OPTIMAL PILOT LENGTH FOR UPLINK MASSIVE MIMO SYSTEMS WITH PILOT REUSE

Qi Zhang[‡], *Shi Jin*[⋆], *David Morales*[§], *Matthew McKay*[§], *and Hongbo Zhu*[‡]

[‡]Jiangsu Key Laboratory of Wireless Communications
Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[⋆]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[§]Department of Electronic and Computer Engineering
Hong Kong University of Science and Technology, Hong Kong

ABSTRACT

We investigate the achievable uplink rate of multi-cell massive multi-input multi-output (MIMO) systems considering pilot reuse across cells. Based on tractable approximations for the achievable rates, we derive the optimal pilot length to maximize the sum rate per cell. Interestingly, it is found that, for moderately large numbers of base station (BS) antennas, the optimal pilot length is above the minimum feasible value, and the gains brought by the extra amount of training diminish as the reuse factor grows. Simulation results confirm the accuracy of our analysis.

***Index Terms*—** Multicell, uplink massive MIMO, pilot reuse, pilot length

1. INTRODUCTION

With increasing demands for higher transmission rates, massive MIMO has attracted much interest and is being investigated as a key technique for 5G [1]. With a very large number of BS antennas, massive MIMO can reap the benefits of conventional MIMO at a greater scale, including improvements in energy and spectral efficiencies, without requiring multiple antennas at the user terminals [2–11]. It has been demonstrated in [2] that when the number of BS antennas grows unlimited, uncorrelated interference and noise can be eliminated. In addition, if the number of BS antennas sufficiently exceeds the number of users, linear receivers become optimal [3].

An important practical issue in massive MIMO is channel estimation, which is typically performed using uplink pilot training. Users employing the same pilot sequence give rise to the *pilot contamination* issue, which persists even with unlimited BS antennas [2]. To diminish this effect, more orthogonal pilot sequences can be used to expand the distance between interference users. This increases the pilot length, following for improved channel estimation, however, it also reduce the time for data transmission. Here, we are interested in determining the optimal pilot length that maximizes the sum rate. A similar problem has been addressed in [12, 13], but pilot reuse is not considered in [12], and [13] does not present optimization results with finite number of antennas. In this paper, accounting for pilot reuse effects, we derive tractable expressions for the achievable uplink rate with maximal ratio combining (MRC) and zero forcing (ZF) receivers. Through these analytical results, the optimal pilot length which maximizes the sum rate per cell is put forward. It is found that a larger pilot reuse factor can improve the rate performance. In the asymptotic massive MIMO regime (i.e., with the infinite number of antennas), the optimal pilot length is shown to be the minimum feasible value which ensures orthogonality within the set of pilots. However, for moderate numbers of BS antennas, the optimal pilot length is above the minimum feasible value, and diminishes towards the minimum feasible value as the pilot length grows. Our analytical results are validated via simulations.

2. SYSTEM MODEL

Consider uplink transmission in a cellular network consisting of L cells, each with one M-antenna BS and N single-antenna users. We assume all users transmit data to their BSs in the same time-frequency resource synchronously. Then, the $M \times 1$ received vector at the ith BS is given by

$$\mathbf{y}_i = \sqrt{p_u}\mathbf{G}_{ii}\mathbf{x}_i + \sqrt{p_u}\sum\nolimits_{\substack{l=1 \\ l\neq i}}^{L} \mathbf{G}_{il}\mathbf{x}_l + \mathbf{n}_i,\ i = 1, \ldots, L, \quad (1)$$

where $\mathbf{G}_{il}$ $(l = 1, \ldots, L)$ is the $M \times N$ MIMO channel matrix between the N users in the lth cell and the M BS antennas in the ith cell, $\sqrt{p_u}\mathbf{x}_l$ denotes the $N \times 1$ vector containing the transmitted signals from all users in the lth cell, while p_u is the average transmitted power of each user, and $\mathbf{n}_i \in \mathbb{C}^{M\times 1}$ represents additive white Gaussian noise (AWGN) with zero mean and unit variance.

The channel transmission coefficient from the nth user in the lth cell to the mth antenna of the ith BS is denoted as g_{minl}, i.e., $[\mathbf{G}_{il}]_{mn} = g_{minl}$, which can be modeled as $g_{minl} = h_{minl}\sqrt{\beta_{inl}}$ [2], where $h_{minl} \sim \mathcal{CN}(0, 1)$ denotes the fast fading element from the nth user in the lth cell to the mth antenna of the ith BS, which is statistically independent across users and cells, and β_{inl} is the large-scale fading coefficient from the nth user in the lth cell to the ith BS. It is assumed constant across the antenna array.

This work was supported in part by the China 973 project under Grant 2013CB329005, the National Natural Science Foundation of China under Grant 6127123 and the China 863 Program under Grant 2014AA01A705. The work of S. Jin was supported by the National Natural Science Foundation of China under Grant 61222102 and the International Science & Technology Cooperation Program of China under Grant 2014DFT10300. The work of D. Morales-Jimenez and M. McKay was supported by the Hong Kong Research Grants Council under grant number 616713.

2.1. Channel Estimation

Channel estimation is performed using uplink pilots. These are of length τ symbols, assumed smaller than the coherence time of the channel. Users in one cell transmit orthogonal pilot sequences, and pilots can be reused among different cells with a reuse factor ρ. Therefore, the L cells are divided into ρ groups, where cells within one group share the same set of pilots, and cells in different groups use orthogonal pilots. If the ith cell belongs to the qth group, we define $f_1(i) = q$, where $f_1(\cdot)$ is an index mapping from cells to groups.

Let the nth user in the ith cell use the $\tau \times 1$ pilot vector $\sqrt{\tau}\boldsymbol{\psi}_{ni}$ ($\tau \geq \rho N$), which satisfies

$$\boldsymbol{\psi}_{ni}^H \boldsymbol{\psi}_{cl} = \begin{cases} 1, & f_1(l) = f_1(i), c = n, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

As a result, the ith BS receives the $M \times \tau$ noisy pilot matrix

$$\mathbf{Y}_i = \sqrt{\tau p_u} \sum_{c=1}^{N} \mathbf{g}_{ici} \boldsymbol{\psi}_{ci}^T + \sqrt{\tau p_u} \sum_{l=1, l\neq i}^{L} \sum_{c=1}^{N} \mathbf{g}_{icl} \boldsymbol{\psi}_{cl}^T + \mathbf{N}_i, \tag{3}$$

where $\mathbf{g}_{icl}$ denotes the cth column of the channel matrix $\mathbf{G}_{il}$, and $\mathbf{N}_i$ represents the $M \times \tau$ AWGN matrix. The BS estimates the channel of the nth user in its own cell by first applying

$$\mathbf{Y}_i \boldsymbol{\psi}_{ni}^* = \sqrt{\tau p_u} \mathbf{g}_{ini} + \sqrt{\tau p_u} \sum\nolimits_{\substack{l=1, l\neq i \\ f_1(l) = f_1(i)}}^{L} \mathbf{g}_{inl} + \mathbf{w}_i, \tag{4}$$

where $\mathbf{w}_i \triangleq \mathbf{N}_i \boldsymbol{\psi}_{ni}^*$. Since $\boldsymbol{\psi}_{ni}^H \boldsymbol{\psi}_{ni} = 1$, the entries of $\mathbf{w}_i$ are $\sim \mathcal{CN}(0, 1)$. Note that the second term in (4) denotes interference from users in other cells that are using the same pilot sequence. From (4), the minimum mean-square-error estimate of the channel vector $\mathbf{g}_{ini}$ admits [14]

$$\hat{\mathbf{g}}_{ini} = \left(\sum\nolimits_{\substack{l=1 \\ f_1(l) = f_1(i)}}^{L} \mathbf{g}_{inl} + \frac{1}{\sqrt{\tau p_u}} \mathbf{w}_i \right) \eta_{ini}, \tag{5}$$

where $\eta_{icl} \triangleq \tau p_u \beta_{icl} / \left(1 + \tau p_u \sum\nolimits_{\substack{l'=1 \\ f_1(l') = f_1(l)}}^{L} \beta_{icl'} \right)$.

2.2. Achievable Uplink Rate

Let $\hat{\mathbf{G}}_{ii} = [\hat{\mathbf{g}}_{i1i}, \ldots, \hat{\mathbf{g}}_{iNi}]$, and $\hat{\mathbf{A}}_{ii}$ be a linear receiver matrix established from $\hat{\mathbf{G}}_{ii}$. Then, with the channel estimation error $\bar{\mathbf{g}}_{ici} \triangleq \hat{\mathbf{g}}_{ici} - \mathbf{g}_{ici}$, the nth element of the detected signal is

$$r_{ni} = \sqrt{p_u} \hat{\mathbf{a}}_{ini}^H \sum_{l=1}^{L} \sum_{c=1}^{N} \mathbf{g}_{icl} x_{cl} - \sqrt{p_u} \hat{\mathbf{a}}_{ini}^H \sum_{c=1}^{N} \bar{\mathbf{g}}_{ici} x_{ci} + \hat{\mathbf{a}}_{ini}^H \mathbf{n}_i, \tag{6}$$

where $\hat{\mathbf{a}}_{ini}$ is the nth column of $\hat{\mathbf{A}}_{ii}$. According to the classical assumption of worst-case uncorrelated Gaussian noise [4], along with the fact that $\texttt{Var}\{\bar{g}_{mici}\} = \beta_{ici}(1 - \eta_{ici})$, and that the variance of noise elements is 1, the achievable uplink rate of the nth user in the ith cell reads

$$R_{ni} = \mathbb{E}\left\{ \log_2 \left(1 + p_u \left| \hat{\mathbf{a}}_{ini}^H \hat{\mathbf{g}}_{ini} \right|^2 / \Omega \right) \right\}, \tag{7}$$

where

$$\Omega \triangleq p_u \sum_{c=1, c\neq n}^{N} \left\{ \left| \hat{\mathbf{a}}_{ini}^H \hat{\mathbf{g}}_{ici} \right|^2 \right\} + p_u \sum_{l=1, l\neq i}^{L} \sum_{c=1}^{N} \left\{ \left| \hat{\mathbf{a}}_{ini}^H \mathbf{g}_{icl} \right|^2 \right\} + \left[p_u \sum_{c=1}^{N} \beta_{ici} \left(1 - \eta_{ici} \right) + 1 \right] \left\{ \| \hat{\mathbf{a}}_{ini} \|^2 \right\}. \tag{8}$$

3. OPTIMAL PILOT LENGTH

Larger τ can increase the accuracy of channel estimation, but it will also squeeze the time left for data transmission. Therefore, we aim to obtain the optimal τ to maximize the uplink sum rate, i.e.,

$$\tau^* = \arg \max_{N\rho \leq \tau \leq T} \frac{T - \tau}{T} \sum_{n=1}^{N} R_{ni}, \tag{9}$$

where T is the channel coherence time in symbols. Note that the subscript i in (9) represents an arbitrary cell, so the optimal τ solved by (9) can be regarded as a general system parameter applied for all cells. Next, we seek solutions to this problem for two different receivers.

3.1. MRC Receiver

For the MRC receiver, $\hat{\mathbf{a}}_{ini} = \hat{\mathbf{g}}_{ini}$. It is difficult to derive exact solutions for the optimization problem due to the expectation in (7). Therefore, we provide a tractable approximation which simplifies the optimization.

Proposition 1 *With MRC receivers, the achievable uplink rate of the nth user in the ith cell can be approximated as*

$$R_{ni}^{\texttt{MRC}} \approx \tilde{R}_{ni}^{\texttt{MRC}} = \log_2 \left(1 + \frac{(M+1) p_u^2 \beta_{ini}^2 \tau}{\Delta_1 \tau + \left(p_u \sum\limits_{l=1}^{L} \sum\limits_{c=1}^{N} \beta_{icl} + 1 \right)} \right), \tag{10}$$

where

$$\Delta_1 \triangleq M p_u^2 \sum\nolimits_{\substack{l=1, l\neq i \\ f_1(l) = f_1(i)}}^{L} \beta_{inl}^2 - p_u^2 \beta_{ini}^2 + p_u^2 \sum\nolimits_{\substack{l=1 \\ f_1(l) = f_1(i)}}^{L} \beta_{inl} \sum_{l=1}^{L} \sum_{c=1}^{N} \beta_{icl} + p_u \sum\nolimits_{\substack{l=1 \\ f_1(l) = f_1(i)}}^{L} \beta_{inl}. \tag{11}$$

Proof: See Appendix A. □

Note that the rate in (10) is conditioned on specific user locations. However, the pilot length we aim to optimize should be independent of this. As such, in the following, for users within cell i, we assume some power control technique is applied to compensate for the large-scale fading, so that $p_u \beta_{ici}$ (for any c) can be set uniformly equal to a user-location independent value, λ_{ii}. For interfering users from cell l ($l \neq i$), we assume a typical user with the average large-scale fading from that cell, which can be regarded as the representation for any user in cell l, so that β_{icl} (for any c) can be approximated by the average large-scale fading, λ_{il}. This is reasonable since the fluctuations in the large-scale fading across different cells are expected to be very small because of the distance between cells. This approximation will be validated with numerical results. Applying these assumptions in (10),

$$\tilde{R}_{ni}^{\texttt{MRC},\lambda} = \log_2 \left(1 + \frac{(M+1) \lambda_{ii}^2 \tau}{\Delta_2 \tau + \left(N \lambda_{ii} + N p_u \sum_{\substack{l=1 \\ l\neq i}}^{L} \lambda_{il} + 1 \right)} \right), \tag{12}$$

where

$$\Delta_2 \triangleq M p_u^2 \sum\nolimits_{\substack{l=1, l\neq i \\ f_1(l) = f_1(i)}}^{L} \lambda_{il}^2 - \lambda_{ii}^2 + \left(\lambda_{ii} + p_u \sum\nolimits_{\substack{l=1, l\neq i \\ f_1(l) = f_1(i)}}^{L} \lambda_{il} \right) \left[N \left(\lambda_{ii} + p_u \sum\nolimits_{\substack{l=1 \\ l\neq i}}^{L} \lambda_{il} \right) + 1 \right]. \tag{13}$$

With (12), the optimization problem (9) becomes

$$\tau^* = \arg \max_{N\rho \le \tau \le T} \mathcal{S}_1(\tau), \tag{14}$$

where $\mathcal{S}_1(\tau) \triangleq N \frac{T-\tau}{T} \tilde{R}_{ni}^{\mathtt{MRC},\lambda}$. In solving this optimization, the following lemma will be important.

Lemma 1 *If a_1, a_2, b_1, b_2 and b_3 are all positive,*

$$s_1(x) = (a_1 - a_2 x) \log_2 \left(1 + \frac{b_1 x}{b_2 x + b_3}\right), \tag{15}$$

is concave on $x \in (0, \infty)$, and has a maximum value at x^* which is the solution of $s_1'(x) = 0$.

Proof: See Appendix B. □

Now, we give the solution of (14).

Proposition 2 *The solution of* (14) *obeys the following:*

- *if $\mathcal{S}'_1(N\rho) > 0$, the optimal τ^* is obtained by Algorithm 1 and rounded to the nearest larger/smaller integer;*
- *if $\mathcal{S}'_1(N\rho) \le 0$, the optimal $\tau^* = N\rho$.*

Proof: With Lemma 1 and $\Delta_2 > 0$, we know $\mathcal{S}_1(\tau)$ is concave. Since $\mathcal{S}'_1(T) < 0$ and $\tau \in [N\rho, T]$, it is deduced that if $\mathcal{S}'_1(N\rho) > 0$, $\mathcal{S}_1(\tau)$ first increases and then decreases. Therefore, there exists an optimal τ^* which satisfies $\mathcal{S}'_1(\tau) = 0$ can maximize $\mathcal{S}_1(\tau)$, and it can be approximately obtained by Algorithm 1. If $\mathcal{S}'_1(N\rho) \le 0$, $\mathcal{S}_1(\tau)$ is a monotonically decreasing function. Therefore, the optimal τ^* is $N\rho$. □

Algorithm 1 Solving equation $\mathcal{S}'_1(\tau) = 0$

Input: $\mathcal{S}_1(\tau)$, initial interval $[a, b]$, and maximal error ϵ
Output: τ^* or Failure due to bad initial interval

```
F(τ) ← S'_1(τ)
if F(a) · F(b) ≤ 0 then
    while (b − a) > ϵ do
        x = (a + b)/2
        if F(x) · F(a) < 0 then
            b ← x
        else
            a ← x
        end if
    end while
    τ* ← x
else
    return 'Failure (bad initial interval)'
end if
```

Corollary 1 *When $M \to \infty$,* (14) *is solved as $\tau^* = N\rho$.*

Note from *Corollary 1* that with infinite M, it is optimal to use the smallest feasible pilot length. This is because when $M \to \infty$, all the uncorrelated interference vanishes. The residual interference (pilot contamination) cannot be removed by increasing the training period, as it is an interference floor inherent to pilot reuse in the multi-cell system. Therefore, increasing the pilot length can only squeeze the data transmission time and impair the capacity performance.

3.2. ZF Receiver

For ZF receivers, $\hat{\mathbf{a}}_{ini}\hat{\mathbf{g}}_{ici} = \delta[n{-}c]$, with $\delta[\cdot]$ the Kronecker delta function. As before, upon substituting into (7), it is difficult to derive exact solutions for the optimization problem (9). Therefore, a tractable lower bound of the achievable uplink rate with ZF receivers is given first.

Proposition 3 *With ZF receiver, the achievable uplink rate of the nth user in the ith cell can be lower bounded as*

$$R_{ni}^{\mathtt{ZF}} \ge \tilde{R}_{ni}^{\mathtt{ZF}} = \log_2 \left(1 + \frac{(M-N) p_u^2 \beta_{ini}^2 \tau}{\Delta_3 \tau + \left(p_u \sum_{l=1}^{L} \sum_{c=1}^{N} \beta_{icl} + 1\right)}\right), \tag{16}$$

where

$$\Delta_3 \triangleq (M-N-1) p_u^2 \sum\nolimits_{\substack{l=1, l\neq i \\ f_1(l)=f_1(i)}}^{L} \beta_{inl}^2 - p_u^2 \sum_{c=1}^{N} \beta_{ici}^2 + p_u^2 \sum\nolimits_{\substack{l=1 \\ f_1(l)=f_1(i)}}^{L} \beta_{inl} \sum_{l=1}^{L} \sum_{c=1}^{N} \beta_{icl} + p_u \sum\nolimits_{\substack{l=1 \\ f_1(l)=f_1(i)}}^{L} \beta_{inl}. \tag{17}$$

Proof: See Appendix C. □

Similar to before, to get the optimal pilot length, we remove the dependence on the large-scale fading in (16). Under the same assumptions made in (12), we have now

$$\tilde{R}_{ni}^{\mathtt{ZF},\lambda} = \log_2 \left(1 + \frac{(M-N) \lambda_{ii}^2 \tau}{\Delta_4 \tau + \left(N p_u \sum_{l=1}^{L} \lambda_{il} + 1\right)}\right), \tag{18}$$

where

$$\Delta_4 \triangleq (M-N-1) p_u^2 \sum\nolimits_{\substack{l=1, l\neq i \\ f_1(l)=f_1(i)}}^{L} \lambda_{il}^2 - N\lambda_{ii}^2 + \left(\lambda_{ii} + p_u \sum\nolimits_{\substack{l=1, l\neq i \\ f_1(l)=f_1(i)}}^{L} \lambda_{il}\right) \left[N\left(\lambda_{ii} + p_u \sum\nolimits_{\substack{l=1 \\ l\neq i}}^{L} \lambda_{il}\right) + 1\right]. \tag{19}$$

With (18), the optimization problem (9) becomes

$$\tau^* = \arg \max_{N\rho \le \tau \le T} \mathcal{S}_2(\tau), \tag{20}$$

where $\mathcal{S}_2(\tau) \triangleq N \frac{T-\tau}{T} \tilde{R}_{ni}^{\mathtt{ZF},\lambda}$.

Proposition 4 *The solution of* (20) *obeys the following:*

- *if $\mathcal{S}'_2(N\rho) > 0$, the optimal τ^* is obtained by Algorithm 1 and rounded to the nearest larger/smaller integer;*
- *if $\mathcal{S}'_2(N\rho) \le 0$, the optimal $\tau^* = N\rho$.*

Proof: Follows similar steps as Proposition 2. □

Corollary 2 *When $M \to \infty$,* (20) *is solved as $\tau^* = N\rho$.*

4. NUMERICAL RESULTS

The cell radius considered in our system is $r_c = 1000$ meters and N users are distributed randomly and uniformly in each cell, with the exclusion of a central disk of radius $r_h = 100$ meters around each BS. We determine the set of cells within $8r_c$. We assume that $\beta_{icl} = z/(r/r_h)^v$, where z is a log-normal random variable with standard deviation $\sigma = 8$dB, r is the distance from the cth user in the lth cell to the ith BS, and the path loss exponent $v = 3.8$. The coherence time of the channel is set to $T = 196$.

In Fig. 1, we compare the simulated uplink sum rate in (7) with the analytical expressions (12) and (18), where λ_{il} is set to the average of β_{icl} over 10^4 realizations. Note that λ_{ii} in (12) and (18) is assumed uniform across users due to power control. Different pilot reuse factors are considered, and the pilot length τ is fixed. We can see a close agreement between the simulation and analytical results, especially for

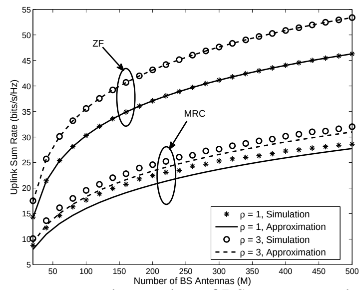

Fig. 1: Uplink sum rate vs. the number of BS antennas, where $p_u = 10$dB, $N = 10$ and $\tau = 30$.

the ZF receiver. It is seen that a larger ρ improves the uplink performance. This is because with fixed N and τ, a larger pilot reuse weakens the interference from other cells without reducing the number of users or the data transmission time. Given the tightness of our analytical results, we will use them to explore the optimal pilot length in the subsequent results.

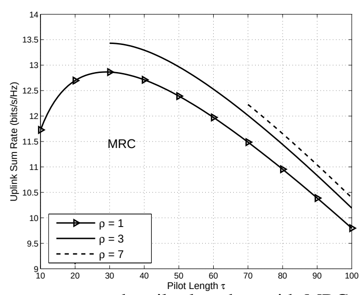

Fig. 2: Uplink sum rate vs. the pilot length τ with MRC receivers, where $M = 100$, $N = 10$ and $p_u = 10$dB.

In Fig. 2 and 3, we analyze the change of uplink sum rate for MRC and ZF receivers, respectively, as the pilot length τ increases. Here, N is fixed, and we are interested in a moderately large number of BS antennas, in which case increasing τ brings some gain in terms of channel estimation. Due to the constraint that $\tau \geq N\rho$, curves with different ρ have a different starting point. It can be found that when $\rho = 1$, the sum rate first increases and then decreases. However, for larger ρ, the sum rate almost monotonically decreases. This can be explained by noting that with fixed N, the benefit of increasing τ is that it provides higher channel estimation accuracy. But for larger ρ, τ starts at larger values which already give a very good channel estimate, and therefore the benefit brought by a slightly better estimation is negligible as compared with the loss in transmission time.

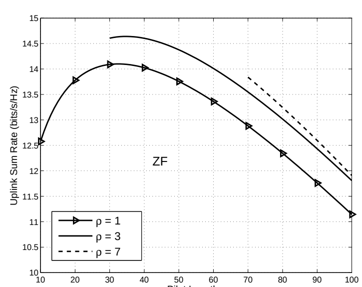

Fig. 3: Uplink sum rate vs. the pilot length τ with ZF receivers, where $M = 100$, $N = 10$ and $p_u = 10$dB.

5. CONCLUSION

In this paper, we studied the achievable uplink rate of multi-cell massive MIMO systems with MRC and ZF receivers. Allowing different cells to use orthogonal pilots, we derived tractable expressions for the achievable rates. Based on these results, we obtained the optimal pilot length to maximize the sum rate per cell. It was found that larger pilot reuse factor can improve the rate performance, and as it grew, the optimal pilot length turned closer to the minimum feasible value.

A. PROOF OF PROPOSITION 1

With some basic algebraic operations, we have $\mathbb{E}\left\{\|\hat{\mathbf{g}}_{ini}\|^2\right\} = M\beta_{ini}\eta_{ini}$, and

$$\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{ini}^H\hat{\mathbf{g}}_{ici}\right|^2\right\} = \begin{cases} (M^2+M)\beta_{ini}^2\eta_{ini}^2, & c = n, \\ M\beta_{ini}\beta_{ici}\eta_{ini}\eta_{ici}, & c \neq n, \end{cases} \tag{21}$$

as well as

$$\sum_{l=1,l\neq i}^{L}\sum_{c=1}^{N}\mathbb{E}\left\{\left|\hat{\mathbf{g}}_{ini}^H\mathbf{g}_{icl}\right|^2\right\} =$$

$$M^2\eta_{ini}^2\sum\nolimits_{\substack{l=1,l\neq i\\ f_1(l)=f_1(i)}}^{L}\beta_{inl}^2 + M\beta_{ini}\eta_{ini}\sum_{l=1,l\neq i}^{L}\sum_{c=1}^{N}\beta_{icl}. \tag{22}$$

Then, using $\hat{\mathbf{a}}_{ini} = \hat{\mathbf{g}}_{ini}$ and [15, Lemma 1] for (7), we can obtain the desired result.

B. PROOF OF LEMMA 1

Through some basic algebra, we know that $s_1''(x)$, the second derivative of $s_1(x)$, is less than 0. Therefore, $s_1(x)$ is concave, and $s_1'(x)$ is monotonically decreasing with x. Moreover, we have $s_1'(0) = a_1b_1/\ln 2b_3 + c_1/\ln 2c_3 > 0$, and when $x \to \infty$, $s_1'(x) \to -a_2\ln\left(1+\frac{b_1}{b_2}\right) < 0$. Hence, we can deduce that with $x \in (0,\infty)$, $s_1(x)$ first increases and then decreases. Therefore, there exists a x_0 which satisfies $s_1'(x_0) = 0$ and maximizes $s_1(x)$.

C. PROOF OF PROPOSITION 2

For ZF receivers, we know that if $f_1(l) \neq f_1(i)$, or $f_1(l) = f_1(i)$ and $c \neq n$, $\hat{\mathbf{a}}_{ini}$ is independent of $\hat{\mathbf{g}}_{icl}$. Then,

$$\mathbb{E}\left\{\left|\hat{\mathbf{a}}_{ini}^H\mathbf{g}_{icl}\right|^2\right\} = \beta_{icl}\mathbb{E}\left\{\|\hat{\mathbf{a}}_{ini}\|^2\right\}. \tag{23}$$

If $f_1(l) = f_1(i)$ and $c = n$, we have

$$\mathbb{E}\left\{\left|\hat{\mathbf{a}}_{ini}^H\mathbf{g}_{icl}\right|^2\right\} = \mathbb{E}\left\{\left|\hat{\mathbf{a}}_{ini}^H\left(\hat{\mathbf{g}}_{inl}-\bar{\mathbf{g}}_{inl}\right)\right|^2\right\}. \tag{24}$$

From (4), we have that $\hat{\mathbf{g}}_{inl} = \frac{\beta_{inl}}{\beta_{ini}}\hat{\mathbf{g}}_{ini}$, and $\bar{g}_{minl} \sim \mathcal{CN}\left(0,\beta_{inl}(1-\eta_{inl})\right)$. According to the property of MMSE estimation [14], $\bar{\mathbf{g}}_{inl}$ is independent of $\hat{\mathbf{g}}_{inl}$. Then, (24) becomes

$$\mathbb{E}\left\{|\hat{\mathbf{a}}_{ini}\mathbf{g}_{icl}|^2\right\} = \mathbb{E}\left\{\left|\frac{\beta_{inl}}{\beta_{ini}} - \hat{\mathbf{a}}_{ini}^H\bar{\mathbf{g}}_{inl}\right|^2\right\}$$

$$= \frac{\beta_{inl}^2}{\beta_{ini}^2} + \beta_{inl}(1-\eta_{inl})\mathbb{E}\left\{\|\hat{\mathbf{a}}_{ini}\|^2\right\}. \tag{25}$$

Let $z_{ini} \triangleq 1/\|\hat{\mathbf{a}}_{ini}\|^2$, which is a Chi-squared random variable [16]. Then, $\mathbb{E}\left\{\|\mathbf{a}_{ini}\|^2\right\} = 1/\beta_{ini}\eta_{ini}(M-N)$. From Jensen's inequality, $\mathbb{E}\left\{\log_2(1+1/x)\right\} \geq \log_2(1+1/\mathbb{E}\{x\})$, we can obtain the desired result.

6. REFERENCES

[1] 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 Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

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

[3] 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, Feb. 2013.

[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] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "On the optimality of single-carrier transmission in large-scale antenna systems," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 276–279, Aug. 2012.

[6] S. Wagner, R. Couillet, D. T. M. Slock, and M. Debbah, "Large system analysis of zero-forcing precoding in MISO broadcast channels with limited feedback," in *Proc. IEEE Int. Work. Signal Process. Adv. Wireless Commun. (SPAWC)*, June 2010.

[7] H. Q. Ngo, E. G. Larsson, and T. L. 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, June 2013.

[8] 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, Jan. 2013.

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

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

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

[12] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Commun. Lett.*, vol. 2, no. 6, pp. 605–608, Sept. 2014.

[13] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, accepted to appear.

[14] S. M. Kay, *Fundamentals of Statistical Signal Srocessing: Estimation Theory (Volume 1)*. Prentice Hall, 1993.

[15] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary-rank channel means," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 966–981, Oct. 2014.

[16] D. Gore, R. W. Heath Jr., and A. Paulraj, "Transmit selection in spatial multiplexing systems," *IEEE Commun. Lett.*, vol. 6, no. 11, pp. 491–493, Nov. 2002.