

Prediction of electroencephalographic spectra from neurophysiology

P. A. Robinson,^{1,*} C. J. Rennie,^{1,2,3} J. J. Wright,⁴ H. Bahramali,³ E. Gordon,³ and D. L. Rowe^{1,3}

¹*School of Physics, University of Sydney, New South Wales 2006, Australia*

²*Department of Medical Physics, Westmead Hospital, Westmead, New South Wales 2145, Australia*

³*Brain Dynamics Center, Department of Psychological Medicine, Westmead Hospital and University of Sydney, Westmead, New South Wales 2145, Australia*

⁴*Mental Health Research Institute, Parkville, Victoria 3052, Australia*

(Received 22 May 2000; revised manuscript received 19 October 2000; published 18 January 2001)

A recent neurophysical model of propagation of electrical waves in the cortex is extended to include a physiologically motivated subcortical feedback loop via the thalamus. The electroencephalographic spectrum when the system is driven by white noise is then calculated analytically in terms of physiological parameters, including the effects of filtering of signals by the cerebrospinal fluid, skull, and scalp. The spectral power at low frequencies is found to vary as f^{-1} when awake and f^{-3} when asleep, with a breakpoint to a steeper power-law tail at frequencies above about 20 Hz in both cases; the f^{-1} range concurs with recent magnetoencephalographic observations of such a regime. Parameter sensitivities are explored, enabling a model with fewer free parameters to be proposed, and showing that spectra predicted for physiologically reasonable parameter values strongly resemble those observed in the laboratory. Alpha and beta peaks seen near 10 Hz and twice that frequency, respectively, in the relaxed wakeful state are generated via subcortical feedback in this model, thereby leading to predictions of their frequencies in terms of physiological parameters, and of correlations in their occurrence. Subcortical feedback is also predicted to be responsible for production of anticorrelated peaks in deep sleep states that correspond to the occurrence of theta rhythm at around half the alpha frequency and sleep spindles at 3/2 times the alpha frequency. An additional positively correlated waking peak near three times the alpha frequency is also predicted and tentatively observed, as are two new types of sleep spindle near 5/2 and 7/2 times the alpha frequency, and anticorrelated with alpha. These results provide a theoretical basis for the conventional division of EEG spectra into frequency bands, but imply that the exact bounds of these bands depend on the individual. Three types of potential instability are found: one at zero frequency, another in the theta band at around half the alpha frequency, and a third at the alpha frequency itself.

DOI: 10.1103/PhysRevE.63.021903

PACS number(s): 87.10.+e, 87.19.La, 87.18.-h

I. INTRODUCTION

Electroencephalograms (EEGs) have diverse forms that are correlated with differing pathologies and states of arousal, as determined independently by behavioral and clinical measures [1], and are inferred to be closely connected to brain dynamics, information processing, and cognition [2,3]. These correlations are widely used diagnostically, where they have been employed to elucidate fundamental mechanisms of overall brain function and dysfunction. Even so, they are not well understood in terms of the underlying physiology, despite 125 years' work [1,4,5]. Prominent in EEGs are the alpha rhythm (narrowband at 7–11 Hz in the relaxed waking state, weakening rapidly with increased attention, or with sleep), and beta rhythm (broader band at 15–25 Hz), and sleep spindles (12–14 Hz, occurring in deeper sleep and sometimes termed sigma activity), as shown in Fig. 1. The experimental setup and procedures for acquisition of these spectra have been described in detail elsewhere, along with the methods used to ensure that artifacts do not significantly contaminate the data [6]; a brief summary is given in Appendix A. The origins of these rhythms are also obscure and have not been quantified in terms of physiology, although the thalamus is widely thought

to be involved in alpha, beta, and sleep spindles [7–12], since experiments show high coherence between EEGs in the two structures [10,13–15]. Underlying the specific rhythms is a smoother background EEG spectrum whose origin and structure are even more cryptic. Recently, analogous magnetoencephalographic (MEG) measurements revealed that the low frequency component of this background was $1/f$ noise, whose power spectrum $P(f)$ decreased as $P(f) \sim f^{-a}$, with $a \approx 1$ [16]; at high frequencies the decrease was much steeper. In deep sleep there is a strong enhancement of the spectrum at low frequencies, in the so-called *delta* ($f < 3.5$ Hz) and *theta* ($4 \text{ Hz} < f < 7.5 \text{ Hz}$) ranges [1,10,17,18].

Scientific and clinical understanding of EEGs would be greatly facilitated by a theory, based on physiology, that could predict features of EEG spectra such as those mentioned above. Consequently, many attempts to model large-scale cortical electrical activity have been made, with varying degrees of physiological realism and incorporating differing aspects of the cortex. One avenue has been to average over microscopic neural structure to develop continuum cortical models on scales from millimeters up to the whole cortex [7,10,15,19–33]. Some continuum models yield wave equations for the propagation of cortical activity [10,21–24,27,31–33], which are particularly convenient for analysis and which incorporate realistic anatomical features such as separate excitatory and inhibitory neural populations

*Electronic address: robinson@physics.usyd.edu.au

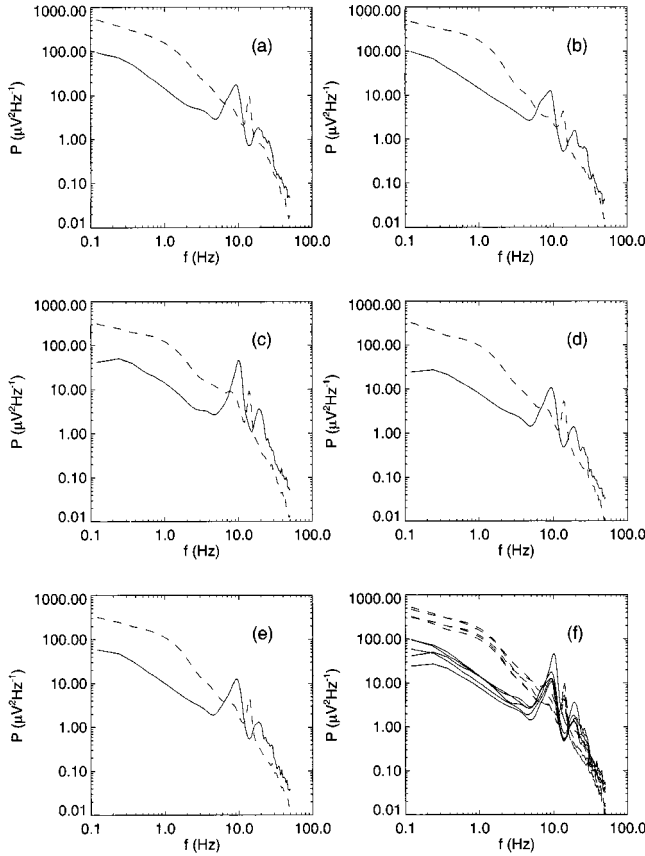


FIG. 1. Illustrative examples of EEG spectra from a normal subject, with waking spectra (eyes closed, solid curves) showing alpha and beta peaks near 9.5 Hz and 19 Hz, while sleep spectra (sleep stage 2, dashed curves) show spindle peaks near 14 Hz. All these spectra have been smoothed for clearer display using a 1 Hz wide window function. (a) Cz electrode. (b) Fz electrode. (c) Pz electrode. (d) C3 electrode. (e) C4 electrode. (f) All spectra superposed to show interelectrode variability.

(e.g., pyramidal cells and interneurons, respectively), nonlinearities, long-range excitatory connections, dendritic signal-integration effects, axonal time delays, and intracortical feedback.

In this paper we study the spectral properties and parameter dependences of our recent physiology-based wave-equation theory of the cortex [21–23,33], concentrating in particular on the consequences of adding a subcortical feedback loop, probably involving the thalamus, with plausible physiological characteristics, and the filtering that occurs between cortex and scalp. Our aims are to formulate a relatively simple, physiologically based model that can reproduce the main features of EEGs such as those in Fig. 1, and to explore its qualitative properties and parameter sensitivities. In future work, we will test the predictions of this model via quantitative comparisons with data. It is also anticipated that the usefulness of our results will be further enhanced by comparison with simultaneous measurements of brain function via positron emission tomography or functional magnetic resonance imaging.

In Sec. II we briefly review the model developed in our previous work and extend it to include corticothalamic feed-

back and filtering by overlying tissues. Analytic predictions for EEG spectra in terms of physiological parameters are then made in Sec. III, and the structure of the underlying spectrum is estimated. Parameter sensitivities are explored in Sec. IV, leading to elimination of several parameters to yield a simplified model that is well suited for comparison with data. This simplified model is used to estimate the locations, sizes, and possible instabilities of spectral peaks in Sec. V, and thereby formulate constraints on the underlying feedback physiology. These results justify the customary division of EEG spectra into distinct bands, but highlight the significance of individual variability in this subdivision. In the case of spectra predicted when the cortex is near marginal stability, a state that has been inferred from MEG observations [16], further simplifications can be obtained; these are discussed in Sec. VI.

II. THEORY

In this section we outline the main relevant results of our recent wave-equation formulation of continuum cortical dynamics [21–23], which incorporates the physiological effects discussed in Sec. I. We then generalize this model in Sec. II B to incorporate a corticothalamic feedback loop, motivated by a range of physiological considerations.

A. Basic model

The mean firing rates (or *pulse densities*) Q_a of excitatory ($a=e$) and inhibitory ($a=i$) neurons are approximately related to the cell-body potentials V_a by

$$Q_a(\mathbf{r}, t) = \frac{Q_a^{\max}}{1 + \exp[-(\pi/\sigma_a\sqrt{3})\{V_a(\mathbf{r}, t) - \theta_a\}]}, \quad (1)$$

where θ_a is the mean firing threshold of neurons of type a , σ_a is the standard deviation in this threshold in the neural population, and $Q_a^{\max} \geq 100 \text{ s}^{-1}$ is the maximum attainable firing rate. The coordinate \mathbf{r} refers to position on the cortex, modeled as a two-dimensional sheet, which is justified by its relative thinness.

The quantity V_a is the potential at the cell body after inputs have been summed and filtered through the dendrites. A good approximation to V_a is [21–23]

$$V_a(\mathbf{r}, t) = \int_{-\infty}^{\infty} L(t-t') P_a(\mathbf{r}, t') dt', \quad (2)$$

with

$$L(u) = \frac{\alpha\beta}{\beta - \alpha} (e^{-\alpha u} - e^{-\beta u}) \Theta(u), \quad (3)$$

where P_a is the mean potential generated by action potentials arriving from other neurons at the dendrites of neurons of type a , Θ is the unit step function, and α and β are dendritic rate constants, with $\beta \geq \alpha$ assumed without loss of generality. The Fourier transform of $L(u)$, used below, is

$$L(\omega) = (1 - i\omega/\alpha)^{-1} (1 - i\omega/\beta)^{-1}, \quad (4)$$

which implies that the dendrites act as a low-pass filter with cutoff frequency α and a steeper fall-off for $\omega > \beta$.

The potential P_a at a particular location comprises contributions from fields $\phi_{e,i}$ that represent signals propagating from other cortical neurons, and subcortical inputs ϕ_s [21,23]:

$$P_a = N_{ae}s_e\phi_e + N_{ai}s_i\phi_i + N_{as}s_s\phi_s. \quad (5)$$

Here, N_{ab} is the mean number of couplings from neurons of type $b=e,i,s$ to those of type a , and s_b is the size of the response to a unit signal from neurons of type b .

Outgoing pulses from each neuron propagate along its axon and axonal tree at a velocity $v \approx 10 \text{ m s}^{-1}$. Assuming an isotropic distribution of axons with a distribution of ranges in accord with experiment (see Ref. [21] for the exact distribution), this propagation can be described by damped wave equations for the fields ϕ_a , which represent axonal signals:

$$D_a\phi_a(\mathbf{r},t) = Q_a(\mathbf{r},t), \quad (6)$$

$$D_a = \frac{1}{\gamma_a^2} \left[\frac{\partial^2}{\partial t^2} + 2\gamma_a \frac{\partial}{\partial t} + \gamma_a^2 - v^2 \nabla^2 \right], \quad (7)$$

where $\gamma_a = v/r_a$ and r_a is the characteristic range of axons of type a [21]. Strictly speaking, very short-range excitatory axons have lower v than the longer range corticocortical ones, because they are not myelinated. This effect is omitted here on the grounds that the difference will not have strong effects on long-range wave propagation, but could be incorporated by splitting excitatory neurons into two populations with a corresponding division of the field ϕ_e into two parts.

Equations (1)–(7) determine the steady states of cortical activation, when the cortex is driven by a constant, spatially uniform stimulus $\phi_s^{(0)}$. Small perturbations relative to these steady states obey a linear wave equation which yields the transfer functions [23]

$$\frac{\phi_e(\mathbf{k},\omega)}{\phi_s(\mathbf{k},\omega)} = \frac{G_{es}L(\omega)}{D_e[1 - G_{ii}L(\omega)] - G_{ee}L(\omega)}, \quad (8)$$

$$\frac{\phi_i(\mathbf{k},\omega)}{\phi_s(\mathbf{k},\omega)} = \frac{D_e G_{is}}{G_{es}} \frac{\phi_e(\mathbf{k},\omega)}{\phi_s(\mathbf{k},\omega)}, \quad (9)$$

$$D_e(\mathbf{k},\omega) = k^2 r_e^2 + (1 - i\omega/\gamma_e)^2, \quad (10)$$

in Fourier space. Here, the gain parameters $G_{ab} = \rho_a N_{ab} s_b$ express the response strength in neurons a due to a unit signal incident from neurons of type b . The parameter $\rho_a = dQ_a^{(0)}/dV_a \approx \pi Q_a^{(0)}/\sigma_a \sqrt{3}$ is evaluated in the steady state where $Q_a^{(0)} = 5 - 10 \text{ s}^{-1} \ll Q_a^{\max}$ is the steady-state firing rate. In writing (8) and (9) we used the short range of inhibitory axons to set $D_i = 1$, the *local inhibition approximation* [21], and assumed that the numbers of interconnections between neural types are proportional to the number of available synapses, the *random connectivity approximation* [21–23,28,34].

TABLE I. Physiological estimates of parameters of the model, as discussed in Refs. [23] and [24]. The limits given are only approximate.

| Quantity | Value | Unit |
|--------------------------|-------------|-------------------------------|
| $\sigma_{e,i}$ | 3–8 | mV |
| $\theta_{e,i}$ | 10–25 | mV |
| α | 50–200 | s^{-1} |
| β/α | 1–5 | |
| r_e | 70–100 | mm |
| r_i | ~ 0.1 | mm |
| v | 7–10 | m s^{-1} |
| t_0 | 25–100 | ms |
| N_{ee}, N_{ie}, N_{se} | 2000–6000 | |
| N_{ei}, N_{ii}, N_{si} | 300–1000 | |
| N_{es}, N_{is} | 30–100 | |
| $s_{e,s}$ | 1–3 | $\mu\text{V s}$ |
| $-s_i$ | 5–8 | $\mu\text{V s}$ |
| γ_e | 70–150 | s^{-1} |
| γ_i | $\sim 10^5$ | s^{-1} |
| $Q_{e,i}^{\max}$ | 100–300 | s^{-1} |
| $Q_{e,i}$ | 5–10 | s^{-1} |
| $\rho_{e,i,s}$ | 1000–6000 | $\text{V}^{-1} \text{s}^{-1}$ |
| G_{ee} | 2–90 | |
| $-G_{ii}$ | 1.3–60 | |
| $-G_{es}$ | 0.03–2 | |

The parameters of the above model are physiologically measurable, although not all have yet been measured precisely. Table I lists typical values inferred from physiology [23,24], plus values of several other quantities that are discussed below.

B. Subcortical feedback

Several suggestions for the production of the resonances that correspond to observed cortical rhythms have been made. These include the possible occurrence of standing waves in the cortex, a potential scenario that is incorporated in the model outlined in Sec. II A, but which does not appear to yield sufficiently sharp resonances to account for observations [21–23] if physiologically realistic input parameters are used in our model. A second possibility is the presence of a thalamic “pacemaker” or “clock” [9,10], which we do not consider here, although it can be incorporated into our model. Third, it has been widely noted that the thalamus also displays alpha rhythm (and possibly beta and sleep-spindle peaks), with a high degree of correlation with the cortical one [7–10,13,14,18], which may result either from pacemaker activity or from some form of corticothalamic feedback. Hence, we are particularly motivated to include corticothalamic feedback in our model.

A key subcortical feedback loop that may underlie production of the alpha rhythm involves feedback via the diffuse thalamocortical projection system [8,9,13,14,35]. This system projects both to other thalamic nuclei and from the thalamus to the whole cortex and thus has the potential to be

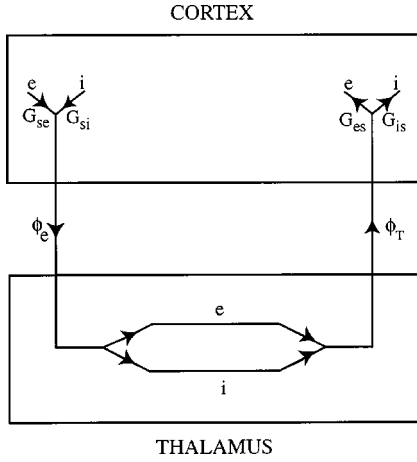


FIG. 2. Schematic block diagram of the corticothalamic feedback loop, showing inputs from excitatory (e) and inhibitory (i) neurons in the cortex, transmission through both thalamic populations, and feedback to both cortical populations. The gain parameters G_{ab} are defined after [10] in the text.

involved in modulating large-scale EEG activity [8,14,18,36]; it also has the correct latency of tens of ms [8]. Other corticothalamic loops pass through the so-called specific nuclei of the thalamus (e.g., the pulvinar, lateral posterior, and ventral anterior nuclei), which themselves project preferentially back to the same cortical areas that stimulate them [1,36]. We stress that, although we assume corticothalamic feedback in the remainder of this paper, for definiteness and simplicity, the results obtained are applicable to other potential feedback loops with like characteristics. Corticothalamic feedback should be significant because the thalamus is known to strongly affect the state of attention of the cortex, implying the existence of influential projections from thalamus to cortex. Strong projections in the reverse direction also exist to complete the loop, involving an even greater number of axons [37–39].

Corticothalamic (CT) feedback is modeled approximately here by assuming that ϕ_s is the sum of a non-CT part ϕ_N and a feedback ϕ_T , which travels along excitatory neurons that project to excitatory and inhibitory neurons in the thalamus before returning to the cortex via other excitatory neurons. This situation is illustrated in Fig. 2. The feedback signal thus passes through $n \sim 1$ additional neurons compared to direct corticocortical connections (which involve more than one neuron on average). This adds a propagation time delay t_0 and $n \sim 1$ extra stages of dendritic filtering with rate constants η_1 and η_2 of the same order as α and β . We thus approximate ϕ_T by

$$\phi_T = T_e(\omega) \phi_e. \quad (11)$$

This incorporates feedback of both excitatory and inhibitory input signals to excitatory neurons (thereby generating the field ϕ_e) that project to the thalamus. After time delays and thalamic low-pass filtering, encapsulated in the transfer function $T_e(\omega)$, this signal returns to the cortex via other excitatory neurons. We also allow for the possibility of both direct feedback, and feedbacks that emphasize changes in cortical

signals by differentiating them in the loop. Differentiation circuits are commonplace in the brain (e.g., for edge detection in the visual field), so the latter type of feedback certainly cannot be ruled out, and has been discussed previously in the context of CT feedback [8]. Limiting cases of truly specific feedback (each point in the cortex feeds back to itself) and totally diffuse feedback with uniform driving (every point feeds back to the whole cortex, making every point equivalent to every other one if the driving is at least statistically uniform, thereby reducing the system to a zero-dimensional one) are included in the form (11), but intermediate types of spatial variation would require a k -dependent transfer function and are not considered here.

The frequency-domain transfer function that corresponds to the above physiology can be written as

$$T_e(\omega) = G_{ee} \Psi(\omega) \tau(\omega) / G_{es}, \quad (12)$$

$$\Psi(\omega) = \psi - i\omega t_0 \psi', \quad (13)$$

$$\tau(\omega) = \frac{e^{i\omega t_0}}{(1 - i\omega/\eta_1)^n (1 - i\omega/\eta_2)^n}, \quad (14)$$

where $\tau(0) = 1$, and ψ and ψ' measure the strengths of direct and differential feedbacks. The factor t_0 in (13) is included to make ψ' dimensionless.

The magnitudes and signs of ψ and ψ' in (13) are expected to vary with state of arousal as the relative responses of the excitatory and inhibitory neurons in the thalamus change. Such changes may result from the action of neurotransmitters or neuromodulators that preferentially affect one population or the other, for example. Hence, if the coupling of thalamic inhibitory neurons to thalamocortical excitatory fibers exceeds that of excitatory ones in some states, the thalamus will invert incoming signals as well as filtering, delaying, and (in part) differentiating them. In Sec. IV we will see that, for many purposes, one may approximate (14) by setting $n = 1$ and $\eta_1 = \eta_2 = \alpha$.

Combining (8)–(14), we find the transfer functions for non-CT stimuli to be given by

$$\frac{\phi_e(\mathbf{k}, \omega)}{\phi_N(\mathbf{k}, \omega)} = G_{es} L(\omega) (D_e [1 - G_{ii} L(\omega)] - G_{ee} [1 + \Psi(\omega) \tau(\omega)] L(\omega))^{-1}, \quad (15)$$

and the analog of (9).

The dispersion relation of waves in our model system is given by setting the denominator of (15) to zero, giving

$$k^2 + q^2(\omega) = 0, \quad (16)$$

$$q^2(\omega) r_e^2 = \left(1 - \frac{i\omega}{\gamma_e}\right)^2 - \frac{G_{ee} [1 + \Psi(\omega) \tau(\omega)] L(\omega)}{1 - G_{ii} L(\omega)}. \quad (17)$$

In (11)–(17) linearization is implicitly carried out relative to steady states of the combined corticothalamic system, not the cortex alone. The steady state values in Table I refer to the combined system.

III. ELECTROENCEPHALOGRAPHIC SPECTRA

In this section we derive the form of the EEG spectrum from the transfer function (15). We then derive analytic forms for quantities such as the $\omega=0$ stability boundaries of the system, the asymptotic forms of the spectrum at high and low frequencies, and breakpoints between different ranges. We also examine the effects on the spectrum of filtering through the cerebrospinal fluid, skull, and scalp.

A. Filtering by overlying structures

In calculating scalp EEG spectra (rather than intracranial ones at the cortical surface), one must consider the possibility of filtering due to shielding that results from volume conduction by the cerebrospinal fluid, skull, and the scalp itself [10,20,40]. Filtering of high spatial frequencies k can be significant, as found by Srinivasan *et al.* [40]. The k dependence of their low-to-moderate wave number results is reasonably well fitted by a spatial filter of the form

$$F(k) = e^{-k^2/k_0^2}, \quad (18)$$

where $F(k)$ is the square of the ratio of scalp to cortical voltage and $k_0 \approx 30 \text{ m}^{-1}$. We show in Appendix B that the precise form of this function does not significantly affect the form of the final spectra calculated, so we use the form (18) because it allows the spectrum to be evaluated in a convenient closed form.

B. Spectra

EEGs result from fields generated as charges flow in and out of neural membranes as neurons fire and the resulting signals propagate. Hence, we assume that the observed signals primarily result from the firing of excitatory neurons, since these are larger and much more numerous than inhibitory ones, and have a higher degree of spatial alignment, which implies they contribute more coherently to the total signal [10]. Further reasons for ignoring ϕ_i are that its transfer function is proportional to the one for ϕ_e at small k (thereby giving only a multiplicative correction), while volume conduction filters out both signals at large k [10,40]. We see below that the bulk of the EEG power is at large scales where both fields have the same spectral profile in any case.

Stimuli to the cortex have complicated temporal and spatial dependences. We approximate these here by assuming that their total ϕ_N consists of white noise in space and time, which is effectively filtered via its interaction with the cortex to yield the EEG spectrum. This picture is consistent with recent results that indicate that at least the alpha frequency range is consistent with linearly filtered noise in most cases [7,41], although in 1.25% of cases possible evidence of non-linear behavior was found.

The resulting spectrum is

$$P(\omega) = \int |\phi_e(\mathbf{k}, \omega)|^2 F(k) d^2\mathbf{k} \quad (19)$$

$$= \frac{|\phi_N|^2}{r_e^4} \left| \frac{G_{es}L(\omega)}{1 - G_{ii}L(\omega)} \right|^2 \int \frac{F(k) d^2\mathbf{k}}{|k^2 + q^2|^2} \quad (20)$$

$$= P_N \left| \frac{L(\omega)}{1 - G_{ii}L(\omega)} \right|^2 \frac{1}{|q^2 r_e^2| \sin \theta} \times \text{Im}[\exp(q^{*2}/k_0^2) E_1(q^{*2}/k_0^2)], \quad (21)$$

$$P_N = \pi |\phi_N|^2 G_{es}^2 / r_e^2, \quad (22)$$

where θ is the complex argument of q^2 , $|\phi_N|^2$ is the white-noise power level in Fourier space, and E_1 is the exponential integral function [42], with

$$E_1(z) = -\gamma - \ln z - \sum_{j=1}^{\infty} \frac{(-z)^j}{j j!}, \quad (23)$$

for $|\arg z| < \pi$, where $\gamma = 0.5772 \dots$ is Euler's constant. The spectrum (21) has been calculated for the specific form (18) of $F(k)$.

The limit $k_0 \rightarrow \infty$ corresponds to the absence of volume conduction and, in the related limit $|q^2|/k_0^2 \rightarrow 0$, we find that (21) simplifies to

$$P(\omega) = P_N \left| \frac{L(\omega)}{1 - G_{ii}L(\omega)} \right|^2 \frac{\theta}{|q^2 r_e^2| \sin \theta}. \quad (24)$$

In the opposite limit, $|q^2|/k_0^2 \gg 1$, expansion of (21) in powers of k_0^2/q^2 yields

$$P(\omega) = P_{\infty}(\omega) \frac{k_0^2 \sin \theta}{|q^2| \theta}, \quad (25)$$

where $P_{\infty}(\omega)$ is the spectrum given by (24). At large frequencies, where $|q^2|$ becomes large and $|\theta| \approx \pi$, this implies an asymptotic frequency filtering function

$$F(\omega) \approx \omega_F^3 / \omega^3, \quad (26)$$

with a turnover frequency of

$$\omega_F \approx \gamma_e (2k_0^2 r_e^2 / \pi)^{1/3}. \quad (27)$$

For the parameters in Table I, the turnover frequency occurs at about 30 Hz—too high to significantly affect the main part of the frequency spectrum. Equation (26) does not have the form $F(\omega) = \exp(-\omega^2 v^2 / k_0^2)$ one might naively expect from (18) for a dispersion relation whose real part has the asymptotic form $\omega = kv$. This difference occurs because of damping, which implies that ω is the solution of the *complex* dispersion relation (16). Hence, the Fourier transform of waves satisfying this relation has a Lorentzian profile in the real part of ω , giving $F(\omega) \propto \omega^{-2}$; the actual ω^{-3} dependence comes about because of the detailed structure of (21), which causes the coefficient of the ω^{-2} term in an expansion in powers of ω^{-1} to be zero.

The shape of the spectrum depends strongly on the locus of q^2 in the complex plane, with instability occurring if this locus crosses the negative real axis, which corresponds to q itself acquiring a negative real part, consistent with earlier

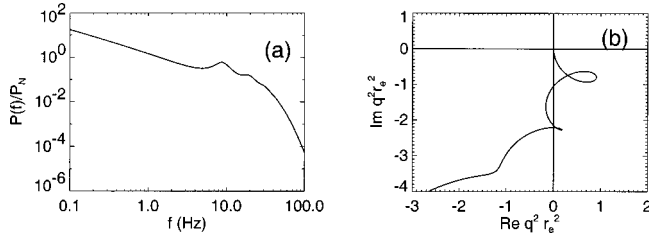


FIG. 3. Theoretical EEG power spectrum $P(f)/P_N$ and $q^2 r_e^2$. Parameters are $\gamma_e = 110 \text{ s}^{-1}$, $\alpha = 70 \text{ s}^{-1}$, $\beta/\alpha = 4$, $t_0 = 0.07 \text{ s}$, $G_{ii} = -1$, $G_{ee} = 1$, $\eta_1 = \eta_2 = 150 \text{ s}^{-1}$, $n = 1$, $\psi = 1$, $\psi' = 0$, and $k_0 = 30 \text{ m}^{-1}$, implying $S = 0$. (a) Spectrum $P(f)/P_N$. (b) Locus of $q^2 r_e^2$ for $\omega > 0$.

work [21,22]. Stability at $\omega = 0$ corresponds to $\text{Re}[q^2(0)] > 0$, $\text{Im}[q^2(0)] = 0$, and $\theta(0) = 0$, implying that we can define a stability parameter S

$$S = 1 - \frac{G_{ee}(1 + \psi)}{1 - G_{ii}}, \quad (28)$$

that must satisfy $S \geq 0$ for the system to avoid instability at $\omega = 0$ (this is a necessary, but not sufficient, condition for overall stability).

C. Illustrative spectrum

Figure 3(a) shows an illustrative spectrum obtained from (21) for parameters typical of those in Table I. Alpha and beta peaks (plus weak, higher-frequency ones), are seen near 8 Hz and 22 Hz, respectively, superposed on a smooth underlying spectrum which has approximately f^{-5} behavior at large f , f^{-1} behavior for $0.4 \text{ Hz} \lesssim f \lesssim 3 \text{ Hz}$, and levels off as $f \rightarrow 0$. Figure 3(b) shows the corresponding locus of $q^2(\omega)$ for $\omega > 0$. After starting near the origin at $f = 0$, q^2 increases in modulus, with $q^2(\omega) \propto \omega$, where $P(f) \propto f^{-1}$. Feedback then causes the curve to loop back toward the origin several times, producing correlated alpha, beta, and higher frequency enhancements. Finally, the locus recedes along an asymptotically parabolic path in the regime with $P(f) \propto f^{-5}$ until filtering via volume conduction cuts in, and $P(f) \propto f^{-8}$ thereafter, although the latter regime is not fully attained even at 100 Hz. The onset of dendritic filtering produces the noticeable knee in the spectrum at $\omega \approx \alpha$, thereby restricting the potential range of the f^{-1} spectrum to $\omega \lesssim \alpha$.

Momentarily ignoring $F(k)$, the integrand in (20) peaks where $k^2 = \max\{0, -\text{Re } q^2\}$, and has a characteristic width of $|\text{Im } q^2|$. Hence, in a cortex of characteristic linear size l , the mode number M at which the integrand peaks is

$$M = \frac{l}{2\pi r_e} [\max\{0, -\text{Re}(q^2 r_e^2)\}]^{1/2}, \quad (29)$$

and the characteristic width of the integrand in M is of order $(l/2\pi r_e)[\text{Im}(q^2 r_e^2)]^{1/2}$. Figure 3(b) thus shows that M is small up to at least the beta frequency. This implies that EEGs in this regime are spatially large scale, thereby reduc-

ing spectral distortions that would otherwise occur due to the effects of volume conduction [10,40]. This point is explored further in Sec. IV.

D. Features of the underlying spectrum

In this section we temporarily ignore the spectral peaks (which are discussed in Sec. V) and examine the properties of the smooth, underlying spectrum. At very low frequencies (21) implies $P(\omega) \approx P(0)$, with $\theta(\omega) \approx 0$ and, hence,

$$P(0) = \frac{P_N}{G_0^2 S}, \quad (30)$$

$$G_0 = 1 - G_{ii}, \quad (31)$$

provided the system is stable ($S > 0$) and $|q^2| \ll k_0^2$. This prediction is in accord with the physical expectation that a small stability parameter will be associated with high EEG power. In this regime, only S and G_0 affect the power for constant P_N ; none of the feedback parameters affect the power level.

If S is small, $q^2(0) \approx 0$ and the behavior of $P(\omega)$ at small ω depends on the leading terms in the expansion of $q^2(\omega)$ in powers of ω . If we write

$$q^2(\omega) r_e^2 = \sum_{j=0}^{\infty} A_j (-i\omega)^j, \quad (32)$$

the structure of (17) implies that all the A_j are real. Explicit expressions for A_0 and A_1 are found in Appendix C.

The case $A_0 = 0$ corresponds to $S = 0$ (see Appendix C) and, hence, to marginal stability, a condition that favors complex dynamics, such as actually seen in EEG recordings [41]. If A_0 is small and $A_1 \neq 0$, there is a small- ω regime in which $q^2(\omega) \approx -i\omega A_1$. In this case, one finds $\theta \approx \pm \pi/2$ for $\omega > 0$ and

$$P(\omega) \approx \frac{P_N}{G_0^2} \frac{\pi}{2\omega |A_1|}. \quad (33)$$

Such a regime with $P(\omega) \propto \omega^{-1}$ was recently observed in magnetoencephalographic observations [16]. Significantly, all the model parameters other than k_0 affect the behavior in this regime, via A_1 .

If A_0 and A_1 are both very small, the $j=2$ term in (32) will dominate in magnitude over the $j=3$ term at small ω unless $A_2 \approx 0$, while the $j=3$ term dominates the imaginary part of q^2 . Assuming that A_2 is not too close to zero, two cases exist: if $A_2 > 0$, $\theta \approx \pm \pi$ and

$$P(\omega) = \frac{P_N}{G_0^2} \frac{\pi}{|A_3| \omega^3}. \quad (34)$$

If $A_2 < 0$, $\theta \approx \pi A_3/A_2$, and

$$P(\omega) = \frac{P_N}{G_0^2} \frac{1}{|A_2| \omega^2}. \quad (35)$$

Regimes exist with these behaviors provided $A_0 \approx 0$ (which implies $S \approx 0$), and $A_1 \approx 0$. Higher order cancellations of the first few terms in (32) may occur, but such occurrences represent increasingly smaller zones of parameter space, so we do not consider them in this paper. If $A_2 > 0$ the condition $A_1 = 0$ defines a stability boundary of the system, since q^2 crosses the negative imaginary axis infinitesimally close to the origin at $\omega = 0$ in this case.

In applying (34) and (35), the condition $S = 0$ implies an interrelationship between G_{ee} , G_{ii} , and ψ , via (28). Likewise, the condition $A_1 = 0$ at $S = 0$ requires

$$\psi \left(\frac{2}{\gamma_e} + \frac{2}{\alpha G_0} + t_0 + \frac{n}{\eta_1} + \frac{n}{\eta_2} \right) - \psi' t_0 = - \left(\frac{2}{\gamma_e} + \frac{2}{\alpha G_0} \right), \quad (36)$$

where Eq. (C2) of Appendix C has been used.

At high frequencies with $\omega > \alpha$ the term $(1 - i\omega/\gamma_e)^2$ on the right-hand side of (17) dominates, $\theta \approx -\pi$ for $\omega > 0$, $|q^2 r_e^2| \sin \theta \approx -2\gamma_e/\omega$, and $|L(\omega)|^2 \approx \alpha^2 \beta^2 / \omega^4$. If volume conduction is not important, these scalings give

$$P(\omega) = \frac{P_N \pi \alpha^2 \beta^2 \gamma_e}{2\omega^5}, \quad (37)$$

in this range, provided $|G_{ii}L(\omega)| \lesssim 1$ [22]. In this regime, the power level is most sensitive to α and β , with a weaker dependence on γ_e , while the feedback parameters have no effect. If volume conduction is strong ($\omega \gg \omega_F$), the filtering factor in (25) must be included, steepening the spectrum to an ω^{-8} dependence. In addition, the existence of an ω^{-1} regime requires $|G_{ii}L(\omega)| \gtrsim 1$. Hence,

$$1 + \omega_{\max}^2 / \beta^2 < |G_{ii}| \lesssim 1 + \omega_{\min}^2 / \beta^2, \quad (38)$$

where ω_{\min} is the minimum frequency of any observed ω^{-5} regime and ω_{\max} is the maximum frequency of any observed ω^{-1} regime.

IV. PARAMETER SENSITIVITIES AND SIMPLIFIED MODEL

In this section we examine how the spectrum (21) varies with each of the model parameters γ_e , α , β , G_{ii} , G_{ee} , ψ , ψ' , t_0 , n , η_m (with $\eta_1 = \eta_2$ assumed for convenience), and k_0 . Except for the parameter being varied at any given time, the parameters are identical to those of Fig. 3 throughout this section and P_N is fixed. We term the values in Fig. 3 “nominal” for convenience. Possible correlations between changes in parameters due to their common dependence on steady state firing rates, for example, are not included here; only one parameter at a time is varied.

Toward the end of this section, we argue that several of the model’s parameters can be absorbed into one another for most purposes, leaving a simplified model with fewer parameters for fitting to the shapes of experimental spectra. We also clarify the *a priori* physiological constraints on these parameters. Further simplifications are made in Sec. VI, where the number of parameters is further reduced by one in the case of marginally stable spectra.

Figure 4(a) shows the effect of varying γ_e from 70 to 150 s^{-1} (nominal value: 110 s^{-1}). At high f , $P \propto \gamma_e$ is satisfied in accord with (37), whereas there is little effect at low f , as expected from (30).

Figure 4(b) shows the effect of varying α from 40 to 100 s^{-1} (nominal value: 70 s^{-1}), with $\beta/\alpha = 4$ fixed. For $f \gtrsim \alpha$, $P \propto \alpha^4$, in accord with the effects of dendritic filtering incorporated in (37). However, dependence on α (via A_1) is weak in the f^{-1} regime, in agreement with (33). The corresponding loci of $q^2 r_e^2$ have loops that expand and approach the origin as α increases, producing the enhanced peaks seen—for a strong alpha peak to be seen at a frequency ω_α , one must have $\alpha \gtrsim \omega_\alpha$. The effects of increasing the ratio β/α from 1 to 7 (nominal value: 4), seen in Fig. 4(c), are quite similar to those of increasing α itself, in the frequency ratio of interest—higher β/α corresponds to less dendritic filtering, more clearly visible spectral peaks, and a shallower (as ω^{-3} rather than ω^{-5}) fall-off at frequencies $\alpha \lesssim \omega \lesssim \beta$, before the onset of an ω^{-5} tail for $\omega \gtrsim \beta$.

Figure 4(d) shows the effect of varying G_{ii} from -0.5 to -1.5 (nominal value: -1). For $G_{ii} = -0.5$, $S \leq 0$ is satisfied, leading to a $1/f$ spectrum, with instability at $f = 0$ in the first case (q^2 intersects the negative real axis at $\omega = 0$), and marginal stability in the second. For $G_{ii} = -1.5$, $S > 0$ holds, and a plateau occurs at low f , as in (30). Below a few Hz, P decreases as G_0 increases in all cases, in accord with (30) and (33). The main effect at higher frequencies is that the alpha and beta resonances weaken with increasing G_0 , a reflection of contracted loops of q^2 in the complex plane. Similar effects are seen in Fig. 4(e), in which G_{ee} is varied from 0.5 to 1.5 (nominal value: 1). This leads primarily to a steepening of the spectrum at low f due to the onset of instability at $f = 0$ for $G_{ee} > 1$, plus some sharpening of the spectral peaks owing to expansion of the q^2 loops.

Figure 4(f) shows the effect of varying the direct corticothalamic feedback parameter ψ from -0.5 to 2.5 (nominal value: 1). Larger $|\psi|$ increases the magnitudes of the spectral peaks, and the low- f component of the spectrum as $f \rightarrow 0$ instability is approached, while negative ψ results in an inversion of the peaks and troughs, with new peaks appearing midway between the original ones, which are replaced by troughs; the new peaks are still weak for $\psi = -0.5$. Note that there is a $1/f$ low-frequency component for $\psi = 0, 2.5$, whereas there is a plateau for -0.5 because $S > 0$ in this case. Variation of the final gain-related parameter ψ' from -0.6 to 0.6 (nominal value: 0) is explored in Fig. 4(g). As $|\psi'|$ increases the peaks sharpen and become more pronounced. As ψ' increases there is an increase in the frequencies of the alpha and beta peaks, with a corresponding reduction in their frequency ratio from just below 3 at large negative ψ' to just above 2 at large positive ψ' , a phenomenon that will be further examined in Sec. V. Note that the second q^2 loop is larger than the first in this case, because the magnitude of the differential feedback term peaks for nonzero ω . Finally, variations in ψ' produce shifts in the normalization of the f^{-1} range, as expected from (33). Although (34) and (36) imply that an f^{-3} regime should be attainable for suitable ψ' , an instability at nonzero f inter-

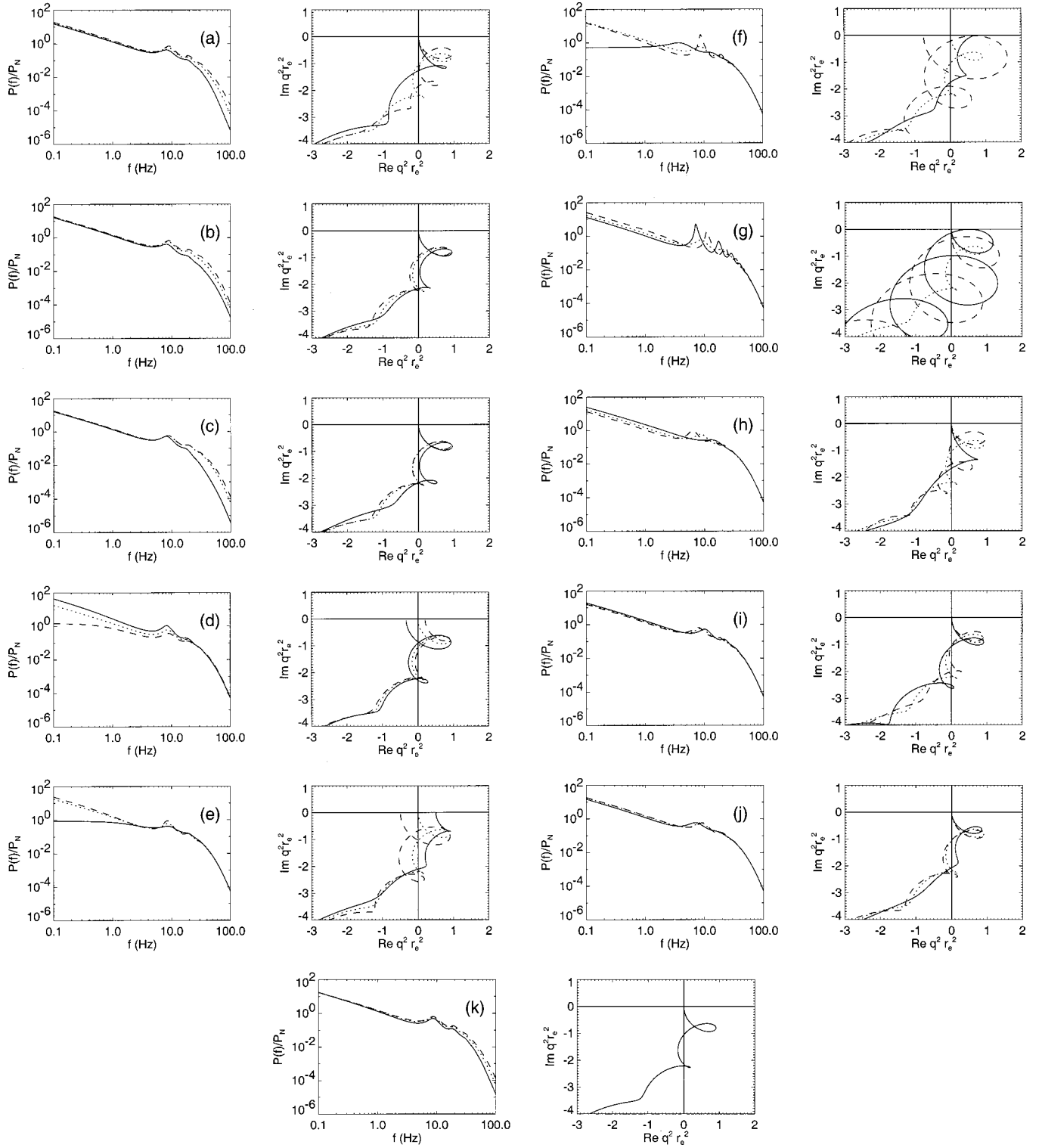


FIG. 4. Dependence of spectra $P(f)/P_N$ and $q^2 r_e^2$ on individual parameters, with other parameters fixed at the values used in Fig. 3. Solid, dotted, and dashed curves are used in order of increasing magnitude of the parameter being varied. In each of the rows (a)–(k) the left frame shows $P(f)/P_N$ and the right shows $q^2 r_e^2$ for $f > 0$. In each case one of the curves is the same as that in the corresponding frame of Fig. 3. (a) $\gamma_e = 70, 110, 150 \text{ s}^{-1}$, (b) $\alpha = 40, 70, 110 \text{ s}^{-1}$, (c) $\beta/\alpha = 1, 4, 7$, (d) $G_{ii} = -0.5, -1.0, -1.5$, (e) $G_{ee} = 0.5, 1.0, 1.5$, (f) $\psi = -0.5, 1, 2.5$, (g) $\psi' = -0.6, 0, 0.6$, (h) $t_0 = 0.03, 0.07, 0.11 \text{ s}$, (i) $n = 0, 1, 2$, (j) $\eta_1, \eta_2 = 50, 150, 250 \text{ s}^{-1}$, (k) $k_0 = 20, 30, 40 \text{ m}^{-1}$.

venes first for the parameters considered when $|\psi'|$ is further increased. We consider this point further in Sec. VI.

The main effects of increasing the time delay t_0 in the corticothalamic loop are to reduce the alpha and beta frequencies, and sharpen the peaks, as shown in Fig. 4(h), the latter effect due to there being less thalamic filtering at low frequencies. Similar effects are seen with changes in n , η_1 , and η_2 , as seen in Figs 4(i) and 4(j). There are also weak effects on the normalization of the f^{-1} regime, as expected from (33). These four parameters leave the high- f part of the spectrum essentially unchanged.

Figure 4(k) shows the effect of varying the characteristic wave number k_0 above which volume conduction filters out cortical signals. There is little effect on the shape of the spectrum below about 25 Hz, but a substantial steepening above this point for small k_0 , in agreement with the semi-quantitative estimate (26).

In summary, Fig. 4 shows that for fixed P_N only α , β , γ_e , and k_0 significantly affect the high frequency spectrum, in accord with (26) and (37). The lowest frequencies are only affected by G_{ii} , G_{ee} , and ψ , via their effects on S , which occurs in (30). The f^{-1} spectrum is affected moderately by all parameters, with its existence depending most strongly on the value of S , which replaces it with a low- f plateau if it is significantly greater than zero. Large values of α , β , G_{ii} , $|\psi|$, $|\psi'|$, η_1 , and η_2 tend to sharpen the peaks, while t_0 , η , n , ψ , and ψ' are the chief determinants of the peak frequencies.

Having examined the variation of the predicted spectrum with each of our model's parameters. We note that when fitting experimental data, a model with fewer parameters would tend to give fits that were simpler to interpret and of higher statistical significance. Published EEG data usually span only the range 1–40 Hz or thereabout, sometimes as great a range as 0.1–100 Hz, or as little as 2–30 Hz. Figure 4 shows that these ranges are unlikely to be sufficient to distinguish all the model parameters unambiguously, particularly once the effects of noise are included. Hence, there is little point in distinguishing all these parameters theoretically. Following this argument, we note that increases in β have effects similar to increasing α in the frequency range of interest, so these two parameters can be replaced by a common effective value of α for most purposes, giving $L(\omega) = (1 - i\omega/\alpha)^{-2}$ in (4). Similarly, the effects of n , η_1 , and η_2 are quite similar to those of t_0 provided n is nonzero (otherwise the differential feedback becomes unphysically large in magnitude at high frequencies). Hence, we can approximate these parameters by setting $n=1$ and $\eta_1 = \eta_2 = \alpha$, which implies $\tau(\omega) = e^{i\omega t_0} L(\omega)$ in (14). The remaining model has only seven parameters (rather than 11), aside from the overall normalization of the spectrum. In Sec. VI, the likely restriction to $S \approx 0$ reduces this to only six parameters aside from normalization. At the cortical surface, k_0 is also irrelevant, leaving one fewer parameter still. There is also the possibility that the dependence of scalp spectra on k_0 will not be experimentally distinguishable from the effects due to γ_e and α .

When fitting to experimental data, we emphasize that the

normalization is set by the observed root-mean-square (rms) signal, while the other parameters are not completely free, since physiology constrains them, in some cases quite strongly (see Table I). For example, the characteristic range r_e of excitatory axons has been measured, as have axonal signal velocities v . These measurements constrain γ_e to within roughly $\pm 40\%$ of its best estimate of 110 s^{-1} . Likewise, the dendritic rate constant α is known to be within roughly a factor of two of 100 s^{-1} , with a major uncertainty arising from the difficulty of assigning a single rate to describe all the different neural and synaptic types. Similarly, $500 \text{ s}^{-1} \geq \beta$ is probably satisfied, and $\beta \geq \alpha$ can be assumed without loss of generality (in the simplified version of our model we set $\beta = \alpha$ and the resulting effective value of α then lies somewhere between 50 and 500 s^{-1}).

Although their signs are strictly known, the values of G_{ee} and G_{ii} are less well constrained by physiology, with typical magnitudes estimated to lie between roughly 1 and 100, based on the values in Table I. The ratio $|G_{ee}/G_{ii}|$ is somewhat better constrained, with a value of 1–5 times the ratio $Q_e^{(0)}/Q_i^{(0)} \sim 1$ of the steady-state firing rates in the two neural populations [see (1)]. Since normal EEG and magnetoencephalographic [16] spectra peak strongly at low frequencies, S must be small and non-negative to ensure a significant f^{-1} , f^{-2} , or f^{-3} spectral range while avoiding instability. Likewise, if $S \approx 0$, $A_1 > 0$ is required whenever $A_2 > 0$ to avoid a low- f instability—a condition that restricts combinations of parameters via (36) and via an analogous expression for the locus on which $A_2 = 0$.

The magnitude of Ψ must be of order unity to be consistent with physiological knowledge that the thalamus has strong, but not overwhelming, effects on the cortex and its state of attention, and that the cortex likewise projects strongly to the thalamus [37–39]. Thus, although corticothalamic connections (in both directions) are much less numerous than corticocortical ones, these connections must be more effective in stimulating the cortex than their numbers imply—e.g., as a result of larger s_b values than for other fibers, or because they have more axonal synapses than average. Hence, $|\psi|$ and/or $|\psi' \alpha t_0|$ are probably of order unity in some states, the latter being the maximal magnitude of the differential feedback term $\omega t_0 \psi' \tau(\omega)$ if $\alpha \sim \beta, \eta_1, \eta_2$.

The value of the delay t_0 (including corticothalamic dendritic delays) must be of order 25–100 ms, since the latency for signals to reach the cortex from the thalamus is $\sim 25 \text{ ms}$ [8] and there is presumably a similar delay in the reverse direction, plus some shorter intrathalamic delays. Although the factors G_{es}^2 and $|\phi_N^2|$ in the normalization (22) are not separately observable, they do not constitute free parameters of the spectral shape, and variations of the magnitude of the normalization $G_{es}^2 |\phi_N^2|$ can be determined experimentally. Finally, the volume conduction turnover wave number k_0 can be estimated from the electrical conductivities and thicknesses of the various tissues overlying the cortex [40], giving $k_0 \approx 30 \text{ m}^{-1}$.

V. SPECTRAL PEAKS

In this section we turn our attention to the EEG spectral peaks, which are of central importance to practical electro-

encephalography, where they are often termed *rhythms*. We use the simplified model introduced in the preceding section to avoid an unnecessarily complicated discussion. The effects of spatial filtering are also ignored, since the preceding section implies that these are only significant for frequencies above those of the major rhythms. We estimate the locations of the peaks and the resulting frequency ratios, the sharpness of the peaks, and the conditions for onset of instabilities. In doing this, we unify all the peaks observed into a single theoretical framework in which each peak is distinguished by a unique number and a theoretical basis is obtained for the customary subdivision of the spectrum into bands.

A. Frequencies and occurrence of spectral peaks

Some numerical experimentation shows that the factor in (24) that dominates in determining the frequencies of spectral peaks is the variation of $\text{Im } q^2 r_e^2 = |q^2 r_e^2| \sin \theta$. We find that the dominant behavior can be approximated in the limits $\omega^2 \ll \alpha^2$ and $\omega^2 \gg \alpha^2$ by

$$\text{Im } q^2 r_e^2 = -\frac{2x}{\gamma_e t_0} - B(\omega) G_{ee}[\psi \sin x - x \psi' \cos x], \quad (39)$$

$$B(\omega) \approx 1/G_0, \quad \omega^2 \ll \alpha^2 \quad (40)$$

$$\sim \alpha^4/\omega^4, \quad \alpha^2 \ll \omega^2, \quad (41)$$

$$x = \omega t_0. \quad (42)$$

Since the first term on the right-hand side of (39) is negative (for positive x , which is assumed henceforth), peaks of $P(\omega)$ occur approximately where the term in square brackets has its maximum negative value of $-|\Psi|$. The factor $B(\omega)$ in (39) and the factor θ in the numerator in (24) act to downshift the frequencies from those we will shortly predict using the above argument, with this shift increasing for $\omega > \alpha$ where the fractional rate of change of $B(\omega)$ becomes large. However, the present approach is adequate to predict the main trends semiquantitatively, while numerical results easily yield more accurate peak frequencies when required.

We find that the maximum negative values of the term in square brackets in (39) are attained for

$$\sin x = -\psi/|\Psi|, \quad (43)$$

$$\cos x = \psi' x/|\Psi|. \quad (44)$$

For $\psi' = 0$, these equations yield peaks at

$$x_m = (m - \frac{1}{2})\pi, \quad (45)$$

with *peak index* $m = 2, 4, \dots$ for $\psi > 0$ and $m = 1, 3, 5, \dots$ for $\psi < 0$. For $|\psi' x_m/\psi| \gg 1$ we find x_m is given by (45) with the addition of a term $\text{sign}(\psi)\pi/2$. An approximate result, valid for all ψ' , is

$$\omega_m t_0 \approx x_m + \sin^{-1}(\psi' x_m/|\Psi_m|) \text{sign}(\psi), \quad (46)$$

$$|\Psi_m| = (\psi^2 + \psi'^2 x_m^2)^{1/2}, \quad (47)$$

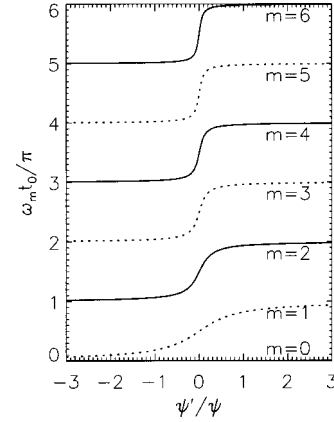


FIG. 5. Peak frequencies predicted from (46) as functions of ψ'/ψ . Solid lines represent peaks that occur for $\psi > 0$ (troughs for $\psi < 0$) while dashed lines represent peaks for $\psi < 0$ (troughs for $\psi > 0$).

with $\text{sign}(u) = 0$ for $u = 0$ here and x_m given by (45). The approximate frequency given by (46) increases monotonically from $(m-1)\pi$ to $m\pi$ as ψ' increases from $-\infty$ to ∞ , as was illustrated in Fig. 4(g). Note that there is a discontinuity in ω_m as a function of ψ' at $\psi' = 0$ in the case $\psi = 0$.

Figure 5 shows the dependence of ω_m on ψ'/ψ for various m , as obtained from (46). The solid and dashed lines correspond to peaks and troughs in the spectrum for $\psi > 0$, and the reverse for $\psi < 0$. On each curve $\omega_m t_0$ passes through a range of π as ψ' increases, with most of the change concentrated in the range $|\psi'| \leq 1/(\omega_m t_0)$. The frequency ratio between successive peaks decreases as ψ' increases, which may enable the ratio ψ'/ψ to be determined experimentally. Typical ratios f_β/f_α of the beta to alpha frequencies are very close to 2 (see Fig. 1, for example). This is consistent with the association $m=2$ for alpha and $m=4$ for beta, if $2\pi\psi'/\psi \gtrsim 1$ and $\psi > 0$ in the waking state, giving a frequency ratio just over 2. It is also consistent with $m=3$ for alpha and $m=5$ for beta, if $2\pi\psi'/\psi \lesssim -1$ and $\psi < 0$, which leads to a frequency ratio just less than 2. (Physiologically, positive feedback with $\psi > 0$ is expected in the waking state, also favoring the first scenario [12].) The sleep spindle peak appears roughly midway between the alpha and beta peaks, as illustrated in Fig. 1, qualitatively consistent with the two scenarios just mentioned, for $m=3$ and $m=4$, respectively. The largest qualitative difference between the two scenarios is that there is another nonzero- f waking peak below the alpha peak in the case $\psi < 0$; however, in practice, this may not be distinguishable from the $1/f$ peak (present in both waking and sleeping states at marginal stability, and to which we assign the value $m=0$ for completeness). It seems that the most promising method of distinguishing the two possibilities is detailed comparison of the ratio f_β/f_α , as determined from the full spectral formula (22), with experiment—if this ratio definitely exceeds 2, one has the simplest possible correspondence between the peak index of observed and predicted peaks, as listed in Table II. In determining this ratio, one must be careful to correct for downshifts due to the decrease of the smooth underlying spectrum, an effect that becomes stronger at high frequencies.

TABLE II. Association between rhythms at $f < 40$ Hz and the resonances discussed in Sec. V. The first column gives the peak index m , the second its designation on the assumption $\psi'/\psi > 1/\pi$, the third its nominal frequency in Hz based on an alpha frequency of 10 Hz, the fourth the corresponding nominal frequency range in Hz of the associated sub-band, and the fifth whether it has a peak in sleeping or waking states. The sixth column gives the peak frequencies in Hz inferred from the data displayed in Figs. 1 and 6 (the nominal ones should be multiplied by 0.95, the ratio of the actual to nominal alpha frequencies, for comparison with these data). The frequency of the delta peak was obtained from Fig. 1. Dashes indicate where no peak was observed in this instance. The conventional bands are delta=0–3.5 Hz, theta=4–7.5 Hz, alpha=8–13 Hz, beta=14–30 Hz, and gamma=above 30 Hz.

| m | Rhythm | f | Band | State | Data |
|-----|---------------|-----|-----------|-------|---------------|
| 0 | delta | 0 | 0–2.5 | both | < 0.2 |
| 1 | theta | 5 | 2.5–7.5 | sleep | 4.6 ± 0.5 |
| 2 | alpha | 10 | 7.5–12.5 | wake | 9.5 ± 0.5 |
| 3 | spindle (sp.) | 15 | 12.5–17.5 | sleep | 14 ± 0.5 |
| 4 | beta | 20 | 17.5–22.5 | wake | 19 ± 1 |
| 5 | beta-2 sp. | 25 | 22.5–27.5 | sleep | 23 ± 1 |
| 6 | beta-2 | 30 | 27.5–32.5 | wake | 26 ± 2 |
| 7 | gamma-1 sp. | 35 | 32.5–37.5 | sleep | 33 ± 2 |
| 8 | gamma-1 | 40 | 37.5–42.5 | wake | |

In either scenario for the ratio ψ'/ψ discussed in the previous paragraphs, the ratio ψ'/ψ has the same sign in both sleeping and waking states, even when ψ changes sign (i.e., ψ and ψ' appear to change sign at roughly the same point). The simplest feedback architecture consistent with this is one in which any thalamic signal inversion occurs at a common structure through which both the direct and differential feedback signals pass. A very likely candidate for such a structure is the thalamic reticular nucleus because there is wide agreement that, in sleep, connections from this nucleus strongly inhibit other thalamic nuclei [12,14,43]. The enhanced inhibitory influence of the thalamic reticular nucleus combined with the existence of an otherwise excitatory feedback loop from the cortex through the thalamic reticular nucleus to other thalamic nuclei and thence back to the cortex, is the reason that both ψ and ψ' might change sign in sleep relative to waking states. A third possible scenario is one in which the even- m resonances alone determine the spectral peaks, with a shift in ψ'/ψ from large positive values in waking to large negative ones in sleep, with the consequent frequency downshift allowing each resonance to account for one waking peak and a sleep peak roughly 5 Hz below it. However, this possibility requires an additional argument as to why the odd- m peaks never seem to contribute. A fourth and final possibility, that suffers from the same problems as the third, is that the opposite trend in ψ'/ψ may occur, allowing each even- m resonance to account for a waking peak and a sleep peak roughly 5 Hz higher in frequency. We discuss the distinction between these scenarios further at the end of Sec. VI.

If we assume the association between m and spectral rhythm given in Table II, we find that the alpha ($m=2$) and

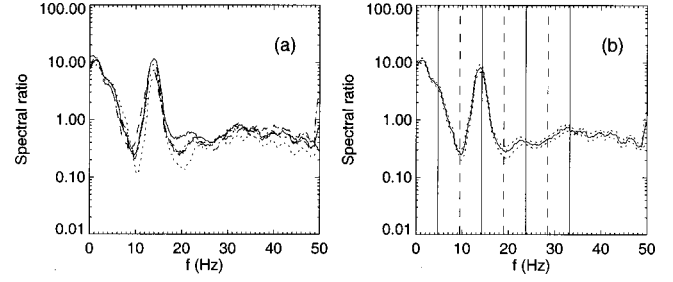


FIG. 6. Ratios of sleep spectra in Fig. 1 to the corresponding waking spectra. The regions at 49–50 Hz should be ignored, as they have been contaminated by the 50 Hz mains frequency. (a) Ratios obtained at Cz (solid curve), Fz (dashed), Pz (dotted), C3 (dashed–dotted pattern), and C4 (dashed–triple-dotted pattern) electrodes. (b) Mean ratio (solid curve) bracketed by the interelectrode standard deviations (dashed curves). Vertical lines are at the first to seventh multiples of 4.75 Hz, the locations of peaks and troughs predicted on the basis of an alpha frequency of 9.5 Hz.

beta ($m=4$) peaks should be positively correlated with each other, and with a new $m=6$ peak at roughly three times the alpha frequency. We term this the *beta-2* peak, a nomenclature that we will justify in the next paragraph. This peak is seen at roughly three times the alpha frequency in Fig. 1, located on the wing of the beta peak, from which it is not well resolved in this example (see the discussion of Fig. 6 below for further analysis). The predicted alpha–beta correlation has been observed in recent work [44]. Similarly, we expect that peaks at $m=1$ (theta or delta rhythm, whose peak may blend smoothly into the $1/f$ peak in many practical situations), $m=3$ (sleep spindles; which we denote by the symbol $\sigma_{\beta 1}$ since they occur in the lower part of what is commonly termed the beta band), and $m=5$ will be positively correlated with each other, and negatively correlated with the alpha, beta, and gamma peaks. Since the latter peaks occur in the waking state, they are obviously anticorrelated with sleep spindles, which are only seen during sleep. Evidence for an anticorrelation between theta and alpha power in the waking state has also been presented recently [45], while theta and delta are well known to be most prominent in deep sleep, especially sleep stage 2 [17,18]. We term the $m=5$ rhythm, which has not previously been discussed, *beta-2 sleep spindles* (symbol $\sigma_{\beta 2}$, since they occupy the upper part of the conventional beta band) in analogy with normal sleep spindles. An enhancement corresponding to this rhythm is seen around 21 Hz in Fig. 1.

The above arguments unify all the rhythms discussed into a single family, with two subfamilies corresponding to waking and sleeping states. This provides a theoretical justification for dividing EEG spectra into bands, but implies that the frequency boundaries of these bands will vary somewhat from individual to individual, with an inverse dependence on t_0 , weaker dependence on ψ'/ψ , and still weaker variation with the other model parameters. Such variations between individuals have been recognized in the literature [18].

The third column of Table II illustrates nominal assignments of frequency band designations based on an alpha frequency of 10 Hz. These correspond quite well with the conventional assignments of the delta, theta, and alpha bands.

The traditional beta band incorporates the spindle, beta, beta-2 spindle, and beta-2 bands of our classification scheme. As conventionally defined, the gamma band extends undifferentiated upward from 30 Hz. In principle, our analysis implies that it can be subdivided into sub-bands of order 5 Hz in width, which we label gamma-1, gamma-1 spindle ($\sigma_{\gamma 1}$), gamma-2, gamma-2 spindle ($\sigma_{\gamma 2}$), etc. However, in practice, the relevant peaks are unlikely to be distinguishable from experimental noise in this range (see the next subsection for details), so it is not unreasonable to label the entire range the gamma band, as is conventional. In the above terminology, the theta rhythm could be labeled the alpha spindle and denoted σ_α , but the old nomenclature is too well established to be likely to make this renaming acceptable.

Figure 6(a) shows the ratios of the pairs of spectra in Figs. 1(a)–1(e). These ratios are remarkably similar at different electrodes, despite the relatively large differences between the actual spectra. In Fig. 6(b) the mean of the five ratios is plotted, bracketed by the interelectrode standard deviation. This ratio oscillates strongly, with an amplitude that decreases rapidly with f . It has peaks at $f \approx 5, 14, 23$, and 33 Hz, as listed in Table II, the last of these being relatively indistinct. Troughs are also seen near 9.5, 19, and 26 Hz. We argue that these peaks and troughs correspond to the theta, spindle, beta-2 spindle, gamma-1 spindle, alpha, beta, and beta-2 rhythms, respectively. The peak at <0.2 Hz, seen in Fig. 1 corresponds to delta rhythm. At high frequencies, the ratios seen in Fig. 6 approach a nearly constant plateau, as expected from (37).

Use of ratios between waking and sleeping spectra removes skull-thickness differences between electrodes, as well as the systematic frequency shifts associated with the shape of the underlying smooth spectrum, and also highlights the differences between the waking and sleeping spectra. The ratios of the frequencies obtained from Figs. 1 and 6 to the alpha frequency are close to integers and half-integers, as predicted from our theory. Figure 6 thus provides evidence of the detection of the $m=5$ (beta-2 spindle) rhythm, with weaker evidence for the $m=6$ (beta-2), and $m=7$ (gamma-1 spindle) rhythms. In future work, these tentative detections will be further tested using EEGs from multiple subjects with longer acquisition times.

B. Sharpness of spectral peaks

The prominence of spectral peaks can be quantified in terms of their Q values (which should not be confused with the firing rates Q_e and Q_i in Sec. II), which are large for strong peaks. We define an effective Q_m to be the ratio of the height of the m th peak to the trough immediately above it in frequency. This gives

$$Q_m \approx \frac{|\text{Im } q_{m+1}^2 r_e^2|}{|\text{Im } q_m^2 r_e^2|} \quad (48)$$

$$\approx \frac{\text{Im}[2x_{m+1} + G_{ee}B(\omega_{m+1})\gamma_e t_0 \Psi_{m+1} e^{i\omega_{m+1}t_0}]}{\text{Im}[2x_m + G_{ee}B(\omega_m)\gamma_e t_0 \Psi_m e^{i\omega_m t_0}]}. \quad (49)$$

Since ω_m and ω_{m+1} correspond to a peak and its neighboring trough, the imaginary parts of $\Psi_m \exp(i\omega_m t_0)$ and $\Psi_{m+1} \exp(i\omega_{m+1} t_0)$ are approximately equal and opposite, with the denominator in (49) being small for a strong peak. Hence, if we approximate x_{m+1} by x_m in the first term in the numerator, and $B(\omega_{m+1})$ by $B(\omega_m)$, we find

$$Q_m \approx \frac{4\omega_m}{G_{ee}B(\omega_m)\gamma_e |\text{Im}[(\Psi_m - \Psi_{res})e^{i\omega_m t_0}]|}, \quad (50)$$

where the exactly resonant value Ψ_{res} is complex and satisfies

$$\text{Im}[\Psi_{res} e^{i\omega_m t}] = 2\omega_m / \gamma_e G_{ee}B(\omega_m). \quad (51)$$

At large and small ω (with some deviations in between), Eq. (50) implies that the m th resonance sharpens as $\Psi \rightarrow \Psi_m$ and that $|\Psi|$ must certainly exceed the modulus of the right-hand side of (51) before resonance can occur. The behavior of $B(\omega)$ given by (40) and (41) implies that high- Q resonances are only possible for $\omega_m \leq \alpha$ and, hence, for low m . The rapid fall-off in Q for high m is seen in the theoretical spectra in Figs. 3 and 4 and in the experimental ones in Figs. 1 and 6. Experimentally, it seems that only the $m=1-4$ resonances can have large Q , consistent with Fig. 6. The rapid decrease of Q_m with m underpins our comments in the preceding section that only the peaks with small m are likely to be discernible in experimental data. It is worth noting that, as was seen in Fig. 6, the $m=1$ solution does not always produce a strong discrete peak; in some cases it blends with the $1/f$ spectrum with no intervening minimum. This behavior is consistent with the preceding discussion, because the latter explicitly ignored the factor θ in the numerator in (24), which cancels the $\sin \theta$ term in the denominator at small θ . This behavior can weaken the theta peak relative to what was implied above.

C. Instabilities

The condition $S \geq 0$ is necessary, but not sufficient, for cortical stability, since it guarantees the absence of instability only at $f=0$, and then only so long as A_1 does not change sign. One possibility for instability at $S=0$ occurs if A_1 changes sign, in which case the spectrum passes through a $1/f^3$ regime, given by (34), followed by an instability that first sets in at $f=0$ then extends to include higher frequencies. Alternatively, it is possible for instability to set in at nonzero frequency if a q^2 loop cuts the negative real axis. Instability at ω_m then sets in approximately when $|\Psi| = |\Psi_m|$, where the spectrum can first develop a singularity. Since $|\Psi_m|$ increases monotonically with m , the first such instability usually sets a stability boundary for the cortical dynamics as a whole. Assuming the simpler association between m and peaks, this boundary is thus set by the $m=2$ (alpha) resonance in the waking state, and the $m=1$ (theta) resonance in sleep (or the $m=3$ resonance if the theta resonance is weakened, as was discussed above).

If an instability sets in, the system will rapidly move to a nonlinear regime in which the linear analysis is invalid.

Hence, in the presence of instabilities, spectra computed from equations in the preceding sections will not be correct.

If the ratio ψ'/ψ is specified, as will be the case in Sec. VI, we can estimate the stability boundaries implied by the above discussion. When $A_1=0$, (36) implies that one must have

$$\psi \left[\frac{2}{\gamma_e} + \frac{2}{\alpha G_0} + \frac{2}{\alpha} + t' \right] > - \left(\frac{2}{\gamma_e} + \frac{2}{\alpha G_0} \right), \quad (52)$$

$$t' = t_0(1 - \psi'/\psi), \quad (53)$$

which implies $\psi > -1$ unless $\psi'/\psi > 1 - 2/\alpha t_0$. The relationship $S=0$ then implies a boundary in the $G_{ee}-G_{ii}$ plane, with the physically allowable region satisfying

$$G_{ii} < \frac{\frac{2}{\gamma_e} + \frac{4}{\alpha} + t' - G_{ee} \left(\frac{2}{\alpha} + t' \right)}{\frac{2}{\gamma_e} + \frac{2}{\alpha} + t'}. \quad (54)$$

This condition is a straight line with slope above -1 unless the denominator in (54) is negative, which occurs for

$$\frac{\psi'}{\psi} > 1 + \frac{2}{t_0} \left(\frac{1}{\gamma_e} + \frac{1}{\alpha} \right). \quad (55)$$

In the latter case, the condition

$$G_{ii} > \frac{\frac{2}{\gamma_e} + \frac{4}{\alpha} + t' - G_{ee} \left(\frac{2}{\alpha} + t' \right)}{\frac{2}{\gamma_e} + \frac{2}{\alpha} + t'}, \quad (56)$$

applies and the slope of the boundary line is more negative than -1 . At large ψ'/ψ , the slope of the boundary is $-[1 + 2/(\gamma_e t_0 \psi'/\psi)]$.

If $|\psi'/\psi|$ is large, one has $|\Psi| \approx |\psi \omega t_0|$ and ψ is forced to be small once ωt_0 exceeds unity, which is the case at all spectral peaks with $m \geq 1$, except at $m=1$ if one also has $\psi'/\psi \leq -1$. The approximate condition for the absence of nonzero- f instabilities of the $m=1$ and 2 rhythms (theta and alpha), $|\Psi| < |\Psi_m|$ then yields

$$1 - G_{ee} \left[1 - \frac{2}{\gamma_e t_0 |\psi'/\psi|} \right] \geq G_{ii} \geq 1 - G_{ee} \left[1 + \frac{2}{\gamma_e t_0 |\psi'/\psi|} \right]. \quad (57)$$

Equation (57) defines lines with negative slopes either side of unity in magnitude. Closer analysis of the instability condition $q^2=0$ implies that the left inequality is imposed by the $m=1$ (theta) instability (or $m=3$ if theta is weakened, a proviso that will be assumed implicitly henceforth), and the right by the $m=2$ (alpha) instability.

If $|\psi'/\psi|$ is sufficiently small, none of the resonances can become unstable. This can be seen by noting that, for $\psi'=0$, Eq. (17) implies that the locus of q^2 is composed of a secular part and an oscillatory part that results from CT feed-

back and gives rise to loops superposed on the secular trend. These loops decrease in radius as $\tau(\omega)$ in (12) decreases in magnitude with frequency, while at zero frequency and marginal stability the radius of the loop is just sufficient for the locus to start at the origin. The only way an instability can occur is for a subsequent loop to reach the origin or the negative real axis, a possibility that is most favored if $|\tau(\omega)|$ stays as large as possible—i.e., for α , β , η_1 , and η_2 as large as possible. If we make these parameters infinite (the optimal case), we find that the locus q_c^2 of the loop centers satisfies

$$|q_c^2 r_e^2|^2 = \frac{\psi^2}{(1+\psi)^2} + \frac{2\omega^2}{\gamma_e^2} + \frac{\omega^4}{\gamma_e^4}, \quad (58)$$

which is monotonic increasing with ω unless $-2 < \psi < -1$, a range that is forbidden by (54), which requires $\psi > -1$ for ψ'/ψ small. Hence, since the loop center moves monotonically away from the origin, and the loop radius is nonincreasing, $q^2=0$ cannot be attained and no instability starting at the origin is possible in this case. Once q_c^2 moves into the left half of the complex plane, the origin is no longer the closest point on the nonpositive real axis and the possibility of an instability arising due to intersection of q^2 with some other point on this axis arises. However, the imaginary part of q_c^2 increases monotonically in magnitude in this regime from a value which is already too large for $q^2=0$ (and, hence, $\text{Im } q^2=0$) to be attained at the point where q_c^2 first crosses into the left half plane. Hence, combining these arguments, no instability is possible for $\psi'=0$.

From the above arguments, we conclude that the physically allowable region of parameter space is bounded below by the higher of the $m=2$ boundary obtained from (57) and the $A_1=0$ boundary (56), if the latter applies in place of (54). It is bounded above by the lower of the $m=1$ boundary from (57) and the $A_1=0$ boundary (54), unless (56) applies as a lower bound and (54) is irrelevant. For small positive ψ'/ψ the $A_1=0$ condition sets the upper bound, but this is supplanted by the $m=1$ boundary at larger values, particularly once (56) applies and the $A_1=0$ boundary becomes a lower bound. There is no lower boundary for very small ψ'/ψ because the $A_1=0$ boundary is an upper bound and no resonances can become unstable. At larger ψ'/ψ the $m=2$ lower boundary cuts in, followed by competition between $A_1=0$ and $m=2$ lower bounds, which both have the same slope at large ψ'/ψ . At small negative ψ'/ψ the $A_1=0$ upper bound is again the only relevant one, with the $m=2$ lower bound appearing at larger negative values. At large negative ψ'/ψ the $m=1$ and $A_1=0$ boundaries compete closely to set the upper bound, since they have the same slopes as $\psi'/\psi \rightarrow \infty$.

We thus find that there are three possible routes to instability from the marginally stable state, apart from S becoming negative: (a) instability starting at $f=0$ due to A_1 changing sign, (b) $m=1$ theta instability at $f \approx 1/2t_0 \approx 5$ Hz when ψ is negative and ψ'/ψ is positive, or lower frequencies if ψ'/ψ is negative, and (c) $m=2$ alpha instability at $f \approx 1/t_0 \approx 10$ Hz when ψ and ψ'/ψ are positive (or at around 5 Hz if ψ is positive and ψ'/ψ is negative). Frequencies in the alpha

range are seen in photic epilepsy and some stages of grand mal seizures, initial frequencies are around 4 Hz in petit mal seizures, and very low frequencies predominate in some stages of grand mal [18]. Hence, we speculate that some of these seizures may represent the nonlinear stages of the instabilities considered here. However, to verify or reject this speculation will require nonlinear simulations of our model, which are beyond the scope of the present paper.

VI. MARGINALLY STABLE EEG SPECTRA

In this section we explore the dependence of spectral structure and instabilities on the gain parameters of our model and impose some restrictions on the possible locus of an arousal sequence that carries the cortex through all the states of arousal from sleep stage 4 to high-vigilance waking states (e.g., while doing complex mental arithmetic under time pressure).

Observed normal EEG and MEG spectra do not appear to exhibit any low- f plateau, except perhaps below about 0.3 Hz (see Fig. 1 and [16]). Equations (30) and (33), and the requirement to avoid $f=0$ instability, then restrict S to very small, non-negative values. Hence, in this section we assume $S=0$ in addition to the other simplifying assumptions made in Secs. IV and V ($\alpha=\beta=\eta_1=\eta_2$ and $n=1$). If we further fix the values of $\alpha=70 \text{ s}^{-1}$, $\gamma_e=110 \text{ s}^{-1}$, and $t_0=0.07 \text{ s}$, we are left with only k_0 and the gain parameters G_{ii} , G_{ee} , ψ , and ψ' . The characteristic wave number k_0 can be ignored here since it does not affect the stability of the cortex. Of the remaining four parameters, ψ can be eliminated via (28) for $S=0$. Physiologically, one requires $G_{ee}>0$ and $G_{ii}<0$, which restricts attention to a quarter of the $G_{ee}-G_{ii}$ plane for fixed ψ'/ψ . We choose the latter ratio as the third free parameter, rather than ψ' itself, since it measures the relative strengths of direct and differential feedbacks.

Figure 7 shows the regions in the $G_{ee}-G_{ii}$ plane in which the model spectrum (21) is marginally stable (white zones) or unstable at some frequencies (gray zones) for $\psi'/\psi=0, \pm 1, \pm 5$; only the white regions represent physically allowable steady states. The locus of $\psi=0$ is shown dotted in each frame and solid contours of the lowest unstable frequency are overplotted in the gray zones.

In Fig. 7(a), $\psi'/\psi=0$ and we see that there is a single zone of instability at upper right, bounded by the $A_1=0$ locus (54). As expected, the instability onset frequency near this boundary is zero. The physically allowable region has no lower bound, consistent with the arguments in Sec. V that no resonant instabilities are possible in this case. Figure 8 displays a sequence of spectra and q^2 loci as the upper instability boundary is crossed, showing the development of a widening f^{-3} regime below about 5 Hz, followed by onset of instability at $f=0$ initially, then moving to higher f , with $f \approx 0.8 \text{ Hz}$ in column (c). In the unstable regime the corresponding q^2 locus loops first up from the origin, then down across the negative real axis very close to the origin. At the boundary, this crossing occurs at the origin.

In Fig. 7(b), $\psi'/\psi=1$ and the stable zone is now bounded above and below by the $m=1$ and $m=2$ instability boundaries, respectively, from (57). The crossover between the

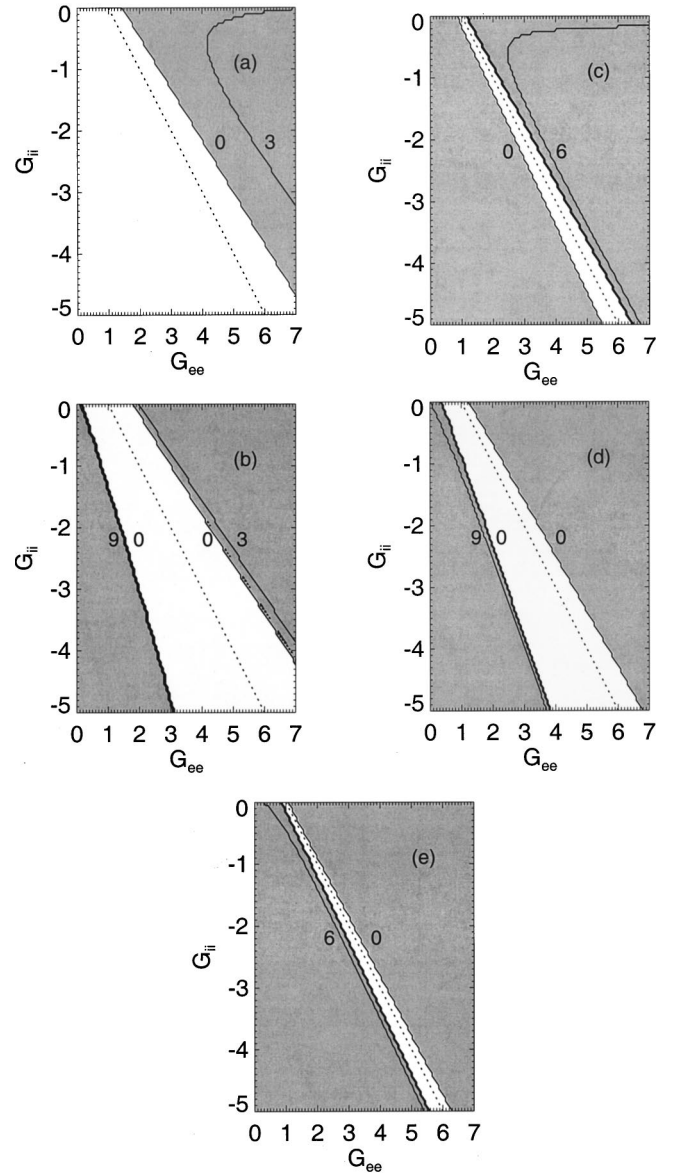


FIG. 7. Zones of instability (gray) in the $G_{ee}-G_{ii}$ plane for $S=0$, $\gamma_e=110 \text{ s}^{-1}$, $\alpha=\beta=\eta_1=\eta_2=70 \text{ s}^{-1}$, $n=1$, and $t_0=0.07 \text{ s}$. The dotted line marks $\psi=0$, while the contours within the gray areas show the minimum unstable frequency in Hz. The irregular appearance of the zone boundaries is due to the finite resolution of the array used in plotting. (a) $\psi'/\psi=0$, (b) $\psi'/\psi=1$, (c) $\psi'/\psi=5$, (d) $\psi'/\psi=-1$, (e) $\psi'/\psi=-5$.

$A_1=0$ and $m=1$ conditions determining the upper bound occurs around $\psi'/\psi=0.6$ for the parameters considered here. The estimate (57) is only approximately correct here because ψ'/ψ is relatively small. Figure 9 displays a sequence of spectra and q^2 loci as the upper instability boundary is crossed, showing that there is no f^{-3} regime in this case; rather the theta peak at around 3.5 Hz sharpens until theta instability sets in. This is reflected in the q^2 locus crossing, then recrossing, the negative real axis in the unstable regime. In this case, the first crossing is near, but not at, the origin, even at the instability boundary. Figure 10 displays spectra and q^2 loci as the lower instability boundary is crossed,

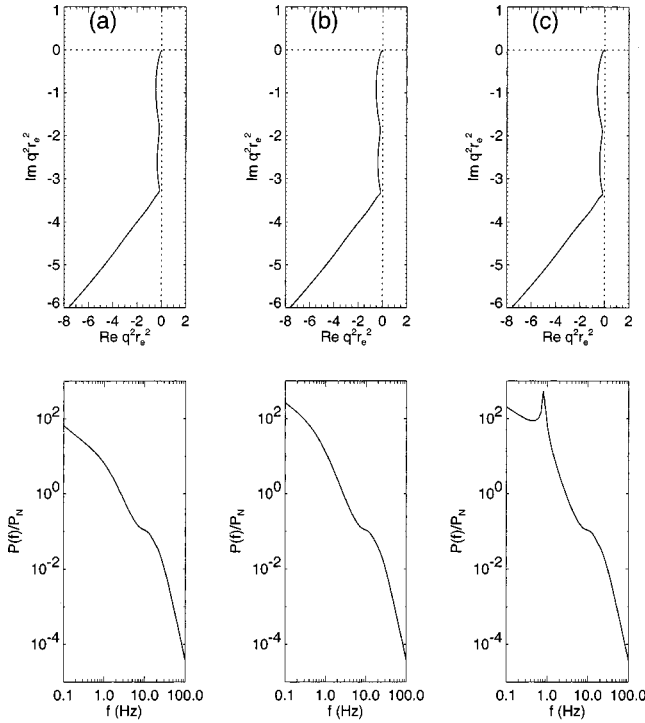


FIG. 8. Loci of $q^2 r_e^2$ (upper frames) and spectra (lower frames) spanning the instability boundary in Fig. 7(a) at $G_{ii} = -2$ and $G_{ee} = 3.7, 3.8, 3.9$ in columns (a)–(c), respectively. In the case of the instability in (c) linear theory breaks down and the shape of the spectrum is no longer reliable.

showing that the alpha peak becomes unstable in this case once the q^2 locus loops right around the origin, crossing the negative real axis in the process. At the instability boundary the first crossing of the negative axis occurs at the origin.

In Fig. 7(c), $\psi'/\psi = 5$. Here the stable zone is bounded above by the $m = 1$ instability boundary, which is reasonably well estimated by (57). The lower bound is set by the $A_1 = 0$ instability, although the $m = 2$ instability sets in at almost the same point.

Figure 7(d) shows the case $\psi'/\psi = -1$, in which the stable zone is bounded above by the $A_1 = 0$ curve, and below by the $m = 2$ instability. This situation persists in Fig. 7(e), in which $\psi'/\psi = -5$.

Taking Fig. 7 as a whole (and also the results presented in Fig. 4), we note that it seems to be very difficult to achieve the conditions required for an ω^{-2} spectrum, given by (35), to manifest itself. This appears to be because A_2 changes sign close to the locus of $\psi = 0$ for physiologically realistic parameters, leaving it with the wrong sign for (35) to apply on the locus where $A_1 = 0$.

Figure 7 allows us to make some comments regarding likely arousal sequences. Specifically, the appearance of a low-frequency f^{-3} spectrum near the $A_1 = 0$ boundary allows us to associate proximity to this boundary with sleep. Hence, unless ψ'/ψ is large and positive, sleep corresponds to negative ψ , and wakefulness to positive ψ , with the reverse association for large positive ψ'/ψ . Since G_{ee} and G_{ii} are both proportional to the mean firing rates of neurons [see the discussion following (10)], one would expect an arousal

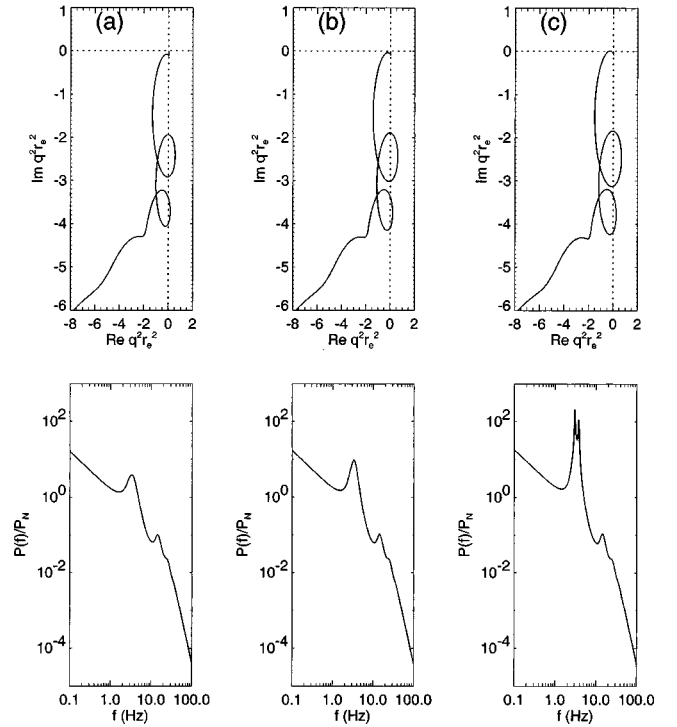


FIG. 9. Loci of $q^2 r_e^2$ (upper frames) and spectra (lower frames) spanning the upper instability boundary in Fig. 7(b) at $G_{ii} = -2$ and $G_{ee} = 4.1, 4.2, 4.3$ in columns (a)–(c), respectively. In the case of the instability in (c) linear theory breaks down and the shape of the spectrum is no longer reliable.

sequence that approximately followed a line passing through the origin in the $G_{ee} - G_{ii}$ plane. The association with $\psi > 0$ in the waking state provides a more natural fit with this expectation, and also accords with the physiological expectation of positive feedback in the waking state. The results presented in Fig. 7 thus support the correspondence between m values and spectral peaks made in Sec. V A.

VII. SUMMARY AND DISCUSSION

In this work we have investigated the spectral properties of a generalized version of our recent model of cortical electrical activity, including a physiologically plausible cortico-thalamic loop with both direct and differential feedbacks, and exploring the effects of volume conduction in tissues overlying the cortex. Differential feedback is included to allow for the possibility that temporal variations are emphasized—likely, since almost all sensory inputs to the cortex first pass through the thalamus, which plays a key role in determining the attentional state of the brain. Unlike many previous models, our model does not presuppose the existence of thalamic pacemakers or “clocks” to regulate the periods of the alpha and beta rhythms. Instead, these emerge directly from the delays in the feedback loop.

One of the main results of this work is the formula (21) for the EEG spectrum that arises from white-noise subcortical inputs, in accord with previous work [7,41], allowing for volume conduction effects. Volume conduction is found to affect the spectrum significantly only above about 25 Hz,

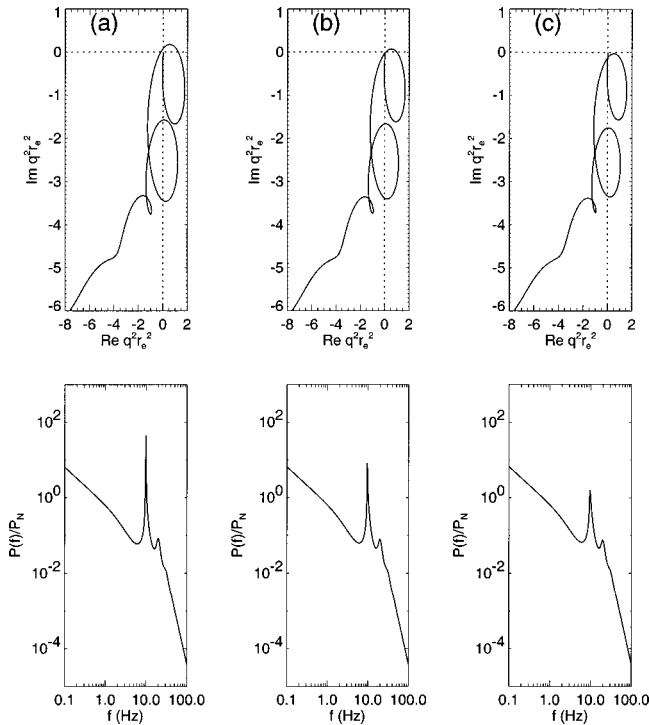


FIG. 10. Loci of $q^2 r_e^2$ (upper frames) and spectra (lower frames) spanning the lower instability boundary in Fig. 7(b) at $G_{ii} = -2$ and $G_{ee} = 1.3, 1.4, 1.5$ in columns (a)–(c), respectively. In the case of the instability in (a) linear theory breaks down and the shape of the spectrum is no longer reliable.

and contributes at most an extra factor of f^{-3} relative to the unfiltered spectrum.

EEG spectra are comprised of a series of peaks superposed on smooth underlying continuum. Equation (21) accounts for the major features of the underlying continuum observed in EEG and MEG experiments, with a low- f $1/f$ or f^{-3} component occurring near marginal stability, and a steepening of the spectrum to f^{-5} (f^{-8} with filtering via volume conduction above about 25 Hz) at frequencies above α , the characteristic dendritic rate constant, where dendritic filtering becomes strong. The observed knee in the spectrum occurs at around 20 Hz, consistent with $\alpha \approx 120 \text{ s}^{-1}$, which is in turn consistent with independent physiological estimates listed in Table I.

Exploration of the parameter sensitivities of the model was carried out in Sec. IV, resulting in the development of a simplified model with fewer parameters than the initial one. This simplified version was used to analyze the frequencies, sharpness, stability, and mutual correlations of spectral peaks predicted by the model. The main results were that each peak can be denoted by a unique “harmonic” number m , with the zeroth peak corresponding to the $1/f$ or $1/f^3$ spectrum, $m = 1$ to theta, $m = 2$ to alpha, $m = 3$ to sleep spindles, and $m = 4$ to beta. The $m = 2, 4, \dots$ peaks are predicted to be positively correlated with one another and prominent in waking. Likewise, the $m = 1, 3, 5, \dots$ peaks should be positively correlated and prominent in sleep, whereas these two sets are anticorrelated overall and the $m = 0$ peak is always present, but stronger in sleep. Comparison with published observa-

tions confirms the predicted correlations, anticorrelations, and states in which certain peaks are expected to be most prominent. The theory is also consistent with the data shown in Figs. 1 and 6, whence we tentatively identify new peaks corresponding to $m = 5$ (beta-2 spindle), $m = 6$ (beta-2), and $m = 7$ (gamma-1 spindle) rhythms. The inferred corticothalamic feedback delay $t_0 \approx 1/f_\alpha \approx 100 \text{ ms}$ is also consistent with physiological constraints on this quantity. We conclude that the simplest association between state of arousal and our parameters is for ψ and ψ' to be positive in waking and negative in sleep, although we cannot yet categorically rule out certain alternatives discussed in Secs. V A and VI.

Our analysis of the spectral peaks leads to the unification of all the major large-scale brain rhythms, and the prediction and tentative confirmation of two new ones. It also provides a theoretical basis for the customary subdivision of the spectrum into bands, but implies that the frequencies of these bands are not universal, but are tied to the alpha frequency of the individual subject, contrary to some conventions, but consistent with the recognition of such variations in the literature [18].

Analysis of the sharpness and stability of the spectral peaks showed that only the first few can have high Q , and that stability of the $m = 1$ (or $m = 3$) and $m = 2$ rhythms can impose bounds on the physically allowable states of the cortex, corresponding to instabilities near 5 Hz (or 15 Hz) and 10 Hz, respectively. Together with the $A_1 > 0$ condition required to avoid instability at $f = 0$, these conditions restrict the range of physically allowable states in parameter space. We speculate that these instabilities may be associated with the linear stages of certain generalized seizures, possibly including grand mal, petit mal, and photic epilepsies. Significantly, an f^{-3} regime is predicted at low f close to the $A_1 = 0$ boundary in sleep. This is consistent with the observed steepening and enhancement of the low- f spectrum in sleep.

The above results indicate that our model provides a promising semiquantitative theoretical explanation for EEG spectra. Most significantly, we explain the entire spectrum in a unified way, a major advance on previous analyses, which have generally concentrated on a single frequency band at a time. In a forthcoming paper we will compare it in detail with observations of multiple subjects in various states of arousal to determine whether it is quantitatively consistent and, in particular, whether the parameters inferred are consistent with those inferred from independent physiological measurements. We hope that these physiological links will prompt experimental physiologists to determine tighter constraints on some of the less well known parameter values.

Our model differs from some in the literature in that, rather than adding numerous parameters to describe ever finer details, it restricts attention to a set of physiologically realistic mechanisms that appear to be sufficient to capture the overall behavior of the EEG, and parametrizes these as simply as possible. This keeps the number of parameters low and their physical significance clear.

One limitation of the present work is that it does not attempt to describe spatial variations of the spectrum, although these are significant in practice (albeit less so when spectral ratios, such as those in Fig. 6 are considered). This is not a

fundamental limitation, as the model can be easily generalized to include such variations if next-order accuracy is required, but at the price of introducing additional parameters. A second limitation is that, although the cortex appears to operate near marginal stability, we have not as yet identified a mechanism to keep it at this point. This does not affect the present discussion, but is an important point of principle. Third, although we argue that corticothalamic feedback is the most important one in determining the observed spectral rhythms, we do not exclude the possibility that other feedback loops with similar time delays may also contribute.

ACKNOWLEDGMENTS

The authors thank K. E. Crowley and I. Colrain for scoring the sleep spectra in Fig. 1.

This work was supported by the Australian Research Council, the National Health and Medical Research Council, and the Ross Trust, Melbourne.

APPENDIX A: BRIEF DESCRIPTION OF THE EXPERIMENT USED TO OBTAIN FIG. 1

In this appendix we briefly outline the experiment used to obtain the data used in Fig. 1, justifying the accuracy of the overall form of the spectrum and the reality of the spectral peaks seen. A fuller description is found in Ref. [6] and more details will be published in a forthcoming work in which a large set of EEG spectra from multiple subjects will be analyzed in detail.

The data in Fig. 1 were obtained from the Cz, Pz, Fz, C3, and C4 electrodes in the International 10–20 system, with linked ears providing the reference potential. The Cz electrode is located at the crown (or *vertex*) of the head, furthest from the reference electrodes and from muscle groups that could generate electrical interference. The Fz and Pz electrodes are located roughly 6 cm in front of and behind this electrode, respectively, while the C3 and C4 electrodes are situated symmetrically ≈ 6 cm either side of it. The data were low-pass filtered to remove frequencies above 50 Hz and sampled at 250 Hz for periods of 2 minutes in waking and 13 minutes in sleep. The spectra were then calculated by Fourier transforming 8 second segments of these data and averaging these transforms.

Narrow band artifacts at the mains frequency of 50 Hz and its subharmonic of 25 Hz were identified due to their presence even in the absence of a subject. As each affected only a single spectral channel, they were simply removed by deleting these channels before plotting the spectra. Ocular artifacts due to eye blinks and pendular eye movements were removed by standard algorithms [46]. When present, muscular artifacts due to jaw, neck, and other movements, were found to contaminate the high-frequency spectrum with noise having an approximately $1/f$ spectrum which is easily identified. Trials with a subset of subjects showed that, with appropriate instructions to minimize movement and relax

jaw muscles, this class of artifact could be suppressed to the point that the high-frequency spectrum was not significantly affected. Cardiac interference at low frequencies was also found to be minimal, with no spectral peak detected at the heartbeat frequency of around 1 Hz, or its harmonics.

The waking spectra in Fig. 1 were for a normal female subject, aged 30, in a relaxed, eyes-closed state, while the sleep spectrum was for the same subject in sleep stage 2.

APPENDIX B: GENERAL EFFECTS OF FILTERING BY OVERLYING STRUCTURES

In this appendix we show that the high frequency effects of filtering are nearly the same for any physically reasonable filter function.

If we assume that $F(k)$ only depends on the magnitude of the wave vector, and falls off rapidly for $k > k_0$, the integral in (20) can be written in the form

$$I = \pi \int_0^\infty dk^2 \frac{F(k^2/k_0^2)}{|k^2 + q^2|^2}. \quad (\text{B1})$$

Upon changing variables to $u = k^2/k_0^2$, we find

$$I = \frac{\pi}{2i \operatorname{Im} q^2} \int_0^\infty du F(u) \left[\frac{1}{u + q^2/k_0^2} - \frac{1}{u + q^{*2}/k_0^2} \right] \quad (\text{B2})$$

$$\sim \frac{\pi}{\operatorname{Im} q^2} \frac{k_0^2 \operatorname{Im} q^2}{|q^4|}, \quad (\text{B3})$$

where (B2) has been expanded in powers of $k_0^2/q^2 u$ and we have used the fact that F falls off rapidly for $k > k_0$. This result is of the same form as the more specialized one, Eq. (26), and shows that the asymptotic ω^{-3} filtering does not depend strongly on the structure of the filter function.

APPENDIX C: EXPLICIT EXPRESSIONS FOR A_0 AND A_1

In this appendix we give explicit expressions for the coefficients A_0 and A_1 in the expansion (32) for $\alpha = \beta$. Expressions for higher A_j are complicated and their exact analytic form is likely to be more strongly model dependent, so we do not reproduce them here.

Direct expansion of (17) yields

$$A_0 = S, \quad (\text{C1})$$

$$A_1 = \frac{2}{\gamma_e} + \frac{G_{ee}}{G_0} \left[\frac{2(1+\psi)}{\alpha G_0} + \left(t_0 + \frac{n}{\eta_1} + \frac{n}{\eta_2} \right) \psi - t_0 \psi' \right], \quad (\text{C2})$$

with S given by (28) and G_0 given by (31).

- [1] M.R. Rosenzweig and A.L. Leiman, *Physiological Psychology* (Random House, New York, 1989).
- [2] M.P. Stryker, *Nature* (London) **338**, 297 (1989).
- [3] W. Singer and C.M. Gray, *Annu. Rev. Neurosci.* **18**, 555 (1995).
- [4] R. Caton, *Br. Med. J.* **1**, 278 (1875).
- [5] H. Berger, *Arch. Psychol. Nervenkr.* **87**, 527 (1929).
- [6] A.R. Haig, E. Gordon, G. Rogers, and J. Anderson, *Electroencephalogr. Clin. Neurophysiol.* **94**, 288 (1995).
- [7] F.H. Lopes da Silva, A. Hoeks, A. Smits, and L.H. Zetterberg, *Kybernetik* **15**, 27 (1974).
- [8] D.L. Robinson, *Int. J. Neurosci.* **22**, 81 (1983).
- [9] G. Buzsáki, *Neuroscience* (Oxford) **41**, 351 (1991).
- [10] P.L. Nunez, in *Neocortical Dynamics and Human EEG Rhythms*, edited by P.L. Nunez (Oxford University Press, Oxford, 1995), Chaps. 1 and 9.
- [11] F.H. Lopes da Silva, *Electroencephalogr. Clin. Neurophysiol.* **79**, 81 (1991).
- [12] A.M.L. Coenen, *Neurosci. Biobehav. Rev.* **19**, 447 (1995).
- [13] F.H. Lopes da Silva, J.E. Vos, J. Mooibroek, and A. van Rotterdam, *Electroencephalogr. Clin. Neurophysiol.* **50**, 449 (1980).
- [14] M. Steriade, P. Gloor, R.R. Llinás, F.H. Lopes da Silva, and M.-M. Mesulam, *Electroencephalogr. Clin. Neurophysiol.* **76**, 481 (1990).
- [15] W.J. Freeman, in *Induced Rhythms of the Brain*, edited by E. Başar and T.H. Bullock (Birkhäuser, Basel, 1991).
- [16] E. Novikov, A. Novikov, D. Shannahoff-Khalsa, B. Schwartz, and J. Wright, *Phys. Rev. E* **56**, R2387 (1997).
- [17] G. Dumermuth, B. Lange, D. Lehmann, C.A. Meier, R. Dinkelmann, and L. Molinari, *Eur. Neurol.* **22**, 322 (1983).
- [18] E. Niedermeyer, in *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 4th ed., edited by E. Niedermeyer and F.H. Lopes da Silva (Williams and Wilkins, Baltimore, 1999), Chaps. 9, 13, and 27.
- [19] H.R. Wilson and J.D. Cowan, *Kybernetik* **13**, 55 (1973).
- [20] P.L. Nunez, *Electric Fields of the Brain* (Oxford University Press, Oxford, 1981).
- [21] P.A. Robinson, C.J. Rennie, and J.J. Wright, *Phys. Rev. E* **56**, 826 (1997).
- [22] P.A. Robinson, J.J. Wright, and C.J. Rennie, *Phys. Rev. E* **57**, 4578 (1998).
- [23] C.J. Rennie, P.A. Robinson, and J.J. Wright, *Phys. Rev. E* **59**, 3320 (1999).
- [24] C.J. Rennie, J.J. Wright, and P.A. Robinson, *J. Theor. Biol.* **205**, 17 (2000).
- [25] A. van Rotterdam, F.H. Lopes da Silva, J. van den Ende, M.A. Viergever, and A.J. Hermans, *Bull. Math. Biol.* **44**, 283 (1982).
- [26] W.J. Freeman, *Mass Action in the Nervous System* (Academic, New York, 1975).
- [27] P.L. Nunez, *IEEE Trans. Biomed. Eng.* **21**, 473 (1974).
- [28] J.J. Wright and D.T.J. Liley, *Network Comput. Neural Syst.* **5**, 191 (1994).
- [29] J.J. Wright, *Biol. Cybern.* **81**, 131 (1999).
- [30] J.J. Wright and D.T.J. Liley, *Behav. Brain Sci.* **19**, 285 (1996).
- [31] V.K. Jirsa and H. Haken, *Phys. Rev. Lett.* **77**, 960 (1996).
- [32] V.K. Jirsa and H. Haken, *Physica D* **99**, 503 (1997).
- [33] P.A. Robinson, C.J. Rennie, J.J. Wright, and P.D. Bourke, *Phys. Rev. E* **58**, 3557 (1998).
- [34] D.T.J. Liley and J.J. Wright, *Network* **5**, 175 (1994).
- [35] G.H. Bishop, *Cold Spring Harbor Symp. Quant. Biol.* **4**, 305 (1936).
- [36] E.R. Kandel, J.H. Schwartz, and T.M. Jessell, *Principles of Neural Science* (Appleton and Lange, Norwalk, Connecticut, 1991).
- [37] V. Braitenberg and A. Shüz, *Anatomy of the Cortex: Statistics and Geometry* (Springer, Berlin, 1991).
- [38] R. Douglas and K. Martin, in *The Synaptic Organization of the Brain*, edited by G. Shepherd (Oxford, New York, 1998), p. 459.
- [39] S.M. Sherman and C. Koch, in *The Synaptic Organization of the Brain*, edited by G. Shepherd (Oxford, New York, 1998), p. 289.
- [40] R. Srinivasan, P.L. Nunez, and R.B. Silberstein, *IEEE Trans. Biomed. Eng.* **45**, 814 (1998).
- [41] C.J. Stam, J.P.M. Pijn, P. Suffczynski, and F.H. Lopes da Silva, *Electroencephalogr. Clin. Neurophysiol.* **110**, 1801 (1999).
- [42] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
- [43] J.A. Gray, *Neural Networks* **10**, 1257 (1997).
- [44] C.L. Lim, R.J. Barry, E. Gordon, A. Sawant, C.J. Rennie, and C. Yiannikas, *Int. J. Psychophysiol.* **21**, 151 (1996).
- [45] W. Klimesch, *Brain Res. Rev.* **29**, 169 (1999).
- [46] G. Gratton, M.G. Coles, and E. Donchin, *Electroencephalogr. Clin. Neurophysiol.* **55**, 468 (1983).