9th MIT.ONR Workshop on C^3 Systems Naval Postgraduate School, 2-5 June 1986

NONLINEAR NONEQUILIBRIUM STATISTICAL MECHANICS APPROACH TO C^3I SYSTEMS

Lester Ingber

National Research Council-NPS Senior Research Associate
and Physics Department and C^3 Group — Code 61IL
Naval Postgraduate School, Monterey, CA 93943

It is proposed to incorporate ''intuition'' into large complex multivariate nonlinear C^3I systems requiring stochastic or probabilistic treatment, i.e., to seek regions of variable-space where more local analytic resources can be optimally allocated. These mathematical techniques have been utilized for a variety of other systems, ranging from neuroscience, to nuclear physics, to financial markets. The experiences gained by detailing each of these systems offers specific insights by which to approach C^3I systems.

I. INTRODUCTION

Even without having agreement on just what is C^3I, there is widespread criticism that we do not spend enough on C^3I relative to what we spend on specific weapons systems [1]. The Eastport Study Group [2] has made this issue its primary concern with regard to the SDI program. There is also an ever-present problem of weighing the political and military aspects between hierarchical and distributed design of C^3I, the former being politically desirable and appropriate for deterministic or modestly stochastic operations, and the latter being more appropriate for severely stochastic systems [3]. Future battle management, e.g., as being investigated by the SDI program, certainly must consider distributive adaptive C^3I for severely stochastic systems [2].

In this paper I will outline an interdisciplinary approach that is attempting to piece together a specific coherent C^3I model that may yield insights into C^3I systems most appropriate for severely stochastic combat operations.

Section II motivates the necessity of formulating ''order parameters'' relevant to specific situations, which views the physiology (function) of C^3I systems as complementary to their anatomy (structure), by outlining a theory of personal combat [4-7]. There can be no pretense that personal combat is equivalent to international combat, but there are some similarities that deserve mention.

Section III outlines a theory I have formulated of mesoscopic and macroscopic brain function, derived from microscopic synaptic chemical-electrical interactions [8-15]. Since many investigators now find it useful to use brain function as a metaphor for other processes they perceive to be present in their own disciplines, it is relevant to discuss the actual brain and the processes by which it performs ''Biological Intelligence'' (BI). The mathematical formalism used turns out to describe a parallel processing of mesoscopic information in a distributed adaptive system that we know exists, and that we know to be robust under many changes in its internal and external environments. Indeed, in many circumstances, especially those requiring pattern recognition under uncertainty [16-18], BI is still superior to AI which typically requires a deterministic and hierarchical spine on which to grow tree- and loop-like structures.

These technical methods are quite general, and I also have applied them to nuclear physics — detailing Riemannian contributions to the binding energy of nucleons interacting via exchanges of mesons [19-22], and to financial markets — defining an approach to explain various phenomena such as leptokurtosis, the biasing of price data [23]. These systems are all quite different in their natures, but they do share a common approach by these methods of nonlinear nonequilibrium statistical mechanics. The nuclear physics system illustrates how patterns of information can be represented by eigenfunctions of the probability distribution. The markets system illustrates how the mesoscopic scale can be formulated phenomenologically, without the luxury of deriving it from a microscopic system as was done for the

neocortical system.

Section IV outlines how the mathematics used for BI can be used to develop a distributive adaptive system capable of processing more general types of information relevant to C^3I.

Section V discusses work in progress in which we are attempting to use BI to fit data from combat training simulations. We are using these simulations because data is available to fit our theory, and because we can then test our theory by seeing if the dynamic probability distribution we develop can be used successfully in future simulations to enhance the chances of victory. Perhaps these methods will be useful for SDI BM/C^3I systems as well.

Section VI discusses how these tools might be used as decision-making aids in a C^3I system in real time combat situations.

I emphasize that these methods have not been previously applied to C^3 systems, although they have been tested in other systems. Very approximately, this approach can be considered a nonlinear stochastic generalization of Lanchester theory [24].

II. ANATOMY VS. PHYSIOLOGY OF C^3I

A typical C^3I organization-chart—e.g., Sense, Process, Decide, Act, Analysis, Environment, etc.—might well be useful for allocating resources to build a system for combat, or even be useful for developing training methods to keep each component fit and ready for battle. However, especially in this simple example, it is easy to intuit that this outline is not directly useable in actual combat, since in a real-time situation, only a small subset of these parameters, possibly even an entirely new subset aggregated from those given, is of immediate concern to a commander.

Establishing the function (physiology) of C^3I systems seems today to be at least as much an art as a science. However, this function is extremely important, as it defines the actual variables, or order parameters, that a commander requires in combat operations.

For example, in the context that there is much to learn for C^3I systems from the function of the human brain, there is a counterpart to the three levels of processing in processes of attention. I refer to three levels of attention required in personal combat between highly skilled opponents. As a first approximation, time resources are roughly equivalent to distance between opponents.

At a ‘‘far’’ distance, i.e., beyond the distance at which either side can touch the other within a single movement, there is so much uncertainty as to future possibilities, that the only realistic techniques called upon are strategic feints and ‘‘themes’’ of sparring, often categorized by five elements (earth, air, fire, water, void), to cause and break rhythms in the opponent [7]. This is akin to a gross macroscopic perception of the engagement.

At a ‘‘medium’’ distance, i.e., just within the distance at which either side can strike each other with a single movement, skills required are strategic and tactical feint-defense-attack combinations composed of arhythmic spurts of several techniques, somewhat similar to the ‘‘middle’’ game of chess, with the dimension of time thrown in [4-7]. This is akin to a mesoscopic perception of the engagement, wherein the order parameters are the individual combinatoric phrases rather than their individual techniques.

At a ‘‘close’’ distance, i.e., within the distance at which either side can reach or lunge with elbows and knees, one must function within critical reaction times of very few tenths of a second. At this microscopic perception it is more sheer power and chance than strategy or even tactics, that determines the outcome, as only simple repeated firings of techniques are realistic.

Some other interesting analogies between C^3I systems and personal combat can be drawn. In order to be effective within tenths of a seconds against strong opponents, one must train to have distributed control at many stages of the C^3I-karate organization. Visual and auditory senses must be trained to receive information in parallel with somatic senses actively seeking information. Imaginary scenarios and forecasts must be made in parallel with decisions being made in real time. The trained body must coordinate itself to perform techniques, just using quite general constraints imposed by these decisions: there are many techniques that might accomplish similar goals, but the choice of technique does not seem to made by one central command center.

Perhaps the most important analogy to stress emphasize, and that I will also stress in my outline of my work in neocortex, is that the concept of ''scaling'' should be applied to C^3I systems to determine the relevant order parameters describing levels of distributed command and control.

III. BIOLOGICAL INTELLIGENCE (BI)

1. Introduction—Theory vs. Model

BI demonstrates the physiology of neocortex. Proper treatment of nonlinearities demonstrates how multiple hypotheses are generated and processed by STM. Similarly, it should be expected that useful decision aids to commanders will require robust C^3I nonlinear models of previous combat operations.

Modern technology has made it possible to detail actual properties of many physical and biological nonlinear nonequilibrium systems, i.e., in contrast to performing otherwise important investigations of (quasi-)linearized approximate models. Typically, the price paid for this detail is that a set of complementary approaches, sometimes mutually exclusive, must be used for particular aspects [25]. C^3I and neocortex present similar challenges.

A series of publications has detailed a statistical mechanics approach to macroscopic regions of neocortex, derived from statistical aggregates of microscopic neurons, i.e., a statistical mechanics of neocortical interactions (SMNI) [8-14]. As found necessary for other nonlinear nonequilibrium systems, a mesoscopic scale is sought to develop a Gaussian-Markovian statistics for further macroscopic development [26, 27]. This mesoscopic scale is found in the observed physiology of columnar interactions. Long-term-memory (LTM) properties and the duration and capacity of short-term-memory (STM), i.e., the ''7 ± 2 rule,'' have been derived from multiple minima of a nonlinear Lagrangian (time-dependent and space-dependent ''cost function''); the alpha frequency and velocity of propagation of columnar information-processing, consistent with observed movements of attention across the visual field, have been derived in linearized ranges within these minima.

Coarse-graining is an important general method of treating nonlinear nonequilibrium statistical systems, e.g., in order to develop Gaussian-Markovian probability distributions. Also, less resources are required to process the coarser variables, which is efficient if that is all that is required for macroscopic function. The theory capable of treating these systems require mathematical tools only developed in the late 1970's [28-39], including quite general nonlinear nonequilibrium structures into previously linear treatments of Gaussian-Markovian systems [40].

This theory is geared to explain macroscopic neocortical activity, retaining as much correct description of underlying microscopic synaptic activity as can be carried by modern mathematical physics, which turns out to be sufficient for several important circumstances. Only after this process is completed, are approximate numerical and algebraic methods applied to solve the resulting mathematics. It is at *this* stage that modelling is most useful. The 1980's already have demonstrated that many systems require the use of several complementary algebraic and numerical algorithms to detail several scales of interaction [25]. Neocortex is not unique in requiring several approaches, nor is it unique in requiring it own unique algorithms.

For example, without sufficient mathematical or physical justification, many models assume (quasi-)linear deterministic rate equations—analogous to conserved quadratic ''Hamiltonians''—to postulate ''average'' neurons, thereby neglecting statistical and stochastic background interactions, nonlinearities induced by interactions among neurons, and spatial-temporal statistics of large ensembles of these interacting neurons. In fact, these nonlinearities and statistics are essential mechanisms of STM [11, 13], and possibly of alpha rhythm observed in electroencephalographic (EEG) and magnetoencephalographic (MEG) [41] activity [12]. These results are not obtained by ''fitting'' theoretical parameters mocking neuronal mechanisms to empirical data. Rather, these results are obtained by taking reasonable synaptic parameters, developing the statistical mechanics of neocortical interactions, and then discovering that indeed they are consistent with the empirical macroscopic data. Other models which have offered plausible brain mechanisms can be processed by this theory, extending their ranges of validity [8, 9].

2. Description of Theory

Microscopic Neurons

When describing the activity of large ensembles of neocortical neurons, each one typically having many thousands of synaptic interactions it is a reasonable assumption that simple algebraic summation of excitatory (E) depolarizations and inhibitory (I) hyperpolarizations at the base of the inner axonal membrane determine the firing depolarization response of a neuron within its absolute and relative refractory periods [42].

This is straightforwardly mathematically summarized. Within τ_j~5−10 msec, the conditional probability that neuron j fires, given its previous interactions with k neurons, is

$$p_{\sigma_j} \approx \Gamma\,\Psi$$

$$\approx \frac{\exp(-\sigma_j F_j)}{\exp F_j + \exp(-F_j)} ,$$

$$F_j = \frac{V_j - \sum_k a^*_{jk} v_{jk}}{\left[\pi \sum_{k'} a^*_{jk'}({v_{jk'}}^2 + {\phi_{jk'}}^2)\right]^{1/2}} ,$$

$$a_{jk} = \frac{1}{2} A_{jk}(\sigma_k + 1) + B_{jk} . \qquad (1)$$

This is true for Γ Poisson, and for Ψ Poisson or Gaussian. V_j is the axonal depolarization threshold, v_{jk} is the induced synaptic polarization of E or I type at the axon, and ϕ_{jk} is its variance. The efficacy a_{jk}, related to the inverse conductivity across synaptic gaps, is composed of a contribution A_{jk} from the connectivity between neurons which is activated if the impinging k-neuron fires, and a contribution B_{jk} from spontaneous background noise.

Mesoscopic Domains

As is found for most nonequilibrium systems, a mesoscopic scale is required to formulate the statistical mechanics of the microscopic system, from which the macroscopic scale can be developed [26]. 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≈100 neurons (about 200 in visual cortex) comprise modular units vertically oriented relative to the warped and convoluted neocortical surface throughout most, if not all, regions of neocortex [43-47]. Clusters of about 100 neurons have been deduced to be reasonable from other considerations as well [48]. The overwhelming majority of neuronal interactions are short-ranged, diverging out via efferent minicolumnar fibers to within ~1 mm, which is the extent of a ''macrocolumn'' comprising ~10^3 minicolumns of N^*≈10^5 neurons. Macrocolumns also exhibit rather specific information-processing features. This theory has retained the divergence:convergence of minicolumn:macrocolumn efferent:afferent interactions by considering domains of minicolumns as having similar synaptic interactions within the extent of a macrocolumn. This dynamically macrocolumnar-averaged minicolumn is designated in this theory as a ''mesocolumn.''

This being the empirical situation, it is interesting that N≈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 1970's [34, 37]. N is small enough to permit nearest-neighbor (NN) 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 [49], 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 [50]. This is a definite mathematical convenience, else a macrocolumn of ~10^3 minicolumns would have to be described by a system of

minicolumns with up to sixteenth order next-nearest neighbors. 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 to 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.

After statistically shaping the microscopic system, the parameters of the mesoscopic system are minicolumnar-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 [10].

The mathematical development of mesocolumns establishes a mesoscopic Lagrangian $\underline{L}$, which may be considered as a ''cost function.'' 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., $|G|$.

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

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

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

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

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

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

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

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

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

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

$$a^G_{G'} = \frac{1}{2}A^G_{G'} + B^G_{G'} , \qquad (2)$$

where $A^G_{G'}$ and $B^G_{G'}$ are minicolumnar-averaged inter-neuronal synaptic efficacies, $v^G_{G'}$ and $\phi^G_{G'}$ are averaged means and variances of contributions to neuronal electric polarizations, and NN interactions $\underline{V}'$ are detailed in other SMNI papers.

Macroscopic Regions

Inclusion of all the above microscopic and mesoscopic features of neocortex permits a true nonphenomenological Gaussian-Markovian formal development for macroscopic regions encompassing $\sim 5\times 10^5$ minicolumns of spatial extent $\sim 5\times 10^9$ μm^2, albeit one that is still highly nonlinear and nonequilibrium. The development of mesocolumnar domains presents conditional probability distributions for

mesocolumnar firings with spatially coupled NN interactions. The macroscopic spatial folding of these mesoscopic domains and their macroscopic temporal folding of tens to hundreds of τ, with a resolution of at least τ/N [11], yields a true path-integral formulation, in terms of a Lagrangian possessing a *bona fide* variational principle for most-probable firing states. At this point in formal development, no continuous-time approximation has yet been made; this is done, with clear justification, for some applications discussed in the next section. This is relevant, e.g., to the possibility of chaotic behavior in neocortex [10], which, neglecting NN interactions, is essentially a time-discretized, two-dimensional (M^G), dissipative, stochastic system. Much of this algebra is greatly facilitated by, but does not require, the use of Riemannian geometry to develop the nonlinear means, variances, and ''potential'' contributions to the Lagrangian [37].

The mathematical macroscopic development proceeds by ''folding'' the mesoscopic probability distribution over and over, in time θ,

$$\dot{M}^G = [M^G(t+\theta) - M^G(t)]/\theta \ , \ \theta < \tau \ , \tag{3}$$

and in a space ~ Λ~5×10^5 macrocolumns ~ 5×10^9 μm^2. For momentary simplicity, consider the folding of just one variable M at just one spatial point over many time epochs: Labelling u intermediate time epochs by s, i.e., $t_s = t_0 + s\Delta t$, in the limits $\lim_{u\to\infty}$ and $\lim_{\Delta t\to 0}$, and assuming $M_{t_0} = M(t_0)$ and $M_t = M(t \equiv t_{u+1})$ are fixed,

$$P[M_t|M_{t_0}] = \int \cdots \int dM_{t-\Delta t} dM_{t-2\Delta t} \cdots dM_{t_0+\Delta t}$$

$$\times P[M_t|M_{t-\Delta t}] P[M_{t-\Delta t}|M_{t-2\Delta t}] \times \cdots P[M_{t_0+\Delta t}|M_{t_0}] \ ,$$

$$P[M_t|M_{t_0}] = \int \cdots \int \underline{D}M \exp(-\sum_{s=0}^{u} \Delta t \underline{L}_s) \ ,$$

$$\underline{D}M = (2\pi \hat{g}_0^2 \Delta t)^{-1/2} \prod_{s=1}^{u} (2\pi \hat{g}_s^2 \Delta t)^{-1/2} dM_s \ ,$$

$$\int dM_s \to \sum_{\alpha=1}^{N} \Delta M_{\alpha s} \ , M_0 = M_{t_0} \ , M_{u+1} = M_t \ , \tag{4}$$

where α labels the range of N values of M. Extension to multiple variables, e.g., G = E and I, and to many cells, e.g., a region of mesocolumns, is discussed in Section IV.2 below.

Mesocolumns were derived in a ''prepoint'' discretization, e.g.,

$$\dot{M}_s^G = [M^G(t+\theta) - M(t)]/\theta \ ,$$

$$g_s^G = g^G[M^G(t), t] \ . \tag{5}$$

There are a number of non-trivial technical points which must be considered when dealing with multivariate nonlinear systems. Very fortunate for this theory, the necessary mathematical techniques for handling such systems were developed by physicists is the late 1970's, and this neuroscience problem is the first physical system that used these methods.

To capture a flavor of some of the mathematical technicalities, consider that there exists a transformation to the midpoint discretization, in which the standard rules of differential calculus hold for the *same* distribution in terms of a transformed $\underline{L}$, defined as a Feynman Lagrangian $\underline{L}_F$.

$$M^G(\bar{t}_s) = \frac{1}{2}(M_{s+1}^G + M_s^G) \ , \ \dot{M}^G(\bar{t}_s) = (M_{s+1}^G - M_s^G)/\theta \ . \tag{6}$$

I.e., expanding all prepoint-discretized functions about the midpoint ($t + \theta/2$ above) introduces many additional terms, which are recognized as having the same structure of a Riemannian geometry induced on the M^G variables. These will be specified in more detail in Section IV.2.

Using the midpoint discretization, the variational principle offers insight, but the prepoint discretization does not contain explicit Riemannian terms. The nonlinear variances considerably complicate the algebra required. Riemannian geometry facilitates, but is not necessary, to derive these results. The Riemannian geometry is a reflection that the probability distribution is invariant under general nonlinear transformations of these variables. In other words, the same information content can be expressed in a variety of ways. For example, sensory cortex may transmit information to motor cortex, although they have somewhat different neuronal structures or neuronal languages. Information can be transmitted between ''different-looking'' regions, e.g., between motor cortex and sensory cortex:

$$I = \int \underline{D}\tilde{M}\ \tilde{P} \ln(\tilde{P}/\bar{P}) \ . \tag{7}$$

3. Applications

Several papers have described in detail how this theory can be used to advantage [8-14]. These applications provide a conceptual framework for treating other similar systems, e.g., those of C^3I.

(A) Intuitive view of statistical analyses. Three-dimensional views over $E - I$ of the stationary Lagrangian offers an intuitive ''potential'' description of neocortical interactions, detailing local minima and maxima [9, 10]. Such pairwise presentation of variables offers an intuitive and accurate estimate of relative probabilities and variances associated with multiple minima.

(B) Inclusion of global circuitry. The path-integral formalism permits straightforward extension of this development to include constraints on short-ranged mesocolumnar interactions induced by long-ranged fibers of greater spatial extent than macrocolumnar distances, e.g., long-ranged excitatory fibers from ipsilateral association, contralateral commissural, and thalamocortical processes [9, 10]. Such constraints may be viewed as global commands issued to mesoscopic domains, which must use their own internal algorithms on their microscopic units to meet these constraints.

(C) Processing of patterned information. Firing states linearized about stationary firing states, give rise to simple eigenfunction expansions of the macroscopic probability distribution [8, 9]. These eigenfunctions are to be identified with the algebraic vector spaces utilized to great advantage by other investigators [51, 52], but not derived by them from realistic synaptic interactions respecting the nonlinear statistical nature of this dynamic system. This identification will permit detailed numerical calculation of associative learning, retrieval and storage of memories, etc. For example, the accuracy of retrieval of a specific pattern is directly proportional to the overlap of a STM ''search''-eigenfunction with a long-term memory (LTM) stored eigenfunction. These eigenfunctions may encompass various degrees of neural mass [50], ranging from minicolumns, to aggregates of mesocolumns coupled by NN interactions, to regions coupled by long-ranged fibers.

More specifically, learning and retrieval mechanisms can be developed by first determining expansion coefficients of eigenfunction expansions of the differential Fokker-Planck distributions, e.g., considering stationary states as Hermite polynomials in neighborhoods of minima. Although this is a reasonably large computer calculation, similar calculations of greater computational difficulty have been performed many years ago, e.g., when calculating quantum states of Schr̈odinger wave-functions of nucleon-nucleon scattering and of nuclear matter, using realistic forces—i.e., quite nonlinear nucleon-nucleon forces derived from meson-exchange forces [19]. The Fokker-Planck equation is quite similar to the Schr̈odinger equation, and this analogy recently has been used to great advantage, to apply the modern methods used here for neocortex to determine Riemannian contributions to nuclear forces [20-22]. These methods can be very useful for classical systems as well.

(D) Phase transitions and Catastrophes. Higher-order polynomial expansions about stationary states yield Ginsburg-Landau expressions, from which first-order and second-order phase transitions can be exhibited, if they exist [10, 53]. The polynomial expansions, with coefficients derived from empirical synaptic parameters, are a starting point from which to apply methods of Catastrophe Theory, e.g., as discussed by Alex Woodcock at this conference. Such investigations can offer insights into mechanisms that severely alter the global context of a system.

(E) Coding of long-term-memory. A precise scenario of neocortical information processing is detailed, from coding of long-ranged firings from stimuli external to a macrocolumn by short-ranged

mesocolumnar firings, to STM storage via hysteresis, and to LTM storage via plastic deformation [10]. In contrast to the appearance of multiple minima in the interior of M^G-space, which are candidates for multiple STM under conditions of sensitive adjustment of synaptic interactions, (see sub-Section G below) [11], typically one or at most a few minima appear at the corners of M^G-space, corresponding to all G-neurons collectively firing or not firing [10]. When these corner minima are present, they are typically much deeper than those found for the interior minima, corresponding to longer-lived states with properties of hysteresis rather than simple jumps. These corner minima are therefore candidates for LTM phenomena. Similar properties of corner minima in simpler models of neocortex have been shown to satisfy properties desirable for multistable perception [54] and for collective computational properties [55]. LTM illustrates the adaptive capabilities of neocortex, a feature very useful for other distributed systems.

(F) Wave-propagation dispersion relations and alpha frequency. Only after the multiple minima are established, then it may be useful to perform linear expansions about specific minima specified by the Euler-Lagrange variational equations. This permits the development of stability analyses and dispersion relations in frequency-wavenumber space [9, 10, 12]. This calculation requires the inclusion of global constraints, discussed in (B) above.

More specifically, the variational principle permits derivation of the Euler-Lagrange equations. These equations are then linearized about a given local minima to investigate oscillatory behavior. Here, long ranged constraints in the form of Lagrange multipliers J_G were used to efficiently search for minima, corresponding to roots of the Euler-Lagrange equations.

$$0 = \hat{\delta}\underline{L}_F = \underline{L}_{F,\dot{G}:t} - \hat{\delta}_G \underline{L}_F$$

$$\approx -\underline{f}_{|G|}\ddot{\underline{M}}^{|G|} + \underline{f}^1_G \dot{\underline{M}}^{G^\neg} - \underline{g}_{|G|}\nabla^2 \underline{M}^{|G|} + \underline{b}_{|G|}\underline{M}^{|G|} + \underline{b}\,\underline{M}^{G^\neg} ,$$

$$[\cdots]_{,\dot{G}:t} = [\cdots]_{,\dot{G}G'}\dot{M}^{G'} + [\cdots]_{,\dot{G}\dot{G}'}\ddot{M}^{G'} , \; G^\neg \neq G ,$$

$$\underline{M}^G = M^G - \ll \bar{M}^G \gg ,$$

$$\underline{M}^G = \mathrm{Re}\,\underline{M}^G_{\mathrm{osc}} \exp[-i(\underline{\xi}\cdot r - \omega t)] , \; \xi = |\underline{\xi}| ,$$

$$\underline{M}^G_{\mathrm{osc}}(r,t) = \int d^2\underline{\xi}\, d\omega\, \hat{\underline{M}}^G_{\mathrm{osc}}(\underline{\xi},\omega)\exp[i(\underline{\xi}\cdot r - \omega t)] ,$$

$$\omega\tau = \pm\{-1.86 + 2.38(\xi\rho)^2; -1.25i + 1.51i(\xi\rho)^2\} . \tag{8}$$

It is calculated that

$$\omega \sim 10^2 \ \mathrm{sec}^{-1} , \tag{9}$$

which is equivalent to

$$\nu = \omega/(2\pi) = 16 \ \mathrm{cps\ (Hz)} , \tag{10}$$

as observed for the alpha frequency.

The propagation velocity v is calculated from

$$v = d\omega/d\xi \approx 1 \ \mathrm{cm/sec} , \; \xi \sim 30\rho , \tag{11}$$

which tests the NN interactions. Thus, within 10^{-1} sec, short-ranged interactions over several minicolumns of 10^{-1} cm may simultaneously interact with long-ranged interactions over tens of cm, since the long-ranged interactions are speeded by myelinated fibers affording velocities of 600−900 cm/sec [56]. In other words, interaction among different neocortical modalities, e.g., visual, auditory, etc., may simultaneously interact within the same time scales, as observed.

This propagation velocity is consistent with the observed movement of attention [57] and with the observed movement of hallucinations across the visual field [58], of ~1/2 mm/sec, about 5 times as slow as v. (I.e., the observed movement is ~ 8 msec/°, and a macrocolumn ~ mm processes 180° of visual

field.) Therefore, NN interactions may play some part, i.e., within several interations of interactions, in disengaging and orienting selective attention.

(G) Short-term-memory capacity. The most detailed and dramatic application of this theory has been to predict a stochastic mechanism underlying the phenomena of human STM capacity [11, 13], transpiring on the order of tenths of a second to seconds, limited to the retention of 7 ± 2 items [59]. 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 also are limited to a STM capacity of 7 ± 2 items [60]. 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 may be limited to as few as two or three chunks [61]. This STM capacity-limited chunking phenomena also has been noted with items requiring varying depths and breadths of processing [5-7, 16, 17]. Another interesting phenomena of STM capacity explained by this theory is the primacy vs. 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 [62].

STM is the mechanism by which neocortex holds multiple hypotheses for further processing. Multiple minima of Lagrangians modeling similar systems can be similarly analyzed. Contour plots of the stationary Lagrangian, $\overline{L}$, for typical synaptic parameters balanced between predominately inhibitory and predominately excitatory firing states, are examined at many scales when the background synaptic noise is only modestly shifted to cause both efferent and afferent mesocolumnar firing states to have a common most-probable firing, centered at [11]

$$M^G = M^{*G} = 0 \ . \tag{12}$$

Within the range of synaptic parameters considered, for values of $\tau \overline{L} \sim 10^{-2}$, this ''centering'' mechanism causes the appearance of from 5 to 10−11 extrema for values of $\tau \overline{L}$ on the order of $\sim 10^{-2}$. In the absence of external constraints and this centering mechanism, no stable minima are found in the interior of M^G space. I.e., the system either shuts down, with no firings, or it becomes epileptic, with maximal firings at the upper limits of excitatory or of excitatory and inhibitory firings. The appearance of these extrema due to the centering mechanism is clearly dependent on the nonlinearities present in the derived Lagrangian, stressing competition and cooperation among excitatory and inhibitory interactions at columnar as well as at neuronal scales.

These number of minima are determined when the resolution of the contours is commensurate with the resolution of columnar firings, i.e., on the order of five to ten neuronal firing per columnar mesh point. Most important contributions to the probability distribution P come from ranges of the time-slice θ and the ''action'' $N \underline{L}$, such that $\theta N \underline{L} \leq 1$. By considering the contributions to the first and second moments of ΔM^G for small time slices θ, conditions on the time and variable meshes can be derived [63, 64].

$$< M^G(t+\theta) - M^G(t) > \approx g^G(t)\theta \ ,$$

$$< [M^G(t+\theta) - M^G(t)]^2 > \approx g^{GG}(t)\theta \ . \tag{13}$$

The time slice is determined by $\theta \leq (N \overline{L})^{-1}$ throughout the ranges of M^G giving the most important contributions to the probability distribution P. The variable mesh, a function of M^G, is optimally chosen such that ΔM^G is measured by the covariance $g^{GG'}$ (diagonal in neocortex due to independence of E and I chemical interactions), or $\Delta M^G \sim (g^{GG}\theta)^{1/2}$ in the notation of the SMNI papers. For $N \sim 10^2$ and $\overline{L} \sim 10^{-2}/\tau$, it is reasonable to pick $\theta \sim \tau$. Then it is calculated that that optimal meshes are $\Delta M^E \sim 7$ and $\Delta M^I \sim 4$, essentially the resolutions used in the coarse contour plots.

Since the extrema appear to lie fairly well along a line in the two-dimensional M^G-space, and since coefficients of slowly varying dM^G/dt terms in the nonstationary $\underline{L}$ are noted to be small perturbations on $\overline{L}$ [10], a solution to the stationary probability distribution was hypothesized to be proportional to $\exp(-\Phi/D)$, where $\Phi = CN^2\overline{L}$, the diffusion $D = N/\tau$, and C a constant.

$$P_{\rm stat} \approx N_{\rm stat} g^{1/2} \exp(-\Phi/D) \ ,$$

$$\Phi = CN^2\overline{\underline{L}} \sim CN^2 \int dM^G L_{,G} \ ,$$

$$D = N/\tau \ . \tag{14}$$

Along the line of the extrema, for $C \approx 1$, this Φ is determined to be an accurate solution to the full two-dimensional Fokker-Planck equation [13], and a weak-noise high-barrier regime defined by $\Delta\Phi/D > 1$, where $\Delta\Phi$ is the difference in Φ from minima to maxima, can be assumed for further analyses [65].

$$0 = \frac{\partial P}{\partial t} = \frac{1}{2}(g^{GG'}P)_{,GG'} - (g^G P)_{,G} + N\underline{V}P \ . \tag{15}$$

This is extremely useful, as a linear stability analysis,

$$\delta \dot{M}^G \approx - N^2 \overline{\underline{L}}_{,GG'} \delta M^{G'} \ , \tag{16}$$

shows that stability with respect to mesocolumnar fluctuations induced by several neurons changing their firings is determined by the second derivatives of $-\Phi$ [66]; here this just measures the parabolic curvature of $\overline{\underline{L}}$ at the extrema. Thus, all the extrema of the stationary Lagrangian are determined to be stable minima of the time-dependent dynamic system. Note however, that it is unlikely that a true potential exists over all M^G-space [67].

This stationary solution is also useful for calculating the time of first passage, t_{vp}, to fluctuate out of a valley in one minima over a peak to another minima.

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

$$\times \exp\{CN\tau[\overline{\underline{L}}(\ll \bar{M} \gg_p) - \overline{\underline{L}}(\ll \bar{M} \gg_v)]\} \ . \tag{17}$$

It turns out that the values of $\tau\overline{\underline{L}} \sim 10^{-2}$ for which the minima exist are just right to give t_{vp} on the order of tenths a second for about 9 of the minima when the maximum of 10−11 are present. The other minima give t_{vp} on the order of many seconds, which is large enough to cause hysteresis to dominate single jumps between other minima [11]. Thus, 7 ± 2 is the capacity of STM, for memories or new patterns which can be accessed in any order during tenths of a second, all as observed empirically [60]. (When the number of neurons/minicolumn is taken to be ~220, modeling visual neocortex [11], then the minima become deeper and sharper, consistent with sharper depth of processing, but several minima become isolated from the main group. This effect might be responsible for the lowering of STM capacity for visual processing, mentioned above.)

This is a very sensitive calculation. If N were a factor of 10 larger, or if $\tau\overline{\underline{L}} \sim 0.1$ at the minima, then t_{vp} is on the order of hours instead of tenths of seconds, becoming unrealistic for STM durations. Oppositely, if t_{vp} were much smaller, i.e., less than $\sim 5\tau$, this would be inconsistent with empirical time scales necessary for formation of any memory trace [68]. In this context, it is noted that the threshold factor of the probability distribution scales as $(N^* N)^{1/2}$, demanding that both the macrocolumnar divergence and minicolumnar convergence of mesocolumnar firings be tested by these calculations.

Yin-Yang Processing of Information

This theory demonstrates that, relatively independent of local information-processing at the submicroscopic synaptic and microscopic neuronal scales, there is statistical global processing of patterns of information at the mesoscopic and macroscopic scales.

This picture represents neocortex as a pattern-processing computer. The underlying mathematical theory, i.e., the path-integral approach, specifies a parallel-processing algorithm which statistically finds those parameter-regions of firing which contribute most to the overall probability distribution: This is a kind of ''intuitive'' algorithm, globally searching a large multivariate data base to find parameter-regions deserving more detailed local information-processing. The derived probability distribution can be thought of as a filter, or processor, of incoming patterns of information. This filter is adaptive, as it can be

modified as it interacts with previously stored patterns of information, changing the mesoscopic synaptic parameters.

IV. APPLICATIONS OF BI TO C^3I

1. A Generic System

(A) Target Variables—Recognition In order to make the mathematics more transparent, consider a grid defined within a given time epoch, where the grid is to be conceived as a generalized ''radar'' screen, representing data being accumulated by multiple sensors. Each cell has information pertaining to relocatable targets that may be moving between cells. Each ''×'' represents a minimal set of targets, e.g., clusters of targets, which have a number of associated variables, e.g., coordinate position, velocity, acceleration, numbers of targets within these categories, etc. The information collected within each time epoch serves to define changes in these variables between neighboring epochs, both within each cell and between neighboring cells.

Thus, large sets of problems are defined by requiring algorithms to recognize and parametrize changing patterns of these target variables.

(B) Decision-Making Variables—Response It must also be assumed, if objective responses to targets are required, that decision-making variables be defined and functionally parametrized. These variables may include properties of actions to be taken, consistently scaled to match target variables.

Thus, larger sets of problems are defined by requiring algorithms to parameterize and to optimally allocate decision-making variables according to the perceived changing patterns of target variables defined in (A). It is also reasonable to expect that any algorithm for response, i.e., in contradistinction to mere recognition, somehow consistently fold in the parameters of both (A) and (B).

(C) Response-Time and Computational Constraints These problems are further exasperated by the real nature of physical systems. Not much time may be available to optimally solve the problems defined in (A) and (B).

Thus, larger sets of problems are defined by requiring algorithms to respond to problems in (A) and (B), but so constrained that they may not be able to always predict the absolutely best response. It may be necessary to settle for a ''good'' response.

(D) Fitting and Predicting Error, Noise and Risk Given the absence of perfect humans and of perfect machines, it is clear that any algorithm addressing the problems in (A), (B) and (C) require some degree of parametrization and modeling. There exist some errors in attempting to match any algorithm to a given genuine complex physical system. In order to minimize these errors to within required tolerances, these errors must be quantified.

By design of the targets or by design of the sensors, there also exists some degree of background noise tending to thwart a completely deterministic description of the target variables. This noise must be quantified, at least in order to assess a measure of credibility given to the identification of changing patterns of target variables.

The size and complexity of real physical systems, and the response-time and computational constraints described in (C), dictate that without always being able to make a best single decision, there exist elements of risk in any response algorithm. This risk must be quantified, at least in order to assess the chances to be taken by alternative responses. The ''expected gain'' of any response is the sum of products of each possible response multiplied by its associated risk, assuming independence among responses; otherwise, cross-correlations must be assessed and folded into this analysis.

Thus, larger sets of problems are defined by requiring algorithms to consistently include fits of variances (error, noise, risk) of all parameters in (A), (B) and (C). Only if variances are consistently fitted, can the mean values (signals), approximately corresponding to the otherwise deterministic parameters in the hypothetical absence of these variances, be extracted. Only if past events include these ''2nd moment'' fits, i.e., only by fitting *bona fide* probability distributions, can the future be optimally predicted, albeit only with some (quantifiable) degree of statistical (un)certainty.

2. Method of Solution

(A) One Variable, One Cell There are three equivalent representations of this stochastic system.

For momentary simplicity, again consider the above ''radar'' grid, but now consider only one parameter, $M(t)$, in just one cell, representing just one of the variables discussed in Section (1A) or Section (1B). The problem of determining the change of M within time Δt is

$$M(t+\Delta t) - M(t) = \Delta t\ f[M(t)]\,, \tag{18}$$

where $f[M]$ is some function to be fit, which describes how M is changing. For small enough Δt, and assuming continuity of M, this is often written as

$$\dot{M} = \frac{dM}{dt} = f\,. \tag{19}$$

If background noise, η, is present, assumed to be Gaussian-Markovian (''white'' noise), then this affects the description of changing M by

$$\dot{M} = f + \hat{g}\eta\,,$$

$$< \eta(t) >_\eta = 0\,,$$

$$< \eta(t)\eta(t') >_\eta = \delta(t-t')\,, \tag{20}$$

where $\hat{g}^2$ is the (constant here) variance of the background noise. Here η is assumed to have a zero mean. Eq. (20) is referred to as a Langevin rate-equation in the scientific literature.

Physicists and engineers, e.g., in fluid mechanics, recognize an equivalent ''diffusion'' equation to Eq. (20), defining a differential equation for the conditional probability distribution, $P[M(t+\Delta t)|M(t)]$, of finding M at the time $t+\Delta t$, given its value at time t.

$$\frac{\partial P}{\partial t} = \frac{\partial(-fP)}{\partial M} + \frac{1}{2}\frac{\partial^2(\hat{g}^2 P)}{\partial M^2} \tag{21}$$

is known as a Fokker-Planck equation.

Some physicists, e.g., in elementary-particle physics, are familiar with yet another representation of Eq. (20) or (21). For small time epochs, the conditional probability P is

$$P[M_{t+\Delta t}|M_t] = (2\pi\hat{g}^2\Delta t)^{-1/2}\exp(-\Delta t L)\,,$$

$$L = (\dot{M} - f)^2/(2\hat{g}^2)\,. \tag{22}$$

L is defined to be the Lagrangian. This representation for P permits a ''global'' path-integral description of the evolution of P from time t_0 to a long time t, i.e., in contradistinction to the ''local'' differential Eq. (21). Labelling u intermediate time epochs by s, i.e., $t_s = t_0 + s\Delta t$, in the limits $\lim_{u\to\infty}$ and $\lim_{\Delta t\to 0}$, and assuming $M_{t_0} = M(t_0)$ and $M_t = M(t \equiv t_{u+1})$ are fixed,

$$P[M_t|M_{t_0}] = \int\cdots\int dM_{t-\Delta t}\,dM_{t-2\Delta t}\cdots dM_{t_0+\Delta t}$$

$$\times P[M_t|M_{t-\Delta t}]P[M_{t-\Delta t}|M_{t-2\Delta t}]\times\cdots P[M_{t_0+\Delta t}|M_{t_0}]\,,$$

$$P[M_t|M_{t_0}] = \int\cdots\int \underline{D}M\exp(-\sum_{s=0}^{u}\Delta t L_s)\,,$$

$$\underline{D}M = (2\pi\hat{g}_0^2\Delta t)^{-1/2}\prod_{s=1}^{u}(2\pi\hat{g}_s^2\Delta t)^{-1/2}dM_s\,,$$

$$\int dM_s \to \sum_{\alpha=1}^{N}\Delta M_{\alpha s}\,,\ M_0 = M_{t_0}\,,\ M_{u+1} = M_t\,, \tag{23}$$

where α labels the range of N values of M. For notational simplicity, the indices s and α often will be dropped in the following, but these time and range discretizations must of course be explicitly programmed in all actual numerical calculations.

There are some advantages to the path-integral representation over its equivalent Fokker-Planck and rate-equation representations. For example, there exists a variational principle wherein a set of Euler-Lagrange differential equations exist for the Lagrangian L, directly yielding those values or trajectories of M which give the largest contribution to the probability distribution P.

Because P is a *bona fide* probability distribution, there exist Monte Carlo numerical algorithms, sampling the M-space without having to calculate all values of M at all intermediate time epochs from t_0 to t to find P. This numerical algorithm also has the nice feature of avoiding traps in local minima when there are deeper minima to be had, representing more probable states. This is so useful that noise is sometimes artificially added to otherwise deterministic systems, e.g., as in simulated annealing [69] to derive optimum circuitry on chips, by hypothesizing a cost function similar to the potential Φ in Eq. (14) in Section III. More efficient simulated annealing algorithms for finding a global minimum of a cost function or set of data have been discussed by Harold Szu at this conference.

In practice, some of these benefits are often illusory. Monte Carlo methods are notoriously poor for most nonstationary systems with multiple minima. However, a new method has been developed for explicitly solving the path integral, thereby obtaining the dynamic evolution of all states (minima) of the system [63, 64]. This cannot be done with the differential equation representations. Calculating P via the path integral facilitates the inclusion of boundary conditions, and the new methods also can take advantage of the Gaussian-Markovian nature of the system to produce an efficient numerical algorithm.

(B) Many Nonlinear Variables It is possible to formulate Langevin equations generalized from Eq. (20),

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

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

$$G = 1, \cdots, \Theta , \tag{24}$$

where G corresponds to any number of Θ variables, e.g., target and decision-making variables in (IA) and (IB), f^G and $\hat{g}_i^G$ are arbitrarily nonlinear functions of any or all M^G, and of t, and the index i corresponds to recognizing that there can be many different sources contributing to the variance of M^G. The time of evaluation of $\hat{g}_{si}^G$ during s-epochs intermediate between t_0 and t, $\bar{t}_s$ between t_s and $t_{s+1} = t_s + \Delta t$, must now be explicitly prescribed. Unless otherwise specified, a midpoint Stratonovich rule will be chosen here, using $M^G(\bar{t}_s) = \frac{1}{2}(M_{s+1}^G + M_s^G)$, $\dot{M}^G(\bar{t}_s) = (M_{s+1}^G - M_s^G)/\Delta t$, and $\bar{t}_s = t_s + \Delta t/2$. This choice is consistent with other physical systems, and allows the use of standard calculus in Eq. (24).

The path integral generalized from Eq. (23) is written as

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

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

$$\int dM_s^G \rightarrow \sum_{\alpha=1}^{N^G} \Delta M_{\alpha s}^G , M_0^G = M_{t_0}^G , M_{u+1}^G = M_t^G ,$$

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

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

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

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

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

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

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

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

$$R_{FJKL} = \frac{1}{2}(g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN}(\Gamma^M_{FK}\Gamma^N_{JL} - \Gamma^M_{FL}\Gamma^N_{JK}) . \tag{25}$$

Note that the variance $g^{GG'}$ is the GG'-matrix inverse of the G-space metric $g_{GG'}$. R is calculated to be the Riemannian curvature scalar, and Γ^F_{JK} is the affine connection in this space.

(C) Many Cells For many cells, i.e., Λ cells indexed by ν, the path integral in Eq. (25) is further generalized, essentially by expanding the parameter space from the set $\{G\}$ to the set $\{G, \nu\}$.

Constraints may be placed on variables by adding them to the potential $\tilde{V}_s$, e.g., as $J_{sG\nu} M^{G\nu}_s$ with Lagrange multipliers $J_{sG\nu}$.

If a prepoint-discretization rule is adopted, transforming from the midpoint-discretized Feynman $\tilde{L}_s$ and $\tilde{g}_{s_+}$, to define $\dot{M}^{G\nu}(\bar{t}_s) = (M^{G\nu}_{s+1} - M^{G\nu}_s)/\Delta t$, $M^{G\nu}(\bar{t}_s) = M^{G\nu}_s$, $\bar{t}_s = t_s$, and $\tilde{g}_{s_+} = \tilde{g}_s$, then a simpler expression is obtained for the Lagrangian, one in which the Riemannian terms are not explicitly present.

$$\tilde{L}' = \frac{1}{2}(\dot{M}^{G\nu} - g^{G\nu}) g_{GG'\nu\nu'} (\dot{M}^{G'\nu'} - g^{G'\nu'}) - \tilde{V} . \tag{26}$$

However, although $\tilde{P}$ is invariant under this transformation, $\tilde{L}'$ does not possess the variational principle possessed by the Feynman Lagrangian $\tilde{L}$, so that if the prepoint-discretized $\tilde{L}'$ and $\tilde{g}_{s_+}$ are used to fit the data, then some tests must still be made to see how efficiently the path integral can be calculated using $\tilde{L}'$ instead of $\tilde{L}$ to globally scan the data.

Eq. (25) (or first its equivalent prepoint discretization) will be fit to the data by assuming functional forms for $\tilde{V}_s$, $g^{G\nu}_s$ and $g^{GG'\nu\nu'}_s$. The convergence of $\tilde{L}$ or $\tilde{L}'$ is expected to be quite good. I.e., even polynomial forms for $g^{G\nu}_s$ and $g^{GG'\nu\nu'}_s$, with coefficients to be fit, define a Padé rational approximate to $\tilde{L}$ usually giving better convergence than obtained for $g^{G\nu}_s$ or $g^{GG'\nu\nu'}_s$ separately. Also, note that $\tilde{L}_s$ is a single scalar function to be fit.

$$g^G = X^G + X^G_{G'} \underline{M}^{G'} + X^G_{G'G''} \underline{M}^{G'} \underline{M}^{G''} + \cdots ,$$

$$g_{GG'} = Y_{GG'} + Y_{GG'G''} \underline{M}^{G''} + Y_{GG'G''G'''} \underline{M}^{G''} \underline{M}^{G'''} + \cdots ,$$

$$\underline{M}^{G\nu}_s = M^{G\nu}_s - \ll M^{G\nu}_s \gg . \tag{27}$$

Once the parameters $\{X, Y, \ll M \gg\}$ are fit, the theory is ready to track or predict. Science is not only empiricism. Modeling and chunking of information is required, not only for aesthetics, but also to reduce required computational resources of brains as well as machines.

3. Future Research and Development

Given a complex system possessing many variables, I believe it appropriate to initially apply some non-parametric statistical methods as a coarse ''macroscopic'' filter to discover, even in real time, some systematics of the system. An example is mentioned in the next Section V.

These macroscopic systematics can form the basis of a first-order set of trial functions for a ''mesoscopic'' filter, e.g., modeled as a parametric nonlinear nonequilibrium Gaussian Markovian statistical

mechanics, as discussed above [25]. This filter can be used to ascertain just what scope of the underlying variable space should be allocated further detailed, more expensive and time-consuming processing by relatively microscopic algorithms. Or, this mesoscopic filter may be sufficient, e.g., for ''shotgun'' responses to clusters of targets.

The final level of detailed processing most likely needs to be performed by a ''microscopic'' fine filter which is *not* explicitly dependent on macroscopic or mesoscopic properties. Markovian or Gaussian properties generally are only appropriate and useful for aggregates of microscopic details. Typically, specific complex systems at the microscopic level exhibit even fewer typical features than the typically novel features discovered even at the mesoscopic level.

My work in neuroscience discussed above suggests an approach for implementing the mesoscopic filter into hardwiring. Consider each cell of the ''radar'' screen above now be represented as one ν-cell at a given time labeled by s. Each circle consists of ~10^2 on-off bits, representing N^G α-states of one G-variable $M^{G\nu}_{\alpha s}$ in that ν-cell at time s, which therefore represents a field rather than a simple binary node. Each circle statistically reacts to the other circles in that cell and in $\nu_{\rm NN}$ cells at time $s-1$, according to an algorithm encoded in each ν-cell. Long-ranged constraints might be added by superimposing (magnetic) fields, i.e., modeling the $J_{sG\nu}$ constraints described in Section (2C) above.

V. COMBAT SIMULATIONS

An important class of problems confronting C^3I systems concerns how to pass through enough, but not too much, timely information to decision-makers to permit them to assess the overall ''macroscopic'' nature of detailed ''microscopic'' operations unfolding in time. Similarly, there must also be a reasonable information-conduit through which their macroscopic decisions can be effectively implemented at the microscopic level.

It is proposed that modern methods of nonlinear nonequilibrium statistical mechanics be utilized to approach such problems, not just to merely model abstract scenarios. Basically, this approach seeks to define a ''mesoscopic'' scale, established between the microscopic and macroscopic scales, specifically appropriate to each C^3I system: nonlinear multivariate functions describing drifts (trends) and diffusions (risks) must be sought. This requires trial and error, intelligence and creativity, and much experience to be gained by dealing with at least several C^3I systems. These functional forms and their coefficients must be fit to real empirical data, e.g., initial, intermediate and final resources, to develop a time-dependent multivariable probability distribution of order parameters defining the mesoscopic scale. Then, after this algebraic and numerical development, there is the possibility that the resulting codes can be implemented on small computers in the field, affording useful software support for decision-making and intelligence-gathering, while being robust against perturbations in these functional fits.

At NPS, Stephen Upton and I are developing statistical mechanical C^3I models of combat simulations. As pointed out in the Introduction Section I, simulations can be an important source of empirical data, only if their assumptions are clearly recognized. I.e., they are at best only as good as they model actual combat [70].

Our primary focus is an NPS simulation reported at this conference by Mike Sovereign and Joe Stewart, Interim Battle Group Tactical Trainer (IBGTT). IBGTT is rather unique in possessing a high degree of human-machine interactions. It is hoped that by fitting nonlinear statistical mechanical models to this data, we may capture the essence of realistic combat operations. The previous work ar NPS has accomplished a coarse macroscopic linear regression of three years of data, e.g., as discussed in Section IV.3. We plan to construct the mesoscopic model.

Another simulation we are investigating for a similar mesoscopic analysis is to model the C^3 system of a Marine Air-Ground Task Force (MAGTF), composed of four elements: Command, Ground Combat, Aviation Combat, Combat Service Support. An example of such a simulation is the Tactical Warfare Simulation, Evaluation and Analysis System (TWSEAS). These are located at: MC Development and Education Center (MCDEC), Quantico, VA; Camp Lejune, NC; Camp Pendelton, CA. There are three types of MAGTF's: Marine amphibious unit (MAU), Marine amphibious brigade (MAB), Marine amphibious force (MAF). The C^3 structure of all MAGTF's is given in the chart below.

The MAGTF order parameters M^G of the air support might include measures of readiness (aggregated by other relatively microscopic algorithms) of: (1.a) weapons carried and (1.b) personnel carried; in turn (1.a) and (1.b) depend on the order parameters defining capabilities of the ground troops and the logistic systems. By establishing a functional probability distribution that might truly describe the dynamic MAGTF, i.e., admitting arbitrarily nonlinear drifts and diffusions, alternative scenarios can be objectively assessed by commanders who are presented with information at a level commensurate with their tasks, and their decisions can be established as constraints on the mesoscopic cells of microscopic networks of the MAGTF.

VI. STATISTICAL BI DECISION-MAKING

A typical scenario that might take advantage of previous analysis that has fit a Lagrangian to previous data follows. For example, assume that in the middle of an engagement, a commander (human or machine) has available data representing measures of readiness of his forces and those of his enemy. He makes a judgement as to which of several established classes of conflict he is engaged in, e.g., possibly severely or moderately stochastic, possibly overwhelming resources in, or not in, his favor, etc. He chooses one of previously established Lagrangians which is a coarse description of his present engagement, and sets the initial time boundary condition according to his present data.

He chooses some time in the future when he feels he will be called on to make a judgement with regard to the deployment of his resources. He uses a small computer to determine the distribution of his variables at the future time. Most likely, he will obtain several possible likely states, with varying degrees of first moments (''probability'') and second moments (''risk'').

He might do this for several alternative initial parameter settings, especially if he can exercise some immediate control of their values, thereby obtaining another possible set of future states of the engagement. He also might have to fold in some constraints, in the form of Lagrange multipliers, to accommodate orders he has received from a higher command. He could also use the associated Euler-Lagrange variational equations to determine the most likely trajectory that his resources would follow enroute from his present state to his selected future state.

Thus, the commander has obtained a valuable source of information to aid him in making decisions, and in determining sets of orders of constraints which he should pass down to his subordinates. Conversely, his subordinates, by aggregating their data into the specified order parameters, can communicate information to their commander in a language readily accessible to his decision-making process.

REFERENCES

1. B.G. Blair, *Strategic Command and Control*, Brooking Institution, Washington, DC, (1985).
2. Eastport Study Group, *A Report to the Director Strategic Defense Initiative Organization*, SDIO, Washington, DC, (1985).
3. G.E. Orr, *Combat Operations C^3I: Fundamentals and Interactions*, Air University, Maxwell Air Force Bace, AL, (1983).
4. L. Ingber, Physics of karate techniques, Sensei Thesis, Japan Karate Association, Tokyo, Japan, (1968).
5. L. Ingber, *The Karate Instructor's Handbook*, Physical Studies Institute-Institute for the Study of Attention, Solana Beach, CA, (1976).
6. L. Ingber, *Karate: Kinematics and Dynamics*, Unique, Hollywood, CA, (1981).
7. L. Ingber, *Elements of Advanced Karate*, Ohara, Burbank, CA, (1985).
8. L. Ingber, Towards a unified brain theory, *J. Social Biol. Struct.* **4**, 211-224 (1981).
9. L. Ingber, Statistical mechanics of neocortical interactions. I. Basic formulation, *Physica D* **5**, 83-107 (1982).
10. L. Ingber, Statistical mechanics of neocortical interactions. Dynamics of synaptic modification, *Phys. Rev. A* **28**, 395-416 (1983).

11. L. Ingber, Statistical mechanics of neocortical interactions. Derivation of short-term-memory capacity, *Phys. Rev. A* **29**, 3346-3358 (1984).
12. L. Ingber, Statistical mechanics of neocortical interactions. EEG dispersion relations, *IEEE Trans. Biomed. Eng.* **32**, 91-94 (1985).
13. L. Ingber, Statistical mechanics of neocortical interactions: Stability and duration of the 7±2 rule of short-term-memory capacity, *Phys. Rev. A* **31**, 1183-1186 (1985).
14. L. Ingber, Towards clinical applications of statistical mechanics of neocortical interactions, *Innov. Tech. Biol. Med.* **6**, 753-758 (1985).
15. L. Ingber, Statistical mechanics of neocortical interactions, *Bull. Am. Phys. Soc.* **31**, 868 (1986).
16. L. Ingber, Editorial: Learning to learn, *Explore* **7**, 5-8 (1972).
17. L. Ingber, Attention, physics and teaching, *J. Social Biol. Struct.* **4**, 225-235 (1981).
18. L. Ingber, Statistical mechanics algorithm for response to targets (SMART), in *Workshop on Uncertainty and Probability in Artificial Intelligence: UC Los Angeles, 14-16 August 1985*, (Edited by P. Cheeseman), pp. 258-264, American Association for Artificial Intelligence, Menlo Park, CA, (1985).
19. L. Ingber, Nuclear forces, *Phys. Rev.* **174**, 1250-1263 (1968).
20. L. Ingber, Riemannian corrections to velocity-dependent nuclear forces, *Phys. Rev. C* **28**, 2536-2539 (1983).
21. L. Ingber, Path-integral Riemannian contributions to nuclear Schrödinger equation, *Phys. Rev. D* **29**, 1171-1174 (1984).
22. L. Ingber, Riemannian contributions to short-ranged velocity-dependent nucleon-nucleon interactions, *Phys. Rev. D* **33**, 3781-3784 (1986).
23. L. Ingber, Statistical mechanics of nonlinear nonequilibrium financial markets, *Math. Modelling* **5** (6), 343-361 (1984).
24. J.G. Taylor, *Lanchester Models of Warfare*, Operations Research Society of America, Arlington, VA, (1983).
25. National Research Council, *Committee on the Applications of Mathematics: Computational Modeling and Mathematics Applied to the Physical Sciences*, National Academy Press, Washington, DC, (1984).
26. H. Haken, *Synergetics*, Springer, New York, (1983).
27. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam, (1981).
28. J.D. Bekenstein and L. Parker, Path integrals for a particle in curved space, *Phys. Rev. D* **23**, 2850-2869 (1981).
29. K.S. Cheng, Quantization of a general dynamical system by Feynman's path integration formulation, *J. Math. Phys.* **13**, 1723-1726 (1972).
30. H. Dekker, Functional integration and the Onsager-Machlup Lagrangian for continuous Markov processes in Riemannian geometries, *Phys. Rev. A* **19**, 2102-2111 (1979).
31. B.S. DeWitt, Dynamical theory in curved spaces. I. A review of the classical and quantum action principles, *Rev. Mod. Phys.* **29**, 377-397 (1957).
32. H. Grabert and M.S. Green, Fluctuations and nonlinear irreversible processes, *Phys. Rev. A* **19**, 1747-1756 (1979).
33. H. Grabert, R. Graham, and M.S. Green, Fluctuations and nonlinear irreversible processes. II., *Phys. Rev. A* **21**, 2136-2146 (1980).
34. R. Graham, Covariant formulation of non-equilibrium statistical thermodynamics, *Z. Physik* **B26**, 397-405 (1977).

35. R. Graham, Lagrangian for diffusion in curved phase space, *Phys. Rev. Lett.* **38**, 51-53 (1977).
36. T. Kawai, Quantum action principle in curved space, *Found. Phys.* **5**, 143-158 (1975).
37. F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions*, Reidel, Dordrecht, The Netherlands, (1982).
38. M.M. Mizrahi, Phase space path integrals, without limiting procedure, *J. Math. Phys.* **19**, 298-307 (1978).
39. L.S. Schulman, *Techniques and Applications of Path Integration*, J. Wiley & Sons, New York, (1981).
40. R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York, (1965).
41. S.J. Williamson, L. Kaufman, and D. Brenner, Evoked neuromagnetic fields of the human brain, *J. Appl. Phys.* **50**, 2418-2421 (1979).
42. G.M. Shepherd, *The Synaptic Organization of the Brain*, Oxford University Press, New York, NY, (1979).
43. 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).
44. 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).
45. 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).
46. 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).
47. V.B. Mountcastle, An organizing principle for cerebral function: The unit module and the distributed system, in *The Mindful Brain*, (Edited by G.M. Edelman and V.B. Mountcastle), pp. 7-50, Massachusetts Institute of Technology, Cambridge, (1978).
48. T.H. Bullock, Reassessment of neural connectivity and its specification, in *Information Processing in the Nervous System*, (Edited by H.M. Pinsker and W.D. Willis, Jr.), Raven Press, New York, NY, (1980).
49. R.W. Dykes, Parallel processing of somatosensory information: A theory, *Brain Res. Rev.* **6**, 47-115 (1983).
50. R.P Erickson, The across-fiber pattern theory: An organizing principle for molar neural function, *Sensory Physiol.* **6**, 79-110 (1982).
51. J.A. Anderson, J.W. Silverstein, S.A. Ritz, and R.S. Jones, Distinctive features, categorical perception and probability learning: Some applications of a neural model, *Psych. Rev.* **84**, 413-451 (1977).
52. L.N. Cooper, A possible organization of animal memory and learning, in *Collective Properties of Physical Systems*, (Edited by B. Lundqvist and S. Lundqvist), pp. 252-264, Academic Press, New York, NY, (1973).
53. S.-K. Ma, *Modern Theory of Critical Phenomena*, Benjamin/Cummings, Reading, MA, (1976).
54. A.H. Kawamoto and J.A. Anderson, A neural model of multistable perception, *Acta Psychologica* **59**, 35-65 (1985).
55. J.J. Hopfield, Neurons with graded responses have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci. USA* **81**, 3088-3092 (1984).
56. P.L. Nunez, *Electric Fields of the Brain: The Neurophysics of EEG*, Oxford University Press, London, (1981).
57. Y. Tsal, Movements of attention across the visual field, *J. Exp. Psychol.* **9**, 523-530 (1983).

58. J.D. Cowan, Spontaneous symmetry breaking in large scale nervous activity, *Int. J. Quant. Chem.* **22**, 1059-1082 (1982).

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

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

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

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

63. 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).

64. 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).

65. S.R. Shenoy and G.S. Agarwal, First-passage times and hysteresis in multivariable stochastic processes: The two-mode ring laser, *Phys. Rev. A* **29**, 1315-1325 (1984).

66. 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).

67. R. Graham and T. Tél, Existence of a potential for dissipative dynamical systems, *Phys. Rev. Lett.* **52**, 9-12 (1984).

68. B. Libet, Brain stimulation in the study of neuronal functions for conscious sensory experience, *Human Neurobiol.* **1**, 235-242 (1982).

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

70. Comptroller General, Models, Data, and War: A Critique of the Foundation for Defense Analyses, Report No. PAD-80-21, U.S. General Accounting Office, Washington, DC, (1980).