Proposed mechanism for learning and memory erasure in a white-noise-driven sleeping cortex

Moira L. Steyn-Ross,[1,*] D. A. Steyn-Ross,[1] J. W. Sleigh,[2] M. T. Wilson,[1] and Lara C. Wilcocks[1]
[1]*Department of Physics and Electronic Engineering, Private Bag 3105, University of Waikato, Hamilton, New Zealand*
[2]*Department of Anaesthetics, Waikato Hospital, Hamilton, New Zealand*
(Received 9 June 2005; revised manuscript received 12 October 2005; published 16 December 2005)

Understanding the structure and purpose of sleep remains one of the grand challenges of neurobiology. Here we use a mean-field linearized theory of the sleeping cortex to derive statistics for synaptic learning and memory erasure. The growth in correlated low-frequency high-amplitude voltage fluctuations during slow-wave sleep (SWS) is characterized by a probability density function that becomes broader and shallower as the transition into rapid-eye-movement (REM) sleep is approached. At transition, the Shannon information entropy of the fluctuations is maximized. If we assume Hebbian-learning rules apply to the cortex, then its correlated response to white-noise stimulation during SWS provides a natural mechanism for a synaptic weight change that will tend to shut down reverberant neural activity. In contrast, during REM sleep the weights will evolve in a direction that encourages excitatory activity. These entropy and weight-change predictions lead us to identify the final portion of deep SWS that occurs immediately prior to transition into REM sleep as a time of enhanced erasure of labile memory. We draw a link between the sleeping cortex and Landauer's dissipation theorem for irreversible computing [R. Landauer, IBM J. Res. Devel. **5**, 183 (1961)], arguing that because information erasure is an irreversible computation, there is an inherent entropy cost as the cortex transits from SWS into REM sleep.

DOI: 10.1103/PhysRevE.72.061910 PACS number(s): 87.19.La, 05.10.Gg, 05.70.Fh

I. INTRODUCTION

In a recent paper the authors have presented a simplified model for the electrical behaviour of a sleeping cortex [1]. Our sleep model uses physiologically realistic parameters and is a generalization of earlier work that presented a theoretical description for the state of forced quiescence brought about by administration of a general anaesthetic agent [2–6]. The simplifying assumptions common to both models are (i) the cortex is never far from a homogeneous equilibrium state; (ii) the cortex is constantly bombarded by a low-level background of white noise; and (iii) the scalp-measured EEG (electroencephalogram) or on-cortex ECoG (electrocorticogram) signal is proportional to the net soma voltage obtained by averaging across the spatial extent of the population of excitatory neurons within a cortical "macrocolumn"—a small volume (~ 1 mm^3) of cortical tissue containing about 85 000 excitatory and 15 000 inhibitory neurons. Thus this is a "mean-field" model, since it ignores fine spatial details of cortical structure. The final assumption is (iv) that although action potentials ("spikes") are a defining nonlinearity for individual neuron communication, these momentary events can be ignored when modelling at the population level, and thus it is sufficient to model the effect on overall neural activity in terms of an average firing *rate*. Therefore any synaptic modification that depends on the arrival times of individual spikes—such as spike-timing plasticity—is beyond the scope of our population-based model.

The presumption of near-equilibrium stochasticity means that we are deliberately focusing on small random fluctuations about steady state for which linearized theory can be expected to be valid. However, if the cortex develops an instability such that small fluctuations grow into large-scale nonlinear behaviors (e.g., a limit cycle), or if the cortex is exposed to a specific (i.e., nonrandom) stimulus, then the response is beyond the region of applicability of our model. Despite these rather strong restrictions, we have been able to demonstrate in Ref. [1] that linearized fluctuation theory is able to make several testable predictions regarding changes in EEG characteristics as the sleeper cycles from slow-wave sleep into REM sleep: (1) There will be a strong surge in electrical activity as REM state is approached; (2) this surge will show a pronounced redistribution of spectral energy towards zero frequency; (3) the voltage fluctuations will become increasingly correlated in time; (4) these changes are abruptly reversed in REM: power drops away as the spectrum becomes flatter and the correlations disappear. All four of these EEG-change predictions appear to be consistent with laboratory recordings of sleeping cat reported by Destexhe *et al.* [7].

In this paper we investigate the noise-driven learning and memory implications of this abrupt transition from SWS into REM sleep. Using standard Ornstein-Uhlenbeck [8] theory, we compute the stationary covariance matrix $\mathbf{\Sigma}(h_e,h_i)$ for the covarying excitatory and inhibitory voltage fluctuations h_e,h_i, and write down expressions for $P(h_e,h_i)$, the bivariate stationary PDF (probability distribution function), and for H, the Shannon entropy of these zero-mean fluctuations. If Hebbian [9] learning (synaptic reinforcement) occurs during sleep, then the synaptic weight changes within the cell assembly will tend to align along the direction that maximizes fluctuation variability (i.e., parallel to the first principal component of the $\mathbf{\Sigma}$ covariance matrix). As the cortex follows its sleep-cycling trajectory, the pattern of learning will be deter-

*Electronic address: msr@waikato.ac.nz; URL: phys.waikato.ac.nz/cortex

mined by the altering character of the covariance statistics: when the excitatory and inhibitory fluctuations are strongly correlated, the PDF broadens, causing Shannon entropy to be maximized. By linking Hebb's cell assembly with Hopfield's associative memory network, we propose that these correlated fluctuations in deep SWS provide a prescription for the weakening and erasure of potentially labile parts of memory. Conversely, when these cross-population fluctuations are weakly correlated—as is the case in REM sleep—the PDF becomes strongly peaked, entropy is minimized, and memory is expected to be strengthened.

We then relate these ideas to Landauer's dissipation theorem for irreversible computation, arguing that the transition to REM sleep is an irreversible erasure process that proceeds via a gradual state-space expansion and entropy increase in SWS, followed by an abrupt state-space compression on entry into REM sleep.

II. THEORY

A. A continuum model for the noise-driven cortex

Instead of considering the detailed interactions of discrete nerve cells, we choose to model at the level of the macrocolumn, a small volume of cortex containing $\sim 10^5$ excitatory and inhibitory neurons. We then presume that, to first approximation, the cortex can be pictured as a homogeneous collection of interconnected macrocolumns. For the purposes of modelling the cycles of natural sleep, we argue that this mean-field or population-based approach can be justified on two grounds. First, the gross states of vigilence—wakefulness, slow-wave sleep, REM sleep, and even anaesthetic unconsciousness—can be distinguished using a *single* EEG scalp electrode, so detailed spatial knowledge is not essential. Second, these gross brain states are not properties of individual neurons: rather, these states emerge as cooperative behaviours of large populations of neurons.

This continuum approach has a rich history in neurophysiological modelling [10–18]. Following the recent models of Wright [19], Liley [20,21], Robinson [22,23], and Rennie [24] we have developed a set of equations to describe the time-evolution of the mean excitatory and inhibitory soma voltages V_e, V_i, subject to excitatory and inhibitory synaptic inputs whose activity levels across the sleep cycle are governed by the varying concentrations of particular neuromodulators and "somnogens" (fatigue agents). The full development of these equations is given in Ref. [1]. In that paper we derived a simplified "adiabatic" equation set which assumes that, relative to the voltage changes that accumulate at the soma, synaptic-input events are fast and rapidly equilibrating. This fast-synapse, slow-neuron approximation appears to give valid descriptions of the major changes in EEG recorded during induction of anaesthesia and during the cycles of natural sleep.

Assuming that the cortex is subject to a continuous flux of low-level white-noise stimulation (originating from subbrain structures that are active during sleep), then in the slow-neuron adiabatic limit the excitatory and inhibitory soma voltages obey a pair of coupled stochastic differential (Langevin) equations,

$$\frac{d}{dt}\begin{bmatrix} V_e \\ V_i \end{bmatrix} = \begin{bmatrix} A_e(V_e, V_i) \\ A_i(V_e, V_i) \end{bmatrix} + \begin{bmatrix} \Gamma_e(t) \\ \Gamma_i(t) \end{bmatrix}, \tag{2.1}$$

where the $A_{e,i}$ are *drift* terms defined by

$$A_e = [V_e^{\text{rest}} - V_e + g_{ee} \cdot (V_e^{\text{rev}} - V_e) + g_{ie} \cdot (V_i^{\text{rev}} - V_e)]/\tau_e, \tag{2.2a}$$

$$A_i = [V_i^{\text{rest}} - V_i + g_{ei} \cdot (V_e^{\text{rev}} - V_i) + g_{ii} \cdot (V_i^{\text{rev}} - V_i)]/\tau_i. \tag{2.2b}$$

$V_{e,i}^{\text{rest}}$ is the cell resting voltage for the e,i population, and $V_{e,i}^{\text{rev}}$ is the synaptic reversal potential: here we take $V_e^{\text{rev}}=0$ mV for AMPA (excitatory) receptors, and $V_i^{\text{rev}}=-70$ mV for GABA (inhibitory) receptors; $\tau_{e,i}$ is the membrane time-constant. The four g_{jk} coefficients can be thought of as dimensionless synaptic conductances, and these depend on both V_e and V_i. Their double-subscripting implies a left-to-right "flow of action," thus g_{ie} is to be read $g_{i\to e}$, and indicates the total strength of synaptic flux being transmitted *from* inhibitory cells *to* excitatory cells. Each synaptic-source term in (2.2) contributes a voltage perturbation that depends on the product of the conductance g_{jk} with the deviation of the cell voltage V_k from the relevant reversal potential V_j^{rev}:

$$g_{ek} \cdot (V_e^{\text{rev}} - V_k) \equiv \rho_e (N_{ek} Q_e(V_e) + \phi_{ek}^{\text{sc}}) \left(\frac{V_e^{\text{rev}} - V_k}{V_e^{\text{rev}} - V_k^{\text{rest}}} \right), \tag{2.3a}$$

$$g_{ik} \cdot (V_i^{\text{rev}} - V_k) \equiv \rho_i (N_{ik} Q_i(V_i) + \phi_{ik}^{\text{sc}}) \left(\frac{V_i^{\text{rev}} - V_k}{V_i^{\text{rev}} - V_k^{\text{rest}}} \right), \tag{2.3b}$$

where $\rho_{e,i}$ is the synaptic strength [units: mV·s], being a signed measure of the charge transferred across the synapse per incoming action potential: $\rho_e > 0$ (excitatory event), $\rho_i < 0$ (inhibitory event); $N_{ek,ik}$ is the number of incoming $e \to k$, $i \to k$ synaptic connections per cell; $\phi_{ek,ik}^{\text{sc}}$ is the rate of spike flux entering from subcortical sources; and $Q_{e,i}$ is a sigmoidal function that maps a cell's membrane voltage to its output firing rate.

The $\Gamma_{e,i}$ *diffusion* terms of (2.1) are Gaussian-distributed white-noise sources entering the e,i macrocolumn populations from the subcortex,

$$\Gamma_e = b_{ee}\xi_1(t) + b_{ie}\xi_2(t), \tag{2.4a}$$

$$\Gamma_i = b_{ei}\xi_3(t) + b_{ii}\xi_4(t), \tag{2.4b}$$

where the four $\xi_m(t)$ stochastic terms are independent, zero-mean, delta-correlated white-noise sources:

$$\langle \xi_m(t) \rangle = 0, \tag{2.5}$$

$$\langle \xi_m(t)\xi_n(t') \rangle = \delta_{mn}\delta(t - t'), \tag{2.6}$$

and the b_{jk} white-noise scale factors are given by

$$b_{ek} = \rho_e \left(\frac{V_e^{\text{rev}} - V_k}{V_e^{\text{rev}} - V_k^{\text{rest}}} \right) \sqrt{\phi_{ek}^{\text{sc}}}/\tau_k, \tag{2.7a}$$

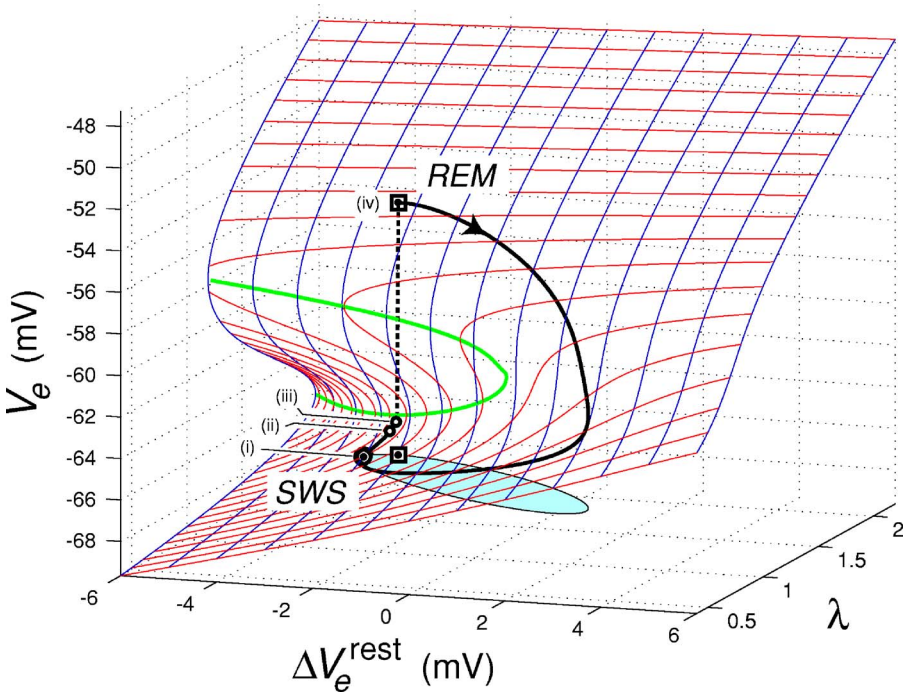


FIG. 1. (Color online) Distribution of steady-state cortical voltages over the sleep domain. The vertical axis is the equilibrium excitatory soma voltage V_e; the sleep domain is defined by a "fatigue" axis ΔV_e^{rest} and a "synaptic efficiency" axis λ. There are three steady-state values for V_e within the green ⊃-shaped region adjoining the left edge, and single steady-states elsewhere. We picture the 90-min sleep cycle as a clockwise elliptical tour of the ΔV_e^{rest}-λ plane, tracing out the arrowed trajectory on the equilibrium manifold. The tour commences in SWS at the point marked (i) (⊙), moves into deeper sleep [(i)→(ii)→(iii)], then makes an abrupt jump transition (dashed vertical line) from SWS into REM at (iv) (⊡).

$$b_{ik} = \rho_i \left(\frac{V_i^{\text{rev}} - V_k}{V_i^{\text{rev}} - V_k^{\text{rest}}} \right) \sqrt{\phi_{ik}^{\text{sc}}} / \tau_k. \tag{2.7b}$$

B. A model for sleep

To bring sleep to this slow-neuron, stochastic model, we incorporate two neuromodulation effects that operate on time scales that are vastly slower than the membrane time constants: (a) *adenosine fatigue* and (b) *acetylcholine activation*. (a) During waking hours, adenosine, one of several somnogens or "fatigue agents," steadily accumulates in the cortex [25]. Then, during subsequent recovery sleep, adenosine concentrations slowly diminish. Adenosine acts to decrease cortical activity by increasing a potassium (K^+) leak current, lowering the neuron resting voltage V_e^{rest}, thus making the population less likely to fire. [We have focused on adenosine since it regarded as the archetypal somnogen controlling sleep homeostasis (self-regulation); however, its modulation of cell resting voltage could also represent the effect of a variety of other circadian sleep neuromodulators.] (b) Acetylcholine (ACh) is an excitatory neuromodulator that is abundant during REM sleep, and absent during other stages of sleep [25]. Thus we identify REM sleep with periods of high ACh concentration, and SWS with periods of low ACh. Acetylcholine activates the cortex by reducing the K^+ leak current, so raising V_e^{rest} and making the cell more excitable. But in addition to its adenosine-antagonizing excitatory action, ACh has the paradoxical effect of simultaneously *reducing* the amplitude of the excitatory postsynaptic potential [26]. In our model, this corresponds to a reduction in ρ_e, the excitatory synaptic strength. Taking both (a) and (b) into account, we apply the following adjustment to Eq. (2.2a):

$$V_e^{\text{rest}} \to V_e^{\text{rest}} + \Delta V_e^{\text{rest}} \tag{2.8}$$

and introduce a modulation of the excitatory synaptic gain appearing in the (2.3a) drift and (2.7a) diffusion equations:

$$\rho_e \to \lambda \rho_e, \tag{2.9}$$

where λ is a dimensionless synaptic scalefactor; an *increase* in λ corresponds to a *decrease* in ACh concentration. These mappings (2.8) and (2.9) define our sleep domain in terms of a "fatigue" axis ΔV_e^{rest} and a "synaptic-gain" axis λ. Because the neuron resting voltage is affected by both neurochemicals, the position on the ΔV_e^{rest} axis at any given time will depend on the concentrations of ACh *and* adenosine.

C. Equilibrium phase space for sleep trajectory

The equilibrium or reference states of the sleeping cortex are located by setting the time-derivatives and noise terms in Eqs. (2.1) to zero, then solving numerically to find the stationary membrane voltages $V_{e,i}^{\text{o}}$ as a function of the sleep-domain parameters ΔV_e^{rest} and λ. The resulting 3D manifold of steady states is displayed in Fig. 1. Although most of the manifold displays a single equilibrium value V_e^{o}, for $\Delta V_e^{\text{rest}} < +1$ mV there is an S-shaped "fold" that contains *three* steady states. A linear stability analysis predicts that within this multiroot region, the midbranch root will be *unstable* with respect to small perturbations, while the top and bottom branches will be stable.

Draped over the Fig. 1 manifold we impose an elliptical tour whose orientation and shape is designed to be qualitatively consistent with our picture of the cortical changes experienced by a sleeper who cycles from SWS to REM and back. The location marked (⊙) is at a hyperpolarized, low-firing position on the bottom branch, presumed to be SWS.

It is not known whether the switchover from SWS to REM sleep is controlled by the sub-brain—e.g., by activation of the REM-on cells in the pontine brainstem that release acetylcholine into the cortex [27], thereby raising the resting voltage—or is initiated by the cortex itself as adenosine levels drop and the resting voltage rises. Either mechanism

could cause the elliptical tour to encounter the bottom-branch turning point, forcing an immediate jump transition into the cortically-activated state, presumed to be REM, marked ($\boxdot$) on the high-firing top branch.

The elevated levels of acetylcholine in REM then lead to a simultaneous reduction in λ (i.e., reduction in synaptic efficiency) and increase in V_e^{rest}. At some point, ACh production ceases and the hyperpolarizing influence of the residual adenosine pushes the tour back into deepening stages of SWS, allowing the cycle to repeat. (Although our proposed tour forms a *closed* elliptical cycle, in fact we would expect the progressive decreases in adenosine to cause the tour to drift to the right, transforming the ellipse into a 2D "corkscrew." We ignore this complication here.)

D. Linearized fluctuation theory: PDF, stationary covariance, and entropy

Of primary interest is not so much the distribution of steady states (since population-average membrane potentials $V_{e,i}^{\text{o}}$ are not accessible via standard ac-coupled EEG electronics), but rather $h_{e,i}(t)$, the small ac (zero-mean) fluctuations *about* these steady states:

$$h_{e,i}(t) = V_{e,i}(t) - V_{e,i}^{\text{o}} \tag{2.10}$$

since sleep-stage-induced changes in the $h_{e,i}$ fluctuation statistics should be detectable in the scalp-measured EEG. Provided these fluctuations remain small, the Langevin equations (2.1) can be replaced by their linearized approximation: a two-variable *Ornstein-Uhlenbeck* (Brownian motion) process,

$$\frac{d}{dt}\begin{bmatrix} h_e \\ h_i \end{bmatrix} = -\mathbf{A}\begin{bmatrix} h_e \\ h_i \end{bmatrix} + \sqrt{\mathbf{D}}\begin{bmatrix} \xi_e(t) \\ \xi_i(t) \end{bmatrix}, \tag{2.11}$$

where $\mathbf{A}$ and $\mathbf{D}$ are time-independent 2×2 matrices that vary as we tour the sleep manifold. $\mathbf{A}$ is the *drift matrix* that determines the rate at which the cortex relaxes to equilibrium; and $\mathbf{D}$ is the *diffusion matrix* that scales ξ_e and ξ_i, a pair of white-noise sources that buffet the excitatory and inhibitory neural populations. $\mathbf{A}$ is the negative of the Jacobian matrix formed from the partial derivatives of the $A_{e,i}$ drift terms of Eq. (2.2) evaluated at equilibrium,

$$-A_{jk} = \left.\frac{\partial A_j}{\partial V_k}\right|_{\text{eq}} \tag{2.12}$$

and $\mathbf{D}$ is the diagonal matrix computed from the covariances of the $\Gamma_{e,i}$ white-noise sources,

$$\mathbf{D} = \begin{bmatrix} b_{ee}^2 + b_{ie}^2 & 0 \\ 0 & b_{ei}^2 + b_{ii}^2 \end{bmatrix}_{\text{eq}}. \tag{2.13}$$

As (2.11) is a two-variable Ornstein-Uhlenbeck process, we may immediately infer that the stationary probability density function (PDF) is a bivariate Gaussian [28,29]:

$$P(h_e,h_i) = N\exp\left[-\tfrac{1}{2}\mathbf{h}^{\text{T}}\mathbf{\Sigma}^{-1}\mathbf{h}\right], \tag{2.14}$$

where $\mathbf{h}=[h_e,h_i]^{\text{T}}$ is the two-element column vector of EEG voltage fluctuations, and N is a normalization factor that ensures $\int P(\mathbf{h})d\mathbf{h}=1$. The shape and orientation of the PDF is determined by the stationary covariance matrix $\mathbf{\Sigma}$:

$$\mathbf{\Sigma} \equiv \begin{bmatrix} \langle h_e^2\rangle & \langle h_e h_i\rangle \\ \langle h_i h_e\rangle & \langle h_i^2\rangle \end{bmatrix} \equiv \begin{bmatrix} \sigma_e^2 & \sigma_{ei} \\ \sigma_{ie} & \sigma_i^2 \end{bmatrix}. \tag{2.15}$$

Here, for example, $\langle h_e^2\rangle \equiv \sigma_e^2$ is the *variance* of the excitatory fluctuations, and $\langle h_e h_i\rangle \equiv \sigma_{ei}=\sigma_{ie}$ denotes the excitatory-inhibitory *covariance*. Following Gardiner [8], because our Ornstein-Uhlenbeck process is two-dimensional, we may write an *exact* expression for the $\mathbf{\Sigma}$ covariance matrix:

$$\mathbf{\Sigma} = \frac{\det(\mathbf{A})\mathbf{D} + [\mathbf{A} - \text{tr}(\mathbf{A})\mathbf{I}]\mathbf{D}[\mathbf{A} - \text{tr}(\mathbf{A})\mathbf{I}]^{\text{T}}}{2\,\text{tr}(\mathbf{A})\det(\mathbf{A})}, \tag{2.16}$$

where $\mathbf{I}$ is the 2×2 identity matrix; $\det(\cdot)$ and $\text{tr}(\cdot)$ are determinant and trace operators respectively.

The bivariate PDF of Eq. (2.14) has its maximum at the origin, $(h_e,h_i)=(0,0)$; this point corresponds to a given equilibrium cortical state $(V_e^{\text{o}},V_i^{\text{o}})$ on the Fig. 1 sleep manifold. The "flatness" of the PDF can be quantified via its Shannon entropy [30] H,

$$H \equiv -\int P(\mathbf{h})\ln(P(\mathbf{h}))d\mathbf{h} = \ln(2\pi e\sqrt{\det\mathbf{\Sigma}}). \tag{2.17}$$

This entropy is a scalar measure of the range of excitatory and inhibitory fluctuation amplitudes available to the noise-driven cortex at a given cortical state. The entropy tracks the profound changes in the PDF distribution and shape that occur in the vicinity of the SWS-to-REM sleep transition.

III. THEORETICAL PREDICTIONS AND NUMERICAL RESULTS

Figure 2 shows the linearized predictions from Eqs. (2.15)–(2.17) for changes in fluctuation statistics over a complete tour around the elliptical sleep trajectory shown in Fig. 1. The transition into REM sleep occurs at cycle ≈ 0.097. We are particularly interested in the abrupt change in behavior as the cortex crosses from SWS into REM sleep, so we have selected three points within SWS [points marked (i), (ii), (iii) on the cycle axis], together with a single point in REM sleep (at cycle=0.1), for closer examination in Figs. 3-5.

As we move (i)$\rightarrow$(ii)$\rightarrow$(iii) through deepening SWS towards the transition point into REM sleep, we find that the power of both the excitatory [σ_e^2: Fig. 2(a)] and inhibitory [σ_i^2: Fig. 2(b)] voltage fluctuations are predicted to increase dramatically, particularly in the immediate vicinity of transition. This is matched by a corresponding increase in the cross-fluctuation power, σ_{ei} [Fig. 2(c)]. From its definition in (2.15), we see that σ_{ei} is a signed quantity. The fact that σ_{ei} is positive here means that the h_e and h_i fluctuations are *positively correlated*, so the excitatory and inhibitory fluctuations will tend to be in phase, with the degree of correlation increasing as the transition point is approached. We can quantify this loss of independence by computing the correlation coefficient r:

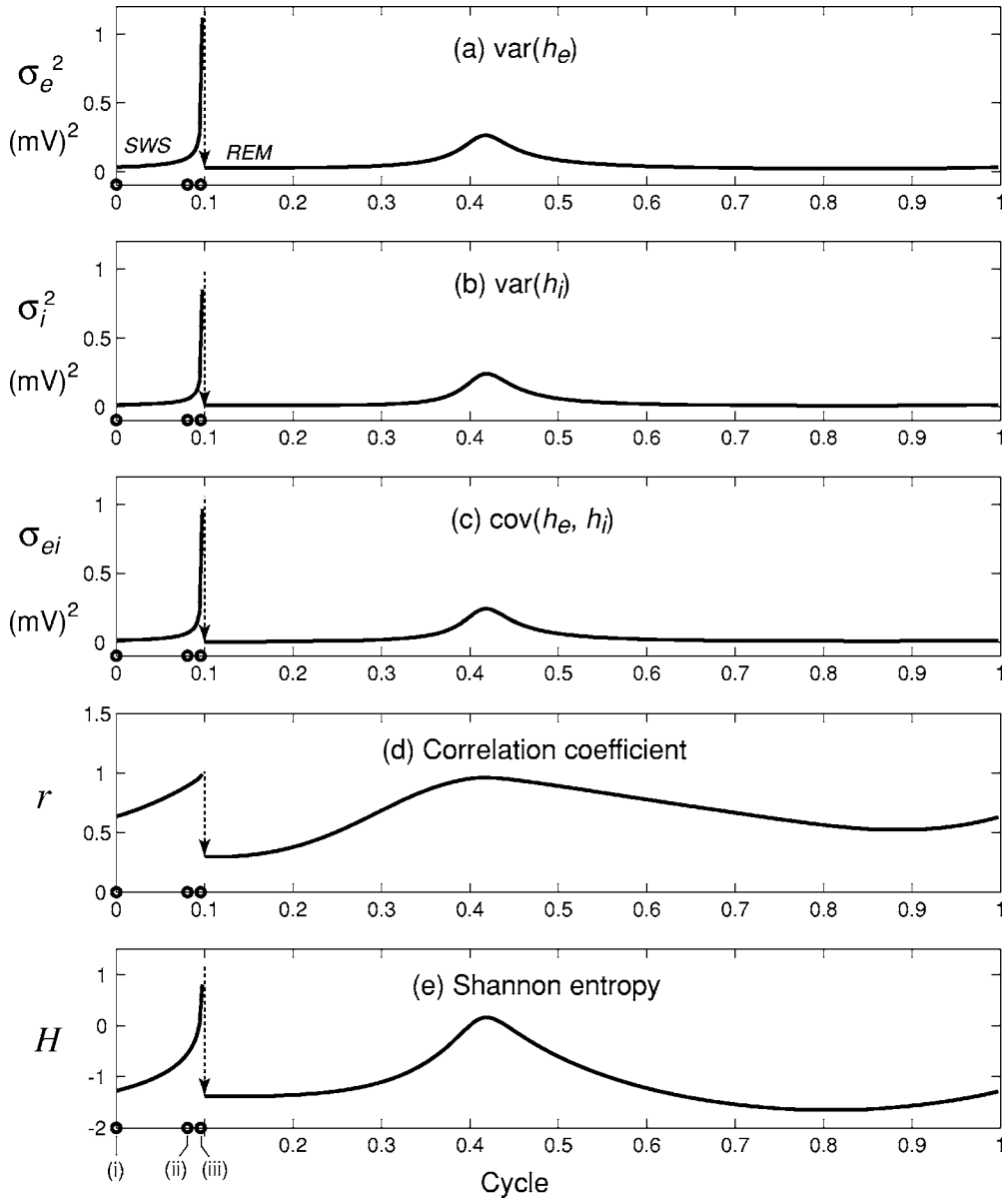


FIG. 2. Predicted changes in soma-voltage fluctuation statistics for the phase-transition model of the sleep cycle: (a) excitatory variance σ_e^2, (b) inhibitory variance σ_i^2, (c) excitatory-inhibitory covariance σ_{ei}, (d) excitatory-inhibitory correlation coefficient r, and (e) Shannon entropy H for the bivariate probability density function $P(h_e,h_i)$ (see Fig. 3). The transition into REM sleep at cycle ≈ 0.097 is heralded by a pronounced increase in the white-noise responsiveness of both the excitatory and inhibitory neural populations. Moreover, these fluctuation responses are predicted to become highly correlated in the vicinity of transition. (The three open circles on the left end of the "Cycle" axis mark the analysis points for the SWS graphs shown in Figs. 3–5.)

$$r \equiv \frac{\sigma_{ei}}{\sigma_e \sigma_i}. \tag{3.1}$$

Its variation with sleep phase is shown in Fig. 2(d). The significant feature is the prediction of *perfect correlation* ($r \to 1$) between the $h_e(t)$, $h_i(t)$ fluctuations immediately prior to the jump into REM sleep at cycle ≈ 0.097.

The origin for these surges in electrical activity near transition becomes clear when we examine the form of Eq. (2.16). At transition, the dominant eigenvalue of the drift matrix $\mathbf{A}$ goes to zero. Recalling that the trace of a matrix is equal to the product of its eigenvalues [31], it follows that the denominator on the RHS of Eq. (2.16) goes to zero while the numerator remains finite, and therefore all four elements of the covariance matrix $\mathbf{\Sigma}$ will diverge at transition. (The symmetric, smooth power surge seen as the macrocolumn leaves REM sleep at cycle ≈ 0.42 arises from the dominant eigenvalue approaching, but not crossing its zero axis; this secondary peak is not discussed further in this paper.)

Concommitant with the peaking of fluctuation and cross-fluctuation activity is the rise in Shannon entropy H [Fig. 2(d)]. Because H quantifies the size of the state-space available to the macrocolumn, it is clear that, in the period leading up to transition into REM sleep, there is an expansion in

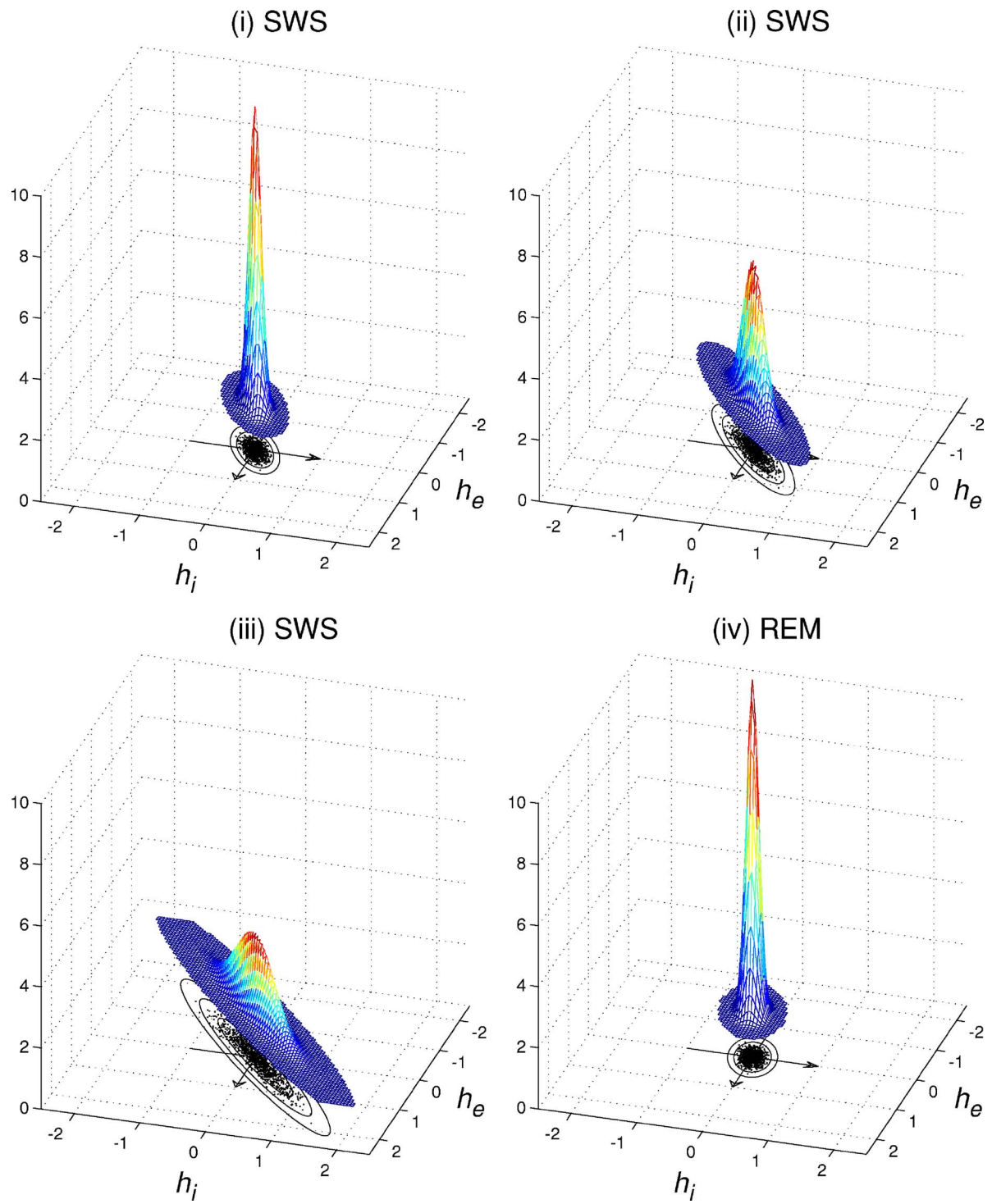


FIG. 3. (Color online) Predicted evolution of the bivariate Gaussian probability density function (PDF) for h_e and h_i voltage fluctations (in mV) during deepening SWS [(i), (ii), (iii)] and across the transition into REM sleep (iv). The correlation coefficients for the four PDFs shown are respectively (i) 0.64, (ii) 0.89, (iii) 0.96, and (iv) 0.30. The elliptical bull's-eye contour lines mark the boundaries for fluctuation excursions that are 1, 2, 3, and 4 standard deviations away from the (0, 0) equilibrium point. (Each PDF mesh has been moved upwards by 1.5 density units $[(\mathrm{mV})^{-2}]$ in order to better view the contours.)

state space that occurs at the same time as the growth of correlated activity in both the excitatory and inhibitory neural populations. After transition, all five measures plotted in Fig. 2 drop away. Fluctuations are now of smaller amplitude and are less correlated. This model suggests that in REM sleep, the excitatory and inhibitory fluctuations are more independent, and, compared with deep slow-wave sleep, occupy a much smaller state space.

Figure 3 shows the evolution of the fluctuations PDF [Eq. (2.14)] at the four points of interest that bracket the

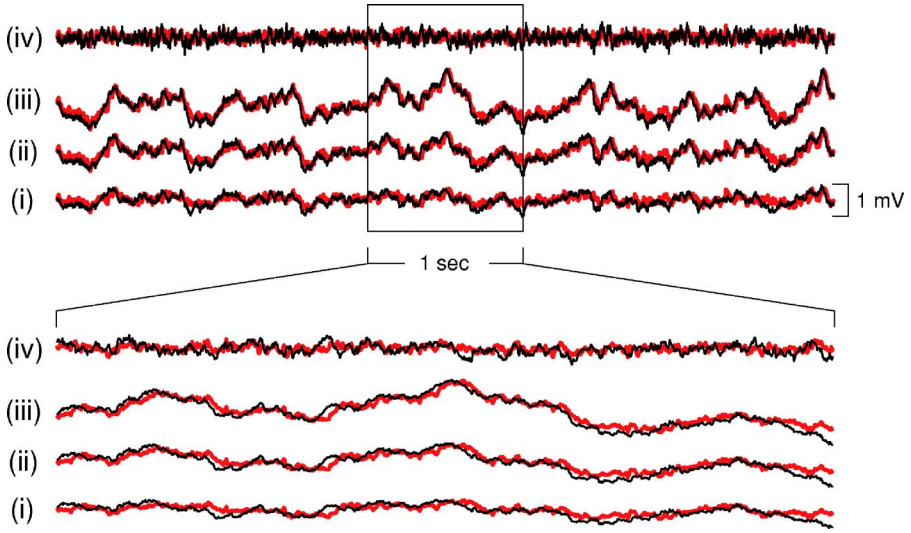


FIG. 4. (Color online) Numerical simulations of the Langevin equations (2.1) at three points within SWS [(i), (ii), (iii)] and one point in REM sleep (iv), each trace showing a 5-sec time series of excitatory (h_e: black) and inhibitory (h_i: red) neural activity. The *lower* set of traces is a zoomed view of the *upper boxed* 1-sec segment. To aid comparison of the altered dynamics, each simulation run was driven by an *identical* sequence of 20 000 random numbers [four white-noise generators $\xi_m(t)$ updated at intervals of $\Delta t=1$ ms for a simulation time of 5 s]. These nonlinear numerical experiments provide a cross-check against the linearized-theory predictions of Fig. 3. Note that the SWS approach to REM sleep [(i)→(ii)→(iii)] is marked by large-amplitude, slow-frequency correlated voltage fluctuations; these slow correlated fluctuations disappear in REM sleep (iv).

SWS→REM sleep transition. The increase in correlated activity during deepening SWS [(i), (ii), (iii)] squeezes the covariance contours so that they become distinctly elongated along the $h_e=h_i$ in-phase direction. The character of the PDF changes abruptly in REM (iv): the extreme contour elongations have disappeared, signalling that the h_e and h_i fluctuations are significantly less correlated. (See also Fig. 1.)

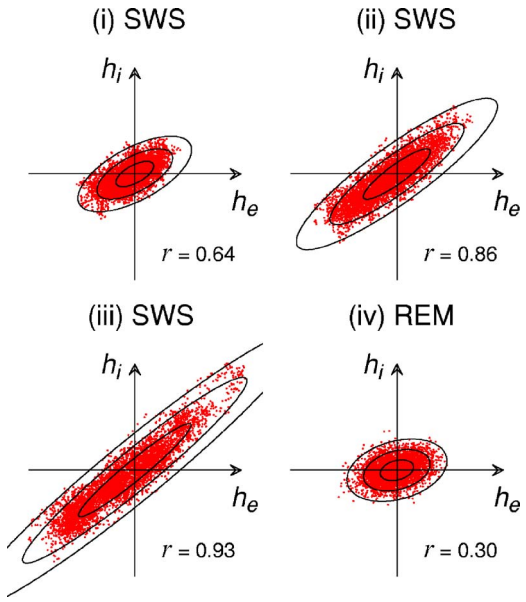


FIG. 5. (Color online) These h_e-vs-h_i 5000-point scatter plots show the point-cloud distribution for the *nonlinear* simulation voltage traces of Fig. 4. Superimposed are the 1-, 2-, and 3-σ elliptical countours predicted by the *linearized* fluctuation theory of Eq. (2.15). The axis arrows provide an amplitude scale, marking the interval −1 to +1 mV. The r values are the correlation coefficients computed from the simulation time series. The agreement between nonlinear simulation and linear-theory prediction (see caption of Fig. 3) is very satisfactory.

To test the accuracy of the linearised predictions, we ran a series of 5-sec numerical simulations of the *nonlinearized* sleep equations (2.1); the resulting time-series appear in Fig. 4. It is clear that the spectral content and temporal character of the theoretical EEG patterns are very different in SWS compared with REM sleep. The SWS [(i), (ii), (iii)] traces are dominated by large, slow voltage variations that show a high degree of phase similarity between the excitatory and inhibitory fluctuations, whereas in REM sleep (iv) the correlated low-frequency content has almost completely disappeared.

The scatter plots of Fig. 5 were derived from the nonlinear simulations shown in Fig. 4. Their (co)variances and correlation coefficients show excellent agreement with the theoretical predictions of Fig. 3, giving us confidence that our linearised theory does provide an accurate description of near-equilibrium fluctuation behavior of the nonlinear Langevin sleep equations.

Equation (2.4) shows that our cortical circuit is bombarded with random stimulation that enters excitatory and inibitory macrocolumn populations, from both excitatory and inhibitory subcortical sources, with all four connection possibilities permitted: $e\to e$, $e\to i$, $i\to e$, $i\to i$. An anonymous referee has pointed out that the vast majority of input pathways from subcortical structures will be of the $e\to e$ and $e\to i$ types, so we ran a series of stochastic simulations in which the number of white-noise sources was reduced from four to two, with Eq. (2.4) being modified to read

$$\Gamma_e = b_{ee}\xi_1(t), \tag{3.2a}$$

$$\Gamma_i = b_{ei}\xi_3(t). \tag{3.2b}$$

Although fluctuation amplitudes in the two-source case were lower (as expected, since there is less stimulation energy), we found no qualitative changes in the simulation behaviors. As for the four-source experiments reported in this paper, the two-source runs gave the expected critically-slowed fluctuation response: an increase in low-frequency power and correlation times on approach to the SWS→REM sleep transition, with abrupt reduction in fluctuation correlations and amplitude after transition.

Use of white-noise stimulation is mathematically convenient (since it allows us to compute exact statistics for the linearised fluctuations), but it is probably not a good representation for structured subcortical inputs that might arise from specific sensory stimulus. To simulate the effect of a structured stimulus we replaced the white-noise drives with correlated ("pink") noise generators, varying both the correlation time for each source, and the degree of cross-source correlation. In all cases, the resulting soma-voltage fluctuations were larger, slower, and more correlated than when white-noise driving was used, with the degree of correlated slowing depending on the closeness of the macrocolumn to the SWS/REM-sleep transition. Because the fluctuations were larger, there was a much greater probability of an "early" jump transition from SWS to REM sleep. We conclude from these structured-noise tests that the prediction of critically slowed fluctuations is a general property of the network, and is not an artifact of the choice of stimulus.

IV. SLOW FLUCTUATIONS AND MEMORY ERASURE

As the slow-wave sleeping cortex approaches transition into REM sleep, our macrocolumn model shows pronounced changes in the fluctuation statistics for the excitatory and inhibitory membrane voltages. We propose that these altered fluctuation characteristics cause systematic alterations in synaptic weights that can lead to the *unlearning* or *erasure* of memory. Our argument is based on Hebb's [9] philosophy for the imprinting (learning) of memories, and on the influential work of Hopfield [32] who was the first to present a mathematical basis for associative memory networks.

A. Synaptic plasticity

One of Hebb's most enduring postulates asserts that sustained local neural activity causes changes in the strength of the synaptic connections linking the participating neurons. Generalising this idea from the local to the *population* level, we can construct a hypothetical schematic for the synaptic weights that set the strength of connections between the excitatory E and inhibitory I populations of the macrocolumn. As illustrated in Fig. 6, we define four mean-field synaptic weights: w_{ee} and w_{ii} are self-action terms for the $e\to e$ and $i\to i$ interactions *within* the excitatory and inhibitory subpopulations respectively; w_{ei} and w_{ie} are $e\to i$ and $i\to e$ cross-weights that give the strength of coupling *between* the E and I populations. In our model, the cross weights are symmetric: $w_{ei}=w_{ie}$.

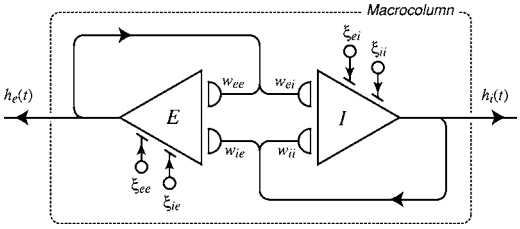


FIG. 6. Mean-field synaptic weights for the linearized macrocolumn. The w_{ee} and w_{ii} weights give the coupling strengths *within* the E and I neural populations; the cross weights $w_{ei}=w_{ie}$ determine the strength of coupling *between* the E and I populations. The four circles labeled ξ_{jk} represent four independent white-noise stimulation sources originating from the subcortex.

In the spirit of the mean-field approach of Bienenstock and Lehmann [33], we assume that the population weights obey a set of Hebbian covariance rules [34], but unlike those authors who work with *instantaneous* values for activity covariance, we write the learning rules in terms of the *steady-state* variances and covariances of the voltage fluctuations in h_e and h_i:

$$\Delta w_{ee} = \eta\langle h_e^2\rangle\Delta t, \quad \Delta w_{ii} = \eta\langle h_i^2\rangle\Delta t, \quad \Delta w_{ei,ie} = \eta\langle h_e h_i\rangle\Delta t. \tag{4.1}$$

Here η is a (small, positive) rate-constant assumed to be common to all four rules. We can write Eq. (4.1) as an update rule for the weights matrix $\mathbf{W}$,

$$\Delta\mathbf{W} \equiv \begin{bmatrix} \Delta w_{ee} & \Delta w_{ei} \\ \Delta w_{ie} & \Delta w_{ii} \end{bmatrix} = \eta \begin{bmatrix} \sigma_e^2 & \sigma_{ei} \\ \sigma_{ie} & \sigma_i^2 \end{bmatrix} \Delta t. \tag{4.2}$$

Thus the rate of change of the $\mathbf{W}$ weights matrix is proportional to $\mathbf{\Sigma}$, the Eq. (2.15) covariance matrix for small fluctuations. The assumption of a Hebbian covariance rule implies that the synaptic weights will evolve along the direction that maximises fluctuation variability, i.e., the weight change will align itself with the long axis (first principal component) [35] of the covariance ellipses illustrated in Figs. 3 and 5. As SWS deepens [(i)→(ii)→(iii)], the h_e and h_i fluctuations become more correlated, causing the point-cloud to elongate along the $h_e=h_i$ mirror axis: the excitatory-inhibitory correlation coefficient r tends to unity as the two neural populations increasingly behave as one.

B. Reverberating cell assemblies

Hebb [9] defined a reverberating cell assembly as a collection of neurons whose synaptic connections enable sustained neural activity even after the activating stimulus has been removed. That is, the stimulus lives on in the assembly as a reverberation. In our model we consider each stationary (equilibrium) state of the macrocolumn to be a *macrostate* that represents a huge number of equivalent, but unknowable, internal activity patterns or *microstates*. We define a *memory* to be any microstate (activity configuration) that, once stimulated, can persist (reverberate) for some time.

In Hopfield's [32] theory of associative networks, each memory is represented as a valley in an energy landscape;

these valleys act as attractors for the network, with deeper valleys representing stronger attractors and more easily recalled memories [36]. By design, our macrocolumn model has no knowledge of its detailed internal structure, so we cannot identify its memories. However, we can glean some knowledge of the overall strength of the family of memories held within a given macrostate by probing its responsiveness to random noise: if a small perturbation produces a small response, then the fluctuations PDF will be narrow and highly peaked [e.g., see Fig. 3(iv)], indicating that the population-average attractor is strong and well-localized about the equilibrium point. At the opposite extreme, if a small perturbation produces an exaggeratedly large fluctuation response [e.g., Fig. 3(iii)], the PDF will be broad, corresponding to a weakened attractor that is only diffusely located at equilibrium. Thus, in the Hopfield interpretation, the fluctuation response in deep SWS (i.e., close to the SWS→REM-sleep boundary) is consistent with weakened or erased memory.

From a Hebbian-plasticity viewpoint (see Sec. IV A) we expect that the coupling between excitatory and inhibitory populations will be strengthened by the increasingly correlated fluctuations exhibited during the approach to the SWS-to-REM sleep transition. The net effect of this tight cross-coupling will be to severely weaken network reverberations, since an excitatory event will stimulate prompt inhibition that will feed back to damp out the excitation. We conclude that, irrespective of whether we view the macrocolumn as a Hebbian cell assembly or a Hopfield associative network, deep SWS is a time of suppressed reverberation and weakened memory.

This situation is reversed following the abrupt transition into REM [see panel (iv) of Figs. 3–5]. The slow, large-amplitude, highly correlated fluctuations of deep SWS are replaced by excitatory and inhibitory variations that are faster, smaller, and less correlated (i.e., more independent), as indicated by covariance contours that are more circular. At this point in REM sleep we find that the σ_{ei} covariance is much smaller than the individual population variances ($\sigma_e^2 \approx 5.4\sigma_{ei}$, $\sigma_i^2 \approx 2.1\sigma_{ei}$). Applying Hebbian rule (4.2) here shows that the $w_{ei,ie}$ anti-reverberatory weight change will be swamped by the combined reverberatory effects of growth in w_{ee} and w_{ii}. (A w_{ii} increase implies that the inhibitory population will become more inhibited, so will exert a weaker moderating influence on the excitatory population.) We conclude that REM sleep is a time when reverberant memory is reinforced, but because the fluctuation variances are smaller, the *rate* of weight change during REM sleep will be considerably slower than for deep SWS.

V. DISCUSSION

In this paper we have described a model for memory erasure in terms of Hebbian-induced synaptic weight changes that will tend to suppress reverberation as the sleeping cortex approaches the SWS-to-REM sleep transition. We can glean a deeper insight into this transition by regarding the cortex as a biological computer that, in a sufficiently abstract sense, might be expected to conform to some of the general principles of computing and information processing. We now consider the entropy and energy implications of information erasure, as interpreted in a landmark paper by Landauer, then relate his findings to our sleeping cortex model.

In 1961 Landauer [37] proposed a general thermodynamic principle of computing, later elaborated by Bennett [38]. We will state his principle in two parts, and will refer to it as *Landauer's dissipation theorem* (LDT).

(1) Any computation step that is *logically irreversible* has an irreducible entropy cost associated with it.

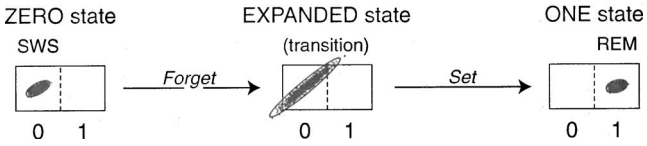
(2) When implemented in a physical device (a computer), logical irreversibility implies physical irreversibility, resulting in an irreducible energy dissipation proportional to the entropy change.

A reversible operation retains sufficient information in its outputs to enable reconstruction of its inputs. For example, NOT (logical negation) and EXCLUSIVE OR are logically reversible since each forms a one-to-one mapping from input to output that conserves the entropy of the bit pattern. In contrast, a logically irreversible operation discards information by forming a many-to-one mapping from input to output, making it is impossible to reconstruct unambiguously the input state from the output state. Thus, in a device containing N binary elements, the erasure operations SET TO ZERO and SET TO ONE are both many-to-one mappings that take 2^N possible input states and map these to a single bit pattern, resulting in an entropy reduction $\Delta S = k \log 2^N - k \log 2^0 = kN \log 2$. When implemented on a computer, LDT asserts that the erasure will cause a quantity of heat $\Delta Q = T\Delta S$ to be dissipated to the environment. (k is Boltzmann's constant; T is absolute temperature).

Landauer distinguished between two broad classes of bistable computing devices which we will characterize as *static* and *active*. Static devices, such as ferrites, ferroelectrics, and magnetic films, can hold information without dissipating energy. Active devices, such as relays, electronic flip-flops, and dynamic read-write memory, rely on continuous energy dissipation to maintain steady state. Landauer's theorem is presumed to apply to both classes. For the class of active bistable devices, LDT quantifies the *additional* dissipation caused by irreversible switching events. It is very clear that the biological system is not dissipationless, even at "steady state." First, each firing of an action potential releases $\sim 10^{11} kT$ units of energy [38]. Second, even the nonfiring hyperpolarized state requires a continuous expenditure of metabolic energy via ion pumps to maintain the ionic concentration gradients across the cell membrane. Thus the macrocolumn may be grouped with Landauer's class of *active* bistable devices.

Bennett [38] explored the connection between logical and thermodynamic irreversibility using a one-molecule gas whose location within its container, left or right, was measured by a device referred to as "Maxwell's demon." Bennett described the essential steps required in order for the demon to erase its knowledge of the location of the molecule: first, there must be an *expansion* (entropy increase) of the molecule's phase space as the molecule is allowed once again to occupy the full volume of its container; and second, there must be a *compression* (entropy decrease) of the occupied volume of the demon's phase space in order to restore the demon to its standard (zero memory) state.

In order to apply the ideas of Landauer and Bennett to the cortical computer, we will impose a crude binary discipline to the macrocolumn: let the ZERO state of the macrocolumn be the hyperpolarized (nonfiring) state, and the ONE state the depolarized (high-firing) state. If we identify slow-wave sleep as the ZERO state, and REM sleep as the ONE state, then we can regard the SWS→REM-sleep transition is an irreversible erasure operation that proceeds via a gradual state-space expansion ("forget") followed by an abrupt compression ("set"), as summarized in the following state-transition map:



[The three shaded elliptical regions are miniatures of the point-clouds and 3-σ contours for SWS(i), SWS(iii), and REM(iv) copied from Fig. 5.] In our sleep model, the erasure occurs during the expansion of the fluctuations phase space [e.g., (i)→(ii)→(iii) of Figs. 3 and 5], and is completed by the phase-space compression immediately following the jump into REM sleep [at (iv) of Figs. 3 and 5]. Biophysically, the "forgetting" is the suppression of reverberations resulting from Hebbian plasticity at the synapses: the growth in correlated activity between excitatory and inhibitory neural populations during the expansion phase will tend to increase the inhibitory synaptic weights. Computationally, the forgetting is represented by the increase in entropy as the macrocolumn explores larger regions of its phase space.

What is being erased or "forgotten" during slow-wave sleep? Crick and Mitchison [39] noted that associative networks can become overloaded when an attempt is made to store too many patterns, resulting in the subsequent recall of an inappropriate response to a given input stimulus. Having classified these undesirable interactions as unwanted or "parasitic" modes, Crick and Mitchison gave a three-step algorithm for their elimination: (1) the system (cortex) should be isolated from its major inputs; (2) it should be driven by random activations that excite any incipient parasitic modes; and (3) these spurious modes should be damped down via some (unspecified) "unlearning" mechanism. Crick and Mitchison conjectured that REM sleep might provide this essential parasitic-cleansing function. Our model suggests that SWS provides a much better substrate for memory erasure: the cortex is disconnected from structured stimulus; the subcortical white-noise driving will tend to excite any memory that is evoked by unstructured input; these memories will be suppressed by a Hebbian "unlearning" rule that strengthens naturally as SWS deepens and excitatory-inhibitory fluctuations become more highly correlated.

There is clinical evidence that slow correlated electrical activity in the cortex interferes with normal memory function. During the induction of general anaesthesia, scalp-measured EEG shows a diffuse increase in low-frequency (e.g., delta band: 1–3.5 Hz) fluctuations [40] that exhibit strong coherence between separated frontal electrodes [41]. Clinicians have reported that when patients are presented with verbal learning tasks during induction, they display a marked reduction in short term memory capacity, and this deficit is apparent for some minutes following recovery of consciousness [42,43]: a patient can maintain a lucid conversation at recovery, but have no memory of it 30 min later. Other amnesic states that are caused by disease (e.g., the postictal state following epileptic seizure) or by anticholinergic drugs (e.g., hyoscine) are almost invariably associated with enhanced delta-band activity in the EEG. Conversely, procholinergic drugs—which cause a shift to REM-like high-frequency uncorrelated EEG—are commonly used to *improve* memory function in patients with Alzheimer's disease.

We need to point out that our equilibrium theory for sleep ignores the fact that deep SWS has a dynamic structure [44] in which the cortex slowly cycles between hyperpolarized quiescence (the "down" state) and depolarized high-firing activity (the "up" state), with these activity cycles occurring on a time-scale of $\lesssim$1 s; this is Steriade's "slow oscillation." However, if each "down"-to-"up" transition within SWS can be thought of as a short-lived SWS-to-REM state change, then it is conceivable that the slow oscillation drives cyclic changes of synaptic weights, with each oscillation defining a miniature erasure-consolidation cycle.

We acknowledge that our working definition of "memory" is rather abstract, and that our model ignores the complex neurobiological processes of acquisition, consolidation, enhancement and recall. Although the role that sleep may (or may not) play in these processes remains a matter of active research and vigorous debate [45,46], it is clear that consolidation requires integration of a new memory within the context of an existing set—a process that is likely to involve erasure of the weaker, more labile components of the memory [47].

In conclusion, our model describes a mechanism by which the large, slow, correlated voltage fluctuations in the cerebral cortex seen during SWS can result in the loss of short-term labile memory. This prediction of memory erasure is consistent not only with clinical observations of the SWS-like state that follows anaesthesia or seizure (described above), but also with the profound short-term amnesia exhibited by subjects transiently awakened after 10 minutes of sleep [48]. The possible significance of SWS-induced erasure with respect to longer-term memories has yet to be established.

ACKNOWLEDGMENTS

This work has been supported by the New Zealand Marsden Fund, Contract No. UOW307.

[1] D. A. Steyn-Ross, M. L. Steyn-Ross, J. W. Sleigh, M. T. Wilson, I. P. Gillies, and J. J. Wright, J. Biol. Phys. **31**, 3 (2005).
[2] M. L. Steyn-Ross, D. A. Steyn-Ross, J. W. Sleigh, and D. T. J. Liley, Phys. Rev. E **60**, 7299 (1999).
[3] M. L. Steyn-Ross, D. A. Steyn-Ross, J. W. Sleigh, and L. C. Wilcocks, Phys. Rev. E **64**, 011917 (2001).
[4] D. A. Steyn-Ross, M. L. Steyn-Ross, L. C. Wilcocks, and J. W. Sleigh, Phys. Rev. E **64**, 011918 (2001).
[5] M. L. Steyn-Ross, D. A. Steyn-Ross, J. W. Sleigh, and D. R. Whiting, Phys. Rev. E **68**, 021902 (2003).
[6] M. L. Steyn-Ross, D. A. Steyn-Ross, and J. W. Sleigh, Prog. Biophys. Mol. Biol. **85**, 369 (2004).
[7] A. Destexhe, D. Contreras, and M. Steriade, J. Neurosci. **19**, 4595 (1999).
[8] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, 3rd ed., Springer Series in Synergetics, Vol. 13 (Springer-Verlag, New York, 2004).
[9] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
[10] R. L. Beurle, Philos. Trans. R. Soc. London, Ser. A **240**, 55 (1956).
[11] J. S. Griffith, Biophys. J. **3**, 299 (1963).
[12] H. R. Wilson and J. D. Cowan, Biophys. J. **12**, 1 (1972).
[13] H. R. Wilson and J. D. Cowan, Kybernetik **13**, 55 (1973).
[14] W. J. Freeman, *Mass Action in the Nervous System* (Academic Press, New York, 1975).
[15] P. L. Nunez, Math. Biosci. **21**, 279 (1974).
[16] P. L. Nunez, *Electric Fields of the Brain: The Neurophysics of EEG* (Oxford University Press, New York, 1981).
[17] V. K. Jirsa and H. Haken, Phys. Rev. Lett. **77**, 960 (1996).
[18] V. K. Jirsa and H. Haken, Physica D **99**, 503 (1997).
[19] J. J. Wright and D. T. J. Liley, Behav. Brain Sci. **19**, 285 (1996).
[20] D. T. J. Liley, P. J. Cadusch, and J. J. Wright, Neurocomputing **26–27**, 795 (1999).
[21] D. T. J. Liley, P. J. Cadusch, and M. P. Dafilis, Network Comput. Neural Syst. **13**, 67 (2002).
[22] P. A. Robinson, C. J. Rennie, and J. J. Wright, Phys. Rev. E **56**, 826 (1997).
[23] P. A. Robinson, C. J. Rennie, J. J. Wright, and P. D. Bourke, Phys. Rev. E **58**, 3557 (1998).
[24] C. J. Rennie, J. J. Wright, and P. A. Robinson, J. Theor. Biol. **205**, 17 (2000).
[25] R. A. España and T. E. Scammell, Sleep **27**, 811 (2004).
[26] M. E. Hasselmo, Behav. Brain Res. **67**, 1 (1995).
[27] E. F. Pace-Schott and J. A. Hobson, Nature Reviews Neuroscience **3**, 591 (2002).
[28] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer Series in Synergetics Vol. 18 (Springer-Verlag, Berlin, 1984).
[29] H. Carmichael, *Statistical Methods in Quantum Optics: Master Equations and Fokker-Planck Equations* (Springer, New York, 1999).
[30] C. E. Shannon and W. Weaver, *The Mathematical Theory of Information* (University of Illinois Press, Urbana, IL, 1949).
[31] D. M. Wiberg, *State Space and Linear Systems, Schaum's Outline Series* (McGraw-Hill, New York, 1971).
[32] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **78**, 2554 (1982).
[33] E. Bienenstock and D. Lehmann, Adv. Complex Syst. **1**, 361 (1998).
[34] T. J. Sejnowski, J. Math. Biol. **4**, 303 (1977).
[35] W. Gerstner and W. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity* (Cambridge University Press, Cambridge, UK, 2002).
[36] D. J. Amit, *Modelling Brain Function: The World of Attractor Neural Networks* (Cambridge University Press, Cambridge, UK, 1990).
[37] R. Landauer, IBM J. Res. Dev. **5**, 183 (1961).
[38] C. H. Bennett, Int. J. Theor. Phys. **21**, 905 (1982).
[39] F. Crick and G. Mitchison, Nature (London) **304**, 111 (1983).
[40] K. Kuizenga, J. M. K. H. Wierda, and C. J. Kalkman, Br. J. Anaesth. **86**, 354 (2001).
[41] E. R. John *et al.*, Conscious Cogn **10**, 165 (2001).
[42] R. A. Veselis *et al.*, Br. J. Anaesth. **69**, 246 (1992).
[43] M. L. Williams and J. W. Sleigh, Anaesth. Intensive Care **27**, 265 (1999).
[44] M. Steriade, A. Núñez, and F. Amzica, J. Neurosci. **13**, 3252 (1993).
[45] R. P. Vertes, Neuron **44**, 135 (2004).
[46] M. P. Walker and R. Stickgold, Neuron **44**, 121 (2004).
[47] M. P. Walker, Cell. Mol. Life Sci. **61**, 3009 (2004).
[48] J. K. Wyatt, R. R. Bootzin, J. J. Allen, and J. L. Anthony, Sleep **20**, 512 (1997).