

# Fast Simulation of Linear Communication Systems via Conditional Monte Carlo Analysis

Lachlan L. H. Andrew, Jonathan H. Manton and Duong H. Pham \*  
ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN)  
Department of Electrical and Electronic Engineering  
University of Melbourne, Vic 3010, Australia  
{L.Andrew, J.Manton, D.Pham}@ee.mu.oz.au

## Abstract

This paper presents a new technique for the fast simulation of the bit error rate and other statistical performance measures of communication systems. Whereas traditional fast simulation techniques are usually based on importance sampling, the proposed technique is based on conditional Monte Carlo analysis. One advantage over importance sampling is that the proposed technique is systematic in its design and is proved to be asymptotically optimal.

## 1. Introduction

Determining the statistical performance of a communication system is almost always done by simulation. However, naive simulation can be exceedingly slow, especially at high signal to noise ratios. The slowness is due to the fact that, if the true bit error rate is of the order of  $10^{-6}$ , then a million simulations are required on average before a bit error occurs, and thus hundreds of millions of simulations are required before the bit error rate can be calculated accurately. Even more simulations may be required if the distribution of the number of bit errors must be computed.

One technique for speeding up bit error rate calculations is importance sampling [1]. The rough idea behind importance sampling is to change the noise distribution so that bit errors are more likely to occur, and then having determined the statistical performance of this new system via simulation, to map the results back to the original system. Unfortunately, simply choosing a new noise distribution which increases the bit error rate is insufficient for importance sampling to be successful. Indeed, choosing a suitable noise distribution is an art and often it is not possible to prove that the resulting choice is optimal in any sense.

Section 2 of paper presents an alternative approach to the fast simulation problem. The idea is to use simulation to compute some components of a multi-dimensional random variable (noise), and then determine in closed form the expected value of

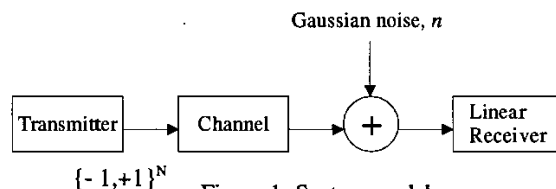


Figure 1: System model.

bit error rate conditional on those components. This technique is called conditional Monte Carlo analysis (see, for example, [2]). The main contribution of this paper is the derivation of a conditional Monte Carlo algorithm for linear communication systems; attention is restricted to linear systems to simplify presentation.

One advantage of this conditional Monte Carlo analysis approach is that the resulting algorithm is asymptotically optimal, as proved in Section 3. This means that as the bit error probability approaches zero, the number of samples required for a given estimation accuracy increases at a rate slower than any power law.

Numerical results in Section 4 demonstrate asymptotic optimality for more general channels than for which it has been formally proved. To date, the proof of optimality is only for the two dimensional linear case. However, the technique itself extends easily to multiple dimensions, and we believe that the optimality also holds in the more general case.

The system model considered in this paper is illustrated in Figure 1, and is essentially the same as that of [3, 4]. It consists of a block source, emitting blocks of length  $N$  symbols. In the numerical examples, bipolar binary symbols are used. These symbols pass through an  $L$ -tap channel, and Gaussian noise,  $n$ , of variance  $\sigma^2$  is added. The resulting signal is processed by a receiver, which may contain a linear equaliser, followed by nearest neighbour decoding.

## 2. Algorithm

It has long been known [5] that the efficiency of Monte Carlo simulation can be improved by making use of the relations

$$E_{X,Y}[Z] = E_X[E_Y[Z|X]]$$

\*This work was supported by the Australian Research Council

and

$$\begin{aligned}\text{Var}_{X,Y}[Z] &= \text{E}_X[\text{Var}_Y[Z|X]] + \text{Var}_X[\text{E}_Y[Z|X]] \\ &> \text{Var}_X[\text{E}_Y[Z|X]].\end{aligned}$$

If  $\text{E}_Y[Z|X]$  is known in closed form, then the variance of the Monte Carlo estimate can be reduced by generating  $X$  randomly, and taking the sample mean of  $\text{E}_Y[Z|X]$ , rather than generating both  $X$  and  $Y$  randomly, and taking the sample mean of  $Z$  itself. In the case of block error rate calculations,  $X$  will denote the direction of the noise vector experienced by a block, and  $Y$  will denote its magnitude, while  $Z$  is the block error rate.

Our proposed conditional Monte Carlo algorithm for estimating the probability of error,  $\varepsilon$ , when transmitting a block,  $B$ , of length  $N$ , over a linear channel, assuming additive white Gaussian noise, is the following.

1. Generate  $\hat{n}$  uniformly on the unit hypersphere.
2. Calculate the minimum positive scaling factor,  $x$ , such that the scaled noise  $x\hat{n}$  would cause  $B$  to be received in error.
3. The error probability estimate is then  $\hat{\varepsilon} = \text{Pr}_X(X > x)$ , where  $X^2 \sim \sigma^2 \chi_N^2$  is the square of the radial component of a noise sample.

The value of  $x$  chosen in step 2 depends on the specific form of the decoder. In additive white Gaussian noise (AWGN) with a known channel, the maximum likelihood estimator is a nearest neighbour detector, and the value of  $x$  is simply the distance to a hyperplane separating the codewords. In more complicated cases, such as when the channel must be estimated by a non-linear procedure, decision boundaries may have more complicated shapes.

This approach is complementary to that of [3]. Whereas there the radial component of the noise is simulated and an expectation is taken over the directional component, in this paper the directional component is simulated and the radial component is integrated explicitly. Although [3] averages in closed form over  $N - 1$  out of  $N$  dimensions, compared to the single radial component used here, it is the radial component which contains most of the variability of the bit error rate. It is this fact which allows the proposed algorithm to achieve asymptotic optimality without importance sampling. However, the performance of the algorithm can be further improved for large block sizes by employing importance sampling in the selection of  $\hat{n}$ .

Note also that, under the proposed algorithm, each random vector,  $\hat{n}$ , yields an entire curve of block error rate versus signal to noise ratio (SNR). This is similar to what is done in a different context in [6], and is in contrast to standard Monte Carlo techniques which require separate simulations to be run for each value of SNR. The benefit of this feature is greater for more complex systems, where the computational effort to find  $x$  can become large.

### 3. Optimality

An estimator,  $A(\sigma)$ , is said to be *asymptotically optimal* [7] for a rare event probability  $p(\sigma)$  if it is unbiased and

$$\frac{\log(\text{Var}[A(\sigma)])}{\log(\text{E}[A(\sigma)])} \rightarrow K \geq 2 \quad (1)$$

as  $\text{E}[A(\sigma)] = p(\sigma) \rightarrow 0$ . The practical upshot of this is that, as the bit error probability approaches zero, the number of samples required for a given statistical accuracy increases at most logarithmically.

**Theorem 1** Consider a function  $f : [0, 1] \times \mathbb{R}^+ \rightarrow \mathbb{R}$  given by  $f(u, \sigma) = \exp(\phi_1(u, \sigma)/\phi_2(\sigma))$ , where, for sufficiently small  $\sigma$ ,

$$\phi_1(u, \sigma) < 0 \quad (2a)$$

$$\phi_2(\sigma) > 0 \quad (2b)$$

$$\phi_1(u_2, \sigma) < \phi_1(u_1, \sigma) \text{ if } u_2 > u_1, \quad (2c)$$

$$\phi_2(\sigma) \rightarrow 0 \text{ as } \sigma \rightarrow 0 \quad (2d)$$

$$\phi_2(\sigma) \log(\phi_2(\sigma)) = o(\phi_1(0, \sigma)), \quad (2e)$$

and  $\phi_1$  satisfies the Lipschitz condition

$$\begin{aligned}(\exists U, \sigma_0, k > 0)(\forall u \in (0, U), \sigma \in (0, \sigma_0)) \\ |\phi_1(0, \sigma) - \phi_1(u, \sigma)| < ku. \quad (2f)\end{aligned}$$

If  $U \sim U(0, 1)$  is uniformly distributed on  $(0, 1)$ , then  $f(U, \sigma)$  is an asymptotically optimal estimator for  $\text{E}[f(U, \sigma)]$  as  $\sigma \rightarrow 0$ .

This theorem is proved in the appendix. To apply it to the above estimation algorithm, let  $F_x$  be the distribution of  $x$  (as a function of  $\hat{n}$ ). That is,  $F_x(x) < u$  with probability  $u$ . Then  $f(u, \sigma) = f(F_x(x), \sigma)$  is the estimated BER from a single noise sample. Let us first consider the case of  $N = 2$  dimensions with binary phase shift keying (BPSK) and a distortionless channel. In this case,  $f(\cdot, \cdot)$  can be found in closed form.

Without further loss of generality, consider the case when the transmitted signal is  $(-1, -1)$ . An error occurs in bit  $i$ ,  $i = 1, 2$ , if noise component  $n_i > 1$ . There are three cases to consider, depending on the octant in which the noise vector lies, as illustrated in Figure 2(a). In the cross-hatched region, the noise is in the same direction as the signal, and thus the probability of error is zero. In the white region, sufficiently large noise will cause an error in both bits, while in the remaining region, only single bit errors are possible.

The magnitude of the noise must exceed  $x_1 = \sec(\theta)$  to cause an error in bit 1, or  $x_2 = \sec(\theta - \pi/2)$  to cause an error in bit 2, as illustrated in Figure 2(b). If negative values of  $\sec(\cdot)$  are deemed infinite, then  $x = \min(x_1, x_2)$ . If  $\arg \hat{n} = \theta \sim U[-\pi, \pi]$ , then it is easily verified that

$$u(\theta) = \begin{cases} 2|\theta/\pi| & \theta \in [-\pi/4, \pi/4] \\ 2|\theta/\pi - 1/2| & \theta \in (\pi/4, 3\pi/4] \\ |\theta/\pi - 1/4| & \theta \in (-\pi/2, -\pi/4) \cup (3\pi/4, \pi) \\ |\theta/2\pi| + 1/2 & \theta \in [-\pi, \pi/2) \end{cases} \quad (3)$$

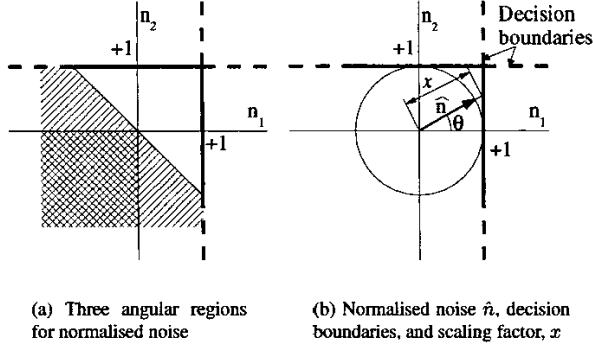


Figure 2: The amount,  $x$ , by which the normalised noise vector,  $\hat{n}$ , must be scaled to cross a decision boundary depends on which region it lies in.

is monotonic in  $x$ , and  $u \sim U[0, 1]$ . Explicitly,

$$1/x = \begin{cases} \cos(u\pi/2) & 0 \leq u \leq 1/2 \\ \cos(u\pi - \pi/4) & 1/2 < u < 3/4 \\ 0 & u \geq 3/4. \end{cases} \quad (4)$$

The cumulative distribution function (cdf) of a  $\chi^2$  random variable with two degrees of freedom is simply  $\Pr(Y < y) = 1 - e^{-y^2/2}$ . Thus

$$f(u, \sigma) = \Pr(X^2 > x/\sigma) = e^{-x^2/2\sigma^2} \quad (5)$$

yielding  $\phi_2(\sigma) = \sigma^2$  and

$$\begin{aligned} \phi_1(u, \sigma) &= -x^2/2 \\ &= \begin{cases} -\sec^2(u\pi/2)/2 & 0 \leq u \leq 1/2 \\ -\sec^2(u\pi - \pi/4)/2 & 1/2 < u < 3/4 \\ -\infty & 3/4 \leq u \leq 1. \end{cases} \end{aligned} \quad (7)$$

These functions satisfy condition (2). To see that (2f) is satisfied, note that the inequality  $\sec^2(\theta) \leq (1 - \theta^2)^{-1}$  for  $\theta < 1$ , leads to the sufficient conditions  $U < 1/2$  and  $k > \pi^2 U/4$ .

The key aspect of the geometry of this system is that the decision boundary is tangential to the sphere corresponding to equiprobable noise. Thus, the BER decreases slowly as the direction of the noise deviates from the direction of maximum noise. This property also holds in higher dimensions, and with distorting channels, although the exact form of  $f(\cdot, \cdot)$  becomes more complicated.

For general  $N$  and general linear time invariant channels,  $f(\cdot, \cdot)$  can be approximated as

$$\begin{aligned} f(F_x(x), \sigma) &= \Pr_X(X > x) \\ &\approx 1 - \Phi(\sqrt{2}x/\sigma - \sqrt{2N-1}) \\ &= \text{erfc}(\sqrt{2}x/\sigma - \sqrt{2N-1})/2 \\ &\approx \exp(-u^2)/2\sqrt{\pi}u \end{aligned}$$

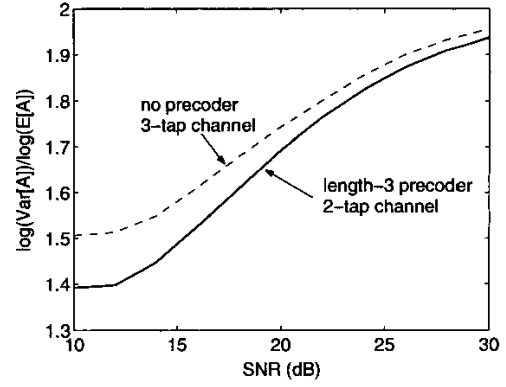


Figure 3: Performance for length-two blocks over random channels.

where  $u = (\sqrt{2}x/\sigma - \sqrt{2N-1})$ , and the first approximation is Fisher's approximation [8]. This gives  $\phi_2(\sigma) = \sigma^2$ , and

$$\begin{aligned} \phi_1(F_x(x), \sigma) &= -(\sqrt{2}x/\sigma - \sqrt{2N-1})^2 \\ &\quad - \sigma^2 \log \left( 2\sqrt{\pi} \left( \frac{\sqrt{2}x}{\sigma} - \sqrt{2N-1} \right) \right). \end{aligned}$$

We believe that these satisfy the hypotheses of Theorem 1.

## 4. Numerical results

Numerical experiments were carried out on blocks of two symbols, with random channels of length three and a zero-forcing equaliser. The dashed line in Figure 3 shows the ratio

$$\frac{\log(\text{Var}[A(\sigma)])}{\log(E[A(\sigma)])}$$

as a function of SNR, where  $A$  is the proposed estimator. The specific estimators used was for the probability of there being exactly one error. The results were averaged over 100 000 samples. Given the definition in (1), this graph clearly demonstrates that the algorithm is asymptotically optimal in this case.

Many block-based communications systems use linear precoders. One very common example is orthogonal frequency division multiplex (OFDM) systems, in which the linear precoder consists of an inverse Fourier transform followed by a cyclic prefix which adds redundancy to the data. The solid line in Figure 3 shows the results for 500 000 random precoders, which map two-symbol data blocks into three-symbol channel blocks (with channel length 2). This demonstrates that the proposed scheme is again asymptotically optimal when precoders are used.

## 5. Conclusion

An algorithm has been proposed for computing the distribution of bit errors in a block-based communication system. The algorithm has been shown to be asymptotically optimal. For simplicity, this has only been shown in the simple case of two-dimensional blocks transmitted over a three-tap channel, but it holds more generally. This is the subject of a forthcoming paper.

This algorithm is closely related to that of [3], but is more flexible since the former relies intrinsically on the decision boundaries being hyperplanes. It is also more efficient than [4], the precursor of [3].

This pilot study can be extended in many ways to improve the efficiency of the algorithm. Most notably, the algorithm can be combined with importance sampling to improve the performance with large blocks.

## References

- [1] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communications systems," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 597–613, May 1997.
- [2] P. Bratley, B. L. Fox, and L. E. Schrage, *A Guide to Simulation*. New York, NY: Springer Verlag, 1987.
- [3] J.-E. Porath and T. Aulin, "Improved technique for quick error rate estimation of multi-dimensional communications schemes," *IEE Proc.-Commun.*, vol. 146, pp. 343–346, Dec. 1999.
- [4] N. C. Beaulieu, E. Bileli, and M. Lai, "Sample rejection and importance sampling in the simulation of multidimensional signalling systems," *IEE Proc.-Commun.*, vol. 140, pp. 445–452, Dec. 1993.
- [5] Y. A. Schreider, *Method of Statistical Testing*. Amsterdam: Elsevier, 1964.
- [6] A. E. Conway and D. E. O'Brien, "Hybrid analysis of response time distributions in queueing networks," *IEEE Trans. Commun.*, vol. 41, pp. 1091–1101, July 1993.
- [7] P. Heidelberger, "Fast simulation of rare events in queueing and reliability models," *ACM Trans. Model. Comput. Simul.*, vol. 5, pp. 43–85, Jan. 1995.
- [8] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, vol. 1 of *Probability and mathematical statistics*. New York, NY: John Wiley & Sons, 2nd ed., 1994.

## Appendix

Theorem 1 will be proved using the following lemmas.

**Lemma 1** A positive bounded random variable,  $Z \in [0, a]$ , with mean  $E[Z] = p$ , has variance bounded by  $\text{Var}[Z] \leq p(a - p)$ .

**Lemma 2** Let

$$\theta(\sigma) = \max \left\{ \theta : f(\theta, \sigma) > \frac{1}{2} f(0, \sigma) \right\}, \quad (8)$$

with  $f(\cdot, \cdot)$  as defined in Theorem 1. Then

$$\theta(\sigma) \geq \frac{\phi_2(\sigma) \log 2}{k}, \quad (9)$$

with  $k$  defined in (2f), for all  $\sigma$  such that

$$\frac{\phi_2(\sigma) \log 2}{k} < U. \quad (10)$$

**Proof:** For all  $\theta \in (0, \phi_2(\sigma) \log 2/k)$  where  $\sigma$  satisfies (10),

$$\begin{aligned} \phi_1(\theta, \sigma) - \phi_1(0, \sigma) &> -k\theta \\ &> -\phi_2(\sigma) \log 2, \end{aligned}$$

whence

$$\begin{aligned} \frac{f(\theta, \sigma)}{f(0, \sigma)} &= \exp \left( \frac{\phi_1(\theta, \sigma) - \phi_1(0, \sigma)}{\phi_2(\sigma)} \right) \\ &> \exp(-\log 2) = 1/2, \end{aligned}$$

giving the result.  $\square$

**Proof of Theorem 1:** By Lemma 1, and since  $f(u, \sigma) \leq f(0, \sigma)$  by (2c), it is sufficient to show that

$$\frac{\log(f(0, \sigma))}{\log(E[f(U, \sigma)])} \rightarrow 1$$

as  $\sigma \rightarrow 0$ . Note also that

$$|\log E[f(U, \sigma)]| < |\log f(0, \sigma)| + |\log \theta(\sigma)| + \log 2, \quad (11)$$

since  $\theta(\sigma) \leq 1$  and  $f(0, \sigma) < 1$  by (2a) and (2b), and that

$$E[f(U, \sigma)] \geq \int_0^{\theta(\sigma)} \frac{1}{2} f(0, \sigma) du.$$

Thus it is sufficient to show that

$$\frac{\log(\theta(\sigma))}{\log(f(0, \sigma))} \rightarrow 0$$

as  $\sigma \rightarrow 0$ . But, by (9) of Lemma 2 and (2e),

$$\begin{aligned} \frac{\log(\theta(\sigma))}{\log(f(0, \sigma))} &< \frac{\log(\log 2) - \log k + \log(\phi_2(\sigma))}{\phi_1(0, \sigma)/\phi_2(\sigma)} \\ &\rightarrow 0. \end{aligned}$$

The result then follows, by the positivity of the left hand side of the inequality.  $\square$