

Draft:

%A L. Ingber

%T Data mining and knowledge discovery via statistical mechanics in nonlinear stochastic systems

%J Mathl. Computer Modelling

%V 27

%N 3

%P 9-31

%D 1998

**Data mining and knowledge discovery
via statistical mechanics in nonlinear stochastic systems**

Lester Ingber

Lester Ingber Research, PO Box 06440, Wacker Dr PO - Sears Tower, Chicago, IL 60606-0440

and

DRW Investments LLC, Chicago Mercantile Exchange Center 30 S Wacker Dr Ste 1516, Chicago, IL 60606

ingber@ingber.com, ingber@alumni.caltech.edu

Abstract—A modern calculus of multivariate nonlinear multiplicative Gaussian-Markovian systems provides models of many complex systems faithful to their nature, e.g., by not prematurely applying quasi-linear approximations for the sole purpose of easing analysis. To handle these complex algebraic constructs, sophisticated numerical tools have been developed, e.g., methods of adaptive simulated annealing (ASA) global optimization and of path integration (PATHINT). In-depth application to three quite different complex systems have yielded some insights into the benefits to be obtained by application of these algorithms and tools, in statistical mechanical descriptions of neocortex (short-term memory and electroencephalography), financial markets (interest-rate and trading models), and combat analysis (baselining simulations to exercise data).

Keywords: simulated annealing; path integration; electroencephalography; financial markets; combat analysis

1. INTRODUCTION

Too often the management of complex systems is ill-served by not utilizing the best tools available. For example, requirements set by decision-makers often are not formulated in the same language as constructs formulated by powerful mathematical formalisms, and so the products of analyses are not properly or maximally utilized, even if and when they come close to faithfully representing the powerful intuitions they are supposed to model. In turn, even powerful mathematical constructs are ill-served, especially when dealing with multivariate nonlinear complex systems, when these formalisms are butchered into quasi-linear approximations to satisfy constraints of numerical algorithms familiar to particular analysts, but which tend to destroy the power of the intuitive constructs developed by decision-makers. These problems are present in many disciplines.

For at least a large class of systems, these problems can be bypassed by using a blend of an intuitive and powerful mathematical-physics formalism to generate “canonical momenta” indicators (CMI), which are used by AI-type rule-based models of management of complex systems. Typically, both the formalism generating the CMI and the rule-based models have quite nonlinear constructs, and they must be “trained” or fit to data subsequent to testing on “out-of-sample” data, before they can be used effectively for “real-time” production runs. To handle these fits of nonlinear models of real-world data, some generic powerful optimization codes have been developed, e.g., Adaptive Simulated Annealing (ASA) [1].

Section 2 presents the mathematical framework for modeling multivariate nonlinear multiplicative Gaussian-Markovian systems. Section 3 presents two major codes which are tools to calculate these systems and to fit them to data. Section 4 presents three major systems in which this approach has been fruitfully applied. Section 5 is the conclusion.

2. MODELING

2.1. Statistical Mechanics of Large Systems

Aggregation problems in nonlinear nonequilibrium systems typically are “solved” (accommodated) by having new entities/languages developed at these disparate scales in order to efficiently pass information back and forth. This is quite different from the nature of quasi-equilibrium quasi-linear

systems, where thermodynamic or cybernetic approaches are possible. These approaches typically fail for nonequilibrium nonlinear systems.

Many systems are aptly modeled in terms of multivariate differential rate-equations, known as Langevin equations,

$$\begin{aligned}\dot{M}^G &= f^G + \hat{g}_j^G \eta^j, (G = 1, \dots, \Lambda), (j = 1, \dots, N), \\ \dot{M}^G &= dM^G/dt, \\ \langle \eta^j(t) \rangle_{\eta} &= 0, \langle \eta^j(t), \eta^{j'}(t') \rangle_{\eta} = \delta^{jj'} \delta(t - t'),\end{aligned}\quad (1)$$

where f^G and \hat{g}_j^G are generally nonlinear functions of mesoscopic order parameters M^G , j is a microscopic index indicating the source of fluctuations, and $N \geq \Lambda$. The Einstein convention of summing over repeated indices is used. Vertical bars on an index, e.g., $|j|$, imply no sum is to be taken on repeated indices.

Via a somewhat lengthy, albeit instructive calculation, outlined in several other papers [2-4], involving an intermediate derivation of a corresponding Fokker-Planck or Schrödinger-type equation for the conditional probability distribution $P[M(t)|M(t_0)]$, the Langevin rate Eq. (1) is developed into the more useful probability distribution for M^G at long-time macroscopic time event $t = (u + 1)\theta + t_0$, in terms of a Stratonovich path-integral over mesoscopic Gaussian conditional probabilities [5-9]. Here, macroscopic variables are defined as the long-time limit of the evolving mesoscopic system.

The corresponding Schrödinger-type equation is [7,8]

$$\begin{aligned}\partial P/\partial t &= \frac{1}{2} (g^{GG'} P)_{,GG'} - (g^G P)_{,G} + V, \\ g^{GG'} &= k_T \delta^{jk} \hat{g}_j^G \hat{g}_k^{G'}, \\ g^G &= f^G + \frac{1}{2} \delta^{jk} \hat{g}_j^{G'} \hat{g}_k^G, \\ [\dots]_{,G} &= \partial[\dots]/\partial M^G.\end{aligned}\quad (2)$$

This is properly referred to as a Fokker-Planck equation when $V \equiv 0$. Note that although the partial differential Eq. (2) contains equivalent information regarding M^G as in the stochastic differential Eq. (1),

all references to j have been properly averaged over. I.e., \hat{g}_j^G in Eq. (1) is an entity with parameters in both microscopic and mesoscopic spaces, but M is a purely mesoscopic variable, and this is more clearly reflected in Eq. (2).

The path integral representation is given in terms of the Lagrangian L .

$$P[M_t|M_{t_0}]dM(t) = \int \cdots \int \underline{D}M \exp(-S)\delta[M(t_0) = M_0]\delta[M(t) = M_t],$$

$$S = k_T^{-1} \min_{t_0} \int_{t_0}^t dt' L,$$

$$\underline{D}M = \lim_{u \rightarrow \infty} \prod_{\rho=1}^{u+1} g^{1/2} \prod_G (2\pi\theta)^{-1/2} dM_\rho^G,$$

$$L(\dot{M}^G, M^G, t) = \frac{1}{2} (\dot{M}^G - h^G) g_{GG'} (\dot{M}^{G'} - h^{G'}) + \frac{1}{2} h^G_{;G} + R/6 - V,$$

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

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

$$g = \det(g_{GG'}),$$

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

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

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

$$R_{FJKL} = \frac{1}{2} (g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN} (\Gamma_{FK}^M \Gamma_{JL}^N - \Gamma_{FL}^M \Gamma_{JK}^N). \quad (3)$$

Mesoscopic variables have been defined as M^G in the Langevin and Fokker-Planck representations, in terms of their development from the microscopic system labeled by j . The Riemannian curvature term R arises from nonlinear $g_{GG'}$, which is a bona fide metric of this space [7]. Even if a stationary solution, i.e., $\dot{M}^G = 0$, is ultimately sought, a necessarily prior stochastic treatment of \dot{M}^G terms gives rise to these Riemannian “corrections.” Even for a constant metric, the term $h^G_{;G}$ contributes to L for a nonlinear

mean h^G . V may include terms such as $\sum_{T'} J_{T'G} M^G$, where the Lagrange multipliers $J_{T'G}$ are constraints on M^G , which are advantageously modeled as extrinsic sources in this representation; they too may be time-dependent. Using the variational principle, J_{TG} may also be used to constrain M^G to regions where they are empirically bound. More complicated constraints may be affixed to L using methods of optimal control theory [10]. With respect to a steady state \bar{P} , when it exists, the information gain in state P is defined by

$$\Upsilon[P] = \int \cdots \int \underline{D}M' P \ln(P/\bar{P}),$$

$$\underline{D}M' = \underline{D}M/dM_{u+1} . \quad (4)$$

In the economics literature, there appears to be sentiment to define Eq. (1) by the Ito, rather than the Stratonovich prescription. It should be noted that virtually all investigations of other physical systems, which are also continuous time models of discrete processes, conclude that the Stratonovich interpretation coincides with reality, when multiplicative noise with zero correlation time, modeled in terms of white noise η^j , is properly considered as the limit of real noise with finite correlation time [11]. The path integral succinctly demonstrates the difference between the two: The Ito prescription corresponds to the prepoint discretization of L , wherein $\theta \dot{M}(t) \rightarrow M_{\rho+1} - M_{\rho}$ and $M(t) \rightarrow M_{\rho}$. The Stratonovich prescription corresponds to the midpoint discretization of L , wherein $\theta \dot{M}(t) \rightarrow M_{\rho+1} - M_{\rho}$ and $M(t) \rightarrow \frac{1}{2}(M_{\rho+1} + M_{\rho})$. In terms of the functions appearing in the Fokker-Planck Eq. (2), the Ito prescription of the prepoint discretized Lagrangian, L_I , is relatively simple, albeit deceptively so because of its nonstandard calculus.

$$L_I(\dot{M}^G, M^G, t) = \frac{1}{2} (\dot{M}^G - g^G) g_{GG'} (\dot{M}^{G'} - g^{G'}) - V . \quad (5)$$

In the absence of a nonphenomenological microscopic theory, if the Ito prescription is proposed rather than the Stratonovich prescription, then this choice must be justified by numerical fits to data for each case considered.

There are several other advantages to Eq. (3) over Eq. (1). Extrema and most probable states of M^G , $\ll M^G \gg$, are simply derived by a variational principle, similar to conditions sought in previous studies [12]. In the Stratonovich prescription, necessary, albeit not sufficient, conditions are given by

$$\delta_G L = L_{,G} - L_{,\dot{G};t} = 0 ,$$

$$L_{,\dot{G};t} = L_{,\dot{G}G'} \dot{M}^{G'} + L_{,\dot{G}\ddot{G}'} \ddot{M}^{G'} . \quad (6)$$

For stationary states, $\dot{M}^G = 0$, and $\partial \bar{L} / \partial \bar{M}^G = 0$ defines $\ll \bar{M}^G \gg$, where the bars identify stationary variables; in this case, the macroscopic variables are equal to their mesoscopic counterparts. [Note that \bar{L} is *not* the stationary solution of the system, e.g., to Eq. (2) with $\partial P / \partial t = 0$. However, in some cases [13], \bar{L} is a definite aid to finding such stationary states.] Many times only properties of stationary states are examined, but here a temporal dependence is included. E.g., the \dot{M}^G terms in L permit steady states and their fluctuations to be investigated in a nonequilibrium context. Note that Eq. (6) must be derived from the path integral, Eq. (3), which is at least one reason to justify its development.

2.2. Algebraic Complexity Yields Simple Intuitive Results

It must be emphasized that the output need not be confined to complex algebraic forms or tables of numbers. Because L possesses a variational principle, sets of contour graphs, at different long-time epochs of the path-integral of P over its variables at all intermediate times, give a visually intuitive and accurate decision-aid to view the dynamic evolution of the scenario. For example, this Lagrangian approach permits a quantitative assessment of concepts usually only loosely defined.

$$\text{“Momentum”} = \Pi^G = \frac{\partial L}{\partial(\partial M^G / \partial t)} ,$$

$$\text{“Mass”} g_{GG'} = \frac{\partial^2 L}{\partial(\partial M^G / \partial t) \partial(\partial M^{G'} / \partial t)} ,$$

$$\text{“Force”} = \frac{\partial L}{\partial M^G} ,$$

$$\text{“} F = ma \text{”}: \delta L = 0 = \frac{\partial L}{\partial M^G} - \frac{\partial}{\partial t} \frac{\partial L}{\partial(\partial M^G / \partial t)} , \quad (7)$$

where M^G are the variables and L is the Lagrangian. These physical entities provide another form of intuitive, but quantitatively precise, presentation of these analyses. For example, daily newspapers use this terminology to discuss the movement of security prices. In this paper, the Π^G serve as canonical momenta indicators (CMI) for these systems.

3. TOOLS

3.1. Adaptive Simulated Annealing (ASA)

ASA [14] fits short-time probability distributions to observed data, using a maximum likelihood technique on the Lagrangian. This algorithm has been developed to fit observed data to a theoretical cost function over a D -dimensional parameter space [14], adapting for varying sensitivities of parameters during the fit. The ASA code can be obtained at no charge, via WWW from <http://www.ingber.com/> or via FTP from <ftp.ingber.com>.

3.1.1. General description

Simulated annealing (SA) was developed in 1983 to deal with highly nonlinear problems [15], as an extension of a Monte-Carlo importance-sampling technique developed in 1953 for chemical physics problems. It helps to visualize the problems presented by such complex systems as a geographical terrain. For example, consider a mountain range, with two “parameters,” e.g., along the North–South and East–West directions. We wish to find the lowest valley in this terrain. SA approaches this problem similar to using a bouncing ball that can bounce over mountains from valley to valley. We start at a high “temperature,” where the temperature is an SA parameter that mimics the effect of a fast moving particle in a hot object like a hot molten metal, thereby permitting the ball to make very high bounces and being able to bounce over any mountain to access any valley, given enough bounces. As the temperature is made relatively colder, the ball cannot bounce so high, and it also can settle to become trapped in relatively smaller ranges of valleys.

We imagine that our mountain range is aptly described by a “cost function.” We define probability distributions of the two directional parameters, called generating distributions since they generate possible valleys or states we are to explore. We define another distribution, called the acceptance distribution, which depends on the difference of cost functions of the present generated valley we are to explore and the last saved lowest valley. The acceptance distribution decides probabilistically whether to stay in a new lower valley or to bounce out of it. All the generating and acceptance distributions depend on temperatures.

In 1984 [16], it was established that SA possessed a proof that, by carefully controlling the rates of cooling of temperatures, it could statistically find the best minimum, e.g., the lowest valley of our example above. This was good news for people trying to solve hard problems which could not be solved by other algorithms. The bad news was that the guarantee was only good if they were willing to run SA forever. In 1987, a method of fast annealing (FA) was developed [17], which permitted lowering the temperature exponentially faster, thereby statistically guaranteeing that the minimum could be found in some finite time. However, that time still could be quite long. Shortly thereafter, in 1987 the author developed Very Fast Simulated Reannealing (VFSR) [14], now called Adaptive Simulated Annealing (ASA), which is exponentially faster than FA.

ASA has been applied to many problems by many people in many disciplines [18-20]. The feedback of many users regularly scrutinizing the source code ensures its soundness as it becomes more flexible and powerful.

3.1.2. Mathematical outline

ASA considers a parameter α_k^i in dimension i generated at annealing-time k with the range

$$\alpha_k^i \in [A_i, B_i], \quad (8)$$

calculated with the random variable y^i ,

$$\alpha_{k+1}^i = \alpha_k^i + y^i(B_i - A_i),$$

$$y^i \in [-1, 1]. \quad (9)$$

The generating function $g_T(y)$ is defined,

$$g_T(y) = \prod_{i=1}^D \frac{1}{2(|y^i| + T_i) \ln(1 + 1/T_i)} \equiv \prod_{i=1}^D g_T^i(y^i), \quad (10)$$

where the subscript i on T_i specifies the parameter index, and the k -dependence in $T_i(k)$ for the annealing schedule has been dropped for brevity. Its cumulative probability distribution is

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

$$G_T^i(y^i) = \frac{1}{2} + \frac{\text{sgn}(y^i)}{2} \frac{\ln(1 + |y^i|/T_i)}{\ln(1 + 1/T_i)} . \quad (11)$$

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

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

$$y^i = \text{sgn}(u^i - \frac{1}{2}) T_i [(1 + 1/T_i)^{|2u^i-1|} - 1] . \quad (12)$$

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

$$T_i(k) = T_{0i} \exp(-c_i k^{1/D}) , \quad (13)$$

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

$$\sum_{k_0}^{\infty} g_k \approx \sum_{k_0}^{\infty} \left[\prod_{i=1}^D \frac{1}{2|y^i|c_i} \right] \frac{1}{k} = \infty . \quad (14)$$

Control can be taken over c_i , such that

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

$$c_i = m_i \exp(-n_i/D) , \quad (15)$$

where m_i and n_i can be considered “free” parameters to help tune ASA for specific problems.

3.1.3. Reannealing

ASA has over 100 OPTIONS available for tuning. A few important ones are described here.

Whenever doing a multi-dimensional search in the course of a complex nonlinear physical problem, inevitably one must deal with different changing sensitivities of the α^i in the search. At any given annealing-time, the range over which the relatively insensitive parameters are being searched can be “stretched out” relative to the ranges of the more sensitive parameters. This can be accomplished by periodically rescaling the annealing-time k , essentially reannealing, every hundred or so acceptance-events (or at some user-defined modulus of the number of accepted or generated states), in terms of the sensitivities s_i calculated at the most current minimum value of the cost function, C ,

$$s_i = \partial C / \partial \alpha^i . \quad (16)$$

In terms of the largest $s_i = s_{\max}$, a default rescaling is performed for each k_i of each parameter dimension, whereby a new index k'_i is calculated from each k_i ,

$$\begin{aligned}
 k_i &\rightarrow k'_i , \\
 T'_{ik'} &= T_{ik}(s_{\max}/s_i) , \\
 k'_i &= (\ln(T_{i0}/T'_{ik'})/c_i)^D .
 \end{aligned} \tag{17}$$

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

3.1.4. Quenching

Another adaptive feature of ASA is its ability to perform quenching in a methodical fashion. This is applied by noting that the temperature schedule above can be redefined as

$$\begin{aligned}
 T_i(k_i) &= T_{0i} \exp(-c_i k_i^{Q_i/D}) , \\
 c_i &= m_i \exp(-n_i Q_i/D) ,
 \end{aligned} \tag{18}$$

in terms of the “quenching factor” Q_i . The sampling proof fails if $Q_i > 1$ as

$$\sum_k \prod^D 1/k^{Q_i/D} = \sum_k 1/k^{Q_i} < \infty . \tag{19}$$

This simple calculation shows how the “curse of dimensionality” arises, and also gives a possible way of living with this disease. In ASA, the influence of large dimensions becomes clearly focussed on the exponential of the power of k being $1/D$, as the annealing required to properly sample the space becomes prohibitively slow. So, if resources cannot be committed to properly sample the space, then for some systems perhaps the next best procedure may be to turn on quenching, whereby Q_i can become on the order of the size of number of dimensions.

The scale of the power of $1/D$ temperature schedule used for the acceptance function can be altered in a similar fashion. However, this does not affect the annealing proof of ASA, and so this may be used without damaging the sampling property.

3.1.5. Self optimization

If not much information is known about a particular system, if the ASA defaults do not seem to work very well, and if after a bit of experimentation it still is not clear how to select values for some of the ASA OPTIONS, then the SELF_OPTIMIZE OPTIONS can be very useful. This sets up a top level search on the ASA OPTIONS themselves, using criteria of the system as its own cost function, e.g., the best attained optimal value of the system's cost function (the cost function for the actual problem to be solved) for each given set of top level OPTIONS, or the number of generated states required to reach a given value of the system's cost function, etc. Since this can consume a lot of CPU resources, it is recommended that only a few ASA OPTIONS and a scaled down system cost function or system data be selected for this OPTIONS.

Even if good results are being attained by ASA, SELF_OPTIMIZE can be used to find a more efficient set of ASA OPTIONS. Self optimization of such parameters can be very useful for production runs of complex systems.

3.1.6. Parallel code

It is quite difficult to directly parallelize an SA algorithm [19], e.g., without incurring very restrictive constraints on temperature schedules [21], or violating an associated sampling proof [22]. However, the fat tail of ASA permits parallelization of developing generated states prior to subjecting them to the acceptance test [23]. The ASA_PARALLEL OPTIONS provide parameters to easily parallelize the code, using various implementations, e.g., PVM, shared memory, etc.

The scale of parallelization afforded by ASA, without violating its sampling proof, is given by a typical ratio of the number of generated to accepted states. Several experts in parallelization suggest that massive parallelization e.g., on the order of the human brain, may take place quite far into the future, that this might be somewhat less useful for many applications than previously thought, and that most useful scales of parallelization might be on scales of order 10's to many 1000's. Depending on the specific problem, such scales are common in ASA optimization, and the ASA code can implement such parallelization.

3.1.7. Widespread use and comparisons

The file http://www.ingber.com/MISC.DIR/asa_examples has several templates of “toy” test problems, especially illustrating how tuning can increase the efficiency of ASA by orders of magnitude.

The file http://www.ingber.com/asa_papers has references to the the use of ASA by some other researchers, e.g., in studies ranging from: comparisons among SA algorithms and between ASA and genetic algorithms, tabu and hillclimbing [24-28], to molecular models [29], to imaging [30], to neural networks [31], to econometrics [32], to geophysical inversion [33], to wide-spread use in financial institutions [18], etc.

3.2. PATHINT

Another code, PATHINT, develops the long-time probability distribution from the Lagrangian fit by the first code. A robust and accurate histogram-based (non-Monte Carlo) path-integral algorithm to calculate the long-time probability distribution has been developed to handle nonlinear Lagrangians [34-40],

The histogram procedure recognizes that the distribution can be numerically approximated to a high degree of accuracy as sum of rectangles at points M_i of height P_i and width ΔM_i . For convenience, just consider a one-dimensional system. The above path-integral representation can be rewritten, for each of its intermediate integrals, as

$$\begin{aligned}
 P(M; t + \Delta t) &= \int dM' [g_s^{1/2} (2\pi\Delta t)^{-1/2} \exp(-L_s\Delta t)] P(M'; t) \\
 &= \int dM' G(M, M'; \Delta t) P(M'; t) , \\
 P(M; t) &= \sum_{i=1}^N \pi(M - M_i) P_i(t) \\
 \pi(M - M_i) &= \begin{cases} 0, & (M_i - \frac{1}{2} \Delta M_{i-1}) \leq M \leq (M_i + \frac{1}{2} \Delta M_i) , \\ 1, & \text{otherwise} , \end{cases} \tag{20}
 \end{aligned}$$

which yields

$$P_i(t + \Delta t) = T_{ij}(\Delta t) P_j(t) ,$$

$$T_{ij}(\Delta t) = \frac{2}{\Delta M_{i-1} + \Delta M_i} \int_{M_i - \Delta M_{i-1}/2}^{M_i + \Delta M_i/2} dM \int_{M_j - \Delta M_{j-1}/2}^{M_j + \Delta M_j/2} dM' G(M, M'; \Delta t). \quad (21)$$

T_{ij} is a banded matrix representing the Gaussian nature of the short-time probability centered about the (varying) drift.

Fitting data with the short-time probability distribution, effectively using an integral over this epoch, permits the use of coarser meshes than the corresponding stochastic differential equation. The coarser resolution is appropriate, typically required, for numerical solution of the time-dependent path-integral: By considering the contributions to the first and second moments of ΔM^G for small time slices θ , conditions on the time and variable meshes can be derived [34]. The time slice essentially is determined by $\theta \leq \bar{L}^{-1}$, where \bar{L} is the “static” Lagrangian with $dM^G/dt = 0$, throughout the ranges of M^G giving the most important contributions to the probability distribution P . The variable mesh, a function of M^G , is optimally chosen such that ΔM^G is measured by the covariance $g^{GG'}$, or $\Delta M^G \sim (g^{GG'} \theta)^{1/2}$.

4. APPLICATIONS

4.1. Statistical Mechanics of Neocortical Interactions (SMNI)

Since the early 1980's, a series of papers on the statistical mechanics of neocortical interactions (SMNI) has been developed to model columns and regions of neocortex, spanning mm to cm of tissue [4,13,23,39-56]. This model was the first physical application of a nonlinear multivariate calculus developed by other mathematical physicists in the late 1970's [57,58].

4.1.1. SMNI development

4.1.1.1. Statistical aggregation

SMNI studies have detailed that the predominant physics of short-term memory and of (short-fiber contribution to) EEG phenomena takes place in a narrow “parabolic trough” in M^G space, roughly along a diagonal line [44]. Here, G represents E or I , M^E represents contributions to columnar firing from excitatory neurons, and M^I represents contributions to columnar firing from inhibitory neurons. About 100 neurons comprise a minicolumn (twice that number in visual cortex); about 1000 minicolumns

comprise a macrocolumn. A mesocolumn is developed by SMNI to reflect the convergence of short-ranged (as well as long-ranged) interactions of macrocolumnar input on minicolumnar structures, in terms of synaptic interactions taking place among neurons (about 10,000 synapses per neuron).

In a prepoint discretization, where the Riemannian geometry is not explicit (but calculated in the first SMNI papers), the distributions of neuronal activities p_{σ_i} is developed into distributions for activity under an electrode site P in terms of a Lagrangian \underline{L} and threshold functions F^G ,

$$P = \prod_G P^G [M^G(r; t + \tau) | M^{\bar{G}}(r'; t)] = \sum_{\sigma_j} \delta \left(\sum_{j^E} \sigma_j - M^E(r; t + \tau) \right) \delta \left(\sum_{j^I} \sigma_j - M^I(r; t + \tau) \right) \prod_j^N P_{\sigma_j}$$

$$\approx \prod_G (2\pi\tau g^{GG})^{-1/2} \exp(-N\tau \underline{L}^G) = (2\pi\tau)^{-1/2} g^{1/2} \exp(-N\tau \underline{L}),$$

$$\underline{L} = \underline{T} - \underline{V}, \quad \underline{T} = (2N)^{-1} (\dot{M}^G - g^G) g_{GG'} (\dot{M}^{G'} - g^{G'}),$$

$$g^G = -\tau^{-1} (M^G + N^G \tanh F^G), \quad g^{GG'} = (g_{GG'})^{-1} = \delta_G^{G'} \tau^{-1} N^G \operatorname{sech}^2 F^G, \quad g = \det(g_{GG'}),$$

$$F^G = \frac{V^G - v_{G'}^{|G|} T_{G'}^{|G|}}{(\pi[(v_{G'}^{|G|})^2 + (\phi_{G'}^{|G|})^2] T_{G'}^{|G|})^{1/2}},$$

$$T_{G'}^{|G|} = a_{G'}^{|G|} N^{G'} + \frac{1}{2} A_{G'}^{|G|} M^{G'} + a_{G'}^{\dagger|G|} N^{\dagger G'} + \frac{1}{2} A_{G'}^{\dagger|G|} M^{\dagger G'} + a_{G'}^{\ddagger|G|} N^{\ddagger G'} + \frac{1}{2} A_{G'}^{\ddagger|G|} M^{\ddagger G'},$$

$$a_{G'}^{\dagger G} = \frac{1}{2} A_{G'}^{\dagger G} + B_{G'}^{\dagger G}, \quad A_E^{\dagger I} = A_I^{\dagger E} = A_I^{\dagger I} = B_E^{\dagger I} = B_I^{\dagger E} = B_I^{\dagger I} = 0, \quad a_E^{\ddagger E} = \frac{1}{2} A_E^{\ddagger E} + B_E^{\ddagger E}, \quad (22)$$

where no sum is taken over repeated $|G|$, $A_{G'}^G$ and $B_{G'}^G$ are macrocolumnar-averaged interneuronal synaptic efficacies, $v_{G'}^G$ and $\phi_{G'}^G$ are averaged means and variances of contributions to neuronal electric polarizations, N^G are the numbers of excitatory and inhibitory neurons per minicolumn, and the variables associated with M^G , $M^{\dagger G}$ and $M^{\ddagger G}$ relate to multiple scales of activities from minicolumns, between minicolumns within regions, and across regions, resp. The nearest-neighbor interactions \underline{V} can be modeled in greater detail by a stochastic mesoscopic neural network [23]. The SMNI papers give more detail on this derivation.

In terms of the above variables, an energy or Hamiltonian density H can be defined,

$$\underline{H} = \underline{T} + \underline{V}, \quad (23)$$

in terms of the M^G and Π^G variables, and the path integral is now defined over all the $\underline{D}M^G$ as well as over the $\underline{D}\Pi^G$ variables.

A mechanical-analog model the string model, is derived explicitly for neocortical interactions using SMNI [51]. In addition to providing overlap with current EEG paradigms, this defines a probability distribution of firing activity, which can be used to further investigate the existence of other nonlinear phenomena, e.g., bifurcations or chaotic behavior, in brain states.

4.1.1.2. Short-term memory

Previous SMNI studies have detailed that maximal numbers of attractors lie within the physical firing space of M^G , consistent with experimentally observed capacities of auditory and visual short-term memory (STM), when a “centering” mechanism is enforced by shifting background conductivities of synaptic interactions, consistent with experimental observations under conditions of selective attention [13,39,40,44,59]. This leads to an effect of having all attractors of the short-time distribution lie along a diagonal line in M^G space, effectively defining a narrow parabolic trough containing these most likely firing states. This essentially collapses the 2 dimensional M^G space down to a 1 dimensional space of most importance.

Thus, the predominant physics of short-term memory and of (short-fiber contribution to) EEG phenomena takes place in a narrow “parabolic trough” in M^G space, roughly along a diagonal line [44].

The object of interest within a short refractory time, τ , approximately 5 to 10 msec, is the Lagrangian \underline{L} for a mesocolumn. $\tau\underline{L}$ can vary by as much as a factor of 10^5 from the highest peak to the lowest valley in M^G space. Therefore, it is reasonable to assume that a single independent firing variable might offer a crude description of this physics. Furthermore, the scalp potential Φ can be considered to be a function of this firing variable. (Here, “potential” refers to the electric potential, not the potential term in the Lagrangian.) In an abbreviated notation subscripting the time-dependence,

$$\Phi_t - \ll \Phi \gg = \Phi(M_t^E, M_t^I) \approx a(M_t^E - \ll M^E \gg) + b(M_t^I - \ll M^I \gg), \quad (24)$$

where a and b are constants, and $\ll \Phi \gg$ and $\ll M^G \gg$ represent typical minima in the trough. In the context of fitting data to the dynamic variables, there are three effective constants, $\{ a, b, \phi \}$,

$$\Phi_t - \phi = aM_t^E + bM_t^I. \quad (25)$$

The mesoscopic probability distributions, P , are scaled and aggregated over this columnar firing space to obtain the macroscopic probability distribution over the scalp-potential space:

$$P_{\Phi}[\Phi] = \int dM^E dM^I P[M^E, M^I] \delta[\Phi - \Phi'(M^E, M^I)] . \quad (26)$$

The parabolic trough described above justifies a form

$$P_{\Phi} = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{\Delta t}{2\sigma^2} \int dx L_{\Phi}\right) ,$$

$$L_{\Phi} = \frac{\alpha}{2} |\partial\Phi/\partial t|^2 + \frac{\beta}{2} |\partial\Phi/\partial x|^2 + \frac{\gamma}{2} |\Phi|^2 + F(\Phi) , \quad (27)$$

where $F(\Phi)$ contains nonlinearities away from the trough, σ^2 is on the order of N given the derivation of L_{Φ} above, and the integral over x is taken over the spatial region of interest. In general, there also will be terms linear in $\partial\Phi/\partial t$ and in $\partial\Phi/\partial x$.

4.1.1.3. EEG development

Previous calculations of EEG phenomena [45], show that the short-fiber contribution to the α frequency and the movement of attention across the visual field are consistent with the assumption that the EEG physics is derived from an average over the fluctuations of the system, e.g., represented by σ in the above equation. I.e., this is described by the Euler-Lagrange equations derived from the variational principle possessed by L_{Φ} (essentially the counterpart to force equals mass times acceleration), more properly by the ‘‘midpoint-discretized’’ Feynman L_{Φ} , with its Riemannian terms [42,43,50], Hence, the variational principle applies,

$$0 = \frac{\partial}{\partial t} \frac{\partial L_{\Phi}}{\partial(\partial\Phi/\partial t)} + \frac{\partial}{\partial x} \frac{\partial L_{\Phi}}{\partial(\partial\Phi/\partial x)} - \frac{\partial L_{\Phi}}{\partial\Phi} . \quad (28)$$

The result is

$$\alpha \frac{\partial^2 \Phi}{\partial t^2} + \beta \frac{\partial^2 \Phi}{\partial x^2} + \gamma \Phi - \frac{\partial F}{\partial \Phi} = 0 . \quad (29)$$

If there exist regions in neocortical parameter space such that $\beta/\alpha = -c^2$, $\gamma/\alpha = \omega_0^2$,

$$\frac{1}{\alpha} \frac{\partial F}{\partial \Phi} = -\Phi f(\Phi) , \quad (30)$$

and x is taken to be one-dimensional, then a nonlinear string is achieved. Terms linear in $\partial\Phi/\partial t$ and in

$\partial\Phi/\partial x$ in L_{Φ} can make other contributions, e.g., giving rise to damping terms.

The SMNI calculations are of minicolumnar interactions among hundreds of neurons, within a macrocolumnar extent of hundreds of thousands of neurons. Such interactions take place on time scales of several τ , where τ is on the order of 10 msec (of the order of time constants of cortical pyramidal cells). This also is the observed time scale of the dynamics of STM. SMNI hypothesizes that columnar interactions within and/or between regions containing many millions of neurons are responsible for these phenomena at time scales of several seconds. That is, the nonlinear evolution at finer temporal scales gives a base of support for the phenomena observed at the coarser temporal scales, e.g., by establishing mesoscopic attractors at many macrocolumnar spatial locations to process patterns in larger regions.

In addition to describing nonlinear attractors of short-term memory [13,39,44], ASA was used to explicitly fit the actual highly nonlinear probability distributions to EEG data [4,55,56]. SMNI proposes that models to be fitted to data include models of activity under each electrode, e.g., due to short-ranged neuronal fibers, as well as models of activity across electrodes, e.g., due to long-ranged fibers. These influences can be disentangled by SMNI fits aided by the use of CMI.

The CMI are more sensitive measures than the energy density, effectively the square of the CMI, or the information which also effectively is in terms of the square of the CMI (essentially integrals over quantities proportional to the energy times a factor of an exponential including the energy as an argument). This is even more important when dealing with oscillatory variables.

4.1.2. EEG data

The 1996 project used evoked potential (EP) EEG data from a multi-electrode array under a variety of conditions, collected at several centers in the United States, sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) project. The earlier 1991 study used only averaged EP data [60].

Each set of results is presented with 6 figures, labeled as [{alcoholic | control}, {stimulus 1 | match | no-match}, subject, {potential | momenta}], abbreviated to {a | c}_{1 | m | n}_subject.{pot | mom} where match or no-match was performed for stimulus 2 after 3.2 sec of a presentation of stimulus 1 [60]. Data includes 10 trials of 69 epochs each between 150 and 400 msec after presentation. For each subjects run, after fitting 28 parameters with ASA, epoch by epoch averages are developed of the raw data and of the

multivariate SMNI CMI. It was noted that much poorer fits were achieved when the centering mechanism was turned off and the denominators in F^G were set to constants, confirming the importance of using the full SMNI model. All stimuli were presented for 300 msec. Note that the subject number also includes the {alcoholic | control} tag, but this tag was added just to aid sorting of files (as there are contribution from co2 and co3 subjects). Each figure contains graphs superimposed for 6 electrode sites (out of 64 in the data) which have been modeled by SMNI using a circuitry of frontal sites (F3 and F4) feeding temporal (sides of head T7 and T8) and parietal (top of head P7 and P8) sites, where odd-numbered (even-numbered) sites refer to the left (right) brain. A time delay of one epoch of 3.906 msec was taken for long-ranged fiber communications between $F3 \rightarrow T7$, $F4 \rightarrow T8$, $T7 \leftarrow T8$, $T7 \rightarrow P7$, $T8 \rightarrow P8$, and $P7 \leftarrow P8$. A time delay of two epochs was taken for $F3 \rightarrow P7$ and $F4 \rightarrow P8$.

4.1.3. Testing Data

When the parameters of a theory of a physical system possess clear relationships to observed physical entities, and the theory fits experimental phenomenon while the parameters stay within experimentally determined ranges of these entities, then generally it is conceded that the theory and its parameters have passed a reasonable test. It is argued that this is the case for SMNI and its parameters, and this approach sufficed for the first study of the present data [55], just as SMNI also has been tested in previous papers.

When a model of a physical system has a relatively phenomenological nature then often such a model is best tested by first “training” its parameters on one set of data, then seeing to what degree the same parameters can be used to match the model to out-of-sample “testing” data. For example, this was performed for the statistical mechanics of financial markets (SMFM) project [2,3,61], applied to trading models [62,63]. The SMFM projects similarly use ASA and the algebra presented here for this SMNI project.

In the present project, there exists barely enough data to additionally test SMNI in this training versus testing methodology. That is, when first examining the data, it was decided to try to find sets of data from at least 10 control and 10 alcoholic subjects, each set containing at least 10 runs for each of the 3 experimental paradigms, as reported in a previous paper [55]. When reviewing this data, e.g., for the example of the one alcoholic and the one control subject which were illustrated in graphs in that previous paper, it was determined that there exists 10 additional sets of data for each subject for each paradigm,

except for the c_n case of the no-match paradigm for the control subject where only 5 additional out-of-sample runs exist. For this latter case, to keep the number of runs sampled consistent across all sets of data, e.g., to keep the relative amplitudes of fluctuations reasonably meaningful, 5 runs of the previous testing set were joined with the 5 runs of the present training set to fill out the data sets required for this study.

4.1.4. Results

The utility of the CMI in such a system can be seen in Figure 1, from a recent study fitting SMNI to EEG data [56]. Neither the energy or the information give details of the components as do the CMI. The information and energy densities are calculated and printed out after fits to data, along with the CMI.

Figure 1

Similar results are seen for other subjects. Sixty PostScript figures, each figure containing four six-electrode graphs of potential and CMI for testing and training data sets, for each of the can be retrieved via WWW from http://www.ingber.com/MISC.DIR/smni97_eeg_cmi.tar.Z, or as file smni97_eeg_cmi.tar.Z via FTP from ftp.ingber.com in the MISC.DIR directory.

4.1.5. Chaos

Given the context of studies in complex nonlinear systems [64], the question can be asked: What if EEG has chaotic mechanisms that overshadow the above stochastic considerations? The real issue is whether the scatter in data can be distinguished between being due to noise or chaos [65]. In this regard, several studies have been proposed with regard to comparing chaos to simple filtered (colored) noise [64,66]. Since the existence of multiplicative noise in neocortical interactions has been derived, then the previous references must be generalized, and further investigation is required to decide whether EEG scatter can be distinguished from multiplicative noise.

A recent study with realistic EEG wave equations using PATHINT strongly suggests that if chaos exists in a deterministic limit, it does not survive in macroscopic stochastic neocortex [67]. I.e., it is important to include stochastic aspects, as arise from the statistics of synaptic and columnar interactions,

in any realistic description of macroscopic neocortex.

4.2. Statistical Mechanics of Financial Markets

4.2.1. Random walk model

The use of Brownian motion as a model for financial systems is generally attributed to Bachelier [68], though he incorrectly intuited that the noise scaled linearly instead of as the square root relative to the random log-price variable. Einstein is generally credited with using the correct mathematical description in a larger physical context of statistical systems. However, several studies imply that changing prices of many markets do not follow a random walk, that they may have long-term dependences in price correlations, and that they may not be efficient in quickly arbitraging new information [69-71]. A random walk for returns, rate of change of prices over prices, is described by a Langevin equation with simple additive noise η , typically representing the continual random influx of information into the market.

$$\dot{\Gamma} = -\gamma_1 + \gamma_2 \eta ,$$

$$\dot{\Gamma} = d\Gamma/dt ,$$

$$\langle \eta(t) \rangle_{\eta} = 0 , \langle \eta(t), \eta(t') \rangle_{\eta} = \delta(t - t') , \quad (31)$$

where γ_1 and γ_2 are constants, and Γ is the logarithm of (scaled) price. Price, although the most dramatic observable, may not be the only appropriate dependent variable or order parameter for the system of markets [72]. This possibility has also been called the “semistrong form of the efficient market hypothesis” [69].

The generalization of this approach to include multivariate nonlinear nonequilibrium markets led to a model of statistical mechanics of financial markets (SMFM) [2].

4.2.2. Application to Term-Structure Models

An example of application of this formalism was to a term structure model developed for interest rates [3,61]. This model was developed with some sound economic theory, but the methodology presented here of using the methods of modern calculus, e.g., using the path-integral representation to

define a maximum likelihood cost function has not yet been recognized by this community [73].

The pioneering Brennan-Schwartz (BS) model [74,75] was used to illustrate how to numerically implement this methodology [3,61]. Since “real time” is used in this model, the variable t will be used instead of Θ .

The BS model is summarized by:

$$dr = [a_1 + b_1(l - r)]dt + r\sigma_1 dz_1 ,$$

$$dl = [l(a_2 + b_2r + c_2l)]dt + l\sigma_2 dz_2 ,$$

$$\langle dz_i \rangle = 0 , i = \{1, 2\} ,$$

$$\langle dz_i(t) dz_j(t') \rangle = dt \delta(t - t') , i = j ,$$

$$\langle dz_i(t) dz_j(t') \rangle = \rho dt \delta(t - t') , i \neq j ,$$

$$\delta(t - t') = \begin{cases} 0 , & t \neq t' , \\ 1 , & t = t' , \end{cases} \quad (32)$$

where $\langle . \rangle$ denotes expectations.

These can be rewritten as Langevin equations (in the Itô prepoint discretization)

$$dr/dt = a_1 + b_1(l - r) + \sigma_1 r (\gamma^+ n_1 + \text{sgn} \rho \gamma^- n_2) ,$$

$$dl/dt = l(a_2 + b_2r + c_2l) + \sigma_2 l (\text{sgn} \rho \gamma^- n_1 + \gamma^+ n_2) ,$$

$$\gamma^\pm = \frac{1}{\sqrt{2}} [1 \pm (1 - \rho^2)^{1/2}]^{1/2} ,$$

$$n_i = (dt)^{1/2} p_i , \quad (33)$$

where p_1 and p_2 are independent [0,1] Gaussian distributions.

The cost function C is defined from the equivalent short-time probability distribution, P , for the above set of equations.

$$\begin{aligned} P &= g^{1/2} (2\pi dt)^{-1/2} \exp(-Ldt) \\ &= \exp(-C) , \end{aligned}$$

$$\begin{aligned}
C &= Ldt + \ln(2\pi dt) - \frac{1}{2} \ln(g) , \\
L &= \frac{1}{2} F^\dagger \underline{g} F , \\
F &= \begin{pmatrix} dr/dt - (a_1 + b_1(l-r)) \\ dl/dt - l(a_2 + b_2r + c_2l) \end{pmatrix} , \\
g &= \det(\underline{g}) , \\
k &= 1 - \rho^2 .
\end{aligned} \tag{34}$$

\underline{g} , the metric in $\{r, l\}$ -space, is the inverse of the covariance matrix,

$$\underline{g}^{-1} = \begin{pmatrix} (r\sigma_1)^2 & \rho r l \sigma_1 \sigma_2 \\ \rho r l \sigma_1 \sigma_2 & (l\sigma_2)^2 \end{pmatrix} . \tag{35}$$

The correct mesh for time, dt , in order that P represent the Langevin equations (to order $dt^{3/2}$) is

$$dt \leq 1/\bar{L} , \tag{36}$$

where \bar{L} is L evaluated with $ds/dt = dl/dt = 0$. If dt is greater than $1/\bar{L}$, then it is inappropriate to use P , and instead the path integral over intermediate states of folded short-time distributions must be calculated. In this context, it should be noted that the correct time mesh for the corresponding differential equations must be at least as small, since typically differentiation is a ‘‘sharpening’’ process. This will be noted in any discipline requiring numerical calculation, when comparing differential and integral representations of the same system.

As reported [3,61], this methodology achieved a cost function a factor of 10 smaller than that calculated using the parameters given in the previous literature.

4.2.3. Fitting SMFM to SP500

For the purposes of this paper, it suffices to consider a two-variable problem, SP500 prices of futures, p^1 , and cash, p^2 [63,76]. Data included 251 points of 1989 and 252 points of 1990 daily closing data. Time between data was taken as real time t , e.g., a weekend added two days to the time between data of a Monday and a previous Friday.

It was decided that relative data should be more important to the dynamics of the SMFM model than absolute data, and an arbitrary form was developed to preprocess data used in the fits,

$$M^i(t) = p^i(t + \Delta t)/p^i(t), \quad (37)$$

where $i = \{1, 2\} = \{\text{futures, cash}\}$, and Δt was the time between neighboring data points, and $t + \Delta t$ is the current trading time. The ratio served to suppress strong drifts in the absolute data, to model other underlying market dynamics.

4.2.3.1. ASA fits of SMFM to data

Two source of noise were assumed, so that the equations of this SMFM model are

$$\frac{dM^G}{dt} = \sum_{G'=1}^2 f_{G'}^G M^{G'} + \sum_{i=1}^2 \hat{g}_i^G \eta^i, \quad G = \{1, 2\}. \quad (38)$$

The 8 parameters, $\{f_{G'}^G, \hat{g}_i^G\}$ were all taken to be constants. As discussed previously, the path-integral representation was used to define an effective cost function. Minimization of the cost function was performed using ASA.

Using 1989 data, the parameters $f_{G'}^G$ were constrained to lie between -1.0 and 1.0. The parameters \hat{g}_i^G were constrained to lie between 0 and 1.0. The values of the parameters, obtained by this fitting process were: $f_1^1 = 0.0686821$, $f_2^1 = -0.068713$, $\hat{g}_1^1 = 0.000122309$, $\hat{g}_2^1 = 0.000224755$, $f_1^2 = 0.645019$, $f_2^2 = -0.645172$, $\hat{g}_1^2 = 0.00209127$, $\hat{g}_2^2 = 0.00122221$.

4.2.3.2. ASA fits of trading rules

A simple model of trading was developed. Two time-weighted moving averages, of wide and narrow windows, a_w and a_n were defined for each of the two momenta variables. During each new epoch of a_w , always using the fits of the SMFM model described in the previous section as a zeroth order estimate, the parameters $\{f_{G'}^G, \hat{g}_i^G\}$ were refit using data within each epoch. Averaged canonical momenta, i.e., using Eq. (7), were calculated for each new set of a_w and a_n windows. Fluctuation parameters $\Delta\Pi_w^G$ and $\Delta\Pi_n^G$, were defined, such that any change in trading position required that there was some reasonable information outside of these fluctuations that could be used as criteria for trading decisions. No trading was performed for the first few days of the year until the momenta could be

calculated. Commissions of \$70 were paid every time a new trade of 100 units was taken. Thus, there were 6 trading parameters used in this example, $\{a_w, a_n, \Delta\Pi_w^G, \Delta\Pi_n^G\}$.

The order of choices made for daily trading are as follows. A 0 represents no positions are open and no trading is performed until enough data is gathered, e.g., to calculate momenta. A 1 represents entering a long position, whether from a waiting or a short position, or a current long position was maintained. This was performed if the both wide-window and narrow-window averaged momenta of both cash and futures prices were both greater than their $\Delta\Pi_w^G$ and $\Delta\Pi_n^G$ fluctuation parameters. A -1 represents entering a short position, whether from a waiting or a long position, or a current short position was maintained. This was performed if the both wide-window and narrow-window averaged momenta of both cash and futures prices were both less than their $\Delta\Pi_w^G$ and $\Delta\Pi_n^G$ fluctuation parameters.

4.2.3.3. In-sample ASA fits of trading rules

For the data of 1989, recursive optimization was performed. The trading parameters were optimized in an outer shell, using the negative of the net yearly profit/loss as a cost function. This could have been weighted by something like the absolute value of maximum loss to help minimize risk, but this was not done here. The inner shell of optimization fine-tuning of the SMFM model was performed daily over the current a_w epoch.

The trading-rule parameters were constrained to lie within the following ranges: a_w integers between 15 and 25, a_n integers between 3 and 14, $\Delta\Pi_w^G$ and $\Delta\Pi_n^G$ between 0 and 200. The trading parameters fit by this procedure were: $a_w = 18$, $a_n = 11$, $\Delta\Pi_w^1 = 30.3474$, $\Delta\Pi_w^2 = 98.0307$, $\Delta\Pi_n^1 = 11.2855$, $\Delta\Pi_n^2 = 54.8492$.

The summary of results was: cumulative profit = \$54170, number of profitable long positions = 11, number of profitable short positions = 8, number of losing long positions = 5, number of losing short positions = 6, maximum profit of any given trade = \$11005, maximum loss of any trade = -\$2545, maximum accumulated profit during year = \$54170, maximum loss sustained during year = \$0.

4.2.3.4. Out-of-sample SMFM trading

The trading process described above was applied to the 1990 out-of-sample SP500 data. Note that 1990 was a “bear” market, while 1989 was a “bull” market. Thus, these two years had quite different

overall contexts, and this was believed to provide a stronger test of this methodology than picking two years with similar contexts.

The inner shell of optimization was performed as described above for 1990 as well. The summary of results was: cumulative profit = \$28300, number of profitable long positions = 10, number of profitable short positions = 6, number of losing long positions = 6, number of losing short positions = 10, maximum profit of any given trade = \$6780, maximum loss of any trade = -\$2450, maximum accumulated profit during year = \$29965, maximum loss sustained during year = -\$5945. Tables of results are available as file markets96_momenta_tbl.txt.Z in <http://www.ingber.com/MISC.DIR/> and <ftp.ingber.com/MISC.DIR>.

Only one variable, the futures SP500, was actually traded, albeit the code can accommodate trading on multiple markets. There is more leverage and liquidity in actually trading the futures market. The multivariable coupling to the cash market entered in three important ways: (1) The SMFM fits were to the coupled system, requiring a global optimization of all parameters in both markets to define the time evolution of the futures market. (2) The canonical momenta for the futures market is in terms of the partial derivative of the full Lagrangian; the dependency on the cash market enters both as a function of the relative value of the off-diagonal to diagonal terms in the metric, as well as a contribution to the drifts and diffusions from this market. (3) The canonical momenta of both markets were used as technical indicators for trading the futures market.

4.2.3.5. Reversing data sets

The same procedures described above were repeated, but using the 1990 SP500 data set for training and the 1989 data set for testing.

For the training phase, using 1990 data, the parameters f_G^G were constrained to lie between -1.0 and 1.0. The parameters \hat{g}_i^G were constrained to lie between 0 and 1.0. The values of the parameters, obtained by this fitting process were: $f_1^1 = 0.0685466$, $f_2^1 = -0.068571$, $\hat{g}_1^1 = 7.52368 \cdot 10^{-6}$, $\hat{g}_2^1 = 0.000274467$, $f_1^2 = 0.642585$, $f_2^2 = -0.642732$, $\hat{g}_1^2 = 9.30768 \cdot 10^{-5}$, $\hat{g}_2^2 = 0.00265532$. Note that these values are quite close to those obtained above when fitting the 1989 data.

The trading-rule parameters were constrained to lie within the following ranges: a_w integers between 15 and 25, a_n integers between 3 and 14, $\Delta\Pi_w^G$ and $\Delta\Pi_n^G$ between 0 and 200. The trading parameters fit by this procedure were: $a_w = 11$, $a_n = 8$, $\Delta\Pi_w^1 = 23.2324$, $\Delta\Pi_w^2 = 135.212$, $\Delta\Pi_n^1 = 169.512$,

$$\Delta\Pi_n^2 = 9.50857,$$

The summary of results was: cumulative profit = \$42405, number of profitable long positions = 11, number of profitable short positions = 8, number of losing long positions = 7, number of losing short positions = 6, maximum profit of any given trade = \$8280, maximum loss of any trade = -\$1895, maximum accumulated profit during year = \$47605, maximum loss sustained during year = -\$2915.

For the testing phase, the summary of results was: cumulative profit = \$35790, number of profitable long positions = 10, number of profitable short positions = 6, number of losing long positions = 6, number of losing short positions = 3, maximum profit of any given trade = \$9780, maximum loss of any trade = -\$4270, maximum accumulated profit during year = \$35790, maximum loss sustained during year = \$0. Tables of results are available as file markets96_momenta_tbl.txt.Z in <http://www.ingber.com/MISC.DIR/> and <ftp.ingber.com/MISC.DIR>.

4.3. Statistical Mechanics of Combat

A series of papers has developed a statistical mechanics of combat (SMC) directed to large-scale forces [37,77-81], where details and the rationale of this presentation can be found. This model provided a common language to quantitatively compare data from exercises to simulations, thereby baselining the simulations.

4.3.1. The U.S. Army National Training Center (NTC)

The NTC is a large maneuver range dedicated to the simulation of desert combat, training battalion and brigade size mechanized units from U.S. Army heavy divisions and separate brigades. The NTC is unique in that it is highly instrumented with the Multiple Integrated Laser Engagement System (MILES) and range instrumentation which follows the location and activity of most vehicles and some dismounted infantry. The NTC also has a dedicated Opposing Force (OPFOR) which acts as the enemy during force-on-force exercises with visiting units. Transfers of data between different databases and computer operating systems were automated for this analysis [82].

Consider a scenario taken from our NTC study: two Red systems, Red T-72 tanks (*RT*) and Red armored personnel carriers (*RBMP*), and three Blue systems, Blue M1A1 and M60 tanks (*BT*), Blue armored personnel carriers (*BAPC*), and Blue tube-launched optically-tracked wire-guided missiles

(*BTOW*), where *RT* specifies the number of Red tanks at a given time *t*, etc. Consider the kills suffered by *BT*, ΔBT , e.g., within a time epoch $\Delta t \approx 5$ min

$$\Delta BT/\Delta t \equiv \dot{BT} = x_{RT}^{BT} RT + y_{RT}^{BT} RT BT + x_{RBMP}^{BT} RBMP + y_{RBMP}^{BT} RBMP BT \quad (39)$$

Here, the *x* terms represent attrition owing to point fire; the *y* terms represent attrition owing to area fire. Note that the algebraic forms chosen are consistent with current perceptions of aggregated large scale combat. The version of the combat simulation used to generate this data does not permit direct-fire fratricide; such terms are set to zero. In most NTC scenarios fratricide typically is negligible.

Now consider sources of noise, e.g., that at least arise from probabilities of detection (PD), acquisition (PA), hit (PH), kill (PK), etc. Furthermore, such noise likely has its own functional dependencies, e.g., possibly being proportional to the numbers of units involved in the combat. For simplicity here, still generating much nonlinearity, only diagonal noise terms are considered. Coupling among the variables takes place in the drift terms (deterministic limit); for simplicity only linear terms in the drifts are taken for this prototype study.

$$\begin{aligned} \frac{\Delta BT}{\Delta t} \equiv \dot{BT} = & x_{RT}^{BT} RT + y_{RBMP}^{BT} RT BT + x_{RBMP}^{BT} RBMP + y_{RBMP}^{BT} RBMP BT \\ & + z_{BT}^{BT} BT \eta_{BT}^{BT} \end{aligned} \quad (40)$$

where the η represent sources of (white) noise. The noise terms are taken to be log normal (multiplicative) noise for the diagonal terms and additive noise for the off-diagonal terms. This induces a high degree of nonlinearity, which can be seen by transforming each variable M^G to X^G

$$X^G = \ln M^G, \quad \dot{X}^G = \dot{M}^G/M^G,$$

$$M^G = \{RT, RBMP, BT, BAPC, BTOW\} \quad (41)$$

yielding X^G equations with constant coefficients of the noise, at the expense of introducing exponential terms in the drifts.

The methodology presented here can accommodate any other nonlinear functional forms, and any other variables that can be reasonably represented by such rate equations, e.g., expenditures of ammunition or bytes of communication [83]. Variables that cannot be so represented, e.g., terrain, C^3 , weather, etc., must be considered as “super-variables” that specify the overall context for the above set of

rate equations.

4.3.2. Janus simulation data

Janus is an interactive, two-sided, closed, stochastic, ground combat simulation. Players direct their elements, executing tactical plans and reacting to enemy actions. The disposition of opposing forces is not completely known to players. Janus models individual systems moving, searching, detecting and engaging other ground or air systems over a three-dimensional terrain representation, using Army-developed algorithms and data to model combat processes.

For this study, data collected during our NTC-Janus(T) project circa 1988 was used to fit the coefficients of the above 5 coupled equations. Time epochs were 5 mins each, and we used data from 6 battle simulations between 30 mins and 75 mins into the battles, for a total of 60 states of data, each state giving the present values of each of the 2 Red and 3 Blue units.

It should be noted that the numbers of units in this particular set of data are barely large enough to be considered large-scale so that the statistical methodology being presented is applicable. At the least, this paper presents a full study to demonstrate the SMC approach for future sets of large-scale data.

4.3.3. Algebraic development

The five coupled stochastic differential equations, for variables $M^G = \{RT, RBMP, BT, BAPC, BTOW\}$, can be represented equivalently by a short-time conditional probability distribution P in terms of a Lagrangian L :

$$P(R \cdot, B \cdot; t + \Delta t | R \cdot, B \cdot; t) = \frac{1}{(2\pi\Delta t)^{5/2} \sigma^{1/2}} \exp(-L\Delta t) \quad (42)$$

where σ is the determinant of the inverse of the covariance matrix, the metric matrix of this space, “ $R \cdot$ ” represents $\{RT, RBMP\}$, and “ $B \cdot$ ” represents $\{BT, BAPC, BTOW\}$. (Here, the prepoint discretization is used, which hides the Riemannian corrections explicit in the midpoint discretized Feynman Lagrangian; only the latter representation possesses a variational principle useful for arbitrary noise.)

This defines a scalar “dynamic cost function,” $C(x, y, z)$,

$$C(x, y, z) = L\Delta t + \frac{5}{2} \ln(2\pi\Delta t) + \frac{1}{2} \ln \sigma \quad (43)$$

which can be used with the adaptive simulated annealing (ASA) algorithm [1,14] to find the best fit of $\{x, y, z\}$ to the data.

4.3.4. Numerical results

A systematic numerical procedure has been developed for fitting parameters in such stochastic nonlinear systems to data using methods of adaptive simulated annealing (ASA) [1,14,19,20], and then integrating the path integral using a non-Monte Carlo technique especially suited for nonlinear systems [34]. This numerical methodology has been applied with success to several systems [4,37,38,40,51,61,63].

4.3.5. Results

CMI can be very useful indicators of combat analysis, but here the fits of the model can be very revealing as well. Table 1 gives the results of ASA fits of the above 5 coupled equations to Janus-generated data. Note that the noise coefficient is roughly the same for all units, being largest for *BTOW*. Note the relative importance of coefficients in “predicting” the immediate next epoch, with *BTOW* larger than *BAPC* larger than *BT* in depleting Red forces (but being multiplied by the total number of units at any time). The coefficients of “prediction” of attrition by Red forces has *RT* larger than *RBMP* against *BTOW*, and *RT* less than *RBMP* against *BT* and *BAPC* (but being multiplied by the total number of units at any time).

	<i>RT</i>	<i>RBMP</i>	<i>BT</i>	<i>BAPC</i>	<i>BTOW</i>	η [.]
<i>RT</i>	-	-	-8.56E-5	-5.91E-3	-3.63E-2	3.72E-3
<i>RBMP</i>	-	-	-2.66E-3	-2.17E-2	-3.14E-2	4.33E-3
<i>BT</i>	-6.66E-4	-4.67E-3	-	-	-	7.93E-3
<i>BAPC</i>	-1.01E-4	-3.95E-3	-	-	-	6.73E-3
<i>BTOW</i>	-2.11E-3	-1.24E-6	-	-	-	1.25E-2

TABLE 1. Entities in the table are the ASA-fitted coefficients of the coupled set of 5 equations representing the dynamics of Red and Blue interactions. Note that the last column coefficients are multiplied by the corresponding variable in the first column. A dash

represents no coefficient present in the equations.

5. CONCLUSIONS

Modern methods of developing multivariate nonlinear multiplicative Gaussian-Markovian systems are quite important, as there are many such systems and the mathematics must be diligently exercised if such models are to faithfully represent the data they describe. Similarly, sophisticated numerical techniques, e.g., global optimization and path integration are important tools to carry out the modeling and fitting to data without compromising the model, e.g., by unwarranted quasi-linear approximations. Three quite different systems have benefited from this approach.

The large-scale modeling of neocortical interactions has benefited from the use of canonical momenta indicators (CMI), intuitive constructs that yet are faithful to the complex algebra describing this multiple-scaled complex system. After global optimization of the Lagrangian describing interactions among large sets of neurons, the derived CMI yield stronger signal resolution of EEG than do raw potential data.

CMI also have been profitably applied to multivariate financial markets. Here, another level of recursive optimization produced optimal trading rules to apply to the fitted CMI.

This approach also has proven useful for combat simulation analyses. The contribution of individual components to the synergistic operation of entire systems can be detailed after global optimization of models baselined to exercise data.

It is clear that ASA optimization and PATHINT path-integral tools are very useful to develop the algebra of statistical mechanics for a large class of nonlinear stochastic systems. However, it also is clear that each system typically presents its own non-typical unique character and this must be included in any such analysis. A virtue of this statistical mechanics approach and these associated tools is they appear to be flexible and robust to handle quite different systems.

REFERENCES

1. L. Ingber, Adaptive Simulated Annealing (ASA), Global optimization C-code, Lester Ingber Research, Chicago, IL, (1993).
2. L. Ingber, Statistical mechanics of nonlinear nonequilibrium financial markets, *Math. Modelling* **5** (6), 343-361 (1984).
3. L. Ingber, M.F. Wehner, G.M. Jabbour, and T.M. Barnhill, Application of statistical mechanics methodology to term-structure bond-pricing models, *Mathl. Comput. Modelling* **15** (11), 77-98 (1991).
4. L. Ingber, Statistical mechanics of neocortical interactions: A scaling paradigm applied to electroencephalography, *Phys. Rev. A* **44** (6), 4017-4060 (1991).
5. K.S. Cheng, Quantization of a general dynamical system by Feynman's path integration formulation, *J. Math. Phys.* **13**, 1723-1726 (1972).
6. H. Dekker, Functional integration and the Onsager-Machlup Lagrangian for continuous Markov processes in Riemannian geometries, *Phys. Rev. A* **19**, 2102-2111 (1979).
7. R. Graham, Path-integral methods on nonequilibrium thermodynamics and statistics, in *Stochastic Processes in Nonequilibrium Systems*, (Edited by L. Garrido, P. Seglar, and P.J. Shepherd), pp. 82-138, Springer, New York, NY, (1978).
8. F. Langouche, D. Roekaerts, and E. Tirapegui, Discretization problems of functional integrals in phase space, *Phys. Rev. D* **20**, 419-432 (1979).
9. F. Langouche, D. Roekaerts, and E. Tirapegui, Short derivation of Feynman Lagrangian for general diffusion process, *J. Phys. A* **113**, 449-452 (1980).
10. P. Hagedorn, *Non-Linear Oscillations*, Oxford Univ., New York, NY, (1981).
11. C.W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, Springer-Verlag, Berlin, Germany, (1983).
12. R. C. Merton, An intertemporal capital asset pricing model, *Econometrica* **41**, 867-887 (1973).
13. L. Ingber, Statistical mechanics of neocortical interactions: Stability and duration of the 7+-2 rule of short-term-memory capacity, *Phys. Rev. A* **31**, 1183-1186 (1985).

14. L. Ingber, Very fast simulated re-annealing, *Mathl. Comput. Modelling* **12** (8), 967-973 (1989).
15. S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, Optimization by simulated annealing, *Science* **220** (4598), 671-680 (1983).
16. 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).
17. H. Szu and R. Hartley, Fast simulated annealing, *Phys. Lett. A* **122** (3-4), 157-162 (1987).
18. M. Wofsey, Technology: Shortcut tests validity of complicated formulas, *The Wall Street Journal* **222** (60), B1 (1993).
19. L. Ingber, Simulated annealing: Practice versus theory, *Mathl. Comput. Modelling* **18** (11), 29-57 (1993).
20. L. Ingber, Adaptive simulated annealing (ASA): Lessons learned, *Control and Cybernetics* **25** (1), 33-54 (1996).
21. K. Kimura and K. Taki, Time-homogeneous parallel annealing algorithm, Report TR-673, Institute for New Generation Computer Technology, Tokyo, Japan, (1991).
22. R. Frost, Ensemble Based Simulated Annealing (EBSA), <ftp.sdsc.edu:/pub/sdsc/math/Ebsa>, University of California San Diego, La Jolla, CA, (1993).
23. L. Ingber, Generic mesoscopic neural networks based on statistical mechanics of neocortical interactions, *Phys. Rev. A* **45** (4), R2183-R2186 (1992).
24. B. Rosen, Function optimization based on advanced simulated annealing, *IEEE Workshop on Physics and Computation - PhysComp '92*, 289-293 (1992).
25. L. Ingber and B. Rosen, Genetic algorithms and very fast simulated reannealing: A comparison, *Mathl. Comput. Modelling* **16** (11), 87-100 (1992).
26. D.G. Mayer, J.A. Belward, and K. Burrage, Use of advanced techniques to optimize a multi-dimensional dairy model, *Agricultural Systems* **50**, 239-253 (1996).
27. D.G. Mayer, J.A. Belward, K. Burrage, and M.A. Stuart, Optimization of a dairy farm model - Comparison of simulated annealing, simulated quenching and genetic algorithms, in *Proceedings 1995 International Congress on Modelling and Simulation, 27-30 November 1995, University of Newcastle*, pp. 33-38, (1995).

28. D.G. Mayer, P.M. Pepper, J.A. Belward, K. Burrage, and A.J. Swain, Simulated annealing - A robust optimization technique for fitting nonlinear regression models, in *Proceedings 'Modelling, Simulation and Optimization' Conference, International Association of Science and Technology for Development (IASTED), 6-9 May 1996 Gold Coast, (1996)*.
29. A. Su, S. Mager, S.L. Mayo, and H.A. Lester, A multi-substrate single-file model for ion-coupled transporters, *Biophys. J* **70**, 762-777 (1996).
30. K. Wu and M.D. Levine, 3-D shape approximation using parametric geons, *Intl J Image and Vision Computing* **15** (2), 143-158 (1997).
31. B. Cohen, Training synaptic delays in a recurrent neural network, M.S. Thesis, Tel-Aviv University, Tel-Aviv, Israel, (1994).
32. S. Sakata, High breakdown point estimation in econometrics, Ph.D. Thesis, University of California at San Diego, La Jolla, CA, (1995).
33. M.K. Sen and P.L. Stoffa, *Global Optimization Methods in Geophysical Inversion*, Elsevier, The Netherlands, (1995).
34. M.F. Wehner and W.G. Wolfer, Numerical evaluation of path-integral solutions to Fokker-Planck equations. I., *Phys. Rev. A* **27**, 2663-2670 (1983).
35. M.F. Wehner and W.G. Wolfer, Numerical evaluation of path-integral solutions to Fokker-Planck equations. II. Restricted stochastic processes, *Phys. Rev. A* **28**, 3003-3011 (1983).
36. 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).
37. 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).
38. L. Ingber, Path-integral evolution of multivariate systems with moderate noise, *Phys. Rev. E* **51** (2), 1616-1619 (1995).
39. L. Ingber, Statistical mechanics of neocortical interactions: Path-integral evolution of short-term memory, *Phys. Rev. E* **49** (5B), 4652-4664 (1994).
40. L. Ingber and P.L. Nunez, Statistical mechanics of neocortical interactions: High resolution path-integral calculation of short-term memory, *Phys. Rev. E* **51** (5), 5074-5083 (1995).

41. L. Ingber, Towards a unified brain theory, *J. Social Biol. Struct.* **4**, 211-224 (1981).
42. L. Ingber, Statistical mechanics of neocortical interactions. I. Basic formulation, *Physica D* **5**, 83-107 (1982).
43. L. Ingber, Statistical mechanics of neocortical interactions. Dynamics of synaptic modification, *Phys. Rev. A* **28**, 395-416 (1983).
44. L. Ingber, Statistical mechanics of neocortical interactions. Derivation of short-term-memory capacity, *Phys. Rev. A* **29**, 3346-3358 (1984).
45. L. Ingber, Statistical mechanics of neocortical interactions. EEG dispersion relations, *IEEE Trans. Biomed. Eng.* **32**, 91-94 (1985).
46. L. Ingber, Towards clinical applications of statistical mechanics of neocortical interactions, *Innov. Tech. Biol. Med.* **6**, 753-758 (1985).
47. L. Ingber, Statistical mechanics of neocortical interactions, *Bull. Am. Phys. Soc.* **31**, 868 (1986).
48. L. Ingber, Applications of biological intelligence to Command, Control and Communications, in *Computer Simulation in Brain Science: Proceedings, University of Copenhagen, 20-22 August 1986*, (Edited by R. Cotterill), pp. 513-533, Cambridge University Press, London, (1988).
49. L. Ingber, Statistical mechanics of mesoscales in neocortex and in command, control and communications (C^3): Proceedings, Sixth International Conference, St. Louis, MO, 4-7 August 1987, *Mathl. Comput. Modelling* **11**, 457-463 (1988).
50. 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).
51. L. Ingber and P.L. Nunez, Multiple scales of statistical physics of neocortex: Application to electroencephalography, *Mathl. Comput. Modelling* **13** (7), 83-95 (1990).
52. L. Ingber, Statistical mechanics of multiple scales of neocortical interactions, in *Neocortical Dynamics and Human EEG Rhythms*, (Edited by P.L. Nunez), pp. 628-681, Oxford University Press, New York, NY, (1995).
53. L. Ingber, Statistical mechanics of neocortical interactions: Constraints on 40 Hz models of short-term memory, *Phys. Rev. E* **52** (4), 4561-4563 (1995).

54. L. Ingber, Statistical mechanics of neocortical interactions: Multiple scales of EEG, in *Frontier Science in EEG: Continuous Waveform Analysis (Electroencephal. clin. Neurophysiol. Suppl. 45)*, (Edited by R.M. Dasheiff and D.J. Vincent), pp. 79-112, Elsevier, Amsterdam, (1996).
55. L. Ingber, Statistical mechanics of neocortical interactions: Applications of canonical momenta indicators to electroencephalography, *Phys. Rev. E* **55** (4), 4578-4593 (1997).
56. L. Ingber, Statistical mechanics of neocortical interactions: Training and testing canonical momenta indicators of EEG, *Mathl. Computer Modelling* **27** (3), 33-64 (1998).
57. R. Graham, Covariant formulation of non-equilibrium statistical thermodynamics, *Z. Physik* **B26**, 397-405 (1977).
58. F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions*, Reidel, Dordrecht, The Netherlands, (1982).
59. V.B. Mountcastle, R.A. Andersen, and B.C. Motter, The influence of attentive fixation upon the excitability of the light-sensitive neurons of the posterior parietal cortex, *J. Neurosci.* **1**, 1218-1235 (1981).
60. X.L. Zhang, H. Begleiter, B. Porjesz, W. Wang, and A. Litke, Event related potentials during object recognition tasks, *Brain Res. Bull.* **38** (6), 531-538 (1995).
61. L. Ingber, Statistical mechanical aids to calculating term structure models, *Phys. Rev. A* **42** (12), 7057-7064 (1990).
62. L. Ingber, Statistical mechanics of nonlinear nonequilibrium financial markets: Applications to optimized trading, *Mathl. Computer Modelling* **23** (7), 101-121 (1996).
63. L. Ingber, Canonical momenta indicators of financial markets and neocortical EEG, in *Progress in Neural Information Processing*, (Edited by S.-I. Amari, L. Xu, I. King, and K.-S. Leung), pp. 777-784, Springer, New York, (1996).
64. R. Pool, Is it chaos, or is it just noise?, *Science* **243**, 25-28 (1989).
65. H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, and L.Sh. Tsimring, The analysis of observed chaotic data in physical systems, *Rev. Mod. Phys.* **65** (4), 1331-1392 (1993).
66. P. Grassberger, Do climatic attractors exist?, *Nature* **323**, 609-612 (1986).

67. L. Ingber, R. Srinivasan, and P.L. Nunez, Path-integral evolution of chaos embedded in noise: Duffing neocortical analog, *Mathl. Computer Modelling* **23** (3), 43-53 (1996).
68. L. Bachelier, Théorie de la Spéculation, *Annales de l'Ecole Normale Supérieure* **3** (1900).
69. M. C. Jensen, Some anomalous evidence regarding market efficiency, an editorial introduction, *J. Finan. Econ.* **6**, 95-101 (1978).
70. B. B. Mandelbrot, When can price be arbitrated efficiently? A limit to the validity of the random walk and martingale models, *Rev. Econ. Statist.* **53**, 225-236 (1971).
71. S. J. Taylor, Tests of the random walk hypothesis against a price-trend hypothesis, *J. Finan. Quant. Anal.* **17**, 37-61 (1982).
72. P. Brown, A. W. Kleidon, and T. A. Marsh, New evidence on the nature of size-related anomalies in stock prices, *J. Fin. Econ.* **12**, 33-56 (1983).
73. B. Dietrich-Campbell and E. Schwartz, Valuing debt options, *J. Finan. Econ.* **16**, 321-343 (1986).
74. M.J. Brennan and E.S. Schwartz, A continuous time approach to the pricing of bonds, *J. Banking Finan.* **3**, 133-155 (1979).
75. M.J. Brennan and E.S. Schwartz, An equilibrium model of bond pricing and a test of market efficiency, *J. Finan. Quant. Anal.* **17**, 301-329 (1982).
76. L. Ingber, Trading markets with canonical momenta and adaptive simulated annealing, Report No. TMWCMASA-96, Lester Ingber Research, McLean, VA, (1996).
77. 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*, pp. 258-264, American Association for Artificial Intelligence, Menlo Park, CA, (1985).
78. L. Ingber, Nonlinear nonequilibrium statistical mechanics approach to C^3 systems, in *9th MIT/ONR Workshop on C^3 Systems: Naval Postgraduate School, Monterey, CA, 2-5 June 1986*, pp. 237-244, MIT, Cambridge, MA, (1986).
79. L. Ingber and D.D. Sworder, Statistical mechanics of combat with human factors, *Mathl. Comput. Modelling* **15** (11), 99-127 (1991).
80. L. Ingber, Statistical mechanics of combat and extensions, in *Toward a Science of Command, Control, and Communications*, (Edited by C. Jones), pp. 117-149, American Institute of

Aeronautics and Astronautics, Washington, D.C., (1993).

81. M. Bowman and L. Ingber, Canonical momenta of nonlinear combat, in *Proceedings of the 1997 Simulation Multi-Conference, 6-10 April 1997, Atlanta, GA*, Society for Computer Simulation, San Diego, CA, (1997).
82. M. Bowman, Integration of the NTC Tactical Database and JANUS(T) Towards a Combat Decision Support System, M.S. Thesis, Naval Postgraduate School, Monterey, CA, (1989).
83. L. Ingber, Mathematical comparison of JANUS(T) simulation to National Training Center, in *The Science of Command and Control: Part II, Coping With Complexity*, (Edited by S.E. Johnson and A.H. Levis), pp. 165-176, AFCEA International, Washington, DC, (1989).

FIGURE CAPTIONS

FIG. 1. For the matching-stimulus c_1 paradigm for control subject co2c0000337, each figure gives data under 6 electrodes marked in the legends. The left hand figures represent data for the training calculations; the right hand figures represent data for the testing calculations. The top figures represent averages over 10 runs of raw evoked potential data; the bottom figures represent averages over 10 calculations of canonical momenta indicators using this data.

