

# Fast Estimation of Outage Probabilities in MIMO Channels

Rajan Srinivasan and George Tiba

**Abstract**—Fast estimation methods for small outage probabilities of signaling in fading multiple-input multiple-output (MIMO) channels are developed. Communication over such channels is of much current interest, and quick and accurate methods for estimating outage capacities are needed. The methods described herein use adaptive importance sampling (IS) techniques as developed in a series of recent publications. Fast algorithms are provided for evaluating “nonergodic” capacities of Rayleigh fading MIMO channels. The methodology can be extended to more general models. Numerical results on outage capacity are provided, and these extend and complement known results in the literature.

**Index Terms**—Fast simulation, information rates, multiple-input multiple-output (MIMO) systems, Monte Carlo methods, outage probabilities.

## I. INTRODUCTION

IN THIS LETTER, we consider the problem of estimating channel capacities in multiple-input multiple-output (MIMO) communication systems operating over random channels. Such systems are of great technical and economic importance for wireless applications, as the signaling techniques that underlie them promise hugely enhanced information transmission rates [1]–[3]. For signal transmission formats wherein the MIMO channel can be considered “nonergodic,” mutual information is presented as a random variable, and its probability of falling below a certain “outage capacity” is referred to as the outage probability. Estimation of outage probabilities and capacities for channel models usually employed is analytically difficult, [3], and researchers resort to Monte Carlo (MC) simulation. For the low outage probabilities desirable in practice, MC simulations are long, as amply evidenced by results in the by-now almost classic publication [1]. The situation is exacerbated when very small probabilities are of interest. In such cases, the usual considerations that necessitate invoking special simulation procedures [such as importance sampling (IS)] in rare-event simulations apply. These fast procedures have been in use for several years and have found a range of employment. A good review article on IS in communication systems is [4]. A recent comprehensive text dealing with system simulation is [5], whereas [8] is concerned with adaptive IS methods for detection and communications.

Paper approved by V. A. Aalo, the Editor for Diversity and Fading Channel Theory of the IEEE Communications Society. Manuscript received November 22, 2002; revised July 9, 2003.

R. Srinivasan is with the Telecommunication Engineering Group, University of Twente, 7500 AE Enschede, The Netherlands (e-mail: r.srinivasan@el.utwente.nl).

G. Tiba was with the Telecommunication Engineering Group, University of Twente, 7500 AE Enschede, The Netherlands. He is now with Nuclear Power Plant Cernavoda Unit 2, Jud. Constanta, Romania (e-mail: tibatg@cne-u2.ro).

Digital Object Identifier 10.1109/TCOMM.2004.826259

The literature on IS is large, and the above references provide a path to available important results.

Herein, we briefly describe and use adaptive IS methods for the application stated above. Whereas a form of adaptive IS has been studied in [6], the methods used here rely entirely on the results in [7], [9], and [10]. A brief introduction to the standard (idealized) fading channel model used and the expression for outage probability is given, followed by description of estimation techniques. The inverse problem, that of determining an outage capacity for specified outage probability and signal-to-noise ratio (SNR), is then studied, also using IS. Numerical results are presented, together with some brief conclusions.

## II. MIMO OUTAGE PROBABILITY

Consider a single-user point-to-point digital communication system consisting of  $n_t$  transmitting and  $n_r$  receiving antennas operating over a linear Gaussian channel. The transmitted signal is denoted by the  $n_t$ -dimensional vector  $\mathbf{x}$ , and the received signal by the  $n_r$ -dimensional vector  $\mathbf{y}$ . The total transmitted power is constrained to  $P_T$ , irrespective of the value of  $n_t$ , and the average power at the output of each receiving antenna is denoted by  $P$ . The additive channel noise, denoted by the  $n_r$ -dimensional vector  $\mathbf{n}$ , is zero-mean complex Gaussian with independent real and imaginary components and having covariance matrix  $N\mathbf{I}_{n_r}$ , where  $N$  denotes the noise power at each receiver branch, and  $\mathbf{I}_{n_r}$  the  $n_r \times n_r$  identity matrix. The average SNR at each receiver antenna is denoted by  $\rho = P/N$ . The channel is represented by the  $n_r \times n_t$  complex input–output matrix  $\mathbf{G}$ . A normalized channel matrix  $\mathbf{H}$  is defined by  $P_T^{1/2}\mathbf{G} = P^{1/2}\mathbf{H}$ . The channel can be deterministic and fixed, or it can be random. In the latter case, it is assumed that it remains unchanged for the entire duration of a communication, but can randomly change to another realization for another channel use. This is the nonergodic case. It is further assumed that the channel realization  $\mathbf{H}$  is known at the receiver, possibly through channel estimation procedures. With this formulation, the maximum mutual information  $\mathcal{I}$  between  $\mathbf{x}$  and the channel output  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$  has been shown in [1] to be given by

$$\mathcal{I} = \log_2 \det \left[ \mathbf{I}_{n_r} + \frac{\rho}{n_t} \mathbf{H}\mathbf{H}^\dagger \right] \frac{\text{bits}}{\text{s}} \text{ Hz} \quad (1)$$

where “det” denotes determinant and  $(\cdot)^\dagger$  conjugate transpose. The usual ergodic (or Shannon) capacity is just  $E\{\mathcal{I}\}$  computed over the fading channel statistics, available in [3].

With the channel modeled as random, capacity is treated as a stochastic variable with a certain probability distribution. The outage probability  $P_{\text{out}}$  is defined as

$$P_{\text{out}} = P(\mathcal{I} \leq R) \quad (2)$$

where  $R$  is a given transmission rate. The simulation problem is then one of estimating, for given values of SNR  $\rho$ , and numbers of transmitting and receiving elements  $n_t$  and  $n_r$ , small values of  $P_{\text{out}}$ . It is also of interest to determine the rate  $R$  for given  $\rho$ ,  $n_t$ , and  $n_r$  to meet a desired (small) value of  $P_{\text{out}}$ . Using (1), (2) can be rewritten as

$$P_{\text{out}} = P \left( \det \left[ I_{n_r} + \frac{\rho}{n_t} \mathbf{H} \mathbf{H}^\dagger \right] \leq 2^R \right). \quad (3)$$

Using MC simulation, capacity results have been obtained in [1] typically down to  $P_{\text{out}}$  values of 0.01. Estimating lower values can entail prohibitively long computations.

### III. FAST ESTIMATION OF $P_{\text{out}}$

To estimate  $P_{\text{out}}$  using IS, we make a few observations. An estimator of (3) has to be implemented, which causes the event therein to occur frequently, thereby ensuring shorter simulation lengths than would be possible with an MC procedure. This can be done by suitably biasing the random variables that make up  $\mathbf{H}$ . From the form of (3), it is clear that biasing should compress the probability density of the determinant, which has support in  $(1, \infty)$ . The compression is toward unity. It follows, therefore, that a simple scaling down of the variables in  $\mathbf{H}$  with a real positive number will achieve this. While this may not be the best possible method of biasing, it is effective, as shown in the following. The chief problem then centers around determining good values of the scaling constant for given sets of parameters of the MIMO system. An optimum value of scaling is that which provides an unbiased IS estimator with minimum variance. Equivalently, it minimizes the simulation length required to achieve the same variance as a conventional MC estimator, resulting in simulation gain.

In adopting a model for the channel matrix, it is usually assumed [1], [3], that the entries are independent and complex Gaussian, with independent and identically distributed (i.i.d.) real and imaginary components, each with variance  $1/2$ . This is the Rayleigh fading channel model. Symmetry suggests that each random variable in  $\mathbf{H}$  contributes equally to the determinant. Therefore, a single scaling parameter is used for biasing all the component random variables. Assume that each variable is scaled by the quantity  $b^{1/2}$ , where  $0 < b \leq 1$ . The IS estimator  $\hat{P}_{\text{out}}$  of the outage probability in (3) can then be written as

$$\hat{P}_{\text{out}} = \frac{1}{K} \sum_1^K \mathbf{1} \left( \det \left[ I_{n_r} + \frac{\rho}{n_t} \mathbf{H} \mathbf{H}^\dagger \right] \leq 2^R \right) W(\boldsymbol{\xi}; b), \quad \boldsymbol{\xi} \sim f_\star \quad (4)$$

where  $\boldsymbol{\xi}$  denotes a  $(n_r n_t \times 1)$  complex vector containing all the elements of  $\mathbf{H}$  in any order,  $\mathbf{1}(\cdot)$  is an indicator function which is one if the event in parentheses occurs and zero otherwise, and  $K$  is the length of the IS simulation. It is an unbiased and consistent estimator. Throughout, hats on quantities indicate estimates. The weighting function  $W$  is defined by

$$W(\boldsymbol{\xi}; b) \equiv \frac{f(\boldsymbol{\xi})}{f_\star(\boldsymbol{\xi}; b)} \quad (5)$$

where  $f_\star$  and  $f$  are density functions of  $\boldsymbol{\xi}$  with and without ( $b = 1$ ) biasing; the density of a complex vector being that of the real vector composed of its real and imaginary parts. The notation  $\boldsymbol{\xi} \sim f_\star$  indicates that vector  $\boldsymbol{\xi}$  is drawn from the distribution corresponding to density  $f_\star$ . Note that this scaling-based IS estimator is applicable to channels characterized by any probability distribution. For the Rayleigh fading channel, the weighting function takes the simple form

$$W(\boldsymbol{\xi}; b) = b^{n_t n_r} e^{-(1-1/b) \boldsymbol{\xi}^\dagger \boldsymbol{\xi}}.$$

The estimator in (4) has variance

$$\text{var} \hat{P}_{\text{out}} = \frac{1}{K} [E_\star \{ \mathbf{1}(\mathcal{A}) W^2(\boldsymbol{\xi}; b) \} - P_{\text{out}}^2] \quad (6)$$

with  $E_\star$  denoting expectation with respect to the biased distribution. In (6) and below,  $\mathcal{A}$  denotes the event in (4). Minimizing this variance is equivalent to minimizing the expectation term, which we denote as  $I$  and rewrite as

$$\begin{aligned} I(b) &= E_\star \{ \mathbf{1}(\mathcal{A}) W^2(\boldsymbol{\xi}; b) \} \\ &= E \{ \mathbf{1}(\mathcal{A}) W(\boldsymbol{\xi}; b) \} \end{aligned} \quad (7)$$

using (5). A stochastic Newton optimization algorithm that minimizes an estimate  $\hat{I}$  of  $I$  is described by

$$b_{j+1} = b_j - \delta \frac{\hat{I}'(b_j)}{\hat{I}''(b_j)}, \quad j = 1, 2, \dots \quad (8)$$

and used for the minimization. This algorithm determines estimates of optimum scaling factors. Here,  $j$  is a recursion index and  $\delta$  a step-size parameter used to achieve a tradeoff between speed and noisiness of convergence. Primes indicate derivatives with respect to  $b$ . These estimators can be set up by differentiating  $I$  in (7) twice, with respect to  $b$ , to obtain

$$\begin{aligned} \hat{I}'(b_j) &= \frac{1}{K} \sum_1^K \mathbf{1}(\mathcal{A}) W(\boldsymbol{\xi}; b_j) W'(\boldsymbol{\xi}; b_j), \quad \boldsymbol{\xi} \sim f_\star \\ \hat{I}''(b_j) &= \frac{1}{K} \sum_1^K \mathbf{1}(\mathcal{A}) W(\boldsymbol{\xi}; b_j) W''(\boldsymbol{\xi}; b_j), \quad \boldsymbol{\xi} \sim f_\star \end{aligned}$$

where

$$\begin{aligned} W'(\boldsymbol{\xi}; b_j) &= \frac{1}{b_j^2} (n_t n_r b_j - \boldsymbol{\xi}^\dagger \boldsymbol{\xi}) \cdot W(\boldsymbol{\xi}; b_j) \\ W''(\boldsymbol{\xi}; b_j) &= \frac{W'(\boldsymbol{\xi}; b_j)^2}{W(\boldsymbol{\xi}; b_j)} - \frac{1}{b_j^3} (n_t n_r b_j - 2\boldsymbol{\xi}^\dagger \boldsymbol{\xi}) \cdot W(\boldsymbol{\xi}; b_j). \end{aligned}$$

As is typical of stochastic approximation procedures, convergence of the  $b$ -algorithm in (8) is characterized by a small random vibration around the optimum value. It may be noted that a choice of  $b$  that minimizes  $I$  need not necessarily minimize the variances of  $\hat{I}'$  and  $\hat{I}''$ . The two IS algorithms (4) and (8) are implemented simultaneously. This method of adaptive IS was proposed in [9] in the context of tail probability estimation of sums of random variables, and also applied therein to detection problems.

#### IV. FAST ESTIMATION OF $R$

The inverse problem, namely, that of determining an  $R$  that achieves a specified outage probability for given SNR, can also be solved using IS. The genesis of the method lies in the so-called *inverse IS* problem first formulated and solved in [9], with applications to threshold determination in (radar) detection algorithms therein and in [10]. It has subsequently been used for parameter optimization in communication systems [7]. The procedure here involves minimizing a stochastic objective function  $J$  with respect to the rate  $R$  to be estimated. It has the form

$$J(R, b) = \left[ \hat{P}_{\text{out}}(R, b) - P_{\text{out}}(R_o) \right]^2.$$

Here  $P_{\text{out}}(R_o)$  is specified, and  $R_o$ , to be estimated, is the rate that achieves the specification. Dependence of  $J$  and  $\hat{P}_{\text{out}}$  on the scaling factor  $b$  is made explicit to emphasize that minimization of  $J$  is concomitant with determination and use of optimum biasing. The algorithm to estimate  $R_o$  is

$$R_{o,j+1} = R_{o,j} + \delta_R \frac{P_{\text{out}}(R_o) - \hat{P}_{\text{out}}(R_{o,j}, b_j)}{\hat{P}'_{\text{out}}(R_{o,j}, b_j)} \quad (9)$$

where  $\delta_R$  is a step-size parameter. The prime in the denominator of the second term on the right-hand side (RHS) of (9) indicates derivative with respect to  $R$ . To estimate this derivative, we turn to (4), which, unfortunately, is not differentiable owing to the indicator. Nevertheless, a derivative can be generated by approximating it with a differentiable function [7], [10]. The sigmoid nonlinearity, famous from its use in the backpropagation algorithm for training artificial neural networks, is employed. In our situation, it takes the form  $\mathcal{S}(R) = (1 + e^{aA})^{-1}$ , where

$$A \equiv \det \left[ I_{n_r} + \left( \frac{\rho}{n_t} \right) \mathbf{H}\mathbf{H}^\dagger \right] - 2^R$$

and  $a$  is a positive constant. Replacing the indicator in (4) with  $\mathcal{S}$  and carrying out the differentiation yields

$$\hat{P}'_{\text{out}}(R_{o,j}, b_j) = \frac{a 2^{R_{o,j}} \ln 2}{K} \sum_1^K \mathcal{S}(R_{o,j}) [1 - \mathcal{S}(R_{o,j})] \cdot \mathcal{W}(\boldsymbol{\xi}; b_j), \quad \boldsymbol{\xi} \sim f_*$$

It is important to stress that the sigmoid is used only for the purpose of estimating the derivative in (9), while the outage probability itself is estimated as in (4). Finally, the algorithms (4), (8), and (9) are implemented simultaneously.

#### V. RESULTS AND DISCUSSION

Results of implementation are in Figs. 1–6. Adaptively determined optimum scaling factors give rise to estimates of outage probabilities, Figs. 1–3, down to around  $10^{-9}$ . In [1], it was shown that at a  $P_{\text{out}}$  of  $10^{-2}$ , a 3-dB increase in SNR at 21 dB produces almost  $n$  ( $n = n_t = n_r$ ) more b/s/Hz in outage capacities. This turns out to be true also for lower  $P_{\text{out}}$  values of  $10^{-3}$ ,  $10^{-4}$ , and  $10^{-5}$ , but the increase is somewhat smaller than  $n$ . At lower SNRs, the increase is smaller. Results for different numbers of transmitter and receiver antennas are in Fig. 4. These

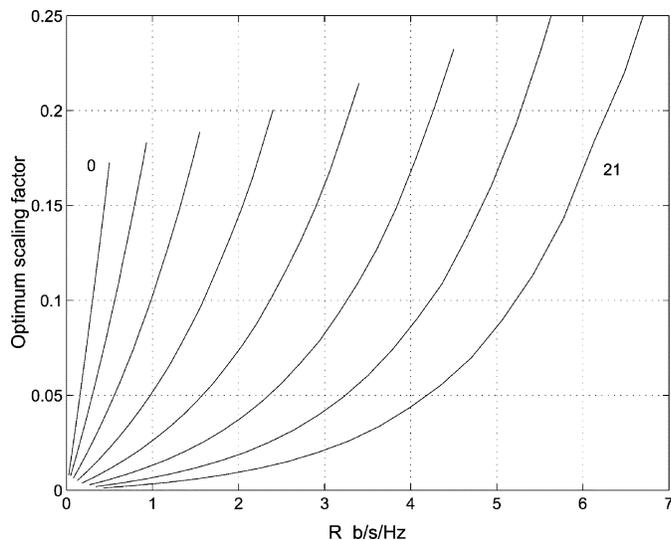


Fig. 1. Optimum biasing versus outage capacities for  $n_t = n_r = 2$ . Parameter is SNR in steps of 3 dB.

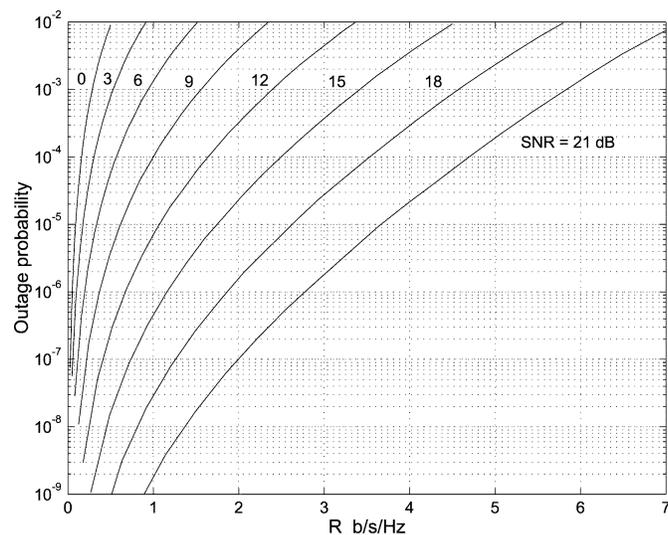


Fig. 2. Outage probabilities versus capacities for  $n_t = n_r = 2$ .

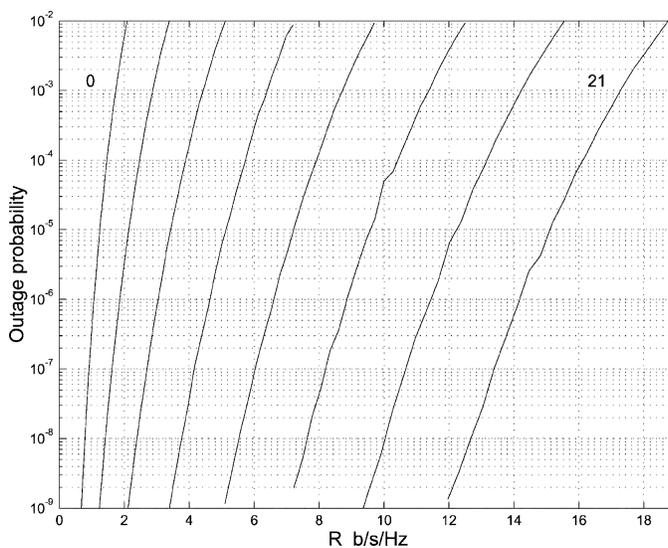


Fig. 3. Outage probabilities versus capacities for  $n_t = n_r = 4$ .

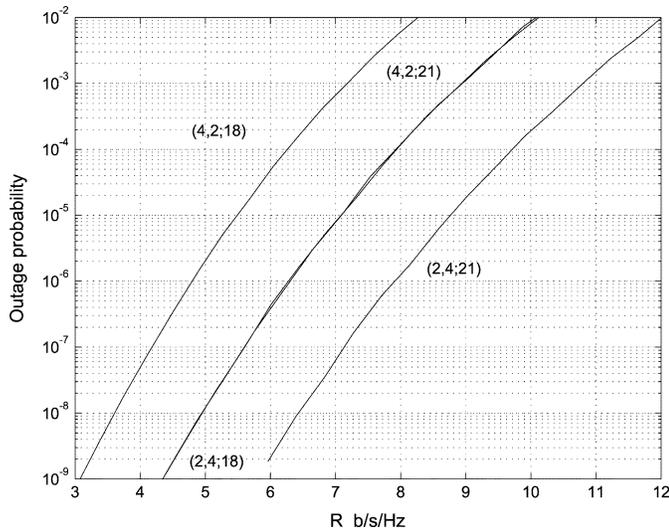


Fig. 4. Outage probabilities versus capacities for  $(n_t, n_r, \text{SNR dB})$ .

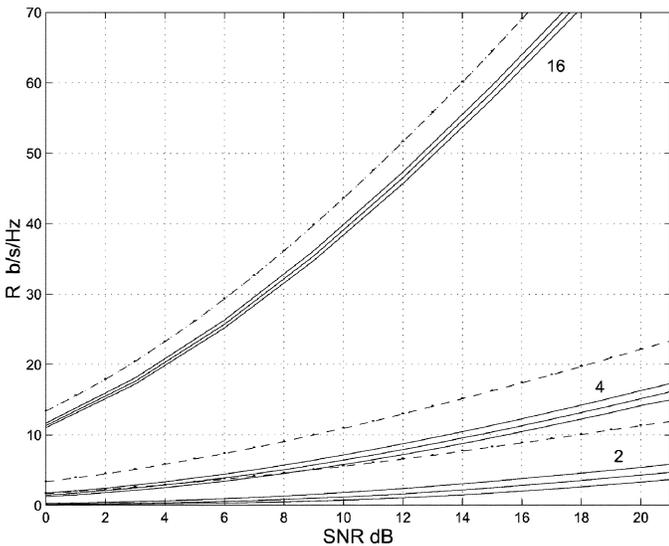


Fig. 5. Outage capacities versus SNR for  $n_t = n_r = 2, 4, 16$ . Sets of three graphs are for  $P_{\text{out}} = 10^{-3}, 10^{-4}, 10^{-5}$ , going downward. Dashed lines are ergodic capacities for the same number of antennas.

show that an  $(n_t, n_r) = (2, 4)$  configuration enjoys an SNR advantage of 3 dB over a  $(4, 2)$  system at 21 dB, which alternately translates to an advantage of almost 2 b/s/Hz, even for ultra-low outage probabilities.

Experiments with inverse estimation lead to results in Fig. 5 for three sets of antennas at three specified outage probabilities. Also shown for comparison are ergodic capacities, obtained through simulation. An interesting finding is that for  $n = 2$ , at 21 dB, SNR increases of less than 3 dB produce reductions of outage probabilities from  $10^{-2}$  to  $10^{-3}$  to  $10^{-4}$  to  $10^{-5}$  at fixed outage capacities. For  $n = 4$  antennas, the additional SNR required is less than 1.5 dB, whereas for  $n = 16$ , it is less than half a decibel. Outage capacity curves for different  $P_{\text{out}}$ 's become closely bunched as the number of antennas increases, and also get closer to Shannon capacities (shown dashed). This implies that with an increasing number of antennas, less loss (in comparison to Shannon capacity) in terms of b/s/Hz is incurred

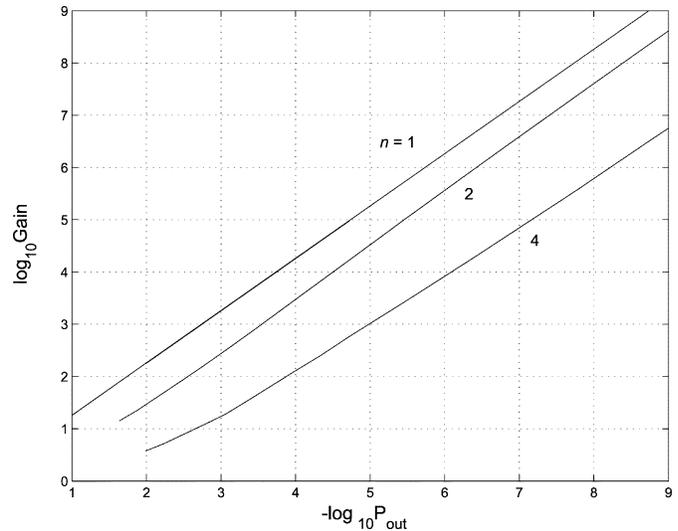


Fig. 6. Simulation gains as a function of outage probabilities. SNR = 21 dB for  $n = 2, 4$ .

if systems are operated at low outage probabilities. Of course, operating at rates equal to the mean capacity would lead to unacceptably high outage probabilities.

Effectiveness of IS algorithms is best quantified by estimating gains (ratios) in simulation lengths  $K$  achieved over conventional MC procedures for equal variances. Actual variances of both procedures decrease inversely with (respective) simulation lengths, as evident from (6). As seen in Fig. 6, a roughly linear behavior of gain versus (low) outage probability is observed for the logarithmic scales used. This implies that simulation gain varies as  $10^c / P_{\text{out}}^m$ , where  $m$  and  $c$  denote slope and  $y$ -axis intercept of the extrapolated line. Assume now that for conventional MC simulation, a length of  $100P_{\text{out}}^{-1}$  (a rule of thumb figure) provides sufficient accuracy; this translates (using an asymptotic normality argument) to a 95% confidence of having an error not exceeding 20%. Therefore, the IS length required for the same variance would be  $K = 10^{2-c} P_{\text{out}}^{m-1}$ . From this, we infer that if the slope is less than, equal to, or greater than unity, then IS simulation lengths would increase, remain constant, or even decrease with decreasing outage (rare-event) probability. This simple though striking fact demonstrates the potential of IS. In Fig. 6, we see that for  $n = 1$  (for which the outage probability can be calculated in closed form) and 2, the length  $K$  is independent of  $P_{\text{out}}$ . The single-antenna case being simple, the gain has been determined analytically (by scaling the exponential magnitude squared channel gain) and evaluated numerically. A closer examination of (3) and (7) reveals that the gain in this case is independent of SNR for a fixed  $P_{\text{out}}$ . For  $n = 2$ , at  $P_{\text{out}} = 10^{-6}$ , the gain is approximately  $3.5 \times 10^5$ , which means that an IS length of  $K = 300$  provides the same accuracy as an MC length of  $10^8$ . This value of  $K$  is sufficient to ensure the fixed above-given accuracy for any value of (low) outage probability. All simulations reported here have employed not less than  $K = 2000$ . IS gains decrease and simulation lengths needed increase with  $n$ . This is presumably due to the shape of the density of the determinant, indicating that biasing by uniform scaling loses its effectiveness with increasing size of the channel matrix.

An important question is how much computation (relative to MC) is finally needed to implement adaptive IS algorithms. Subtleties often arise in actual implementation of fast simulation algorithms. It is beyond the remit of this small letter to study them all. Therefore, we address this general issue briefly and qualitatively. Typically, computational saving afforded by the sometimes enormous IS simulation gains is used to generate estimates of rare-event probabilities for various values of system parameters. Adaptive biasing parameters usually do not change drastically with small changes in system parameters. This allows the design of adaptive algorithms which converge in relatively few iterations, following an initial search. Moreover, given the iterative nature of the stochastic recursions, it is possible at every iteration to use the same set of (unbiased) generated random variates till completion of the recursions. However, some caution must be exercised here. The set of random numbers must at least be tested by validation on problems with *a priori* known answers. This has been used by the first author in several applications previously. These artifices speed up computations by several orders of magnitude, and usually produce smooth results. An additional and, by all means, important facility of these IS algorithms is the possibility of simultaneously performing parameter optimization in situations that involve the occurrence of rare events. Be all this as it may, for probabilities that are not low, conventional MC simulations certainly involve far less analytical effort, less computational complexity, and slightly larger simulation lengths.

## VI. CONCLUSIONS

The use of adaptive IS for the estimation of outage capacities of Rayleigh faded MIMO channels has been briefly described and demonstrated. A few observations have been made that could have bearing on implementation of such systems. For reasons of economy, more detailed results have not been sought. As an initial aid to researchers, we have provided some results

on biasing parameters that can be used to validate simulations. In this regard, it should be noted that biasing by scaling is a robust though conservative technique. More effective schemes can be developed, but this would require intimate study of the interplay of random variables involved in expressions for outage probabilities, especially for situations that embody correlation properties of signals and antennas. Nevertheless, the methods described herein are general enough in their applicability, and can be used to study practical channel models available in the literature. As a final comment, being a forced MC procedure, IS should be viewed not as an alternative to MC simulation, but a complement.

## REFERENCES

- [1] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, pp. 315–335, 1998.
- [2] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
- [3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Dec. 1999.
- [4] P. J. Smith, M. Shafi, and H. Gao, "Quick simulation: A review of importance sampling techniques in communication systems," *IEEE J. Select. Areas Commun.*, vol. 15, pp. 597–613, May 1997.
- [5] M. C. Jeruchim, P. Balaban, and K. S. Shanmugan, *Simulation of Communication Systems: Modeling, Methodology, and Techniques*, 2nd ed. New York: Kluwer/Plenum, 2000.
- [6] W. A. Al-Qaq, M. Devetsikiotis, and J. K. Townsend, "Importance sampling methodologies for simulation of communication systems with time-varying channels and adaptive filters," *IEEE J. Select. Areas Commun.*, vol. 11, pp. 317–326, Apr. 1993.
- [7] D. Remondo, R. Srinivasan, V. F. Nicola, W. Van Etten, and H. E. P. Tattje, "Adaptive importance sampling methods for performance evaluation and parameter optimization of communication systems," *IEEE Trans. Commun.*, vol. 48, pp. 557–565, Apr. 2000.
- [8] R. Srinivasan, *Importance Sampling – Applications in Communications and Detection*. Berlin, Germany: Springer-Verlag, 2002.
- [9] —, "Some results in importance sampling and an application to detection," *Signal Processing*, vol. 65, pp. 73–88, Feb. 1998.
- [10] —, "Simulation of CFAR detection algorithms for arbitrary clutter distributions," *Proc. IEE Radar, Sonar, Navigation*, vol. 147, pp. 31–40, Feb. 2000.