COVERAGE AND AREA SPECTRAL EFFICIENCY IN DOWNLINK RANDOM CELLULAR NETWORKS WITH CHANNEL ESTIMATION ERROR

Yueping Wu[*], *Matthew R. McKay*[*], *Robert W. Heath Jr.*[†]

[*]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong

[†]Department of Electrical and Computer Engineering, The University of Texas at Austin, USA

ABSTRACT

We investigate the impact of channel estimation on the performance of downlink random cellular networks. First, we derive a new closed-form expression for the coverage probability under certain practical conditions. We show that the coverage probability is dependent on the user and base station (BS) densities solely through their ratio for arbitrary pilot-training length. Next, we derive the optimal pilot-training length that maximizes the area spectral efficiency (ASE) in several asymptotic regimes, and capture the dependence of this optimal length on the ratio between the user and BS densities. The ASE loss due to training is shown to be less significant in small cell networks with a larger base station density.

Index Terms— Channel estimation, downlink random cellular networks, coverage probability, area spectral efficiency

1. INTRODUCTION

Pilot-training is a practical way to obtain channel state information (CSI) at the receiver, which is essential to perform coherent data detection. Because the channels are estimated, errors in the estimation impact subsequent communication performance. The impact of channel estimation (CE) errors has been extensively investigated in different network architectures including simple network settings [1, 2] and cellular networks [3–6]. A limitation of prior works [3–6] is that a deterministic network configuration was configured. Unfortunately, deterministic networks do not capture the irregularity found in realistic cellular deployments.

Recently, a stochastic model for cellular networks was proposed in [7–11], where the locations of the base stations (BSs) are modeled as a Poisson point process (PPP). This model facilitates the use of mathematical tools from stochastic geometry [12] to derive network-wide performance measures. A main advantage of the stochastic geometry framework is that it permits the evaluation of closed-form network-wide performance measures. The model was also found to predict performance that was a good fit with measured BS locations [11], and has been widely used in other works [13–17]. Prior works, however, did not address the impact of CE error.

In this paper, we investigate the impact of CE on the performance of downlink random cellular networks, where the locations of BSs are modeled as a PPP and each user employs the minimum-mean-square-error (MMSE) estimator to obtain an estimate of the channel to its serving BS. First, we derive a new closed-form expression for the coverage probability under interference-limited condition. Then we show that the coverage probability is dependent on the user and BS densities solely through their ratio for arbitrary pilot-training length. Next, we derive the optimal pilot-training length that maximizes the area spectral efficiency (ASE) in several asymptotic regimes, and capture the dependence of this optimal length on the ratio between the user and BS densities. We show that the ASE loss due to training is less significant in small cell networks with a larger BS density. In our recent paper [18], we investigated the impact of CE in decentralized wireless ad hoc networks, where each transmitter has an intended receiver with a fixed distance away. In this paper, however, we consider cellular networks, and also vary the number of users each BS connects to and the distance between each BS and its user.

The work of Y. Wu and M. R. McKay was supported by the Hong Kong Research Grants Council under grant no. 616910. The work of R. Heath was supported in part by the National Science Foundation under grant no. 1218338. This work was carried out while Y. Wu was visiting UT Austin.

2. SYSTEM MODEL

We assume the locations of the BSs in the cellular networks are spatially distributed as a PPP Φ_b of density λ_b and transmit at power P. The locations of the users are also modeled by an independent PPP Φ_m of density λ_m. Due to Slivnyak's theorem [12], we conduct analysis on a typical mobile user located at the origin. We assume that each user connects to the nearest BS, i.e., the users in the Voronoi cell of a BS are associated with it. The transmitted signals are attenuated by a factor $\frac{1}{r^\alpha}$ with distance r where $\alpha > 2$ is the path loss exponent. At the physical layer, each user obtains an estimate of CSI to its serving BS via pilot-training symbols sent from this BS. The pilot-training symbols, which are initially known at both the BS and users in the same cell, are broadcasted by each BS to its connecting users. During the data transmission stage, we assume that the orthogonal multiple access technique, e.g., time devision multiple access (TDMA), is utilized within a cell, so that each BS serves one user at a time and no intra-cell interference exists.

We consider the block fading channel, where the transmitter-receiver channels are constant over a block comprising of W_cT_c channel uses, with one symbol per channel use, and evolves independently from block to block. The factor T_c corresponds to the coherence time, while the factor W_c refers to the coherence bandwidth. During each block, each transmitter sends a frame of length $L = W_cT_c$ (i.e., assumed equal to the block length, as in [1, 19]) to its corresponding receiver, which comprises of L_T pilot-training symbols, followed by $L - L_T$ data symbols. The pilot-training symbols are utilized by the user to obtain an estimate of the channel. This channel estimate is then used to detect the data symbols transmitted from the corresponding BS. This procedure is repeated over all subsequent frames.

Before describing the CE and data detection procedures in more detail, we first note that if the user density λ_m is not sufficiently large, some BSs may possibly connect to no user. In this case, these unconnected BSs do not transmit signals in the coming time slots

until the locations of the users change, and they have some users to serve. Consequently, we need to provide the transmission probability of each interfering BS, which is as follows.

Lemma 1. *The transmission probability of each interfering BS is*

$$p_t \approx 1 - \left(\frac{3.5}{3.5 + \frac{\lambda_m}{\lambda_b}} \right)^{3.5} . \quad (1)$$

Proof. See [20]. □

We observe from (1) that the transmission probability of each interfering BS increases with $\frac{\lambda_m}{\lambda_b}$. This indicates that for a sufficiently large user density λ_m (compared to the BS density λ_b), p_t approaches one, i.e., each interfering BS connects to at least one user. This is consistent with the intuition and is also the case considered in [13]. Note that the approximation in (1) has been shown to be quite accurate. As explained in [20], the active interfering BSs are approximated as a PPP Φ of density $\lambda \approx p_t \lambda_b$. Further, we see from (1) that when the density ratio $\frac{\lambda_m}{\lambda_b}$ is not sufficiently high, p_t is small, so that only a small fraction of the BSs are active. For example, when $\lambda_m = \lambda_b$, then $\lambda \approx 0.585\lambda_b$, i.e., only about half of the BSs are actually active.

2.1. Channel Estimation Stage

The $L_T \times 1$ baseband equivalent received vector $\mathbf{y}_0$ at the typical user, formed by concatenating the L_T received symbols during the first L_T channel uses, is given by

$$\mathbf{y}_0 = \sqrt{\frac{L_T P}{r_0^\alpha}} h_{00}\mathbf{t}_{00} + \sum_{B_\ell \in \Phi(\lambda)/B_0} \sqrt{\frac{P}{|D_{\ell 0}|^\alpha}} h_{\ell 0}\mathbf{x}_\ell + \mathbf{n}_0 \quad (2)$$

where[1] $h_{00} \overset{d}{\sim} \mathcal{CN}(0,1)$ is the channel between the typical user and its serving BS, $h_{\ell 0} \overset{d}{\sim} \mathcal{CN}(0,1)$ is the channel of the ℓth BS with respect to (w.r.t.) the typical user, r_0 is the distance between the typical user and its serving BS, B_ℓ is the location of the ℓth BS, B_0 denotes the location of the typical user's serving BS, $\mathbf{t}_{00}$ is a $L_T \times 1$ training symbol vector satisfying $\mathbf{t}_{00}^\dagger \mathbf{t}_{00} = 1$ [21], $\mathbf{x}_\ell \overset{d}{\sim} \mathcal{CN}_{L_T \times 1}(0, \mathbf{I}_{L_T})$ is a transmission vector[2] from node ℓ, and $\mathbf{n}_0 \overset{d}{\sim} \mathcal{CN}_{L_T \times 1}(0, N_0 \mathbf{I}_{L_T})$ is the additive white Gaussian noise (AWGN) vector. Note that h_{00}, $h_{\ell 0}$, $\mathbf{t}_{00}$, $\mathbf{x}_\ell$ and $\mathbf{n}_0$ are independent. The first step of CE is to assemble the observation scalar signal $y = \mathbf{t}_{00}^\dagger \mathbf{y}_0$ at the typical user as

$$y = \sqrt{\frac{L_T P}{r_0^\alpha}} h_{00} + \sum_{B_\ell \in \Phi(\lambda)/B_0} \sqrt{\frac{P}{|D_{\ell 0}|^\alpha}} h_{\ell 0} q_{\ell 0} + z_0 \quad (3)$$

where $q_{\ell 0} = \mathbf{t}_{00}^\dagger \mathbf{x}_\ell \overset{d}{\sim} \mathcal{CN}(0,1)$, and $z_0 = \mathbf{t}_{00}^\dagger \mathbf{n}_0 \overset{d}{\sim} \mathcal{CN}(0, N_0)$.

To obtain an estimate of the channel, we use the low-complexity linear MMSE estimator. According to standard MMSE estimation [22], conditioned on the distance between the typical user and its serving BS r_0, the estimate of h_{00} is $\widehat{h}_{00}|_{r_0} = \frac{\mathrm{E}[h_{00} y^*]}{\mathrm{E}[y y^*]} y = \frac{\sqrt{\frac{L_T P}{r_0^\alpha}}}{\frac{L_T P}{r_0^\alpha} + N_0 + P\mathrm{Var}(I)} y$, where $I = \sum_{B_\ell \in \Phi(\lambda)/B_0} \sqrt{\frac{1}{|D_{\ell 0}|^\alpha}} h_{\ell 0} q_{\ell 0}$ and $\mathrm{Var}(I)$ is calculated conditioning on r_0. According to Campbell's Theorem [12] and conditioning on r_0, $\mathrm{Var}(I)$ is calculated as

$$\mathrm{Var}(I) = 2\pi\lambda \int_{r_0}^{\infty} r^{1-\alpha} \mathrm{d}r = 2\pi\lambda \frac{r_0^{2-\alpha}}{\alpha - 2} . \quad (4)$$

The CE error can be expressed as $e_{00} = h_{00} - \widehat{h}_{00}$, where $e_{00} \overset{d}{\sim} \mathcal{CN}\left(0, \sigma_e^2\right)$, and $\widehat{h}_{00}$ and e_{00} are uncorrelated [2]. After some algebraic manipulations, the variance of the CE error, given that the distance between the typical user and its serving BS is r_0, is

$$\sigma_e^2|_{r_0} = \mathrm{E}\left[\left(h_{00} - \widehat{h}_{00}|_{r_0} \right)^2 \right] = \frac{1}{1 + \frac{L_T}{r_0^\alpha} \frac{1}{\frac{1}{\rho} + 2\pi\lambda \frac{r_0^{2-\alpha}}{\alpha - 2}}} \quad (5)$$

where $\rho = \frac{P}{N_0}$ is the transmit signal-to-noise ratio (SNR).

2.2. Data Transmission Stage

After the CE stage, the BS then sends data for the rest of the frame duration. The received signal at the typical receiver during the nth channel use, for $n = L_T + 1, \ldots, L$, is given by

$$d_0[n] = \sqrt{\frac{1}{r_0^\alpha}} \widehat{h}_{00} s_0[n] + \underbrace{\sqrt{\frac{1}{r_0^\alpha}} e_{00} s_0[n] + \sum_{B_\ell \in \Phi(\lambda)/B_0} \sqrt{\frac{1}{|D_\ell|^\alpha}} h_{\ell 0} s_\ell[n] + n_0[n]}_{\text{unknown at the receiver}} \quad (6)$$

where $s_0[n]$ and $s_\ell[n]$ are independent Gaussian distributed data symbols from the typical and the ℓth transmitting node, respectively, satisfying $\mathrm{E}\left[|s_0[n]|^2\right] = P$ and $\mathrm{E}\left[|s_\ell[n]|^2\right] = P$, and $n_0[n] \overset{d}{\sim} \mathcal{CN}(0, N_0)$ is AWGN. As the terms in (6) which are unknown at the receiver are treated as noise, an estimate of $s_0[n]$ is then formed as $\widehat{s}_0[n] = \sqrt{r_0^\alpha} \frac{\widehat{h}_{00}^*}{|\widehat{h}_{00}|^2} d_0[n]$, from which the signal-to-interference-plus-noise ratio (SINR) can be written as

$$\mathrm{SINR}_0 = \frac{\frac{\rho}{r_0^\alpha} |\widehat{h}_{00}|^2}{\sum_{B_\ell \in \Phi(\lambda)/B_0} \frac{\rho}{|D_\ell|^\alpha} |h_{\ell 0}|^2 + \frac{\rho}{r_0^\alpha} \sigma_e^2|_{r_0} + 1} . \quad (7)$$

3. PERFORMANCE ANALYSIS: COVERAGE PROBABILITY AND AREA SPECTRAL EFFICIENCY

In this section, we will investigate the impact of CE on the coverage probability and ASE.

3.1. Coverage Probability

The coverage probability is defined as the probability that the mutual information of the channel between the typical user and its serving BS is larger than a target rate R_{eff} data bits/node/channel use, and is given by

$$\begin{aligned} p_c(\beta, L_T) &:= \Pr\left(\left(1 - \frac{L_T}{L}\right) \frac{\log_2(1 + \mathrm{SINR})}{M} > R_{\text{eff}} \right) \\ &= \Pr\left(\left(1 - \frac{L_T}{L}\right) \frac{\log_2(1 + \mathrm{SINR})}{M} > \left(1 - \frac{L_T}{L}\right) \frac{R}{M} \right) \\ &= \Pr(\mathrm{SINR} > \beta) \end{aligned} \quad (8)$$

[1] The notation $X \overset{d}{\sim} Y$ means that X *is distributed as* Y.

[2] The Gaussian assumption for the interfering symbols is well justified. As we will show, the optimal pilot-training length is typically small compared to the frame length. In the optimal scenario, the majority of the frame is thus used for data transmission, as opposed to pilot training, during which the data symbols transmitted from all nodes are Gaussian distributed.

$$p_c\left(\beta, L_T\right) = 2\pi\lambda_b \int_0^\infty \exp\left(-2\pi\lambda\left(\frac{\pi r_0^2 \beta^{\frac{2}{\alpha}}}{\alpha \sin\left(\frac{2\pi}{\alpha}\right)\left(1-\sigma_e^2|_{r_0}\right)^{\frac{2}{\alpha}}} - \frac{1}{2} r_0^2\, {}_2F_1\left(1, \frac{2}{\alpha}; 1+\frac{2}{\alpha}; -\frac{1-\sigma_e^2|_{r_0}}{\beta}\right)\right)\right)$$
$$\times \exp\left(-\frac{\beta\left(\sigma_e^2|_{r_0} - \frac{r_0^\alpha}{\rho}\right)}{1-\sigma_e^2|_{r_0}}\right) r_0 \exp\left(-\lambda_b \pi r_0^2\right) \mathrm{d}r_0 \tag{9}$$

where M ($M \geq 1$) is the number of users connected to the typical user's serving BS, R is the rate in which the data is encoded at the transmitter over the data transmission stage, $\beta = 2^R - 1$ is the SINR operating value, and the second line of (8) follows by noting that $R_{\text{eff}} = \left(1 - \frac{L_T}{L}\right)\frac{R}{M}$, which is taken over the whole frame and the whole service duration. Note that the factor $(1 - \frac{L_T}{L})$ represents the fractional amount of time (relative to the total frame length) used for data transmission [18], and the factor $\frac{1}{M}$ represents the fractional amount of time used for serving the typical user (due to the use of TDMA). Based on (7), we first state the most general result for coverage probability, from which all other results in this section follow.

Theorem 1. *The coverage probability of a randomly selected user is given by (9) at the top of this page, where ${}_2F_1\left(a_1, a_2; b_1; z\right)$ is the Gaussian hypergeometric function [23, Eq. (15.1.1)].*

Proof. Follows by first conditioning on r_0 and averaging out the aggregate interference by utilizing the probability generating functional of a PPP [24]. Then (9) is obtained by getting rid of the condition on r_0. □

The integral in (9) is fairly easy to be evaluated numerically. In particular, it is much easier than doing the Monte Carlo simulation, where it is required to first select the serving BS for the typical user and calculate the aggregate interference from all the other active BSs over an infinite plane. Because solving the integral in closed-form appears intractable, to obtain more insights, we evaluate performance in asymptotic regimes. We first note that for cellular conditions, it is reasonable that $\lambda_b \ll 1$ (implying $\lambda \ll 1$). In this case, further considering the interference-limited scenario where the noise is negligible[3] (i.e., $\rho \to \infty$) and $\alpha = 4$, we have the following.

Corollary 1. *When $\alpha = 4$, $\lambda \ll 1$ and $\rho \to \infty$, the coverage probability in (9) is approximated as*

$$p_c\left(\beta, L_T\right) \approx \frac{\sqrt{2\pi L_T\left(1+\frac{1}{\beta}\right)}}{p_t} \exp\left(\eta^2\right) Q\left(\sqrt{2}\eta\right) \tag{10}$$

where $\eta = \sqrt{\frac{L_T\left(1+\frac{1}{\beta}\right)}{2}}\left(\frac{\sqrt{\beta}\left(\pi - 2\arctan\left(\frac{1}{\sqrt{\beta}}\right)\right)}{2} + \frac{\beta}{L_T} + \frac{1}{p_t}\right)$, and $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp\left(-\frac{t^2}{2}\right)\mathrm{d}t$ is the Q-function.

Proof. Omitted due to space limitations. □

We see from (10) that when $\rho \to \infty$, the dependence of $p_c\left(\beta, L_T\right)$ on λ_m and λ_b is solely through p_t, or equivalently, $\frac{\lambda_m}{\lambda_b}$. This is confirmed in Fig. 1, which plots the coverage probability $p_c(\beta, L_T)$ vs. the BS density λ_b for $\frac{\lambda_m}{\lambda_b} = 1$ by using (9).

[3]Note that the noise can be neglected (i.e., $N_0 \to 0$) in the cell interior because it is very small compared to the desired signal power, and also at the cell edge due to the much larger interference power [11].

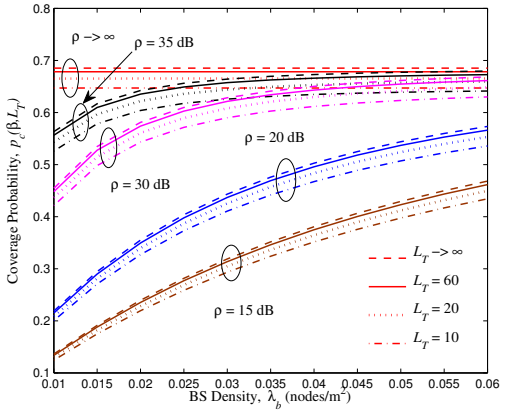

Fig. 1. Coverage probability $p_c\left(\beta, L_T\right)$ vs. BS density λ_b for different transmit SNR ρ and different pilot-training lengths L_T, and with $\alpha = 4$, $\beta = 0$ dB, and $\lambda_m = \lambda_b$.

We see from Fig. 1 that when ρ is very high, e.g., $\rho \geq 30$ dB, the trend of $p_c(\beta, L_T)$ going to be independent of λ_b is evident as ρ increases. This can be explained by noting that when $\frac{\lambda_m}{\lambda_b}$ is fixed, a larger λ_b leads to a larger aggregate interference (and also a worse CE quality), but meanwhile, it decreases the average distance between the typical user and its serving BS (which also implies a better CE quality), which is beneficial. When ρ is sufficiently high, these two competing factors are counter-balanced, which holds true for arbitrary L_T. Therefore, in this case, the coverage probability can not be improved by simply deploying more BSs. Instead, some interference mitigation techniques should be employed to improve the coverage. Note that in [11, Eq. (14)], it was shown that the coverage probability is independent of the BS density in the perfect CSI scenario ($L_T \to \infty$) when $\rho \to \infty$ and $p_t = 1$. This can be easily seen from (10) by letting $p_t = 1$. Fig. 1 also indicates the result in *Corollary 1* holds true for moderately large λ.

When ρ is low or moderately high, we see from Fig. 1 that a larger λ_b is beneficial for a higher coverage, even when $\frac{\lambda_m}{\lambda_b}$ is fixed. This can be explained by noting that when ρ is not sufficiently high, due to the existence of noise, the negative impact of more aggregate interference caused by a larger BS density becomes less significant. Therefore, the positive impact dominates the coverage probability. Moreover, we note from Fig. 1 that for a fixed ρ, to achieve a certain coverage improvement, less BSs are required to be deployed for a larger L_T.

3.2. Area Spectral Efficiency

The ASE is defined as the average total number of successfully transmitted data bits/channel use/unit area in the downlink. Taking into

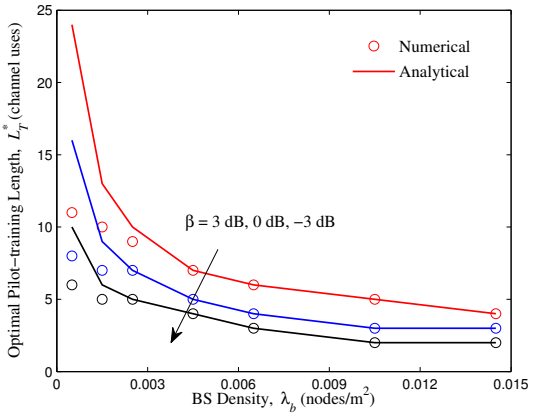

Fig. 2. Optimal pilot-training length L_T^* vs. BS density λ_b for different SINR thresholds β in the interference-limited scenario, and with $\alpha = 4$, $\lambda_m = 0.0001$ nodes/m^2, and $L = 1250$.

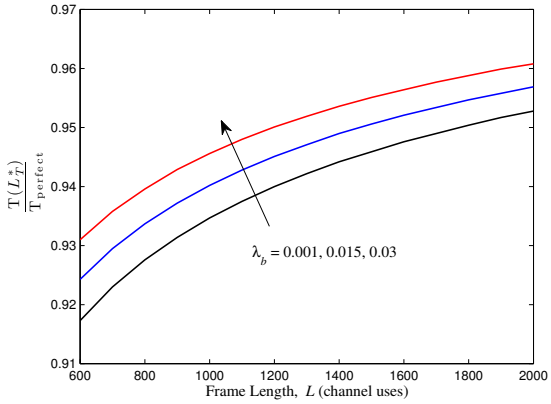

Fig. 3. $\frac{\mathrm{T}(\widetilde{L_T^*})}{\mathrm{T}_{\mathrm{perfect}}}$ vs. frame length L, and with $\alpha = 4$, $\lambda_m = 0.001$ nodes/m^2, $\rho = 15$ dB, $\alpha = 4$, and $\beta = 0$ dB.

account the effect of CE, the ASE is given by

$$\mathrm{T}\left(L_T\right) = R_{\mathrm{eff}}\left(L_T, R\right)\lambda p_c\left(\beta, L_T\right). \tag{11}$$

Substituting the coverage probability expression in (9) into (11), we obtain the ASE in general environment of the cellular networks.

From (11), we observe that increasing L_T has both a positive and a negative effect on the ASE. The positive effect occurs since the coverage probability $p_c\left(\beta, L_T\right)$ increases with L_T. The negative effect occurs since for a fixed frame length L, the time spent for data transmission decreases with L_T. A natural question then arises as to the optimal pilot-training length L_T^* which maximizes the ASE. To obtain some insights into the optimal fraction of training, we consider the asymptotic regime.

Theorem 2. *When $\alpha = 4$ and $\rho \to \infty$, if $\frac{\lambda_m}{\lambda_b}$ is sufficiently small, the optimal pilot-training length that maximizes the ASE is*

$$L_T^* = \begin{cases} \left\lfloor \widetilde{L_T^*} \right\rfloor & \text{if} \quad \mathrm{T}\left(\left\lceil \widetilde{L_T^*} \right\rceil\right) \leq \mathrm{T}\left(\left\lfloor \widetilde{L_T^*} \right\rfloor\right) \\ \left\lceil \widetilde{L_T^*} \right\rceil & \text{otherwise} \end{cases} \tag{12}$$

where[4] $\widetilde{L_T^*} = \sqrt{\frac{2\beta p_t L}{2 - p_t\sqrt{\beta}\left(\pi - 2\arctan\left(\frac{1}{\sqrt{\beta}}\right)\right)}}$.

Proof. Follows by first applying [23, Eq. (7.1.23)] to (10) to obtain an asymptotic coverage probability, and substituting it into (11) to obtain the asymptotic ASE. We then take Taylor series expansion on ASE for small p_t as described in [25], and take derivatives to calculate $\widetilde{L_T^*}$. □

From (12), we see that L_T^* increases with the frame length L, and in particular for large L, L_T^* scales as $O\left(\sqrt{L}\right)$. This implies that for large L, the fraction of the total frame length for CE, $\frac{L_T^*}{L}$, scales as $O\left(\frac{1}{\sqrt{L}}\right)$, which can be quite small. The practical interpretation is that for large L, it is preferable to dedicate a larger proportion of the frame for data transmission. Similar results were obtained in [26] for a time-division duplexing multiuser MIMO downlink scenario, and in [18] for an ad hoc network scenario.

Fig. 2 plots the optimal pilot-training length L_T^* vs. the BS density λ_b. We observe that the 'Analytical' curves plotted using (12) closely match the numerical results, which are obtained by numerically solving L_T^* based on (9), for sufficiently large λ_b (or equivalently, for sufficiently small p_t). Further, we observe that L_T^* decreases with λ_b, i.e., L_T^* increases with p_t, as predicted by (12). It is thus optimal to use a larger fraction of the frame for data transmission in the small cell network which has a relatively large BS density, e.g., femto cells. Note that although (12) is not very accurate for small λ_b, the 'Analytical' curves show the same trend as the numerical curves, as can be seen from Fig. 2. Moreover, for large cell networks with a small BS density, e.g., macro cells, the optimal pilot-training length can be calculated numerically based on (9).

We now compare the ASE when using L_T^* with an ideal (impractical) scenario where perfect CSI is obtained without the need for any training. In this ideal scenario, the ASE is given by $\mathrm{T}_{\mathrm{perfect}} = \log_2\left(1+\beta\right)\lambda p_c\left(\beta, \infty\right)$. For large L, it can be shown that

$$\frac{\mathrm{T}\left(\widetilde{L_T^*}\right)}{\mathrm{T}_{\mathrm{perfect}}} = 1 - 2\sqrt{\frac{2\beta p_t}{2 - p_t\sqrt{\beta}\left(\pi - 2\arctan\left(\frac{1}{\sqrt{\beta}}\right)\right)}}\sqrt{\frac{1}{L}} + O\left(\frac{1}{L}\right). \tag{13}$$

One key insight drawn from (13) is that for channels with a sufficiently long coherence time (i.e., the channel remains constant during the transmission of a packet of length L, whose value may be large), there is a negligible performance loss resulting from channel estimation compared to the perfect CSI scenario.

We also observe from (13) that $\frac{\mathrm{T}(\widetilde{L_T^*})}{\mathrm{T}_{\mathrm{perfect}}}$ decreases with the transmission probability p_t, or equivalently, increases with the BS density λ_b for a fixed λ_m. This is confirmed in Fig. 3, where the curves are plotted by numerically solving $\mathrm{T}\left(L_T^*\right)$ based on (9). The implication is that for small cell networks with a larger BS density, the ASE loss due to training is less significant.

[4]Note that $\lceil\cdot\rceil$ and $\lfloor\cdot\rfloor$ are the ceiling and floor functions respectively.

4. REFERENCES

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

[2] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, June 2010.

[3] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *To appear in IEEE J. Sel. Areas Commun.*

[4] C. K. Wen, "Performance analysis of MIMO cellular network with channel estimation errors," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3414–3424, Nov. 2010.

[5] M. Biguesh and A. B. Gershman, "Downlink channel estimation in cellular systems with antenna arrays at base stations using channel probing with feedback," *EURASIP Journal on Applied Signal Processing*, pp. 1330–1339, Sept. 2004.

[6] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath, "Channel estimation and linear precoding in multiuser multiple-antenna TDD systems," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2102–2116, June 2011.

[7] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev, "Stochastic geometry and architecture of communication networks," *J. Telecommunication Systems*, vol. 7, no. 1, pp. 209–227, 1997.

[8] F. Baccelli and S. Zuyev, "Stochastic geometry models of mobile communication networks," in *Frontiers in queueing*, Boca Raton, FL: CRC Press, 1997, pp. 227–243.

[9] T. X. Brown, "Cellular performance bounds via shotgun cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2443–2455, Nov. 2000.

[10] P Madhusudhanan, J. G. Restrepo, Y. Liu, and T. X. Brown, "Carrier to interference ratio analysis for the shotgun cellular system," in *Proc. of IEEE Global Telecommunications Conference (GLOBECOM)*, Honolulu, Hawaii, Dec. 2009, pp. 1–6.

[11] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[12] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, John Wiley and Sons, England, 2nd edition, 1995.

[13] S. Akoum and R. W. Heath Jr, "Interference coordination: Random clustering and adaptive limited feedback," *Submitted to IEEE Trans. Signal Process.*, 2012.

[14] R. W. Heath Jr. and M. Kountouris, "Modeling heterogeneous network interference," in *Proc. of Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2012, pp. 17–22.

[15] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.

[16] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *Submitted to IEEE Trans. on Wireless. Commun.*

[17] P. Xia, H. S. Jo, and J. G. Andrews, "Fundamentals of inter-cell overhead signaling in heterogeneous cellular networks," *IEEE Journal on Special Topics in Signal Processing*, vol. 6, no. 3, pp. 257–269, June 2012.

[18] Y. Wu, R. H. Y. Louie, and M. R. McKay, "Analysis and design of wireless ad hoc networks with channel estimation errors," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1447–1459, Mar. 2013.

[19] M. Kobayashi, G. Caire, and N. Jindal, "How much training and feedback are needed in MIMO broadcast channels?," in *Proc. of IEEE Int. Symp. Inform. Theory (ISIT)*, Toronto, Canada, July 2008, pp. 2663–2667.

[20] S. Lee and K. Huang, "Coverage and economy of cellular networks with many base stations," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1038–1040, July 2012.

[21] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, Apr. 2006.

[22] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Academic Press, New York, 2nd edition, 1981.

[23] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, 9th edition, 1970.

[24] M. Haenggi and R. K. Ganti, "Interference in large wireless networks," *Foundations and Trends in Networking*, vol. 3, no. 2, pp. 127–248, 2009.

[25] N. Jindal and A. Lozano, "A unified treatment of optimum pilot overhead in multipath fading channels," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2939–2948, Oct. 2010.

[26] M. Kobayashi, N. Jindal, and G. Caire, "Training and feedback optimization for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2228–2240, Aug. 2011.