%A L. Ingber
%T Statistical mechanics of neocortical interactions: Path-integral evolution of short-term memory
%J Phys. Rev. E
%V 49
%N 5B
%D 1994
%P 4652-4664

Statistical mechanics of neocortical interactions: Path-integral evolution of short-term memory

Lester Ingber*

(Received 13 September 1993; revised manuscript received 1 February 1994)

Lester Ingber Research, P.O. Box 857, McLean, Virginia 22101

Previous papers in this series of statistical mechanics of neocortical interactions (SMNI) have detailed a development from the relatively microscopic scales of neurons up to the macroscopic scales as recorded by electroencephalography (EEG), requiring an intermediate mesocolumnar scale to be developed at the scale of minicolumns ($\approx 10^2$ neurons) and macrocolumns ($\approx 10^5$ neurons). Opportunity was taken to view SMNI as sets of statistical constraints, not necessarily describing specific synaptic or neuronal mechanisms, on neuronal interactions, on some aspects of short-term memory (STM), e.g., its capacity, stability, and duration. A recently developed C-language code, PATHINT, provides a non-Monte Carlo technique for calculating the dynamic evolution of arbitrary-dimension (subject to computer resources) nonlinear Lagrangians, such as derived for the two-variable SMNI problem. Here, PATHINT is used to explicitly detail the evolution of the SMNI constraints on STM.

PACS Nos.: 87.10.+e, 05.40.+j, 02.50.-r, 02.70.-c

*Electronic address: ingber@alumni.caltech.edu

I. INTRODUCTION

A. SMNI modeling

To learn more about complex systems, functional models are inevitably formed to represent huge sets of data. In nature, complex systems often present different phenomena at different scales. In this context, a plausible model of statistical mechanics of neocortical interactions (SMNI) has been developed over the past decade [1-16]. Some recent experimental work further justifies the SMNI mathematical development of the microscopic scale into mesocolumns [17]. The focus of SMNI is to explain phenomena at spatial scales of millimeters to centimeters, much larger than spatial scales of neuronal interactions. In this paper, to place all calculations in this context, it is helpful to consider the utility of SMNI as directed towards phenomena such as measured by electroencephalography (EEG). For example, fitted SMNI functional forms to EEG data may help to explicate some underlying biophysical mechanisms responsible for the normal and abnormal behavioral states being investigated [13,15].

However, like many nonlinear nonequilibrium systems, in the course of the SMNI development from the relatively microscopic scales of neurons up to the macroscopic scales of EEG, an intermediate mesocolumnar scale had to be developed at the scale of minicolumns ($\approx 10^2$ neurons) and macrocolumns ($\approx 10^5$ neurons). Then, opportunity was taken to view SMNI as sets of statistical constraints, not necessarily describing specific synaptic or neuronal mechanisms, on neuronal interactions, e.g., on some aspects of short-term memory (STM).

A quite different approach to neuronal systems is taken by artificial neural networks (ANN). Both ANN and SMNI structures are represented in terms of units with algebraic properties greatly simplifying specific realistic neuronal components [18]. Of course, there is a clear logical difference between considering a small ensemble of simple ANN units (each unit representing an “average” neuron) to study the properties of small ensembles of neurons, versus considering distributions of interactions between model neurons to develop ensembles of units (each unit representing a column of neurons) developed by SMNI to study properties of large ensembles of columns. Only the latter has a chance for any statistical justification. Unlike SMNI, ANN models may yield insights into specific mechanisms of learning, memory, retrieval, and information processing among small ensembles of model neurons, etc. However, consider that there are several million neurons located under a cm^2 area of neocortical surface. Current estimates are that 1 to several percent of coherent neuronal firings may account for the amplitudes of electric potential measured on the scalp. This translates into measuring firings of hundreds of thousands of neurons as contributing to activity measured under a typical electrode. Even when EEG recordings are made directly on the brain surface, tens of thousands of neurons are contributing to activity measured under electrodes. ANN models cannot approach the order of magnitude of neurons participating in phenomena at the scale of EEG, just as neither ANN nor SMNI can detail relatively smaller scale activity at the membrane or atomic levels. Attempts by ANN to do so likely would require statistical interpretations such as are made by SMNI; otherwise the output of the models would just replace the data collected from huge numbers of neuronal firings—a regression from 20th century science back to empiricism. Thus, as is the case in many physical sciences, the SMNI approach is to perform prior statistical analyses up to the scale of interest (here at EEG scales). The ANN approach must perform statistical analyses after processing its units.

While ANN models use simplified algebraic structures to represent real neurons, SMNI models develop the statistics of large numbers of realistic neurons representing huge numbers of synaptic interactions—there are 10^4 to 10^5 synapses per neuron. Furthermore, unlike most ANN approaches, SMNI accepts constraints on all its macrocolumnar averaged parameters to be taken from experimentally determined ranges of synaptic and neuronal interactions; there are no unphysical parameters. The stochastic and nonlinear nature of SMNI development is directly derived from experimentally observed synaptic interactions and from the mathematical development of observed minicolumns and macrocolumns of neurons. SMNI has required the use of mathematical physics techniques first published in the late 1970s in the context of developing an approach to multivariate nonlinear nonequilibrium statistical mechanics.

B. Outline of paper

Section II gives a short description of SMNI as it has been focused to EEG analyses [13] and generic algorithms for nonlinear systems [14]. Section III gives a short description of that part of the SMNI algebra of the mesoscopic scale relevant to this paper, where confirmation of experimental data was obtained with systematics of STM phenomena. Section IV presents a path-integral algorithm, PATHINT, applied to SMNI, with detailed calculations of the evolution of STM. Section V concludes with a brief outlook to future work that now can be reasonably accomplished given the PATHINT algorithm together with the adaptive simulated annealing (ASA) code [19], previously called very fast simulated reannealing (VFSR) [20], which was used to fit EEG data [13].

II. OUTLINE OF SMNI

A. Top-down versus bottom-up

In order to detail a model of EEG phenomena, it is useful to seek guidance from "top-down" models; e.g., the nonlinear string model representing nonlinear dipoles of neuronal columnar activity [21]. In order to construct a more detailed "bottom-up" model that can give reasonable algebraic functions with physical parameters to be fitted by data, a wealth of empirical data and modern techniques of mathematical physics across multiple scales of neocortical activity are developed up to the scale described by the top-down model. At each of these scales, reasonable procedures and submodels for climbing from scale to scale are derived. Each of these submodels was tested against some experimental data to see if the theory was on the right track. For example, at the mesoscopic scale the consistency of SMNI was checked with known aspects of visual and auditory STM, e.g., the 4 ± 2 and 7 ± 2 STM capacity rules, respectively, the detailed duration and stability of such states, and the primacy versus recency rule of error rates of learned items in STM [4,6]. At the macroscopic scale, SMNI consistency was checked with most stable frequencies being in the high α to low β range, and the velocities of propagation of information across minicolumns being consistent with other experimental data [3,5]. SMNI has demonstrated that the currently accepted dipole EEG model can be derived as the Euler-Lagrange equations of an electric-potential Lagrangian, describing the trajectories of most likely states, making it possible to return to the top-down EEG model, but now with a derivation and detailed structure given to the dipole model [12,13]. The SMNI approach, of fitting scaled nonlinear stochastic columnar activity directly to EEG data, goes beyond the dipole model, making it possible to extract more signal from noise.

The theoretical and experimental importance of specific scaling of interactions in the neocortex has been quantitatively demonstrated: It has been shown that the explicit algebraic form of the probability distribution for mesoscopic columnar interactions is driven by a nonlinear threshold factor of the same form taken to describe microscopic neuronal interactions, in terms of electrical-chemical synaptic and neuronal parameters all lying within their experimentally observed ranges; these threshold factors largely determine the nature of the drifts and diffusions of the system. This mesoscopic probability distribution has successfully described STM phenomena and, when used as a basis to derive the most likely trajectories using the Euler-Lagrange variational equations, it also has described the systematics of EEG phenomena. More recently, the mesoscopic form of the full probability distribution has been taken more seriously for macroscopic interactions, deriving macroscopic drifts and diffusions linearly related to sums of their (nonlinear) mesoscopic counterparts, scaling its variables to describe interactions among regional interactions correlated with observed electrical activities measured by electrode recordings of scalp EEG, with apparent success [13]. These results give strong quantitative support for an accurate intuitive picture, portraying neocortical interactions as having common algebraic or physics mechanisms that scale across quite disparate spatial scales and functional or behavioral phenomena, i.e., describing interactions among neurons, columns of neurons, and regional masses of neurons.

B. Generic application

The SMNI methodology also defines an algorithm to construct a mesoscopic neural network (MNN), based on realistic neocortical processes and parameters, to record patterns of brain activity and to compute the evolution of this system [14]. MNN makes it possible to add a finer minicolumnar scale to the explicit SMNI development at the mesoscopic and regional scales.

Furthermore, this new 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 [22], ASA [19,20], and parallel-processing computation.

MNN generalizes the ASA code [20]. When applied to systems like SMNI, ASA fits short-time probability distributions to observed data, using a maximum likelihood technique on the "effective" Lagrangian (including the exponential prefactor). This algorithm 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 [23], $T_i = T_0/k$, and much faster than Boltzmann annealing [24], $T_i = T_0/\ln k$. To be more specific, the kth estimate of parameter α^i,

$$\alpha^i_k \in [A_i, B_i] \ , \tag{1}$$

is used with the random variable x^i to get the $(k+1)$th estimate,

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

$$x^i \in [-1, 1] \ . \tag{2}$$

The generating function is defined as

$$g_T(x) = \prod_{i=1}^{D} \frac{1}{2 \ln(1 + 1/T_i)(|x^i| + T_i)} \equiv \prod_{i=1}^{D} g^i_T(x^i) \ ,$$

$$T_i = T_{i0} \exp(-c_i k^{1/D}) \ . \tag{3}$$

The cost function C used here is defined by

$$C = Ldt + \frac{1}{2} \ln(2\pi dt) - \ln(g) \ , \tag{4}$$

in terms of the Lagrangian L and the determinant of the metric g.

As discussed in the Conclusion, ASA is a natural partner with PATHINT for fitting and evolving multivariate nonlinear Gaussian-Markovian systems such as described by SMNI. ASA has been extremely useful to many researchers world-wide in many other kinds of applications.

III. SMNI MODEL OF STM

A. Basic assumptions

The most detailed and dramatic application of the theory outlined here is to predict stochastic bounds for the phenomena of human STM capacity during focused selective attention [4,6,25-27], transpiring on the order of tenths of a second to seconds, limited to the retention of 7 ± 2 items [28]. This is true even for apparently exceptional memory performers who, while they may be capable of more efficient encoding and retrieval of STM, and while they may be more efficient in "chunking" larger patterns of information into single items, nevertheless are limited to a STM capacity of 7 ± 2 items [29]. Mechanisms for various STM phenomena have been proposed across many spatial scales [30]. This "rule" is verified for acoustical STM, but for visual or semantic STM, which typically require longer times for rehearsal in an hypothesized articulatory loop of individual items, STM capacity appears to be limited to 2–4 [31]. Another interesting phenomenon of STM capacity explained by SMNI is the primacy versus recency effect in STM serial processing, wherein first-learned items are recalled most error-free, with last-learned items still more error-free than those in the middle [32].

The basic assumption being made is that a pattern of neuronal firing that persists for many τ cycles is a candidate to store the ‘‘memory’’ of activity that gave rise to this pattern. If several firing patterns can simultaneously exist, then there is the capability of storing several memories. The short-time probability distribution derived for the neocortex is the primary tool to seek such firing patterns. Since this distribution is exponentially sensitive to (minus) the Lagrangian function $\underline{L}$, sometimes it is more convenient to deal directly with $\underline{L}$, whereby its minima specify the most likely states that can be sustained at a given time. Then, several important features of these patterned states can be investigated, as is done for other physical systems [33], e.g., the evolution of these states, the ‘‘time of first passage’’ to jump from one state to another state, hysteresis between states that have different depths (values of the Lagrangian at these local minima), the stability of each state under external forces, etc.

B. SMNI mesoscopic propagator

As is found for most nonequilibrium systems, e.g., for lasers, chemical systems, fluids, and ecological systems [33,34], a mesoscopic scale is required to formulate the statistical mechanics of the microscopic system, from which the macroscopic scale can be developed [33]. The neocortex is particularly interesting in this context in that a clear scale for the mesoscopic system exists, both anatomically (structurally) and physiologically (functionally). ‘‘Minicolumns’’ of about $N \approx 110$ neurons (about 220 in the visual cortex) comprise modular units vertically oriented relative to the warped and convoluted neocortical surface throughout most, if not all, regions of the neocortex [35-40]. Clusters of about 100 neurons have been deduced to be reasonable from other considerations as well [41]. Since the short-ranged interactions between neurons take place within ~1 mm, which is the extent of a ‘‘macrocolumn’’ comprising $\sim 10^3$ minicolumns of $N^* \approx 10^5$ neurons, and since macrocolumns also exhibit rather specific information-processing features, this theory has retained the divergence-convergence of macrocolumn-minicolumn, efferent-afferent interactions by considering domains of minicolumns as having similar synaptic interactions within the extent of a macrocolumn. This macrocolumnar-averaged minicolumn is designated in this theory as a ‘‘mesocolumn.’’

This being the observed situation, it is interesting that $N \approx 10^2$ is just the right order of magnitude to permit a formal analysis using methods of mathematical physics just developed for statistical systems in the late 1970s [22,42]. N is small enough to permit nearest-neighbor interactions to be formulated, such that interactions between mesocolumns are small enough to be considered gradient perturbations on otherwise independent mesocolumnar firing states. This is consistent with rather continuous spatial gradient interactions observed among columns [43], and with the basic hypothesis that nonrandom differentiation of properties among broadly tuned individual neurons coexists with functional columnar averages representing superpositions of patterned information [44]. This is a definite mathematical convenience; otherwise, a macrocolumn of $\sim 10^3$ minicolumns would have to be described by a system of minicolumns with up to 16th-order next-nearest neighbors. (Consider 1000 minicolumns spread out in a two-dimensional grid about 33 by 33 minicolumns, and focus attention on the center minicolumn.) The MNN algorithm described above can replace this nearest-neighbor approximation, introducing an additional finer scale of direct minicolumnar interactions.

Also, N is large enough to permit the derived binomial distribution of afferent minicolumnar firing states to be well approximated by a Gaussian distribution, a luxury not afforded an ‘‘average’’ neuron, even in this otherwise similar physical context. Finally, mesocolumnar interactions are observed to take place via one to several relays of neuronal interactions, so that their time scales are similarly $\tau \approx 5 - 10$ msec. Even after statistically shaping the microscopic system, the parameters of the mesoscopic system are still macrocolumnar-averaged synaptic parameters, i.e., reflecting the statistics of millions of synapses with regard to their chemical and electrical properties. Explicit laminar circuitry, and more complicated synaptic interactions, e.g., dependent on all combinations of presynaptic and postsynaptic firings, can be included without loss of detailed analysis [3]. The mathematical development of mesocolumns establishes a mesoscopic Lagrangian $\underline{L}$, which may be considered as a ‘‘cost function’’ with variables M^G, $\dot{M}^G$, and ∇M^G, and with parameters defined by the macrocolumnar-averaged chemical-electrical entities developed below.

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 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 $M^I = \sum_{j\in I} \sigma_j$.

$$P = \prod_G P^G[M^G(r;t+\tau)|M^{\bar{G}}(r';t)]$$

$$= \sum_{\sigma_j} \delta\left(\sum_{j\in E} \sigma_j - M^E(r;t+\tau)\right)\delta\left(\sum_{j\in I} \sigma_j - M^I(r;t+\tau)\right)\prod_j^N p_{\sigma_j}$$

$$\approx \prod_G (2\pi\tau g^{GG})^{-1/2} \exp(-N\tau \underline{L}^G) , \tag{5}$$

where the final form is derived using the fact that $N > 100$. $\bar{G}$ represents contributions from both E and I sources. This defines the Lagrangian, in terms of its first-moment drifts g^G, its second-moment diffusion matrix $g^{GG'}$, and its potential $\underline{V}'$, all of which depend sensitively on threshold factors F^G,

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

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

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

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

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

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

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

$$a^G_{G'} = \frac{1}{2} A^G_{G'} + B^G_{G'} , \tag{6}$$

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. For neocortex, due to chemical independence of excitatory and inhibitory interactions, the diffusion matrix $g^{GG'}$ is diagonal.

C. Previous SMNI treatment of STM

1. STM capacity

Three cases of neuronal firings were considered [4]. Since STM duration is still long relative to τ, stationary solutions of $\bar{\underline{L}}$, derived from $\underline{L}$ in Eq. (6), were investigated to determine how many stable minima, $\ll \bar{M}^G \gg$, may simultaneously exist within this duration. Also, individual mesocolumns were studied. I.e., take the uniform limit of $\dot{\bar{M}}^G = 0 = \nabla\bar{M}^G$. Although the $\dot{\bar{M}}^G = 0$ limit should only be taken for the midpoint-discretized Lagrangian $\underline{L}_F$, this is a small difference here [4]. Section IV below will use these results as a zeroth order basis for more detailed path-integral calculations.

A model of dominant inhibition describes how minicolumnar firings are suppressed by their neighboring minicolumns. For example, this could be effected by developing nearest-neighbor (NN) mesocolumnar interactions [3], but the averaged effect is established by inhibitory mesocolumns (IC) by setting $A_E^I = A_I^E = 2A_E^E = 0.01N^*/N$. Since there appears to be relatively little I—I connectivity, set $A_I^I = 0.0001N^*/N$. The background synaptic noise is taken to be $B_I^E = B_E^I = 2B_E^E = 10B_I^I = 0.002N^*/N$. As minicolumns are observed to have ~110 neurons (the visual cortex appears to have approximately twice this density) [40], and as there appear to be a predominance of E over I neurons [45], here take $N^E = 80$ and $N^I = 30$. Use $N^*/N = 10^3$, $J_G = 0$ (absence of long-ranged interactions), and V^G, $v_{G'}^G$, and $\phi_{G'}^G$ as estimated previously, i.e., $V^G = 10$ mV, $|v_{G'}^G| = 0.1$ mV, $\phi_{G'}^G = 0.1$ mV. The "threshold factors" $F_{\rm IC}^G$ for this IC model are then

$$F_{\rm IC}^E = \frac{0.5\bar{M}^I - 0.25\bar{M}^E + 3.0}{\pi^{1/2}(0.1\bar{M}^I + 0.05\bar{M}^E + 9.80)^{1/2}} ,$$

$$F_{\rm IC}^I = \frac{0.005\bar{M}^I - 0.5\bar{M}^E - 45.8}{\pi^{1/2}(0.001\bar{M}^I + 0.1\bar{M}^E + 11.2)^{1/2}} . \tag{7}$$

In the prepoint-discretized deterministic limit, the threshold factors determine when and how smoothly the "step functions" $\tanh F_{\rm IC}^G$ in $g^G(t)$ change $M^G(t)$ to $M^G(t+\theta)$. $F_{\rm IC}^I$ will cause afferent $\bar{M}^I$ to fire for most of its values, as $\bar{M}^I \sim -N^I \tanh F_{\rm IC}^I$ will be positive for most values of $\bar{M}^G$ in $F_{\rm IC}^I$, which is already weighted heavily with a term -45.8. Looking at $F_{\rm IC}^E$, it is seen that the relatively high positive values of efferent $\bar{M}^I$ require at least moderate values of positive efferent $\bar{M}^E$ to cause firings of afferent $\bar{M}^E$.

It is discovered that more minima of $\overline{L}$ are created, or "restored," if the numerator of F^G contains terms only in $\bar{M}^G$, tending to center $\overline{L}$ about $\bar{M}^G = 0$. Of course, any mechanism producing more as well as deeper minima is statistically favored. However, this particular "centering" mechanism has plausible support: $M^G(t+\tau) = 0$ is the state of afferent firing with highest statistical weight. I.e., there are more combinations of neuronal firings, $\sigma_j = \pm 1$, yielding this state than any other $M^G(t+\tau)$; e.g., $\sim 2^{N^G+1/2}(\pi N^G)^{-1/2}$ relative to the states $M^G = \pm N^G$. Similarly, $M^{*G}(t)$ is the state of efferent firing with highest statistical weight. Therefore, it is natural to explore mechanisms that favor common highly weighted efferent and afferent firings in ranges consistent with favorable firing threshold factors $F^G \approx 0$.

The centering effect of the IC model of dominant inhibition, labeled here as the IC′ model, is quite easy for the neocortex to accommodate. For example, this can be accomplished simply by readjusting the synaptic background noise from B_E^G to B'^G_E,

$$B'^G_E = \frac{V^G - (\frac{1}{2}A_I^G + B_I^G)v_I^G N^I - \frac{1}{2}A_E^G v_E^G N^E}{v_E^G N^G} \tag{8}$$

for both $G = E$ and $G = I$. This is modified straightforwardly when regional influences from long-ranged firings $M^{\ddagger E}$ are included [13]. In general, B_E^G and B_I^G (and possibly A_E^G and A_I^G due to actions of neuromodulators, and J_G or $M^{\ddagger E}$ constraints from long-ranged fibers) are available to force the constant in the numerator to zero, giving an extra degree(s) of freedom to this mechanism. (If B'^G_E would be negative, this leads to unphysical results in the square-root denominator of F^G. Here, in all examples where this occurs, it is possible to instead find positive B'^G_I to appropriately shift the numerator of F^G.) In this context, it is experimentally observed that the synaptic sensitivity of neurons engaged in selective attention is altered, presumably by the influence of chemical neuromodulators on postsynaptic neurons [46].

By this centering mechanism, $B'^E_E = 1.38$ and $B'^I_I = 15.3$, and $F_{\rm IC}^G$ is transformed to $F_{\rm IC'}^G$,

$$F_{\rm IC'}^E = \frac{0.5\bar{M}^I - 0.25\bar{M}^E}{\pi^{1/2}(0.1\bar{M}^I + 0.05\bar{M}^E + 10.4)^{1/2}} ,$$

$$F^I_{\mathrm{IC}'} = \frac{0.005\bar{M}^I - 0.5\bar{M}^E}{\pi^{1/2}(0.001\bar{M}^I + 0.1\bar{M}^E + 20.4)^{1/2}} . \tag{9}$$

Aside from the enforced vanishing of the constant terms in the numerators of $F^G_{\mathrm{IC}'}$, the only other change in $F^G_{\mathrm{IC}'}$ relative to F^G_{IC} is to moderately affect the constant terms in the denominators. This increases the number of minima of $\tau \underline{L}_{\mathrm{IC}'}$ to 4. The two minima clustered close to the origin are no longer discernible for $\tau \underline{L}_{\mathrm{IC}'} > 0.03$.

The other "extreme" of normal neocortical firings is a model of dominant excitation, effected by establishing excitatory mesocolumns (EC) by using the same parameters $\{ B^G_{G'}, v^G_{G'}, \phi^G_{G'}, A^I_I \}$ as in the IC model, but setting $A^E_E = 2A^I_E = 2A^E_I = 0.01N^*/N$. This yields

$$F^E_{\mathrm{EC}} = \frac{0.25\bar{M}^I - 0.5\bar{M}^E - 24.5}{\pi^{1/2}(0.05\bar{M}^I + 0.10\bar{M}^E + 12.3)^{1/2}} ,$$

$$F^I_{\mathrm{EC}} = \frac{0.005\bar{M}^I - 0.25\bar{M}^E - 25.8}{\pi^{1/2}(0.001\bar{M}^I + 0.05\bar{M}^E + 7.24)^{1/2}} . \tag{10}$$

The negative constant in the numerator of F^I_{EC} inhibits afferent $\bar{M}^I$ firings. Although there is also a negative constant in the numerator of F^E_{EC}, the increased coefficient of $\bar{M}^E$ (relative to its corresponding value in F^E_{IC}), and the fact that $\bar{M}^E$ can range up to $N^E = 80$, readily permits excitatory firings throughout most of the range of $\bar{M}^E$. This permits three minima.

Applying the centering mechanism to EC, $B'^E_I = 10.2$ and $B'^I_I = 8.62$. The net effect in $F^G_{\mathrm{EC}'}$, in addition to removing the constant terms in the numerators of F^G_{EC}, is to change the constant terms in the denominators: 12.3 in F^E_{EC} is changed to 17.2 in $F^E_{\mathrm{EC}'}$, and 7.24 in F^I_{EC} is changed to 12.4 in $F^I_{\mathrm{EC}'}$. Now six prominent minima are possible along a line through $\bar{M}^G = 0$, and two others are at $\bar{M}^G = \pm N^G$. Each pair of minima above and below the $\bar{M}^I = 0$ axis merge into single minima for $\tau \underline{L}_{\mathrm{EC}'} > 0.02$, and these lose resolution for $\tau \underline{L}_{\mathrm{EC}'} > 0.03$.

Now it is natural to examine a balanced case intermediate between IC and EC, labeled BC. This is accomplished by changing $A^E_E = A^I_E = A^E_I = 0.005N^*/N$. This yields

$$F^E_{\mathrm{BC}} = \frac{0.25\bar{M}^I - 0.25\bar{M}^E - 4.50}{\pi^{1/2}(0.050\bar{M}^E + 0.050\bar{M}^I + 8.30)^{1/2}} ,$$

$$F^I_{\mathrm{BC}} = \frac{0.005\bar{M}^I - 0.25\bar{M}^E - 25.8}{\pi^{1/2}(0.001\bar{M}^I + 0.050\bar{M}^E + 7.24)^{1/2}} . \tag{11}$$

Three minima are possible, on the boundaries of $\bar{M}^G$ space.

Applying the centering mechanism to BC, $B'^E_E = 0.438$ and $B'^I_I = 8.62$. The net effect in $F^G_{\mathrm{BC}'}$, in addition to removing the constant terms in the numerators of F^G_{BC}, is to change the constant terms in the denominators: 8.30 in F^E_{BC} is changed to 7.40 in $F^E_{\mathrm{BC}'}$, and 7.24 in F^I_{BC} is changed to 12.4 in $F^I_{\mathrm{BC}'}$. Now ten minima are possible. The nine minima along the diagonal line lose resolution for $\tau \underline{L}_{\mathrm{BC}'} > 0.01$ above $\bar{M}^I = 0$ and for $\tau \underline{L}_{\mathrm{BC}'} > 0.02$ below $\bar{M}^I = 0$.

The effects of using the full Feynman Lagrangian $\underline{L}_F$ were considered, including all the Riemannian and other nonlinear corrections discussed in previous SMNI papers. The net effect is to slightly raise the threshold at which minima dissipate, to about $\tau \underline{L}_{\mathrm{BC}'} \geq 0.03$, which is relevant for the duration of STM, discussed subsequently. However, the minima structure is essentially the same.

If N^* is scaled larger or smaller, this effectively scales $A^G_{G'} = A^{*G}_{G'} N^*/N$ and $B^G_{G'} = B^{*G}_{G'} N^*/N$, disturbing the relatively sensitive balance that permits a few percent of efferent firings to affect their afferents. Then, the number of possible minima is typically reduced to one or two. If N is scaled larger or smaller, the number of minima is altered and the duration of STM is affected, as discussed subsequently. However, for N still in the range of a few hundred, the number of possible minima is not severely reduced. The case $N = 220$, e.g., the visual cortex was considered: For model BC′, the number of prominent minima found is 11, but they form clusters, with higher peaks between clusters than between minima within a cluster. The larger N sharpens the minima and therefore the resolution of visual information processing.

The sharpness of the tanh F^G step-function contribution to the mean firing is sensitive to a factor of $N^{\frac{1}{2}}$ in F^G. Additionally, the strength of coupling between mesocolumns scales as $N^{3/2}$. Thus the neuronal size of mesocolumns directly affects the breadth and depth of the information processing capability of the neocortex. It is interesting to note that the human visual cortex, which may be assumed to require the finest tuning in the neocortex, is unique in having twice the number of neurons per minicolumn than other regions of the neocortex [40].

2. STM stability and duration

The calculation of stability and time of duration in most likely states of firing starts by using the differential-equation Hamiltonian formulation of the path-integral Lagrangian, called the Fokker-Planck equation. The Fokker-Planck equation for the region Ω is

$$\frac{\partial \tilde{P}}{\partial t} \approx \Omega^{-1} \int d^2 r [\frac{1}{2} (g^{GG'} \tilde{P})_{,GG'} - (g^G \tilde{P})_{,G} + N\underline{V}' \tilde{P}] \ ,$$

$$(\cdots)_{,G} \equiv \partial(\cdots)/\partial M^G \ . \tag{12}$$

The true Fokker-Planck equation is actually more general, e.g., if long-ranged spatial structures are included, where the independent variables M^G are fields which themselves may depend on space and time coordinates. The above equation is derived in the nearest-neighbor approximation from the general equation using functional derivatives [4],

$$\partial(\cdots)/\partial M^G \rightarrow \delta(\cdots)/\delta M^G \ ,$$

$$\delta(\cdots)/\delta M^G = (\cdots)_{,G} - \nabla_i (\cdots)_{,\nabla_i G} + \nabla_i^2 (\cdots)_{,\nabla_i^2 G} \ , \tag{13}$$

where we have used the compacted notation introduced previously [4].

An estimate of a stationary solution P_{stat} to the Fokker-Planck differential equation for the probability distribution P of M^G firings for an uncoupled mesocolumn, i.e., $\underline{V}' = 0$, is given by the stationary limit of the short-time propagator,

$$P_{\mathrm{stat}} \approx N_{\mathrm{stat}} g^{1/2} \exp(-CN\tau\underline{L}) \ ,$$

$$g = \det(g^{GG'})^{-1} \equiv \det(g_{GG'}) = g_{EE} g_{II} \ , \tag{14}$$

where N_{stat} and C are constant factors. An estimate of the approximation made is estimated by seeking values of constants C, such that the stationary Fokker-Planck equation is satisfied exactly. Contour plots of C versus $\bar{M}^G$ demonstrate that there exists real positive C which may only range from $\sim 10^{-1}$ to ~ 1, for which there exists unbroken contours of C which pass through or at least border the line of minima [6]. At each point $\bar{M}^G$, this leaves a quadratic equation for C to be solved. Dropping the $g^{1/2}$ factor results in C not being real throughout the domain of $\bar{M}^G$.

Thus we have defined an approximate solution with potential $N^2 \underline{L} = \int A \, dM$, drift A, and diffusion N/τ. Stability of transient solutions, defined for $\underline{\delta} M^G$ about a stationary state by

$$\underline{\delta} \dot{M}^G \approx -A_{,G} \underline{\delta} M^G = -N^2 \underline{L}_{,GG} \underline{\delta} M^G \ , \tag{15}$$

is therefore equivalent to $\ll \bar{M} \gg$ being a minimum of $\underline{L}$.

Since the minima of the Lagrangian lie deep in a valley along a line, a parabolic trough, the time for first passage, t_{vp}, is estimated in analogy to a one-dimensional system as [47]

$$t_{vp} \approx \pi N^{-2} [|\underline{L}_{,GG'}(\ll \bar{M} \gg_p)| \underline{L}_{,GG'}(\ll \bar{M} \gg_v)]^{-1/2}$$

$$\times \exp \{ CN\tau[\bar{L}(\ll \bar{M} \gg_p) - \bar{L}(\ll \bar{M} \gg_v)] \} , \tag{16}$$

where $\ll \bar{M} \gg_v$ is the minimum at the valley of $\bar{L}$ in question, and $\ll \bar{M} \gg_p$ is the maximum at a peak separating two minima. These equations are reasonable but crude estimates, and future numerical work must be done to detail the extent of their validity. Section IV begins this program by transforming to axes that can take advantage of this parabolic trough.

The exponential factor can be quite large in some instances, and quite small in others. As noted previously [3], differences in $\bar{L}$ from valleys to peaks are still large relative to the Riemannian correction terms and relative to differential spatial-temporal contributions, thereby permitting this simpler analysis. However, values of $\tau\bar{L}$ at maxima separating the far minima may be greater than 1, thereby yielding a very large t_{vp}, typical of many physical systems undergoing hysteresis [3]. Relaxation times t_r about this stationary state are estimated by $|g^G_{,G}|^{-1}$ [47], and are on the order of τ. For changes ΔZ in synaptic parameters $Z = \{ A^*_{jk}, B^*_{jk}, V_j, v_{jk}, \phi_{jk}, N^{*G} \}$ that transpire within a Δt of several tenths of a second to seconds, e.g., during typical attention spans, hysteresis is more probable than simple jumps between minima if the following inequalities are satisfied.

To address the issue of limited capacity of STM, it is reasonable to require that within time spans of tenths of a second to tens of seconds, simple jumps among minima are more probable than hysteresis. This permits all minima to be readily accessible during STM duration, in any ordering [29], at least more so than if hysteresis were more probable. In agreement with this empirical requirement, it is found that $\tau[\bar{L}(\ll \bar{M} \gg_p) - \bar{L}(\ll \bar{M} \gg_v)] \sim 0.01$—$0.03$ for these models using empirical values for synaptic parameters. Then for $|\tau\bar{L}_{,GG'}| \sim 10^{-3}$, $t_{vp} \sim 10\tau$—100τ, on the order of several tenths of a second to a second. Use of the full Feynman Lagrangian $\bar{L}_F$ increases t_{vp} slightly. For these relatively short t_{vp} the second inequality above is violated, and simple jumps are more probable than hysteresis, as required for STM.

Under conditions of serial processing, the deeper valleys of $\bar{L}$ representing the more likely firing states will be occupied first. In all cases considered here, some valleys are deeper than the others. This implies that the last several items in STM should be harder to encode (learn) and retain, with the possible exception of the last one or two items, which represent the most recent shifting of firing patterns $\bar{M}^G$ to these minima $\ll \bar{M} \gg_v$ of $\bar{L}$. These conclusions are consistent with empirical observations, and are obtained independent of any other rehearsal mechanisms that may exist.

Calculations in these models establish that the prefactor most often is $\sim \tau$. However, points close to the corners $\bar{M}^G = \pm(N^E, N^I)$ have much more rapid variations. Therefore, minima at these corners, even when $\tau\bar{L}(\ll \bar{M} \gg_p) \sim 0.01$—$0.03$, because of their sharp peaks, typically have t_{vp} on the order of tens of seconds to jump to minima clustered on the diagonal. This is within the range where hysteresis is more probable for these minima. Therefore, minima at the corners of $\bar{M}^G$ space most likely do not contribute to STM, bringing the number of available minima down to 7 ± 2 as empirically observed.

These previous results were obtained by looking at the space-time independent "uniform" Lagrangian and the Fokker-Planck associated with this Lagrangian in the continuous M^G limit. The present study does not require any of these assumptions, but deals directly with the originally derived mesoscopic propagator.

IV. PATH-INTEGRAL CALCULATION OF EVOLUTION OF STM

A. Path-integral algorithm

The path-integral C-language code, PATHINT, developed by the author calculates the long-time probability distribution from the Lagrangian, e.g., as fit by the ASA 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 [48-50], which was extended to two-dimensional problems [51]. The code used here was developed for use in arbitrary dimensions, with additional code to handle general Neumann and Dirichlet conditions, as well as the possibility of including time-dependent potentials, drifts, and diffusions. Such calculations are useful for many kinds of financial instruments [52,53]. Monte Carlo algorithms for path integrals are well known to have extreme difficulty in evolving nonlinear systems with multiple optima [54].

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

$$P(M;t+\Delta t) = \int dM'[g_s^{1/2}(2\pi\Delta t)^{-1/2}\exp(-L_s\Delta t)]P(M';t)$$

$$= \int dM' G(M,M';\Delta t)P(M';t) ,$$

$$P(M;t) = \sum_{i=1}^{N} \pi(M-M_i)P_i(t) ,$$

$$\pi(M-M_i) = \begin{cases} 1\,, & (M_i - \frac{1}{2}\Delta M_{i-1}) \le M \le (M_i + \frac{1}{2}\Delta M_i)\ , \\ 0\,, & \text{otherwise}\,. \end{cases} \tag{17}$$

This yields

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

$$T_{ij}(\Delta t) = \frac{2}{\Delta M_{i-1} + \Delta M_i} \int_{M_i-\Delta M_{i-1}/2}^{M_i+\Delta M_i/2} dM \int_{M_j-\Delta M_{j-1}/2}^{M_j+\Delta M_j/2} dM' G(M,M';\Delta t) . \tag{18}$$

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

This histogram procedure has been extended to two dimensions, i.e., using a matrix T_{ijkl} [51]. Explicit dependence of L on time t also can be included without complications. Care must be used in developing the mesh in ΔM^G, which is strongly dependent on the diagonal elements of the diffusion matrix, e.g.,

$$\Delta M^G \approx (\Delta t g^{|G||G|})^{1/2} . \tag{19}$$

Presently, this constrains the dependence of the covariance of each variable to be a (nonlinear) function of that variable, in order to present a straightforward rectangular underlying mesh.

Since integration is inherently a smoothing process [52], 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. For example, 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 conditions on the time and variable meshes can be derived [48]. The time slice essentially is determined by $\theta \le \bar{L}_\nu^{-1}$, where $\bar{L}_\nu$ is the uniform Lagrangian, respecting ranges giving the most important contributions to the probability distribution P. Thus, θ is roughly measured by the diffusion divided by the square of the drift.

The code here was tested against the test problems given in previous one-dimensional systems [48,49], and it was established that the method of images for both Dirichlet and Neumann boundary conditions is as accurate as boundary element methods for the systems investigated. Two-dimensional runs were tested by using cross products of one-dimensional examples whose analytic solutions are known.

B. PATHINT applied to SMNI

The accuracy of this histogram path-integral algorithm relies heavily on Eq. (19), i.e., being able to use the diagonal diffusions to select a proper rectangular mesh. However, for the SMNI problem, both g^{EE} and g^{II} are highly nonlinear in both M^E and M^I, preventing a rectangular mesh from being developed in M^G space.

To confront this problem, use is made of the previous observations [4,6], that the most likely states of the "centered" systems lie along diagonals in M^G space, a line determined by the numerator of the threshold factor, essentially

$$A^E_E M^E - A^E_I M^I \approx 0 \ , \tag{20}$$

where for neocortex A^E_E is on the order of A^E_I. Along this line, for a "centered" system, the threshold factor $F^E \approx 0$, and $\underline{L}^E$ is a minimum. However, looking at $\underline{L}^I$, in F^I the numerator $(A^I_E M^E - A^I_I M^I)$ is typically small only for small M^E, since for neocortex $A^I_I \ll A^I_E$.

General transformations of variables must be treated with some care [22,42]. In the Stratonovich midpoint representation, the invariance of multiplicative-noise Gaussian Markovian systems is made explicit, inducing a Riemannian geometry with the metric being the inverse diffusion matrix. For example, in the path integral representation, the conditional probability density is written as

$$P = \int \cdots \int \underline{D}M \exp\left(-\sum_{s=0}^{u} \Delta t \underline{L}_{Fs} \right) ,$$

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

$$\int dM^G_s \to \sum_{\iota=1}^{N^G} \Delta M^G_{\iota s} \ , \ M^G_0 = M^G_{t_0} \ , \ M^G_{u+1} = M^G_t \ ,$$

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

$$(\cdots)_{,G} = \frac{\partial(\cdots)}{\partial M^G} \ ,$$

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

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

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

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

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

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

$$R_{FJKL} = \frac{1}{2} (g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN} (\Gamma^M_{FK} \Gamma^N_{JL} - \Gamma^M_{FL} \Gamma^N_{JK}) \ , \tag{21}$$

where R is the Riemannian curvature, and we also have explicitly noted the discretization in the mesh of $M^G_{\iota s}$ by ι. If M is a field, e.g., also dependent on a spatial variable x discretized by ν, then the variables M^G_s is increased to $M^{G\nu}_s$, e.g., as we have prescribed for the macroscopic neocortex. The term $R/6$ in $\underline{L}_F$ includes a contribution of $R/12$ from the WKB approximation to the same order of $(\Delta t)^{3/2}$ [22].

A prepoint discretization for the same probability distribution P gives a much simpler algebraic form, e.g., as given in Eq. (6) for SMNI. Under a general transformation $M' = M'(M)$, the Fokker-Planck equation,

$$\frac{\partial P}{\partial t} = \frac{1}{2} (g^{GH} P)_{,GH} - (g^G P)_{,G} + VP \ ,$$

$$(\cdots)_{,G} \equiv \partial(\cdots)/\partial M^G \ , \tag{22}$$

becomes

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

$$(\cdots)_{,G} \equiv \partial(\cdots)/\partial M'^G \ ,$$

$$g'^G = \frac{\partial M'^G}{\partial M^H} g^H + \frac{1}{2} g^{HJ} \frac{\partial^2 M'^G}{\partial M^H \partial M^J} \ ,$$

$$g'^{GH} = \frac{\partial M'^G}{\partial M^J} \frac{\partial M'^H}{\partial M^K} g^{JK} \ ,$$

$$dM' = \left|\left|\frac{\partial M'}{\partial M}\right|\right| dM \ ,$$

$$PdM = P'dM' \ . \tag{23}$$

Now, in the context of this paper, to prepare a mesh determined primarily on values of the diffusions along the diagonal slope a,

$$a = \frac{A_I^E}{A_E^E} \ , \tag{24}$$

a transformation is used from $M^G = (M^E, M^I)$ to $M'^H = (M'^X, M'^Y)$,

$$M'^X = \frac{1}{(2a)^{1/2}}(M^E - aM^I) \ ,$$

$$M'^Y = \frac{1}{(2a)^{1/2}}(M^E + aM^I) \ ,$$

$$M^E = \frac{a}{(2a)^{1/2}}(M'^Y + M'^X) \ ,$$

$$M^I = \frac{1}{(2a)^{1/2}}(M'^Y - M'^X) \ , \tag{25}$$

preserving

$$dM' = \left|\left|\frac{\partial M'}{\partial M}\right|\right| dM = dM \ . \tag{26}$$

This defines the drifts,

$$g'^X = \frac{1}{(2a)^{1/2}}(g^E - ag^I) \ ,$$

$$g'^Y = \frac{1}{(2a)^{1/2}}(g^E + ag^I) \ , \tag{27}$$

and the diffusions,

$$g'^{XX} = g'^{YY} = \frac{1}{2a}(g^{EE} + a^2 g^{II}) \ ,$$

$$g'^{XY} = g'^{YX} = \frac{1}{2a}(g^{EE} - a^2 g^{II}) . \tag{28}$$

Creating this rectangular grid in M' space, encompassing the boundaries of M space, also enlarges the range, e.g., from 161×61 to 155.56×155.56 for $a = 1$. Reflecting boundary conditions were imposed along the "diagonal" four walls of M space in M' space, and the values of the contributions to the Green's function from points outside these walls were set to a small number 10^{-20}.

After the numerical calculations in M' space, PATHINT prints out values in the original physical M space. A bandwidth of 5 ΔM_j's was used, as this gave essentially the same results as three and four such units on each side of the diagonal terms for both M'^X and M'^Y. (The smaller bands gave runs much faster using much smaller matrices, but the larger band was used in the interest of erring on the side of caution.) The mesh for each M' variable was determined by evaluating its respective diagonal diffusion along its axis. Experience with the BC′ and EC′ models showed that coarser calculations with $\Delta t = \tau$ gave sufficiently different answers than with $\Delta t = 0.5\tau$, so that only these latter models are reported here.

C. PATHINT results From SMNI STM calculation

The results of these explicit calculations corroborate the relatively intuitively derived results obtained previously [4,6]. An initial δ-function disturbance at $M^G = 0$ excites all modes in M^G space. The normalization of the discrete probability density is kept fixed. If a distribution of firings were to be presented to the system, this initial state would be filtered and processed according to the set of attractors defined by the Lagrangian.

Figures 1(a) and 1(b) show the evolution of the BC′ model at 0.05 and 0.5 sec, at 5τ and 50τ. The slope $a = 1.0$. The results are at least consistent with that intuited by the earlier SMNI studies [4,6], in that there are close to ten peaks that can be considered candidates to store STM. Plots were prepared using GNUPLOT [55]. As discussed above, the large peaks in the distributions in the corners are not candidates for STM, as memories trapped in theses states are not accessible from other local peaks with time scales of 1/10 sec. Furthermore, it can be assumed that these models have not taken into account other biological mechanisms in normal neocortex that tend to keep firings from extremes of all firing and all not firing.

To represent a macrocolumnar-averaged minicolumn within a time of epoch τ, i.e, a "mesocolumn," with 80 excitatory neurons and 30 inhibitory neurons, a transition matrix would contain $80 \times 30 \times 80 \times 30 \approx 6 \times 10^6$ elements. Using the above algorithm, for the balanced case, 403,929 elements in the larger M' space were needed.

One hundred foldings were considered sufficient to observe systematics of response to a delta-function disturbance at $M^G = 0$. On the author's Sun SPARCstation-2, the calculation of a transition matrix took about 3 CPU min, and about 6 sec for each subsequent time-folding.

Figure 1

Figures 2(a) and 2(b) show the evolution of the EC′ model at 0.05 and 0.5 sec. The slope $a = 0.5$. The number of peaks that can be considered candidates for STM storage are much less than for the BC′ model. A matrix of 505,800 elements was required for this calculation.

Figure 2

Figures 3(a) 3(b) show the evolution of the IC′ model at 0.05 and 0.5 sec. The slope $a = 2.0$. Similar to the EC′ model, the number of peaks that can be considered candidates for STM storage are less than for the BC′ model, although more than for the EC′ model. A small diffusion in the corners required smaller meshes there, resulting in larger matrix of 1,850,330 elements. This run was performed on a SPARCstation-10MP.

Figure 3

Figures 4(a) and 4(b) show the evolution of the BC′_VIS model, the BC′ model for a visual cortex, at 0.05 and 0.5 sec. The slope $a = 0.5$ as in the BC′ model, but the number of neurons/minicolumn is increased from $(N^E, N^I) = (80, 30)$ to $(N^E, N^I) = (160, 60)$. The larger number of neuronal states resulted in a matrix of 1,479,993 elements required for this calculation. This run was performed on a SPARCstation-10MP.

Figure 4

It is clear that models BC′ and BC′_VIS support multiple stable states in the interior physical firing M^G space for time scales of a few tenths of a second. Models EC′ and IC′ do not possess these attributes. This is in accord with earlier intuitions obtained by examining stability of the static Lagrangian [4,6].

Figures 5(a) and 5(b) examine the interior of M^G space a little closer by applying a cut-off to all points of the probability density higher than 0.0001, for models BC′ and BC′_VIS. Model BC′_VIS demonstrates the clustering effect noted earlier [4,6], which is easiest seen in the contour maps at the base of Fig. 5(b). I.e., all likely states cannot be as readily accessed in model BC′_VIS as in model BC′ within time scales of 1/10 sec.

These calculations must be considered as a starting point for future investigations. For example, before these can be scrutinized to calculate accurate times of first passage, etc., further development of PATHINT should be explored to handle the complications of multivariate nonlinear diffusions. However, these calculations do explicitly demonstrate the stability of multiple memory states within experimentally observed time epochs. There is no suggestion at all from these calculations that there is any marked sensitivity to initial conditions of neuronal firing states, but there is a marked sensitivity to the synaptic parameters defining the three classes of models presented here.

V. CONCLUSION

When approaching a system at a given scale, science typically requires that there is at least some understanding of this scale from some underlying finer-grained scales. The natural scale underlying macroscopic neocortical phenomena at the level of EEG is at the columnar scales. Then, the SMNI theory must be tested here as well, and STM is a natural phenomena to use for such tests. SMNI cannot do any justice at all to smaller scales, but it seems that artificial neural networks can barely approach the columnar scales as well. For example, just to keep track of the states of a minicolumn of only 100 grossly simplified neurons would require an accounting of $2^{110} \approx 10^{33}$ states; an accounting of the transition matrix of these states would require the square of this astronomical number. Even an SMNI statistical aggregation, keeping vital nonlinear dynamics derived at the finer neuronal level, into combinations of 80 excitatory and 30 inhibitory states would require a transition matrix of $\approx 6 \times 10^6$. All these estimates increase by a factor of 16 for visual cortex with 220 neurons per minicolumn. The use of PATHINT greatly reduces the number of elements required for an accurate calculation of the evolution of the SMNI mesoscopic system, and such calculations are given here.

There are several factors in the SMNI development that support optimism for extracting more signal from noise in EEG data than is currently possible. While SMNI is most logically tested using data collected from brain surface recordings, the necessity and utility of performing noninvasive EEG scalp recordings argues strongly for further developing SMNI to extract better signal out of noise from scalp recordings.

(a) In the course of a logical, nonlinear, stochastic development of aggregating neuronal and synaptic interactions to larger and larger scales, opportunities are taken to use techniques of mathematical physics to overcome several technical hurdles. Paradigms and metaphors from other disciplines do not substitute for logical SMNI development.

(b) The SMNI theoretical model has independent validity in describing EEG dispersion relations, systematics of short-term memory, velocities of propagation of information across neocortical fields, recency versus primacy effects, etc. Fits of such models to data should do better in extracting signal from noise than *ad hoc* phenomenological models.

(c) ASA enables the fitting of quite arbitrary nonlinear stochastic models to such data as presented by EEG systems. This means that functional dependences in the noise itself (the diffusion matrix) as well as the functional dependences in the driving terms (the drift vector) can be fit directly. Once fitted, PATHINT can evolve the system, testing long-time correlations between the model(s) and the data, as well as serving to predict events.

(d) 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.

(e) Yet to explore are the ramifications of using the derived (not hypothesized) Riemannian metric induced by multivariate Fokker-Plank-type systems. This seems to form a natural invariant measure of information, that could or should be used to explore flows of information between neocortical regions.

(f) The SMNI approach shows how to "renormalize" the spatial activity to get a model that more closely matches the experimental situation of scalp measurement, wherein there is attenuation of ranges of wave numbers [45].

(g) The MNN parallel algorithm may offer real-time processing of nonlinear modeling and fitting of EEG data for clinical use. Regional EEG data can be interpreted as mechanisms occurring at the minicolumnar scales, scales which overlap with other work being performed by ANN.

(h) This PATHINT code is an important partner to the ASA code. ASA has made it possible to perform fits of complex nonlinear SMNI distributions to EEG data [13]. Now, using ASA, the parameters of the fitted SMNI distribution can be used to determine a distribution of firings in a short initial time epoch of EEG. Then, PATHINT can be used to predict the evolution of the system, possibly to predict oncoming states, e.g., epileptic seizures of patients baselined to an SMNI fitted distribution.

FIGURE CAPTIONS

FIG. 1. Model BC′. (a) is the evolution at 5τ. (b) is the evolution at 50τ.

FIG. 2. Model EC′. (a) is the evolution at 5τ. (b) is the evolution at 50τ.

FIG. 3. Model IC′. (a) is the evolution at 5τ. (b) is the evolution at 50τ.

FIG. 4. Model BC′_VIS. (a) is the evolution at 5τ. (b) is the evolution at 50τ.

FIG. 5. Examination of interiors of models BC′ and BC′_VIS. Probability densities are cut off for values higher than 0.0001 at time 50τ. (a) Model BC′. (b) Model BC′_VIS.

REFERENCES

[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] L. Ingber, "Generic mesoscopic neural networks based on statistical mechanics of neocortical interactions," *Phys. Rev. A* **45**, R2183-R2186 (1992).

[15] L. Ingber, "Statistical mechanics of neocortical interactions: Multiple scales of EEG," *Electroencephal. clin. Neurophysiol.* (to be published) (1996).

[16] L. Ingber, "Statistical mechanics of multiple scales of neocortical interactions," in *Neocortical Dynamics and Human EEG Rhythms*, ed. by P.L. Nunez (Oxford University Press, New York, NY, 1995), p. 628-681.

[17] B. Jagadeesh, H.S. Wheat, and D. Ferster, "Linearity of summation of synaptic potentials underlying direction selectivity in simple cells of the cat visual cortex," *Science* **262**, 1901-1904 (1993).

[18] J. Hertz, A. Krogh, and R.G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA, 1991).

[19] L. Ingber, "Adaptive Simulated Annealing (ASA)," [ftp.alumni.caltech.edu: /pub/ingber/ASA-shar, ASA-shar.Z, ASA.tar.Z, ASA.tar.gz, ASA.zip], Lester Ingber Research, McLean, VA, 1993.

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

[21] P.L. Nunez, "Localization of brain activity with Electroencephalography," in *Advances in Neurology, Vol. 54: Magnetoencephalography*, ed. by S. Sato (Raven Press, New York, NY, 1990), p. 39-65.

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

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

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

[25] L. Ingber, "Editorial: Learning to learn," *Explore* **7**, 5-8 (1972).

[26] L. Ingber, *Karate: Kinematics and Dynamics* (Unique, Hollywood, CA, 1981).

[27] L. Ingber, *Elements of Advanced Karate* (Ohara, Burbank, CA, 1985).

[28] G.A. Miller, "The magical number seven, plus or minus two," *Psychol. Rev.* **63**, 81-97 (1956).

[29] K.A. Ericsson and W.G. Chase, "Exceptional memory," *Am. Sci.* **70**, 607-615 (1982).

[30] H. Eichenbaum, "Thinking about brain cell assemblies," *Science* **261**, 993-994 (1993).

[31] G. Zhang and H.A. Simon, "STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses," *Memory & Cognition* **13**, 193-201 (1985).

[32] B.B. Murdock, Jr., "A distributed memory model for serial-order information," *Psychol. Rev.* **90**, 316-338 (1983).

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

[34] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).

[35] C.D. Gilbert and T.N. Wiesel, "Functional organization of the visual cortex," *Prog. Brain Res.* **58**, 209-218 (1983).

[36] 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).

[37] 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).

[38] T.J. Imig and R.A. Reale, "Patterns of cortico-cortical connections related to tonotopic maps in cat auditory cortex," *J. Comp. Neurol.* **192**, 293-332 (1980).

[39] E.G. Jones, J.D. Coulter, and S.H.C. Hendry, "Intracortical connectivity of architectonic fields in the somatic sensory, motor and parietal cortex of monkeys," *J. Comp. Neurol.* **181**, 291-348 (1978).

[40] 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.

[41] T.H. Bullock, "Reassessment of neural connectivity and its specification," in *Information Processing in the Nervous System*, ed. by H.M. Pinsker and W.D. Willis, Jr. (Raven Press, New York, NY, 1980).

[42] R. Graham, "Covariant formulation of non-equilibrium statistical thermodynamics," *Z. Physik* **B26**, 397-405 (1977).

[43] R.W. Dykes, "Parallel processing of somatosensory information: A theory," *Brain Res. Rev.* **6**, 47-115 (1983).

[44] R.P Erickson, "The across-fiber pattern theory: An organizing principle for molar neural function," *Sensory Physiol.* **6**, 79-110 (1982).

[45] P.L. Nunez, *Electric Fields of the Brain: The Neurophysics of EEG* (Oxford University Press, London, 1981).

[46] 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).

[47] G.S. Agarwal and S.R. Shenoy, "Observability of hysteresis in first-order equilibrium and nonequilibrium phase transitions," *Phys. Rev. A* **23**, 2719-2723 (1981).

[48] 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).

[49] 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).

[50] 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).

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

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

[53] 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).

[54] K. Binder and D. Stauffer, "A simple introduction to Monte Carlo simulations and some specialized topics," in *Applications of the Monte Carlo Method in Statistical Physics*, ed. by K. Binder (Springer-Verlag, Berlin, 1985), p. 1-36.

[55]

Figure 1a

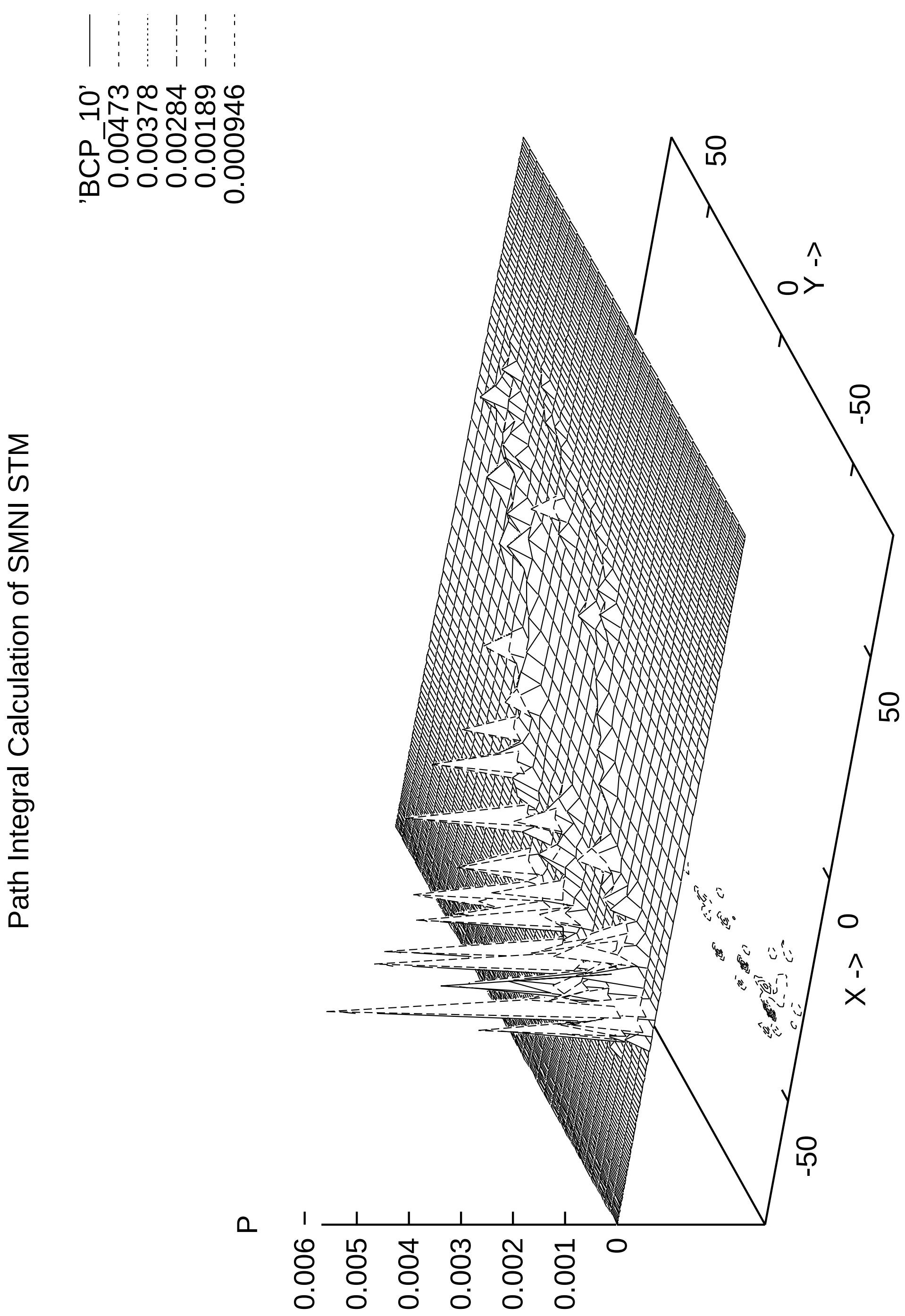

Path Integral Calculation of SMNI STM
'BCP_10'
0.00473
0.00378
0.00284
0.00189
0.000946
P
0.006
0.005
0.004
0.003
0.002
0.001
0
X ->
-50
0
50
Y ->
-50
0
50

Figure 1b

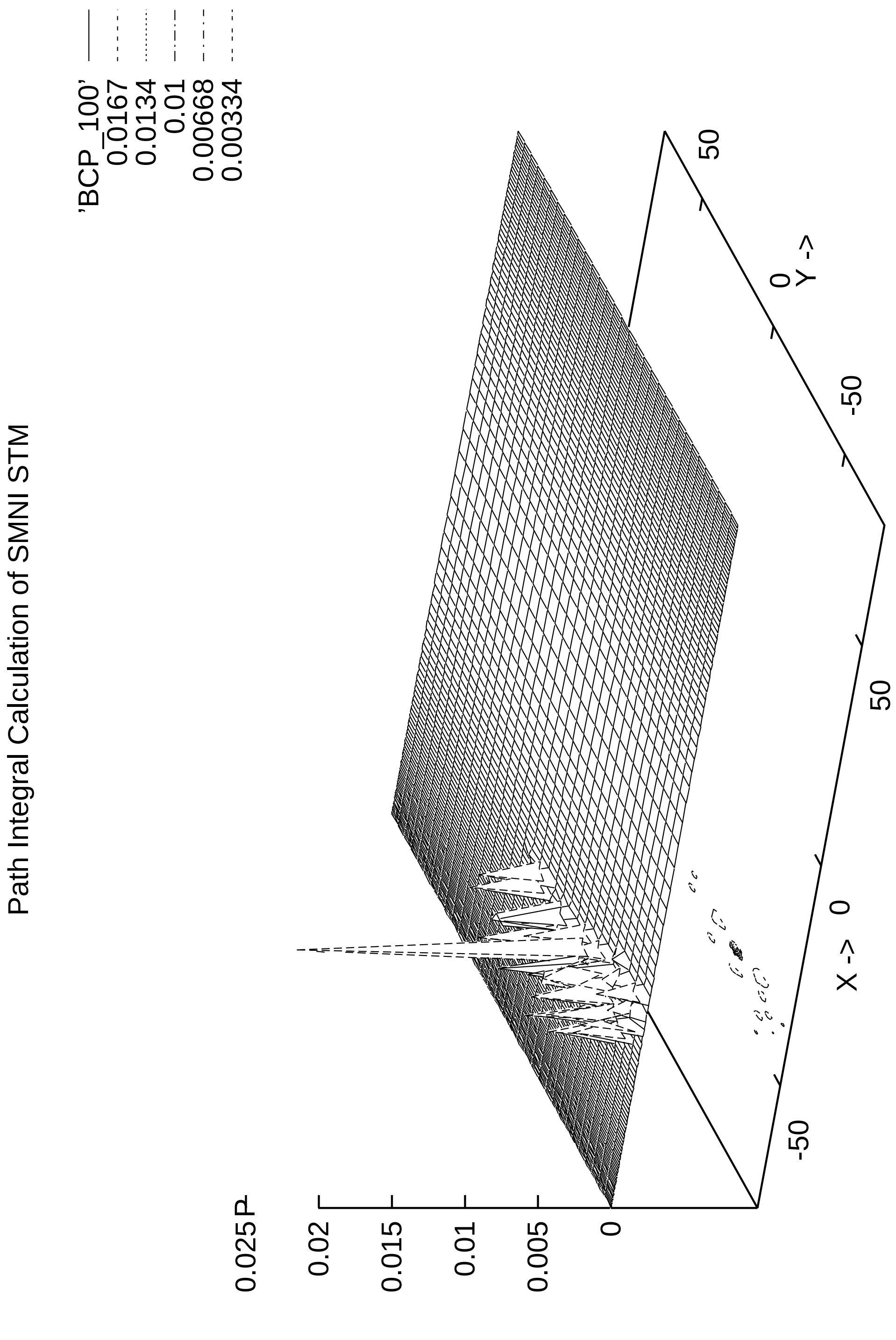

Path Integral Calculation of SMNI STM
'BCP_100'
0.0167
0.0134
0.01
0.00668
0.00334
P
0.025
0.02
0.015
0.01
0.005
0
X ->
-50
0
50
Y ->
-50
0
50

Figure 2a

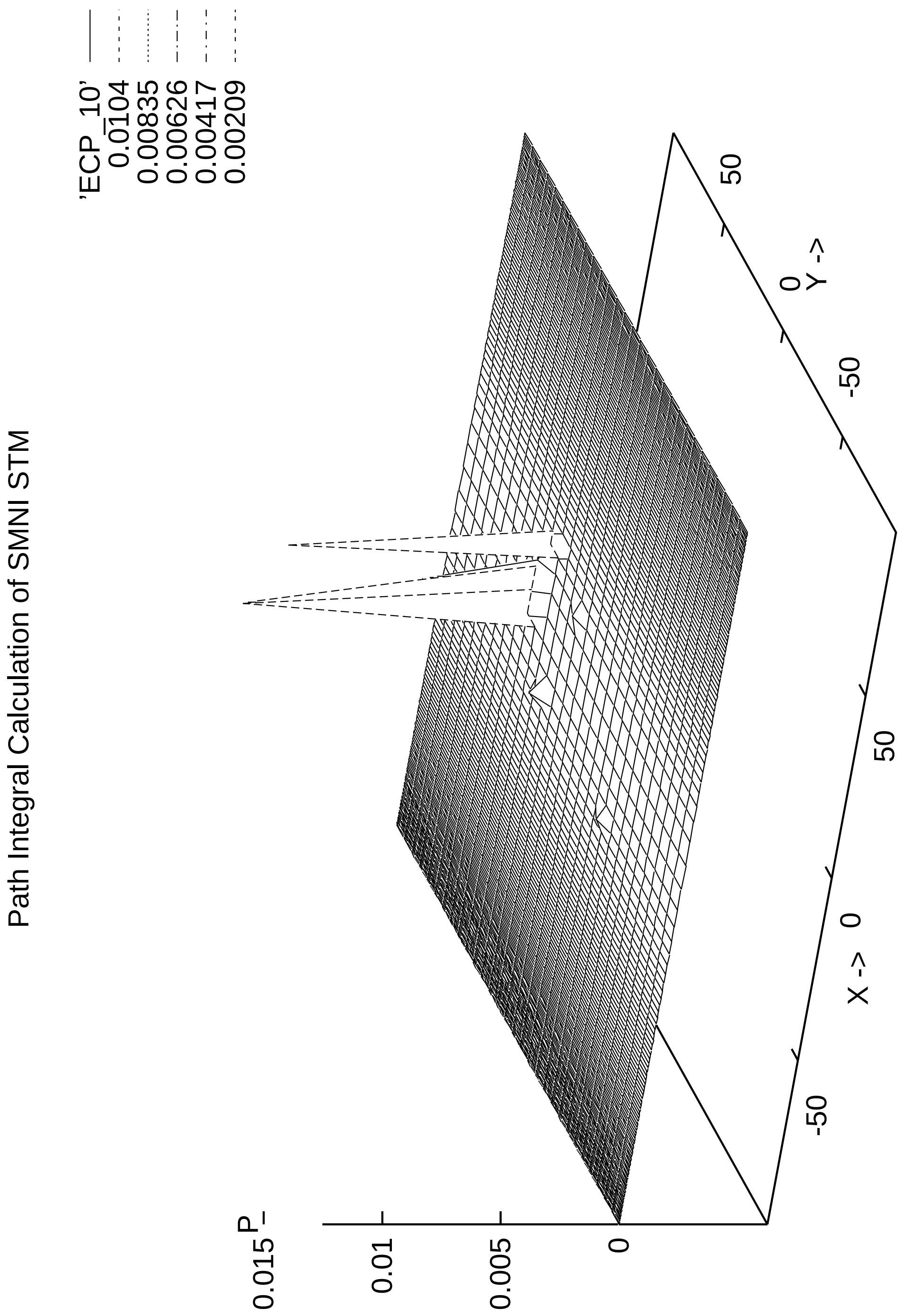

Path Integral Calculation of SMNI STM
'ECP_10'
0.0104
0.00835
0.00626
0.00417
0.00209
P
0.015
0.01
0.005
0
X ->
-50
0
50
Y ->
-50
0
50

Figure 2b

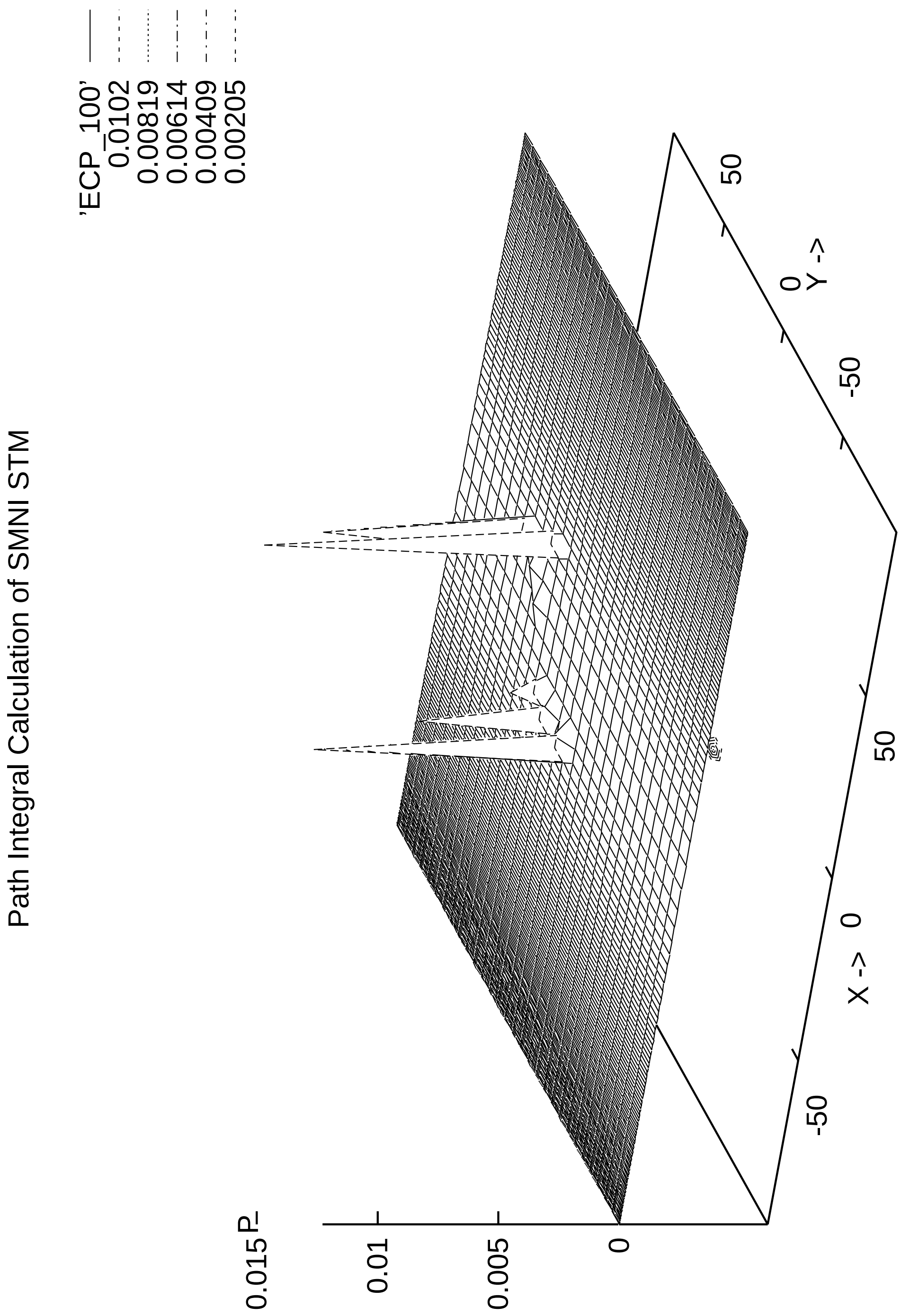

Path Integral Calculation of SMNI STM
'ECP_100'
0.0102
0.00819
0.00614
0.00409
0.00205
P
0.015
0.01
0.005
0
-50
0
50
X ->
-50
0
50
Y ->

Figure 3a

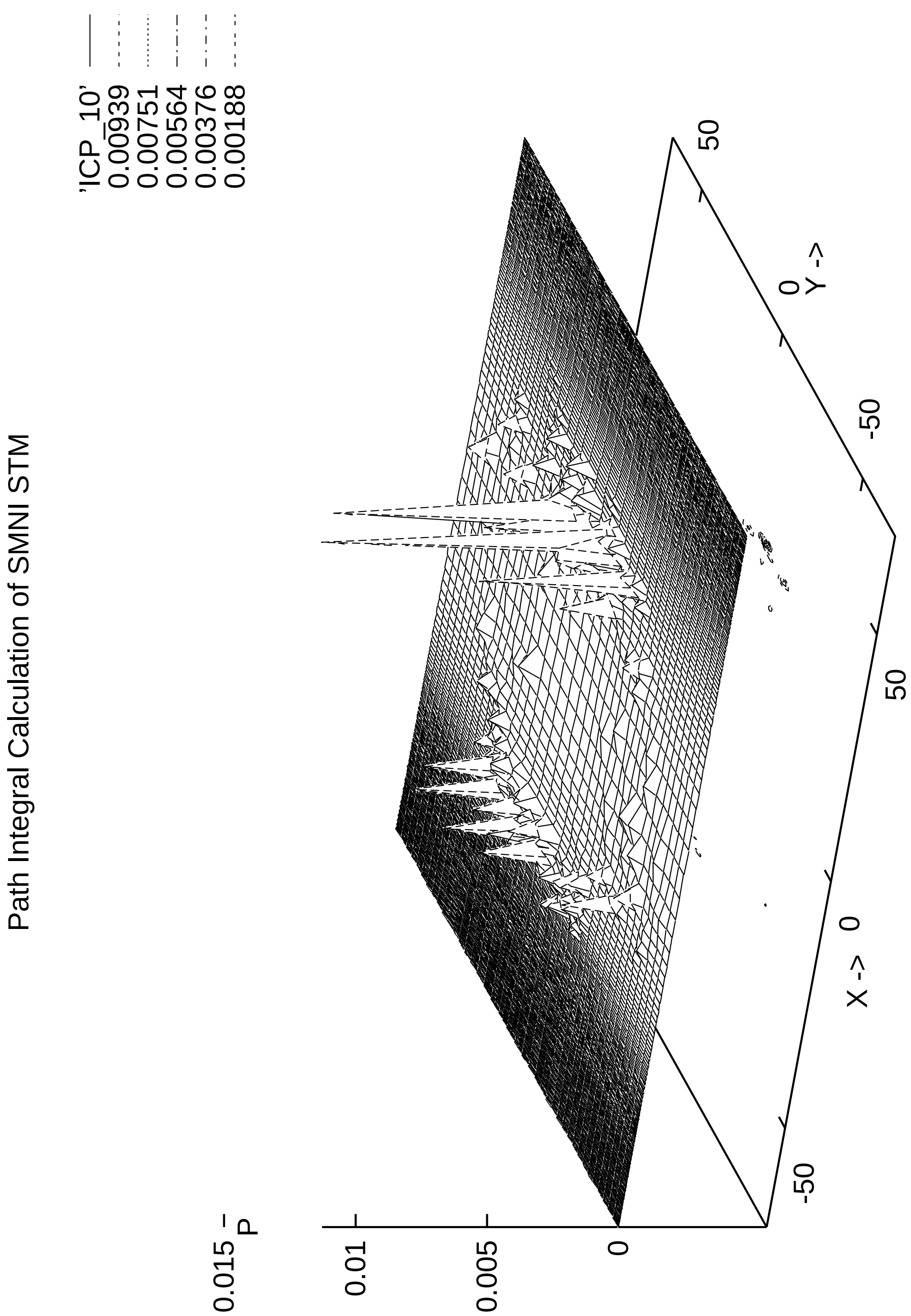

Path Integral Calculation of SMNI STM
'ICP_10'
0.00939
0.00751
0.00564
0.00376
0.00188
P
0.015
0.01
0.005
0
X ->
-50
0
50
Y ->
-50
0
50

Figure 3b

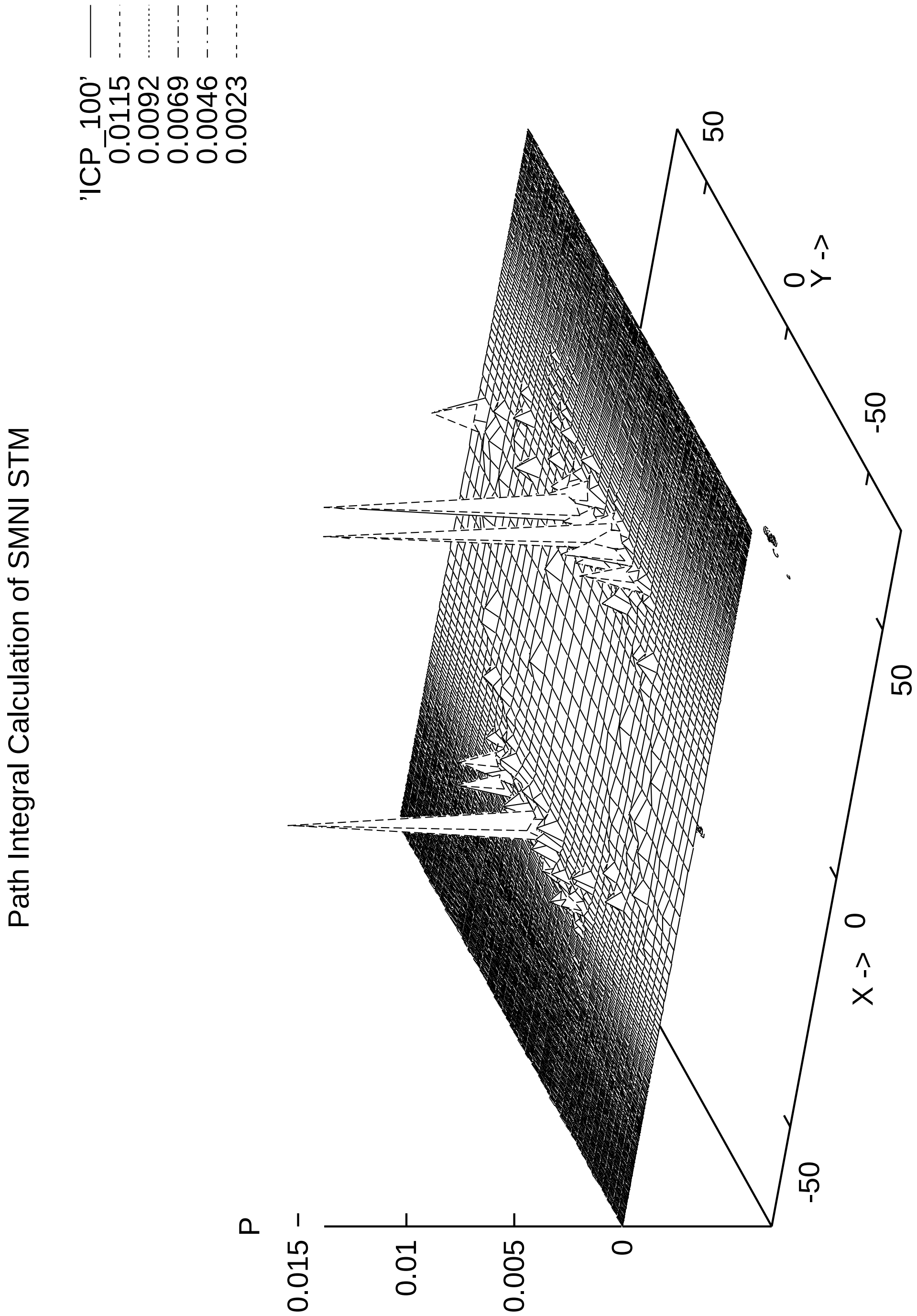

Path Integral Calculation of SMNI STM
'ICP_100'
0.0115
0.0092
0.0069
0.0046
0.0023
P
0.015
0.01
0.005
0
-50
0
50
X ->
-50
0
50
Y ->

Figure 4a

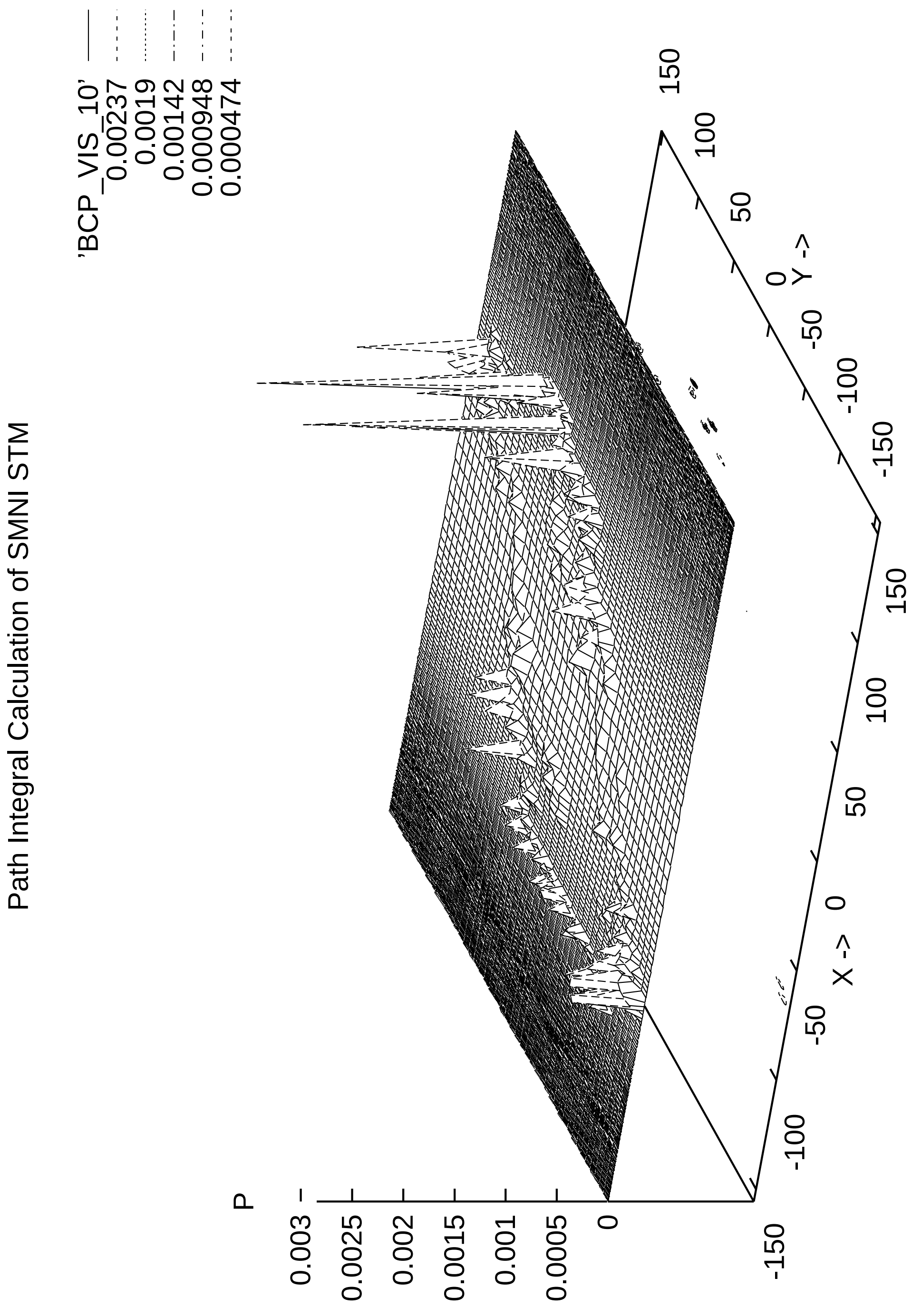

Path Integral Calculation of SMNI STM
'BCP_VIS_10'
0.00237
0.0019
0.00142
0.000948
0.000474
P
0.003
0.0025
0.002
0.0015
0.001
0.0005
0
X ->
-150
-100
-50
0
50
100
150
Y ->
-150
-100
-50
0
50
100
150

Figure 4b

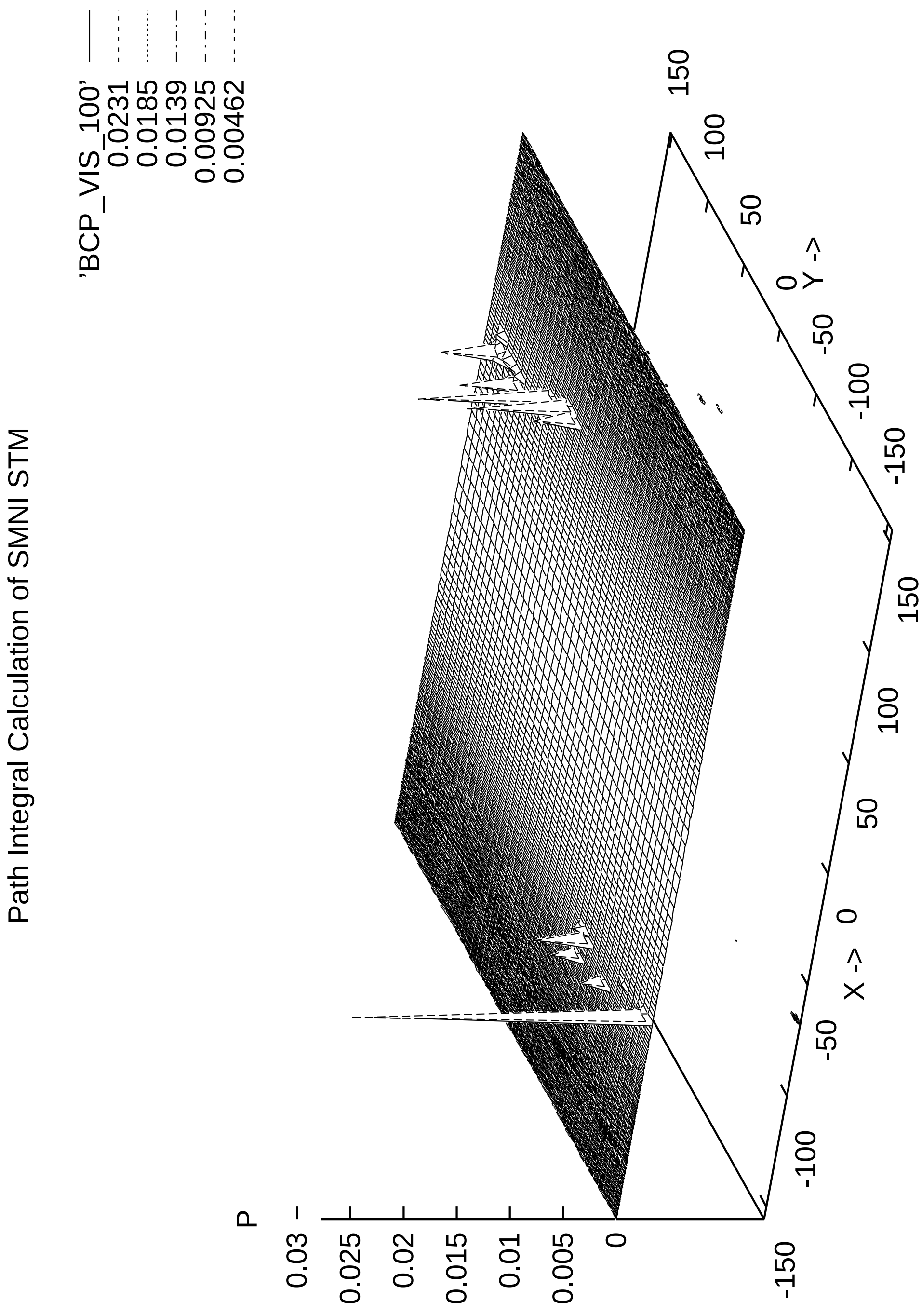

Path Integral Calculation of SMNI STM
'BCP_VIS_100'
0.0231
0.0185
0.0139
0.00925
0.00462
P
0.03
0.025
0.02
0.015
0.01
0.005
0
X ->
-150
-100
-50
0
50
100
150
Y ->
-150
-100
-50
0
50
100
150

Figure 5a

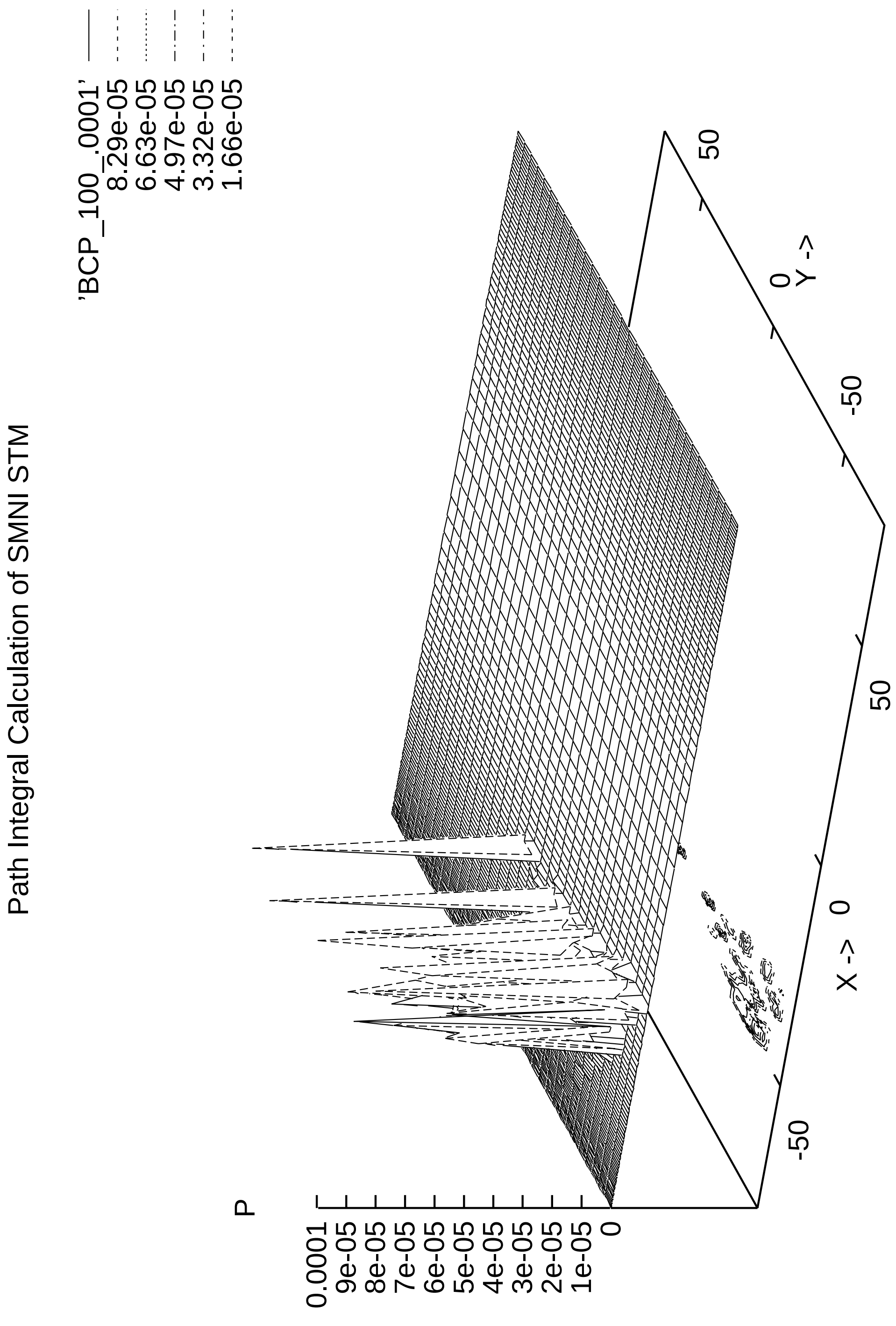
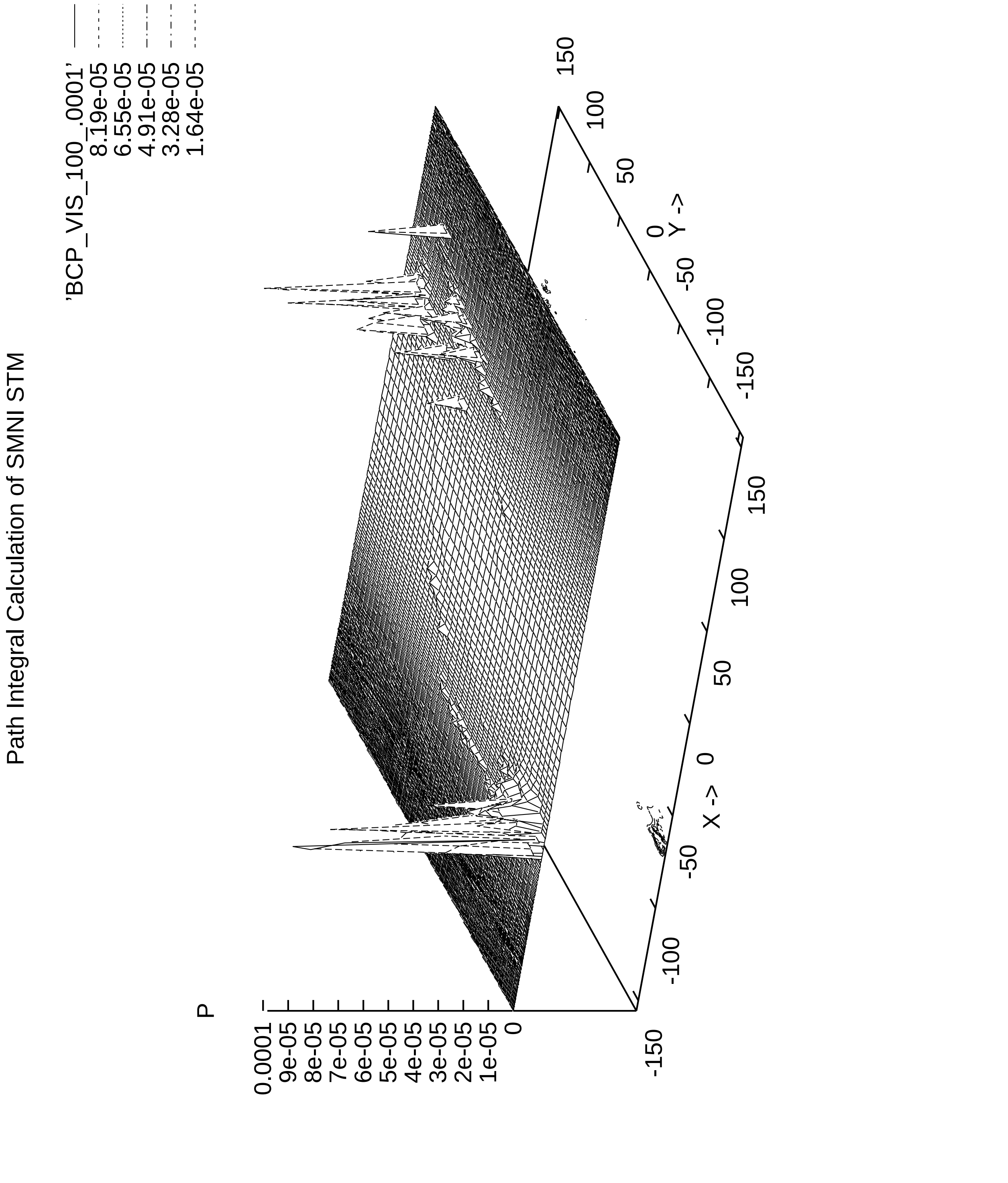

Path Integral Calculation of SMNI STM
'BCP_100_.0001'
8.29e-05
6.63e-05
4.97e-05
3.32e-05
1.66e-05
P
0.0001
9e-05
8e-05
7e-05
6e-05
5e-05
4e-05
3e-05
2e-05
1e-05
0
X ->
-50
0
50
Y ->
-50
0
50

Figure 5b


Path Integral Calculation of SMNI STM
'BCP_VIS_100_.0001'
8.19e-05
6.55e-05
4.91e-05
3.28e-05
1.64e-05
P
0.0001
9e-05
8e-05
7e-05
6e-05
5e-05
4e-05
3e-05
2e-05
1e-05
0
-150
-100
-50
0
50
100
150
X ->
-150
-100
-50
0
50
100
150
Y ->