Draft of paper appearing in:
%A L. Ingber
%T Very fast simulated re-annealing
%J Mathl. Comput. Modelling
%V 12
%P 967-973
%D 1989

Present Address: P.O. Box 857, McLean, VA 22101

Very Fast Simulated Re-Annealing

Lester Ingber

Physics Department, Naval Postgraduate School
Monterey, CA 93943
and
U.S. Army Concepts Analysis Agency
8120 Woodmont Avenue, Bethesda, MD 20814-2797

An algorithm is developed to statistically find the best global fit of a nonlinear non-convex cost-function over a D-dimensional space. It is argued that this algorithm permits an annealing schedule for "temperature" T decreasing exponentially in annealing-time k, $T = T_0 \exp(-c\mathrm{k}^{1/D})$. The introduction of re-annealing also permits adaptation to changing sensitivities in the multi-dimensional parameter-space. This annealing schedule is faster than fast Cauchy annealing, where $T = T_0/\mathrm{k}$, and much faster than Boltzmann annealing, where $T = T_0/\ln \mathrm{k}$. Applications are being made to fit empirical data to Lagrangians representing nonlinear Gaussian-Markovian systems.

I. Introduction

Modeling phenomena is as much a cornerstone of 20th century Science as is collection of empirical data [1]. Modeling is essential to go beyond current knowledge, to better understand new or complex phenomena. Many instances arise in essentially all fields of Science when mathematical models of the real world become tested by fitting some parameters to empirical data. Since the real world is often nonlinear and stochastic, it is not surprising that often this process involves fitting statistical, nonlinear, non-convex functional forms to data. Physical methods of ‘‘simulated annealing’’ have been found to be extremely useful tools for this purpose in a wide variety of examples.

This paper contributes to this methodology by presenting an improvement over previous algorithms. Sections II and III give a short outline of previous Boltzmann annealing (BA) and fast Cauchy fast annealing (FA) algorithms. Section IV presents the new very fast algorithm. Section V enhances this algorithm with a re-annealing modification found to be extremely useful for multi-dimensional parameter-spaces. This method will be referred to here as very fast re-annealing (VFR).

Section VI gives an outline of some applications presently using this new algorithm. Section VII gives a short discussion.

II. Boltzmann Annealing

Boltzmann annealing was essentially introduced as a Monte Carlo importance-sampling technique for doing large-dimensional path integrals arising in statistical physics problems [2]. This method was generalized to apply more generally to fitting non-convex cost-functions arising in a variety of problems, e.g., finding the optimal wiring for a densely wired computer chip [3].

The method of simulated annealing consists of three functional relationships.

1. $g(x)$: Probability density of state-space of D parameters $x = \{x^i; i = 1, D\}$.
2. $h(x)$: Probability density for acceptance of new cost-function given the just previous value.
3. $T(\mathrm{k})$: schedule of ‘‘annealing’’ the ‘‘temperature’’ T in annealing-time steps k, i.e., of changing the volatility or fluctuations of the two previous probability densities.

Based on functional form derived for many physical systems belonging to the class of Gaussian-Markovian systems, the algorithm chooses for g,

$$g(x) = (2\pi T)^{-D/2} \exp[-\Delta x^2/(2T)] \ , \tag{1}$$

where $\Delta x = x - x_0$ is the deviation of x from x_0 (usually taken to be the just-previously chosen point to test), and where T is clearly a measure of the fluctuations of the Boltzmann distribution g in the D-dimensional x-space.

The acceptance probability is based on the chances of obtaining a new state $E_{\mathrm{k}+1}$ relative to a previous state E_{k},

$$h(x) = \frac{\exp(-E_{\mathrm{k}+1}/T)}{\exp(-E_{\mathrm{k}+1}/T) + \exp(-E_{\mathrm{k}}/T)} = \frac{1}{1 + \exp(\Delta E/T)} \ , \tag{2}$$

where ΔE represents the ‘‘energy’’ difference between the present and previous values of the cost-function appropriate to the physical problem, i.e., $\Delta E = E_{\mathrm{k}+1} - E_{\mathrm{k}}$. This essentially is the Boltzmann distribution contributing to the statistical mechanical partition function of the system.

Given $g(x)$, it has been proven [4] that it suffices to obtain a global minimum of $E(x)$ if T is selected to be not faster than

$$T(\mathrm{k}) = \frac{T_0}{\ln \mathrm{k}} \tag{3}$$

For the purposes of this paper, a heuristic demonstration follows, to show that Eq. (3) will suffice to give a global minimum of $E(x)$ [5].

In order to statistically assure, i.e., requiring many trials, that any point in x-space can be sampled infinitely often in annealing-time (IOT), it suffices to prove that the products of probabilities of not generating a state x IOT for all annealing-times successive to k_0 yield zero,

$$\prod_{\mathrm{k}_0}^{\infty} (1 - g_{\mathrm{k}}) = 0 \ . \tag{4}$$

This is equivalent to

$$\sum_{\mathrm{k}_0}^{\infty} g_{\mathrm{k}} = \infty \ . \tag{5}$$

The problem then reduces to finding $T(\mathrm{k})$ to satisfy Eq. (5).

For BA, if $T(\mathrm{k})$ is selected to be Eq. (3), then Eq. (1) gives

$$\sum_{\mathrm{k}_0}^{\infty} g_{\mathrm{k}} \ge \sum_{\mathrm{k}_0}^{\infty} \exp(-\ln \mathrm{k}) = \sum_{\mathrm{k}_0}^{\infty} 1/\mathrm{k} = \infty \ . \tag{6}$$

Although there are sound physical principles underlying the choices of Eqs. (1) and (2) [2], it was noted that this method of finding the global minimum in x-space was not limited to physics examples requiring *bona fide* ‘‘temperatures’’ and ‘‘energies.’’ Rather, this methodology can be readily extended to any problem for which a reasonable probability density $h(x)$ can be formulated [3].

III. Fast Annealing

It was also noted that this methodology can be readily extended to use any reasonable generating function $g(x)$, without relying on the principles underlying the ergodic nature of statistical physics. Specifically, it was noted that the Cauchy distribution has some definite advantages over the Boltzmann form [5]. The Cauchy distribution,

$$g(x) = \frac{T}{(\Delta x^2 + T^2)^{(D+1)/2}} \ , \tag{7}$$

has a ‘‘fatter’’ tail than the Gaussian form of the Boltzmann distribution, permitting easier access to test local minima in the search for the desired global minimum.

It is instructive to note the similar corresponding heuristic demonstration, that the Cauchy $g(x)$ statistically finds a global minimum. If Eq. (3) is replaced by

$$T(\mathrm{k}) = \frac{T_0}{\mathrm{k}} \ , \tag{8}$$

then

$$\sum_{\mathrm{k}_0}^{\infty} g_{\mathrm{k}} \approx \frac{T_0}{\Delta x^{D+1}} \sum_{\mathrm{k}_0}^{\infty} \frac{1}{\mathrm{k}} = \infty \ . \tag{9}$$

Note that the ‘‘normalization’’ of g has introduced the annealing-time index k.

The method of FA is thus statistically seen to have an annealing schedule exponentially faster than the method of BA. This method has been tested in a variety of problems [5].

IV. Very Fast Annealing

In a variety of physical problems, e.g., as outlined in Sec. VI below, we have a D-dimensional parameter-space. Different parameters have different finite ranges, fixed by physical considerations, and different annealing-time-dependent sensitivities, measured by the curvature of the cost-function at local minima. BA and FA have g distributions which sample infinite ranges, and there is no provision for considering differences in each parameter-dimension, e.g., different sensitivities might require different annealing schedules. Also, there is no quick algorithm for calculating a D-dimensional Cauchy random generator.

For example, one might choose a D-product of one-dimensional Cauchy distributions, because the one-dimensional Cauchy has a few quick algorithms. This would also permit different T_0's to account different sensitivities.

$$g_{i\mathrm{k}} = \frac{T_{i0}}{\Delta x^{i2} + T^2} \ . \tag{10}$$

But then we would require an annealing schedule going as

$$T_i(\mathrm{k}) = T_0/\mathrm{k}^{1/D} \ , \tag{11}$$

which, although faster than BA, is still quite slow. E.g., consider $D = 6$ and assume a final temperature $T_f = 10^{-4} T_0$ is desired.

The above problems provide motivation for the development of a new algorithm. Consider a parameter α^i_k in dimension i generated at annealing-time k with the range

$$\alpha^i_\mathrm{k} \in [A_i, B_i] \ , \tag{12}$$

calculated with the random variable y^i,

$$\alpha^i_{\mathrm{k}+1} = \alpha^i_\mathrm{k} + y^i (B_i - A_i) \ ,$$

$$y^i \in [-1, 1] \ . \tag{13}$$

Define the generating function

$$g_T(y) = \prod_{i=1}^{D} \frac{1}{2(|y^i| + T_i) \ln(1 + 1/T_i)} \equiv \prod_{i=1}^{D} g^i_T(y^i) \ . \tag{14}$$

Its cumulative probability distribution is

$$G_T(y) = \int_{-1}^{y^1} \cdots \int_{-1}^{y^D} dy'^1 \cdots dy'^D \, g_T(y') \equiv \prod_{i=1}^{D} G^i_T(y^i) \ ,$$

$$G^i_T(y^i) = \frac{1}{2} + \frac{\mathrm{sgn}(y^i)}{2} \frac{\ln(1 + |y^i|/T_i)}{\ln(1 + 1/T_i)} \ . \tag{15}$$

y^i is generated from a u^i from the uniform distribution

$$u^i \in U[0, 1] \ ,$$

$$y^i = \mathrm{sgn}(u - \frac{1}{2}) T_i [(1 + 1/T_i)^{|2u^i - 1|} - 1] \ . \tag{16}$$

It is straightforward to calculate that for an annealing schedule for T_i

$$T_i(\mathrm{k}) = T_{0i} \exp(-c_i \mathrm{k}^{1/D}) \ , \tag{17}$$

a global minima statistically can be obtained. I.e.,

$$\sum_{k_0}^{\infty} g_k \approx \sum_{k_0}^{\infty} [\prod_{i=1}^{D} \frac{1}{2|y^i|c_i}] \frac{1}{k} = \infty \ . \tag{18}$$

It seems sensible to choose control over c_i, such that

$$T_{fi} = T_{0i} \exp(-m_i) \text{ when } k_f = \exp n_i \ ,$$

$$c_i = m_i \exp(-n_i/D) \ . \tag{19}$$

The cost-functions $\underline{L}$ we are exploring are of the form

$$h(M;\alpha) = \exp(-\underline{L}/T) \ ,$$

$$\underline{L} = L\Delta t + \frac{1}{2} \ln(2\pi\Delta t g_t^2) \ , \tag{20}$$

where L is a Lagrangian with dynamic variables $M(t)$, and parameter-coefficients α to be fit to data. g_t is the determinant of the metric, defined below. It has proven fruitful to use the same annealing schedule for this acceptance function h as used for the generating function g, i.e., Eqs. (17) and (19).

New parameters α^i_{k+1} are generated from old parameters α^i_k from

$$\alpha^i_{k+1} = \alpha^i_k + y^i(B_i - A_i) \ , \tag{21}$$

constrained by

$$\alpha^i_{k+1} \in [A_i, B_i] \ . \tag{22}$$

I.e., y^i's are generated until a set of D are obtained satisfying these constraints.

V. Re-Annealing

Whenever doing a multi-dimensional search in the course of a real-world nonlinear physical problem, inevitably one must deal with different changing sensitivities of the α^i in the search. At any given annealing-time, it seems sensible to attempt to ‘‘stretch out’’ the range over which the relatively insensitive parameters are being searched, relative to the ranges of the more sensitive parameters.

It has proven fruitful to accomplish this by periodically rescaling the annealing-time k, essentially re-annealing, every hundred or so acceptance-events, in terms of the sensitivities s_i calculated at the most current minimum value of $\underline{L}$,

$$s_i = (A_i - B_i)\partial\underline{L}/\partial\alpha^i \ . \tag{23}$$

In terms of the largest $s_i = s_{\max}$, it has proven fruitful to re-anneal by using a linear rescaling,

$$k'_i = ((\ln[(T_{i0}/T_{ik})(s_{\max}/s_i)])/c_i)^D \ . \tag{24}$$

T_{i0} is set to unity to begin the search, which is ample to span each parameter dimension.

The acceptance temperature is similarly rescaled. In addition, since the initial acceptance temperature is set equal to a trial value of $\underline{L}$, this is typically very large relative to the global minimum. Therefore, when this rescaling is performed, the initial acceptance temperature is reset to the most current minimum of $\underline{L}$, and the annealing-time associated with this temperature is set to give a new temperature equal to the lowest value of the cost-function encountered to annealing-date.

A few other twists can be added, and such searches undoubtedly will never be strictly by rote. Physical systems are so different, some experience with each one is required to develop a truly efficient algorithm.

The above two Sections define this method of very fast re-annealing (VFR).

VI. Applications

Some explanation of the problems we are solving using VFR contribute a better understanding of its value.

Many large-scale nonlinear stochastic systems can be described within the framework of Gaussian-Markovian systems. For example, (deceptively) simple Langevin rate equations describe the evolution of a set of variables M^G. The Einstein summation convention is used, whereby repeated indices in a term are to be summed over. In the midpoint Stratonovich representation [6],

$$\dot{M}^G = dM^G/dt = f^G + \hat{g}_i^G \eta^i \ ,$$

$$< \eta_{t'}^{i'} \eta_t^i >_\eta = \delta(t - t')\delta_{ii'} \ ,$$

$$i = 1, \cdots, \Xi \ ,$$

$$G = 1, \cdots, \Theta \ . \tag{25}$$

Expanded sets of equations can represent a field $M^G(r, t)$, and the discussion below generalizes as well [7].

Another mathematically equivalent representation is given by the Fokker-Planck equation, in terms of the "drifts" g^G and "diffusions" $g^{GG'}$,

$$\frac{\partial P}{\partial t} = VP + \frac{\partial(-g^G P)}{\partial M^G} + \frac{1}{2} \frac{\partial^2(g^{GG'} P)}{\partial M^G \partial M^{G'}} \ ,$$

$$g^G = f^G + \frac{1}{2} \hat{g}_i^{G'} \frac{\partial \hat{g}_i^G}{\partial M^{G'}} \ ,$$

$$g^{GG'} = \hat{g}_i^G \hat{g}_i^{G'} \ , \tag{26}$$

where the "potential" V might arise directly from the physics or by simulating the boundary conditions.

In many problems of interest, the drifts and diffusions are also parametrized. For example, these parameters can enter as expansion coefficients of polynomials describing accepted models of particular systems, e.g., modeling economic markets [8], or combat scenarios [9,10]. In combat systems, such equations appear as

$$\dot{r} = x_b^r b + y_{br}^r br + z^r \eta_r \ ,$$

$$\dot{b} = x_r^b r + y_{rb}^b rb + z^b \eta_b \ , \tag{27}$$

where the M^G are Red (r) and Blue (b) force levels, and where $\{x, y, z\}$ are parameters to be fit to data. Such modeling is essential to compare computer models to field data, or to qualify computer models to augment training, or to feed information from battalion-level computer scenarios into corps- and theater-level computer scenarios which must rely on highly aggregated models to run in real-time.

To perform fits to data, it is most useful to consider yet another mathematically equivalent representation to Eqs. (25) and (26), a Lagrangian (L) representation in a path-integral context [6]. (This reference contains many references to other published works.) The long-time probability distribution P at time $t_f = u\Delta t + t_0$, evolving from time t_0, in terms of the discrete index s, is given by

$$P = \int \cdots \int \underline{D} M \exp(-\sum_{s=0}^{u} \Delta t L_s) \ ,$$

$$\underline{D} M = g_{0_+}^{1/2} (2\pi\Delta t)^{-1/2} \prod_{s=1}^{u} g_{s_+}^{1/2} \prod_{G=1}^{\Theta} (2\pi\Delta t)^{-1/2} dM_s^G \ ,$$

$$L = \frac{1}{2} (\dot{M}^G - h^G) g_{GG'} (\dot{M}^{G'} - h^{G'}) + \frac{1}{2} h^G{}_{;G} + R/6 - V \ ,$$

$$[\cdots]_{,G} = \frac{\partial[\cdots]}{\partial M^G} \ ,$$

$$h^G = g^G - \frac{1}{2} g^{-1/2} (g^{1/2} g^{GG'})_{,G'} \ ,$$

$$g_{GG'} = (g^{GG'})^{-1} \ ,$$

$$g_s[M^G(\bar{t}_s), \bar{t}_s] = \det(g_{GG'})_s \ , \ g_{s_+} = g_s[M^G_{s+1}, \bar{t}_s] \ ,$$

$$M^G(\bar{t}_s) = \frac{1}{2} (M^G_{s+1} + M^G_s) \ , \ \dot{M}^G(\bar{t}_s) = (M^G_{s+1} - M^G_s)/\Delta t \ ,$$

$$\bar{t}_s = t_s + \Delta t/2 \ ,$$

$$h^G{}_{;G} = h^G_{,G} + \Gamma^F_{GF} h^G = g^{-1/2} (g^{1/2} h^G)_{,G} \ ,$$

$$\Gamma^F_{JK} \equiv g^{LF} [JK, L] = g^{LF} (g_{JL,K} + g_{KL,J} - g_{JK,L}) \ ,$$

$$R = g^{JL} R_{JL} = g^{JL} g^{JK} R_{FJKL} \ ,$$

$$R_{FJKL} = \frac{1}{2} (g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN} (\Gamma^M_{FK} \Gamma^N_{JL} - \Gamma^M_{FL} \Gamma^N_{JK}) \ , \tag{28}$$

Such a path-integral representation may be derived directly for many systems, ranging from nuclear physics [11], to neuroscience [12-15]. (These references for both systems contain many references to other published papers.) In nuclear physics, the parameters include coupling constants and masses of mesons. In neuroscience, the parameters include chemical and electrical synaptic parameters, obtained by averaging over millions of synapses within minicolumnar structures of hundreds of neurons.

In the combat models, even relatively simple functional drifts and diffusions give rise to Lagrangians nonlinear in their underlying parameters. Even extremely simple Lagrangians can present subtle nonlinearities [16]. In the nuclear physics and neuroscience systems, the drifts and diffusions are quite nonlinear in their underlying variables (M^G), as well as being nonlinear in their underlying parameters. (For the nuclear physics quantum-mechanical problem, cost-functions are typically constucted as a mean-square difference between scattering data and the theoretical calculation derived from L.)

As written in Eq. (28), L possesses a variational principle. The path-integral representation for the short-time probability distribution over the variable-space, also presents a natural cost-function over the parameter-space. Then, a maximum likelihood fit is achieved by taking a product of this ‘‘probability distribution,’’ a form similar to the acceptance function h above, over many replications of empirical data for the M^G’s, and searching over the parameter-space for the largest value of P, essentially the smallest value of $\underline{L}$.

It should be pointed out that, when performing the parameter-fits, we take advantage of the Riemannian invariance of Eq. (28) under nonlinear point transformations, and utilize the prepoint Itô discretization. This permits the fitting of P using a much simpler Lagrangian, albeit one that does not possess the variational principle over its variable-space [6].

$$L = \frac{1}{2}(\dot{M}^G - g^G) g_{GG'} (\dot{M}^{G'} - g^{G'}) - V \ ,$$

$$M^G(\bar{t}_s) = M_s^G \ , \ \dot{M}^G(\bar{t}_s) = (M_{s+1}^G - M_s^G)/\Delta t \ . \tag{29}$$

Note that the measure of the volatility, i.e., the scale of the covariance matrix $g_{GG'}$, can be fit using VFR because it enters nonlinearly in Eq. (20).

In all the above systems, the parameters have finite ranges, constrained by reasonable physical considerations. We are presently fitting combat systems to exercise data [10], and plan to soon fit human EEG data to a Lagrangian [15] derived from mesoscopic neocortical interactions [17], and macroscopic considerations [18].

The path-integral representation for the long-time probability distribution over the variable-space also affords another luxury. Many empirical systems currently present data which is quite noisy and incomplete, yet still require modeling efforts. In such systems, it is often possible to find approximately equal minimum values for different algebraic cost-functions, obtained by fitting Lagrangians of the short-time probability distributions over the parameter-space.

In these cases offering alternative models (Lagrangians), a calculation of the long-time path integral provides a sensitive separation of these models as they evolve in time, i.e., providing a measure of the importance of long-time correlations in the comparison of empirical data to models developed by short-time fits. This also provides an internal check on whether is it reasonable to even begin modeling a given physical system as a nonlinear multivariate Gaussian-Markovian system. Standard Monte Carlo techniques typically fail for highly nonlinear problems. Only recently has it been possible to accurately calculate the evolution of a nonlinear path integral with complex boundary conditions [19]. We are currently extending these algorithms to two dimensions, and hope to extend them even further.

Thus, the cost-function $\underline{L}$ in the form in Eq. (20) is statistically fit to data by inserting a simulated temperature $T \neq 1$ into the short-time probability distribution with Lagrangian L given in Eq. (29). The lowering of T with the appropriate annealing schedule makes $\underline{L}$ more sensitive to the depths and breadths of its minima than it might otherwise be, permitting an efficient fitting algorithm. Then the true path integral, with $T \equiv 1$, is calculated using Eq. (28) or (29), depending on the numerical algorithm used, to find the long-time evolution of the system. Details of the above parameter-fits of Lagrangians to empirical data will be published at a later date [20].

VII. Discussion

An algorithm of very fast simulated re-annealing has been developed to fit a empirical data to a theoretical cost-function over a D-dimensional parameter-space. The annealing schedule for the temperatures T_i decrease exponentially in annealing-time k, i.e., $T_i = T_{i0} \exp(-c_i \mathrm{k}^{1/D})$. Of course, the fatter the tail of the generating function, the smaller the ratio of acceptance to generated points in the fit. However, in practice, it is found that for a given generating function, this ratio is approximately constant as the fit finds a global minimum. Therefore, for a large parameter space, the efficiency of the fit is determined by the annealing schedule of the generating function.

No rigorous proofs have been given. It is expected that the obvious utility of this algorithm will motivate such proofs. However, actual fits to data are a finite process, and often even only heuristic guides to algorithms that obviously fit many classes of data are important. Heuristic arguments have been given here that this algorithm is faster than the fast Cauchy annealing, where $T_i = T_0/\mathrm{k}$, and much faster than Boltzmann annealing, where $T_i = T_0/\ln \mathrm{k}$.

References

1. M. Jammer, *The Philosophy of Quantum Mechanics*, Wiley & Sons, New York, NY, (1974).

2. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21** (6), 1087-1092 (1953).

3. S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, Optimization by simulated annealing, *Science* **220** (4598), 671-680 (1983).

4. S. Geman and D. Geman, Stochastic relaxation, Gibbs distribution and the Bayesian restoration in images, *IEEE Trans. Patt. Anal. Mac. Int.* **6** (6), 721-741 (1984).

5. H. Szu and R. Hartley, Fast simulated annealing, *Phys. Lett. A* **122** (3-4), 157-162 (1987).

6. F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions*, Reidel, Dordrecht, (1982).

7. L. Ingber, Statistical mechanics algorithm for response to targets (SMART), in *Workshop on Uncertainty and Probability in Artificial Intelligence: UC Los Angeles, 14-16 August 1985*, (Edited by P. Cheeseman), pp. 258-264, American Association for Artificial Intelligence, Menlo Park, CA, (1985).

8. L. Ingber, Statistical mechanics of nonlinear nonequilibrium financial markets, *Math. Modelling* **5** (6), 343-361 (1984).

9. L. Ingber, C^3 decision aids: Statistical mechanics application of biological intelligence, in *1987 Symposium on C^3 Research: National Defense University, Washington, DC, 16-18 June 1987*, pp. 49-57, National Defense University, Washington, DC, (1987).

10. L. Ingber, Mathematical comparison of computer models to exercise data: Comparison of JANUS(T) to National Training Center data, in *1988 JDL C^2 Symposium: Naval Postgraduate School, Monterey, CA, 7-9 June 1988*, pp. 541-549, SAIC, McLean, VA, (1988).

11. L. Ingber, Riemannian contributions to short-ranged velocity-dependent nucleon-nucleon interactions, *Phys. Rev. D* **33**, 3781-3784 (1986).

12. L. Ingber, Statistical mechanics of neocortical interactions. Dynamics of synaptic modification, *Phys. Rev. A* **28**, 395-416 (1983).

13. L. Ingber, Statistical mechanics of neocortical interactions. Derivation of short-term-memory capacity, *Phys. Rev. A* **29**, 3346-3358 (1984).

14. L. Ingber, Statistical mechanics of neocortical interactions. EEG dispersion relations, *IEEE Trans. Biomed. Eng.* **32**, 91-94 (1985).

15. L. Ingber and P.L. Nunez, Multiple scales of statistical physics of neocortex: Application to electroencephalography, *Mathl. Comput. Modelling* **13** (7), 83-95 (1990).

16. L. Ingber, Noise-induced extrema in time-dependent Ginsburg-Landau systems, *Math. Modelling* **7**, 525-528 (1986).

17. L. Ingber, Mesoscales in neocortex and in command, control and communications (C^3) systems, in *Systems with Learning and Memory Abilities: Proceedings, University of Paris 15-19 June 1987*, (Edited by J. Delacour and J.C.S. Levy), pp. 387-409, Elsevier, Amsterdam, (1988).

18. P.L. Nunez, *Electric Fields of the Brain: The Neurophysics of EEG*, Oxford University Press, New York, (1981).

19. M.F. Wehner and W.G. Wolfer, Numerical evaluation of path integral solutions to Fokker-Planck equations. III. Time and functionally dependent coefficients, *Phys. Rev. A*

35, 1795-1801 (1987).

20. L. Ingber, H. Fujio, and M.F. Wehner, Mathematical comparison of combat computer models to exercise data, *Mathl. Comput. Modelling* **15** (1), 65-90 (1991).