

Adaptive Importance Sampling and Quasi-Monte Carlo Methods for 6G URLLC Systems

Xiongwen Ke, Houying Zhu, Kai Yi, Gaoning He, Ganghua Yang, Yu Guang Wang

Abstract—In this paper, we propose an efficient simulation method based on adaptive importance sampling, which can automatically find the optimal proposal within the Gaussian family based on previous samples, to evaluate the probability of bit error rate (BER) or word error rate (WER). These two measures, which involve high-dimensional black-box integration and rare-event sampling, can characterize the performance of coded modulation. We further integrate the quasi-Monte Carlo method within our framework to improve the convergence speed. The proposed importance sampling algorithm is demonstrated to have much higher efficiency than the standard Monte Carlo method in the AWGN scenario.

Index Terms—adaptive importance sampling, Monte Carlo, quasi-Monte Carlo, bit error rate, error probability

I. INTRODUCTION

In wireless communication, reliability is assessed by measuring the correct rate of packets of a certain size delivered within a specified time. For a single user, N packets can be specified to be sent according to business requirements, the packet transmission delay can be measured, and reliability can be evaluated by recording whether the packet transmission delay is within the specified time. The reliability (or success rate) of packet transmission in traditional mobile broadband (MBB) systems (2G/3G/4G) is usually at the level of 0.1 or 0.01. This is because they mainly serve for voice and video transmission between humans who are relatively delay tolerant. In addition to MBB, 5G/6G introduces URLLC systems that are designed for communications between “intelligence” including humans and machines which ask for higher reliability at the level of 10^{-6} or even to 10^{-9} [1]. This enables many future applications such as telesurgery, robot cooperation, and remote control of machines. However, the efficiency of evaluating a high-reliability system is becoming a crucial problem since generating error packets at very low probability becomes a rare event and therefore performing Monte-Carlo simulation is time-consuming. For example, suppose the reliability target value is 99.999% (32bytes @ x ms), assuming that 100,000 packets are sent, if the delay when transmitting 99,999 packets is less than or equal to x ms, then the reliability of transmitting 32-byte packets within x ms is 99.999%. For the whole system, the required statistic is the proportion of users who satisfy 99.999% of the total number of simulated users. Then, 10^6 sampling points are needed to find the incorrect packets

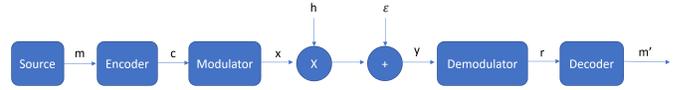


Fig. 1. Illustration of system model

with traditional sampling methods, which would exert an impractical big cost.

Importance sampling (IS) technology provides a remedy, which can reduce the variance of the MC estimator by tiling or scaling the original target distribution as a proposal. The target function in practice is usually complex, like modulated codewords with noise. Using importance sampling, the resulting biased distribution would have a higher frequency of occurrence of rare events. This biased distribution will be corrected by using the importance weight. The optimal proposal distribution is not known since it depends on the black-box integral function. In fact, even if the optimal proposal distribution is known, it is usually difficult to sample. Thus, the proper choice of the family of proposal distribution and its parameter is crucial to find a sub-optimal proposal. For a gentle introduction to the importance of sampling in communications systems, see [2].

Two families of distributions have often been used as proposals in the literature on communications systems. [3], [4] used the Bernoulli distribution as a proposal for importance sampling to evaluate the performance of binary linear block codes. [3] also developed an adaptive approach to update the parameters in the Bernoulli proposal. [5] used the Gaussian proposal for importance sampling to estimate the random coding Union (RCU) bound and gave the asymptotic analysis of the estimator. However, the RCU bound [6] is only the theoretical maximal channel coding rate achievable at a given block length and error probability, which is a bit far away from practice. The Gaussian proposal has also been used to evaluate the performance of Low-density Parity Check (LDPC) Codes in an AWGN channel [7] and Rayleigh Fading Channels [8]. However, both of them didn't discuss how to determine the parameter for the Gaussian proposal.

Our paper makes two main contributions. First, we introduce an adaptive importance sampling scheme based on the work of Bugallo et al. [9] to update the parameters of our Gaussian proposal. Second, instead of relying on traditional random number generators, we use Quasi-Monte Carlo (QMC) and randomized QMC methods to sample from our Gaussian proposal. QMC and RQMC offer a substitute for the widely

X. Ke and K. Yi are with The University of New South Wales, Australia; H. Zhu is with Macquarie University, Australia; G. He and G. Yang are with Huawei Technologies, Co. Ltd., Shanghai, China; Y. G. Wang is with Shanghai Jiao Tong University, China. Email: yuguang.wang@sjtu.edu.cn

utilized Monte Carlo method in scientific computing and financial analysis [10]–[12]. Unlike Monte Carlo, which uses a random sampling approach, QMC uses a deterministic rule to generate points. As a result, QMC can achieve the same level of integration error with fewer points, leading to increased efficiency in the integration process, thus higher efficiency in importance sampling. In this paper, we focus on additive white Gaussian noise (AWGN) in link-level simulation when the encoder and decoder are omitted, see Figure 1.

II. MODEL SETTING AND PERFORMANCE MEASURE

Let $m \in \mathcal{X}^k$ be the message vector. A codeword $c \in \mathcal{X}^n$ is obtained by encoding this message word. The signal X is the sequence after modulation of codeword c . At the output of the channel, a word $r \in \mathcal{X}^n$ perturbed by some noise before demodulation, is what we can observe in practice. The decoder will try to recover the observation r to the original message m . In communication system modelling, we can have the binary phase-shift keying (BPSK) modulation with additive white Gaussian noise (AWGN) channels or Rayleigh fading channels. For example, in Figure 1, when we set $h \equiv 1$ and ϵ as Gaussian distribution, we obtain the AWGN channel.

The performance of the codes is measured by some metric, averaged over all words r generated by some independent drawings from some joint distribution $\pi(Y|X)\pi(X)$, and then transferred by the demodulator. For example, by averaging the 0-1 loss (i.e. if the received word is erroneously decoded, then the loss is 1 and 0 otherwise), we get the probability of word error (WER). In general, we can write $P_e = E_{\pi(Y)}[I(Y)] \approx \frac{1}{N} \sum_{i=1}^N I(y_i)$, where $y_i \sim \pi(Y)$. Due to the complex structure of both the demodulator and decoder and the length of the code, we cannot explicitly write down the active area of indicator function $I(Y)$. We need to use Monte Carlo simulation to evaluate this high-dimensional integration. When the SNR is high, P_e could be very small. For the case of WER, the relative error can be approximated by $\frac{\sqrt{\text{Var}[\hat{P}_e]}}{P_e} \approx \frac{1}{\sqrt{P_e N}}$, which indicates that, for rare event sampling, a large sample size is required to maintain accuracy.

III. IMPORTANCE SAMPLING

To sample the rare event, a good candidate solution is importance sampling (IS)

$$\hat{P}_e \approx \frac{1}{N} \sum_{i=1}^N I(y_i) \frac{\pi(y_i)}{g(y_i)} = \frac{1}{N} \sum_{i=1}^N I(y_i) w(y_i), \quad (1)$$

where $w(y_i) = \frac{\pi(y_i)}{g(y_i)}$, $y_i \sim g(Y)$ and $g(Y)$ is the density of proposal distribution, which has higher probability to incur the error. There also exists another estimator for IS, which only requires both the proposal and target distribution known up to a normalizing constant by normalizing the importance weight: $\hat{P}_e \approx \frac{\sum_{i=1}^N I(y_i) w(y_i)}{\sum_{i=1}^N w(y_i)}$. Note that $\pi(Y) = \int \pi(Y|X)\pi(X)dx$ has no closed form in practice. For example, in AWGN channel, $\pi(Y|X)$ is the pdf of $\prod_{i=1}^n N(y_i, \sigma^2)$ and X can follow the discrete uniform distribution. Then, we have two strategies:

- 1) We jointly sample Y and X based on the proposal $g(Y, X) = g(Y|X)\pi(X)$. In this case the importance weight change to $w(y_i, x_i) = \frac{\pi(y_i|x_i)}{g(y_i|x_i)}$.
- 2) We apply the conditional importance sampling approach [13], [14] by first sampling $x_i \sim \pi(X)$, then conditional on x_i , we use importance sampling to evaluate $E_{\pi(Y)}[I(Y)|X]$. In this case, we have the IS estimator
$$P_e = E_{\pi(X)}[E_{\pi(Y|X)}[I(Y)|X]] \approx \frac{1}{S} \sum_{j=1}^S \frac{\sum_{i=1}^N I(Y_{ij})w(y_{ij}, x_j)}{\sum_{i=1}^N w(y_{ij}, x_j)}. \quad (2)$$

This estimator will work well if the error is dominated by the noise ϵ and the codeword x has a minor impact on the error. For example, when the loss is WER, we could have

$$E_{\pi(Y|X)}[I(Y)^2|X] = E_{\pi(Y|X)}[I(Y)|X] \approx p_e. \quad (3)$$

Then, by the law of total variance,

$$\begin{aligned} \text{Var}_{\pi(Y)}[I(Y)] &= E_{\pi(X)}[\text{Var}_{\pi(Y|X)}[I(Y)|X]] \\ &\quad + \text{Var}_{\pi(X)}[E_{\pi(Y|X)}[I(Y)|X]] \\ &\approx E_{\pi(X)}[\text{Var}_{\pi(Y|X)}[I(Y)|X]] \\ &\approx \text{Var}_{\pi(Y|X)}[I(Y)|X] = p_e - p_e^2. \end{aligned} \quad (4)$$

In other words, the variance of the IS estimator in equation (2) is close to the case that we have the analytic form of $\pi(y_i)$. This implies that we can sample x_i such that $S \ll N$. In the extreme case, $S = 1$.

Conditional on X , the optimal proposal is given by

$$g^*(Y|X) = \frac{I(Y)\pi(Y|X)}{E_{\pi(Y|X)}[I(Y)|X]}, \quad (5)$$

which is impossible to sample. The key problem in importance sampling is designing a sub-optimal proposal. We need to balance the following four points.

- 1) The proposal distribution can efficiently sample rare events compared to the target distribution.
- 2) The importance weight can be evaluated quickly.
- 3) Sampling from the proposal distribution is straightforward.
- 4) The effective sample size should not be too small.

IV. IMPORTANCE SAMPLING WITH AWGN CHANNEL

If codeword x is sent over the AWGN channel, then the conditional target distribution is $y|x \sim N(x, \sigma^2 I_n)$. To tackle the rare event simulation, we consider two approaches to construct the proposal distribution.

A. Exponential Tilting

Exponential Tilting is used in Monte Carlo estimation for rare-event simulation, which will shift the target distribution. Suppose $\pi(Y, X)$ is the pdf of normal distribution $N(x, \sigma^2 I_n)$, the tilted density $g_\theta(y)$ is $N(x + \theta, \sigma^2 I_n)$, where $\theta = (\theta_1, \dots, \theta_n)$ is the vector of tilting parameters we need to tune. Then, the unnormalize importance weight reads

$$\frac{\pi(y_i|x_i)}{g_\theta(y_i|x_i)} = \exp \left[-\frac{(y_i - x_i)^T \theta}{\sigma^2} \right] \exp \left[\frac{\|\theta\|^2}{2\sigma^2} \right]. \quad (6)$$

The resulting IS estimator is

$$\hat{p}_e \approx \frac{\sum_{i=1}^N \exp\left[-\frac{(y_i - x_i)^T \theta}{\sigma^2}\right] I(y_i)}{\sum_{i=1}^N \exp\left[-\frac{(y_i - x_i)^T \theta}{\sigma^2}\right]}, \quad (7)$$

or if we use conditional importance sampling,

$$\hat{p}_e \approx \frac{1}{S} \sum_{j=1}^S \frac{\sum_{i=1}^N \exp\left[-\frac{(y_{ij} - x_j)^T \theta}{\sigma^2}\right] I(y_{ij})}{\sum_{i=1}^N \exp\left[-\frac{(y_{ij} - x_j)^T \theta}{\sigma^2}\right]}. \quad (8)$$

The variance of importance sampling is

$$\begin{aligned} & \text{Var}_{g(Y|X)\pi(X)} \left[I(Y) \frac{\pi(Y|X)}{g(Y|X)} \right] \\ &= E_{\pi(Y|X)\pi(X)} \left[I(Y) \frac{\pi(Y|X)}{g(Y|X)} \right]^2 - p_e^2, \end{aligned} \quad (9)$$

which is not possible to obtain. But our problem is to find the tilting parameter θ to minimize the variance above. Since only the first term of the equation (9) depends on θ , we can find a surrogate objective function, which is the second-moment estimator for IS.

Lemma 1. *The second moment estimator for IS in the exponential tilting case is*

$$E_{\theta}[\hat{p}_e^2] = \frac{\sum_{i=1}^N I(y_i)^2 \exp\left[\frac{-2(y_i - x_i)^T \theta + \|\theta\|_2^2}{2\sigma^2}\right]}{N} \quad (10)$$

with its derivative

$$\begin{aligned} \frac{\partial E_{\theta}[\hat{p}_e^2]}{\partial \theta} &= \frac{1}{N} \sum_{i=1}^N I(y_i)^2 \exp\left[\frac{-2(y_i - x_i)^T \theta + \|\theta\|_2^2}{2\sigma^2}\right] \\ &\quad \times \left[\frac{-2(y_i - x_i) + 2\theta}{2\sigma^2} \right], \end{aligned} \quad (11)$$

where (x_i, y_i) is sampled from the target distribution $\pi(Y, X)$. In addition, $\frac{\partial^2 E_{\theta}[\hat{p}_e^2]}{\partial \theta^2} > 0$ for all θ_i , which implies that the second moment estimator for IS in (10) is a convex function with respect to the variable θ .

Therefore, we have the following theorem.

Theorem 2. *The optimal tilting parameter $\hat{\theta}$ that minimizes the variance of the IS estimator given by (7) or (8) is*

$$\hat{\theta} = \frac{\sum_{i=1}^N I(y_i)^2 (y_i - x_i) \exp\left[-\frac{(y_i - x_i)^T \hat{\theta}}{\sigma^2}\right]}{\sum_{i=1}^N I(y_i)^2 \exp\left[-\frac{(y_i - x_i)^T \hat{\theta}}{\sigma^2}\right]}, \quad (12)$$

where (x_i, y_i) is sampled from $\pi(Y, X)$. Thus, as $N \rightarrow \infty$,

$$\begin{aligned} \hat{\theta} &\xrightarrow{a.s.} \frac{E_{g(Y,X)} \left[I(Y)^2 (Y - X) \frac{\pi^2(Y,X)}{g^2(Y,X)} \right]}{E_{g(Y,X)} \left[I(Y)^2 \frac{\pi^2(Y,X)}{g^2(Y,X)} \right]} \\ &\approx \frac{\sum_{i=1}^N I(y_i)^2 (y_i - x_i) \exp\left[-\frac{2(y_i - x_i)^T \hat{\theta}}{\sigma^2}\right]}{\sum_{i=1}^N I(y_i)^2 \exp\left[-\frac{2(y_i - x_i)^T \hat{\theta}}{\sigma^2}\right]}, \end{aligned} \quad (13)$$

where $g(Y|X)$ could be the pdf of $N(x + \theta, \sigma^2 I_n)$ for any θ and (x_i, y_i) is sampled from $g(Y|X)\pi(X)$.

Remark. *Lemma 1 and equation (12) in Theorem 2 tell us that it is not possible to obtain a closed-form expression for $\hat{\theta}$, but the numerical result can be obtained by using gradient descent with convergence guaranteed by the convexity of $E_{\theta}[\hat{p}_e^2]$. However, evaluating the gradient from equation (11) needs to sample the rare event from the target distribution. When the SNR is high, the Monte Carlo gradient from (11) might be unreliable. A remarkable consequence of equation (13) is that we can obtain an unbiased IS estimator for $\hat{\theta}$ by using any exponential tilting normal distribution. This provides two merits: 1. Sampling the rare event from the tilting distribution is easier. 2. We can update the estimator $\hat{\theta}$ based on samplers from the previous proposal.*

B. Scaling Variance

Another way to sample the rare event is by scaling the variance of the target distribution. In this case, the scale density $g_c(y)$ is $N(x, c\sigma^2 I_n)$, where $c > 1$ is the scaling parameter to be tuned. The unnormalize importance weight then reads

$$\frac{\pi(y_i|x_i)}{g_{\theta}(y_i|x_i)} = c^{\frac{n}{2}} \exp\left[\left(\frac{1}{c} - 1\right) \frac{\|y_i - x_i\|_2^2}{2\sigma^2}\right]. \quad (14)$$

Similar to exponential tilting, we can obtain the optimal IS estimator using scaling variance.

Lemma 3. *The second moment estimator for IS in the scaling variance case is*

$$E_c[\hat{p}_e^2] = \frac{c^{\frac{n}{2}} \sum_{i=1}^N I(y_i)^2 \exp\left[\left(\frac{1}{c} - 1\right) \frac{\|y_i - x_i\|_2^2}{2\sigma^2}\right]}{N} \quad (15)$$

with its derivative given by

$$\begin{aligned} \frac{\partial E_c[\hat{p}_e^2]}{\partial c} &= \frac{n}{2} c^{\frac{n}{2}-1} \frac{\sum_{i=1}^N I(y_i)^2 \exp\left[\left(\frac{1}{c} - 1\right) \frac{\|y_i - x_i\|_2^2}{2\sigma^2}\right]}{N} \\ &\quad - c^{\frac{n}{2}-2} \frac{\sum_{i=1}^N I(y_i)^2 \exp\left[\left(\frac{1}{c} - 1\right) \frac{\|y_i - x_i\|_2^2}{2\sigma^2}\right] \frac{\|y_i - x_i\|_2^2}{2\sigma^2}}{N}, \end{aligned} \quad (16)$$

where (x_i, y_i) is sampled from the target distribution $\pi(Y, X)$.

Theorem 4. *By setting $\frac{\partial E_c[\hat{p}_e^2]}{\partial c} = 0$, the stationary point \hat{c} satisfies*

$$\hat{c} = \frac{2 \sum_{i=1}^N I(y_i)^2 \exp\left[\left(\frac{1}{c} - 1\right) \frac{\|y_i - x_i\|_2^2}{2\sigma^2}\right] \frac{\|y_i - x_i\|_2^2}{2\sigma^2}}{n \sum_{i=1}^N I(y_i)^2 \exp\left[\left(\frac{1}{c} - 1\right) \frac{\|y_i - x_i\|_2^2}{2\sigma^2}\right]}, \quad (17)$$

where (x_i, y_i) is sampled from $\pi(Y, X)$. Thus, as $N \rightarrow \infty$,

$$\begin{aligned} \hat{c} &\xrightarrow{a.s.} \frac{2E_{g(Y,X)} \left[I(Y)^2 \frac{\|Y-X\|_2^2}{2\sigma^2} \frac{\pi^2(Y,X)}{g^2(Y,X)} \right]}{nE_{g(Y,X)} \left[I(Y)^2 \frac{\pi^2(Y,X)}{g^2(Y,X)} \right]} \\ &\approx \frac{2 \sum_{i=1}^N I(y_i)^2 \exp\left[\left(\frac{1}{c} - 1\right) \frac{\|y_i - x_i\|_2^2}{\sigma^2}\right] \frac{\|y_i - x_i\|_2^2}{2\sigma^2}}{n \sum_{i=1}^N I(y_i)^2 \exp\left[\left(\frac{1}{c} - 1\right) \frac{\|y_i - x_i\|_2^2}{\sigma^2}\right]}, \end{aligned} \quad (18)$$

where $g(Y|X)$ is the pdf of $N(x, c\sigma^2 I_n)$ for any c and (x_i, y_i) is sampled from $g(Y|X)\pi(X)$.

Remark. *Theorem 4 is slightly weaker than Theorem 2 because $\frac{\partial^2 E_c[\hat{p}_c^2]}{\partial c^2} > 0$ can not be guaranteed. Hence, \hat{c} in equation (18) is only the stationary point. In each update, we let $c^* = \max\{1 + \delta, \hat{c}\}$, where $\delta > 0$. This guarantees that the proposal distribution always has a larger variance than the target distribution.*

C. Adaptive Importance Sampling

The adaptive importance sampling [15] is an iterative process to generate the proposal densities to reduce the variance of importance sampling. The procedure consists of three basic steps: generating samples from proposals, calculating the importance weight of each sample (weighting), and updating the parameters in the distribution family of the proposal to obtain new proposals for the next iteration (adapting). The way to make adaption is tailored to the specific problem as mentioned. Many methods have been proposed [16]–[18] for combining the samples from different proposals. Here, we choose a straightforward way, which is to normalize the weighted cross of all samples from all proposals. Algorithm 1 below implements our adaptive importance sampling strategy.

Algorithm 1 Adaptive Importance Sampling

Input Number of proposals J , Number of samples from each proposal N

Initialization: η_1

for $j = 1 \cdots J$ **do**

Sampling: Draw $\{x_{ij}, y_{ij}\}_{i=1}^N$ from $g_{\eta_j}(Y|X)\pi(X)$

Weighting: Calculate the weight $\{w_{ij}\}_{i=1}^N$

Adaptation: Updating η_j based on equation (13) or (18)

end for

Reweighting: $\bar{w}_{ij} = \frac{w_{ij}}{\sum_{j=1}^J \sum_{i=1}^N w_{ij}}$

Output: $\sum_{j=1}^J \sum_{i=1}^N \bar{w}_{ij} I(y_i)$

V. QUASI-MONTE CARLO METHOD

The Monte Carlo method is one of the most widely used numerical methods for simulating probability distributions. It is a decisive step in overcoming the curse of dimensionality problem as computing operations do not grow exponentially but remains manageable and proportional to the dataset's size. A major drawback, however, is that Monte Carlo methods produce probabilistic predictions rather than definitive answers. Another critical point in applying Monte Carlo methods is the generation of random samples. Although Monte Carlo commonly assumes that true random numbers are used as inputs, in practice, they are, in fact, deterministic. These inputs are called pseudo-random numbers and are designed to mimic true random numbers. Quasi-Monte Carlo methods are a deterministic counterpart to the Monte Carlo methods, which rely on using low-discrepancy sequences with more evenly space-filling properties rather than random sequences [10]. As a result, there is a significant improvement when

Monte Carlo	Quasi-Monte Carlo
random	deterministic
$N^{-1/2}$ convergence	close to N^{-1} convergence
Probabilistic error bounds	deterministic error bounds

using quasi-Monte Carlo for high-dimensional integration to achieve a much faster explicit convergence rate. Furthermore, when a similar problem needs to be solved repeatedly, quasi-Monte Carlo is more reliable, as is often the case in pricing applications, where quasi-Monte Carlo methods have gained tremendous attention [19]–[21].

Both MC and QMC take the same format when approximating an integral, i.e., $\int_{[0,1]^s} f(\mathbf{x}) d\mathbf{x} \approx \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i)$, where $\mathbf{x}_i \sim U([0,1]^s)$ are independent random samples for MC methods. In contrast, QMC methods utilize a deterministic low-discrepancy sequence in place of these random samples. The comparison between the two methods is summarized in the table above for quick reference.

QMC methods can integrate continuous functions in a hypercube of d -dimensional Euclidean space using equal weights with convergence rate $\mathcal{O}(N^{-1}(\log N)^{d-1})$, that is, for a continuous function f on $[0,1]^d$,

$$\left| \frac{1}{N} \sum_{i=1}^N f(x_i) - \int_{[0,1]^d} f(x) dx \right| \leq C_f \frac{(\log N)^{d-1}}{N}, \quad (19)$$

where $x_{i=1}^N$ are points from a low-discrepancy sequence and C_f is a constant that depends only on the function f . Roughly speaking, low-discrepancy sequences are designed in such a way that there are no large empty regions or clustered points. The sequence systematically fills gaps in any initial segment, unlike random points. This leads to improved performance compared to traditional Monte Carlo methods.

Recall that codeword x is sent over the AWGN channel, then the conditional target distribution $y|x \sim N(x, \sigma^2 I_n)$, where $y_i \sim \pi(Y)$. Moving to the QMC way, we will use the low discrepancy point in the sampling procedure for the Gaussian distribution. Low-discrepancy sequences are a type of sequence widely used in numerical computation and statistical analysis. They include classical constructions such as Halton-type sequences, Sobol points, digital nets [22], and lattice rules [23] along with their variations. The seminal work on low-discrepancy sequences can be found in [24]. When randomization techniques such as random shift or scrambling are applied to these sequences, they become randomized Quasi-Monte Carlo (RQMC) sequences. RQMC estimators are unbiased and can lead to a reduction in variance compared to ordinary Monte Carlo methods. Furthermore, RQMC provides a probabilistic framework for analyzing QMC-related methods, as discussed in [12] and references therein. The difference between MC and some QMC points is shown in Figure 2. It is evident that the QMC points are more uniformly distributed within the unit square, as compared to MC points.

Low-discrepancy sequences have also been integrated into Markov Chain Monte Carlo (MCMC) methods, as described

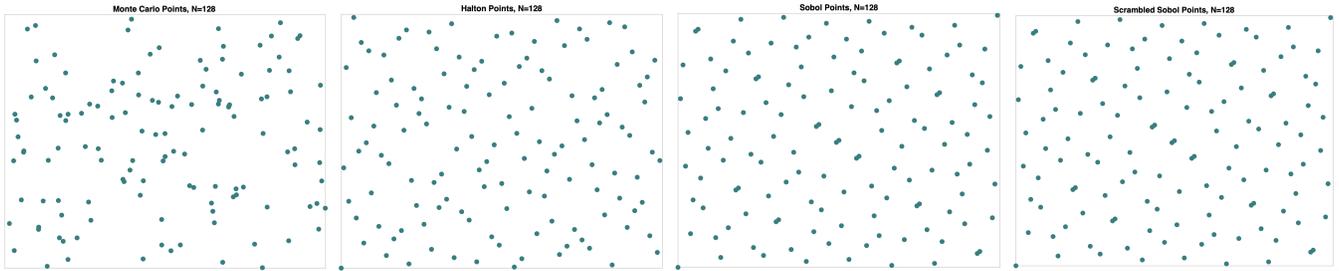


Fig. 2. From left to right: MC Points, and QMC Points including Halton, Sobol, and Scrambled Sobol. In each case, the number of points is $N = 128$.

in [25], [26]. The deterministic versions of statistical samplers, such as acceptance-rejection samplers [27] and importance sampling [28], have been studied, and it has been observed that the use of QMC points can lead to improved performance in various applications.

A. Using QMC in Importance Sampling

In the importance sampling framework, there are at least two ways to improve the algorithm’s performance: by designing a more effective proposal distribution and by obtaining better samples from the proposal. This section focuses on the latter aspect, specifically, obtaining better samples from the chosen proposal distribution. To achieve this, the quasi-Monte Carlo method will be used. As a result, we will obtain low-discrepancy samples that are specifically designed for the chosen proposal distribution and have better space-filling properties compared to those obtained using the Monte Carlo method.

The use of QMC points in the importance sampling framework has been explored in previous studies, such as [28], where a significant improvement in numerical integration tasks was observed. More recent research, such as [29, Theorem 3.1], has shown that using randomized QMC in the importance sampling framework for a Gaussian proposal can lead to an RMSE of the order $N^{-1+\epsilon}$ for arbitrarily small $\epsilon > 0$. In the importance sampling framework, QMC points can be used to sample from the proposal distribution, replacing Monte Carlo points. The higher integration accuracy of QMC methods compared to MC methods means that fewer sampling points are needed to estimate the bit error. However, different types of QMC points may result in slightly different performances. In our experiment, we will consider several commonly used QMC point sets, and further research could be done to construct the optimal QMC/RQMC points for the importance sampling algorithm. This is left for future work.

VI. EXPERIMENTS

In this section, we demonstrate how different sampling methods perform estimating the bit error in a 6G wireless communication scenario when the number of sampling is large. We consider only the AWGN channel with the proposed methods. We first randomly generate some code words x and modulate them to latent space. Then we sample y from $g_\eta(y | x)$ with parameter η . This is done by adding Gaussian noise based on the proposal distribution on modulated code

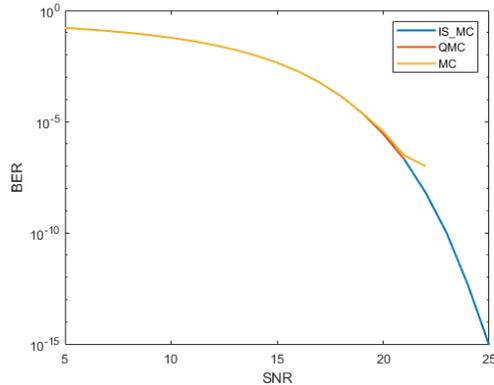


Fig. 3. SNR vs. BER for three sampling methods: Monte Carlo, Quasi-Monte Carlo and MC-based Importance Sampling.

words. We repeat this process several times. Each time we updated the parameter η based on equation (18). After we finish sampling, we reweight all the important weights of the samples. Next, we demodulate the noisy signal back and compute the bit error rate for each sampler. The importance sampling estimator is obtained by calculating the weighted sum of the bit error rate. See Algorithm 1. We evaluate the BER by repeating the simulation multiple times. We compare three methods in the experiment: vanilla Monte Carlo, Quasi-Monte Carlo and importance sampling based on scaling variance. All experiments are simulated in Matlab.

A. MC and QMC points

In the experiments, we use both Monte Carlo and Quasi-Monte Carlo sampling points. Figure 2 illustrates the MC points by pseudo-random algorithm and QMC points by Sobol, Halton, and scrambled methods from left to right. All the methods sample $n = 128$ points in unit square $[0, 1]^2$. The MC points in the leftmost are scattered randomly in $[0, 1]^2$. The QMC points are more regularly aligned in the unit square than MC points but with higher integration accuracy with the same number of sampling points. Sobol and Halton are typical QMC points, while scrambled Sobol points are randomized QMC methods. Here the “randomized” means that each scrambled point is a Sobol point with a small perturbation. The scrambled points have the integration accuracy between MC and QMC methods as mentioned.

B. Importance Sampling by Monte Carlo Method

We present the simulation results of biasing the noise pdf using variance scaling, as well as replacing the Monte Carlo method with the Quasi-Monte Carlo method.

We evaluate the BER for sampling methods for signal-to-noise ratio (SNR, db) from 5 to 25, where the bit error rate is calculated by $BER = \frac{\sum_{i=1}^N |m_i - m'_i|}{N}$. We assume that the input signal x has 5000 codewords and 100 packets, noise ϵ is Gaussian distributed with mean zero and variance $0.5 \times 10^{-\frac{SNR}{10}}$. Three methods are evaluated in the experiments: the plain Monte Carlo (MC) Method, Quasi-Monte Carlo (QMC) method, and Importance Sampling with MC (IS-MC). In IS-MC, the proposal SNR is the original SNR reduced by 5 using Algorithm 1.

We show the SNR vs. BER in the semilog plot in Figure 3, where the x-axis is SNR from 0 up to 25, and the y-axis is the BER. In the simulation, we use $N_s = 100$ packs and $N = 5000$ packets and qammod order is 16. The number of sampling points is then 500,000. We observe that the importance sampling method achieves great improvement on fixed sampling numbers. When SNR is less than 20, those three methods show similar performance. However, neither the Monte Carlo method nor the Quasi-Monte Carlo method can estimate BER when SNR is larger than 20. The importance sampling method can precisely estimate BER roughly equal to 10^{-15} by using 500,000 samplings.

C. Comparison of MC and QMC Methods

We evaluate MC and QMC methods by comparing each of them with importance sampling. As the IS method shows a small variance for each SNR up to 25, we treat the IS result as the ground truth and compute the relative error for the plain MC and the QMC and RQMC methods. The QMC (or RQMC) method provides (nearly) equal-area points which can be used to sample Gaussian distributed points by dividing the norm of each point. Figure 4 shows the difference of the bit error rates between each of MC sampling, Sobol-based QMC sampling, and scrambled-Sobol-based RQMC sampling with IS-MC sampling. Both MC and QMC methods have a tiny error when SNR is less than 15. When SNR is large and the probability of error is small, the variance of both of the two methods increases. When the number of sampling points is limited (and then SNR is high), the QMC shows a smaller variance than the plain MC method. The RQMC method has a similar performance as the QMC.

VII. CONCLUSION AND DISCUSSION

We propose a new and efficient sampling method for evaluating wireless communication that requires high reliability, specifically for 6G systems. Our approach is based on an adaptive importance sampling algorithm with an updating rule that has a theoretically guaranteed minimal variance for the simplified AWGN case. Our results show that the proposed importance sampling approach is significantly more efficient than traditional Monte Carlo methods, achieving a BER of 10^{-15} with 500,000 samplings while Monte Carlo fails at

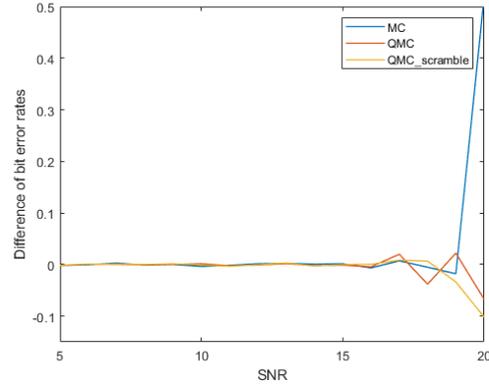


Fig. 4. Comparison of MC with QMC (left), and MC with RQMC (right). The y-axis is the difference of bit error rates between MC, QMC or RQMC with IS-MC. The QMC sampling uses Sobol points and the RQMC is based on scrambled Sobol points.

10^{-6} . Quasi-Monte Carlo and randomized Quasi-Monte Carlo methods, which have demonstrated higher accuracy in high-dimensional integration, are also potential replacements for Monte Carlo in our importance sampling framework.

While our main focus is to present a novel concept and framework for simulating rare events in 6G Ultra-Reliable Low-Latency Communications systems, further empirical evaluations comparing the performance and efficiency of our proposed approach to existing methods will be conducted in future work. Moreover, we believe that our framework can be generalized to more complicated scenarios, such as Rayleigh and more general fading channels.

REFERENCES

- [1] W. Tong and P. Zhu, *6G: The next horizon: From connected people and things to connected intelligence*. Cambridge University Press, 2021.
- [2] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communications systems," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 4, pp. 597–613, 1997.
- [3] G. Romano and D. Ciuonzo, "Minimum-variance importance-sampling Bernoulli estimator for fast simulation of linear block codes over binary symmetric channels," *IEEE transactions on Wireless Communications*, vol. 13, no. 1, pp. 486–496, 2013.
- [4] A. Mahadevan and J. M. Morris, "SNR-invariant importance sampling for hard-decision decoding performance of linear block codes," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 100–111, 2007.
- [5] J. Font-Segura, A. Martinez, and A. G. i Fàbregas, "Importance sampling for coded-modulation error probability estimation," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 289–300, 2019.
- [6] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.
- [7] E. Cavus, C. L. Haymes, and B. Daneshrad, "Low BER performance estimation of LDPC codes via application of importance sampling to trapping sets," *IEEE Transactions on Communications*, vol. 57, no. 7, pp. 1886–1888, 2009.
- [8] S.-K. Ahn, K. Yang, and D. Har, "Evaluation of the low error-rate performance of LDPC codes over Rayleigh fading channels using importance sampling," *IEEE Transactions on Communications*, vol. 61, no. 6, pp. 2166–2177, 2013.
- [9] 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 Processing Magazine*, vol. 34, no. 4, pp. 60–79, 2017.

- [10] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, vol. 63.
- [11] G. Leobacher and F. Pillichshammer, *Introduction to quasi-Monte Carlo integration and applications*. Birkhäuser/Springer, Cham, 2014.
- [12] C. Lemieux, *Monte Carlo and quasi-Monte Carlo sampling*. Springer, New York, 2009.
- [13] J. A. Bucklew, "Conditional importance sampling estimators," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 143–153, 2005.
- [14] M. Rowland, A. Harutyunyan, H. Hasselt, D. Borsa, T. Schaul, R. Munos, and W. Dabney, "Conditional importance sampling for off-policy learning," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 45–55.
- [15] M. F. Bugallo, L. Martino, and J. Corander, "Adaptive importance sampling in signal processing," *Digital Signal Processing*, vol. 47, pp. 36–49, 2015.
- [16] O. Cappé, A. Guillin, J.-M. Marin, and C. P. Robert, "Population Monte Carlo," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.
- [17] J.-M. Cornuet, J.-M. Martin, A. Mira, and C. P. Robert, "Adaptive multiple importance sampling," *Scandinavian Journal of Statistics*, vol. 39, no. 4, pp. 798–812, 2012.
- [18] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from uncertainty," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4422–4437, 2015.
- [19] J. Dick, F. Y. Kuo, and I. H. Sloan, "High-dimensional integration: the quasi-Monte Carlo way," *Acta Numerica*, vol. 22, pp. 133–288, 2013.
- [20] J. Xiang and X. Wang, "Quasi-Monte Carlo simulation for American option sensitivities," *Journal of Computational and Applied Mathematics*, vol. 413, p. 114268, 2022.
- [21] C. Zhang, X. Wang, and Z. He, "Efficient importance sampling in quasi-Monte Carlo methods for computational finance," *SIAM Journal on Scientific Computing*, vol. 43, no. 1, pp. B1–B29, 2021.
- [22] J. Dick and F. Pillichshammer, *Digital nets and sequences*. Cambridge University Press, Cambridge, 2010.
- [23] I. H. Sloan and S. Joe, *Lattice methods for multiple integration*. Oxford University Press, New York, 1994.
- [24] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*. Wiley-Interscience [John Wiley & Sons], New York-London-Sydney, 1974.
- [25] S. Chen, J. Dick, and A. B. Owen, "Consistency of Markov chain quasi-Monte Carlo on continuous state spaces," *The Annals of Statistics*, vol. 39, no. 2, pp. 673–701, 2011.
- [26] J. Dick, D. Rudolf, and H. Zhu, "Discrepancy bounds for uniformly ergodic Markov chain quasi-Monte Carlo," *The Annals of Applied Probability*, vol. 26, no. 5, pp. 3178–3205, 2016.
- [27] H. Zhu and J. Dick, "A discrepancy bound for deterministic acceptance-rejection samplers beyond $N^{-1/2}$ in dimension 1," *Statistics and Computing*, vol. 27, no. 4, pp. 901–911, 2017.
- [28] J. Dick, D. Rudolf, and H. Zhu, "A weighted discrepancy bound of quasi-Monte Carlo importance sampling," *Statistics & Probability Letters*, vol. 149, pp. 100–106, 2019.
- [29] Z. He, Z. Zheng, and X. Wang, "On the error rate of importance sampling with randomized quasi-Monte Carlo," *Arxiv: 2203.03220*, 2022.