On Accuracy and Efficiency of Monte-Carlo BER Simulations for Fading Channels

Sergey Loyka*
School of Information Technology and Engineering,
University of Ottawa,
161 Louis Pasteur, Ottawa,
Ontario, Canada, K1N 6N5
E-mail: sergey.loyka@ieee.org

Francois Gagnon
Department of Electrical Engineering
Ecole de Technologie Superieure
1100, Notre-Dame St. West, Montreal
Quebec, H3C 1K3, Canada
E-mail: francois.gagnon@etsmtl.ca

***Abstract*— Accuracy and efficiency of Monte-Carlo (MC) bit error rate (BER) simulation for a fading channel is discussed. While the simulation error is minimized for given total number of trials by using one symbol/noise realization per channel realization ($M=1$), it does not result in the best efficiency. The latter can be further improved by vectorizing the "noise part" of the simulation and using $M>1$. Extensive numerical simulations validate these conclusions. Efficient numerical procedure of finding optimum *M* is suggested. The proposed technique preserves the problem-independence of the original MC method, and can also be combined with any known importance sampling method for better simulation efficiency.**

***Index Terms*— Monte-Carlo simulation**

I. Introduction

Numerical Monte-Carlo (MC) simulation is a popular technique to estimate bit error rate (BER) of a digital communication system and various signal processing algorithms, which is especially valuable when the system/algorithm is complex enough so that analytical analysis is not feasible or too complex [1-3]. There are many improved versions of the basic MC method, including various variance reduction techniques (e.g. importance sampling) [6-10]. However, while increasing the simulation efficiency, all these techniques have a common drawback: the generality appeal of the original MC method (problem-independence) is lost as all improved techniques turn out to be problem-specific [2,8].

Despite of the fact that the MC method is widely used for the error rate simulation of digital systems, accuracy of such simulations is not widely discussed in the literature. While the accuracy of MC BER simulation for a fixed AWGN channel is well understood [1], the BER simulation accuracy and its efficiency for fading channels is not studied well.

In this paper, we address the problem by evaluating analytically the MC BER simulation accuracy (in terms of the estimation variance and the confidence interval/probability) in a fading channel. Based on this study, an efficient way to carry out such a simulation is suggested, which can also be combined with any known importance sampling technique [6-10] for better simulation efficiency. Furthermore, contrary to importance sampling techniques, our modification of the original MC method preserves its problem-independence.

II. MC BER Simulation in an AWGN Channel

In this section, we review well-known MC simulation method and its accuracy when applied to a fixed (non-fading) AWGN channel [1]. The standard baseband system model in this case is,

$$r = a \cdot s + \xi \tag{1}$$

where r, s are the received and transmitted signals respectively, ξ is the additive noise, a is the constant channel gain. The receiver estimates the Tx symbols based on r ,

$$\hat{s} = Q(r) \tag{2}$$

according to its decision function *Q*. Introducing the error function,

$$\varepsilon(s,\xi) = \begin{cases} 1, & \hat{s} \neq s \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

BER can be expressed as

$$P_e = E_{\xi,s}\left[\varepsilon(s,\xi)\right] = \bar{\varepsilon}, \tag{4}$$

where $E_{s,\xi}[\ ]$ is the expectation over the noise ξ and the Tx symbol s . Using multiple realizations of the Tx symbol and noise, the MC method estimates BER using ensemble average instead of the true expectation,

$$\hat{P}_e = \frac{1}{N}\sum_{i=1}^{N} \varepsilon(s_i,\xi_i) = \hat{\varepsilon}, \tag{5}$$

where *N* is the number of realizations. The estimation error is

$$\Delta = \hat{\varepsilon} - \bar{\varepsilon} = \frac{1}{N}\sum_{i=1}^{N} (\varepsilon_i - \bar{\varepsilon}), \tag{6}$$

where $\varepsilon_i = \varepsilon(s_i,\xi_i)$. Its variance is the measure of the estimation accuracy [1,2],

$$\sigma_\Delta^2 = \frac{P_e(1-P_e)}{N}, \tag{7}$$

Specifically, the normalized estimation error is,

$$\sigma_n = \frac{\sigma_\Delta}{P_e} = \sqrt{\frac{1-P_e}{P_e N}} \approx \frac{1}{\sqrt{N_e}}, \quad (8)$$

where N_e is the expected number of errors in the ensemble. The last equality holds true for small BER, $P_e << 1$. Clearly, small P_e requires for large N if the estimation error is to be small,

$$N \approx \frac{1}{\sigma_n^2 P_e}, \quad (9)$$

For example, 10% error results in $N \approx 100/P_e$. Note also that σ_n can be used to specify confidence intervals for given confidence probability [1-3,10].

It should be pointed out that the analysis above presumes that the errors are independent, i.e. the system is memoryless. If this is not the case, the estimation error will increase [1].

Various variance reduction techniques can be used to decrease the error without increasing the number of MC trials [1-3,6-10]. Their implementation is, however, more complex and, often, problem-dependant too. Hence, a preliminary analytical or numerical analysis is required. Simplicity and problem-independence appeal of the original MC method is lost in this case.

III. MC BER Simulation in a Fading Channel

In this section, we give an explicit evaluation of the MC BER estimation error in a fading channel, which will be further used to optimize the simulation method. In a fading channel, there are two BERs : the instantaneous (i.e., for a given fading channel gain *a*) and the average. The instantaneous BER is still given by (4), but it is now a function of *a*, $P_e(a) = E_{\xi,s}\left[\varepsilon(s,\xi,a)\right] = \varepsilon(a)$. The average BER (ABER) is

$$\overline{P}_e = E_a\left[\varepsilon(a)\right] = \bar{\varepsilon}, \quad (10)$$

Using the analogy with double integration (expectation) required to find the ABER analytically (as indicated above), it is tempting to use the same technique for the MC method, i.e. to estimate the instantaneous BER for each channel realization first and then to average it over the channel realizations. As we demonstrate below, this is not the best approach in terms of accuracy or number of trials required (simulation time). Specifically, the MC ABER estimation is

$$\hat{P}_e = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}\varepsilon_{ij} = \hat{\varepsilon}, \quad (11)$$

where $\varepsilon_{ij} = \varepsilon(s_{ij}, \xi_{ij}, a_i)$, $M^{-1}\sum_{j=1}^{M}\varepsilon_{ij}$ is the instantaneous BER estimation; *M* noise and Tx symbol realizations are used to estimate the instantaneous BER and *N* channel realizations are used to estimate the ABER, each channel realization has its own set of Tx symbols and noise. The estimation error is

$$\Delta = \hat{\varepsilon} - \bar{\varepsilon} = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}(\varepsilon_{ij} - \bar{\varepsilon}) = \Delta_1 + \Delta_2, \quad (12)$$

where

$$\Delta_1 = \frac{1}{N}\sum_{i=1}^{N}(\varepsilon_i - \bar{\varepsilon}), \quad \Delta_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M}(\varepsilon_{ij} - \varepsilon_i) \quad (13)$$

and $\varepsilon_i = E_{s,\xi}(\varepsilon_{ij}) = P_e(a_i)$. In fact, Δ_1 is the error in estimating the ABER when the instantaneous BER is known exactly (for example, if it is found analytically), and Δ_2 is the average error in estimating the instantaneous BER. It is straightforward to show that

$$E\left[\Delta_1\right] = E\left[\Delta_2\right] = 0, \quad E\left[\Delta_1\Delta_2\right] = 0 \quad (14)$$

where, if not otherwise indicated, the expectation is taken over all random variables. Hence, the variance of the estimation is

$$\sigma_\Delta^2 = \sigma_1^2 + \sigma_2^2, \quad (15)$$

where, after some manipulations (see Appendix for details),

$$\sigma_1^2 = \frac{E\left[P_e^2(a)\right] - \overline{P}_e^{\,2}}{N}, \quad \sigma_2^2 = \frac{\overline{P}_e - E\left[P_e^2(a)\right]}{NM} \quad (16)$$

For given total number of MC trials $N_t = NM$, the variance is clearly minimized when $N = N_t$ and, hence, $M = 1$: the best strategy (in terms of minimum simulation error) is to use 1 noise/symbol realization per channel realization, and not to estimate instantaneous BER first. The normalized error for this strategy is

$$\sigma_n = \sqrt{\frac{1-\overline{P}_e}{\overline{P}_e N}} \approx \frac{1}{\sqrt{N_e}}, \quad (17)$$

Remarkably, this is the same as (8) with the ABER used instead of BER in (8). The ABER estimation accuracy is still determined by the expected number of errors, as one would expect. The required number of trials for given accuracy can be determined using (9), where $\overline{P}_e$ should be used instead of P_e. It should be noted that the analysis above is valid for any type of noise, fading and modulation/coding.

The effect of *M* for fixed $N_t = NM$ can be better seen by expressing (15)-(16) in the following form:

$$\sigma_\Delta^2 = \frac{\overline{P}_e}{N_t}\left(1 + (M-1)\frac{E\left[P_e^2(a)\right]}{\overline{P}_e} - M\overline{P}_e\right) \geq \frac{\overline{P}_e}{N_t}\left(1-\overline{P}_e\right), \quad (18)$$

since $E\left[P_e^2(a)\right] \geq \overline{P}_e^{\,2}$. Note that for coherently-demodulated BPSK in a frequency-flat Rayleigh fading channel,

$$E\left[P_e^2(a)\right] = \frac{1}{4} - \frac{1}{\pi}\sqrt{\frac{\gamma_0}{\gamma_0+1}}\tan^{-1}\left(\sqrt{\frac{\gamma_0+1}{\gamma_0}}\right) \approx \frac{1}{8\gamma_0}\left(1-\frac{2}{\pi}\right) \quad (19)$$

where the approximate equality holds true for large average SNR (low BER), $\gamma_0 >> 1$ (in this case, large N_t is required since $\overline{P}_e$ is low, and, hence, simulation efficiency is of importance). The ABER is $\overline{P}_e \approx 1/(4\gamma_0)$. Hence, the normalized error is

$$\sigma_n \approx \sqrt{4\gamma_0 b_M / N_t}, \quad (20)$$

where $b_M = 1 + (M-1)(1/2 - 1/\pi)$. For given σ_n, N_t can be estimated as follows:

$$N_t \approx 4\gamma_0 b_M / \sigma_n^2, \quad (21)$$

Since, for most modulation types, the ABER in a Rayleigh fading channel can be expressed as $\overline{P_e} \approx c/\gamma_0$, where the constant c depends on the modulation, the results similar to (20)-(21) hold true for other modulation types as well. When n-th order diversity combining is used, the ABER becomes $\overline{P_e} \approx c_n / \gamma_0^n$, where c_n depends on the diversity combining and modulation type, and, similarly to (21), the required number of samples can be estimated as

$$N_t \approx \frac{b'_M \gamma_0^n}{c_n \sigma_n^2}, \quad (22)$$

where b'_M depends on M and the modulation/combining type. (20)-(22) provide a guideline for controlled-accuracy simulations in a fading channel.

When averaging over the noise takes approximately the same time as averaging over the channel, setting $M = 1$ significantly reduces the simulation time for a given simulation error. For example, consider BPSK in a Rayleigh fading channel at $\gamma_0 = 30dB$ and $\sigma_n = 0.1$, i.e. 10% accuracy. Setting $M = 1$ results in $N_t \approx 4 \cdot 10^5$ while setting $M = N = \sqrt{N_t}$ (i.e. estimating BER for a given channel and then averaging over channel realizations) results in $N_t \approx 5 \cdot 10^9$, i.e. four orders of magnitude difference in simulation time.

Clearly, setting $M = 1$ minimizes the simulation error for given N_t (the minimum error is given by (17)). Does it however mean that one should always use $M = 1$ to perform an "optimum" MC simulation? A detailed analysis below demonstrates that it is not so.

IV. Optimizing the Simulation Method

An optimum simulation method should minimize the simulation error for given simulation time (or vice versa). Hence, a simulation efficiency measure can be defined as follows [2,3]:

$$F = \sigma^2 T, \quad (23)$$

where σ^2 is the simulated estimation variance, and T is the simulation time. Clearly, the simulation method with the least F is the best one, i.e. it provides the least variance for given time (or vice versa). When the simulation time is linear in the number of trials, $T = cN$ (which is an accurate approximation for most simulation methods when N is large), F is roughly independent of N, $F = cP_e(1 - P_e)$. Hence, this definition of F is roughly independent of N when the variance can be expressed as in (7) or (17) (i.e. simple increase in N does not affect the efficiency). This, however, assumes that $M = 1$. Hence, one may improve the efficiency by finding an optimum M. This optimization is possible because the BER simulation for given channel (i.e. averaging over the noise) can be easily vectorized and, hence, this part of the simulation runs much faster (for example, using popular software tool Matlab [4], the vectorization results in order of magnitude reduction in the simulation time). Hence, using M moderately larger than 1 will not increase significantly the simulation time. On the contrary, vectorization of the "fading part" of the simulation (i.e. averaging over the channel realizations) does not bring significant advantage. This is so because of mainly two reasons: (i) some complex processing algorithms (e.g. V-BLAST [5]) do not allow reasonably simple vectorization of the "fading part", (ii) the second vectorization within the method produces matrices of large size, which has negative impact on the efficiency.

An analytical solution for the optimum M is hardly possible as it depends on a structure of the software code and simulation package used as well as on a computing hardware platform. This, in turn, would require building an analytical model for F incorporating all these particularities, which, if successful, would result in a case-specific solution anyway, with unavoidable loss of generality. An efficient numerical procedure to find an optimum M and run an efficient simulation in a general case is suggested instead:

1. A complete simulation is run for different M with only moderately large N_t (say, $10^3 \ldots 10^4$), which requires short time only (from a fraction of a second to few minutes on a modern PC), to generate a curve similar to that on Fig. 1.
2. Optimum M is found numerically based on the curve generated.
3. A full-scale simulation is run with optimum M and as large N_t as required for desired accuracy at a given BER level. For example, an accurate estimation of BER on order of 10^{-6} requires approximately $N_t \approx 10^8$, which results in simulation time on the order of few hours to few days.

Note that the function of step 1 and 2 is to find optimum M only and, hence, poor accuracy of BER estimation (due to not sufficiently large N_t) at these steps is not an issue. Step 3 relies on the fact that optimum M roughly does not depend on N_t (unless $N = N_t / M$ is too small; that is why $N_t = 10^3 \ldots 10^4$ is required at step 1).

V. An Example

Extensive numerical simulations have been carried out to validate the points above. Some of the results are presented below. As an example, we consider coherently-demodulated BPSK operating on a frequency-flat Rayleigh fading channel. Fig. 1 shows the normalized MC BER estimation error (estimated by running the simulation many times and finding empirical variance of the estimated BER) and the analytical result (17). Clearly, there is no significant difference between the two. We also note that using $M \neq 1$ for given N_t results in increased error. The optimum M is about 10, and the optimum simulation runs about 3 times faster than that for $M = 1$. Since F varies slowly about the optimum value, slightly different M will not result in significant deterioration. It should be noted that specific values (optimum M, performance improvement etc.) depend on the system model and

the method implementation. It is clear, however, that the optimum M is not necessarily 1. Similar results hold true when N is kept fixed and M increases, as it should be since simple increase in N_t roughly does not affect F. These observations are also true when a more complex system is simulated (e.g. V-BLAST [5]).

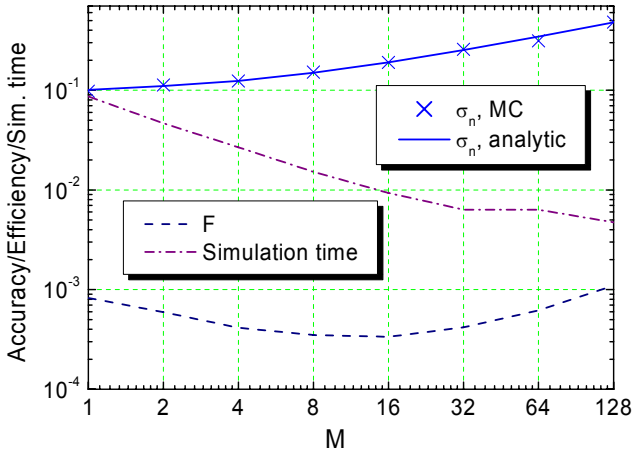

Fig 1. Accuracy, efficiency and simulation time versus M for coherently-demodulated BPSK in a Rayleigh fading channel; $N_t = 4096$, average $SNR = 10\text{dB}$.

VI. Appendix

(14) is obtained by inspection. (15) follows directly from (14), where σ_1^2 and σ_2^2 are the variance of Δ_1 and Δ_2 respectively. Using (12)-(14), one obtains:

$$\sigma_1^2 = E\left[\Delta_1^2\right] = E\left[\frac{1}{N^2}\sum_{i_1,i_2=1}^{N}(\varepsilon_{i_1}-\bar{\varepsilon})(\varepsilon_{i_2}-\bar{\varepsilon})\right], \quad \text{(A1)}$$
$$= \frac{1}{N^2}\sum_{i_1,i_2=1}^{N}\left(E\left[\varepsilon_{i_1}\varepsilon_{i_2}\right]-\bar{\varepsilon}^2\right)$$

Noting that,

$$E\left[\varepsilon_{i_1}\varepsilon_{i_2}\right] = \begin{cases} E\left[P_e^2(a)\right], & i_1 = i_2 \\ \left(\bar{P}_e\right)^2, & i_1 \neq i_2 \end{cases}, \quad \text{(A2)}$$

σ_1^2 is finally expressed as in (16). Let us now consider σ_2^2,

$$\sigma_2^2 = E\left[\Delta_2^2\right] = \frac{1}{N^2M^2}E\left[\sum_{i_1,j_1}\sum_{i_2,j_2}(\varepsilon_{i_1j_1}-\varepsilon_{i_1})(\varepsilon_{i_2j_2}-\varepsilon_{i_2})\right],$$
$$= \frac{1}{N^2M^2}\sum_{i_1,j_1}\sum_{i_2,j_2} z_{i_1j_1i_2j_2} \quad \text{(A3)}$$

where $z_{i_1j_1i_2j_2} = E\left[\varepsilon_{i_1j_1}\varepsilon_{i_2j_2}\right] - E\left[\varepsilon_{i_1}\varepsilon_{i_2}\right]$. We consider further 2 cases.

Case I, $i_1 = i_2$, $j_1 = j_2$:

$$z_{i_1j_1i_2j_2} = E\left[\varepsilon_{ij}^2\right] - E\left[\varepsilon_i^2\right] = E\left[\varepsilon_{ij}\right] - E\left[\varepsilon_i^2\right]$$
$$= \bar{P}_e - E_a\left[P_e^2(a)\right], \quad \text{(A4)}$$

where we used the fact that $\varepsilon_{ij}^2 = \varepsilon_{ij}$. Case II, $i_1 \neq i_2$ or $j_1 \neq j_2$:

$$z_{i_1j_1i_2j_2} = E\left[\varepsilon_{i_1j_1}\varepsilon_{i_2j_2}\right] - E\left[\varepsilon_{i_1}\varepsilon_{i_2}\right]$$
$$= E_a\left[E_{s,\xi}\left[\varepsilon_{i_1j_1}\varepsilon_{i_2j_2}\right]\right] - E\left[\varepsilon_{i_1}\varepsilon_{i_2}\right]$$
$$= E_a\left[E_{s,\xi}\left[\varepsilon_{i_1j_1}\right]E_{s,\xi}\left[\varepsilon_{i_2j_2}\right]\right] - E\left[\varepsilon_{i_1}\varepsilon_{i_2}\right], \quad \text{(A5)}$$
$$= E_a\left[\varepsilon_{i_1}\varepsilon_{i_2}\right] - E\left[\varepsilon_{i_1}\varepsilon_{i_2}\right] = 0$$

where the 3rd equality is due to the fact that the noise and symbol realizations are always independent in this case (recall that each channel realization is assumed to have its own set of noise and symbol realizations). Combining (A3)-(A5), one finally obtains (16).

References

[1] M.C. Jeruchim, P. Balaban, K.S. Shanmugan, Simulation of Communication Systems, Kluwer, New York, 2000.

[2] Y.A. Shreider et al, Method of Statistical Trials (Method Monte-Carlo), GIFML, Moscow, 1962.

[3] J.M. Hammersley, D.C. Handscomb, Monte Carlo Methods, Methuen & Co., London, 1967.

[4] D.J Higham, N.J. Higham, Matlab Guide, SIAM, Philadelphia, 2000.

[5] S. Loyka, F. Gagnon, On BER Analysis of the BLAST Without Optimal Ordering over Rayleigh Fading Channel, 2004 IEEE Vehicular Technology Conference, Los Angeles, CA, Sep. 26-29.

[6] J. Cavers, P. Ho, Reducing the Computation Time in Simulations of Fading Channels, 1992 IEEE Vehicular Technology Conference, Denver, CO, 10-13 May 1992, v. 1, pp. 359 – 363.

[7] R. Srinivasan, Importance Sampling, Springer, Berlin, 2002.

[8] P.J. Smith, M. Shafi, H. Gao, Quick Simulation: A Review of Importance Sampling Techniques in Communications, IEEE J. Selected Areas Comm., v. 15, N. 4, pp. 597-613, May 1997.

[9] N.C. Beaulieu, An Investigation of Gaussian Tail and Rayleigh Tail Density Functions for Importance Sampling Digital Communication System Simulation, IEEE Trans. Comm., v. 38, N. 9, pp. 1288-1292, Sep. 1990.

[10] M.C. Jeruchim, Techniques for Estimating the Bit Error Rate in the Simulation of Digital Communication Systems, IEEE J. Selec. Areas Comm., v. 2, N. 1, pp. 153-170, Jan. 1984.