ELSEVIER

Progress in Biophysics & Molecular Biology 85 (2004) 369–385

Progress in Biophysics & Molecular Biology

www.elsevier.com/locate/pbiomolbio

Modelling general anaesthesia as a first-order phase transition in the cortex

Moira L. Steyn-Ross[a,*], D.A. Steyn-Ross[a], J.W. Sleigh[b]

[a] *Department of Physics and Electronic Engineering, University of Waikato, Private Bag 3105, Hamilton, New Zealand*
[b] *Department of Anaesthetics, Waikato Hospital, Hamilton, New Zealand*

Abstract

Since 1997 we have been developing a theoretical foundation for general anaesthesia. We have been able to demonstrate that the abrupt change in brain state brought on by anaesthetic drugs can be characterized as a first-order phase transition in the population-average membrane voltage of the cortical neurons. The theory predicts that, as the critical point of phase change into unconsciousness is approached, the electrical fluctuations in cortical activity will grow strongly in amplitude while slowing in frequency, becoming more correlated both in time and in space. Thus the bio-electrical change of brain-state has deep similarities with thermodynamic phase changes of classical physics. The theory further predicts the existence of a second critical point, hysteretically separated from the first, corresponding to the return path from comatose unconsciousness back to normal responsiveness. There is a steadily accumulating body of clinical evidence in support of all of the phase-transition predictions: low-frequency power surge in EEG activity; increased correlation time and correlation length in EEG fluctuations; hysteresis separation, with respect to drug concentration, between the point of induction and the point of emergence.

PACS: 87.19.La; 05.10.Gg; 05.70.Fh; 87.22.Jb

Keywords: Anaesthesia; Cortical modelling; EEG; Phase transition; Hysteresis

1. Introduction

After 20 years working as a clinical anaesthetist, one of us (JWS) had observed that the process of "putting people to sleep"—that is, rendering the patient controllably and reversibly

*Corresponding author. Tel.: +64-7-8384585.
E-mail addresses: msr@phys.waikato.ac.nz (M.L. Steyn-Ross), asr@waikato.ac.nz (D.A. Steyn-Ross), sleighj@hwl.co.nz (J.W. Sleigh).
URL: http://www.phys.waikato.ac.nz/cortex.

0079-6107/$ - see front matter
doi:10.1016/j.pbiomolbio.2004.02.001

unconscious via application of anaesthetic drugs—is neither smooth nor gradual. Rather, the transition into anaesthetic unconsciousness is both dramatic and abrupt—as if a switch inside the brain is being flipped to the "sleep" position at a critical drug concentration.

This motivated us to ask whether this biological change of state might be analogous to the more familiar thermodynamic phase changes of physics, such as water freezing and ice melting. To investigate this possibility, we adopted the Liley et al. (1999) continuum model of the cortex, and incorporated the effect of a general anaesthetic agent into the model by assuming that the effectiveness of *inhibitory* synaptic events increases as the concentration of the anaesthetic increases. We found that, for certain ranges of anaesthetic concentration, the model predicts the existence of multiple steady states for brain activity, leading to the possibility that, at a critical level of anaesthetic, there would be a sudden switch-over from normal levels of cortical activity to a quiescent, low-firing state.

We then added white-noise driving to the model, simulating the effect of random stimulation of the cortex from subcortical sources. Using stochastic calculus techniques, we computed a theoretical spectrum for the white-noise-driven cortex in the adiabatic limit of fast inputs and slow membrane response, and discovered the surprising prediction of a pronounced *increase* in low-frequency cortical activity as the point of transition into the quiescent state was approached. We were gratified to learn, subsequent to those calculations, that a surge in brain activity during induction of anaesthesia is well-known in the anaesthetics research community, being described as the drug *biphasic effect*. However, prior to our work, there had been no convincing explanation for the puzzling and apparently paradoxical behaviour that a drug intended to inhibit brain activity can have precisely the opposite effect at low concentrations.

In this paper we review the progress to date in our development of an anaesthetic phase-transition theory (Steyn-Ross et al., 1999, 2001a, 2001b, 2003), placing its multiple theoretical predictions in the context of some encouraging recent clinical findings. We conclude with a brief discussion of a recent experimental report that supports the tantalizing speculation that the cycles of *natural sleep* might share several of the phase-transition characteristics of anaesthesia.

2. Macrocolumn model for the cortex

The cerebral cortex is the convoluted outer $\sim$3-mm-thick layer of "grey-matter" neurons responsible for the higher brain functions such as reasoning, awareness, judgement. Because small volumes of neurons are known to act cooperatively, it is not unreasonable to model the cortex as a collection of *macrocolumns*—organized assemblies of cooperating neurons occupying a cylindrical volume of diameter $\sim$1 mm and depth 3 mm, and containing about 100 000 neurons. A macrocolumn schematic is drawn in Fig. 1.

A given neuron is classified according to its effect on the destination nerve cells that receive its electrical impulses (*action potentials*): if the incoming impulse makes a destination cell more likely to fire, then the sending cell is classified as *excitatory*; if the impulse makes the receiving cells less likely to fire, then the transmitting neuron is *inhibitory*. The ratio of excitatory to inhibitory neurons in the cortex is about 85% : 15%. Excitatory neurons tend to be arranged with their *dendrites* (receiving fibres) and *axons* (transmitting fibres) oriented parallel to one another to form aligned current dipoles, while inhibitory neurons have their fibres oriented at random with

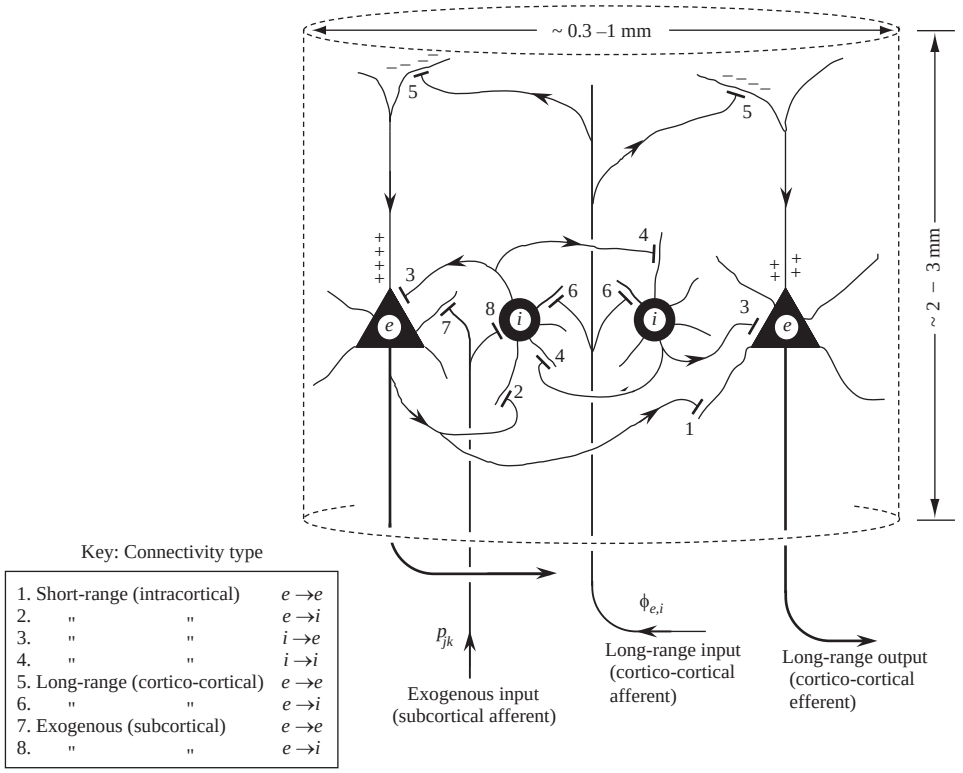

Fig. 1. Schematic representation of the connective topology within a cortical macrocolumn. Only four of the ~100 000 neurons are shown. Triangles are excitatory (pyramidal) cells which receive excitatory input via apical dendrites (e.g., connection type 5) and basal dendrites (1, 7); and inhibitory input directly at the cell body (3). Circles are inhibitory (stellate or basket) cells receiving input from dendritic connections (2, 4, 6) and at the cell body (8). Excitatory output from the macrocolumn is via trunk-lines (axons) shown bold. The symbol $\phi_{e,i}$ represents $e \to e$, $e \to i$ input from distant macrocolumns, and p_{jk} represents input from the subcortex (e.g., thalamus and brainstem). (For clarity, we have omitted p_{ie} and p_{ii} exogenous inputs corresponding to connection types 9 and 10, respectively.)

approximately spherical symmetry; consequently it is generally accepted that the scalp-measured electroencephalogram (EEG) voltage signal is generated by fluctuations in the spatially-averaged excitatory membrane voltage. Although the population of inhibitory neurons has negligible *direct* effect on the EEG, the inhibitory neurons play a crucial moderating role on the behaviour of the excitatory population, so cannot be ignored in any physiologically plausible description of cortical activity.

The Liley et al. (1999) model consists of eight coupled partial differential equations (PDEs) that describe the interacting behaviours of the excitatory and inhibitory neural *populations* within a cortical macrocolumn. The cortex is represented as a 1-D continuum of macrocolumn mass (i.e., a "cortical rod"). Because the emphasis is on population-averaged behaviours, no attempt is made to model the separate time-courses of the 100 000 neurons within a macrocolumn.

The primary variables of interest are the macrocolumn-averaged excitatory soma voltage h_e and inhibitory voltage h_i, both of which can vary in time and space. These population soma voltages

are assumed to obey the following equations of motion:

$$\tau_e \frac{\partial h_e}{\partial t} = (h_e^{rest} - h_e) + \psi_{ee}(h_e) I_{ee}(h_e) + \psi_{ie}(h_e) I_{ie}(h_i), \tag{1a}$$

$$\tau_i \frac{\partial h_i}{\partial t} = (h_i^{rest} - h_i) + \psi_{ei}(h_i) I_{ei}(h_e) + \psi_{ii}(h_i) I_{ii}(h_i). \tag{1b}$$

The symbols $\tau_{e,i}$ are the excitatory and inhibitory membrane time-constants (see Table 1 for values). The ψ_{jk} (where $j,k \in \{e,i\}$) are dimensionless weighting factors that account for the fact that the effectiveness of a presynaptic input I_{jk} (from a neuron of type j) impinging on a postsynaptic neuron (of type k) depends on the signed displacement of the h_k soma voltage from the ionic reversal potential h_j^{rev},

$$\psi_{jk}(h_k) = \frac{h_j^{rev} - h_k}{|h_j^{rev} - h_k^{rest}|}. \tag{2}$$

For example, for I_{ee}, representing excitatory synaptic input to an excitatory cell, the ψ_{ee} weight will be zero if $h_e = h_e^{rev}$, and will change sign as h_e crosses h_e^{rev}. The weights are normalized with

Table 1
Symbol definitions and model constants for the 1-D Liley neural macrocolumn

Symbol	Description	Value	Unit
e,i	(as subscript) excitatory, inhibitory cell populations		
$h_{e,i}$	population mean soma voltage		mV
$\tau_{e,i}$	membrane time constant	0.040, 0.040	s
$h_{e,i}^{rest}$	cell resting potential	−70, −70	mV
$h_{e,i}^{rev}$	cell reversal potential (Nernst potential)	45, −90	mV
$I_{ee,ie}$	total $e \to e, i \to e$ voltage input to excitatory synapses		mV
$I_{ei,ii}$	total $e \to i, i \to i$ voltage input to inhibitory synapses		mV
ψ_{jk}	weighting factors for the I_{jk} inputs $(j,k \in \{e,i\})$		
$p_{ee,ie}$	exogenous (subcortical) spike input to e population	1100, 1600	s^{-1}
$p_{ei,ii}$	exogenous (subcortical) spike input to i population	1600, 1100	s^{-1}
$\phi_{ee,ei}$	long-range (cortico-cortical) spike input to e, i populations		s^{-1}
$\Lambda_{ee,ei}$	characteristic cortico-cortical inverse-length scale	0.040, 0.065	$(mm)^{-1}$
EPSP, IPSP	excitatory, inhibitory post-synaptic potential		mV
$\gamma_{e,i}$	neurotransmitter rate constant for EPSP, IPSP	300, 65	s^{-1}
$G_{e,i}$	peak amplitude of EPSP, IPSP	0.18, 0.37	mV
e	(e.g., Eqs. (3), (11)) base of natural logarithms	2.71828...	
$N_{ee,ei}^{\beta}$	total number of local $e \to e, e \to i$ synaptic connections	3034, 3034	
$N_{ie,ii}^{\beta}$	total number of local $i \to e, i \to i$ synaptic connections	536, 536	
$N_{ee,ei}^{\alpha}$	total number of synaptic connections from distant e-populations	4000, 2000	
$\bar{v}$	mean axonal conduction speed	7000	mm s^{-1}
$S_e(h_e), S_i(h_i)$	sigmoid function mapping soma voltage to firing rate		s^{-1}
S_e^{max}, S_i^{max}	maximum value for sigmoid function	100, 100	s^{-1}
$\theta_{e,i}$	inflexion-point voltage for sigmoid function	−60, −60	mV
$g_{e,i}$	sigmoid slope at inflexion point	0.28, 0.14	$(mV)^{-1}$
ℓ	length of macrocolumn "cell"	1	mm

respect to resting voltage so that if an excitatory cell is at its resting potential h_e^{rest}, then $\psi_{ee} = +1$ and $\psi_{ie} = -1$, thus excitatory input I_{ee} will tend to raise the soma voltage, and inhibitory input I_{ie} will tend to lower the soma voltage.

When an action potential from the sending neuron reaches the synapse (the axon–dendrite junction) connecting it to the receiving neuron, a momentary flux of neurotransmitter is released into the synaptic cleft, causing a temporary alteration in the ionic conductance of the receiving neuron. The resulting current pulse is integrated in the receiving neuron, giving a brief voltage change referred to as the *post-synaptic potential* or (PSP). Liley models the time-course for the PSP as an alpha-function impulse of the form $\gamma t \exp(-\gamma t)$, where γ is a rate constant and $1/\gamma$ is the time-to-peak; see Fig. 2. These excitatory (I_{ek}) and inhibitory (I_{ik}) alpha-function PSPs are defined by the following second-order differential equations:

$$\left(\frac{\partial}{\partial t} + \gamma_e\right)^2 I_{ee}(h_e) = [N_{ee}^{\beta} S_e(h_e) + \phi_{ee}(h_e) + p_{ee}] G_e \gamma_e e, \tag{3a}$$

$$\left(\frac{\partial}{\partial t} + \gamma_e\right)^2 I_{ei}(h_e) = [N_{ei}^{\beta} S_e(h_e) + \phi_{ei}(h_e) + p_{ei}] G_e \gamma_e e, \tag{3b}$$

$$\left(\frac{\partial}{\partial t} + \gamma_i\right)^2 I_{ie}(h_i) = [N_{ie}^{\beta} S_i(h_i) + p_{ie}] G_i \gamma_i e, \tag{3c}$$

$$\left(\frac{\partial}{\partial t} + \gamma_i\right)^2 I_{ii}(h_i) = [N_{ii}^{\beta} S_i(h_i) + p_{ii}] G_i \gamma_i e. \tag{3d}$$

The $G_{e,i}$ are, respectively, the peak amplitudes (in mV) for the EPSP and IPSP response functions (see Table 1). The four p_{jk} are spike-rate inputs (units: s^{-1}) entering the macrocolumn

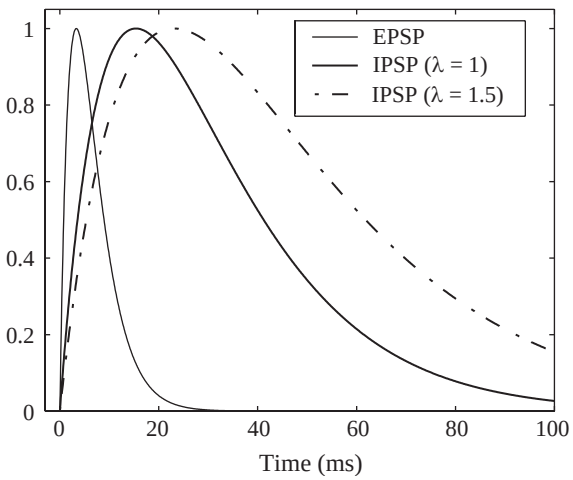

Fig. 2. Model impulse responses $\gamma t \exp(-\gamma t)$ for excitatory (light curve), inhibitory (bold), and anaesthetic-modified inhibitory (bold–dashed) postsynaptic membranes. The symbol λ is the dimensionless anaesthetic-effect scale-factor giving the lengthening of the IPSP duration.

from subcortical sources; the two $\phi_{ee,ei}$ are long-range (cortico–cortical) excitatory spike inputs from distant macrocolumns elsewhere in the cortex; and the four $N_{jk}^{\beta}S_j$ are the local spike-rate contributions generated by within macrocolumn activity. The N_{jk}^{β} constants are the number of $j\to k$ local connections, and the S_j are sigmoidal transfer functions that map from soma voltage to firing-rate,

$$S_e(h_e) = \frac{S_e^{\max}}{1+\exp[-g_e(h_e-\theta_e)]}, \tag{4a}$$

$$S_i(h_i) = \frac{S_i^{\max}}{1+\exp[-g_i(h_i-\theta_i)]}. \tag{4b}$$

Here, $\theta_{e,i}$ is the inflexion-point voltage for the $S_{e,i}$ sigmoid, and $g_{e,i}$ is the slope of the sigmoid function at inflexion. These transfer functions model the voltage-dependent coupling strength between neurons: when the firing rate is high, the coupling strength is high, and vice versa. In our early papers we followed Liley in setting the saturated firing rate $S_{e,i}^{\max} = 1000\ \mathrm{s}^{-1}$; more recently (Steyn-Ross et al., 2003) we lowered these maximum firing rates to a more physiologically plausible $100\ \mathrm{s}^{-1}$.

The ϕ_{ek} long-range inputs from distant macrocolumns (see Eqs. (3a),(3b)) are governed by a pair of wave equations,

$$\left[\left(\frac{\partial}{\partial t}+\bar{v}\Lambda_{ee}\right)^2-\bar{v}^2\frac{\partial^2}{\partial x^2}\right]\phi_{ee}(h_e) = \bar{v}\Lambda_{ee}N_{ee}^{\alpha}\left(\frac{\partial}{\partial t}+\bar{v}\Lambda_{ee}\right)S_e(h_e), \tag{5a}$$

$$\left[\left(\frac{\partial}{\partial t}+\bar{v}\Lambda_{ei}\right)^2-\bar{v}^2\frac{\partial^2}{\partial x^2}\right]\phi_{ei}(h_e) = \bar{v}\Lambda_{ei}N_{ei}^{\alpha}\left(\frac{\partial}{\partial t}+\bar{v}\Lambda_{ei}\right)S_e(h_e), \tag{5b}$$

where $\bar{v}$ is the mean axonal conduction speed, and Λ_{ek} is the characteristic inverse-length scale for cortico-cortical connections. N_{ek}^{α} is the number of long-range connections of type $e\to k$. (No equations for ϕ_{ik} are required since all long-range cortical connections are exclusively from excitatory sources.)

3. Anaesthesia model

In order to bring the effect of anaesthetic agents to the Liley model, we need to know how general anaesthetic agents operate at the cellular level. There is an increasing body of evidence that propofol and other commonly used GABAergic (GABA = gamma-amino butyric acid) induction anaesthetics enhance the inhibitory effect of the GABA neurotransmitter by holding the chloride channels of the post-synaptic (i.e., receiving) neuron open longer, allowing a larger negative charge to accumulate within the cell, thus making it less likely to fire (Franks and Lieb, 1997). We therefore incorporated anaesthetic effect by lengthening the duration of the inhibitory post-synaptic potential (IPSP) by a dimensionless factor λ_{GABA} assumed proportional to anaesthetic concentration: this is done by replacing the IPSP rate constant γ_i in Eqs. (3c)–(3d) by $\gamma_i/\lambda_{\mathrm{GABA}}$. This alteration means that the effectiveness of each inhibitory PSP event is strengthened as anaesthetic concentration increases. See Fig 2.

The second change to the Liley model was to introduce white-noise fluctuations into the four p_{jk} spike-rate fluxes [unit: s^{-1}] entering the cortex from nonspecific excitatory and inhibitory subcortical sources (see Eqs. (3a)–(3d)). Thus each p_{jk} term is rewritten as the sum of a mean value plus a stochastic variation about the mean,

$$p_{jk} \to \langle p_{jk} \rangle + \alpha \sqrt{\langle p_{jk} \rangle}\, \xi_n(t), \tag{6}$$

where α is a dimensionless scale-factor introduced to ensure that the stochastic fluctuations always remain small. Each $\xi_n(t)$ $(n = 1, \ldots, 4)$ is an independent, Gaussian-distributed white-noise generator of zero mean and delta-function covariance,

$$\langle \xi_n(t) \rangle = 0, \quad \langle \xi_n(t)\xi_m(t') \rangle = \delta_{mn}\delta(t - t'). \tag{7}$$

The $\xi_n(t)$ carry units of $s^{-1/2}$. While the motivation for introducing Gaussian-white-noise driving is primarily mathematical—it allows us to compute a theoretical EEG spectrum and to investigate how the spectrum changes in response to changes in anaesthetic concentration—there is compelling physiological evidence that the cortex seems to require a background "wash" of input noise to function normally: it has been known for half a century that the conscious state can only be maintained while the cortex receives an ongoing flux of non-specific excitatory input from the reticular activating system within the brainstem (e.g. Kelly, 1991, p. 815). This brainstem input can be modelled as the sum of many individual Poisson-distributed neuronal processes, which, by the central limit theorem, we approximate as Gaussian-distributed noise.

The final model change was to invoke a "homogeneous slow membrane" approximation in which we assumed (i) that the cortex is spatially homogeneous, so that the $\partial^2/\partial x^2$ spatial-derivative terms in Eqs. (5) for the long-range spike-rates $\phi_{\mathrm{ee,ei}}$ could be set to zero and (ii) that relative to the slow time-scale of the membrane soma at which dendritic inputs are integrated, all synaptic inputs can be treated as "fast" variables that have already reached their steady state. This separation of time-scales (the *adiabatic* approximation) permits a dramatic simplification from the original Liley model (eight nonlinear partial differential equations in time and space) to the adiabatic "Waikato" model—a pair of first-order stochastic ordinary differential equations (i.e., Langevin equations) for h_e and h_i that can be solved analytically using tools from stochastic calculus (Gardiner, 1985). The system equations for the macrocolumn are now,

$$\frac{\mathrm{d}h_e}{\mathrm{d}t} = F_1(h_e, h_i) + \Gamma_e(t), \tag{8a}$$

$$\frac{\mathrm{d}h_i}{\mathrm{d}t} = F_2(h_e, h_i) + \Gamma_i(t), \tag{8b}$$

where the F_1, F_2 are *drift* terms defined by

$$F_1(h_e, h_i) = \frac{1}{\tau_e}\{(h_e^{\mathrm{rest}} - h_e) + \psi_{ee}(h_e)[(N_{ee}^{\alpha} + N_{ee}^{\beta})S_e(h_e) + \langle p_{ee} \rangle]G_e e/\gamma_e + \lambda_{\mathrm{GABA}}\psi_{ie}(h_e)[N_{ie}^{\beta}S_i(h_i) + \langle p_{ie} \rangle]G_i e/\gamma_i\}, \tag{9a}$$

$$F_2(h_e, h_i) = \frac{1}{\tau_i}\{(h_i^{\mathrm{rest}} - h_i) + \psi_{ei}(h_i)[(N_{ei}^{\alpha} + N_{ei}^{\beta})S_e(h_e) + \langle p_{ei} \rangle]G_e e/\gamma_e + \lambda_{\mathrm{GABA}}\psi_{ii}(h_i)[N_{ii}^{\beta}S_i(h_i) + \langle p_{ii} \rangle]G_i e/\gamma_i\} \tag{9b}$$

and the Γ_e, Γ_i are *diffusive noise* terms,

$$\Gamma_e(t) = b_{ee}\xi_1(t) + b_{ie}\xi_3(t), \tag{10a}$$

$$\Gamma_i(t) = b_{ei}\xi_2(t) + b_{ii}\xi_4(t) \tag{10b}$$

whose b_{jk} coefficients, like the $F_{1,2}$ drift terms, depend on GABA anaesthetic effect λ_{GABA} and (h_e, h_i) soma-voltage coordinate,

$$b_{ee} = \psi_{ee}(h_e)\alpha\langle p_{ee}\rangle^{1/2} G_e e/\gamma_e\tau_e, \tag{11a}$$

$$b_{ie} = \lambda_{GABA}\psi_{ie}(h_e)\alpha\langle p_{ie}\rangle^{1/2} G_i e/\gamma_i\tau_e, \tag{11b}$$

$$b_{ei} = \psi_{ei}(h_i)\alpha\langle p_{ei}\rangle^{1/2} G_e e/\gamma_e\tau_i, \tag{11c}$$

$$b_{ii} = \lambda_{GABA}\psi_{ii}(h_i)\alpha\langle p_{ii}\rangle^{1/2} G_i e/\gamma_i\tau_i. \tag{11d}$$

4. Theoretical analysis of model

Our analysis of the resulting anaesthesia model proceeds as follows:

- Identify the distribution of homogeneous *steady-state* voltages as a function of anaesthetic effect.
- Determine the small-perturbation *stability* of these cortical steady states.
- Evaluate the white-noise-driven *spectrum* and *correlation time* of the homogeneous model cortex for small voltage fluctuations about steady state.
- Calculate the *spatial covariance* predictions for an infinitely-long 1-D rod of cortical mass as a function of anaesthetic effect.

4.1. Distribution of steady states

We located the steady states of the "homogeneous slow-membrane" cortex by setting to zero the time-derivatives in Eq. (8) and turning off the white-noise driving (i.e., setting $\Gamma_e = \Gamma_i = 0$). The numerically determined distribution of excitatory and inhibitory steady-state voltages (h_e^0, h_i^0) as a function of anaesthetic effect λ is shown in Fig. 3. For intermediate values of anaesthetic effect (region II), we found three steady-state solutions; at higher levels (region I: "coma") and at lower levels (region III: "anti-anaesthetic" → seizure) only a single state is available to the brain.

We analyzed the stability of the steady states, and found that the middle-branch of region-II is unstable with respect to small perturbations. Therefore the model predicts that a first-order (i.e., discontinuous) transition to the low-firing branch can occur at the right-hand knee A_3 (LOC: loss of consciousness); similarly a distinct jump-return from quiescent to active branch can occur at the left-hand knee Q_1 (ROC: recovery of consciousness). Note that the LOC and ROC critical points correspond to different levels of anaesthetic effect. In other words, the model predicts a *drug hysteresis* effect: the patient will awaken at a *lower* concentration of anaesthetic than that required to put her to sleep. This is a significant and testable prediction.

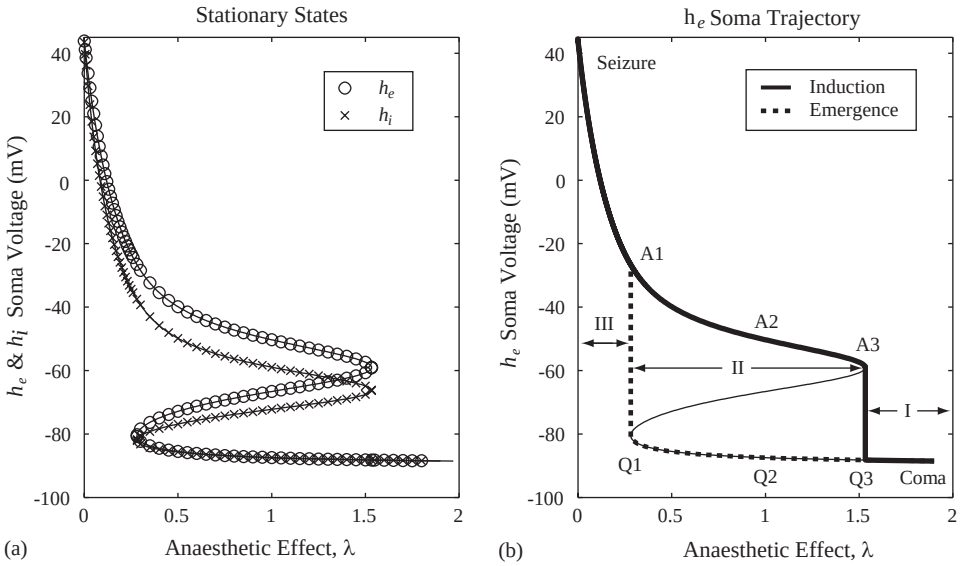

Fig. 3. (a) Model predictions for the stationary states for h_e (circles) and h_i (crosses) as a function of anaesthetic effect λ. (b) In region II, only the upper (active) and lower (quiescent) branch points are stable. For $\lambda > 1.53$ (region I), neural firing is strongly suppressed ("coma"); for $\lambda < 0.3$ (region III), neural firing is maximized ("seizure") [Sigmoid saturation set at $S_{e,i}^{max} = 1000\,s^{-1}$; all other values as in Table 1].

4.2. Theoretical EEG spectrum

We linearized the stochastic differential equations (8) about homogeneous steady state to derive a two-variable Ornstein–Uhlenbeck (Brownian motion) description for the macrocolumn,

$$\frac{d}{dt}\begin{bmatrix}\delta h_e\\ \delta h_i\end{bmatrix} = -\mathbf{A}\begin{bmatrix}\delta h_e\\ \delta h_i\end{bmatrix} + \sqrt{\mathbf{D}}\begin{bmatrix}\xi_e(t)\\ \xi_i(t)\end{bmatrix}, \tag{12}$$

where the $\delta h_{e,i}$ are the small excitatory and inhibitory soma-voltage fluctuations about steady state. $\mathbf{A}$ is the 2×2 drift matrix evaluated at the equilibrium point,

$$\mathbf{A} = -\begin{bmatrix}\frac{\partial F_1}{\partial h_e} & \frac{\partial F_1}{\partial h_i}\\ \frac{\partial F_2}{\partial h_e} & \frac{\partial F_2}{\partial h_i}\end{bmatrix}_{eq} \tag{13}$$

and $\mathbf{D}$ is a diffusion matrix defined such that

$$\sqrt{D_{11}}\xi_e(t) \equiv \Gamma_e(t) = b_{ee}\xi_1(t) + b_{ie}\xi_2(t), \tag{14a}$$

$$\sqrt{D_{22}}\xi_i(t) \equiv \Gamma_i(t) = b_{ei}\xi_3(t) + b_{ii}\xi_4(t). \tag{14b}$$

After applying the delta-correlation property for the four noise sources (Eq. (7)), we obtained a diagonal diffusion matrix,

$$\mathbf{D} = \begin{bmatrix}b_{ee}^2 + b_{ei}^2 & 0\\ 0 & b_{ie}^2 + b_{ii}^2\end{bmatrix}_{eq}. \tag{15}$$

We were then able to compute, using standard Ornstein–Uhlenbeck analysis (Gardiner, 1985), the 2×2 fluctuation spectrum matrix $\mathbf{S}(\omega)$,

$$\mathbf{S}(\omega) = \frac{1}{2\pi}(\mathbf{A} + \mathrm{i}\omega\mathbf{I})^{-1}\mathbf{D}\left(\mathbf{A}^{\mathrm{T}} - \mathrm{i}\omega\mathbf{I}\right)^{-1}, \tag{16}$$

where the superscript T signifies matrix transpose and $\mathbf{I}$ is the identity matrix. The S_{11} element of this matrix returns the fluctuation spectrum for $\delta h_{\mathrm{e}}(t)$, presumed to be the source of the scalp-measured EEG. [The off-diagonal entries in the spectrum matrix correspond to the $\delta h_{\mathrm{ei}}(t)$ cross-spectra; and the S_{22} entry gives the spectrum for the *inhibitory* voltage fluctuations $\delta h_{\mathrm{i}}(t)$.]

Using Eq. (16) we calculated the theoretical $S_{11}(\omega)$ spectrum. We were surprised by the result: despite the fact that anaesthetic is administered with the end-goal of *inhibiting* brain function, our stochastic theory predicted that as anaesthetic concentration is increased towards a critical value, the brain response to white noise driving will *increase* dramatically. It turns out that this paradoxical boost in brain activity just prior to induction is, in fact, well-known in the anaesthetics community, and is referred to as drug *biphasic response* (Kuizenga et al., 1998, 2001b).

Further, the theory predicts that there will be a second biphasic surge in EEG power as the patient emerges from unconsciousness, and, consistent with the drug hysteresis prediction of Fig. 3, that this emergence surge in brain activity will occur at a *lower* concentration of anaesthetic than that recorded for the induction surge. Fig. 4 shows the model predictions for changes in EEG power and spectral content for the induction and emergence trajectories.

Kuizenga and colleagues have demonstrated convincingly that there are two biphasic peaks per induction–emergence cycle for patients undergoing propofol anaesthesia (Kuizenga et al., 2001a). Data from one of Kuizenga's clinical experiments is shown in Fig. 5. Anaesthetic drug concentration is measured in the blood, while the "effect" site is the brain; it takes about 2 min for the drug to diffuse across the blood–brain barrier to generate an altered EEG response, and this delay is clearly evident in Fig. 5(a). Even after compensating for this diffusion delay in Fig. 5(b), an obvious drug-hysteresis effect remains: the induction surge in brain activity occurs at a significantly higher propofol concentration ($\sim$7–8 mg/l) than that ($\sim$1.5–3 mg/l) at which the emergence surge appears.

The pronounced increase in low-frequency power as either critical point is approached (see Fig. 4) is analogous to the *critical slowing down* phenomenon that is the hallmark of thermodynamic phase transitions in physics: near transition the fluctuations become strongly correlated in time. We show in Steyn-Ross et al. (2001b) that the change in EEG correlation time τ can be quantified from the frequency spectrum using spectral entropy, H_ω, defined

$$H_\omega = -\int_0^\infty p(\omega)\ln[p(\omega)]\,\mathrm{d}\omega, \tag{17}$$

where $p(\omega)$ is the probability density function for the spectrum. This particular entropy is a measure of *spectral flatness*, and is maximum when the spectrum is white. Thus an increase in correlation time corresponds to a *reduction* in spectral entropy. The EEG fluctuation spectrum in our anaesthesia model is approximately Lorentzian, and we find that $H_\omega \sim \ln[1/\tau]$. In Steyn-Ross et al. (2001b) we prove that this exponential scaling relationship between correlation time and spectral entropy becomes exact for an ideal Lorentzian spectrum, giving $H_\omega = \ln[2\pi/\tau]$. Thus the

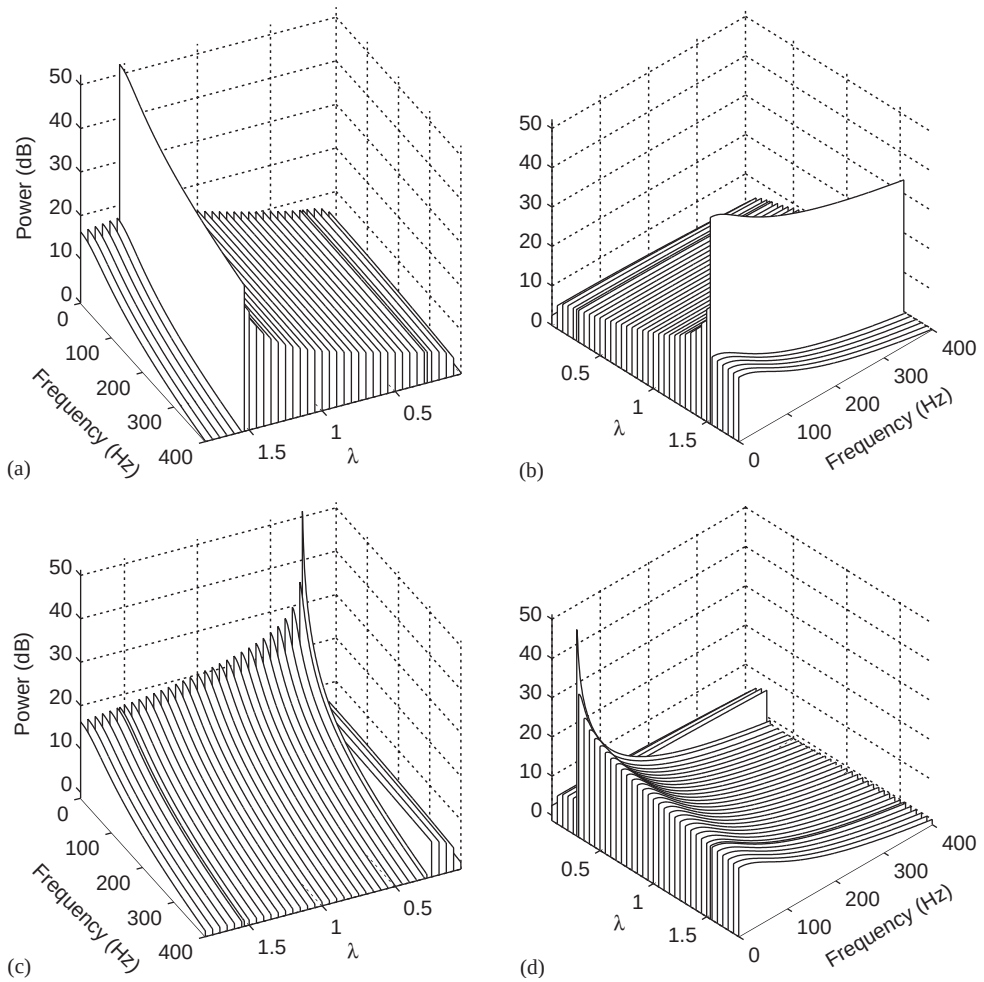

Fig. 4. Spectral power plots for (a, b) *induction* and (c, d) *emergence*. Two views are shown for each trajectory. All spectra have maximum power at DC, with the DC-peakiness becoming more pronounced prior to induction at $\lambda = 1.53$ (a, b), and even more pronounced just prior to emergence at $\lambda = 0.28$ (c, d) [Sigmoid saturation set at $S_{e,i}^{max} = 1000\,s^{-1}$; all other values as in Table 1].

approach to anaesthetic transition can be detected either in the EEG time-domain (increase in correlation time of voltage fluctuations) or in the frequency-domain (increase in low-frequency power and decrease in spectral entropy).

4.3. Numerical simulations

In order to verify the numerical correctness of the model predictions, we ran nonlinearized stochastic simulations of both the full Liley equations and the simplified adiabatic equations, and were able to confirm the number and nature of the steady states (Wilcocks 2001; Steyn-Ross et al., 2001b) as a function of anaesthetic effect. A simulation started on the unstable mid-branch of

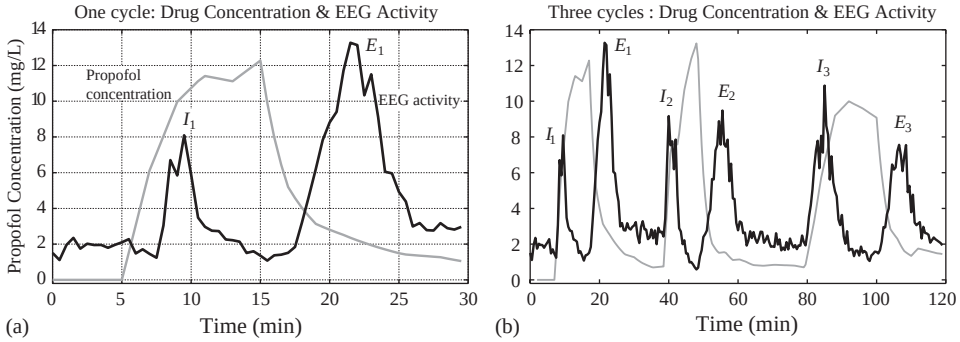

Fig. 5. Multiple slow anaesthesia inductions using propofol. Grey curve: propofol blood concentration in mg/l; black curve: EEG slew-rate activity in μV/s for 11–15 Hz frequency band (EEG activity has been divided by 30 to aid comparison with propofol peaks). Patient response was recorded for three consecutive induction–emergence cycles. (a) First induction–emergence cycle. Note that the EEG response lags behind propofol concentration changes by $\sim$2 min, and that there are two distinct activity peaks: the first at I_1, the induction of unconsciousness, and the second at E_1, the emergence from unconsciousness. (b) Three consecutive cycles. Concentration trace (grey) has been delayed (i.e., moved to right) by 2 min to compensate for lag in brain response [Data provided courtesy of K. Kuizenga].

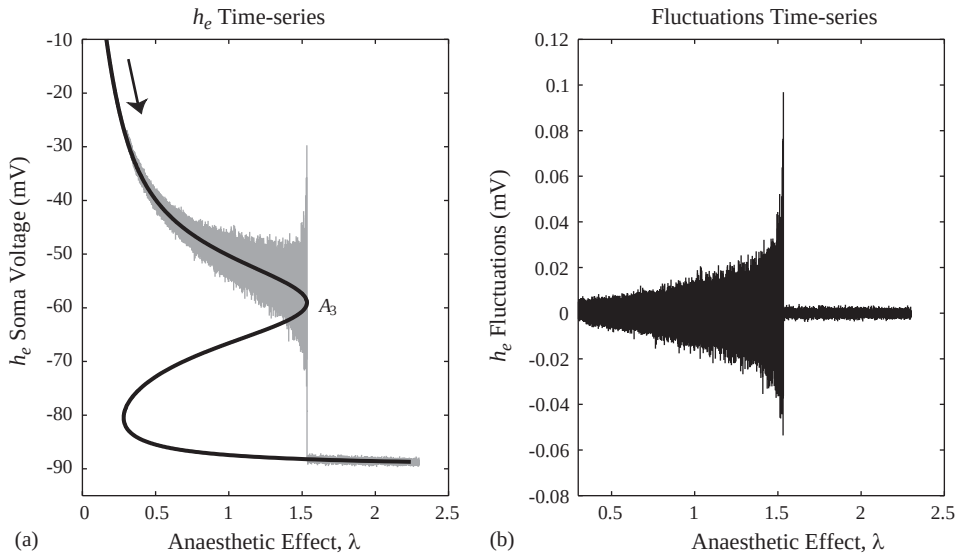

Fig. 6. Simulation time-series of excitatory soma voltage h_e for induction into unconsciousness. Fluctuation amplitude grows strongly as the critical point A_3 is approached. (a) Time development (in grey) along the equilibrium curve (black); (b) fluctuations as the AC residual after subtraction of the DC equilibrium component. In (a), the fluctuations are displayed at 300× actual size in order to make them visible on the equilibrium voltage scale. Their true scale is shown in (b) [Sigmoid saturation set at $S_{e,i}^{max} = 1000\,s^{-1}$; all other values as in Table 1].

Fig. 3 would "fall off" the potential hill there, and would settle, with equal probability, into either the "active" upper branch or the "quiescent" lower-branch potential valley.

Fig. 6 shows a time-series generated by a simulation of the adiabatic equations that was started on the top-branch of Fig. 3 near A_1, with anaesthetic effect λ slowly increasing from 0.3 to 2.3

during the course of the run. It nicely illustrates the growth in fluctuation amplitude as the A_3 induction point at $\lambda = 1.53$ is approached. Fourier analysis of these fluctuations (not shown) confirms quantitatively the theoretical prediction of spectral redistribution towards zero frequency on approach to the critical point.

4.4. Spatial covariance of EEG

Our three earlier papers (Steyn-Ross et al., 1999, 2001a, b) assumed a *homogeneous* cortex in which the anaesthetic properties of a single macrocolumn could be taken as a proxy for the entire cerebral cortex. This is a "single-electrode" theory in the sense that the overall average state of the cortex can, in principle, be determined by a single electrode pair—one wire located in the intracellular fluid to define the zero reference, and the sensing wire located in the "population average" excitatory neuron.

In our most recent paper (Steyn-Ross et al., 2003) we have made a first step towards a *multiple-electrode* theory by modelling the cerebral cortex as an infinite 1-D rod of macrocolumn "mass". We now explicitly allow (weak) spatial variation in the system by retaining the $\partial^2/\partial x^2$ terms in the long-range coupling from distant macrocolumns distributed along a rod. We refer to this treatment as "spatio-adiabatic" since we continue to assume an adiabatic separation of time-scales in which synaptic inputs are fast processes, so can be replaced by their steady-state values.

This inclusion of a spatial dimension alters the $F_{1,2}$ drift terms of Eq. (9), which now read,

$$F_1'(h_e,h_i) = \frac{1}{\tau_e}\Bigg\{(h_e^{\rm rest}-h_e)+\psi_{ee}(h_e)\left[(N_{ee}^{\alpha}+N_{ee}^{\beta})S_e(h_e)+\frac{1}{\Lambda_{ee}^2}\frac{\partial^2\phi_e}{\partial x^2}+\langle p_{ee}\rangle\right]G_e e/\gamma_e$$
$$+\lambda_{\rm GABA}\psi_{ie}(h_e)[N_{ie}^{\beta}S_i(h_i)+\langle p_{ie}\rangle]G_i e/\gamma_i\Bigg\}, \tag{18a}$$

$$F_2'(h_e,h_i) = \frac{1}{\tau_i}\Bigg\{(h_i^{\rm rest}-h_i)+\psi_{ei}(h_i)\left[(N_{ei}^{\alpha}+N_{ei}^{\beta})S_e(h_e)+\frac{1}{\Lambda_{ei}^2}\frac{\partial^2\phi_i}{\partial x^2}+\langle p_{ei}\rangle\right]G_e e/\gamma_e$$
$$+\lambda_{\rm GABA}\psi_{ii}(h_i)[N_{ii}^{\beta}S_i(h_i)+\langle p_{ii}\rangle]G_i e/\gamma_i\Bigg\}. \tag{18b}$$

Similarly, because the cortex is now driven by 1-D spatio-temporal white noise, the delta-correlation property of Eq. (7) must be rewritten to include a delta-function over 1-D space,

$$\langle \xi_n(x,t)\rangle = 0, \quad \langle \xi_n(x,t)\xi_m(x',t')\rangle = \delta_{mn}\delta(x-x')\delta(t-t'). \tag{19}$$

This extra delta-function implies that the $\xi(x,t)$ noise sources now have dimension length$^{-1/2}\cdot$ time$^{-1/2}$, so we choose to scale the $\xi(x,t)$ by $\sqrt{\ell}$, where ℓ is the effective length of the macrocolumn "cell", here taken as $\ell = 1$ mm. This ensures that the product $\sqrt{\ell}\xi(x,t)$ has units of s^{-1}, independent of space. [For the case of 3-D spatio-temporal white noise $\xi(\vec{r},t)$, the appropriate product would be $\sqrt{V}\xi(\vec{r},t)$ where V is the nominal *volume* of the cellular unit.] Thus the b_{jk} coefficients appearing in the diffusion matrix (Eq. (15)) and defined in Eq. (11) are redefined for the 1-D cortex as

$$b_{jk} \to \sqrt{\ell}\, b_{jk}. \tag{20}$$

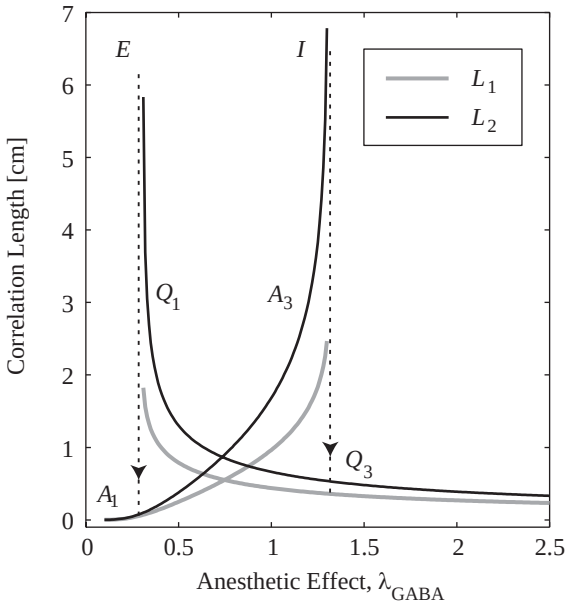

Fig. 7. Variation of L_1 and L_2 correlation lengths as a function of anaesthetic effect λ_{GABA} for the spatio-adiabatic 1-D cortex. Points A_1 and A_3 lie on the active, high-firing branch; Q_1 and Q_3 are on the quiescent, low-firing branch. Fluctuations become more correlated in space on approach to the transition points [For this and next figure, sigmoid saturation firing-rate set at $S_{e,i}^{max} = 100\,s^{-1}$].

To quantify the degree to which the voltage fluctuations at separated points x and x' on the cortex are correlated, we calculated the theoretical steady-state spatial covariance $G(x, x')$ under the assumption that the covariance depends only on the separation $|x - x'|$ of the sensing electrodes (and not on their absolute positions x and x'). Because of the presence of the space-derivative terms, this calculation cannot be done directly in x-space. Instead, we Fourier transform the Langevin equations to wavenumber q-space (thus $\partial^2/\partial x^2 \to -q^2$), compute the wavenumber covariance $\tilde{G}(q, q')$, then Fourier invert to retrieve the x-space covariance $G(|x - x'|)$. The resulting spatial covariance is predicted to be the difference of two exponential decays in space,

$$G(|x - x'|) = a_1 L_1 \exp(-|x - x'|/L_1) \quad - a_2 L_2 \exp(-|x - x'|/L_2), \tag{21}$$

where the values of the $a_{1,2}$ and $L_{1,2}$ coefficients are determined by the chosen operating point on the trajectory of steady states shown in Fig. 3. L_1 and L_2 can be interpreted as *correlation lengths* since each gives the electrode separation at which its respective exponential has decayed to $1/e$ times its the zero-lag maximum.

Fig. 7 shows the predicted variation in L_1 and L_2 as a function of anaesthetic effect for both the induction and emergence trajectories. The significant result is the *increase in correlation length* as the critical points are approached. This means that EEG fluctuations are predicted to become much more correlated in space on approach to induction (I) of unconsciousness, and again for the return path on approach to emergence (E) back into consciousness. We also note that the bottom-branch correlation lengths are *larger* than that for the A_1 state of (presumed) normal

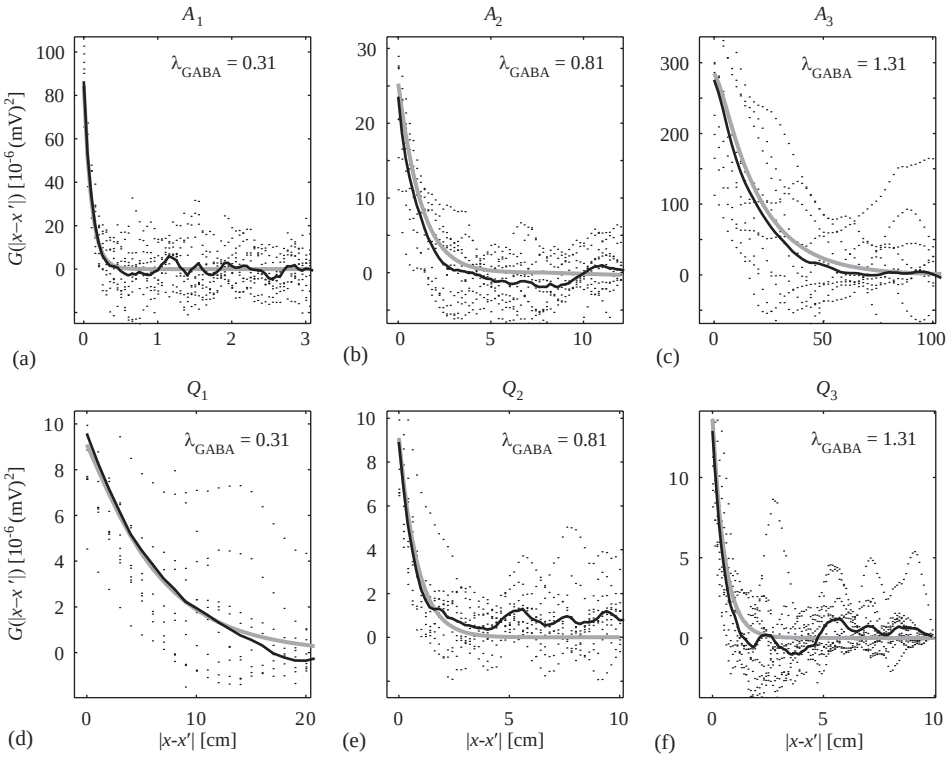

Fig. 8. Comparison of Eq. (21) EEG covariance predictions (thick-grey curves), stochastic simulation results (black dots), and 10-run averages (black curves) for a 1-D rod of cortical mass. The top three panels (a, b, c) trace the induction trajectory $A_1 \rightarrow A_2 \rightarrow A_3$ along the active branch of Fig. 3; the bottom three panels (d, e, f) trace the emergence trajectory $Q_1 \leftarrow Q_2 \leftarrow Q_3$ along the quiescent branch. The spatial width of the covariance curve increases strongly on approach to induction (a)→(c) [note change in x-axis scale], and also on approach to emergence near (d).

consciousness, thus the unconscious state is more ordered, so is expected to have lower entropy (⇒ fewer microstates available to the cortex). This prediction of a reduction in cortical entropy is consistent with our earlier (Steyn-Ross et al., 2001a) statistical mechanics treatment of the anaesthetic phase transition in which we assume that anaesthetic effect can be mapped to an *inverse* thermodynamic temperature.

We ran numerical simulations of the spatio-adiabatic equations, computing the spatial covariance for the voltage fluctuations. We demonstrated that, after averaging the results over a sufficiently large number of runs, the simulation covariance estimates converged to the linearized theoretical prediction of Eq. (21) (Whiting, 2003; Steyn-Ross et al., 2003). This is illustrated in Fig. 8.

There is recent clinical support for the model predictions of increased correlation length and increased correlation time near the anaesthesia transition points. John et al. (2001) analyzed the EEG changes for 176 patients undergoing general anaesthesia for a range of different anaesthetic agents, and found a marked increase in EEG *coherence* (normalized cross-spectral density for separated points x, x') just prior to loss of consciousness (LOC), and again at recovery of

consciousness. (He also reported a large increase in low-frequency power on approach to LOC—thus confirming the biphasic drug response reported earlier by Kuizenga et al. (1998) and others.)

5. Summary and future work

We have developed a mean-field model that describes the bulk electrical changes in the cerebral cortex induced by a GABAergic general anaesthetic agent. The primary prediction is that there will be a first-order phase transition in the population average neuron voltage at a critical level of anaesthetic concentration. Heralding this step change in brain state will be a pronounced increase in low-frequency power of the EEG signal—this is the "critical slowing down" of classical phase transitions, identified as the "biphasic effect" in the anaesthetics literature. Because the trajectory of cortical steady states forms a reverse-S shape with an unstable mid-branch, the emergence path out of unconsciousness cannot retrace the induction path into unconsciousness: there will be a hysteresis separation between the induction and the emergence critical points.

By analyzing a 1-D model of the cortex we have established that spatially-separated electrodes should register increasingly correlated EEG signals in the vicinity of the critical points. Model results suggest that the unconscious state is characterized by voltage fluctuations that are more strongly correlated—both in time and in space—than is the case in the normal conscious state, leading to the prediction that cortical entropy will be lower when unconscious, and higher when conscious. Our construction of a thermodynamics analogy presumes that anaesthetic acts like an inverse temperature, reducing cortical excitability to bring about a more ordered state.

The clinical anaesthesia results surveyed here (Kuizenga et al., 1998, 2001a, b; John et al., 2001) provide good supporting evidence for all of the major EEG-change predictions of the anaesthetodynamic phase transition theory.

5.1. Extension to natural sleep

The point of contact between natural sleep and our phase-transition theory of anaesthetic sleep is provided by clinical reports of work by Destexhe, Contreras and Steriade (Destexhe et al., 1999) investigating sleep cycles in a cat as the animal transits between slow-wave sleep (SWS) and rapid-eye-movement (REM) sleep (also known as *paradoxical* or dream sleep). Their EEG measurements show pronounced increases both in low-frequency power and in the cortical "space constant" (correlation length for EEG fluctuations) as the animal approaches the SWS-to-REM transition point. These dramatic changes in natural-sleep EEG patterns have persuasive similarities to the cortical changes predicted by our first-order phase-transition model for anaesthesia. This motivates us to examine the possibility of generalizing the theory to construct a phase-change description for transitions between the stages of natural sleep.

Acknowledgements

Fig. 5 was generated using clinical data supplied by Dr. Karel Kuizenga; we thank Dr. Kuizenga for his prompt and courteous response to our request for access to his measurements.

References

Destexhe, A., Contreras, D., Steriade, M., 1999. Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. J. Neurosci. 19 (11), 4595–4608.

Franks, N.P., Lieb, W.R., 1997. Anaesthetics set their sites on ion channels. Nature 389, 334–335.

Gardiner, C.W., 1985. Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences. Springer Series in Synergetics, Vol. 13. Springer, Berlin.

John, E.R., Prichep, L.S., Kox, W., Valdés-Sosa, P., Bosch-Bayard, J., Aubert, E., Tom, M., diMichele, F., Guginoi, L.D., 2001. Invariant reversible QEEG effects of anesthetics. Consciousness and Cognition 10, 165–183.

Kelly, D.D., 1991. Disorders of sleep and consciousness. In: Kandel, E.R., Schwartz, J.H., Jessell, T.M. (Eds.), Principles of Neural Science, 3rd Edition. Prentice-Hall International, Toronto, pp. 805–819 (Chapter 52).

Kuizenga, K., Kalkman, C.J., Hennis, P.J., 1998. Quantitative electroencephalographic analysis of the biphasic concentration—effect relationship of propofol in surgical patients during extradural analgesia. Br. J. Anaesthesia 80, 725–732.

Kuizenga, K., Proost, J.H., Wierda, J.M.K.H., Kalkman, C.J., 2001a. Predictability of processed electroencephalography effects on the basis of pharmacokinetic–pharmacodynamic modeling during repeated propofol infusions in patients with extradural analgesia. Anesthesiology 95, 607–615.

Kuizenga, K., Wierda, J.M.K.H., Kalkman, C.J., 2001b. Biphasic EEG changes in relation to loss of consciousness during induction with thiopental, propofol, etomidate, midazolam or sevoflurane. Br. J. Anaesthesia 86, 354–360.

Liley, D.T.J., Cadusch, P.J., Wright, J.J., 1999. A continuum theory of electro-cortical activity. Neurocomputing 26–27, 795–800.

Steyn-Ross, M.L., Steyn-Ross, D.A., Sleigh, J.W., Liley, D.T.J., 1999. Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: evidence for a general anesthetic-induced phase transition. Phys. Rev. E 60, 7299–7311.

Steyn-Ross, M.L., Steyn-Ross, D.A., Sleigh, J.W., Wilcocks, L.C., 2001a. Toward a theory of the general anesthetic-induced phase transition of the cerebral cortex: I. A statistical mechanics analogy. Phys. Rev. E 64, 011917.

Steyn-Ross, D.A., Steyn-Ross, M.L., Wilcocks, L.C., Sleigh, J.W., 2001b. Toward a theory of the general anesthetic-induced phase transition of the cerebral cortex: II. Stochastic numerical simulations, spectral entropy, and correlations. Phys. Rev. E 64, 011918.

Steyn-Ross, M.L., Steyn-Ross, D.A., Sleigh, J.W., Whiting, D.R., 2003. Theoretical predictions for spatial covariance of the EEG signal during the anesthetic-induced phase transition: increased correlation length and emergence of self-organization. Phys. Rev. E 68, 021902.

Whiting, D.R., 2003. Modelling the spatial effects of the anaesthetic-induced phase transition in the cerebral cortex. Master's thesis, University of Waikato, Hamilton, New Zealand.

Wilcocks, L.C., 2001. Investigation of a stochastic model of the electrical behaviour of the cerebral cortex. Master's thesis, University of Waikato, Hamilton, New Zealand.