Uplink Linear Receivers for Multi-Cell Multiuser MIMO With Pilot Contamination: Large System Analysis

Narayanan Krishnan, *Member, IEEE*, Roy D. Yates, *Fellow, IEEE*, and Narayan B. Mandayam, *Fellow, IEEE*

Abstract—Base stations with a large number of transmit antennas can potentially serve a large number of users at high rates. However, the receiver processing in the uplink relies on channel estimates, which are known to suffer from pilot interference. In this paper, making use of the similarity of the uplink received signal in CDMA with that of a multi-cell multi-antenna system, we perform a large system analysis when the receiver employs an MMSE filter with a pilot contaminated estimate. We assume a Rayleigh fading channel with different received powers from users. We find the asymptotic signal to interference plus noise ratio (SINR) as the number of antennas and number of users per base station grow larger while maintaining a fixed ratio. Through the SINR expression we explore the scenario where the number of users being served are comparable to the number of antennas at the base station. The SINR explicitly captures the effect of pilot contamination and is found to be the same as that employing a matched filter with a pilot contaminated estimate. We also find the exact expression for the interference suppression obtained using an MMSE filter, which is an important factor when there are a significant number of users in the system as compared to the number of antennas. In a typical set up, in terms of the five percentile SINR, the MMSE filter is shown to provide significant gains over matched filtering and is within 5 dB of MMSE filter with perfect channel estimate. Simulation results for achievable rates are close to large system limits for even a 10-antenna base station with 3 or more users per cell.

Index Terms—Massive MIMO, MU-MIMO, large scale antenna systems (LSAS), pilot contamination, channel estimation, land mobile radio cellular systems, OFDM, multi-user detection, large system analysis.

I. INTRODUCTION

CELLULAR systems with large number of base station antennas have been found to be advantageous in mitigating the fading effects of the channel [3] while increasing system capacity. It is shown in [3] that in an infinite antenna regime, and in a bandwidth of 20 MHz, a time division duplexing

system has the potential to serve 40 single antenna users with an average throughput of 17 Mbps per user. However, any advantages offered by multiple antennas at the base station can be utilized only by gaining the channel knowledge between the base station and all the users. This requires training data to be sent from the users. In a typical system the time-frequency resources are divided into Physical Resource Blocks (PRBs) of coherence-time coherence-bandwidth product. For each user, it is necessary and sufficient to estimate the channel in every PRB. Thus, some resources (time slots or equivalently frequency channels) are used for channel estimation and the rest are used for transmission in uplink or downlink. However, in [2], it has been shown that the number of pilot symbols required is proportional to the total number of users in the system. Hence, as the system scales with the number of users, the dedicated training symbols may take up a significant portion of the PRB. As this is undesirable, only a part of the coherence time is utilized to learn the channel. In this case, the pilot sequences in different cells overlap over time-frequency resources and, as a consequence, the channel estimates are corrupted. This *pilot interference* is found to be a limiting factor as we increase the number of antennas [5].

It is shown in [3] that in the limit of large number of antennas, the SINR using a matched filter receiver is limited by interference power due to pilot contamination. While the result assumes a regime with finite number of users, we can also envision a regime where the number of users may be comparable to the number of antennas such as a system with 50 antenna base stations serving 50 users simultaneously. In this paper, we do a large system analysis of uplink multi-cell, multi-antenna system when the receiver employs an MMSE filter to decode the received signal. We investigate the SINR in a regime where the number of users per cell is comparable to the number of antennas at the base station. The MMSE filter which is designed to maximize the SINR is evaluated when we have a pilot corrupted channel estimate. We let the number of antennas and the number of users per base station grow large simultaneously while maintaining a fixed users/antennas ratio α and observe the SINR for the above two cases as a function of α. To do so we make use of the similarity of the uplink received signal in a MIMO system to that of the received signal in a CDMA system [11].

Much of the research in large MIMO systems with Rayleigh channel can be borrowed from the considerable literature for CDMA systems. The channel vector with i.i.d entries for the

large MIMO system is analogous to the signature sequence in a CDMA system so that antennas contribute to the processing gain. For example, the uplink analysis of an asymptotic regime [11] with both users and signature sequences tending to infinity translates directly to results in a large MIMO system when signatures are replaced by antennas. In both systems it is observed that the asymptotic analysis is a good approximation for practical number of antennas (signatures) and users. While in a CDMA system we assume that the signature sequences are known, there are practical limitations in learning a mobile radio multi-antenna channel (antenna signatures) in a multi-cellular system, as shown in [3]. In this paper, we explore this limitation when users simultaneously estimate the channel and the estimates are subject to pilot contamination. We focus our results in the regime $\alpha > 0.1$ as opposed to recent works such as [3], [8], [14], [16] which are found to be approximately in the regime of $0 \leq \alpha < 0.1$. Further, we compare the results of the asymptotic SINR expression so obtained with that of the performance of the matched filter.

A. Related Work

A similar large system analysis in the context of a Network-MIMO architecture was presented in [20]. The authors concluded that high spectral efficiencies can be realized even with 50 antennas in their architecture, paralleling the existing literature results in CDMA systems. In [16], the results obtained suggest to scale the transmission power by the square root of the number of base station antennas, as opposed to scaling by the number of antennas. This assumes that the transmission power during training and data are same. In general we take the approach in [1] where the transmission power can be different for the training and data symbols during a coherence time. Joint channel estimation and multi-user detection was considered in [25] for a single cell multi-user MIMO DS-CDMA systems. For fixed number of transmit antennas per user and receive antennas per base station, the authors employ replica method for CDMA large system analysis to obtain lower bounds on achievable rate for different feedback based receiver strategies. Mathematically, our model can be viewed as an extension of the linear receiver scheme considered in their model to a multi-cell scenario.

In the MU-MIMO literature the asymptotic SINR is also called the "deterministic equivalent" of the SINR. Recent work in [10] finds the deterministic equivalent of SINR with distributed sets of correlated antennas in the uplink. Authors in [14], have done a considerable work in providing the deterministic equivalent for beamforming/maximum ratio combining and regularized zero forcing/MMSE in the downlink/uplink with a generalized channel model taking into account the effect of pilot contaminated channel estimate. They find the number of antennas required to match a fixed percentage of the rate of an infinite antenna regime. Also, the number of extra of antennas required for the matched filter to equal the rate obtained out of the MMSE filter is shown, implicitly showing the interference suppression capability of MMSE filter. We derive in our work the exact amount by which the MMSE filter suppresses the interference for a Rayleigh fading channel and provide

some fresh engineering insights in the regime with $\alpha > 0.1$. Additionally, we derive those results under a stochastic rather than a deterministic received power model as presented in [14, Sec. 4]. A summary of the results is given in Section I-B.

There have been significant attempts to mitigate pilot contamination in the recent works in "Massive MIMO" systems which are, however, specific to the case when antennas far exceed the number of users served i.e., $\alpha < 0.1$. In [8], time shifted pilot schemes were introduced to reduce pilot contamination. There, simultaneous transmission of pilots was avoided by scheduling only a subset of base stations to transmit uplink pilots. Simultaneously, other base stations transmit in the downlink to their users and it is shown that the interference created by these downlink transmission can be cancelled with a large number of antennas at the base station estimating the channel. Pilot contamination is now restricted to base stations in a group that simultaneously transmit uplink pilots. However, this requires that the number of antennas far exceed the number of users. In their recent work, authors in [18], [19] show that pilot contamination can be avoided using subspace based channel estimation techniques. They show that the eigenvalues corresponding to the other-cell interference subspace can be separated from the in-cell users in a regime where α is below a threshold. Their analyses assume an ideal power controlled situation with strict user scheduling and antennas far exceeding the number of users. By contrast, we examine the operating regime in which $0.1 < \alpha$ and determine the effect of pilot contamination on interference and interference suppression capability of MMSE receiver. Also, even with power control, pilot contamination is prevalent when linear MMSE channel estimation is employed.

B. Contributions of Our Work

We develop a large system asymptotic expression for the SINR in a Rayleigh fading environment when using an matched filter and MMSE filter with a pilot contaminated channel estimate for a arbitrary user in the system. The SINR expression is dependent on the number of users and the number of antennas only through their ratio α. If $\tilde{P}_{\text{signal}}$, $\tilde{P}_{\text{contam}}$, $\tilde{P}_{\text{inter}}(\mathbf{c})$ and σ^2 represent the signal power, interference power due to pilot contaminated estimate, the filter dependent interference power for a linear filter $\mathbf{c}$ and the noise variance, respectively, we show that the expression for asymptotic SINR for both matched filter and MMSE filter can be generalized to

$$\text{SINR}(\mathbf{c}) = \frac{\tilde{P}_{\text{signal}}}{\sigma^2 + \tilde{P}_{\text{contam}} + \alpha \tilde{P}_{\text{inter}}(\mathbf{c})}. \quad (1)$$

If $\hat{\mathbf{c}}^{\text{MF}}$ and $\hat{\mathbf{c}}^{\text{MMSE}}$ denote the matched filter and the MMSE filter with a pilot contaminated estimate, then the following result summarizes our contribution:

- From the expression for SINR it is derived that the total interference is the sum of two terms. The first term given by $\tilde{P}_{\text{contam}}$ is due to employing pilot contaminated channel estimate and the second is the filter dependent interference term $\alpha \tilde{P}_{\text{inter}}(\mathbf{c})$ due to comparable K and M. Also, the the the

filter dependent term contributes to interference only when $\alpha \neq 0$.

- We show that $\tilde{P}_{inter}(\hat{\mathbf{c}}^{MF}) - \tilde{P}_{inter}(\hat{\mathbf{c}}^{MMSE}) = \mathcal{C}(\alpha) \geq 0$, where $\mathcal{C}(\alpha)$ is called the interference suppression term. We find a closed form expression for $\mathcal{C}(\alpha)$ showing the interference suppression capability of the MMSE filter when $\alpha > 0$.
- It is derived that $\tilde{P}_{contam}(\hat{\mathbf{c}}^{MF}) = \tilde{P}_{contam}(\hat{\mathbf{c}}^{MMSE}) = \tilde{P}_{contam}$. The contribution of the pilot contaminated channel estimate to the pilot interference given by $\tilde{P}_{contam}$ is same for both matched filter and MMSE filter.
- Although the filter $\hat{\mathbf{c}}^{MMSE}$ depends on the pilot contaminated channel estimate of all the K users in a cell to its base station, the $\tilde{P}_{contam}$ is independent of K or M or α for all values of $\alpha > 0$ and therefore is the same as what was found at $\alpha = 0$.

As per the authors' knowledge, these contributions have not been reported yet in the literature.

We validate the theoretical results with simulations. We show that even a system with 50 base station antennas each serving some number of users sufficiently qualifies for the term large system as the users' SINRs are close to the asymptotic limit. Simulation results for achievable rates are close to theory for even a 10-antenna base station with 3 or more users per cell. The following summarizes the key contributions through simulation:

- The theoretical results are derived assuming that the same set of in-cell orthogonal training signals are repeated across the cells. However, we also show through simulations that in the case of independently generated but non-orthogonal training signals(with orthogonal in-cell training), the resulting SINR performance is close to the asymptotic limit.
- In an example seven cell set up, the MMSE filter performs the best in the absence of pilot contamination. We also show an intermediate regime where the MMSE filter with pilot contamination obtains around 7 dB gain over the matched filter with a pilot contaminated estimate.
- In terms of the five percentile SINR, the MMSE receiver is shown to provide significant gains over matched filtering. Also, in most of the operating points α, the performance of the MMSE receiver with pilot estimate is within 5 dB of the MMSE filter with perfect estimate.
- We also show that the achievable rates are within a 1 bit/symbol of the MMSE filter with perfect estimate when the number of users are comparable to the number of antennas.

II. SYSTEM MODEL

We consider a system similar to that in [3] with B non-cooperating base stations and K users per base station. We assume that all KB users in the system are allocated the same time-frequency resource by a scheduler. Also, each base station is equipped with M antennas. The channel vector representing the small scale fading between user k in cell j and the an-tennas in base station l is given by a $M \times 1$ vector $\mathbf{h}_{jk}^{(l)}$. The entries of $\mathbf{h}_{jk}^{(l)}$ are assumed to be independent zero mean i.i.d Gaussian random variables with variance $1/M$ corresponding to the scaling of transmit power by the number of receiver antennas at the base station. This corresponds to an ideal and favorable propagation medium with rich scattering. A large scale fading coefficient, which represents the power attenuation due to distance and effects of shadowing between base station l and kth user in jth cell is given by $\beta_{jk}^{(l)}$. We assume that $\beta_{jk}^{(l)} < 1$ as we do not expect the received power to be greater than what is transmitted. This is constant across the antennas of the cell l. Accordingly, overall channel vector is given by $\sqrt{\beta_{jk}^{(l)}}\mathbf{h}_{jk}^{(l)}$.

A. Uplink Transmission

We assume that all users' transmission are perfectly synchronized. Also, while a user's transmission is intended for its own base station, other base stations also hear the transmission. Defining q_{jk} as the symbol transmitted by user k in cell j, $\mathbf{w}^{(l)}$ as the $M \times 1$ noise vector with zero mean circularly symmetric Gaussian entries such that $\mathbb{E}[\mathbf{w}^{(l)}\mathbf{w}^{(l)H}] = \sigma^2\mathbf{I}$, the received signal at base station l is given by,

$$\mathbf{y}^{(l)} = \sum_{j=1}^{B}\sum_{k=1}^{K}\sqrt{\beta_{jk}^{(l)}}\mathbf{h}_{jk}^{(l)}q_{jk} + \mathbf{w}^{(l)}. \qquad (2)$$

Here, the signal-to-noise ratio (SNR) of the received signal per receive dimension is given by $1/\sigma^2 M$. To utilize the advantages offered by multiple antennas, the base station has to have an estimate of the channel to all users prior to detection of uplink signals. In a system employing an OFDM physical layer with time-frequency resources, we can divide the resources into physical resource blocks (PRBs) contained in the coherence-time coherence-bandwidth product. Although the channel vector $\mathbf{h}_{jk}^{(l)}$ of each user has to be relearned by the base station at the start of PRB, once learnt for a subcarrier it remains the same for all subcarriers within that PRB. Let the number of coherent symbols be given by T_c and coherent subcarriers be N_c. We define a Resource Element(RE) to be a subcarrier at a symbol time. Therefore, if we fix the number of resource elements used for estimation to be T such that $T \leq T_cN_c$, a total of T user's channel can be learnt. This observation was noted in [3]. We would like to point out that it is relevant here as the number of users that can be supported depends on total available coherent resource elements N_cT_c. Depending on its value the number of users per cell K that could be supported can be comparable to M. We define the load on the system as $\alpha = K/M$ throughout this paper. For example, in a single cell set up even in very conservative scenarios of short coherence time with $T_c = 7$ symbols and frequency selective channel with $N_c = 14$, around 49 users per cell can be supported at $\alpha = 0.5$ if half the resources are used for channel estimation. This is also illustrated in Fig. 1. Therefore, it is worthwhile to investigate not only the $M \gg K$ scenario but also the case when M and K large and comparable.
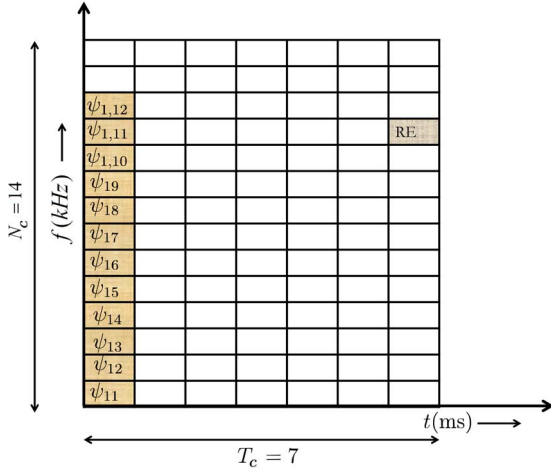
Fig. 1. PRB is composed of 98 orthogonal Resource Elements (RE) and 12 RE for training. The first user in the first base station transmits pilots ψ_{11} in the first RE and remain silent for the rest of training duration. Similarly, other users in the first cell transmit pilots during the RE allocated for its training. If orthogonal pilots are allocated for all the users in the system, a typical 7 cell system 14 users per cell would consume all the coherent RE without any time for data transmission. In our model we let the training symbols of other cell users ψ_{jk} for all j and k simultaneously during ψ_{1k} leading to pilot contamination problem.

B. Limitations in Gaining Channel Knowledge

During each coherence time, users in a cell spend some pilot symbol times in each PRB for channel estimation at the base station and then data transmission ensues until the end of the block. At base station l, the number of channel vectors $\mathbf{h}_{jk}^{(l)}$ that needs to be learnt is equal to the number of users in the system which is KB. To accomplish that, the number of pilots required must at least be KB symbol times for the pilot sequences to be orthogonal across the users in the system. However, such a system will not be scalable as there exists some large B for which the product KB will occupy all the coherent resource elements. In the example illustrated in Fig. 1 a 7 cell system with 14 users per cell would end up using all the coherent resource elements if orthogonal channel training is provided for all the users in the system. This is clearly undesirable as pilot training is consuming a significant part of PRB.

In one of the approaches taken in [3], the base station is concerned with only knowing the channel to its own K users and spends only K resource elements for channel estimation instead of KB. Every base station similarly spends K resource elements for channel estimation for its K users. The pilot signals are processed and an MMSE based channel estimate of the channel is formed. MMSE channel estimation is the commonly employed in multiuser MIMO systems [1], [4], [15]. Let $\Psi_{jk} \in \mathbb{C}^{K \times 1}$ denote the training sequence of user k in cell j of duration K symbols. Also, assume that the in-cell training sequences are orthogonal i.e., $\Psi_{jk}^H \Psi_{jn} = 0$ if $k \neq n$ and 1 otherwise. We assume that the training sequences across the cells are independently generated and hence in general $\Psi_{jk}^H \Psi_{in} \neq 0$ if $i \neq j$ and for all k and n. We assume a constant average transmit SNR per symbol during a coherence time. If ρ_d is the transmit SNR of data symbol and ρ_{avg} the average

transmit SNR per symbol then with K resource elements for channel estimation we have

$$\rho_{\text{avg}} = \frac{\rho_p K + \rho_d (T_c N_c - K)}{T_c N_c}. \quad (3)$$

If the transmit power of data symbols is scaled by the number of antennas then $\rho_d = 1/\sigma^2 M$. Therefore, $\rho_d \to 0$ for large M and $\rho_p \approx \rho_{\text{avg}} T_c N_c / K$. Therefore, in the constant average transmit SNR per symbol model the pilot transmission is not scaled by M which is the number of antennas at the base station. With $\mathbf{N} \in \mathbb{C}^{M \times K}$ denoting the additive complex Gaussian noise matrix with i.i.d entries with variance $1/M$, and ρ_p denoting the SNR during pilot transmission, the received signal at base station l across the K training resources is given by

$$\mathbf{Y}^{(l)} = \sum_{j=1}^{B} \sum_{k=1}^{K} \sqrt{\beta_{jk}^{(l)}} \mathbf{h}_{jk}^{(l)} \Psi_{jk}^H + \frac{\mathbf{N}^{(l)}}{\sqrt{\rho_p}}. \quad (4)$$

Since the transmission power of pilot symbols are not scaled by the number of antennas we assume that noise variance $1/M$ for the entries of the $\mathbf{N}^{(l)}$. This is required because we have defined the entries of $\mathbf{h}_{jk}^{(l)}$ to be complex i.i.d with variance $1/M$.

Without loss of generality we assume that the receiving base station is indexed $l = 1$ throughout this paper. Consequently, to simplify the exposition we drop the superscript $(.)^{(l)}$ from terms in (4) and denote $\mathbf{Y}^{(1)} \triangleq \mathbf{Y}$, $\mathbf{N}^{(1)} \triangleq \mathbf{N}$, $\beta_{jk}^{(1)} \triangleq \beta_{jk}$, $\mathbf{h}_{jk}^{(1)} \triangleq \mathbf{h}_{jk}$. Hence, β_{jk} and $\mathbf{h}_{jk}$ represent the large scale fading and the small scale channel vector between the user k in cell j to cell 1. Also, in (2), if $l = 1$ then $\mathbf{y}^{(1)} \triangleq \mathbf{y}$ and $\mathbf{w}^{(1)} \triangleq \mathbf{w}$. The MMSE channel estimate for user k in the first cell is then given by

$$\hat{\mathbf{h}}_{1k} = \mathbf{Y} \left(\frac{\mathbf{I}}{\rho_p} + \sum_{j=1}^{B} \sum_{k=1}^{K} \beta_{jk} \Psi_{jk} \Psi_{jk}^H \right)^{-1} \Psi_{1k} \sqrt{\beta_{1k}}. \quad (5)$$

Here, $\hat{\mathbf{h}}_{1k} (\triangleq \hat{\mathbf{h}}_{1k}^{(1)})$ is the channel estimate of the user k in cell 1 to the base station 1. Although it is not common in practice, we assume as in [3]–[5] that the in-cell pilots are repeated across the cells, to get some analytic insight; this implies that $\Psi_{jk} = \Psi_{ik}$ for all k. The MMSE channel estimate [23] with pilot contamination when the in-cell orthogonal pilots [4] are repeated across the cells is given by

$$\hat{\mathbf{h}}_{1k} = \frac{\sqrt{\beta_{1k}}}{\beta^{(k)} + \frac{1}{\rho_p}} \left(\sum_{j=1}^{B} \sqrt{\beta_{jk}} \mathbf{h}_{jk} + \frac{\mathbf{N} \Psi_{1k}}{\sqrt{\rho_p}} \right) \quad (6)$$

where, $\beta^{(k)} = \sum_{j=1}^{B} \beta_{jk}$ and if $\mathbf{h}_{1k} = \hat{\mathbf{h}}_{1k} + \tilde{\mathbf{h}}_{1k}$, then $\tilde{\mathbf{h}}_{1k}$ is zero mean with covariance

$$\mathbb{E}\left[\tilde{\mathbf{h}}_{1k} \tilde{\mathbf{h}}_{1k}^H \right] = \frac{1}{M} \left(\frac{\sum_{j=2}^{B} \beta_{jk} + 1/\rho_p}{\beta^{(k)} + 1/\rho_p} \right) \mathbf{I}.$$

This estimate is used to design linear detectors to filter the received signal. Later we show that even with actual training given by (5) the SINRs are very close to when in-cell pilots are reused across the cells. In a power controlled system with $\beta_{1k} = 1$, for all $k = 1, \ldots K$, if the target average SNR is $\rho_{\text{avg}} = 20$ dB, then in the example in Fig. 1, the SNR of pilot transmission is given by $\rho_p \approx \rho_{\text{avg}} T_c = 28$ dB. As the above example shows, even with a very conservative coherence time T_c, the SNR of pilot transmission is high enough such that $\sum_{j=2}^{B} \beta_{jk} \gg 1/\rho_p$ unless all the other cell received powers of interferers contributing to pilot contamination are approximately 28 dB below the in-cell user. As we will see in simulations, having other-cell received powers 30 dB below in-cell corresponds to a situation in which pilot contamination is insignificant. In general since $\rho_p \approx \rho_{\text{avg}} T_c N_c / K$, we can always design a system based on fixing a percentage of resource elements in a PRB for training such that noise in the channel estimate is not the significant contributor. Therefore, assuming we have high enough pilot power, we ignore the additive noise affecting the channel estimation to focus our results on the pilot contamination problem. Therefore, with $\rho_p \to \infty$ the channel estimate is given by

$$\hat{\mathbf{h}}_{1k} = \frac{\sqrt{\beta_{1k}}}{\beta^{(k)}} \sum_{j=1}^{B} \sqrt{\beta_{jk}} \mathbf{h}_{jk}. \tag{7}$$

C. Linear Receivers

We assume that the received signal is projected onto a linear filter $\mathbf{c}_{lk}^{(l)} \in \mathbb{C}^{M \times 1}$ for the kth user in the lth cell. Since the SINR analysis is identical for all users in the system we focus only on user $k = 1$ in base station indexed $l = 1$. We also drop the superscript $(.)^{(1)}$ for notational convenience. Consequently, if coherent detection is employed then $\mathbf{c} = \hat{\mathbf{h}}_{11}$. Alternatively, using the channel estimates for all users of the first cell, the MMSE filter for user 1 in the cell 1 is defined as $\arg \min_{\mathbf{c}} \mathbb{E}[|q_{11} - \mathbf{c}^H \mathbf{y}|^2 | \hat{\mathbf{h}}_{1k} \forall k]$. Defining

$$\mathbf{z} = \sum_{j=2}^{B} \sum_{k=1}^{K} \sqrt{\beta_{jk}} \mathbf{h}_{jk} q_{jk} \tag{8}$$

which represents the other cell interference and $\mathbf{h}_{1k} = \hat{\mathbf{h}}_{1k} + \tilde{\mathbf{h}}_{1k}$ the received signal can be rewritten as

$$\mathbf{y} = \sum_{k=1}^{K} \sqrt{\beta_{1k}} \hat{\mathbf{h}}_{1k} q_{1k} + \sum_{k=1}^{K} \sqrt{\beta_{1k}}, \tilde{\mathbf{h}}_{1k} q_{1k} + \mathbf{z} + \mathbf{w} \tag{9}$$

where, $\tilde{\mathbf{h}}_{1k}$ is the result of pilot contamination. The MMSE filter is then given by the expression

$$\hat{\mathbf{c}} = \left(\mathbb{E}[\mathbf{y}\mathbf{y}^H | \hat{\mathbf{h}}_{1k} \forall k] \right)^{-1} \mathbb{E}[\mathbf{y} q_{11}^* | \hat{\mathbf{h}}_{1k} \forall k]$$

$$= \mathbf{S}^{-1} \sqrt{\beta_{11}} \hat{\mathbf{h}}_{11}, \tag{10}$$

where

$$\mathbf{S} = \left(\sum_{k=2}^{K} \beta_{1k} \hat{\mathbf{h}}_{1k} \hat{\mathbf{h}}_{1k}^H + (\theta_1 + \theta_2 + \sigma^2) \mathbf{I} \right) \tag{11}$$

$$\theta_1 \mathbf{I} = \mathbb{E}[\mathbf{z}\mathbf{z}^H] = \sum_{j=2}^{B} \left[\frac{1}{M} \sum_{k=1}^{K} \beta_{jk} \right] \mathbf{I}, \tag{12}$$

$$\theta_2 \mathbf{I} = \sum_{k=1}^{K} \beta_{1k} \mathbb{E}[\tilde{\mathbf{h}}_{1k} \tilde{\mathbf{h}}_{1k}^H]$$

$$= \sum_{j=2}^{B} \left[\frac{1}{M} \sum_{k=1}^{K} \beta_{jk} \left(\frac{\beta_{1k}}{\beta^{(k)}} \right) \right] \mathbf{I}. \tag{13}$$

As seen from the expression for the filter in (10), the lack of channel knowledge of other-cell users and only a partial channel knowledge of in-cell users shows up as effective noise terms θ_1 and θ_2, respectively. To obtain the expression we also use the properties of the MMSE estimate that $\mathbb{E}[\hat{\mathbf{h}} \tilde{\mathbf{h}}^H] = 0$. In an ideal situation, the channel estimation incurs no error and $\hat{\mathbf{h}}_{1k} = \mathbf{h}_{1k}$ for all k, then

$$\mathbf{c}^* = \left(\sum_{k=1}^{K} \beta_{1k} \mathbf{h}_{1k} \mathbf{h}_{1k}^H + (\theta_1 + \sigma^2) \mathbf{I} \right)^{-1} \sqrt{\beta_{11}} \mathbf{h}_{11}. \tag{14}$$

This is an optimistic scenario which will serve as a benchmark for the performance of the MMSE filter with pilot contamination.

III. MMSE FILTER WITH PILOT CONTAMINATED ESTIMATE

After processing the received signal using the linear filter $\mathbf{c}$, let P_{signal}, $P_{\text{noise}}(\mathbf{c})$, $P_{\text{contam}}(\mathbf{c})$, $P_{\text{inter}}(\mathbf{c})$ denote the signal power, noise power, pilot interference power and interference power, respectively as a function of the filter $\mathbf{c}$. It follows that:

$$P_{\text{signal}}(\mathbf{c}) = \beta_{11} \mathbf{c}^H \mathbf{h}_{11} \mathbf{h}_{11}^H \mathbf{c} \tag{15}$$

$$P_{\text{noise}}(\mathbf{c}) = \sigma^2 \mathbf{c}^H \mathbf{c} \tag{16}$$

$$P_{\text{contam}}(\mathbf{c}) = \mathbf{c}^H \left(\sum_{j=2}^{B} \beta_{j1} \mathbf{h}_{j1} \mathbf{h}_{j1}^H \right) \mathbf{c} \tag{17}$$

$$P_{\text{inter}}(\mathbf{c}) = \mathbf{c}^H \left(\sum_{j=1}^{B} \sum_{k=2}^{K} \beta_{jk} \mathbf{h}_{jk} \mathbf{h}_{jk}^H \right) \mathbf{c}. \tag{18}$$

The received SINR is then given by the expression

$$\text{SINR} = \frac{P_{\text{signal}}(\mathbf{c})}{P_{\text{noise}}(\mathbf{c}) + P_{\text{contam}}(\mathbf{c}) + P_{\text{inter}}(\mathbf{c})}. \tag{19}$$

The motivation to define $P_{\text{contam}}(\mathbf{c})$ as in (17) is because the user 1 of other cells is sending pilots in the same resource element as the user 1 of the first base station. Hence, user 1 for $j = 2, \ldots B$ contribute to the interference in a different way as compared to the rest of the users in the system. This interference contribution is termed as pilot interference and is

due to pilot contaminated channel estimate being used to design linear filters. As per the definition the $P_{\text{contam}}(\mathbf{c})$ is dependent on the linear filter $\mathbf{c}$ and could be different for $\hat{\mathbf{h}}_{11}$ and $\hat{\mathbf{c}}$.

Define $\boldsymbol{\beta}_j$ as the random variable representing the large scale fading gain from an arbitrary user in the jth cell. Therefore, β_{jk} can be interpreted as the realization of $\boldsymbol{\beta}_j$ for the user k and let $\boldsymbol{\beta} = \sum_{i=1}^{B} \boldsymbol{\beta}_i$. Next, we state the main theorem of the paper which gives the expression of SINR for a large system when an MMSE filter with a pilot contaminated estimate is used to decode the received signal.

Theorem 1: As $M, K \to \infty$, with $K/M = \alpha$, the SINR at the output of filter $\hat{\mathbf{c}}$ given in (10) converges almost surely to

$$\widehat{\text{SINR}} = \frac{\frac{\beta_{11}}{1+\left(\sum_{j=2}^{B} \beta_{j1}\right)/\beta_{11}}}{\sigma^2 + \frac{\left(\sum_{j=2}^{B} \beta_{j1}^2\right)/\beta_{11}}{1+\left(\sum_{j=2}^{B} \beta_{j1}\right)/\beta_{11}} + \alpha\left(\mathbb{E}[\boldsymbol{\beta}] - \mathcal{C}(\alpha)\right)} \quad (20)$$

where, the constants $\mathcal{C}(\alpha), \eta_1, \eta_2$ are given by

$$\mathcal{C}(\alpha) = \mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2 \eta_1}{1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1}\right] + \frac{\eta_2}{\eta_1}\mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\left(\sum_{j=2}^{B}\frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}}\right)}{1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1}\right]$$

$$+ \frac{\eta_2}{\eta_1}\mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\left(\sum_{j=2}^{B}\frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}}\right)}{\left(1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1\right)^2}\right], \quad (21)$$

$$\eta_1 = \left(\sigma^2 + \alpha\mathbb{E}[\boldsymbol{\beta}] - \alpha\mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2 \eta_1}{1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1}\right]\right)^{-1}, \quad (22)$$

$$\eta_2 = \left(\eta_1^{-2} - \alpha\mathbb{E}\left[\left(\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}}{1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1}\right)^2\right]\right)^{-1}. \quad (23)$$

Proof: Proof given in Appendix B. ∎

We will see in a large system that Theorem 1 characterizes the effect of pilot interference power and interference averaging. Specifically, in order to put Theorem 1 into proper perspective we state two propositions which are the results for SINR of a large system with MMSE filter employing a perfect estimate and a matched filter with pilot contaminated estimate, respectively.

Proposition 2: As $M, K \to \infty$, with $K/M = \alpha$, the SINR at the output of filter $\mathbf{c}^*$ given in (14) converges almost surely to

$$\text{SINR}^* = \beta_{11}\eta_1 = \frac{\beta_{11}}{\sigma^2 + \alpha\sum_{j=2}^{N}\mathbb{E}[\boldsymbol{\beta}_j] + \alpha\mathbb{E}\left[\frac{\boldsymbol{\beta}_1}{1+\boldsymbol{\beta}_1\eta_1^*}\right]} \quad (24)$$

where, $\eta_1^* = \left(\sigma^2 + \alpha\sum_{j=2}^{B}\mathbb{E}[\boldsymbol{\beta}_j] + \alpha\mathbb{E}[\boldsymbol{\beta}_1/(1 + \boldsymbol{\beta}_1\eta_1^*)]\right)^{-1}$.

Proof: We state the proposition without proof as it is straightforward to obtain it from the large system analysis techniques used for CDMA systems in [11], [13]. ∎

SINR^* is the SINR with MMSE filtering with a perfect channel estimate to its own users. This is best case scenario as compared to the MMSE with a channel estimate. We do not expect the SINR of MMSE filter with estimate to exceed this SINR^*.

Proposition 3: As $M, K \to \infty$, with $K/M = \alpha$, the SINR at the output of filter $\mathbf{c} = \hat{\mathbf{h}}_{11}$ converges almost surely to

$$\overline{\text{SINR}} = \frac{\frac{\beta_{11}}{1+\left(\sum_{j=2}^{B} \beta_{j1}\right)/\beta_{11}}}{\sigma^2 + \left[\frac{\left(\sum_{j=2}^{B} \beta_{j1}^2\right)/\beta_{11}}{1+\left(\sum_{j=2}^{B} \beta_{j1}\right)/\beta_{11}} + \alpha\mathbb{E}[\boldsymbol{\beta}]\right]}. \quad (25)$$

Proof: Proof given in Appendix C. ∎

It is interesting to see that the expression for $\overline{\text{SINR}}$ converges to a similar expression to the result in matched filtering $\widehat{\text{SINR}}$. If $\tilde{P}_{\text{signal}} = \beta_{11}^2 / \sum_{j=1}^{B} \beta_{j1}$, $\tilde{P}_{\text{contam}} = \sum_{j=2}^{B} \beta_{j1}^2 / \sum_{j=1}^{B} \beta_{j1}$ and

$$\tilde{P}_{\text{inter}}(\mathbf{c}) = \begin{cases} \mathbb{E}[\boldsymbol{\beta}], & \mathbf{c} = \hat{\mathbf{h}}_{11} \\ \mathbb{E}[\boldsymbol{\beta}] - \mathcal{C}(\alpha), & \mathbf{c} = \hat{\mathbf{c}} \end{cases} \quad (26)$$

then we can define the expression for the asymptotic SINR with matched filtering and MMSE filtering with pilot contaminated channel estimate as

$$\text{SINR}(\mathbf{c}) = \frac{\tilde{P}_{\text{signal}}}{\sigma^2 + \tilde{P}_{\text{contam}} + \alpha\tilde{P}_{\text{inter}}(\mathbf{c})}. \quad (27)$$

Here, $\tilde{P}_{\text{signal}}$ is the effective signal power, $\tilde{P}_{\text{contam}}$ is termed as the pilot interference power and is the consequence of employing pilot contaminated channel estimate. In the same lines as in [11], the filter dependent term $\tilde{P}_{\text{inter}}(\mathbf{c})$ is called the interference averaging term and is relevant when $\alpha \neq 0$. Although, $M \to \infty$ implying that the channel between the user 1 is asymptotically orthogonal to channel between any other users in the system, because $K \to \infty$ with $K/M = \alpha$, the contribution to sum interference from all the users is non-zero and is given by $\alpha\tilde{P}_{\text{inter}}(\mathbf{c})$ for a linear filter $\mathbf{c}$. If $\nu_1 = (\beta_{11}/\beta^{(1)})^2$, then the following equations shows the relationships between the $\tilde{P}_{\text{signal}}$ and $\tilde{P}_{\text{contam}}$ to that of definitions in (15) and (17):

$$\frac{P_{\text{signal}}(\hat{\mathbf{h}}_{11})}{\sqrt{\nu_1}} = \frac{P_{\text{signal}}(\hat{\mathbf{c}})}{\tilde{P}_{\text{signal}}\eta_1^2} = \tilde{P}_{\text{signal}} \quad (28)$$

$$\frac{P_{\text{contam}}(\hat{\mathbf{h}}_{11})}{\sqrt{\nu_1}} = \frac{P_{\text{contam}}(\hat{\mathbf{c}})}{\tilde{P}_{\text{signal}}\eta_1^2} = \tilde{P}_{\text{contam}}. \quad (29)$$

It is seen that for both the filters $\hat{\mathbf{h}}_{11}$ and $\hat{\mathbf{c}}$, their respective signal powers given by $P_{\text{signal}}(\hat{\mathbf{h}}_{11})$ and $P_{\text{signal}}(\hat{\mathbf{c}})$ are just scaled versions of $\tilde{P}_{\text{signal}}$. Similarly, $P_{\text{contam}}(\hat{\mathbf{h}}_{11})$ and $P_{\text{contam}}(\hat{\mathbf{c}})$ are the scaled versions of $\tilde{P}_{\text{contam}}$. The following can be concluded for both matched filter as well as MMSE filter with a pilot contaminated estimate:

- The contribution of the pilot contaminated channel estimate to interference is given by $\tilde{P}_{\text{contam}}$ is same for both matched filter and MMSE filter.
- The contribution of pilot interference is independent of α and is equal to $\tilde{P}_{\text{contam}}$ for all values of α.

As trivial as the above comments may seem it is not obvious for the filter $\hat{\mathbf{c}}$ from the definition of filter dependent pilot interference power P_{contam} in (17). This is because the matrix $\mathbf{S}$ in filter $\hat{\mathbf{c}}$ is dependent upon the channel of all the users in the

system through the pilot contaminated channel estimate $\hat{\mathbf{h}}_{1k}$ for all $k = 1, \ldots, K$. However, as we see in the Appendix B, the contribution of the matrix $\mathbf{S}$ can be summed up into the constant η_1 for a large system. Further, the following relations can also be obtained on the interference averaging term $\tilde{\mathrm{P}}_{\mathrm{inter}}(\mathbf{c})$ when $\mathbf{c} = \hat{\mathbf{h}}_{11}$ and $\mathbf{c} = \hat{\mathbf{c}}$

$$\frac{P_{\mathrm{noise}}(\hat{\mathbf{c}}) + P_{\mathrm{inter}}(\hat{\mathbf{c}})}{\tilde{\mathrm{P}}_{\mathrm{signal}} \, \eta_1^2} = \sigma^2 + \alpha \tilde{\mathrm{P}}_{\mathrm{inter}}(\hat{\mathbf{c}}), \qquad (30)$$

$$\frac{P_{\mathrm{noise}}(\hat{\mathbf{h}}_{11}) + P_{\mathrm{inter}}(\hat{\mathbf{h}}_{11})}{\sqrt{\nu_1}} = \sigma^2 + \alpha \tilde{\mathrm{P}}_{\mathrm{inter}}(\hat{\mathbf{h}}_{11}). \qquad (31)$$

We have shown that $\tilde{\mathrm{P}}_{\mathrm{inter}}(\hat{\mathbf{h}}_{11}) - \tilde{\mathrm{P}}_{\mathrm{inter}}(\hat{\mathbf{c}}) = \mathcal{C}(\alpha) \geq 0$ implying that interference suppression of amount $\mathcal{C}(\alpha)$ can be always achieved. The amount of interference suppression $\mathcal{C}(\alpha)$ is of course depended on α through (21).

The SINR expression in the limit of infinite antennas but finite number of users per cell are obtained when we put $\alpha = 0$ in (20) and (25) or in (27). This corresponds to a similar expression as for downlink SINR in [5, eq. (16)]. Since the SNR of the uplink received signal in (2) is given by $\mathsf{SNR} = 1/\sigma^2$, it is seen that the SINR expression so obtained is limited by the pilot interference powers at high SNRs and we obtain [3, eq. (13)]. This need not necessarily be achieved by transmitting at high power per symbol or equivalently by having $\sigma^2 \to 0$. Had the transmission power of data symbols be not scaled by the number of antennas, the SNR of the received signal would be linearly increasing in M and when $\alpha = 0$ we again obtain [3, eq. (13)].

The deterministic equivalent for SINR in Rayleigh fading for MMSE filter and the corresponding expression $\mathcal{C}(\alpha)$ representing the interference suppression power are our main contribution of this paper. The terms $\mathbb{E}[\boldsymbol{\beta}]$ and $\mathcal{C}(\alpha)$ can be computed offline for a system with the knowledge of large scale fading distribution. It can also be estimated without the knowledge of the large scale fading distribution from user realizations over time. Depending on the value of $\mathbb{E}[\boldsymbol{\beta}] - \mathcal{C}(\alpha)$ and the operating point α a decision can be made whether to use an MMSE filter or an matched filter. As we will see in the next section, the interference suppression obtained with an MMSE filter is necessary in increasing the outage SINR and achievable rate of the system, when there are considerable number of users as represented by the ratio $\alpha > 0.1$. This is as opposed to the regime in which antennas far outnumber users. In this operating point $\alpha \approx 0$, and then MMSE filter itself may not be necessary as pilot signals are the main contributor to interference. This regime with $\alpha \approx 0$ has been well explored in recent studies in [3], [8], [14], [16]. As mentioned earlier most of our focus for performance analysis is on the regime $\alpha > 0.1$ although the results are perfectly valid for any $\alpha \geq 0$. Across the users in the system the SINR is a random variable by virtue of different received powers of both the signal and the interferers contributing to pilot contamination. Also, the pilot interference power is random by virtue of the choice of the interferers contributing to pilot contamination.
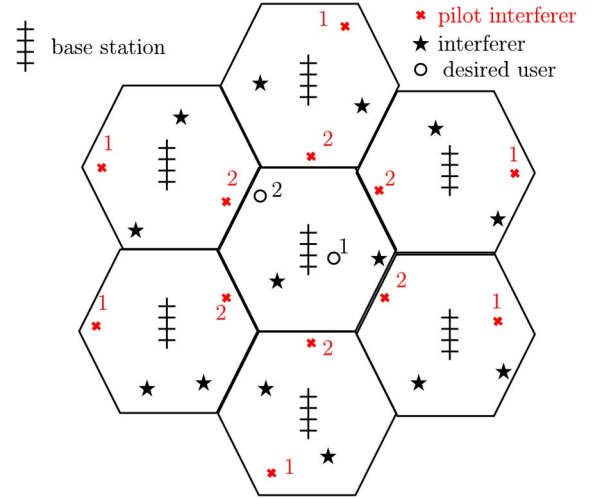

Fig. 2. In the favorable case, the sum of the received powers of interferers contributing to pilot contamination are very less as compared to that of the desired user. User 1 in the center cell represents such a scenario. The SINR with a pilot corrupted estimate is then comparable to that of perfect estimate. On the other hand for user 2 in the center cell, the pilot interferers received powers are comparable to that of the desired user and represents the worst case scenarios.

IV. PERFORMANCE ANALYSIS

For the numerical evaluation, we consider hexagonal cells with users uniformly distributed in each of the cells, as shown in Fig. 2. We consider a scenario where 6 closest cells are interfering with the center cell. We assume $\beta_{1k} = 1$ so that received powers from all the users within a cell are unity. We consider a high SNR of 20 dB and the received powers from all the users in other cells are assumed to take a constant value of $\beta_{jk} = 0.001$, or 0.01, or 0.1 for $j \neq 1$. These represent the contribution of other cell interference for three different idealized scenarios. The interference from other cells is strong as β_{jk} is close to 1. We consider the SINR for the user one in the center cell. Figs. 3 and 4 plot the asymptotic $\widehat{\mathsf{SINR}}$ of the MMSE with a pilot corrupted estimate given by $\widehat{\mathsf{SINR}}$ for the case of different received powers. Although in theory the effect for small scale fading vanishes only with infinite number of antennas it is seen in Fig. 4 that even for a 50-antenna base station the actual SINR realizations obtained through simulations are very near to the asymptotic limit. We also plot $\overline{\mathsf{SINR}}$ and SINR^* as baseline for performance comparisons. In Fig. 3, it is seen that SINR^* is already affected by other cell interference due $\beta_{jk} = 0.1$ for $j \neq 1$. Hence, the $\widehat{\mathsf{SINR}}$ is not expected to perform better than that and there is further 4 dB loss due to pilot contamination. However, in the other extreme case when the other cell β_{jk}'s are close to zero, the channel estimate is already better and the $\widehat{\mathsf{SINR}}$ performs close to SINR^*. Useful gains employing an MMSE filter with pilot contaminated channel estimate can be obtained when the other cell β_{jk}'s are neither close to zero or close to unity. In this example when the $\beta_{jk} = 0.01$ for $j \neq 1$, around 7 dB gains are possible in comparison with matched filter with pilot estimate when operating at $\alpha = 0.5$ as seen from Fig. 4. While there is a loss of 3 dB with respect to the perfect MMSE due nature of channel estimate, the reader is reminded that this is a worse case loss. The curves closes in
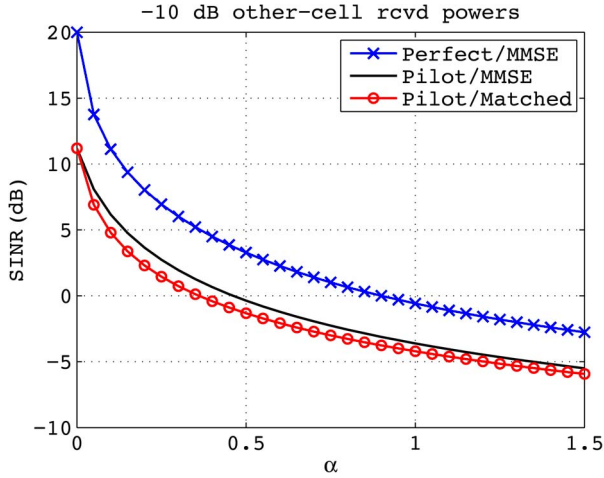
Fig. 3. The plot shows the asymptotic SINR of the first user in the first cell when MMSE filter with pilot contaminated estimate is used to decode the received signal along with the baseline comparison criterion of MMSE with a perfect estimate and matched filter with a pilot contaminated estimate for an idealized seven cell set up. It is seen that when the other cell received power is just 10 dB lower than that in cell users the MMSE filter with pilot contaminated estimate performs close to its corresponding matched filter. This is because the limitation is now the other cell interference which MMSE filter is not designed to suppress. Hence, we do not expect the MMSE filter with pilot estimate to be useful when $\beta_j > 0.1$.
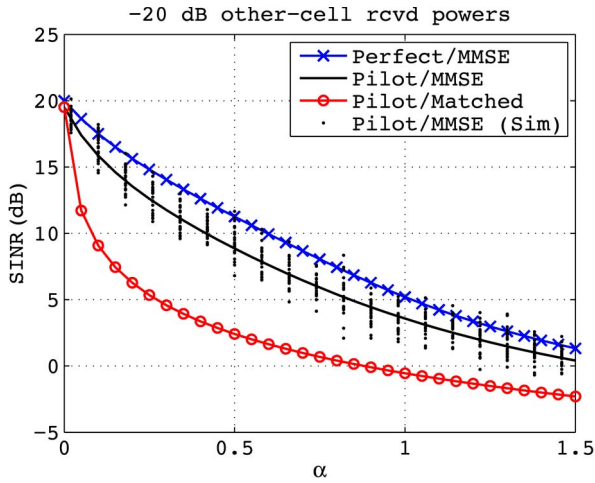

Fig. 4. The plot shows the asymptotic SINR of the first user in the first cell when MMSE filter with pilot contaminated estimate is used to decode along with the baseline comparison criterion of MMSE with a perfect estimate and matched filter with a pilot contaminated estimate for an idealized seven cell set up. In this case when the other cell received powers is 20 dB lower than that of in-cell received powers significant gains are obtained as compared to a matched filter with pilot contaminated estimate.

as we decrease α which represents the $M \gg K$ scenario and also when α increases as in that case the limitation is now the averaged interference term.

In Fig. 5 we plot the achievable sum rate for users in the first cell with each users SNR being 20 dB. We assume large enough coherence time so that the training time need not be taken into account. This is because our focus is on the sum rate achievable with variation in α. However, if necessary the sum rate can be easily adjusted based on training overhead when coherence time is a significant factor. We fix the number of antennas and calculate the sum rate with varying α as $\alpha M \log_2(1 + \widehat{\text{SINR}})$.
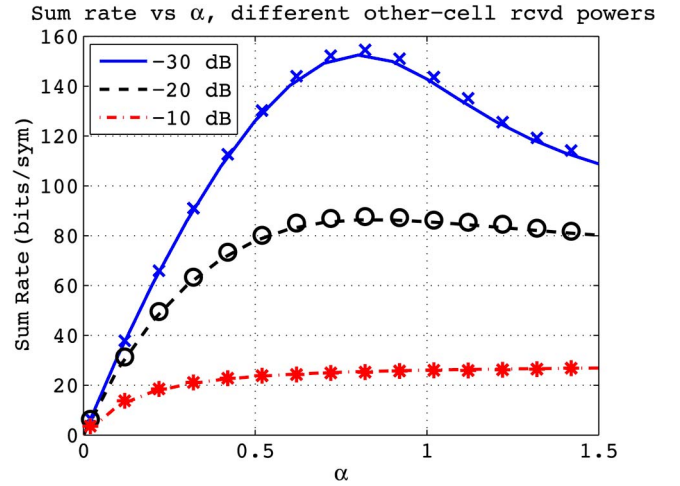

Fig. 5. The plot shows the sum rate of the users in first cell for different operating points of α corresponding to a SNR of 20 dB. The three curves correspond to three different received powers from other cell users. The markers correspond to simulation with 50-antenna base stations and match the theoretical predictions.

The three curves corresponds to the received powers of all users from other cells being either $\beta_{jk} = 0.001$, or 0.01, or 0.1 for $j \neq 1$ assuming unit received power from the in-cell users. Sum rates of over 20 bits/symbol are achieved for users when the other cell received powers are below 10 dB of the in-cell received powers. Also, the simulation with 50-antenna base station is seen to match the theoretical rates predicted for this set up. The interference limited system has the flexibility to serve a large number of users at low SINR or a few number of users at a high SINR depending on the operating point α. The plot suggest a optimal operating point α for which the sum rate is maximum. For example when $\beta_j = 0.01$ and $\alpha = 0.8$ gives a sum rate of around 88 bits/symbol. Larger α causes the $\widehat{\text{SINR}}$ to be lower so that the α term outside the $\log_2$ is ineffective to increase sum rate while a lower α implies that less users are served and hence lesser sum rate. When the other-cell received powers are large, the curve flattens and the sum rate is constant for most of the operating points α.

In Fig. 6, we plot the difference of achievable rate per user between MMSE filter with a perfect estimate and MMSE filter with a pilot estimate for different values of other cell interference power. We do not take into account the training overhead for comparison assuming we obtain the perfect estimate with the same training time. We limit ourselves to $\beta_j < 0.1$ since $\widehat{\text{SINR}}$ is already close to $\overline{\text{SINR}}$ otherwise. Also, in $\beta_j > 0.1$ regime, there is significant other cell interference which both the perfect estimate based and the pilot based MMSE filter are not designed to suppress, thereby affecting the achievable rates. We plot five different curves corresponding to system operating points α. As seen earlier the total interference with filter $\hat{\mathbf{c}}$ is given by $\tilde{\text{P}}_{\text{contam}} + \alpha\tilde{\text{P}}_{\text{inter}}(\hat{\mathbf{c}})$ and with $\mathbf{c}^*$ the interference power is given by $\alpha(\sum_{j=2}^{B} \mathbb{E}[\boldsymbol{\beta}_j] + \mathbb{E}[\boldsymbol{\beta}_1/(1 + \boldsymbol{\beta}_1\eta_1^*)])$. When α is small then the significant loss of rate with $\hat{\mathbf{c}}$ is due to $\tilde{\text{P}}_{\text{contam}}$ as the filter dependent interference power for both filters in negligible. On the hand, when α is large $\alpha\tilde{\text{P}}_{\text{inter}}$ dominates the $\tilde{\text{P}}_{\text{contam}}$ and since both the filter are affected by $P_{\text{inter}}(\mathbf{c})$
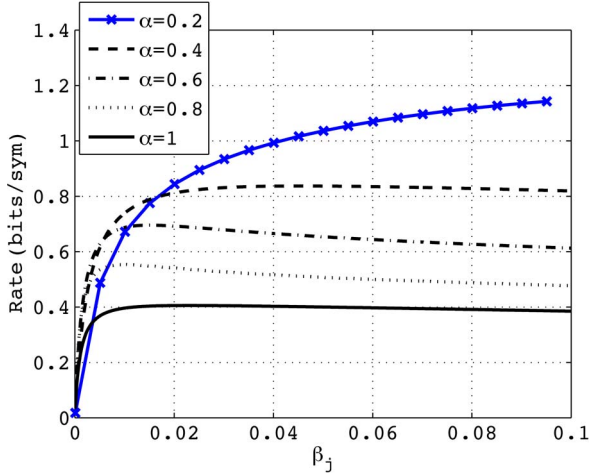
Fig. 6. The plot shows the loss of rate due to pilot contamination when an MMSE filter with pilot contaminated estimate is used to decode the received signal as compared to perfect estimate MMSE filter. The different curves correspond to different values of α from 0.2 to 1. It is seen that smaller the α the MMSE filter with pilot contaminated estimate performs worse than the ideal MMSE filter. However, the sum rate for users per base station is different.

the rate difference is small. Although when $\alpha = 1$ there is only a 0.4 bits/symbol difference the sum rate will be affected differently. For example in a 50-antenna base station with 50 users at $\beta_j = 0.1$ this could mean that sum rate with pilot contaminated MMSE filter is 20 bits/symbol lower than that perfect MMSE filter. On the other hand when $\alpha = 0.2$, the sum rate difference is 12 bits/symbol.

A. Effect of Pilot Contamination

Through a couple of typical possible realizations of user positions, we explain the effect of pilot interference in $\widehat{\text{SINR}}$, $\overline{\text{SINR}}$ and compare it with that of the SINR with a perfect estimate. For illustration, in Fig. 2 consider only distance based pathloss in large scale fading although the result holds when shadowing is also present. This is applicable to both matched filter and MMSE filter. Consider the first scenario when

$$\frac{\sum_{j=2}^{B} \beta_{j1}}{\beta_{11}} \ll 1 \Rightarrow \frac{\sum_{j=2}^{B} \beta_{j1}^2}{\beta_{11}} \ll 1. \qquad (32)$$

This corresponds to the fact that sum of received powers of the interferers are much less that that of desired user power. Under these conditions the SINR of the received signal in (25) is

$$\text{SINR} \approx \frac{\beta_{11}}{\sigma^2 + \alpha \left(\mathbb{E}[\boldsymbol{\beta}] - \mathcal{I} \right)}$$

where, $\mathcal{I} = 0$ if matched filter is employed or $\mathcal{I} = \mathcal{C}(\alpha)$ if MMSE filter $\hat{\mathbf{c}}$ is employed. As we will see in the next section, typically scenarios show that the interference suppression power $\mathcal{C}(\alpha)$ is almost same as what could have been with a filter $\mathbf{c}^*$. Hence, the SINR of the filter with the corrupt channel estimate is as good as the SINR with a perfect channel estimate. In Fig. 2, the situation of user 1 in the center cell represents the favorable scenario with the interferers contributing to the pilot contaminated channel estimate are far such that the condition
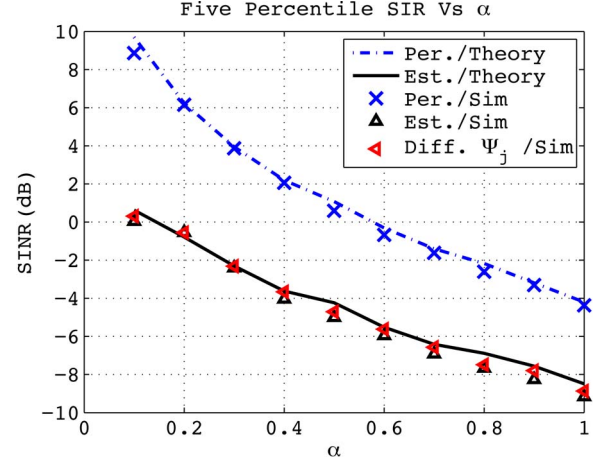

Fig. 7. The plot shows the five percentile SINR of the first user in the first cell for a seven cell set up in a non-idealized scenario. The received powers from users can be different depending on their positions and shadowing and hence the received SINR is random. The theoretical curves are matched with simulation. The details are described in Section IV-B.

(32) is satisfied. On the other hand if gains of all the interferers are comparable to that of the desired users, i.e.,

$$\frac{\sum_{j=2}^{B} \beta_{j1}}{\beta_{11}} \approx B - 1 \qquad (33)$$

then pilot interference contributes negatively to the SINR in addition to interference averaging. This is represented by realization of user 2 of the center cell in Fig. 2. Therefore, we can conclude that, as compared to the linear filter with perfect estimate, the filter with a pilot estimate has higher probability that it is less than a given SINR.

B. Five Percentile SINR

In the earlier section we showed that pilot contamination has the effect of reducing the outage SINR. In order to get more intuition under practical scenarios of large scale fading gains, we consider the seven cell model with cell radius is $R = 1$ km, and assume a COST231 model for propagation loss between the base station and the users. The noise power is assumed to be -174 dBm and user transmit power of 23 dBm. We plot the five percentile of the SINR in Fig. 7 for the perfect MMSE filtering given by SINR^* and MMSE filter with pilot contamination given by $\widehat{\text{SINR}}$, for varying values of α. To that extend, we compute the interference terms $\mathbb{E}[\boldsymbol{\beta}]$ and $\mathcal{C}(\alpha)$ offline by averaging over a sufficient number of user positions. Also, η_1 and η_2 can be computed offline for different values of α as they are constant for the system and dependent on the large scale fading characteristics. Notice that $\widehat{\text{SINR}}$ is devoid of the small scale fading parameters. Also, for SINR^* we compute the terms, η_1^* and the average interference $\mathbb{E}[\boldsymbol{\beta}] - \mathbb{E}[\boldsymbol{\beta}_1^2 \eta_1^*/(1 + \boldsymbol{\beta}_1 \eta_1^*)]$. It is found that $\mathbb{E}[\boldsymbol{\beta}] - \mathcal{C} = 38$ dB and $\mathbb{E}[\boldsymbol{\beta}] - \mathbb{E}[\boldsymbol{\beta}_1^2 \eta_1^*/(1 + \boldsymbol{\beta}_1 \eta_1^*)] = 36$ dB. This shows that in terms of the interference suppression the performance of both filters $\mathbf{c}^*$ and $\hat{\mathbf{c}}$ are almost same. In order to compare the theoretical expression we also plot the five percentile SINR which is generated using simulations. These involves computing the SINRs

TABLE I
TABLE SHOWS THE ACHIEVABLE RATE (BITS/SYMBOL) WITH MMSE
FILTERING FOR A 50 AND 10-ANTENNA BASE STATION SERVING
DIFFERENT NUMBER OF USERS. R_{per} CORRESPONDS TO RATE
WITH A PERFECT ESTIMATE AND R_{pilot} CORRESPONDS TO
THE PILOT CONTAMINATED FILTER

α	$M = 50$		$M = 10$	
	R_{per}	R_{pilot}	R_{per}	R_{pilot}
0.1	6.0	4.7		
0.2	5.0	4.0		
0.3	4.3	3.4	4.4	3.6
0.4	3.8	3.0	4.0	3.2
0.5	3.4	2.7	3.6	2.9
0.6	3.1	2.5	3.3	2.7
0.7	2.8	2.3	3.0	2.4
0.8	2.64	2.1	2.7	2.2
0.9	2.4	1.9	2.4	2.0
1.0	2.2	1.9	2.3	2.0

for various small scale fading channel realizations along with large scale fading. For the simulation we use 50 antenna base stations each serving a different number of users corresponding to different α. The channel estimate is based on same in-cell orthogonal training sequences being repeated across the cells. Further, we also compare the five percentile SINR obtained out of actual channel estimation. In order for that we assume different independently generated in-cell orthogonal training sequences which are non-orthogonal across the cells. We perform an MMSE estimation of the channel and generate MMSE filter using the actual channel estimate given by (5).

It is seen through Fig. 7, that $\widehat{SINR}$ in (20) matches the five percentile SINR obtained through simulation and is typically less by about 0.3 dB of the theoretical expression. This is true for both in-cell pilot sequences being repeated across the cells as well as different and independent pilot training sequences across the cells. This also implies that the even a 50 antenna base station is large enough for the theoretical predictions to be effective in addition to being independent of the effect of small scale fading in the resulting SINR. As we increase the number of antennas the theoretical expression exactly matches the SINR obtained through simulation. Also, the MMSE filter with pilot contamination performs just 5 dB below the MMSE filter with perfect channel estimate and this gap is unambiguously a result of the pilot contaminated channel estimate. Also, it is seen that even at $\alpha = 1$, which implies a heavily loaded system the five percentile $\widehat{SINR}$ is -9 dB which is well within the sensitivity of base station receivers.

Table I shows the achievable rates per symbol for a user in the central cell using MMSE based detection with perfect estimate and pilot contaminated estimate. The achievable rate is given by $R = \mathbb{E}[\log(1 + \widehat{SINR})]$ which here is calculated by averaging the instantaneous rate over 2×10^3 realizations of user positions for user 1. This is the same for all users in the central cell. Both theory and simulations based on 50 base stations antennas serving different number of users agree to the numbers shown in the table. It is seen that the difference between them is approximately 1 bit/symbol for small $\alpha > 0.1$ and closes in when it increases. However, for $\alpha \ll 1$, the difference between the rates increases as effect of pilot interference will never let the SINR approach SNR.

Further, Table I, also shows the simulated results for the achievable rate for a 10-antenna base station with 3 to 10 users. It is seen that even for a 10 antenna base station, the simulated results agree closely to the earlier results obtained from theory in Table I. This highlights the usefulness of the large system analysis in providing accurate predictions for achievable rates for not necessarily large number of antennas but also contemporary MIMO systems. However, we would like to point out that with more number of antennas we can serve more users at the same rate given below.

V. CONCLUSION

In this paper, we found the expression for SINR for a large system when a MMSE filter with a pilot contaminated estimate is employed to decode the received signal. We validated the expression through simulations and showed that a 50-antenna base station serving different number of users is sufficient enough to employ our large system results. We characterized the effect of pilot contamination in that it has the effect of reducing the five percentile SINRs as compared to the MMSE with perfect estimate for all values of α. We also found an explicit expression for the interference suppression power due to MMSE filter and compared it with that of matched filter. We showed that five percentile SINR of the MMSE with a pilot contaminated estimate is within 5 dB of MMSE filter with a perfect estimate. We also found that the results with actual channel estimation match the theoretical results. In this paper, we have assumed that the training time allocated to users is K symbols. In future work, we wish to study the training overhead for different values of training time and users depending on the coherence time. We believe this can be done through a large system analysis of training time versus the number of users per cell. It would also be interesting to see if the considerable work done by the authors in [14], in getting a generalized expression for the deterministic equivalent of the SINR can be further simplified into intuitive expressions for other channel models. This will be of help in realizing engineering conclusions tailored for different channels models like distributed antennas, correlated antennas, distributed sets of correlated antennas [10] to name a few. Also, recent work has proposed that pilot contamination as an artefact of linear channel estimation techniques [18]. While they have provided a theoretical understanding in an ideal situation, practical solutions applicable to a regime with large number of users are still to be found. Algorithms from multi-user detection for CDMA systems are a useful tool when we have enough coherence time [21]. We are currently looking into such adaptive algorithms that could be implemented for short coherence time scenarios.

APPENDIX A
RESULTS FROM LITERATURE

In this section we briefly describe the necessary results from literature to derive the asymptotic SINR expressions. These results were used previously in the context of CDMA systems in [11]–[13].

Lemma 4 [13, Lemma 1]: If $\mathbf{S}$ is a deterministic $M \times M$ matrix with uniformly bounded spectral radius for all M. Let $\mathbf{q} = (1/\sqrt{M})[q_1 \; q_2 \; \ldots \; q_M]^T$ where q_i's are i.i.d complex random variables with zero mean, unit variance and finite eight moment. Let $\mathbf{r}$ be a similar vector independent of $\mathbf{q}$. Then

$$\mathbf{q}^H \mathbf{S} \mathbf{q} \to \frac{1}{M} \text{trace}\{\mathbf{S}\}, \tag{34}$$

$$\mathbf{q}^H \mathbf{S} \mathbf{r} \to 0 \tag{35}$$

almost surely as $M \to \infty$.

Results in linear MMSE filters for large dimensions have been obtained using Stieltjes transform result on symmetric matrices in [11]–[13]. For completeness and clarity of understanding of MMSE for multi-user MIMO with pilot contamination we first define the Stieltjes transform of a random variable and state the result without proof here.

Definition 5 [24, Sec. 2.2.1]: Let a real valued random variable be given by the distribution G. Then, the Stieltjes transform $m(z)$ with complex argument z and positive imaginary part is defined as

$$m(z) = \int \frac{1}{\lambda - z} dG(\lambda). \tag{36}$$

Theorem 6 [13, Theorem 6]: Let $\mathbf{X} \in \mathbb{C}^{M \times K}$ be a matrix with independent and identically distributed complex entries each with variance $1/M$. Also, let $\mathbf{T} \in \mathbb{C}^{K \times K}$ be a random Hermitian non-negative definite matrix independent of $\mathbf{X}$ such that the empirical distribution of its eigenvalues converges to a fixed distribution F as $M \to \infty$. Then as $K \to \infty$ and $M \to \infty$ with $K/M = \alpha$, almost surely the empirical distribution of eigenvalues of $\mathbf{X}\mathbf{T}\mathbf{X}^H$ converges to a non-random distribution function G whose Stieltjes transform $m(z)$ satisfies

$$m(z) = \frac{1}{-z + \alpha \int \frac{p}{1 + pm(z)} dF(p)} \tag{37}$$

for z complex with positive imaginary part.

Next we state a corollary from [12] which is also a consequence of Stieltjes transform result in Theorem 26.

Corollary 7: Where Theorem 26 is applicable

$$\alpha \mathbb{E} \left[\frac{p \frac{d}{dz} m(z)}{(1 + pm(z))^2} \right] = \int \frac{\lambda}{(\lambda - z)^2} dG(\lambda). \tag{38}$$

Proof:

$$\alpha \mathbb{E} \left[\frac{p \frac{d}{dz} m(z)}{(1 + pm(z))^2} \right] \overset{(a)}{=} -\alpha \frac{d}{dz} \mathbb{E} \left[1 - \frac{pm(z)}{1 + pm(z)} \right],$$

$$= \frac{d}{dz} \alpha \mathbb{E} \left[\frac{pm(z)}{1 + pm(z)} \right],$$

$$\overset{(b)}{=} \frac{d}{dz} (1 + zm(z)),$$

$$= \frac{d}{dz} \int \frac{\lambda}{\lambda - z} dG(\lambda), \tag{39}$$

where (a) is due to dominated convergence theorem and (b) is due to (37). ∎

APPENDIX B
PROOF OF THEOREM 1

Let the overall channel matrix representing the system be defined as, $\mathbf{H} = [\mathbf{H}_1 \; \mathbf{H}_2 \; \ldots \; \mathbf{H}_K]$, where, $\mathbf{H}_i = [\mathbf{h}_{1i} \; \mathbf{h}_{2i} \; \ldots \; \mathbf{h}_{Bi}]$. Also, define the large scale fading coefficient vector to be $\mathbf{a}_i = [\sqrt{\beta_{1i}} \; \sqrt{\beta_{2i}} \; \ldots \; \sqrt{\beta_{Bi}}]^T$, and $\mathbf{e}_i \in \mathbb{C}^{B \times 1}$ as a unit vector with 1 in the ith position for $i \in \{1, 2, \ldots, B\}$. With the above definitions the channel estimate $\hat{\mathbf{h}}_{1k} = (\sqrt{\beta_{1k}}/\beta^{(k)})\mathbf{H}_k \mathbf{a}_k$ and $\hat{\mathbf{c}} = (\beta_{11}/\beta^{(1)})\mathbf{S}^{-1}\mathbf{H}_1 \mathbf{a}_1$. Also, if $\nu_k = (\beta_{1k}/\beta^{(k)})^2$, then the signal power, noise power, and pilot interference power, are, respectively given by

$$P_{\text{signal}}(\hat{\mathbf{c}}) = \beta_{11} \nu_1 \left| \mathbf{a}_1^H \mathbf{H}_1^H \mathbf{S}^{-1} \mathbf{H}_1 \mathbf{e}_1 \right|^2 \tag{40}$$

$$P_{\text{noise}}(\hat{\mathbf{c}}) = \nu_1 \sigma^2 \mathbf{a}_1^H \mathbf{H}_1^H (\mathbf{S}^{-1})^2 \mathbf{H}_1 \mathbf{a}_1, \tag{41}$$

$$P_{\text{contam}}(\hat{\mathbf{c}}) = \nu_1 \sum_{j=2}^{B} \beta_{j1} \left| \mathbf{a}_1^H \mathbf{H}_1^H \mathbf{S}^{-1} \mathbf{H}_1 \mathbf{e}_j \right|^2. \tag{42}$$

For further analysis let us define

$$\mathbf{Z} = \mathbf{S}^{-1} \left(\sum_{j=1}^{B} \sum_{k=2}^{K} \beta_{jk} \mathbf{h}_{jk} \mathbf{h}_{jk}^H \right) \mathbf{S}^{-1} \tag{43}$$

then the interference power is given by

$$P_{\text{inter}}(\hat{\mathbf{c}}) = \nu_1 \mathbf{a}_1^H \mathbf{H}_1^H \mathbf{Z} \mathbf{H}_1 \mathbf{a}_1. \tag{44}$$

Define a real block diagonal matrix $\mathbf{D}_1 \in \mathbb{R}^{(K-1)B \times (K-1)B}$ and $\mathbf{S}_1 \in \mathbb{C}^{M \times (K-1)B}$ as

$$\mathbf{D}_1 = \text{diag}\{\nu_2 \mathbf{a}_2 \mathbf{a}_2^H, \nu_3 \mathbf{a}_3 \mathbf{a}_3^H, \ldots, \nu_K \mathbf{a}_K \mathbf{a}_K^H\}, \tag{45}$$

$$\mathbf{S}_1 = [\mathbf{H}_2 \; \ldots \; \mathbf{H}_K]. \tag{46}$$

Then the matrix $\mathbf{S}$ can be rewritten as

$$\mathbf{S} = \sum_{k=2}^{K} \nu_k \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H + (\theta_1 + \theta_2 + \sigma^2)\mathbf{I}$$

$$= \mathbf{S}_1 \mathbf{D}_1 \mathbf{S}_1^H + (\theta_1 + \theta_2 + \sigma^2)\mathbf{I}. \tag{47}$$

For the matrix $\mathbf{D}_1$ there are $(K-1)(B-1)$ eigenvalues which are equal to zero and $K-1$ non-zero values given by $\beta_{1k}^2/\beta^{(k)}$, for all $k \in \{2, \ldots, K\}$. Also, define $\boldsymbol{\beta}_j$ as the random variable representing the large scale fading gain from an arbitrary user in the jth cell. Therefore, β_{jk} can be interpreted as the realization of $\boldsymbol{\beta}_j$ for the kth user. Therefore, Theorem 26 takes the form

$$m(z) = \left(-z + \alpha \mathbb{E} \left[\frac{\boldsymbol{\beta}_1^2/\boldsymbol{\beta}}{1 + \boldsymbol{\beta}_1^2 m(z)/\boldsymbol{\beta}} \right] \right)^{-1} \tag{48}$$

where, the expectation is now over the joint distribution of the $\boldsymbol{\beta}_j$s and $\boldsymbol{\beta}$.

Also, notice that the spectral radius of $\mathbf{S}$ is bounded by $(\theta_1 + \theta_2 + \sigma^2)^{-1}$. Therefore, with $\boldsymbol{\beta} = \sum_{j=1}^B \boldsymbol{\beta}_j$, $\bar{\theta}_1 = \alpha \sum_{j=2}^B \mathbb{E}[\boldsymbol{\beta}_j]$, $\bar{\theta}_2 = \alpha \sum_{j=2}^B \mathbb{E}[\boldsymbol{\beta}_j(\boldsymbol{\beta}_1/\boldsymbol{\beta})]$ and using Lemma 4 we can conclude that

$$\mathbf{H}_1^H \mathbf{S}^{-1} \mathbf{H}_1 \to \frac{1}{M} \text{trace}\{\mathbf{S}^{-1}\}\mathbf{I} = \eta_1 \mathbf{I}, \qquad (49)$$

$$\mathbf{H}_1^H (\mathbf{S}^{-1})^2 \mathbf{H}_1 \to \frac{1}{M} \text{trace}\{\mathbf{S}^{-2}\}\mathbf{I} = \eta_2 \mathbf{I} \qquad (50)$$

almost surely as $M \to \infty$ where, if G is the non-random limiting distribution of the eigenvalues λ of the matrix $\mathbf{S}_1 \mathbf{D}_1 \mathbf{S}_1^H$, then

$$\eta_1 = \int \frac{1}{\lambda + \bar{\theta}_1 + \bar{\theta}_2 + \sigma^2} dG(\lambda) \qquad (51)$$

$$\eta_2 = \int \frac{1}{(\lambda + \bar{\theta}_1 + \bar{\theta}_2 + \sigma^2)^2} dG(\lambda). \qquad (52)$$

From (51) and (52) and using the definition of Stieltjes transform [24] we can find that $\eta_1 = \lim_{-z \to \bar{\theta}_2 + \bar{\theta}_2 + \sigma^2} m(z)$, and since $m(z)$ is complex analytic $\eta_2 = \lim_{-z \to \bar{\theta}_2 + \bar{\theta}_2 + \sigma^2} (d/dz)m(z)$. The values of η_1 and η_2 can then also be obtained from solving (48) and its derivative. The equations are given by

$$\eta_1 = \left(\sigma^2 + \alpha \mathbb{E}[\boldsymbol{\beta}] - \alpha \mathbb{E}\left[\frac{\left(\frac{\beta_1^2}{\boldsymbol{\beta}}\right)^2 \eta_1}{1 + \frac{\beta_1^2}{\boldsymbol{\beta}} \eta_1} \right] \right)^{-1}, \qquad (53)$$

$$\eta_2 = \left(\eta_1^{-2} - \alpha \mathbb{E}\left[\left(\frac{\frac{\beta_1^2}{\boldsymbol{\beta}}}{1 + \frac{\beta_1^2}{\boldsymbol{\beta}} \eta_1} \right)^2 \right] \right)^{-1}. \qquad (54)$$

Similarly, to evaluate the expression for interference P_{inter}, trace$\{\mathbf{Z}\}$ can be expanded as in (58), shown at the bottom of the page where, in step (a) we use trace$\{\mathbf{qq}^H\} = \mathbf{q}^H \mathbf{q}$ and if

$$\mathbf{S}_k = \sum_{k \neq 1, k} \nu_k \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H + (\theta_1 + \theta_2 + \sigma^2)\mathbf{I} \qquad (59)$$

then in step (b) use matrix inversion lemma as

$$\mathbf{S}^{-1} = \mathbf{S}_k^{-1}\left(\mathbf{I} - \frac{\nu_k \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1}}{1 + \nu_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{a}_k} \right). \qquad (60)$$

Notice that using Lemma 4, the terms $\mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \xrightarrow{a.s.} \eta_1 \mathbf{I}$ and $\mathbf{H}_k^H \mathbf{S}_k^{-2} \mathbf{H}_k \xrightarrow{a.s.} \eta_2 \mathbf{I}$ and it appears repeatedly in (58). Therefore, in the limit of infinite number of antennas, and for a given α, using Lemma 4 we have

$$\mathbf{H}_1^H \mathbf{Z} \mathbf{H}_1 \xrightarrow{a.s.} \frac{\text{trace}\{\mathbf{Z}\}}{M} \mathbf{I} \qquad (61)$$

and trace$\{\mathbf{Z}\}/M$ in turn converges to the expression in (63), shown at the top of the next page. In (62), shown at the top of the next page, the third term $\mathbb{E}[(\boldsymbol{\beta}_1^2/\boldsymbol{\beta})\eta_2/(1 + (\boldsymbol{\beta}_1^2/\boldsymbol{\beta})\eta_1)^2]$ is equal to $\int_0^\infty (\lambda/(\lambda + \bar{\theta}_1 + \bar{\theta}_2 + \sigma^2)^2)dG(\lambda)$. This follows from corollary 7.

Using equations (49), (50), (63) in expressions (40)–(42) and (44) the SINR given in (19) converges almost surely to $\widehat{\text{SINR}}$ as in expression (20).

APPENDIX C
PROOF OF PROPOSITION 3

With $\nu_1 = (\beta_{11}/\beta^{(1)})^2$, using the matched filter given by $\mathbf{c} = (\sqrt{\beta_{11}}/\beta^{(1)})\mathbf{H}_1 \mathbf{a}_1$, the signal power, noise power, pilot interference power is given by $P_{\text{signal}} = \nu_1 |\mathbf{a}_1 \mathbf{H}_1^H \mathbf{H}_1 \mathbf{e}_1|^2$, $P_{\text{noise}} = (\nu_1/\beta_{11})\sigma^2 \mathbf{a}_1^H \mathbf{H}_1^H \mathbf{H}_1 \mathbf{a}_1$, $P_{\text{contam}} = (\nu_1/\beta_{11}) \sum_{j=1}^B \beta_{j1} |\mathbf{a}_1^H \mathbf{H}_1^H \mathbf{H}_1 \mathbf{e}_j|^2$. Since, $\mathbf{H}^H \mathbf{H} \xrightarrow{a.s.} \mathbf{I}$ as $M \to \infty$, we have, $P_{\text{signal}} \xrightarrow{a.s.} \nu_1 \beta_{11}$, $P_{\text{noise}} \xrightarrow{a.s.} (\nu_1/\beta_{11})\beta^{(1)}\sigma^2$,

$$\text{trace}\{\mathbf{Z}\} = \text{trace}\left\{ \sum_{j=1}^B \sum_{k=2}^K \beta_{jk} \mathbf{S}^{-1} \mathbf{H}_k \mathbf{e}_j \mathbf{e}_j^H \mathbf{H}_k^H \mathbf{S}^{-1} \right\} \qquad (55)$$

$$\stackrel{(a)}{=} \sum_{j=1}^B \sum_{k=2}^K \beta_{jk} \mathbf{e}_j^H \mathbf{H}_k^H (\mathbf{S}^{-1})^2 \mathbf{H}_k \mathbf{e}_j \qquad (56)$$

$$\stackrel{(b)}{=} \sum_{j=1}^B \sum_{k=2}^K \beta_{jk} \mathbf{e}_j^H \mathbf{H}_k^H \mathbf{S}_k^{-2}\left(\mathbf{I} - 2\frac{\nu_k \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1}}{1 + \nu_k^2 \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{a}_k} + \frac{\nu_k^2 \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1}}{(1 + \nu_k^2 \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{a}_k)^2} \right) \mathbf{H}_k \mathbf{e}_j \qquad (57)$$

$$= \sum_{j=1}^B \left(\sum_{k=2}^K \beta_{jk} \mathbf{e}_j^H \mathbf{H}_k^H \mathbf{S}_k^{-2} \mathbf{H}_k \mathbf{e}_j - 2\sum_{k=2}^K \beta_{jk} \frac{\nu_k \mathbf{e}_j^H \mathbf{H}_k^H \mathbf{S}_k^{-2} \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{e}_j}{1 + \nu_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{a}_k} \right.$$

$$\left. + \sum_{k=2}^K \beta_{jk} \frac{\nu_k^2 \mathbf{e}_j^H \mathbf{H}_k^H \mathbf{S}_k^{-2} \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{a}_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{e}_j}{(1 + \nu_k \mathbf{a}_k^H \mathbf{H}_k^H \mathbf{S}_k^{-1} \mathbf{H}_k \mathbf{a}_k)^2} \right) \qquad (58)$$

$$\frac{\text{trace}\{\mathbf{Z}\}}{M} \xrightarrow{a.s.} \alpha \sum_{j=1}^{B} \mathbb{E}\left[\boldsymbol{\beta}_j \left(\eta_2 - 2 \frac{\left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right)^2 \boldsymbol{\beta}_j \eta_1 \eta_2}{1 + \eta_1 \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}} + \frac{\left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right)^4 \boldsymbol{\beta}_j \boldsymbol{\beta} \eta_2 \eta_1^2}{\left(1 + \eta_1 \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2} \right) \right]$$

$$= \alpha \sum_{j=2}^{B} \mathbb{E}[\boldsymbol{\beta}_j]\eta_2 + \alpha \sum_{j=1}^{B} \mathbb{E}\left[\boldsymbol{\beta}_j \frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}} \right] \eta_2 - \alpha \mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right)^2 \sum_{j=1}^{B} \boldsymbol{\beta}_j \eta_1 \eta_2}{1 + \eta_1 \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}} \right] - \alpha \mathbb{E}\left[\frac{\left(\frac{\boldsymbol{\beta}_1}{\boldsymbol{\beta}}\right)^2 \sum_{j=1}^{B} \boldsymbol{\beta}_j \eta_1 \eta_2}{\left(1 + \eta_1 \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\right)^2} \right]$$

$$= (\bar{\theta}_1 + \bar{\theta}_2)\eta_2 + \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_2}{\left(1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1\right)^2} \right] - \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\left(\sum_{j=2}^{B}\frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}}\right)\eta_1\eta_2}{1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1} \right] - \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\left(\sum_{j=2}^{B}\frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}}\right)\eta_1\eta_2}{\left(1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1\right)^2} \right] \tag{62}$$

$$= \int_0^\infty \frac{\lambda + \bar{\theta}_1 + \bar{\theta}_2}{(\lambda + \bar{\theta}_1 + \bar{\theta}_2 + \sigma^2)^2} dG(\lambda) - \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\left(\sum_{j=2}^{B}\frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}}\right)\eta_1\eta_2}{1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1} \right] - \alpha \mathbb{E}\left[\frac{\frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\left(\sum_{j=2}^{B}\frac{\boldsymbol{\beta}_j^2}{\boldsymbol{\beta}}\right)\eta_1\eta_2}{\left(1 + \frac{\boldsymbol{\beta}_1^2}{\boldsymbol{\beta}}\eta_1\right)^2} \right] \tag{63}$$

$P_{\text{contam}} \xrightarrow{a.s.} (\nu_1/\beta_{11}) \sum_{j=2}^{B} \beta_{j1}^2$. Using Lemma 4 in the interference term we have

$$P_{\text{inter}} = \frac{\nu_1}{\beta_{11}} \mathbf{a}_1^H \mathbf{H}_1^H \left(\sum_{j=1}^{B} \sum_{k=2}^{K} \beta_{jk} \mathbf{h}_k \mathbf{h}_k^H \right) \mathbf{H}_1 \mathbf{a}_1,$$

$$\xrightarrow{a.s.} \frac{\nu_1}{\beta_{11}} \mathbf{a}_1^H \left(\frac{1}{M} \text{trace}\left\{ \sum_{j=1}^{B} \sum_{k=2}^{K} \beta_{jk} \mathbf{h}_k \mathbf{h}_k^H \right\} \mathbf{I} \right) \mathbf{a}_1,$$

$$= \frac{\nu_1}{\beta_{11}} \alpha \sum_{j=1}^{B} \beta_{j1} \frac{1}{K} \left(\sum_{j=1}^{B} \sum_{k=2}^{K} \beta_{jk} \mathbf{h}_k^H \mathbf{h}_k \right),$$

$$= \frac{\nu_1}{\beta_{11}} \alpha \sum_{j=1}^{B} \beta_{j1} \left(\sum_{j=1}^{B} \mathbb{E}[\boldsymbol{\beta}_j] \right). \tag{64}$$

Rearranging the terms in (19) we get the expression for $\overline{\text{SINR}}$.

REFERENCES

[1] B. Hassibi and B. M. Hochwald, "How much training is needed for multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, pp. 951–963, Apr. 2003.

[2] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. Asilomar Conf. Signal Syst. Comput.*, Nov. 2006, pp. 359–363.

[3] T. L. Marzetta, "Non-cooperative multiantenna base stations with unlimited number of antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[4] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *Proc. IEEE ISIT*, Jun. 2009, pp. 2184–2188.

[5] 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.

[6] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[7] 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.

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

[9] H. Q. Ngo, M. Matthaiou, T. Q. Duong, and E. G. Larsson, "Uplink performance analysis of multicell MU-MIMO systems with ZF receivers," in *Proc. IEEE Swedish Commun. Technol. Workshop*, Oct. 2011, pp. 40–45.

[10] J. Zhang, C. K. Wen, S. Jin, X. Gao, and K. Wong, "On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 133–148, Feb. 2013.

[11] D. Tse and S. Hanly, "Linear multiuser receivers: Effective interference, Effective bandwidth and User capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.

[12] J. Zhang, E. K. P. Chong, and D. N. C. Tse, "Ouput MAI distributions of linear MMSE multiuser receivers in DS-CDMA systems," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1128–1144, Mar. 2001.

[13] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2059–2077, Sep. 2000.

[14] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in 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.

[15] D. Samardzija and N. B. Mandayam, "Impact of pilot design on achievable data rates in multiple antenna multiuser TDD system," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1370–1379, Sep. 2007.

[16] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Uplink power efficiency of multiuser MIMO with very large antenna arrays," in *Proc. 49th Allerton Conf. Commun., Control, Comput.*, Sep. 2011, pp. 1272–1279.

[17] N. Krishnan, R. D. Yates, and N. B. Mandayam, "Cellular system with many antennas: Large system analysis under pilot contamination," in *Proc. 50th Allerton Conf. Commun., Control, Comput.*, Oct. 2012, pp. 1220–1224.

[18] R. Mueller, M. Vekhapera, and L. Cottatellucci, "Blind pilot decontamination," in *Proc. 17th Int. ITG WSA*, Mar. 2013, pp. 1–6.

[19] R. Mueller, M. Vehkapera, and L. Cottatellucci, "Analysis of pilot decontamination based on power control," in *Proc. IEEE Veh. Technol. Conf.*, Jun. 2013, pp. 1–5.

[20] H. Huh, G. Caire, H. C. Papadopoulos, and S. A. Ramprashad, "Achieving "Massive MIMO" spectral efficiency with a not-so-large number of antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3226–3239, Sep. 2012.

[21] S. Verdu, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[22] D. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[23] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, NY, USA: Springer-Verlag, 1998, ser. Springer Texts in Electrical Engineering.

[24] A. M. Tulino and S. Verdu, "Random matrix theory for wireless communications," in *Foundations and Trends in Communications and Information Theory*, vol. 1. Delft, The Netherlands: Now, 2004.

[25] K. Takeuchi, M. Vehkapera, T. Tanaka, and R. R. Muller, "Large-system analysis of joint channel and data estimation for MIMO DS-CDMA systems," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1385–1412, Mar. 2012.

Narayanan Krishnan (M'14) received the B.Tech. degree in electronics and communications engineering from College of Engineering Trivandrum in Thiruvananthapuram, India, in 2004, the M.S. degree from Kansas State University, Manhattan, KS, USA, and the Ph.D. degree in electrical and computer engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, North Brunswick, NJ, USA. From 2004 to 2007, he was with Infosys Technologies Ltd., Bangalore, India, as a Network Test Engineer. His research interests are in resource allocation for next generation cellular systems, communication theory, multi-user detection. He is currently with Qualcomm Technologies Inc. San Diego, CA, USA.

Roy D. Yates (F'11) received the B.S.E. degree from Princeton University, Princeton, NJ, USA, in 1983, and the M.S. and Ph.D. degrees from Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1986 and 1990, respectively, all in electrical engineering. Since 1990, he has been with the Wireless Information Networks Laboratory (WINLAB), Rutgers University, North Brunswick, NJ, USA and the Department of Electrical and Computer Engineering (ECE) at Rutgers University, Piscataway, NJ, USA where he is currently a Distinguished Professor and Associate Director of WINLAB. In 1999–2000, he served as Director of WINLAB. Dr. Yates is a former Associate Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATION Series in Wireless Communication and also a past Associate Editor for Communications Networks of the IEEE TRANSACTIONS ON INFORMATION THEORY. He is an author of three editions of the text *Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers?* (New York, NY, USA: Wiley, 2014). Dr. Yates was a co-recipient of the Marconi Paper Prize for Best Paper in the IEEE TRANSACTIONS IN WIRELESS COMMUNICATION for 2002, and was awarded the Rutgers University Scholar-Teacher Award in 2011. His research interests include wireless resource allocation, information status updating, and spectrum regulation.

Narayan B. Mandayam (F'09) is a Distinguished Professor at Rutgers University, Camden, NJ, USA. He received the B.Tech (Hons.) degree in 1989 from the Indian Institute of Technology Kharagpur, Kharagpur, India, and the M.S. and Ph.D. degrees in 1991 and 1994, respectively, from Rice University Houston, TX, USA, all in electrical engineering. From 1994 to 1996, he was a Research Associate at the Wireless Information Network Laboratory (WINLAB), Rutgers University, North Brunswick, NJ, USA before joining the Faculty of the Electrical and Computer Engineering Department at Rutgers University, Piscataway, NJ, USA where he became Associate Professor in 2001 and Professor in 2003. Currently, he also serves as Associate Director at WINLAB. He was a Visiting Faculty Fellow with the Department of Electrical Engineering, Princeton University, Princeton, NJ, USA in 2002 and a Visiting Faculty Member with the Indian Institute of Science in 2003. Dr. Mandayam's research interests are in various aspects of wireless data transmission with emphasis on techniques for cognitive radio networks including their implications for spectrum policy. Using constructs from game theory, communications and networking, his work focusses on radio resource management as well as signal processing for enabling wireless technologies to support various applications. His recent interests also include modelling and analysis of trustworthy knowledge creation on the Internet.

Dr. Mandayam is a co-recipient of the 2014 IEEE Donald G. Fink Award for his IEEE Proceedings paper titled "Frontiers of Wireless and Mobile Communications" and the 2009 Fred W. Ellersick Prize from the IEEE Communications Society for his work on dynamic spectrum access models and spectrum policy. He is also a recipient of the Peter D. Cherasia Faculty Scholar Award from Rutgers University (2010), the National Science Foundation CAREER Award (1998) and the Institute Silver Medal from the Indian Institute of Technology (1989). He is a co-author of the books: *Principles of Cognitive Radio* (Cambridge Univ. Press, Cambridge, U.K., 2012) and *Wireless Networks: Multiuser Detection in Cross-Layer Design* (Springer, New York, NY, USA, 2004). He has served as an Editor for IEEE COMMUNICATION LETTERS and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He has also served as a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATION Special Issues on Adaptive, Spectrum Agile and Cognitive Radio Networks (2007) and Game Theory in Communication Systems (2008). He currently serves as a Distinguished Lecturer of the IEEE.