite M. The number of terminals served is $K = 10$.

C. Mitigating Pilot Contamination

Pilot contamination impairs the system performance and does not vanish as the number of BS antennas increases. Therefore, mitigating the pilot contamination effect is an important issue and several approaches have be proposed.

The first category is protocol-based. For example, a time-shifted (asynchronous) protocol to stagger the pilot transmission is proposed in [14] [15] [16]. In such protocol, uplink pilots are transmitted when other conflict users are receiving downlink signals. However, one concern of this approach is that the transmit power of uplink pilot is generally much lower than the downlink transmit power, resulting higher channel estimation errors.

The second category is by precoding methods. For example, [17] propose a distributed single-cell precoding method in which the precoding matrix at one BS is designed to minimize the sum of the squared error of its own users and interference to the users in all other cells. Other works also consider multi-cell cooperation [18] [19]. However, the cooperation does not scale with the number of antennas since the information exchange overhead required among the BSs increases with the number of antennas.

The third category is AOA-based methods. [20] [21] [22] shows that users with mutually non-overlapping AOA PDFs hardly contaminate each other even if they use the same pilot sequence. Therefore, one can assign identical pilot sequences only to users of this type, and a coordinated scheme is discussed in [22].

The fourth category is blind methods. For example, [23] propose an eigenvalue-decomposition-based (EVD) channel estimation. The EVD-based estimation is based on the assumption that the channel vectors from different users are orthogonal. This assumption allows one to estimate channel vectors using the statistics of the received data.

VI. Conclusion

In this report, we study the multi-user downlink precoding method, for both the capacity achieving DPC and a set of frequent-use linear precoders (matched precoder, zero-forcing precoder, and regularized-ZF precoder). Also, we discuss the advantage of using a large transmit antenna array for both single user and multi-user scenario. In addition, we study the performance of downlink linear precoders under limiting case in which the number of BS antennas and the number of users approaches to infinity but with a fixed ratio. Next, we also discuss the pilot contamination problem in multi-cell scenario and some proposed approaches to mitigate the effect of pilot contamination.

References

[1] H. Weingarten, Y. Steinberg, and S. S. Shamai, "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE transactions on information theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[2] B. Clerckx and C. Oestges, *MIMO wireless networks: Channels, techniques and standards for multi-antenna, multi-user and multi-cell systems*. Academic Press, 2013.

[3] M. Costa, "Writing on dirty paper (corresp.)," *IEEE transactions on information theory*, vol. 29, no. 3, pp. 439–441, 1983.

[4] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Second order statistics of nlos indoor mimo channels based on 5.2 ghz measurements," in *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*, vol. 1. IEEE, 2001, pp. 156–160.

[5] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[6] J. Lee and N. Jindal, "Dirty paper coding vs. linear precoding for mimo broadcast channels," in *Signals, Systems and Computers, 2006. ACSSC'06. Fortieth Asilomar Conference on*. IEEE, 2006, pp. 779–783.

[7] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part i: channel inversion and regularization," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 195–202, 2005.

[8] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[9] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell labs technical journal*, vol. 1, no. 2, pp. 41–59, 1996.

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

[11] M. Matthaiou, M. R. McKay, P. J. Smith, and J. A. Nossek, "On the condition number distribution of complex wishart matrices," *IEEE Transactions on Communications*, vol. 58, no. 6, pp. 1705–1717, 2010.

[12] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.

[13] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum rate," in *in Proc. 40th Annual Allerton Conf. Communications, Control and Computing*. Citeseer, 2002.

[14] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user tdd systems," in *Communications (ICC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–5.

[15] F. Fernandes, A. Ashikhmin, and T. Marzetta, "Interference reduction on cellular networks with large antenna arrays," in *Proc. of IEEE International Conference on Communications (ICC)*, 2012.

[16] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative tdd large scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 192–201, 2013.

[17] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell tdd systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, 2011.

[18] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell mimo downlink with cell cooperation and fair scheduling: A large-system limit analysis," *IEEE Transactions on Information Theory*, vol. 57, no. 12, pp. 7771–7786, 2011.

[19] H. Huh, A. M. Tulino, and G. Caire, "Network mimo with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2911–2934, 2012.

[20] M. Filippou, D. Gesbert, and H. Yin, "Decontaminating pilots in cognitive massive mimo networks," in *Wireless Communication Systems (ISWCS), 2012 International Symposium on*. IEEE, 2012, pp. 816–820.

[21] H. Yin, D. Gesbert, M. C. Filippou, and Y. Liu, "Decontaminating pilots in massive mimo systems," in *Communications (ICC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 3170–3175.

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

[23] H. Q. Ngo and E. G. Larsson, "Evd-based channel estimation in multicell multiuser mimo systems with very large antenna arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3249–3252.