Multiple Importance Sampling for Symbol Error Rate Estimation of Maximum-Likelihood Detectors in MIMO Channels

Víctor Elvira, *Senior Member, IEEE,* Ignacio Santamaria, *Senior Member, IEEE,*

Abstract— In this paper we propose a multiple importance sampling (MIS) method for the efficient symbol error rate (SER) estimation of maximum likelihood (ML) multiple-input multiple-output (MIMO) detectors. Given a transmitted symbol from the input lattice, obtaining the SER requires the computation of an integral outside its Voronoi region in a high-dimensional space, for which a closed-form solution does not exist. Hence, the SER must be approximated through crude or naive Monte Carlo (MC) simulations. This practice is widely used in the literature despite its inefficiency, particularly severe at high signal-to-noise-ratio (SNR) or for systems with stringent SER requirements. It is well-known that more sophisticated MC-based techniques such as MIS, when carefully designed, can reduce the variance of the estimators in several orders of magnitude with respect to naive Monte Carlo in rare-event estimation, or equivalently, they need significantly less samples for attaining a desired performance. The proposed MIS method provides unbiased SER estimates by sampling from a mixture of components that are carefully chosen and parametrized. The number of components, the parameters of the components, and their weights in the mixture, are automatically chosen by the proposed method. As a result, the proposed method is flexible, easy-to-use, theoretically sound, and presents a high performance in a variety of scenarios. We show in our simulations that SERs lower than 10^{-8} can be accurately estimated with just 10^4 random samples.

Index Terms— Multiple importance sampling, symbol error rate, Monte Carlo, multiple-input multiple-output (MIMO), maximum likelihood (ML) detection.

I. INTRODUCTION

Monte Carlo (MC) simulation is a key methodology for performance evaluation in wireless communication links, for many of which an exact computation is typically infeasible. Monte Carlo simulations are routinely used to estimate the symbol error rate (SER), the bit error rate (BER), or the outage probability in wireless systems. However, naive Monte Carlo simulation can be very inefficient and requires very long simulation runs when the target probability is small. For instance, the typical BER for lightwave communication systems or satellite links (after error correction coding) can be on the order of 10^{-9} to 10^{-12} [1] [2], the probability of a packet loss in packet switching systems is also on the order of 10^{-9}, and ultra-reliable low-latency communications (URLLC) have requirements in terms of outage probability lower than 10^{-6} [3]. A common rule of thumb is to simulate 10 or 100 times more samples than the inverse of the probability we are trying to estimate. For systems with such a stringent requirements, the MC simulation is infeasible due to the need of performing an unreasonable large number of trials. But even for more conventional systems with BER requirements on the order of $10^{-3} - 10^{-5}$, MC simulation can be costly in time and computational resources, especially when it is necessary to evaluate the system performance in a large number of scenarios due to, for example, user mobility.

In this paper, we develop an efficient simulation scheme to study the SER of multiple-input multiple-output (MIMO) systems using maximum likelihood decoding. The performance of suboptimal linear MIMO detectors, such as the zero-forcing (ZF) or the minimum mean-squared error (MMSE) detectors, have been analyzed in [4] based on a high signal-to-noise-ratio (SNR) approximation. The authors of [5] derived approximate bit error rate (BER) expressions for the MMSE MIMO receiver based on the Gamma distribution approximation for the signal-to-interference-plus-noise ratio (SINR). The exact error performance of MMSE detection, possibly with successive interference cancellation (SIC), was investigated in [6], and closed-form SER expressions for the matched filter (MF) detector in MIMO systems have recently been derived in [7]. Based on the union bound, asymptotic SER expressions for ML-MIMO systems decoder have been derived in [8] [9]. However, these results only provide an upper bound, which may be loose for scenarios involving a large number of antennas, so to characterize the performance of ML-MIMO decoders one must resort to computer simulations. When the targeted SER is low and the number of antennas increases, MC simulation becomes rapidly infeasible because of excessively long run times required to generate errors (rare events) in sufficiently large numbers for obtaining statistically significant results.

To overcome this limitation of naive MC simulation, this paper studies efficient simulation schemes based on importance sampling (IS) methods [10]–[12], which are particularly suitable for rare-events estimation problems. The IS methodology has been used for variance reduction in SER or BER simulations in a wide range of scenarios since the late seventies [13]–[17]. However, many digital communication researchers are still unaware of the potential benefits of IS

V. Elvira is with School of Mathematics, University of Edinburgh, United Kingdom (e-mail victor.elvira@ed.ac.uk).

I. Santamaria is with the Department of Communications Engineering, University of Cantabria, Santander, Spain (e-mail: i.santamaria@unican.es).

techniques to characterize the statistical performance of digital communication systems, mostly due to the difficulty to find an adequate IS algorithm [18].

In IS, the samples are simulated from a proposal probability density function (pdf), different from the target pdf, with the goal of increasing the number of errors during simulation. Then, the samples are properly weighted using the ratio of the target and the proposal pdfs in such a way the resulting estimators are unbiased and consistent with the number of samples [12]. In IS, the key for a successful performance is in the choice of this proposal pdf. With a well-chosen proposal, IS can reduce the variance of the estimators by several orders of magnitude with respect to naive Monte Carlo (or equivalently IS can work at a similar performance by using significantly less number of samples). Many works in the literature are devoted for the choice of the IS proposal, using usually a mixture of components as proposal in *multiple IS* (MIS) [19] and through an iterative adaptation of this proposal in *adaptive IS* (AIS) [20]. In [21] a biased channel distribution is proposed for the simulation of orthogonal space-time block codes (OSTBCs) on Nakagami channels. The authors of [22] used IS for the estimation of the outage probability of multi-antenna receivers with generalized selection combining. In a different setup, IS has been also used for approximating outage capacity over generalized fading channels [23]–[25]. Adaptive importance sampling has been used in [26] for error rate estimation in multiple access systems, and a nested IS estimator to estimate the random-coding error probability of coded modulation has been proposed in [27].

A multiple importance sampling technique called ALOE ("At Least One rare Event") [28], has recently been applied to SER estimation in single-input single-ouput systems [29] transmitting non-square 2D constellations [30]. ALOE is extremely efficient to estimate the integral of a Gaussian in a region defined by a union of half-spaces, which is precisely the error event in a digital communications system. Conditioned to a transmitted symbol, an error occurs when the observation falls in a union of half-spaces or, equivalently, outside a given Voronoi region. The proposal in ALOE simulates the system conditionally on an error taking place, which makes it more efficient than other importance sampling techniques. However, ALOE requires a perfect knowledge of the Voronoi decision regions. This information is not available in MIMO systems, where the Voronoi regions may be determined by a very large number of unknown hyperplanes, which precludes a direct application of ALOE to ML-MIMO decoders. A simple alternative that uses just a few hyperplanes obtained from the closest points to a given lattice point has been explored in [31], but it provides biased estimators especially at low SNRs.

In this paper, we propose a novel IS-based method for rare-event estimation, called *defensive adaptive* ALOE or DA^{2}LOE , particularly suited for SER estimation of the ML-MIMO decoders. DA^{2}LOE is a MIS method with a carefully designed mixture proposal. In particular, we consider a proposal with $K + 1$ components that are truncated versions of the targeted distribution (this truncation is key for a more efficient use of the computational budget). In this mixture, K components are defined in a region that guarantees that each sample produces at least one error (unlike in naive Monte Carlo, where especially at medium-high SNR, most samples produce zero errors), taking into account the K closest neighbors to the transmitted signal in the transformed lattice. The $(K+1)$-th proposal is designed in such a way the estimator is unbiased, ensuring that all integration area is covered. DA^{2}LOE automatically selects the number of components, their parameters, and the component weights in the mixture, all by taking into account the dimension (number of antennas) and the SNR. As a consequence, DA^{2}LOE is a flexible, easy-to-use, interpretable, and high-performance method that requires very limited tuning.

A. Outline

The outline of this paper is as follows. In Section II we present the ML detection problem in MIMO systems. In Section III we briefly describe standard Monte Carlo, importance sampling, and multiple importance sampling methods. The proposed algorithm DA^{2}LOE is presented in Section IV. Simulations results are provided in Section V. The paper is concluded with a final discussion in Section VI.

B. Notation

In this paper, matrices are denoted by bold-faced upper case letters, bold-faced lower case letters denote column vectors, and scalars are denoted by light-face lower case letters. A real matrix of dimension $M \times N$ is denoted $\mathbf{A} \in \mathbb{R}^{M \times N}$ and $\mathbf{x} \in \mathbb{R}^M$ indicates that $\mathbf{x}$ is a real vector of dimension M. The notation $\mathbf{x} \sim \mathcal{N}_M(\boldsymbol{\mu}, \mathbf{R})$ indicates that $\mathbf{x}$ is an M-dimensional Gaussian random vector of mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$ and $\mathbf{x} \sim \mathcal{CN}_M(\boldsymbol{\mu}, \mathbf{R})$ is used for complex normal vectors. $\mathrm{E}[\cdot]$ represents the expectation operator and the superscript $(\cdot)^\top$ denotes transpose.

II. MIMO DETECTION

We consider a multiple-input multiple-output (MIMO) spatial multiplexing system with T transmit and R receive antennas, where the channel is unknown at the transmiter but perfectly known at the receiver side. We assume $R \geq T$. The received signal follows the well-known baseband model

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{H}$ is an $R \times T$ matrix whose columns represent the known complex channel gains from each transmit antenna to the R receive antennas, $\mathbf{s} = (s_1, \ldots, s_T)^\top$ is the vector with the unknown transmitted symbols, and $\mathbf{n} = (n_1, \ldots, n_R)^\top$ is the noise vector. The noise is modeled as $\mathbf{n} \sim \mathcal{CN}_R(\mathbf{0}, \sigma^2\mathbf{I})$. Each symbol $\mathbf{s}$ belongs to a discrete constellation that we will assume to be a square M-QAM (Quadrature Amplitude Modulation) signal. The average energy per symbol is normalized so that $\mathrm{E}[|s_k|^2] = 1$.

We will find convenient to reformulate the problem in terms of real variables as follows

$$\underbrace{\begin{pmatrix} Re(\mathbf{x}) \\ Im(\mathbf{x}) \end{pmatrix}}_{\mathbf{x}} = \underbrace{\begin{pmatrix} Re(\mathbf{H}) & -Im(\mathbf{H}) \\ Im(\mathbf{H}) & Re(\mathbf{H}) \end{pmatrix}}_{\mathbf{H}} \underbrace{\begin{pmatrix} Re(\mathbf{s}) \\ Im(\mathbf{s}) \end{pmatrix}}_{\mathbf{s}} + \underbrace{\begin{pmatrix} Re(\mathbf{n}) \\ Im(\mathbf{n}) \end{pmatrix}}_{\mathbf{n}}. \tag{2}$$

Then, without loss of generality we can focus on the system model given by (1) but with real vectors and matrices, in this case the symbols take values on a finite alphabet of integers, $\mathbf{n} \sim \mathcal{N}_{2R}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I})$, and the vector $\mathbf{s} \in \mathcal{D}^{2T}$ belongs to a $2T$-dimensional square real lattice with cardinality $|\mathcal{D}^{2T}| = M^T$. For instance, if the number of antennas is $T = 10$ and we transmit 64-QAM symbols, the number of lattice points is $64^{10} \approx 1.15 \cdot 10^{18}$.

The MIMO detection problem consists of estimating the symbol vector $\hat{\mathbf{s}} \in D^{2T}$ in (1) that minimizes the symbol error probability. Under white Gaussian noise, the optimal maximum likelihood (ML) decoder is obtained by solving an *Integer Least Squares* problem

$$\hat{\mathbf{s}} = \min_{\mathbf{s} \in \mathcal{D}^{2T}} \|\mathbf{x} - \mathbf{H}\mathbf{s}\|_2^2. \tag{3}$$

Note that this problem is equivalent to constructing the Voronoi diagram with all transformed points $\mathbf{Hs}$ with $\mathbf{s} \in \mathcal{D}^{2T}$, and estimating $\hat{\mathbf{s}}$ as the associated point to the region where $\mathbf{x}$ falls. Maximum likelihood decoding therefore amounts to finding the closest lattice point $\mathbf{Hs}$ to a given noisy observation, $\mathbf{x}$. This problem is known to be NP-hard for generic channels [32], [33] meaning that the problem complexity is exponential on the dimension of the lattice [34]. Using the Finke-Phost sphere-decoding (SD) algorithm [35], ML detection can be achieved at an average complexity that is cubic in the number of transmit antennas, as was shown in [36]. Later Agrell et al. showed that the Schnorr-Euchner (SE) enumeration strategy [37] reduces the complexity of SD algorithms in comparison to [35], and very efficient SD implementations have been proposed in [38]. Our goal in this work is to develop efficient symbol error rate (SER) estimation techniques for ML MIMO detection. Although we focus on SER estimation, all methods in this paper can be generalized to estimate the uncoded bit error rate (BER).

A. Problem statement

Let us assume that $\mathbf{s}_q$ is transmitted and let $\overline{\mathcal{H}}_q$ be its associated Voronoi region, which is a polytope defined by the intersection of finitely many hyperplanes in $\mathbb{R}^{2R}$. The observation $\mathbf{x} = \mathbf{H}\mathbf{s}_q + \mathbf{n}$ is the $2R$-dimensional lattice point distorted by the MIMO channel and perturbed by additive Gaussian noise. Assuming that all symbols are transmitted with equal probability, the symbol error probability (SER) is

$$P_e = \frac{1}{M^T} \sum_{q=1}^{M^T} p_{e|\mathbf{s}_q}, \tag{4}$$

where

$$p_{e|\mathbf{s}_q} = \sum_{j \neq q} \frac{T_{j,q}}{T} p(\mathbf{s}_j|\mathbf{s}_q) = \sum_{j \neq q} \frac{T_{j,q}}{T} \text{Prob}(\mathbf{H}\mathbf{s}_q + \mathbf{n} \in \overline{\mathcal{H}}_j) \tag{5}$$

is the conditional SER when $\mathbf{s}_q$ is transmitted, and $T_{j,q}$ is the number of complex symbol errors when $\mathbf{s}_q$ is transmitted and $\mathbf{s}_j$ is detected. Each term $\text{Prob}(\mathbf{H}\mathbf{s}_q + \mathbf{n} \in \overline{\mathcal{H}}_j)$ is the integral of a Gaussian distribution centered at $\mathbf{H}\mathbf{s}_q \in \mathbb{R}^{2R}$ in the polytope region $\overline{\mathcal{H}}_j$. Note that transmitting $\mathbf{s}_q$ and

deciding $\mathbf{s}_j$ can result in a different number of complex symbol errors $T_{j,q} \in \{1, ..., T\}$; therefore each term in (5) is weighted by $T_{j,q}/T$. If we weight each term in (5) by the ratio of uncoded bits in error instead of $T_{j,q}/T$, then we would get an estimate of the uncoded BER. The integrals required to evaluate (5) are in general difficult to compute in closed-form, and therefore Monte Carlo simulation is typically used to estimate the SER. However, naive Monte Carlo simulation (also called crude Monte Carlo) can be very inefficient and requires very long simulation runs, especially at high signal-to-noise ratios or when either the constellation size, M, or the MIMO configuration, $R \times T$, is large. Our goal thus is to develop efficient sampling schemes to estimate $p_{e|\mathbf{s}_q}$ in (5).

III. SER ESTIMATION VIA IMPORTANCE SAMPLING

A. Standard Monte Carlo

For notational simplicity let us denote the transmitted vector as $\mathbf{s}$, with associated decision region $\overline{\mathcal{H}} \subset \mathbb{R}^{2R}$, and let $p_e \triangleq p_e(\mathbf{s})$ be the conditional SER of interest, which can be expressed as

$$p_e = \int_{\mathcal{H}} h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}, \tag{6}$$

where $\widetilde{\pi}(\mathbf{x}) \triangleq \mathcal{N}_{2R}(\mathbf{Hs}, \frac{\sigma^2}{2}\mathbf{I})$ is the Gaussian distribution of the observation, and the function $h(\mathbf{x}) = \ell(\mathbf{x})/T$ is the ratio of wrongly detected symbols, i.e., $\ell(\mathbf{x})$ is the number of symbol errors when $\mathbf{s}$ is transmitted and $\mathbf{x}$ is received. Note that $h(\mathbf{x}) = 0$ when $\mathbf{x} \in \overline{\mathcal{H}}$, i.e., when it falls in the Voronoi decision region for $\mathbf{s}$, whereas $1/T \leq h(\mathbf{x}) \leq 1$ when $\mathbf{x} \in \mathcal{H}$. The standard Monte Carlo procedure consists on simulating N i.i.d. samples from the targeted distribution and approximating Eq. (6) as

$$\widehat{p}_e^{(\text{MC})} = \frac{1}{N} \sum_{n=1}^{N} h(\mathbf{x}_n), \quad \mathbf{x}_n \sim \widetilde{\pi}(\mathbf{x}), \quad n = 1, ..., N. \tag{7}$$

This procedure is reasonably efficient when $h(\mathbf{x}_n) > 0$ for most samples. Standard Monte Carlo is known to be very inefficient for rare events estimation [10], [11], i.e., when p_e is small, as it happens at high SNR in MIMO detection.

B. Importance sampling

Importance sampling (IS) is a Monte Carlo technique used when sampling from $\widetilde{\pi}(\mathbf{x})$ is either not possible or not efficient [11], [12]. The N samples are simulated instead from a so-called *proposal* distribution, $q(\mathbf{x})$, and the estimator of p_e is built as

$$\widehat{p}_e^{(\text{IS})} = \frac{1}{N} \sum_{n=1}^{N} w_n h(\mathbf{x}_n), \quad \mathbf{x}_n \sim q(\mathbf{x}), \quad n = 1, ..., N, \tag{8}$$

where $w_n = \frac{\widetilde{\pi}(\mathbf{x}_n)}{q(\mathbf{x}_n)}$ is the importance weight (also called likelihood ratio). The estimator $\widehat{p}_e^{(\text{IS})}$ is unbiased if $q(\mathbf{x}) > 0$ for all $\mathbf{x}$ where $h(\mathbf{x})\widetilde{\pi}(\mathbf{x}) > 0$. Note that the particular case with $q(\mathbf{x}) = \widetilde{\pi}(\mathbf{x}) = \mathcal{N}_{2R}(\mathbf{Hs}, \frac{\sigma^2}{2}\mathbf{I})$ yields the standard Monte Carlo estimator of Eq. (7). In this way, IS can also be seen as a generalization of standard Monte Carlo. The key of IS is in

the appropriate choice of the proposal $q(\mathbf{x})$. The variance of (8) is

$$\text{Var}\left[\widehat{p}_e^{(\text{IS})}\right] = \frac{1}{N}\int_{\mathcal{H}} h^2(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x} - \frac{p_e^2}{N}. \qquad (9)$$

and the optimal proposal, which achieves zero variance, is given by $q^*(\mathbf{x}) = \frac{h(\mathbf{x})\widetilde{\pi}(\mathbf{x})}{p_e}$. The optimal proposal requires to know p_e, which is what we wish to estimate, hence it is rarely available in practice [11], [12]. Intuitively, the variance of the IS estimator increases when too many samples from $q(\mathbf{x})$ fall in $\overline{\mathcal{H}}$, and hence $h(\mathbf{x})$ takes value zero. This is exactly what happens with the MC estimator when σ^2 decreases, making MC simulation infeasible to estimate very low error probabilities.

C. Multiple importance sampling and the ALOE algorithm

Multiple importance sampling (MIS) is a natural extension of IS allowing for the simulation from a set of J proposals, $\{q_j(\mathbf{x})\}_{j=1}^J$, instead of just one [19]. However, the extension from one to several proposals is not straightforward and many sampling and weighting schemes can be devised (see [19] for a review). A conventional way to proceed is to simulate from the mixture proposal as

$$\mathbf{x}_n \sim q_{\boldsymbol{\alpha}} = \sum_{j=1}^J \alpha_j q_j(\mathbf{x}), \qquad n = 1, ..., N, \qquad (10)$$

where $\boldsymbol{\alpha} = [\alpha_1, ..., \alpha_J]$ is a simplex vector with all non-negative weights in the mixture such that $\sum_{j=1}^J \alpha_j = 1$. The MIS extension of Eq. (8) is

$$\widehat{p}_e^{(\text{MIS})} = \frac{1}{N}\sum_{n=1}^N \frac{h(\mathbf{x}_n)\widetilde{\pi}(\mathbf{x}_n)}{q_{\boldsymbol{\alpha}}(\mathbf{x}_n)}, \qquad (11)$$

where the importance weight is now $w_n = \frac{\widetilde{\pi}(\mathbf{x}_n)}{q_{\boldsymbol{\alpha}}(\mathbf{x}_n)}$.

We recall that we aim at integrating $h(\mathbf{x})\widetilde{\pi}(\mathbf{x})$ in the region $\mathcal{H}$, where $h(\mathbf{x}) = \frac{\ell(\mathbf{x})}{T}$ is the ratio of symbol errors when $\mathbf{x}$ is received. The region $\mathcal{H}$ can be described as the union of all half-spaces in $\mathbb{R}^{2R}$ generated by the hyperplanes that define the border of $\overline{\mathcal{H}}$.

A natural way to adapt a MIS scheme to our problem is to follow the choice of proposals in [28] called ALOE for "At Least One rare Event". In ALOE the number of proposals is the number of hyperplanes that define the polytope $\overline{\mathcal{H}}$. Let $\mathcal{V} = \{\mathbf{v}_j\}_{j=1}^J$ be the set containing the J *neighbors* of $\mathbf{Hs}$, where we define a *neighbor* of $\mathbf{Hs}$ as a point in the transformed (by the MIMO channel) lattice whose Voronoi region shares a border with the Voronoi region of $\mathbf{Hs}$. Each proposal in ALOE is a truncated version of a Gaussian centered at $\mathbf{Hs}$ beyond a hyperplane that defines the border between $\mathbf{Hs}$ and its j-th neighbor $\mathbf{v}_j$.

More specifically, $q_j(\mathbf{x}) = \frac{\mathbb{I}_{\mathcal{S}_j(\mathbf{x})}\widetilde{\pi}(\mathbf{x})}{P_j}$, where $P_j = \int \mathbb{I}_{\mathcal{S}_j(\mathbf{x})}\widetilde{\pi}(\mathbf{x})d\mathbf{x}$ is the integral of the target distribution in the half-space $\mathcal{H}_j = \{\mathbf{x} \in \mathbb{R}^{2R} \,|\, \mathbf{x}^\top\boldsymbol{\gamma}_j \geq \beta_j\}$ where the hyperplane $\mathbf{x}^T\boldsymbol{\gamma}_j = \beta_j$ is equidistant from $\mathbf{Hs}$ and $\mathbf{v}_j$. Under white Gaussian noise, it is a simple exercise to show that the integral in the half-space $\mathcal{H}_j$ concides with the integral in the half-space $\mathbb{R}^{2R-1} \times [d_j/2, \infty)$, where d_j is the Euclidean distance

between the j-th neighbor $\mathbf{v}_j$ and the transformed symbol $\mathbf{Hs}$. Since the noise variance is $\sigma^2/2$ then $P_j = Q(d_j/(\sqrt{2}\sigma))$, where $Q(\cdot) = 1 - \Phi(\cdot)$ is the Q-function or tail distribution of the standard normal pdf, and $\Phi(\cdot)$ is the associated cdf.

The union set of the J proposals in ALOE covers the integration area,

$$\mathcal{H} = \bigcup_{j=1}^J \mathcal{H}_j,$$

therefore the ALOE estimator is unbiased. Note also that each draw from the mixture $\mathbf{x}_n$ falls in a region where $\ell(\mathbf{x}_n) > 0$, so there is at least one symbol error per draw.

In ALOE, the samples $\mathbf{x}_n$, $n = 1, ..., N$, are independently simulated from $q_{\boldsymbol{\alpha}}$. The procedure for the efficient simulation from a truncated Gaussian distribution is described in [28] and summarized in Appendix I.

Substituting the ALOE proposals in Eq. (11) yields the MIS estimator

$$\widehat{p}_e^{(\text{ALOE})} = \frac{1}{N}\sum_{n=1}^N \frac{h(\mathbf{x}_n)\widetilde{\pi}(\mathbf{x}_n)}{q_{\boldsymbol{\alpha}}(\mathbf{x}_n)} \qquad (12)$$

$$= \frac{1}{N}\sum_{n=1}^N \frac{h(\mathbf{x}_n)}{\sum_{j=1}^J \alpha_j \mathbb{I}_{\mathcal{H}_j}(\mathbf{x}_n)P_j^{-1}}. \qquad (13)$$

We first define $\overline{p} = \sum_{j=1}^J P_j$, which is a union upper bound of p_e. The weights in the mixture proposal in Eq. (10) are chosen as $\alpha_j = P_j/\overline{p}$, for $j = 1, ..., J$ (see [28] for a justification). Then, the ALOE estimator is

$$\widehat{p}_e^{(\text{ALOE})} = \frac{\overline{p}}{N}\sum_{n=1}^N \frac{h(\mathbf{x}_n)}{\sum_{j=1}^J \mathbb{I}_{\mathcal{H}_k}(\mathbf{x}_n)} = \frac{\overline{p}}{N}\sum_{n=1}^N \frac{h(\mathbf{x}_n)}{C(\mathbf{x}_n)}, \quad (14)$$

where $C(\mathbf{x}_n) = \sum_{j=1}^J \mathbb{I}_{\mathcal{H}_j}(\mathbf{x}_n)$ is the number of half-spaces $\mathcal{H}_j$ where $\mathbf{x}_n$ is present. ALOE is unbiased, and its variance can be bounded as

$$\text{Var}\left(\widehat{p}_e^{(\text{ALOE})}\right) \leq \frac{p_e(\overline{p} - p_e)}{N},$$

i.e., the bound of the variance decays with the number of samples, N, and the scale factor depends on the true probability, p_e. It is straightforward to understand the high performance of ALOE for high SNR scenarios for which p_e is very small. A deeper theoretical analysis of ALOE can be found in [28].

ALOE has recently been used for SER estimation of single-input single-output (SISO) channels with non-square two-dimensional constellations in [29]. With two-dimensional lattices or constellations the Voronoi regions are determined by just a few hyperplanes, which can easily be computed, and hence an exact implementation of ALOE is possible. However, for MIMO systems the number of hyperplanes J determining a Voronoi region can be very large. Therefore, finding all neighbors $\mathcal{V} = \{\mathbf{v}_j\}_{j=1}^J$ of $\mathbf{Hs}$ (i.e., those whose associated polytopes share one hyperplane with the polytope that encloses $\overline{\mathcal{H}}$) for each lattice point under test, $\mathbf{Hs}$, is in general not feasible. The best we can hope for is to select, applying a SD-like algorithm, a set $\mathcal{C}_K = \{\mathbf{c}_k\}_{k=1}^K$ containing the K closest

points to $\mathbf{Hs}$ in the transformed lattice. We form in this way a new estimator that we call ALOE-K

$$\widehat{p}_e^{\text{(ALOE-K)}} = \frac{1}{N} \sum_{n=1}^{N} \frac{h(\mathbf{x}_n)}{\sum_{k=1}^{K} \alpha_k \mathbb{I}_{\mathcal{H}_k}(\mathbf{x}_n) P_k^{-1}} \quad (15)$$

$$= \frac{\overline{p}^{\text{(ALOE-K)}}}{N} \sum_{n=1}^{N} \frac{h(\mathbf{x}_n)}{C(\mathbf{x}_n)} \quad (16)$$

where $\alpha_k = P_k$, for $k = 1, ..., K$, as in ALOE and $\overline{p}^{\text{(ALOE-K)}} = \sum_{j=1}^{K} P_j$. For ALOE-K to be unbiased, the proposal support $\mathcal{Q}^K = \bigcup_{k=1}^{K} \mathcal{H}_k$ should contain the whole integration area $\mathcal{H}$. If one neighbor of the set $\mathcal{V} = \{\mathbf{v}_j\}_{j=1}^{J}$ is missed, then the mixture proposal will not cover the whole integration area and hence ALOE-K will be negatively biased with

$$\text{bias}^{\text{(ALOE-K)}} = \mathrm{E}[\widehat{p}_e^{\text{(ALOE-K)}}] - p_e \quad (17)$$

$$= -\int_{\mathcal{H} \setminus \mathcal{Q}^K} h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x} < 0, \quad (18)$$

i.e., $\widehat{p}_e^{\text{(ALOE-K)}}$ underestimates the probability of error. In the next section we propose an alternative MIS estimator that incorporates a defensive component in the mixture to recover the unbiasedness of the estimate, while adaptively selecting the number of hyperplanes K.

IV. DEFENSIVE ADAPTIVE ALOE

Here we present the *defensive adaptive* ALOE (DA^{2}LOE) algorithm with three important features. First, DA^{2}LOE is unbiased thanks to the introduction of an extra proposal in the mixture that plays a similar role as the defensive component in defensive IS [39]. However, in DA^{2}LOE the defensive proposal is not the targeted distribution as in conventional defensive IS, but a more efficient distribution that takes into account the geometry on the problem as well as the rest of proposals already included in the mixture. Moreover, the weight of this proposal in the mixture is also automatically set. Second, in addition to the defensive component the proposal mixture in DA^{2}LOE is composed of K Gaussians truncated beyond a hyperplane, as in ALOE. Finally, DA^{2}LOE allows for an automatic choice of K jointly with the design of the aforementioned defensive proposal.

A. Mixture proposal

The DA^{2}LOE mixture proposal is

$$q_{\boldsymbol{\alpha}}(\mathbf{x}) = \sum_{k=1}^{K+1} \alpha_k q_k(\mathbf{x}), \quad (19)$$

with $K + 1$ components $q_k(\mathbf{x})$ and non-negative associated mixture weights α_k. For each lattice point under test, $\mathbf{Hs}$, we choose its $K+1$ nearest neighbors $\mathcal{C}_{K+1} = \{\mathbf{c}_k\}_{k=1}^{K+1}$, defined in decreasing order of Euclidean distance, e.g., $\mathbf{c}_1$ is the closest point to $\mathbf{Hs}$ (below we discuss the choice of K jointly with the other parameters). The sphere decoding algorithm [38] can be applied to find the $K+1$ nearest neighbors, $\mathcal{C}_{K+1} = \{\mathbf{c}_k\}_{k=1}^{K+1}$, that belong to a hypersphere of appropriate radius centered at $\mathbf{Hs}$. To this end, an increasing radius search procedure can be

applied [40]. The $K+1$ DA^{2}LOE components are selected as follows.

- K **ALOE-like components.** We design the set of proposals $\{q_k(\mathbf{x})\}_{k=1}^{K}$ similarly as in ALOE. Each proposal $q_k(\mathbf{x})$ is a truncated version of a Gaussian centered at $\mathbf{Hs}$ beyond each hyperplane that defines the pairwise border between $\mathbf{Hs}$ and its k-th neighbor $\mathbf{c}_k$. In Appendix I, we describe the procedure to both simulate from $q_k(\mathbf{x})$ and find its associated normalizing constant P_k.

- **One defensive component.** The defensive proposal, $q_{K+1}(\mathbf{x})$, is designed as follows. Let $\mathbf{c}_{K+1}$ be the $(K+1)$-th closest point to $\mathbf{Hs}$ at Euclidean distance d_{K+1}. Note that $\mathbf{c}_{K+1}$ is the next closest point after the K neighbors that determine the ALOE-like proposals. Then, the defensive proposal is a truncated Gaussian distribution conditionally on $||\mathbf{x} - \mathbf{Hs}||_2 > d_{K+1}/2$, i.e.,

$$q_{K+1}(\mathbf{x}) = P_{K+1}^{-1}\widetilde{\pi}(\mathbf{x})\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}), \quad (20)$$

where

$$\mathcal{S}_{K+1} = \{\mathbf{x} \in \mathbb{R}^{2R}, ||\mathbf{x} - \mathbf{Hs}||_2 \geq d_{k+1}/2\}$$

is the outside of a hypersphere of radius $d_{K+1}/2$ and $P_{K+1} = \int_{\mathcal{S}_{K+1}} \widetilde{\pi}(\mathbf{x})d\mathbf{x}$ is its normalizing constant. The choice of the *defensive* proposal ensures that the mixture covers all the support where the integrand of Eq. (6) takes values different from zero, and hence DA^{2}LOE is unbiased. In Appendix I-C, we describe how to sample from a multivariate normal truncated outside a hypersphere, as well as the computation of P_{K+1}.

B. Algorithm, choice of mixture weights, and estimator

Algorithm 1 summarizes the DA^{2}LOE method. In Step 1 (Proposal adaptation), DA^{2}LOE algorithm starts with the defensive component in the mixture proposal, and iteratively incorporates ALOE-like components. The choice of K is done by sequentially adding points in increasing order of Euclidean distance until a stopping rule is fulfilled, forming the set $\mathcal{C}_{K+1}$ with the $K + 1$ closest points. Since the distances to the neighbors $d_1 \geq d_2 \geq \ldots \geq d_K$ are increasing, and $P_k = Q(d_j/(\sqrt{2}\sigma))$ is a monotonically decreasing function, then $P_1 \geq P_2 \geq \ldots \geq P_K$. The algorithm stops incorporating proposals when the K-th component has received a small mixture weight, i.e., $\alpha_K < \gamma$, where γ is a threshold chosen *a priori* (see the discussion in Section IV-D). In Step 2 (Sampling), N independent samples, $\{\mathbf{x}_n\}_{n=1}^{N}$, are simulated from the designed mixture proposal $q_{\boldsymbol{\alpha}}(\mathbf{x})$ as in Eq. (19). The standard importance weights are computed in Step 3 (Weighting). Finally, the (unnormalized) IS estimator is built in Step 4. Figure 1 displays a graphical representation of DA^{2}LOE and naive Monte Carlo.

The DA^{2}LOE estimator is derived as follows. First, by

standard IS arguments,

$$\widehat{p}_e^{\,(\mathrm{DA^2LOE})} = \frac{1}{N}\sum_{n=1}^{N}\frac{h(\mathbf{x}_n)\widetilde{\pi}(\mathbf{x}_n)}{q_{\boldsymbol{\alpha}}(\mathbf{x}_n)} \tag{21}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\frac{h(\mathbf{x}_n)}{\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n)P_{K+1}^{-1} + \sum_{k=1}^{K}\alpha_k\mathbb{I}_{\mathcal{H}_k}(\mathbf{x}_n)P_k^{-1}}. \tag{22}$$

The mixture weights of the K ALOE-like proposals are chosen as $\alpha_k = (1-\alpha_{K+1})\frac{P_k}{\sum_{k=1}^{K}P_k}$, $k=1,...,K$, i.e., $\alpha_k \propto P_k$, following arguments as in [28]. Note that $\sum_{k=1}^{K+1}\alpha_k = 1$, i.e., $\{\alpha_k\}_{k=1}^{K+1}$ represents a probability mass function. The choice of α_{K+1} is discussed in Section IV-D.

Substituting the choice of the mixture weights in Eq. (22), the estimator can be written as

$$\widehat{p}_e^{\,(\mathrm{DA^2LOE})} = \frac{1}{N}\sum_{n=1}^{N}\frac{h(\mathbf{x}_n)}{\frac{\alpha_{K+1}}{P_{K+1}}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n) + \sum_{k=1}^{K}\frac{\alpha_k}{P_k}\mathbb{I}_{\mathcal{H}_k}(\mathbf{x}_n)}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\frac{h(\mathbf{x}_n)}{\frac{\alpha_{K+1}}{P_{K+1}}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n) + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}\sum_{k=1}^{K}\mathbb{I}_{\mathcal{H}_k}(\mathbf{x}_n)}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\frac{h(\mathbf{x}_n)}{\frac{\alpha_{K+1}}{P_{K+1}}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n) + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x}_n)}, \tag{23}$$

where we recall that $\overline{p}^{(\mathrm{ALOE\text{-}K})} = \sum_{k=1}^{K}P_k$ and $C(\mathbf{x}_n) = \sum_{k=1}^{K}\mathbb{I}_{\mathcal{H}_k}(\mathbf{x}_n)$ is the number of half-spaces $\mathcal{H}_k$ where $\mathbf{x}_n$ is present.

C. Analysis of DA²LOE

In the following we discuss the unbiasedness of DA²LOE and provide an upper bound for its variance.

Theorem 1: The DA²LOE estimation $\widehat{p}_e^{\,(\mathrm{DA^2LOE})}$ is unbiased, i.e., $\mathrm{E}[\widehat{p}_e^{\,(\mathrm{DA^2LOE})}] = p_e$.

Proof: Note that, by construction, the proposal of DA²LOE $q_{\boldsymbol{\alpha}}(\mathbf{x})$ fully covers $\mathcal{H}$, the support where the integrand of Eq. (6) takes values different from zero. Hence, because of IS arguments, the DA²LOE estimator is unbiased since it corresponds to the unnormalized IS estimator. See Appendix II-A for a formal proof. $\square$

Theorem 2: The variance $\mathrm{Var}[\widehat{p}_e^{\,(\mathrm{DA^2LOE})}]$ of DA²LOE can be upper-bounded as

$$\mathrm{Var}[\widehat{p}_e^{\,(\mathrm{DA^2LOE})}] = \int_{\mathcal{H}}\frac{h^2(\mathbf{x})\widetilde{\pi}(\mathbf{x})}{\frac{\alpha_{K+1}}{P_{K+1}}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}) + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x})}d\mathbf{x} - p_e^2 \tag{24}$$

$$< \frac{1}{N}\left(\frac{P_{K+1}^2}{\alpha_{K+1}} + p_e\left(\frac{\overline{p}^{(\mathrm{ALOE\text{-}K})}}{1-\alpha_{K+1}} - p_e\right)\right). \tag{25}$$

Proof: See Appendix II-B for a proof. $\square$

Algorithm 1 DA²LOE

Input: γ and ξ

1: **1. Proposal adaptation**. Set $k = 0$
2: **while** $\alpha_k > \gamma$ **do**
3:     set $k := k+1$
4:     find $\mathbf{c}_k$ and $\mathbf{c}_{k+1}$, the k-th and $(k+1)$-th closest points to $\mathbf{Hs}$, at distances d_k and d_{k+1}, respectively
5:     define the half-space

$$\mathcal{H}_k = \{\mathbf{x} \in \mathbb{R}^{2R} \,|\, \mathbf{x}^{\top}\boldsymbol{\gamma}_k \geq \beta_k\}$$

    which is equidistant from $\mathbf{s}$ and $\mathbf{c}_k$, and the set

$$\mathcal{S}_{k+1} = \{\mathbf{x} \in \mathbb{R}^{2R} \,|\, ||\mathbf{x} - \mathbf{Hs}||_2 \geq d_{k+1}/2\}$$

6:     compute $P_k = \int \mathbb{I}_{\mathcal{H}_k(\mathbf{x})}\widetilde{\pi}(\mathbf{x})d\mathbf{x}$, i.e., the integral of the target distribution in the half-space $\mathcal{H}_k$
7:     Compute $P_{k+1} = \int \mathbb{I}_{\mathcal{S}_{k+1}(\mathbf{x})}\widetilde{\pi}(\mathbf{x})d\mathbf{x}$, the integral of the target distribution outside the hypersphere $\mathcal{S}_{k+1}$, and set $\alpha_{k+1} = \min(P_{k+1}, \xi)$
8:     recompute the mixture weights

$$\alpha_j = (1-\alpha_{k+1})\frac{P_j}{\sum_{i=1}^{k}P_i}, j = 1,...,k \tag{26}$$

9:
10: **end while**
11: Set $K := k$ as the final number of ALOE proposals
12: **2. Sampling.** Simulate independently

$$\mathbf{x}_n \sim q_{\boldsymbol{\alpha}}(\mathbf{x}) = \sum_{k=1}^{K+1}\alpha_k q_k(\mathbf{x}) \tag{27}$$

13: **3. Weighting.** The importance weight of the n-th sample is

$$w_n = \frac{1}{\frac{\alpha_{K+1}}{P_{K+1}}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n) + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x}_n)}. \tag{28}$$

14: **4. Estimator.** The IS estimator is given by

$$\widehat{p}_e^{\,(\mathrm{DA^2LOE})} = \frac{1}{N}\sum_{n=1}^{N}w_n h(\mathbf{x}_n) \tag{29}$$

$$= \frac{1}{N}\sum_{n=1}^{N}\frac{h(\mathbf{x}_n)}{\frac{\alpha_{K+1}}{P_{K+1}}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n) + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x}_n)}. \tag{30}$$

15: **Output**: $\widehat{p}_e^{\,(\mathrm{DA^2LOE})}$

The bound in (25) provides useful insights on the choice of K and α_{K+1} that are further discussed in the next section.

We now analyze the variance of ALOE-K in Eq. (16). Let us first consider the case when the support of the mixture proposal covers the integration area (i.e., when $\mathcal{Q}^K \triangleq \bigcup_{k=1}^{K}\mathcal{H}_k = \mathcal{H}$). Then, the variance can be bounded as $\mathrm{Var}[\widehat{p}_e^{\,(\mathrm{ALOE\text{-}K})}] \leq \frac{p_e(\overline{p}^{(\mathrm{ALOE\text{-}K})}-p_e)}{N} \triangleq \overline{V}^{(\mathrm{ALOE})}$, since it would particularize to the ALOE estimator (see [28, Theorem 2]). In the case with $\mathcal{Q}^K \subset \mathcal{H}$ (the mixture proposal covers only a subset of the support of the integrand), the ALOE-K variance is bounded

as $\mathrm{Var}[\widehat{p}_e{}^{(\text{ALOE-K})}] \leq \frac{p_e^{(\text{ALOE-K})}(\overline{p}^{(\text{ALOE-K})} - p_e^{(\text{ALOE-K})})}{N} \triangleq \overline{V}^{(\text{ALOE-K})}$, where $p_e^{(\text{ALOE-K})} = \mathrm{E}[\widehat{p}_e{}^{(\text{ALOE-K})}] < p_e$ is the expectation of the ALOE-K estimator (which is negatively biased in this case).

Regarding DA^{2}LOE, we recall that P_{K+1} is the integral of the Gaussian out of a hypersphere with a radius that depends on K. When K grows, P_{K+1} gets smaller. Then, in the bound of (25), for a fixed α_{K+1}, we can make the term $\frac{P_{K+1}^2}{\alpha_{K+1}}$ arbitrarily small by increasing K. Moreover, we can also reduce α_{K+1} in such a way that the ratio $\frac{P_{K+1}^2}{\alpha_{K+1}}$ is still negligible with respect to the second term in (25). In summary, by increasing K and carefully reducing α_{K+1}, we can cover better the integration area with the K ALOE-like proposals, while allowing us to reduce the weight of the defensive component. In this scenario, the bound in (25) is dominated by its second term as

$$\mathrm{Var}[\widehat{p}_e{}^{(\text{DA}^2\text{LOE})}] < \frac{1}{N}\left(\frac{P_{K+1}^2}{\alpha_{K+1}} + p_e\left(\frac{\overline{p}^{(\text{ALOE-K})}}{1-\alpha_{K+1}} - p_e\right)\right)$$
$$\approx \frac{1}{N}\left(p_e\left(\frac{\overline{p}^{(\text{ALOE-K})}}{1-\alpha_{K+1}} - p_e\right)\right) \triangleq \overline{V}^{(\text{DA}^2\text{LOE})}.$$
(31)

First, note that by increasing K, we would eventually cover the whole support of the integrand with the K ALOE-like proposals, and then we could get $\overline{V}^{(\text{DA}^2\text{LOE})}$ arbitrarily close to $\overline{V}^{(\text{ALOE})}$ by reducing α_{K+1}. In other words, we would recover ALOE as a limiting case, since in this scenario, ALOE-K becomes ALOE. Moreover, in this case, $\overline{V}^{(\text{DA}^2\text{LOE})}$ can be re-written as

$$\overline{V}^{(\text{DA}^2\text{LOE})} = \frac{1}{1 + \frac{p_e(1-\alpha_{K+1})}{\overline{p}^{(\text{ALOE-K})} - p_e(1-\alpha_{K+1})}}\overline{V}^{(\text{ALOE})}.$$
(32)

The constant multiplying in the right-hand side of Eq. (32) can be interpreted as the price to pay in DA^{2}LOE to ensure that the estimator is unbiased. Moreover, we note that (25) can be a loose bound, and in practice the MSE in DA^{2}LOE will be in general smaller than in ALOE-K, especially at low SNR where the bias of ALOE-K is particularly significant (see the numerical validation in Section V). The extra cost per sample of DA^{2}LOE compared to ALOE-K is negligible, and it is associated to checking whether $\mathbf{x}_n$ is outside of the hypersphere that defines the defensive component.

The variance of the naive Monte Carlo estimator is exactly $V^{(\text{MC})} = \frac{p_e(1-p_e)}{N}$ (see for instance [11]). ALOE-K guarantees a variance reduction with respect to naive Monte Carlo when $\overline{p}^{(\text{ALOE-K})} < 1$, which happens at medium/high SNR for a reasonable choice of K. Following similar arguments, when the choice of K and α_{K+1} makes the first term of (25) negligible, a variance reduction in DA^{2}LOE is guaranteed if $\frac{\overline{p}^{(\text{ALOE-K})}}{1-\alpha_{K+1}} < 1$ (and, since the estimator is unbiased, also a MSE reduction is ensured). In practice, for most ranges of SNR, DA^{2}LOE largely outperforms naive Monte Carlo in terms of variance for the same N, which is equivalent to say that in general it can obtain better results with much fewer samples.

D. Adaptive choice of K and α_{K+1}

The number of ALOE-like components in the mixture, K, and the weight assigned to the defensive component,

α_{K+1}, determine the accuracy of the estimator as well as its complexity. In addition, at high SNR it is expected that only a few closest symbols (hyperplanes) will contribute to the conditional symbol error probability [41], so both K and α_{K+1} should be chosen as a function of the noise variance. Finally, note that DA^{2}LOE has only two parameters, ξ and γ, that are closely connected to the choice of K and α_{K+1}.

First, note that α_{K+1} modulates the importance of the spherically truncated proposal, and balances the tradeoff between defensiveness (to give enough mass to regions not covered by the ALOE proposals) and efficiency loss (to not give too much mass to already covered regions or to regions in $\overline{\mathcal{H}}$). For moderate/high SNR, the non-covered probability mass is smaller, and it is possible to reduce the weight of the defensive proposal. A sensible choice is $\alpha_{K+1} = P_{K+1}$. In this way, we guarantee that the mixture proposal pdf $q_{\boldsymbol{\alpha}}(\mathbf{x}) \geq \widetilde{\pi}(\mathbf{x})$ for those points not covered by at least one of the K ALOE-like proposals, i.e., for all $\{\mathbf{x} : ||\mathbf{x} - \mathbf{Hs}||_2 > d_{K+1}/2\}$. However, for high-dimensional problems P_{K+1} can be very close to one (even for high K), which can compromise the efficiency of DA^{2}LOE when most of the mass is assigned to the defensive proposal. For this reason, we propose to limit the defensive weight $\alpha_{K+1} \leq \xi$. For higher dimensions, P_{K+1} needs a higher SNR to drop significantly from 1, and hence other less-defensive strategies could be devised. Regarding, the choice of γ, which controls the adaptation of K, we recommend it to be selected on the order of magnitude of $1/N$. The rationale is that when α_K is smaller than $1/N$, on average the proposal would be chosen to simulate less than one sample out of the N random draws, i.e., adding more components in the mixture would not significantly change it. We recall that the use of the defensive proposal attenuates the impact of missing potentially relevant hyperplanes/proposals, making DA^{2}LOE unbiased.

V. SIMULATION RESULTS

In this section, we study the performance of the proposed MIS estimators for MIMO systems with M-QAM transmitted signals by means of Monte Carlo simulations. We consider Rayleigh MIMO channels with i.i.d. zero-mean and unit variance complex Gaussian entries $h_{ij} \sim \mathcal{CN}(0,1)$. In each simulation, the SER for a given E_b/N_0 is estimated as follows:

1) Q codewords are generated by sampling uniformly the $2T$ dimensional squared real lattice defined by the $M-$QAM constellation: $\mathbf{s}_q \in \mathcal{D}^{2T}$, $q = 1, \ldots, Q$.
2) For each noiseless received codeword, $\mathbf{Hs}_q$, we draw N samples from the proposal distribution $\mathbf{x}_{q,n} \sim q(\mathbf{x})$, $n = 1, \ldots, N$. For the standard naive Monte Carlo algorithm, $q(\mathbf{x}) = \mathcal{N}_{2R}\left(\mathbf{Hs}_q, \frac{\sigma^2}{2}\mathbf{I}\right)$. For ALOE-K, $q(\mathbf{x})$ is the K component mixture proposal in (16). For DA^{2}LOE, $q(\mathbf{x})$ is the $K+1$ component mixture proposal in (23), where the last term of the mixture corresponds to the defensive proposal.
3) We apply sphere decoding (SD) or any other approximate ML detection to each received vector. When the number of antennas is large, even efficient SD implementations can be computationally costly. For
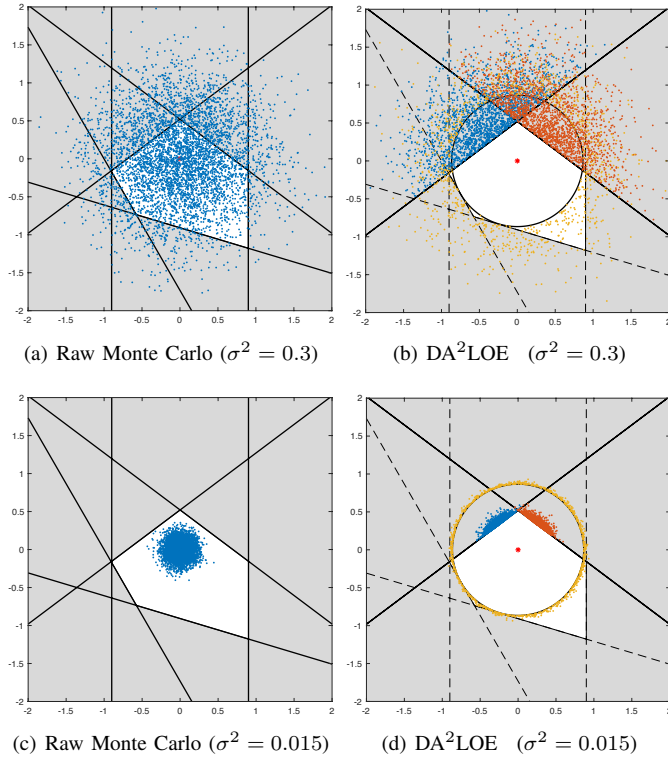
Fig. 1. In this 2-dimensional toy example we show the integration region (gray background), the transmitted symbol (red cross), and the hyperplanes delimiting the Voronoi region (black lines). The samples of naive Monte Carlo and DA^{2}LOE (with $K = 2$) algorithms are displayed with dots (different colors mean different proposals). In DA^{2}LOE , the solid lines describe the truncation border of the three proposals ($K = 2$ hyperplanes and one hypersphere).

this reason, we apply instead an approximate fixed-complexity ML-decoding algorithm that takes advantage of the fact that the number of visited lattice points, Q, is in general much smaller than the total number of received vectors NQ. The idea is the following: for each of the Q visited lattice points, $\mathbf{H}\mathbf{s}_q$, we find its S nearest neighbors using a SD-like algorithm. In our simulations we take $S = 300$. The approximate ML-decoding algorithm then finds for each noisy received signal, $\mathbf{x}_{qn} \sim q(\mathbf{x})$, the closest lattice point to $\mathbf{x}_{qn}$ among the fixed-size subset of S nearest neighbors. Note that for both MC and DA^{2}LOE it is necessary to run the SD algorithm, or its approximate version, for each of the samples generated. In addition, the computational complexity of SD increases with the number of symbols in the constellation. Therefore, for large constellations such as 1024-QAM, used for example for both 5G+ and beyond-WiFi-6 applications, a simulation scheme such as DA^{2}LOE that provides unbiased, low-variance SER estimates with as few samples as possible is even more important.

4) Finally, the SER is obtained by averaging the conditional SER estimates as

$$\widehat{P}_e = \frac{1}{Q} \sum_{q=1}^{Q} \widehat{p}_{e|\mathbf{s}_q}, \qquad (33)$$

where $\widehat{p}_{e|\mathbf{s}_q}$ is the conditional error probability when $\mathbf{s}_q$ is transmitted, which is estimated as (7), (16), and (23) for naive Monte Carlo (denoted as MC), ALOE-K and DA^{2}LOE, respectively.

With the proposed simulation setup, the SER is estimated at each E_b/N_0 with a total of $N_t = QN$ samples. For the proposed MIS estimators a number of samples on the order of $N_t \approx 10^4$ (we use in our simulations $Q = 100$ lattice points and $N = 100$ random samples of the mixture for each lattice point), allows us to obtain low variance estimates for error probabilities lower than 10^{-10}. Similar accuracies cannot be reached with standard MC or previously proposed importance sampling schemes.

As figures of merit to characterize the statistical performance of the different estimators, we use the relative root mean squared error (RRMSE) defined as

$$\text{RRMSE} = \frac{\sqrt{\text{E}\left[(\widehat{P}_e - P_e)^2\right]}}{P_e}, \qquad (34)$$

and the bias

$$\text{Bias} = \left| \text{E}\left[\widehat{P}_e\right] - P_e \right|, \qquad (35)$$

where P_e is the true symbol error probability. The expectations in (34) and (35) are estimated by averaging the results of 500 independent simulations, whereas the true P_e is estimated running ALOE with a high number of mixture proposals and draws per mixture (we set $K = 300$).

A. Performance comparison: MC, ALOE-K and DA^{2}LOE

In this subsection we compare the performance of naive Monte Carlo (denoted as MC in the figures), ALOE-K, and DA^{2}LOE. For the three methods, the $\widehat{P}_e$ at each E_b/N_0 is estimated by generating $Q = 100$ codewords and $N = 100$ random draws of the mixture per codeword, for a total of $N_t = QN = 10^4$ random samples. The parameters for DA^{2}LOE are $\xi = 0.85$ and $\gamma = 1/(10N)$.

Experiment 1: Performance comparison for a 6×6 MIMO system transmitting 64-QAM signals. We consider in the first example a 6×6 MIMO channel transmitting 64-QAM signals (i.e., $T = R = 6$, $M = 64$). Figure 2 shows the SER curves for MC, ALOE-K with $K = 10$, $K = 50$, and $K = 100$ mixture components and DA^{2}LOE. Fig. 3 shows (a) the RRMSE and (b) the bias of the estimates. Figure Fig. 3(b) only shows the results of ALOE with K mixture components, since naive Monte Carlo and DA^{2}LOE are unbiased estimators.

At low or moderate E_b/N_0, ALOE with K mixture components provides biased SER estimates. Note that the Voronoi regions could be determined by many more hyperplanes than the prefixed value K. In other words, the intersection of the K half-spaces does not cover the whole integrating area where the integrand takes significant values and hence the estimator is biased. At high E_b/N_0, however, the conditional symbol error probability has an asymptotic behavior proportional to $Q\left(\frac{d_{min}}{\sqrt{2}\sigma}\right)$, where $Q(\cdot)$ denotes the Gaussian Q-function and d_{min} is the distance to the closest lattice point [41]. Therefore,
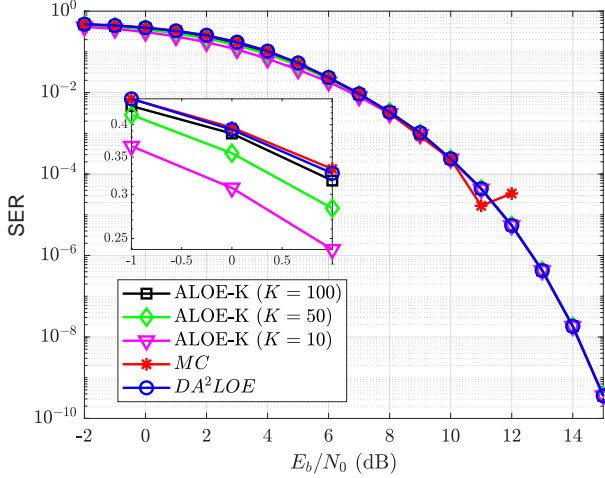
Fig. 2. SER vs. E_b/N_0 curves for MC, ALOE-K, and DA^{2}LOE. 6×6 MIMO system transmitting 64-QAM signals.

in this regime only the closest hyperplanes contribute to the integral and ALOE-K is practically unbiased.

The defensive component of DA^{2}LOE eliminates the bias at low E_b/N_0 values, so DA^{2}LOE and MC are both unbiased estimators. In fact, in this regime MC is almost optimal and, as we observe in Fig. 3(a), DA^{2}LOE replicates its behavior. As the SNR increases, the weight assigned to the defensive component decreases and the number of components of the mixture varies as shown in Fig. 4. When the E_b/N_0 is low, DA^{2}LOE retains all possible hyperplanes (in this case $K = 100$), whereas when the E_b/N_0 is high only a few hyperplanes are kept in the mixture. For intermediate values, the behaviour can be non-monotonic due to the interplay between the defensive and the non-defensive components on the mixture.

In summary, the DA^{2}LOE mixture proposal composed of an adaptive number of hyperplanes plus a defensive component gets unbiased and low-variance estimates with affordable computational cost over the whole E_b/N_0 range.

The SER curves for MC and DA^{2}LOE along with their confidence intervals are depicted in Fig. 5. The confidence interval has been magnified to $[-20\sigma, 20\sigma]$ for visualization purposes. Even so, for high E_b/N_0 values the confidence interval region of DA^{2}LOE is so small in comparison to that of the MC estimate that it is not visible in the figure.

Experiment 2: Performance comparison for a 25×25 MIMO system transmitting 16-QAM signals. In the second example we consider a 25×25 MIMO channel transmitting 16-QAM signals (i.e., $T = R = 25$, $M = 16$). Fig. 6 shows (a) the RRMSE and (b) the bias of the estimates. Again, DA^{2}LOE provides unbiased and extremely accurate SER estimates with just $N_t = 10^4$ random samples per E_b/N_0 point. The number of component mixtures selected by DA^{2}LOE is shown in Fig. 7. Finally, Fig. 8 shows the SER curves for MC and DA^{2}LOE along with their confidence intervals. Again we see the good performance of DA^{2}LOE,

particularly good for the challenging range of high SNR values.

Let us finally remark that DA^{2}LOE estimates the SER for a fixed MIMO channel realization, regardless of its structure (e.g., correlation between antennas) or fading distribution [42]. Its final performance and computational complexity, however, may depend on the channel characteristics. For instance, the number of hyperplanes that determine the decision regions increases as the number of antennas grows, or as the correlation between the matrix entries increases. This, in turn, increases the complexity of DA^{2}LOE , since more hyperplanes need to be included in the mixture to ensure a satisfactory performace. Note also that DA^{2}LOE is specifically designed for ML MIMO detectors. If a linear MIMO decoder, followed by a componentwise decisor, is used, then the decision regions are hypercubes determined by a fixed number of $2T$ hyperplanes. Consequently, for efficient SER estimation of linear MIMO detectors there is no need to use DA^{2}LOE and the simpler ALOE-K estimator can be used instead.

VI. CONCLUSIONS

In this paper we have proposed an efficient SER estimator for ML-MIMO decoders. The so-called DA^{2}LOE generates N samples from a mixture of $K + 1$ truncated Gaussians: each of the first K components is a Gaussian truncated to a half-space that ensures that each random sample produces at least one error, while the last $K + 1$ component is a Gaussian truncated to the outside of a hypersphere and acts as a defensive component that makes the estimator unbiased. The advantage of DA^{2}LOE is that all samples from the first K components of the mixture produce at least one symbol error, which is a key feature of the algorithm, particularly for high SNR when most samples in naive Monte Carlo would give zero errors. DA^{2}LOE is unbiased, and our simulations show that it provides low variance estimates for SER values lower than 10^{-8} with as few as $N = 10^4$ samples, largely reducing the computational cost with respect to naive Monte Carlo. The paper has focused on SER estimates for a fixed MIMO channel realization. Nevertheless, the SER performance of fading MIMO channels could be analyzed by first sampling a channel realization from the fading distribution (probably using some IS methodology adapted to the fading distribution) and then applying DA^{2}LOE . In this case, more efficient nested IS procedures might be worth studying. Another future line of research is to adapt DA^{2}LOE to characterize the BER performance of coded modulations. It would be also possible to design other suitable defensive proposals that could be sequentially adapted.
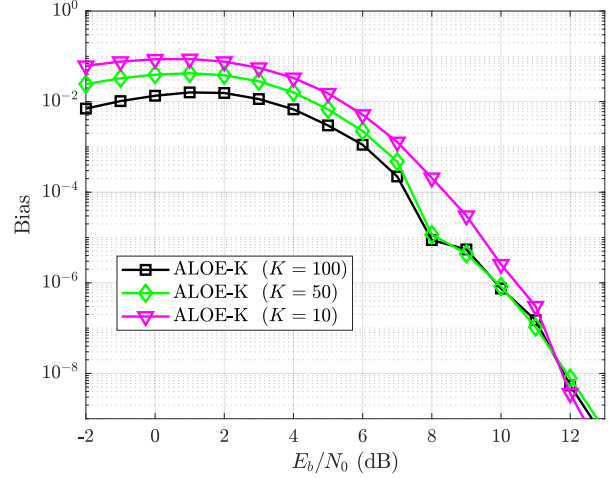
APPENDIX I
SAMPLING FROM TRUNCATED GAUSSIANS

A. Simulation from a truncated Gaussian $\mathcal{N}_{2L}(\boldsymbol{0}, \boldsymbol{I})$ in a half-space

The proposals in the MIS implementation are truncated Gaussian distributions. Let us first describe the simulation of a truncated Gaussian $\widetilde{\pi}(\mathbf{x}) = \mathcal{N}_{2R}(\mathbf{x}; \boldsymbol{0}, \boldsymbol{I})$ in the half-space
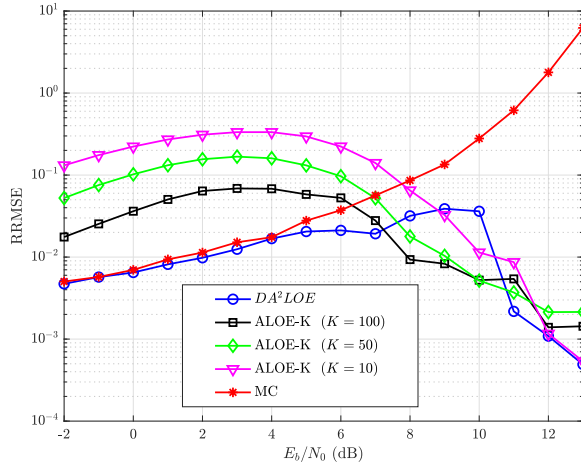
(a) Relative Root MSE (RRMSE)

(b) Bias for ALOE-K with different number of mixture proposals. MC and DA^{2}LOE are unbiased and hence are not shown in the figure.

Fig. 3.   Performance of MC, ALOE-K and DA^{2}LOE for a 6×6 MIMO system transmitting 64-QAM signals.
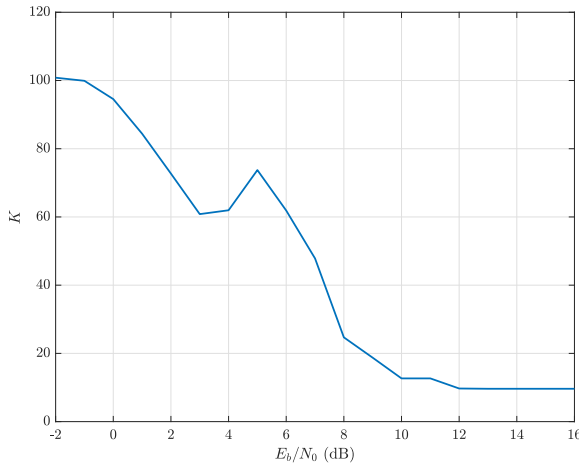

Fig. 4.   Number of DA^{2}LOE hyperplane components for a 6×6 MIMO system transmitting 64-QAM signals.
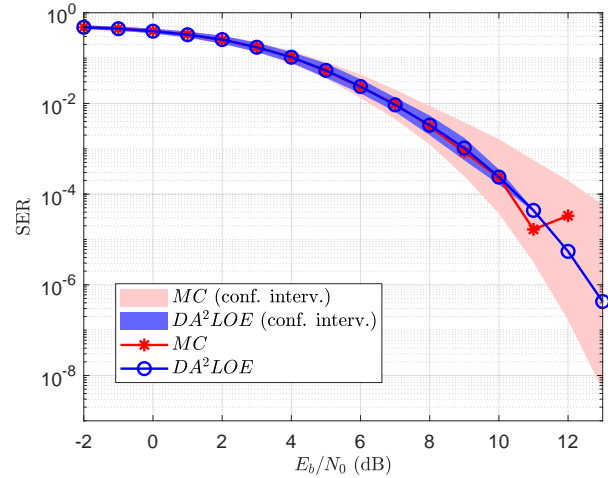
Fig. 5.   SER vs. Eb/N_0 for MC and DA^{2}LOE with confidence intervals. 6×6 MIMO system, 64-QAM signals.

described by $\mathbf{x}^\top \boldsymbol{\omega} \geq \tau$, i.e., $q(\mathbf{x}) = \frac{\widetilde{\pi}(\mathbf{x}) \mathbb{I}_{\mathbf{x}^\top \boldsymbol{\omega} \geq \tau}(\mathbf{x})}{Z}$. It first proceeds by simulating the sample $\mathbf{x}$ from the complementary half-space (i.e., $\mathbf{x}^\top \boldsymbol{\omega} < \tau$), and then delivering the $-\mathbf{x}$ for numerical stability. The algorithm described in [43] proceeds as follows:

1) Simulate $\mathbf{z} \sim \mathcal{N}_{2R}(\mathbf{0}, \mathbf{I})$
2) Simulate $u \sim \mathcal{U}(0, 1)$
3) Let $y = \Phi^{-1}(u\Phi(-\tau))$, where $\Phi(\cdot)$ denotes the cumulative distribution function for the standard Gaussian
4) Let $\mathbf{x} = \boldsymbol{\omega} y + (\mathbf{I} - \boldsymbol{\omega}\boldsymbol{\omega}^T) \mathbf{z}$
5) Output $\mathbf{x} := -\mathbf{x}$

Interestingly, $Z = \Phi(-\tau) = \int \mathbb{I}_{\mathbf{x}^\top \boldsymbol{\omega} \geq \tau}(\mathbf{x}) \mathcal{N}_{2L}(\mathbf{x}; \mathbf{0}, \mathbf{I}) d\mathbf{x}$ is the normalizing constant of the distribution. This normalizing constant is used by the ALOE-K and DA^{2}LOE algorithms.

### B. Extension to a generic truncated Gaussian $\mathcal{N}_{2R}(\mu, \Sigma)$ in a half-space

When the target distribution is a generic truncated Gaussian $\mathcal{N}_{2R}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ that must be integrated over the union of the half-spaces described by the hyperplanes $\{\boldsymbol{\gamma}_k, \beta_k\}_{k=1}^K$, one can transform the problem as follows. First, we describe K half-spaces with the hyperplanes $\boldsymbol{x}^\top \boldsymbol{\omega}_k \geq \tau_k$, where

$$\boldsymbol{\omega}_k = \frac{\boldsymbol{\gamma}_k^T \boldsymbol{\Sigma}^{1/2}}{\sqrt{\boldsymbol{\gamma}_k^T \boldsymbol{\Sigma} \boldsymbol{\gamma}_k}}, \quad \text{and} \quad \tau_k = \frac{\beta_k - \boldsymbol{\gamma}_k^T \boldsymbol{\mu}}{\sqrt{\boldsymbol{\gamma}_k^T \boldsymbol{\Sigma} \boldsymbol{\gamma}_k}}, \quad (36)$$

for $k = 1, \ldots, K$. Then, a Gaussian $\mathcal{N}_{2R}(\mathbf{0}, \mathbf{I})$ is integrated over the union of the transformed half-spaces as described in previous section.
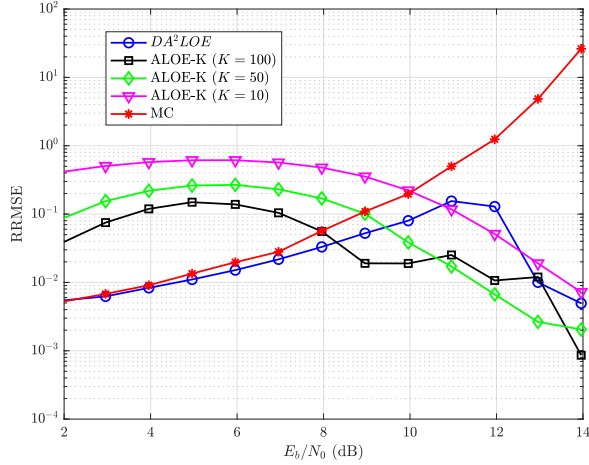
(a) Relative Root MSE (RRMSE)
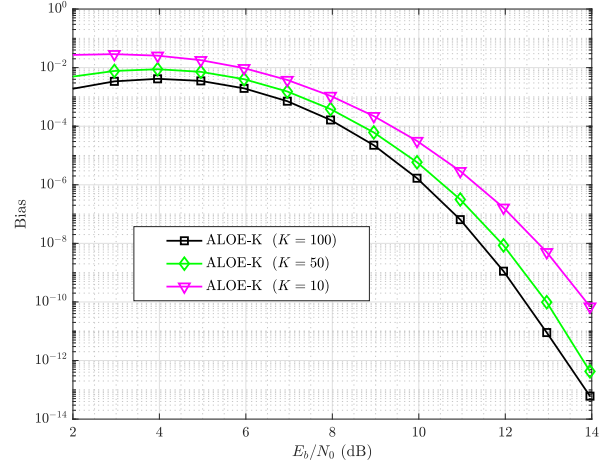
(b) Bias for ALOE-K with different number of mixture proposals. MC and DA²LOE are unbiased and hence are not shown in the figure.

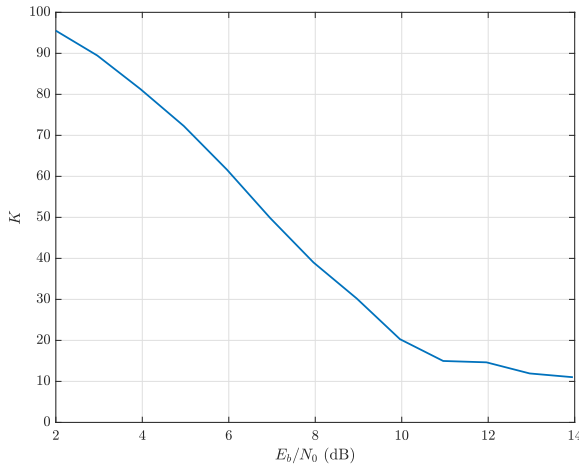Fig. 6. Performance of MC, ALOE-K and DA²LOE for a 25×25 MIMO system transmitting 16-QAM signals.

Fig. 7. Number of DA²LOE hyperplane components for a 25×25 MIMO system transmitting 16-QAM signals.
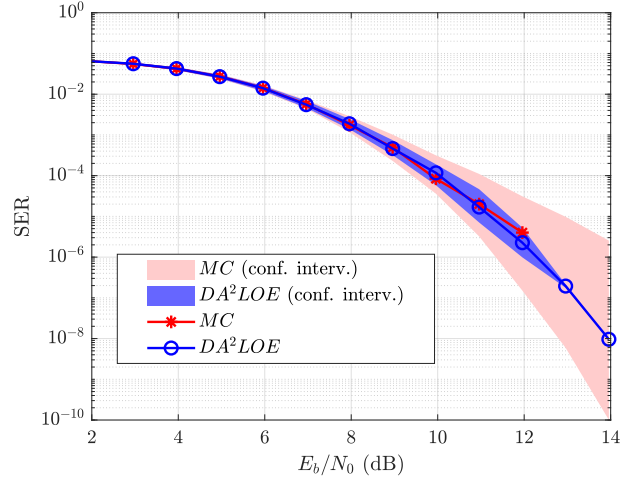
Fig. 8. SER vs. E_b/N_0 for MC and DA²LOE with confidence intervals. 25×25 MIMO system, 16-QAM signals.

C. Simulation from a truncated Gaussian outside a hypersphere

We denote the defensive DA²LOE proposal as

$$q(\mathbf{x}) = \frac{\widetilde{\pi}(\mathbf{x})\mathbb{I}_{\mathcal{S}(a)}(\mathbf{x})}{Z}, \qquad (37)$$

where $\widetilde{\pi}(\mathbf{x}) = \mathcal{N}_{2R}(\mathbf{x}; \boldsymbol{\mu}, \eta^2\mathbf{I})$, and

$$\mathcal{S}(a) = \{\mathbf{x} \in \mathbb{R}^{2R} \,|\, ||\mathbf{x} - \boldsymbol{\mu}||_2 \geq a$$

is the region outside of a hypersphere of radius a (with $a = d_{K+1}/2$ ind DA²LOE) and $Z = \int_{\mathcal{S}(a)} \widetilde{\pi}(\mathbf{x})d\mathbf{x}$ is its normalizing constant. We can efficiently sample from $q(\mathbf{x})$ by using standard methods used in directional statistics [44]. We proceed as follows:

1) Simulate $r \sim \chi^2_{2R}(r|r > a)$, i.e., from the truncated χ^2 distribution with $2R$ degrees of freedom, defined in the interval (a, ∞). This can be efficiently done as follows:

   - Find c as the cdf of the distribution χ^2_{2R} at $\left(\frac{a}{\eta}\right)^2$ (e.g., using in MATLAB the function `chi2inv`).
   - Apply a modified inverse transform sampling by first sampling uniformly $u \sim \mathcal{U}(c, 1)$, and then obtaining r as the evaluation of u in the inverse cdf of χ^2_{2R}.

2) Simulate $\mathbf{z} \sim \mathcal{N}_{2R}(\mathbf{0}, \mathbf{I})$ of dimension $2R$.
3) Project $\mathbf{z}$ in the $2R - 1$ unit sphere $\boldsymbol{\theta} = \frac{\mathbf{z}}{||\mathbf{z}||}$.
4) Output $\mathbf{x} = \boldsymbol{\mu} + \eta\boldsymbol{\theta}\sqrt{r}$.

The normalizing constant is simply $Z = 1 - c$.

APPENDIX II
DA^{2}LOE ANALYSIS

A. Unbiasedness of DA^{2}LOE

The expectation of the DA^{2}LOE estimator, from Eq. (22), is given by

$$\mathrm{E}[\widehat{p}_e^{(\mathrm{DA}^2\mathrm{LOE})}] \tag{38}$$

$$= \mathrm{E}\left[\frac{1}{N}\sum_{n=1}^{N} \frac{h(\mathbf{x}_n)}{\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n)P_{K+1}^{-1} + \sum_{k=1}^{K}\alpha_k\mathbb{I}_{\mathcal{H}_k}(\mathbf{x}_n)P_k^{-1}}\right] \tag{39}$$

where the expectation is with respect to the mixture proposal $q_{\boldsymbol{\alpha}}$, with support $\mathcal{Q}^{K+1} = \bigcup_{k=1}^{K}\mathcal{H}_k \bigcup \mathcal{S}_{K+1}$. Hence, from (39),

$$\mathrm{E}[\widehat{p}_e^{(\mathrm{DA}^2\mathrm{LOE})}]$$

$$= \frac{1}{N}\sum_{n=1}^{N}\int_{\mathcal{Q}^{K+1}} \frac{h(\mathbf{x})}{\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n)P_{K+1}^{-1} + \sum_{k=1}^{K}\alpha_k\mathbb{I}_{\mathcal{H}_k}(\mathbf{x})P_k^{-1}}$$
$$\cdot q_{\boldsymbol{\alpha}}(\mathbf{x})d\mathbf{x} \tag{40}$$

$$= \int_{\mathcal{Q}^{K+1}} \frac{h(\mathbf{x})}{\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n)P_{K+1}^{-1} + \sum_{k=1}^{K}\alpha_k\mathbb{I}_{\mathcal{H}_k}(\mathbf{x})P_k^{-1}} \tag{41}$$

$$\cdot \left(\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x}_n)P_{K+1}^{-1}\widetilde{\pi}(\mathbf{x}) + \sum_{k=1}^{K}\alpha_k\mathbb{I}_{\mathcal{H}_k}(\mathbf{x})P_k^{-1}\widetilde{\pi}(\mathbf{x})\right)d\mathbf{x}$$

$$= \int_{\mathcal{Q}^{K+1}} h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x} \tag{42}$$

where we have used that

$$q_{\boldsymbol{\alpha}}(\mathbf{x}) = \sum_{k=1}^{K+1}\alpha_k q_k(\mathbf{x}) \tag{43}$$

$$= \alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x})P_{K+1}^{-1}\widetilde{\pi}(\mathbf{x}) + \sum_{k=1}^{K}\alpha_k\mathbb{I}_{\mathcal{H}_k}(\mathbf{x})P_k^{-1}\widetilde{\pi}(\mathbf{x}). \tag{44}$$

Next,

$$\mathrm{E}[\widehat{p}_e^{(\mathrm{DA}^2\mathrm{LOE})}] = \int_{\mathcal{H}} h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x} + \int_{\mathcal{Q}^{K+1}\setminus\mathcal{H}} h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x} \tag{45}$$

$$= \int_{\mathcal{H}} h(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x} \tag{46}$$

$$= p_e, \tag{47}$$

where we have used $\mathcal{H} \subseteq \mathcal{Q}^{K+1}$ by construction of the proposal, and $h(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathcal{Q}^{K+1} \setminus \mathcal{H}$.

B. Variance of DA^{2}LOE

Let us express the variance of the 1-sample DA^{2}LOE estimator ($N = 1$, for $N > 1$, simply divide by N) as

$$\mathrm{Var}[\widehat{p}_e^{(\mathrm{DA}^2\mathrm{LOE})}] = \mathrm{E}[w(\mathbf{x})^2 h(\mathbf{x})^2] - \mathrm{E}[w(\mathbf{x})h(\mathbf{x})]^2 \tag{48}$$

where

$$w(\mathbf{x}) = \frac{1}{\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x})P_{K+1}^{-1} + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x})}. \tag{49}$$

Next, we make use of an auxiliary variable z, where $z = 0$ whenever the sampling is done from the defensive proposal, and $z = 1$ otherwise (from the ALOE-like proposals). Hence,

$$\mathrm{E}[w(\mathbf{x})^2 h(\mathbf{x})^2] = \mathbb{P}(z=0)\mathrm{E}[w(\mathbf{x})^2 h(\mathbf{x})^2|z=0] \tag{50}$$

$$+ \mathbb{P}(z=1)\mathrm{E}[w(\mathbf{x})^2 h(\mathbf{x})^2|z=1] \tag{51}$$

$$= \alpha_{K+1}\mathrm{E}[w(\mathbf{x})^2 h(\mathbf{x})^2|z=0] \tag{52}$$

$$+ (1-\alpha_{K+1})\mathrm{E}[w(\mathbf{x})^2 h(\mathbf{x})^2|z=1] \tag{53}$$

Next, since $q_{K+1}(\mathbf{x}) = \mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x})P_{K+1}^{-1}\widetilde{\pi}(\mathbf{x})$,

$$\mathrm{E}_{q_{K+1}}[w(\mathbf{x})^2 h(\mathbf{x})^2|z=0]$$

$$= \int \frac{h(\mathbf{x})^2}{\left(\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x})P_{K+1}^{-1} + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x})\right)^2}q_{K+1}(\mathbf{x})d\mathbf{x} \tag{54}$$

$$= P_{K+1}^{-1}\int_{\mathcal{S}_{K+1}} \frac{h(\mathbf{x})^2\widetilde{\pi}(\mathbf{x})}{\left(\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x})P_{K+1}^{-1} + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x})\right)^2}d\mathbf{x} \tag{55}$$

$$\leq P_{K+1}^{-1}\int_{\mathcal{S}_{K+1}} \frac{1}{\left(\alpha_{K+1}P_{K+1}^{-1}\right)^2}d\mathbf{x} \tag{56}$$

$$= \frac{P_{K+1}}{\alpha_{K+1}^2}\int_{\mathcal{S}_{K+1}} d\mathbf{x} \tag{57}$$

$$= \frac{P_{K+1}^2}{\alpha_{K+1}^2} \tag{58}$$

where we have used that $h(\mathbf{x}) \leq 1$ and that, when $z = 0$, the sample is at least present in $\mathcal{S}_{K+1}$ (and could be also present in the K half-spaces $\mathcal{H}$). Next,

$$\mathrm{E}_{q^{(\mathrm{ALOE})}}[w(\mathbf{x})^2 h(\mathbf{x})^2|z=1]$$

$$= \int \frac{h(\mathbf{x})^2}{\left(\alpha_{K+1}\mathbb{I}_{\mathcal{S}_{K+1}}(\mathbf{x})P_{K+1}^{-1} + \frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x})\right)^2}q^{(\mathrm{ALOE})}(\mathbf{x})d\mathbf{x} \tag{59}$$

$$\leq \int \frac{1}{\left(\frac{(1-\alpha_{K+1})}{\overline{p}^{(\mathrm{ALOE\text{-}K})}}C(\mathbf{x})\right)^2}q^{(\mathrm{ALOE})}(\mathbf{x})d\mathbf{x} \tag{60}$$

$$= \left(\frac{\overline{p}^{(\mathrm{ALOE\text{-}K})}}{1-\alpha_{K+1}}\right)^2 \int \frac{1}{C(\mathbf{x})^2}q^{(\mathrm{ALOE})}(\mathbf{x})d\mathbf{x} \tag{61}$$

where we have used again that $h(\mathbf{x}) \leq 1$. Then, we use [28, Theorem 2], establishing

$$\int \frac{1}{C(\mathbf{x})^2}q^{(\mathrm{ALOE})}(\mathbf{x})d\mathbf{x} \leq \left(\overline{p}^{(\mathrm{ALOE\text{-}K})}\right)^{-1}p_e^{(\mathrm{ALOE\text{-}K})} \tag{62}$$

$$\leq \left(\overline{p}^{(\mathrm{ALOE\text{-}K})}\right)^{-1}p_e, \tag{63}$$

where $p_e^{(\mathrm{ALOE\text{-}K})} = \mathrm{E}[\widehat{p}_e^{(\mathrm{ALOE\text{-}K})}]$. Note that $p_e^{(\mathrm{ALOE\text{-}K})} \leq p_e$ since ALOE uses a subset of the set of half-spaces whose union defines the integration set $\mathcal{H}$. Then,

$$\mathrm{E}_{q^{(\mathrm{ALOE})}}[w(\mathbf{x})^2 h(\mathbf{x})^2|z=1] \leq \left(\frac{\overline{p}^{(\mathrm{ALOE\text{-}K})}}{1-\alpha_{K+1}}\right)^2 \left(\overline{p}^{(\mathrm{ALOE\text{-}K})}\right)^{-1}p_e \tag{64}$$

$$= p_e\frac{\overline{p}^{(\mathrm{ALOE\text{-}K})}}{(1-\alpha_{K+1})^2}. \tag{65}$$

Finally, we are able to bound the variance as

$$\text{Var}[\widehat{p}_e^{(\text{DA}^2\text{LOE})}]$$

$$\leq \alpha_{K+1}\frac{P_{K+1}^2}{\alpha_{K+1}^2} + (1-\alpha_{K+1})p_e\frac{\overline{p}^{(\text{ALOE-K})}}{(1-\alpha_{K+1})^2} - p_e^2 \quad (66)$$

$$= \frac{P_{K+1}^2}{\alpha_{K+1}} + p_e\frac{\overline{p}^{(\text{ALOE-K})}}{1-\alpha_{K+1}} - p_e^2 \quad (67)$$

$$= \frac{P_{K+1}^2}{\alpha_{K+1}} + p_e\frac{\sum_{k=1}^K P_k}{1-\alpha_{K+1}} - p_e^2. \quad (68)$$

REFERENCES

[1] R. Papannanreddy, *Lightwave Communication Systems: A Practical Perspective*. Penram Int. Publishing, 2014.

[2] T. Pratt and J. E. Allnut, *Satellite Communications*. John Wiley and Sons, 2019.

[3] G. Durisi, T. Koch, and P. Popovsky, "Towards massive, ultra-reliable, and low-latency wireless communication with short packets," *Proc. of the IEEE*, vol. 104, no. 9, pp. 1711–1726, 2016.

[4] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: An in-depth study of the high SNR regime," *IEEE Trans. on Inf. Theory*, vol. 57, no. 4, pp. 2008–2026, Apr. 2011.

[5] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis," *IEEE Trans. on Inf. Theory*, vol. 52, no. 1, pp. 271–286, Jan. 2006.

[6] P. Liu and I.-M. Kim, "Exact and closed-form error performance analysis for hard MMSE-SIC detection in MIMO systems," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2463–2477, Sept. 2011.

[7] Y. Hama and H. Ochiao, "Performance analysis of matched-filter detector for MIMO spatial multiplexing over Rayleigh fading channels with imperfect channel estimation," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3220–3233, May. 2019.

[8] R. Van Nee, A. Van Zelst, and G. Awater, "Maximum likelihood decoding in a space division multiplexing system," in *IEEE Vehicular Technology Conf.*, Tokio, Japan, May 2000.

[9] X. Zhu and R. D. Murch, "Performance analysis of maximum likelihood detection in a MIMO antenna system," *IEEE Trans. on Communications*, vol. 50, no. 2, pp. 28–38, Feb. 2002.

[10] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer, 2004.

[11] A. B. Owen, *Monte Carlo theory, methods and examples*, 2013.

[12] V. Elvira and L. Martino, "Advances in importance sampling," *Wiley StatsRef: Statistics Reference Online*, pp. 1–22, 2020.

[13] P. Balaban, "Statistical evaluation of the error rate of the fiber-guide repeater using importance sampling," *Bell Syst. Tech. J.*, vol. 55, no. 6, pp. 745–766, 1976.

[14] K. S. Shanmugam and P. Balaban, "A modified Monte-Carlo simulation technique for the evaluation of error rate in digital communication systems," *IEEE Trans. Commun.*, vol. 28, no. 11, pp. 1916–1924, 1980.

[15] D. Lu and K. Yao, "Improved importance sampling technique for efficient simulation of digital communication systems," *IEEE Journal on Sel. Area Comm.*, vol. 6, no. 1, pp. 67–75, 1988.

[16] W. Al-Qaq and J. Townsend, "A stochastic importance sampling methodology for the efficient simulation of adaptive systems in frequency nonselective Rayleigh fading channels,," *IEEE Journal on Sel. Area Comm.*, vol. 15, no. 4, pp. 614–625, 1997.

[17] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communications systems," *IEEE Journal on Sel. Area Comm.*, vol. 15, no. 4, pp. 597–613, 1997.

[18] R. Srinivasan, *Importance Sampling. Applications in Communications and Detection*. Springer Verlag, 2002.

[19] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized multiple importance sampling," *Statistical Science*, vol. 34, no. 1, pp. 129–155, 2019.

[20] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Míguez, and P. M. Djuric, "Adaptive importance sampling: The past, the present, and the future," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 60–79, 2017.

[21] L. Jacobs and M. Moeneclaey, "Efficient BER simulation of Orthogonal Space-Time Block codes in Nakagami-m fading," in *IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications*, Washington, DC, USA, 2014, pp. 676–680.

[22] N. Ben Rached, Z. Botev, A. Kammoun, M.-S. Alouini, and R. Tempone, "Importance sampling estimator of outage probability under generalized selection combining model," in *Int. Conf Acoust., Speech, Signal Process. (ICASSP)*, Calgary, Alberta, Canada, 2018, pp. 3909–3913.

[23] N. B. Rached, A. Kammoun, M.-S. Alouini, and R. Tempone, "Unified importance sampling schemes for efficient simulation of outage capacity over generalized fading channels," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 376–388, 2015.

[24] ——, "On the efficient simulation of outage probability in a log-normal fading environment," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2583–2593, 2017.

[25] M.-S. Alouini, N. B. Rached, A. Kammoun, and R. Tempone, "On the efficient simulation of the left-tail of the sum of correlated log-normal variates," *Monte Carlo Methods and Applications*, vol. 24, no. 2, pp. 101–115, 2018.

[26] F. Han, S. Zhao, H. Jiang, H. Chen, and C. Zhang, *IEEE Access*.

[27] J. Font-Segura, A. Martinez, and A. Guillén i Fàbregas, "Importance sampling for coded-modulation error probability estimation," *IEEE Trans. Comm.*, vol. 68, no. 1, pp. 289–300, 2020.

[28] A. B. Owen, Y. Maximov, and M. Chertkov, "Importance sampling the union of rare events with an application to power systems analysis," *Electronic Journal of Statistics*, vol. 13, no. 1, pp. 231–254, 2019.

[29] V. Elvira and I. Santamaria, "Multiple importance sampling for efficient symbol error rate estimation," *IEEE Signal Proc. Letters*, vol. 26, no. 3, pp. 420–424, 2019.

[30] I. Santamaria, P. M. Crespo, C. Lameiro, and P. J. Schreier, "Information-theoretic analysis of a family of improper discrete constellations," *Entropy*, vol. 20, no. 1, 2018.

[31] V. Elvira and I. Santamaria, "Efficient SER estimation for MIMO detectors via importance sampling schemes," in *Proc. 53th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, 2019.

[32] M. Grotschel, L. Lovasz, and A. Schriver, *Geometric Algorithms and Combinatorial Optimization*. Springer Verlag, 2nd ed., 1993.

[33] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, 2003.

[34] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014.

[35] U. Fincke and M. Phost, "Improved methods for calculating vectors of short length in a lattice, including a complexity analysis," *Math. Comput.*, vol. 44, pp. 463–471, 1985.

[36] B. Hassibi and H. Vikalo, "On the expected complexity of sphere decoding," in *Proc. 35th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, 2001, pp. 1051–1055.

[37] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subset sum problems," *Math. Programm.*, vol. 66, no. 2, pp. 181–199, 1994.

[38] A. Ghasemmehdi and E. Agrell, "Faster recursions in sphere decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3530–3536, 2011.

[39] A. Owen and Y. Zhou, "Safe and effective importance sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 135–143, 2000.

[40] W. Zhao and G. B. Giannakis, "Sphere decoding algorithms with improved radius search," *IEEE Trans. Comm.*, vol. 53, no. 7, pp. 1104–1109, July 2005.

[41] A. Alvarado, F. Brännström, E. Agrell, and T. Koch, "High-SNR asymptotics of mutual information for discrete constellations with applications to BICM," *IEEE Trans. on Inf. Theory*, vol. 60, pp. 1061–1076, 2014.

[42] V. Raghavan, A. M. Sayeed, and V. V. Veeravalli, "Semiunitary precoding for spatially correlated mimo channels," *IEEE transactions on Information Theory*, vol. 57, no. 3, pp. 1284–1298, 2011.

[43] A. B. Owen, Y. Maximov, and M. Chertkov, "Importance sampling the union of rare events with an application to power systems analysis," *arXiv preprint arXiv:1710.06965*, 2017.

[44] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009, vol. 494.