

Performance of Vector Perturbation Multiuser MIMO Systems with Limited Feedback

Daniel J. Ryan^{1,2}, Iain B. Collings², I. Vaughan L. Clarkson³ and Robert W. Heath Jr.⁴

¹School of Electrical and Information Engineering, The University of Sydney, AUSTRALIA, dan@ee.usyd.edu.au

²Wireless Technologies Laboratory, CSIRO ICT Centre, AUSTRALIA

³School of Information Technology and Electrical Engineering, The University of Queensland, AUSTRALIA

⁴Wireless Networking and Communications Group (WNCG), Department of Electrical and Computer Engineering,
The University of Texas at Austin, USA

Abstract

This paper considers the multiuser multiple-input multiple-output (MIMO) Rayleigh fading broadcast channel. We consider the case where the multiple transmit antennas are used to deliver independent data streams to multiple users via a multi-user technique known as vector perturbation. We propose lattice-theoretic and rate-distortion based approaches to analyze the performance of these systems, taking into account the practical restrictions imposed by limited feedback and training. We show that performance is primarily determined by the ratio between the number of users and the number of transmit antennas. We then propose a new practical low-complexity low-rate feedback scheme, and show that the performance approaches the ideal rate-distortion based scheme.

Index Terms

Vector perturbation, vector precoding, limited feedback, MIMO systems, multiuser, broadcast channel, wireless communications.

This work has been submitted in part to the 2008 IEEE International Conference on Communications, Beijing, China, May 2008.

The work of R. W. Heath was supported in part by the National Science Foundation under grant CCF-514194, CNS-626797, and the DARPA IT-MANET program, Grant W911NF-07-1-0028.

I. INTRODUCTION

Cellular basestations and wireless LAN access points that transmit using multiple antennas to K non-collocated users are examples of the multiuser multiple-input multiple-output (MIMO) Rayleigh fading broadcast channel. In this paper, we consider the case where each of the users has a single receive antenna. Recent theoretical results have shown that in this multiuser case, the data rate scales with the number of antennas in the same way as for point-to-point MIMO communications, *e.g.* [1, 2].

A promising practical transmission method known as vector perturbation has been proposed for this multiuser MIMO channel model, that can achieve rates near capacity [3]. It has superior performance to linear precoding techniques, such as channel inversion and zero-forcing beamforming, as well as non-linear Tomlinson-Harashima precoding [3]. It approaches the performance of ideal dirty-paper coding (DPC) methods [2] which are too complex to be implemented in practice. When using vector perturbation, the data vector to be transmitted is constrained to lie within a $2K$ -dimensional hypercube of side length 1, and is modified by the addition of a *perturbation vector* consisting of complex-valued integers. The perturbation vector is chosen so as to minimize the transmit power for each channel use. This minimization is an instance of the well-studied NP-hard problem of finding the closest lattice point, where here the lattice is determined by the channel. A common method to perform the search is the sphere-decoding algorithm [4]. At each receiver, a modulo decoding process is applied independently at each user to completely remove the perturbation.

The performance of vector perturbation systems are a function of the average power of the sphere-encoded signals, \mathcal{E}_{se} , as this determines the noise gain at the output of each user's demodulator. Due to the fact that sphere-encoding is an NP-hard problem, it is unsurprising that calculating \mathcal{E}_{se} exactly has proved intractable, although some useful partially numerical results have been derived. A lower bound on \mathcal{E}_{se} for arbitrary K and constellation was proposed in [3]. This bound requires estimation by numerical simulation of the system (so \mathcal{E}_{se} can be estimated at the same time), however it does provide insight into the resulting direction of the perturbation vector. More recently, an expression for \mathcal{E}_{se} as $K \rightarrow \infty$, was derived for arbitrary constellations using the replica analysis method of statistical physics [5]. However, this result requires numerical evaluation of a pair of coupled fixed-point equations, and therefore the insight into \mathcal{E}_{se} is limited.

In this paper we provide a new lower bound on \mathcal{E}_{se} for the case of uniformly distributed inputs over i.i.d. Gaussian fading channels, using a lattice-theoretic approach. We propose that this new result can be used as a good approximation to \mathcal{E}_{se} for discrete constellations. We then derive an even simpler approximation for \mathcal{E}_{se} by finding the limit of the lower bound as the number of users goes to infinity, which turns out to be a function

of only the ratio between the number of transmit antennas and the number of users. We provide new upper and lower bounds on the resulting bit error rate for the modulo decoder, given \mathcal{E}_{se} , and then provide simulation results to show that these bounds are tight.

We next turn our attention to some practical implementation issues. Vector perturbation requires channel state information (CSI) at the transmitter. Specifically, a channel estimate is required from each user so that the inverse channel matrix may be used as the linear precoder. In practical systems, this can be achieved by a process of first sending training symbols to each user, and then quantizing the resulting channel vector estimate. The quantized vector is mapped to index bits which are fed back to the transmitter via a low-rate feedback control channel. The transmitter then creates the channel estimate using the index bits. This is superior to the case of providing CSI by analog means using a reciprocal channel, as analyzed for the case of zero-forcing beamforming in [6].

Limited feedback codebook design and search strategies, for point-to-point beamforming and spatial multiplexing systems, have been well studied. In these scenarios the feedback is generally of unit-norm beamforming vectors or unitary precoding matrices *e.g.* [7–9]. More recently, the effect of limited feedback on some multi-user systems has also been considered. In [10], a scheme known as zero-forcing beamforming was proposed where only a row-normalized form of the channel inverse matrix is used as a precoding matrix. This requires only feedback of the shape of each user’s measured channel vector. In [10], a system analysis was provided assuming random vector quantization (RVQ) feedback codebooks, initially considered in [11]. It was shown that the number of feedback bits required to maintain a given rate loss increases linearly with the SNR (in dB). This is due to the fact that imperfect CSI results in interference between users. Similar feedback scaling results were obtained in [12–14] for zero-forcing beamforming, and also for standard channel inversion and zero-forcing with DPC in [15].

In this paper, we consider the effect of limited feedback in vector perturbation systems. We derive expressions for the effective noise power at the output of the modulo demodulator, due to limited feedback. We show that to maintain a given effective noise power as the SNR increases, the number of feedback bits per channel coefficient must grow linearly with the SNR (in decibels). We then remove the assumption of perfect receiver CSI and derive the increase in effective noise power due to both training and quantization. The resulting expression shows that the degradation due to training is independent of the SNR, and that in the high SNR regime the effects of training and quantization become purely additive.

Finally, we propose a new low-complexity low-rate feedback scheme as a practical method for providing CSI at the transmitter. The proposed quantization codebook requires storage of only a small number of channel

magnitude levels, and can be searched with complexity $O(N_T \log N_T)$, where N_T is the number of transmit antennas. This is much faster than a $O(e^{N_T})$ search of an unstructured vector quantization codebook, such as that produced by the generalized Lloyd-Max algorithm [16]. We provide simulation results showing that the proposed scheme performs close to what would be achieved using the ideal rate-distortion based feedback scheme, and significantly outperforms linear precoding schemes which do not use vector perturbation.

II. SYSTEM MODEL

We consider the downlink of a narrowband multi-user MIMO system with N_T transmit antennas broadcasting to $K \leq N_T$ spatially dispersed users. Each user has a single receive antenna. Each channel realization $\mathbf{H} \in \mathbb{C}^{K \times N_T}$ is assumed independent from one packet to the next, and is assumed to be constant for the duration of a packet. The elements $h_{k,t}$ of \mathbf{H} represent the channel between the k th user and t th transmit antenna and are chosen according to an independent and identically distributed (i.i.d.) zero-mean circularly symmetric (ZMCS) complex Gaussian distribution $\mathcal{CN}(0, N_0)$. We denote the k th row as \mathbf{h}_k .

We use $(\cdot)'$ to denote matrix transpose, $(\cdot)^\dagger$ to denote matrix conjugate transpose and $\text{Vol}(\cdot)$ to denote the Jordan-measurable volume of a region. We use $(\cdot)^+$ to denote Moore-Penrose pseudoinverse, and since $N_T \geq K$, it follows that $\mathbf{H}^+ = \mathbf{H}^\dagger(\mathbf{H}\mathbf{H}^\dagger)^{-1}$. We also denote the set of Gaussian (complex) integers as $\mathbb{Z}[j]$.

Given the transmitted vector $\mathbf{x} = [x_1 \dots x_{N_T}]' \in \mathbb{C}^{N_T \times 1}$, the received symbol at user k is given by

$$y_k = \mathbf{h}_k \mathbf{x} + n_k$$

where each n_k is $\mathcal{CN}(0, N_0)$, where N_0 is the noise power. The received symbols can be combined as $\mathbf{y} = [y_1 \dots y_K]' \in \mathbb{C}^{K \times 1}$ to give:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where $\mathbf{n} = [n_1 \dots n_K]'$. The transmitted vector \mathbf{x} is a modified ‘‘perturbed’’ and ‘‘precoded’’ form of the data vector. We first consider \mathbf{a} , and then how \mathbf{x} is generated from it.

The data vector is $\mathbf{a} = [a_1 \dots a_K] \in \mathbf{B}^K$ where \mathbf{B}^K is the K -ary Cartesian product of the region

$$\mathbf{B} \triangleq \{ a : |\text{Re}\{a\}| < 0.5, |\text{Im}\{a\}| < 0.5\}. \quad (2)$$

Clearly, $\text{Vol}(\mathbf{B}^K) = 1$. A practical choice for the distribution of the data vector \mathbf{a} is to assume that each element is an identically and independently distributed (i.i.d) point from an μ^2 -ary square QAM constellation, where the real and imaginary components of each data symbol a_k , $k = 1, \dots, K$ belong to the set $\{\pm x/(2\mu) : x = 1, 3, \dots, \mu - 1\}$. We will see that this structure corresponds to the case of the signal points

having symmetric decoding regions. For analytical purposes we will also consider the case of uniformly distributed inputs where \mathbf{a} is an i.i.d. random variable with probability distribution function $p(\mathbf{a}) = \chi_{\mathbf{B}^K}(\mathbf{a})$ where $\chi(\cdot)$ is the characteristic (indicator) function.

To generate \mathbf{x} , \mathbf{a} is first perturbed and then precoded to create the sphere-encoded signal vector, \mathbf{s} , *i.e.*

$$\mathbf{s} = \mathbf{H}^+(\mathbf{a} + \mathbf{p}) \quad (3)$$

where \mathbf{p} is the Gaussian (complex) integer-valued perturbation vector given by

$$\mathbf{p} = \underset{\mathbf{q} \in \mathbb{Z}[j]^K}{\operatorname{argmin}} \left\| \mathbf{H}^+(\mathbf{a} + \mathbf{q}) \right\|^2. \quad (4)$$

Note that \mathbf{p} are unscaled integers. This approach is in contrast to [3] where the \mathbf{p} are scaled by a factor τ , determined by the constellation size. Here, we effectively set τ equal to one, and then the constellation (e.g. 64-QAM) for each user is scaled so that each constellation point lies within \mathbf{B} , *i.e.* $\mathbf{a} \in \mathbf{B}^K$. Before discussing the distribution of \mathbf{a} we will first discuss the choice of \mathbf{p} .

The minimization in (4) required to choose \mathbf{p} is the well-studied problem of finding the closest lattice point in an unbounded lattice. An optimal approach will have complexity exponential in K *e.g.* the sphere-decoding algorithm of [4]. As K increases, suboptimal methods of polynomial complexity may be employed, such as the lattice reduction based approach of [17], and the singular value decomposition based approach of [18]. It was shown in [19] that LLL-reduction [17] can achieve the full-diversity order. For the purposes of analytical tractability, we assume that the algorithm used to solve (4) gives the optimal solution. An optimal solution to (4) implies that \mathbf{s} will lie within the Voronoi region of the lattice point at the origin. That is, $\mathbf{s} \in \mathcal{V}(\Lambda_{\mathbf{H}^+}, \mathbf{0})$ denotes the Voronoi region about the origin of the lattice $\Lambda_{\mathbf{H}^+}$, which is the lattice with complex-valued generator matrix \mathbf{H}^+ . We note that without the sphere-encoding operation (*i.e.* when \mathbf{p} is $\mathbf{0}$), this corresponds to a simple channel inversion.

The final step in generating \mathbf{x} is to scale \mathbf{s} as follows:

$$\mathbf{x} = \sqrt{\frac{\mathcal{E}_x}{\mathcal{E}_{\text{se}}}} \mathbf{s} \quad (5)$$

where \mathcal{E}_x is the average signal power, *averaged over all packets*, and

$$\mathcal{E}_{\text{se}} \triangleq E_{\mathbf{a}, \mathbf{H}}[\|\mathbf{s}\|^2] = E \left[\min_{\mathbf{q} \in \mathbb{Z}[j]^K} \left\| \mathbf{H}^+(\mathbf{a} + \mathbf{q}) \right\|^2 \right] \quad (6)$$

is the expected power of the sphere-encoded vector \mathbf{s} , where the expectation is taken over \mathbf{a} and \mathbf{H} . Similarly, $\mathcal{E}_{\text{se}|\mathbf{H}} \triangleq E[\|\mathbf{s}\|^2 | \mathbf{H}]$ is the expected power of the sphere-encoded vector \mathbf{s} given \mathbf{H} . Since the same value of \mathcal{E}_{se}

is used for each packet, this implies that the expected transmitted energy is different for each channel realization, as was considered in [3, 5]. The value of \mathcal{E}_{se} obtained by channel inversion is much larger than for the vector perturbation case, and is actually infinite for the case of $N_T = K$ [20]. We also define the following signal-to-noise ratio (SNR), ρ , as the ratio between the average transmit power and the average noise power, *i.e.*

$$\rho = \frac{\mathcal{E}_x}{N_0}. \quad (7)$$

At the k th user's receiver, the data is recovered using a modulo demodulator [3]:

$$\hat{a}_k = \left[\sqrt{\frac{\mathcal{E}_{\text{se}}}{\mathcal{E}_x}} y_k \right]_{\text{mod } B} = \left[a_k + p_k + \sqrt{\frac{\mathcal{E}_{\text{se}}}{\mathcal{E}_x}} n_k \right]_{\text{mod } B} = [a_k + \eta_k]_{\text{mod } B}. \quad (8)$$

where $\eta_k \triangleq \sqrt{\mathcal{E}_{\text{se}}/\mathcal{E}_x} n_k$ for all k is the effective noise, and therefore $\text{Var}\{\eta_k\} = \mathcal{E}_{\text{se}} N_0 / \mathcal{E}_x$.

III. ANALYSIS OF \mathcal{E}_{se} AND BIT ERROR RATES

To be able to fully analyze a vector perturbation system with limited feedback, here we provide simple closed-form approximations to the value of \mathcal{E}_{se} . We first consider the case of uniformly distributed inputs and provide numerical simulation results to demonstrate that this approximates the value of \mathcal{E}_{se} well for practical QAM constellations. Using lattice-theoretic techniques we provide an analytic lower bound on $\mathcal{E}_{\text{se}|\mathbf{H}}$ under the assumption of uniformly distributed inputs. We then apply a random matrix theory result to derive a lower bound on \mathcal{E}_{se} which we propose to use as an approximation. Finally, we derive a simpler equation to approximate the value of \mathcal{E}_{se} by considering the limit as the number of users and transmit antennas go to infinity with constant ratio. We note that in contrast to previous results which require numerical evaluation [3, 5], our results are in closed form.

A. Uniformly Distributed Inputs

We now provide numerical simulation results to justify the consideration of uniformly distributed inputs used in our further analysis. This is achieved by showing that \mathcal{E}_{se} does not vary much with the choice of constellation, and is well approximated by uniformly distributed inputs. In addition we will see that \mathcal{E}_{se} does not vary much with K , as also observed in [3].

Figure 1 provides values for \mathcal{E}_{se} obtained by Monte Carlo simulation for systems with $N_T = K$ and $N_T = 2K$ transmit antennas where we consider $K = 2$ to 7 users. We simulate using QPSK and 64-QAM constellations, and also for uniformly distributed inputs. For each plot over 100000 channel usages are simulated using at least 1000 independently generated channels.

We see that \mathcal{E}_{se} does not vary much with the choice of constellation, but approaches the value of \mathcal{E}_{se} for the uniform input case as the constellation size increases. This can be explained intuitively by noting that the constellation structure is distorted by precoding, so that the elements of \mathbf{s} are placed in a distorted fashion over $\mathcal{V}(\Lambda_{\mathbf{H}^+}, \mathbf{0})$, giving a value of $\mathcal{E}_{\text{se}|\mathbf{H}}$ similar to that obtained by choosing the elements uniformly over $\mathcal{V}(\Lambda_{\mathbf{H}^+}, \mathbf{0})$. We also confirm the observation of [3] that \mathcal{E}_{se} does not vary much with K , and converges as K increases.

B. A Lower Bound on $\mathcal{E}_{\text{se}|\mathbf{H}}$

We now derive a lower bound on $\mathcal{E}_{\text{se}|\mathbf{H}}$ under the assumption of uniformly distributed inputs. We then provide simulation results which show that this bound is less than 1.5dB of the actual value for approximately 90% of \mathbf{H} for $N_T = 2$ transmit antennas, and indicate that this improves as K increases. This lower bound will be used subsequently to derive a lower bound on \mathcal{E}_{se} , which will be shown to be a good approximation via simulation.

Lemma 1: The expected sphere-encoded signal power, $\mathcal{E}_{\text{se}|\mathbf{H}}$, of a wireless MIMO vector perturbation system with N_T transmit antennas and K independent users where the data vector \mathbf{a} is uniformly distributed satisfies

$$\mathcal{E}_{\text{se}|\mathbf{H}} \geq \mathcal{E}_{\text{se}|\mathbf{H},\text{LB}} = \frac{K\Gamma(K+1)^{1/K}}{(K+1)\pi} \det(\mathbf{W})^{-1/K}$$

where $\mathbf{W} \triangleq \mathbf{H}\mathbf{H}^\dagger$ and $\Gamma(\cdot)$ denotes the gamma function.

Proof: First we consider a QR-type decomposition of \mathbf{H}^+ namely $\mathbf{H}^+ = \mathbf{U}\mathbf{R}$ where $\mathbf{U} \in \mathbb{C}^{N_T \times K}$ and $\mathbf{U}^\dagger\mathbf{U} = \mathbf{I}_K$, and $\mathbf{R} \in \mathbb{C}^{K \times K}$ is upper triangular. Now, we propose to generate the vector \mathbf{t} from the data vector in the same way as (4), i.e. $\mathbf{t} = \mathbf{R}(\mathbf{a} + \mathbf{p})$, where $\mathbf{p} \in \mathbb{Z}[j]^K$ is chosen to minimize $\|\mathbf{t}\|$. Since $\|\mathbf{s}\| = \|\mathbf{U}\mathbf{t}\| = \|\mathbf{t}\|$, $\mathcal{E}_{\text{se}|\mathbf{H}}$, and generally \mathbf{p} , are identical to the original system.

Now, we note that if sphere-decoding is performed optimally, it follows that \mathbf{t} is within the Voronoi region $\mathcal{V}(\Lambda_{\mathbf{R}}, \mathbf{0})$. Moreover, since \mathbf{a} is uniformly distributed, it follows that \mathbf{t} is uniformly distributed over $\mathcal{V}(\Lambda_{\mathbf{R}}, \mathbf{0})$.

The volume of the Voronoi region of a lattice in \mathbb{R}^n with non-singular generator matrix $\mathbf{G} \in \mathbb{R}^n$ is $\sqrt{\det(\mathbf{G}^\dagger\mathbf{G})}$ [21]. However, as \mathbf{R} is complex, the volume of the Voronoi region can be found by considering the transformation of \mathbf{R} from \mathbb{C}^K to \mathbb{R}^{2K} (see e.g. [22]). It follows that the volume of the Voronoi region is squared, i.e. $\text{Vol}(\mathcal{V}(\Lambda_{\mathbf{R}}, \mathbf{0})) = \det(\mathbf{R}^\dagger\mathbf{R}) = \det(\mathbf{R}^\dagger\mathbf{U}^\dagger\mathbf{U}\mathbf{R}) = \det((\mathbf{H}^+)^\dagger\mathbf{H}^+) = \det((\mathbf{H}\mathbf{H}^\dagger)^+) = \det(\mathbf{W}^{-1})$.

Now, for any polytope \mathcal{R} , its second moment $\sigma^2(\mathcal{R})$ is defined as $\sigma^2(\mathcal{R}) \triangleq \frac{1}{V(\mathcal{R})} \int_{\mathcal{R}} \|\mathbf{s}\|^2 ds$. Therefore, the expected magnitude of \mathbf{s} for a particular channel instance \mathbf{H} is given by $\mathcal{E}_{\text{se}|\mathbf{H}} = \sigma^2(\mathcal{V}(\Lambda_{\mathbf{H}^+}, \mathbf{0}))$. To obtain a lower bound on $\mathcal{E}_{\text{se}|\mathbf{H}}$, we note that if \mathbf{s} is uniformly distributed over a polytope \mathcal{R} , then the second moment is lower bounded by the second moment of a hypersphere \mathcal{S} of dimension $\dim(\mathcal{R})$ and volume $V(\mathcal{R})$. This is the

same principle used to derive Zador's quantization lower bound [23]. Now, the volume and second moment of a region \mathcal{R} can always be related as follows:

$$G(\mathcal{R}) \triangleq \frac{\sigma^2(\mathcal{R})}{\dim(\mathcal{R})V(\mathcal{R})^{2/\dim(\mathcal{R})}}$$

where $G(\mathcal{R})$ is the dimensionless second moment [21, Ch. 21] and $\dim(\mathcal{R})$ is the dimension of \mathcal{R} . Now, denote \mathcal{S}_L as the L -dimensional sphere in \mathbb{R}^L of unit volume. The normalized second moment is given by $G(\mathcal{S}_L) = \Gamma(L/2 + 1)^{2/L}/((L + 2)\pi)$ [21, Ch. 21]. Hence,

$$E[\|\mathbf{s}\|^2 | \mathbf{H}] \geq 2KG(S_{2K}) \det(\mathbf{W})^{-1/K} = \frac{K\Gamma(K + 1)^{1/K}}{(K + 1)\pi} \det(\mathbf{W})^{-1/K}.$$

■

In Figure 2 we plot the ratio $\mathcal{E}_{\text{se}|\mathbf{H}}/\mathcal{E}_{\text{se}|\mathbf{H},\text{LB}}$ for uniformly distributed inputs, obtained via numerical simulation. We consider the following $N_T \times K$ systems: 2×2 , 4×2 and 4×4 systems. For each configuration of users and transmit antennas, we simulate 20000 channels, where for each channel instance a packet consisting of 100 i.i.d. symbols is transmitted to each user. We see that for the scenarios considered $\mathcal{E}_{\text{se}|\mathbf{H}}/\mathcal{E}_{\text{se}|\mathbf{H},\text{LB}} = 1.5\text{dB}$ all the curves are below 10%, implying that our proposed lower bound $\mathcal{E}_{\text{se}|\mathbf{H},\text{LB}}$ is within 1.5dB of the actual value $\mathcal{E}_{\text{se}|\mathbf{H}}$ for at least 90% of the channel instances.

To derive a lower bound on \mathcal{E}_{se} we note that the h th moment of a complex Wishart distributed matrix $\mathbf{W} \sim \mathcal{W}_K(N_T, \mathbf{I})$ is given by [24, Thm 2.11]

$$E[\det(\mathbf{W})^h] = \prod_{\ell=0}^{K-1} \frac{\Gamma(N_T + h - \ell)}{\Gamma(N_T - \ell)}. \quad (9)$$

We set $h = -1/K$ and take the expectation of the lower bound in (9) to obtain the following corollary.

Corollary 1: The expected sphere-encoded signal power, \mathcal{E}_{se} , of a wireless MIMO vector perturbation system with N_T transmit antennas and K independent users with uniformly distributed inputs satisfies

$$\mathcal{E}_{\text{se}} \geq \mathcal{E}_{\text{se, LB}} \triangleq \frac{K\Gamma(K + 1)^{1/K}}{(K + 1)\pi} \prod_{\ell=0}^{K-1} \frac{\Gamma(N_T - \frac{1}{K} - \ell)}{\Gamma(N_T - \ell)}. \quad (10)$$

Corollary 1 provides a closed-form lower bound for \mathcal{E}_{se} for arbitrary K for uniformly distributed inputs. We note that this is in contrast to the lower bound for arbitrary K presented in [3] and the asymptotic result in [5], both of which require simulation or solving of coupled equations, respectively. Importantly, returning to Figure 1, we see that the lower bound is a good approximation for the actual value of \mathcal{E}_{se} , and hence its use as an approximation to \mathcal{E}_{se} is valid.

C. Asymptotic Lower Bound on \mathcal{E}_{se}

We now derive an even simpler approximation to \mathcal{E}_{se} using an asymptotic analysis. It is desirable to do this as it is not immediately clear from Corollary 1 to ascertain the behavior of \mathcal{E}_{se} as a function of K and N_T without further numerical evaluation. To do this, we now derive the limiting value of $\mathcal{E}_{\text{se,LB}}$ as $K \rightarrow \infty$ with the ratio $\alpha = K/N_T$ fixed. The partially numeric result obtained in [5] was also obtained by considering $K \rightarrow \infty$. We will see that this gives a simple good approximation to the actual value of \mathcal{E}_{se} , from which we can now easily determine the effect of changing the number of antennas.

Theorem 1: The expected sphere-encoded signal power, \mathcal{E}_{se} , of a wireless MIMO vector perturbation system with N_T transmit antennas and K independent users where the data vector \mathbf{a} is uniformly distributed satisfies

$$\lim_{K \rightarrow \infty} \mathcal{E}_{\text{se,LB}} = \frac{\alpha}{\pi} (1 - \alpha)^{\frac{1-\alpha}{\alpha}}$$

for all α such that $1 \leq \alpha < \infty$, where $\alpha \triangleq K/N_T$.

Proof: See Appendix A. ■

It can be shown that for the case that $\alpha = 1$, that $\lim_{K \rightarrow \infty} \mathcal{E}_{\text{se,LB}} = \frac{1}{\pi}$.

In Figure 1, we have also plotted the asymptotic lower bound for $\alpha = 0.5, 1$. We see that although Theorem 1 is an asymptotic lower bound, it is a good approximation to the actual value of \mathcal{E}_{se} for all values of $\alpha \leq 1$, regardless of the actual constellation used. Theorem 1 confirms our original observation from Figure 1 that the actual value of \mathcal{E}_{se} is dependent primarily on α , which we recall is the ratio of the number of users to transmit antennas, and not K or N_T individually, or the constellation used. This dependence is also observed for the simulation-derived bound from [3], and also the asymptotic replica-analysis result, \mathcal{E}_{RA} from [5].

In Figure 3 we compare the values of \mathcal{E}_{se} for both Corollary 1 and Theorem 1 bounds with the real value of \mathcal{E}_{se} , for $K = 4$ when QPSK and 64-QAM constellations are used. We also compare to the asymptotic result \mathcal{E}_{RA} from [5]. We see that for practical values of α , that is when N_T is not significantly greater than K , that the proposed bound of Corollary 1 provides the tightest approximation, whereas for extreme values of N_T , we see that \mathcal{E}_{RA} provides the better approximation. Our new simple asymptotic result is consistently close. Note however that \mathcal{E}_{RA} requires an iterative process to evaluate, where each step requiring numerical integrations, whereas the proposed bounds from Corollary 1 and Theorem 1 can be easily evaluated.

Also Theorem 1 and Figure 3 demonstrate that the value of $\mathcal{E}_{\text{se,LB}}$ decreases rapidly as the number of transmit antennas is increased for a fixed number of users K . By assuming that Theorem 1 provides a good approximation to \mathcal{E}_{se} , we see that by using $\alpha = 2/3, 1/2$ or $1/4$, we decrease $\mathcal{E}_{\text{se,LB}}$ by approximately 4.2, 6.0 and 10.2 dB

respectively, however the gains of adding additional antennas diminishes as N_T increases.

D. Bit Error Rates

We now derive the bit error rate (BER) for the multi-user vector perturbation system using Gray-coded square-QAM constellations using the modulo demodulator in (8) under the assumption of perfect transmitter CSI. We then discuss the implications of training and limited feedback on calculation of the bit error rate.

Lemma 2: The probability of bit error, P_b for a vector perturbation system with N_T transmit antennas and K users using an μ^2 -ary square QAM constellation satisfies

- 1) $P_b \geq P_{b, \text{LB}} \triangleq \frac{2}{\log_2 \mu} [Q(\sqrt{2}\varphi) - Q((2\mu - 1)\sqrt{2}\varphi)],$
- 2) $P_b \leq P_{b, \text{UB}} \triangleq \frac{2}{\log_2 \mu} [Q(\sqrt{2}\varphi) + (\log_2 \mu - 1)Q(3\sqrt{2}\varphi)]$
- 3) $\lim_{\rho \rightarrow \infty} P_b = \frac{2}{\log_2 \mu} Q(\sqrt{2}\varphi).$

where $\varphi \triangleq \frac{1}{2\mu} \sqrt{\frac{\mathcal{E}_m}{\mathcal{E}_{\text{se}}}}$ and $Q(\cdot)$ is the Q -function.

Proof: See Appendix B. ■

In Fig. 4 we plot the bit error rate as a function of SNR for various vector perturbation systems. We consider the case of $K = 2$ users using either $N_T = 2$ or 4 transmit antennas. The values of \mathcal{E}_{se} in the bounds are obtained using the Monte Carlo simulation method of Fig. 4. We see that the upper and lower bounds of Lemma 2 are tight, and the limit is reached. It is clear from Lemma 2 and the figure that the effect of changing \mathcal{E}_{se} is simply to shift the error rate curve and not change the slope. We see that for this scenario, doubling the number of transmit antennas improves the performance by a factor of approximately 7dB. Comparing this to the effect of increasing the constellation size, we see that we can increase the data rate by 2 bits per user per channel usage, and maintain the same BER by doubling the number of transmit antennas.

We now consider the bit error rate for the case of imperfect CSI. In this case the effective noise term $\boldsymbol{\eta}$ defined in (15) is not Gaussian since the interference component of $\boldsymbol{\eta}$ is not Gaussian. Therefore, the above BER analysis can not be immediately applied, especially at high SNR where the effective noise is dominated by the non-Gaussian interference term. An analysis is beyond the scope of this work, however we now briefly examine the effect of training and limited feedback via simulation.

IV. VECTOR PERTURBATION WITH LIMITED FEEDBACK

In this section we consider the practical case where perfect CSI is not available at the transmitter. First, we consider the case where a quantized version of the channel vector for each user is sent back to the transmitter via

a limited feedback control channel. We show that the number of feedback bits required to maintain a constant rate loss compared to the ideal perfect feedback case scales linearly with the SNR. We then extend this analysis to consider the channel estimation error due to training, and show that the performance loss is determined by the ratio between the energy dedicated to training, and the energy dedicated to the data. We show that in the high SNR regime and with sufficiently large number of bits the interference due to the channel error from the training and quantization phases are additive.

A. CSI Obtained via Limited Feedback

In the limited feedback case, each of the K users feeds back the CSI to the transmitter via a low-rate control channel, which is assumed to be error-free with zero-delay [7, 25]. For vector perturbation schemes we propose that the feedback channel is used to convey the CSI in the form of a quantized version of each channel vector \mathbf{h}_k from each user, as the pseudoinverse of \mathbf{H} is required by the vector perturbation approach. Therefore we consider the quantization of the entire vector according to a mean-square error criterion. This is in contrast to other multiuser MIMO schemes such as zero-forcing beamforming [10, 12] in which only the direction of \mathbf{h}_k is used *i.e.* any phase rotation of $\mathbf{h}_k/\|\mathbf{h}_k\|$. This implies that the proposed scheme essentially requires feedback of an extra two real-valued dimensions, the positive magnitude \mathbf{h}_k and also the phase of $\mathbf{h}_k/\|\mathbf{h}_k\|$, or equivalently a single complex dimension. However, we will see that the vastly superior performance of vector perturbation over such schemes means that much less feedback is required to achieve the same performance.

We apply the results of rate distortion theory to give a lower bound on the distortion, which is approached as the length of the vector to be quantized tends to infinity. In practice, this applies to orthogonal frequency division multiplexing (OFDM), where the channel coefficient vectors for each tone can be concatenated. For our analysis, we assume that the number of transmit antennas is finite and the number of quantized channel coefficients is infinite. This gives a lower bound on the distortion for a finite number of parallel channels. We will show by simulation that the rate distortion results can be approached for a relatively small number of channel coefficients.

We now proceed to determine the loss in SNR due to quantization. We first determine the statistics of the quantized channel and the channel error. We then use this to characterize the resulting self-interference due to quantization, and the resulting SNR loss.

We denote \mathbf{H}^Q as the quantized version of the channel obtained at the receiver, and $\mathbf{\Delta}^Q = \mathbf{H} - \mathbf{H}^Q$ as the quantization error. From rate-distortion theory [26, Section 9.7], it is known that if \mathbf{H} is a complex ZMCS i.i.d. Gaussian matrix, the quantization error (in terms of Euclidean mean-squared error (MSE)) for each channel

element $\delta_{k,t}^Q = h_{k,t} - h_{k,t}^Q$ can be modeled as a complex ZMCS i.i.d. Gaussian random variable with variance

$$\text{Var} \left\{ \delta_{k,t}^Q \right\} = 2^{-B}$$

where B is the number of feedback bits used to quantize each channel vector (as the number of channels to quantize goes to infinity). Note that from now on we will drop the subscript from the argument of the $\text{Var} \{ \cdot \}$ operator as it is redundant. Similarly, each element $h_{k,t}^Q$ of the quantized channel matrix is a ZMCS complex Gaussian random variable which is i.i.d. with respect to the other elements of \mathbf{H}^Q , and has variance [26, Section 9.7]

$$\text{Var} \{ h^Q \} = \text{Var} \{ h \} - \text{Var} \{ \delta^Q \} = 1 - 2^{-B}.$$

Note that this additivity property of variance of the quantized channel and the error also applies to any quantizer which satisfies the centroid condition [27].

Using these results, we now consider the effect quantization error has on performance. We have the following system model modified from that presented in Section II,

$$\mathbf{y} = (\mathbf{H}^Q + \mathbf{\Delta}^Q) \sqrt{\frac{\mathcal{E}_x}{\mathcal{E}_{\text{se}}^Q}} \mathbf{s}^Q + \mathbf{n}$$

where $\mathbf{s}^Q \triangleq (\mathbf{H}^Q)^+ (\mathbf{a} + \mathbf{p})$, $\mathbf{p} \in \mathbb{Z}[j]^K$ is chosen to minimize $\|\mathbf{s}^Q\|$, and

$$\mathcal{E}_{\text{se}}^Q \triangleq E_{\mathbf{a}, \mathbf{H}} \left[\left\| (\mathbf{H}^Q)^+ (\mathbf{a} + \mathbf{p}) \right\|^2 \right]. \quad (11)$$

This results in the following equation describing the output of the modulo demodulator,

$$\hat{\mathbf{a}} = [\mathbf{a} + \boldsymbol{\eta}^Q]_{\text{mod } B} \quad (12)$$

where $\boldsymbol{\eta}^Q$ is the self-interference plus noise term defined as

$$\boldsymbol{\eta}^Q \triangleq \mathbf{\Delta}^Q \mathbf{s}^Q + \sqrt{\frac{\mathcal{E}_{\text{se}}^Q}{\mathcal{E}_x}} \mathbf{n}. \quad (13)$$

Clearly, comparing (12) and (8), the only difference between vector perturbation with quantized CSI and perfect CSI is the introduction of the extra noise term $\mathbf{\Delta}^Q \mathbf{s}^Q$. Note that this term does not necessarily have a Gaussian distribution.

We now characterize $\boldsymbol{\eta}^Q$. First, we calculate $\mathcal{E}_{\text{se}}^Q$. Note that \mathbf{H} and \mathbf{H}^Q have the same distribution apart from the variance of the elements. It follows from (6) and (11) that $\mathcal{E}_{\text{se}} \propto 1/\text{Var} \{ h \}$ and $\mathcal{E}_{\text{se}}^Q \propto 1/\text{Var} \{ h^Q \}$. Therefore, applying this to (11) we obtain

$$\mathcal{E}_{\text{se}}^Q = \frac{\text{Var} \{ h \}}{\text{Var} \{ h^Q \}} \mathcal{E}_{\text{se}} = \frac{1}{1 - 2^{-B}} \mathcal{E}_{\text{se}}.$$

We now consider the interference term $\Delta^Q \mathbf{s}$. We make the weak assumption that $p(\mathbf{a}) = p(-\mathbf{a})$ which is satisfied by standard rectangular QAM constellations and uniformly distributed inputs. It follows that $p(\mathbf{s}|\Delta) = p(-\mathbf{s}|\Delta)$ and hence Δ and \mathbf{s} are uncorrelated.

Therefore, noting that \mathbf{s} and Δ have zero mean it follows that

$$\text{Var} \{ \eta^Q \} = \sum_{t=1}^{N_T} E \left[\left| \delta_{k,t}^Q \right|^2 \right] \frac{\mathcal{E}_{\text{se}}^Q}{N_T} + \frac{\mathcal{E}_{\text{se}}^Q N_0}{\mathcal{E}_x} = \mathcal{E}_{\text{se}}^Q \left(\text{Var} \{ \delta^Q \} + \frac{1}{\rho} \right) = \mathcal{E}_{\text{se}} \frac{2^{-B} + \frac{1}{\rho}}{1 - 2^{-B}}.$$

Also, we note that

$$\lim_{B, \rho \rightarrow \infty} \Pr \{ \boldsymbol{\eta}^Q \notin \mathbf{B}^K \} = 0.$$

That is, for sufficiently large B and ρ the distribution of the noise entirely between the constellation nearest neighbor boundaries. It follows that in this scenario the noise variance at the output of the demodulator is unaffected by the modulo operation.

We now consider the noise enhancement, as the ratio of the noise variance with quantized CSI over the noise variance with perfect CSI, *i.e.*

$$v = \frac{\text{Var} \{ \eta^Q \}}{\text{Var} \{ \eta \}} = \frac{\mathcal{E}_{\text{se}}(1 - 2^{-B})}{\left(2^{-B} + \frac{1}{\rho} \right)} \frac{\rho}{\mathcal{E}_{\text{se}}} = \frac{1 + 2^{-B} \rho}{1 - 2^{-B}}. \quad (14)$$

Note that if the extra noise term in (13), $\Delta^Q \mathbf{s}^Q$, had a Gaussian distribution then v would exactly represent an SNR offset in performance due to imperfect CSI (measured in terms of either bit error rate or achievable data rate).

In Figure 5 we plot v for a limited feedback vector perturbation system. We consider systems where the original SNR is 10 and 20dB, with $N_T = 4$ transmit antennas and $K = 4$ users. In the figure, we compare our analytical curves to two practical vector quantization schemes. The analytical curves are calculated under the assumption of using ideal rate-distortion based codebooks. For the first practical scheme, for small B , we use a codebook designed using the generalized (vector) Lloyd-Max algorithm [16], that returns a codebook 2^{BN_T} N_T -dimensional vectors which is a locally optimal MSE quantizer for i.i.d. Gaussian variables. For larger B , the storage and search complexity of such a codebook becomes impractical. Later in Section V we propose a new zero-storage codebook and low-complexity search algorithm for high-rate vector quantization, which we call square lattice angular quantization (SLAQ), and the results for this are also shown in this figure. We see that the curves for the practical codebooks are less than 1dB from the optimal for the case of $\rho = 10$ dB, and 2dB for the case of $\rho = 20$ dB.

We now discuss the degradation in performance as a function of the number of feedback bits. We first note that naturally, as $B \rightarrow \infty$, that $\text{Var}\{\eta^Q\}$ approaches $\mathcal{E}_{\text{se}}/\rho$. However, if B is kept constant while the SNR is increased, a noise floor is approached asymptotically,

$$\lim_{\rho \rightarrow \infty} \text{Var}\{\eta^Q\} = \mathcal{E}_{\text{se}} \frac{2^{-B}}{1 - 2^{-B}}.$$

Therefore the number of feedback bits must be increased with the SNR to avoid this noise floor. To quantify this increase in feedback, we calculate the value of B required so that the noise enhancement, ν , is kept constant. The number of feedback bits required can be obtained by rearranging (14),

$$B \approx \log_2 \left(1 + \frac{\rho + 1}{\nu - 1} \right) \approx \log_2 \rho - \log_2(\nu - 1),$$

where the approximations follow if ρ and B are sufficiently large. *Therefore, the number of feedback bits per channel element must be scaled linearly with $\log_2 \rho$ to maintain the same effective noise power.*

B. CSI Obtained via Limited Feedback with Channel Estimation Error

We now turn to the realistic case where perfect CSI is not available at the receiver. We assume that an estimate of each user's channel vector \mathbf{h}_k is calculated at the receiver and is obtained by the transmission of a training sequence of finite length. Specifically, we propose that each user calculates a minimum mean square error (MMSE) estimate of the channel using a sequence of L training symbols so that the distortion is minimized. The resulting channel estimate is then quantized using a codebook that can achieve the rate distortion lower bound and then sent back to the receiver. This results in an accumulation of error in the channel estimate channel at the transmitter due to training and quantization.

We denote the channel estimate obtained via training as \mathbf{H}^τ . It is obtained by transmitting a training sequence of L symbols simultaneously from the N_T transmit antennas to the K users. We denote this training sequence as $\mathbf{X}^\tau \in \mathbb{C}^{N_T \times L}$ and assume that $L \geq N_T$, and $\text{tr}\{\mathbf{X}^\tau (\mathbf{X}^\tau)^\dagger\} = \sqrt{\mathcal{E}_\tau} L$. Note that the training sequence power \mathcal{E}_τ may be chosen to be greater than the average power of the data transmission phase (*i.e.* \mathcal{E}_x), so as to improve the quality of the channel estimate. The resulting system equation for the training phase is therefore

$$\mathbf{Y}^\tau = \mathbf{H}\mathbf{X}^\tau + \mathbf{N}^\tau$$

where $\mathbf{Y}^\tau = [\mathbf{y}^{(1)} \dots \mathbf{y}^{(L)}]$, $\mathbf{N}^\tau = [\mathbf{n}^{(1)} \dots \mathbf{n}^{(L)}]$ and $(\cdot)^\ell$ denotes time instant ℓ .

We now present the relevant results concerning MIMO channel estimation [28]. Given the training sequence \mathbf{X}^τ , the MMSE channel estimate in the presence of additive white Gaussian noise is the linear MMSE receiver

given by $\mathbf{H}^\tau = \mathbf{Y}^\tau (\mathbf{X}^\tau)^\dagger \left(\sigma_n^2 \mathbf{I} + \mathcal{E}_\tau \mathbf{X}^\tau (\mathbf{X}^\tau)^\dagger \right)^{-1}$. The MMSE error is minimized according for a given average power constraint, \mathcal{E}_τ , by choosing $\mathbf{X}^\tau = \sqrt{\mathcal{E}_\tau L / N_T} \mathbf{Q}$ where $\mathbf{Q} \in \mathbb{C}^{N_T \times L}$ satisfies $\mathbf{Q} \mathbf{Q}^\dagger = \mathbf{I}_{N_T}$ [29, 30]. It follows that the elements of \mathbf{H}^τ are i.i.d. ZMCS complex Gaussian variables with variance

$$\text{Var} \{h^\tau\} = \frac{\rho_\tau L}{N_T + \rho_\tau L}.$$

where $\rho_\tau = \mathcal{E}_\tau / N_0$. Also, the channel estimation error $\mathbf{\Delta}^\tau = \mathbf{H} - \mathbf{H}^\tau$ has elements $\delta_{k,t}^\tau$. Due to the orthogonality principle of MMSE estimation, $\mathbf{\Delta}^\tau$ and \mathbf{H}^τ are uncorrelated, and therefore $\mathbf{\Delta}^\tau$ consists of i.i.d. ZMCS complex Gaussian variables with variance

$$\text{Var} \{\delta^\tau\} = \text{Var} \{h\} - \text{Var} \{h^\tau\} = \frac{N_T}{N_T + \rho_\tau L}.$$

The MMSE channel estimate is subsequently quantized using a codebook that can achieve the rate distortion lower bound as was used in the previous subsection. We denote the resulting quantized channel estimate obtained at the receiver as $\mathbf{H}^{\tau,Q}$. We denote the quantization error between the channel estimate obtained by training and the quantized channel estimate as $\mathbf{\Delta}^{\tau,Q} = \mathbf{H}^\tau - \mathbf{H}^{\tau,Q}$. The elements of $\mathbf{\Delta}^{\tau,Q}$ are i.i.d. ZMCS complex Gaussian random variables with variance

$$\text{Var} \{\delta^{\tau,Q}\} = 2^{-B} \text{Var} \{h^\tau\}.$$

Noting that the channel estimation and quantization error terms are additive, it follows that the effective noise term $\boldsymbol{\eta}^{\tau,Q}$ is

$$\boldsymbol{\eta}^{\tau,Q} \triangleq (\mathbf{\Delta}^{\tau,Q} + \mathbf{\Delta}^\tau) \mathbf{s} + \sqrt{\mathcal{E}_{\text{se}}^{\tau,Q}} \mathbf{n}. \quad (15)$$

We now consider the statistics of $\boldsymbol{\eta}^{\tau,Q}$. Again, noting that \mathcal{E}_{se} is inversely proportional to $\text{Var} \{h\}$, it follows that in this case, the average sphere-encoded signal power is increased to

$$\mathcal{E}_{\text{se}}^{\tau,Q} = \frac{N_T + \rho_\tau L}{\rho_\tau L (1 - 2^{-B})} \mathcal{E}_{\text{se}}.$$

which increases variance of the pure noise component. To examine the self-interference component $(\mathbf{\Delta}^{\tau,Q} + \mathbf{\Delta}^\tau) \mathbf{s}$, we first note that $\mathbf{\Delta}^{\tau,Q} + \mathbf{\Delta}^\tau$ consists of i.i.d. Gaussian entries of variance $\text{Var} \{\delta^{\tau,Q}\} + \text{Var} \{\delta^\tau\}$. The independence of the variance can be explained using the forward test channel interpretation of the rate distortion optimal quantization of a Gaussian source [26, pp. 478-9]. Hence the same assumptions used in the perfect receiver CSI case of the previous section can be applied in the same manner to obtain a Gaussian approximation to $(\mathbf{\Delta}^{\tau,Q} + \mathbf{\Delta}^\tau) \mathbf{s}$.

We now calculate the variance of $\eta^{\tau,Q}$,

$$\begin{aligned}
 \text{Var} \{ \eta^{\tau,Q} \} &= \mathcal{E}_{\text{se}}^{\tau,Q} \left(E \left[\left| \delta_{k,t}^{\tau,Q} \right|^2 \right] + E \left[\left| \delta_{k,t}^{\tau} \right|^2 \right] + \text{Var} \{ n \} \right) \\
 &= \frac{\mathcal{E}_{\text{se}}}{\text{Var} \{ h^{\tau} \} (1 - 2^{-B})} \left(\text{Var} \{ h^{\tau} \} 2^{-B} + \frac{N_{\text{T}}}{N_{\text{T}} + \rho_{\tau} L} + \frac{1}{\rho} \right) \\
 &= \frac{\mathcal{E}_{\text{se}}}{1 - 2^{-B}} \left(2^{-B} + \frac{N_{\text{T}}}{\rho_{\tau} L} + \frac{1}{\rho} \frac{N_{\text{T}} + \rho_{\tau} L}{\rho_{\tau} L} \right) \\
 &= \frac{\mathcal{E}_{\text{se}}}{1 - 2^{-B}} \left(2^{-B} + \frac{1}{\rho} + \frac{\rho + 1}{\rho} \frac{N_{\text{T}}}{\rho_{\tau} L} \right) \tag{16}
 \end{aligned}$$

We can calculate the noise enhancement as the ratio between the noise variance with training and limited feedback, and the noise variance with perfect CSI as follows,

$$v = \frac{1}{1 - 2^{-B}} \left(1 + \rho 2^{-B} + \frac{\rho + 1}{\rho_{\tau}} \frac{N_{\text{T}}}{L} \right). \tag{17}$$

Again, we note that if the extra noise term in (15) had a Gaussian distribution then v would exactly represent an SNR offset in performance due to imperfect CSI.

In Figure 6 we plot v for a limited feedback vector perturbation system with imperfect channel estimation. The same scenario as was used in Figure 5 is used, with $L = 2N_{\text{T}}$ training symbols and $\mathcal{E}_{\tau} = \mathcal{E}_x$. In the figure, we compare our analytical curves to the two practical vector quantization schemes, namely the generalized Lloyd-Max algorithm and SLAQ. We again see that the curves for the practical codes are less than 1dB from the optimal for the case of $\rho = 10\text{dB}$, and 2dB for the case of $\rho = 20\text{dB}$. Note that in this case, v does not approach zero as the number of feedback bits increases, due to the channel estimation error.

We consider the implications of (17) in the case of perfect feedback, *i.e.* as $B \rightarrow \infty$. In this case, the resulting v in the high SNR regime is

$$\lim_{B \rightarrow \infty} v = 1 + \frac{\rho + 1}{\rho_{\tau}} \frac{N_{\text{T}}}{L}.$$

This implies that to maintain a certain v as the SNR fluctuates the total energy in the training symbol $\mathcal{E}_{\tau} L T$ (where T is the symbol time) must be chosen so that the ratio $(\rho + 1)/\rho_{\tau}$ is kept constant. At high SNR, $v \rightarrow 1 + \mathcal{E}_x N_{\text{T}} / \mathcal{E}_{\tau} L$. Therefore, except for small values of SNR, a given v can be maintained by keeping the ratio of the energy dedicated to training and data fixed.

We now briefly consider the relationship between the allocation of bits to limited feedback, and the energy expended by training. To do this we simply state that (17) implies that for sufficiently large B and SNR, that

$$v \approx \left(1 + \rho 2^{-B} + \frac{\mathcal{E}_x N_{\text{T}}}{\mathcal{E}_{\tau} L} \right)$$

Hence, the effect of quantization and channel estimation error becomes simply additive. The allocation of system resources to training and feedback can be balanced according to this relationship to maintain a small value of ν .

V. NEW LOW-COMPLEXITY ALGORITHM FOR CHANNEL VECTOR QUANTIZATION

In this section we propose a new practical quantization scheme for quantization of each user's channel vector \mathbf{h}_k , for low-rate feedback. We propose to quantize the gain and direction of \mathbf{h}_k separately, where the direction of the vector is defined as $\mathbf{u}_k = \mathbf{h}_k / \|\mathbf{h}_k\|$.

For the quantization of the channel gain, we propose to quantize from a codebook of N_G gain levels by choosing the level closest to the channel gain $\|\mathbf{h}_k\|$ in terms of Euclidean distance. We propose to generate the gain codebook using the Lloyd-Max algorithm [27]. This algorithm provides a set of quantization levels given a p.d.f., that are locally-optimal when using the Euclidean distance for quantization. The p.d.f. of $\|\mathbf{h}_k\|$ is available in [31].

For the quantization of the direction, we propose a quantization scheme which we call square-lattice angular quantization (SLAQ). This scheme consists of a direction vector codebook based on a square lattice and a low-complexity search algorithm that finds the closest direction vector in *angle* to the actual direction vector.

The SLAQ codebook is formed by first generating a square integer lattice in \mathbb{R}^{2N_T} . The dimensions of this lattice allow for the real and imaginary components of the N_T -dimensional direction vector to be quantized separately. We propose to allow only lattice points consisting of odd integers in the range $[-M, M]$. This results in M quantization levels for each of the $2N_T$ dimensions. Each direction vector in the codebook corresponds to a lattice point. Due to the structure in this codebook, it is not necessary to store the direction vectors explicitly. This results in significant memory savings compared to other codebooks. It also leads to computational savings, as we are able to exploit the structure when performing the quantization of \mathbf{h}_k , as we describe in the following paragraph.

The quantization of the direction vector proceeds as follows. We propose to convert the complex-valued direction vector $\mathbf{h}_k / \|\mathbf{h}_k\| \in \mathbb{C}^{N_T}$, into \mathbb{R}^{2N_T} by concatenating the real and imaginary components into a single vector, and denote this vector as $\underline{\mathbf{u}}$. It is then quantized to the nearest $2N_T$ -dimensional vector from our SLAQ codebook according to the chordal distance (*i.e.* angular) criterion:

$$\underline{\mathbf{u}}^Q = \underset{\underline{\mathbf{v}}}{\operatorname{argmin}} d(\underline{\mathbf{u}}, \underline{\mathbf{v}}) = 1 - \frac{(\underline{\mathbf{u}}' \underline{\mathbf{v}})^2}{\|\underline{\mathbf{u}}\|^2 \|\underline{\mathbf{v}}\|^2} \quad (18)$$

with the restriction that the elements of $\underline{\mathbf{u}}$ and $\underline{\mathbf{u}}^Q$ each have the same sign. Note that, performing this minimization as an exhaustive search is generally impractical, as the size of the codebook may be very large. For example,

a system operating at $\rho = \rho_\tau = 20\text{dB}$ and $L = 2N_T$ training symbols requires $B = 7$ to be operating with an effective noise variance within 4dB of the perfect CSI case. This results in a codebook size of $2^{7N_T} \approx 10^{2N_T}$ vectors. Therefore, a faster approach is required, and it is here that the structure of the SLAQ codebook can be exploited.

To elaborate, first consider the line in the direction of \mathbf{u} , passing through the origin. Now, consider a test point situated at the origin, that travels along this line towards $\underline{\mathbf{u}}$. As the test point travels along the line, we maintain a note of the lattice point which is closest to the test point in terms of Euclidean distance, and keep a record of all such lattice points. The recording of lattice points terminates when the test point exits one of the $[-M, M]$ boundaries. The minimization in (18) is then performed over only this set of recorded lattice points. In previous related work, we have shown that the direction vector closest in angle is always found with complexity $O(N_T \log N_T)$ by using this approach, see [32, Section IV.A] where we have also provided pseudo-code for the algorithm. The chosen vector is then scaled to have unit norm. Note that a related technique for quantizing beamforming vectors in single-user systems was introduced in [9].

We now consider the performance of this scheme compared to the ideal rate-distortion based codebooks, and with other vector perturbation precoding systems. In Fig. 7 we plot the BER as a function of SNR for a multi-user MIMO system with various precoding methods, namely vector perturbation, and the linear precoding schemes of zero-forcing beamforming [10] and regularized channel inversion [20]. For regularized channel inversion, the precoding matrix is given by $\mathbf{H}^\dagger(\mathbf{H}\mathbf{H}^\dagger + K/(5\rho)\mathbf{I})^{-1}$. The regularization parameter, $K/(5\rho)$ is chosen according to the discussion in [20], as the optimal choice of this value remains an open problem. We compare the systems with or without training (where $L = 4N_T$ and $\mathcal{E}_\tau = \mathcal{E}_x$), and with or without limited feedback. For limited feedback, we consider the SLAQ codebook using 6-ary PAM constellations to quantize the (real-valued, $2N_T$ -dimensional) direction vector, and 8 levels for the gain quantization. We also consider an ideal codebook of the same size.

For the vector perturbation technique, we note that the performance of the practical SLAQ codebook is close to that of an ideal rate-distortion codebook, despite having much lower complexity. The plots also indicate that a relatively large number of feedback bits are required to obtain a BER close to that of an ideal perfect feedback system.

We see that for the vector perturbation techniques at high SNR, training incurs a 2dB loss in terms of BER in this scenario. We see that for the limited feedback scenarios, error floors appear.

Finally, we see from the figure that vector perturbation significantly outperforms the linear precoders. This

is partially due to the fact that that the transmitted energy is allowed to vary between packets, whereas for zero-forcing beamforming this is not possible, as this would require channel inversion, which has $\mathcal{E}_{\text{se}} = \infty$.

VI. DISCUSSION AND CONCLUSION

We have derived a closed-form approximation to the average sphere-encoded signal power \mathcal{E}_{se} for vector perturbation systems, which determines the SNR at the output of the demodulator. This was obtained by noting that \mathcal{E}_{se} is dependent on the second moment of the Voronoi region of the lattice which has generator matrix \mathbf{H}^+ , and not the constellation. We considered the effect that training and limited feedback have on the system, and showed that the noise power gain is approximately proportional to $(1 + \rho 2^{-B})$, indicating that the number of feedback bits must increase linearly with the transmit SNR to maintain the same effective noise power. We also showed that the noise power gain due to training is to the first-order at least, independent of the transmit SNR. Finally, we proposed a new practical low-complexity low-rate feedback scheme that performs close to the ideal rate-distortion based scheme.

APPENDIX A

PROOF OF THEOREM 1

Proof: The proof proceeds by finding upper and lower bounds on \mathcal{E}_{se} , and then showing that the bounds converge as $K \rightarrow \infty$. We make use of the following identity of Gautschi [33],

$$e^{(x-1)\Psi(n+1)} < \frac{\Gamma(n+x)}{\Gamma(n+1)} < n^{x-1}, \quad n \in \mathbb{N}^+, 0 < x < 1 \quad (19)$$

where $\Psi(z)$ is the digamma function defined as $\Psi(z) \triangleq \Gamma'(z)/\Gamma(z)$ and \mathbb{N}^+ denotes the set of positive integers.

We obtain the upper bound on $\mathcal{E}_{\text{se,LB}}$ in (10) by substituting $n = N_T - \ell + 1$ and $x - 1 = -1/K$ into (19) to get

$$\mathcal{E}_{\text{se,LB}} < 2KG(S_{2K}) \frac{\Gamma(N_T - K + 1 - \frac{1}{K})}{\Gamma(N_T - K + 1)} \left[\prod_{\ell=0}^{K-2} (N_T - \ell - 1) \right]^{-\frac{1}{K}} \quad (20)$$

where the case of $\ell = K - 1$ in the summation is treated separately so that the bound in (19) is valid for the case when $N_T = K$. We now take limits as $K \rightarrow \infty$ of the different components of (20). First, we have that as $K \rightarrow \infty$, $2KG(S_{2K}) = \frac{K}{\pi e}$ since $\lim_{L \rightarrow \infty} G(S_L) = 1/(2\pi e)$ [21, pg. 452]. Also, Stirling's formula,

$\Gamma(z) = e^{-z} z^{z-\frac{1}{2}} \frac{1}{\sqrt{2\pi}} (1 + O(z^{-1}))$ can be used to show that

$$\lim_{K \rightarrow \infty} \frac{\Gamma(N_T - K + 1 - \frac{1}{K})}{\Gamma(N_T - K + 1)} = 1$$

We now turn to the rightmost term of (20) to get

$$\prod_{\ell=0}^{K-2} (N_T - \ell - 1) = \frac{(N_T - 1)!}{(N_T - K)!} = \frac{\Gamma(N_T)}{\Gamma(N_T - K + 1)}$$

and using Stirling's formula again we have

$$\begin{aligned} \lim_{K \rightarrow \infty} \left[\frac{\Gamma(N_T)}{\Gamma(N_T - K + 1)} \right]^{-\frac{1}{K}} &= \lim_{K \rightarrow \infty} \frac{e^{-\frac{N_T - K - 1}{K}} N_T^{\frac{N_T}{K} - \frac{1}{2K}}}{e^{-\frac{N_T}{K}} (N_T - K + 1)^{\frac{N_T - K}{K} - \frac{1}{2K}}} \\ &= \frac{e}{K} \alpha (1 - \alpha)^{\frac{1-\alpha}{\alpha}}. \end{aligned} \quad (21)$$

By combining the limiting terms we obtain

$$\lim_{K \rightarrow \infty} \mathcal{E}_{\text{se, LB}} \leq \frac{\alpha}{\pi} (1 - \alpha)^{\frac{1-\alpha}{\alpha}}. \quad (22)$$

We now turn to deriving an asymptotic lower bound on $\mathcal{E}_{\text{se, LB}}$. Applying the identity in (19) and using the same variable substitution as in (20) we obtain

$$\mathcal{E}_{\text{se, LB}} > 2KG(S_{2K}) \frac{\Gamma(N_T - K + 1 - \frac{1}{K})}{\Gamma(N_T - K + 1)} \left[\prod_{\ell=0}^{K-2} e^{-\psi(N_T - \ell)} \right]^{\frac{1}{K}} \quad (23)$$

Now, for integer n we have the relationship [34, Eq. 6.3.2]

$$\Psi(n) = -\gamma + \sum_{p=1}^{n-1} \frac{1}{p} = -\gamma + H_{n-1}$$

where γ is the Euler-Mascheroni constant and H_n is an harmonic number. Now, from the identity [35, pg. 76]:

$$\frac{1}{24n^2} < H_{n-1} - \log\left(n - \frac{1}{2}\right) - \gamma < \frac{1}{24(n-1)^2} \quad (24)$$

it follows that $\Psi(n) \leq \log\left(n - \frac{1}{2}\right) + \frac{1}{24(n-1)^2} \leq \log(n) + \frac{1}{24(n-1)^2}$. Applying this to (23) we have

$$\left[\prod_{\ell=0}^{K-2} e^{-\psi(N_T - \ell)} \right]^{\frac{1}{K}} \geq \left[\prod_{\ell=0}^{K-2} \frac{e^{-\frac{1}{24(N_T - \ell - 1)^2}}}{(N_T - \ell)} \right]^{\frac{1}{K}} = e^{-\frac{1}{24K(N_T - 1)^2}} \left[\frac{\Gamma(N_T - K)}{\Gamma(N_T)} \right]^{\frac{1}{K}}$$

The limit as $K \rightarrow \infty$ can be performed similarly as for the case of upper bound on $\mathcal{E}_{\text{se, LB}}$ in (21). Substituting Stirling's series into this equation, and taking limits as $K \rightarrow \infty$ provides us with the asymptotic lower bound

$$\lim_{K \rightarrow \infty} \mathcal{E}_{\text{se, LB}} \geq \frac{\alpha}{\pi} (1 - \alpha)^{\frac{1-\alpha}{\alpha}}. \quad (25)$$

Noting that (22) and (25) are equal completes the proof. ■

APPENDIX B

PROOF OF LEMMA 2

Proof: Without loss of generality, we consider just the real component of the received signal at a single user. Denote $a = \text{Re}\{a_1\}$, $\hat{a} = \text{Re}\{\hat{a}_1\}$ and $\eta = \text{Re}\{\eta_1\}$. The resulting probability of bit error is obtained by summing the contribution of the probability of a bit error event over the repeated PAM constellation of infinite extent. We obtain

$$P_b = \sum_{q=-\infty}^{\infty} \sum_{\ell=1}^{\mu-1} d(q\mu + \ell) \Pr\left(q + \frac{2\ell-1}{2\mu} < \eta < q + \frac{2\ell+1}{2\mu}\right) \quad (26)$$

where $d(k)$ is the average Hamming distance between the m th and $(m+k)_{\text{mod } \mu}$ th symbol in the Gray-coded μ -ary PAM constellation, *i.e.*

$$d(k) = \frac{1}{\mu} \sum_{m=0}^{M-1} d(m, (m+k)_{\text{mod } \mu}).$$

Since η is a zero-mean Gaussian random variable with variance $\mathcal{E}_{\text{sc}}/(2\rho)$, we can define

$$G_{q,\ell} \triangleq \frac{1}{\log_2 \mu} \Pr\left(q + \frac{2\ell-1}{2\mu} < \eta < q + \frac{2\ell+1}{2\mu}\right) = \frac{1}{2\log_2 \mu} [\text{erf}((2\mu q + 2\ell + 1)\varphi) - \text{erf}((2\mu q + 2\ell - 1)\varphi)]$$

where $\text{erf}(\cdot)$ denotes the error function, and the second equality results from recalling from the theorem that we defined $\varphi = \frac{1}{2\mu} \sqrt{\frac{\rho}{\mathcal{E}_{\text{sc}}}}$. We now consider the lower bound on P_b ,

$$\begin{aligned} P_b &= \sum_{q=-\infty}^{\infty} \sum_{\ell=1}^{\mu-1} d(q\mu + \ell) G_{q,\ell} \\ &\geq \sum_{\ell=1}^{\mu-1} G_{0,\ell} + G_{-1,\ell} = 2G_{0,\ell} \\ &= \frac{2}{\log_2 \mu} \sum_{\ell=1}^{\mu-1} [\text{erf}((2\ell+1)\varphi) - \text{erf}((2\ell-1)\varphi)] \\ &= \frac{1}{\log_2 \mu} [\text{erf}((2\mu-1)\varphi) - \text{erf}(\varphi)] = \frac{2}{\log_2 \mu} [Q(\sqrt{2}\varphi) - Q(\sqrt{2}(2\mu-1)\varphi)] \end{aligned}$$

where the first inequality follows from the fact that $d(\cdot) \geq 1$ for all the error events, and by considering only $q = -1, 0$ in the summation. The final equality follows from the fact that $\text{erf}(x) = 1 - 2Q(\sqrt{2}x)$. We now consider the upper bound.

$$\begin{aligned} P_b &= G_{0,1} + \sum_{\ell=2}^{\mu-1} d(\ell) G_{0,\ell} + \sum_{q=1}^{\infty} \sum_{\ell=1}^{\mu-1} d(q\mu + \ell) G_{q,\ell} + G_{-1,\mu-1} + \sum_{\ell=1}^{\mu-2} d(\ell) G_{-1,\ell} + \sum_{q=-\infty}^{-1} \sum_{\ell=1}^{\mu-1} d(q\mu + \ell) G_{q,\ell} \\ &\leq 2G_{0,1} + 2\log_2 \mu \sum_{\ell=2}^{\mu-1} G_{0,\ell} + 2\log_2 \mu \sum_{q=1}^{\infty} \sum_{\ell=0}^{\mu-1} G_{q,\ell} \\ &= \frac{1}{\log_2 \mu} [\text{erf}(3\varphi) - \text{erf} \varphi] + 1 - \text{erf}(3\varphi) = \frac{2}{\log_2 \mu} [Q(\sqrt{2}\varphi) + (\log_2 \mu - 1)Q(\sqrt{2}\varphi)] \end{aligned}$$

where the inequality follows from the fact that adjacent symbols have a Hamming distance of 1, and that all other errors have a Hamming distance of at most $\log_2 \mu$. The third result of the theorem follows by taking the limit of the upper and lower bounds as $\varphi \rightarrow \infty$. ■

REFERENCES

- [1] S. Vishnawath, N. Jindal, and A. Goldsmith, "Duality, achievable rates and sum capacity of Gaussian MIMO channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.
- [2] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.
- [3] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication - Part II: Perturbation," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 537–544, Mar. 2005.
- [4] U. Fincke and M. Pohst, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Mathematics of Computation*, vol. 44, no. 170, pp. 463–471, Apr. 1985.
- [5] R. Müller, D. Guo, and A. Moustakas, "Vector precoding for wireless MIMO systems and its replica analysis," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 530–540, Apr. 2008.
- [6] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Quantized vs. analog feedback for the MIMO broadcast channel: A comparison between zero-forcing based achievable rates," in *Proc. Intern. Symposium on Inform. Theory (ISIT)*, Nice, France, Jun. 2007.
- [7] D. J. Love and R. W. Heath Jr., "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2735–2745, Oct. 2003.
- [8] D. J. Love and R. W. Heath Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
- [9] D. J. Ryan, I. V. L. Clarkson, I. B. Collings, D. Guo, and M. L. Honig, "QAM codebooks for low-complexity limited feedback MIMO beamforming," in *Proc. IEEE Intern. Conf. on Commun. (ICC)*, Glasgow, Scotland, Jun. 2007, pp. 4162–4167.
- [10] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.
- [11] W. Santipach and M. L. Honig, "Signature optimization for CDMA with limited feedback," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3475–3492, Oct. 2005.
- [12] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.
- [13] C. Swannack, E. Uysal-Biyikoglu, and G. G. Wornell, "MIMO broadcast scheduling with limited channel state information," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT)*, Monticello, IL, USA, Oct. 2006.
- [14] K. Huang, R. W. Heath Jr., and J. G. Andrews, "Multiuser limited feedback for wireless multi-antenna communication," in *Proc. Intern. Symposium on Inform. Theory (ISIT)*, Nice, France, Jun. 2007.
- [15] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3417–3428, Jul. 2007.
- [16] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. Boston: Kluwer Academic Publishers, 1992.

- [17] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Trans. Commun.*, vol. 52, pp. 2057–2060, Dec. 2004.
- [18] M. Airy, S. Bhadra, R. W. Heath Jr., and S. Shakkottai, "Transmit precoding for the multiple antenna broadcast channel," in *Proc. of the IEEE Veh. Tech. Conference*, vol. 3, Melbourne, Australia, May 2006, pp. 1396–1400.
- [19] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Communication over mimo broadcast channels using lattice-basis reduction," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4567–4582, Dec. 2007.
- [20] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication - Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.
- [21] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*, 2nd ed. Springer-Verlag, 1993.
- [22] I. E. Teletar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs, Tech. Rep., 1995.
- [23] P. L. Zador, "Development and evaluation of procedures for quantizing multivariate distributions," Ph.D. dissertation, Stanford Univ., 1963.
- [24] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.
- [25] D. J. Love, R. W. Heath Jr., W. Santipach, and M. L. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Comms. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.
- [26] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
- [27] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inform. Theory*, vol. 28, pp. 129–137, Mar. 1982.
- [28] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [29] S. D. Silverstein, "Application of orthogonal codes to the calibration of active phased array antennas for communication satellites," *IEEE Trans. Signal Process.*, pp. 206–218, Jan. 1997.
- [30] T. L. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Communications, Control, and Computing*, Sep. 1999, pp. 958–966. [Online]. Available: <http://cm.bell-labs.com/who/hochwald/papers/training/>
- [31] J. G. Proakis, *Digital Communications*. McGraw-Hill, 2000.
- [32] D. J. Ryan, I. B. Collings, and I. V. L. Clarkson, "GLRT-optimal noncoherent lattice decoding," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3773–3786, Jul. 2007.
- [33] W. Gautschi, "Some elementary inequalities relating to the gamma and incomplete gamma function," *J. Math. Phys.*, pp. 77–81, 1959.
- [34] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions, 5th printing*. New York: Dover, 1968.
- [35] J. Havil, *Gamma: Exploring Euler's Constant*. Princeton University Press, 2003.

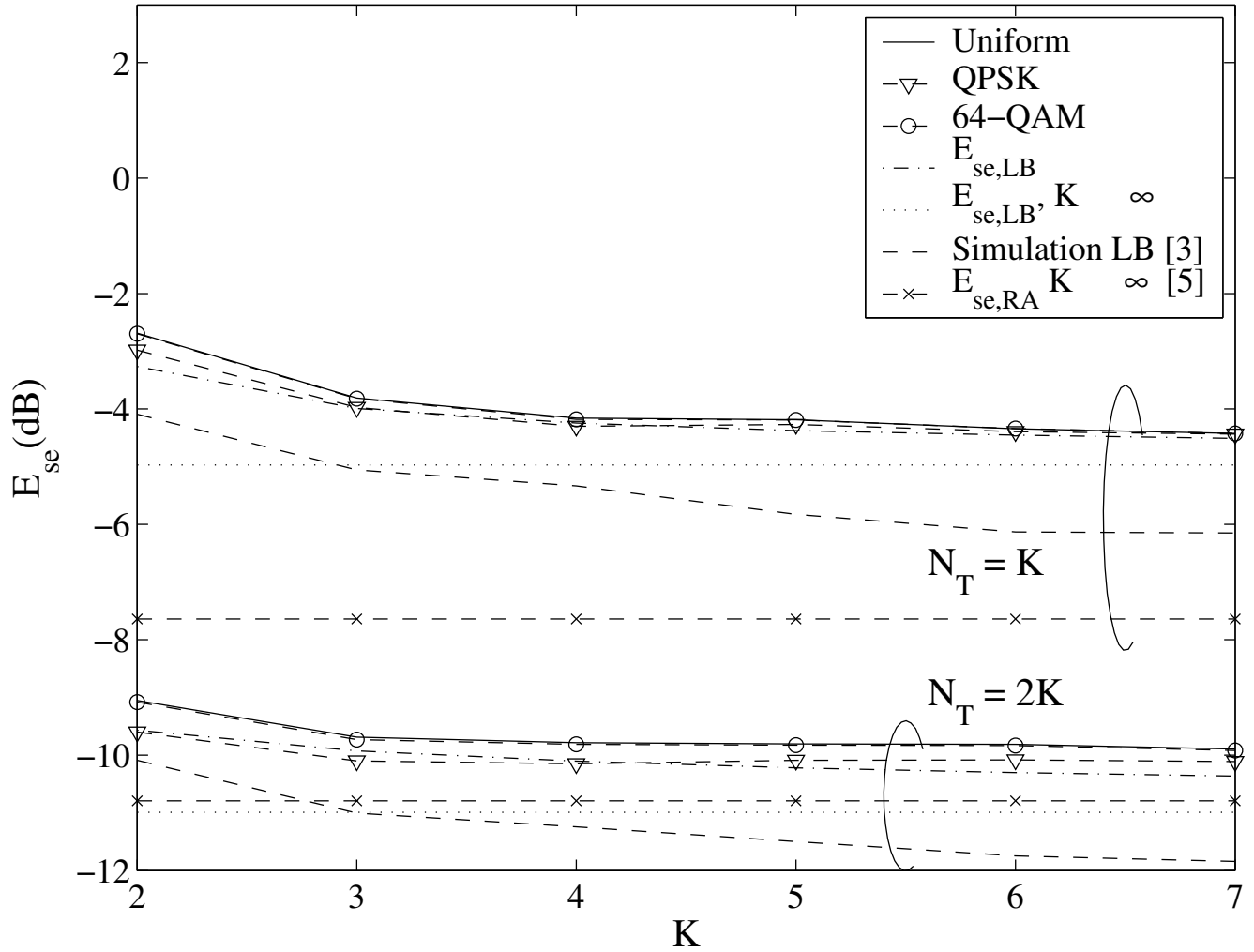


Fig. 1. Plot of \mathcal{E}_{se} , the average power of the sphere-encoded vector, as a function of the number of users K . Plots are shown for $N_T = K$ and $2K$ transmit antennas for uniformly distributed inputs as well as some square QAM constellations. Also plotted is the uniformly distributed inputs lower bound of Corollary 1.

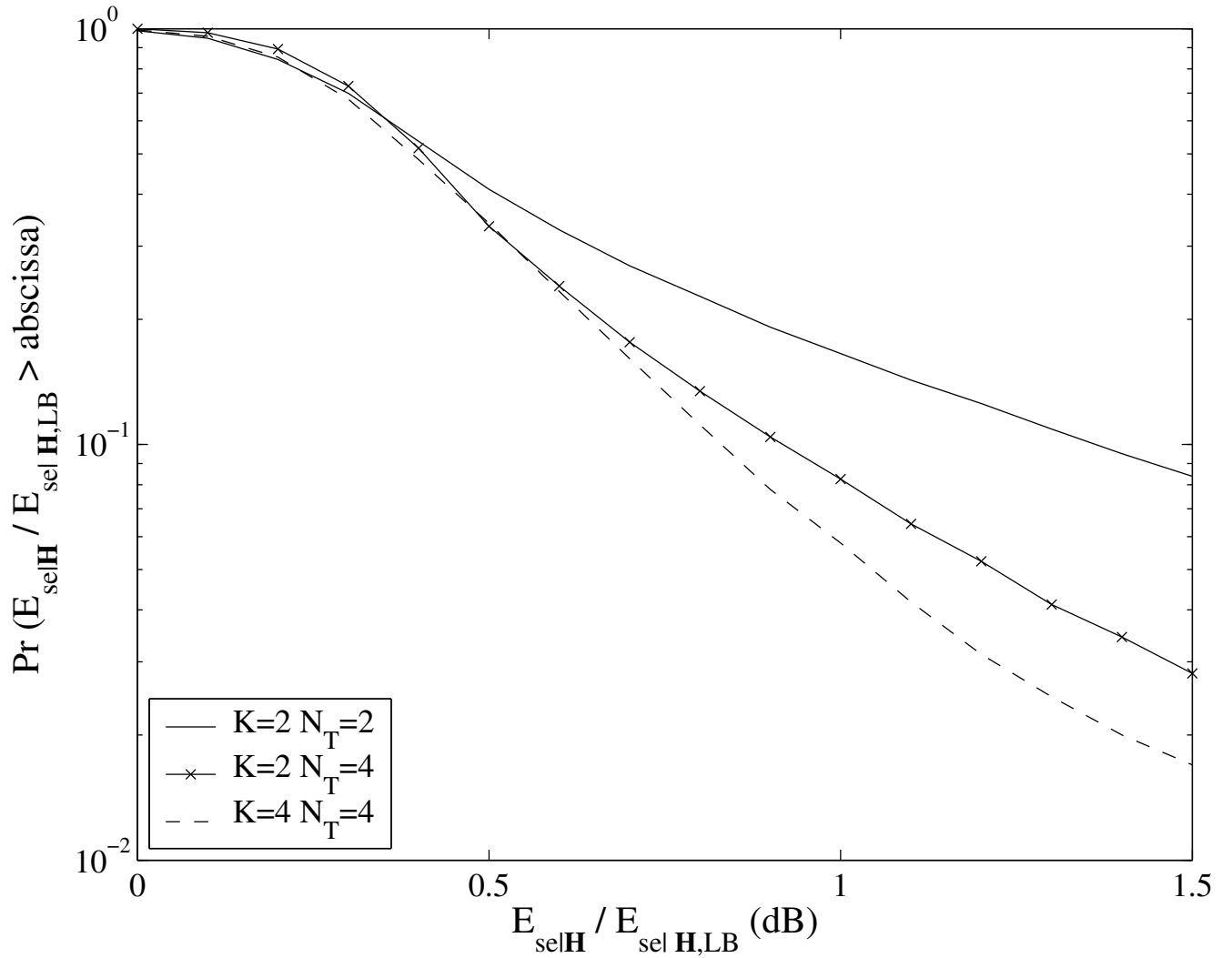


Fig. 2. Plot of the complementary c.d.f. of the ratio of $\mathcal{E}_{\text{selH,LB}}/\mathcal{E}_{\text{selH}}$ for uniformly distributed inputs, where $\mathcal{E}_{\text{selH,LB}}$ is obtained from Lemma 1. Plots are shown for various $N_T \times K$ multi-user systems, specifically 2×2 , 4×2 and 4×4 systems.

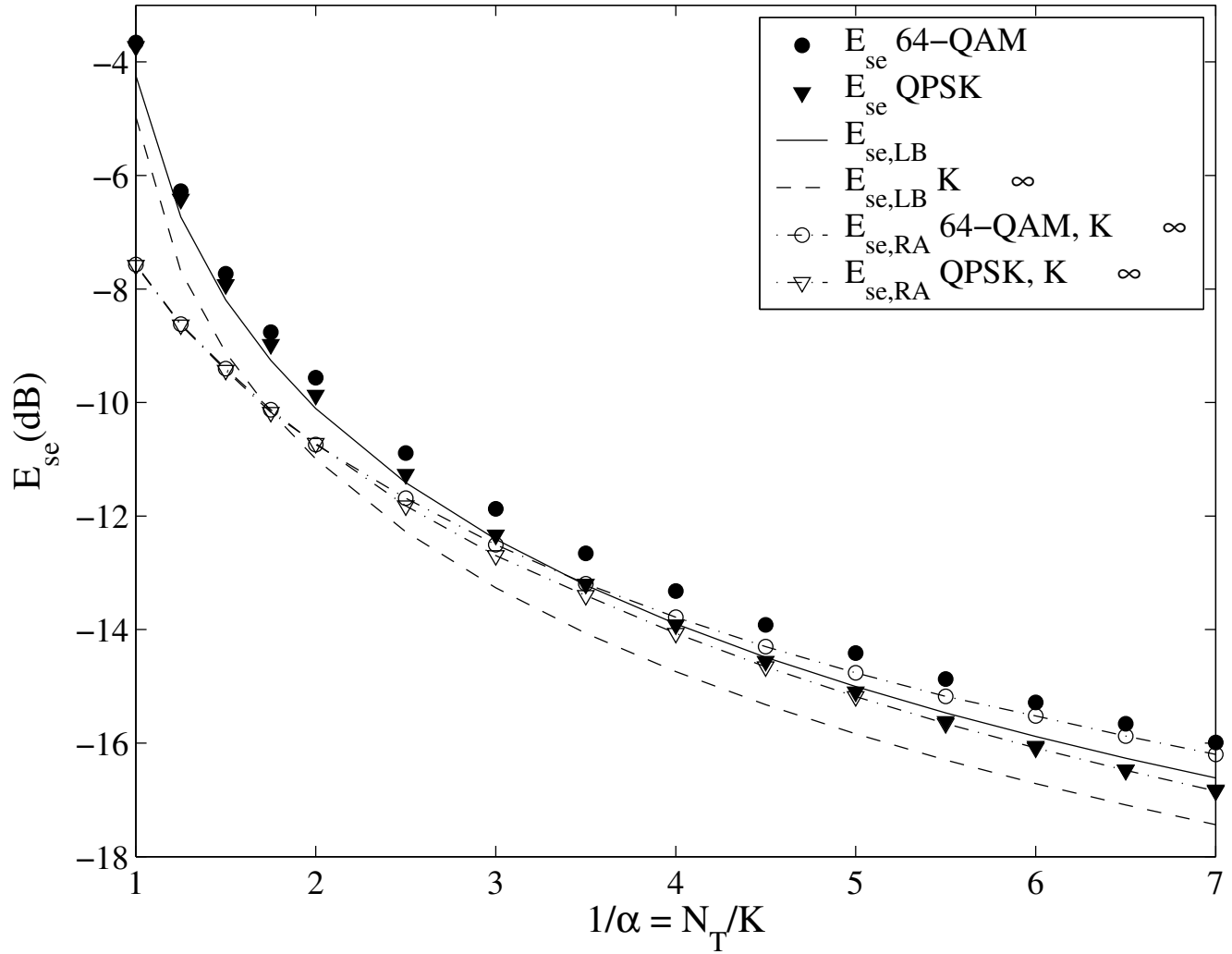


Fig. 3. Plot of the asymptotic lower bound of sphere-encoded signal power \mathcal{E}_{se} in dB derived in Theorem 1 as a function α , which is the ratio of the number of users K to the number of transmit antennas N_T .

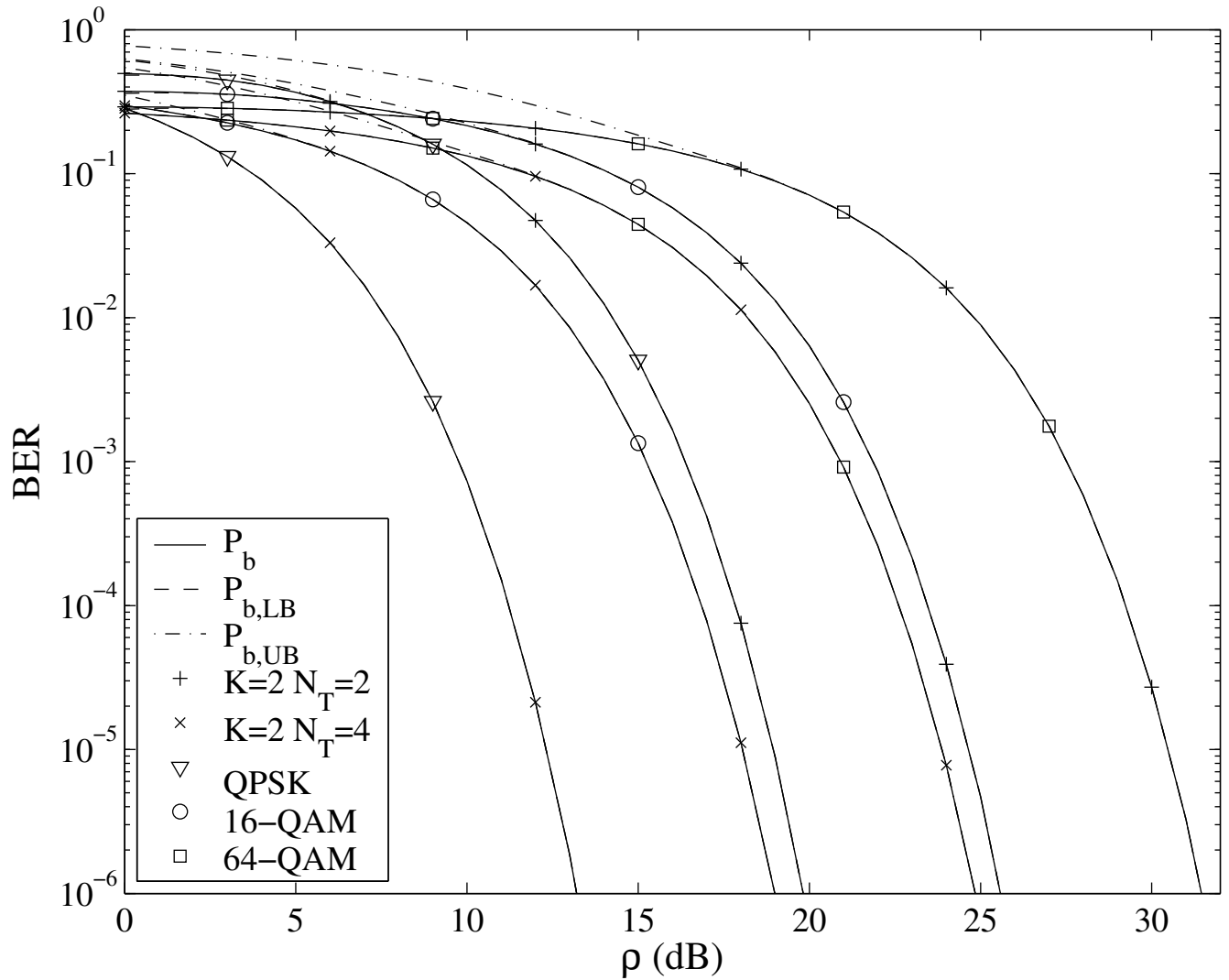


Fig. 4. Plot of the BER as a function of SNR for various vector perturbation systems with perfect channel state information at the transmitter. We consider $K = 2$ users with either $N_T = 2$ or 4 transmit antennas. We use QPSK, 16-QAM and 64-QAM constellations. We also plot the lower and upper bounds of Lemma 2. The lower bound is very tight and not visible.

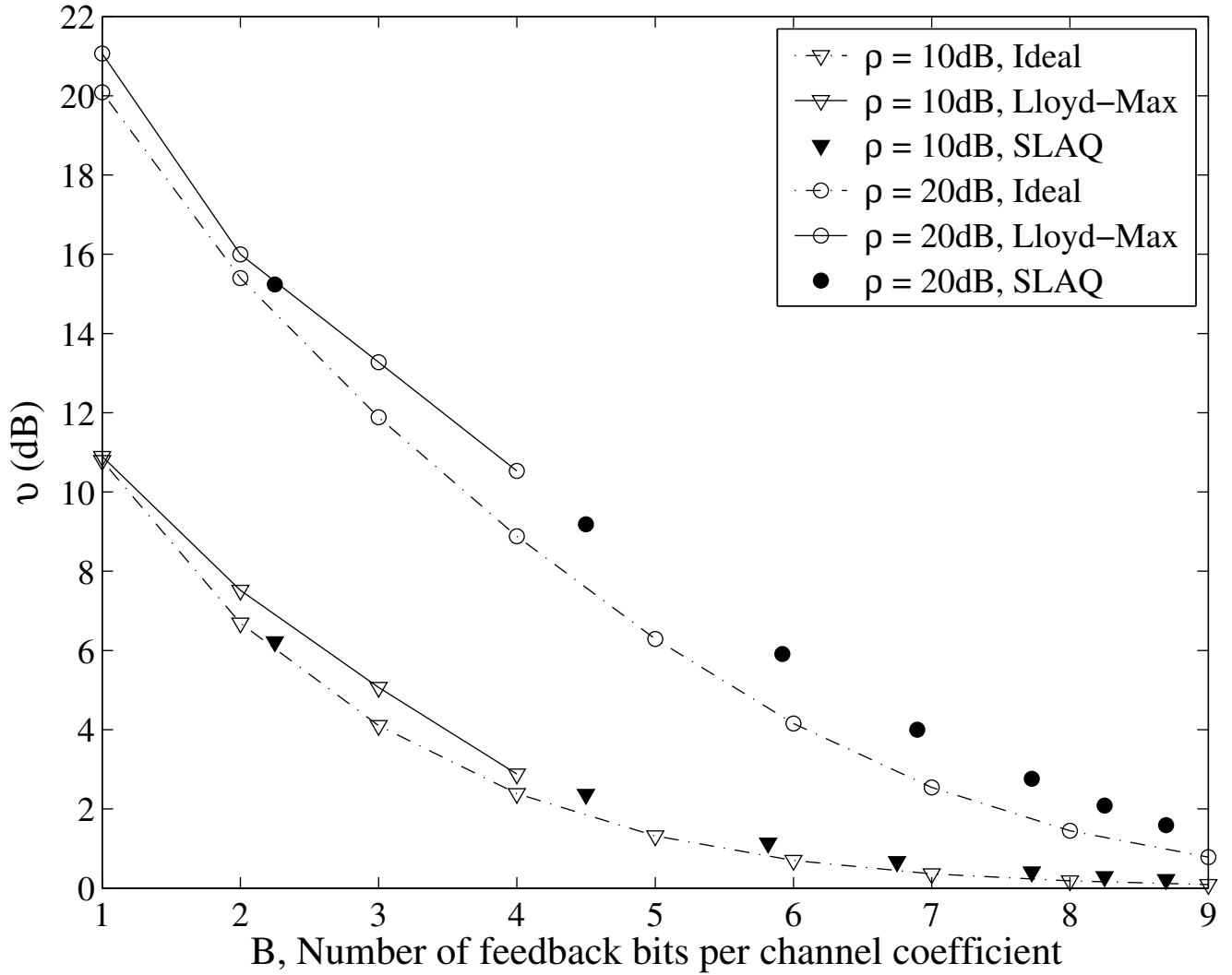


Fig. 5. Plot of the SNR loss as a function of the number of feedback bits per channel element for a vector perturbation system with perfect receiver CSI. The system has $N_T = 4$ transmit antennas, and we consider SNRs of $\rho = 10$ and 20dB. Plots are shown for ideal codebooks, along with Lloyd-Max codebooks and our proposed low-complexity SLAQ scheme.

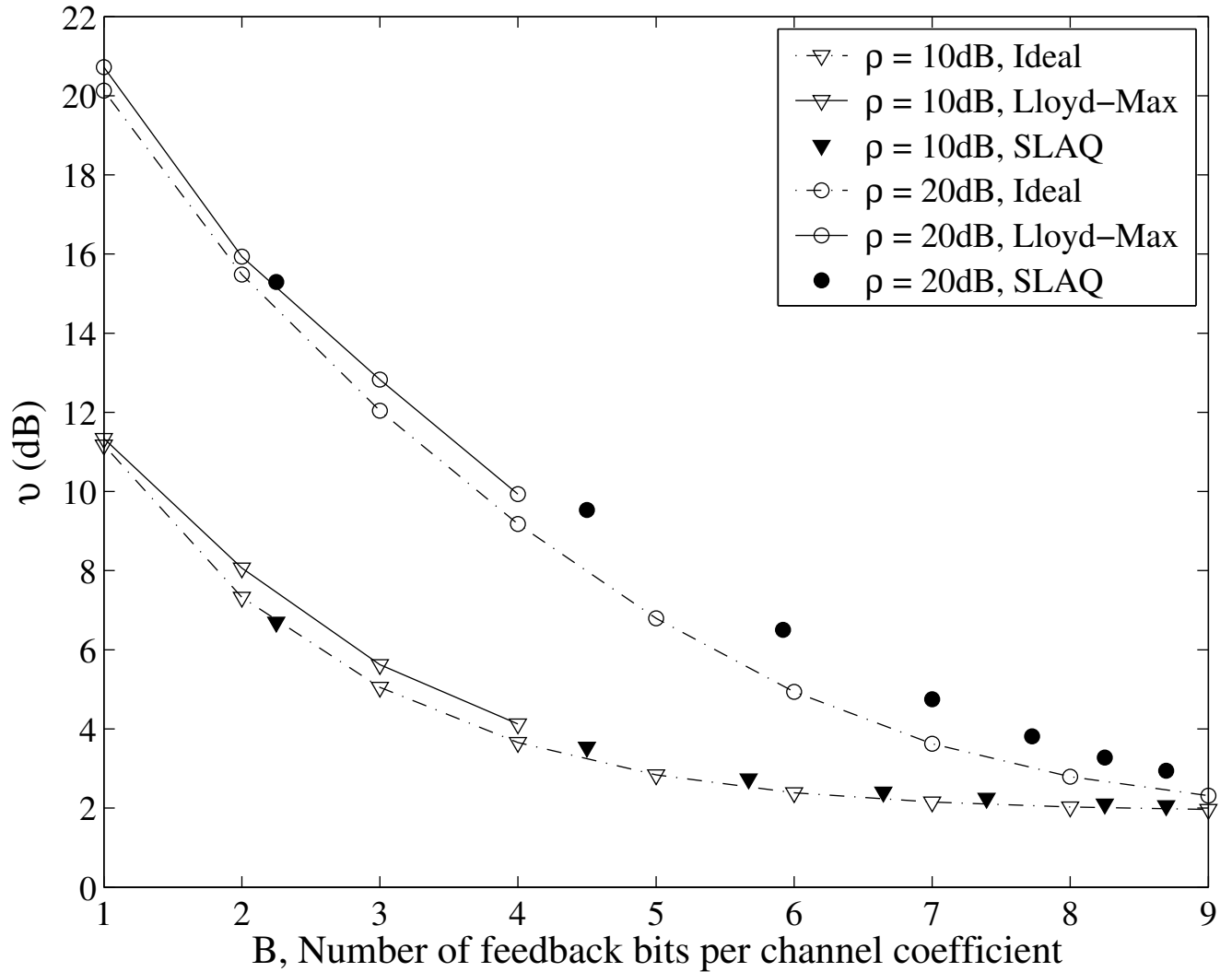


Fig. 6. Plot of the SNR loss as a function of the number of feedback bits per channel element for a vector perturbation system with imperfect receiver CSI. The system has $N_T = 4$ transmit antennas, and we consider SNRs of $\rho = 10$ and 20dB. The training stage is $L = 2N_T = 8$ symbols long. Plots are shown for ideal codebooks, along with Lloyd-Max codebooks and our proposed low-complexity SLAQ scheme.

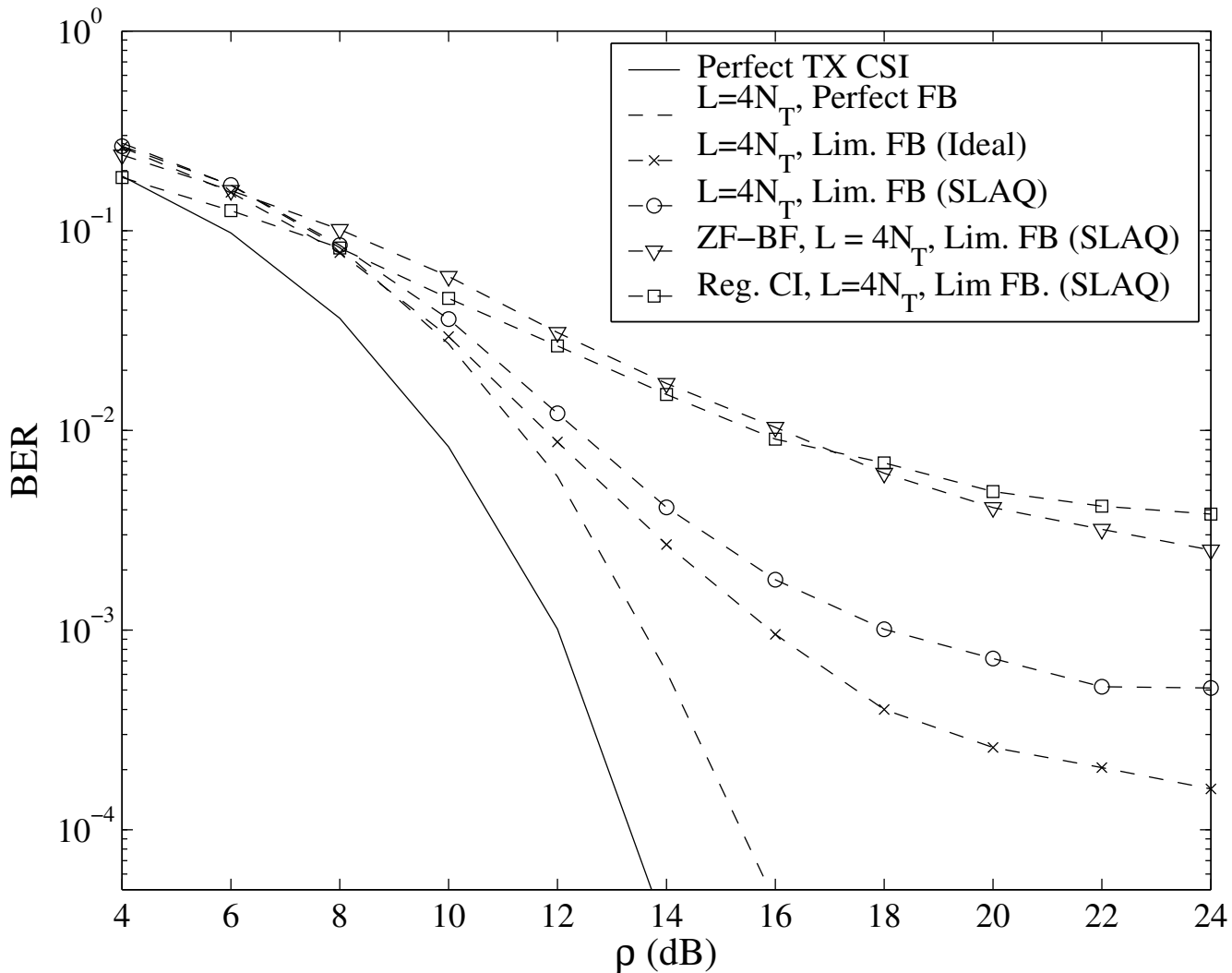


Fig. 7. Plot of the BER as a function of SNR for vector perturbation and some linear precoders systems with perfect and imperfect channel state information at the transmitter. We consider $K = 2$ users with $N_T = 3$ transmit antennas using QPSK constellations. We consider perfect transmitter CSI, training using $L = 4$ symbols and perfect feedback, and also training using $L = 2$ symbols with limited feedback. For the limited feedback case 6.17 bits is used per channel coefficient using either an ideal rate-distortion codebook or a SLAQ codebook. We also consider zero-forcing beamforming and regularized channel inversion.