User Profiling: A Method for Limited Feedback in OFDMA Systems

Vinay Majjigi, Rajiv Agarwal, and John Cioffi
{vmajjigi,rajivag,cioffi@stanford.edu}

***Abstract*—In the OFDMA downlink, obtaining Channel State Information (CSI) from users is necessary for the Base Station (BS) to optimize network performance by intelligently allocating resources and scheduling mobile stations (MS). However, the overhead of obtaining CSI is a large burden and therefore, schemes to reduce CSI are necessary for a realizable system. By considering a MS's coherence time Δt_{coh} and coherence bandwidth B_{coh}, and exploiting this redundancy in the MS's CSI, the amount of feedback can be tailored to the specific user's profile and greatly decreased. Typically, reducing feedback results in more uncertainty in CSIT and performance degradation. However, in the proposed scheme, the CSIT's deviation from the true channel state is bounded, and thus can provide robust scheduling. Specifically, two BS schemes are suggested that either fix the average BER or the average outage probability regardless of user mobility or delay spread.**

I. INTRODUCTION

Current and future wireless technologies are required to support high-data rate applications with stringent Quality of Service (QoS) constraints. In Metropolitan Area Networks (MAN) such as WiMAX, users are expected to have widely varying channel conditions that make reliable transmission difficult. To avoid a loss in system performance, the Base Station (BS) acquires channel state information (CSI) from the users and then allocates resources and schedules users based on some performance criterion. Considerable research efforts have focused on resource allocation when the transmitter has acquired all the CSI from each user and its entire spectrum known as full CSIT [1],[2],[3]. More recently, efforts have focused on different forms of partial CSIT. The partial CSIT research vary in scope with the common goal being to reduce the number of bits being fed back and aid in resource allocation [4],[5],[6]. However, many of the feedback schemes rely only on users' instantaneous CSI, and ignore longer term statistics and structure of the CSI.

Users in a MAN are expected to have large variations in mobilities and channel delay spreads which implies widely varying channel coherence time Δt_{c} and bandwidth B_{coh}. Δt_{c} varies inversely with velocity, and B_{coh} varies inversely with delay spread. In the proposed work, this redundancy in CSI is investigated as a means to reduce a user's feedback load.

A stationary user terminal has a large Δt_{c}, while a fast moving will have a short Δt_{c}. Intuitively, the stationary user should feedback its CSI much less often than the fast user. Likewise, a user in an urban environment will experience a large channel delay spread, as compared with a rural user. These differences result in a channel response that is frequency-selective or frequency-flat, respectively. By allowing the user to estimate its channel coherence time and bandwidth, it can vary the amount of feedback information to reflect the amount of variation in its channel.

By adapting the feedback, mobile stations (MS) reduce their uplink transmissions, and therefore save transmission power and increase spectral efficiency. For a MS, conserving battery power is an important criteria, therefore reducing uplink transmissions is a practical design objective.

In this paper, two ideas are presented:

(1) A MS-centric user profiling algorithm is proposed that reduces the necessary bits of feedback and the frequency of uplink transmissions by adapting to the channel's coherence time and bandwidth. In doing so, the MS ensures a bounded statistical correlation between the true channel state and the BS's CSIT.

(2) With the CSIT's variation to the true CSI bounded, two transmission schemes are suggested that either provide a fixed average BER or a fixed average outage probability regardless of user velocity or delay spread.

In essence, the proposed algorithm minimizes the feedback necessary for different environments while guaranteeing the accuracy of the CSIT. This allows the users to feedback robust CSI and allow accurate CSIT-based scheduling in widely varying channel conditions. Simulation results are presented to validate the performance bounds as well as compare to a non-adaptive feedback scheme.

II. SYSTEM MODEL

In this paper, an OFDMA downlink system is considered. In the frequency domain, there are G tones (subcarriers) separated by f_s Hz. In the time domain, the smallest quanta is a frame of T_s seconds. For a single user, the entire channel spectrum undergoes correlated frequency-selective fading, but an individual tone is considered to undergo frequency-flat fading. Likewise, quasi-stationary fading is assumed where time correlation exists between multiple frames, but individual frames have a constant channel value. We denote the channel value on any individual tone i and frame index j as $h(i,j)$. It is assumed to be Rayleigh distributed, so that the channel power-gains $\gamma(i,j) = |h(i,j)|^2$ are exponential random variables. The fading is correlated both in time and frequency.

The proposed algorithm is not specific to a particular correlation model for fading, but specific correlation functions are mentioned for clarity. Jakes model is often used to characterize time correlation and results in the autocorrelation function being the zeroth-order Bessel function of the first kind $A_C(\Delta t) = J_o(2\pi f_D \Delta t)$, where f_D is the maximum Doppler shift, and Δt is the time separation [7]. The autocorrelation in frequency is related to the delay spread of the channel τ_{rms}. The delay spread is often modeled as an exponential decaying channel response, and thus the autocorrelation function is

$A_C(\Delta f) = 1/(1+4\pi^2\tau_{rms}^2\Delta f^2)$, where Δf is the frequency separation [8].

Borrowing from the WiMAX TDD nomenclature, a frame consists of an uplink period used for feedback of CSIT, followed by a downlink period used for data transmission. Full CSIT is defined as obtaining channel gain information for all G tones every frame. Partial CSIT is obtained by grouping L tones into a subchannel, and W frames into a time-block. This creates a subchannel-time block that has L contiguous tones and W contiguous frames. At the beginning of the time-block, the MS sends a single channel gain value γ_{SB}^l for the l^{th} **S**ubchannel-time **B**lock to the BS. The BS uses γ_{SB}^l as the channel gain for the entire subchannel-time block, therefore the reduction in CSIT is $\frac{1}{LW}$.

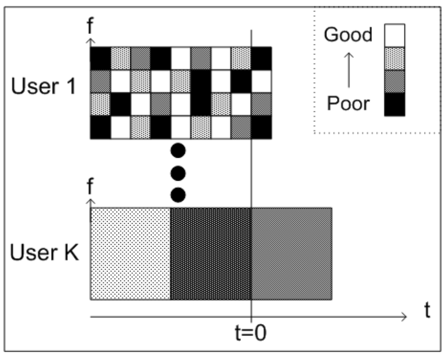

Fig. 1. Example of CSIT at BS for different users. The shade of a block represents its channel gain value.

This feedback reduction is different for users because they experience different channel environments. In general, different sized subchannel-time blocks may be used by users, as visualized in Figure 1. L and W depend on the long-term statistics of the user, which is assumed to remain the same over many subchannel-time blocks. Hence, feedback of L and W is infrequent and this additional feedback is negligible.

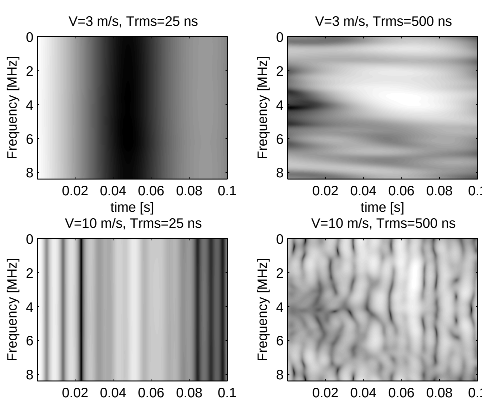

Fig. 2. Example of CSI for different user velocities and delay spreads. The dark pixels represent deep fades, likewise the white pixels are strong channel states.

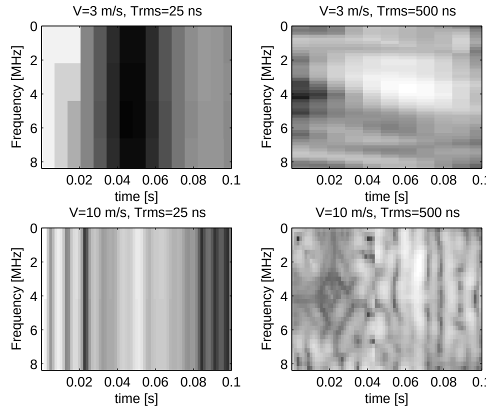

Fig. 3. Example of CSIT based on the channel states in Figure 2 after the user profiling algorithm. The dark pixels represent deep fades, likewise the white pixels are strong channel states.

III. Proposed Scheme

In this section, the user profiling algorithm is described. Then, the statistical correlation between the CSIT and the true channel state is bounded. Finally, two BS transmit schemes are presented that guarantee: fixed $\overline{\text{BER}}$ and fixed $\overline{P_{out}}$.

A. User Profiling Algorithm

The user profiling algorithm is MS-centric with only an infrequent communication to the BS to specify CSI subchannel-time block sizes and then periodic CSIT updates.

(1) MS estimates B_{coh} and Δt_{coh} using past CSIR values
(2) MS determines subchannel-time block size, i.e. L and W
(3) MS informs BS of L and W
(4) MS estimates γ_{SB}^l on $\frac{G}{L}$ subchannels
(5) MS transmits γ_{SB}^l for $\frac{G}{L}$ subchannels every WT_s seconds
(6) BS assigns modulation level to subchannel-time block based on γ_{SB}^l

Each MS estimates B_{coh} and Δt_{coh} by correlating their channel states against frequency and time, respectively. This requires the MS to use past channel state values to determine the coherence structure of the CSI. Let N denote the number of frames of past CSIR values that the MS uses to compute B_{coh} and Δt_{coh}. Assuming N is chosen large enough so that the time average approaches the ensemble mean of the process, B_{coh} is given as,

$$B_{coh} := \arg\max_f \frac{E[\gamma(i,j)\gamma(i+\lceil \frac{f}{f_s} \rceil, j)]}{E[|\gamma(i,j)|^2]} \geq \hat{\rho} \qquad (1)$$

$0 \leq \hat{\rho} \leq 1$ is a correlation threshold design parameter. Likewise Δt_{coh} is given as,

$$\Delta t_{coh} := \arg\max_t \frac{E[\gamma(i,j)\gamma(i, j+\lceil \frac{t}{T_s} \rceil)]}{E[|\gamma(i,j)|^2]} \geq \hat{\rho} \qquad (2)$$

In [8, §3.3.2], the definition of B_{coh} and Δt_{coh} uses $\hat{\rho} > 0.9$.

After estimating B_{coh} and Δt_{coh}, each user designates a size for its subchannel-time blocks to feedback to the BS. The subchannel block size is given as the maximum number of tones within B_{coh}, i.e. $L = \lfloor B_{coh}/f_s \rfloor$ tones and the periodicity of feedback is $W = \lfloor \Delta t_{coh}/T_s \rfloor$ frames. For clarity, integer number of subchannels are being ignored in the derivations but can be easily incorporated. For the purpose of illustration, the channel states of four different user mobilities and delay spreads are simulated in Figure 2. The corresponding output of the user profiling algorithm is shown in Figure 3, note the similarities between the four images. The user's mobility and environment are expected to be changing far slower than WT_s, therefore B_{coh} and Δt_{coh} are assumed constant for many subchannel-time blocks.

With L and W determined, let Γ^l denoted the channel gains for the l^{th} subchannel-time block,

$$\Gamma^l = \begin{pmatrix} \gamma^l_{1,1} & \gamma^l_{1,2} & \cdots & \gamma^l_{1,W} \\ \gamma^l_{2,1} & \gamma^l_{2,2} & \cdots & \gamma^l_{2,W} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma^l_{L,1} & \gamma^l_{L,2} & \cdots & \gamma^l_{L,W} \end{pmatrix} \tag{3}$$

The MS estimates the overall CSI for subchannel l with the middle tone at the beginning of the subchannel time-block, i.e. $\gamma^l_{\text{SB}} = \gamma^l_{L/2,1}$. The MS transmits γ^l_{SB} for each subchannel to the BS.

The MS and BS will agree on a transmission scheme for the subchannel-time block for the next WT_s seconds. After this time, the MS will update the BS with new CSI. The subchannel-time block forms a larger feedback quanta than individual tones and frames, and thus a reduction in feedback load.

B. Bound on CSIT's Correlation with True CSI

In a subchannel-time block, the correlation of the channel gains between any two tones for a given frame is at least $\hat{\rho}$, likewise the channel gains at two different frames for the same tone is at least $\hat{\rho}$ from Eqn (1) and Eqn (2). Under this condition, the correlation between any two channel gains within a subchannel-time block is $\rho > \hat{\rho}^2$. To show this, consider Γ^l, specifically the four corners of the subchannel-time block $\boldsymbol{\gamma} = [\gamma_{1,1}, \gamma_{L,1}, \gamma_{1,W}, \gamma_{L,W}]^T$. Assuming a monotonic autocorrelation function[1], by showing the correlation between any of these four corners has a minimum correlation, all points within the subchannel-time block will also have a minimum correlation. The covariance matrix $\Sigma_{\boldsymbol{\gamma}}$ normalized by the variance is,

$$\Sigma_{\boldsymbol{\gamma}} = \begin{pmatrix} 1 & \hat{\rho} & \hat{\rho} & \rho \\ \hat{\rho} & 1 & \rho & \hat{\rho} \\ \hat{\rho} & \rho & 1 & \hat{\rho} \\ \rho & \hat{\rho} & \hat{\rho} & 1 \end{pmatrix} \tag{4}$$

A covariance matrix must be positive definite, i.e. $|\Sigma_{\boldsymbol{\gamma}}| > 0$, so there is an implicit constraint on ρ. While the algebra is not shown here, this condition forces $\rho > \hat{\rho}^2$.

[1] While Jakes Model is not strictly monotonically decreasing, over practical ranges of $\hat{\rho}$, e.g. $\hat{\rho} > 0.5$, it is monotonically decreasing.

C. Fixed $\overline{BER}$ Adaptive Rate

The BS assigns the MS's subchannel-time block l a constellation size $M(\gamma^l_{\text{SB}})$ that corresponds to the the spectral efficiency of continuous M-QAM,

$$M(\gamma^l_{\text{SB}}) = 1 + \frac{3\gamma^l_{\text{SB}}}{2K_0} \tag{5}$$

Where $K_0 = -\ln(5\text{BER}_0)$ and BER_0 is the target bit error rate for perfect CSIT [9].

For continuous rate M-QAM constellations with two-dimension Gray coding in AWGN, the BER is well approximated by

$$\text{BER}(M, \gamma) \simeq 0.2 \text{ exp } \left(-\frac{3\gamma}{2(M-1)}\right) \tag{6}$$

Indeed, Eqn (6) is an upper bound to the true BER for $M \geq 4$ and BER $\leq 10^{-2}$ [10]. The BS sets the constellation size $M(\gamma^l_{\text{SB}})$ as given in Eqn (5) for each tone and frame in that subchannel-time block. Under perfect CSIT, the constellation the BS assigns exactly corresponds to the true channel gain, and therefore Eqn (6) specifies the BER for each tone and frame. Ignoring the subchannel-block indexing l for clarity, with partial CSIT, the true channel gain γ_{ij} will differ from γ_{SB}, and results in a mismatch between the rate assigned by the BS and true channel state, resulting in a $\text{BER}_{ij} = \text{BER}(\gamma_{\text{SB}}, \gamma_{ij})$. Where

$$\text{BER}(\gamma_{\text{SB}}, \gamma_{ij}) = 0.2(5\text{BER}_0)^{\gamma_{ij}/\gamma_{\text{SB}}} \tag{7}$$

In the proposed scheme, γ_{SB} is correlated to γ_{ij} by at least ρ for each i and j in the corresponding subchannel-time block. To develop an upper-bound on the average BER over the different tones and frames, $\overline{\text{BER}}$, we assume that each γ_{ij} is only correlated to γ_{SB} by ρ.

To find $\overline{\text{BER}}$, a short derivation is shown below, whereas the full derivation is found in [9, Eq (39)-(49)]. First, an expression for $\text{BER}(\gamma_{\text{SB}})$,

$$\text{BER}(\gamma_{\text{SB}}) = \int_0^\infty \text{BER}(\gamma_{\text{SB}}, \gamma_{ij}|\gamma_{\text{SB}}) p_{\gamma_{ij}|\gamma_{\text{SB}}}(\gamma_{ij}|\gamma_{\text{SB}}) d\gamma_{ij} \tag{8}$$

where $p_{\gamma_{ij}|\gamma_{\text{SB}}}(\gamma_{ij}|\gamma_{\text{SB}})$ is the conditional pdf derived from the fading model, and specified in [9, Eq (41)]. Finally, $\overline{\text{BER}}$ is found by the following,

$$\overline{BER} = \int_{\gamma_1}^\infty \text{BER}(\gamma_{\text{SB}}) p_{\gamma_{\text{SB}}}(\gamma_{\text{SB}}) d\gamma_{\text{SB}} \tag{9}$$

where $\gamma_1 = [\text{erfc}^{-1}(2\text{BER}_0)]^2$ and $p_{\gamma_{\text{SB}}}(\gamma_{\text{SB}})$ is given by an exponential distribution for Rayleigh fading. This leads to the $\overline{\text{BER}}$ bound that this scheme guarantees.

$$\overline{BER} = 0.2(1-\rho)K_0 \int_{u_1}^1 \frac{u}{(1-u)^2} \text{ exp} \left(-\frac{K_0 u(1-\rho u)}{1-u}\right) du \tag{10}$$

Where $u_1 = \gamma_1/(\gamma_1 + (1-\rho)K_0\bar{\gamma})$ and $\bar{\gamma}$ is the average channel gain.

It is shown that $\overline{\text{BER}}$ is constant as given in Eqn (10) for a given ρ, and ρ is bounded in Section III-B. Therefore, the proposed scheme can maintain the same average BER performance for widely varying channel and environmental conditions.

In a multiuser scenario, non-overlapping subchannel-time blocks will be assigned to different users. However, the above analysis requires the channel gains within a subchannel-time block to be correlated to within ρ, and is not user specific. While specific scheduling algorithms are not mentioned in this paper, the proposed algorithm only requires $\hat{\rho}$ and BER_0 to be specified, and then guarantees a constant $\overline{\text{BER}}$ regardless of velocity or delay spread. By varying the feedback load for different users based on their channel conditions, their $\overline{\text{BER}}$ can be kept constant regardless of the BS's scheduling algorithm.

D. Fixed $\overline{P_{out}}$ Adaptive Rate

In extreme fading conditions, the BS will declare outage when the channel enters a deep fade given by a threshold γ_{th}. In this case, the BS transmits the constellation $M(\gamma_{\text{SB}})$ if $\gamma_{\text{SB}} \geq \gamma_{th}$ to the subchannel-time block, and declares outage and stops transmitting when $\gamma_{\text{SB}} < \gamma_{th}$. If CSIT is assumed perfect, the outage probability for Rayleigh fading is given as,

$$P_{out} = 1 - \exp(-\gamma_{th}/\bar{\gamma}) \tag{11}$$

However, the average outage probability $\overline{P_{out}}$ over the different tones and frames will deviate from this nominal value because the true channel gain γ_{ij} will differ from γ_{SB}. This, for example, would occur when the BS has CSIT that is incorrectly low, as the BS would suspend transmission even though the channel could support a non-zero rate, resulting in additional outage. Let the additional outage due to incorrect CSIT be P_{out}^{CSIT}, while P_{out}, as defined in Eqn (11) is simply the outage due to the channel entering a fade. Similar to Section III-C, the bounded correlation between true CSI and the CSIT results in a bounded P_{out}^{CSIT}. Using a similar technique to [9, Eq (43)-(45)], it can be shown that P_{out}^{CSIT} is given as,

$$P_{out}^{\text{CSIT}} = \int_{\gamma_{th}}^{\infty} \left(1 - Q_1\left(\frac{\rho\gamma}{(1-\rho)\bar{\gamma}}, \gamma_{th}\right)\right) \frac{1}{\bar{\gamma}} e^{(-\gamma/\bar{\gamma})} d\gamma \tag{12}$$

Where $Q_1(...)$ is the generalized Marcum Q-function of order 1 [11]. Similar to Section III-C, by guaranteeing the CSIT's correlation to the true channel state, a fixed P_{out}^{CSIT} is maintained under varying channel conditions.

The average outage probability is simply the sum of the individual outage probabilities $\overline{P_{out}} = P_{out} + P_{out}^{\text{CSIT}}$. P_{out} is a only a function of the channel and is not affected by the feedback mechanism. Therefore, to minimize the outage probability, a robust feedback scheme is required to minimize P_{out}^{CSIT}. As shown in Section IV, the proposed scheme has significant advantages to non-adaptive schemes, one of which is bounding P_{out}^{CSIT}, and likewise $\overline{P_{out}}$, in widely varying channel environments.

IV. Simulation Results

The simulation results serve two purposes. First, they support the analytical results provided in Section III that bound the average BER and average outage probability. Second, the results show the advantage of an adaptive subchannel-time block size versus a naive scheme that fixes the subchannel-time block regardless of MS velocity or delay spread. The results presented are bounds on: $\overline{\text{BER}}$ and P_{out}^{CSIT}, and feedback overhead: number of feedback bits and frequency of uplink transmissions.

The results presented are Monte Carlo simulations that are compared to the analytic bounds Eqn (10) and Eqn (12). The common simulation parameters are listed in Table I. The fading is correlated Rayleigh fading, where the correlation in time is given by Jakes model $A_C(\Delta t) = J_o(2\pi f_D \Delta t)$, and the frequency correlation is derived from an exponentially decaying delay spread, and given by $A_C(\Delta f) = 1/(1 + 4\pi^2\tau_{rms}^2\Delta f^2)$. In the presented figures, 100 simulation runs were conducted.

TABLE I
SIMULATION PARAMETERS

f_c	3 GHz
Bandwidth	8.4 MHz
G	512
f_s	16.4 kHz
$\hat{\rho}$	0.9
BER_0	10^{-5}
L *(Naive)*	6
W *(Naive)*	13

For the practical range of MS velocities used in our simulation, we found that the bound $\rho > \hat{\rho}^2$ was too conservative, and the true CSI within the subchannel-time block is very close to the CSIT. Hence, in Figure 4 and Figure 5, Eqn (10) and Eqn (12) are plotted respectively, using $\rho = \hat{\rho}$ instead of $\rho = \hat{\rho}^2$.

In Figure 4, the simulation results support the upper bound on $\overline{\text{BER}}$ for the proposed scheme. As is shown, the proposed scheme is able to maintain a relatively constant $\overline{\text{BER}}$ regardless of velocity or delay spread. In contrast, the naive scheme clearly suffers at high velocity. The average channel gain is $\overline{\gamma} = 20$ dB.

In Figure 5, P_{out}^{CSIT} for the proposed scheme is shown to be below the bound. At low velocities, the channel coherence time is large and the naive scheme performs well, but it does not adapt to shorter coherence times. The proposed scheme maintains a nearly constant P_{out}^{CSIT} regardless of delay spread or velocity. The average channel gain is $\overline{\gamma} = 10$ dB, and the threshold is $\gamma_{th} = 0$ dB that implies a nominal outage probability of $P_{out} = 0.095$.

By adapting to the MS and its environment, the algorithm varies the subchannel-time block to match the channel's coherence structure. This has the effect of increasing feedback as the MS velocity increases as evidenced by Figure 6 and Figure 7. In Figure 6, the curves for the proposed scheme are close together because the frequency of uplink transmissions does not depend on the delay spread. In Figure 7, as the delay spread decreases, the user experiences flatter fading and thus can afford to feedback less bits as evidenced in Figure 7. The curve for feedback overhead for full CSIT is also shown.

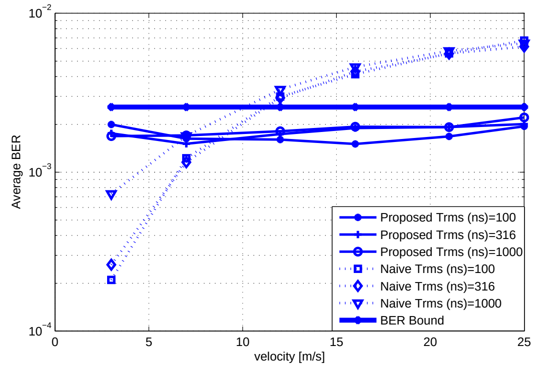

Fig. 4. Bounded $\overline{\text{BER}}$ for partial CSIT

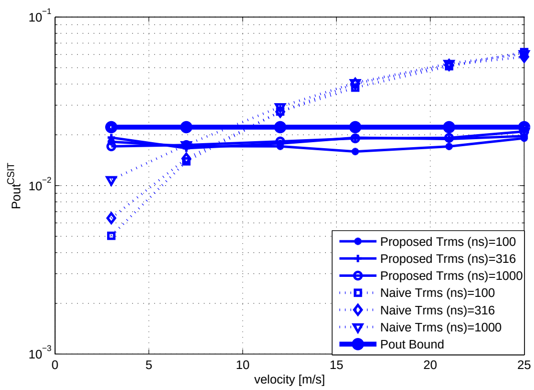

Fig. 5. Bounded P_{out}^{CSIT} for partial CSIT

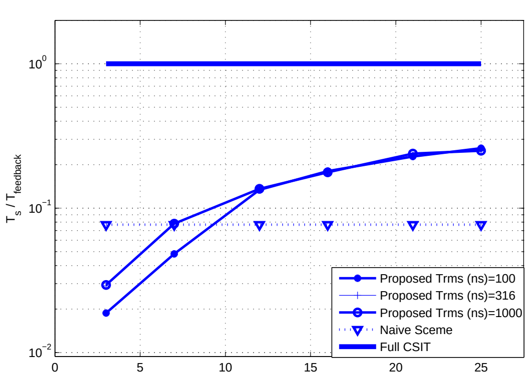

Fig. 6. Feedback frequency for varying velocities

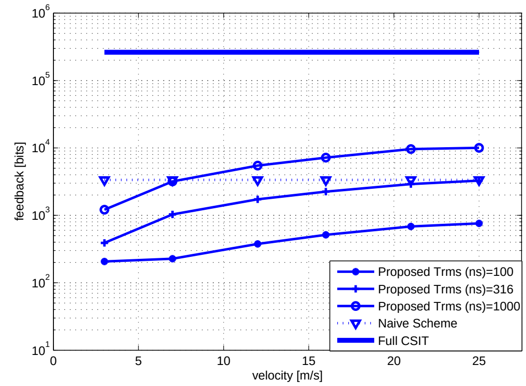

Fig. 7. Feedback bits for varying velocities

V. Conclusion

This paper proposed a new method to quantize CSIT for the OFDMA downlink. By considering a user's coherence time and bandwidth to determine subchannel block size and feedback periodicity, the amount of feedback bits as well as uplink transmissions can be substantially reduced saving both bandwidth and power. By adapting the subchannel-time block size to the channel conditions, the deviation of the CSIT to the true channel states is bounded leading to robust transmit schemes. That is, the BS can make accurate scheduling decisions regardless of user mobility and delay spread. Lastly, through numerical simulations, the proposed algorithm is shown to adapt to user's conditions better than fixed subchannel-time block schemes that are often discussed in industry and literature.

References

[1] K. S. M. Mohseni and J. M. Cioffi, "Optimal resource allocation for ofdma downlink systems," in *IEEE ISIT*, 2006.

[2] Ahmed K. F. Khattab and Khaled M. F. Elsayed, "Opportunistic scheduling of delay sensitive traffic in OFDMA-based wireless networks," *IEEE Computer Society*, 2006.

[3] C. Wong, R. Cheng, K. Letaief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, pp. 1747–1758, Oct. 1999.

[4] J. Chen, R. A. Berry, and M. L. Honig, "Performance of limited feedback schemes for downlink ofdma with finite coherence time," in *IEEE ISIT*, 2007.

[5] R.Agarwal, R. Vannithamby, and J. Cioffi, "Optimal allocation of feedback bits for downlink OFDMA systems," in *IEEE ISIT*, 2008.

[6] V. K. N. Lau and T. Wu, "Optimal transmission and limited feedback design for OFDM/MIMO systems in frequency selective block fading channels," *IEEE Trans. on Wireless Comm.*, vol. 6, pp. 1569–1573, May 2007.

[7] W. Jakes, *Microwave Mobile Communications*. IEEE Press, 1994.

[8] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[9] M.-S. Alouini and A. Goldsmith, "Adaptive modulation over nakagami fading channels," *Wireless Personal Communications*, vol. 13, pp. 119–143, May 2000.

[10] A. Goldsmith and S. Chua, "Variable-rate variable power m-qam for fading channels," *IEEE Trans. Commun.*, vol. COM-45, pp. 1218–1230, October 1997.

[11] N. Temme, *Special Functions-An Introduction to the Classical Functions of Mathematical Physics*. John Wiley and Sons, 1996.