Statistical Mechanics of Neocortical Interactions (SMNI): Multiple Scales of Short-Term Memory and EEG Phenomena

Lester Ingber

ingber@ingber.com • ingber@alumni.caltech.edu
http://www.ingber.com/

SMNI RATIONALE—"TOP DOWN"

SMNI DEVELOPMENT—"BOTTOM UP"

MATHEMATICAL AND NUMERICAL ASPECTS

GENERIC MESOSCOPIC NEURAL NETWORKS (MNN)

SHORT-TERM MEMORY (STM)

ELECTROENCEPHALOGRAPHY (EEG)

CANONICAL MOMENTA INDICATORS (CMI) — EEG

CHAOS IN EEG?

SMNI CORRELATES OF REACTION TIMES

SMNI FEATURES

$Id: smni01_lecture,v 1.30 2002/03/14 14:44:36 ingber Exp ingber $

SMNI RATIONALE—"TOP DOWN"

Noninvasive Recordings of Brain Activity

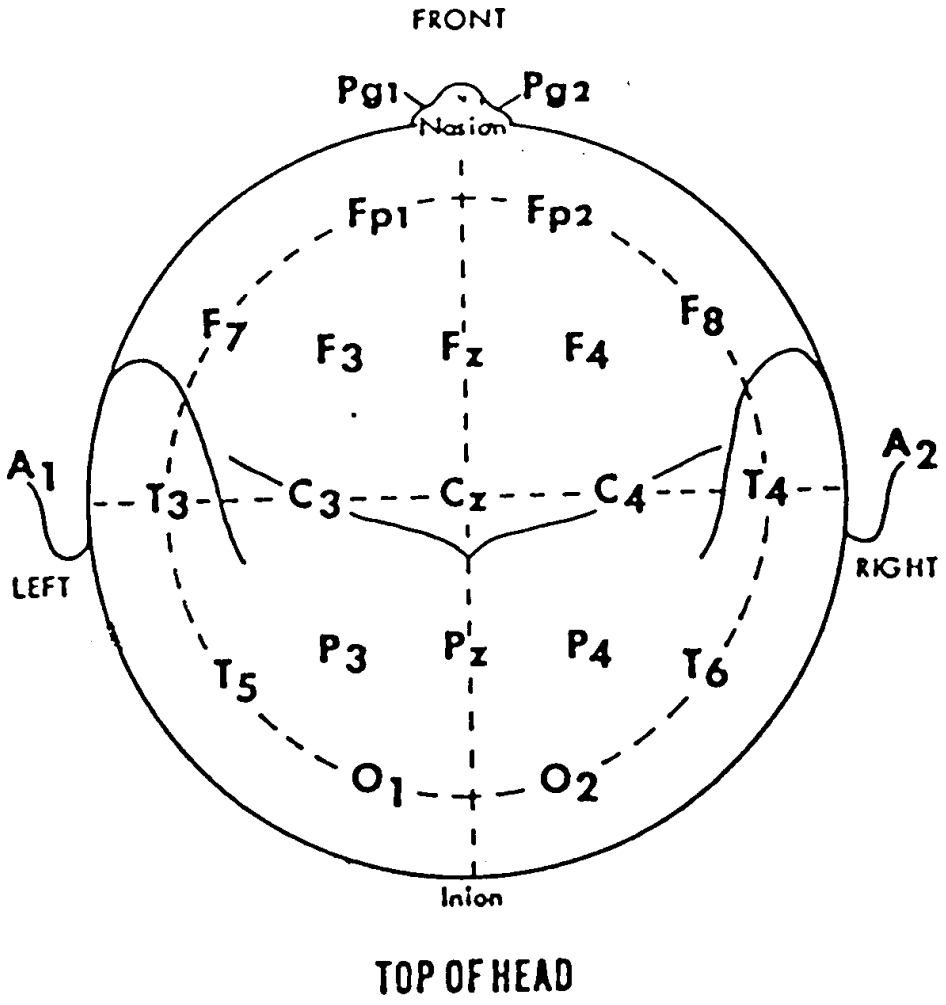
There are several noninvasive experimental or clinical methods of recording brain activity, e.g.,

electroencephalography (EEG)
magnetoencephalography (MEG)
magnetic resonance imaging (MRI)
positron-emission tomography (PET)
single-photon-emission-computed tomography (SPECT)

While MRI, PET, and SPECT offer better three-dimensional presentations of brain activity, EEG and MEG offer superior temporal resolutions on the order of neuronal relaxation times, i.e., milliseconds.

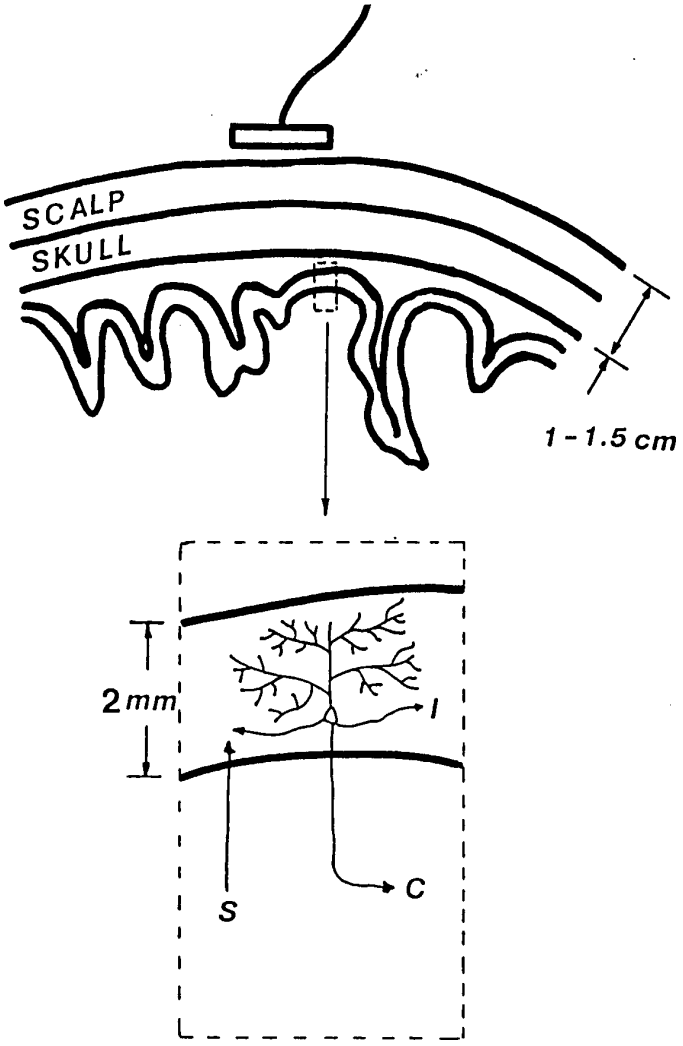
EEG Electrodes

A typical map of EEG electrode sites is given as below. Many neuroscientists are becoming aware that higher electrode densities are required for many studies. For example, if each site below represented 5 closely spaced electrodes, a numerical Laplacian can offer relatively reference-free recordings and better estimates of localized sources of activity.



Single Electrode Recording of Dipole Activity

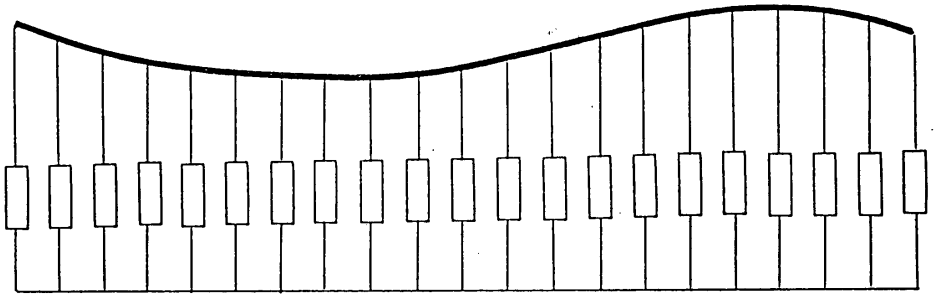
Macrocolumns may be considered as "point sources" of dipole-like interactions, mainly due to coherent current flow of top-layer afferent interactions to bottom-layer efferent interactions. However, there is a problem of non-uniqueness of the electric potential that arises from such source activity; Laplacian measurements can help to address this problem.



EEG of Mechanical String

The mechanical string has linear properties and is connected to local nonlinear oscillators. Local cortical dynamics in dipole layers is here considered analogous to the nonlinear mechanical oscillators which influence global modes. Macroscopic scalp potentials are analogous to the lower modes of string displacement.

For purposes of illustration, a linear string with attached oscillators, e.g., nonlinear springs may be compared to a one-dimensional strip of neocortex:



String Equation

The following equation describes the string displacement Φ

$$\frac{\partial^2\Phi}{\partial t^2} - c^2\,\frac{\partial^2\Phi}{\partial x^2} + [\omega_0^2 + f(\Phi)]\Phi = 0\ ,$$

for a linear array (length l) of sensors (electrodes) of size s. Thus, wave-numbers in the approximate range

$$\frac{\pi}{l} \le k \le \frac{\pi}{s}$$

can be observed. If the center to center spacing of sensors is also s, $l = Ms$, where M = (number of sensors - 1), $k = 2n\pi/R$ for $n = \{1, 2, 3, \cdots\}$ (string forms closed loop), and sensors span half the string (brain), $l = R/2$, then

$$1 \le n \le M$$

for some maximum M, which is on the order of 3 to 7 in EEG studies using 16 to 64 electrodes in two-dimensional arrays on the cortical surface.

For scalp recordings, the wavenumber restriction is more severe. For example, a typical circumference of the neocortex following a coordinate in and out of fissures and sulci is $R = 100$ cm (about 50 cm along the scalp surface). If EEG power is mostly restricted to $k < 0.5$ cm^{-1}, only modes $n < 4$ are observed, independent of the number of electrodes.

Theory should be able to be similarly "filtered," e.g., in order to properly fit EEG data.

String Observables

The string displacement (potential within the cortex) is given by

$$\Phi(x,t) = \sum_{n=1}^{\infty} G_n(t) \sin k_n x \ ,$$

but the observed Φ is given by

$$\Phi^\dagger(x,t) = \sum_{n=1}^{M} G_n(t) \sin k_n x \ .$$

In the linear case, where $f(\Phi) = 0$ (equal linear oscillators to simulate local circuit effects in cortical columns), then

$$\frac{\partial^2 \Phi}{\partial t^2} - c^2 \frac{\partial^2 \Phi}{\partial x^2} + \omega_0^2 \Phi = 0 \ ,$$

$$\Phi = \sum_{n=1}^{\infty} A_n \cos \omega_n t \sin k_n x \ ,$$

$$\omega_n^2 = \omega_0^2 + c^2 k_n^2 \ ,$$

giving a dispersion relation $\omega_n(k_n)$. For the nonlinear case, $f(\Phi) \neq 0$, the restoring force of each spring is amplitude-dependent. In fact, local oscillators may undergo chaotic motion.

What can be said about

$$\Phi^\dagger(x,t) = \sum_{n=1}^{M} G_n(t) \sin k_n x \ ,$$

the macroscopic observable displacement potential on the scalp or cortical surface?

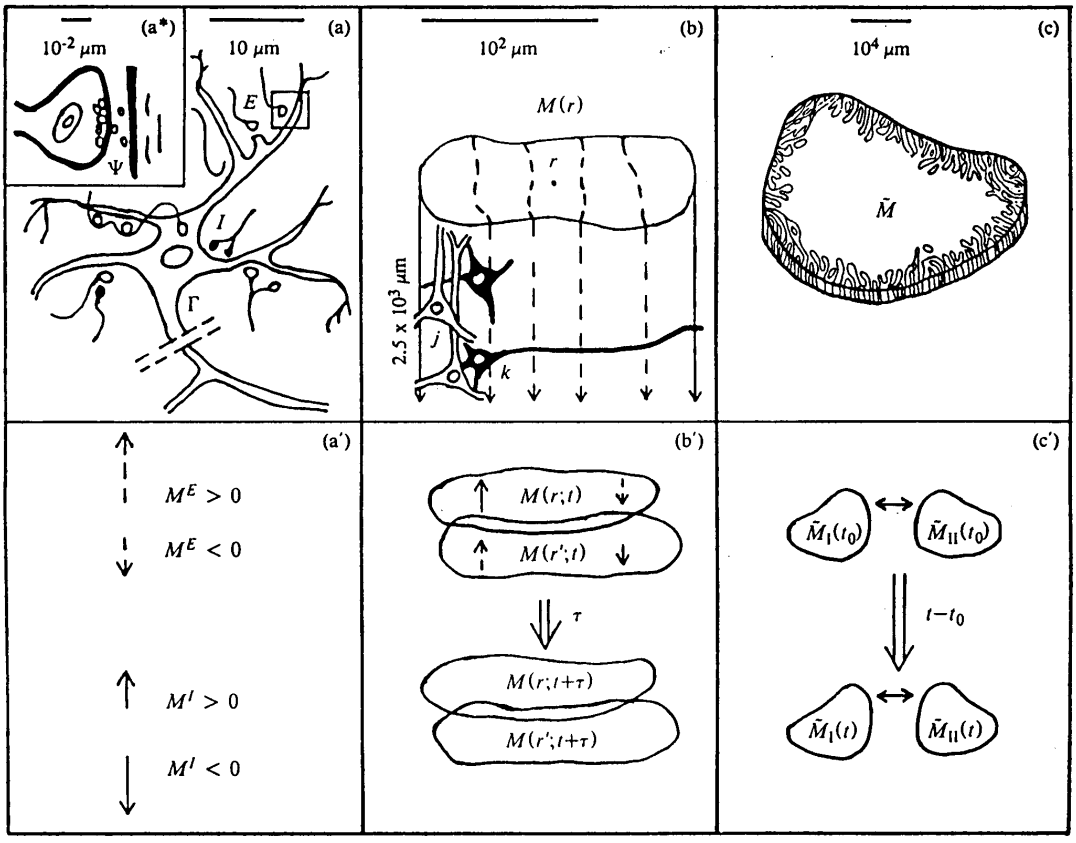
It would seem that $\Phi^\dagger$ should be described as a linear or quasi-linear variable, but influenced by the local nonlinear behavior which crosses the hierarchical level from mesoscopic (columnar dipoles) to macroscopic.

How can this intuition be mathematically articulated, for the purposes of consistent description as well as to lay the foundation for detailed numerical calculations?

SMNI DEVELOPMENT—"BOTTOM UP"

Scales Illustrated

Illustrated are three biophysical scales of neocortical interactions: (a)-(a*)-(a') microscopic neurons; (b)-(b') mesocolumnar domains; (c)-(c') macroscopic regions. In (a*) synaptic interneuronal interactions, averaged over by mesocolumns, are phenomenologically described by the mean and variance of a distribution Ψ. Similarly, in (a) intraneuronal transmissions are phenomenologically described by the mean and variance of Γ. Mesocolumnar averaged excitatory (E) and inhibitory (I) neuronal fi rings are represented in (a'). In (b) the vertical organization of minicolumns is sketched together with their horizontal stratifi cation, yielding a physiological entity, the mesocolumn. In (b') the overlap of interacting mesocolumns is sketched. In (c) macroscopic regions of neocortex are depicted as arising from many mesocolumnar domains. These are the regions designated for study here. (c') sketches how regions may be coupled by long-ranged interactions.



Microscopic Neurons

A derivation has been given of the physics of chemical inter-neuronal and electrical intra-neuronal interactions. This derivation generalized a previous similar derivation. The derivation yields a short-time probability distribution of a given neuron firing due to its just-previous interactions with other neurons. Within τ_j~5−10 msec, the conditional probability that neuron j fires ($\sigma_j = +1$) or does not fire ($\sigma_j = -1$), given its previous interactions with k neurons, is

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

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

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

Γ represents the "intra-neuronal" probability distribution, e.g., of a contribution to polarization achieved at an axon given activity at a synapse, taking into account averaging over different neurons, geometries, etc. Ψ represents the "inter-neuronal" probability distribution, e.g., of thousands of quanta of neurotransmitters released at one neuron's postsynaptic site effecting a (hyper-)polarization at another neuron's presynaptic site, taking into account interactions with neuromodulators, etc. This development is true for Γ Poisson, and for Ψ Poisson or Gaussian.

V_j is the depolarization threshold in the somatic-axonal region, 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.

Even at the microscopic scale of an individual neuron, with soma ≈ 10 μm, this conceptual framework assumes a great deal of statistical aggregation of molecular scales of interaction, e.g., of the biophysics of membranes, of thickness $\approx 5 \times 10^{-3}$ μm, composed of biomolecular leaflets of phospholipid molecules.

Mesoscopic Aggregation

This microscopic scale itself represents a high aggregation of sub-microscopic scales, aggregating effects of tens of thousands of quanta of chemical transmitters as they influence the 5×10^{-3} μm This microscopic scale is aggregated up to the mesoscopic scale, using

$$P_q(q) = \int dq_1 dq_2 P_{q_1 q_2}(q_1, q_2) \delta[q - (q_1 + q_2)] \; .$$

The SMNI approach can be developed without recourse to borrowing paradigms or metaphors from other disciplines. Rather, 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. After such development, advantage can be taken of associated collateral descriptions and intuitions afforded by such mathematical and physics techniques as they have been used in other disciplines, but paradigms and metaphors do not substitute for logical SMNI development.

Mesoscopic Interactions

Microscopic Scale

Retain independence of excitatory (E) and inhibitory (I) interactions

Retain nonlinear development of probability densities

Mesoscopic Scale

Convergence<->Divergence —minicolumnar<->macrocolumnar

Nearest-neighbor (NN) interactions summarize $N^{16}N$ interactions

Macroscopic scale

Include long-ranged interactions, constraints on mesocolumns

For the purposes of mesoscopic and macroscopic investigation, this biological picture can be cast into an equivalent network. However, the above aspects must not be simply cast away.

Mathematical Development

A derived mesoscopic Lagrangian $\underline{L}_M$ defines the short-time probability distribution of firings in a minicolumn, composed of ~10^2 neurons, given its just previous interactions with all other neurons in its macrocolumnar surround. G is used to represent excitatory (E) and inhibitory (I) contributions. $\bar{G}$ designates contributions from both E and I.

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

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

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

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

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

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

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

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

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

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

$$a_{G'}^G = \frac{1}{2}A_{G'}^G + B_{G'}^G ,$$

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. $M^{G'}$ 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, ~10^5. Similarly, $A_G^{G'}$ and $B_G^{G'}$ have been scaled by $N^*/N \sim 10^3$ to keep F^G invariant.

Inclusion of Macroscopic Circuitry

The most important features of this development are described by the Lagrangian $\underline{L}^G$ in the negative of the argument of the exponential describing the probability distribution, and the "threshold factor" F^G describing an important sensitivity of the distribution to changes in its variables and parameters.

To more properly include long-ranged fibers, when it is possible to numerically include interactions among macrocolumns, the J_G terms can be dropped, and more realistically replaced by a modified threshold factor F^G,

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

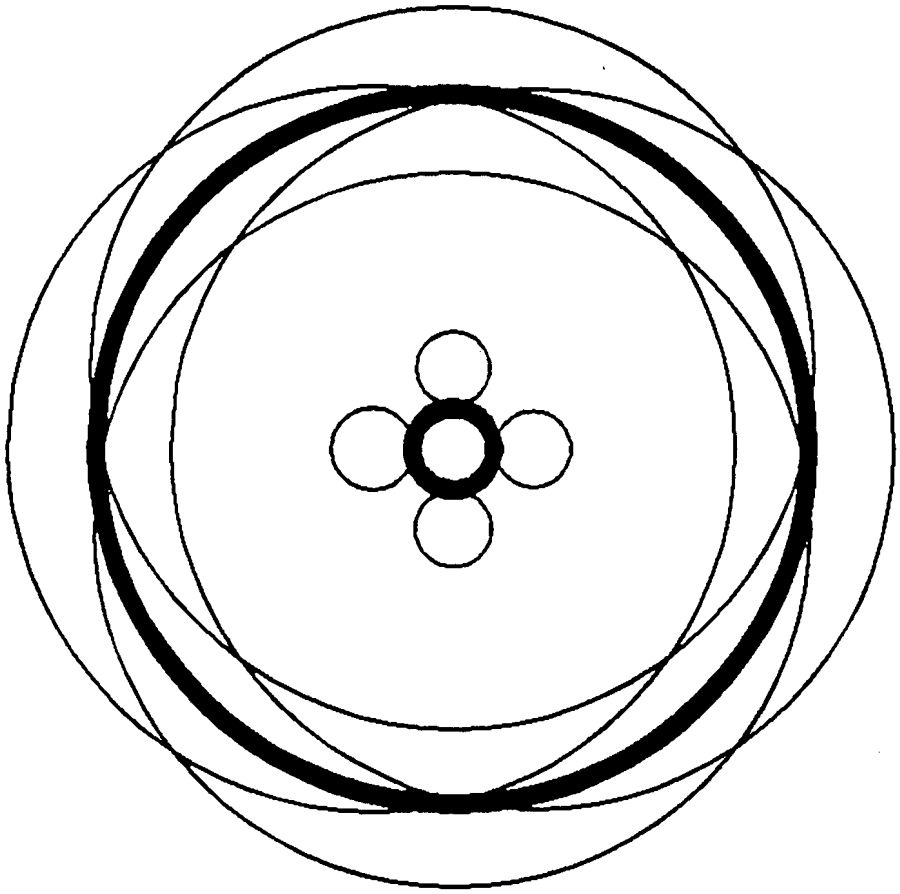
$$a^{\ddagger E}_{E'} = \frac{1}{2} A^{\ddagger E}_{E'} + B^{\ddagger E}_{E'} .$$

Here, afferent contributions from $N^{\ddagger E}$ long-ranged excitatory fibers, e.g., cortico-cortical neurons, have been added, where $N^{\ddagger E}$ might be on the order of 10% of N^*: Of the approximately 10^{10} to 10^{11} neocortical neurons, estimates of the number of pyramidal cells range from 1/10 to 2/3. Nearly every pyramidal cell has an axon branch that makes a cortico-cortical connection; i.e., the number of cortico-cortical fibers is of the order 10^{10}.

Equivalent Nearest-Neighbor Interactions

Nearest-neighbor (NN) interactions between mesocolumns are illustrated. Afferent minicolumns of ~10^2 neurons are represented by the inner circles, and efferent macrocolumns of ~10^5 neurons by the outer circles. Illustrated are the NN interactions between a mesocolumn, represented by the thick circles, and its nearest neighbors, represented by thin circles. The area outside the outer thick circle represents the effective number of efferent macrocolumnar nearest-neighbor neurons. I.e., this is the number of neurons outside the macrocolumnar area of influence of the central minicolumn.

This approximation, albeit successful in the 1983 calculations, can be replaced by a more sophisticated algorithm, MNN, published in 1992.



MATHEMATICAL AND NUMERICAL ASPECTS

Induced Riemannian Geometry

A Riemannian geometry is derived as a consequence of nonlinear noise, reflecting that the probability distribution is invariant under general nonlinear transformations of these variables.

This becomes explicit under a transformation to the midpoint discretization, in which the standard rules of differential calculus hold for the *same* distribution:

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

$$\tilde{P} = \prod_\nu P \ , \ P = \int \cdots \int \underline{D}M \exp(-\sum_{s=0}^{u} \Delta t L_{Fs}) \ ,$$

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

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

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

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

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

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

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

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

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

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

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

Measures of Nonlinear Nonequilibrium

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

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

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

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

where M^G are the variables and L is the Lagrangian. These physical entities provide another form of intuitive, but quantitatively precise, presentation of these analyses.

Representations of Path Integral

The Langevin Rate-Equation exhibits a stochastic equation, wherein drifts can be arbitrarily nonlinear functions, and multiplicative noise is added.

$$M(t+\Delta t) - M(t) \sim \Delta t\ f[M(t)] ,$$

$$\dot{M} = \frac{dM}{dt} \sim f ,$$

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

$$< \eta(t) >_\eta = 0 , < \eta(t)\eta(t') >_\eta = \delta(t-t') .$$

The Diffusion Equation is another equivalent representation of Langevin equations. The first moment "drift" is identified as f, and the second moment "diffusion," the variance, is identified as $\hat{g}^2$.

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

The Path-Integral Lagrangian represents yet another equivalent representation of Langevin equations. Recently it has been demonstrated that the drift and diffusion, in addition to possibly being quite general nonlinear functions of the independent variables and of time explicitly, may also be explicit functions of the distribution P itself.

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

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

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

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

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

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

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

This representation is useful for fitting stochastic data to parameters in L.

Adaptive Simulated Annealing (ASA)

This algorithm fits empirical data to a theoretical cost function over a D-dimensional parameter space, adapting for varying sensitivities of parameters during the fit. This algorithm was first published in 1989, and made publicly available in November 1992.

Heuristic arguments have been developed to demonstrate that this algorithm is faster than the fast Cauchy annealing, $T_i = T_0/k$, and much faster than Boltzmann annealing, $T_i = T_0/\ln k$.

For parameters

$$\alpha_k^i \in [A_i, B_i] ,$$

sampling with the random variable x^i,

$$x^i \in [-1, 1] ,$$

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

define the generating function

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

in terms of parameter "temperatures"

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

The cost-functions $\underline{L}$ under consideration are of the form

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

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

where L is a Lagrangian with dynamic variables $M(t)$, and parameter-coefficients α to be fit to data. g_t is the determinant of the metric, and T is the cost "temperature."

For several test problems, ASA has been shown to be orders of magnitude more efficient than other similar techniques. ASA has been applied to several complex systems, including specific problems in neuroscience, finance and combat systems.

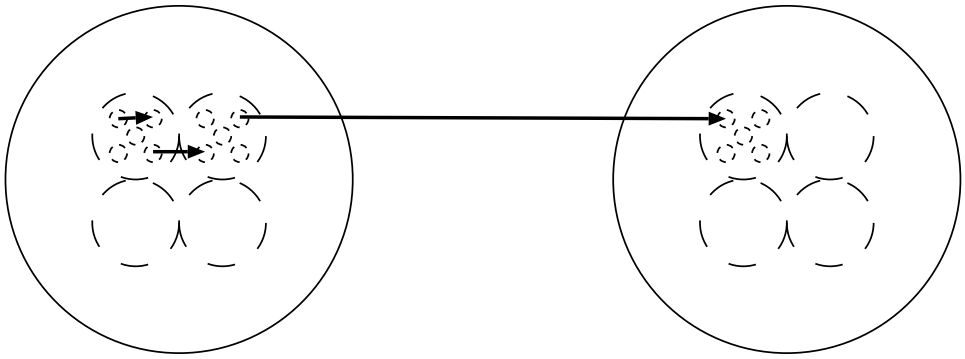
GENERIC MESOSCOPIC NEURAL NETWORKS (MNN)

Applications

Modern stochastic calculus permits development of alternative descriptions of path-integral Lagrangians, Fokker-Planck equations, and Langevin rate equations. The induced Riemannian geometry affords invariance of probability distribution under general nonlinear transformations.

ASA presents a powerful global optimization that has been tested in a variety of problems defi ned by nonlinear Lagrangians.

Parallel-processing computations can be applied to ASA as well as to a neural-network architecture.



MNN Learning

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

The cost function is a sum of effective Lagrangians from each node and over each time epoch of data.

This program of fitting coefficients in Lagrangian uses methods of adaptive simulated annealing (ASA). This maximum likelihood procedure (statistically) avoids problems of trapping in local minima, as experienced by other types of gradient and regression techniques.

MNN Prediction

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

$$M^G(t+\Delta t) - M^G(t) = dM^G = g^G dt + \hat{g}_i^G dW^i \ ,$$

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

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

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

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

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

A coarse deterministic estimate to "predict" the evolution can be applied using the most probable path

$$dM^G/dt = g^G - g^{1/2}(g^{-1/2} g^{GG'})_{,G'} \ .$$

PATHINT, even when parallelized, typically can be too slow for "predicting" evolution of these systems. However, a new algorithm, PATHTREE holds some promise.

MNN Parallel Processing

The use of parallel processors can make this algorithm even more efficient, as ASA lends itself well to parallelization.

During "learning," blocks of random numbers are generated in parallel, and then sequentially checked to find a generating point satisfying all boundary conditions.

Advantage is taken of the low ratio of acceptance to generated points typical in ASA, to generate blocks of cost functions, and then sequentially checked to find the next best current minimum.

Additionally, when fitting dynamic systems, e.g., the three physical systems examined to date, parallelization is attained by independently calculating each time epoch's contribution to the cost function.

Similarly, during "prediction," blocks of random numbers are generated to support the Langevin-equation calculations, and each node is processed in parallel. PATHINT or PATHTREE also possess features to promote fast calculations.

SHORT-TERM MEMORY (STM)

Derivation of Short-Term Memory (STM)

At this mesoscopic scale, properties of STM— its capacity, duration and stability—have been calculated, and found to be consistent with empirical observations. The first publications on this approach to STM appeared in 1984.

The maximum STM capacity, consistent with the 7 ± 2 rule, is obtained when a ''centering mechanism'' is invoked. This occurs when the threshold factor F^G takes minima in the interior of M^G firing-space (i.e., not the corners of this space), as empirically observed.

In the SMNI papers, the background noise $B^G_{G'}$ was reasonably adjusted to center F^G, with $J_G = 0$, but similar results could have been obtained by adjusting the influence of the long-ranged fibers $M^{\ddagger G}$.

Within a time scale of several seconds, the human brain can store only about 7±2 auditory *chunks* of information (4±2 visual chunks).

To derive this, choose empirical ranges of synaptic parameters corresponding to a predominately excitatory case (EC), predominately inhibitory case (IC), and a balanced case (BC) in between. For each case, also consider a ''centering mechanism'' (EC', IC', BC'), whereby some synaptic parameter is internally manipulated, e.g., some chemical neuromodulation or imposition of patterns of firing, such that there is a maximal efficiency of matching of afferent and efferent firings:

$$M^G \approx M^{*G} \approx 0 \ .$$

This sets conditions on other possible minima of the *static* Lagrangian $\underline{\overline{L}}$.

Centering Mechanism

The centering effect 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,

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

for both $G = E$ and $G = I$.

This is modified straightforwardly when regional influences from $M^{\ddagger E}$ are included. 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.

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.

An important side result is to drive most probable states, i.e., small L which is driven largely by small F^G, to regions where

$$v_E^G A_E^G M^E \approx |v_I^G| A_I^G M^I \ .$$

Since I−I efficacies typically are relatively quite small, the probability density under the centering mechanism is strongly peaked along the line

$$v_E^E A_E^E M^E \approx |v_I^E| A_I^E M^I \ .$$

Applying the Centering Mechanism—"Inhibitory" State

A model of dominant inhibition describes how minicolumnar firings are suppressed by their neighboring minicolumns. For example, 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 nonvisual minicolumns are observed to have ~110 neurons and as there appear to be a predominance of E over I neurons, 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

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

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

$F_{\rm IC}^I$ will cause efferent $\bar{M}^I(t+\Delta t)$ to fire for most afferent input firings, as it will be positive for most values of $\bar{M}^G(t)$ 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 weights of afferent $\bar{M}^I$ require at least moderate values of positive afferent $\bar{M}^E$ to cause firings of efferent $\bar{M}^E$, diminishing the influence of $\bar{M}^E$.

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

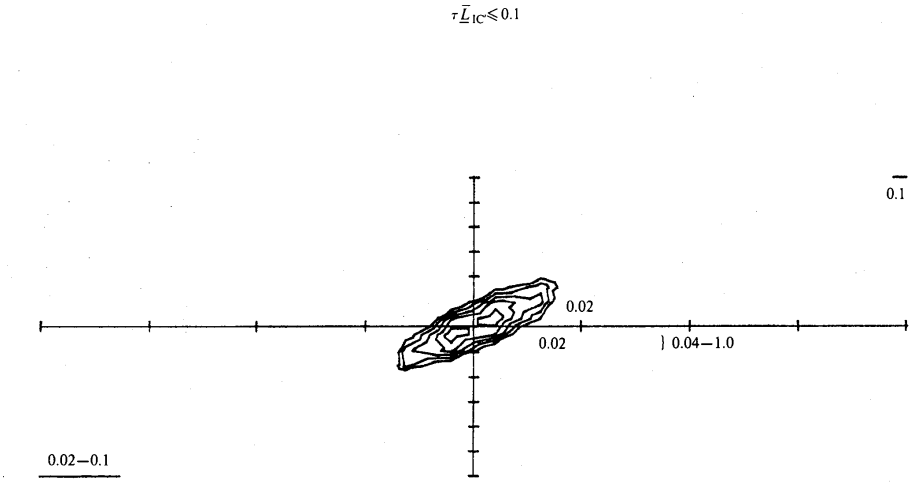
$$F_{\rm IC'}^E = \frac{0.5\bar{M}^I - 0.25\bar{M}^E}{\pi^{1/2}(0.1\bar{M}^I + 0.05\bar{M}^E + 10.4)^{1/2}} ,$$

$$F_{\rm IC'}^I = \frac{0.005\bar{M}^I - 0.5\bar{M}^E}{\pi^{1/2}(0.001\bar{M}^I + 0.1\bar{M}^E + 20.4)^{1/2}} .$$

Contours of "Inhibitory" State

Contours of the Lagrangian illustrate "valleys" that trap firing-states of mesocolumns. ($\tau \overline{\underline{L}}$ can be as large as 10^3.)

No interior stable states are observed at scales of $\tau \overline{\underline{L}}$ ranging from 10^3 down to 10^{-2}, until the "centering mechanism" is turned on.



Applying the Centering Mechanism—"Excitatory" State

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 $A^E_E = 2A^I_E = 2A^E_I = 0.01 N^*/N$. This yields

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

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

The negative constants in the numerators of $F^G_{\rm EC}$ enhance efferent firings for both E and I afferent inputs. However, the increased coefficient of $\bar{M}^E$ in $F^E_{\rm EC}$ (e.g., relative to its corresponding value in $F^E_{\rm IC}$), and the fact that $\bar{M}^E$ can range up to $N^E = 80$, readily enhance excitatory relative to inhibitory firings throughout most of the range of $\bar{M}^E$. This is only a first approximation, and the full Lagrangian must be used to determine the actual evolution.

Using the centering mechanism, $B'^E_E = 10.2$ and $B'^I_I = 8.62$, and $F^G_{\rm EC}$ is transformed to $F^G_{\rm EC'}$,

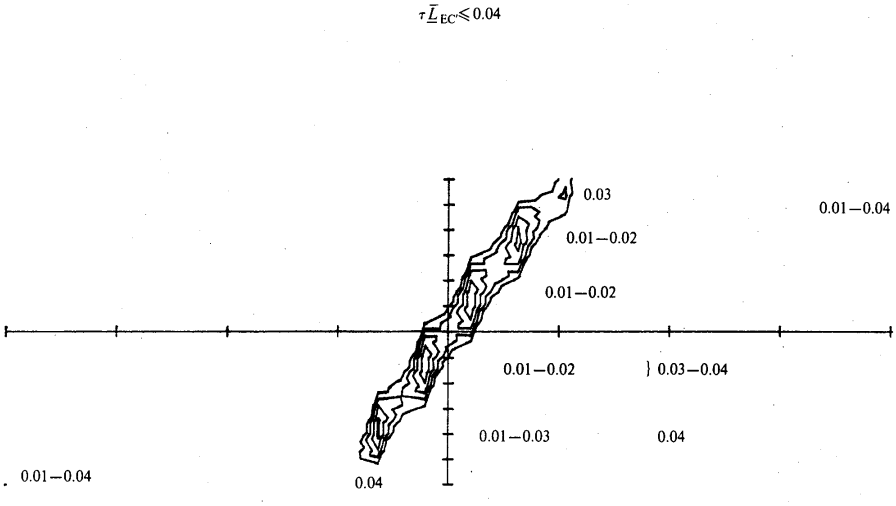
$$F^E_{\rm EC'} = \frac{0.25\bar{M}^I - 0.5\bar{M}^E}{\pi^{1/2}(0.05\bar{M}^I + 0.10\bar{M}^E + 17.2)^{1/2}} ,$$

$$F^I_{\rm EC'} = \frac{0.005\bar{M}^I - 0.25\bar{M}^E}{\pi^{1/2}(0.001\bar{M}^I + 0.05\bar{M}^E + 12.4)^{1/2}} .$$

Contours of "Excitatory" State

Contours of the Lagrangian illustrate "valleys" that trap firing-states of mesocolumns. ($\tau \underline{\overline{L}}$ can be as large as 10^3.)

No interior stable states are observed at scales of $\tau \underline{\overline{L}}$ ranging from 10^3 down to 10^{-2}, until the "centering mechanism" is turned on.



Applying the Centering Mechanism—"Balanced" State

Now it is natural to examine a balanced case intermediate between IC and EC, labeled BC. This is accomplished by changing $A_E^E = A_E^I = A_I^E = 0.005 N^*/N$. This yields

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

$$F_{\rm BC}^I = \frac{0.005\bar{M}^I - 0.25\bar{M}^E - 25.8}{\pi^{1/2}(0.001\bar{M}^I + 0.050\bar{M}^E + 7.24)^{1/2}} .$$

Here the constant in the numerator of $F_{\rm BC}^E$, while still negative to promote E efferent firings, is much greater than that in $F_{\rm EC}^E$, thereby decreasing the net excitatory activity to a more moderate level. A similar argument applies in comparing $F_{\rm BC}^I$ to $F_{\rm IC}^I$, permitting a moderate level of inhibitory firing.

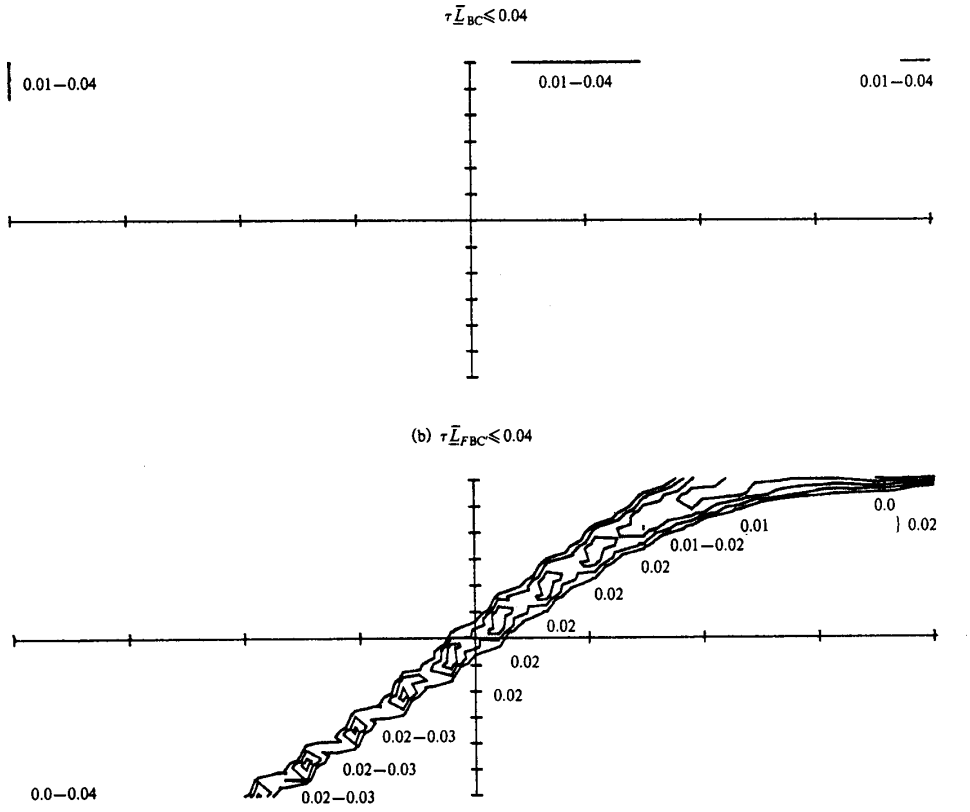
Applying the centering mechanism to BC, $B'^E_E = 0.438$ and $B'^I_I = 8.62$, and $F_{\rm BC}^G$ is transformed to $F_{\rm BC'}^G$,

$$F_{\rm BC'}^E = \frac{0.25\bar{M}^I - 0.25\bar{M}^E}{\pi^{1/2}(0.050\bar{M}^I + 0.050\bar{M}^E + 7.40)^{1/2}} ,$$

$$F_{\rm BC'}^I = \frac{0.005\bar{M}^I - 0.5\bar{M}^E}{\pi^{1/2}(0.001\bar{M}^I + 0.050\bar{M}^E + 12.4)^{1/2}} .$$

Contours of "Balanced" State

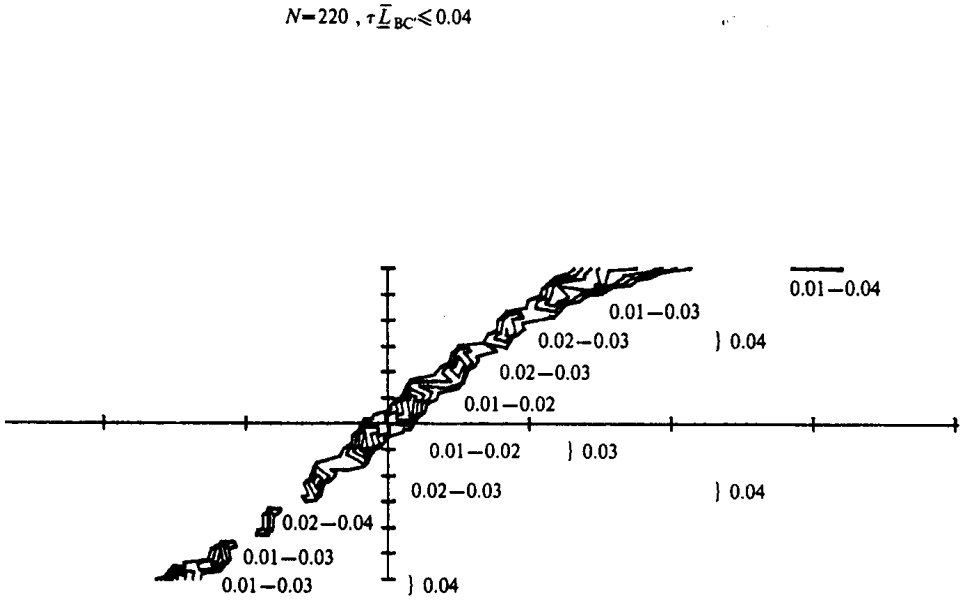
No interior stable states are observed at scales of $\tau\bar{\underline{L}}$ ranging from 10^3 down to 10^{-2}, until the "centering mechanism" is turned on.



(a) Contours for values less than 0.04 are drawn for $\tau\bar{\underline{L}}_{\mathrm{BC}}$. The $\bar{M}^E$ axis increases to the right, from $-N^E = -80$ to $N^E = 80$. The $\bar{M}^I$ axis increases to the right, from $-N^I = -30$ to $N^I = 30$. In each cluster, the smaller values are closer to the center. (b) Contours for values less than 0.04 are drawn for $\tau\bar{\underline{L}}_{\mathrm{BC}'}$.

Modeling Visual Cortex STM

When $N = 220$, modeling the number of neurons per minicolumn in visual neocortex, then only 5-6 minima are found, consistent with visual STM. These minima are narrower, consistent with the sharpness required to store visual patterns.



STM Stability and Duration

The attractors of these models can be identified. Possible hysteresis and/or jumps between local minima can be explicitly calculated within the limitations of studying a specific attractors.

Detailed calculations identify the inner valleys of the parabolic trough with stable short-term-memory states having durations on the order of tenths of a second.

Stability is investigated by

$$\delta \dot{M}^G \approx -N^2 \underline{\bar{L}}_{,GG'} \delta M^{G'} .$$

Therefore, minima of the static Lagrangian $\underline{\bar{L}}$ are minima of the dynamic transient system defined by L. The time of first passage is calculated as

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

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

For $\tau \underline{\bar{L}} \sim 10^{-2}$, the *only* values found for all three cases of firing, the time of first passage t_{vp} is found to be several tenths of second for jumps among most minima, up to 9. There is hysteresis for deeper valleys at 10th-11th minima of $\underline{\bar{L}}_{F\mathrm{BC}'}$ at the corners of the $\bar{M}^G$ plane. The hysteresis occurs in about a few minutes, which is too long to affect the 7 ± 2 rule. This result is *exponentially* sensitive to N in Φ/D, and *exponentially* sensitive to $(N^* N)^{1/2}$ in F^G, the ‘‘threshold factor.’’

Use is made in later development of EEG analyses of the discovered nature of the line of stable minima lying in a deep parabolic trough, across a wide range of cases of extreme types of firings.

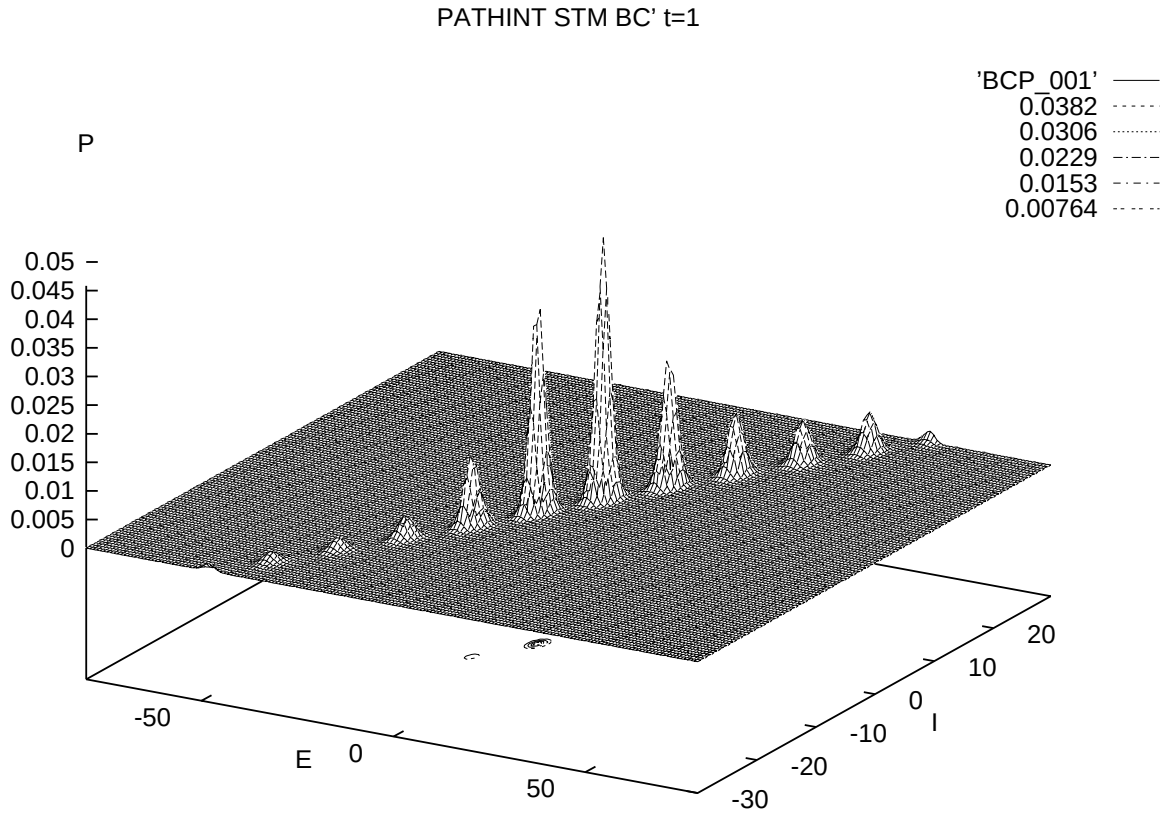
PATHINT Calculations of STM

PATHINT is an important partner to the ASA code. ASA has made it possible to perform fits of complex nonlinear short-time probability distributions to EEG data. PATHINT details the evolution of the attractors of these short-time distributions, e.g., as studied in 1984.

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.

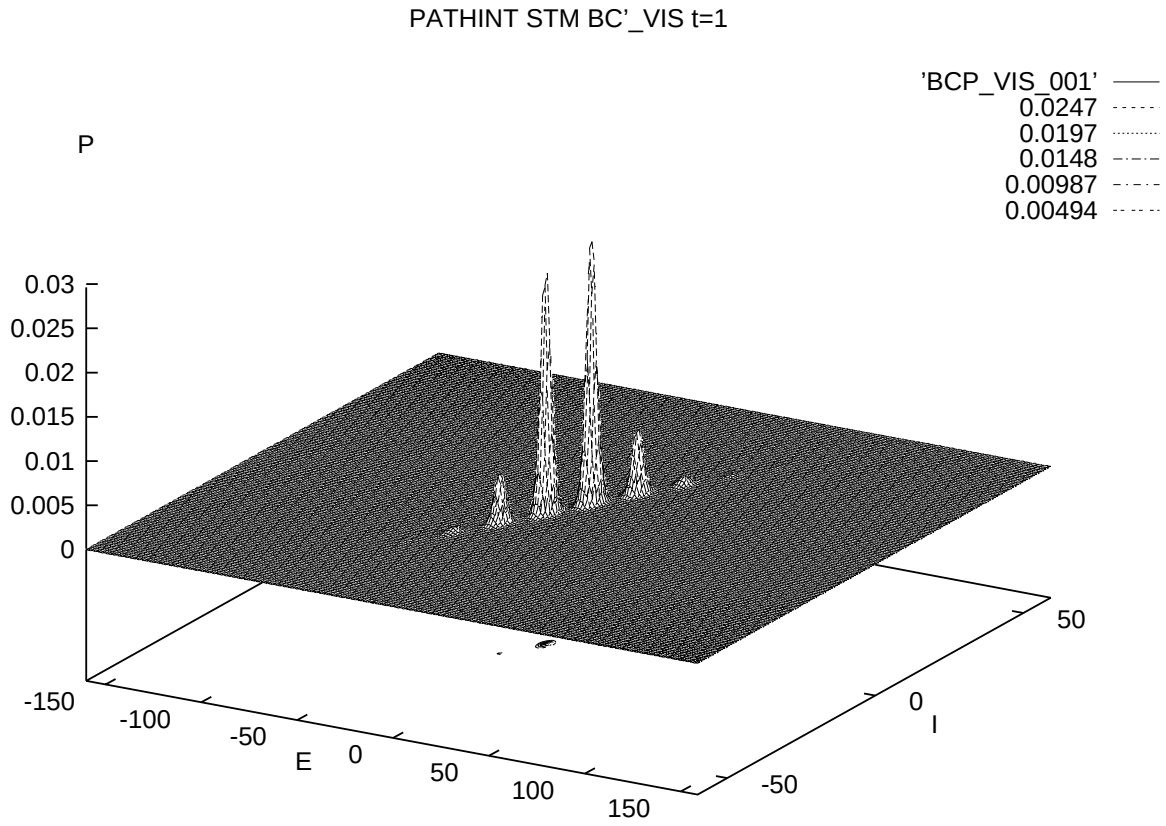
Then, PATHINT can be used to predict the evolution of the system, possibly to predict oncoming states, e.g., epileptic seizures of patients baselined to a fitted distribution.

Below is the evolution of model BC′ at 0.01 seconds = τ, after 100 foldings of the path integral. In agreement with previous studies, 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.



PATHINT Calculations of STM BC′_VIS

The interior of M^G-space of model BC′_VIS is examined at 0.01 seconds = τ.



These high resolution calculations were published in 1995.

Primacy Versus Recency Rule

SMNI also presents an explanation, or at least an upper statistical constraint, on the primacy versus recency rule observed in serial processing.

First-learned items are recalled most error-free, and last-learned items are still more error-free than those in the middle. I.e., the deepest minima are more likely first accessed, while the more recent memories or newer patterns have synaptic parameters most recently tuned or are more actively rehearsed.

Note that for visual cortex, presentation of 7±2 items would have memories distributed among different clusters, and therefore the recency effect should not be observed. Instead the 4±2 rule should dictate the number of presented items.

40 Hz Models of STM

An alternate theory of STM, based on after-depolarization (ADP) at synaptic sites, has been proposed.

Feature	SMNI	ADP
7 ± 2 Rule	attractors of L	40 Hz subcycles
4 ± 2 Rule	attractors of visual L	?
Primacy versus Recency	statistics of attractors	?
Large-Scale Influences	consistent with EEG	?
Duration	local interactions	neuromodulators

ADP proposes a "refresher" mechanism of 40 Hz to sustain memories for time scales on the order of tenths of seconds within cycles of 5−12 Hz, even under the influence of long-ranged regional firings and neuromodulators. SMNI PATHINT calculations show a rapid deterioration of attractors in the absence of external influences.

ADP and SMNI together forge a stronger theory of STM than either separately.

ELECTROENCEPHALOGRAPHY (EEG)

Local and Global EEG

The derived mesoscopic dispersion relations also are consistent with other global macroscopic dispersion relations, described by long-range fibers interacting across regions.

This SMNI model yields oscillatory solutions consistent with the alpha rhythm, i.e., $\omega \approx 10^2$ sec^{-1}, equivalent to $\nu = \omega/(2\pi) \approx 16$ Hz. This suggests that these complementary local and global theories may be confluent, considered as a joint set of dispersion relations evolving from the most likely trajectories of a joint Lagrangian, referred to as the ''equations of motion,'' but linearly simplified in neighborhoods of minima of the stationary Lagrangian.

These two approaches, i.e., local mesocolumnar versus global macrocolumnar, give rise to important alternative conjectures:

(1) Is the EEG global resonance of primarily long-ranged cortical interactions? If so, can relatively short-ranged local firing patterns effectively modulate this frequency and its harmonics, to enhance their information processing across macroscopic regions?

(2) Or, does global circuitry imply boundary conditions on collective mesoscopic states of local firing patterns, and is the EEG a manifestation of these collective local firings?

(3) Or, is the truth some combination of (1) and (2) above? For example, the possibility of generating EEG rhythms from multiple mechanisms at multiple scales of interactions, e.g., as discussed above, may account for weakly damped oscillatory behavior in a variety of physiological conditions.

This theory has allowed the local and global approaches to complement each other at a common level of formal analysis, i.e., yielding the same dispersion relations derived from the ''equations of motion,'' analogous to $\sum(\text{forces}) = d(\text{momentum})/dt$ describing mechanical systems.

EEG Phenomena—Euler-Lagrange Approximation

The variational principle permits derivation of the Euler-Lagrange equations. These equations are then linearized about a given local minima to investigate oscillatory behavior. This calculation was first published in 1983.

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. This illustrates how macroscopic constraints can be imposed on the mesoscopic and microscopic systems.

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

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

$$G^\neg \neq G ,$$

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

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

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

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

It is calculated that

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

which is equivalent to

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

This is approximately within the experimentally observed ranges of the *alpha* and *beta* frequencies.

E-L Propagation of Information

The propagation velocity v is calculated from

$$v = d\omega/d\xi \approx 1 \text{ cm/sec} , \xi \sim 30\rho ,$$

which tests the NN interactions, the weakest part of this theory.

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 and have velocities of 600−900 cm/sec. 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 and with the observed movement of hallucinations across the visual field which moves at ~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.)

Macroscopic Linearization Aids Probability Development

The fitting of the full SMNI probability distribution to EEG data was published in 1991.

Previous STM studies have detailed that the predominant physics of short-term memory and of (short-fiber contribution) to EEG phenomena takes place in a narrow ‘‘parabolic trough’’ in M^G-space, roughly along a diagonal line. I.e., $\tau \underline{L}_M$ can vary by as much as 10^5 from the highest peak to the lowest valley in M^G-space. Therefore, it is reasonable to assume that a single independent firing variable might offer a crude description of this physics. Furthermore, the scalp potential Φ can be considered to be a function of this firing variable.

In an abbreviated notation subscripting the time-dependence,

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

where a and b are constants of the same sign, and $\ll \Phi \gg$ and $\ll M^G \gg$ represent a minima in the trough.

Laplacian techniques help to better localize sources of activity, and thereby present data more suitable for modeling. E.g., then Φ is more directly related to columnar firings, instead of representing the electric potential produced by such activity.

This determines an SMNI approach to study EEG under conditions of selective attention.

EEG Macrocolumnar Lagrangian

Again, aggregation is performed,

$$P_\Phi[\Phi_{t+\Delta t}|\Phi_t] = \int dM^E_{t+\Delta t} dM^I_{t+\Delta t} dM^E_t dM^I_t P_M[M^E_{t+\Delta t}, M^I_{t+\Delta t}|M^E_t, M^I_t]$$
$$\delta[\Phi_{t+\Delta t} - \Phi(M^E_{t+\Delta t}, M^I_{t+\Delta t})]\delta[\Phi_t - \Phi(M^E_t, M^I_t)] \; .$$

Under conditions of selective attention, within the parabolic trough along a line in $\bar{M}^G$ space, the parabolic shapes of the multiple minima, ascertained by the stability analysis, justifies a form

$$P_\Phi = (2\pi\sigma^2 dt)^{-1/2} \exp[-(dt/2\sigma^2) \int dx L_\Phi] \; ,$$

$$L_\Phi = \frac{1}{2} |\partial\Phi/\partial t|^2 - \frac{1}{2} c^2 |\partial\Phi/\partial x|^2 - \frac{1}{2} \omega_0^2 |\Phi|^2 - F(\Phi) \; ,$$

where $F(\Phi)$ contains nonlinearities away from the trough, and where σ^2 is on the order of N, given the derivation of L_M above.

EEG Variational Equation

Previous calculations of EEG phenomena showed that the (short-fiber contribution to the) alpha frequency and the movement of attention across the visual field are consistent with the assumption that the EEG physics is derived from an average over the fluctuations σ of the system. I.e., this is described by the Euler-Lagrange equations derived from the variational principle possessed by L_Φ, more properly by the ''midpoint-discretized'' L_Φ, with its Riemannian terms. Hence,

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

When expressed in the firing state variables, this leads to the same results published in 1983.

The result for the Φ equation is:

$$\frac{\partial^2\Phi}{\partial t^2} - c^2\frac{\partial^2\Phi}{\partial x^2} + \omega_0^2\Phi + \frac{\partial F}{\partial\Phi} = 0 \ .$$

If the identification

$$\frac{\partial F}{\partial\Phi} = \Phi f(\Phi) \ ,$$

is made, then

$$\frac{\partial^2\Phi}{\partial t^2} - c^2\frac{\partial^2\Phi}{\partial x^2} + [\omega_0^2 + f(\Phi)]\Phi = 0 \ ,$$

is recovered, i.e., the dipole-like string equation.

The previous application of the variational principle was at the scale of minicolumns and, with the aid of nearest-neighbor interactions, the spatial-temporal Euler-Lagrange equation gave rise to dispersion relations consistent with STM experimental observations.

Here, the scale of interactions is at the macrocolumnar level, and spatial interactions must be developed taking into account specific regional circuitries.

Macroscopic Coarse-Graining

Now the issue posed previously, how to mathematically justify the intuitive coarse-graining of Φ to get $\Phi^\dagger$, can be approached.

In L_Φ above, consider terms of the form

$$\int \Phi^2 dx = \int dx \sum_n^\infty \sum_m^\infty G_n G_m \sin k_n x \sin k_m x$$

$$= \sum_n \sum_m G_n G_m \int dx \sin k_n x \sin k_m x$$

$$= (2\pi/R) \sum_n G_n^2 \ .$$

By similarly considering all terms in L_Φ, a short-time probability distribution for the change in node n is defined,

$$p_n[G_n(t+\Delta t)|G_n(t)] \ .$$

Note that in general the $F(\Phi)$ term in L_Φ will require coupling between G_n and G_m, $n \neq m$. This defines

$$P_\Phi = p_1 p_2 \cdots p_\infty \ .$$

Now a coarse-graining can be defined that satisfies some physical and mathematical rigor:

$$P_{\Phi^\dagger} = \int dk_{M+1} dk_{M+2} \cdots dk_\infty \ p_1 p_2 \cdots p_M p_{M+1} p_{M+2} \cdots p_\infty \ .$$

I.e., since SMNI is developed in terms of *bona fide* probability distributions, variables which are not observed can be integrated out.

The integration over the fine-grained wave-numbers tends to smooth out the influence of the k_n's for $n > M$, effectively ''renormalizing''

$$G_n \to G^\dagger_n \ ,$$

$$\Phi \to \Phi^\dagger \ ,$$

$$L_\Phi \to L^\dagger_\Phi \ .$$

This development shows how this probability approach to EEG specifically addresses experimental issues at the scale of the more phenomenological dipole model.

Development of Macrocolumnar EEG Distribution

Advantage can be taken of the prepoint discretization, where the postpoint $M^G(t+\Delta t)$ moments are given by

$$m \equiv < \Phi_\nu - \phi > = a < M^E > + b < M^I > = a g^E + b g^I \ ,$$

$$\sigma^2 \equiv < (\Phi_\nu - \phi)^2 > - < \Phi_\nu - \phi >^2 = a^2 g^{EE} + b^2 g^{II} \ .$$

Note that the macroscopic drifts and diffusions of the Φ's are simply linearly related to the mesoscopic drifts and diffusions of the M^G's. For the prepoint $M^G(t)$ firings, the same linear relationship in terms of $\{ \phi, a, b \}$ is assumed.

The data being fit are consistent with invoking the "centering" mechanism. Therefore, for the prepoint $M^E(t)$ firings, the nature of the parabolic trough derived for the STM Lagrangian is taken advantage of, and

$$M^I(t) = c M^E(t) \ ,$$

where the slope c is determined for each electrode site. This permits a complete transformation from M^G variables to Φ variables.

Similarly, as appearing in the modified threshold factor F^G, each regional influence from electrode site μ acting at electrode site ν, given by afferent firings $M^{\ddagger E}$, is taken as

$$M^{\ddagger E}_{\mu\to\nu} = d_\nu M^E_\mu(t - T_{\mu\to\nu}) \ ,$$

where d_ν are constants to be fitted at each electrode site, and $T_{\mu\to\nu}$ is the delay time estimated for inter-electrode signal propagation, typically on the order of one to several multiples of $\tau = 5$ msec. In future fits, some experimentation will be performed, taking the T's as parameters.

This defines the conditional probability distribution for the measured scalp potential Φ_ν,

$$P_\nu[\Phi_\nu(t+\Delta t)|\Phi_\nu(t)] = \frac{1}{(2\pi\sigma^2\Delta t)^{1/2}} \exp(-L_\nu \Delta t) \ ,$$

$$L_\nu = \frac{1}{2\sigma^2} (\dot{\Phi}_\nu - m)^2 \ .$$

The probability distribution for all electrodes is taken to be the product of all these distributions:

$$P = \prod_\nu P_\nu \ ,$$

$$L = \sum_\nu L_\nu \ .$$

Development of EEG Dipole Distribution

The model SMNI, derived for $P[M^G(t+\Delta t)|M^{\bar{G}}(t)]$, is for a macrocolumnar-averaged minicolumn; hence it is expected to be a reasonable approximation to represent a macrocolumn, scaled to its contribution to Φ_v. Hence L is used to represent this macroscopic regional Lagrangian, scaled from its mesoscopic mesocolumnar counterpart $\underline{L}$.

However, the expression for P_v uses the dipole assumption to also use this expression to represent several to many macrocolumns present in a region under an electrode: A macrocolumn has a spatial extent of about a millimeter. Often most data represents a resolution more on the order of up to several centimeters, many macrocolumns.

A scaling is tentatively assumed, to use the expression for the macrocolumnar distribution for the electrode distribution, and see if the fits are consistent with this scaling. One argument in favor of this procedure is that it is generally acknowledged that only a small fraction of firings, those that fire coherently, are responsible for the observed activity being recorded.

The results obtained here seem to confirm that this approximation is in fact quite reasonable.

Key Indicators of EEG Correlates to Brain States

The SMNI probability distribution can be used directly to model EEG data, instead of using just the variational equations. Some important features not previously considered in this field that were used the 1991 were:

- *Intra-Electrode Coherency* is determined by the standard deviations of excitatory and inhibitory firings under a given electrode as calculated using SMNI. Once the SMNI parameters are fit, then these firings are calculated as transformations on the EEG data, as described in terms of the SMNI derived probability distributions. This is primarily a measure of coherent columnar activity.
- *Inter-Electrode Circuitry* is determined by the fraction of available long-ranged fibers under one electrode which actively contribute to activity under another electrode, within the resolution of time given in the data (which is typically greater than or equal to the relative refractory time of most neurons, about 5−10 msec). This is primarily a measure of inter-regional activity/circuitry. Realistic delays can be modeled and fit to data.

The electrical potential of each electrode, labeled by G, is represented by its dipole-like nature, $M^G(t)$, which is influenced by its underlying columnar activity as well as its interactions with other electrodes, $M^{G'}$, $G \neq G'$. This can be expressed as:

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

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

$$\hat{g}_i^G = (N^G/\tau)^{1/2} \mathrm{sech} F^G \ ,$$

$$F^G = \frac{(V^G - a_{G'}^{|G|} v_{G'}^{|G|} N^{G'} - \frac{1}{2} A_{G'}^{|G|} v_{G'}^{|G|} M^{G'} - a_{E'}^{\ddagger E} v_{E'}^{E} N^{\ddagger E'} - \frac{1}{2} A_{E'}^{\ddagger E} v_{E'}^{E} M^{\ddagger E'})}{(\pi[(v_{G'}^{|G|})^2 + (\phi_{G'}^{|G|})^2](a_{G'}^{|G|} N^{G'} + \frac{1}{2} A_{G'}^{|G|} M^{G'} + a_{E'}^{\ddagger E} N^{\ddagger E'} + \frac{1}{2} A_{E'}^{\ddagger E} M^{\ddagger E'}))^{1/2}} \ .$$

The equivalent Lagrangian is used for the actual fits.

Pilot Study—EEG Correlates to Behavioral States

In a 1991 paper, sets of EEG data were obtained from subjects while they were reacting to pattern-matching "odd-ball"-type tasks requiring varying states of selective attention taxing their short-term memory. Based on psychiatric and family-history evaluations, 49 subjects were classified into two groups, 25 possibly having high-risk and 24 possibly having low-risk genetic propensities to alcoholism.

Although M^G were permitted to roam throughout their physical ranges of $\pm N^E = \pm 80$ and $\pm N^I = \pm 30$ (in the nonvisual neocortex, true for all these regions), their observed effective regional-averaged firing states were observed to obey the centering mechanism. I.e., this numerical result is consistent with the assumption that the most likely firing states are centered about the region $M^G \approx 0 \approx M^{*E}$ in F^G.

Fitted parameters were used to calculate equivalent columnar firing states and time delays between regions. No statistical differences were observed between the total group, the high-risk group, and the low-risk group.

Lessons Learned

The previous study used data collected under the assumptions that:
- there is a genetic predisposition to alcoholism, and
- that this predisposition could be correlated to EEG activity.

These assumptions were negated by the SMNI study: E.g., there were no statistical differences in intra-electrode coherencies or in inter-electrode circuitry, or in any other parameter, between the two groups. Especially in light of other studies, it seems that if such a predisposition exists, it is a multifactorial issue that requires a very large subject population to resolve the many parameters, more than was available for this EEG study.

A later 1997 study including more details of stochastic distributions had more success.

CANONICAL MOMENTA INDICATORS (CMI) —EEG

Canonical Momenta Indicators (CMI)

Some 1996 papers illustrated how canonical momenta derived from fitted nonlinear stochastic processes, using ASA to fit models to S&P 500 data, can be useful indicators of nonequilibrium behavior of financial markets.

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

Training Phase

These techniques are quite generic, and can be applied to the SMNI model. In a 1997 paper, a given SMNI model is fit to EEG data, e.g., as performed in 1991. This develops a zeroth order guess for SMNI parameters for a given subject's training data. Next, ASA is used recursively to seek parameterized predictor rules, e.g., modeled according to guidelines used by clinicians. The parameterized predictor rules form an outer ASA shell, while regularly fine-tuning the SMNI inner-shell parameters within a moving window (one of the outer-shell parameters). The outer-shell cost function is defined as some measure of successful predictions of upcoming EEG events.

Testing Phase

In the testing phase, the outer-shell parameters fit in the training phase are used in out-of-sample data. Again, the process of regularly fine-tuning the inner-shell of SMNI parameters is used in this phase.

Utility

These momenta indicators should be considered as supplemental to other clinical indicators. This is how they are being used in financial trading systems. In the context of other invariant measures, the CMI transform covariantly under Riemannian transformations, but are more sensitive measures of neocortical activity than other invariants such as the energy density, effectively the square of the CMI, or the information which also effectively is in terms of the square of the CMI (essentially integrals over quantities proportional to the energy times a factor of an exponential including the energy as an argument). Neither the energy or the information give details of the components as do the CMI. EEG is measuring a quite oscillatory system and the relative signs of such activity are quite important.

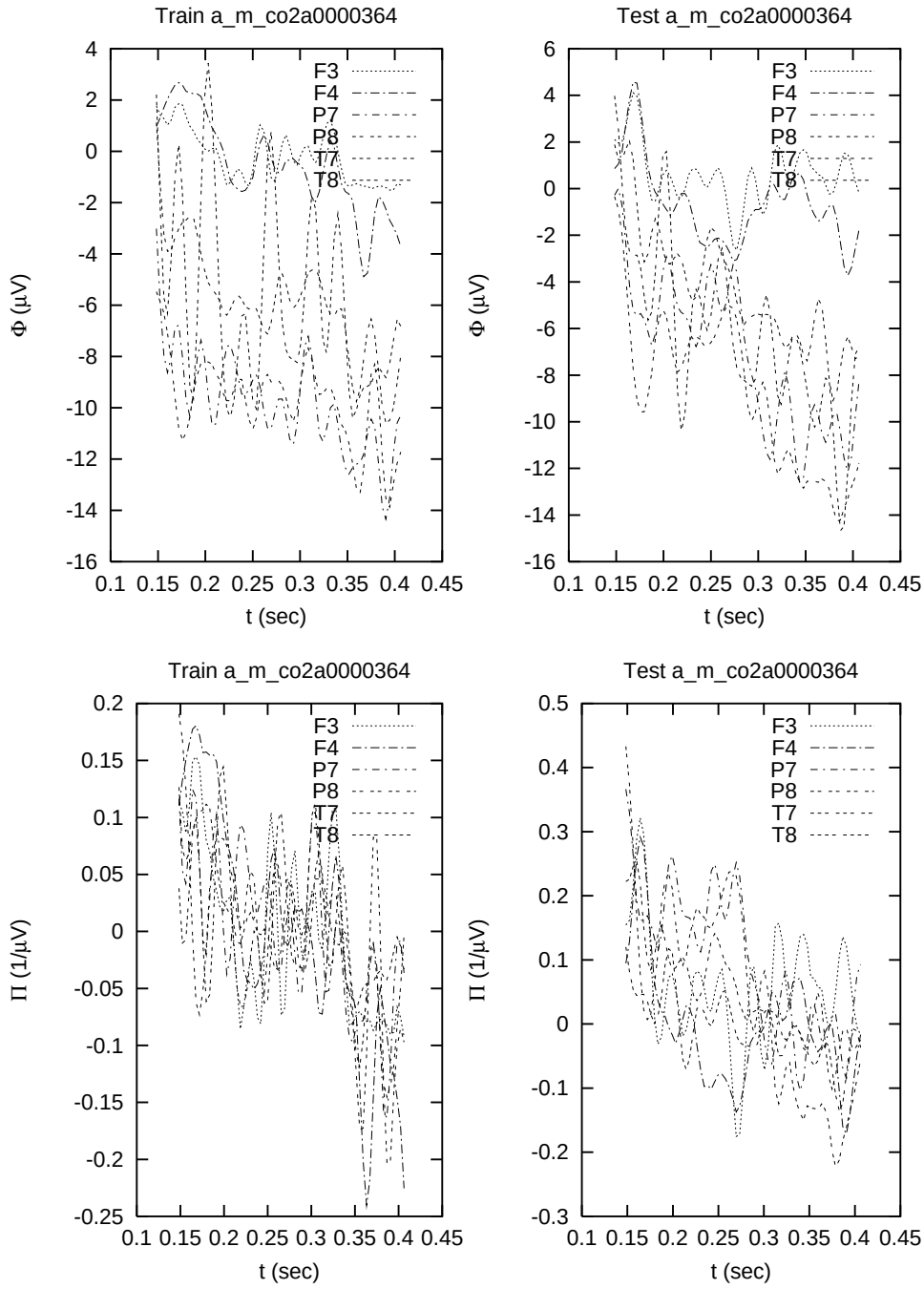
SMNI CMI of Genetic Predisposition to Alcoholism

Each set of results is presented with 6 figures, labeled as [{alcoholic|control}, {stimulus 1|match|no-match}, subject, {potential|momenta}], abbreviated to {a|c}_{1|m|n}_subject.{pot|mom} where match or no-match was performed for stimulus 2 after 3.2 sec of a presentation of stimulus 1. Data includes 10 trials of 69 epochs each between 150 and 400 msec after presentation. For each subjects run, after fitting 28 parameters with ASA, epoch by epoch averages are developed of the raw data and of the multivariate SMNI canonical momenta. There are fits and CMI calculations using data sets from 10 control and 10 alcoholic subjects for each of the 3 paradigms. For some subjects there also are out-of-sample CMI calculations. All stimuli were presented for 300 msec. Note that the subject number also includes the {alcoholic|control} tag, but this tag was added just to aid sorting of files (as there are contribution from co2 and co3 subjects). Each figure contains graphs superimposed for 6 electrode sites (out of 64 in the data) which have been modeled by SMNI using the circuitry:

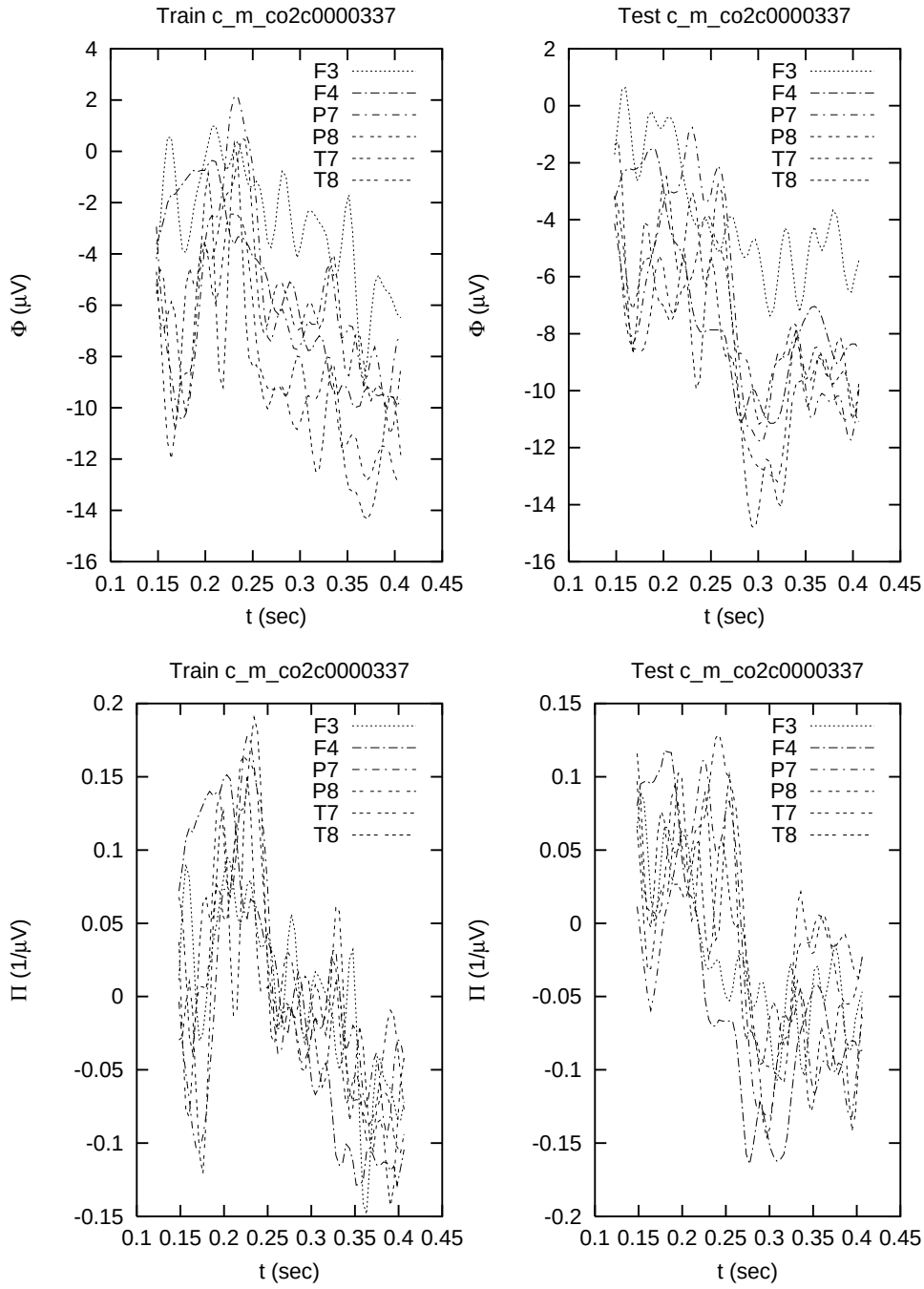
Site	Contributions From	Time Delays (3.906 msec)
F3		
F4		
T7	F3	1
T7	T8	1
T8	F4	1
T8	T7	1
P7	T7	1
P7	P8	1
P7	F3	2
P8	T8	1
P8	P7	1
P8	F4	2

The SMNI CMI give more signal to noise presentation than the raw data, especially for the significant matching tasks between the control and the alcoholic groups. The CMI can be processed further as is the raw data, and also used to calculate "energy" and "information/entropy" densities.

Data vs SMNI CMI for Alcoholic Group —S2 Match



Data vs SMNI CMI for Control Group —S2 Match



CHAOS IN EEG?

What if EEG has chaotic mechanisms that overshadow the above stochastic considerations? The real issue is whether the scatter in data can be distinguished between being due to noise or chaos.

The SMNI-derived probability distributions can be used to help determine if chaos is a viable mechanism in EEG. The probability distribution itself is a mathematical measure to which tests can be applied to determine the existence of other nonlinear mechanisms.

The path integral has been used to compare long-time correlations in data to predictions of models, while calculating their sensitivity, e.g., of second moments, to initial conditions. This also helps to compare alternative models, previously having their short-time probability distributions fit to data, with respect to their predictive power over long time scales.

Similar to serious work undertaken in several fields, the impulse to identify ''chaos'' in a complex system often has been premature. It is not supported by the facts, tentative as they are because of sparse data. Similar caution should be exercised regarding neocortical interactions.

Duffing EEG Analog —Chaos in Noise

A study of chaos in a model of EEG was cast into a Duffi ng analog.

$$\ddot{x} = f(x,t) \ ,$$

$$f = -\alpha \dot{x} - \omega_0^2 x + B\cos t \ .$$

This can be recast as

$$\dot{x} = y \ ,$$

$$\dot{y} = f(x,t) \ ,$$

$$f = -\alpha y - \omega_0^2 x + B\cos t \ .$$

Note that this is equivalent to a 3-dimensional autonomous set of equations, e.g., replacing $\cos t$ by $\cos z$, defi ning $\dot{z} = \beta$, and taking β to be some constant.

We studied a model embedding this deterministic Duffi ng system in moderate noise, e.g., as exists in such models as SMNI. Independent Gaussian-Markovian ("white") noise is added to both $\dot{x}$ and $\dot{y}$, η_i^j, where the variables are represented by $i = \{x, y\}$ and the noise terms are represented by $j = \{1, 2\}$,

$$\dot{x} = y + \hat{g}_x^1 \eta_1 \ ,$$

$$\dot{y} = f(x,t) + \hat{g}_y^2 \eta_2 \ ,$$

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

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

In this study, we take moderate noise and simply set $\hat{g}_i^j = 1.0 \delta_i^j$.

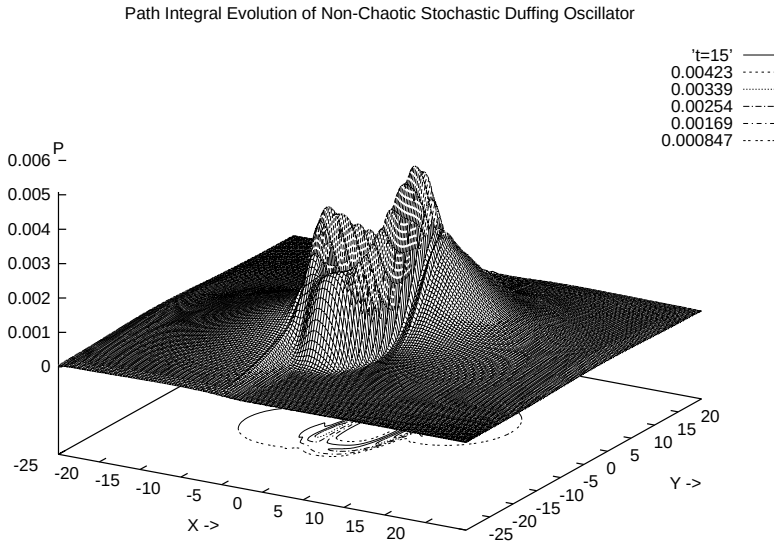
The equivalent short-time conditional probability distribution P, in terms of its Lagrangian L, corresponding to these Langevin rate-equations is

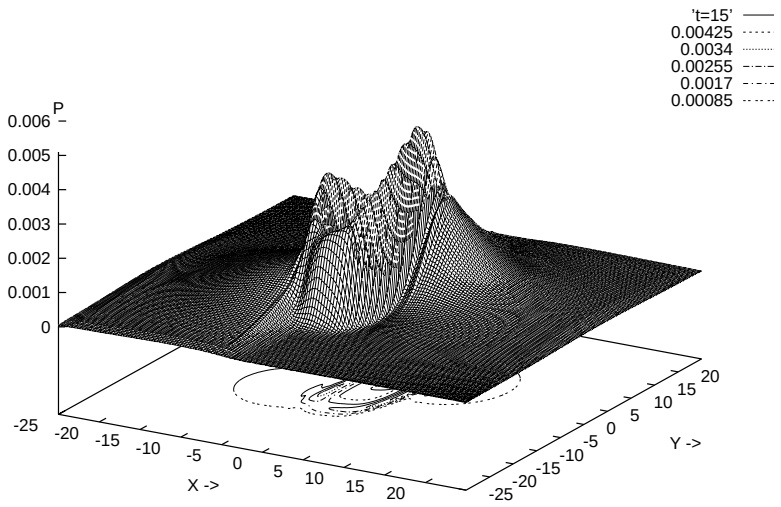
$$P[x, y; t + \Delta t | x, y, t] = \frac{1}{(2\pi\Delta t)(\hat{g}^{11}\hat{g}^{22})} \exp(-L\Delta t) \ ,$$

$$L = \frac{(\dot{x} - y)^2}{2(\hat{g}^{11})} + \frac{(\dot{y} - f)^2}{2(\hat{g}^{22})} \ .$$

Duffing EEG Analog —Preliminary Indications

No differences were seen in the stochastic system, comparing regions of Duffing parameters that give rise to chaotic and non-chaotic solutions. More calculations must be performed for longer durations to draw more definitive conclusions.





SMNI CORRELATES OF REACTION TIMES

Hick's Law

SMNI has given detailed descriptions of short-term memory (STM) phenomena and some aspects of evoked potential EEG.

A natural extension of these applications is to examine the relationship between STM and reaction time (RT). Hick's Law (observation) states

$$RT \propto n \ln n$$

where n is the number of items present in STM.

Given the audit trail back to averaged neuronal variables and that SMNI affords a unified description of STM, EEG and RT, there is motivation to pursue RT as a non-invasive diagnostic tool.

Time of First Passage Estimate of RT

The RT necessary to "visit" the states under control during the span of STM can be calculated as the mean time of "first passage" between multiple states of this distribution, in terms of the probability P as an outer integral $\int dt$ (sum) over refraction times of synaptic interactions during STM time t, and an inner integral $\int dM$ (sum) taken over the mesocolumnar firing states M,

$$RT = R - \int dt\ t \int dM\ \frac{dP}{dt}\ ,$$

where R is the time for preprocessing stimuli, on the order of tenths of a second.

Within tenths of a second, the conditional probability of visiting one state from another P, can be well approximated by a short-time probability distribution expressed in terms of the Lagrangian L as

$$P = \frac{1}{\sqrt{2\pi dtg}} \exp(-Ldt)\ ,$$

where g is the determinant of the covariance matrix of the distribution P in the space of columnar firings.

Calculation of Hick's Law

This expression for RT can be approximately rewritten as

$$RT \approx R + K \int dt \int dM \; P \ln P \; ,$$

where K is a constant when the Lagrangian is approximately constant over the time scales observed. Since the peaks of the most likely M states of P are to a very good approximation well-separated Gaussian peaks, these states by be treated as independent entities under the integral. This last expression is essentially the "information" content weighted by the time during which processing of information is observed. The calculation of the heights of peaks corresponding to most likely states includes the combinatoric factors of their possible columnar manifestations as well as the dynamics of synaptic and columnar interactions. In the approximation that we only consider the combinatorics of items of STM as contributing to most likely states measured by P, i.e., that P measures the frequency of occurrences of all possible combinations of these items, we obtain Hick's Law, the observed linear relationship of RT versus STM information storage.

For example, when the bits of information are measured by the probability P being the frequency of accessing a given number of items in STM, the bits of information in 2, 4 and 8 states are given as approximately multiples of $\ln 2$ of items, i.e., $\ln 2$, $2 \ln 2$ and $3 \ln 2$, resp. (The limit of taking the logarithm of all combinations of independent items yields a constant times the sum over $p_i \ln p_i$, where p_i is the frequency of occurrence of item i.)

SMNI FEATURES

Increasing Signal to Noise/Audit Trail to Sources

Logical and Testable Development Across Multiple Scales

SMNI is a logical, nonlinear, stochastic development of aggregating neuronal and synaptic interactions to larger and larger scales. Paradigms and metaphors from other disciplines do not substitute for logical SMNI development.

Validity Across Multiple Scales

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

Use of ASA and PATHINT on Nonlinear Stochastic Systems

ASA enables the fi tting of quite arbitrary nonlinear stochastic models to such data as presented by EEG systems. Once fi tted, PATHINT, or a newer algorithm PATHTREE, can evolve the system, testing long-time correlations between the model(s) and the data, as well as serving to predict events.

Inclusion of Short-Range and Long-Range Interactions

SMNI proposes that models to be fi tted to data include models of activity under each electrode, e.g., due to short-ranged neuronal fi bers, as well as models of activity across electrodes, e.g., due to long-ranged fi bers.

Riemannian Invariants

Yet to explore are the ramifi cations 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/should be used to explore fbws of information between neocortical regions.

Renormalization of Attenuated Frequencies

The SMNI approach shows how to "renormalize" the spatial activity to get a model that more closely matches the experimental situation, wherein there is attenuation of ranges of wave numbers.

MNN Real-Time Processing and Audit Trail to Finer Scales

The MNN parallel algorithm may offer real-time processing of nonlinear modeling and fi tting of EEG data for clinical use. Regional EEG data can be interpreted as mechanisms occurring at the minicolumnar scales.

Recursive ASA Optimization of Momenta Indicators + Clinical Rules

Similar to codes developed for fi nancial systems, recursive ASA optimizations of inner-shell SMNI indicators and outer-shell clinical guides should improve predictions of and decisions on clinical observations.