2013

Performance of Uplink Multiuser Massive MIMO Systems

Zhengdao Wang
Iowa State University, zhengdao@iastate.edu

Abstract
We study the performance of uplink transmission in a large-scale (massive) MIMO system, where all the transmitters have single antennas and the receiver (base station) has a large number of antennas. Specifically, we analyze achievable degrees of freedom of the system without assuming channel state information at the receiver. Also, we quantify the amount of power saving that is possible with increasing number of receive antennas.

PERFORMANCE OF UPLINK MULTIUSER MASSIVE MIMO SYSTEMS

Zhengdao Wang

Dept. of Electrical & Computer Engineering, Iowa State University. Email: zhengdao@iastate.edu

ABSTRACT

We study the performance of uplink transmission in a large-scale (massive) MIMO system, where all the transmitters have single antennas and the receiver (base station) has a large number of antennas. Specifically, we analyze achievable degrees of freedom of the system without assuming channel state information at the receiver. Also, we quantify the amount of power saving that is possible with increasing number of receive antennas.

1. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems are a type of cellular communication where the base station is equipped with a large number of antennas. The base station serves multiple mobile stations that are usually equipped with a small number of antennas, typically one. Massive MIMO holds good potentials for improving future communication system performance. There are several challenges with designing such massive MIMO systems, including e.g., channel state information (CSI) acquisition, base station received signal processing, downlink beamforming with imperfect CSI, etc. For multi-cell system, pilot contamination and inter-cell interference also need to be dealt with. There is already a body of results in the literature about the analysis and design of large MIMO systems; see e.g., the overview article [5] and references there in.

To reveal the potential that is possible with massive MIMO systems, it is important to quantify the achievable performance of such systems in realistic scenarios. For example, it is too optimistic to assume that perfect CSI can be acquired at the base station in the uplink, because such acquisition takes time, energy, and channel estimation error will always exist. For the downlink, in order to perform effective beamforming, CSI is again needed, which needs to be either estimated by the mobile stations and then fed back to the base station, which is a non-trivial task, or, acquired by the base station by exploiting channel reciprocity in a time-division duplexing setup.

In this paper, we are interested in performance of the *uplink* transmission in a *single-cell* system. In particular, we ask what rates can be achieved in the uplink by the mobile users if we assume realistic channel estimation at the base station. Similar analysis has been performed in [4], but the analysis therein assumes equal power transmission during the channel training phase and the data transmission phase. Also the training duration has not be optimized, and the effect of channel coherence interval was not incorporated.

Our main analysis framework is similar to that of [3]. The system we are analyzing can be viewed as a point-to-point MIMO channel if the mobile stations are allowed to fully cooperate. Therefore, the rates obtained in [3], and the stronger result on non-coherent MIMO channel capacity in [7] can serve as upper bound for the system sum rate. Furthermore, the training strategy as optimized in [3] involves orthogonal signaling from the multiple transmit antennas and therefore is directly applicable to the multi-user single-cell system of interest.

For a system with K mobile users, M base station antennas, and block fading channel with coherence interval T, we derive achievable rate using linear channel estimation and linear base station (front-end) processing; see Section 3. The total degrees of freedom (DoF), to be defined more precisely later, is quantified in Theorem. 1. We also quantify the needed transmission power for achieving a given rate, when $M \gg 1$, which is an refinement of the corresponding result in [4].

2. SYSTEM MODEL

Notation: We use $A^\dagger$ to denote the Hermitian transpose of a matrix A, I_K to denote a $K \times K$ identity matrix, $\mathbb{C}$ to denote the complex number set, $\lfloor \cdot \rfloor$ to denote the integer floor operation, i.i.d. to denote "independent and identically distributed", and $\mathcal{CN}(0,1)$to denote circularly symmetric complex Gaussian distribution with zero mean and unit variance.

We consider a single-cell uplink system, where there are K mobile users and one base station. Each user has one transmit antenna, and the base station has M receive antennas. The received signal at the base station is

$$y = Hs + n \tag{1}$$

where $H \in \mathbb{C}^{M\times K}$ is the channel matrix, $s \in \mathbb{C}^{K\times 1}$ is the transmitted signals from all the K users; $n \in \mathbb{C}^{M\times 1}$ is the additive noise, $y \in \mathbb{C}^{M\times 1}$ is the received signal. We make the following assumptions:

A1) The channel is block fading such that within a *coherence interval* of T channel uses or time slots, the channel remains constant. The entries of H are i.i.d. and taken from $\mathcal{CN}(0,1)$. The channel changes independently from block to block. The CSI is neither available at the transmitters nor at the receiver.

A2) Entries of the noise vector n are i.i.d. and from $\mathcal{CN}(0,1)$. Noises in different channel uses are independent.

A3) The average transmit power per user is P. So within a coherence interval the total transmitted energy is PT. We do not impose a peak power constraint.

In summary, the system has four parameters, (M, K, T, P). We will allow the system to operate in the ergodic regime, so coding and decoding can occur over multiple coherent intervals.

3. ACHIEVABLE RATES

We assume that $K \le M$ and $K < T$ in this section. To derive the achievable rates for the users, we use a well-known scheme that consists of two phases (see e.g., [3]):

Training Phase. This phase consists of K time intervals. The training signal transmitted by the users can be represented by a $K \times K$ matrix Φ such that $\Phi\Phi^\dagger = EI_K$, where E is the total training energy per user per coherent interval.

Data Transmission Phase. Information-bearing symbols are transmitted by the users in the remaining $T-K$ time intervals. The average energy per symbol per user is $P_\mathrm{d} = (PT - E)/(T-K)$.

3.1. Channel estimation

In the training phase, we will choose $\Phi = \sqrt{E}I_K$ for simplicity. Other scaled unitary matrix can also be used without affecting the achievable rate. Note that the transmission power is allowed to vary from the training phase to the data transmission phase. Also, setting the training period equal to the total number of transmit antennas possesses certain optimality as derived in [3]. With our choice of Φ, the received signal $Y_\mathrm{p} \in \mathbb{C}^{M\times K}$ during the training phase can be written as

$$Y_\mathrm{p} = H\Phi + N = \sqrt{E}H + N \tag{2}$$

where $N \in \mathbb{C}^{M\times K}$ is the additive noise. The equation describes $M \times K$ independent identities, one for each channel coefficient. The (linear) minimum mean-squared error (MMSE) estimate for the channel H is given by

$$\hat{H} = \frac{\sqrt{E}}{E+1}Y_\mathrm{p} = \frac{E}{E+1}H + \frac{\sqrt{E}}{E+1}N. \tag{3}$$

The channel estimation error is defined as

$$\tilde{H} = H - \hat{H} = \frac{1}{E+1}H - \frac{\sqrt{E}}{E+1}N. \tag{4}$$

It is well known and easy to verify that the elements of $\hat{H}$ are i.i.d. complex Gaussian with zero mean and variance

$$\sigma^2_{\hat{H}} = \frac{E}{E+1}, \tag{5}$$

and the elements of $\tilde{H}$ are i.i.d. complex Gaussian with zero mean and variance

$$\sigma^2_{\tilde{H}} = \frac{1}{E+1}. \tag{6}$$

Moreover, $\hat{H}$ and $\tilde{H}$ are in general uncorrelated as a property of linear MMSE estimator, and in this case independent thanks to the Gaussian assumptions.

3.2. Equivalent channel

Once the channel is estimated, the base station has $\hat{H}$ and will decode the users' information using $\hat{H}$. We can write the received signal as

$$y = \hat{H}s + \tilde{H}s + n := \hat{H}s + v \tag{7}$$

where $v := \tilde{H}s + n$ is the new equivalent noise containing actual noise n and self interference $\tilde{H}s$ caused by inaccurate channel estimation. Assuming that each element of s has variance P_d during the data transmission phase, and there is no cooperation among the users, the variance of each component of v is

$$\sigma_v^2 = \frac{KP_\mathrm{d}}{E+1} + 1. \tag{8}$$

If we replace v with a zero-mean complex Gaussian noise with equal variance σ_v^2, but independent of s, then the system described in (7) can be viewed as MIMO system with perfect CSI at the receiver, and equivalent signal to noise ratio (SNR)

$$\rho := \frac{P_\mathrm{d}\sigma^2_{\hat{H}}}{\sigma_v^2} = \frac{P_\mathrm{d}E}{KP_\mathrm{d} + E + 1} = \frac{P_\mathrm{d}}{1 + \frac{KP_\mathrm{d}+1}{E}}. \tag{9}$$

The SNR is the signal power from a single transmitter per receive antenna divided by the noise variance per receive antenna. It is a standard argument that a noise equivalent to v but assumed independent of s is "worse"; see e.g., [3]. As a result, the derived rate based on such assumption is achievable. In the following, for notational brevity, we assume that v in (7) is independent of s without introducing a new symbol to represent the equivalent *independent* noise.

Note that the effective SNR ρ is the actual SNR P_d divided by a loss factor $1+(KP_\mathrm{d}+1)/E$. The loss factor can be made small if the energy E used in the training phase is large.

3.3. Energy splitting optimization

The energy in the training phase can be optimized to maximize the effective SNR ρ in (9), as has been done in [3, Theorem 2]. We adapt the result below for our case because it is relevant to our discussion. Specifically, let $\alpha := E/(PT)$ denote the percentage of energy devoted to training within one coherent interval. Define an auxiliary variable when $T \neq 2K$:

$$\gamma := \frac{(1+PT)(T-K)}{PT(T-2K)} \tag{10}$$

which is positive if $T > 2K$ and negative if $T < 2K$. The optimal value for α that maximizes ρ is given as follows:

$$\alpha = \begin{cases} \gamma - \sqrt{\gamma(\gamma-1)}, & T > 2K \\ \frac{1}{2}, & T = 2K \\ \gamma + \sqrt{\gamma(\gamma-1)}, & T < 2K \end{cases} \tag{11}$$

The maximized effective SNR ρ is given as

$$\rho = \begin{cases} \frac{PT}{T-2K}(\sqrt{\gamma} - \sqrt{\gamma-1})^2, & T > 2K \\ \frac{(PT)^2}{2T(1+PT)}, & T = 2K \\ \frac{PT}{2K-T}(\sqrt{-\gamma} - \sqrt{-\gamma+1})^2, & T > 2K \end{cases} \tag{12}$$

At high SNR ($P \gg 1$), the optimal values are

$$\alpha = \frac{\sqrt{T-K}}{\sqrt{T-K}+\sqrt{K}}, \quad \rho = \frac{T}{(\sqrt{T-K}+\sqrt{K})^2}P. \tag{13}$$

At low SNR ($P \ll 1$), the optimal values are

$$\alpha = \frac{1}{2}, \quad \rho = \frac{(PT)^2}{4(T-K)}. \tag{14}$$

3.4. Achievable rates

Given the channel model (7), linear processing can be applied to y to recover s, as in e.g., [4]. Let $A \in \mathbb{C}^{K\times M}$ denote the linear processing matrix. The processed signal is

$$\hat{s} := Ay = A\hat{H}s + Av. \tag{15}$$

The Maximum Ratio Combining (MRC) processing is obtained by setting $A = \hat{H}^\dagger$. The Zero-Forcing (ZF) processing is obtained by setting $A = (\hat{H}^\dagger\hat{H})^{-1}\hat{H}^\dagger$.

Based on the equivalent channel model, viewed as a multi-user MIMO systems with perfect receiver CSI and equivalent SNR ρ, the achievable rates lower bounds derived in [4, Propositions 2 and 3] can then be applied. Specifically, for MRC the following ergodic rate per user is achievable:

$$R^{(\text{MRC})} := \left(1-\frac{K}{T}\right)\log_2\left(1+\frac{\rho(M-1)}{\rho(K-1)+1}\right). \tag{16}$$

For ZF, the following rate per user is achievable:

$$R^{(\text{ZF})} := \left(1-\frac{K}{T}\right)\log_2\left[1+\rho(M-K)\right]. \tag{17}$$

Note that the factor $(1-\frac{K}{T})$ is due to the fact that during one coherence interval of length T, K time slots have been used for the training purpose. The number of data transmission slots is $T-K$, and the achieved rate needs to be averaged over T channel uses.

4. DEGREES OF FREEDOM

We define the DoF of the system as

$$d(M,K,T) := \sup \lim_{P\to\infty} \frac{R^{(\text{total})}(P)}{\log_2(P)} \tag{18}$$

where the supremum is taken over the totality of all reliable communication schemes for the system, and $R^{(\text{total})}$ denotes the sum rate of the K users under the power constraint P. We may also speak of the (achieved) degree of freedom of one user for a particular achievability scheme, which is the achieved rate of the user normalized by $\log_2(P)$ in the limit of $P \to \infty$. The DoF measures the multiplexing gain offered by the system when compared to a reference point-to-point single-antenna communication link, in the high SNR regime.

Theorem 1 *For an (M,K,T) MIMO uplink system with M receive antennas, K users, and coherence interval T, the total DoF of the system is*

$$d(M,K,T) = K^*\left(1-\frac{K^*}{T}\right). \tag{19}$$

where $K^* := \min(M, K, \lfloor T/2 \rfloor)$. □

Proof: To prove the converse, we observe that if we allow the K transmitters to cooperate, then the system is a point-to-point MIMO system with K transmit antennas, M receive antennas, and with no CSI at the receiver. The DoF of this channel has been quantified in [7], in the same form as in the theorem. Without cooperation, the users can at most achieve a rate as high as in the cooperation case.

To prove the achievability, we first look at the case $K^* < M$. In this case, we note that if we allow only K^* users to transmit, and let the remaining users be silent, then using the achievability scheme describe in Section 3, each of the K^* users can achieve a rate per user using the zero-forcing receiver given as follows (cf. (17))

$$\left(1-\frac{K^*}{T}\right)\log_2\left[1+\rho(M-K^*)\right]. \tag{20}$$

Note that the condition $K^* < M$ is needed. If we choose $E = KP$ and $P_\text{d} = P$, then the effective SNR in (9) becomes

$$\rho = \frac{P}{1+\frac{KP+1}{KP}}. \tag{21}$$

It can be seen that as $P\to\infty$, $\log(\rho)/\log(P) \to 1$ and a DoF per user of $(1-K^*/T)$ is achieved. The total achieved DoF is therefore $K^*(1 - K^*/T)$. Although better energy splitting is possible, as in Section 3.3, it will not improve the DoF.

When $K^* = M$, the case is more subtle. In this case the zero-forcing receive is no longer sufficient. In fact, even the optimal linear processing, which is the MMSE receiver [4, eq. (31)], is not sufficient. The insufficiency can be established by using the results in [2, Sec.IV.C] to show that as $P \to \infty$, the effective SNR at the output of MMSE receiver has a limit distribution that is independent of SNR. We skip the details here.

Instead, we notice that the equivalent channel (7) has SNR given by (21), which for $KP > 1$ is greater than $P/3$. So, the MIMO system can be viewed as a Multiple Access Channel (MAC) with K^* single-antenna transmitters, and one receiver with M receive antennas. Perfect CSI is known at the receiver, and the SNR between $P/3$ and P. Using the MAC capacity region result [1, Theorem 14.3.1], [6, Sec. 10.2.1], it can be shown that a total DoF of K^* can be achieved over $T - K^*$ the time slots. $\square$

Remark 1. The DoF is the same as that of a point-to-point MIMO channel with K transmit antennas and M receive antennas without transmit- or receive-side CSI [7]. This is a bit surprising because optimal signaling over non-coherent MIMO channel generally requires cooperation among the transmit antennas. It turns out that as far as DoF is concerned, transmit antenna cooperation is not necessary. This is the new twist compared to the point-to-point case.

Remark 2. It can be seen from the achievability proof that for $M > K$, which is generally applicable for “massive” MIMO systems, zero-forcing at the base station is sufficient for achieving the optimal DoF. However, the MRC is not sufficient because ρ shows up both in the numerator and denominator of (16). So as $\rho \to \infty$, the achieved rate is limited. This is due to the interference among the users.

Remark 3. For the case $K^* = M$, non-linear decoding such as successive interference cancellation is needed.

Remark 4. When T is large, a per-user DoF close to 1 is achievable, as long as $K \leq M$.

Remark 5. When M is larger than K^*, increasing M further has no effect on the DoF. However, it is clear that more receive antennas is useful because more energy is collected by additional antennas. We will discuss the benefit of energy savings in the next section.

5. DISCUSSION

5.1. Power savings for fixed rate

As more antennas are added to the base station, more energy can be collected. Therefore, it is possible that less energy is needed to be transmitted from the mobile stations. When there is perfect CSI at the base station, it has been shown in [4] that the transmission power can be reduced by a factor $1/M$ to maintain the same rate, compared to a single-user single-antenna system.

When there is no CSI at the receiver, however, it was observed in [4] that the power savings factor is $1/\sqrt{M}$ instead of $1/M$. In the following we do a slightly finer analysis of the effected power savings when M is large, assuming the training phase has been optimized as in Section 3.

Consider $M \gg K > 1$. Because the received power is linearly proportional to M, the transmitted power can be smaller when M is larger. When $M \gg 1$, the system is operating in power-limited regime. It can be seen from (16) and (17) that when ρ is small, MRC performs better than ZF, which has been previously observed, e.g., [4]. On the other hand, in the low-SNR regime the difference between them is a constant factor $(M-1)/(M-K)$ in the SNR term within the logarithmic functions in (16) and (17). The difference becomes negligible when M is large. Using either result, and the effective SNR in (14), we see that if we fix the per-user rate at $R = (1 - K/T)\log_2(1+\rho_0)$, then the required power P can be found by setting $\rho M = \rho_0$, resulting in

$$P = \sqrt{\frac{4\rho_0(T-K)}{MT^2}} + o\left(\frac{1}{\sqrt{M}}\right) \qquad (22)$$

It is interesting to note that increasing T has a similar effect as increasing M on the required transmission power, reducing the power by $1/\sqrt{M}$ or $1/\sqrt{T}$. The reason is the if T is increased, then the energy that can be expended on training is increased, improving the quality of channel estimation. On the other hand, for (22) to be applicable, we need $M \gg K$.

5.2. Peak power limited case

If the peak power is limited rather than the average power, then our DoF result still holds because the achievability proof actually uses equal power in the training and data transmission phases. The power savings discussion in the previous subsection still applies, because the system is limited by the total amount of energy available, and not how the energy is expended. In the regime where the SNR is neither very high or very low, the peak power constraint will affect the rate. A detailed analysis is not included here.

5.3. MMSE and optimal processing

If MMSE processing is used at the base station, then the performance can be improved compared to MRC and ZF. However, at low SNR, MRC is near optimal and at high SNR, ZF is near optimal. So MMSE processing will not change the nature of the results that we have obtained, although a slightly higher rate is possible. It is also possible to analyze the achievable rate with optimal non-linear processing, using known MAC capacity region results.

6. REFERENCES

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., 1991.

[2] H. Gao, P. J. Smith, and M. V. Clark, "Theoretical reliability of MMSE linear diversity combining in rayleigh-fading additive interference channels," *IEEE Trans. Commun.*, vol. 46, no. 5, pp. 666–672, May 1998.

[3] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Info. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

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

[5] 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, Jan. 2013.

[6] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005.

[7] L. Zheng and D. Tse, "Communicating on the Grassmann manifold: A geometric approach to the non-coherent multiple antenna channel," *IEEE Trans. Info. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.