

# Efficient SER Estimation for MIMO Detectors via Importance Sampling Schemes

Víctor Elvira

The University of Edinburgh, United Kingdom

Ignacio Santamaria

University of Cantabria, Spain

**Abstract**—In this paper we propose two importance sampling methods for the efficient symbol error rate (SER) estimation of maximum likelihood (ML) multiple-input multiple-output (MIMO) detectors. Conditioned to a given transmitted symbol, computing the SER requires the evaluation of an integral outside a given polytope in a high-dimensional space, for which a closed-form solution does not exist. Therefore, Monte Carlo (MC) simulation is typically used to estimate the SER, although a naive or raw MC implementation can be very inefficient at high signal-to-noise-ratios or for systems with stringent SER requirements. A reduced variance estimator is provided by the Truncated Hypersphere Importance Sampling (THIS) method, which samples from a proposal density that excludes the largest hypersphere circumscribed within the Voronoi region of the transmitted vector. A much more efficient estimator is provided by the existing ALOE (which stands for “At Least One rare Event”) method, which samples conditionally on an error taking place. The paper describes in detail these two IS methods, discussing their advantages and limitations, and comparing their performances.

**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 of performance evaluation for wireless communication links, for many of which an exact computation is typically unfeasible. Monte Carlo simulations are routinely used to estimate the symbol error rate (SER), the bit error rate (BER), or the outage probability in block fading channels. 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 is of the order of  $10^{-9}$  to  $10^{-12}$  [1], while ultra-reliable low-latency communications (URLLC) have stringent requirements in terms of outage probability (e.g. lower than  $10^{-6}$ ) [2]. For systems with such a stringent requirements in terms of SER/BER or outage probability, the need to perform in excess of  $10^{12}$  trials makes MC simulation infeasible.

To overcome this limitation of raw MC simulation, this paper studies efficient SER estimation techniques for multiple-input multiple-output (MIMO) communication systems based on importance sampling (IS) techniques. IS has been used as a method for variance reduction in SER or BER simulations in a wide range of scenarios since the late seventies [3]–[7]. Despite this, many digital communication researchers are still unaware of the potential benefits of IS techniques to char-

acterize the statistical performance of digital communication systems.

The basic IS methodology samples from a proposal distribution that increases the number of errors during simulation, and then weights the samples by the ratios of the target to the proposal densities [8]. As a single-proposal distribution for the MIMO detection problem, we propose in this paper to use a truncated Gaussian where the mass around its center has been removed. We call this IS method as *truncated hypersphere importance sampling* (THIS). However, covering the region where both the target density and the function to be evaluated take significant values with a single proposal is not possible in general. An alternative is to use more than one proposal through the so-called multiple importance sampling (MIS) technique [9], [10]. In this paper we use a MIS technique called ALOE (“At Least One rare Event”) [11], which is particularly well suited for SER estimation. 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 polytope. 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 region in the lattice of symbols transformed by the channel.

## II. MIMO DETECTION

We consider a multiple-input multiple-output (MIMO) spatial multiplexing system with  $L$  transmit and receive antennas, where the channel is unknown at the transmitter but perfectly known at the receiver side.<sup>1</sup> The received signal follows the well-known baseband model

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

where  $\mathbf{H}$  is an  $L \times L$  matrix whose columns represent the known complex channel gains from each transmit antenna to the  $L$  receive antennas,  $\mathbf{s} = (s_1, \dots, s_L)^T$  is the vector with the unknown transmitted symbols, and  $\mathbf{n} = (n_1, \dots, n_L)^T$  is the noise vector. The noise is modeled as  $\mathbf{n} \sim \mathcal{CN}_L(\mathbf{0}, \sigma^2 \mathbf{I})$ , where  $\mathcal{CN}_L(\mathbf{0}, \mathbf{R})$  denotes a complex Gaussian distribution

<sup>1</sup>The extension to MIMO systems with different number of Tx/Rx antennas is straightforward, but we stick to the square  $L \times L$  MIMO system for notational simplicity as well as to ease exposition.

in  $\mathbb{C}^L$  with zero mean and covariance  $\mathbf{R}$ . Each symbol  $\mathbf{s}$  belongs to a discrete constellation that we will assume to be an square  $M$ -QAM (Quadrature Amplitude Modulation) signal. The average energy per symbol is normalized so that  $E[|s_k|^2] = 1$ .

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

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

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,  $\mathbf{n} \sim \mathcal{N}_{2L}(\mathbf{0}, \frac{\sigma^2}{2}\mathbf{I})$ , and the vector  $\mathbf{s} \in \mathcal{D}^{2L}$  belongs to a  $2L$ -dimensional square real lattice with cardinality  $|\mathcal{D}^{2L}| = M^L$ .

The MIMO detection problem consists of estimating the complex symbol vector  $\hat{\mathbf{s}} \in \mathcal{D}^{2L}$  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

$$\hat{\mathbf{s}} = \min_{\mathbf{s} \in \mathcal{D}^{2L}} \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2. \quad (3)$$

Note that this problem is equivalent to constructing the Voronoi diagram with all transformed points  $\mathbf{H}\mathbf{s}$  with  $\mathbf{s} \in \mathcal{D}^{2L}$ , 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  $\mathbf{H}\mathbf{s}$  to a given noisy observation,  $\mathbf{x}$ , which is known to be NP-hard for generic channels [12], [13] meaning that the problem complexity is exponential on the dimension of the lattice [14]. Using the Finke-Phost sphere-decoding (SD) algorithm [15], ML detection can be achieved at an average complexity that is cubic in the number of transmit antennas, as was shown in [16]. Later Agrell et al. showed that the Schnorr-Euchner (SE) enumeration strategy [17] reduces the complexity of SD algorithms in comparison to [15], and very efficient SD implementations have been proposed in [18].

#### A. Problem statement

Let us assume that  $\mathbf{s}_m$  is transmitted and let  $\mathcal{V}_m$  be its associated Voronoi region, which is a polytope defined by the intersection of finitely many hyperplanes in  $\mathbb{R}^{2L}$ . The observation  $\mathbf{x} = \mathbf{H}\mathbf{s}_m + \mathbf{n}$  is the  $2L$ -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

$$P_e = \frac{1}{M^L} \sum_{m=1}^{M^L} p(e|\mathbf{s}_m), \quad (4)$$

where

$$p(e|\mathbf{s}_m) = \sum_{j \neq m} \frac{L_{j,m}}{L} p(\mathbf{s}_j|\mathbf{s}_m) = \sum_{j \neq m} \frac{L_{j,m}}{L} \text{Prob}(\mathbf{H}\mathbf{s}_m + \mathbf{n} \in \mathcal{V}_j) \quad (5)$$

is the conditional SER when  $\mathbf{s}_m$  is transmitted, and  $L_{j,m}$  is the number of complex symbol errors when  $\mathbf{s}_m$  is transmitted and  $\mathbf{s}_j$  is detected. Each term  $\text{Prob}(\mathbf{H}\mathbf{s}_m + \mathbf{n} \in \mathcal{V}_j)$  is the integral of a Gaussian distribution centered at  $\mathbf{r}_m = \mathbf{H}\mathbf{s}_m$  in the polytope region  $\mathcal{V}_j$ . Note that transmitting  $\mathbf{s}_m$  and deciding  $\mathbf{s}_j$  can result in a different number of complex symbol errors  $L_{j,m} \in \{1, \dots, L\}$ ; therefore each term in (5) is weighted by  $L_{j,m}/L$ . 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 raw 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 number of antennas,  $L$ , is large. Our goal thus is to develop efficient sampling schemes to estimate  $p(e|\mathbf{s}_m)$  in (5).

### III. EFFICIENT SER ESTIMATION VIA IMPORTANCE SAMPLING

For notational simplicity let us denote the transmitted vector as  $\mathbf{s}$ , with associated decision region  $\mathcal{V} \subset \mathbb{R}^{2L}$ , and let  $p = p(e|\mathbf{s})$  be the conditional SER of interest, which can be expressed as

$$p = \int_{\mathcal{V}} h(\mathbf{x}) \tilde{\pi}(\mathbf{x}) d\mathbf{x}, \quad (6)$$

where  $\tilde{\pi}(\mathbf{x}) \triangleq \mathcal{N}_{2L}(\mathbf{H}\mathbf{s}, \frac{\sigma^2}{2}\mathbf{I})$  is the Gaussian distribution of the observation, and the function  $h(\mathbf{x}) = \ell(\mathbf{x})/L$  is the ratio of wrongly detected symbols, i.e.,  $\ell(\mathbf{x})$  is the number of errors when  $\mathbf{s}$  is transmitted and  $\mathbf{x}$  is received. Note that  $h(\mathbf{x}) = 0$  when  $\mathbf{x}$  falls in the Voronoi decision region for  $\mathbf{s}$ .

#### A. Single-proposal importance sampling

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

$$\hat{p}^{(\text{IS})} = \frac{1}{N} \sum_{n=1}^N w_n h(\mathbf{x}_n), \quad \mathbf{x}_n \sim q(\mathbf{x}), \quad n = 1, \dots, N, \quad (7)$$

where  $w_n = \frac{\tilde{\pi}(\mathbf{x}_n)}{q(\mathbf{x}_n)}$  is the importance weight.

Let us first consider a Gaussian proposal  $q(\mathbf{x}) = \tilde{\pi}(\mathbf{x}) = \mathcal{N}_{2L}(\mathbf{H}\mathbf{s}, \frac{\sigma^2}{2}\mathbf{I})$ . This particular case yields the standard raw Monte Carlo algorithm whose efficiency decays when  $\sigma^2$  decreases. This is an illustrative example of how IS is an efficient technique if  $q(\mathbf{x})$  is close to  $h(\mathbf{x})\tilde{\pi}(\mathbf{x})$ . However, the variance of the IS estimator increases when too many samples from  $q(\mathbf{x})$  fall in  $\mathcal{V}$ , and hence  $h(\mathbf{x})$  takes value zero.

We present here a single-proposal IS method that increases the efficiency by avoiding the simulated samples to be close to  $\mathbf{H}\mathbf{s}$ . We call the method as *truncated hypersphere importance sampling* (THIS). The method only requires to know the distance  $d_{\min}$  to the closest neighbor of  $\mathbf{H}\mathbf{s}$ . Then, we set the IS proposal to a truncated Gaussian distribution  $q(\mathbf{x}; d_{\min}/2) \propto \tilde{\pi}(\mathbf{x}) \mathbb{I}_{\overline{\mathcal{R}}_{d_{\min}/2}}(\mathbf{x})$ , where  $\overline{\mathcal{R}}_{d_{\min}/2}$  is the hypersphere of radius  $d_{\min}/2$ , i.e.,  $q(\mathbf{x}; d_{\min}/2)$  is a distribution proportional to the

targeted Gaussian where the mass around its center has been removed. Note that by construction,  $q(\mathbf{x}; d_{\min}/2)$  covers all the support where the integrand of Eq. (6) takes values different from zero because the hypersphere is inside the Voronoi decision region of  $\mathbf{H}_s$ , i.e.,  $\mathcal{R}_{d_{\min}/2} \subset \mathcal{V}$ . The estimator of THIS is built as

$$\hat{p}^{(\text{THIS})} = \frac{Z_q}{N} \sum_{n=1}^N h(\mathbf{x}_n), \quad \mathbf{x}_n \sim q(\mathbf{x}; d_{\min}/2), \quad n = 1, \dots, N, \quad (8)$$

where  $Z_q$  is the normalizing constant of the proposal, i.e.,  $Z_q = \int_{\mathcal{R}_{d_{\min}/2}} \mathcal{N}_{2L}(\mathbf{H}_s, \frac{\sigma^2}{2} \mathbf{I}) d\mathbf{x}$  and it can be easily computed. In Appendix A we describe how to sample from the truncated Gaussian and the computation of  $Z_q$ . Note that, for every sample  $\mathbf{x}_n$ , the SD algorithm is run in order to find the ML estimate of the transmitted vector of symbols, and hence, the value of the ratio of errors  $h(\mathbf{x}_n)$ . The estimator  $\hat{p}^{(\text{THIS})}$  is unbiased, consistent, and guarantees a variance reduction w.r.t. the raw MC estimator, i.e.,  $\text{Var}[\hat{p}^{(\text{THIS})}] < \text{Var}[\hat{p}^{(\text{MC})}]$ . See Appendix B for more details.

### B. Multiple importance sampling

Multiple importance sampling (MIS) is a natural extension of IS allowing for the simulation from a set of  $K$  proposals,  $\{q_k(\mathbf{x})\}_{k=1}^K$ , instead of just one [9]. However, the extension from one to several proposals is not straightforward and many sampling and weighting schemes can be devised (see [10] for a review). A conventional way to proceed is to simulate from the mixture proposal as

$$\mathbf{x}_n \sim q_\alpha = \sum_{k=1}^K \alpha_k q_k(\mathbf{x}), \quad n = 1, \dots, N, \quad (9)$$

where  $\alpha = [\alpha_1, \dots, \alpha_K]$  is a simplex vector with all non-negative weights in the mixture such that  $\sum_{k=1}^K \alpha_k = 1$ . The MIS extension of Eq. (7) is

$$\hat{p}^{(\text{MIS})} = \frac{1}{N} \sum_{n=1}^N \frac{h(\mathbf{x}_n) \tilde{\pi}(\mathbf{x}_n)}{q_\alpha(\mathbf{x}_n)}, \quad (10)$$

i.e., the new importance weight is now  $w_n = \frac{\tilde{\pi}(\mathbf{x}_n)}{q_\alpha(\mathbf{x}_n)}$ .

We recall that we aim at integrating  $\tilde{\pi}(\mathbf{x})$  in region  $\bar{\mathcal{V}}$ , weighting each sample with  $h(\mathbf{x}) = \frac{\ell(\mathbf{x})}{L}$ , the ratio of errors in the detector when  $\mathbf{x}$  is received. The region  $\bar{\mathcal{V}}$  can be described as the union of all half-spaces in  $\mathbb{R}^{2L}$  generated by the hyperplanes that define the border of  $\mathcal{V}$ .

A natural way to adapt this MIS scheme to our problem is to follow the choice of proposals in [11] called ALOE for ‘‘At Least One rare Event’’. In ALOE the number of proposals is the number of hyperplanes defining the Voronoi region and each proposal is a truncated version of a Gaussian centered at the received symbol beyond each hyperplane. More specifically,  $q_k(\mathbf{x}) = \frac{\mathbb{I}_{\mathcal{S}_k}(\mathbf{x}) \tilde{\pi}(\mathbf{x})}{P_k}$ , where  $P_k = \int \mathbb{I}_{\mathcal{S}_k}(\mathbf{x}) \tilde{\pi}(\mathbf{x}) d\mathbf{x}$  is the integral of the target distribution beyond the hyperplane  $\mathcal{S}_k = \{\mathbf{x} \in \mathbb{R}^{2L} \mid \mathbf{x}^T \boldsymbol{\gamma}_k \geq \beta_k\}$  (parametrized by  $\boldsymbol{\gamma}_k$  and  $\beta_k$ ), and  $\bar{\mathcal{V}} = \bigcup_{k=1}^K \mathcal{S}_k$ . The procedure for the efficient simulation

from a generic truncated Gaussian distribution is described in [11]. ALOE has recently been used for the SER estimation of single-input single-output (SISO) AWGN channels with non-square two-dimensional constellations in [19]. 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 determining a Voronoi region can be very large and computing all hyperplanes is in general not feasible. In this work we use the SD algorithm to find  $K$  lattice points that belong to a hypersphere centered at  $\mathbf{H}_s$ , where  $K$  is a fixed number chosen to limit the complexity of this step. These points determine the  $K$  hyperplanes to be used in the mixture proposal (9). Our experiments suggest that selecting  $K = \lceil 2 \log_2(M) L^2 \rceil$  provides accurate SER estimates.

Let us summarize the implementation of a modified ALOE for approximating the integral of Eq. (6). We first define  $\bar{p} = \sum_{k=1}^K P_k$ , which is an upper union bound of  $p$ . Then, in ALOE, Eq. (10) yields

$$\hat{p}^{(\text{ALOE})} = \frac{1}{N} \sum_{n=1}^N \frac{h(\mathbf{x}_n) \tilde{\pi}(\mathbf{x}_n)}{q_\alpha(\mathbf{x}_n)} \quad (11)$$

$$= \frac{1}{N} \sum_{n=1}^N \frac{h(\mathbf{x}_n)}{\sum_{k=1}^K \alpha_k \mathbb{I}_{\mathcal{S}_k}(\mathbf{x}_n) P_k^{-1}}. \quad (12)$$

The weight of each proposal in the mixture defined in Eq. (9) is chosen as  $\alpha_k = P_k/\bar{p}$ , for  $k = 1, \dots, K$ . Then, ALOE estimator is

$$\hat{p}^{(\text{ALOE})} = \frac{\bar{p}}{N} \sum_{n=1}^N \frac{h(\mathbf{x}_n)}{\sum_{k=1}^K \mathbb{I}_{\mathcal{S}_k}(\mathbf{x}_n)} = \frac{\bar{p}}{N} \sum_{n=1}^N \frac{h(\mathbf{x}_n)}{C(\mathbf{x}_n)}, \quad (13)$$

where  $C(\mathbf{x}_n) = \sum_{k=1}^K \mathbb{I}_{\mathcal{S}_k}(\mathbf{x}_n)$  is the number of half-spaces  $\mathcal{S}_k$  where  $\mathbf{x}_n$  is present. ALOE estimator is unbiased, and

$$\text{Var}(\hat{p}^{(\text{ALOE})}) \leq \frac{p(\bar{p} - p)}{N},$$

i.e., the bound of the variance decays with the number of samples, and the constant factor depends on the true probability,  $p$ . It is straightforward to understand the high performance of ALOE for high SNR scenarios for which  $p$  is small. A deeper theoretical analysis of ALOE can be found in [11].

## IV. SIMULATION RESULTS

We consider an example with  $L = 6$  antennas and  $M = 64$ . As a figure of merit for the different estimators we use the relative root mean square error, defined as

$$\text{RRMSE}(\hat{p}) = \frac{\sqrt{\text{Var}(\hat{p})}}{p}, \quad (14)$$

where  $p$  is the true probability. We compare the performance of raw Monte Carlo (Raw-MC), ALOE and THIS. For the three methods, the SER at each  $Eb/N_0$  is estimated with a total of  $10^5$  random samples. More specifically, we generate 100 different vectors of symbols (i.e., lattice points) and, for each lattice point,  $N = 1000$  samples are drawn from each

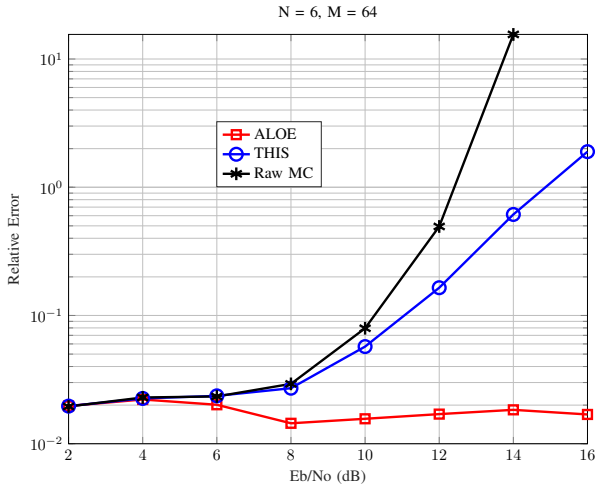


Fig. 1: RRMSE with Raw-MC, ALOE and THIS. Number of antennas  $L = 6$  and 64-QAM symbols.

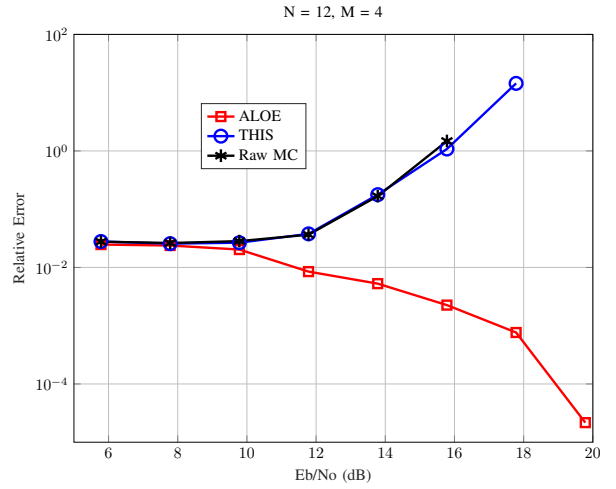


Fig. 3: RRMSE with Raw-MC, ALOE and THIS. Number of antennas  $L = 12$  and QPSK symbols.

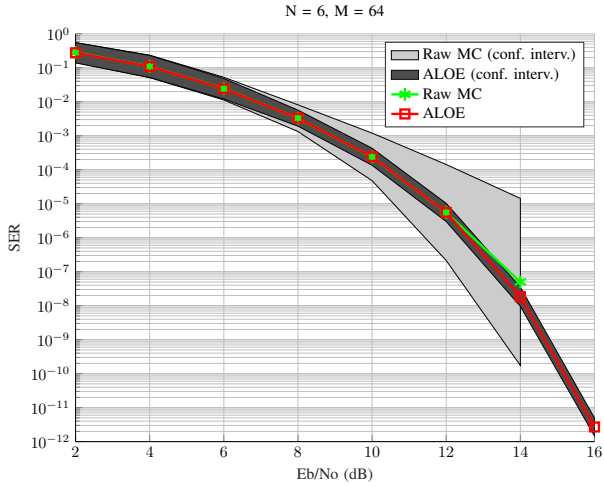


Fig. 2: SER vs.  $Eb/N_0$  for Raw-MC (confidence interval in grey) and ALOE (confidence interval in black).

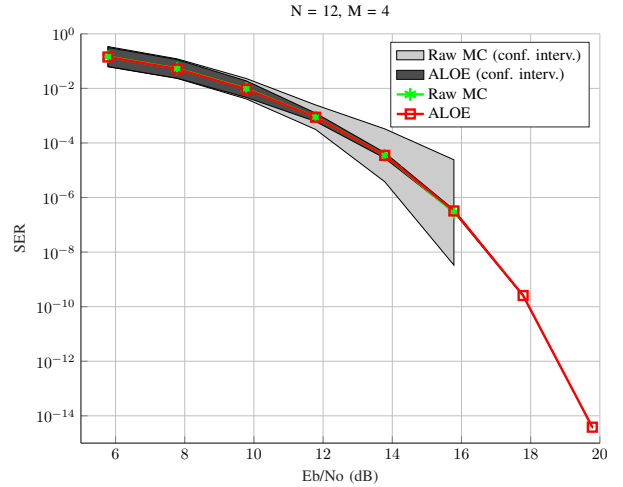


Fig. 4: SER vs.  $Eb/N_0$  for Raw-MC (confidence interval in grey) and ALOE (confidence interval in black).

proposal. The RRMSE results are shown in Fig. 1. We observe that the gap between raw Monte Carlo, THIS and ALOE increases as the SNR increases. The SER curves for MC and ALOE along with their confidence intervals are depicted in Fig. 2. Moreover, Figs. 3 and 4 show the RRMSE and the SER curves for the different estimators when  $L = 12$  and  $M = 4$ . Again, ALOE provides extremely accurate SER estimates with just  $10^5$  random samples.

## V. DISCUSSION

The integral in Eq. (6) can be approximated using the raw Monte Carlo by sampling directly from  $\tilde{\pi}(\mathbf{x})$ . However, the efficiency of the method decays when most of the samples fall inside the Voronoi decision region of  $\mathbf{H}_s$ . Importance sampling can alleviate this problem by avoiding or at least decreasing the waste of samples that fall in  $\mathcal{V}$ . The proposed THIS

estimator increases the efficiency by avoiding to sample in the hypersphere centered at  $\mathbf{H}_s$  with radius  $d_{\min}/2$ . The advantage is that the only information it requires is the closest symbol to  $\mathbf{H}_s$  and its distance  $d_{\min}$ . The disadvantage is that in higher dimension, the volume of the  $2L$ -dimensional hypersphere of radius  $d_{\min}/2$  can be small w.r.t. the total area of the Voronoi decision region of  $\mathbf{H}_s$ . In that case, the efficiency of THIS can deteriorate, although it will always be better than raw Monte Carlo. The advantage of ALOE is that all samples have at least one error, i.e.,  $h(\mathbf{x}_n) > 0$  for all  $\mathbf{x}_n$ , which is a key feature of the algorithm, particularly for high SNR when most samples in raw Monte Carlo give  $h(\mathbf{x}_n) = 0$ .

## ACKNOWLEDGMENT

The work of V. Elvira was partially supported by *Agence Nationale de la Recherche* of France under PISCES project

(ANR-17-CE40-0031-01) and the French-American Fulbright Commission. The work of I. Santamaria was partly supported by the Ministerio de Economía y Competitividad (MINECO) of Spain, and AEI/FEDER funds of the E.U., under grant TEC2016-75067-C4-4-R (CARMEN).

#### APPENDIX A: SAMPLING IN THIS

THIS samples from the proposal  $q(\mathbf{x}; a) = \frac{\mathcal{N}_{2L}(\mathbf{H}\mathbf{s}, \frac{\sigma^2}{2}\mathbf{I})\mathbb{I}_{\overline{\mathcal{R}}_a}(\mathbf{x})}{Z_q}$ , i.e., an isotropic Gaussian distribution truncated out of the hyper-sphere  $a = d_{\min}/2$ , by using standard methods used in directional statistics [20]. We proceed as follows:

- 1) Simulate  $r \sim \chi_{(2L)}^2(r|r > a)$ , i.e., from the truncated  $\chi^2$  distribution with  $2L$  degrees of freedom, defined in the interval  $(a, \infty)$ . This can be efficiently done as follows:
  - Find  $c$  as the cdf of the distribution  $\chi_{(2L)}^2$  at  $(\frac{a}{\sigma})^2$ .
  - 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  $\chi_{(2L)}^2$ .
- 2) Simulate  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  of dimension  $2L$ .
- 3) Project  $\mathbf{z}$  in the  $2L - 1$  unit sphere  $\boldsymbol{\theta} = \frac{\mathbf{z}}{\|\mathbf{z}\|}$ .
- 4) Output  $\mathbf{x} = \sigma\boldsymbol{\theta}\sqrt{r}$ .

The normalizing constant of the THIS proposal can be computed as  $Z_q = \int_{\overline{\mathcal{R}}_{d_{\min}/2}} \mathcal{N}_{2L}(\mathbf{H}\mathbf{s}, \frac{\sigma^2}{2}\mathbf{I})d\mathbf{x} = 1 - c$ , where  $c$  is the volume of hypersphere that has been removed from the Gaussian distribution.

#### APPENDIX B: THEORETICAL ANALYSIS OF THIS

The unbiasedness of THIS can be shown as

$$\begin{aligned} \mathbf{E}_{q(\mathbf{x}; d_{\min}/2)}[\widehat{p}^{(\text{THIS})}] &= \mathbf{E}_{q(\mathbf{x}; d_{\min}/2)}\left[\frac{Z_q}{N}\sum_{n=1}^N h(\mathbf{x}_n)\right] \\ &= \int_{\overline{\mathcal{V}}} \frac{Z_q}{N}\sum_{n=1}^N h(\mathbf{x}_n)q(\mathbf{x}_n; d_{\min}/2)d\mathbf{x}_n \\ &= \int_{\overline{\mathcal{V}}} \frac{Z_q}{N}\sum_{n=1}^N h(\mathbf{x}_n)\frac{\tilde{\pi}(\mathbf{x}_n)\mathbb{I}_{\overline{\mathcal{R}}_{d_{\min}/2}}(\mathbf{x}_n)}{Z_q}d\mathbf{x}_n \\ &= \frac{1}{N}\sum_{n=1}^N \int_{\overline{\mathcal{V}}} h(\mathbf{x}_n)\tilde{\pi}(\mathbf{x}_n)\mathbb{I}_{\overline{\mathcal{R}}_{d_{\min}/2}}(\mathbf{x}_n)d\mathbf{x}_n \\ &= \int_{\overline{\mathcal{V}}} h(\mathbf{x}_n)\tilde{\pi}(\mathbf{x}_n)d\mathbf{x}_n = p, \end{aligned}$$

since  $\overline{\mathcal{V}} \subseteq \overline{\mathcal{R}}_{d_{\min}/2}$ . The variance of the estimator of THIS is

$$\begin{aligned} \text{Var}[\widehat{p}^{(\text{THIS})}] &= \frac{1}{N^2}\sum_{n=1}^N \left[ \int_{\overline{\mathcal{V}}} \frac{h^2(\mathbf{x})\tilde{\pi}^2(\mathbf{x})}{q(\mathbf{x}; d_{\min}/2)}d\mathbf{x} - p^2 \right] \\ &= \frac{1}{N}\int_{\overline{\mathcal{V}}} \frac{h^2(\mathbf{x})\tilde{\pi}^2(\mathbf{x})}{q(\mathbf{x}; d_{\min}/2)}d\mathbf{x} - \frac{p^2}{N} \\ &= \frac{Z_q}{N}\int_{\overline{\mathcal{V}}} \frac{h^2(\mathbf{x})\tilde{\pi}^2(\mathbf{x})}{\tilde{\pi}(\mathbf{x})\mathbb{I}_{\overline{\mathcal{R}}_{d_{\min}/2}}(\mathbf{x})}d\mathbf{x} - \frac{p^2}{N} \\ &= \frac{Z_q}{N}\int_{\overline{\mathcal{V}}} h^2(\mathbf{x})\tilde{\pi}(\mathbf{x})d\mathbf{x} - \frac{p^2}{N}. \end{aligned} \quad (15)$$

Similarly, the variance of the raw MC estimator with  $q(\mathbf{x}) = \tilde{\pi}(\mathbf{x})$  can be computed as

$$\begin{aligned} \text{Var}[\widehat{p}^{(\text{MC})}] &= \frac{1}{N^2}\sum_{n=1}^N \left[ \int_{\overline{\mathcal{V}}} \frac{h^2(\mathbf{x})\tilde{\pi}^2(\mathbf{x})}{\tilde{\pi}(\mathbf{x})}d\mathbf{x} - \mathbf{E}_{\tilde{\pi}(\mathbf{x})}^2[\widehat{p}^{(\text{MC})}] \right] \\ &= \frac{1}{N}\int_{\overline{\mathcal{V}}} \frac{h^2(\mathbf{x})\tilde{\pi}^2(\mathbf{x})}{\tilde{\pi}(\mathbf{x})}d\mathbf{x} - \frac{p^2}{N} \\ &= \frac{1}{N}\int_{\overline{\mathcal{V}}} h^2(\mathbf{x})\tilde{\pi}(\mathbf{x})d\mathbf{x} - \frac{p^2}{N}. \end{aligned} \quad (16)$$

By simply comparing Eqs. (15) and (16), and since  $Z_q < 1$ , it follows that  $\text{Var}[\widehat{p}^{(\text{THIS})}] < \text{Var}[\widehat{p}^{(\text{MC})}]$ .

#### REFERENCES

- [1] R. Papannanreddy, *Lightwave Communication Systems: A Practical Perspective*. Penram Int. Publishing, 2014.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Efficient multiple importance sampling estimators," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1757–1761, 2015.
- [10] —, "Generalized multiple importance sampling," *Statistical Science*, vol. 34, no. 1, pp. 129–155, 2019.
- [11] 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.
- [12] M. Grotscchel, L. Lovasz, and A. Schriver, *Geometric Algorithms and Combinatorial Optimization*. Springer Verlag, 2nd ed., 1993.
- [13] 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.
- [14] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014.
- [15] 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.
- [16] 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.
- [17] C. P. Schnorr and M. Euchner, "Lattice basis reduction: Improved practical algorithms and solving subsets sum problems," *Math. Programm.*, vol. 66, no. 2, pp. 181–199, 1994.
- [18] A. Ghasemmehdi and E. Agrell, "Faster recursions in sphere decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3530–3536, 2011.
- [19] 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.
- [20] K. V. Mardia and P. E. Jupp, *Directional statistics*. John Wiley & Sons, 2009, vol. 494.