International Journal of Systems Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsys20

A hybrid strategy for Gilbert's channel characterization using gradient and annealing techniques

TAN-HSU TAN [a] & WEN-WHEI CHANG [a]
[a] Department of Communication Engineering , National Chiao Tung University , Hsinchu, Taiwan Phone: Tel: 0021-886-3-5731826 Fax: Tel: 0021-886-3-5731826
Published online: 16 May 2007.

To cite this article: TAN-HSU TAN & WEN-WHEI CHANG (1998) A hybrid strategy for Gilbert's channel characterization using gradient and annealing techniques, International Journal of Systems Science, 29:6, 579-585, DOI: 10.1080/00207729808929549

To link to this article: http://dx.doi.org/10.1080/00207729808929549

International Journal of Systems Science, 1998, volume 29, number 6, pages 579–585

A hybrid strategy for Gilbert's channel characterization using gradient and annealing techniques

TAN-HSU TAN† and WEN-WHEI CHANG†‡

This paper presents a hybrid algorithm for estimating Gilbert's channel model parameters from an experimental error-gap distribution. A stochastic simulated annealing algorithm is applied to determine automatically a set of good starting points, which are then used by the deterministic gradient algorithm for faster convergence to the global optimum. Simulation results indicate that, for channel characterization, this hybrid strategy provides an ideal compromise between modelling accuracy and convergence time.

1. Introduction

For many years there has been considerable interest in reliable transmission of bit-rate reduced signals over noisy channels. One general solution is to select a fixed coder configuration that meets the error-tolerance capacity of the worst channel to be expected. The main drawback to this approach is that the coders cannot adapt to channel variations. To complicate matters further, the proposed coder candidates were usually ranked according to their average bit-error-rate (BER) performances. The basic problem with this ranking method is that the BER has difficulty distinguishing between random and burst characteristics of sample error sequences, perhaps the most efficient coding system should take the intrinsic natures of typical error occurrences into consideration. Therein lies the motivation for channel modelling, which lends itself easily to parametrization from experimental error sequences, and to further utilization in adaptive error-control design.

Transmission errors encountered in digital communication channels tend to exhibit various degrees of statistical dependences between successive samples. Attempts to characterize such compound channels have included numerous parametrized probabilistic models proposed to assess some of the most relevant aspects of error statistics (Kanal and Sastry 1978). Most studies have emphasized finite-state Markov chain models (Sato *et al.* 1991, Fritchman 1967). The principal difficulty encountered in channel characterization is that model parameters are not directly observable; so methods of deducing them from easily measured error statistics must be considered. The use of exponential curve fitting allows channel modelling to be formulated as a combinatorial optimization problem in which the squared error distortion between the measured error-gap distribution and its modelled fit is a cost function to be minimized. This task can be done by using an estimation method based on the gradient-descent algorithm (Chouinard *et al.* 1988). Although the gradient-descent method has a fast convergence rate, its simple downhill search transitions can easily become trapped in local minima and thus miss finding the globally optimal solution. Thus, it is common to run gradient searches from a large number of starting points, and the best solution is chosen from among those obtained. In addition, quasi-Newton methods (Powell 1984) have also been successfully applied in solving the multidimensional optimization problems. Although the quasi-Newton approach is conceptually satisfying in its ability to converge rapidly, it is based on the assumption that the cost function can be locally approximated by the quadratic form with a convex condition. However, for channel modelling problems, the cost function tends to exhibit many different convex regions and has been found difficult to optimize by means of the quasi-Newton method. An alternative approach to function optimization is based on the stochastic simulated annealing algorithm (Kirkpatrick *et al.* 1983). It aims to benefit through acceptance of permutations that move uphill in a con-

Received 7 July 1997. Revised 9 September 1997. Accepted 3 October 1997.

† Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan. Tel: 0021-886-3-5731826; Fax: 0021-886-3-5710116.

‡ e-mail: wwchang@cc.nctu.edu.tw.

0020-7721/98 $12.00

trolled fashion. The main drawback to simulated annealing is that convergence can be very slow for complicated optimization problems. This suggests a hybrid strategy involving merging the most appealing features of these two methods, the first component of which helps to identify the smallest number of starting points through simulated annealing. The second component, the gradient-descent method, then uses these starting points to ensure that the actual optimum is found efficiently.

This paper is organized as follows. An overall view of the investigation has been presented in this section. In section 2, we present the basic aspects of channel modelling and algorithms for statistical error characterization. Section 3 presents an estimation algorithm based on simulated annealing for optimal identification of Gilbert's channel model parameters. In section 4, we explore the benefits of using a hybrid estimation approach that combines the stochastic annealing algorithm with the deterministic gradient-descent algorithm. Comparative performance results for estimating Gilbert's model parameters in conjunction with various optimization algorithms are also included. Section 5 presents a short summary and a list of conclusions.

2. Error occurrence model

Error control code design and performance analysis require that probabilistic models be used to describe the statistical distribution of errors due to channel impairments. While the binary symmetric channel model is simple, it has some limitations in simulating the behaviour of error bursts over channels with memory. To match error sources more closely, Gilbert (1960) introduced the application of finite-state Markov chain models to the representation of digital channels with memory. Gilbert's channel model consists of a Markov chain having a good state G, which is error free, and a bad state B, in which the channel has an error probability of $1-h$. The model state transition diagram is shown in figure 1. A wide range of channels can be represented by appropriate definitions of the transition probabilities among all states. To illustrate this, consider the example of burst channels. The state transition probabilities P and p are so small that the probabilities Q and q of remaining in G and B states will be high.

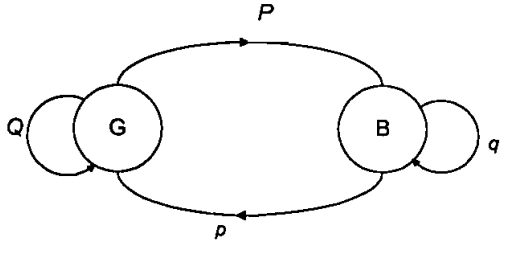

Figure 1. Gilbert's model.

The principal difficulty encountered in parameter estimation is that Gilbert's model parameters $\{P,h,p\}$ are not directly observable; so methods of deducing them from more easily measured error statistics must be derived. The measurement data considered here are the error-gap distributions, denoted by $P(0^m|1)$, that give the probability that at least m successive error-free bits will be encountered next on the condition that an error bit has just occurred. In many applications (Chouinard *et al.* 1988), it suffices to postulate that the experimental error-gap distribution can be well approximated by the sum of two exponential functions:

$$P(0^m|1) = \alpha_1\beta_1^m + \alpha_2\beta_2^m. \tag{1}$$

This is because the number of exponentials correspond to the number of distinct line segments embedded in the measured value, expressed logarithmically, of the error-gap distribution. In the case of curve fitting, $\{\alpha_2,\beta_2\}$ are chosen to match the correct behaviour of $P(0^m|1)$ for large m, and $\{\alpha_1,\beta_1\}$ are chosen to improve the fit for small m. Proceeding in this way, the original descriptive modelling issue can be formulated as a combinatorial optimization problem in which the parameters $\{\alpha_i,\beta_i\}$ are the optimization variables to be identified. For this investigation, a suitable cost function is the sum of the squared errors between the measured error-gap distribution values and the modelled fits. This minimization leads to a constrained nonlinear optimization problem that can be stated as follows:

$$\min_{\alpha,\beta} E = \min_{\alpha,\beta}\left\{\sum_{m=1}^{M}\frac{1}{m}\left[y(m) - \log_{10}\left(\sum_{i=1}^{2}\alpha_i\beta_i^m\right)\right]^2\right\} \tag{2}$$

subject to the following constraints:

(1) $\alpha_1 + \alpha_2 = 1$ and

(2) $0 \leqslant \alpha_i \leqslant 1,\ 0 \leqslant \beta_i \leqslant 1$, for $i = 1,2$,

where M is the longest interval between two consecutive errors and $y(m)$ is the measured value, expressed logarithmically, of the error-gap distribution.

Traditional methods of model parameter estimation consist of exponential curve fitting (Gilbert 1960) and the iterative gradient-descent techniques (Chouinard *et al.* 1988). In the case of curve fitting, the parameters are estimated by determining the number of straight-line segments required to approximate the measured error-gap distribution, and then determining approximate values for the slopes of these segments. The main draw-

back to curve fitting is that it must be augmented by human judgments and so may become subjective and unreliable. Alternatively, the iterative gradient-descent method tries to find optimum solutions by performing successive corrections on parameter estimates in the direction opposite to the gradient of the cost function. According to the steepest-descent method, the updated values of the model parameters at time $n+1$ are computed using the simple recursive relationship

$$(\alpha_1(n+1), \alpha_2(n+1), \beta_1(n+1), \beta_2(n+1)) = (\alpha_1(n), \alpha_2(n), \beta_1(n), \beta_2(n)) - \lambda(n)\,\nabla E(n), \quad (3)$$

where $\lambda(n)$ indicates the step size and $\nabla E(n)$ the gradient of the cost function. The detailed expressions for the step size and cost-function gradient are presented in the Appendix.

Given that the values of $\{\alpha_1, \alpha_2, \beta_1, \beta_2\}$ have been determined, Gilbert's model parameters $\{P, h, p\}$ can be calculated as follows (Gilbert 1960):

$$h = \frac{\beta_1\beta_2}{\beta_1 - \alpha_1(\beta_1 - \beta_2)}, \quad (4)$$

$$P = \frac{(1-\beta_1)(1-\beta_2)}{1-h}, \quad (5)$$

$$p = \alpha_1(\beta_1 - \beta_2) + \frac{(1-\beta_1)(\beta_2 - h)}{1-h}. \quad (6)$$

Although the gradient-descent method converges rapidly, its simple downhill transitions can easily trap its final solution into a local optimum when multiple optima are present. To highlight the problems encountered by the gradient-descent method, we show in figure 2 the error-performance surface that results from using a typical error burst for a range of values of β_1 and β_2.

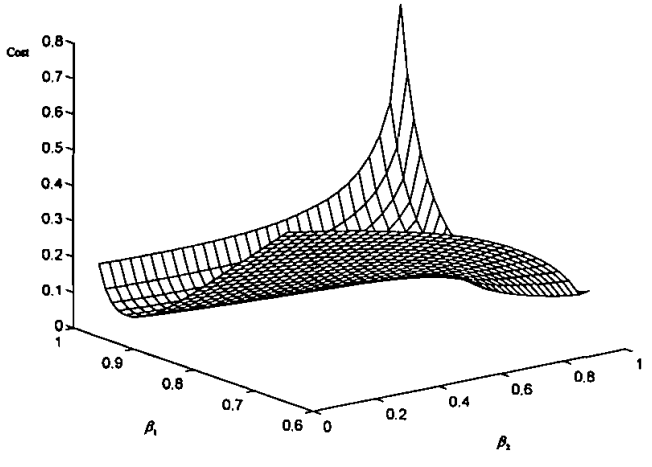

Figure 2. Cost function of burst error for a BER of 1.3% with $\alpha_1 = 0.78$ and $\alpha_2 = 0.22$.

This surface has been found difficult to optimize because the global minimum is in close proximity to a number of local minima, ridges and saddles. One possible solution is to perform iterative improvement starting from a number of initial configurations and to choose the best outcome from all those obtained. However, the computational burden could be intolerable and we would still have no guarantee of finding an optimal solution.

3. Simulated annealing for estimation

This section discusses the basic formulations of the stochastic simulated annealing algorithm for solving combinatorial optimization problems (Kirkpatrick *et al.* 1983). It has been shown that, from almost any starting point, successive iterations of the annealing algorithm would converge asymptotically to the global optimum with probability one sense (Aats and Korst 1989). The study of simulated annealing coincides with the assumption that the equilibrium condition leads to a Boltzmann state transition rule for state updates. This suggests that the relative probabilities of the two global states is determined solely by their cost difference and temperatures, and the probability of being in a given state follows a Boltzmann distribution. Viewed in this respect, temperature provides a new free parameter to help to steer the search direction and step size towards the global optimum solution. In the initial phase, a higher temperature should be used to allow random searches so that it is easier for the states to escape from local optima. The search will move towards some feasibility regions likely to contain the global optimum as the temperature gradually decreases, but escape from local optima is still possible since uphill state transitions are allowed. When the temperature is too low to move uphill, the annealing algorithm becomes a simple downhill-search algorithm and the average state should be very close to the globally optimal solution.

The proposed estimation method based on simulated annealing consists of two nested loops; the inner loop proceeds until the equilibrium condition at each temperature is satisfied, while the outer loop is terminated at a very low temperature. We summarize the relevant aspects of the standard simulated annealing algorithm here; more comprehensive accounts can be found in the paper by Tan and Chang (1996).

(i) Start with a high temperature $T_0 = 10\sigma_\infty$, where σ_∞ is the standard deviation of the costs. Also, choose an initial state $u_0 = (\alpha_1^0, \alpha_2^0, \beta_1^0, \beta_2^0)$ with the associated cost function $E(u_0)$.

(ii) Generate a new state following a random perturbation mechanism. Let ΔE denote the cost difference between the new and old states, and r a random

number uniformly distributed between 0 and 1. The new state is accepted if either $\Delta E < 0$ or the Boltzmann state transition rule $e^{-\Delta E/T} > r$ is satisfied; otherwise, the new state is rejected.

(iii) In the inner loop, for each temperature T, if the equilibrium condition has been reached at the nth iteration,

$$\sum_{i=1}^{n} \frac{\exp\{[\eta_a - E(u_i)]/T\}}{n} < \zeta,$$

then the inner loop stops and goes to step (iv); otherwise, steps (ii)–(iii) are repeated. The threshold ζ is empirically determined and η_a is the average cost of the accepted states.

(iv) Decrease the temperature according to a cooling schedule and return to step (ii) until a desired low temperature is reached.

Having a proper cooling schedule is critical for both the convergence rate and the final performance of the annealing techniques. For this reason, three different temperature schedules were considered here. In standard annealing, the control temperature is decreased according to $T(k) = \gamma T(k-1)$, where k is the index of the outer loop and the values of γ lie between 0.9 and 0.95. However, it has been proven (Aats and Korst 1989) that the necessary and sufficient condition for converging to a global optimum requires the cooling schedule to be inversely proportional to the logarithmic function of time: $T(k) = T_0/\ln(k+1)$. Unfortunately, this approach results in slow convergence because of constraints caused by the bounded variance of the Boltzmann process. To compensate for this shortcoming, Szu and Hartley (1987) proposed a fast simulated annealing using a generation mechanism based on the Cauchy distribution. The Boltzmann distribution has the same general shape as the Cauchy distribution but the latter has a fatter tail at high energies. Proceeding in this way, the cooling schedule will follow the relationship $T(k) = T_0/(k+1)$. In our earlier work (Tan and Chang 1996), we presented preliminary experimental results that substantiate the superiority of fast simulated annealing used for Gilbert's parameter estimation.

To test the validity of the proposed estimation scheme, extensive computer simulations were conducted on a Pentium-133 PC using sample error sequences with different characteristics. Two basic types of error-source model were considered: uniformly distributed random errors and error bursts. Each sample error sequence was 100 000 bits long. For all test samples, we first evaluated the measured values of $P(0^m|1)$ by computing the ratio of consecutive series of error-free bits with lengths equal to or greater than m to the total number of error bursts. Table 1 presents the results of descriptive statistics from different error sources using the simulated annealing method. For our study, the annealing parameters T_0 and ζ were empirically determined to be 5 and 1 respectively. In addition, the annealing algorithm was terminated when the temperature dropped below 0.05. Figure 3 illustrates the basic difference in the error-gap distributions between random and burst error sources, even with the same BER of 1.0%. Compared with the burst errors, the error-gap distribution of random errors tended to decay more slowly. This translates directly into smaller values of β_1 when the error process exhibits predominantly clustered trends. Viewed in this context, the value of β_1 provides an ideal framework for choosing between the random-error-correcting and burst-error-correcting codes. For purposes of comparison, we also show in figure 4 the learning curves of Gilbert's model parameter estimation using the gradient-descent method and the annealing method. Our general conclusion is that the annealing method is preferable to the gradient-descent method, but only at the expense of extra convergence time.

Table 1. Estimated descriptive statistics for different error sources using the simulated annealing method

Source type	BER (%)	α_1	α_2	β_1	β_2
Random error	0.5	0.8010	0.1990	0.9942	0.9945
	1.0	0.7998	0.2002	0.9905	0.9877
	1.5	0.7991	0.2009	0.9879	0.9783
	2.0	0.8000	0.2000	0.9793	0.9806
Burst error	0.5	0.8294	0.1706	0.8566	0.9990
	1.0	0.7989	0.2011	0.7948	0.9992
	1.5	0.7883	0.2117	0.7683	0.9994
	2.0	0.7858	0.2142	0.7330	0.9986

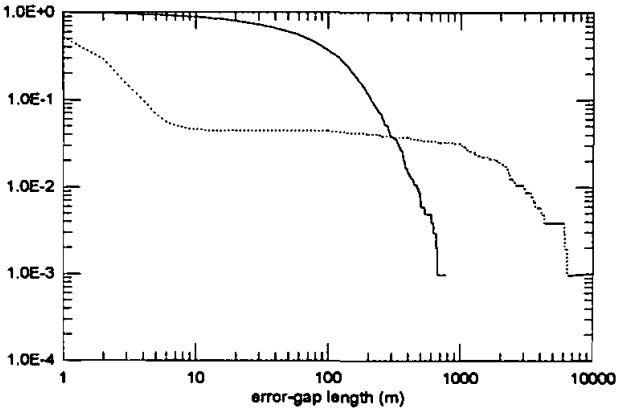

Figure 3. Experimental error-gap distributions of random error (——) and burst error (· · · · · ·) sources with a BER of 1%.

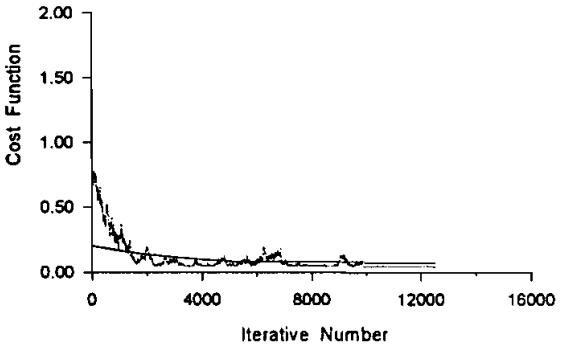

Figure 4. Learning curve for the iterative gradient-descent method (——) and the simulated annealing method (· · · · ·).

Table 2. Results of good points and costs obtained by the proposed hybrid algorithm using a burst error source of BER 0.5%

Good points found by first component		Corresponding cost	Final gradient cost
(α_1, α_2)	(β_1, β_2)	E_{SA}	$E_{gradient}$
0.6, 0.4	0.6555, 0.9987	0.5330	0.5006
0.6, 0.4	0.8350, 0.9986	0.5252	0.5407
0.7, 0.3	0.6560, 0.9988	0.2887	0.1940
0.7, 0.3	0.8460, 0.9988	0.2431	0.2452
0.8, 0.2	0.8025, 0.9989	0.0570	0.0334
0.8, 0.2	0.9050, 0.9989	0.0675	0.0533
0.9, 0.1	0.8240, 0.9992	0.2484	0.2439
0.9, 0.1	0.9190, 0.9992	0.1616	0.1607

4. Hybrid estimation algorithm

While stochastic simulated annealing is conceptually useful in converging to a global optimum, it has some limitations as far as its time-consuming optimization process is concerned. On the other hand, the selection of good starting points (Brooks and Morgan 1994) for the efficient gradient-descent algorithm can be difficult. To overcome these problems, we propose using a hybrid optimization algorithm that merges the most appealing features of these two algorithms. The first component, based on simulated annealing, automatically determines the smallest number of starting points, which are then used by the second component, a gradient-descent method, for rapid convergence to the global optimum.

The tuning of a hybrid algorithm demands both suitable use of all the knowledge available on the actual problem and a suitably designed set of experiments to find an appropriate set of parameters. In our hybrid algorithm, the first component consists of an annealing algorithm that is stopped prematurely after N_{out} temperature reductions. At each temperature, the annealing algorithm searches for an equilibrium point until the maximum number of iterations exceeds N_{in}. At each new temperature, the iteration always starts with the final equilibrium state reached at the previous temperature. On the one hand, the values of N_{in} and N_{out} should be large enough to ensure the equilibrium condition is reached and convergence occurs. On the other hand, overly large values may lead to excessive exploration of the parameter space. In our study, we found suitable values for N_{in} and N_{out} empirically and determined that 1000 and 6 respectively worked best.

The next step of the present investigation concerns selecting good starting points for use by the gradient-descent algorithm. Thus, over a grid of α_1 values containing $(0.6, 0.7, 0.8, 0.9)$, 20 different starting points for $\{\beta_1, \beta_2\}$ were examined using simulated annealing to find the two best solutions for each value of α_1. As an illustrative example, the resultant good points and their corresponding costs using burst errors of BER 0.5% are shown in table 2. With each of these starts, iterative gradient-descent routines were executed in order to choose the best solution from those obtained. The costs incurred using those starts are also given in table 2, where the best solution is shown to be 0.0334, a value near the global optimum. As presented in the discussion above, the test example has demonstrated the ability of the proposed algorithm to find global or near-global optimum solutions. Table 3 presents a comparison of

Table 3. Comparison of results from various estimation algorithms

BER (%)	Method	α_1	α_2	β_1	β_2	Final cost	Central processing unit time (s)
0.5	Gradient	0.8001	0.1999	0.8006	0.9990	0.0680	457
0.5	Annealing	0.8294	0.1706	0.8566	0.9990	0.0277	9537
0.5	Hybrid	0.8127	0.1873	0.8643	0.9990	0.0334	7710
1.3	Gradient	0.7012	0.2988	0.6030	0.9965	0.1095	165
1.3	Annealing	0.7843	0.2157	0.7983	0.9969	0.0295	5136
1.3	Hybrid	0.7833	0.2167	0.7946	0.9969	0.0284	2880

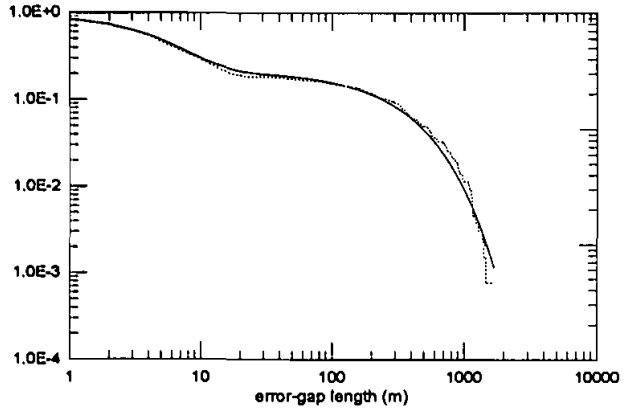

Figure 5. Experimental error-gap distribution (·····) and resulting Gilbert's model fit (——).

scores associated with various estimation algorithms. The first result shows that the gradient-descent method leads to an unsatisfactory local minimum. By contrast, the next set of results demonstrates how the annealing component of the hybrid algorithm can help to identify a far better starting point for use with the gradient-descent search. Using the parameter estimates, the resulting modelled fit for error-gap distribution is plotted in figure 5. Also shown in the figure is the experimentally measured error-gap distribution. The good agreement between them provides justification for asserting the proposed hybrid algorithm's ability to estimate Gilbert's model parameters and demonstrates its usefulness in channel characterization.

5. Conclusions

This paper has explored the benefits of a hybrid strategy combining stochastic simulated annealing and deterministic gradient-descent algorithms for use in modelling Gilbert's channels. Our method coincides with the optimization process of fitting mixtures of exponential distributions to experimental error-gap distribution. We first emphasized the importance of selecting good starting points that allow the iterative gradient-descent method to converge towards the global optimum. This task was accomplished by using simulated annealing to minimize the quadratic error betwen the measured error-gap distribution and its modelled fit. As shown, this nonlinear optimization method helps to identify the feasibility region which is thought to contain the global optimum.

Acknowledgments

This work was supported by the National Science Council, Taiwan, under grant NSC85-2221-E009-029.

Appendix

In the iterative gradient-descent method the cost-function gradient and the step size respectively are given as follows:

$$\nabla E = \left(\frac{\partial E}{\partial \alpha_1}, \frac{\partial E}{\partial \alpha_2}, \frac{\partial E}{\partial \beta_1}, \frac{\partial E}{\partial \beta_2}\right), \tag{A 1}$$

$$\lambda = \left[\sum_{i=1}^{2}\left(\frac{\partial E}{\partial \alpha_i}\right)^2 + \sum_{i=1}^{2}\left(\frac{\partial E}{\beta_i}\right)^2\right]\left[\sum_{i=1}^{2}\sum_{j=1}^{2}\left(\frac{\partial^2 E}{\partial \alpha_i \partial \alpha_j}\frac{\partial E}{\partial \alpha_i}\frac{\partial E}{\partial \alpha_j} + \frac{\partial^2 E}{\partial \alpha_i \partial \beta_i}\frac{\partial E}{\partial \alpha_i}\frac{\partial E}{\partial \beta_j} + \frac{\partial^2 E}{\partial \beta_i \partial \alpha_j}\frac{\partial E}{\partial \beta_i}\frac{\partial E}{\partial \alpha_j} + \frac{\partial^2 E}{\partial \beta_i \partial \beta_j}\frac{\partial E}{\partial \beta_i}\frac{\partial E}{\partial \beta_j}\right)\right]^{-1}, \tag{A 2}$$

where

$$\frac{\partial E}{\partial \alpha_1} = -2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m} \times \frac{[y(m) - \log_{10}(\alpha_1\beta_1^m + \alpha_2\beta_2^m)](\beta_1^m - \beta_2^m)}{\alpha_1\beta_1^m + \alpha_2\beta_2^m}, \tag{A 3}$$

$$\frac{\partial E}{\partial \alpha_2} = -\frac{\partial E}{\partial \alpha_1}, \tag{A 4}$$

$$\frac{\partial E}{\partial \beta_i} = -2\alpha_i(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m} \times \frac{[y(m) - \log_{10}(\alpha_1\beta_1^m + \alpha_2\beta_2^m)]m\beta_i^{m-1}}{\alpha_1\beta_1^m + \alpha_2\beta_2^m}, \tag{A 5}$$

$$\frac{\partial^2 E}{\partial \alpha_1 \partial \alpha_1} = 2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m} \times \frac{[y(m) - \log_{10}(\alpha_1\beta_1^m + \alpha_2\beta_2^m) + \log_{10} e](\beta_1^m - \beta_2^m)^2}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2}, \tag{A 6}$$

$$\frac{\partial^2 E}{\partial \alpha_1 \partial \alpha_2} = -\frac{\partial^2 E}{\partial \alpha_1 \partial \alpha_1}, \tag{A 7}$$

$$\frac{\partial^2 E}{\partial \alpha_2 \partial \alpha_2} = \frac{\partial^2 E}{\partial \alpha_1 \partial \alpha_1}, \tag{A 8}$$

$$\frac{\partial^2 E}{\partial \alpha_1 \partial \beta_1} = -2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m} \times \frac{[y(m) - \log_{10}(\alpha_1\beta_1^m + \alpha_2\beta_2^m)](m\beta_1^{m-1}\beta_2^m)}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2} + 2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m}\frac{(\log_{10} e)(m\alpha_1\beta_1^{m-1})(\beta_1^m - \beta_2^m)}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2}, \tag{A 9}$$

$$\frac{\partial^2 E}{\partial\alpha_2\,\partial\beta_2} = -2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m} \times \frac{[y(m) - \log_{10}(\alpha_1\beta_1^m + \alpha_2\beta_2^m)](m\beta_2^{m-1}\beta_1^m)}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2} - 2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m}\frac{(\log_{10} e)(m\alpha_2\beta_2^{m-1})(\beta_1^m - \beta_2^m)}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2} \tag{A 10}$$

$$\frac{\partial^2 E}{\partial\alpha_1\,\partial\beta_2} = 2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m} \times \frac{[y(m) - \log_{10}(\alpha_1\beta_1^m + \alpha_2\beta_2^m)](m\beta_1^m\beta_2^{m-1})}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2} + 2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m}\frac{(\log_{10} e)(m\alpha_2\beta_2^{m-1})(\beta_1^m - \beta_2^m)}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2} \tag{A 11}$$

$$\frac{\partial^2 E}{\partial\alpha_2\,\partial\beta_1} = 2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m} \times \frac{[y(m) - \log_{10}(\alpha_1\beta_1^m + \alpha_2\beta_2^m)](m\beta_1^{m-1}\beta_2^m)}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2} - 2(\log_{10} e)\sum_{m=1}^{M}\frac{1}{m}\frac{(\log_{10} e)(m\alpha_1\beta_1^{m-1})(\beta_1^m - \beta_2^m)}{(\alpha_1\beta_1^m + \alpha_2\beta_2^m)^2} \tag{A 12}$$

References

AATS, E., and KORST, J., 1989, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing* (New York: Wiley).

BROOKS, S. P., and MORGAN, B. J. T., 1994, Automatic starting point selection for function optimization. *Statistics and Computing*, **4**, 173–177.

CHOUINARD, J.-Y., LECOURS, M., and DELISLE, G. Y., 1988, Estimation of Gilbert's and Fritchman's models parameters using the gradient method for digital mobile radio channels. *IEEE Transactions on Vehicular Techology*, **37**, 158–166.

FRITCHMAN, B. D., 1967, A binary channel characterization using partitioned Markov chains. *IEEE Transactions on Information Theory*, **13**, 221–227.

GILBERT, E. N., 1960, Capacity of a burst-noise channel. *The Bell System Technical Journal*, **39**, 1253–1265.

KANAL, L. N., and SASTRY, A. R. K., 1978, Models for channels with memory and their applications to error control. *Proceedings of the IEEE*, **66**, 724–744.

KIRKPATRICK, S., GELARTT, C. D., and VECCHI, M. P., 1983, Optimization by simulated annealing. *Science*, **220**, 671–680.

POWELL, M. J. D., 1984, Some properties of the variable metric algorithms for unconstrained optimization. *Proceedings of the International Congress of Mathematicians*.

SATO, T., TOKUDA, K., KAWABE, M., and KATO, T., 1991, Simulaton of burst error models and an adaptive error control scheme for high speed data transmission over analog cellular systems. *IEEE Transactions on Vehicular Technology*, **40**, 443–452.

SZU, H., and HARTLEY, R., 1987, Fast simulated annealing. *Physics Letters A*, **122**, 157–162.

TAN, T. H., and CHANG, W. W., 1996, Estimation of Gilbert's model parameters using the simulated annealing method. *Proceedings of the Seventh IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications*, pp. 1226–1229.