Physical Review A **45**(4), pp. R2183-R2186 (1992). A “Rapid Communications.”

Generic mesoscopic neural networks based on statistical mechanics of neocortical interactions

Lester Ingber*
Science Transfer Corporation, P.O. Box 857, McLean, Virginia 22101
(Received 26 November 1991)

A series of papers over the past decade [the most recent being L. Ingber, Phys. Rev. A **44**, 4017 (1991)] has developed a statistical mechanics of neocortical interactions (SMNI), deriving aggregate behavior of experimentally observed columns of neurons from statistical electrical-chemical properties of synaptic interactions, demonstrating its capability in describing large-scale properties of short-term memory and electroencephalographic systematics. This methodology also defines an algorithm to construct a mesoscopic neural network, based on realistic neocortical processes and parameters, to record patterns of brain activity and to compute the evolution of this system. Furthermore, this algorithm is quite generic, and can be used to similarly process information in other systems, especially, but not limited to, those amenable to modeling by mathematical physics techniques alternatively described by path-integral Lagrangians, Fokker-Planck equations, or Langevin rate equations. This methodology is made possible and practical by a confluence of techniques drawn from SMNI itself, modern methods of functional stochastic calculus defining nonlinear Lagrangians, very fast simulated reannealing, and parallel-processing computation.

PACS Nos.: 87.10.+e, 05.40.+j, 02.50.+s, 02.70.+d

I. MOTIVATION

During the past decade, a statistical mechanics of neocortical interactions (SMNI) has been developed to address mesoscopic phenomena in the human neocortex, based on a physically and mathematically justifiable nonlinear stochastic development [1-13]. The basic approach of the SMNI has been to statistically aggregate synaptic and neuronal interactions, from microscopic systematics to mesoscopic interactions among minicolumns of hundreds of neurons, to macroscopic macrocolumnar interactions among thousands of minicolumns, to approach regional spatial scales of several millimeters to several centimeters. The theory has been tested by verifying observations at the mesoscopic scale, e.g., short-term-memory phenomena [4,6], and at the macroscopic scale, e.g., electroencephalography (EEG) [4,5,13]. A current description of the theory to date is in Ref. [13], where extensions were made to the SMNI to correlate human behavioral states to circuitries measured by EEG electrode recordings, an ongoing project. While the experimental resolution of EEG typically is on the order of several centimeters, new work has shown that under some circumstances this resolution can be sharpened to several millimeters, the spatial extent of macrocolumns [14].

As noted early in this development [1,15], basing information processing at macroscopic scales on processes at mesoscopic scales leads to new kinds of computational algorithms, such as those proposed here. This approach is consistent with the view that many complex nonlinear nonequilibrium systems develop mesoscopic intermediate structures, e.g., to develop a Gaussian-Markovian statistics to maximize the flow of information [16,17]. Viewing the neocortex as the prototypical information processor, a mesoscopic-neural-network (MNN) algorithm can be developed to store and predict patterns of activity inherent in many such nonlinear stochastic multivariate systems.

II. SMNI MNN

The SMNI theory has developed a nonlinear Lagrangian dynamics describing mesoscopic neocortical interactions. The mathematical development of mesocolumns establishes a mesoscopic Lagrangian $\underline{L}$, defining the short-time probability distribution of firings in a minicolumn, composed of $\sim 10^2$ neurons [18-20], given its just previous interactions with all other neurons in its macrocolumnar surround. In the prepoint discretization this is considerably simplified, and the Riemannian geometry induced by the nonlinear metric (inverse diffusion matrix) is not explicitly displayed as it is in the midpoint discretization [21]. The relevance of this Riemannian geometry in the neocortex has been discussed previously in the SMNI papers, and will not be further addressed here.

The Einstein summation convention is used for compactness, whereby any index appearing more than once among factors in any term is assumed to be summed over, unless otherwise indicated by vertical bars, e.g., $|G|$. The mesoscopic probability distribution P is given by the product of microscopic neuronal conditional probability distributions p_{σ_i}, constrained such that the aggregate mesoscopic excitatory firings $M^E = \sum_{j \in E} \sigma_j$, and the aggregate mesoscopic inhibitory firings $M^I = \sum_{j \in I} \sigma_j$. The SMNI further develops this into a Lagrangian dynamics in terms of first-moment drifts g^G, a second-moment diffusion matrix $g^{GG'}$, a metric $g_{GG'}$ as the inverse of the diffusion matrix, and a potential $\underline{V}'$, where $G = \{E, I\}$.

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

$$\underline{L} = (2N)^{-1}(\dot{M}^G - g^G) g_{GG'} (\dot{M}^{G'} - g^{G'}) + M^G J_G/(2N\tau) - \underline{V}' ,$$

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

$$g = \det(g_{GG'}) . \quad (1)$$

In the SMNI, the drifts, diffusions, and potential are given by

$$g^G = -\tau^{-1}(M^G + N^G \tanh F^G) ,$$

$$g^{GG'} = (g_{GG'})^{-1} = \delta_G^{G'} \tau^{-1} N^G \mathrm{sech}^2 F^G ,$$

$$\underline{V}' = \sum_G \underline{V}''^G_{G'} (\rho \nabla M^{G'})^2 \ . \tag{2}$$

These all depend sensitively on "threshold factors" F^G,

$$F^G = \frac{V^G - v^{|G|}_{G'} (a^{|G|}_{G'} N^{G'} + \frac{1}{2} A^{|G|}_{G'} | M^{G'})}{\{\pi[(v^{|G|}_{G'})^2 + (\phi^{|G|}_{G'})^2](a^{|G|}_{G'} N^{G'} + \frac{1}{2} A^{|G|}_{G'} M^{G'})\}^{1/2}} \ ,$$

$$a^G_{G'} = \frac{1}{2} A^G_{G'} + B^G_{G'} \ , \tag{3}$$

where $A^G_{G'}$ 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, and nearest-neighbor interactions $\underline{V}'$ are detailed in other SMNI papers [2,4]. $M^{G'}$ and $N^{G'}$ in F^G are afferent macrocolumnar firings, scaled to efferent minicolumnar firings by $N/N^* \sim 10^{-3}$, where N^* is the number of neurons in a macrocolumn. Similarly, $A^{G'}_G$ and $B^{G'}_G$ have been scaled by $N^*/N \sim 10^3$ to keep F^G invariant. This scaling is for convenience only.

In a recent project correlating EEG circuitries to human behavioral states [13], in order to more properly include long-ranged fibers so that interactions among macrocolumns could be included, the J_G Lagrange constraint terms were dropped and more realistically replaced by a modified threshold factor F^G including afferent contributions from $N^{\ddagger E}$ long-ranged excitatory fibers. For example, cortico-cortical neurons were added, where $N^{\ddagger E}$ might be on the order of 10–20 % of N^*: Nearly every pyramidal cell has an axon branch that makes a cortico-cortical connection; i.e., the number of cortico-cortical fibers may be as high as 10^{10}.

While the development of nearest-neighbor interactions into a potential term V' was useful to explore EEG dispersion relations [4,5], for present purposes this is not necessary and, as permitted in the development of SMNI, we simply incorporate nearest-neighbor interactions with firings $M^{\dagger G}$ by again redefining F^G:

$$F^G = \frac{V^G - v^{|G|}_{G'} T^{|G|}_{G'}}{(\pi[(v^{|G|}_{G'})^2 + (\phi^{|G|}_{G'})^2] T^{|G|}_{G'})^{1/2}} \ ,$$

$$T^{|G|}_{G'} = a^{|G|}_{G'} N^{G'} + \frac{1}{2} A^{|G|}_{G'} M^{G'} + a^{\dagger |G|}_{G'} N^{\dagger G'} + \frac{1}{2} A^{\dagger |G|}_{G'} M^{\dagger G'} + a^{\ddagger |G|}_{G'} N^{\ddagger G'} + \frac{1}{2} A^{\ddagger |G|}_{G'} M^{\ddagger G'} \ ,$$

$$a^{\dagger G}_{G'} = \frac{1}{2} A^{\dagger G}_{G'} + B^{\dagger G}_{G'} \ ,$$

$$A^{\ddagger I}_E = A^{\ddagger E}_I = A^{\ddagger I}_I = B^{\ddagger I}_E = B^{\ddagger E}_I = B^{\ddagger I}_I = 0 \ ,$$

$$a^{\ddagger E}_E = \frac{1}{2} A^{\ddagger E}_E + B^{\ddagger E}_E \ . \tag{4}$$

This result presents an SMNI MNN as a set of nodes, each described by a short-time probability distribution interacting with the other nodes. A set of 1000 such nodes represents a macrocolumn, scaled to represent a "dipole" current source [13]. A circuitry among patches of macrocolumns represents a typical circuit of activity correlated to specific behavioral states as recorded by EEG under specific experimental or clinical conditions. We are applying SMNI MNN to learn patterns of an individual's EEG, and then to predict possible future states of brain activities or anomalies, etc. The methodology used is quite general, and will be described generically.

III. GENERIC MNN

We now generalize this SMNI MNN, generated from Eq. (1), to model other large-scale nonlinear stochastic multivariate systems, by considering general drifts and diffusions to model such systems, now letting G represent an arbitrary number of variables. Ideally, these systems inherently will be of the Fokker-Planck type,

$$\frac{\partial P}{\partial t} = \frac{\partial(-g^G P)}{\partial M^G} + \frac{1}{2} \frac{\partial^2 (g^{GG'} P)}{\partial M^G \partial M^{G'}} \ . \tag{5}$$

The topology, geometry, and connectivity of the MNN can of course be generalized. E.g., we need not be restricted to nearest-neighbor interactions, although this is simpler to implement especially on parallel processors. Also, we can include "hidden layers" to increase the complexity of the MNN, although the inclusion of nonlinear structure in the drifts g^G and diffusions $g^{GG'}$ may make this unnecessary for many systems.

A. Learning

"Learning" takes place by presenting the MNN with data, and parametrizing the data in terms of the "firings," or multivariate M^G "spins." The "weights," or coefficients of functions of M^G appearing in the drifts and diffusions, are fit to incoming data, considering the joint "effective" Lagrangian (including the logarithm of the prefactor in the probability distribution) [22] as a dynamic cost function. The cost function is a sum of effective Lagrangians from each node and over each time epoch of data.

This program of fitting coefficients in Lagrangian cost functions to data, has been accomplished in several systems, i.e., in combat analyses [23,24], finance [25,26], and neuroscience [13], using methods of very fast simulated reannealing (VFSR) [27,28]. This maximum likelihood procedure (statistically) avoids problems of trapping in local minima, as experienced by other types of gradient and regression techniques. VFSR has been shown to be orders of magnitude more efficient than other similar techniques, e.g., genetic algorithms, and we have described how it is extremely well suited to our present project by parallelizing this code on a Connection Machine [28,29].

VFSR has been developed to fit observed data to a large class of theoretical cost function over a D-dimensional parameter space, adapting for varying sensitivities of parameters during the fit. The annealing schedule for the "temperatures" (artificial fluctuation parameters) T_i decrease exponentially in "time" (cycle number of iterative process) k, i.e., $T_i = T_{i0} \exp(-c_i k^{1/D})$. Heuristic arguments have been developed to demonstrate that this algorithm is faster than the fast Cauchy annealing [30], $T_i = T_0/k$, and much faster than Boltzmann annealing [31], $T_i = T_0 / \ln k$.

B. Prediction

"Prediction" takes advantage of a mathematically equivalent representation of the Lagrangian path-integral algorithm, i.e., a set of coupled Langevin rate-equations. The Itô (prepoint-discretized) Langevin equation is analyzed in terms of the Wiener process dW^i, which is rewritten in terms of Gaussian noise $\eta^i = dW^i/dt$ in the limit [21]:

$$dM^G = g^G dt + \hat{g}_i^G dW^i \ ,$$

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

$$M = \{ \, M^G ; G = 1, \cdots, \Lambda \, \} \ \ , \ \ \eta = \{ \, \eta^i ; i = 1, \cdots, N \, \} \ ,$$

$$< \eta^j(t) >_\eta = 0 \ \ , \ \ < \eta^j(t), \eta^{j'}(t') >_\eta = \delta^{jj'} \delta(t - t') \ . \tag{6}$$

Moments of an arbitrary function $F(\eta)$ over this stochastic space are defined by a path integral over η^i. The Lagrangian diffusions are calculated as

$$g^{GG'} = \sum_{i=1}^{N} \hat{g}_i^G \hat{g}_i^{G'} \ . \tag{7}$$

The calculation of the evolution of Langevin systems has been implemented in the above-mentioned systems using VFSR [13,23,25]. It has been used as an aid to debug the VFSR fitting codes, by first generating data from coupled Langevin equations, relaxing the coefficients, and then fitting this data with the effective Lagrangian cost-function algorithm to recapture the original coefficients within the diffusions defined by $g^{GG'}$. Thus, these previous projects have been simple demonstrations of the MNN, considering each of many nodes as having the same set of weights, but different random generators residing at each node during generation of data using the Langevin systems in the "prediction" phase. That is, in principle, the reverse procedure could have been exercised: Considering a given Lagrangian as representing a "learned" set of patterns, e.g., in the sense of a sum over eigenfunctions of its probability distribution, then a "prediction" procedure is well defined by developing aggregated statistics over many trajectories produced by the Langevin equations.

When discretized, the Langevin equations give the evolution of the system from state $M(t)$ to time $M(t + \Delta t)$ at each node, in terms of weights fit in the learning phase described above. The inverse "static" Lagrangian (setting $\dot{M}^G = 0$) is a reasonable measure of the time mesh required to integrate the Langevin equations [25]. The ensemble of many nodes gives a reasonable statistical evolution of the system.

IV. DISCUSSION

We have described a computational algorithm faithful to a model of neocortical interactions that has been baselined to experimental observations. The present focus of our neuroscience project is constructing a clinical tool to be used for real-time diagnoses of brain function.

Similarly to the neocortex, in many complex systems, as spatial-temporal scales of observation are increased, new phenomena arise by virtue of synergistic interactions among smaller-scale entities—perhaps more properly labeled "quasientities"—which serve to explain much observed data in a parsimonious, usually mathematically aesthetic, fashion [17,32]. Many complex systems are in nonequilibrium, being driven by nonlinear and stochastic interactions of many external and internal degrees of freedom. For these systems, classical thermodynamical approaches typically do not apply. Such systems are best treated by respecting some intermediate mesoscale as "fundamental" to drive larger macroscopic processes.

It seems reasonable to approach the computation of such systems in a similar fashion. For example, the limitations of artificial neural networks (ANN's) are now better understood, especially with regard to large-scale systems [33]. There is some work in progress purposely adding noise during the learning phase to prevent overtraining to make ANN's more robust, as likely arises in some circumstances of human learning, but the noise is not considered as an integral part of ANN's. It has been previously recognized that training ANN's is equivalent to an optimal control problem [34]. The MNN, in fact more closely mimicking the neocortex, starts with mesoscopic units at its nodes. The use of VFSR permits the processing of quite general nonlinear, stochastic, and multivariate descriptions of systems, without being limited to equilibrium energy-type cost functions. Furthermore, many complex systems have structure in their "noise," which can be modeled by nonconstant diffusions, permitting better signal resolution. The use of the effective Lagrangian encases the parameters in the diffusion both in the Lagrangian itself and in the logarithm of the prefactor; this sets up a competition permitting the fitting of all parameters in the diffusion. This has been the case in the previous systems studied using VFSR.

This construction of MNN to describe nonlinear stochastic multivariate systems would not be practical without the ability to perform detailed computations. To do this, we appeal to Fokker-Planck models, known to be quite robust for many large-scale systems, to model the MNN nodes. This enables us to apply VFSR to a joint conditional probability distribution defined in the Lagrangian representation, fitting the weights in the MNN cost function over all nodes and all epochs. Then, we use the equivalent Langevin rate-equation representation to facilitate the testing or prediction phase of MNN. MNN might be useful for general path-integral calculations, but this methodology must be tested. The use of these various mathematical representations to develop this computational algorithm would not be possible without the recent development of mathematical physics techniques in functional analysis [21,35].

We use parallel processors to make this algorithm even more efficient. During "learning," VFSR lends itself well to parallelization. Blocks of random numbers are generated in parallel, and then sequentially checked to find a generating point satisfying all boundary conditions. Advantage is taken of the low ratio of acceptance to generated points typical in VFSR, to generate blocks of cost functions, and then sequentially checked to find the next best current minimum. Additionally, when fitting dynamic systems, e.g., the three physical systems examined to date, parallelization is attained by independently calculating each time epoch's contribution to the cost function. Similarly, during "prediction," blocks of random numbers are generated to support the Langevin-equation calculations, and each node is processed in parallel.

As observed in the real neocortex and with ANN's, it is expected that the MNN will afford a phenomenological computational approach even to systems not directly amenable to the algebraic model selected at any one node, e.g., a Fokker-Planck description. The merit of such applications must be decided for each class of systems investigated.

Finally, it should be possible to hardwire the MNN to process nonlinear stochastic information faithful to specific systems, e.g., the neocortex. For example, if a computer medium possesses noise that can be well modeled by an MNN, e.g., consider temperature-induced noise or quantum-mechanical uncertainty, then this circumstance might be incorporated in the construction of an MNN for a specific stochastic system.

ACKNOWLEDGMENT

I thank Bruce Rosen for several discussions, a reading of an early draft, and for demonstrating how my present fitting algorithms can be used on parallel processors.

REFERENCES

*Electronic address: ingber@alumni.caltech.edu

[1] L. Ingber, "Towards a unified brain theory," *J. Social Biol. Struct.* **4**, 211-224 (1981).

[2] L. Ingber, "Statistical mechanics of neocortical interactions. I. Basic formulation," *Physica D* **5**, 83-107 (1982).

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

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

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

[6] 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).

[7] L. Ingber, "Towards clinical applications of statistical mechanics of neocortical interactions," *Innov. Tech. Biol. Med.* **6**, 753-758 (1985).

[8] L. Ingber, "Statistical mechanics of neocortical interactions," *Bull. Am. Phys. Soc.* **31**, 868 (1986).

[9] 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*, ed. by R. Cotterill (Cambridge University Press, London, 1988), p. 513-533.

[10] 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).

[11] 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*, ed. by J. Delacour and J.C.S. Levy (Elsevier, Amsterdam, 1988), p. 387-409.

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

[13] L. Ingber, "Statistical mechanics of neocortical interactions: A scaling paradigm applied to electroencephalography," *Phys. Rev. A* **44**, 4017-4060 (1991).

[14] D. Cohen, B.N. Cuffin, K. Yunokuchi, R. Maniewski, C. Purcell, G.R. Cosgrove, J. Ives, J. Kennedy, and D. Schomer, "MEG versus EEG localization test using implanted sources in the human brain," *Ann. Neurol.* **28**, 811-817 (1990).

[15] 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*, ed. by P. Cheeseman (American Association for Artificial Intelligence, Menlo Park, CA, 1985), p. 258-264.

[16] R. Graham, "Path-integral methods on nonequilibrium thermodynamics and statistics," in *Stochastic Processes in Nonequilibrium Systems*, ed. by L. Garrido, P. Seglar and P.J. Shepherd (Springer, New York, NY, 1978), p. 82-138.

[17] H. Haken, *Synergetics*, 3rd ed. (Springer, New York, 1983).

[18] D.H. Hubel and T.N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.* **160**, 106-154 (1962).

[19] P.S. Goldman and W.J.H. Nauta, "Columnar distribution of cortico-cortical fibers in the frontal association, limbic, and motor cortex of the developing rhesus monkey," *Brain Res.* **122**, 393-413 (1977).

[20] V.B. Mountcastle, "An organizing principle for cerebral function: The unit module and the distributed system," in *The Mindful Brain*, ed. by G.M. Edelman and V.B. Mountcastle (Massachusetts Institute of Technology, Cambridge, 1978), p. 7-50.

[21] F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions* (Reidel, Dordrecht, The Netherlands, 1982).

[22] E.S. Abers and B.W. Lee, ‘‘Gauge theories,’’ *Phys. Rep. C* **9**, 1-141 (1973).

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

[24] L. Ingber and D.D. Sworder, ‘‘Statistical mechanics of combat with human factors,’’ *Mathl. Comput. Modelling* **15**, 99-127 (1991).

[25] L. Ingber, ‘‘Statistical mechanical aids to calculating term structure models,’’ *Phys. Rev. A* **42**, 7057-7064 (1990).

[26] 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**, 77-98 (1991).

[27] L. Ingber, ‘‘Very fast simulated re-annealing,’’ *Mathl. Comput. Modelling* **12**, 967-973 (1989).

[28] L. Ingber and B. Rosen, ‘‘Genetic algorithms and very fast simulated reannealing: A comparison,’’ *Mathl. Comput. Modelling* **16**, 87-100 (1992).

[29] W.D. Hillis, *The Connection Machine* (The MIT Press, Cambridge, MA, 1985).

[30] H. Szu and R. Hartley, ‘‘Fast simulated annealing,’’ *Phys. Lett. A* **122**, 157-162 (1987).

[31] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, ‘‘Optimization by simulated annealing,’’ *Science* **220**, 671-680 (1983).

[32] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Wiley, New York, NY, 1973).

[33] Defense Advanced Research Projects Agency, *DARPA Neural Network Study* (AFCEA Int’l., Fairfax, VA, 1988).

[34] A. Kehagias, ‘‘Optimal control for training: The missing link between hidden Markov models and connectionist networks,’’ *Mathl. Comput. Modelling* **14**, 284-289 (1990).

[35] R. Graham, ‘‘Lagrangian for diffusion in curved phase space,’’ *Phys. Rev. Lett.* **38**, 51-53 (1977).