Base Station-Side Rate Estimation for Threshold-Based Feedback, and Design Implications in Multi-User OFDM Systems

Vineeth Kumar, *Student Member, IEEE*, and Neelesh B. Mehta, *Senior Member, IEEE*

***Abstract*— Rate adaptation and scheduling are essential in ensuring that contemporary orthogonal frequency division multiplexing systems achieve high downlink spectral efficiencies. They depend upon reduced feedback schemes to efficiently feedback channel state information from the users to the base station (BS). In the popular threshold-based quantized feedback scheme, a user feeds back to the BS the quantized value of the signal-to-noise ratio for each subchannel. For this scheme, we derive a novel, throughput-optimal discrete rate adaptation (TORA) policy, which enables a system designer to reduce the feedback overhead. We present it in closed form for different multi-antenna diversity modes for the exponentially correlated subchannel gains model. We also develop a computationally simpler suboptimal variant of it. We derive an insightful lower bound for the fading- and user location-averaged throughput gain achieved by TORA over conventional rate adaptation for 1-bit feedback. We present extensive results to benchmark the system-level performance of TORA for different numbers of feedback bits and modulation and coding schemes available at the BS, and various schedulers, quantizers, and multi-antenna modes.**

***Index Terms*— OFDM, discrete rate adaptation, quantized feedback, scheduling, correlation.**

I. Introduction

ORTHOGONAL frequency division multiplexing (OFDM) is the physical layer technology of choice in contemporary wideband communication standards such as 4G Long Term Evolution (LTE), IEEE 802.11 wireless local area networks (WLAN), and IEEE 802.16 WiMAX [1], [2]. In it, the system bandwidth is divided into several narrow bandwidth orthogonal subcarriers, which are aggregated to form subchannels (SCs). For example, in LTE, twelve adjacent subcarriers are grouped together into a physical resource block (PRB), which has a bandwidth of 180 kHz [2].

In order to achieve high spectral efficiencies, these standards employ rate adaptation and scheduling, which exploit the different channel fades seen by different users and also the time-variations in them. In scheduling, the base station (BS) determines which user to allocate to each SC. In rate adaptation, the BS determines the modulation and coding scheme (MCS) to be used for transmission to the scheduled user for each SC. To facilitate this, the rate of each SC for each user ideally needs to be fed back to the BS. Such feedback is needed not just in frequency-division duplexing systems but also in time division duplexing systems, when the uplink and downlink interferences are asymmetric.

Even in current generation OFDM systems, the uplink bandwidth available for feedback is scarce and cannot support the feeding back of all the above channel state information (CSI). In order to significantly reduce the feedback overhead, several reduced or partial CSI feedback schemes have been proposed in the literature and adopted in standards such as LTE [2]. These include best-m feedback [3]–[6], threshold-based quantized feedback [4], [7]–[12], SC clustering [13], and opportunistic splitting [14]. For example, in the best-m scheme, each user reports the m SCs that can support the highest rates among all the SCs and the corresponding rates. In threshold-based quantized feedback, each user feeds back a quantized value of the signal-to-noise-ratio (SNR) of each SC. In the SC clustering scheme, non-overlapping clusters of SCs are formed and one bit is fed back per SC cluster [13].

LTE employs a multitude of feedback schemes that are adaptations and combinations of the above schemes. Specifically, in *wideband feedback*, only one 4-bit wideband channel quality indicator (CQI) value is fed back for the entire system bandwidth. In *eNodeB-configured sub-band feedback*, multiple PRBs are grouped into sub-bands, and CQI is fed back for each sub-band. In *UE-selected sub-band feedback*, only the CQIs of a few dynamically selected sub-bands are fed back [2, Ch. 10]. Thus, the feedback overhead per PRB is much less than 4 bits.

A. Literature on Threshold-Based Quantized Feedback

Threshold-based quantized feedback schemes have been widely studied in the literature given their simplicity and effectiveness, and are the focus of this paper. A comprehensive survey of the literature on threshold-based quantized feedback is provided in [3]. In [4], the average and outage capacities of a multiple-input-multiple-output (MIMO)-OFDM system with B-bit quantized feedback are derived for the round-robin (RR) scheduler. In [7], the effect of feeding back quantized SNR on the throughput of the greedy scheduler, which exploits multi-user diversity, is studied. In [8], bit error rate (BER)

Manuscript received October 28, 2016; revised April 18, 2017 and July 13, 2017; accepted September 6, 2017. Date of publication September 19, 2017; date of current version November 9, 2017. This work was supported in part by the DST-Swaranajayanti Fellowship Award under Grant DST/SJF/ETA-01/2014-15 and in part by a research grant from the Ministry of Communications & Information Technology, India. The associate editor coordinating the review of this paper and approving it for publication was X. Zhu. *(Corresponding author: Vineeth Kumar.)*

The authors are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India (e-mail: vineethkumar01@gmail.com; neeleshbmehta@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2017.2752154

expressions are derived for binary phase shift keying with B-bit quantized feedback that is optimized to minimize the BER.

The 1-bit feedback scheme, which has the least feedback overhead, has attracted considerable attention by itself. In [9], it is shown that it is sum-rate optimal and achieves the double-logarithmic capacity growth when the number of users is sufficiently large. In [10], a two-round feedback scheme for cumulative distribution function (CDF)-based scheduling is proposed. 1-bit feedback is used in the first round to select a user. And, in the second round, the selected user feeds back its unquantized SNR to the BS.

While [7], [9], and [10] assume continuous rate adaptation, [4], [11], and [12] assume discrete rate adaptation. In the former, the Shannon capacity formula is used to map the SNR to the transmit rate, which can take any positive real value. However, in discrete rate adaptation, the BS transmits using only a pre-specified set of MCSs. Each MCS has a rate and an associated SNR threshold, which we shall refer to as the *MCS threshold*. A transmission using an MCS can be decoded successfully so long as the SNR exceeds its MCS threshold. In [11], a multi-stage threshold-based feedback scheme for selecting the user with the highest rate is proposed. In [12], CDF scheduling is considered, and a method to fine-tune the quantizer that generates feedback for each user is proposed. However, the sets of MCSs that the BS chooses from can be different for different users. In practice, the set of MCSs available at the BS is not user-specific.

It is often implicitly assumed in the above literature [4], [11], [12], [15], [16] that when the BS receives quantized feedback from a user, which indicates that its SNR lies between a lower and an upper quantization level, it chooses the highest rate MCS whose MCS threshold does not exceed the lower level. We shall refer to this as *conventional rate adaptation (CRA)*.

B. Focus and Contributions

In this paper, we propose a novel discrete rate adaptation policy for OFDM systems that use threshold-based quantized feedback. The innovation in our approach is that the BS no longer relies only on the lower quantization level indicated by the feedback to determine the MCS. Instead, it determines the *throughput-optimal* MCS based on the fed back CSI and the statistical information about each user, such as mean SNR and correlation between SCs. We show that the throughput-optimal MCS maximizes the feedback-conditioned goodput, which is the number of bits that can be correctly decoded by the user conditioned on the quantized CSI it fed back to the BS. Consequently, without additional feedback, it leads to a higher average cell throughput than with CRA, which chooses the MCS conservatively using only the fed back CSI.

We also address several new design questions that consequently arise, such as which statistical information is most useful and what form of quantization is most effective for threshold-based feedback. We make the following specific contributions:

- We derive the feedback-conditioned goodput for different multi-antenna diversity modes in closed-form for the widely used exponentially correlated SC gains model, which captures the decrease in correlation as the spacing between SCs increases [17]–[19]. We also present a novel, low complexity Monte Carlo method that computes it efficiently. This serves as a fundamental benchmark to establish the near-optimality of other lower-complexity, practically-amenable MCS selection policies such as those that ignore correlation and can, therefore, be easily implemented at a BS.
- We derive an insightful lower bound on the throughput gain achieved by the throughput-optimal policy over CRA for 1-bit feedback.
- We extend our approach to the multi-user setting for which we present new adaptations of three popular schedulers, which trade-off differently between cell throughput and user fairness.
- Our study also brings out the key role played by the manner in which the quantization levels, which map the real-valued SNRs to feedback bits, are defined. We show that BS-side rate estimation delivers significant throughput gains for the widely assumed model in the literature in which the quantization levels are the same as the MCSs thresholds [12], [15]. However, the gains with percentile threshold-based feedback [4], [20], in which the quantization levels are user-specific and a function of the CDF of the SNR of the user, are relatively less.

We note that BS-side estimation has been studied for best-m feedback in [21]. A throughput-optimal rate adaptation policy for best-m feedback has also been developed in [6]. However, these works differ from ours in their models, analytical techniques, and results. A scheme that uses a notion of goodput similar to ours and adapts the transmit rate and power of the BS based on quantized feedback received in the past is studied in [16]. However, OFDM is not modeled. A goodput-based, subcarrier-level fair resource allocation policy is developed for multi-user OFDM systems in [22]. However, perfect CSI at the BS is assumed.

C. Organization and Notations

This paper is organized as follows. Section II presents the system model. Section III develops the throughput-optimal rate adaptation policy. The overall system throughput implications and simulation results are presented in Section IV. Our conclusions follow in Section V.

Notations: The probability of an event A is denoted by $\Pr(A)$. The conditional probability of A given event B is denoted by $\Pr(A \mid B)$. The joint probability of events A and B is denoted by $\Pr(A, B)$. The probability density function (PDF) and CDF of a random variable (RV) X are denoted by $f_X(\cdot)$ and $F_X(\cdot)$, respectively. We denote the expectation with respect to RV X by $\mathbb{E}_X[\cdot]$. Similarly, the conditional expectation with respect to RV X given an event A is denoted by $\mathbb{E}_X[\cdot|A]$. Matrices are denoted using boldface characters. For a complex number c, c^* and $|c|$ denote its complex conjugate and absolute value, respectively. The multinomial $\binom{t}{w_1,\ldots,w_N}$ denotes $t!/(w_1!\ldots w_N!)$. The indicator function $1_{\{a\}}$ equals 1 if a is true and is 0 otherwise.

II. System Model

We consider an OFDM-based cellular system with one BS and K users. The system bandwidth is divided into N orthogonal SCs. The BS has N_t transmit (Tx) antennas and each user has N_r receive (Rx) antennas. The channel gains between different transmit antennas of the BS and different receive antennas of a user are independent and identically distributed (i.i.d.). This model is valid when the antennas are spaced sufficiently far apart [23], [24]. We focus on the single-cell scenario given that our approach is novel even for it and given the considerable attention that this scenario has received in the literature [4], [6], [12], [16], [22].

Channel Model: Let $H_{kn}(i, j)$ denote the complex baseband gain of downlink SC n for user k between the i^{th} Tx antenna of the BS and the j^{th} Rx antenna of the user. We consider Rayleigh fading. This combined with the simplified path-loss model [23, Ch. 2] implies that the SC power gain $|H_{kn}(i, j)|^2$ is an exponential RV with mean $\beta x_k^{-\alpha}$, where α is the path loss exponent, β is a path loss constant, and x_k is the distance between the user and the BS.

For single-input-multiple-output (SIMO) ($N_t = 1$ and $N_r \geq 1$) with maximal ratio combining, the instantaneous SNR γ_{kn} is given by

$$\gamma_{kn} = \frac{P_T \sum_{j=1}^{N_r} |H_{kn}(1, j)|^2}{\sigma^2}, \tag{1}$$

where P_T is the transmit power of the BS per SC and σ^2 is the additive white Gaussian noise power per SC. γ_{kn} can be written as a gamma RV with PDF $f_{\gamma_{kn}}(\upsilon) = \upsilon^{d-1} e^{-\frac{\upsilon}{a}} / \left(\Gamma(d)\, a^d\right)$, for $\upsilon \geq 0$, where $\Gamma(d) = \int_0^\infty x^{d-1} e^{-x} dx$. It has a shape parameter $d = N_r$ and a scale parameter $a = P_T \beta x_k^{-\alpha} / \sigma^2$ [25]. Single-input-single-output (SISO) ($N_t = 1$ and $N_r = 1$) is a special case of SIMO.

Similarly, for multiple-input-single-output (MISO) with maximal ratio transmission ($N_t > 1$ and $N_r = 1$), the SNR is $\gamma_{kn} = P_T \sum_{i=1}^{N_t} |H_{kn}(i, 1)|^2 / \sigma^2$. It is also a gamma RV with parameters $d = N_t$ and $a = P_T \beta x_k^{-\alpha} / \sigma^2$. For single-stream MIMO ($N_t \geq 2$ and $N_r \geq 2$), the SNR is $\gamma_{kn} = P_T \zeta_{kn} / \sigma^2$, where ζ_{kn} is the largest singular value of the channel matrix $\mathbf{H}_{kn}$. Approximating ζ_{kn} as a gamma RV and using the formulae in [26] for computing the mean of ζ_{kn}, we get that γ_{kn} is a gamma RV with $d = N_t N_r$ and $a = P_T \beta x_k^{-\alpha} \left(\frac{N_t + N_r}{N_t N_r + 1}\right)^{\frac{2}{3}} / \sigma^2$ [25]. From above, the fading-averaged SNR $\bar{\gamma}_k$ of user k for any SC is equal to $P_T N_r \beta x_k^{-\alpha} / \sigma^2$ for SIMO, $P_T \beta x_k^{-\alpha} / \sigma^2$ for SISO, $P_T N_t \beta x_k^{-\alpha} / \sigma^2$ for MISO, and $P_T N_t N_r \beta x_k^{-\alpha} \left(\frac{N_t + N_r}{N_t N_r + 1}\right)^{\frac{2}{3}} / \sigma^2$ for single-stream MIMO.

Note: For supporting MISO and single-stream MIMO, two additional quantities, namely, precoding matrix indicator (PMI) and rank indicator (RI) are fed back in OFDM systems such as LTE [2, Ch. 11]. Given our focus on quantized CSI for rate adaptation, we do not model the precoding matrix quantization and its impact on the SNR. The SNR formulae above for MISO and single-stream MIMO serve as upper bounds on the SC SNR that can be achieved in practice. This ensures tractability and enables us to glean new insights about the system performance.

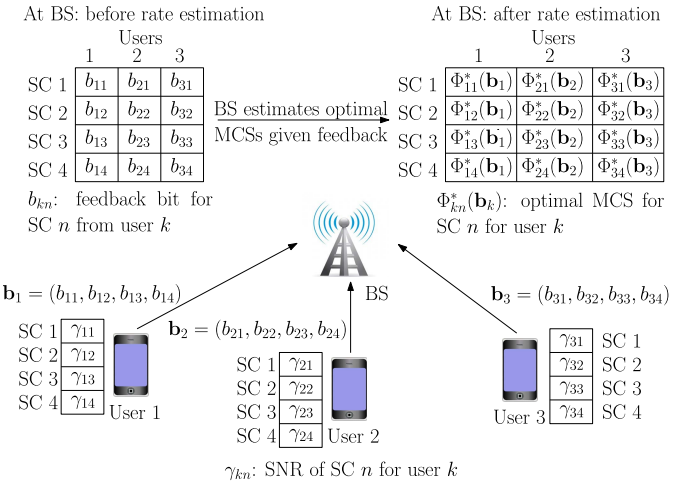

Fig. 1. Illustration of BS-side rate estimation and threshold-based quantized feedback for $K = 3$ users and $N = 4$ SCs. The information available at the BS before and after rate estimation are shown.

The SNRs of all the SCs of a user are statistically identical, but correlated. This follows from the uncorrelated scatterers assumption [23, Ch. 3]. To model the correlation between the SCs, we employ the widely used exponential correlation model, in which covariance of γ_{kn} and γ_{kl} is $\bar{\gamma}_k^2 \rho^{2|n-l|}$, for $1 \leq n, l \leq N$ and $0 \leq \rho \leq 1$ [17]–[19]. The SNRs of the SCs for different users are mutually independent since the users are located far apart from each other compared to the transmission wavelength. The statistical parameters $\bar{\gamma}_k$ and ρ are assumed to be known at the BS. It can determine them using long-term observations of the signals received from the users since they change at a timescale that is several orders of magnitude slower than short-term fading [9], [27].

Threshold-Based Quantized Feedback: User k estimates the downlink SNR γ_{kn} for each SC n and quantizes it to b_{kn} using a B-bit quantizer that has $L = 2^B$ regions. We denote the quantization levels by $0 = Q_1 < \cdots < Q_L < Q_{L+1} = \infty$. We shall refer to the set $\mathcal{Q} = \{Q_1, \ldots, Q_L\}$ as the *quantizer*. The vector $\mathbf{b}_k = (b_{k1}, \ldots, b_{kN})$ is then fed back to the BS.

Discrete Rate Adaptation and MCS Set at BS: The BS maps the received feedback for each SC from each user to an MCS drawn from an *MCS set* $\mathcal{M} = \{1, 2, \ldots, M\}$. MCS m has a rate R_m and an MCS threshold $T_m = (2^{R_m} - 1)/\eta$, where η is the coding loss of a practical code [28]. A transmission using MCS m is successful if $\gamma_{kn} \geq T_m$. Else, an outage occurs [5], [16]. Without loss of generality, let $0 = R_1 < R_2 < \cdots < R_M$ and $0 = T_1 < T_2 < \cdots < T_M < T_{M+1} = \infty$. In general, let $\Phi_{kn}(\mathbf{b}_k)$ denote a rate adaptation policy at the BS for SC n for user k, which maps the feedback vector $\mathbf{b}_k$ to an MCS $m \in \mathcal{M}$. The above model is illustrated in Figure 1.

III. Throughput-Optimal Rate Adaptation (TORA)

For ease of notation, we drop the user index k in this section. Hence, we use $\Phi_n(\mathbf{b})$ instead of $\Phi_{kn}(\mathbf{b}_k)$ to denote a rate adaptation policy. We first characterize in Lemma 1 the throughput-optimal rate adaptation policy. It highlights the importance

of the feedback-conditioned goodput, which we define precisely below. We then derive closed-form expressions for it in Result 1. Thereafter, we investigate low computational complexity approaches to evaluate the feedback-conditioned goodput. While the expressions in Result 1 turn out to be involved, they serve as a fundamental benchmark to assess the efficacy of the above low complexity approaches.

A. Characterization of Optimal Policy

The following lemma gives the throughput-optimal MCS for any MCS set $\mathcal{M}$ at the BS.

Lemma 1: The throughput-optimal MCS $\Phi_n^*(\mathbf{b})$ for SC n given feedback vector $\mathbf{b}$ is

$$\Phi_n^*(\mathbf{b}) = \arg\max_{m\in\mathcal{M}} \{R_m \Pr(\gamma_n \geq T_m|\mathbf{b})\}. \tag{2}$$

Proof: The proof is given in Appendix A. ∎

We shall refer to $\Psi_n^{(m)}(\mathbf{b}) = R_m \Pr(\gamma_n \geq T_m|\mathbf{b})$ as the *feedback-conditioned goodput* of MCS m. Thus, TORA maximizes the feedback-conditioned goodput. It is unlike CRA, in which the MCS is determined entirely by the lower quantization level conveyed by the feedback.

We now derive $\Psi_n^{(m)}(\mathbf{b})$ for exponentially correlated SC SNRs.

Result 1: For SC n, let q_n denote the index of the lower quantization level of the region in which γ_n lies, i.e., $\gamma_n \in [Q_{q_n}, Q_{q_n+1})$. The feedback-conditioned goodput $\Psi_n^{(m)}(\mathbf{b})$ of MCS m for SISO, SIMO, MISO, and single-stream MIMO is given by

$$\Psi_n^{(m)}(\mathbf{b}) = \begin{cases} \dfrac{R_m C_m(n)}{D_m(n)}, & T_m < Q_{q_n+1}, \\ 0, & T_m \geq Q_{q_n+1}, \end{cases} \tag{3}$$

where $C_m(n)$ is the joint probability of the events $\gamma_n \geq T_m$ and $Q_{q_1} \leq \gamma_1 < Q_{q_1+1}, \ldots, Q_{q_N} \leq \gamma_N < Q_{q_N+1}$, and $D_m(n)$ is the joint probability of the events $Q_{q_1} \leq \gamma_1 < Q_{q_1+1}, \ldots, Q_{q_N} \leq \gamma_N < Q_{q_N+1}$. To represent $C_m(n)$ and $D_m(n)$ compactly, let $U(y,u) = \left(\int_y^\infty x^{u-1}e^{-x}\,dx\right)/\Gamma(u)$ denote the upper incomplete gamma function and let $\nu_j = Q_j d/\left(\bar{\gamma}(1-\rho^2)\right)$, $\Theta_{11} = U\left(\nu_{q_1}, w_1+d\right)$, $\Theta_{12} = U\left(\nu_{q_1+1}, w_1+d\right)$, $\Theta_{j1} = U\left(\nu_{q_j}(1+\rho^2), w_{j-1}+w_j+d\right)$, $\Theta_{j2} = U\left(\nu_{q_j+1}(1+\rho^2), w_{j-1}+w_j+d\right)$, $\Theta_{N1} = U\left(\nu_{q_N}, w_{N-1}+d\right)$, $\Theta_{N2} = U\left(\nu_{q_N+1}, w_{N-1}+d\right)$, $\tau_t = (\Theta_{N1}-\Theta_{N2})/(N-1)^t$, and $\Omega = \prod_{j'=2}^{N-1} \frac{\Gamma(w_{j'-1}+w_{j'}+d)}{\Gamma(w_{j'}+d)\Gamma(w_1+d)}$.

For $n=1$, $C_m(1)$ is given by

$$C_m(1) = \sum_{t=0}^{\infty} \frac{\rho^{2t}(N-1)^t}{t!} \sum_{\substack{w_1\geq 0,\ldots,w_{N-1}\geq 0,\\ w_1+\cdots+w_{N-1}=t}} \binom{t}{w_1,\ldots,w_{N-1}} \times \tau_t\Omega\left[U\left(\frac{\max\{T_m, Q_{q_1}\}d}{\bar{\gamma}(1-\rho^2)}, w_1+d\right)-\Theta_{12}\right] \times \left[\prod_{j=2}^{N-1}\frac{\Theta_{j1}-\Theta_{j2}}{(1+\rho^2)^{w_{j-1}+w_j+d}}\right]. \tag{4}$$

For $2 \leq n \leq N-1$, $C_m(n)$ is given by

$$C_m(n) = \sum_{t=0}^{\infty} \frac{\rho^{2t}(N-1)^t}{t!} \sum_{\substack{w_1\geq 0,\ldots,w_{N-1}\geq 0,\\ w_1+\cdots+w_{N-1}=t}} \binom{t}{w_1,\ldots,w_{N-1}} \times \frac{\tau_t\Omega}{(1+\rho^2)^{w_{n-1}+w_n+d}} \times \left[\prod_{j=2, j\neq n}^{N-1}\frac{\Theta_{j1}-\Theta_{j2}}{(1+\rho^2)^{w_{j-1}+w_j+d}}\right] \times \left[U\left(\frac{\max\{T_m, Q_{q_n}\}(1+\rho^2)d}{\bar{\gamma}(1-\rho^2)}, w_{n-1}+w_n+d\right)-\Theta_{n2}\right]. \tag{5}$$

For $n=N$, $C_m(N)$ is given by

$$C_m(N) = \sum_{t=0}^{\infty} \frac{\rho^{2t}(N-1)^t}{t!} \sum_{\substack{w_1\geq 0,\ldots,w_{N-1}\geq 0,\\ w_1+\cdots+w_{N-1}=t}} \binom{t}{w_1,\ldots,w_{N-1}} \times \left[U\left(\frac{\max\{T_m, Q_{q_N}\}d}{\bar{\gamma}(1-\rho^2)}, w_{N-1}+d\right)-\Theta_{N2}\right] \times \frac{(\Theta_{11}-\Theta_{12})\Omega}{(N-1)^t}\left[\prod_{j=2}^{N-1}\frac{\Theta_{j1}-\Theta_{j2}}{(1+\rho^2)^{w_{j-1}+w_j+d}}\right]. \tag{6}$$

And, $D_m(n)$, for $1 \leq n \leq N$, is given by

$$D_m(n) = \sum_{t=0}^{\infty} \frac{\rho^{2t}(N-1)^t}{t!} \sum_{\substack{w_1\geq 0,\ldots,w_{N-1}\geq 0,\\ w_1+\cdots+w_{N-1}=t}} \binom{t}{w_1,\ldots,w_{N-1}} \times (\Theta_{11}-\Theta_{12})\,\tau_t\Omega\left[\prod_{j=2}^{N-1}\frac{\Theta_{j1}-\Theta_{j2}}{(1+\rho^2)^{w_{j-1}+w_j+d}}\right]. \tag{7}$$

Proof: The proof is given in Appendix B. ∎

Comments: In short, in TORA, the BS uses Result 1 to calculate the feedback-conditioned goodput for each MCS given the feedback vector $\mathbf{b}$ for all SCs. It then transmits using the MCS that maximizes the feedback-conditioned goodput, on each SC. The expressions for $C_m(n)$ and $D_m(n)$ above capture the fact that due to SC correlation, the feedback bits for other SCs also carry information about the SNR of a given SC. The expressions for $C_m(n)$ are different for different SCs because, in the exponential correlation model, the conditional probability in (2) is a function of the SC index. The infinite series in $C_m(n)$ and $D_m(n)$ are typical of analyses involving the exponential correlation model [17]–[19] and arise due to the involved forms of the joint PDF and CDF of the SC SNRs. We have found that 40 terms for $\rho = 0.5$ and 90 terms for $\rho = 0.9$ are sufficient to ensure numerical accuracy for the parameters of interest. The number of terms increases with ρ because of the presence of the $(1-\rho^2)$ term in the denominators of the arguments inside the incomplete gamma functions.

The closed-form expression in Result 1 brings out how TORA uses the channel statistics parameters $\bar{\gamma}$ and ρ, and also

why CRA is inherently sub-optimal. CRA selects the highest rate MCS m that has $C_m(n) = D_m(n)$. However, this is only one among the M possibilities that TORA chooses from.

To gain more insights, we now study the simpler and analytically insightful case when the SCs are i.i.d., which happens when $\rho = 0$. This will also lead to a simpler, computationally inexpensive, and near-optimal variant of TORA in Section III-C.

Corollary 1: For $\rho = 0$, the expressions for $C_m(n)$ and $D_m(n)$, for $1 \le n \le N$, simplify to

$$C_m(n) = U\left(\frac{\max(T_m, Q_{q_n})d}{\bar{\gamma}}, d\right) - U\left(\frac{Q_{q_n+1}d}{\bar{\gamma}}, d\right),$$
$$D_m(n) = U\left(\frac{Q_{q_n}d}{\bar{\gamma}}, d\right) - U\left(\frac{Q_{q_n+1}d}{\bar{\gamma}}, d\right). \quad (8)$$

B. Computational Complexity Reduction

We observe that the main computational difficulty in (4)–(7) arises due to the inner summation that requires enumerating all n-tuples that sum to t and evaluating the summand for each of them. The number of such n-tuples is $\binom{N+t-2}{N-2}$, which is large even for small values of N and t [29, Ch. 6]. In order to address this challenge, we develop a technique below to compute $D_m(n)$ in (7). The technique for computing $C_m(n)$ is similar and is not shown to avoid repetition.

Consider a multinomial random vector $\mathbf{W} = (W_1, \ldots, W_{N-1})$. Let $\mathbf{w} = (w_1, \ldots, w_{N-1})$ be a realization of $\mathbf{W}$ such that $\sum_{i=1}^{N-1} w_i = t$. The probability mass function of $\mathbf{W}$ for $w_1 \ge 0, \ldots, w_{N-1} \ge 0$, in terms of probabilities $p_1, \ldots, p_{N-1}$, where $\sum_{i=1}^{N-1} p_i = 1$, is [29, Ch. 6]

$$\Pr(W_1 = w_1, \ldots, W_{N-1} = w_{N-1}) = \binom{t}{w_1, \ldots, w_{N-1}} \prod_{i=1}^{N-1} p_i^{w_i}. \quad (9)$$

Then, from (7), it follows that $D_m(n)$ can be rewritten as

$$D_m(n) = \sum_{t=0}^{\infty} \frac{\rho^{2t}(N-1)^t}{t!} \mathbb{E}_{\mathbf{W}}[h_m(\mathbf{b}, \mathbf{w})], \quad (10)$$

where $h_m(\mathbf{b}, \mathbf{w}) = \left[\prod_{j=2}^{N-1} \frac{\Theta_{j1} - \Theta_{j2}}{(1+\rho^2)^{w_{j-1}+w_j+d}}\right](\Theta_{11} - \Theta_{12})\,\Omega \times (\Theta_{N1} - \Theta_{N2})$ and $p_1 = \cdots = p_{N-1} = 1/(N-1)$. The expression in (10) can now be easily computed using Monte Carlo methods as follows. For each t, we generate S_{MC} realizations of $\mathbf{W}$. Let these be denoted by $\mathbf{w}_\xi$, for $1 \le \xi \le S_{MC}$. Then $\mathbb{E}_{\mathbf{W}}[h_m(\mathbf{b}, \mathbf{w})]$ is well approximated by

$$\mathbb{E}_{\mathbf{W}}[h_m(\mathbf{b}, \mathbf{w})] \approx \frac{1}{S_{MC}} \sum_{\xi=1}^{S_{MC}} h_m(\mathbf{b}, \mathbf{w}_\xi). \quad (11)$$

The approximation error decreases as $O\left(1/\sqrt{S_{MC}}\right)$ [30, Ch. 2]. We have found that $S_{MC} = 300$ for $\rho = 0.5$ and $S_{MC} = 1500$ for $\rho = 0.9$ are sufficient to ensure numerical accuracy. The computational complexity of this method is several orders of magnitude lower than numerically computing $\Pr(\gamma_n \ge T_m \mid \mathbf{b})$.

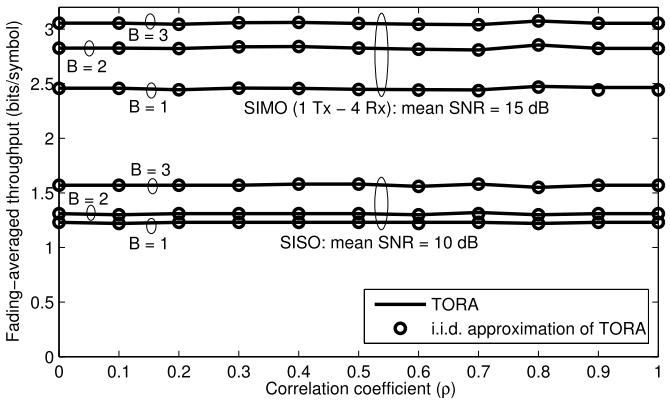

Fig. 2. Fading-averaged throughputs of TORA and its i.i.d. approximation as a function of ρ for SISO and SIMO ($M = 16$ and $N = 10$).

C. Sensitivity of Average Throughput to ρ and $\bar{\gamma}$

We now investigate the sensitivity of the fading-averaged throughput of TORA to ρ and $\bar{\gamma}$ for different multi-antenna modes. Figure 2 plots it for SISO and SIMO (4 Rx antennas) as a function of ρ at mean SNRs of $\bar{\gamma} = 10$ dB and 15 dB, respectively. The results for MISO and single-stream MIMO are similar, and are not shown. Results are shown for $B = 1$, 2, and 3 bits, whose corresponding quantization levels are derived using the LTE MCS set in [2, Table 10.1], which consists of 16 MCSs. For $B = 3$, quantization levels correspond to the MCS thresholds of rates taken alternately from [2, Table 10.1]. Similarly, for $B = 2$, the quantization levels are taken alternately from the above set of thresholds, and so on. The MCS set at the BS consists of all the above 16 MCSs. Also plotted using markers are the corresponding results for the i.i.d. approximation of TORA that uses (8) to compute the feedback-conditioned goodput.

We observe that the average throughput is insensitive to ρ given B, $\bar{\gamma}$, and the multi-antenna mode. This implies that using (8) instead of the involved expressions in (4)–(7) for evaluating feedback-conditioned goodput is near-optimal. This can be intuitively understood by studying the extreme cases when ρ takes values 0 and 1. When $\rho \to 1$, for each fade realization, all the SC SNRs of a user lie within the same quantization region with high probability. Therefore, given the feedback bits for an SC, no additional information is provided about its SNR by the feedback bits of the other SCs, as is the case with $\rho = 0$. Thus for any mean SNR, the fading-averaged throughput is the same for $\rho = 0$ and 1.

We note that this result is unique to threshold-based quantized feedback, and is not known in the literature to the best of our knowledge. This behavior is different from best-m feedback, whose performance is indeed sensitive to ρ [21].

D. Insights: One-Bit Feedback Scheme for SISO

In order to gain more insights, we now analyze 1-bit feedback for SISO for $\rho = 0$, given its simpler structure and given the considerable literature devoted to it [3], [9], [10]. The quantizer is $\mathcal{Q} = \{0, Q_2, \infty\}$. For SC n, the feedback bit b_n is 0, if $\gamma_n < Q_2$, and is 1, otherwise. From Corollary 1,

we see that, in TORA, the throughput-optimal MCS $\Phi_n^*(\mathbf{b})$ for SC n is chosen as a function of the feedback bit b_n as follows:

$$\Phi_n^*(\mathbf{b}) = \begin{cases} \arg\max\limits_{m\in\mathcal{M}} \left\{ R_m \left[\dfrac{e^{-\frac{\min\{T_m, Q_2\}}{\bar{\gamma}}} - e^{-\frac{Q_2}{\bar{\gamma}}}}{1 - e^{-\frac{Q_2}{\bar{\gamma}}}} \right] \right\}, & b_n = 0, \\ \arg\max\limits_{m\in\mathcal{M}} \left\{ R_m \left[e^{-\frac{(\max\{T_m, Q_2\} - Q_2)}{\bar{\gamma}}} \right] \right\}, & b_n = 1. \end{cases} \tag{12}$$

On the other hand, CRA sets the MCS index as 1, which has a rate $R_1 = 0$, if $b_n = 0$. Else, the MCS index is set as m', which has a rate $R_{m'}$. Here, $m' = \arg\max\limits_{m\in\mathcal{M}} \left\{ R_m 1_{\{T_m \le Q_2\}} \right\}$.

We see that even though both TORA and CRA use only two MCSs, they differ in their choice of the MCS for both $b_n = 0$ and 1. For $b_n = 0$, while CRA does not transmit on the SC, TORA transmits at a non-zero rate by exploiting statistical knowledge. For $b_n = 1$, TORA chooses an MCS that has a rate that is greater than or equal to that chosen by CRA.

Average Throughput Gains: We now use (12) to characterize the average throughput gain of a user, which is defined as the difference between the fading-averaged throughputs of TORA and CRA. From (12), we see that the throughput-optimal MCS chosen by TORA depends on $\bar{\gamma}$, which, in turn, is a function of the distance x of the user from the BS. The following lower bound characterizes the average throughput gain as a function of x.

Lemma 2: For $Q_2 < T_M$, the average throughput gain $\Delta R(x)$ of a user at a distance x from the BS for $\rho = 0$ is lower bounded by

$$\Delta R(x) \ge \begin{cases} R_i e^{-\frac{T_i \sigma^2 x^\alpha}{P_T \beta}} - R_{m'} e^{-\frac{Q_2 \sigma^2 x^\alpha}{P_T \beta}}, \\ \quad \text{if } x_i \le x < x_{i-1}, \quad \text{for } m'+1 \le i \le M, \\ 0, \ \text{if } x \ge x_{m'}, \end{cases} \tag{13}$$

where $x_{m'} = \left(P_T \beta \ln\left(\frac{R_{m'+1}}{R_{m'}}\right) / \left(\sigma^2 \left(T_{m'+1} - Q_2\right)\right) \right)^{\frac{1}{\alpha}}$, $x_i = \left(P_T \beta \eta \ln\left(\frac{R_{i+1}}{R_i}\right) / \left(\sigma^2 \left(2^{R_{i+1}} - 2^{R_i}\right)\right) \right)^{\frac{1}{\alpha}}$, for $i \in \{m'+1, \ldots, M-1\}$, and $x_M = 0$.

Proof: The proof is given in Appendix C. ∎

Using Lemma 2, we now derive a lower bound for the average throughput gain of TORA that is also averaged over the user location.

Result 2: When the user is located with uniform probability over a circular cell area of radius r_c, the fading- and user location-averaged cell throughput gain ΔR_{cell} for $Q_2 < T_M$ is lower bounded by

$$\Delta R_{\text{cell}} \ge \frac{2\Gamma\left(\frac{2}{\alpha}\right)}{\alpha r_c^2} \sum_{i=m'+1}^{M} \left(R_i c_i^{\frac{2}{\alpha}} \left[U\left(\frac{x_i^\alpha}{c_i}, \frac{2}{\alpha}\right) - U\left(\frac{x_{i-1}^\alpha}{c_i}, \frac{2}{\alpha}\right) \right] - R_{m'} c_{m'}^{\frac{2}{\alpha}} \left[U\left(\frac{x_i^\alpha}{c_{m'}}, \frac{2}{\alpha}\right) - U\left(\frac{x_{i-1}^\alpha}{c_{m'}}, \frac{2}{\alpha}\right) \right] \right), \tag{14}$$

where $c_{m'} = P_T \beta / \left(Q_2 \sigma^2\right)$ and $c_i = P_T \beta / \left(T_i \sigma^2\right)$, for $i \in \{m'+1, \ldots, M\}$.

Proof: The proof is given in Appendix D. ∎

For $Q_2 \ge T_M$, we can only show that $\Delta R_{\text{cell}} \ge 0$. However, this case is of limited interest since Q_2 for 1-bit feedback will typically be less than the largest MCS threshold T_M. The lower bound for $\Delta R(x)$ can be interpreted as follows. The cell area is partitioned into concentric circles, which correspond to the different MCSs being used. Specifically, x_{i-1} is the maximum distance from the BS at which MCS i, for $m'+1 \le i \le M$, will be allocated by TORA when $b_n = 1$. From (13), we see that $\Delta R(x)$ decays exponentially in x^α / P_T. Therefore, the closer the user is to the BS or the higher the BS transmit power, the higher is its throughput gain for a given Q_2. Note that $\Delta R(x)$ and ΔR_{cell} are functions of the MCS set $\mathcal{M}$ at the BS and the quantizer Q at the user. With multiple users in a cell, they will also be a function of the scheduler used. However, characterizing it is not analytically tractable.

IV. System-Level Performance Evaluation

We now present Monte Carlo simulation results to evaluate the efficacy of BS-side rate estimation and its sensitivity to various system parameters. We drop $K = 10$ users uniformly in a circular cell area with radius $r_c = 1$ km. The transmit power of the BS is fixed such that a user at the cell edge has a mean SNR of 0 dB with the noise power spectral density set as -174 dBm/Hz. There are $N = 10$ SCs, each with bandwidth 180 kHz. The path-loss parameters are $\alpha = 4$ and $\beta = 0.01$. The results are averaged over 1000 independent drops of the 10 users and 2000 fade realizations per user per drop. We first present results for the exponential correlation model. Thereafter, we present results for the typical urban (TU) channel model, in which the SCs are not exponentially correlated [24, Ch. 2]. To determine the MCSs for transmission, we use the near-optimal, low-complexity i.i.d. approximation of TORA in Section III-C.

A. MCS Set at BS

In order to benchmark the performance of the different policies over a wide range of uplink feedback bandwidths and BS rate adaptation capabilities, we present results for $M = 2$, 4, 8, and 16 MCSs. These are generated as follows from the LTE standard, which defines 16 MCSs [2, Table 10.1] with rates 0, 0.15, 0.23, 0.38, 0.60, 0.88, 1.18, 1.48, 1.91, 2.41, 2.73, 3.32, 3.90, 4.52, 5.12, and 5.55 bits/symbol. The $M = 16$ case directly corresponds to the MCSs specified in the table. The $M = 8$ case corresponds to rates taken alternately from the table, which are 0, 0.23, 0.60, 1.18, 1.91, 2.73, 3.90, and 5.12 bits/symbol. Similarly, $M = 4$ corresponds to rates 0, 0.60, 1.91, and 3.90 bits/symbol and $M = 2$ corresponds to rates 0 and 1.91 bits/symbol. As mentioned, the MCS threshold for each rate is given by $T_m = (2^{R_m} - 1)/\eta$, with $\eta = 0.398$ [21]. The number of MCSs at the BS is greater than or equal to the number of quantization levels.

B. Quantizers

In order to evaluate the impact of the quantizer on BS-side rate estimation, we also study percentile threshold-based

quantized feedback [4], [20]. In it, the quantization levels for a user for B-bit feedback are obtained from the CDF $F_{\mathrm{SNR}}(\cdot)$ of its SC SNR as follows: $Q_l = F_{\mathrm{SNR}}^{-1}\left((l-1)/L\right)$, for $2 \le l \le L+1$. This ensures that the SNR lies in the different quantization regions with equal probability. For example, for a 1-bit quantizer, the quantization level is the 50$^{\text{th}}$ percentile value. Similarly, for a 2-bit quantizer, the levels are the 25$^{\text{th}}$, 50$^{\text{th}}$, and 75$^{\text{th}}$ percentile values. For both types of quantizers, the MCS sets are derived from LTE and are described in Section IV-A.

C. Scheduling Algorithms and Their Adaptations to Feedback-Conditioned Goodput

In order to benchmark the performance of TORA as a function of the scheduler, we present results for the following three schedulers, which span a wide range of the trade-off between cell throughput and user fairness. For this, we recall the following notation. The optimal MCS assigned to SC n for user k as per Lemma 1 is $\Phi_{kn}^*(\mathbf{b}_k)$. It has an MCS threshold $T_{\Phi_{kn}^*(\mathbf{b}_k)}$, rate $R_{\Phi_{kn}^*(\mathbf{b}_k)}$, and feedback-conditioned goodput $\Psi_{kn}^*(\mathbf{b}_k) = R_{\Phi_{kn}^*(\mathbf{b}_k)}\Pr\left(\gamma_{kn} > T_{\Phi_{kn}^*(\mathbf{b}_k)} | \mathbf{b}_k\right)$. For B-bit feedback, $\Psi_{kn}^*(\mathbf{b}_k)$ is a discrete RV, which takes $L = 2^B$ values $s_1, \ldots, s_L$, with the probability that it takes the value s_j being p_j. Let i_n denote the user allocated to SC n by the scheduler.

1) *RR Scheduler:* It allocates SCs to users in a predetermined, channel-agnostic manner. For example, in the first scheduling interval, all SCs are allocated to the first user. In the second scheduling interval, all SCs are allocated to the second user, and so on.
2) *Feedback-Conditioned Goodput-Based CDF Scheduler:* It exploits multi-user diversity and is known to maximize cell throughput among all schedulers that allocate equal time on average to each user. Conventionally, CDF scheduling has been defined when $R_{\Phi_{kn}^*(\mathbf{b}_k)}$ is a continuous RV with CDF $F_{R_{\Phi_{kn}^*(\mathbf{b}_k)}}(\cdot)$. In this case, i_n is given by [31]

$$i_n = \underset{1\le k\le K}{\operatorname{argmax}} \left\{ \left(F_{R_{\Phi_{kn}^*(\mathbf{b}_k)}}(r_{kn}) \right)^{\frac{1}{w_k}} \right\}, \tag{15}$$

where r_{kn} is the realization of $R_{\Phi_{kn}^*(\mathbf{b}_k)}$ and $0 \le w_k \le 1$ is a weight factor that is used to give different preferences to different users, subject to the constraint $\sum_{k=1}^{K} w_k = 1$. Setting $w_k = 1/K$ ensures that each user is chosen on average for a fraction of time $1/K$.

Since $R_{\Phi_{kn}^*(\mathbf{b}_k)}$ is discrete in our model, we adapt the above scheduler in terms of the feedback-conditioned goodput as follows. This is motivated by the approach used in [15]. Let $F_{\Psi_{kn}^*(\mathbf{b}_k)}(s_j) = \sum_{i=1}^{j} p_j$ be the cumulative mass function (CMF) of $\Psi_{kn}^*(\mathbf{b}_k)$, and let $F_{\Psi_{kn}^*(\mathbf{b}_k)}(0) = 0$. The CMF is computed empirically since no closed-form formula is available for it. The feedback-conditioned goodput-based CDF scheduler is implemented as follows:

- *Step 1:* In each scheduling interval, the BS computes $\Psi_{kn}^*(\mathbf{b}_k)$ for all users and SCs.
- *Step 2:* The scheduler generates an RV U_{kn}, called scheduling metric, that is uniformly distributed in the interval $\left[F_{\Psi_{kn}^*(\mathbf{b}_k)}(s_{j-1}),\, F_{\Psi_{kn}^*(\mathbf{b}_k)}(s_j) \right)$.
- *Step 3:* SC n is allocated to the user with the largest weighted scheduling metric:

$$i_n = \underset{1\le k\le K}{\operatorname{argmax}} \left\{ (U_{kn})^{\frac{1}{w_k}} \right\}. \tag{16}$$

3) *Feedback-Conditioned Goodput-Based Greedy Scheduler:* It maximizes the cell throughput without considering fairness. SC n is allocated to the user with the largest $\Psi_{kn}^*(\mathbf{b}_k)$:

$$i_n = \underset{1\le k\le K}{\operatorname{argmax}} \left\{ \Psi_{kn}^*(\mathbf{b}_k) \right\}. \tag{17}$$

D. Performance Benchmarking

We compare the fading- and user location-averaged cell throughput, which we shall henceforth refer to as the cell throughput, of TORA with that of the following:

1) *Full CSI Feedback (F-CSI):* In it, $B = 4$ bits are fed back per SC and the number of MCSs M is set to the largest value 16. The feedback bits for an SC indicate the MCS to be used by the BS for downlink transmission on that SC. This approach achieves the highest cell throughput, but at the expense of maximum feedback overhead.
2) *CRA [4], [11], [12], [15], [16]:* As mentioned in Section I-A, CRA selects an MCS for each SC based only on the lower level of the quantization region in which the SC SNR lies.

We note that a meaningful and fair comparison with the approaches in [4], [11], [12], [15], and [16] is not possible due to fundamental differences in the models assumed.

E. Evaluation of Lower Bounds

We first study the fading-averaged throughput gain $\Delta R(x)$ for one user with $M = 16$ MCSs at the BS for SISO. Figure 3 plots $\Delta R(x)$ and its lower bound in (13) as a function of x for the 1-bit quantizer $\mathcal{Q} = \{0, Q_2, \infty\}$ with $Q_2 = 5$, 10, and 15 dB for $\rho = 0$. We see that the trends depend upon the choice of Q_2. For a given Q_2, $\Delta R(x)$ initially decreases as x increases. This is because for small x, for which b_n is 1 with high probability, the bulk of the throughput gain comes from the fact that TORA chooses a high-rate MCS while CRA does not. However, as x increases further, e.g., beyond 300 m for $Q_2 = 15$ dB, $\Delta R(x)$ increases due to the non-negligible contribution from the MCS selected by TORA for $b_n = 0$. As x increases even further, e.g., beyond 500 m for $Q_2 = 15$ dB, $\Delta R(x)$ decreases because $b_n = 0$ occurs with even higher probability, which, in turn, forces low-rate MCSs to be chosen with higher probability. We see that as Q_2 is lowered, the range of x for which the bound is tight increases. This is because, the contribution to $\Delta R(x)$ from the $b_n = 0$ event, which the bound neglects, decreases.

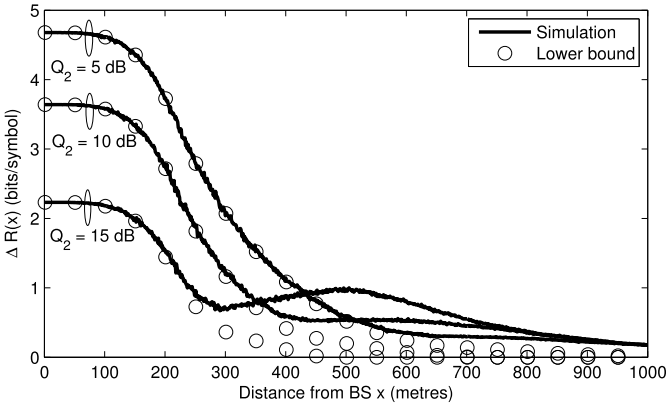

Fig. 3. $\Delta R(x)$ and its lower bound for 1-bit feedback with $Q_2 = 5$, 10, and 15 dB ($M = 16$, $K = 1$, $\rho = 0$, and $r_c = 1$ km).

TABLE I

COMPARISON OF CELL THROUGHPUTS FOR SISO FOR MCS THRESHOLD-BASED FEEDBACK WITH DIFFERENT SCHEDULERS ($\rho = 0.74$, $N = 10$, AND $K = 10$)

B	M	Cell throughput (bits/symbol)								
		Greedy scheduler			CDF scheduler			RR scheduler		
		CRA	TORA	F-CSI	CRA	TORA	F-CSI	CRA	TORA	F-CSI
1	2	1.88	1.88	4.76	1.22	1.22	2.74	0.65	0.65	1.64
1	4	1.88	3.10	4.76	1.22	1.51	2.74	0.65	1.05	1.64
1	8	1.88	3.58	4.76	1.22	1.58	2.74	0.65	1.13	1.64
1	16	1.88	3.72	4.76	1.22	1.60	2.74	0.65	1.16	1.64
2	4	3.51	3.51	4.76	2.01	2.01	2.74	1.16	1.16	1.64
2	8	3.51	3.96	4.76	2.01	2.07	2.74	1.16	1.27	1.64
2	16	3.51	4.10	4.76	2.01	2.08	2.74	1.16	1.30	1.64
3	8	4.40	4.40	4.76	2.51	2.51	2.74	1.48	1.48	1.64
3	16	4.40	4.52	4.76	2.51	2.52	2.74	1.48	1.50	1.64
4	16	4.76	4.76	4.76	2.74	2.74	2.74	1.64	1.64	1.64

F. Cell Throughput Benchmarking

Next, we present a comprehensive benchmarking of the cell throughput of TORA that takes into account different number of quantization levels and quantizers at the user, MCS sets at the BS, schedulers, and multi-antenna diversity modes.

1) MCS Threshold-Based Feedback: Table I presents a comprehensive tabulation of the cell throughputs for MCS threshold-based feedback for SISO. The SC SNRs of each user are exponentially correlated with $\rho = 0.74$.[1]

We observe that, for any given B and M, the greedy scheduler achieves the highest cell throughput while the RR scheduler has the lowest. The cell throughput of TORA exceeds that of CRA except when $L = 2^B$ is equal to M. Even with 1-bit feedback, TORA achieves 78.15%, 58.39%, and 70.73% of the throughput of F-CSI for the greedy, CDF, and RR schedulers, respectively. The corresponding percentages for CRA are only 39.50%, 44.53%, and 39.63%. Note that the TORA can achieve a higher cell throughput even with a lower feedback overhead if more MCSs can be provisioned at the BS. For example, 1-bit feedback with 16 MCSs at the BS achieves 5.98% more throughput than 2-bit feedback with 4 MCSs at the BS for the greedy scheduler. These relative gains diminish as the number of feedback bits increases.

TABLE II

COMPARISON OF CELL THROUGHPUTS FOR SISO FOR PERCENTILE THRESHOLD-BASED FEEDBACK WITH DIFFERENT SCHEDULERS ($\rho = 0.74$, $N = 10$, AND $K = 10$)

B	M	Cell throughput (bits/symbol)								
		Greedy scheduler			CDF scheduler			RR scheduler		
		CRA	TORA	F-CSI	CRA	TORA	F-CSI	CRA	TORA	F-CSI
1	2	1.59	1.85	4.76	0.62	0.77	2.74	0.31	0.65	1.64
1	4	2.71	3.25	4.76	1.14	1.38	2.74	0.57	1.03	1.64
1	8	3.37	3.91	4.76	1.47	1.66	2.74	0.73	1.19	1.64
1	16	3.66	4.15	4.76	1.64	1.75	2.74	0.82	1.24	1.64
2	4	3.19	3.38	4.76	1.56	1.63	2.74	0.86	1.09	1.64
2	8	3.98	4.15	4.76	1.94	2.01	2.74	1.10	1.31	1.64
2	16	4.31	4.45	4.76	2.13	2.19	2.74	1.23	1.41	1.64
3	8	4.20	4.27	4.76	2.17	2.21	2.74	1.29	1.39	1.64
3	16	4.55	4.60	4.76	2.40	2.43	2.74	1.43	1.51	1.64
4	16	4.66	4.68	4.76	2.55	2.56	2.74	1.53	1.57	1.64

2) Percentile Threshold-Based Feedback: Table II presents the corresponding results for percentile threshold-based feedback for SISO. In it, the quantizer of a user is customized to the statistics of its SNR. This is unlike MCS threshold-based feedback, in which the quantizer is the same for all the users. For a given B, the cell throughput increases as M increases for both CRA and TORA. This behavior is different from that in Table I for MCS threshold-based feedback, in which the cell throughput did not vary with M for CRA. TORA again outperforms CRA for all B and M. This increase in cell throughput for TORA with M is more pronounced for percentile threshold-based feedback. For example, for TORA with $B = 1$, the increase in cell throughput when M increases from 2 to 16 is 124.32%, 136.00%, and 90.77% for the greedy, CDF, and RR schedulers, respectively, in Table II. The corresponding increases in Table I for MCS threshold-based feedback are 97.87%, 31.15%, and 78.46%.

For $B = 1$, 2, and 3, percentile threshold-based feedback yields a higher cell throughput than MCS threshold-based feedback except for $M = L$. For example, for $B = 1$ and $M = 16$, percentile threshold-based feedback achieves 11.56%, 9.37%, and 6.90% more throughput for the greedy, CDF, and RR schedulers, respectively, compared to MCS threshold-based feedback. The cell throughput gains with lower feedback overhead when more MCSs are provisioned at the BS is also more for percentile threshold-based feedback.

G. One-Bit Feedback and Optimization of Quantization Level

We now focus on 1-bit feedback. Instead of setting the quantization level Q_2 as one of the MCSs thresholds in MCS threshold-based feedback, we optimize it to see how much the performance can be improved further.[2] We consider the TU channel model, in order to illustrate the utility of BS-side rate estimation for general channel models.

Figures 4 and 5 plot the cell throughputs of CRA, TORA, and F-CSI for the greedy and CDF schedulers, respectively,

[1]The correlation coefficient is chosen as 0.74 since it is equal to the correlation between two subcarriers that are apart by a bandwidth of 180 kHz of a PRB in the TU channel model. This enables us to also study the effect of the correlation model.

[2]Such a study for $B \geq 2$ is difficult because the number of quantization levels to be optimized increases exponentially with B. An optimization of Q_i for $B \geq 2$ is considered in [28]. However, discrete rate adaptation is not modeled.

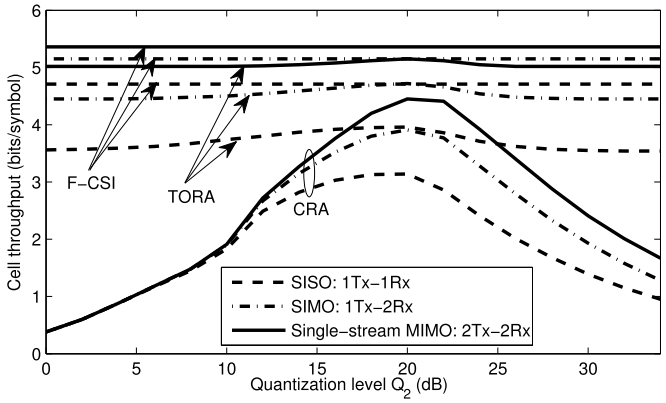

Fig. 4. Optimizing the quantization level Q_2 for 1-bit feedback for the greedy scheduler: Cell throughput as a function of Q_2 for different rate adaptation policies and different multi-antenna modes for the TU channel model ($N = 10$ and $K = 10$).

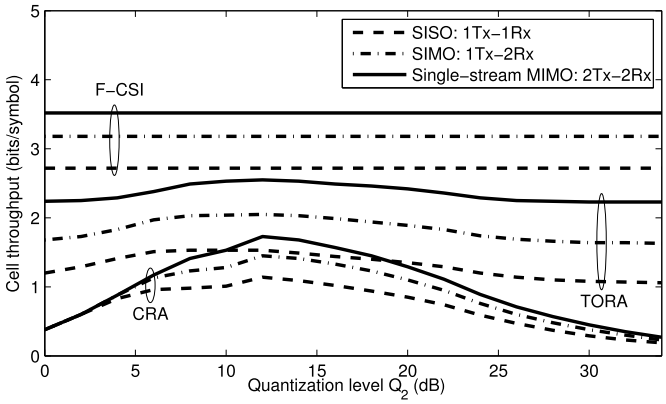

Fig. 5. Optimizing the quantization level Q_2 for 1-bit feedback for the CDF scheduler: Cell throughput as a function of Q_2 for different rate adaptation policies and different multi-antenna modes for the TU channel model ($N = 10$ and $K = 10$).

as a function of Q_2 for the SISO, SIMO, and MIMO modes. Results for the RR scheduler are qualitatively similar and are not shown to conserve space. For both CRA and TORA, the cell throughput first increases, reaches a maximum at the optimal Q_2, and then decreases as Q_2 increases further. This is because, at low values of Q_2, the rates allocated by both policies are correspondingly low, which leads to low cell throughputs. For high values of Q_2, the chances that the SC gain exceeds Q_2 are small, as a result of which both policies declare an outage with high probability. However, the cell throughput of TORA is less sensitive to Q_2 because it utilizes statistical information along with the fed back CSI to determine the transmit rate. We also observe that the optimal Q_2 is insensitive to the multi-antenna mode and the rate adaptation policy for both schedulers.

For the greedy scheduler, the cell throughput of TORA exceeds that of CRA by 26.11%, 21.26%, and 15.73% for SISO, SIMO, and MIMO, respectively, at the optimal Q_2, which is 20 dB. At this Q_2, it also achieves 84.08%, 91.65%, and 96.08% of the cell throughput of F-CSI for SISO, SIMO, and MIMO, respectively. For the CDF scheduler, the corresponding throughput gains are 34.21%, 41.38%, and 47.40%

TABLE III

COMPARISON OF OUTAGE PROBABILITIES OF TORA AND CRA FOR SISO FOR DIFFERENT QUANTIZERS AND SCHEDULERS ($\rho = 0.74$, $N = 10$, AND $K = 10$)

MCS threshold-based feedback							
B	M	Greedy scheduler		CDF scheduler		RR scheduler	
		CRA	TORA	CRA	TORA	CRA	TORA
1	2	0.02	0.02	0.36	0.36	0.33	0.33
1	4	0.02	0.11	0.36	0.24	0.33	0.17
1	8	0.02	0.14	0.36	0.25	0.33	0.19
1	16	0.02	0.15	0.36	0.25	0.33	0.19
2	4	0.00	0.00	0.00	0.00	0.15	0.15
2	8	0.00	0.05	0.00	0.01	0.15	0.10
2	16	0.00	0.06	0.00	0.03	0.15	0.10
3	8	0.00	0.00	0.00	0.00	0.06	0.06
3	16	0.00	0.01	0.00	0.00	0.06	0.05
4	16	0.00	0.00	0.00	0.00	0.04	0.04
Percentile threshold-based feedback							
B	M	Greedy scheduler		CDF scheduler		RR scheduler	
		CRA	TORA	CRA	TORA	CRA	TORA
1	2	0.17	0.07	0.68	0.60	0.42	0.33
1	4	0.01	0.07	0.27	0.27	0.32	0.19
1	8	0.00	0.07	0.00	0.15	0.25	0.15
1	16	0.00	0.06	0.00	0.10	0.25	0.14
2	4	0.00	0.04	0.02	0.06	0.23	0.18
2	8	0.00	0.03	0.00	0.06	0.15	0.11
2	16	0.00	0.02	0.00	0.04	0.13	0.09
3	8	0.00	0.01	0.00	0.03	0.10	0.09
3	16	0.00	0.01	0.00	0.02	0.08	0.07
4	16	0.00	0.00	0.00	0.01	0.06	0.05

at the optimal Q_2, which is 12 dB. This is 56.25%, 64.47%, and 72.44% of the cell throughput of F-CSI for SISO, SIMO, and MIMO, respectively. Thus, substantial cell throughput gains over CRA are achieved.

H. Outage Probability Comparisons

In order to gain different insights into the system behavior, Table III tabulates the outage probabilities of CRA, TORA, and F-CSI for SISO as a function of the number of feedback bits B, number of MCSs M, and quantizers. The results are averaged over fading and user locations. The outage probability is defined as the probability that no user is scheduled to transmit at a non-zero rate on an SC or the transmission by the scheduled user cannot be decoded. For CRA, an outage occurs for an SC when the SNR of the user scheduled on it is lower than T_2. However, for TORA, an outage occurs when the BS estimates a rate that is higher than that can be supported on the SC.

1) MCS Threshold-Based Feedback: For CRA, we observe that, for a given B, the outage probability is insensitive to M for all three schedulers. For $B = 1$, the CDF scheduler has the highest outage probability followed by the RR and greedy schedulers. However, for $B = 2$, 3, and 4, the outage probabilities of the greedy and CDF schedulers are nearly zero due to the higher resolution of feedback. The corresponding values for the RR scheduler are higher, but they decrease as B increases.

The behavior of the outage probability of TORA is different. It is now sensitive to both B and M. For $M = L$, it is the same as that of CRA. However, for $M > L$, it is relatively higher than that of CRA for the greedy scheduler, but lower

for the RR scheduler. For a given B, for $M > L$, it increases as M increases.

2) Percentile Threshold-Based Feedback: Unlike MCS threshold-based feedback, the outage probability of CRA is sensitive to both B and M for all three schedulers. In all cases, for a given B, the outage probability of CRA decreases as M increases. For TORA, as well, the outage probability is sensitive to both B and M for all three schedulers. For a given B, it decreases as M increases. It is relatively higher than that of CRA for the greedy and CDF schedulers except when $B = 1$ and $M = 2$. However, for the RR scheduler, the outage probability is lower than that of CRA, as was also the case for MCS threshold-based feedback. Thus, the trends and sensitivities observed for outage probability are different from those for cell throughput.

V. Conclusions

We proposed BS-side rate estimation for throughput-optimal discrete rate adaptation with threshold-based quantized feedback in OFDM systems. It decoupled the resolution of the feedback from the number of MCSs available at the BS and enabled the system designer to reduce the uplink feedback overhead at the expense of more computations at the BS. We also saw that this can be done without lowering the cell throughput by provisioning more MCSs at the BS. We derived the feedback-conditioned goodput in closed-form for the exponentially correlated SC gains model that applied to various multi-antenna diversity modes. While the expressions for it were involved, they provided a fundamental benchmark to establish the near-optimality of a low-complexity i.i.d. approximation approach that ignored correlation in determining the MCS. For 1-bit feedback, we derived a lower bound for the cell throughput gain, which brought out the dependence of the MCS chosen by TORA on the distance of the user from the BS, and also the extent by which CRA is sub-optimal. Our benchmarking results for MCS threshold-based feedback and percentile threshold-based feedback showed that the cell throughput gains are a function of the quantizer employed by the user in addition to the MCS set and scheduler employed by the BS. The largest gains were observed when the number of feedback bits is small, which is a regime of practical interest in OFDM systems.

Several interesting avenues for future work exist given the importance of optimizing system performance with limited feedback. These include modeling multi-user MIMO and multi-cell scenarios, outdated CSI, and quantization of the beamforming weights for multi-antenna modes.

Appendix

A. Proof of Lemma 1

The proof follows directly from first principles. If MCS m is chosen, then the rate achieved is equal to $R_m 1_{\{\gamma_n \geq T_m\}}$. The indicator function arises because such a transmission will succeed only if $\gamma_n \geq T_m$. Therefore, the fading-averaged throughput conditioned on the feedback vector $\mathbf{b}$ is equal to

$$\mathbb{E}\left[R_m 1_{\{\gamma_n \geq T_m\}} | \mathbf{b}\right] = R_m \Pr(\gamma_n \geq T_m | \mathbf{b}). \quad (18)$$

Hence, in order to maximize the average throughput, the MCS m that maximizes $R_m \Pr(\gamma_n \geq T_m | \mathbf{b})$ must be chosen.

B. Derivation of Result 1

Given $\mathbf{b}$, we know that $Q_{q_1} \leq \gamma_1 < Q_{q_1+1}, \ldots, Q_{q_n} \leq \gamma_n < Q_{q_n+1}, \ldots, Q_{q_N} \leq \gamma_N < Q_{q_N+1}$. Recall that Q_{q_n}, as defined in Section III-A, is the lower level of the quantization region in which γ_n lies. Therefore,

$$\Pr(\gamma_n \geq T_m \mid \mathbf{b}) = \Pr\big\{\gamma_n \geq T_m \mid Q_{q_1} \leq \gamma_1 < Q_{q_1+1}, \ldots, Q_{q_N} \leq \gamma_N < Q_{q_N+1}\big\}. \quad (19)$$

Upon applying Bayes' theorem, (19) can be written as

$$\Pr(\gamma_n \geq T_m \mid \mathbf{b}) = \frac{\Pr\left(\gamma_n \geq T_m, Q_{q_1} \leq \gamma_1 < Q_{q_1+1}, \ldots, Q_{q_N} \leq \gamma_N < Q_{q_N+1}\right)}{\Pr\left(Q_{q_1} \leq \gamma_1 < Q_{q_1+1}, \ldots, Q_{q_N} \leq \gamma_N < Q_{q_N+1}\right)}. \quad (20)$$

We have $\Pr(\gamma_n \geq T_m | \mathbf{b}) = 0$, for $T_m \geq Q_{q_n+1}$. For $T_m < Q_{q_n+1}$, writing (20) in terms of the joint PDF $f_{\boldsymbol{\gamma}}(\cdot)$ of the vector $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)$ of SC SNRs yields

$$\Pr(\gamma_n \geq T_m \mid \mathbf{b}) = \frac{\int_{Q_{q_1}}^{Q_{q_1+1}} \cdots \int_{\max\{T_m, Q_{q_n}\}}^{Q_{q_n+1}} \cdots \int_{Q_{q_N}}^{Q_{q_N+1}} f_{\boldsymbol{\gamma}}(\mathbf{v})\, d\mathbf{v}}{\int_{Q_{q_1}}^{Q_{q_1+1}} \cdots \int_{Q_{q_n}}^{Q_{q_n+1}} \cdots \int_{Q_{q_N}}^{Q_{q_N+1}} f_{\boldsymbol{\gamma}}(\mathbf{v})\, d\mathbf{v}}. \quad (21)$$

The joint PDF of the exponentially correlated SC SNRs is given by [17], [18]

$$f_{\boldsymbol{\gamma}}(\mathbf{v}) = \frac{e^{-\frac{d\left[v_1 + v_N + \left(1+\rho^2\right)\sum_{i=2}^{N-1} v_i\right]}{\bar{\gamma}\left(1-\rho^2\right)}}}{\Gamma(d)\left(\frac{\bar{\gamma}}{d}\right)^d \left(1-\rho^2\right)^{(N-1)d}} \sum_{t=0}^{\infty} \lambda^t \times \sum_{\substack{w_1 \geq 0, \ldots, w_{N-1} \geq 0 \\ w_1 + \cdots + w_{N-1} = t}} \frac{v_1^{w_1+d-1} v_N^{w_{N-1}+d-1} \left(\prod_{i=2}^{N-1} z_i\right)}{\prod_{j=1}^{N-1} w_j!\, \Gamma\left(w_j + d\right)}, \quad (22)$$

where $\lambda = \left(d\rho / \left(\bar{\gamma}\left(1-\rho^2\right)\right)\right)^2$ and $z_i = v_i^{w_{i-1}+w_i+d-1}$. Substituting (22) in (21), evaluating the integrals in terms of incomplete gamma functions, and using $\Psi_n^{(m)}(\mathbf{b}) = R_m \Pr(\gamma_n \geq T_m | \mathbf{b})$ yields Result 1.

C. Proof of Lemma 2

Let $R(x \mid b_n = i)$ denote the fading-averaged throughput of TORA conditioned on $b_n = i$, for $i \in \{0, 1\}$, for a user located at a distance x from the BS. Similarly, for CRA, let $R_{\text{conv}}(x \mid b_n = i)$ denote the conditional fading-averaged throughput.

By definition of m' in Section III-D, $R_{\text{conv}}(x \mid b_n = 1) = R_{m'}$. Furthermore, $R_{\text{conv}}(x \mid b_n = 0) = 0$. Hence, the throughput gain $\Delta R(x)$ of a user located at a distance x from the BS equals

$$\begin{aligned} \Delta R(x) &= \left[R(x \mid b_n = 1) - R_{m'}\right] \Pr(b_n = 1 | X = x) \\ &\quad + R(x \mid b_n = 0) \Pr(b_n = 0 | X = x) \\ &\geq \left[R(x \mid b_n = 1) - R_{m'}\right] \Pr(b_n = 1 | X = x). \end{aligned} \quad (23)$$

From the SNR expression in (1), we get

$$\Pr(b_n = 1|X = x) = \Pr(\gamma_n > Q_2|X = x) = \exp\left(-\frac{Q_2\sigma^2 x^\alpha}{P_{\mathrm{T}}\beta}\right). \quad (24)$$

We now evaluate $R(x \mid b_n = 1)$. From (12), it follows that the feedback-conditioned goodput of MCS m when $b_n = 1$ is given by

$$\Psi_n^{(m)}(\mathbf{b}) = R_m e^{-\frac{(\max\{T_m, Q_2\} - Q_2)}{\bar{\gamma}}}. \quad (25)$$

Therefore, when $Q_2 < T_M$, it follows that $\Psi_n^{(i)}(\mathbf{b}) = R_i$, for $1 \le i \le m'$.

MCS M is chosen by TORA if its feedback-conditioned goodput exceeds that of all the other MCSs, i.e., if

$$R_M e^{\frac{(Q_2 - T_M)}{\bar{\gamma}}} > \max\left\{R_1, \ldots, R_{m'}, R_{m'+1}e^{\frac{(Q_2 - T_{m'+1})}{\bar{\gamma}}}, \ldots, R_{M-1}e^{\frac{(Q_2 - T_{M-1})}{\bar{\gamma}}}\right\}. \quad (26)$$

Using $R_1 < R_2 < \cdots < R_{m'}$, rearranging terms, and substituting $T_m = (2^{R_m} - 1)/\eta$, the inequality in (26) can be reformulated in terms of $\bar{\gamma}$ as

$$\bar{\gamma} > \max\left\{\frac{2^{R_M} - 2^{R_{M-1}}}{\eta \ln\left(\frac{R_M}{R_{M-1}}\right)}, \cdots, \frac{2^{R_M} - 2^{R_{m'+1}}}{\eta \ln\left(\frac{R_M}{R_{m'+1}}\right)}, \frac{2^{R_M} - 2^{\log_2(1+\eta Q_2)}}{\eta \ln\left(\frac{R_M}{R_{m'}}\right)}\right\}. \quad (27)$$

By the definition of m', we know that $R_{m'} \le \log_2(1 + \eta Q_2) \le R_{m'+1}$. It can be shown that the function $f(y) = \left(2^{R_M} - 2^y\right) / \left(\eta \ln\left(\frac{R_M}{y}\right)\right)$, for $y \in (0, R_M)$ and $\eta > 0$, is a monotonically increasing function in y. Hence, (27) simplifies to

$$\bar{\gamma} > \frac{2^{R_M} - 2^{R_{M-1}}}{\eta \ln\left(\frac{R_M}{R_{M-1}}\right)}. \quad (28)$$

Writing $\bar{\gamma}$ in terms of the distance x using (1), we see that MCS M will be chosen if $x < x_{M-1}$, where $x_{M-1} = \left(P_T\beta\eta \ln\left(\frac{R_M}{R_{M-1}}\right) / \left(\sigma^2\left(2^{R_M} - 2^{R_{M-1}}\right)\right)\right)^{\frac{1}{\alpha}}$. Since MCS M is used, it follows from (25) that

$$R(x \mid b_n = 1) = R_M \exp\left(-\frac{\sigma^2 (T_M - Q_2) x^\alpha}{P_T\beta}\right). \quad (29)$$

Upon substituting (29) and the expression for $\Pr(b_n = 1|X = x)$ from (24) in (23), we get

$$\Delta R(x) \ge R_M e^{-\frac{T_M\sigma^2 x^\alpha}{P_{\mathrm{T}}\beta}} - R_{m'}e^{-\frac{Q_2\sigma^2 x^\alpha}{P_{\mathrm{T}}\beta}}, \quad \text{for } 0 \le x < x_{M-1}. \quad (30)$$

Similarly, MCS $M - 1$ will be chosen if $\bar{\gamma} \le \left(2^{R_M} - 2^{R_{M-1}}\right) / \left(\eta \ln\left(\frac{R_M}{R_{M-1}}\right)\right)$ and

$$\bar{\gamma} > \max\left\{\frac{2^{R_{M-1}} - 2^{R_{M-2}}}{\eta \ln\left(\frac{R_{M-1}}{R_{M-2}}\right)}, \cdots, \frac{2^{R_{M-1}} - 2^{R_{m'+1}}}{\eta \ln\left(\frac{R_{M-1}}{R_{m'+1}}\right)}, \frac{2^{R_{M-1}} - 2^{\log_2(1+\eta Q_2)}}{\eta \ln\left(\frac{R_{M-1}}{R_{m'}}\right)}\right\}. \quad (31)$$

As above, it can be shown that this is equivalent to $\bar{\gamma} > \left(2^{R_{M-1}} - 2^{R_{M-2}}\right) / \left(\eta \ln\left(\frac{R_{M-1}}{R_{M-2}}\right)\right)$. Hence, MCS $M - 1$ is chosen if $\left(2^{R_{M-1}} - 2^{R_{M-2}}\right) / \left(\eta \ln\left(\frac{R_{M-1}}{R_{M-2}}\right)\right) < \bar{\gamma} \le \left(2^{R_M} - 2^{R_{M-1}}\right) / \left(\eta \ln\left(\frac{R_M}{R_{M-1}}\right)\right)$. Writing $\bar{\gamma}$ in terms of x, we get that MCS $M - 1$ is chosen if $x_{M-1} \le x < x_{M-2}$, where $x_{M-2} = P_T\beta\eta \ln\left(\frac{R_{M-1}}{R_{M-2}}\right) / \left(\sigma^2\left(2^{R_{M-1}} - 2^{R_{M-2}}\right)\right)$. The feedback-conditioned goodput for MCS $M - 1$, for $x_{M-1} \le x < x_{M-2}$, is given by

$$R(x \mid b_n = 1) = R_{M-1} \exp\left(-\frac{\sigma^2 (T_{M-1} - Q_2) x^\alpha}{P_T\beta}\right). \quad (32)$$

Therefore,

$$\Delta R(x) \ge R_{M-1}e^{-\frac{T_{M-1}\sigma^2 x^\alpha}{P_{\mathrm{T}}\beta}} - R_{m'}e^{-\frac{Q_2\sigma^2 x^\alpha}{P_{\mathrm{T}}\beta}}, \quad \text{for } x_{M-1} \le x < x_{M-2}. \quad (33)$$

Proceeding similarly for MCSs $M - 2, \ldots, m'$ yields (13).

D. Proof of Result 2

The fading- and user location-averaged cell throughput gain ΔR_{cell} is given by

$$\Delta R_{\mathrm{cell}} = \int_0^{r_c} \Delta R(x) f_X(x)\, dx, \quad (34)$$

where $f_X(x) = 2x/r_c^2$, for $0 \le x \le r_c$. Substituting the lower bound for $\Delta R(x)$ for $Q_2 < T_M$ from Lemma 2, we get

$$\Delta R_{\mathrm{cell}} \ge \sum_{i=m'+1}^{M} \int_{x_i}^{x_{i-1}} \left[R_i \exp\left(-\frac{T_i\sigma^2 x^\alpha}{P_{\mathrm{T}}\beta}\right) - R_{m'} \times \exp\left(-\frac{Q_2\sigma^2 x^\alpha}{P_{\mathrm{T}}\beta}\right)\right] \frac{2x}{r_c^2}\, dx. \quad (35)$$

Evaluating the above integral in terms of incomplete gamma functions yields (14).

REFERENCES

[1] *IEEE Standard for Local and Metropolitan Area Networks—Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Standard 802.16-2004, Oct. 2004.

[2] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution, From Theory to Practice*, 2nd ed. Hoboken, NJ, USA: Wiley, 2009.

[3] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[4] J. Leinonen, J. Hamalainen, and M. Juntti, "Capacity analysis of downlink MIMO-OFDMA resource allocation with limited feedback," *IEEE Trans. Commun.*, vol. 61, no. 1, pp. 120–130, Jan. 2013.

[5] Y.-J. Choi and S. Rangarajan, "Analysis of best channel feedback and its adaptive algorithms for multicarrier wireless data systems," *IEEE Trans. Mobile Comput.*, vol. 10, no. 8, pp. 1071–1082, Aug. 2011.
[6] H. Ahmed, K. Jagannathan, and S. Bhashyam, "Queue-aware optimal resource allocation for the LTE downlink with best *M* subband feedback," *IEEE Trans. Wireless Commun.*, vol. 14, no. 9, pp. 4923–4933, Sep. 2015.
[7] F. Flóren, O. Edfors, and B.-A. Molin, "The effect of feedback quantization on the throughput of a multiuser diversity scheme," in *Proc. GLOBECOM*, Dec. 2015, pp. 497–501.
[8] L. Li, M. Pesavento, and A. B. Gershman, "Downlink opportunistic scheduling with low-rate channel state feedback: Error rate analysis and optimization of the feedback parameters," *IEEE Trans. Commun.*, vol. 58, no. 10, pp. 2871–2879, Oct. 2010.
[9] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4363–4372, Nov. 2007.
[10] H. Jin and V. C. M. Leung, "One bit feedback for CDF-based scheduling with resource sharing constraints," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 6281–6291, Dec. 2013.
[11] V. Hassel, D. Gesbert, M.-S. Alouini, and G. E. Oien, "A threshold-based channel state feedback algorithm for modern cellular systems," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2422–2426, Jul. 2007.
[12] A. H. Nguyen and B. D. Rao, "CDF scheduling methods for finite rate multiuser systems with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3086–3096, Jun. 2015.
[13] J. Chen, R. A. Berry, and M. L. Honig, "Limited feedback schemes for downlink OFDMA based on sub-channel groups," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1451–1461, Oct. 2008.
[14] A. Karthik and N. B. Mehta, "An opportunistic, fast, and distributed subchannel and user-pairing algorithm for OFDMA," *IEEE Trans. Commun.*, vol. 60, no. 3, pp. 767–778, Mar. 2012.
[15] D. Park and B. G. Lee, "QoS support by using CDF-based wireless packet scheduling in fading channels," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 2051–2061, Nov. 2006.
[16] B. Makki and T. Eriksson, "Feedback subsampling in temporally-correlated slowly-fading channels using quantized CSI," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2282–2294, Jun. 2013.
[17] V. A. Aalo and T. Piboongungon, "On the multivariate generalized gamma distribution with exponential correlation," in *Proc. GLOBECOM*, Nov. 2005, pp. 1229–1233.
[18] J. Zhang, M. Matthaiou, G. K. Karagiannidis, and L. Dai, "On the multivariate gamma–gamma distribution with arbitrary correlation and applications in wireless communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3834–3840, May 2016.
[19] J. Francis, N. B. Mehta, and S. N. Ananya, "Best-*M* feedback in OFDM: Base-station-side estimation and system implications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3616–3627, May 2016.
[20] L. Pinals and M. Vu, "Maximum entropy quantization for link-state adaptation in two-way relaying," in *Proc. MILCOM*, Oct. 2015, pp. 1691–1696.
[21] J. Francis and N. B. Mehta, "Throughput-optimal scheduling and rate adaptation for reduced feedback best-*M* scheme in OFDM systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 3053–3065, Jul. 2017.
[22] R. Andreotti, T. Wang, V. Lottici, F. Giannetti, and L. Vandendorpe, "Resource allocation via max–min goodput optimization for BIC-OFDMA systems," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2412–2426, Jun. 2016.
[23] A. J. Goldsmith, *Wireless Communications*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[24] G. L. Stüber, *Principles of Mobile Communications*, 2nd ed. New York, NY, USA: Springer, 2011.
[25] Y. Li, L. Zhang, L. J. Cimini, and H. Zhang, "Statistical analysis of MIMO beamforming with co-channel unequal-power MIMO interferers under path-loss and Rayleigh fading," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3738–3748, Aug. 2011.
[26] J. B. Andersen, "Array gain and capacity for known random channels with multiple element arrays at both ends," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 11, pp. 2172–2178, Nov. 2000.
[27] R. Aggarwal, M. Assaad, C. E. Koksal, and P. Schniter, "Joint scheduling and resource allocation in the OFDMA downlink: Utility maximization under imperfect channel-state information," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5589–5604, Nov. 2011.
[28] A. P. T. Lau and F. R. Kschischang, "Feedback quantization strategies for multiuser diversity systems," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1386–1400, Apr. 2007.
[29] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 1, 3rd ed. Hoboken, NJ, USA: Wiley, 2008.
[30] J. M. Hammersley and D. C. Handscomb, *Monte Carlo Methods*, 1st ed. Dordrecht, The Netherlands: Springer, 2008.
[31] D. Park, H. Seo, H. Kwon, and B. G. Lee, "Wireless packet scheduling based on the cumulative distribution function of user transmission rates," *IEEE Trans. Commun.*, vol. 53, no. 11, pp. 1919–1929, Nov. 2005.

Vineeth Kumar (S'16) received the B.Tech. degree in electronics and communications engineering from Amrita University, Amritapuri, Kerala, in 2011, and the M.Tech. degree in signal processing from the National Institute of Technology, Calicut, in 2014. He is currently pursuing the Ph.D. degree with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore. His research interests include the design and performance analysis of limited feedback schemes for rate adaptation and resource allocation in OFDM systems.

Neelesh B. Mehta (S'98–M'01–SM'06) received the B.Tech. degree in electronics and communications engineering from IIT Madras, Chennai, in 1996, and the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology, Pasadena, USA, in 1997 and 2001, respectively. He is currently a Professor with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore. He is a fellow of the Indian National Academy of Engineering and the National Academy of Sciences India. He serves as the Chair of the Executive Editorial Committee of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS. He served on the Board of Governors of the IEEE ComSoc from 2012 to 2015.