The Cost of Noncoherence: Avoiding Channel Estimation Through Differential Encoding in Phase Quantized Systems

SAMIRU GAYAN [1] **(Member, IEEE), HAZER INALTEKIN** [2] **(Member, IEEE),**
RAJITHA SENANAYAKE [3] **(Member, IEEE), AND JAMIE EVANS** [3] **(Senior Member, IEEE)**

[1] Department of Electronic and Telecommunication Engineering, University of Moratuwa, Katubedda 10400, Sri Lanka

[2] School of Engineering, Macquarie University, North Ryde, NSW 2109, Australia

[3] Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3100, Australia

CORRESPONDING AUTHOR: S. GAYAN (e-mail: samirug@uom.lk)

ABSTRACT This paper proposes a novel approach that utilizes differential encoding to overcome the channel estimation problem in communication systems with low-resolution quantization receivers. For differentially encoded data, we derive the maximum likelihood detection rule for the canonical block-2 detectors, employing just two consecutive quantized observations at the channel output and without any receiver-side channel state information. We establish the optimality of this maximum likelihood detection rule within the class of block-L detectors, where $L \geq 3$, under the condition that $n = \log_2 M$, with n and M denoting the number of quantization bits and input alphabet size, respectively. The derived detector has a simple and easily implementable structure, comparing the quantization region indices of consecutive observations to determine the transmitted message index. By leveraging the structure of the derived optimum detector, we obtain the expression for the message error probability in Rayleigh fading wireless channels. Through asymptotic analysis in the high signal-to-noise ratio regime, we reveal a crucial finding that achieving the same diversity order as infinite bit quantization with full channel knowledge requires an additional two bits at the quantizer, in addition to the minimum requirement of $\log_2 M$ bits. One bit compensates for the low-resolution effect, while the other addresses the lack of channel knowledge. Finally, we conduct an extensive simulation study to demonstrate the performance of the optimum detectors and quantify the performance loss resulting from the absence of channel knowledge at the receiver.

INDEX TERMS Low-resolution quantization, ML detectors, D-MPSK modulation, symbol error probability, diversity order.

I. INTRODUCTION
A. BACKGROUND AND MOTIVATION

MILLIMETER Wave (mmWave) and Terahertz (THz) communications have recently gained substantial attention as pivotal technologies in next-generation wireless networks [1], [2], [3]. These systems employ significantly broader bandwidths and extensive antenna arrays at both transmitter and receiver ends to facilitate beamforming and spatial multiplexing [4]. Consequently, the utilization of high-resolution analog-to-digital converters (ADCs) with approximately 12-16 bits, coupled with high sampling rates, results in a considerable power consumption burden at both

the transmitter and receiver components of these systems [5]. This is primarily due to the fact that the power consumption of ADCs increases exponentially with their resolution level, and linearly with their sampling rate [6], [7].

To address this challenge, there are two potential solutions: the utilization of high-resolution but low-speed sub-ADCs or the deployment of high-speed but low-resolution ADCs [5]. Given that the use of a high-resolution, low-speed sub-ADC architecture can introduce error floors due to discrepancies among sub-ADCs [8], current research in this domain has predominantly focused on adopting low-resolution ADCs. Consequently, the concept of substituting

power-intensive high-resolution ADCs with low-resolution counterparts offers a promising solution for power consumption concerns in upcoming wireless applications such as mmWave [4] and THz communications [9], integrated sensing and communication systems [10], [11], UAV deployments, and Low Earth Orbit (LEO) satellites. These application areas exhibit either high energy consumption due to demanding sampling rates and extensive antenna arrays (as seen in massive multiple-input multiple-output (MIMO)) or inherent power limitations when serving a substantial number of users dispersed over diverse geographical locations (as observed in UAVs and LEO satellites).

To achieve high data rates in quantized systems, the precise estimation of channel state information (CSI) plays a pivotal role [12]. However, obtaining accurate CSI poses a formidable challenge in such systems due to non-linearities resulting from low-resolution quantization at the receiver, hindering the realization of the full benefits of low-resolution ADCs for efficient data communication [13]. Moreover, in high-mobility scenarios such as those involving UAVs and LEO satellites, the channel coherence times are typically short, and the Doppler effect is pronounced, further exacerbating the difficulty of channel estimation when low-resolution ADCs are employed.

B. RELATED WORK

Numerous channel estimation algorithms have been proposed for low-resolution quantization based wireless systems [14], [15], [16], [17], [18], [19], [20], [21], [22], [23]. The maximum likelihood channel estimator for one-bit massive MIMO is presented in [14], where it can effectively estimate both the direction and the norm of the channel with one-bit ADCs. It has been shown that the mean squared error (MSE) in the proposed ML channel estimator is better than the zero-forcing (ZF) type estimator and the expectation-maximization (EM) method in [15]. The paper [24] combines the EM algorithm with iterative hard thresholding (IHT) to estimate sparse channels of one-bit quantized systems. In [16] and [17], generalized approximate message passing (GAMP) algorithms have been used for channel estimation with one-bit ADCs exploiting the sparsity of the target vector. A linear minimum mean squared error (LMMSE) and Bussgang LMMSE (BLMMSE) estimators for one-bit massive MIMO systems are introduced in [18] and [19], respectively. The BLMMSE estimator in [19] exhibits superior performance than the LMMSE estimator in [18] in terms of MSE. The BLMMSE estimation method has been expanded to multi-bit resolution systems in [20]. In [25], authors develop the low-resolution aware linear minimum mean-squared error (LRA-LMMSE) channel estimator for a multiuser MIMO system with comparator network aided receivers with one-bit ADCs. In [21], a joint channel and data estimation algorithm has been proposed and it demonstrates a notable enhancement in performance when compared to the conventional approach of separately handling channel estimation and data detection. Nonetheless,

the algorithm's complexity is prohibitively high for practical implementations [21].

The training overhead and complexity of all of these pilot-based channel estimators above are too high. Additionally, the pilot-assisted approach necessitates a greater number of base station (BS) antennas to achieve comparable performance to infinite-resolution systems [26]. Therefore, it is necessary to explore alternatives for pilot-assisted schemes for channel estimation.

The papers [27], [28] proposed a mixed-ADC architecture where most antennas are equipped with low-resolution ADCs, while a few have high-resolution ADCs to acquire precise CSI. This approach significantly reduces the error floor associated with one-bit ADCs [29]. In [27], mixed-ADC architecture with a relatively small number of high-resolution ADCs is able to achieve a large fraction of the channel capacity of conventional architecture where channel estimates are acquired through the high-resolution ADCs. In [28], a mixed-ADC architecture has been adopted for frequency-selective channels, where it uses high-resolution ADCs in the channel training phase. Mixed-ADC architecture relies on the assumption that each high-resolution ADC can be linked to several RF chains through a switch. However, this architecture comes with the drawback of extensive hardware complexity due to the presence of ADC switches. Additionally, the time required to obtain channel estimates experiences a significant increase [20].

Instead of relying on conventional channel estimation techniques as discussed earlier, researchers have recently delved into the use of deep learning (DL) methods for channel estimation in low-resolution quantization based wireless systems [13], [30], [31], [32], [33], [34], [35], [36], [37]. In [30], deep multilayer perceptrons (MLPs) are employed, while [31], [32] utilize convolutional neural networks (CNNs) to learn the non-trivial mapping from quantized received observations to channels for massive MIMO channel estimation employing low-resolution ADCs. The paper [33] employs generative supervised learning for channel estimation, a method that can be trained with an adequate number of pilots in an orthogonal frequency division multiplexing receiver context.

The paper [34] employs conditional generative adversarial networks (cGAN) to predict more realistic channels in the context of one-bit multiuser massive MIMO systems. In contrast, [35] proposes an innovative LSTM-based DL model for channel matrix estimation using one-bit ADCs in massive MIMO. Their results indicate that this proposed model outperforms MLP, CNN, and cGAN-based channel estimation schemes across various numbers of antennas and SNR values. However, it is important to note that [35] involves training two simultaneous models and the design of a complex loss function. In another approach, [36] introduces an autoencoder-inspired end-to-end architecture for the joint tasks of channel estimation and mixed-ADCs allocation in massive MIMO systems. It is worth mentioning that these estimation techniques are currently constrained to scenarios

with low noise levels. Consequently, their practicality and applicability in real network settings may be limited due to this restriction [35].

It is also important to highlight that when employing existing deep learning-based methods, generating a more realistic channel matrix can be challenging due to the loss of information that occurs with successive layers in neural networks [34]. Neural networks are well-suited for tasks like classification or recognition, where the output is a label and information loss may not significantly impact performance [34]. However, in channel estimation which is related to data generation, information loss can lead to poor performance [35]. Therefore, it is challenging and important to carefully design and investigate the loss functions of neural networks to mitigate information loss during the learning process when tackling channel estimation.

Despite these efforts, the channel estimation error in low-resolution quantization based systems remains considerable compared to that in traditional high-resolution systems [12], [38]. This discrepancy becomes even more pronounced when the number of available pilot symbols is restricted [39]. Moreover, when faced with hardware imperfections, the receiver is required not only to estimate channel parameters but also to determine the quantization level of the ADCs [40]. This significantly increases the complexity of signal-processing overhead for channel estimation [13]. When the receiver relies on the estimated channel for data detection (i.e., coherent detection), the substantial channel estimation error further exacerbates the loss in data detection performance.

Given these obstacles for channel estimation with low-resolution ADCs, a fundamental question arises: Is it possible to have reliable communication using low-resolution ADCs without relying on channel estimation? We address this crucial question in the current paper and develop novel detector architectures, based on differential encoding at the transmitter, that operate optimally without depending on the availability of CSI at the receiver.

C. KEY CONTRIBUTIONS

In this paper, we develop a novel approach to overcome the channel estimation problem in wireless systems with low-resolution ADCs. Our approach involves an implementation of differential encoding at the transmitter, specifically focusing on differential M-ary phase shift keying (D-MPSK) modulation [41], [42], [43]. At the receiver, we employ a non-coherent detection scheme that does not rely on CSI. The detector, equipped with a low-resolution ADC, observes *coarsely* quantized channel outputs for data detection. This problem differs *significantly* from its high-resolution counterpart, which is extensively researched in the literature [44], [45], [46], in terms of system set-up, mathematical analysis tools, and optimal detection rules.

Our main contributions are summarized as follows:
- We derive the maximum likelihood estimator (MLE) for block-2 detectors that utilize two consecutive quantized channel outputs to determine the transmitted message index. We obtain the MLE for a general n (number of quantization bits) and M (input alphabet size), satisfying $n \geq \log_2 M$ for single-antenna systems. One form of the MLE computes the most likely estimate for the input message in a constant amount of time, which is independent of n and M. These results extend the canonical block-2 maximum likelihood detector with high-resolution ADCs, and we derive them for the first time in this paper for the low-resolution scenario.
- We investigate a scenario in which the detector can use a block of $L \geq 3$ quantized channel outputs to jointly estimate the channel and decode data. When $n = \log_2 M$, we show that the optimum block-2 MLE continues to be the *best* estimator for the transmitted message indices within the class of block-L detectors. Using the derived MLE expressions, we obtain the message error probability (MEP) formula as a performance metric and conduct numerical analyses to demonstrate system performance as a function of signal-to-noise ratio (SNR) for various values of n and M.
- To evaluate the performance of the MLE in the high SNR regime, we derive the system diversity order (DVO) and show that two additional bits (in addition to the minimum requirement of $\log_2 M$ bits) are required to achieve the same DVO as infinite bit quantization and full CSI. One bit compensates for the impact of low resolution, while the other addresses the absence of channel knowledge. This is the cost of noncoherence in phase-quantized communications systems, which is derived for the first time in the current paper. By establishing the minimum number of bits required at the quantizer to achieve the best DVO, our results complement the previous information-theoretic work [47], [48] that focuses on achievable data rates of phase-quantized non-coherent receivers. Especially, when the link reliability is at stake, our results provide pragmatic design rules and quantization penalty metrics revealing how to trade-off link reliability and the system energy consumption by optimizing the number of quantization bits.

The material of this paper has been published in part in [49], where the MEP analysis is focused only on the D-QPSK modulation and the paper did not include the proofs for the key theorems. The present paper expands upon the results presented in [49], provides full details of the proofs for the main theorems and also provides a rigorous error probability performance and DVO analysis under D-MPSK modulation with n-bit phase quantization. Our analytical results in Theorems 4-5 along with the numerical results presented in Section VII establish a fundamental asymptotic reliability characterization for low-resolution ADC based wireless systems with differential modulation in the high SNR regime. These results do not appear in our previous work [49], nor do they exist in any other previous paper in the literature.

Furthermore, the current manuscript presents a notable difference from our prior studies in [50] and [5]. In those earlier works, we studied the optimal detectors and error rates within the context of *coherent detection* for wireless systems with low-resolution quantizers. In [5], we derived the MLE for single-input-single-output (SISO) configurations employing M-ary phase shift keying (MPSK) modulation with *n*-bit quantization when the receiver has full CSI. By using the structure of the derived detector, we obtained expressions for the SEP, and for scenarios characterized by Nakagami-*m* fading, we demonstrated that a transceiver architecture becomes asymptotically optimal if $n \geq \log_2 M + 1$. Subsequently, in [50], we extended the foundational concepts from [5] to single-input-multiple-output (SIMO) wireless communication systems with low-resolution phase quantizers, and again when the receiver has full CSI. Both papers [5] and [50] share the assumption of full CSI accessibility at the receiver. In contrast, the present study relaxes the full CSI assumption by introducing D-MPSK modulation. This approach allows us to probe into optimal detectors and error rates while operating under a scenario where full CSI is not presumed.

II. SYSTEM SETUP

A. TRANSMITTER ARCHITECTURE
We consider a memoryless message source at the transmitter, which generates message indices uniformly at random from a given alphabet $\Lambda = \{0, \ldots, M-1\}$. The sequence of message indices $\{\lambda_t\}_{t=1}^{\infty}$ over time is independent and identically distributed (iid), where $t \in \mathbb{N}$ represents the time index. The messages are converted into symbols for transmission over the channel by using differential encoding, as described below.[1]

At time $t = 0$, the modulator generates a symbol S_0 uniformly at random from the constellation set $\mathcal{C} = \{e^{J\frac{\pi}{M}(2k+1)}\}_{k=0}^{M-1}$. This symbol initiates the communication process, and the transmitted symbol at time t is given by

$$S_t = S_{t-1}e^{J\frac{2\pi}{M}\lambda_t} \tag{1}$$

for all $t \geq 1$. We note that S_t belongs to $\mathcal{C}$ for all $t \in \mathbb{N}$ (with phase modulo 2π). It can also be seen that $\{S_t\}_{t=0}^{\infty}$ is an iid sequence of symbols over time with uniform marginal distribution over $\mathcal{C}$.

B. CHANNEL MODEL
We consider a block-fading wireless channel which remains constant over a block of N symbols and changes from

---

1. The terms *symbol* and *message* are used to distinguish between what is transmitted over the channel and the index of the information symbol created by the source.

---

one block to another independently [51]. The input-output relationship of the channel for the kth block is given by

$$Y_t = \sqrt{\mathsf{SNR}}\, H_k S_t + W_t, \tag{2}$$

where $t \in [kN : (k+1)N - 1]$, $k \in \mathbb{N}$, $H_k \in \mathbb{C}$ is the unit-power Rayleigh fading coefficient, W_t is the circularly-symmetric zero-mean unit-variance additive white Gaussian noise (AWGN) (i.e., both H_k and W_t are distributed according to $\mathcal{CN}(0,1)$) and SNR is the system signal-to-noise ratio. The sequence of received symbols $\{Y_t\}_{t=1}^{\infty}$ forms a dependent collection of random variables (a random sequence over time) due to common randomness introduced by the channel.

The following lemma characterizes the joint distribution of Y_{t_0} and Y_{t_1} for any two consecutive time indices t_0 and t_1.

Lemma 1: For any two consecutive time indices t_0 and $t_1 = t_0 + 1$, let $\boldsymbol{Y} = (Y_{t_0}, Y_{t_1})^{\top} \in \mathbb{C}^2$. Then, conditioned on $\lambda_{t_1} = \lambda \in \Lambda$, $\boldsymbol{Y}$ is a jointly Gaussian complex random vector with circularly symmetric distribution $\mathcal{CN}(0, \boldsymbol{K})$, where the covariance matrix $\boldsymbol{K} = \mathsf{E}[\boldsymbol{Y}\boldsymbol{Y}^{\dagger}]$ is given by (3), shown at the bottom of the page and $\boldsymbol{I}_2$ in (3) is the 2-by-2 identity matrix.

Proof: The proof of this lemma follows by first establishing joint Gaussianity for Y_{t_0} and Y_{t_1}, and then calculating their covariance matrix. See Appendix A for details. ∎

The statistical structure established in this lemma will be key in our derivation for the MLE when the detector utilizes two consecutive *quantized* channel output samples to decide on the transmitted message index at the current time. Using Lemma 1, it can also be seen that $Y_{t_1}|Y_{t_0} = y_0 \sim \mathcal{CN}\left(\frac{\mathsf{SNR}}{1+\mathsf{SNR}}e^{J\frac{2\pi}{M}\lambda}y_0, \frac{1+2\mathsf{SNR}}{1+\mathsf{SNR}}\right)$ if $\lambda_{t_1} = \lambda$, and Y_{t_0} and Y_{t_1} lie on the same channel fading block.

C. RECEIVER ARCHITECTURE
Our receiver architecture is based on a low-resolution ADC that quantizes the phase of the channel output [52], [53]. This structure is motivated by the fact that the information is encoded in the phase of the transmitted symbols in our setting. Since the channel is circularly symmetric with independent phase and amplitude [54], any form of amplitude quantization will not provide useful information that will assist the estimation process in this setting.

We note that there are several important reasons why we focus on phase modulation and quantization in this paper. Phase modulation is historically known to be optimum up to modulation order 16 under peak power limitations [55]. It is also the optimum modulation scheme for achieving the channel capacity with phase quantized outputs [56]. Phase quantizers can be implemented using one-bit ADCs that consist of simple comparators, and they attain low power consumption levels (in the order of milliwatts). As given

---

$$\boldsymbol{K} = \begin{cases} (1+\mathsf{SNR})\begin{bmatrix} 1 & \frac{\mathsf{SNR}}{1+\mathsf{SNR}}e^{-J\frac{2\pi}{M}\lambda} \\ \frac{\mathsf{SNR}}{1+\mathsf{SNR}}e^{J\frac{2\pi}{M}\lambda} & 1 \end{bmatrix} & \text{if } t_0 \mod N \in [0 : N-2] \\ (1+\mathsf{SNR})\boldsymbol{I}_2 & \text{if } t_0 \mod N = N-1 \end{cases} \tag{3}$$
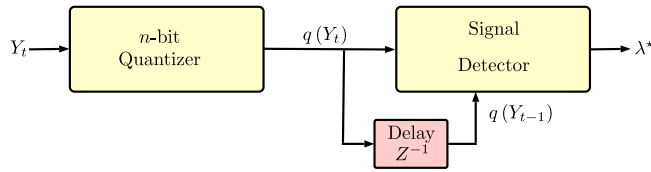
FIGURE 1. The receiver architecture with *n*-bit phase quantizer. The figure is depicted for a block-2 detector for clarity, utilizing two consecutive symbols to estimate the message indices.

in [57], the implementation based on time-to-digital converters (TDCs) can also be adopted to further reduce the area and power consumption of the phase quantizer.

As illustrated in Fig. 1, the received symbol Y_t is first sent through a low-resolution quantizer, and then the quantized symbols at times $t - 1$ and t are used to determine the transmitted message λ_t at time $t \geq 1$. We assume that n bits are used for quantization and $n \geq \log_2 M$.

The quantization operation can be described formally as follows. Let Q_k be the kth quantization region (Q-region), which is defined as the convex cone satisfying

$$Q_k \triangleq \left\{ z \in \mathbb{C} : \frac{2\pi}{2^n} k \leq \mathsf{Arg}(z) < \frac{2\pi}{2^n} (k+1) \right\} \quad (4)$$

for $k \in [0 : 2^n - 1]$ and the principle argument $\mathsf{Arg}(z)$ is always taken in $[0, 2\pi)$. The quantizer q is defined as a mapping from $\mathbb{C}$ to $[0 : 2^n - 1]$ that satisfies $q(z) = k$ if $z \in Q_k$. Note that q is well-defined since the Q-regions are mutually exclusive and collectively cover the complex plane.

For each Q-region, we also define an attraction region (A-region) A_k according to (5), shown at the bottom of the next page for $k \in [0 : 2^n - 1]$. In this equation, $B \mod a$ for a set $B \subseteq \mathbb{R}$ and $a > 0$ is defined as $B \mod a \triangleq \{b \mod a : b \in B\}$. We note that the kth A-region is a convex cone having angular separation $\frac{2\pi}{M}$ and centered around Q_k. The importance of A-regions for our analysis is given by the following lemma.

Lemma 2: Let U and V be two complex Gaussian random variables with distributions $U \sim \mathcal{CN}(u, \sigma^2)$ and $V \sim \mathcal{CN}(v, \sigma^2)$ satisfying $|u| = |v|$. Then, if $u \in A_k$ and $v \notin \mathsf{cl}(A_k)$, we have $\mathsf{Pr}\{U \in Q_k\} > \mathsf{Pr}\{V \in Q_k\}$, where $\mathsf{cl}(A_k)$ represents the topological closure of A_k.

Proof: The proof follows from [5, Lemma 6]. ∎

We will use this lemma to maximize the likelihood function parameterized by the message indices $\lambda \in \Lambda$. Intuitively, the distribution of U is better centered around Q_k than that of V if the conditions in Lemma 2 are satisfied, giving us the inequality $\mathsf{Pr}\{U \in Q_k\} > \mathsf{Pr}\{V \in Q_k\}$. The following decomposition of A-regions into Q-regions will also be used in our MLE derivation.

Lemma 3: For $k \in [0 : 2^n - 1]$, A_k can be decomposed into Q-regions according to (6), shown at the bottom of the next page, where Q_k^U is the upper kth Q-region defined as $Q_k^U \triangleq \{z \in \mathbb{C} : \frac{2\pi}{2^n} k + \frac{\pi}{2^n} \leq \mathsf{Arg}(z) < \frac{2\pi}{2^n}(k+1)\}$ and Q_k^L is the lower kth Q-region defined as $Q_k^L \triangleq \{z \in \mathbb{C} : \frac{2\pi}{2^n} k \leq \mathsf{Arg}(z) < \frac{2\pi}{2^n} k + \frac{\pi}{2^n}\}$.

Proof: The proof of this lemma follows from the following intuitive idea. A_k is a cone with an angular width

of $\frac{2\pi}{M}$ centered around Q_k. Therefore, it can contain $\frac{2^n}{2M}$ Q-regions (each of width $\frac{2\pi}{2^n}$) in the clockwise and counter-clockwise directions from the center of Q_k. The proof of the lemma depends on making this idea formal to have the decomposition given in (6). See Appendix B for details. ∎

III. OPTIMUM SIGNAL DETECTION

In this part, we will present the optimum signal detection problem in the general form, and then specialize it to the practical case where the detector utilizes only two consecutive quantized channel outputs (which we call *block-2* detector) to decide on the transmitted message index. The block-2 detector is the canonical detector used in high-resolution quantization based D-MPSK systems [44], [45]. We will derive the MLE for the class of block-2 detectors for *low-resolution* quantization and also show that it is optimum among the class of block-*L* detectors in an important case of $n = \log_2 M$, which includes DQPSK (i.e., $M = 4$) with one-bit quantization for in-phase and quadrature channels (i.e., $n = 2$).

This problem differs significantly from its high-resolution counterpart. In a low-resolution quantization based system, the output conforms to a nonlinear channel model due to the nonlinearity introduced by the quantization process. In contrast, traditional high-resolution systems are built on a linear channel model. Therefore, signal detection rules developed for receivers with high-resolution ADCs often become sub-optimal for receivers with low-resolution ADCs [8]. Utilizing low-resolution ADCs fundamentally alters both the theoretical framework and practical implementation of a communication setup [58].

A. GENERAL FORMULATION

In general, the detector can utilize a block of L quantized *observed* channel outputs $\boldsymbol{k} = [k_0, \ldots k_{L-1}]^\top$, taken at consecutive times from t_0 to t_{L-1} (i.e., $k_i = q(Y_{t_i})$), to form a block of $L - 1$ message indices $\hat{\boldsymbol{\lambda}} = [\hat{\lambda}_1, \ldots, \hat{\lambda}_{L-1}]^\top$ as an estimate for the actual message sequence $\boldsymbol{\lambda} = [\lambda_{t_1}, \ldots, \lambda_{t_{L-1}}]^\top$ transmitted from t_1 to t_{L-1}. The block size L must be taken in the range $[2 : N]$ since we need at least two quantized channel outputs to estimate the input messages due to differential encoding and having L larger than N does not help for the estimation problem since we lose correlation structure completely at channel transition boundaries (i.e., see Lemma 1 for two consecutive channel outputs and the same idea extends to the general case $L \geq 2$).

Then, for $L \in [2 : N]$, the MLE $\boldsymbol{\lambda}^\star(\boldsymbol{k})$ (also called maximum a posteriori (MAP) detector [59]) is the estimator that satisfies

$$\boldsymbol{\lambda}^\star(\boldsymbol{k}) \in \arg \max_{\hat{\boldsymbol{\lambda}} \in \Lambda^{L-1}} \mathsf{Pr}\left\{ \boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}} \big| \boldsymbol{q} = \boldsymbol{k} \right\} \quad (7)$$

for all $\boldsymbol{k} \in [0 : 2^n - 1]^L$, where $\boldsymbol{q} = [q(Y_{t_0}), \ldots, q(Y_{t_{L-1}})]^\top$ is the *random* quantized channel output vector of size L.[2]

---

2. We do not have any differentiation between random quantities and their realizations in notation since it will be clear from the context which quantities are random and which are their observed realizations.

Since all input message sequences are equiprobable, we can express (7) equivalently according to

$$\lambda^\star(\boldsymbol{k}) \in \arg \max_{\hat{\lambda} \in \Lambda^{L-1}} \Pr\left\{\boldsymbol{q} = \boldsymbol{k} \,\middle|\, \lambda = \hat{\lambda}\right\}. \qquad (8)$$

In this formulation, the larger block sizes can help the receiver to jointly estimate the channel and decode the data. However, they also introduce extra decoding complexity and processing delay. As seen in (7) and (8), the size of the search space for $\lambda^\star(\boldsymbol{k})$ increases exponentially with the block size according to $O(M^L)$, and the size of the input space over which $\lambda^\star(\boldsymbol{k})$ must be well-defined also grows exponentially according to $O(2^{nL})$. Thus, in addition to the search for the optimum decoder, these observations also render potential numerical learning models such as soft-max regression useless for large L given the usually fast time scales (e.g., 2-3 ms [51]) of the wireless channel coherence time.

B. MLE FOR THE BLOCK-2 CASE

For the block-2 case, using (8), we can write the likelihood function $\mathcal{L}(\lambda; \boldsymbol{k})$ of parameter λ (i.e., the estimate of the transmitted message index) according to

$$
\begin{aligned}
\mathcal{L}(\lambda; \boldsymbol{k}) &= \Pr\left\{\left[q(Y_{t_0}), q(Y_{t_1})\right]^\top = \boldsymbol{k} \,\middle|\, \lambda_{t_1} = \lambda\right\} \\
&= \Pr\left\{Y_{t_0} \in Q_{k_0} \text{ and } Y_{t_1} \in Q_{k_1} \,\middle|\, \lambda_{t_1} = \lambda\right\} \\
&= \int_{Q_{k_0}} f_{Y_{t_0}}(y_0) \int_{Q_{k_1}} f_{Y_{t_1}|Y_{t_0}}(y_1|y_0; \lambda) \, dy_1 \, dy_0
\end{aligned}
$$

for any $\boldsymbol{k} = [k_0, k_1]^\top \in [0 : 2^n - 1]^2$, where $f_{Y_{t_0}}(y_0) = \frac{1}{\pi(1+\text{SNR})} e^{-\frac{|y_0|^2}{1+\text{SNR}}}$ and $f_{Y_{t_1}|Y_{t_0}}(y_1|y_0; \lambda) = \frac{1+\text{SNR}}{\pi(1+2\text{SNR})} e^{-\frac{1+\text{SNR}}{1+2\text{SNR}}|y_1 - \frac{\text{SNR}}{1+\text{SNR}} e^{J\frac{2\pi}{M}\lambda} y_0|^2}$ by using Lemma 1. The following theorem establishes an MLE for the estimation problem in (8) for the block-2 case.

Theorem 1: Let $\lambda^\star(\boldsymbol{k})$ be given by

$$\lambda^\star(\boldsymbol{k}) = \min\left\{\lambda \in \Lambda : e^{J\frac{2\pi}{M}\lambda} Q_{k_0} \bigcap A_{k_1} \neq \emptyset\right\}, \qquad (9)$$

where the set $e^{J\frac{2\pi}{M}\lambda} Q_{k_0}$ is defined as $e^{J\frac{2\pi}{M}\lambda} Q_{k_0} \triangleq \{z e^{J\frac{2\pi}{M}\lambda} : z \in Q_{k_0}\}$. Then, $\lambda^\star(\boldsymbol{k})$ is well-defined and satisfies

$$\lambda^\star(\boldsymbol{k}) \in \arg \max_{\lambda \in \Lambda} \mathcal{L}(\lambda; \boldsymbol{k})$$

for all $\boldsymbol{k} \in [0 : 2^n - 1]^2$.

Proof: The key term which we focus on to establish the MLE formula in (9) is the integral expression $\int_{Q_{k_1}} f_{Y_{t_1}|Y_{t_0}}(y_1|y_0; \lambda) dy_1$ in the likelihood function $\mathcal{L}(\lambda; \boldsymbol{k})$. The proof of this theorem follows by showing that this integral is maximized for any $y_0 \in Q_{k_0}$ whenever λ is chosen according to (9). See Appendix C for formal arguments and details. ∎

Theorem 1 establishes a solution for (8) as the minimum rotation required (in terms of message indices) to have a non-empty intersection between Q_{k_0} (i.e., Q-region containing the first observation) and the A-region corresponding to Q_{k_1} (i.e., Q-region containing the second observation). In the proof of Theorem 1, we showed that $e^{J\frac{2\pi}{M}\lambda} Q_{k_0} = Q_{(k_0 + \frac{2^n}{M}\lambda) \mod 2^n}$. Hence, using Lemma 3, an equivalent representation of $\lambda^\star(\boldsymbol{k})$ can be given as in Corollary 1 below, which makes the structure of the solution more lucid.

Corollary 1: The MLE $\lambda^\star(\boldsymbol{k})$ can also be expressed as in (10), shown at the bottom of the next page.

A notable observation based on the MLE given in Corollary 1 is the *symbol-by-symbol* detection interpretation when $n = \log_2 M$. In this case, the Q-region and message indices match each other (i.e., both are $\{0, \ldots, M-1\}$), and the detector behaves as if it first estimates $e^{J\frac{\pi}{M}(2k_0+1)}$ and $e^{J\frac{\pi}{M}(2k_1+1)}$ for the symbols S_{t_0} and S_{t_1}, respectively, and then calculates their phase difference to estimate λ_{t_1}. We will use this idea in the next section to show that the block-2 MLE also gives us the optimum detection rule for $L \geq 3$ when $n = \log_2 M$.

In addition, the structure observed for $n = \log_2 M$ motivates a search for another MLE having a closed form expression in terms of the difference between k_1 and k_0 even when $n \geq \log_2 M + 1$. In the next theorem, we provide another MLE that estimates the message index based on the difference between k_1 and k_0 directly. This MLE coincides with $\lambda^\star(\boldsymbol{k})$ given in Theorem 1 except for $(k_0 + \frac{2^n}{M}\lambda^\star(\boldsymbol{k})) \mod 2^n = (k_1 - \frac{2^n}{2M}) \mod 2^n$ and $(k_0 + \frac{2^n}{M}\lambda^\star(\boldsymbol{k})) \mod 2^n = (k_1 + \frac{2^n}{2M}) \mod 2^n$. In these two cases, the set $\arg \max_{\lambda \in \Lambda} \mathcal{L}(\lambda; \boldsymbol{k})$ is not a singleton, and the MLE $\lambda^{\star\star}(\boldsymbol{k})$ in Theorem 2 chooses the other message index in this set as an estimate for λ_{t_1}.

Theorem 2: Let $\lambda^{\star\star}(\boldsymbol{k})$ be defined as

$$\lambda^{\star\star}(\boldsymbol{k}) = \left\lceil \frac{M}{2^n}(k_1 - k_0) - \frac{1}{2} \right\rceil \mod M. \qquad (11)$$

Then, for all $n \geq \log_2 M$, $\lambda^{\star\star}(\boldsymbol{k})$ satisfies

$$\lambda^{\star\star}(\boldsymbol{k}) \in \arg \max_{\lambda \in \Lambda} \mathcal{L}(\lambda; \boldsymbol{k})$$

for all $\boldsymbol{k} \in [0 : 2^n - 1]^2$.

$$A_k = \left\{ z \in \mathbb{C} : \mathsf{Arg}(z) \in \left[\frac{2\pi}{2^n}k + \frac{\pi}{2^n} - \frac{\pi}{M}, \frac{2\pi}{2^n}k + \frac{\pi}{2^n} + \frac{\pi}{M}\right) \mod 2\pi \right\} \qquad (5)$$

$$A_k = \begin{cases} Q_k & \text{if } n = \log_2 M \\ Q^U_{\left(k - \frac{2^n}{2M}\right) \mod 2^n} \bigcup \left(\bigcup_{i=k-\frac{2^n}{2M}+1}^{k+\frac{2^n}{2M}-1} Q_{i \mod 2^n}\right) \bigcup Q^L_{\left(k+\frac{2^n}{2M}\right) \mod 2^n} & \text{if } n \geq \log_2 M + 1 \end{cases}, \qquad (6)$$

Proof: To prove this theorem, we only consider $n \geq \log_2 M + 1$ since (10) and (11) agree with each other for $n = \log_2 M$. The proof for the case $n \geq \log_2 M + 1$ follows from considering three different sub-cases for the range of k_1 : (i) $k_1 \in [\frac{2^n}{2M} : 2^n - \frac{2^n}{2M} - 1]$, (ii) $k_1 \in [0 : \frac{2^n}{2M} - 1]$ and (iii) $k_1 \in [2^n - \frac{2^n}{2M} : 2^n - 1]$, and showing that $\lambda^{\star\star}(k)$ maximizes $\mathcal{L}(\lambda; k)$ in each sub-case. The details are provided in Appendix D. ∎

We first note that if $n = \log_2 M$, then

$$\lambda^{\star\star}(k) = \left\lceil k_1 - k_0 - \frac{1}{2} \right\rceil \mod M$$
$$= (k_1 - k_0) \mod M,$$

which coincides with $\lambda^{\star}(k)$ exactly in this case.

Secondly, the MLE given in Theorem 2 has a nice intuitive interpretation. $k_1 - k_0$ gives a measure of "distance" between two quantized channel outputs in terms of the number of quantization regions. The scaling coefficient $\frac{M}{2^n}$ converts this measure to another measure of "distance" in terms of the input symbols. Finally, the ceiling operation and the term $-\frac{1}{2}$ act as the correction factors to convert this final measure to an estimate of the message index. We will use $\lambda^{\star\star}(k)$ in our numerical analysis to calculate message error probabilities as it is very handy to implement both in numerical analysis and also in practical embedded software implementations.

Remark: We note that the complexity of the optimum detector in Theorem 2 is $O(1)$ because it only compares two quantization region indices to determine the most likely input message index.

IV. OPTIMALITY OF THE BLOCK-2 MLE WITHIN THE CLASS OF BLOCK-*L* DETECTORS

As discussed in the previous section, processing quantized channel output blocks of size larger than 2 can help the receiver to learn the channel states and simultaneously decode the messages based on the learned channel conditions. However, there is an extra decoding delay and exponential computational complexity associated with this approach.

On the other hand, the block-2 MLE given in Section III has a simple structure, which is easy to implement in practical applications and process data in real time. It outputs the most likely input message index in *constant* time for any observation $k \in [0 : 2^n - 1]^2$, and any M and n satisfying $n \geq \log_2 M$ (i.e., see Theorem 2). In the next theorem, we will show that this simple structure continues to be *optimum* among the class of block-*L* detectors when $n = \log_2 M$, which is the minimum number of bits required to resolve input messages. The proof idea uses a genie-aided detector where the genie reveals perfect information about the channel

states to the receiver. The transmitter still uses differential encoding to encode the messages.

Theorem 3: For $n = \log_2 M$, the block-2 MLEs given in Theorems 1 and 2 are the optimum detectors among the class of block-*L* detectors.

Proof: To prove this theorem, we consider a genie-aided decoder where the genie provides perfect channel information $\{H_k\}_{k=0}^{\infty}$ and the decoder utilizes this information to maximize the likelihood of message indices $\hat{\lambda} = [\hat{\lambda}_1, \ldots, \hat{\lambda}_{L-1}]^\top$ as an estimate for the actual message sequence $\lambda = [\lambda_{t_1}, \ldots, \lambda_{t_{L-1}}]^\top$ transmitted from t_1 to t_{L-1}. The analysis of the genie-aided decoder is easier since the channel outputs are conditionally independent given the channel information. In Appendix E, we first establish the MLE for the genie-aided detector and then show that this MLE is exactly the same MLE we arrived at in Theorems 1 and 2 for $n = \log_2 M$ when $L = 2$. This concludes the proof. ∎

Intuitively, given the channel knowledge, the genie-aided detector defaults to the symbol-by-symbol estimation of the sequence $\{S_t\}_{t=0}^{\infty}$, and then the message sequence $\{\lambda_t\}_{t=1}^{\infty}$ is estimated based on the phase difference between two consecutive symbols in the estimated symbol sequence.

V. MESSAGE ERROR PROBABILITY

In this section, we calculate the message error probability (MEP) for low-resolution quantization systems with differential encoding by using the MLE expressions derived in Section III. To that end, we write the MEP as

$$p(\mathsf{SNR}) = \mathsf{Pr}\{\lambda_t \neq \lambda^{\star\star}(q(Y_{t-1}), q(Y_t))\}. \quad (12)$$

We consider the D-MPSK modulation and the case where both observations are on the same channel fading block. For observations belonging to different fading blocks, we lose the correlation structure and $p(\mathsf{SNR})$ is equal to $1 - \frac{1}{M}$.

Theorem 4: For D-MPSK modulation with $n \geq \log_2 M$, $p(\mathsf{SNR})$ is given according to (13), shown at the bottom of the next page.

Proof: We first focus on the case where $n = \log_2(M)$. Without loss of generality, for any two consecutive time indices $t - 1$ and t, let us assume that $Y_{t-1} \in Q_k$ given according to (4) and $\lambda_t = 0$ (i.e., $S_t = S_{t-1}$). By applying Theorem 2 and setting $k_1 = k_0$, we can ensure that there will be no error in estimating the message $\lambda^{\star\star}$, since this implies that $q(Y_t) = q(Y_{t-1})$. Then, the MEP can be expressed as follows:

$$p(\mathsf{SNR}) = \sum_{k=0}^{2^n-1} \mathsf{Pr}\{Y_{t-1} \in Q_k, Y_t \notin Q_k \mid \lambda_t = 0\}$$
$$\overset{(a)}{=} 2^n \mathsf{Pr}\{Y_{t-1} \in Q_0, Y_t \notin Q_0 \mid \lambda_t = 0\}, \quad (14)$$

$$\lambda^{\star}(k) = \begin{cases} (k_1 - k_0) \mod M & \text{if } n = \log_2 M \\ \min\left\{\lambda \in \Lambda : \left(k_0 + \frac{2^n}{M}\lambda\right) \mod 2^n \in \left[k_1 - \frac{2^n}{2M} : k_1 + \frac{2^n}{2M}\right] \mod 2^n\right\} & \text{if } n \geq \log_2 M + 1 \end{cases}. \quad (10)$$

where (a) follows from the circular symmetry of the problem.

In order to derive $p(\mathsf{SNR})$ in equation (14), we need to use the joint distribution of real and imaginary parts of Y_{t-1} and Y_t. To this end, let $\boldsymbol{Y} = [Y_{t-1}^{\mathrm{re}} \quad Y_{t-1}^{\mathrm{im}} \quad Y_t^{\mathrm{re}} \quad Y_t^{\mathrm{im}}]^\top$, where Y_t^{re} is the real part of Y_t and Y_t^{im} is the imaginary part of Y_t.

Using Lemma 1, the pdf of $\boldsymbol{Y}$ can then be expressed as (15), shown at the bottom of the page, where $\boldsymbol{y} = [y_{t-1}, y_t]^\top$.

Then, by using (15), we can express (14) as (16), shown at the bottom of the page which is equivalent to the MEP expression in Theorem 4 for $n = \log_2(M)$.

The proof for $n \geq \log_2(M) + 1$ follows from similar arguments, with the main difference being how we express $p(\mathsf{SNR})$ in this case. The MEP, in this case, is given by

$$p(\mathsf{SNR}) = 2^n \Pr\{Y_{t-1} \notin A_k, Y_t \in Q_k \mid \lambda_t = 0\},$$

where Q_k and A_k are given according to (4) and (5), respectively.[3] Using the joint distribution for Y_{t-1} and Y_t given in Lemma 1 and its circular symmetry property, we arrive at (13). Therefore, we get the MEP expression in Theorem 4 for $n \geq \log_2(M) + 1$ as well. ∎

For the special case of D-QPSK (i.e., $M = 4$) modulation with $n = 2, 3$, Theorem 4 reduces to the following corollary.

Corollary 2: For D-QPSK modulation with $n = 2, 3$, $p(\mathsf{SNR})$ is given according to (17), shown at the bottom of the page, where $\mathcal{Q}(\cdot)$ is the complementary distribution function of the standard normal random variable.

Proof: See Appendix F. ∎

---

3. This can be seen by a randomization argument selecting one of the message indices randomly with probability 0.5 to maintain symmetry in the problem when there are two messages having the same maximum likelihood.

VI. THE DECAY EXPONENT FOR THE AVERAGE SYMBOL ERROR PROBABILITY

In this section, we will analyze the communication robustness that can be achieved with low-resolution ADCs by focusing on the decay exponent of $p(\mathsf{SNR})$, which is given by[4]

$$\mathsf{DVO} = -\lim_{\mathsf{SNR} \to \infty} \frac{\log p(\mathsf{SNR})}{\log \mathsf{SNR}}. \tag{18}$$

Following the convention in the field, we will call DVO the *diversity order*.

The following theorem establishes the DVO results for D-MPSK modulation under Rayleigh fading.

Theorem 5: For D-MPSK modulation with $n \geq \log_2 M$, the DVO under Rayleigh fading is given according to

$$\mathsf{DVO} = \begin{cases} \frac{1}{2} & n = \log_2 M \text{ and } n = \log_2 M + 1 \\ 1 & n \geq \log_2 M + 2 \end{cases}. \tag{19}$$

Proof: The proof follows by analyzing the decay rate of $p(\mathsf{SNR})$ for $n = \log_2 M$, $n = \log_2 M + 1$ and $n \geq \log_2 M + 2$ case-by-case. In particular, for the case of $n = \log_2 M$, we use the genie-aided receiver in Theorem 3 to show that $\mathsf{DVO} = \frac{1}{2}$. For $n = \log_2 M + 1$, we use the D-QPSK MEP result in (17) to establish that $\mathsf{DVO} = \frac{1}{2}$. For the final case $n \geq \log_2 M + 2$, we apply monotone convergence theorem to show that $p(\mathsf{SNR})$ in (13) decays according to SNR^{-1}. The details are given in Appendix G. ∎

From Theorem 5, we know that the DVO of the low-resolution quantization based systems is the same as that of infinite-bit quantization based systems, which is equal to 1

---

4. The limit in (18) exists for the optimum detectors that we study in the current paper. Hence, there is no ambiguity in the definition of DVO.

$$p(\mathsf{SNR}) = 1 - \frac{2^n}{\pi^2(2\mathsf{SNR}+1)} \int_0^\infty \int_0^{y_t^{\mathrm{re}} \tan\left(\frac{\pi}{2^{n-1}}\right)} \int_0^\infty \int_{y_{t-1}^{\mathrm{re}} \tan(\psi)}^{y_{t-1}^{\mathrm{re}} \tan\left(\frac{2\pi}{M}+\psi\right)} \mathcal{G}(\boldsymbol{y}) \, dy_{t-1}^{\mathrm{im}} \, dy_{t-1}^{\mathrm{re}} \, dy_t^{\mathrm{im}} \, dy_t^{\mathrm{re}} \tag{13}$$

where $\mathcal{G}(\boldsymbol{y}) = \exp\{-\frac{\mathsf{SNR}+1}{2\mathsf{SNR}+1}\|\boldsymbol{y}\|_2^2 + \frac{2\mathsf{SNR}}{2\mathsf{SNR}+1}\mathsf{Re}(y_{t-1}y_t^*)\}$, $\boldsymbol{y} = [y_{t-1}, y_t]^\top$, and

$$\psi = \begin{cases} 0 & \text{if } n = \log_2 M \text{ or } n = \log_2(M) + 1 \\ -\left(2(n - 1 - \log_2 M) - 1\right)\frac{\pi}{2^{n-1}} & \text{if } n \geq \log_2(M) + 2. \end{cases}$$

$$f_Y(\boldsymbol{y}) = \frac{1}{\pi^2(2\mathsf{SNR}+1)} \exp\left\{-\frac{\mathsf{SNR}+1}{2\mathsf{SNR}+1}\|\boldsymbol{y}\|_2^2 + \frac{2\mathsf{SNR}}{2\mathsf{SNR}+1}\mathsf{Re}(y_{t-1}y_t^*)\right\}, \tag{15}$$

$$p(\mathsf{SNR}) = 1 - \frac{2^n}{\pi^2(2\mathsf{SNR}+1)} \int_0^\infty \int_0^{y_t^{\mathrm{re}} \tan\left(\frac{\pi}{2^{n-1}}\right)} \int_0^\infty \int_0^{y_{t-1}^{\mathrm{re}} \tan\left(\frac{\pi}{2^{n-1}}\right)} f_Y(\boldsymbol{y}) \, dy_{t-1}^{\mathrm{im}} \, dy_{t-1}^{\mathrm{re}} \, dy_t^{\mathrm{im}} \, dy_t^{\mathrm{re}} \tag{16}$$

$$p(\mathsf{SNR}) = 1 - \frac{4}{\pi} \int_0^{2\pi} \int_0^\infty \mathcal{Q}^2\left(-\sqrt{2\mathsf{SNR}}r\cos\theta\right) \mathcal{Q}^2\left(-\sqrt{2\mathsf{SNR}}r\sin\theta\right) re^{-r^2} \, dr \, d\theta. \tag{17}$$

---

when $n \geq \log_2 M + 2$ under Rayleigh fading. Since this is the optimum DVO, we can conclude that it is enough to use $\log_2 M + 2$ bits at the quantizer to achieve the optimum DVO with D-MPSK modulation, which is two additional bits on top of the minimum requirement of $\log_2 M$ bits to resolve input messages. The first additional bit is required to overcome the quantization noise in the system. In comparison with [5], the other additional bit at the quantizer is required to overcome the lack of channel knowledge and thereby to achieve the optimum DVO when the channel knowledge is not available at the receiver.

VII. NUMERICAL RESULTS

In this section, we present the numerical results that will help us to illustrate the key theorems of the paper and to comprehensively investigate the MEP performance of the optimum detectors derived in the previous sections as well as several quantization penalties.

A. ERROR PROBABILITY PERFORMANCE
In Fig. 2, a numerical comparison of the MEP for D-MPSK is presented, considering various values of n, the number of quantization bits. The impact of n on the error probability performance, as stated in Theorem 5, is crucial in determining the decay exponent for the MEP, and hence the link reliability, of a low-resolution ADC-based system operating without channel information at the receiver.

We have used the Monte Carlo simulation method to generate the simulated results in Fig. 2, while the analytical results are obtained by using the error probability expression we derived in Theorem 4. The simulated and analytical MEP curves we present in Fig. 2 demonstrate that the analytical results accurately track the simulated results for all cases. These performance curves further indicate that the MEP for $n = \log_2 M$ and $n = \log_2 M + 1$ cases are nearly identical, and the decay slope of MEP is $\frac{1}{2}$. This is in accordance with one of the key results in Theorem 5. On the other hand, when n is increased to $\log_2 M + 2$, there is a significant improvement in the MEP with a corresponding change in the decay slope of MEP. This observation suggests a change in the diversity order and that D-MPSK requires two additional bits (besides $\log_2 M$) to achieve the maximum diversity order of 1 under Rayleigh fading, as emphasized in Theorem 5. One of these bits serves to mitigate the impact of low-resolution quantization at the receiver, while the other bit compensates for the absence of channel knowledge.
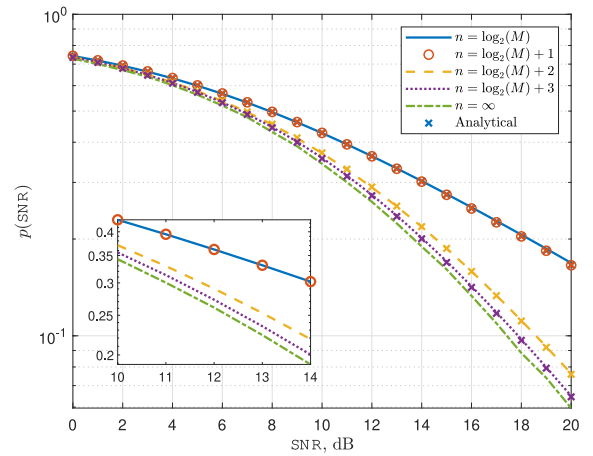
B. QUANTIZATION PENALTY
While phase quantization with less number of quantization bits is desirable, due to less processing complexity at the receiver, it deteriorates the MEP performance of the system. In this part, we will study a quantization penalty metric to quantify the increase in the MEP. To that end, we define $\Psi(\text{SNR}, M, n)$ as
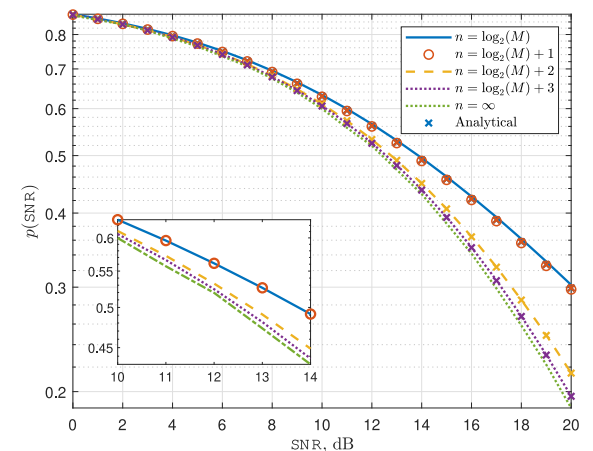
$$\Psi(\text{SNR}, M, n) = 10\log\left(\frac{p(\text{SNR}, M, n)}{p(\text{SNR}, M, \infty)}\right), \quad (20)$$
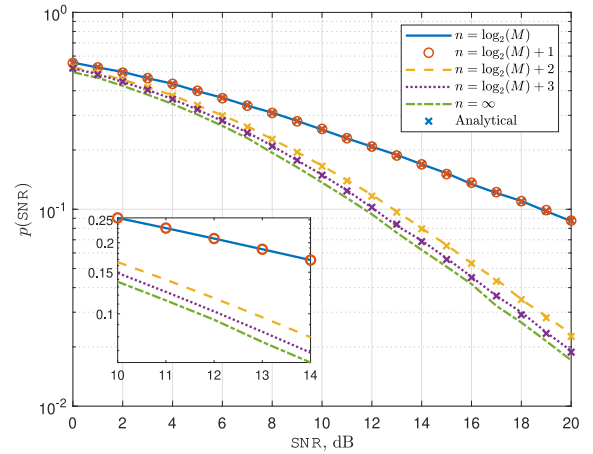
(a) $M = 4$

(b) $M = 8$

(c) $M = 16$

FIGURE 2. MEP curves as a function of SNR for D-MPSK. $M = 4, 8, 16$ and $n = \log_2 M, \log_2 M + 1, \log_2 M + 2, \log_2 M + 3, \infty$.

where we used notation $p(\text{SNR}, M, n)$ to indicate the MEP for D-MPSK modulation and n-bit quantization. We note that the quantity $\Psi(\text{SNR}, M, n)$ quantifies the increment in the
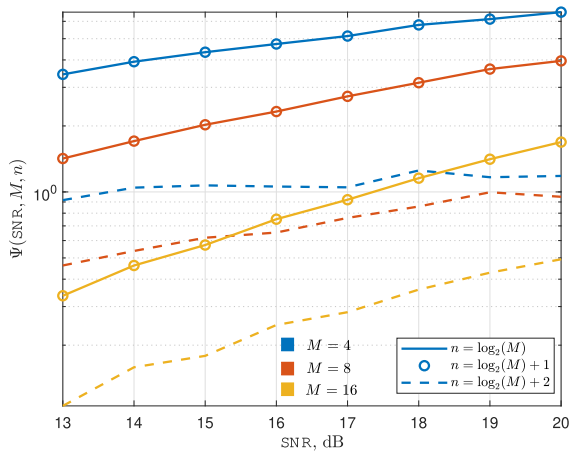
FIGURE 3. Ψ(SNR, *M*, *n*) as a function of SNR. *M* = 4, 8, 16 and *n* = log₂ *M*, log₂ *M* + 1, log₂ *M* + 2.

MEP (in dB scale) when employing a low-resolution ADC with *n* bits compared to the scenario of using a full-resolution ADC ($n = \infty$).

To illustrate this point further, we have generated the MEP curves for $M = 4, 8, 16$ and $n = 2, 3, 4, 5, \infty$ by using our analytical error probability expression in Theorem 4 and zooming in the SNR = 12 dB region in Fig. 2. As demonstrated in this figure, when the SNR is kept fixed at 12 dB, changing from $n = \log_2(M)$ to ∞-bit quantization incurs a quantization penalty of around $\Psi(12\text{dB}, 4, 2) \approx 3.44$, $\Psi(12\text{dB}, 8, 3) \approx 1.42$, $\Psi(12\text{dB}, 16, 4) \approx 0.34$ resulting in a 2.2-fold, 1.4-fold and 1.1-fold increase in the MEP, respectively.

It is critical to compare these values with the scenario where we increase *n* from $n = \log_2 M + 2$ to ∞. The value $n = \log_2 M + 2$ is significant because it is the point where the system attains the full DVO, resulting in a phase transition of the decay exponent in the MEP. This phenomenon has been established and demonstrated in Theorem 5. When the SNR is fixed at 12 dB, transitioning from $n = \log_2(M) + 2$ to an infinite number of bits, the utilization of a low-resolution ADC at the receiver incurs a quantization penalty of approximately $\Psi(12, \text{dB}, 4, 4) \approx 0.35$, $\Psi(12, \text{dB}, 8, 5) \approx 0.19$, and $\Psi(12, \text{dB}, 16, 6) \approx 0.04$. Consequently, this leads to an increase in the MEP by a factor of 1.08, 1.04, and 1.01, respectively. This result indicates that the increase in MEP resulting from using $n = \log_2 M + 2$ instead of ∞ is exceedingly minor, as opposed to using $n = \log_2 M$ or $n = \log_2 M + 1$. Importantly, this finding also provides a key design guideline and a practical rule of thumb for utilizing low-resolution ADCs at the receiver. It provides guidance on determining the sufficient number of bits needed to achieve nearly identical performance to that of infinite-bit quantization. We demonstrate this finding further in Fig. 3 for a wide range of SNR values.
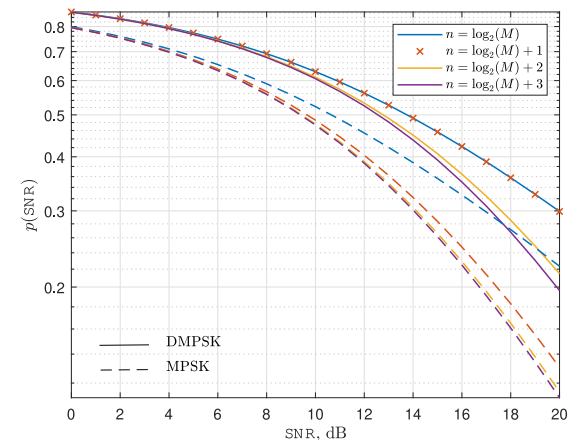
Another perspective on this matter involves comparing the additional SNR required to attain the same MEP when utilizing $(n-1)$-bit quantization instead of *n*-bit quantization. We
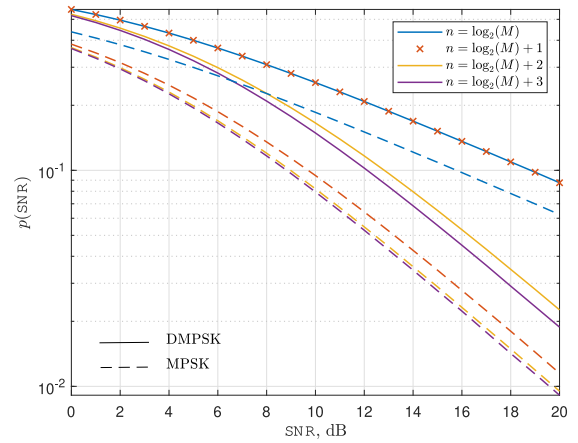
(a) $M = 4$
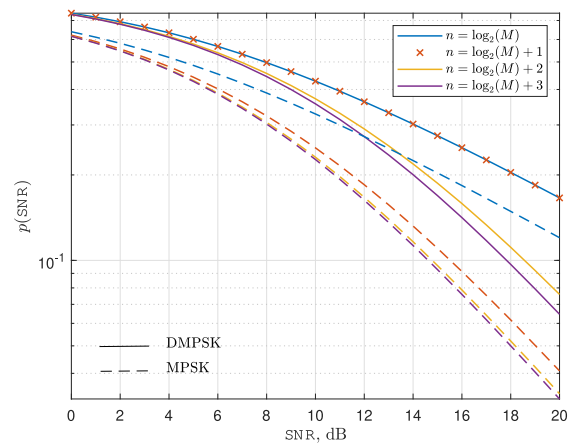
(b) $M = 8$

(c) $M = 16$

FIGURE 4. MEP curves as a function of SNR under coherent detection with MPSK modulation [5] and non-coherent detection with D-MPSK modulation. *M* = 4, 8, 16 and *n* = log₂ *M*, log₂ *M* + 1, log₂ *M* + 2, log₂ *M* + 3.

quantify the corresponding increment in SNR according to

$$\Phi(\text{MEP}, M, n) = 10 \log\left(\frac{\text{SNR}_{n-1}}{\text{SNR}_n}\right), \qquad (21)$$

where SNR_n and SNR_{n-1} are the SNR values required to achieve a specific MEP with n and $n-1$ quantization bits, respectively. As an example case, if we fix the MEP at 0.03 and transition from $n = 4$ to $n = 5$ for $M = 4$, we find that $\Phi(0.03, 4, 5) \approx 0.8$ dB. This indicates that to achieve an MEP of 0.03 with 4-bit quantization, we require 0.8 dB higher transmit power compared to using 5-bit quantization. This metric provides a means to balance the MEP and power consumption to achieve a desired reliability level for the communication link. By utilizing the number of quantization bits as a lever, we can effectively trade off between MEP and power consumption.

C. EFFECT OF THE AVAILABILITY OF CSI AT THE RECEIVER

Next, we focus on the effect of CSI availability at the receiver on the MEP performance of the system. As we established in Theorem 5, the optimum D-MPSK detector with low-resolution quantization at the receiver attains the full DVO when $n = \log_2 M + 2$ and performs as good as the optimum coherent detector in the DVO sense (see [5] for the coherent detection case). Besides the asymptotic DVO analysis provided in Theorem 5, the performance comparison of non-coherent and coherent systems requires further investigation to discover the additional SNR requirement for the D-MPSK to overcome the lack of channel knowledge at the receiver for finite SNR values. This is what we will investigate in this part.

Fig. 4 displays the plot of MEP versus SNR for MPSK with coherent detection and D-MPSK with non-coherent detection. In order to generate the MEP curves for the MPSK modulation with coherent detection, we use the symbol error probability expressions given in [5]. The MEP curves for D-MPSK modulation with non-coherent detection are obtained by using Theorem 4.

It is evident that when $n = \log_2 M$, the value of DVO is $\frac{1}{2}$ regardless of whether CSI is available at the receiver or not. However, for $n = \log_2 M + 1$, the value of DVO varies considerably depending on the availability of CSI. Specifically, to attain the maximum DVO with non-coherent detection, the quantizer must possess at least $\log_2 M + 2$ bits, whereas, with coherent detection, $\log_2 M + 1$ bits are sufficient. Therefore, in the absence of CSI at the receiver, an extra bit is necessary at the quantizer to achieve the full DVO. This additional bit can be considered a penalty for not having CSI at the receiver.

We also observe that there is an increase in MEP for a given SNR with D-MPSK modulation compared to MPSK modulation. We can quantify this increase in the MEP as

$$\Psi'(\mathsf{SNR}, M, n) = 10 \log \left(\frac{p(\mathsf{SNR}, M, n)_{\text{D-MPSK}}}{p(\mathsf{SNR}, M, n)_{\text{MPSK}}} \right), \quad (22)$$

where $p(\mathsf{SNR}, M, n)_{\text{D-MPSK}}$ and $p(\mathsf{SNR}, M, n)_{\text{MPSK}}$ is the MEP for M-DPSK and MPSK modulations with n-bit quantization, respectively.
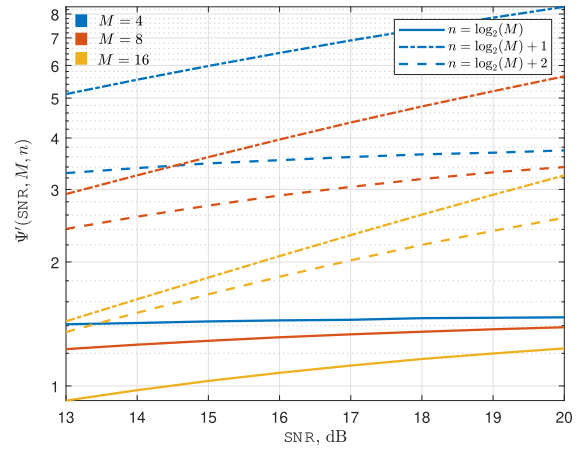

FIGURE 5. $\Psi'(\mathsf{SNR}, M, n)$ as a function of SNR. $M = 4, 8, 16$ and $n = \log_2 M, \log_2 M + 1, \log_2 M + 2$.

As depicted in Figure 5, when the SNR is maintained at a constant level of 18 dB, the transition from QPSK to D-QPSK (i.e., $M = 4$) yields $\Psi'(18\text{dB}, 4, 2) \approx 1.46$, $\Psi'(18\text{dB}, 4, 3) \approx 7.37$, and $\Psi'(18\text{dB}, 4, 4) \approx 3.35$, resulting in a 1.4-fold, 5.46-fold, and 2.16-fold increase in MEP. This example clearly demonstrates that the maximum discrepancy in MEP between QPSK and D-QPSK modulations occurs at $n = 3$. This is because we can achieve the full DVO of 1 with $n = 3$ bits if CSI is available at the receiver, while the $\mathsf{DVO} = \frac{1}{2}$ when the receiver does not have access to CSI and $n = 3$ (see Theorem 5). This finding can be generalized to modulation orders M greater than 4 since the coherent system achieves the full DVO at $n = \log_2 M + 1$ while the non-coherent system still operates at $\mathsf{DVO} = \frac{1}{2}$ at $n = \log_2 M + 1$.

Furthermore, we observe that to attain an equivalent MEP in D-MPSK modulation with n-bit quantization compared to MPSK modulation, a higher SNR must be maintained. In this context, we will examine another penalty metric to quantitatively measure the additional SNR required. To that end, we define

$$\Phi'(\mathsf{MEP}, M, n) = 10 \log \left(\frac{\mathsf{SNR}_{\text{D-MPSK}}}{\mathsf{SNR}_{\text{MPSK}}} \right), \quad (23)$$

where $\mathsf{SNR}_{\text{D-MPSK}}$ and $\mathsf{SNR}_{\text{MPSK}}$ are the SNR required to achieve a certain MEP with D-MPSK and MPSK modulations, respectively. If we consider a fixed MEP value of 0.1, the transition from QPSK to D-QPSK results in $\Phi'(0.1, 4, 2) \approx 3$ dB, $\Phi'(0.1, 4, 3) \approx 9$ dB, and $\Phi'(0.1, 4, 4) \approx 3$ dB. This indicates that achieving a MEP of 0.1 with D-QPSK and 3-bit quantization requires eight times more transmit power compared to QPSK modulation with 3-bit quantization. However, when employing 4-bit quantization, only twice the transmit power of QPSK modulation is needed with D-QPSK modulation. That is, when we have enough number of bits to achieve the full DVO in both cases, the SNR penalty reduces to the usual 3 dB penalty that we have in the ∞-bit quantization case.
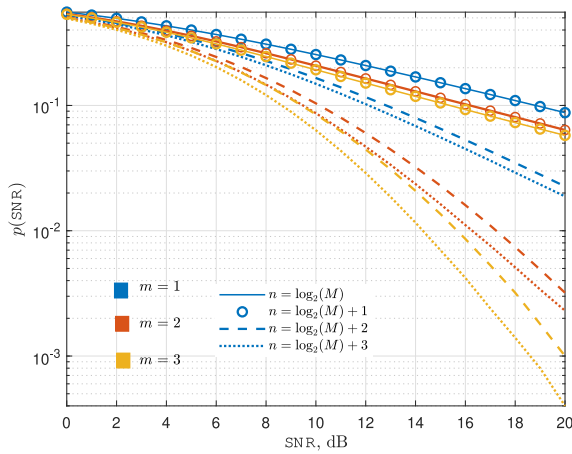
FIGURE 6. MEP curves as a function of SNR with D-QPSK modulation under Nakagami-m fading. $n = 2, 3, 4, 5$, $m = 1, 2, 3$.

FIGURE 7. PEP curves as a function of SNR with D-QPSK modulation. $n = 2, 3, 4, 5$. $t_e = 12, 27$.

Surprisingly, this is also the case when both coherent and non-coherent systems do not have enough number of quantization bits to achieve the full DVO. The maximum SNR penalty occurs at $n = \log_2 M + 1$, where the coherent system operates at the full DVO $= 1$ whereas the non-coherent system still needs an additional bit to achieve DVO $= 1$.

D. COMPARISON UNDER NAKAMAGI-M FADING

In this section, we will compare and contrast the system performance results of the detector presented in Theorem 2 under different fading conditions to Rayleigh fading under which the optimum detector was derived. To that end, we will consider the Nakagami-m fading model which characterizes a broad range of fading phenomena ranging from severe to moderate and no fading conditions as the shape parameter m varies over $[0.5, \infty)$ [60] and it reduces to Rayleigh fading for $m = 1$. We set the spread parameter $\Omega = 1$ in our simulations to make sure that H has unit power.

We have used the Monte Carlo simulation method to generate the simulated results in Fig. 6 which provides performance comparison curves for Nakagami-m fading under D-QPSK modulation with $n = 2, 3, 4, 5$ and $m = 1, 2, 3$.

These performance curves indicate that the MEP for $n = \log_2 M$ and $n = \log_2 M + 1$ cases are nearly identical, and the decay slope of MEP is similar to that under Rayleigh fading (i.e., $\frac{1}{2}$ for $m = 1, 2, 3$). On the other hand, when n is increased to $\log_2 M + 2$, there is a significant improvement in the MEP with a corresponding change in the decay slope of MEP to m. This observation suggests a change in the diversity order and that D-MPSK with two additional bits (besides $\log_2 M$) achieves the maximum diversity order of m under Nakagami-m fading. One of these bits serves to mitigate the impact of low-resolution quantization at the receiver [5], while the other bit compensates for the absence of channel knowledge.
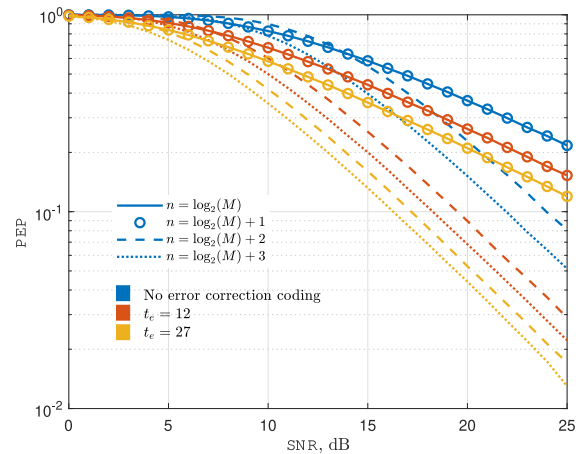
E. EFFECT OF CHANNEL CODING AND PACKET ERROR PROBABILITY

Finally, we investigate the performance of the optimum detectors with channel coding. To this end, we focus on the simple linear Bose-Chaudhuri-Hocquenghem (BCH) block codes [61]. The packet error probability (PEP) for D-QPSK modulation with BCH codes, used as a means of forward error correction, is plotted as a function of SNR in Fig. 7. The results are obtained by using the optimum block-2 detector in Theorem 2 by decoding each transmitted packet symbol-by-symbol and using a BCH code with a 255-bit codeword that can correct up to t_e bit errors. The curves in this figure clearly demonstrate that communication reliability increases significantly by using coded transmissions. However, we note that achieving greater error-correcting capability comes at the cost of an increase in the number of redundant bits and a decrease in the system goodput.

VIII. CONCLUSION

In this work, we have studied differential encoding to overcome the channel estimation challenge in low-resolution ADC-based communication systems. We derived the optimum maximum likelihood block-2 detectors, characterized their message error probability performance and established the optimality of block-2 detectors among the class of block-L detectors when $n = \log_2 M$. We performed an analytical and numerical study to illustrate the performance of the optimum detectors. We derived MEP for D-MPSK modulation under Rayleigh fading and conducted an asymptotic analysis to investigate the corresponding change in the slope of decay for MEP with SNR. Our results indicate that D-MPSK requires two extra bits on top of $\log_2 M$ to achieve the maximum diversity order. The future extensions of the current work will include an investigation of the system performance under more general fading scenarios and multi-antenna transmissions.

APPENDIX A
PROOF OF LEMMA 1

First, consider the case $t_0 \mod N = N - 1$. Then, we have Y_{t_0} and Y_{t_1} in two different fading blocks and are therefore subject to two different iid fading coefficients H_k and H_{k+1} for some k satisfying $t_0 = (k+1)N - 1$. Hence, we can express Y_{t_0} and Y_{t_1} in this case according to

$$Y_{t_0} = \sqrt{\mathsf{SNR}}H_k S_{t_0} + W_{t_0}$$

and

$$Y_{t_1} = \sqrt{\mathsf{SNR}}H_{k+1} S_{t_0} e^{J\frac{2\pi}{M}\lambda} + W_{t_1}.$$

Let $\tilde{H}_k = H_k S_{t_0}$ and $\tilde{H}_{k+1} = H_{k+1} S_{t_0} e^{J\frac{2\pi}{M}\lambda}$. Both $\tilde{H}_k$ and $\tilde{H}_{k+1}$ are distributed according to $\mathcal{CN}(0,1)$ since the symbols belong to $\mathcal{C} = \{e^{J\frac{\pi}{M}(2k+1)}\}_{k=0}^{M-1}$ for all $t \in \mathbb{N}$ (i.e., symbols only introduce a phase rotation) and the channel coefficients are circularly symmetric. $\tilde{H}_k$ and $\tilde{H}_{k+1}$ are also independent with joint distribution $\mathcal{CN}(0,\boldsymbol{I}_2)$ since H_k and $H_{k+1}e^{J\frac{2\pi}{M}\lambda}$ are jointly distributed according to $\mathcal{CN}(0,\boldsymbol{I}_2)$ and the phase rotation introduced by S_{t_0} does not change the form of this distribution. This shows that Y_{t_0} and Y_{t_1} are iid complex random variables with common marginal distribution $\mathcal{CN}(0, 1+\mathsf{SNR})$, and their joint distribution is given by $\mathcal{CN}(0, (1+\mathsf{SNR})\boldsymbol{I}_2)$. This completes the proof for the case $t_0 \mod N = N - 1$.

For the other case $t_0 \mod N \in [0 : N-2]$, Y_{t_0} and Y_{t_1} lie in the same fading block, and they can be represented by

$$Y_{t_0} = \sqrt{\mathsf{SNR}}\tilde{H}_k + W_{t_0}$$

and

$$Y_{t_1} = \sqrt{\mathsf{SNR}}\tilde{H}_k e^{J\frac{2\pi}{M}\lambda} + W_{t_1}$$

for some $k \in \mathbb{N}$ satisfying $t_0 \in [kN : (k+1)N - 2]$ and $\tilde{H}_k = H_k S_{t_0}$. This establishes that Y_{t_0} and Y_{t_1} are jointly Gaussian since they are a linear combination of $\mathcal{CN}(0,1)$ random variables [59]. The circularly symmetry property follows from being the mean vector $\mathsf{E}[Y]$ and the pseudo-covariance matrix $\mathsf{E}[YY^\top]$ both equal to zero. After straightforward calculations, the covariance matrix $\boldsymbol{K} = \mathsf{E}[YY^\dagger]$ can be derived as in (3).

APPENDIX B
PROOF OF LEMMA 3

The following lemma will be used in our proof.

Lemma 4: Let $I = [a, b)$ be an interval satisfying $b - a < 2\pi$ and $b \mod 2\pi \geq a \mod 2\pi$. Then, $I \mod 2\pi = [a \mod 2\pi, b \mod 2\pi)$.

Proof: Let $\beta = b \mod 2\pi$ and $\alpha = a \mod 2\pi$. Since $b - a < 2\pi$ and $\beta \geq \alpha$, we can write $a = k \cdot 2\pi + \alpha$ and $b = k \cdot 2\pi + \beta$ for some $k \in \mathbb{Z}$. For the same k, any $e \in I$ can also be expressed as $e = k \cdot 2\pi + \eta$ for some $\eta \in [\alpha, \beta)$.. This shows $e \mod 2\pi \in [a \mod 2\pi, b \mod 2\pi)$. for all $e \in I$ and therefore $I \mod 2\pi \subseteq$

$[a \mod 2\pi, b \mod 2\pi)$. The reverse direction $[a \mod 2\pi, b \mod 2\pi) \subseteq I \mod 2\pi$ follows from the same arguments. ∎

For $n = \log_2 M$, the proof follows directly by inspecting the definitions of Q-regions and A-regions. Thus, we will focus on the case $n \geq \log_2 M + 1$ for the rest of the proof.

For $n \geq \log_2 M + 1$, we express the kth A-region according to $A_k = \bigcup_{i=k-\frac{2^n}{2M}}^{k+\frac{2^n}{2M}} B_i$, where $B_i = \{z \in \mathbb{C} : \mathsf{Arg}(z) \in I_i \mod 2\pi\}$ for intervals I_i in the form

$$I_i = \left[\frac{2\pi}{2^n}i, \frac{2\pi}{2^n}(i+1)\right)$$

for $i \in \left[k + 1 - \frac{2^n}{2M} : k - 1 + \frac{2^n}{2M}\right]$,

$$I_{k-\frac{2^n}{2M}} = \left[\frac{2\pi}{2^n}\left(k - \frac{2^n}{2M}\right) + \frac{\pi}{2^n}, \frac{2\pi}{2^n}\left(k + 1 - \frac{2^n}{2M}\right)\right)$$

for $i = k - \frac{2^n}{2M}$, and

$$I_{k+\frac{2^n}{2M}} = \left[\frac{2\pi}{2^n}\left(k + \frac{2^n}{2M}\right), \frac{2\pi}{2^n}\left(k + \frac{2^n}{2M}\right) + \frac{\pi}{2^n}\right)$$

for $i = k + \frac{2^n}{2M}$.

For $i \in [k + 1 - \frac{2^n}{2M} : k - 1 + \frac{2^n}{2M}]$, let $\alpha_i = \frac{2\pi}{2^n}i \mod 2\pi$ and $\beta_i = \frac{2\pi}{2^n}(i+1) \mod 2\pi$. This implies $i = \hat{k} \cdot 2^n + \frac{2^n}{2\pi}\alpha_i$ for some $\hat{k} \in \mathbb{Z}$, which further implies $\alpha_i = \frac{2\pi}{2^n}(i \mod 2^n)$. Similarly, $\beta_i = \frac{2\pi}{2^n}(i+1 \mod 2^n)$.

If $i \mod 2^n \in [0 : 2^n - 2]$, $i+1 \mod 2^n = i \mod 2^n + 1$ and $\beta_i \geq \alpha_i$. Hence, using Lemma 4, we can write

$$B_i = \left\{z \in \mathbb{C} : \mathsf{Arg}(z) \in \frac{2\pi}{2^n}[i \mod 2^n, i \mod 2^n + 1)\right\}$$
$$= Q_{i \mod 2^n} \tag{24}$$

If $i \mod 2^n = 2^n - 1$, we can write $I_i \mod 2\pi$ as

$$I_i \mod 2\pi = \bigcup_{m=n}^{\infty}\left[\frac{2\pi}{2^n}i, \frac{2\pi}{2^n}(i+1) - \frac{1}{m}\right) \mod 2\pi$$
$$= \bigcup_{m=n}^{\infty}\left[\frac{2\pi}{2^n}(i \mod 2^n), \frac{2\pi}{2^n}(i \mod 2^n + 1) - \frac{1}{m}\right)$$
$$= \frac{2\pi}{2^n}[i \mod 2^n, i \mod 2^n + 1)$$
$$= Q_{i \mod 2^n},$$

where the second equality follows from application of Lemma 4 and writing $2\pi - \frac{1}{m}$ as $\frac{2\pi}{2^n}(i \mod 2^n + 1) - \frac{1}{m}$ since $i \mod 2^n = 2^n - 1$.

For $i = k - \frac{2^n}{2M}$ and $i = k + \frac{2^n}{2M}$, similar arguments show that $B_{k-\frac{2^n}{2M}} = Q^U_{(k-\frac{2^n}{2M}) \mod 2^n}$ and $B_{k+\frac{2^n}{2M}} = Q^L_{(k+\frac{2^n}{2M}) \mod 2^n}$. Combining all these results, we arrive at the decomposition of A_k given in (6) for all $k \in [0 : 2^n - 1]$.

APPENDIX C
PROOF OF THEOREM 1

We will first obtain some preliminary results that will assist in the proof of this theorem. Below, for a given set $A \subseteq \mathbb{C}$, we use $\mathsf{int}(A)$ to represent its interior points.

Lemma 5: For any $z \in \mathbb{C}$ and $k \in [0 : 2^n - 1]$, there exists a unique $\lambda \in \Lambda$ such that $z e^{J \frac{2\pi}{M} \lambda} \in A_k$. Further, if $z e^{J \frac{2\pi}{M} \lambda} \in \text{int}(A_k)$ for some $\lambda \in \Lambda$, then $z e^{J \frac{2\pi}{M} \beta} \notin \text{cl}(A_k)$ for any $\beta \in \Lambda$ with $\beta \neq \alpha$.

Proof: The existence follows from the fact that A_k is a cone with angular width $\frac{2\pi}{M}$ and it includes one of its boundaries. Thus, there must exist at least one $\lambda \in \Lambda$ such that $z e^{J \frac{2\pi}{M} \lambda} \in A_k$. The uniqueness follows from the fact that the angle between any $u \in A_k$ and $v \in A_k$ is strictly smaller than $\frac{2\pi}{M}$.[5] If $z e^{J \frac{2\pi}{M} \lambda} \in A_k$ and $z e^{J \frac{2\pi}{M} \beta} \in A_k$ for two different λ and β in Λ, then the angle between these two points is greater than or equal to $\frac{2\pi}{M}$, which is a contradiction. The last part follows from the same angle arguments. ∎

An important implication of this lemma, which we will use in our proof, is that if $B \subseteq A_k$, then $e^{J \frac{2\pi}{M} \lambda} B \bigcap A_k = \emptyset$ for all $\lambda \in [1 : M - 1]$.

Lemma 6: For $k \in [0 : 2^n - 1]$ and $\lambda \in \Lambda$, $e^{J \frac{2\pi}{M} \lambda} Q_k = Q_{(k + \frac{2^n}{M} \lambda) \bmod 2^n}$.

Proof: By definition,

$$e^{J \frac{2\pi}{M} \lambda} Q_k = \left\{ z e^{J \frac{2\pi}{M} \lambda} : z \in Q_k \right\}$$

$$= \left\{ z \in \mathbb{C} : \text{Arg}(z) \in \left[\frac{2\pi}{2^n} i, \frac{2\pi}{2^n} (i+1) \right) \bmod 2\pi \right\},$$

where $i = k + \frac{2^n}{M} \lambda$. Hence, $e^{J \frac{2\pi}{M} \lambda} Q_k$ is in the form of the sets B_i analyzed in Appendix B. The same arguments show that $e^{J \frac{2\pi}{M} \lambda} Q_k = Q_{(k + \frac{2^n}{M} \lambda) \bmod 2^n}$. ∎

For the proof, we will only consider the case $t_0 \bmod N \in [0 : N - 2]$, i.e., Y_{t_0} and Y_{t_1} are in the same fading block. In the other case, there is nothing to prove since channel outputs are independent (i.e., see Lemma 1) and all input messages have the same likelihood.

Assume $n \geq \log_2 M + 1$. For $n = \log_2 M$, the arguments are the same and simpler. Now, by Lemma 5, there exists at least one $\lambda \in \Lambda$ such that $e^{J \frac{2\pi}{M} \lambda} Q_{k_0} \bigcap A_{k_1} \neq \emptyset$, and hence $\lambda^\star(\boldsymbol{k})$ is well-defined. By Lemma 6, $e^{J \frac{2\pi}{M} \lambda^\star(\boldsymbol{k})} Q_{k_0} = Q_{(k_0 + \frac{2^n}{M} \lambda^\star(\boldsymbol{k})) \bmod 2^n}$. By Lemma 3, we have (25), shown at the bottom of the next page.

First, consider the case $(k_0 + \frac{2^n}{M} \lambda^\star(\boldsymbol{k})) \bmod 2^n \in [k_1 - \frac{2^n}{2M} + 1 : k_1 + \frac{2^n}{2M} - 1] \bmod 2^n$. Then, we have

$$e^{J \frac{2\pi}{M} \lambda^\star(\boldsymbol{k})} Q_{k_0} = Q_{\left(k_0 + \frac{2^n}{M} \lambda^\star(\boldsymbol{k})\right) \bmod 2^n} \subseteq \text{int}(A_{k_1}).$$

Thus, by Lemma 5, $e^{J \frac{2\pi}{M} \lambda} Q_{k_0} \bigcap \text{cl}(A_{k_1}) = \emptyset$ for all $\lambda \in \Lambda$ with $\lambda \neq \lambda^\star(\boldsymbol{k})$. This observation has the following important implication. Let $g(y_0; \lambda) \triangleq \int_{Q_{k_1}} f_{Y_{t_1} | Y_{t_0}}(y_1 | y_0; \lambda) dy_1$. Using Lemma 2, we have $g(y_0; \lambda^\star(\boldsymbol{k})) > g(y_0; \lambda)$ point-wise for all $y_0 \in Q_{k_0}$ and for all $\lambda \neq \lambda^\star(\boldsymbol{k})$. This shows

$$\mathcal{L}(\lambda^\star(\boldsymbol{k}); \boldsymbol{k}) = \int_{Q_{k_0}} f_{Y_{t_0}}(y_0) g(y_0; \lambda^\star(\boldsymbol{k})) dy_0$$

$$> \int_{Q_{k_0}} f_{Y_{t_0}}(y_0) g(y_0; \lambda) dy_0$$

$$= \mathcal{L}(\lambda; \boldsymbol{k})$$

---

5. As is usual, the angle between two complex numbers u and v is defined as $\text{Angle}(u, v) = \cos^{-1}\left(\frac{\text{Re}(uv^*)}{|u||v|} \right)$.

if $\lambda \in \Lambda$ and $\lambda \neq \lambda^\star(\boldsymbol{k})$. This analysis also shows that $\lambda^\star(\boldsymbol{k})$ is the unique MLE in this case.

If $(k_0 + \frac{2^n}{M} \lambda^\star(\boldsymbol{k})) \bmod 2^n = (k_1 - \frac{2^n}{2M}) \bmod 2^n$, then $e^{J \frac{2\pi}{M} \lambda^\star(\boldsymbol{k})} Q_{k_0} = Q_{(k_1 - \frac{2^n}{2M}) \bmod 2^n}$ and only one half of $e^{J \frac{2\pi}{M} \lambda^\star(\boldsymbol{k})} Q_{k_0}$ (upper one) lies in A_{k_1}. In this case, we must have $\lambda^\star(\boldsymbol{k}) \in [0 : M - 2]$ since if $\lambda^\star(\boldsymbol{k}) = M - 1$, we have the following chain of implications for some $p, q \in \mathbb{Z}$ and $k' \in [0 : 2^n - 1]$:

$$k_0 - \frac{2^n}{M} = p \cdot 2^n + k'$$

$$k_1 - \frac{2^n}{2M} = q \cdot 2^n + k'$$

$$\Rightarrow k_0 = \left(k_1 + \frac{2^n}{2M} \right) + (p - q) 2^n$$

$$\Rightarrow k_0 = \left(k_1 + \frac{2^n}{2M} \right) \bmod 2^n.$$

Therefore, $Q_{k_0} \bigcap A_{k_1} \neq \emptyset$ and this contradicts with the definition of $\lambda^\star(\boldsymbol{k})$ as being the minimum element in Λ such that $e^{J \frac{2\pi}{M} \lambda} Q_{k_0}$ intersects with A_{k_1}. Using modular arithmetic, it can also be seen that $(k_0 + \frac{2^n}{M}(\lambda^\star(\boldsymbol{k}) + 1)) \bmod 2^n = (k_1 + \frac{2^n}{2M}) \bmod 2^n$, and therefore again only one half of $e^{J \frac{2\pi}{M}(\lambda^\star(\boldsymbol{k}) + 1)} Q_{k_0}$ (lower one) lies in A_{k_1}. Using these observations and the circular symmetry in the problem, we conclude that $\mathcal{L}(\lambda^\star(\boldsymbol{k}); \boldsymbol{k}) = \mathcal{L}(\lambda^\star(\boldsymbol{k}) + 1; \boldsymbol{k})$ when $(k_0 + \frac{2^n}{M} \lambda^\star(\boldsymbol{k})) \bmod 2^n = (k_1 - \frac{2^n}{2M}) \bmod 2^n$. For other values of λ in Λ not equal to $\lambda^\star(\boldsymbol{k})$ or $\lambda^\star(\boldsymbol{k}) + 1$, the arguments similar to the ones above show that $\mathcal{L}(\lambda^\star(\boldsymbol{k}); \boldsymbol{k}) > \mathcal{L}(\lambda; \boldsymbol{k})$. Hence, $\lambda^\star(\boldsymbol{k})$ is an MLE for this case as well.

If $(k_0 + \frac{2^n}{M} \lambda^\star(\boldsymbol{k})) \bmod 2^n = (k_1 + \frac{2^n}{2M}) \bmod 2^n$, then $\lambda^\star(\boldsymbol{k}) = 0$ and

$$\mathcal{L}(0; \boldsymbol{k}) = \mathcal{L}(M - 1; \boldsymbol{k}) > \mathcal{L}(\lambda; \boldsymbol{k})$$

for all $\lambda \in \Lambda$ not equal to 0 or $M - 1$ (using the same arguments). Hence, $\lambda^\star(\boldsymbol{k})$ is also an MLE in this final case. Combining all of these results, we conclude that $\lambda^\star(\boldsymbol{k})$ is an MLE for the message estimation problem (8) for $L = 2$.

APPENDIX D
PROOF OF THEOREM 2

We will only consider $n \geq \log_2 M + 1$ since (10) and (11) agree with each other for $n = \log_2 M$. The proof for $n \geq \log_2 M + 1$ follows from considering three difference cases for the range of k_1 : (i) $k_1 \in [\frac{2^n}{2M} : 2^n - \frac{2^n}{2M} - 1]$, (ii) $k_1 \in [0 : \frac{2^n}{2M} - 1]$ and (iii) $k_1 \in [2^n - \frac{2^n}{2M} : 2^n - 1]$.

If $k_1 \in [\frac{2^n}{2M} : 2^n - \frac{2^n}{2M} - 1]$, then we can write $[k_1 - \frac{2^n}{2M} : k_1 + \frac{2^n}{2M}] \bmod 2^n = [k_1 - \frac{2^n}{2M} : k_1 + \frac{2^n}{2M}]$. Assume first $k_0 \in [0 : k_1 - \frac{2^n}{2M} - 1]$. Using Corollary 1, we need

$$k_0 + \frac{2^n}{M} \lambda \geq k_1 - \frac{2^n}{2M}$$

$$\Rightarrow \frac{2^n}{M} \lambda \geq k_1 - k_0 - \frac{2^n}{2M}$$

$$\Rightarrow \lambda \geq \frac{M}{2^n}(k_1 - k_0) - \frac{1}{2}. \tag{26}$$

The minimum integer value satisfying (26) can be equivalently written according to

$$\lambda^{\star\star}(\boldsymbol{k}) = \left\lceil \frac{M}{2^n}(k_1 - k_0) - \frac{1}{2} \right\rceil \mod M.$$

We note that $\left\lceil \frac{M}{2^n}(k_1 - k_0) - \frac{1}{2} \right\rceil$ is always an element of $[1 : M - 1]$, and the mod M operation does not change anything in this case.

Secondly, assume that $k_0 \in [k_1 - \frac{2^n}{2M} : k_1 + \frac{2^n}{2M} - 1]$. Then, $\lambda^{\star}(\boldsymbol{k}) = 0$, and we can upper and lower bound $\left\lceil \frac{M}{2^n}(k_1 - k_0) - \frac{1}{2} \right\rceil$ as

$$\left\lceil -1 + \frac{M}{2^n} \right\rceil = 0 \leq \left\lceil \frac{M}{2^n}(k_1 - k_0) - \frac{1}{2} \right\rceil \leq \lceil 0 \rceil = 0,$$

where the lower bound is obtained for $k_0 = k_1 + \frac{2^n}{2M} - 1$ and the upper bound is obtained for $k_0 = k_1 - \frac{2^n}{2M}$. Hence, $\lambda^{\star\star}(\boldsymbol{k})$ coincides with $\lambda^{\star}(\boldsymbol{k})$ in this case as well, and therefore it is an MLE.

For $k_0 = k_1 + \frac{2^n}{2M}$, we have $\lambda^{\star\star}(\boldsymbol{k}) = M - 1 \neq \lambda^{\star}(\boldsymbol{k}) = 0$. However, we have proven in Appendix C that both message indices 0 and $M - 1$ maximize the likelihood function in this case, and hence $\lambda^{\star\star}(\boldsymbol{k})$ is still an MLE.

Finally, we consider the case $k_0 \in [k_1 + \frac{2^n}{M} + 1 : 2^n - 1]$. Using Corollary 1, we require to have

$$k_0 + \frac{2^n}{M}\lambda - 2^n \geq k_1 - \frac{2^n}{2M}$$

in this case. This implies we must have $\lambda^{\star}(\boldsymbol{k}) - M = \left\lceil \frac{M}{2^n}(k_1 - k_0) - \frac{1}{2} \right\rceil$ for the first time the above inequality is satisfied. This is equivalent to writing $\lambda^{\star}(\boldsymbol{k}) = \left\lceil \frac{M}{2^n}(k_1 - k_0) - \frac{1}{2} \right\rceil \mod M$, which coincides with $\lambda^{\star\star}(\boldsymbol{k})$ given in Theorem 2.

The remaining cases $k_1 \in [0 : \frac{2^n}{2M} - 1]$ and $k_1 \in [2^n - \frac{2^n}{2M} : 2^n - 1]$ can be analyzed similarly to complete the proof.

APPENDIX E
PROOF OF THEOREM 3
Consider a genie-aided decoder where the genie provides the perfect channel information $\{H_k\}_{k=0}^{\infty}$ and the decoder utilizes this information to maximize the likelihood of message indices $\hat{\boldsymbol{\lambda}} = [\hat{\lambda}_1, \ldots, \hat{\lambda}_{L-1}]^{\top}$ as an estimate for the actual message sequence $\boldsymbol{\lambda} = [\lambda_{t_1}, \ldots, \lambda_{t_{L-1}}]^{\top}$ transmitted from t_1 to t_{L-1}. At fading block transition times, the decoder makes a random guess about the corresponding input message. Below, we will assume that the quantized channel output vector $\boldsymbol{k} = [k_0, \ldots, k_{L-1}]$ corresponds to the observations on the same channel fading block, i.e., $t_0 \in [kN : (k+1)N - L]$ for some $k \in \mathbb{N}$. Otherwise, we can repeat the same arguments by dividing $\boldsymbol{k} = [k_0, \ldots, k_{L-1}]$ into two different blocks at channel transition times.

We write the likelihood function $\mathcal{L}\left(\hat{\boldsymbol{\lambda}}, s; h\right)$ for the genie-aided detector as

$$\mathcal{L}\left(\hat{\boldsymbol{\lambda}}, s; h\right) = \Pr\left\{\boldsymbol{q} = \boldsymbol{k} | \boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}, H_k = h, S_{t_0} = s\right\}.$$

We note that the value of the first symbol in the block is also a parameter for the likelihood function for the genie-aided detector since when channel knowledge is available at the detector, the most likely scenario that will lead to the observation $\boldsymbol{q} = \boldsymbol{k}$ also depends on our estimate for S_{t_0}. This was not the case for the detectors investigated in Section III, and hence was not shown explicitly in that part of the paper. See Lemma 1 as well that shows the channel output distribution for consecutive samples in time does only depend on the phase difference between S_{t_1} and S_{t_0} when no channel state information is available at the receiver.

Given $\boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}, H_k = h$ and $S_{t_0} = s$, the channel outputs $Y_{t_0}, \ldots, Y_{t_{L-1}}$ are independent, and we can further write $\mathcal{L}(\hat{\boldsymbol{\lambda}}, s; h)$ according to

$$\mathcal{L}\left(\hat{\boldsymbol{\lambda}}, s; h\right)$$
$$= \prod_{i=0}^{L-1} \Pr\left\{Y_{t_i} \in Q_{k_i} | \boldsymbol{\lambda} = \hat{\boldsymbol{\lambda}}, H_k = h, S_{t_0} = s\right\}.$$

Given this information, we also have $Y_{t_0} \sim \mathcal{CN}(hs, 1)$ and $Y_{t_i} \sim \mathcal{CN}(hse^{J\frac{2\pi}{M}\sum_{j=1}^{i}\hat{\lambda}_{t_j}}, 1)$ for $i \in [1 : L - 1]$. Hence, using Lemma 2 and Lemma 3 (i.e., $Q_k = A_k$ for $k \in [0 : 2^n - 1]$ in this case), we can see that $\mathcal{L}(\hat{\boldsymbol{\lambda}}, s; h)$ is maximized when $hs \in Q_{k_0}$ and $hse^{J\frac{2\pi}{M}\sum_{j=1}^{i}\hat{\lambda}_{t_j}} \in Q_{k_i}$ for $i \in [1 : L - 1]$. Using Lemma 5, for each $h \in \mathbb{C}$, we know that there exists a unique $s^{\star} \in \mathcal{C}$ and $\boldsymbol{\lambda}^{\star} \in \Lambda^{L-1}$ that will make the means of Y_{t_i} lie in Q_{k_i} for all $i \in [0 : L - 1]$. This shows

$$\mathcal{L}\left(\hat{\boldsymbol{\lambda}}, s; h\right) \leq \prod_{i=0}^{L-1} \Pr\left\{Y_{t_i} \in Q_{k_i} | \boldsymbol{\lambda} = \boldsymbol{\lambda}^{\star}, H_k = h, S_{t_0} = s^{\star}\right\},$$

where $\boldsymbol{\lambda}^{\star}$ and $s^{\star}$ are chosen as described above for given $h \in \mathbb{C}$. The genie-aided detector achieves the upper bound on the likelihood function by first estimating S_{t_0} as $s^{\star}$ and then calculating the optimum differential phase shifts to have the means of channel outputs in Q_{k_i}'s.

We conclude the proof by observing iteratively that if $hs \in Q_{k_0}$ and $hse^{J\frac{2\pi}{M}\sum_{j=1}^{i}\lambda_{t_j}^{\star}} \in Q_{k_i}$, then $\lambda_{t_i}^{\star} = (k_i - k_{i-1}) \mod M$ for $i \in [1 : L - 1]$. The last step follows from the similar modular arithmetic arguments used in Appendix C and hence skipped to avoid repetitions. This is exactly the same MLE we arrived at in Theorems 1 and 2 for $n = \log_2 M$ when $L = 2$.

$$\left(k_0 + \frac{2^n}{M}\lambda^{\star}(\boldsymbol{k})\right) \mod 2^n \in \left[k_1 - \frac{2^n}{2M} : k_1 + \frac{2^n}{2M}\right] \mod 2^n. \tag{25}$$

APPENDIX F
PROOF OF COROLLARY 2

To start with, we first focus on the case where $n = 2$ and from equation (14), we have

$$p(\mathsf{SNR}) = 4\Pr\{Y_{t-1} \in Q_0, Y_t \notin Q_0 \mid \lambda_t = 0\}.$$

Conditioned on a particular realization of fading coefficient $H = re^{J\theta_h}$ (i.e., $|H| = r$ and $\mathsf{Arg}(H) = \theta_h$) and assuming $S_{t-1} = S_t = e^{J\theta_s}$, we write the conditional MEP as (27), shown at the bottom of the page, where $\theta = \theta_h + \theta_s$ and $\mathcal{Q}^2(\cdot)$ is the second power of $\mathcal{Q}(\cdot)$. The expression for $p(\mathsf{SNR})$ can be obtained by taking the average of the above expression with respect to the Rayleigh fading distribution $f_{|H|}(r) = 2re^{-r^2}$ for $r \geq 0$ and the uniform distribution $f_\Theta(\theta) = \frac{1}{2\pi}$ for the phase angle Θ. This averaging leads to the MEP expression presented in Corollary 2 for $n = 2$. Since $p(\mathsf{SNR})$ is the same for $n = 2$ and $n = 3$, as given in Theorem 4, the same MEP expression holds for $n = 3$ as well.

APPENDIX G
PROOF OF THEOREM 5

We will first start with a definition that will simplify the notation below.

Definition 1: We say a function f is *exponentially equal* to SNR^d if $\lim_{\mathsf{SNR} \to \infty} \frac{\log f(\mathsf{SNR})}{\log \mathsf{SNR}} = d$ for some $d \in \mathbb{R}$. We write $f(\mathsf{SNR}) \doteq \mathsf{SNR}^d$ to indicate exponential equality whenever this limit exists. Similarly, we also write $f(\mathsf{SNR}) \stackrel{e}{\leq} \mathsf{SNR}^d$ and $f(\mathsf{SNR}) \stackrel{e}{\geq} \mathsf{SNR}^d$ if $\lim_{\mathsf{SNR} \to \infty} \frac{\log f(\mathsf{SNR})}{\log \mathsf{SNR}} \leq d$ and $\lim_{\mathsf{SNR} \to \infty} \frac{\log f(\mathsf{SNR})}{\log \mathsf{SNR}} \geq d$, respectively.

We first focus on the scenario where $n = \log_2(M)$. In this case, we proved in Appendix E that the genie-aided

detector with perfect CSI and the block-2 optimum detector in Theorem 2 produce the same estimated sequence of message indices. Given perfect CSI, the channel outputs are independent and the genie-aided detector boils down to symbol-by-symbol detection to estimate the most likely input message indices. This is a similar setup investigated in detail in [5] for the coherent detection scenario. The same asymptotic analysis shows that the $\mathsf{DVO} = \frac{1}{2}$ for D-MPSK modulation with $n = \log_2(M)$.

Next, we focus on the case where $n = \log_2(M) + 1$. In this case, we will first provide the proof for the D-QPSK modulation and then extend it to general $M \geq 4$. From Corollary 2, we already know that $p(\mathsf{SNR})$ is identical for D-QPSK modulation with $n = 2$ and $n = 3$. Consequently, we can deduce that $\mathsf{DVO} = \frac{1}{2}$ for D-QPSK with $n = 3$.

In order to extend the proof for $n = \log_2(M) + 1$ for $M \geq 4$, we introduce both lower and upper bounds for MEP in equation (13) as

$$
\begin{aligned}
p(\mathsf{SNR}, 4, 3) &\leq p(\mathsf{SNR}, M, \log_2(M) + 1) \\
&\leq p(\mathsf{SNR}, M, \log_2 M),
\end{aligned}
\tag{28}
$$

where $p(\mathsf{SNR}, M, n)$ is obtained by re-writing $p(\mathsf{SNR})$ in (13) as a function of M and n. The bounds in (28) play a critical role in proving Theorem 5 for the case where $n = \log_2(M) + 1$. Specifically, these bounds exhibit an identical decay rate as SNR tends to infinity. This is evident as $p(\mathsf{SNR}, 4, 3) \stackrel{e}{=} \mathsf{SNR}^{-\frac{1}{2}}$ and $p(\mathsf{SNR}, M, \log_2 M) \stackrel{e}{=} \mathsf{SNR}^{-\frac{1}{2}}$. Consequently, we can establish that $\mathsf{SNR}^{-\frac{1}{2}} \stackrel{e}{\leq} p(\mathsf{SNR}, M, \log_2(M) + 1) \stackrel{e}{\leq} \mathsf{SNR}^{-\frac{1}{2}}$, which implies that $p(\mathsf{SNR}, M, \log_2(M) + 1) \stackrel{e}{=} \mathsf{SNR}^{-\frac{1}{2}}$. Therefore, we can

$$
\begin{aligned}
p(\mathsf{SNR}, r, \theta) &= 1 - 4\Pr\left\{\sqrt{\mathsf{SNR}}\, r\, e^{J\theta} + W_{t-1} \in Q_0, \sqrt{\mathsf{SNR}}\, r\, e^{J\theta} + W_t \in Q_0\right\} \\
&= 1 - 4\mathcal{Q}^2\left(-\sqrt{2\mathsf{SNR}}\, r\cos\theta\right)\mathcal{Q}^2\left(-\sqrt{2\mathsf{SNR}}\, r\sin\theta\right),
\end{aligned}
\tag{27}
$$

$$p(\mathsf{SNR}) = p_1(\mathsf{SNR}) + p_2(\mathsf{SNR}) + p_3(\mathsf{SNR}),\tag{29}$$

where

$$p_1(\mathsf{SNR}) = \frac{2^n}{\pi^2(2\mathsf{SNR}+1)}\int_0^\infty \int_0^{y_t^{\mathrm{re}}\tan\left(\frac{\pi}{2^{n-1}}\right)}\int_0^\infty \int_{y_{t-1}^{\mathrm{re}}\tan\left(\frac{2\pi}{M}+\psi\right)}^\infty \mathcal{G}(\mathbf{y})\, dy_{t-1}^{\mathrm{im}}\, dy_{t-1}^{\mathrm{re}}\, dy_t^{\mathrm{im}}\, dy_t^{\mathrm{re}},$$

$$p_2(\mathsf{SNR}) = \frac{2^n}{\pi^2(2\mathsf{SNR}+1)}\int_0^\infty \int_0^{y_t^{\mathrm{re}}\tan\left(\frac{\pi}{2^{n-1}}\right)}\int_0^\infty \int_{-\infty}^{y_{t-1}^{\mathrm{re}}\tan(\psi)} \mathcal{G}(\mathbf{y})\, dy_{t-1}^{\mathrm{im}}\, dy_{t-1}^{\mathrm{re}}\, dy_t^{\mathrm{im}}\, dy_t^{\mathrm{re}},$$

$$p_3(\mathsf{SNR}) = \frac{2^n}{\pi^2(2\mathsf{SNR}+1)}\int_0^\infty \int_0^{y_t^{\mathrm{re}}\tan\left(\frac{\pi}{2^{n-1}}\right)}\int_0^{-\infty} \int_\infty^{-\infty} \mathcal{G}(\mathbf{y})\, dy_{t-1}^{\mathrm{im}}\, dy_{t-1}^{\mathrm{re}}\, dy_t^{\mathrm{im}}\, dy_t^{\mathrm{re}}.$$

$$\Theta(1) = \int_0^\infty \int_0^{y_t^{\mathrm{re}}\tan\left(\frac{\pi}{2^{n-1}}\right)}\int_0^\infty \int_{y_{t-1}^{\mathrm{re}}\tan\left(\frac{2\pi}{M}+\psi\right)}^\infty \mathcal{G}'(\mathbf{y})\, dy_{t-1}^{\mathrm{im}}\, dy_{t-1}^{\mathrm{re}}\, dy_t^{\mathrm{im}}\, dy_t^{\mathrm{re}},\tag{30}$$

conclude that $\mathsf{DVO} = \frac{1}{2}$ for D-MPSK modulation with $n = \log_2(M) + 1$.

Finally, we move to the proof for $n \geq \log_2(M) + 2$ case. To that end, we can rewrite the expression in (13) as (29), shown at the bottom of the previous page.

Then, as $\mathsf{SNR} \to \infty$, it can be seen that the term outside the integral in the $p_1(\mathsf{SNR})$ expression decays to zero according to $\frac{2^{n-1}}{\pi^2}\mathsf{SNR}^{-1}$. Additionally, the integral expression itself, by using the monotone convergence theorem [62], can be written as (30), shown at the bottom of the previous page, where $\mathcal{G}'(\boldsymbol{y}) = \exp\{-\frac{1}{2}\|\boldsymbol{y}\|_2^2 + \mathsf{Re}(y_{t-1}y_t^*)\}$. This shows $p_1(\mathsf{SNR}) \overset{e}{=} \mathsf{SNR}^{-1}$. The arguments leading to $p_2(\mathsf{SNR}) \overset{e}{=} \mathsf{SNR}^{-1}$ and $p_3(\mathsf{SNR}) \overset{e}{=} \mathsf{SNR}^{-1}$ are the same. Finally, using [5, Lemma 2], we conclude that $p(\mathsf{SNR}) \overset{e}{=} \mathsf{SNR}^{-1}$, which implies that $\mathsf{DVO} = 1$ for $n \geq \log_2(M) + 2$.

REFERENCES

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[3] H.-J. Song and N. Lee, "Terahertz communications: Challenges in the next decade," *IEEE Trans. THz Sci. Technol.*, vol. 12, no. 2, pp. 105–117, Mar. 2022.

[4] A. Khalili, F. Shirani, E. Erkip, and Y. C. Eldar, "MIMO networks with one-bit ADCs: Receiver design and communication strategies," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1580–1594, Mar. 2022.

[5] S. Gayan, R. Senanayake, H. Inaltekin, and J. Evans, "Low-resolution quantization in phase modulated systems: Optimum detectors and error rate analysis," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1000–1021, 2020.

[6] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[7] B. Murmann. "ADC performance survey 1997–2020." Accessed: Oct. 18, 2023. [Online]. Available: https://github.com/bmurmann/ADC-survey

[8] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.

[9] D. He, Z. Wang, T. Q. S. Quek, S. Chen, and L. Hanzo, "Deep learning-assisted TeraHertz QPSK detection relying on single-bit quantization," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8175–8187, Dec. 2021.

[10] P. Kumari, A. Mezghani, and R. W. Heath, "A low-resolution ADC proof-of-concept development for a fully-digital millimeter-wave joint communication-radar," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Jan. 2020, pp. 8619–8623.

[11] O. Dizdar, A. Kaushik, B. Clerckx, and C. Masouros, "Rate-splitting multiple access for joint radar-communications with low-resolution DACs," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jan. 2021, pp. 1–6.

[12] J. Park, N. Lee, S.-N. Hong, and Y.-S. Jeon, "Learning from noisy labels for MIMO detection with one-bit ADCs," *IEEE Wireless Commun. Lett.*, vol. 12, no. 3, pp. 456–460, Mar. 2023.

[13] B. Fesl, M. Koller, and W. Utschick, "On the mean square error optimal estimator in one-bit quantized systems," *IEEE Trans. Signal Process.*, vol. 71, pp. 1968–1980, Jun. 2023.

[14] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.

[15] M. T. Ivrlac and J. A. Nossekh, "On MIMO channel estimation with single-bit signal-quantization," in *Proc. Int. ITG/IEEE Workshop Smart Antennas (WSA)*, Feb. 2007, pp. 1–7.

[16] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. 48th Asilom. Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2014, pp. 957–961.

[17] A. Mezghani and J. A. Nossek, "Efficient reconstruction of sparse vectors from quantized observations," in *Proc. Int. ITG Workshop Smart Antennas (WSA)*, 2012, pp. 193–200.

[18] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[19] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

[20] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.

[21] C. Wen, C. Wang, S. Jin, K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.

[22] F. Liu, X. Shang, and H. Zhu, "Efficient majorization-minimization-based channel estimation for one-bit massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3444–3457, Jun. 2021.

[23] F. Liu, X. Shang, Y. Cheng, and G. Zhang, "Computationally efficient maximum likelihood channel estimation for coarsely quantized massive MIMO systems," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 444–448, Feb. 2022.

[24] C. Stöckle, J. Munir, A. Mezghani, and J. A. Nossek, "Channel estimation in massive MIMO systems using 1-bit quantization," in *Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2016, pp. 1–6.

[25] A. B. L. B. Fernandes, Z. Shao, L. T. N. Landau, and R. C. de Lamare, "Multiuser-MIMO systems using comparator network-aided receivers with 1-bit quantization," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 908–922, Feb. 2023.

[26] B. Srinivas, P. Priya, D. Sen, and S. Chakrabarti, "Channel estimation in sub-6 GHz and hybrid millimeter wave MIMO systems with low-resolution ADCs," *IEEE Trans. Green Commun. Netw.*, vol. 7, no. 2, pp. 707–723, Jun. 2023.

[27] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, Apr. 2016.

[28] N. Liang and W. Zhang, "Mixed-ADC massive MIMO uplink in frequency-selective channel," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4651–4656, Nov. 2016.

[29] H. Pirzadeh and A. L. Swindlehurst, "Spectral efficiency of mixed-ADC massive MIMO," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3599–3613, Jul. 2018.

[30] Y. Zhang, M. Alrabeiah, and A. Alkhateeb, "Deep learning for massive MIMO with 1-bit ADCs: When more antennas need fewer pilots," *IEEE Trans. Wireless Commun. Lett.*, vol. 9, no. 8, pp. 1273–1277, Aug. 2020.

[31] S. Gao, P. Dong, Z. Pan, and G. Y. Li, "Deep learning-based channel estimation for massive MIMO with mixed-resolution ADCs," *IEEE Commun. Lett.*, vol. 23, no. 11, pp. 1989–1993, Jul. 2022.

[32] P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, "Deep CNN-based channel estimation for mmWave massive MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 13, no. 5, pp. 989–1000, Sep. 2019.

[33] E. Balevi and J. G. Andrews, "One-bit OFDM receivers via deep learning," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4326–4336, Jun. 2019.

[34] Y. Dong, H. Wang, and Y.-D. Yao, "Channel estimation for one-bit multiuser massive MIMO using conditional GAN," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 854–858, Mar. 2021.

[35] I. Helmy, P. Tarafder, and W. Choi, "LSTM-GRU model-based channel prediction for one-bit massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 72, no. 8, pp. 11053–11057, Aug. 2023.

[36] L. Xu, F. Gao, T. Zhou, S. Ma, and W. Zhang, "Joint channel estimation and mixed-ADCs allocation for massive MIMO via deep learning," *IEEE Trans. Wireless Commun.*, vol. 22, no. 2, pp. 1029–1043, Feb. 2023.

[37] L. V. Nguyen, D. H. N. Nguyen, and A. L. Swindlehurst, "Deep learning for estimation and pilot signal design in few-bit massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 379–392, Jan. 2023.

[38] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, Dec. 2013.

[39] Y.-S. Jeon, D. Kim, S.-N. Hong, N. Lee, and R. W. Heath, "Artificial intelligence for physical-layer design of MIMO communications with one-bit ADCs," *IEEE Commun. Mag.*, vol. 60, no. 7, pp. 76–81, Jul. 2022.

[40] M. S. Stein, S. Bar, J. A. Nossek, and J. Tabrikian, "Performance analysis for channel estimation with 1-bit ADC and unknown quantization threshold," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2557–2571, May 2018.

[41] G. Bottomley, H. Arslan, R. Ramesh, and G. Brismark, "Coherent MAP detection of DQPSK bits," *IEEE Commun. Lett.*, vol. 4, no. 11, pp. 354–356, Nov. 2000.

[42] K. Kiasaleh and T. He, "On the performance of DQPSK communication systems impaired by timing error, mixer imbalance, and frequency nonselective slow Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 46, no. 3, pp. 642–652, Aug. 1997.

[43] C.-L. Liu and K. Feher, "Bit error rate performance of $\frac{\pi}{4}$-DQPSK in a frequency-selective fast Rayleigh fading channel," *IEEE Trans. Veh. Technol.*, vol. 40, no. 3, pp. 558–568, Aug. 1991.

[44] J. G. Proakis, *Digital Communications*, 4th ed. New York, NY, USA: McGraw-Hill Educ., 2000.

[45] I. A. A. Falujah, "Performance analysis of DPSK systems with MIMO employing EGC over wireless fading channels," Ph.D. dissertation, Facul. Grad. School, Univ. Texas, Arlington, Arlington, TX, USA, 2007.

[46] A. Schenk and R. F. H. Fischer, "Noncoherent detection in massive MIMO systems," in *Proc. 17th Int. ITG Workshop Smart Antennas*, 2013, pp. 1–8.

[47] J. Singh and U. Madhow, "Phase-quantized block noncoherent communication," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2828–2839, Jul. 2013.

[48] S. Wang, L. Zhang, Y. Li, J. Wang, and E. Oki, "Multiuser MIMO communication under quantized phase-only measurements," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1083–1099, Mar. 2016.

[49] S. Gayan, H. Inaltekin, R. Senanayake, and J. Evans, "Differential MPSK with n-bit phase quantization," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2023, pp. 2553–2558.

[50] S. Gayan, R. Senanayake, H. Inaltekin, and J. Evans, "Reliability characterization for SIMO communication systems with low-resolution phase quantization under Rayleigh fading," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 2660–2679, 2021.

[51] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. New York, NY, USA: Cambridge Univ. Press, 2005.

[52] S. Gayan, H. Inaltekin, R. Senanayake, and J. Evans, "Phase modulated communication with low-resolution ADCs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.

[53] S. Gayan, R. Senanayake, H. Inaltekin, and J. Evans, "Selection combining for multi-antenna communication with low-resolution ADCs," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2021, pp. 3343–3348.

[54] K.-T. Fang, S. Kotz, and K.-W. Ng, *Symmetric Multivariate and Related Distributions*. New York, NY, USA: Springer, 1990.

[55] R. W. Lucky and J. C. Hancock, "On the optimum performance of n-ary systems having two degrees of freedom," *IRE Trans. Commun. Syst.*, vol. 10, no. 2, pp. 185–192, Jun. 1962.

[56] N. I. Bernardo, J. Zhu, and J. Evans, "Is phase shift keying optimal for channels with phase-quantized output?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2021, pp. 634–639.

[57] P. Nazari, B. Chun, F. Tzeng, and P. Heydari, "Polar quantizer for wireless receivers: Theory, analysis, and CMOS implementation," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 61, no. 3, pp. 877–887, Mar. 2014.

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

[59] R. G. Gallager, *Principles of Digital Communication*. New York, NY, USA: Cambridge Univ. Press, 2008.

[60] M. Nakagami, "The *m*-distribution—A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation*. Oxford, U.K.: Pergamon Press, 1960, pp. 7–36.

[61] R. Bose and D. Ray-Chaudhuri, "On a class of error correcting binary group codes," *Inf. Control*, vol. 3, no. 1, pp. 68–79, Mar. 1960.

[62] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1987.

SAMIRU GAYAN (Member, IEEE) received the B.Sc. and M.Phil. degrees in electronic and telecommunication engineering from the University of Moratuwa, Sri Lanka, in 2011 and 2015, respectively, and the Ph.D. degree in electrical and electronic engineering from The University of Melbourne, Australia, in 2021. He is currently a Senior Lecturer with the Department of Electronic and Telecommunication Engineering, University of Moratuwa. His research interests include communications theory, satellite communications, and signal processing for wireless communications. He is a recipient of the Melbourne Research Scholarship in 2017.

HAZER INALTEKIN (Member, IEEE) received the B.S. degree (High Hons.) in electrical and electronics engineering from Bogazici University, Istanbul, Turkey, and the M.S./Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, USA. He is a Senior Lecturer with Macquarie University. Prior to joining Macquarie University, he held various researcher and faculty positions in Australia, Europe, and USA. His research interests include airborne networks, fog/edge computing, IoT, wireless communications, and information theory.

RAJITHA SENANAYAKE (Member, IEEE) received the B.E. degree in electrical and electronics engineering from the University of Peradeniya, Sri Lanka, in 2009, the BIT degree in information technology from the University of Colombo, Sri Lanka, in 2010, and the Ph.D. degree in electrical and electronic engineering from The University of Melbourne, Australia, in 2015. From 2015 to 2016, she was with the Department of Electrical and Computer Systems Engineering, Monash University, Australia. She is currently serving as a Senior Lecturer with the Department of Electrical and Electronics Engineering, The University of Melbourne. Her research interests are in cooperative communications, distributed antenna systems, and joint radar and communications. She is a recipient of the Australian Research Council Discovery Early Career Researcher Award.

JAMIE EVANS (Senior Member, IEEE) was born in Newcastle, Australia, in 1970. He received the B.S. degree in physics and the B.E. degree in computer engineering from the University of Newcastle in 1992 and 1993, respectively, where he received the University Medal upon graduation, and the M.S. and Ph.D. degrees in electrical engineering from The University of Melbourne, Australia, in 1996 and 1998, respectively. From March 1998 to June 1999, he was a Visiting Researcher with the Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley. Since returning to Australia in July 1999, he has held academic positions with the University of Sydney, The University of Melbourne, and Monash University. He is currently a Professor with the Department of Electrical and Electronic Engineering, The University of Melbourne. His research interests are in communications theory, information theory, and statistical signal processing with a focus on wireless communications networks. He was awarded the Chancellor's Prize for excellence for his Ph.D. thesis.