Exploiting Quantized Channel Norm Feedback Through Conditional Statistics in Arbitrarily Correlated MIMO Systems

EMIL BJÖRNSON, DAVID HAMMARWALL, AND BJÖRN OTTERSTEN

Stockholm 2009

KTH Royal Institute of Technology
ACCESS Linnaeus Center
Signal Processing Lab

Exploiting Quantized Channel Norm Feedback Through Conditional Statistics in Arbitrarily Correlated MIMO Systems

Emil Björnson, *Student Member, IEEE*, David Hammarwall, *Member, IEEE*, and Björn Ottersten, *Fellow, IEEE*

Abstract—In the design of narrowband multi-antenna systems, a limiting factor is the amount of channel state information (CSI) available at the transmitter. This is especially evident in multi-user systems, where the spatial user separability determines the multiplexing gain, but it is also important for transmission-rate adaptation in single-user systems. To limit the feedback load, the unknown and multi-dimensional channel needs to be represented by a limited number of bits. When combined with long-term channel statistics, the norm of the channel matrix has been shown to provide substantial CSI that permits efficient user selection, linear precoder design, and rate adaptation. Herein, we consider quantized feedback of the squared Frobenius norm in a Rayleigh fading environment with arbitrary spatial correlation. The conditional channel statistics are characterized and their moments are derived for both identical, distinct, and sets of repeated eigenvalues. These results are applied for minimum mean square error (MMSE) estimation of signal and interference powers in single- and multi-user systems, for the purpose of reliable rate adaptation and resource allocation. The problem of efficient feedback quantization is discussed and an entropy-maximizing framework is developed where the post-user-selection distribution can be taken into account in the design of the quantization levels. The analytic results of this paper are directly applicable in many widely used communication techniques, such as space-time block codes, linear precoding, space division multiple access (SDMA), and scheduling.

Index Terms—Channel gain feedback, estimation, MIMO systems, norm-conditional statistics, quantization, Rayleigh fading, space division multiple access (SDMA).

I. INTRODUCTION

WIRELESS communication systems with antenna arrays at both the transmitter and receiver have the ability of greatly improving the capacity over single-antenna systems.

The potential gains have been shown for narrowband channels in [1] and [2], under the assumption of independent and identically distributed zero-mean complex Gaussian channel coefficients between the transmit and receive antennas. Such channels are often referred to as uncorrelated Rayleigh fading, since there is no correlation in the spatial dimension and the envelope of the received signal is Rayleigh distributed. From a mathematical point of view, uncorrelated Rayleigh fading channels occur naturally when the antenna separation is large and the scattering in the propagation channel is sufficiently rich. However, it has been shown experimentally that the channel coefficients are often spatially correlated in outdoor scenarios [3], and correlation frequently occurs in indoor environments as well [4], [5]. This motivates the analysis of the more general case of Rayleigh fading where the channel coefficients are arbitrarily correlated.

Channel variations are normally characterized by small-scale and large-scale fading [6]. The former describes changes in the signal paths of the order of the carrier wavelength and is time- and frequency-dependent. To avoid the frequency dependency we consider narrowband block-fading channels; that is, the channel matrix is constant for a block of symbols and then updated independently from the assumed Gaussian distribution for the next block. The large-scale fading corresponds to variations in the channel statistics due to effects like shadowing by buildings and power decay due to propagation distance. These effects are typically frequency independent and slowly varying in time. Hence, the transmitter and receiver can keep track on the statistics by reverse-link estimation or a negligible feedback overhead.

In single-user multiple-input multiple-output (MIMO) systems, the small-scale fading can be mitigated with using orthogonal space-time block codes (OSTBCs) [7]–[9]. Using only statistical channel state information (CSI) at the transmitter, the capacity can be unexpectedly good if linear precoding takes care of the spatial correlation [9]–[12]. In practice, a small amount of channel gain feedback is however necessary for rate adaptation to achieve this performance. In multi-user MIMO systems the situation is somewhat different, because the multi-user diversity gain depends on the amount of instantaneous CSI available at the transmitter [13], [14]. This CSI can be exploited to schedule users for transmission on time-frequency slots and spatial directions in which they experience particularly strong gains. Unfortunately, the amount of feedback needed to achieve full CSI is prohibitive in many realistic scenarios. Therefore, the design of

limited feedback systems that capture most of the performance has been an active research topic.

Many multi-user limited feedback systems are based on linear precoding. Although this approach is only asymptotically optimal in the number of users [15], the loss in performance comes with a substantial decrease in complexity compared with non-linear precoding (e.g., optimal dirty-paper coding [16]). One approach to linear precoding in space division multiple access (SDMA) is to allocate users to a set of beams based on feedback of their achieved channel gains. These beams can either be generated randomly [14] or belong to a fixed grid of beams [17]. Another approach is to design and adapt the precoder matrix to statistical user information and feedback of instantaneous CSI. This can be implemented in a zero-forcing fashion [18]–[20], where the co-user interference is made zero (for full CSI) or statistically small and manageable (for partial CSI). Although this strong zero-forcing condition is suboptimal, it provides a simple design structure and can achieve close-to-optimal performance if the amount of feedback is correctly scaled with the signal-to-interference-and-noise ratio (SINR) [18]. In general, the type of approach that is most favorable depends on various system parameters, such that coherence time, number of users, spatial correlation, and average SINR.

Feedback of quantized gain information plays an important role in the design of both user-selection algorithms and linear precoders. In [21], channel norm based user-selection was shown to provide close-to-optimal performance asymptotically in the number of transmit antennas. When considering zero-forcing precoding and limited feedback, it was proposed in [18] that each user should feed back its normalized channel vector using a codebook and calculate a regular zero-forcing precoder. Additional feedback of the instantaneous channel norm is however required to estimate the SINR and perform reliable rate adaptation [22]. In spatially correlated systems, the long-term statistics provide directional information and feedback of the channel norm is sufficient to perform efficient statistical zero-forcing [19] and estimate the instantaneous SINR that is used for rate adaptation [23]. In neither of these papers, channel gain quantization or multi-antenna receivers are considered. With multiple antennas at both sides, more degrees of freedom are available in the interference cancellation, but the precoder and receiver combining design problem becomes considerably more difficult. Some of these problems were addressed in [20].

Herein, we analyze the impact of channel gain information on Rayleigh fading MIMO systems with arbitrary spatial correlation. The conditional statistics and minimum mean square error (MMSE) framework derived in [23] for correlated systems with single-antenna users are generalized to cover general fading environments, multi-antenna users, and quantized gain information. The contributions to communication are an entropy-maximizing quantization framework that can be applied to gain feedback and the derivations of closed-form estimators of the instantaneous SINR in single- and multi-user systems, using such gain feedback. These results can be applied to handle gain feedback

and rate adaptation in system both with and without additional feedback of directional channel information.

Notations

For notational convenience we use boldface (lower case) for column vectors, $\mathbf{x}$, and (upper case) for matrices, $\mathbf{X}$. With $\mathbf{X}^T$, $\mathbf{X}^H$, and $\mathbf{X}^*$ we denote the transpose, the conjugate transpose, and the conjugate of $\mathbf{X}$, respectively. The Kronecker product of two matrices $\mathbf{X}$ and $\mathbf{Y}$ is denoted $\mathbf{X} \otimes \mathbf{Y}$, $\mathrm{vec}(\mathbf{X})$ is the column vector obtained by stacking the columns of $\mathbf{X}$, and $\mathrm{diag}(x_1, \ldots, x_N)$ is the N-by-N diagonal matrix with $x_1, \ldots, x_N$ at the main diagonal. If the ijth element of a matrix $\mathbf{X}$ is x_{ij}, then $[\mathbf{X}]_{ij} = x_{ij}$. The distribution of circularly symmetric complex Gaussian vectors is denoted $\mathcal{CN}(\bar{\mathbf{x}}, \mathbf{R})$, with mean value $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{R}$.

The notation $\triangleq$ is used for definitions. The squared 2-norm of a vector $\mathbf{x}$ is denoted $\|\mathbf{x}\|^2$ and the squared Frobenius norm of a matrix $\mathbf{X}$ is denoted $\|\mathbf{X}\|^2$, and both are defined as the sum of the squared absolute values of all the elements. The sum of absolute values of all the elements in $\mathbf{x}$ is denoted $\|\mathbf{x}\|_1$. If $\mathcal{S}$ is a set, then the set members are denoted $\mathcal{S}(1), \ldots, \mathcal{S}(|\mathcal{S}|)$, where $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$.

Let $\mathbf{x} = [x_1, \ldots, x_n]^T$. The generalized Heaviside step function $H_a(\mathbf{x})$ is 1 if $x_i \geq 0$ for all i and $\sum_{i=1}^n x_i \geq a$, and 0 otherwise. The function $H_{a,b}(\mathbf{x})$ is 1 if $x_i \geq 0$, for all i, and $a \leq \sum_{i=1}^n x_i < b$, and 0 otherwise. Finally, $\delta(x)$ denotes Dirac's delta function.

A. System Model

Consider the downlink of a communication system with a single base station equipped with an array of n_T antennas and several mobile users, each with an array of n_R antennas. The symbol-sampled complex baseband equivalent of the narrowband flat-fading channel to user k is represented by $\mathbf{H}_k \in \mathbb{C}^{n_R \times n_T}$. The elements of $\mathbf{H}_k$ are modeled as Rayleigh fading with arbitrary correlation, and thus we assume that $\mathrm{vec}(\mathbf{H}_k) \in \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$. The received vector $\mathbf{y}_k(t) \in \mathbb{C}^{n_R}$ of user k at symbol slot t is modeled as

$$\mathbf{y}_k(t) = \mathbf{H}_k \mathbf{x}(t) + \mathbf{n}_k(t) \qquad (1)$$

where the vector of transmitted signals is denoted $\mathbf{x}(t) \in \mathbb{C}^{n_T}$ and the power of the system is normalized such that $\mathbf{n}_k(t) \in \mathbb{C}^{n_R}$ is white noise with elements that are distributed as $\mathcal{CN}(0, 1)$.

The system model in (1) depends on three different time scales. The variations in the matrix $\mathbf{H}_k$ are modeled by quasi-static block-fading; that is, the channel realization is constant for a block of symbols and then modeled as independent in the next block. Within a block, only the noise $\mathbf{n}_k(t)$ and the transmitted signal $\mathbf{x}(t)$ are changing. The statistics change very slowly, measured in the number of blocks, and it is therefore assumed that the current correlation matrix $\mathbf{R}_k$ is known to both the base station and user k.

B. Feedback-Based Estimation of Weighted Channel Norms

To achieve reliable rate estimation and exploit the spatial and multi-user diversity, the transmitter often needs more information than just the channel statistics. Such partial and instantaneous CSI can be estimated at the receiver side and then fed back to the transmitter [24]. When the channel conditions change rapidly with time, the number of feedback symbols spent on achieving partial CSI not only reduces the time the information can be used at the transmitter before it is outdated but also the number of symbols available for data transmission on the reverse link. Hence, the feedback needs to represent some limited amount of information that can be described efficiently by a small number of bits.

In a block-fading environment, the feedback system can in principle be described as a cyclical system that estimates and feeds back partial CSI in the beginning of each block to improve the system performance during the rest of the block. The results herein are however not limited to this type of fading. For simplicity, we assume that there exists an error-free feedback channel from each mobile user to the base station.

The instantaneous CSI can be divided into directional information and gain information, herein the latter will be considered. Throughout this paper, we consider the estimation of weighted squared Frobenius norms of the channel at the transmitter [20], [23], where the weights are known at the transmitter but not necessarily at the receiver. On the contrary, the channel is only perfectly known to the receiver and any instantaneous CSI exploited at the transmitter must be conveyed over the limited feedback link. The generic estimation problem that we focus on is

$$\text{Estimate} \quad \|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2$$
$$\text{given} \quad \rho \triangleq \|\widetilde{\mathbf{H}}\|^2 \text{ or a quantized version } a \leq \rho < b. \quad (2)$$

In this formulation, we have the weighting matrix $\mathbf{B} \in \mathbb{C}^{n_B \times n_C n_D}$ and the effective channel $\widetilde{\mathbf{H}} \triangleq \mathbf{CHD}$, where $\mathbf{C} \in \mathbb{C}^{n_C \times n_R}$ and $\mathbf{D} \in \mathbb{C}^{n_T \times n_D}$ are matrices known to the receiver. In the area of communication, two interesting feedback and estimation scenarios can be formulated in terms of the generic problem.

1) The receive combiner matrix and precoder matrix are known to the receiver and are used as $\mathbf{C}$ and $\mathbf{D}$, respectively. The squared norm of the effective channel $\widetilde{\mathbf{H}} = \mathbf{CHD}$ is fed back to the transmitter. This information is used to estimate the weighted squared norm $\|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2$, which is either the total channel gain ($\mathbf{B} = \mathbf{I}$) or the gain in a certain spatial subspace.

2) Either the receive combiner matrix, the precoder matrix, or both matrices are unknown to the receiver at the time of feedback. In these cases, the effective channel becomes $\widetilde{\mathbf{H}} = \mathbf{HD}$, $\widetilde{\mathbf{H}} = \mathbf{CH}$, or $\widetilde{\mathbf{H}} = \mathbf{H}$, respectively, and the squared norm $\|\widetilde{\mathbf{H}}\|^2$ is fed back. This information is used to estimate the weighted squared norm $\|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2$, where $\mathbf{B}$ may represent receive combiner and/or precoder matrices that are known to the transmitter.

The results of this paper are independent of the quantization, but a quantization framework is proposed in Section III and adapted to multi-user systems in Section IV-B.

C. Outline

In Section II, we analyze the special case of feedback of $\rho_k = \|\mathbf{H}_k\|^2$ with a diagonal correlation matrix $\mathbf{R}_k$. Closed-form expressions of the conditional moments of the elements in $\mathbf{H}_k$ are derived for both exact norm feedback and a quantized norm. A short overview of the applications of these results in renewal theory is provided. In Section III, the results are generalized for communication purposes. A general entropy-maximizing quantization framework is presented and the results of Section II are used to characterize the distribution of the effective squared channel norm and to derive an MMSE estimator of weighted squared norms, given quantized norm information. Section IV shows how these results are applicable on MMSE estimation of signal/interference powers and rate adaptation in single- and multi-user systems. Some of the results are illustrated numerically in Section V and conclusions are drawn in Section VI.

II. ANALYSIS OF ZERO-MEAN COMPLEX GAUSSIAN VECTORS WITH NORM INFORMATION

In this section, we consider an N-dimensional vector $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathbb{C}^N$, for $N \geq 1$, with zero-mean and independent complex Gaussian entries—that is, $\mathbf{v} \in \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$. First, the distribution of the squared norm $\rho \triangleq \|\mathbf{v}\|^2$ will be presented. Then, expressions of the pth order conditional moment and (p_i, p_j)th order conditional cross-moment are derived for the cases of either an exactly known norm ρ or a known interval $a \leq \rho < b$ (representing a quantization of $\|\mathbf{v}\|^2$). These moments will be used in Section IV to derive a MMSE estimator of weighted squared norms as formulated in (2), and their corresponding mean squared errors (MSEs).

Without loss of generality, we assume that the diagonal elements, $\lambda_i > 0$, of the positive definite correlation matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_N)$ are ordered such that elements with identical distributions have adjacent indices. When analyzing $\|\mathbf{v}\|^2$, we distinguish between three different cases, depending on the distinctness of the variances λ_i (hereafter called eigenvalues):

- identical eigenvalues: $\lambda_1 = \cdots = \lambda_N = \mu_1$, for some μ_1;
- distinct eigenvalues: $\lambda_i \neq \lambda_j$, for all $i \neq j$;
- one or more sets of repeated eigenvalues among λ_j.

While the former two cases are clearly structured and commonly treated in literature, the third case needs further specification [25]. Let the M distinct values among the eigenvalues be $\mu_1, \ldots, \mu_M$ ($1 \leq M \leq N$), with the strictly positive multiplicities $r_1, \ldots, r_M$ ($\sum_{i=1}^M r_i = N$). Then, we have the characterization

$$\lambda_1 = \cdots = \lambda_{r_1} = \mu_1$$
$$\vdots$$
$$\lambda_{r_1 + \cdots + r_{M-1} + 1} = \cdots = \lambda_{r_1 + \cdots + r_M} = \mu_M. \quad (3)$$

To simplify the notation, we gather the eigenvalue multiplicities in a vector $\mathbf{r} = [r_1, \ldots, r_M]^T$ and define the function $r(j) = m$ that gives the group index of λ_j from $\{1, \ldots, M\}$

(i.e., $1 \leq m \leq M$ is the integer that satisfies $\sum_{i=1}^{m-1} r_i < j \leq \sum_{i=1}^{m} r_i$).

These three cases are directly applicable to systems with uncorrelated fading (identical eigenvalues), correlated fading (distinct eigenvalues), and Kronecker-structured systems (see Section III) with correlation at either the transmitter or receiver (repeated eigenvalues with either multiplicity n_R or n_T).

Next, the probability density function (pdf) of the squared norm $\|\mathbf{v}\|^2 \triangleq \sum_{i=1}^{N} |v_i|^2$ will be given for the three cases described above. Since $v_i \in \mathcal{CN}(0, \lambda_i)$ for all i, then $|v_i|^2 \in \mathrm{Exp}(1/\lambda_i)$ and the squared norm $\|\mathbf{v}\|^2$ is the sum of independent exponentially distributed variables (each with the rate $1/\lambda_i$). In the case of identical eigenvalues, the pdf is that of a scaled χ_{2N}^2-distribution (i.e., an Erlang distribution):

$$f_{\|\mathbf{v}\|^2}^{\mathrm{ident}}(\rho) = \frac{\rho^{N-1} e^{-\rho/\mu_1}}{(N-1)! \mu_1^N} H_0(\rho) \qquad (4)$$

where $H_0(\rho)$ is the Heaviside step function. In the case of distinct eigenvalues, the pdf of $\|\mathbf{v}\|^2$ is well-known in the field of renewal theory [25] and was derived for communications purposes in [23]:

$$f_{\|\mathbf{v}\|^2}^{\mathrm{dist}}(\rho) = \sum_{k=1}^{N} \frac{e^{-\rho/\lambda_k}}{\lambda_k \prod_{\substack{l=1 \\ l \neq k}}^{N} \left(1 - \frac{\lambda_l}{\lambda_k}\right)} H_0(\rho). \qquad (5)$$

In the third case, with repeated eigenvalues that satisfy (3), the pdf was derived in [25] and [26]:

$$f_{\|\mathbf{v}\|^2}^{\mathrm{repeat}}(\rho) = H_0(\rho) \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \sum_{k=1}^{M} \sum_{l=1}^{r_k} \frac{\Psi_{k,l,\mathbf{r}}}{(r_k - l)!} (-\rho)^{r_k - l} e^{-\rho/\mu_k} \qquad (6)$$

where

$$\Psi_{k,l,\mathbf{r}} = (-1)^{r_k - 1} \sum_{\mathbf{i} \in \Omega_{k,l}} \prod_{j \neq k} \binom{i_j + r_j - 1}{i_j} \left(\frac{1}{\mu_j} - \frac{1}{\mu_k}\right)^{-(r_j + i_j)} \qquad (7)$$

with $\mathbf{i} = [i_1, \ldots, i_M]^T$ from the set $\Omega_{k,l}$ of all partitions of $l - 1$ (with $i_k = 0$) defined as

$$\Omega_{k,l} = \left\{ [i_1, \ldots, i_M] \in \mathbb{Z}^M; \sum_{j=1}^{M} i_j = l - 1, i_k = 0, i_j \geq 0 \ \forall j \right\}. \qquad (8)$$

One remark is that the pdf in (6) actually becomes that in (4) if $M = 1$ and that in (5) if $M = N$. Since the expressions with identical and distinct eigenvalues are simpler and very useful in practice, we will distinguish between all three cases throughout the paper.

A. Conditional Statistics: Known Norm Value or Interval

Next, we will consider the conditional statistics of the elements of $\mathbf{v}$ when its squared norm $\|\mathbf{v}\|^2$ is known exactly or in a quantized way. The absolute value and the phase of a complex Gaussian variable are independent [16]. Thus, $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$ can be identically expressed as $\mathbf{v} = $

$[|v_1|e^{j\phi_1}, \ldots, |v_N|e^{j\phi_N}]^T$, where the phase ϕ_i is uniformly distributed in $[0, 2\pi)$ and $|v_i|^2 \in \mathrm{Exp}(1/\lambda_i)$ for all i. Observe that information regarding $\|\mathbf{v}\|^2 \triangleq \sum_{i=1}^{N} |v_i|^2$ will not provide any knowledge of the phases. The squared magnitudes of the individual elements, $|v_i|^2$, will however depend on $\|\mathbf{v}\|^2$.

In this section, we will derive closed-form expressions of the pth-order conditional moment of $|v_i|^2$ and (p_i, p_j)th order conditional cross-moment of $|v_i|^2$ and $|v_j|^2$. This will be done in two different cases, namely when the squared norm $\rho \triangleq \|\mathbf{v}\|^2$ is either known exactly or when a quantization is known. We denote the quantized squared norm with ϱ and it represents the information $a \leq \|\mathbf{v}\|^2 < b$, for some real-valued interval parameters. This type of quantized information can, for example, be achieved by feedback. The conditional moments derived in the section will be used in Section IV for MMSE estimation and MSE calculation of weighted squared norms in systems with either perfect or quantized squared norm feedback.

The following theorem gives closed-form expressions of the conditional moments in the case of an exactly known squared norm ρ. Although the expressions are quite simple in their structure, two elementary functions $G_{N,M,i,j}^{(1)}(\rho)$ and $G_{N,M,L,i,j,m}^{(2)}(\rho)$ are introduced to achieve a more convenient presentation. These are defined and discussed in Appendix A. Observe that the mean value of an element is given by $p = 1$, the quadratic mean by $p = 2$, and that $p_i = p_j = 1$ gives the cross-correlation.

Theorem 1 (Conditional Moments With Known Norm): Let $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ has strictly positive eigenvalues λ_i and $N \geq 2$. Define $\rho \triangleq \|\mathbf{v}\|^2$. In the case of identical eigenvalues (i.e., $\lambda_i = \mu_1$ for all i), the pth order conditional moment of $|v_i|^2$ and (p_i, p_j)th order conditional cross-moment between $|v_i|^2$ and $|v_j|^2$ ($i \neq j$) are

$$\mathbb{E}\{|v_i|^{2p}|\rho\} = \frac{(N-1)}{\rho^{N-1}}(-1)^{N-2} G_{p,N-2,1,1}^{(1)}(\rho),$$

$$\mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\rho\} = \begin{cases} \frac{(N-1)(N-2)}{\rho^{N-1}}(-1)^{N-3} G_{p_i,p_j,N-3,1,1,1}^{(2)}(\rho), & N > 2, \\ \frac{(-1)^{p_j}}{\rho} G_{p_i,p_j,1,1}^{(1)}(\rho), & N = 2. \end{cases} \qquad (9)$$

In the case of distinct eigenvalues, the corresponding moments are

$$\mathbb{E}\{|v_i|^{2p}|\rho\} = \frac{\frac{1}{\lambda_i}}{f_{\|\mathbf{v}\|^2}^{\mathrm{dist}}(\rho)} \sum_{\substack{k=1 \\ k \neq i}}^{N} \frac{e^{-\rho/\lambda_k} G_{p,0,i,k}^{(1)}(\rho)}{\lambda_k \prod_{l \notin \{i,k\}} \left(1 - \frac{\lambda_l}{\lambda_k}\right)},$$

$$\mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\rho\} = \frac{\frac{1}{\lambda_i \lambda_j}}{f_{\|\mathbf{v}\|^2}^{\mathrm{dist}}(\rho)} \begin{cases} \sum_{\substack{k=1 \\ k \notin \{i,j\}}}^{N} \frac{e^{-\rho/\lambda_k} G_{p_i,p_j,0,i,j,k}^{(2)}(\rho)}{\lambda_k \prod_{l \notin \{i,j,k\}} \left(1 - \frac{\lambda_l}{\lambda_k}\right)}, & N > 2, \\ (-1)^{p_j} e^{-\rho/\lambda_j} G_{p_i,p_j,i,j}^{(1)}(\rho), & N = 2. \end{cases} \qquad (10)$$

Finally, if the eigenvalues are nondistinct and nonidentical, let $\tilde{\mathbf{r}} = [\tilde{r}_1, \ldots, \tilde{r}_M]^T$ be the eigenvalue multiplicities when the

elements involved in the moments have been removed. The pth order conditional moment of $|v_i|^2$ and (p_i, p_j)th order conditional cross-moment between $|v_i|^2$ and $|v_j|^2$ ($i \neq j$) are

$$\mathbb{E}\{|v_i|^{2p}|\rho\} = \frac{1}{f_{\|\mathbf{v}\|^2}(\rho)} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}}$$
$$\times \sum_{k=1}^{M} \sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k - l)!} e^{-\rho/\mu_k} G_{p,\tilde{r}_k-l,r(i),k}^{(1)}(\rho)$$
$$\mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\rho\}$$
$$= \frac{1}{f_{\|\mathbf{v}\|^2}(\rho)} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}}$$
$$\times \sum_{k=1}^{M} \sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k - l)!} e^{-\rho/\mu_k} G_{p_i,p_j,\tilde{r}_k-l,r(i),r(j),k}^{(2)}(\rho). \tag{11}$$

Proof: The proof is given in Appendix C. ∎

Observe that Theorem 1 only handles the case of $N \geq 2$, but the solution in the special case of $N = 1$ is trivial: $|v_1|^{2p} = \rho^p$. The theorem generalizes the previous results of [23], where expressions of the first and second order moment and cross-correlation were derived in the special case of distinct eigenvalues.

Next, we proceed with deriving closed-form expressions of the same conditional moments and cross-moments as in the previous theorem but in the case of quantized norm information. Once again, the expressions contain some functions that are defined in Appendix A.

Theorem 2 (Conditional Moments With Known Norm Interval): Let $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ has strictly positive eigenvalues λ_i and $N \geq 2$. Let ϱ contain the quantized information $a \leq \|\mathbf{v}\|^2 < b$ (where $0 \leq a \leq b$). In the case of identical eigenvalues (i.e., $\lambda_i = \mu_1$ for all i), the pth order conditional moment of $|v_i|^2$ and (p_i, p_j)th order conditional cross-moment between $|v_i|^2$ and $|v_j|^2$ ($i \neq j$) are

$$\mathbb{E}\{|v_i|^{2p}|\varrho\}$$
$$= \frac{1}{g_\varrho^{\mathrm{ident}}} \frac{(-1)^{N-2}}{\mu_1^N (N-2)!} \widetilde{G}_{p,N-2,1,1}^{(1)}(a,b),$$
$$\mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\varrho\}$$
$$= \frac{1}{g_\varrho^{\mathrm{ident}}} \begin{cases} \frac{(-1)^{N-3}}{\mu_1^N (N-3)!} \widetilde{G}_{p_i,p_j,N-3,1,1,1}^{(2)}(a,b), & N>2, \\ \frac{(-1)^{p_j}}{\mu_1^2} \widetilde{G}_{p_i,p_j,1,1}^{(1)}(a,b), & N=2 \end{cases} \tag{12}$$

where

$$g_\varrho^{\mathrm{ident}} = \sum_{j=0}^{N-1} \frac{\left(\frac{a}{\mu}\right)^j e^{-a/\mu} - \left(\frac{b}{\mu}\right)^j e^{-b/\mu}}{j!}. \tag{13}$$

In the case of distinct eigenvalues, the corresponding moments are

$$\mathbb{E}\{|v_i|^{2p}|\varrho\} = \frac{\frac{1}{\lambda_i}}{g_\varrho^{\mathrm{dist}}} \sum_{\substack{k=1 \\ k \neq i}}^{N} \frac{\widetilde{G}_{p,0,i,k}^{(1)}(a,b)}{\lambda_k \prod_{l \notin \{i,k\}} \left(1 - \frac{\lambda_l}{\lambda_k}\right)},$$

$$\mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\varrho\} = \frac{\frac{1}{\lambda_i \lambda_j}}{g_\varrho^{\mathrm{dist}}} \begin{cases} \sum_{\substack{k=1 \\ k \notin \{i,j\}}}^{N} \frac{\widetilde{G}_{p_i,p_j,0,i,j,k}^{(2)}(a,b)}{\lambda_k \prod_{l \notin \{i,j,k\}} \left(1 - \frac{\lambda_l}{\lambda_k}\right)}, & N>2, \\ (-1)^{p_j} \widetilde{G}_{p_i,p_j,i,j}^{(1)}(a,b), & N=2 \end{cases} \tag{14}$$

where

$$g_\varrho^{\mathrm{dist}} = \sum_{k=1}^{N} \frac{e^{-a/\lambda_k} - e^{-b/\lambda_k}}{\prod_{i \neq k} \left(1 - \frac{\lambda_i}{\lambda_k}\right)}. \tag{15}$$

Finally, if the eigenvalues are nondistinct and nonidentical, let $\tilde{\mathbf{r}} = [\tilde{r}_1, \ldots, \tilde{r}_M]^T$ be the eigenvalue multiplicities when the elements involved in the moments have been removed. The pth order conditional moment of $|v_i|^2$ and (p_i, p_j)th order conditional cross-moment between $|v_i|^2$ and $|v_j|^2$ ($i \neq j$) are

$$\mathbb{E}\{|v_i|^{2p}|\varrho\}$$
$$= \frac{1}{g_\varrho^{\mathrm{repeat}}} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}}$$
$$\times \sum_{k=1}^{M} \sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k - l)!} \widetilde{G}_{p,\tilde{r}_k-l,r(i),k}^{(1)}(a,b)$$
$$\mathbb{E}\{|v_i|^{2p_i}|v_j|^{2p_j}|\varrho\}$$
$$= \frac{1}{g_\varrho^{\mathrm{repeat}}} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}}$$
$$\times \sum_{k=1}^{M} \sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k - l)!} \widetilde{G}_{p_i,p_j,\tilde{r}_k-l,r(i),r(j),k}^{(2)}(a,b) \tag{16}$$

where

$$g_\varrho^{\mathrm{repeat}} = \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \sum_{k=1}^{M} \sum_{l=1}^{r_k} \frac{\Psi_{k,l,\mathbf{r}}}{\left(-\frac{1}{\mu_k}\right)^{r_k-l+1}}$$
$$\times \sum_{j=0}^{r_k-l} \frac{\left(\frac{b}{\mu_k}\right)^j e^{-b/\mu_k} - \left(\frac{a}{\mu_k}\right)^j e^{-a/\mu_k}}{j!}. \tag{17}$$

Proof: The proof is given in Appendix C. ∎

This section will be concluded by Theorem 3 that gives the MMSE estimator of $\rho \triangleq \|\mathbf{v}\|^2$ from the quantized information $a \leq \rho < b$. Observe that the theorem completes Theorem 2 for $N = 1$.

Theorem 3 (Norm Estimation From Known Norm Interval): Let $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ has strictly positive eigenvalues λ_i. Let $\rho \triangleq \|\mathbf{v}\|^2$ and let ϱ contain the quantized information $a \leq \rho < b$ (where $0 \leq a \leq b$). The conditional pth order moment of $\|\mathbf{v}\|^2$, given ϱ, is

$$\mathbb{E}\{\rho^p|\varrho\} = \frac{1}{g_\varrho^{\mathrm{ident}}} \frac{\widetilde{G}_{N+p-1,1}^{(3)}(a,b)}{\mu_1^N (N-1)!}, \tag{18}$$

$$\mathbb{E}\{\rho^p|\varrho\} = \frac{1}{g_\varrho^{\mathrm{dist}}} \sum_{k=1}^{N} \frac{\widetilde{G}_{p,k}^{(3)}(a,b)}{\lambda_k \prod_{j \neq k} \left(1 - \frac{\lambda_j}{\lambda_k}\right)}, \tag{19}$$

and

$$\mathbb{E}\{\rho^p|\varrho\} = \frac{1}{g_\varrho^{\text{repeat}}} \prod_{\tilde{l}=1}^{M} \frac{1}{\mu_{\tilde{l}}^{r_{\tilde{l}}}} \sum_{k=1}^{M} \sum_{l=1}^{r_k} \frac{\Psi_{k,l,\mathbf{r}} \widetilde{G}_{r_k+p-l,k}^{(3)}(a,b)}{(-1)^{r_k-l}(r_k-l)!}$$

(20)

when the eigenvalues are identical (i.e., $\lambda_i = \mu_1$ for all i), distinct, or neither identical nor distinct, respectively. The variables g_ϱ^{ident}, g_ϱ^{dist}, and $g_\varrho^{\text{repeat}}$ are given in (13), (15), and (17), respectively.

Proof: The proof is given in Appendix C. ∎

In the remaining sections, the analytic results of Theorem 1, 2, and 3 will be applied to problems in wireless communications. The results of this section are however general and have important applications in other areas, for example in the analysis of N-out-of-M systems with exponential failure rates in renewal theory [25], [27]. In principle, these systems consist of M components and the system will keep running until N of them have break down. The time between the $(i-1)$th and ith component failure is distributed as $\text{Exp}(1/\lambda_i)$ (i.e., failures may change the failure rates of the surviving components). Thus, $\|\mathbf{v}\|^2$ is the time to system failure. The results herein can be used for MMSE estimation of the time between component failures, given the exact time of system failure or a time interval (e.g., if the functionality is tested only at certain occasions). Similarly, the MSE and the correlation between component failures can be calculated, and the time of system failure can be MMSE estimated, given a time interval.

III. NORM FEEDBACK AND MMSE ESTIMATION OF WEIGHTED SQUARED CHANNEL NORMS

In this section, we return to the generic estimation problem in (2) and the system model in (1). Thus, the effective channel used for norm feedback is $\widetilde{\mathbf{H}}_k = \mathbf{C}_k \mathbf{H}_k \mathbf{D}_k$, where $\text{vec}(\mathbf{H}_k) \in \mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$ and $\mathbf{C}_k, \mathbf{D}_k$ are arbitrary matrices known at the receiver. In this section, we will first develop a general entropy-maximizing quantization framework. Then, the results of Section II will be used to derive the distribution of the squared norm $\rho_k = \|\widetilde{\mathbf{H}}_k\|^2$ of the effective channel, which is necessary to apply the quantization framework to the problem at hand. Finally, we solve the estimation problem in (2) by deriving the MMSE estimator, and its MSE, of the weighted squared norm $\|\mathbf{B}_k \text{vec}(\widetilde{\mathbf{H}}_k)\|^2$, conditioned on exact or quantized feedback of ρ_k. As described in Section I-B, the weighting matrix $\mathbf{B}_k$ can represent receive combining and precoding matrices. The applications of this section on user-selection, link-adaptation, and linear precoding will be considered in Section IV. The user index will be dropped in this section for brevity.

The results herein are derived for a general positive semi-definite correlation matrix $\mathbf{R}$, but we will also give the corresponding expressions in the special case of Kronecker-structured correlation. In this widely used model, the transmit and receive side correlation are separable as $\mathbf{R} = \mathbf{R}_{\text{Tx}}^T \otimes \mathbf{R}_{\text{Rx}}$, where $\mathbf{R}_{\text{Tx}} \in \mathbb{C}^{n_T \times n_T}$ and $\mathbf{R}_{\text{Rx}} \in \mathbb{C}^{n_R \times n_R}$ are the positive semi-definite transmit and receive correlation matrices, respectively. As a result, the matrix $\mathbf{H}$ can in this case be decomposed as

$$\mathbf{H} = \mathbf{R}_{\text{Rx}}^{1/2} \mathbf{G} \mathbf{R}_{\text{Tx}}^{1/2}$$

(21)

where the elements of $\mathbf{G}$ are independent and identically distributed (i.i.d.) as $\mathcal{CN}(0,1)$. The eigenvalues of $\mathbf{R}$ become the products of any two eigenvalues of $\mathbf{R}_{\text{Tx}}$ and $\mathbf{R}_{\text{Rx}}$, respectively. Depending on the amount of spatial correlation at the transmitter and receiver, the eigenvalues of $\mathbf{R}$ are either identical (e.g., if $\mathbf{R}_{\text{Tx}} = \mathbf{R}_{\text{Rx}} = \mathbf{I}$), distinct (e.g., if distinct eigenvalues at both sides), or consist repeated eigenvalues (e.g., when one of the sides is spatially uncorrelated with either $\mathbf{R}_{\text{Rx}} = \mathbf{I}$ or $\mathbf{R}_{\text{Tx}} = \mathbf{I}$). Eigenvalues that are measured in practice are naturally distinct, but clustering of those that are close-to-equal may be necessary to achieve numerical stability. Recall that these three cases correspond to those in Section II.

A. General Entropy-Maximizing Quantization Framework

Next, we will present a general framework for quantization of a stochastic variable $X \in [0, \infty)$, with the cumulative distribution function (cdf) $F(x)$, for the purpose of finite rate feedback. This variable may represent the weighted squared norm of a communication system, but the results are valid for any continuous cdf that fulfills $F(x) = 0$, for $x \le 0$, and $F(x) < 1$, for $x < \infty$.

With quantization, we mean the process of dividing a continuous range of values into a finite number of intervals. Herein, we consider L-bits quantization of the range $[0, \infty)$ of X, which means that the range is divided into 2^L disjoint intervals $[a_{i-1}, a_i)$, $1 \le i \le 2^L$. In our context, the purpose of the quantization is feedback and storage of the variable using L bits. Note that each interval, $\mathcal{Q}$, should be seen as a representative for all values of the original variable that lies in the interval. The actual value in the interval that best represents the quantized information, $\mathcal{Q}$, will change depending on the application (e.g., estimation of X or some function of it). When designing the quantization, we need to choose the decision boundaries a_i, for $0 \le i \le 2^L$, so that some design criteria is fulfilled. There is no over-all optimal criteria, but from an information-theoretical perspective it makes sense to maximize the entropy of the quantization and thereby the average amount of channel information that is fed back.

Lemma 1 (Entropy-Maximizing Quantization): Let X be a stochastic variable with a continuous cdf $F(x)$, that fulfills $F(x) = 0$, for $x \le 0$, and $F(x) < 1$, for $x < \infty$. Assume that the sample space, $[0, \infty)$, of X is quantized into 2^L disjoint intervals ($1 \le i \le 2^L$), where the ith interval is $[a_{i-1}, a_i)$ with $a_0 = 0$ and $a_{2^L} = \infty$. The interval boundaries that maximizes the entropy of X are given by

$$a_i = F^{-1}\left(\frac{i}{2^L}\right), \quad i = 1, \ldots, 2^L - 1.$$

(22)

This quantization will make the outcome of X equally probable in all the quantization intervals.

Let $\mathcal{Q} \in \{1, \ldots, 2^L\}$ denote the index such that the outcome $X \in [a_{\mathcal{Q}-1}, a_{\mathcal{Q}})$. The quantization maximizes the mutual

information between $\mathcal{Q}$ and $g(X)$, for any invertible function $g(\cdot) : \mathbb{R}^+ \to \mathbb{R}$.

Proof: The lemma follows from a division of the cdf of X into 2^L disjoint intervals of equal probability, and from the observation that X and $g(X)$ contain the same information. ∎

The inverses of cdfs can in general not be given in closed form, but since the function is bijective and nondecreasing the quantization boundaries in the lemma can be calculated efficiently using line search.

An important result of Lemma 1 is that even if we are interested in some function of X (e.g., the capacity if X represents the SNR), the entropy-maximizing quantization is still that of (22). Next, we will derive the distribution of the squared norm and apply this quantization framework.

B. Distribution and Feedback of the Squared Channel Norm

Consider feedback of the squared norm $\rho \triangleq \|\widetilde{\mathbf{H}}\|^2$ of the effective channel. Although we have assumed full CSI at the receiver [24], it is unreasonable to feedback the positive real-valued squared norm ρ with unlimited accuracy in a fading environment (if it still should provide information on the current channel conditions at the time of reception). Hence, we will quantize the squared norm so it can be described by a finite bit sequence. In order to apply the entropy-maximizing quantization framework in Lemma 1 we need to derive the cdf of ρ, which is given by the following corollary.

Corollary 1 (Distribution of the General Squared Norm): Let the channel $\mathbf{H} \in \mathbb{C}^{n_R \times n_T}$ be distributed as $\mathrm{vec}(\mathbf{H}) \in \mathcal{CN}(\mathbf{0}, \mathbf{R})$. Let $\mathbf{C} \in \mathbb{C}^{n_C \times n_R}$ and $\mathbf{D} \in \mathbb{C}^{n_T \times n_D}$ be arbitrary matrices such that the effective channel has the distribution $\mathrm{vec}(\mathbf{CHD}) \in \mathcal{CN}(\mathbf{0}, \widetilde{\mathbf{R}})$, where $\widetilde{\mathbf{R}} = (\mathbf{D}^T \otimes \mathbf{C})\mathbf{R}(\mathbf{D}^* \otimes \mathbf{C}^H)$. If the nonzero eigenvalues of $\widetilde{\mathbf{R}}$ are denoted $\lambda_1, \dots, \lambda_N$ (for $N \leq n_T n_R$), then the pdf of $\rho \triangleq \|\mathbf{CHD}\|^2$ is given by (4), (5), or (6), in the cases of identical, distinct, or nonidentical nondistinct eigenvalues, respectively. The corresponding cdf, $F(\rho)$, is given by (13), (15), and (17), respectively, using $a = 0$ and $b = \rho$.

Proof: The proof is given in Appendix C. ∎

In the Kronecker-structured case, $\mathbf{R} = \mathbf{R}_{\mathrm{Tx}}^T \otimes \mathbf{R}_{\mathrm{Rx}}$, the effective channel inherits this property: $\widetilde{\mathbf{R}} = \widetilde{\mathbf{R}}_{\mathrm{Tx}}^T \otimes \widetilde{\mathbf{R}}_{\mathrm{Rx}}$, where $\widetilde{\mathbf{R}}_{\mathrm{Tx}} = \mathbf{D}^H \mathbf{R}_{\mathrm{Tx}} \mathbf{D}$ and $\widetilde{\mathbf{R}}_{\mathrm{Rx}} = \mathbf{C} \mathbf{R}_{\mathrm{Rx}} \mathbf{C}^H$. The N nonzero eigenvalues of $\widetilde{\mathbf{R}}$ are given as the product of any two nonzero eigenvalues of $\widetilde{\mathbf{R}}_{\mathrm{Tx}}$ and $\widetilde{\mathbf{R}}_{\mathrm{Rx}}$, respectively.

To summarize, the distribution of the squared norm, ρ, of the effective channel is given by Corollary 1. Using Lemma 1, this distribution can be used to calculate the entropy-maximizing quantization of ρ.

C. MMSE Estimation of Weighted Squared Channel Norms

Next, we assume that the receiver has fed back information regarding the squared norm $\rho = \|\widetilde{\mathbf{H}}\|^2$ of the effective channel and the transmitter wants to estimate the weighted squared norm $\|\mathbf{B}\mathrm{vec}(\widetilde{\mathbf{H}})\|^2$. This corresponds to the generic estimation problem in (2). Using the conditional moments and cross-moments derived in Theorem 1 and 2, we will solve this problem by deriving the MMSE estimator of $\|\mathbf{B}\mathrm{vec}(\widetilde{\mathbf{H}})\|^2$ and its corresponding MSE. The following corollary extends results of [20], [23] by deriving the first

two conditional moments of the weighted squared norm for arbitrary eigenvalue structure of the effective channel. Observe that the first moment, $\mathbb{E}\{\|\mathbf{B}\mathrm{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\}$, is the MMSE estimator, while the corresponding MSE can be calculated as $\mathbb{E}\{\|\mathbf{B}\mathrm{vec}(\widetilde{\mathbf{H}})\|^4 | \varrho\} - \mathbb{E}\{\|\mathbf{B}\mathrm{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\}^2$.

Corollary 2 (MMSE Estimation of Weighted Squared Norms): Let the effective channel $\widetilde{\mathbf{H}} \in \mathbb{C}^{n_C \times n_D}$ be distributed as $\mathrm{vec}(\widetilde{\mathbf{H}}) \in \mathcal{CN}(\mathbf{0}, \widetilde{\mathbf{R}})$. Let $\widetilde{\mathbf{R}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ be the eigenvalue decomposition of the correlation matrix, where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$ is positive semi-definite. If the weighting matrix $\mathbf{B} \in \mathbb{C}^{n_B \times n_C n_D}$ is independent of $\widetilde{\mathbf{H}}$ and if ϱ contains information regarding the squared norm $\|\widetilde{\mathbf{H}}\|^2$, then

$$\mathbb{E}\{\|\mathbf{B}\mathrm{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\} = \mathbf{1}^T \mathbf{A}_{|\cdot|^2} \widehat{\mathbf{\Lambda}} \mathbf{1},$$

$$\mathbb{E}\{\|\mathbf{B}\mathrm{vec}(\widetilde{\mathbf{H}})\|^4 | \varrho\} = \frac{1}{2}\mathbf{p}^T \mathbf{T}\mathbf{p} + \frac{1}{2}\sum_{k_1=1}^{M}\sum_{k_2=1}^{M} \tilde{\mathbf{p}}_{k_1 k_2}^H \mathbf{T}\tilde{\mathbf{p}}_{k_1 k_2} \tag{23}$$

where $\mathbf{A} = \mathbf{BU}$, $[\mathbf{A}_{|\cdot|^2}]_{ij} = |a_{ij}|^2$, $\mathbf{1} = [1, \dots, 1]^T$, $\mathbf{p} = \mathbf{A}_{|\cdot|^2}^T \mathbf{1}$, $\tilde{\mathbf{p}}_{k_1 k_2} = [a_{k_1 1}a_{k_2 1}^*, \dots, a_{k_1 N}a_{k_2 N}^*]^T$, and

$$[\widehat{\mathbf{\Lambda}}]_{ij} = \begin{cases} \mathbb{E}\{|v_i|^2 | \varrho\}, & i = j, \\ 0, & i \neq j, \end{cases}$$

$$[\mathbf{T}]_{ij} = \begin{cases} \mathbb{E}\{|v_i|^4 | \varrho\}, & i = j, \\ 2\mathbb{E}\{|v_i|^2 |v_j|^2 | \varrho\}, & i \neq j. \end{cases} \tag{24}$$

For all i such that $\lambda_i = 0$, we have that $\mathbb{E}\{|v_i|^2 | \varrho\} = \mathbb{E}\{|v_i|^4 | \varrho\} = \mathbb{E}\{|v_i|^2 |v_j|^2 | \varrho\} = 0$. If ϱ represents the exact value of $\rho = \|\widetilde{\mathbf{H}}\|^2$ or the quantized information $a \leq \|\widetilde{\mathbf{H}}\|^2 < b$, then the remaining conditional moments of (24) are given by Theorem 1 and 2 (by removing all zero-valued eigenvalues), respectively.

Proof: The proof is given in Appendix C. ∎

In the Kronecker-structured case, $\widetilde{\mathbf{R}} = \widetilde{\mathbf{R}}_{\mathrm{Tx}}^T \otimes \widetilde{\mathbf{R}}_{\mathrm{Rx}}$, let $\mathbf{R}_{\mathrm{Tx}} = \mathbf{U}_{\mathrm{Tx}} \mathbf{\Lambda}_{\mathrm{Tx}} \mathbf{U}_{\mathrm{Tx}}^H$ and $\mathbf{R}_{\mathrm{Rx}} = \mathbf{U}_{\mathrm{Rx}} \mathbf{\Lambda}_{\mathrm{Rx}} \mathbf{U}_{\mathrm{Rx}}^H$ be the eigenvalue decompositions of the effective transmit and receive correlation, respectively. Then, we have $\mathbf{U} = (\mathbf{U}_{\mathrm{Tx}}^T \otimes \mathbf{U}_{\mathrm{Rx}})$ and $\mathbf{\Lambda} = (\mathbf{\Lambda}_{\mathrm{Tx}}^T \otimes \mathbf{\Lambda}_{\mathrm{Rx}})$. Thus, the nonzero eigenvalues of $\mathbf{\Lambda}$ are the products of any two nonzero eigenvalues of $\widetilde{\mathbf{R}}_{\mathrm{Tx}}$ and $\widetilde{\mathbf{R}}_{\mathrm{Rx}}$, respectively. If the weighting matrix is also Kronecker-structured, $\mathbf{B} = (\mathbf{B}_1^T \otimes \mathbf{B}_2)$, then the weighted squared norm can be expressed as $\|\mathbf{B}_1 \widetilde{\mathbf{H}} \mathbf{B}_2\|^2$.

To summarize the section, the entropy-maximizing quantization of an arbitrary nonnegative random variable was given in Lemma 1. The distribution of the squared norm $\rho \triangleq \|\mathbf{CHD}\|^2$ of the efficient channel was derived in Corollary 1 and this distribution is sufficient to calculate the entropy-maximizing quantization of ρ. Finally, the MMSE estimator of weighted squared norms with the structure $\|\mathbf{B}\mathrm{vec}(\widehat{\mathbf{H}})\|^2$, and their corresponding MSEs, was derived in Corollary 2 when feedback of either the exact value or a quantization of $\|\mathbf{CHD}\|^2$ is available.

IV. APPLICATIONS IN SINGLE- AND MULTI-USER SYSTEMS

In both single-user and multi-user systems, there is a need of feeding back a limited number of bits to shape the transmission to the spatial properties of the multi-antenna channel, adapt the symbol constellations to current conditions, and to perform efficient user-selection (in the multi-user case). There is a tight

connection between these goals and the SINR; we want to select users for transmission in spatial directions that permit high transmissions rates and the Shannon capacity, which gives an upper bound on the achievable rate, is a function of the SINR [2]. Hence, it is important to choose a feedback parameter that provides a reliable way of estimating the SINR, and to quantize this parameter efficiently by maximizing the amount of information per bit.

In this section, we consider norm based feedback for the purpose of rate adaptation and MMSE estimation of signal/interference powers. It will be shown that the results of Section III fit naturally into both single-user systems with OSTBCs and multi-user SDMA systems with beamforming.

A. Orthogonal Space-Time Block Codes With Precoding

We consider linear precoded OSTBCs, which should be regarded as the general type of OSTBCs that can exploit the spatial properties of the channel to improve the performance [9]–[12]. Recall the system model in (1) and assume that there is only one active user, so the user index can be dropped. Assume that a OSTBC is used to transmit K symbols over T symbols slots (i.e., the coding rate is K/T). Let $\mathbf{s} = [s_1, \ldots, s_K]^T \in \mathbb{C}^K$ be the vector of data symbols, where we have normalized such that $\mathbb{E}\{|s_i|^2\} = 1$ for all i. These symbols are coded in a matrix $\mathbf{C}(\mathbf{s}) \in \mathbb{C}^{B \times T}$ that fulfills the orthogonality property $\mathbf{C}(\mathbf{s})\mathbf{C}(\mathbf{s})^H = \|\mathbf{s}\|^2\mathbf{I}$ [8]. In addition, we use an arbitrary precoding matrix $\mathbf{W} \in \mathbb{C}^{n_T \times B}$ that projects the code into B spatial directions and is known to both the transmitter and the receiver [9]. The transmitted signals over T consecutive symbol slots is thus $[\mathbf{x}(t), \ldots, \mathbf{x}(t+T-1)] = \mathbf{W}\mathbf{C}(\mathbf{s})$ and the corresponding system model is

$$\mathbf{Y} = \mathbf{H}\mathbf{W}\mathbf{C}(\mathbf{s}) + \mathbf{N} \qquad (25)$$

where $\mathbf{Y} = [\mathbf{y}(t), \ldots, \mathbf{y}(t+T-1)] \in \mathbb{C}^{n_R \times T}$, $\mathbf{N} \in \mathbb{C}^{n_R \times T}$ contains i.i.d. noise samples with $\mathcal{CN}(0,1)$, and the data symbols are present in the entries of $\mathbf{C}(\mathbf{s})$. From [11], [28], it is known that OSTBCs provide the possibility of decomposing (25) into K independent and virtual single-antenna systems as

$$y_i' = \|\mathbf{H}\mathbf{W}\|s_i + n_i', \quad 1 \le i \le K, \qquad (26)$$

where $n_i' \in \mathcal{CN}(0,1)$. The corresponding SNR and maximum rate per source symbol are

$$\mathrm{SNR}^{\mathrm{OSTBC}} = \|\mathbf{H}\mathbf{W}\|^2, \quad \mathrm{R}^{\mathrm{OSTBC}} = \frac{K}{T}\log_2(1+\|\mathbf{H}\mathbf{W}\|^2). \qquad (27)$$

The exact SNR and rate values are known at the receiver, while the transmitter only knows the statistics. The SNR can be estimated at the transmitter as the average $\mathbb{E}\{\|\mathbf{H}\mathbf{W}\|^2\}$, but the estimation error will typically be large if no instantaneous CSI feedback is available. More robust performance is achieved by simply feeding back a quantized version of $\|\mathbf{H}\mathbf{W}\|^2$ to improve the estimation.

The effective channel is $\widetilde{\mathbf{H}} = \mathbf{H}\mathbf{W}$. The entropy-maximizing quantization of $\rho = \|\mathbf{H}\mathbf{W}\|^2$ is given by Lemma 1, with the cdf of ρ given by Corollary 1 (with $\mathbf{C} = \mathbf{I}$ and $\mathbf{D} = \mathbf{W}$). The quantization boundaries are functions of the precoder and the channel statistics, and need only to be updated at the relatively slow rate

that these are changing. Given the quantized feedback information of $a \le \rho < b$, the MMSE estimator (and the corresponding MSE) of the SNR is given by Corollary 2 (with $\mathbf{B} = \mathbf{I}$).

When estimation is used to choose an appropriate transmission rate, it might be necessary to include a fade-margin to achieve a target frame error rate, denoted p_{out}. Observe that packages sent in outage should not be considered lost since the information in them can still be utilized using, for example, hybrid ARQ. To control the error rate, we propose to include a fade-margin parameter α that is designed such that $\mathbb{P}(\widehat{\mathrm{SNR}} > \mathrm{SNR}^{\mathrm{OSTBC}}) \le p_{\mathrm{out}}$, where the SNR estimate $\widehat{\mathrm{SNR}}$ is determined as

$$\widehat{\mathrm{SNR}} = \max\left(\mathbb{E}\{\|\mathbf{H}\mathbf{W}\|^2|\varrho\} - \alpha\sqrt{\mathrm{MSE}\{\|\mathbf{H}\mathbf{W}\|^2|\varrho\}}, a\right) \qquad (28)$$

where $\mathrm{MSE}\{\|\mathbf{H}\mathbf{W}\|^2|\varrho\} = \mathbb{E}\{\|\mathbf{H}\mathbf{W}\|^4|\varrho\} - \mathbb{E}\{\|\mathbf{H}\mathbf{W}\|^2|\varrho\}^2$ and ϱ contains the quantized information $a \le \|\mathbf{H}\mathbf{W}\|^2 < b$. Hence, the SNR estimate can be calculated directly, using Corollary 2. MMSE estimation of the maximum rate $\mathrm{R}^{\mathrm{OSTBC}}$ can be treated in a similar manner [29].

To summarize, the framework in Lemma 1 can be used for entropy-maximizing quantization of the channel gain with linear precoded OSTBCs. Using Corollary 1, the SNR can be estimated either in the MMSE sense or in an outage-robust way as proposed in (28).

B. Beamforming for SDMA

Next, we consider a downlink multi-user SDMA system with beamforming transmission. The problem of efficient precoding and receive combining will be discussed, but the main focus will be on adapting the quantization framework of Section III-A to systems with user-selection and on developing a robust SINR estimation framework with feedback of norm based channel information.

Assume that K users have been scheduled for transmission and let the transmit beamforming vector and the data symbol intended for user k be denoted $\mathbf{w}_k \in \mathbb{C}^{n_T}$ and $s_k(t)$, respectively. Without loss of generality, we assume that $\mathbb{E}\{|s_k(t)|^2\} = 1$. Using the system model in (1), the transmitted signal is

$$\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{w}_k s_k(t) = \mathbf{W}\mathbf{s}(t) \qquad (29)$$

where $\mathbf{W} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_K]$ is the precoding matrix and $\mathbf{s}(t) \triangleq [s_1(t), \ldots, s_K(t)]^T$ is the vector of all transmitted symbols. Linear combining is assumed at the receiver side; that is, each user k uses a receive beamforming vector $\mathbf{v}_k \in \mathbb{C}^{n_R}$, with $\|\mathbf{v}_k\| = 1$, to achieve a scalar received signal $y_k(t) = \mathbf{v}_k^H \mathbf{y}_k(t)$.

In principle, the purpose of the precoding matrix $\mathbf{W}$ is to transmit K simultaneous data streams with an acceptably low co-user interference, while the linear combining at each receiver is used to further reduce both the inter- and intra-cell interference. With the notation $\mathbf{W}_k^{\mathrm{intra}} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_K]$, the SINR (when averaging over the noise and transmitted symbols) of user k is

$$\mathrm{SINR}_k = \frac{|\mathbf{v}_k^H \mathbf{H}_k \mathbf{w}_k|^2}{\|\mathbf{v}_k^H \mathbf{H}_k \mathbf{W}_k^{\mathrm{intra}}\|^2 + 1}. \qquad (30)$$

In order to optimize the system performance, we want to choose the beamforming vectors to maximize the sum rate of the selected users, possibly under some fairness condition. The optimal user-selection and beamforming scheme is very difficult to obtain in practice since base stations and users have asymmetric information; herein, the base station knows the channel statistics and some quantized feedback from each user, while each user knows its own channel perfectly but has limited information regarding the co-users. The main difficulty lies in the design of the limited feedback; it should reflect the channel properties when an SINR maximizing receive beamformer has been applied. Such a receive beamformer can in general not be designed until the user-selection and precoder design is finished, which is a stage when the transmitter truly needs instantaneous channel information. To resolve the receive beamformer ambiguity, for the sake of feedback design, we propose a two-step approach.

- **Stage 1, Feedback and Transmitter Design:** A reasonable, but suboptimal, virtual receive beamformer is assumed which is derived such that the efficient channel $\widetilde{\mathbf{H}}_k = \mathbf{v}_{k,\mathrm{virtual}}^H \mathbf{H}_k$ has statistical properties which may be derived at both the receiver and the transmitter (changes on a slow basis). The squared norm $\rho_k = \|\widetilde{\mathbf{H}}_k\|^2$ is quantized and fed back. Using this feedback information the transmitter selects users and design its precoder, assuming that all receivers uses their $\mathbf{v}_{k,\mathrm{virtual}}$ as receive beamformers. Additional directional feedback might be necessary if the spatial correlation is weak.

- **Stage 2, Data Transmission and Receiver Design:** The base station transmits data using the selected precoding. The receivers are free to select more beneficial receive beamformers if they desire, which could potentially increase their SINRs. These receive beamformers may for example be functions of the own channel matrix, $\mathbf{H}_k$, and some overhead or measurement of the interference. The SINRs estimated by the transmitter will then act as slightly pessimistic estimates.

Next, we will describe the first stage in greater detail. User selection and precoding design was thoroughly analyzed in [19] with similar prerequisites. Hence, our focus will be on feedback design and estimation of the SINR for a given precoder matrix and set of users. First, the entropy-maximizing framework of Section III-A will be adapted to multi-user systems. Then, the design of virtual receive beamformers will be discussed. Finally, observe that the signal and interference powers in (30) are weighted squared norms and therefore we will show how Corollary 2 can be used to estimate these from quantized feedback of $\|\widetilde{\mathbf{H}}_k\|^2 = \|\mathbf{v}_{k,\mathrm{virtual}}^H \mathbf{H}_k\|^2$. The user indexes will be dropped for brevity.

1) Post-User-Selection Quantization: The entropy-maximizing quantization framework in Lemma 1 can be used to calculate an efficient quantization of the squared norm $\|\widetilde{\mathbf{H}}\|^2$. In multi-user systems, user-selection can however change the statistics of the norm. If the scheduler takes its decisions based on, for example, the instantaneous sum rate, then users that experience strong channel norms are more likely to be selected.

Hence, the post-user-selection cdf of the squared norm will be the result of a transformation from the pre-scheduling cdf, $F(\cdot)$, that shifts the probability mass towards larger values.

The feedback information can be used both in the process of selecting users and in subsequent precoding design for the selected users. As discussed in [30], less CSI is required to choose appropriate users than to design a precoder that guarantees high and robust throughput. Thus, it makes more sense to maximize the post-user-selection entropy, than the pre-user-selection entropy as was done in Lemma 1.

The post-user-selection distribution depends strongly on the type of selection criterion, and is often difficult to derive analytically. In [31], the distribution was derived in a single-antenna system with known co-channel statistics, but the latter assumption is unreasonable in most multi-user scenarios. Observe that the post-user-selection cdf can be written as $G(F(\cdot))$, for some transformation function $G(\cdot)$. Using this notation, the following theorem gives the entropy-maximizing post-user-selection quantization.

Theorem 4 (Entropy-Maximizing Post-User-Selection Quantization): Let X have the continuous pre-user-selection cdf $F(x)$, which fulfills the properties in Lemma 1. Let the post-user-selection cdf be denoted $G(F(\cdot))$ for some continuous transformation function $G(\cdot) : [0,1] \rightarrow [0,1]$, which will be increasing and bijective on $(0,1)$ if the probability of selecting a user increases with its value X.

If the sample space, $[0,\infty)$, of X is quantized into 2^L disjoint intervals ($1 \leq i \leq 2^L$), where the ith interval is $[a_{i-1}, a_i)$ with $a_0 = 0$ and $a_{2^L} = \infty$, then the entropy-maximizing post-user-selection quantization is given by

$$a_i = F^{-1}\left(G^{-1}\left(\frac{i}{2^L}\right)\right), \quad i = 1, \ldots, 2^L - 1. \qquad (31)$$

Proof: The theorem follows directly from Lemma 1. ∎

To illustrate the usefulness of the notation with a transformation function $G(\cdot)$, we consider the following scheduler for which $G(\cdot)$ can be derived in closed form.

Definition 1 (Greatest Quality Probability Scheduler): Consider a scheduler that selects M users out of N. Let the channel quality of user k be measured by $x_k \geq 0$ and let its cdf be $F_k(\cdot)$, for all users $k = 1, \ldots, N$. Then, the Greatest Quality Probability (GQP) scheduler selects those M users that have the largest cdf values $F_k(x_k)$ of their current realization of x_k.

The proposed scheduler selects users based on the cdf values of their current channel quality (i.e., the percentage of realizations with worse performance). The quality, x_k, may represent the squared norm, or some other suitable measure. An important property of the proposed scheduler is that it provides fairness in terms of selecting users with identical probability, because $F_k(x_k) \in U[0,1)$ for all users k. The spatial separability between users is however ignored, but this is of minor importance when the number of transmit antennas grows [21]. When the users have identical statistics and x_k represents the SNR, then the GQP scheduler coincides with maximum throughput sched-

uling [13] (i.e., the users with the highest rates are selected). For the proposed scheduler, the transformation function becomes

$$G(z) = \sum_{i=0}^{M-1} \binom{N}{i} \frac{M-i}{M} z^{N-i} (1-z)^i. \qquad (32)$$

This is shown by observing that the cdf values, $F_k(x_k)$, are identically distributed among all users and that a selected user has any of the Mth largest $F_k(x_k)$ with equal probability. It is worth noting that the selection scheme in Definition 1 becomes idealized when quantization is introduced; the exact values of $F_k(x_k)$ are unknown and have to be estimated based on the available feedback information. The point is however that the transformation function $G(\cdot)$ can be determined explicitly for certain schedulers. In general, the function depends on all users and will therefore be unavailable at the receivers. It can however be approximated in various ways. In Section V, it will be illustrated numerically that even a simple parametrization as $G(x) = (e^{cx} - 1)/(e^c - 1)$, for some parameter c, can significantly improve the performance. Thus, the gain of post-user-selection quantization can be exploited by simple means.

2) Design of Virtual Receive Beamformers: The virtual receive beamformer $\mathbf{v}_{\text{virtual}}$ should be designed such that the statistics of the effective channel $\widetilde{\mathbf{H}} = \mathbf{v}_{\text{virtual}}^H \mathbf{H}$ can be derived deterministically at both the receiver and transmitter. At first sight, this assumption seems to lead to the conclusion that $\mathbf{v}_{\text{virtual}}$ needs to be independent of the realization $\mathbf{H}$. This requirement can however be relaxed, since the effective channel will be deterministic in eigendirections with eigenvalues that become zero. Thus, the system can be designed such that the transmitter knows that $\mathbf{v}_{\text{virtual}}$ always will cancel out the channel in some predefined eigendirections (e.g., such that are expected to contain much interference).

As an example, the following virtual receive beamformer was proposed in [20] for Kronecker-structured systems with $\mathbf{R}_{\text{Rx}} = \mathbf{I}$, but can be generalized for arbitrary receiver side correlation. Let the eigenvalue decomposition of the transmit side correlation matrix be partitioned as

$$\mathbf{R}_{\text{Tx}} = \underbrace{[\mathbf{U}_1 \ \ \mathbf{U}_2]}_{=\mathbf{U}} \text{diag}(\lambda_1, \ldots, \lambda_{n_T}) \underbrace{[\mathbf{U}_1 \ \ \mathbf{U}_2]^H}_{=\mathbf{U}^H} \qquad (33)$$

where $\mathbf{U}_1 \in \mathbb{C}^{n_T \times n_T - n_I}$ and $\mathbf{U}_2 \in \mathbb{C}^{n_T \times n_I}$ contain eigenvectors, and the eigenvalues are ordered in some (predefined) arbitrary way. If $n_I < n_R$, then there exist a receive beamformer $\mathbf{v}_{\text{virtual}}$ that will completely cancel out the power in the eigensubspace $\mathbf{U}_2$ such that the experienced channel $\widetilde{\mathbf{H}} = \mathbf{v}_{\text{virtual}}^H \mathbf{H}$ has the distribution $\text{vec}(\widetilde{\mathbf{H}}) \in \mathcal{CN}(\mathbf{0}, \widetilde{\mathbf{R}})$, with $\widetilde{\mathbf{R}} =$ $\mathbf{U}\text{diag}(\lambda_1, \ldots, \lambda_{n_R - n_I}, 0, \ldots, 0)\mathbf{U}^H$. To achieve this, the receive beamformer should be chosen arbitrarily in the null space of $(\mathbf{HU}_2)^H$ (i.e., $\mathbf{v}_{\text{virtual}}^H \mathbf{HU}_2 = \mathbf{0}$). Using this virtual receive beamformer, the transmitter knows that the experienced channel will have the correlation matrix $\widetilde{\mathbf{R}}$.

In practice, the virtual receive beamformer can be designed in various ways depending on the environment. The design can also be relaxed such that the effective channel only becomes approximately Gaussian; the important thing is that the first and second order statistics are approximately known at the transmitter.

3) Estimation of the SINR: Finally, we consider estimation of the SINR in (30) at the transmitter (e.g., for the purpose of user-selection and rate adaptation). Apart from the channel statistics, the transmitter has received quantized feedback of $\|\widetilde{\mathbf{H}}\|^2$, the squared norm of the effective channel with the virtual receive beamformer. The unknown quantities in the SINR expression are the signal and interference powers, which both are weighted squared norms: $\|\mathbf{v}_{\text{virtual}}^H \mathbf{H} \widetilde{\mathbf{W}}\|^2 = \|(\widetilde{\mathbf{W}}^T \otimes \mathbf{I})\text{vec}(\widetilde{\mathbf{H}})\|^2$, where the weighting matrix $\widetilde{\mathbf{W}}$ contains one or several transmit beamformers. These beamformers are either directly known to transmitter or they should be selected in the precoder design to maximize the (weighted) sum rate. In any way, the SINR can be estimated as a function of the transmit beamformers.

Similar to [19], [20], [23], we propose to use the pessimistic SINR estimator in (34), at the bottom of the page. In this estimator, $\mathbf{B}_1 = (\mathbf{w}^T \otimes \mathbf{I})$ and $\mathbf{B}_2 = ((\mathbf{W}^{\text{intra}})^T \otimes \mathbf{I})$. The MSEs are calculated as

$$\text{MSE}\{\|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\}$$
$$= \mathbb{E}\{\|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^4 | \varrho\} - \mathbb{E}\{\|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\}^2$$

and ϱ represents either exact norm information $\rho = \|\widetilde{\mathbf{H}}\|^2$ or the quantized feedback information $a \leq \|\widetilde{\mathbf{H}}\|^2 < b$. The design parameter α in (34) can be used to achieve a target frame error rate, p_{out}. This adaptive fade-margin is similar to the one in Section IV-A and is an essential control-feature in most systems, including those with advanced error control.

If the virtual receive beamformer, $\mathbf{v}_{\text{virtual}}$, is designed as described in the previous section, the signal and interference powers (and their MSEs) in (34) can be MMSE estimated using Corollary 2. If $\mathbf{v}_{\text{virtual}}$ only approximately fulfills the requirements and/or an improved receive beamformer is used in the actual data transmission, then the SINR estimate in (34) will not be the ideal one. The performance loss is however limited in many practical systems, as illustrated in [23]. The explanation is that small estimation errors have limited consequences since the adaptive fade-margin in (34) is used to adapt the SINR estimate to control the error rate.

$$\widehat{\text{SINR}} = \max\left(0, \frac{\mathbb{E}\{\|\mathbf{B}_1\text{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\} - \alpha\sqrt{\text{MSE}\{\|\mathbf{B}_1\text{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\}}}{\mathbb{E}\{\|\mathbf{B}_2\text{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\} + \alpha\sqrt{\text{MSE}\{\|\mathbf{B}_2\text{vec}(\widetilde{\mathbf{H}})\|^2 | \varrho\}} + 1}\right). \qquad (34)$$

To summarize the section, we have considered a multi-user system with beamforming at both the base station and the receiving users. The entropy-maximizing quantization framework in Lemma 1 has been extended to take the effect of user-selection into account. A virtual receive beamformer was proposed to overcome the receive beamformer ambiguity in the feedback design. Finally, it was shown how Corollary 2 can be used to perform robust SINR estimation in systems with norm feedback.

V. NUMERICAL EXAMPLES

This section will illustrate how the analytic results of the paper can be used to improve the performance of MIMO communication systems. Two numerical examples will be given, corresponding to the single-user case in Section IV-A and the multi-user case Section IV-B, respectively. In the single-user case, we consider transmission from a four-antenna transmitter to a two-antenna receiver, using OSTBCs ($K = T = 2$) and the linear precoder $\mathbf{W}$ in [10] that adapts the coding to the channel statistics. The average SNR (defined as $\mathbb{E}\{|[\mathbf{H}]_{ij}|^2\}$) is 10 dB and we assume the Kronecker channel model in (21) with uncorrelated receive antennas ($\mathbf{R}_{\mathrm{Rx}} = \mathbf{I}$). The transmit correlation follows the exponential model of [32], which models a uniform linear array (ULA) with the correlation between adjacent antennas as a parameter. The instantaneous SNR of this system is $\|\mathbf{HW}\|^2$ and is quantized and fed back using the entropy-maximizing framework in Lemma 1.

The average throughput over $2 \cdot 10^6$ realizations for different numbers of feedback bits is shown in Fig. 1 with varying antenna correlation (absolute value of the coefficient in [32]) and with an outage probability of 5%. Observe that a logarithmic scale has been used on the x-axis. The corresponding throughputs with half a wavelength antenna separation and different angular spreads (standard deviation of Gaussian distributed scatterers, as seen from the transmitter) are given as a reference to show that many measured systems in fact have antenna correlations around 0.9 (cf. [3]). From Fig. 1, it is clear that just a few bits of norm feedback are sufficient to achieve performance close to that of full CSI; 52% of the feedback gain is achieved with one bit of feedback, while three bits gives 84% and five bits 95%. The amount of correlation has little impact on the percentage of feedback gain. Finally, observe that the performance of this precoded system increases with the transmit antenna correlation, as expected from [12].

In the multi-user case, we consider downlink zero-forcing SDMA communication from a transmitter with an eight-antenna uniform circular array (UCA) to 20 users, each equipped with four uncorrelated receive antennas. The angular spread is 10 degrees and the transmit antenna separation is half a wavelength. The users are uniformly distributed in the area $0.2R \leq r \leq R$ of a circular cell of radius R. The average SNR is 10 dB at the cell boundary and the power decay is proportional to r^{-4}. The scheduling is performed using the greedy user selection [19], [33] and with proportional fairness as scheduling criterion [34]. The performance is measured in terms of the cdf of the average cell throughput over different scenarios. Each of the 10^4 considered scenarios represent a unique random constellation of mobiles with fixed statistics, while the average cell throughput is calculated over 150 scheduling decisions.
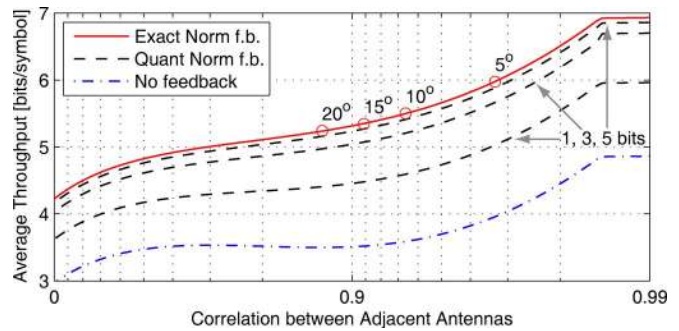

Fig. 1. The average throughput as a function of the absolute value of the correlation between adjacent antennas at the transmitter. The performance is given for the cases with exact SNR/norm feedback, with quantized feedback using 1, 3, or 5 bits (increasing performance), and without feedback. The performance with different amounts of angular spread is marked with circles as a reference. Observe the logarithmic scale of the x-axis.
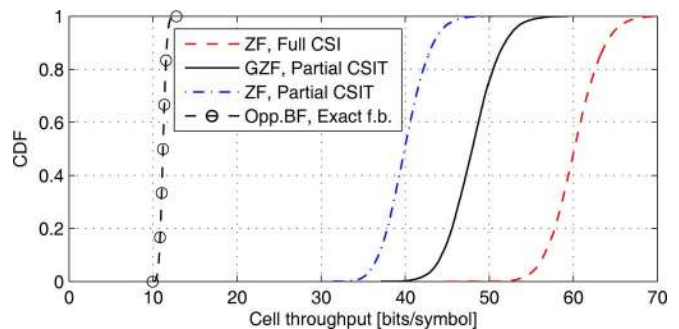

Fig. 2. The cumulative distribution functions (cdfs) of the average cell throughput over scenarios with 20 uniformly distributed users in a circular cell. The performance of zero-forcing with full CSI is compared with generalized zero-forcing with gain feedback [19], directional-quantized zero-forcing [18], and multi-user opportunistic beamforming [14]. The GZF uses 3 bits of gain feedback, while the latter two schemes uses 3 bits of directional feedback and perfect gain feedback.

In Fig. 2, the performance of zero-forcing (ZF) precoding with full CSI is compared with 1) directional-quantized ZF [18] with a Grassmannian codebook [35]; 2) multi-user opportunistic beamforming [14]; and 3) the generalized zero-forcing (GZF) scheme in [19]. The GZF uses the receive antennas to suppress the interference sensitive subspace of [19] and uses 3 bits of norm feedback (with post-user-selection quantization and the transformation function in Theorem 4 approximated as $G(x) = (e^{1.1x} - 1)/(e^{1.1} - 1)$). The outage probability is 5%. The quantized ZF and opportunistic beamforming schemes are based on 3 bits directional information and perfect gain feedback. It is however seen in Fig. 2 that the GZF scheme outperforms the other partial CSI schemes, although it is based on a considerably smaller feedback load. Observe that the framework derived herein can be applied to handle quantized gain feedback in the two competing schemes.

Finally, in Fig. 3 the performance of the GZF scheme is shown for different numbers of feedback bits and with both pre- and post-user-selection quantization. With one bit of feedback, 71–73% of the feedback gain is achieved, depending on the type of quantization. The corresponding interval is 90–92% for three bits and 97–98% for five bits. It is clear that a few bits of feedback are sufficient to achieve most of the feedback gain, and that the benefit of considering the post-scheduling
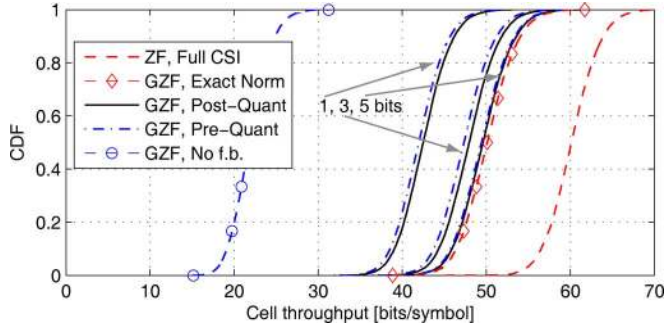
Fig. 3. The cumulative distribution functions (cdfs) of the average cell throughput over scenarios with 20 uniformly distributed users in a circular cell. The performance of the generalized zero-forcing scheme [19] is shown for different types of norm feedback: no feedback, pre- or post-scheduling quantization with 1, 3, or 5 bits, and perfect feedback (increasing performance). Zero-forcing precoding with full CSI is given as a reference.

distribution in the quantization is nonnegligible. The lack of instantaneous directional information will however make the scheme suboptimal, even for perfect norm feedback.

VI. CONCLUSION

For arbitrarily correlated zero-mean complex Gaussian matrices, closed-form expressions for the conditional matrix distribution and moments of individual elements have been derived when the squared Frobenius norm of the matrix is either known exactly or known to lie in a specific quantization interval. In addition, MMSE estimators (and their resulting MSEs) of weighted squared norms have been derived, given quantized norm information. This mathematical contribution has clear applications in renewal theory, but herein the main focus has been on the applications in wireless communication systems with statistical CSI and limited feedback. In these systems, the signal and interference powers are weighted squared norms. An entropy-maximizing framework was proposed for feedback quantization and it has been shown how feedback of quantized norm information enables robust estimation of the SINR in an MMSE based framework. The usefulness of the results were exemplified in single-user systems with linearly precoded OSTBCs and in multi-user SDMA systems with beamforming and quantization that takes the post-user-selection distribution into account.

APPENDIX A
SOME USEFUL ELEMENTARY FUNCTIONS

Throughout the paper, a few nonstandard elementary functions have been used extensively. Specifically, they appear in the derivation of Theorem 1, 2, 3, and Lemma 4. This appendix will first define the functions and then provide integral expressions that have these functions as solutions.

Definition 2: For nonnegative integers N, M, L, i, j, m, we define the five functions in (35)–(39). These are shown at the bottom of the next page.

The next two lemmas show how the functions in Definition 2 appear as the solutions to certain integrals.

Lemma 2: Let N and M be two nonnegative integers and let C be a nonzero real-valued scalar. Then,

$$\int x^N (x-\rho)^M e^{-Cx} dx$$

$$= -\sum_{k=0}^{M} \binom{M}{k} \frac{(-\rho)^{M-k}(N+k)! e^{-Cx}}{C^{N+k+1}} \sum_{j=0}^{N+k} \frac{(Cx)^j}{j!} + D \tag{40}$$

$$\int x^N (x-\rho)^M dx$$

$$= \sum_{k=0}^{M} \binom{M}{k} \frac{x^{N+k+1}(-\rho)^{M-k}}{N+k+1} + D \tag{41}$$

where ρ is a real-valued constant and D is some arbitrary constant.

Proof: First, assume that $M = 0$ and observe that (40) holds for $N = 0$, since $\int e^{-Cx} dx = -e^{-Cx}/C + D$. Then by the principle of induction, we have that

$$\int e^{-Cx} x^N dx = -\frac{N! e^{-Cx}}{C^{N+1}} \sum_{j=0}^{N} \frac{(Cx)^j}{j!} + D$$

where we integrated by parts and used (40) for $N-1$ and $M = 0$. Then, for $M > 0$, the expression in (40) follows by the Binomial series expansion $(x-\rho)^M = \sum_{k=0}^{M} \binom{M}{k} x^M (-\rho)^{M-k}$. The expression in (41) follows by the same kind of Binomial series expansion and pure integration. ∎

Lemma 3: Let μ_i, μ_j, μ_m be strictly positive scalars. The functions introduced in Definition 2 satisfy

$$G_{N,M,i,j}^{(1)}(\rho) = \int_0^\rho x^N (x-\rho)^M e^{-x(1/\mu_i - 1/\mu_j)} dx$$

$$G_{N,M,L,i,j,m}^{(2)}(\rho) = \int_0^\rho x^N e^{-x(1/\mu_i - 1/\mu_m)} G_{M,L,j,m}^{(1)}(\rho-x) dx$$

$$\widetilde{G}_{N,M,i,j}^{(1)}(a,b) = \int_a^b e^{-\rho/\mu_j} G_{N,M,i,j}^{(1)}(\rho) d\rho$$

$$\widetilde{G}_{N,M,L,i,j,m}^{(2)}(a,b) = \int_a^b e^{-\rho/\mu_m} G_{N,M,L,i,j,m}^{(2)}(\rho) d\rho$$

$$\widetilde{G}_{N,i}^{(3)}(a,b) = \int_a^b \rho^N e^{-\rho/\mu_i} d\rho.$$

Proof: The results for $G_{N,M,i,j}^{(1)}(\rho)$ and $G_{N,M,L,i,j,m}^{(2)}(\rho)$ follow from Lemma 2, when the lower and upper bound are 0 and ρ, respectively. Similarly, the results for $\widetilde{G}_{N,M,i,j}^{(1)}(a,b)$, $\widetilde{G}_{N,M,L,i,j,m}^{(2)}(a,b)$, and $\widetilde{G}_{N,i}^{(3)}(a,b)$ are achieved from the lemma when the lower and upper bounds are a and b, respectively. ∎

APPENDIX B
JOINT CONDITIONAL DISTRIBUTIONS

Consider $\mathbf{v} = [v_1, \ldots, v_N]^T = [|v_1|e^{j\phi_1}, \ldots, |v_N|e^{j\phi_N}]^T$, as defined in Section II. Joint conditional pdfs of sets of $|v_i|^2$

(with known $\rho \triangleq \|\mathbf{v}\|^2$) are used in the proofs of several theorems. Since $|v_j|^2 = \rho - \sum_{i \neq j} |v_i|^2$ for any choice of j, note that the joint conditional pdf of $\mathbf{v}$ can be factorized as

$$
\begin{aligned}
f_{\mathbf{v}\|\|\mathbf{v}\|^2}(\mathbf{v}|\rho) &= \frac{1}{2^N} f_{\bar{\mathbf{v}}\|\|\mathbf{v}\|^2}(\bar{\mathbf{v}}|\rho) f_{\boldsymbol{\phi}}(\boldsymbol{\phi}) \\
&= \frac{1}{2^N} f_{\bar{\mathbf{v}}_{\widetilde{\mathcal{S}}}\|\|\mathbf{v}\|^2}(\bar{\mathbf{v}}_{\widetilde{\mathcal{S}}}|\rho) \delta\left(\rho - \sum_{i=1}^{N} |v_i|^2\right) f_{\boldsymbol{\phi}}(\boldsymbol{\phi})
\end{aligned}
$$

(42)

where $\widetilde{\mathcal{S}} = \{1, \ldots, N\} \setminus \{j\}$ and $f_{\boldsymbol{\phi}}(\boldsymbol{\phi}) = \prod_{i=1}^{N} H_{0,2\pi}(\phi_i)/2\pi$. While the joint conditional pdf $f_{\mathbf{v}\|\|\mathbf{v}\|^2}(\mathbf{v}|\rho)$ is a function of $2N$ complex-valued variables, the expression in (42) has separated it into an N-dimensional uniform phase distribution and the $(N-1)$-dimensional conditional distribution $f_{\bar{\mathbf{v}}_{\widetilde{\mathcal{S}}}\|\|\mathbf{v}\|^2}(\bar{\mathbf{v}}_{\widetilde{\mathcal{S}}}|\rho)$. The following lemma derives closed-form expressions for this pdf in the three cases of identical, distinct, and neither identical nor distinct eigenvalues.

Lemma 4: Let $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathcal{CN}(\mathbf{0}, \boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$ has strictly positive eigenvalues λ_i, and

define $\rho \triangleq \|\mathbf{v}\|^2$. Let $\mathcal{S}$ be a nonempty set with distinct indexes from $\{1, \ldots, N\}$ and with cardinality $1 \leq |\mathcal{S}| < N$. If the eigenvalues λ_i are identical (i.e., $\lambda_i = \mu_1$ for all i), then the joint conditional pdf of $\bar{\mathbf{v}}_{\mathcal{S}} = [|v_{\mathcal{S}(1)}|^2, \ldots, |v_{\mathcal{S}(|\mathcal{S}|)}|^2]^T$, when ρ is known, is

$$
f^{\mathrm{ident}}_{\bar{\mathbf{v}}_{\mathcal{S}}\|\|\mathbf{v}\|^2}(\boldsymbol{\xi}_{\mathcal{S}}|\rho) = \frac{(\rho - \|\boldsymbol{\xi}_{\mathcal{S}}\|_1)^{N-|\mathcal{S}|-1}}{\rho^{N-1}} \frac{(N-1)! H_{0,\rho}(\boldsymbol{\xi}_{\mathcal{S}})}{(N-|\mathcal{S}|-1)!}
$$

(43)

where $\boldsymbol{\xi}_{\mathcal{S}} = [\xi_{\mathcal{S}(1)}, \ldots, \xi_{\mathcal{S}(|\mathcal{S}|)}]^T$. If the eigenvalues λ_i are distinct, then the joint conditional pdf is

$$
\begin{aligned}
& f^{\mathrm{dist}}_{\bar{\mathbf{v}}_{\mathcal{S}}\|\|\mathbf{v}\|^2}(\boldsymbol{\xi}_{\mathcal{S}}|\rho) \\
& = \frac{H_{0,\rho}(\boldsymbol{\xi}_{\mathcal{S}})}{f^{\mathrm{dist}}_{\|\mathbf{v}\|^2}(\rho) \prod_{j \in \mathcal{S}} \lambda_j} \sum_{\substack{k=1 \\ k \notin \mathcal{S}}}^{N} \frac{e^{-\rho/\lambda_k} \prod_{l \in \mathcal{S}} e^{-(\xi_j/\lambda_l)(1-\lambda_l/\lambda_k)}}{\lambda_k \prod_{l \notin \mathcal{S} \cup \{k\}} \left(1 - \frac{\lambda_l}{\lambda_k}\right)}.
\end{aligned}
$$

(44)

Finally, if the eigenvalues are nondistinct and nonidentical, then assume that they are ordered such that the characterization

$$
G^{(1)}_{N,M,i,j}(\rho) \triangleq
\begin{cases}
\sum_{k=0}^{M} \binom{M}{k} \frac{(N+k)!(-\rho)^{M-k}}{(\frac{1}{\mu_i} - \frac{1}{\mu_j})^{N+k+1}} \left(1 - \sum_{m=0}^{N+k} \frac{\rho^m e^{-\rho(1/\mu_i - 1/\mu_j)}}{m!(\frac{1}{\mu_i} - \frac{1}{\mu_j})^{-m}}\right), & i \neq j, \\
\rho^{N+M+1} \sum_{k=0}^{M} \binom{M}{k} \frac{(-1)^{M-k}}{N+k+1}, & i = j,
\end{cases}
$$

(35)

$$
G^{(2)}_{N,M,L,i,j,m}(\rho) \triangleq
\begin{cases}
\sum_{k=0}^{L} \binom{L}{k}(M+k)! \left(\frac{G^{(1)}_{N,L-k,i,m}(\rho)}{(\frac{1}{\mu_j} - \frac{1}{\mu_m})^{M+k+1}} - \sum_{l=0}^{M+k} \frac{e^{-\rho(1/\mu_j - 1/\mu_m)} G^{(1)}_{N,L-k+l,i,j}(\rho)}{(-1)^l l!(\frac{1}{\mu_j} - \frac{1}{\mu_m})^{M+k-l+1}}\right), & i \neq m \neq j, \\
G^{(1)}_{M,N+L+1,j,m}(\rho) \sum_{k=0}^{L} \binom{L}{k} \frac{(-1)^{N+k+1}}{N+k+1}, & i = m \neq j, \\
G^{(1)}_{N,M+L+1,i,m}(\rho) \sum_{k=0}^{L} \binom{L}{k} \frac{(-1)^{M+k+1}}{M+k+1}, & i \neq m = j, \\
\rho^{N+M+L+2} \left(\sum_{k=0}^{N} \binom{N}{k} \frac{(-1)^{L+k}}{M+L+k+2}\right) \left(\sum_{k=0}^{M} \binom{M}{k} \frac{(-1)^k}{L+k+1}\right), & i = j = m,
\end{cases}
$$

(36)

$$
\widetilde{G}^{(1)}_{N,M,i,j}(a,b) \triangleq
\begin{cases}
\sum_{k=0}^{M} \binom{M}{k} \frac{(N+k)!(-1)^{M-k}}{(\frac{1}{\mu_i} - \frac{1}{\mu_j})^{N+k+1}} \left(\widetilde{G}^{(3)}_{M-k,j}(a,b) - \sum_{m=0}^{N+k} \frac{\widetilde{G}^{(3)}_{M+m-k,i}(a,b)}{m!(\frac{1}{\mu_i} - \frac{1}{\mu_j})^{-m}}\right), & i \neq j, \\
\widetilde{G}^{(3)}_{N+M+1,j}(a,b) \sum_{k=0}^{M} \binom{M}{k} \frac{(-1)^{M-k}}{N+k+1}, & i = j,
\end{cases}
$$

(37)

$$
\widetilde{G}^{(2)}_{N,M,L,i,j,m}(a,b) \triangleq
\begin{cases}
\sum_{k=0}^{L} \binom{L}{k}(M+k)! \left(\frac{\widetilde{G}^{(1)}_{N,L-k,i,m}(a,b)}{(\frac{1}{\mu_j} - \frac{1}{\mu_m})^{M+k+1}} - \sum_{l=0}^{M+k} \frac{\widetilde{G}^{(1)}_{N,L-k+l,i,j}(a,b)}{(-1)^l l!(\frac{1}{\mu_j} - \frac{1}{\mu_m})^{M+k-l+1}}\right), & i \neq m \neq j, \\
\widetilde{G}^{(1)}_{M,N+L+1,j,m}(a,b) \sum_{k=0}^{L} \binom{L}{k} \frac{(-1)^{N+k+1}}{N+k+1}, & i = m \neq j, \\
\widetilde{G}^{(1)}_{N,M+L+1,i,m}(a,b) \sum_{k=0}^{L} \binom{L}{k} \frac{(-1)^{M+k+1}}{M+k+1}, & i \neq m = j, \\
\widetilde{G}^{(3)}_{N+M+L+2,m}(a,b) \left(\sum_{k=0}^{N} \binom{N}{k} \frac{(-1)^{L+k}}{M+L+k+2}\right) \left(\sum_{k=0}^{M} \binom{M}{k} \frac{(-1)^k}{L+k+1}\right), & i = j = m,
\end{cases}
$$

(38)

$$
\widetilde{G}^{(3)}_{N,i}(a,b) \triangleq \frac{N!}{(\frac{1}{\mu_i})^{N+1}} \sum_{m=0}^{N} \frac{\left(\frac{a}{\mu_i}\right)^m e^{-a/\mu_i} - \left(\frac{b}{\mu_i}\right)^m e^{-b/\mu_i}}{m!}.
$$

(39)

in (3) is fulfilled. Let $\tilde{\mathbf{r}} = [\tilde{r}_1, \ldots, \tilde{r}_M]^T$ be the eigenvalue multiplicities when the elements in $\bar{\mathbf{v}}_{\mathcal{S}}$ have been removed. Then, the joint conditional pdf is

$$f_{\bar{\mathbf{v}}_{\mathcal{S}} \| \|\mathbf{v}\|^2}^{\text{repeat}}(\boldsymbol{\xi}_{\mathcal{S}}|\rho) = \frac{H_{0,\rho}(\boldsymbol{\xi}_{\mathcal{S}})}{f_{\|\mathbf{v}\|^2}(\rho)} \prod_{\bar{l}=1}^{M} \frac{1}{\mu_{\bar{l}}^{r_{\bar{l}}}}$$

$$\times \sum_{k=1}^{M} \sum_{l=1}^{\tilde{r}_k} \frac{\Psi_{k,l,\tilde{\mathbf{r}}}}{(\tilde{r}_k - l)!} e^{-\rho/\mu_k} (\|\boldsymbol{\xi}_{\mathcal{S}}\| - \rho)^{\tilde{r}_k - l} \prod_{j \in \mathcal{S}} e^{-\xi_j (1/\mu_{\tau(j)} - 1/\mu_k)}. \tag{45}$$

Proof: The expression for the joint conditional distribution $f_{\bar{\mathbf{v}}_{\mathcal{S}} \| \|\mathbf{v}\|^2}(\boldsymbol{\xi}_{\mathcal{S}}|\rho)$ was proved in [23], in the case of distinct eigenvalues, using induction. Herein, all three cases will be proved using a somewhat shorter approach based on the law of total probability. Let $\mathbf{v}_{\mathcal{S}^c}$ be the vector with all elements that remain when those with indexes in $\mathcal{S}$ have been removed. By using the law of total probability to condition on $\|\bar{\mathbf{v}}_{\mathcal{S}}\|_1$ and then Bayes' formula and that $f_{\bar{\mathbf{v}}_{\mathcal{S}}, \|\bar{\mathbf{v}}_{\mathcal{S}}\|_1}(\boldsymbol{\xi}_{\mathcal{S}}, x) = f_{\bar{\mathbf{v}}_{\mathcal{S}}}(\boldsymbol{\xi}_{\mathcal{S}})\delta(x - \|\boldsymbol{\xi}_{\mathcal{S}}\|_1)$, we have that

$$f_{\bar{\mathbf{v}}_{\mathcal{S}} \| \|\mathbf{v}\|^2}(\boldsymbol{\xi}_{\mathcal{S}}|\rho) = \int_0^\rho f_{\bar{\mathbf{v}}_{\mathcal{S}} \| \|\bar{\mathbf{v}}_{\mathcal{S}}\|_1}(\boldsymbol{\xi}_{\mathcal{S}}|x) f_{\|\bar{\mathbf{v}}_{\mathcal{S}}\|_1 \| \|\mathbf{v}\|^2}(x|\rho)dx$$

$$= \int_0^\rho f_{\bar{\mathbf{v}}_{\mathcal{S}}}(\boldsymbol{\xi}_{\mathcal{S}})\delta(x - \|\boldsymbol{\xi}_{\mathcal{S}}\|_1) \frac{f_{\|\mathbf{v}\|^2 \| \|\bar{\mathbf{v}}_{\mathcal{S}}\|_1}(\rho|x)}{f_{\|\mathbf{v}\|^2}(\rho)}dx$$

$$= f_{\bar{\mathbf{v}}_{\mathcal{S}}}(\boldsymbol{\xi}_{\mathcal{S}}) f_{\|\mathbf{v}_{\mathcal{S}^c}\|^2}(\rho - \|\boldsymbol{\xi}_{\mathcal{S}}\|_1) \frac{H_{0,\rho}(\boldsymbol{\xi}_{\mathcal{S}})}{f_{\|\mathbf{v}\|^2}(\rho)}. \tag{46}$$

Then, the theorem follows from observing that $f_{\bar{\mathbf{v}}_{\mathcal{S}}}(\boldsymbol{\xi}_{\mathcal{S}}) = \prod_{j \in \mathcal{S}} e^{-\xi_j/\lambda_j}/\lambda_j$, and that $f_{\|\mathbf{v}_{\mathcal{S}^c}\|^2}(\rho - \|\boldsymbol{\xi}_{\mathcal{S}}\|_1)$ and $f_{\|\mathbf{v}\|^2}(\rho)$ are given in (4), (5), and (6) for the three different cases. ∎

APPENDIX C
COLLECTION OF PROOFS

Proof of Theorem 1: By definition, the conditional pth order moment is

$$\mathbb{E}\{|v_i|^{2p}|\rho\} = \int_0^\rho \xi_i^p f_{|v_i|^2 \| \|\mathbf{v}\|^2}(\xi_i|\rho)d\xi_i$$

where the conditional distribution is given in Lemma 4. In all three eigenvalue cases, the integral can be solved using Lemma 3, by observing that the terms that depend on ξ_i form an integral that equals the function $G_{N,M,i,j}^{(1)}(\rho)$, from Definition 2, for different values of N, M, i, and j.

The (p_i, p_j)th order cross-moment is defined as

$$\mathbb{E}\{|v_i|^{2p}|v_j|^{2p}|\rho\}$$
$$= \int_0^\rho \xi_i^{p_i} \int_0^{\rho - \xi_i} \xi_j^{p_j} f_{|v_i|^2, |v_j|^2 \| \|\mathbf{v}\|^2}(\xi_i, \xi_j|\rho)d\xi_j d\xi_i.$$

The joint conditional distribution $f_{|v_i|^2, |v_j|^2 \| \|\mathbf{v}\|^2}(\xi_i, \xi_j|\rho)$ is given by Lemma 4 for $N > 2$. In this case, the double integral can be determined (using Lemma 3) by observing that the terms that depend on ξ_i and ξ_j form a double integral that equals the function $G_{N,M,L,i,j,m}^{(2)}(\rho)$ for different values of N, M, L, i, j, and m. In the special case of $N = 2$, the joint conditional distribution becomes degenerate, since $|v_j|^2 = \rho - |v_i|^2$. Using

the second equality in (42), the cross-moment can be expressed as

$$\mathbb{E}\{|v_i|^{2p}|v_j|^{2p}|\rho\} = \int_0^\rho \xi_i^{p_i}(\rho - \xi_i)^{p_j} f_{|v_i|^2 \| \|\mathbf{v}\|^2}(\xi_i|\rho)d\xi_i$$

which is solved by a similar identification.

Proof of Theorem 2: Using the law of total probability, the conditional moment can be expressed as

$$\mathbb{E}\{|v_i|^{2p}|\varrho\} = \int_a^b \mathbb{E}\{|v_i|^{2p}|\rho\} f_{\|\mathbf{v}\|^2|\varrho}(\rho|\varrho)d\rho$$
$$= \frac{\int_a^b \mathbb{E}\{|v_i|^{2p}|\rho\} f_{\|\mathbf{v}\|^2}(\rho)d\rho}{\int_a^b f_{\|\mathbf{v}\|^2}(\rho)d\rho} \tag{47}$$

where ρ represents the exact value of $\|\mathbf{v}\|^2$. Observe that the conditional moment $\mathbb{E}\{|v_i|^{2p}|\rho\}$ is given by Theorem 1, and that the pdf of the norm is given in (4), (5), and (6) for the three different cases of eigenvalue structure. The integral in the numerator of (47) can be solved directly (using Lemma 3), while the integral in the denominator can be solved by straightforward integration using Lemma 2. The conditional cross-moments can be derived by the same approach.

Proof of Theorem 3: The theorem follows from observing that

$$\mathbb{E}\{\rho^p|\varrho\} = \frac{\int_a^b \rho^p f_{\|\mathbf{v}\|^2}(\rho)d\rho}{\int_a^b f_{\|\mathbf{v}\|^2}(\rho)d\rho}. \tag{48}$$

The numerator is given by Lemma 3 and the denominator was calculated in the proof of Theorem 2.

Proof of Corollary 1: The expression $\widetilde{\mathbf{R}} = (\mathbf{D}^T \otimes \mathbf{C})\mathbf{R}(\mathbf{D}^* \otimes \mathbf{C}^H)$ is obtained by using the rule $\text{vec}(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X})\text{vec}(\mathbf{Y})$, for general matrices $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$. The distribution of the squared norm ρ of the effective channel is achieved by using that the Frobenius norm and that the distribution of a complex Gaussian vector is invariant under unitary matrix transformations. Since zero-valued eigenvalues have no impact on the norm, the distribution of ρ is equivalent to that of $\|\mathbf{v}\|^2$, where $\mathbf{v} \in \mathcal{CN}(\mathbf{0}, \boldsymbol{\Lambda})$ and $\boldsymbol{\Lambda}$ is a diagonal matrix with all nonzero (i.e., strictly positive) eigenvalues of $\widetilde{\mathbf{R}}$.

Proof of Corollary 2: Let $\mathbf{v} = \mathbf{U}^H \text{vec}(\widetilde{\mathbf{H}})$ and observe that $\mathbf{v} \in \mathcal{CN}(\mathbf{0}, \boldsymbol{\Lambda})$. The corollary follows from straightforward and tedious expansion of $\|\mathbf{B}\text{vec}(\widetilde{\mathbf{H}})\|^2 = \|\mathbf{Av}\|^2 = \sum_{k=1}^{M} |\sum_{l=1}^{N} a_{kl} v_l|^2$ and $\|\mathbf{Av}\|^4 = (\|\mathbf{Av}\|^2)^2$. The expectation is evaluated using Theorem 1 and 2.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 311–335, 1998.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, 1999.

[3] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 321–331, 2003.

[4] J. Wallace and M. Jensen, "Measured characteristics of the MIMO wireless channel," in *Proc. IEEE Vehicular Technology Conf. (VTC)*, 2001, vol. 4, pp. 2038–2042.

[5] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wideband MIMO radio channels based on NLOS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, May 2004.

[6] B. Sklar, "Rayleigh fading channels in mobile digital communication systems. I. Characterization," *IEEE Commun. Mag.*, vol. 35, pp. 90–100, 1997.

[7] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[8] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[9] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[10] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading correlations," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 239–241, Jun. 2002.

[11] A. Hjørungnes and D. Gesbert, "Precoding of orthogonal space-time block codes in arbitrarily correlated MIMO channels: Iterative and closed-form solutions," *IEEE Trans. Wireless Commun.*, vol. 6, no. 3, pp. 1072–1082, Mar. 2007.

[12] E. Björnson, E. Jorswieck, and B. Ottersten, "On the impact of spatial correlation and precoder design in MIMO systems with space-time block coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2009, pp. 2741–2744.

[13] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 1995, vol. 1, pp. 331–335.

[14] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[15] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, Mar. 2006.

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

[17] J. Ramiro-Moreno, K. I. Pedersen, and P. Mogensen, "Capacity gain of beamforming techniques in a WCDMA system under channelization code constraints," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1199–1208, Jul. 2004.

[18] N. Jindal, "Antenna combining for the MIMO downlink channel," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3834–3844, Oct. 2008.

[19] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Utilizing the spatial information provided by channel norm feedback in SDMA systems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3278–3293, Jul. 2008.

[20] E. Björnson and B. Ottersten, "Exploiting long-term statistics in spatially correlated multi-user MIMO systems with quantized channel norm feedback," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2008, pp. 3117–3120.

[21] X. Zhang, E. Jorswieck, and B. Ottersten, "User selection schemes in multiple antenna broadcast channels with guaranteed performance," presented at the IEEE SPAWC, Helsinki, Finland, Jun. 17–20, 2007.

[22] M. Kountouris, R. de Francisco, D. Gesbert, D. Slock, and T. Sälzer, "Efficient metrics for scheduling in MIMO broadcast channels with limited feedback," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2007, vol. 3, pp. 109–112.

[23] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Acquiring partial CSI for spatially selective transmission by instantaneous channel norm feedback," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 1188–1204, Mar. 2008.

[24] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, submitted for publication.

[25] E. Scheuer, "Reliability of an m-out-of-n system when component failure induces higher failure rates in survivors," *IEEE Trans. Rel.*, vol. 37, no. 1, pp. 73–74, Apr. 1988.

[26] S. Amari and R. Misra, "Closed-from expressions for distribution of sum of exponential random variables," *IEEE Trans. Rel.*, vol. 46, no. 4, pp. 519–522, Dec. 1997.

[27] E. Cramer and U. Kamps, *Sequential k-Out-of-n Systems*. New York: Elsevier, 2001, vol. 20, ch. 12, pp. 693–739.

[28] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 3, pp. 451–460, Mar. 1999.

[29] E. Björnson and B. Ottersten, "Post-user-selection quantization and estimation of correlated Frobenius and spectral channel norms," in *Proc. IEEE PIMRC*, 2008.

[30] R. Zakhour and D. Gesbert, "A two-stage approach to feedback design in multi-user MIMO channels with limited channel state information," presented at the IEEE PIMRC, Athens, Greece, Sep. 3–7, 2007.

[31] A. Marques, G. Giannakis, F. Digham, and F. Ramos, "Power-efficient wireless OFDMA using limited-rate feedback," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 685–696, Feb. 2008.

[32] S. Loyka, "Channel capacity of MIMO architecture using the exponential correlation matrix," *IEEE Commun. Lett.*, vol. 5, no. 9, pp. 369–371, Sep. 2001.

[33] M. Trivellato, F. Boccardi, and F. Tosato, "User selection schemes for MIMO broadcast channels with limited feedback," in *Proc. IEEE Vehicular Technology Conf. (VTC)—Spring*, 2007, pp. 2089–2093.

[34] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[35] D. Love and R. Heath, "Grassmannian beamforming on correlated MIMO channels," in *Proc. IEEE GLOBECOM*, 2004, pp. 106–110.

Emil Björnson (S'07) was born in Malmö, Sweden, in 1983. He received the M.S. degree in engineering mathematics from Lund University, Lund, Sweden, in 2007. He is currently working towards the Ph.D. degree in telecommunications at the Signal Processing Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden.

His research interests include wireless communications, resource allocation, estimation theory, stochastic signal processing, and mathematical optimization.

David Hammarwall (S'03–M'07) was born in Stockholm, Sweden, in 1977. He received the M.S. degree (with highest hons.) in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 2003. In 2001-02 he pursued M.S.-level studies at the Department of Electrical Engineering, Stanford University, Stanford, CA (as part of the KTH degree). He was awarded an "Excellent Graduate Student Position," from the president's office at KTH, and received the Ph.D. degree in telecommunications from the same university, in 2007.

He has since joined Ericsson Research, Stockholm, Sweden. His research interests include wireless communications, resource optimization, beamforming, and scheduling.

Björn Ottersten (S'87–M'89–SM'99–F'04) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986 and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1989.

He has held research positions at the Department of Electrical Engineering, Linköping University; the Information Systems Laboratory, Stanford University; and the Katholieke Universiteit Leuven, Leuven, Belgium. During 1996–1997, he was Director of Research at ArrayComm Inc., San Jose, CA, a start-up company based on Ottersten's patented technology. In 1991, he was appointed Professor of Signal Processing at the Royal Institute of Technology (KTH), Stockholm, Sweden. From 2004 to 2008, he was Dean of the School of Electrical Engineering at KTH, and from 1992 to 2004 he was head of the Department for Signals, Sensors, and Systems at KTH. He is also Director of security and trust at the University of Luxembourg. His research interests include wireless communications, stochastic signal processing, sensor array processing, and time-series analysis.

Dr. Ottersten has coauthored papers that received an IEEE Signal Processing Society Best Paper Award in 1993, 2001, and 2006. He has served as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and on the Editorial Board of the *IEEE Signal Processing Magazine*. He is currently Editor-in-Chief of the *EURASIP Signal Processing Journal* and a member of the Editorial Board of the *EURASIP Journal of Advances Signal Processing*. He is a Fellow of EURASIP. He is a first recipient of the European Research Council advanced research grant.