All-Digital LoS MIMO with Low-Precision Analog-to-Digital Conversion

Ahmet Dundar Sezer, *Member, IEEE,* and Upamanyu Madhow, *Fellow, IEEE*

***Abstract*—Line-of-sight (LoS) multi-input multi-output (MIMO) systems exhibit attractive scaling properties with increase in carrier frequency: for a fixed form factor and range, the spatial degrees of freedom increase quadratically for 2D arrays, in addition to the typically linear increase in available bandwidth. In this paper, we investigate whether modern all-digital baseband signal processing architectures can be devised for such regimes, given the difficulty of analog-to-digital conversion for large bandwidths. We propose low-precision quantizer designs and accompanying spatial demultiplexing algorithms, considering 2×2 LoS MIMO with QPSK for analytical insight, and 4×4 MIMO with QPSK and 16QAM for performance evaluation. Unlike prior work, channel state information is utilized only at the receiver (i.e., transmit precoding is not employed). We investigate quantizers with regular structure whose high-SNR mutual information approaches that of an unquantized system. We prove that amplitude-phase quantization is necessary to attain this benchmark; phase-only quantization falls short. We show that quantizers based on maximizing per-antenna output entropy perform better than standard Minimum Mean Squared Quantization Error (MMSQE) quantization. For spatial demultiplexing with severely quantized observations, we introduce the novel concept of virtual quantization which, combined with linear detection, provides reliable demodulation at significantly reduced complexity compared to maximum likelihood detection.**

***Index Terms*—Millimeter wave, THz, LoS, MIMO, ADC, quantization, quantizer design, spatial demultiplexing.**

I. Introduction

The immense potential of the millimeter wave (mmWave) and terahertz (THz) bands for next generation wireless communications is well recognized in the research community [1]–[3]. There are vast swaths of underutilized spectrum in these bands, and the number of available spatial degrees of freedom (DoF) for a given aperture is higher at smaller carrier wavelengths. In this paper, we consider line of sight (LoS) multi-input multi-output (MIMO) communication, which exploits this attractive scaling of spatial DoF and bandwidth in these bands. For 1D apertures with a horizontal distance, or range, R between transmit and receive arrays having lengths of L_T and L_R, the number of spatial DoF based on information-theoretic considerations, given by [4]

$$DoF \approx \frac{L_T L_R}{R\lambda} + 1 \tag{1}$$

scales inversely with the carrier wavelength λ, and therefore linearly with the carrier frequency $f_c = c/\lambda$, where c is the speed of light. The result in (1) can be extended for 2D apertures and rewritten as [4]

$$DoF \approx \frac{A_T A_R}{R^2\lambda^2} + 1 \tag{2}$$

where A_T and A_R are the areas occupied by the 2D arrays at the transmitter and the receiver, respectively, so that the scaling with f_c becomes quadratic. Since transmission bandwidth typically scales linearly with carrier frequency, the overall data rates can potentially scale cubically with carrier frequency.

Advances in mmWave radio frequency integrated circuits (RFIC) in low-cost silicon semiconductor processes open up the possibility of deploying LoS MIMO at scale, for example, to boost link capacities in wireless backhaul mesh networks for urban picocells [5]. Consider 4×4 LoS MIMO with a link distance of 100 m. At a carrier frequency of 140 GHz, the form factor required for a well-conditioned spatial channel is small enough to permit opportunistic deployment (e.g., on lamp-posts): the inter-antenna spacing for orthogonal eigenmodes is 33 cm. With QPSK modulation and $10 - 20$ GHz bandwidth, we can achieve $80 - 160$ Gbps uncoded data rates. Therefore, with lightweight channel coding, 100 Gbps becomes a feasible target. However, can we leverage the economies of scale in digital computation to realize such transceivers at reasonable cost and power consumption, using all-digital baseband signal processing? While this is standard in modern communication receivers operating at lower bandwidths (typically below 1 GHz), as signaling bandwidths increase, realizing high-precision analog-to-digital converters (ADCs) is a challenge [6], [7]. Motivated by these considerations, we investigate in this paper whether it is possible to use all-digital processing in LoS MIMO receivers with severely quantized samples.

Contributions: We investigate the design of low-precision quantizers and of spatial demultiplexing with heavily quantized observations.

Quantizer Design: Rather than trying to design optimal quantizers, our first goal is to design quantizers with regular structure which approach the same Shannon limit as an unquantized system at high SNR.

• Our first result is negative. As bandwidth increases, a particularly attractive approach is phase-only quantization: this can be implemented by passing linear combinations of the

This work was supported in part by ComSenTer, one of six centers in JUMP, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. Use was made of computational facilities purchased with funds from the National Science Foundation (CNS-1725797) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 1720256) at UC Santa Barbara.

A. D. Sezer and U. Madhow are with the Department of Electrical and Computer Engineering, University of California Santa Barbara, CA, 93106 USA (e-mail: adsezer@ece.ucsb.edu, madhow@ece.ucsb.edu)

real and imaginary parts of the sample through sign detectors (one-bit ADCs), and therefore does not require automatic gain control (see [8]). However, we prove for the 2×2 QPSK system that phase-only quantizers cannot meet the unquantized benchmark at high SNR.

• Our second result shows that amplitude-phase quantization with a relatively small number of bins does attain the unquantized benchmark. Specifically, we prove for the 2×2 QPSK system that 2-level amplitude and 8-level phase quantization works.

• For the 4×4 system, we obtain practical guidelines and design prescriptions for quantizer design. We show via mutual information computations that per-antenna quantization into equal probability regions (which maximizes per-antenna output entropy) performs better than conventional MMSQE quantization, and that I/Q quantization performs better than amplitude/phase quantization. In particular, we show that equal probability I/Q quantization with 2 bits per real dimension, designed using a Gaussian approximation for the received samples, achieves the unquantized benchmark at high SNR for QPSK modulation, attaining a maximum data rate of 8 bits per channel use.

Spatial demultiplexing: For a 4×4 QPSK system, we investigate spatial demultiplexing with observations quantized using the 2 bit I/Q quantizer that we have designed. We considered well-conditioned LoS MIMO channels for which linear zero-forcing detection provides near-optimal performance with unquantized observations. Our goal is to attain uncoded error probabilities of 10^{-3} or better, for which reliable communication can be obtained with lightweight, high-rate error correcting codes.

• We show that linear detection with quantized observations leads to an error floor. Since mutual information computations show that the maximum rate of 8 bits per channel use is attainable with moderate SNR penalty for the quantizer design used, we expect maximum likelihood detection not to exhibit an error floor. We show that this is indeed the case, but the prohibitive complexity (exponential in the number of transmitted bits) motivates design of lower-complexity spatial demultiplexing schemes.

• We introduce the concept of *virtual quantization,* modeling the uncertainty created by quantization as a nuisance parameter, so that the task of estimating the transmitted symbols can be approached using the tools of composite hypothesis testing. We employ a Generalized Likelihood Ratio Test (GLRT) approach which leverages the efficacy of linear detection for the well-conditioned MIMO channels considered here. A key computational advantage of the proposed approach is that, unlike maximum likelihood detection, its complexity does not scale with constellation size. In addition, we show that our proposed virtual quantization concept is resistant to the changes in channel condition due to suboptimal separation of the transmitter and the receiver (up to $\pm 20\%$ variations of the nominal range).

• While the bulk of our numerical examples are for QPSK, we also present results for 16QAM, demonstrating that our prescriptions for quantizer design and our proposed virtual quantization approach extend to larger constellations.

Notation: Throughout the paper, random variables are denoted by capital letters and small letters are used for the specific value that the random variables take. Bold letters are used to denote vectors and matrices. $\mathbb{E}_Z$ denotes the expectation operator over the random variable Z. $|Z|$ and $\angle Z$ represent the amplitude and the phase of Z, respectively. $\Re(Z)$ and $\Im(Z)$ denote real and imaginary part of complex number Z, respectively. $\mathbf{X}^\intercal$ and $\mathbf{X}^\dagger$ are the transpose and Hermitian transpose of $\mathbf{X}$, respectively. $\mathbf{I}_n$ is the identity matrix of size n.

II. Related Work

The DoF for LoS MIMO as a function of transceiver form factor and antenna placement, range and carrier frequency are by now well known [4], [9]. It is worth contrasting the motivation for our work with a recently developed LoS MIMO system [10] which employs $2.5\,\text{GHz}$ bandwidth in E-band ($70-80\,\text{GHz}$ carrier frequency), and achieves $100+$ Gbps at a distance of 1.5 km using 8-fold multiplexing (spatial degrees of freedom along with dual polarization) and alphabets as large as 64QAM. The optimal antenna separation is $1.72\,\text{m}$, requiring bulky antenna structures and careful installation. We envision higher frequencies and shorter ranges to reduce form factor to enable opportunistic deployment. The goal of our investigation of severely quantized LoS MIMO, therefore, is to examine how far we can push the paradigm of using larger available bandwidths (which limits the precision of available ADCs) to reduce the required constellation size (which potentially enables reduction in ADC precision). Our work also contrasts with recent efforts in the research literature based on analog-centric [11], [12] or hybrid analog-digital [13], [14] processing in an attempt to sidestep the ADC bottleneck.

Since LoS MIMO is often envisioned for quasi-static links (e.g., wireless backhaul), it is natural to consider precoding with channel state information at the transmitter, as in a number of theoretical studies [14]–[16]. Transmit precoding can also significantly reduce the dynamic range at the receiver, easing the task of analog-to-digital conversion. Indeed, prior studies of MIMO capacity with low-precision ADC assume transmit precoding [13], [17], [18]. Channel capacity with 1-bit ADC is studied in [17], which provides capacity bounds and a convex optimization based algorithm to obtain capacity-achieving constellations. In [13], the capacity with transmit precoding, together with hybrid analog-digital processing at the receiver, where analog linear combinations of the signals received at different antennas are quantized, is studied. In [18], joint transmit power and ADC allocation problem is studied for throughput maximization, which results in that using few one-bit ADCs with the adaptive threshold receiver is enough to achieve near optimal performance.

Transmit precoding leads to increased dynamic range at the transmitter, which aggravates the already difficult problem of producing power at higher frequencies, such as the millimeter wave or THz bands. In this paper, therefore, we explore LoS MIMO *without* transmit precoding, in contrast to the cited prior work. We assume that the receiver has ideal channel estimates. Channel estimation with low-precision ADC is not

as challenging as demodulation: [19] is an early example for a SISO dispersive channel, while [20] and [21] propose effective estimation techniques for massive MIMO with 1-bit quantization at the receive antennas.

There have been prior studies of demodulation [22]–[24] based on quantized samples for MIMO systems without precoding, but these consider Rayleigh faded channel models associated with rich scattering environments, unlike the LoS MIMO setting considered here. The computational intractability of maximum likelihood detection is pointed out in [22], [23], while large system analysis for suboptimal loopy belief propagation is considered in [24].

Shannon limits for an ideal SISO discrete-time additive white Gaussian noise (AWGN) channel with low-precision ADC are studied in [25]. It is shown that the optimal input distribution is discrete and can be computed numerically, but standard constellations are near-optimal. Further, the use of ADCs with 2-3 bits precision results in only a small reduction in channel capacity even at moderately high SNR. Our model is perhaps the simplest possible extension of this framework to MIMO systems.

This paper builds on our preliminary results on quantizer design in an earlier conference paper [26]. We provide proofs and technical details, as well as more detailed insights and numerical results, for quantizer design here. The results on spatial demultiplexing, including the proposed virtual quantization concept, are entirely new.

III. System Model and Problem Formulation

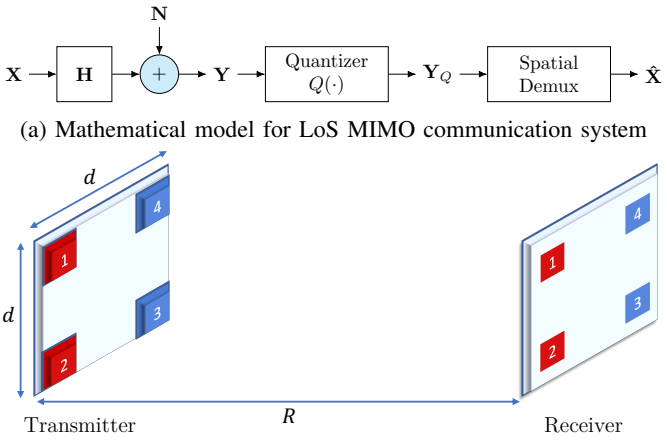
We consider a symmetric 4×4 LoS MIMO communication system, with equal inter-antenna spacings at transmitter and receiver, as shown in Fig. 1. Each transmit/receive antenna may be a fixed beam antenna [10] or an electronically steerable "subarray" [4]. Fixed beam antennas are typically manually aligned along the LoS, while subarrays can be aligned using feedback-based beam alignment algorithms [27], [28]. In either case, we obtain a highly directive beam along the LoS, so that multipath can be ignored. The received signal vector $\mathbf{Y} \triangleq [Y_1 \cdots Y_4]^\mathsf{T} \in \mathbb{C}^{4\times 1}$ is given by

$$\mathbf{Y} = \mathbf{H}\,\mathbf{X} + \mathbf{N}\,, \tag{3}$$

where $\mathbf{X} \triangleq [X_1 \cdots X_4]^\mathsf{T} \in \mathbb{C}^{4\times 1}$ is the transmitted symbol vector, $\mathbf{H} \in \mathbb{C}^{4\times 4}$ is the normalized channel matrix (with each column normalized to unit norm), and $\mathbf{N} \sim \mathcal{CN}(0, \sigma^2\,\mathbf{I}_4)$ is AWGN. Under this normalization, the SNR for the kth data stream is given by $SNR = \mathbb{E}\{|X_k|^2\}/\sigma^2$.

Input: We consider QPSK modulation unless otherwise stated (results for 16QAM are included in Section VI). For QPSK modulation, $\{X_i\}_{i=1}^4$ are independent and identically distributed symbols taking values $\{e^{j\pi/4}, e^{j3\pi/4}, e^{j5\pi/4}, e^{j7\pi/4}\}$ with equal probability. Thus, $SNR = \mathbb{E}\{|X_k|^2\}/\sigma^2 = 1/\sigma^2$ where $\mathbb{E}\{|X_k|^2\} = 1$ for all $k \in \{1, \ldots, 4\}$.

Channel: For the pure LoS channel we consider, the elements of $\mathbf{H}$ in (3) are calculated by employing *ray-tracing* in consideration of the spherical nature of the wave propagation [29], [30] and the columns of $\mathbf{H}$ are normalized to unit norm. Since the path loss differences among different transmit-receive antenna pairs are negligible, the normalized channel matrix for the symmetric 4×4 LoS MIMO communication system is given by



(a) Mathematical model for LoS MIMO communication system

(b) Geometric configuration for the 4×4 LoS MIMO system (with blue and red colored antennas) and the 2×2 LoS MIMO system (with only red colored antennas)

Fig. 1: LoS MIMO communication system model

$$\mathbf{H} = \frac{1}{2} e^{-j\Phi} \begin{bmatrix} 1 & e^{-j\theta} & e^{-j2\theta} & e^{-j\theta} \\ e^{-j\theta} & 1 & e^{-j\theta} & e^{-j2\theta} \\ e^{-j2\theta} & e^{-j\theta} & 1 & e^{-j\theta} \\ e^{-j\theta} & e^{-j2\theta} & e^{-j\theta} & 1 \end{bmatrix}, \tag{4}$$

where the random variable Φ denotes the common phase change along the path between the transmitter and the receiver, and the "cross-over phase" depends on the inter-antenna spacing d and link distance R as follows [4], [31]:

$$\theta = \frac{2\pi}{\lambda}(\sqrt{R^2 + d^2} - R) \approx \frac{\pi d^2}{\lambda R} \text{ for } R \gg d \tag{5}$$

where λ denotes the carrier wavelength. We note that the channel matrix in (4) is symmetric due to the symmetry of the distances between transmit-receive antenna pairs. We would like our quantizer designs to be robust to variations in the common phase Φ, which is assumed to be uniformly distributed over $[0, 2\pi)$.

The common phase Φ is well modeled as uniform over $[0, 2\pi]$: even small variations in the link distance R lead to large phase changes in the carrier because of the small wavelength, and the relative phase offset between the local oscillators (LOs) at the transmitter and receiver drifts across time.

The simplicity of the frequency nonselective LoS channel model in (4) is a consequence of our assumption of a strongly directive beam along the LoS path. For the scenarios of interest to us (e.g., lamppost-to-lamppost or rooftop-to-rooftop backhaul communication), reflections from scattering surfaces (e.g., from a wall or the ground), which are already attenuated due to scattering, arrive at an angle significantly offset from the LoS and are further attenuated by the transmit and receive beams pointed along the LoS.

Quantizer: We consider identical quantizers at each receive antenna. The quantized output of the ith receive antenna can be expressed as

$$\bar{Y}_i = Q(Y_i)\,, \tag{6}$$

for $i \in \{1,\ldots,4\}$. $Q(\cdot)$ in (6) represents the quantizer function at each receive antenna and for a given input y, $Q(y)$ can be characterized as

$$Q(y) = \tilde{y}_j\,,\ \text{if}\ \ y \in \Gamma_j\,, \tag{7}$$

for $j \in \{1,\ldots,T\}$, where $\tilde{y}_j$ for $j \in \{1,\ldots,T\}$ is a design parameter and Γ_1, …, Γ_T denote the decision regions for the quantizer, with T denoting the number of quantizer bins at each receive antenna.

Spatial demultiplexer: Based on the quantized observations; that is, $\mathbf{Y}_Q \triangleq \left[\bar{Y}_1 \cdots \bar{Y}_4\right]^\intercal \in \mathbb{C}^{4\times 1}$, the receiver performs the spatial demultiplexing and provides the estimate of the transmitted symbol as $\hat{\mathbf{X}} \triangleq \left[\hat{X}_1 \cdots \hat{X}_4\right]^\intercal \in \mathbb{C}^{4\times 1}$.

We begin with the 2×2 system depicted in Fig. 1b with red colored antennas in order to make some fundamental theoretical observations regarding quantization. For this scheme, $\mathbf{X}$ and $\mathbf{Y}$ in (3) are defined as $\mathbf{X} \triangleq [X_1\, X_2]^\intercal \in \mathbb{C}^{2\times 1}$ and $\mathbf{Y} \triangleq [Y_1\, Y_2]^\intercal \in \mathbb{C}^{2\times 1}$, respectively. Also, $\mathbf{N} \sim \mathcal{CN}(0, \sigma^2\,\mathbf{I}_2)$. The channel matrix for this scheme corresponds to the renormalized version of the upper-left 2×2 submatrix of (4) and is given by

$$\mathbf{H} = \frac{1}{\sqrt{2}} e^{-j\Phi} \begin{bmatrix} 1 & e^{-j\theta} \\ e^{-j\theta} & 1 \end{bmatrix}\,, \tag{8}$$

where $\theta \approx \frac{\pi d^2}{\lambda R}$ for $R \gg d$ as in (5).

One possible formulation of optimal quantization is to minimize the "information distortion"

$$D(\mathbf{X}, \mathbf{Y}_Q, \theta) \triangleq \mathbb{E}_\Phi\{I(\mathbf{X}; \mathbf{Y} \mid \Phi, \theta) - I(\mathbf{X}; \mathbf{Y}_Q \mid \Phi, \theta)\} \tag{9}$$

where $\mathbf{Y}_Q \triangleq \left[\bar{Y}_1\, \bar{Y}_2\right]^\intercal$ and the function $I(\bar{\mathbf{X}}; \bar{\mathbf{Y}} \mid \Phi, \theta)$ represents the mutual information between the random variables $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$ for given Φ and θ. Based on the data processing equality, $D(\mathbf{X}, \mathbf{Y}_Q, \theta) \geq 0$ since $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Y}_Q$ form a Markov chain; that is, $\mathbf{X} \to \mathbf{Y} \to \mathbf{Y}_Q$. Also, $I(\mathbf{X}; \mathbf{Y} \mid \Phi, \theta)$ in (9) does not depend on any parameter related to quantizer. For that reason, the problem of minimizing $D(\mathbf{X}, \mathbf{Y}_Q, \theta)$ in (9) is equivalent to

$$\max_{\{\Gamma_j\}_{j=1}^{T}} \ \mathbb{E}_\Phi\{I(\mathbf{X}; \mathbf{Y}_Q \mid \Phi, \theta)\}\,. \tag{10}$$

In the optimization problem in (10), the mutual information between $\mathbf{X}$ and $\mathbf{Y}_Q$ must be maximized over the set of all possible quantization regions of the quantizer at the receive antennas. The number of quantization bins for the quantizer is not fixed in (10), and must also be optimized. Thus, it is difficult to solve (10). Furthermore, the optimal quantizers may correspond to irregular regions, leading to implementation difficulties. In this paper, therefore, we opt for designing regular quantizers with the goal of ensuring that $D(\mathbf{X}, \mathbf{Y}_Q, \theta) \to 0$ at high SNR.

IV. Analytical Insights from 2×2 MIMO

In this section, we provide detailed insight regarding quantization via the 2×2 system. By analysis of a limiting noiseless regime, we prove that phase-only quantization cannot yield unique decodability, while amplitude-phase quantization can.

A. Phase-only Quantization

The following lemma establishes a negative result for phase-only quantization. Consider a pair of possible transmitted symbol vectors $\mathbf{X}^{(1)} = (X_1^{(1)}, X_2^{(1)})$ and $\mathbf{X}^{(2)} = (X_1^{(2)}, X_2^{(2)})$. The corresponding noise-free received samples are given by $Y_1^{(i)} \triangleq e^{-j\phi}(X_1^{(i)} + e^{-j\theta}X_2^{(i)})/\sqrt{2}$ and $Y_2^{(i)} \triangleq e^{-j\phi}(e^{-j\theta}X_1^{(i)} + X_2^{(i)})/\sqrt{2}$ for $i \in \{1,2\}$. The lemma specifies choices for $\mathbf{X}^1$ and $\mathbf{X}^2$ for which the noise-free received samples prior to quantization have the same phase. Thus, these pairs cannot be distinguished based on phase-only quantization.

Lemma 1: *For* $(X_1^{(1)}, X_2^{(1)}) = (e^{j\pi(2i-1)/4}, e^{j\pi(2i+1)/4})$ *where* $i \in \{1,\ldots,4\}$*, the following statements hold:*

(i) For $\theta \in (-\pi/2, \pi/2)$ *and* $(X_1^{(2)}, X_2^{(2)}) = (X_2^{(1)}, X_1^{(1)})$,

$$\angle Y_1^{(1)} = \angle Y_2^{(1)},\ \angle Y_1^{(2)} = \angle Y_2^{(2)},\ \angle Y_1^{(1)} = \angle Y_1^{(2)} \tag{11}$$

(ii) For $\theta \in (\pi/2, 3\pi/2)$ *and* $(X_1^{(2)}, X_2^{(2)}) = (X_2^{(1)}, X_1^{(1)})$,

$$\angle Y_1^{(1)} = \angle Y_2^{(1)} + \pi,\ \angle Y_1^{(2)} = \angle Y_2^{(2)} + \pi,\ \angle Y_1^{(1)} = \angle Y_1^{(2)} + \pi \tag{12}$$

(iii) For $\theta \in (-\pi/2, \pi/2)$ *and* $(X_1^{(2)}, X_2^{(2)}) = (e^{j\pi}X_2^{(1)}, e^{j\pi}X_1^{(1)})$,

$$\angle Y_1^{(1)} = \angle Y_2^{(1)},\ \angle Y_1^{(2)} = \angle Y_2^{(2)},\ \angle Y_1^{(1)} = \angle Y_1^{(2)} + \pi \tag{13}$$

(iv) For $\theta \in (\pi/2, 3\pi/2)$ *and* $(X_1^{(2)}, X_2^{(2)}) = (e^{j\pi}X_2^{(1)}, e^{j\pi}X_1^{(1)})$,

$$\angle Y_1^{(1)} = \angle Y_2^{(1)} + \pi,\ \angle Y_1^{(2)} = \angle Y_2^{(2)} + \pi,\ \angle Y_1^{(1)} = \angle Y_1^{(2)} \tag{14}$$

Proof: The result in the lemma can simply be shown by using Euler's formula and Pythagorean trigonometric identity. ■

This results in the following proposition stating that phase-only quantization cannot achieve the unquantized benchmark.

Proposition 1: *For any phase-only quantization scheme with any number of bins,* $D(\mathbf{X}, \mathbf{Y}_Q, \theta) > 0$ *for all* $\theta \in [0, 2\pi)$ *as* $\sigma \to 0$.

Proof: In order to show that $D(\mathbf{X}, \mathbf{Y}_Q, \theta) > 0$ for all $\theta \in [0, 2\pi)$ as $\sigma \to 0$, $\mathbb{E}_\Phi\{I(\mathbf{X}; \mathbf{Y}_Q \mid \Phi, \theta)\} < 4$ should be proved for $\theta \in [0, 2\pi)$ as $\sigma \to 0$ since $\mathbb{E}_\Phi\{I(\mathbf{X}; \mathbf{Y} \mid \Phi, \theta) \to 4$ as $\sigma \to 0$. Before proving that, first, it is shown that $p(\mathbf{X} = \mathbf{x} \mid \mathbf{Y}_Q = \mathbf{y}_Q, \Phi = \phi, \theta)$ satisfies for some $\mathbf{Y}_Q = \mathbf{y}_Q$ that $0 < p(\mathbf{X} = \mathbf{x} \mid \mathbf{Y}_Q = \mathbf{y}_Q, \Phi = \phi, \theta) < 1$ as $\sigma \to 0$. In other words, we show that $p(\mathbf{X} = \mathbf{x} \mid \mathbf{Y}_Q = \mathbf{y}_Q, \Phi = \phi, \theta)$ cannot be $p(\mathbf{X} = \mathbf{x} \mid \mathbf{Y}_Q = \mathbf{y}_Q, \Phi = \phi, \theta) \in \{0, 1\}$ for all $\mathbf{Y}_Q = \mathbf{y}_Q$. Based on the statement in Lemma 1, it can be stated that the noise-free outputs corresponding to inputs

$(X_1^{(1)}, X_2^{(1)}) = (e^{j\pi(2i-1)/4}, e^{j\pi(2i+1)/4})$ for $i \in \{1,\dots,4\}$ and $(X_1^{(2)}, X_2^{(2)}) = (X_2^{(1)}, X_1^{(1)})$ have the same phase and fall in the same quantization bin for $\theta \in (-\pi/2, \pi/2)$. Similarly, for $(X_1^{(1)}, X_2^{(1)}) = (e^{j\pi(2i-1)/4}, e^{j\pi(2i+1)/4})$ for $i \in \{1,\dots,4\}$ and $(X_1^{(2)}, X_2^{(2)}) = (e^{j\pi} X_2^{(1)}, e^{j\pi} X_1^{(1)})$, the outputs without additive noise stay in the same bin of any given phase-only quantization mapping since they have the same phase for $\theta \in (\pi/2, 3\pi/2)$. In addition, for $\theta = \pi/2$ and $\theta = 3\pi/2$, the amplitude of one of the noise-free outputs is zero and the same ambiguity occurs for those cases as well. Without loss of generality, say that those noiseless outputs (i.e., the outputs with additive Gaussian noise as $\sigma \to 0$) after quantization is $\mathbf{Y}_Q = \bar{\mathbf{y}}_Q$ for the inputs $\mathbf{X} = \mathbf{x}_1$ and $\mathbf{X} = \mathbf{x}_2$. Then, the following statements hold for $i \in \{1, 2\}$ based on Bayes' theorem:

$$0 < p(\mathbf{x}_i \mid \mathbf{Y}_Q = \bar{\mathbf{y}}_Q, \phi, \theta) = \frac{p(\mathbf{Y}_Q = \bar{\mathbf{y}}_Q \mid \mathbf{x}_i, \phi, \theta)}{\sum_{\bar{\mathbf{x}}} p(\mathbf{Y}_Q = \bar{\mathbf{y}}_Q \mid \bar{\mathbf{x}}, \phi, \theta)} < 1 \quad (15)$$

since $p(\mathbf{Y}_Q = \bar{\mathbf{y}}_Q \mid \mathbf{x}_1, \phi, \theta) \to 1$ and $p(\mathbf{Y}_Q = \bar{\mathbf{y}}_Q \mid \mathbf{x}_2, \phi, \theta) \to 1$ as $\sigma \to 0$. Then, as $\sigma \to 0$, $H(\mathbf{X} \mid \mathbf{Y}_Q, \Phi = \phi, \theta) > 0$ for all $\theta \in [0, 2\pi)$ and consequently $I(\mathbf{X}; \mathbf{Y}_Q \mid \Phi = \phi, \theta) < 4$ for all $\phi \in [0, 2\pi)$ based on the definition of mutual information and $\mathbb{E}_\Phi\{I(\mathbf{X}; \mathbf{Y}_Q \mid \Phi, \theta)\} < 4$ is satisfied for all $\theta \in [0, 2\pi)$. ■

While the unquantized benchmark cannot be achieved, it is still of interest to ask how many phase quantization bins are enough to reach the high-SNR asymptote for phase-only quantization. We now establish that, for our system, 8 phase quantization bins suffice. We begin with the following lemma.

Lemma 2: *For any possible* $(X_1^{(1)}, X_2^{(1)})$ *and* $(X_1^{(2)}, X_2^{(2)})$ *input pairs,* $\angle Y_1^{(1)} - \angle Y_1^{(2)} = 0 \pmod{\pi/4}$ *and* $\angle Y_2^{(1)} - \angle Y_2^{(2)} = 0 \pmod{\pi/4}$*, where* $Y_1^{(i)}$ *and* $Y_2^{(i)}$ *are as defined in Lemma 1. Also,* $\angle Y_1^{(1)} - \angle Y_1^{(2)}$ *and* $\angle Y_2^{(1)} - \angle Y_2^{(2)}$ *can take* 8 *different values.*

Proof: By using $\arctan(x) - \arctan(y) = \arctan(\frac{x-y}{1+xy})$ and Euler's formula, the proof is straightforward. ■

Based on Lemma 2, we can derive the following proposition stating that 8 phase quantization bins suffice.

Proposition 2: *As* $\sigma \to 0$*, any phase-only quantization schemes with more than 8 regions cannot achieve higher data rate than phase-only quantization scheme with 8 equally partitioned sectors.*

Proof: Consider a phase-only quantization scheme having more than 8 bins and let $L > 8$ denote the number of bins of that scheme. Lemma 2 implies that the noise-free outputs (i.e., the outputs as $\sigma \to 0$) of the all possible inputs can take 8 different phase values for given ϕ and θ. Then, based on the pigeonhole principle, at least $L - 8$ bins of the phase-only quantizer do not contain any outputs for given ϕ and θ as $\sigma \to 0$. In other words, none of the outputs corresponding to all possible inputs fall into those bins as $\sigma \to 0$. Let $\bar{\mathcal{S}}$ denote the index set of those empty bins. Then, define a new set, $\mathcal{S}_E$ as $\mathcal{S}_E = \{\mathbf{y} = [i, j] \mid (i \in \bar{\mathcal{S}} \wedge j \in \{1,\dots,L\}) \vee (j \in \bar{\mathcal{S}} \wedge i \in \{1,\dots,L\})\}$. It can be stated that $p(\mathbf{Y}_Q = \mathbf{y}_Q) = 0$ for all $\mathbf{y}_Q \in \mathcal{S}_E$ as $\sigma \to 0$. Now, two cases should be analyzed separately. First, if exactly $L - 8$ bins of the phase-only quantizer are empty as $\sigma \to 0$; then, any two different noise-free outputs having different phases cannot be in the same bin due to the result in Lemma 2, which also holds for the phase-only quantization scheme with equally divided 8 regions. Let $\mathcal{S}_Q^L$ and $\mathcal{S}_Q$ denote the sets of all possible quantized outputs for the quantization scheme having $L > 8$ bins and 8 bins, respectively. There exists a one-to-one correspondence between $\mathcal{S}_Q^L \backslash \mathcal{S}_E$ and $\mathcal{S}_Q$ and if $\mathbf{m} \in \mathcal{S}_Q^L \backslash \mathcal{S}_E$ and $\mathbf{n} \in \mathcal{S}_Q$ are the paired elements, it is stated that the input producing quantized output $\mathbf{m}$ in the quantization scheme with more than 8 bins produces $\mathbf{n}$ in the quantization scheme with equally partitioned 8 regions as $\sigma \to 0$. Thus,

$$\sum_{\mathbf{y}_Q \in \mathcal{S}_Q^L} H(\mathbf{X} \mid \mathbf{y}_Q) p(\mathbf{y}_Q) = \sum_{\mathbf{y}_Q \in \mathcal{S}_Q^L \backslash \mathcal{S}_E} H(\mathbf{X} \mid \mathbf{y}_Q) p(\mathbf{y}_Q) \quad (16)$$

$$= \sum_{\mathbf{y}_Q \in \mathcal{S}_Q} H(\mathbf{X} \mid \mathbf{y}_Q) p(\mathbf{y}_Q), \quad (17)$$

where (16) is due to $p(\mathbf{Y}_Q = \mathbf{y}_Q) = 0$ for all $\mathbf{y}_Q \in \mathcal{S}_E$. Therefore, both of the schemes achieve the same data rate if $L - 8$ bins of the phase-only quantizer are empty as $\sigma \to 0$. Next, consider the case that more than $L - 8$ bins are empty. Since the noise-free outputs can have 8 different phase values for given ϕ and θ, some of those outputs having different phases are in the same bin, which is not a possible case for the phase-only quantization scheme with equally sized regions. For that reason, it can be calculated that $H(\mathbf{X} \mid \mathbf{Y}_Q^{(1)}, \Phi = \phi, \theta) - H(\mathbf{X} \mid \mathbf{Y}_Q^{(2)}, \Phi = \phi, \theta) \geq 0$ as $\sigma \to 0$, where $\mathbf{Y}_Q^{(1)}$ and $\mathbf{Y}_Q^{(2)}$ denote the quantized outputs under the quantization schemes with more than 8 regions and equally partitioned exactly 8 regions, respectively. Therefore, $I(\mathbf{X}; \mathbf{Y}_Q^{(1)} \mid \Phi = \phi, \theta) \leq I(\mathbf{X}; \mathbf{Y}_Q^{(2)} \mid \Phi = \phi, \theta)$. ■

B. Amplitude-Phase Quantization

For K-ary amplitude and M-ary phase quantization, the quantization set of $(m + M(k-1))$th-bin of a quantizer can be written as

$$\bar{\Gamma}_{m+M(k-1)} = \{\bar{Y} \mid A_{k-1} \leq |\bar{Y}| < A_k, \frac{2\pi}{M}(m-1) \leq \angle \bar{Y} < \frac{2\pi}{M} m\}, \quad (18)$$

for $m \in \{1,\dots,M\}$ and $k \in \{1,\dots,K\}$, where $A_1,\dots,A_{K-1}$ are the amplitude thresholds (we set $A_0 = 0$ and $A_K = \infty$ to maintain a unified notation across quantization bins).

The following proposition states that $K = 2$ and $M = 8$ suffices to attain the unquantized benchmark.

Proposition 3: *As* $\sigma \to 0$*, circularly symmetric quantization with* 2*-level amplitude and* 8*-level phase quantization attains* $D(\mathbf{X}, \mathbf{Y}_Q, \theta) \to 0$ *for* $\theta \in [0, 2\pi)$.

Proof: Proposition 1 is based on the observation that the outputs of some input pairs have the same phase at both of the antennas as $\sigma \to 0$, so that those outputs cannot be differentiated by employing any phase-only quantization scheme. On the other hand, the proof of Proposition 2 shows that a phase-only scheme with equally partitioned 8 regions

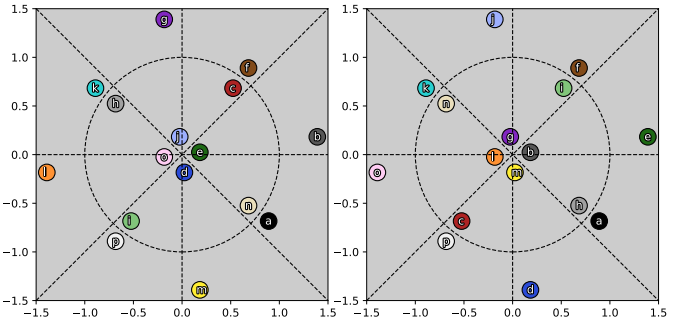


Fig. 2: All possible noise-free outputs before the quantization, $e^{-j\phi}(X_1 + e^{-j\theta}X_2)/\sqrt{2}$ (Left) and $e^{-j\phi}(e^{-j\theta}X_1 + X_2)/\sqrt{2}$ (Right), for $\theta = 5\pi/12$ and $\phi = \pi/4$.

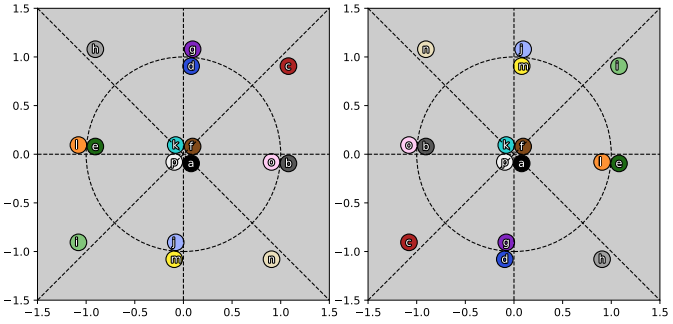


Fig. 3: All possible noise-free outputs before the quantization, $e^{-j\phi}(X_1 + e^{-j\theta}X_2)/\sqrt{2}$ (Left) and $e^{-j\phi}(e^{-j\theta}X_1 + X_2)/\sqrt{2}$ (Right), for $\theta = 17\pi/18$ and $\phi = \pi/18$.

can distinguish noise-free outputs having two different phases, due to the result in Lemma 2. In this proof, the aim is to show that considering a 2-level amplitude quantization together with phase quantization resolves the ambiguities leading to the result in Proposition 1. First, it can be shown that only the outputs corresponding to the input pairs discussed in Lemma 1 cannot be distinguished via phase-only scheme having equally partitioned 8 regions. For that reason, consider the input pairs in Lemma 1. For $(X_1^{(1)}, X_2^{(1)}) = (e^{j\pi(2i-1)/4}, e^{j\pi(2i+1)/4})$ and $(X_1^{(2)}, X_2^{(2)}) = (X_2^{(1)}, X_1^{(1)})$, where $i \in \{1, \ldots, 4\}$, $|Y_1^{(2)}| < 1 < |Y_1^{(1)}|$ and $|Y_2^{(1)}| < 1 < |Y_2^{(2)}|$ for $\theta \in (0, \pi/2]$, $|Y_1^{(1)}| = |Y_1^{(2)}| = 1$ and $|Y_2^{(1)}| = |Y_2^{(2)}| = 1$ for $\theta = 0$, and $|Y_1^{(1)}| < 1 < |Y_1^{(2)}|$ and $|Y_2^{(2)}| < 1 < |Y_2^{(1)}|$ for $\theta \in [-\pi/2, 0)$. Due to the symmetry, the same approach can be applied for other input pairs (i.e., $(X_1^{(1)}, X_2^{(1)}) = (e^{j\pi(2i-1)/4}, e^{j\pi(2i+1)/4})$ and $(X_1^{(2)}, X_2^{(2)}) = (e^{j\pi}X_2^{(1)}, e^{j\pi}X_1^{(1)})$ for $i \in \{1, \ldots, 4\}$) when $\theta \in (\pi/2, 3\pi/2)$. Since the amplitude of the outputs does not depend on $\Phi = \phi$ and a circularly symmetric quantization scheme is employed, a phase quantization scheme including a 2-level amplitude quantization with $A_0 = 0$, $A_1 = 1$, and $A_2 = \infty$ resolves the ambiguity between those outputs. It is easy to now conclude that $D(\mathbf{X}, \mathbf{Y}_Q, \theta) \to 0$ for $\theta \in [0, 2\pi)$ as $\sigma \to 0$. ■

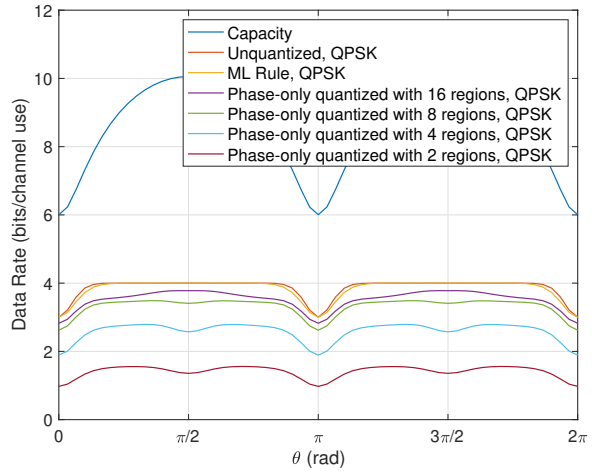
C. Numerical Results

In this section, numerical examples are provided to illustrate the theoretical results. We first illustrate the statements in the lemmas and the propositions via example noise-free outputs prior to quantization. We then compute and compare Shannon limits for different quantization schemes.

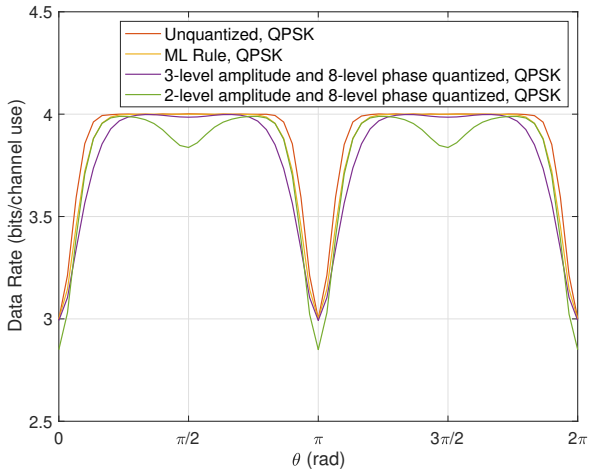
We illustrate the geometry behind the proofs by presenting noiseless outputs prior to quantization for a well-conditioned and a poorly conditioned channel in Fig. 2 and Fig. 3, respectively. We see that some output pairs (e.g., (b, e), (d, m), (g, j) and (l, o) in Fig. 2 and (b, o), (d, g), (e, l) and (j, m) in Fig. 3) have the same phase at *both* receive antennas, and hence cannot be distinguished based on phase-only quantization, as stated in Lemma 1. On the other hand, the other outputs can indeed be distinguished based on phase-only quantization. In addition, for given θ and ϕ, the phase of noise-free outputs can have 8 different values, and two different outputs having two different phase values cannot be in the same bin for phase-only quantization with 8 equal sectors. This is the intuitive basis for Proposition 2. Lastly, noise-free output pairs having the same phase at both receive antennas, such as (b, e) in Fig. 2 can be separated by employing an amplitude quantization scheme with 2 regions as illustrated in Fig. 2 and Fig. 3. This is the intuition behind Proposition 3.

Next, we plot the data rate (mutual information) attained by different quantization schemes (see Appendix A for details). Two benchmarks are considered: an unquantized system, and a quantizer based on Voronoi regions separating the outputs at each antenna. We may view the latter as an ML decision rule at each antenna, where input-pairs that fall on top of each other are interpreted as a single point, and it is easy to see that it attains the unquantized benchmark at high SNR. However, it depends on θ and Φ, and is an irregular quantizer, which is unattractive in practice.

For a 2×2 MIMO system, Fig. 4a and Fig. 4b plot data rate versus $\theta \in [0, 2\pi)$ at 15 dB SNR for phase-only and amplitude-phase quantization, respectively. Similarly, Fig. 5a and Fig. 5b plot data rates versus SNR, fixing $\theta = \pi/2$ (the best conditioned channel). For 2-level amplitude and 8-level phase quantization, the amplitude threshold is set to $A_1 = 1$, whereas the thresholds are $A_1 = 0.75$ and $A_2 = 1.25$ for 3-level amplitude and 8-level phase quantization. The plots illustrate the trends predicted by our theoretical results: phase-only quantization does not attain the unquantized or ML benchmarks, while amplitude-phase quantization does attain these at high enough SNR. However, the performance at moderate SNR can benefit from a larger number of quantization bins than those indicated by high-SNR asymptotics. For example, while 8 phase quantization bins are as good as any other phase-only quantization scheme asymptotically, using 16 quantization bins does provide better performance at moderate SNRs (Fig. 4a and Fig. 5a). Similarly, while 2-level amplitude quantization suffices, there is a gain at moderate SNRs with 3-level quantization (Fig. 4b and Fig. 5b). In particular, Fig. 5b shows that for a well-conditioned channel, while 2-level amplitude quantization attains the unquantized

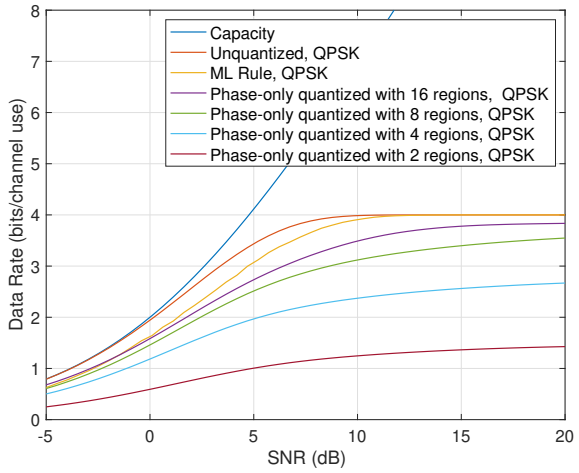


(a)

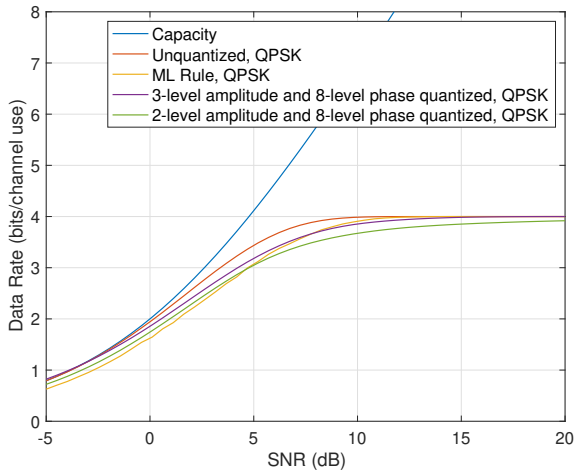


(b)

Fig. 4: Data rate versus θ for various scenarios including (a) the phase-only quantization schemes and (b) the amplitude and phase quantization schemes with different number of regions for 2×2 LoS MIMO, where SNR is $15\,$dB.



(a)



(b)

Fig. 5: Data rate versus SNR for various scenarios including (a) the phase-only quantization schemes and (b) the amplitude and phase quantization schemes with different number of regions for 2×2 LoS MIMO, where $\theta = \pi/2$.

benchmark at high enough SNR, 3-level amplitude quantization has a significant advantage at moderate SNRs, reaching unquantized performance at around 12.5 dB.

Armed with these insights, we consider quantizer design for a 4×4 system in the next section.

V. Quantization for 4×4 LoS MIMO

In this section, we investigate design of regular quantizers for a 4×4 LoS MIMO system in which the transmit and receive antennas are configured in a two-dimensional (2D) planar array as in Fig. 1b.

A. Quantizer Design

We seek to design regular quantizers which are identical for each receive antenna. For per-stream QPSK modulation, there are 4^4 possible noise-free values for the received sample at each antenna, and detailed analysis as in the 2×2 MIMO system is no longer feasible. However, as we shall see, a relatively simple approximation for the distribution of the received samples provides an effective approach for quantizer design.

The distribution functions of $\{Y_i\}_{i=1}^4$ can be expressed as

$$f_{Y_i}(y_i) = \int_0^{2\pi} \frac{1}{|\mathcal{S}|^4} \sum_{\substack{\{x_i\}_{i=1}^4 \\ x_i \in \mathcal{S}}} f(y_i \mid \phi, \{X_i = x_i\}_{i=1}^4)\, p(\phi)\, d\phi \tag{19}$$

for all $i \in \{1, \ldots, 4\}$ with

$$f(y_i \mid \phi, \{X_i = x_i\}_{i=1}^4) = \frac{1}{\pi\sigma^2} e^{-\frac{\|y_i - (\mathbf{H}\mathbf{x})_i\|^2}{\sigma^2}} \tag{20}$$

where $\mathcal{S} = \{e^{j\pi/4}, e^{j3\pi/4}, e^{j5\pi/4}, e^{j7\pi/4}\}$, $p(\phi) = 1/(2\pi)$ for $\phi \in [0, 2\pi)$ and zero otherwise by the assumption,

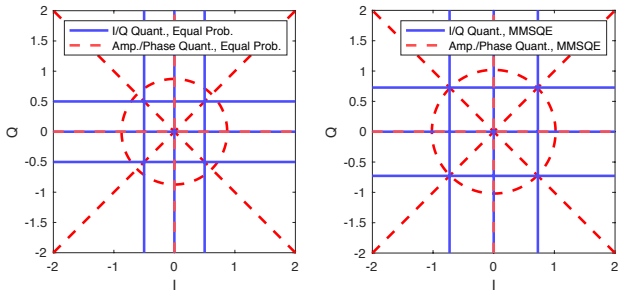


Fig. 6: Quantization schemes for 4×4 LoS MIMO at 10 dB SNR

$\mathbf{x} \triangleq [x_1\, x_2\, x_3\, x_4]^\intercal$, $\mathbf{H}$ is as in (4) with $\Phi = \phi$ and θ, and $(\mathbf{Hx})_i$ selects the ith element of $\mathbf{Hx}$.

Due to the symmetry, it is clear that the outputs before quantization (i.e., $\{Y_i\}_{i=1}^4$) have the same probability density function. Hence, without loss of generality, we focus on one of the outputs before quantization (e.g., say Y_1) to design the corresponding quantizer and employ the same quantizer for all outputs. As seen in (19), Y_1 has a complex and intractable distribution: conditioned on the common phase ϕ, it is a mixture of 4^4 Gaussians, and this conditional density then needs to be averaged over the continuum $[0, 2\pi)$ of values taken by ϕ. We therefore approximate this distribution by a circularly-symmetric complex Gaussian distribution, $\tilde{Y} \sim \mathcal{CN}(\tilde{\mu}, \tilde{\sigma}^2)$, with parameters chosen to minimize the Kullback-Leibler (KL) divergence between the distributions of Y_1 and $\tilde{Y}$, which is given by

$$D_{KL}\left(Y_1 \,\middle\|\, \tilde{Y}\right) = \int_{y \in \mathbb{C}} f_{Y_1}(y) \log\left(\frac{f_{Y_1}(y)}{f_{\tilde{Y}}(y)}\right) dy\,. \tag{21}$$

The optimal $\tilde{Y}$ that minimizes the KL divergence in (21) can be found by *moment matching* [32]: the first and second moments of $\tilde{Y}$ and Y_1 are matched to obtain the optimal Gaussian approximation. Therefore, $\tilde{\mu}$ and $\tilde{\sigma}^2$ can be calculated, respectively, as

$$\tilde{\mu} = \mathbb{E}\{Y_1\} = 0 \tag{22}$$

and

$$\tilde{\sigma}^2 = \mathbb{E}\{Y_1\, Y_1^\dagger\} \tag{23}$$

$$= \sum_{i=1}^{4} \frac{1}{4}\, \mathbb{E}\{X_i\, X_i^\dagger\} + \mathbb{E}\{N_1\, N_1^\dagger\} \tag{24}$$

$$= 1 + \sigma^2 \tag{25}$$

where $\mathbb{E}\{X_i\, X_i^\dagger\} = 1$ for all $i \in \{1, \ldots, 4\}$ by definition. As a result, we consider the complex Gaussian approximation with $\tilde{\mu} = 0$ and $\tilde{\sigma}^2 = 1 + \sigma^2$ to design quantizers for 4×4 LoS MIMO.

We consider two regular quantization schemes: I/Q quantization and amplitude/phase quantization.

- **I/Q quantization:** For I/Q quantization scheme with S^2 regions, the quantization set of $(j + S(i-1))$th-bin of a quantizer can be written as

$$\bar{\Gamma}_{j+S(i-1)} = \{\bar{Y} \mid I_{i-1} \le \Re(\bar{Y}) < I_i\,, \quad Q_{j-1} \le \Im(\bar{Y}) < Q_j\}\,, \tag{26}$$

  for $i, j \in \{1, \ldots, S\}$, where $I_1, \ldots, I_{S-1}$ and $Q_1, \ldots, Q_{S-1}$ are the thresholds for in-phase and quadrature, respectively. We set $I_0 = -\infty$, $I_S = \infty$, $Q_0 = -\infty$, and $Q_S = \infty$ in order to go along with the unified notation.
- **Amplitude/phase quantization:** As for 2×2 LoS MIMO, the quantization set for this scheme is specified as in (18).

We determine the quantizer regions (i.e., the thresholds in (18) and (26)) based on the following two different metrics:

- **Minimum mean squared quantization error (MMSQE)-based regions:** This is the conventional approach to quantizer design based on minimization of the mean squared error given by

$$\mathbb{E}\{(\tilde{Y} - Q(\tilde{Y}))^2\} \tag{27}$$

  where $Q(\cdot)$ is the quantizer function, whose set is defined as either (18) or (26). The optimal decision boundaries in (18) and (26) are obtained as usual, by applying the Lloyd-Max algorithm [33], [34].
- **Equal probability-based regions:** The quantizer regions here are obtained by partitioning the fitted complex Gaussian distribution into equal probability regions. In other words, the quantizer boundaries maximize the entropy of $\tilde{Y}$; that is, $H(\tilde{Y})$. For the circular Gaussian distribution, the boundaries can be specified analytically. For I/Q quantization, the entropy-maximizer thresholds can be calculated as

$$I_i = Q_i = \tilde{\mu} + \frac{\tilde{\sigma}}{\sqrt{2}}\Phi^{-1}\left(\frac{i}{S}\right) = \sqrt{\frac{1+\sigma^2}{2}}\Phi^{-1}\left(\frac{i}{S}\right) \tag{28}$$

  for $i \in \{1, \ldots, S-1\}$, where Φ^{-1} is the inverse distribution function (i.e., the quantile function) for the standard Gaussian distribution with a mean of 0 and a standard deviation of 1. For amplitude/phase quantization, the phase quantization is uniform, and the amplitude thresholds that maximize the entropy can be found as

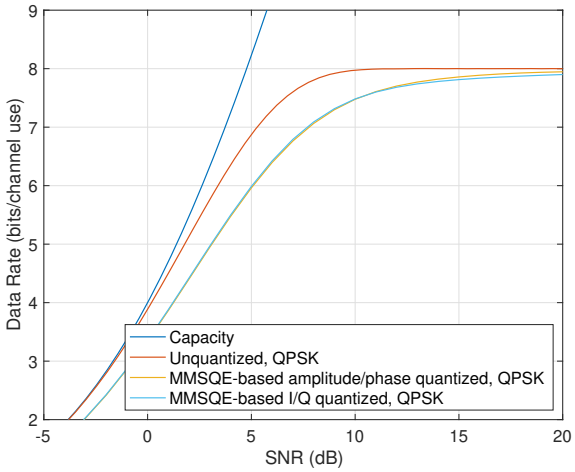
$$A_i = \sqrt{(1+\sigma^2)\log\left(\frac{K}{K-i}\right)} \tag{29}$$

  for $i \in \{1, \ldots, K-1\}$.

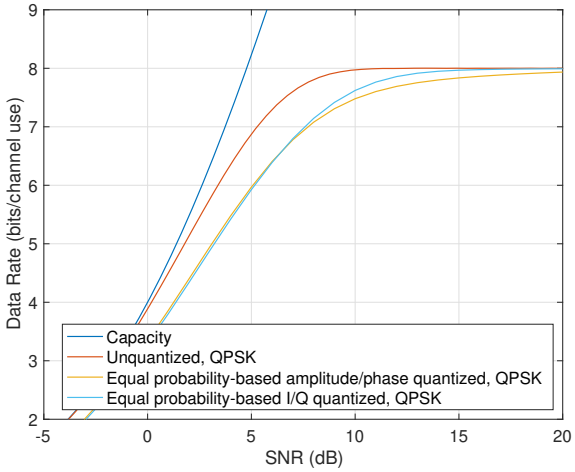
For those two different metrics, Fig. 6 shows the I/Q and amplitude/phase quantizers at 10 dB SNR, each having a total of 16 regions (i.e., $K = 2$, $M = 8$, and $S = 4$).

B. Numerical Results

We now investigate Shannon limits for different quantizer designs; in each case, there are 16 quantization regions in the complex plane (see Appendix A for details). Fig. 7a shows



(a)



(b)

Fig. 7: Data rate versus SNR for (a) MMSQE-based quantization and (b) equal probability-based quantization for a 4×4 system.

that, for conventional MMSQE-based quantization, amplitude-phase and I/Q quantization are both unable to reach the maximum data rate (8 bits per channel use) even at SNR as high as 20 dB, compared to the unquantized system, which saturates at 10 dB SNR. On the other hand, Fig. 7b shows that, for equal probability regions, I/Q quantization attains the maximum data rate of 8 bits per channel use at around 15 dB, while amplitude/phase quantization continues to exhibit a gap to the unquantized limit even at SNR of 20 dB. We conclude that 2 bit quantization on I and Q based on equal probability regions should suffice to attain acceptable performance at moderate SNR, and focus on this setting for investigation of spatial demultiplexing algorithms in the next section.

VI. Spatial Demultiplexing under Severe Quantization

We begin with spatial demultiplexing for 4×4 LoS MIMO system with QPSK modulation with 2 bit I/Q quantization as designed in the previous section. In order to highlight the impact of quantization, consider an ideally conditioned channel with $\theta = \frac{\pi}{2}$ in (4), for which the received antenna responses for different transmitted streams are orthogonal, so that matched filter, linear ZF and maximum likelihood detection all yield the same performance with *unquantized* observations. We shall see, however, that drastic quantization can have a severe impact on the performance of linear detection even in such an ideal setting because of the common channel phase Φ in (4), which can move the observations close to quantization boundaries. On the other hand, the mutual information plot in Figure 7b shows that reliable communication at the maximum data rate of 8 bits per channel use should be possible at an SNR of about 15 dB in this setting. Our goal, therefore, is to devise spatial demultiplexing with reasonable complexity that can approach this performance. We consider an uncoded BER target of about 10^{-3}, which yields reliable communication with high-rate bit interleaved coded modulation, since a binary symmetric channel with this cross-over probability has capacity close to 0.99 bits per channel use.

From the point of view of minimizing the probability of error, the optimal detector based on the quantized output is the maximum likelihood (ML) detector [35], given by

$$\hat{\mathbf{X}}(\mathbf{Y}_Q) = \underset{\mathbf{X} \in \mathcal{S}^{N_R}}{\operatorname{argmax}}\, p(\mathbf{Y}_Q \mid \mathbf{X})\,. \tag{30}$$

The ML detector in (30) selects the best vector from a set of all possible transmitted vectors, which maximizes the conditional probability density function of the quantized observation given the transmitted vector. Since the minimization in (30) is over all possible transmitted vectors, the problem in (30) has prohibitive complexity. As a low-complexity alternative, we consider linear ZF detection. We have also verified by simulations that linear minimum mean squared error (MMSE) detector [36] and sphere decoding [37] also achieve the same performance as linear ZF with quantized observations for the well-conditioned LoS MIMO channel considered here.

Recall that the ZF solution minimizes

$$\tilde{\mathbf{X}}(\mathbf{Y}_Q) = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{Y}_Q - \mathbf{H}\mathbf{X}\| \tag{31}$$

and can be obtained by pre-multiplying $\mathbf{Y}_Q$ by the Moore-Penrose pseudo-inverse of $\mathbf{H}$ to compute

$$\tilde{\mathbf{X}}(\mathbf{Y}_Q) = (\mathbf{H}^\dagger \mathbf{H})^{-1} \mathbf{H}^\dagger \mathbf{Y}_Q \tag{32}$$

first, and then finding ith element of $\hat{\mathbf{X}}(\mathbf{Y}_Q)$ as

$$\hat{X}_i(\mathbf{Y}_Q) = \underset{X_i \in \mathcal{S}}{\operatorname{argmin}} |\tilde{X}_i(\mathbf{Y}_Q) - X_i| \tag{33}$$

for all $i \in \{1, \ldots, 4\}$, where $\tilde{X}_i(\cdot)$ is the ith element of $\tilde{\mathbf{X}}(\cdot)$. Note that $\mathbf{Y}_Q$ in (32) represents the quantized output at the receiver and the quantizer outputs are set to the centroids of the quantizer regions, which are obtained based on the complex Gaussian approximation. Mathematically, the quantized output of ith antenna for the received signal Y_i is equal to

$$\bar{Y}_i = \frac{\int_{\tilde{\Gamma}_i} y f_{\tilde{Y}}(y)\, dy}{\int_{\tilde{\Gamma}_i} f_{\tilde{Y}}(y)\, dy}\,, \tag{34}$$

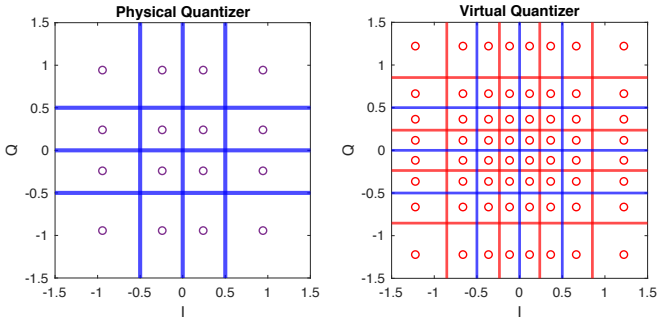


Fig. 8: Physical and virtual quantizers with the outputs corresponding to the centroids of the regions for 10 dB SNR.

where $\tilde{\Gamma}_i$ denotes the quantizer region that Y_i falls into; that is, $Y_i \in \tilde{\Gamma}_i$.

A. Virtual Quantization

Linear ZF detection based on the centroids codebook performs poorly when the unquantized outputs are far from the centroids of the regions that they belong to. On the other hand, we know that linear ZF detection with unquantized outputs yields excellent performance for a well-conditioned MIMO channel. This motivates viewing the unquantized output vector as a hidden variable, or nuisance parameter, for our hypothesis testing problem of estimating the transmitted symbols. We can now apply any of the standard tools of composite hypothesis testing to this problem. We choose here a Generalized Likelihood Ratio Test (GLRT) approach, in which we jointly estimate the unquantized output and the transmitted symbols given the quantized outputs. That is, the transmitted symbols are estimated by jointly maximizing the probability of unquantized output and transmitted symbols given the quantized output:

$$(\hat{\mathbf{X}}(\mathbf{Y}_Q), \mathbf{Y}(\mathbf{Y}_Q)) = \underset{\mathbf{X}\in\mathcal{S}^{N_R},\mathbf{Y}}{\operatorname{argmax}}\; p(\mathbf{X}, \mathbf{Y} \mid \mathbf{Y}_Q)\,. \tag{35}$$

A key difficulty in this optimization problem is that we need to search over a continuum of values for the hidden unquantized output $\mathbf{Y}$. We therefore consider a grid-based approximation of $\mathbf{Y}$, where the grid is finer than that provided by the quantizer. We term this approach *virtual quantization.*

Let $Q_v(\cdot)$ denote the element-wise virtual quantizer function. The virtual quantized hidden variable can be obtained as $\mathbf{Y}_V = Q_v(\mathbf{Y})$. Then, the quantized output $\mathbf{Y}_Q$ can be written as $\mathbf{Y}_Q = Q_c(\mathbf{Y}_V)$ as a coarsening of virtual quantized hidden variable, where $Q_c(\cdot)$ is the coarsening function. Thus, $\mathbf{Y}_Q = Q_c(Q_v(\mathbf{Y})) = Q(\mathbf{Y})$, where $Q(\cdot)$ is the element-wise actual quantizer function defined in (7) for the ith receive antenna and $Q_c(\cdot)$ can be considered as $Q_c(\cdot) = Q(\cdot)$. Thus, $\mathbf{X} \to \mathbf{Y} \to \mathbf{Y}_V \to \mathbf{Y}_Q$ is a Markov chain. Figure 8 shows an example of physical and virtual quantizers used in our numerical results (detailed description is provided later in this section).

Quantizing the nuisance parameter $\mathbf{Y}$ in the joint estimation problem in (35) using the virtual quantizer, we seek to jointly estimate the transmitted symbols and the virtual quantizer outputs:

$$(\hat{\mathbf{X}}(\mathbf{Y}_Q), \mathbf{Y}(\mathbf{Y}_Q)) = \underset{\mathbf{X}\in\mathcal{S}^{N_R},\mathbf{Y}}{\operatorname{argmax}}\; p(\mathbf{X}, Q_v(\mathbf{Y}) \mid \mathbf{Y}_Q) \tag{36}$$

$$= \underset{\mathbf{X}\in\mathcal{S}^{N_R},\mathbf{Y}_V\in\mathcal{T}}{\operatorname{argmax}}\; p(\mathbf{X}, \mathbf{Y}_V \mid \mathbf{Y}_Q) \tag{37}$$

where (37) follows from $\mathbf{Y}_V = Q_v(Y)$ and $\mathcal{T} = \mathcal{T}(\mathbf{Y}_Q)$ is a discrete set including all possible combinations of virtual quantized output for observed $\mathbf{Y}_Q$:

$$\mathcal{T} = \{\mathbf{Y}_V = Q_v(\mathbf{Y}) \mid Q(\mathbf{Y}_V) = \mathbf{Y}_Q\}$$

In order to solve the problem in (37) and estimate the transmitted symbols, we maximize the objective function in (37) with respect to $\mathbf{X}$ for a given $\mathbf{Y}_V \in \mathcal{T}$ first, then substitute the obtained $\mathbf{X}$ to the objective function to solve for $\mathbf{Y}_V$. To begin with, for a given $\mathbf{Y}_V \in \mathcal{T}$, the maximization of the optimization problem in (37) over $\mathbf{X}$ can be expressed as follows:

$$\hat{\mathbf{X}}(\mathbf{Y}_V, \mathbf{Y}_Q) = \underset{\mathbf{X}\in\mathcal{S}^{N_R}}{\operatorname{argmax}}\; p(\mathbf{X}, \mathbf{Y}_V \mid \mathbf{Y}_Q) \tag{38}$$

$$= \underset{\mathbf{X}\in\mathcal{S}^{N_R}}{\operatorname{argmax}}\; p(\mathbf{X} \mid \mathbf{Y}_V, \mathbf{Y}_Q)\, p(\mathbf{Y}_V \mid \mathbf{Y}_Q) \tag{39}$$

$$= \underset{\mathbf{X}\in\mathcal{S}^{N_R}}{\operatorname{argmax}}\; p(\mathbf{X} \mid \mathbf{Y}_V)\, p(\mathbf{Y}_V \mid \mathbf{Y}_Q) \tag{40}$$

$$= \underset{\mathbf{X}\in\mathcal{S}^{N_R}}{\operatorname{argmax}}\; p(\mathbf{X} \mid \mathbf{Y}_V) \tag{41}$$

$$= \underset{\mathbf{X}\in\mathcal{S}^{N_R}}{\operatorname{argmax}}\; p(\mathbf{Y}_V \mid \mathbf{X}) \tag{42}$$

where (39) is a standard conditional probability computation, (40) is based on the Markov property $\mathbf{X} \to \mathbf{Y}_V \to \mathbf{Y}_Q$, (41) is obtained since $p(\mathbf{Y}_V \mid \mathbf{Y}_Q)$ does not depend on $\mathbf{X}$ for given $\mathbf{Y}_V$ and $\mathbf{Y}_Q$, and (42) follows from Bayes' theorem and equally likely transmitted symbols.

The optimization problem in (42) is in the form of (30), except that it considers virtual quantized hidden variable instead of actual quantized output. As described in (32) and (33), (42) can be approximated by linear ZF detection. Therefore, for given $\mathbf{Y}_V$ and observed quantized output $\mathbf{Y}_Q$, $\hat{\mathbf{X}}(\mathbf{Y}_V, \mathbf{Y}_Q)$ can be obtained as in (32) and (33).

Next, we substitute $\hat{\mathbf{X}}(\mathbf{Y}_V, \mathbf{Y}_Q)$ to the optimization problem in (37) and maximize it over $\mathbf{Y}_V$. Substituting $\hat{\mathbf{X}}(\mathbf{Y}_V, \mathbf{Y}_Q)$ to (37),

$$\hat{Y}_v(\hat{\mathbf{X}}, \mathbf{Y}_Q) = \underset{\mathbf{Y}_V\in\mathcal{T}}{\operatorname{argmax}}\; p(\hat{\mathbf{X}}, \mathbf{Y}_V \mid \mathbf{Y}_Q) \tag{43}$$

$$= \underset{\mathbf{Y}_V\in\mathcal{T}}{\operatorname{argmax}}\; p(\mathbf{Y}_V \mid \hat{\mathbf{X}}, \mathbf{Y}_Q)\, p(\hat{\mathbf{X}} \mid \mathbf{Y}_Q) \tag{44}$$

$$\approx \underset{\mathbf{Y}_V\in\mathcal{T}}{\operatorname{argmin}}\; \left\|\mathbf{Y}_V - \bar{\mathbf{Y}}_V\right\|^2 \tag{45}$$

where (44) follows from Bayes' theorem and (45) is obtained by approximating the impact of quantization and noise as Gaussian and modeling $p(\mathbf{Y}_V \mid \hat{\mathbf{X}}, \mathbf{Y}_Q)$ as Gaussian with mean $\bar{\mathbf{Y}}_V$, where $\bar{\mathbf{Y}}_V$ denotes the noiseless reconstruction and is equal to $\bar{\mathbf{Y}}_V = Q_v(\mathbf{H}\hat{\mathbf{X}})$. Note that we ignore $p(\hat{\mathbf{X}} \mid \mathbf{Y}_Q)$ term in (45) even though $\hat{\mathbf{X}}$ depends on $\mathbf{Y}_V$, which means that we are not necessarily attaining the true maximum for (43). We expect the impact of this approximation on (45) to

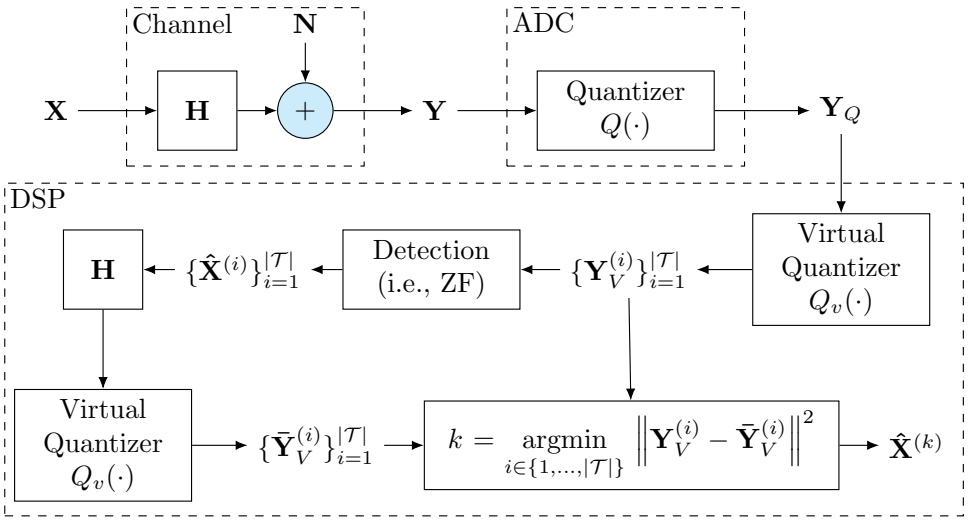


Fig. 9: Flow diagram of virtual quantization based spatial demultiplexing.

be small compared to the term $p(\mathbf{Y}_V \mid \hat{\mathbf{X}}, \mathbf{Y}_Q)$, which decays exponentially according to our Gaussian approximation for the sum of the virtual quantization noise and thermal noise.

The proposed virtual quantization method can now be summarized as follows:

1) Generate set $\mathcal{T}$ based on the physical quantized observation and the virtual quantizer function such that $\mathcal{T} = \{\mathbf{Y}_V = Q_v(\mathbf{Y}) \mid Q(\mathbf{Y}_V) = \mathbf{Y}_Q\}$.
2) For each $\mathbf{Y}_V \in \mathcal{T}$, calculate the ZF solution to obtain $\hat{\mathbf{X}}(\mathbf{Y}_V, \mathbf{Y}_Q)$ by treating $\mathbf{Y}_V$ as if it is the observed output at the receiver.
3) Find $\mathbf{Y}_V \in \mathcal{T}$ that minimizes the Euclidean distance between $\mathbf{Y}_V$ and $Q_v(\mathbf{H}\hat{\mathbf{X}})$; that is, $\|\mathbf{Y}_V - Q_v(\mathbf{H}\hat{\mathbf{X}})\|^2$.
4) Declare the corresponding $\hat{\mathbf{X}}(\mathbf{Y}_V, \mathbf{Y}_Q)$ as the estimated symbol vector.

Fig. 9 illustrates this procedure via a flow diagram.

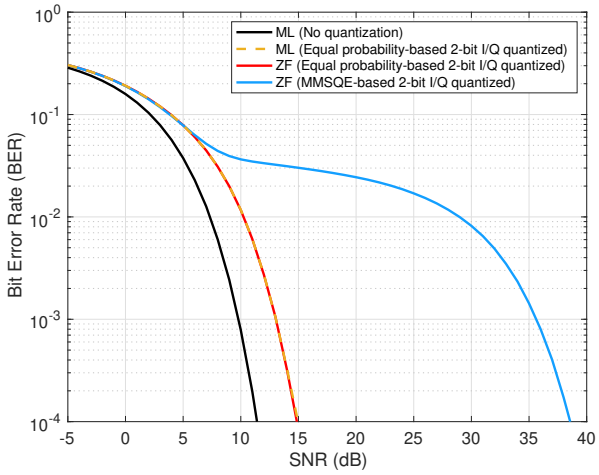
The result of the estimation depends on the virtual quantization function (i.e., $Q_v(\cdot)$) and the virtual quantization output (i.e., $\mathbf{Y}_V$), both of which determine the elements in $\mathcal{T}$ and contribute to the cost function used to specify the estimated symbols. Different virtual quantization functions can be designed for our proposed method. However, in this paper, we consider the same method that we use for the design of the actual physical quantizer. Considering the Gaussian approximation discussed in the design of the actual quantizer, we design an I/Q quantization-based virtual quantizer having equal-probability regions. For the 2-bit per I/Q physical quantizer having 16 regions, we consider a virtual quantizer with 64 regions (i.e., $S = 8$) whose outputs are decided based on the centroids of the corresponding regions similar to those of the physical quantizer, as shown in Fig. 8. In this case, the virtual quantizer divides each physical quantizer bin into 4 regions, which can be considered as a virtually created 1-bit quantizer per each I/Q for each physical quantization bin.

B. Numerical Results

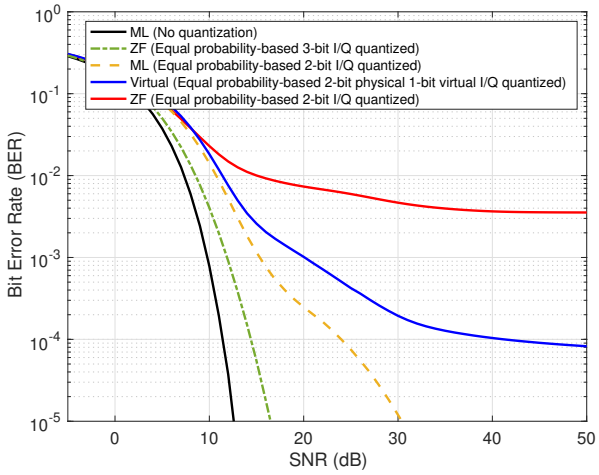
We consider two benchmarks: ML with no quantization and ML with quantization. The former assumes there is no quantization in the system; that is $\mathbf{Y}_Q = \mathbf{Y}$ and provides an unquantized benchmark as if ADCs at the receiver have infinite precision. The latter considers the quantized outputs at the receiver and is obtained based on (30).

For a 4×4 MIMO system, Fig. 10a plots bit error rate (BER) versus SNR, setting $\theta = \pi/2$, for different detection methods when the common phase is fixed to $\Phi = 0$. We may also view this as equivalent to a hybrid analog-digital processing scheme in which the common channel phase Φ is removed by analog derotation *prior* to quantization. The plot indicates that linear ZF detection with equal probability quantization and the centroids codebook remains near-optimal, achieving the same performance as ML reception, as long as the common channel phase is removed prior to quantization. On the other hand, as expected from our mutual information computations, MMSQE-based quantization performs significantly worse than the equal probability-based quantizer. Note that we have simulated other detection methods such as the linear MMSE detector and the sphere decoder based on the quantized outputs at the receiver, and have verified that they achieve the same performance as linear ZF detection.

We now turn to the scenario of interest for us: fully digital processing without derotation of the common phase Φ prior to quantization. The receiver processing with quantized observations does employ knowledge of Φ, but performance is adversely affected by points being rotated close to quantization boundaries. Fig. 10b plots BER averaged over the common phase Φ versus SNR, setting $\theta = \pi/2$, for different detection methods, where Φ is uniformly distributed over $[0, 2\pi)$. We now see that linear ZF detection with the centroids codebook performs significantly worse, with an error floor that stays higher than our target 10^{-3} BER. The proposed virtual quantization approach performs significantly better: its performance is close to that of ML detection at BER of 10^{-2}, while being 5 dB worse at the target BER of 10^{-3}. It still exhibits an error floor at 10^{-4}, motivating additional effort in devising low-complexity strategies for approaching maximum likelihood performance.



(a)



(b)

Fig. 10: Bit error rate versus SNR for different detection methods when the common phase is (a) fixed to $\Phi = 0$; (b) uniformly distributed over $[0, 2\pi)$, for a 4×4 MIMO system with QPSK.

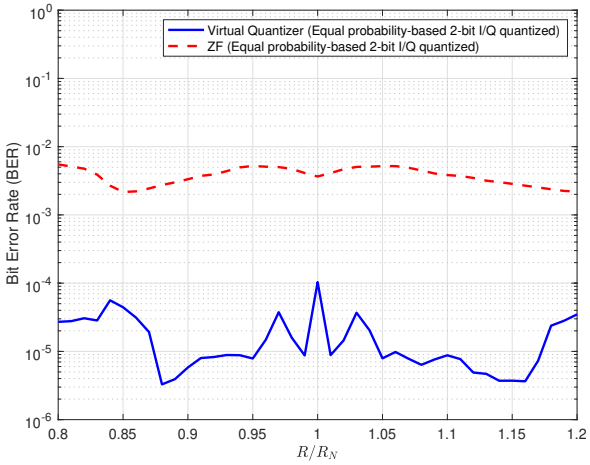


Fig. 11: Bit error rate averaged over Φ versus R/R_N for different detection methods for a 4×4 MIMO system with QPSK, where SNR is $40\,\text{dB}$.

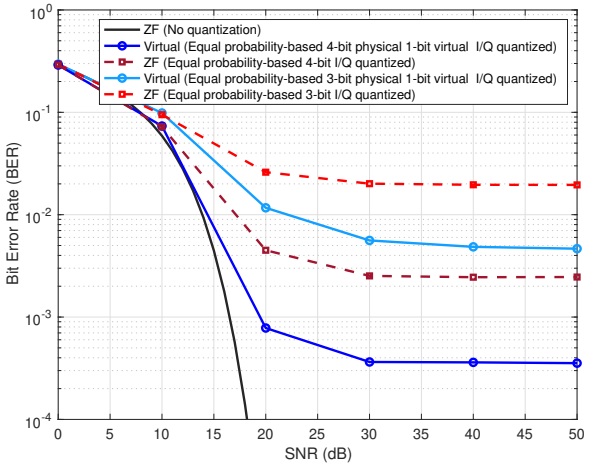


Fig. 12: Bit error rate averaged over Φ versus SNR for different detection methods including our proposed virtual quantization approach when the common phase is uniformly distributed over $[0, 2\pi)$ for a 4×4 MIMO system with 16QAM.

Beyond the ideal model: Since severe quantization destroys the orthogonality of the received signals for different data streams, we expect that our all-digital receiver should be robust to changes in link distance R around the nominal range R_N. Fig. 11 plots BER vs R/R_N at $40\,\text{dB}$ SNR for our proposed virtual quantization approach and linear ZF detection with the centroids codebook, where $R \in [0.8R_N, 1.2R_N]$. The poor performance of linear ZF detection also persists as we vary R.

Scaling to larger constellations: A key advantage of virtual quantization is that its complexity does not scale with constellation size. We verify this by evaluating performance for 16QAM modulation. Given the higher dynamic range of 16QAM, we consider 3 bit and 4 bit I/Q physical quantization, and then add 1 bit virtual quantization as before. Fig. 12 plots BER averaged over the common phase Φ versus SNR, setting $\theta = \pi/2$, for different detection methods. We see that neither 4 bit physical quantization (without virtual quantization) nor 3 bit physical quantization with 1 bit virtual quantization achieve our target 10^{-3} BER even at very high SNR. However, 4 bit physical quantization with 1 bit virtual quantization does achieve our BER target at approximately $20\,\text{dB}$ SNR.

VII. Conclusion

Our study of ideal 4×4 LoS MIMO system at high SNR yields fundamental insight into the impact of severe quantization. We show that equal probability quantization, which maximizes per-antenna output entropy, outperforms standard MMSQE quantization in such regimes. For spatial demultiplexing, we introduce the novel concept of virtual quantization, which may be viewed as approximate joint estimation of

the transmitted symbols and the unquantized received signal, and show that linear detection with virtual quantization is an effective low-complexity alternative to maximum likelihood detection, which requires complexity exponential in the number of transmitted bits. It remains an open issue as to whether the gap to maximum likelihood detection at higher SNR can be further reduced, and the error floor eliminated, while maintaining reasonable complexity. In that regard, the use of look-up tables for reducing the complexity is an interesting direction to explore [23].

An important direction for future research is to investigate quantization-constrained LoS MIMO in more complex settings, including understanding the impact of dispersion due to geometric misalignments and potential performance advantages of spatial oversampling [11]. While we do not consider transmit precoding here, joint transmit-receive optimization subject to dynamic range constraints and nonlinearities at both ends is of great interest. At a fundamental level, the concept of virtual quantization, which treats the unquantized output as a hidden variable, may be worth exploring for other system models.

APPENDIX A
DATA RATES FOR DIFFERENT QUANTIZATION SCHEMES

The data rates shown in Fig. 4, Fig. 5, and Fig. 7 are obtained by calculating

$$\mathbb{E}_{\Phi}\{I(\mathbf{X};\mathbf{Y}_Q \mid \Phi,\theta)\}, \tag{46}$$

for given SNR, θ, and quantization scheme. The mutual information between $\mathbf{X}$ and $\mathbf{Y}_Q$ in (46) can be given as

$$\begin{aligned} &I(\mathbf{X};\mathbf{Y}_Q \mid \Phi=\phi,\theta) \\ &\quad= \sum_{\mathbf{x}\in\mathcal{X}} \sum_{\mathbf{y}_Q\in\mathcal{Y}_{\mathcal{Q}}} p(\mathbf{x})\, p(\mathbf{y}_Q \mid \mathbf{x}) \log_2 \frac{p(\mathbf{y}_Q \mid \mathbf{x})}{p(\mathbf{y}_Q)} \end{aligned} \tag{47}$$

where set $\mathcal{X}$ and set $\mathcal{Y}_{\mathcal{Q}}$ include all possible transmit vectors and quantized output vectors, respectively, and $p(\mathbf{x}) = 1/|\mathcal{X}|$. Based on the quantization scheme, $p(\mathbf{y}_Q \mid \mathbf{x})$ can be calculated as

$$p(\mathbf{y}_Q \mid \mathbf{x}) = \int_{Q(\mathbf{y})=\mathbf{y}_Q} f(\mathbf{y} \mid \mathbf{x})\, d\mathbf{y} \tag{48}$$

$$= \int_{Q(\mathbf{y})=\mathbf{y}_Q} \prod_i f(y_i \mid \mathbf{x})\, dy_i \tag{49}$$

with

$$f(y_i \mid \mathbf{x}) = \frac{1}{\pi\sigma^2} e^{-\frac{\|y_i-(\mathbf{H}\mathbf{x})_i\|^2}{\sigma^2}}, \tag{50}$$

where (49) follows from the fact that y_i's are independent given $\mathbf{x}$ and $\mathbf{H}$ denote the channel matrix described in the manuscript for given $\Phi=\phi$ and θ. We note that $p(\mathbf{y}_Q \mid \mathbf{x})$ is also obtained by performing Monte Carlo simulation.

REFERENCES

[1] T. S. Rappaport, Y. Xing, O. Kanhere, S. Ju, A. Madanayake, S. Mandal, A. Alkhateeb, and G. C. Trichopoulos, "Wireless communications and applications above 100 Ghz: Opportunities and challenges for 6G and beyond," *IEEE Access*, vol. 7, pp. 78 729–78 757, 2019.

[2] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[3] H. Elayan, O. Amin, B. Shihada, R. M. Shubair, and M.-S. Alouini, "Terahertz band: The last piece of RF spectrum puzzle for communication systems," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 1–32, 2020.

[4] E. Torkildson, U. Madhow, and M. Rodwell, "Indoor millimeter wave MIMO: Feasibility and performance," *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4150–4160, Dec. 2011.

[5] M. Eslami Rasekh, D. Guo, and U. Madhow, "Joint routing and resource allocation for millimeter wave picocellular backhaul," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 783–794, 2020.

[6] B. Murmann. ADC Performance Survey 1997-2020. [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html.

[7] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[8] J. Singh and U. Madhow, "On block noncoherent communication with low-precision phase quantization at the receiver," in *2009 IEEE International Symposium on Information Theory*, 2009, pp. 2199–2203.

[9] F. Bohagen, P. Orten, and G. E. Oien, "Design of optimal high-rank line-of-sight MIMO channels," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1420–1425, 2007.

[10] Ericsson Press Release, "Deutsche Telekom and Ericsson top 100Gbps over microwave link," May 2019. [Online]. Available: https://www.ericsson.com/en/press-releases/2019/5/deutsche-telekom-and-ericsson-top-100gbps-over-microwave-link

[11] M. Sawaby, B. Mamandipoor, U. Madhow, and A. Arbabian, "Analog processing to enable scalable high-throughput mm-Wave wireless fiber systems," in *50th Asilomar Conference on Signals, Systems and Computers*, Nov. 2016, pp. 1658–1662.

[12] Y. Yan, P. Bondalapati, A. Tiwari, C. Xia, A. Cashion, D. Zhang, Q. Tang, and M. Reed, "11-Gbps broadband modem-agnostic line-of-sight MIMO over the range of 13 km," in *IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–7.

[13] A. Khalili, S. Rini, L. Barletta, E. Erkip, and Y. C. Eldar, "On MIMO channel capacity with output quantization constraints," in *IEEE International Symposium on Information Theory (ISIT)*, June 2018, pp. 1355–1359.

[14] L. Zhu, S. Wang, and J. Zhu, "Adaptive beamforming design for millimeter-wave line-of-sight MIMO channel," *IEEE Communications Letters*, vol. 23, no. 11, pp. 2095–2098, 2019.

[15] K. Kobayashi, T. Ohtsuki, and T. Kaneko, "Precoding for MIMO systems in line-of-sight (LOS) environment," in *IEEE Global Telecommunications Conference*, 2007, pp. 4370–4374.

[16] L. Zhou and Y. Ohashi, "Low complexity millimeter-wave LOS-MIMO precoding systems for uniform circular arrays," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2014, pp. 1293–1297.

[17] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[18] A. Khalili, S. Shahsavari, F. Shirani, E. Erkip, and Y. C. Eldar, "On throughput of millimeter wave MIMO systems with low resolution ADCs," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5255–5259.

[19] O. Dabeer and U. Madhow, "Channel estimation with low-precision analog-to-digital conversion," in *IEEE International Conference on Communications (ICC)*, May 2010, pp. 1–6.

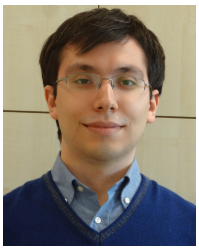
[20] C. Stöckle, J. Munir, A. Mezghani, and J. A. Nossek, "Channel estimation in massive MIMO systems using 1-bit quantization," in *IEEE 17th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, July 2016, pp. 1–6.

[21] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *48th Asilomar Conference on Signals, Systems and Computers*, 2014, pp. 957–961.

[22] A. Mezghani, M. Khoufi, and J. A. Nossek, "Maximum likelihood detection for quantized MIMO systems," in *2008 International ITG Workshop on Smart Antennas*, 2008, pp. 278–284.

[23] F. A. Monteiro and I. J. Wassell, "Euclidean distances in quantized spaces with pre-stored components for MIMO detection," in *2007 European Conference on Wireless Technologies*, 2007, pp. 150–153.

[24] A. Mezghani and J. A. Nossek, "Belief propagation based MIMO detection operating on quantized channel output," in *IEEE International Symposium on Information Theory*, 2010, pp. 2113–2117.
[25] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Transactions on Communications*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.
[26] A. D. Sezer and U. Madhow, "Near-optimal quantization for LoS MIMO with QPSK modulation," in *53rd Asilomar Conference on Signals, Systems, and Computers*, 2019, pp. 1015–1020.
[27] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Transactions on Communications*, vol. 61, no. 10, pp. 4391–4403, 2013.
[28] R. Zhang, H. Zhang, W. Xu, and X. You, "Subarray-cooperation-based multi-resolution codebook and beam alignment design for mmwave backhaul links," *IEEE Access*, vol. 7, pp. 18 319–18 331, 2019.
[29] P. F. Driessen and G. J. Foschini, "On the capacity formula for multiple input-multiple output wireless channels: a geometric interpretation," *IEEE Transactions on Communications*, vol. 47, no. 2, pp. 173–176, 1999.
[30] F. Bohagen, P. Orten, and G. E. Oien, "On spherical vs. plane wave modeling of line-of-sight MIMO channels," *IEEE Transactions on Communications*, vol. 57, no. 3, pp. 841–849, 2009.
[31] M. H. Castañeda Garcia, M. Iwanow, and R. A. Stirling-Gallacher, "LOS MIMO design based on multiple optimum antenna separations," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 2018, pp. 1–5.
[32] C. E. Rasmussen, *Gaussian Processes in Machine Learning*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 63–71. [Online]. Available: https://doi.org/10.1007/978-3-540-28650-9_4
[33] S. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.
[34] J. Max, "Quantizing for minimum distortion," *IRE Transactions on Information Theory*, vol. 6, no. 1, pp. 7–12, 1960.
[35] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 1941–1988, 2015.
[36] M. L. Honig, *Advances in Multiuser Detection*. John Wiley & Sons, Ltd, 2009.
[37] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.



Ahmet Dundar Sezer received the B.S., M.S., and Ph.D. degrees in Electrical and Electronics Engineering from Bilkent University, Ankara, Turkey, in 2011, 2013, and 2018, respectively. He is currently a Post-Doctoral Researcher at the University of California, Santa Barbara, CA, USA. His current research interests include signal processing, wireless communications, and optimization.

Upamanyu Madhow is Professor of Electrical and Computer Engineering at the University of California, Santa Barbara. His current research interests focus on next generation communication, sensing and inference infrastructures centered around millimeter wave systems, and on robust machine learning. He received his bachelor's degree in electrical engineering from the Indian Institute of Technology, Kanpur, in 1985, and his Ph. D. degree in electrical engineering from the University of Illinois, Urbana-Champaign in 1990. He has worked as a research scientist at Bell Communications Research, Morristown, NJ, and as a faculty at the University of Illinois, Urbana-Champaign. Dr. Madhow is a recipient of the 1996 NSF CAREER award, and co-recipient of the 2012 IEEE Marconi prize paper award in wireless communications. He has served as Associate Editor for the IEEE Transactions on Communications, the IEEE Transactions on Information Theory, and the IEEE Transactions on Information Forensics and Security. He is the author of two textbooks published by Cambridge University Press, *Fundamentals of Digital Communication* (2008) and *Introduction to Communication Systems* (2014).