Determination of effective brain connectivity from functional connectivity with application to resting state connectivities

P. A. Robinson
School of Physics, University of Sydney, New South Wales 2006, Australia
Center for Integrative Brain Function, University of Sydney, New South Wales 2006, Australia
Cooperative Research Center for Alertness, Safety, and Productivity, University of Sydney, New South Wales 2006, Australia
Neurosleep, 431 Glebe Point Rd, Glebe, New South Wales 2037, Australia
Center for Integrative Research and Understanding of Sleep, 431 Glebe Pt Rd, Glebe, New South Wales 2037, Australia
and Brain Dynamics Center, Westmead Millennium Institute, Darcy Rd, Westmead, New South Wales 2145, Australia

S. Sarkar
School of Physics, University of Sydney, New South Wales 2006, Australia and Design Lab, Faculty of Architecture, Design, and Planning, University of Sydney, New South Wales 2006, Australia

Grishma Mehta Pandejee and J. A. Henderson
School of Physics, University of Sydney, New South Wales 2006, Australia
(Received 20 May 2013; revised manuscript received 19 February 2014; published 30 July 2014)

Neural field theory insights are used to derive effective brain connectivity matrices from the functional connectivity matrix defined by activity covariances. The symmetric case is exactly solved for a resting state system driven by white noise, in which strengths of connections, often termed effective connectivities, are inferred from functional data; these include strengths of connections that are underestimated or not detected by anatomical imaging. Proximity to criticality is calculated and found to be consistent with estimates obtainable from other methods. Links between anatomical, effective, and functional connectivity and resting state activity are quantified, with applicability to other complex networks. Proof-of-principle results are illustrated using published experimental data on anatomical connectivity and resting state functional connectivity. In particular, it is shown that functional connection matrices can be used to uncover the existence and strength of connections that are missed from anatomical connection matrices, including interhemispheric connections that are difficult to track with techniques such as diffusion spectrum imaging.

DOI: 10.1103/PhysRevE.90.012707 PACS number(s): 87.19.lj, 87.19.ll, 89.75.Fb, 89.75.Hc

I. INTRODUCTION

Complex networks occur in many fields including neuroscience, communications, the internet, scientific collaborations, genetics, ecology, econophysics, statistical mechanics, and economics [1–6], and relationships between their structure and dynamics are central to many applications. Some networks have connectivities that are hard to observe directly, or can only be studied via activity they support. This is especially true if one is interested in *effective* connectivities that incorporate changes in strength of connections that are physically present but may or may not be active. Prominent among such networks are those of the brain, which have anatomical structure that can be observed only imperfectly via techniques like diffusion imaging that can miss connections, especially long-range ones such as those between brain hemispheres [3,4,7], strengths of connectivity (often called effective connectivities, or gains) that vary even for a given anatomical connection, and functional connections set by correlations of network activity itself [2–4]. These issues of function vs structure are highly topical, and are currently being addressed by the US National Institutes of Health Human Connectome Project, the US BRAIN Initiative, and the European Human Brain Project, for example, which recognize the need for a multidisciplinary approach, including physical sciences. Additionally, since the brain provides key examples of dynamical networks, results pertaining to it are more widely applicable.

Many current issues regarding brain networks center on how structure determines dynamics and vice versa [2–4,8–17]. These have implications for understanding how brain architectures and effective connectivities underpin information processing and task execution [2,4,18], how they develop and evolve, and whether their dynamics exhibit features like marginal stability or criticality [12,13,19–28]. The forward problem has been addressed at levels ranging from computations of functional connectivity for specific structures and connection strengths to more general approaches based on linear responses or neural field theory (NFT) [10,16,19]. The more difficult inverse problem of determining effective connectivity from function is of particular interest because of the potential to use brain activity, which is relatively easy to measure, to infer strengths of connections. The results carry over to other types of dynamical networks [5,6,25].

Brain connectivity is often quantified via connection matrices (CMs) whose rows and columns represent brain areas and whose entries measure connectivities between them [2–4]. Anatomical CMs (aCMs) show anatomical connections, regardless of whether they are active, as in Fig. 1(a), which shows a symmetric aCM inferred from diffusion imaging [3,4,7,17]; direct effective CMs (deCMs, which have also been termed gain matrices, eCMs, and strength-of-connection matrices in the literature) embody the strength of each *direct* connection between points in a given brain state; total effective CMs (teCMs, termed eCMs in [19])

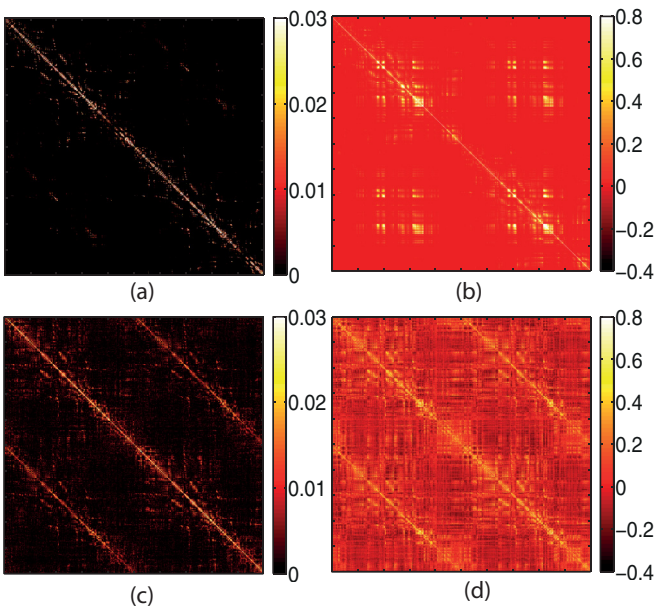

FIG. 1. (Color online) Exemplar connectivity matrices (CMs). Each of 998 nodes into which the cortex is divided corresponds to one row and one column. Entries denote connections between nodes, with largest connections near the main diagonal and two diagonals that correspond to interhemispheric connections between homologous regions. Figures (a) and (d) are adapted from [17]. (a) Experimental aCM from diffusion imaging. (b) fCM derived by assuming the structure in the aCM in (a) is an approximation to that of the corresponding deCM. (c) deCM derived from the fCM in (d). (d) Experimental fCM from fMRI covariances after global signal removal.

measure the total connectivity between points via both direct and indirect paths; and functional CMs (fCMs) are determined from correlations between activity at different points. Most commonly, fCMs are determined from equal-time correlations (to within measurement resolution; i.e., covariances, which are symmetric) of low-frequency functional magnetic resonance imaging (fMRI) signals [2–4] that indirectly measure neural activity [2–4,7,29]—see Fig. 1(d). Using NFT [30], we recently showed that the deCM and teCM correspond to bare and dressed propagators, respectively, and can be used to compute the fCM [19], but the inverse problem is unsolved.

Of particular current interest are the architecture and role of networks that are active when the brain is not performing specific experimental tasks—so-called "resting state" (a misnomer because the normal brain is always active) or "default-mode" networks, which also appear to be activated by many tasks [2–4,8–19,30–33]. These networks are central to much brain activity, but it is not clear to what extent they involve specialized regional neural connections as opposed to simply being the lowest eigenmodes of network activity.

In this paper we combine NFT [19] with matrix analysis and stability criteria to solve the inverse problem of determining deCMs and teCMs from fCMs. Because fCMs determined from covariances are symmetric by definition, and most aCMs are only measured in symmetric form, we focus mostly on the symmetric case. Implications of the results for brain connectivity, stability, and criticality are outlined and illustrated at the proof-of-principle level by means of the data used to produce Fig. 1. We exclude cases in which CMs change on time scales comparable to the activity itself, thus ruling out some task-related dynamic connectivity problems, which remain for future investigation. Section II outlines the relevant theory and derives the necessary inversion formulas. Section III then applies these to illustrative examples of anatomical and functional connection matrices and relates the results to other network measures.

II. THEORY

We write the synaptic activity that dominates brain metabolism and fMRI [7,34] as a column vector $\mathbf{\Phi}(t)$ with one entry per physical location. This follows the approach in [19], but we show the time dependence via the argument t.

In [19] it was pointed out that a relatively weak signal is sufficient to establish the existence of a connection between two points, and most normal large scale brain activity has been established to be approximately linear [19,30], a point that we further establish below. Linear NFT of discrete networks then implies

$$\mathbf{\Phi}(t) = \int \mathbf{\Lambda}^{(0)}(t-t')\mathbf{\Phi}(t')dt' + \mathbf{N}(t), \tag{1}$$

where $\mathbf{N}$ is a vector of external inputs; the causal propagator $\mathbf{\Lambda}^{(0)}$ is identified as a spatiotemporal deCM [19] and depends only on $t-t'$. Equation (1) is equally valid for spatially continuous systems if the index that labels elements is replaced by a continuous position vector, most conveniently written as an argument alongside time.

Fourier transforming (1) vs t gives

$$\mathbf{\Phi}(\omega) = [\mathbf{I} - \mathbf{\Lambda}^{(0)}(\omega)]^{-1}\mathbf{N}(\omega), \tag{2}$$

$$= \mathbf{T}(\omega)\mathbf{N}(\omega), \tag{3}$$

$$= [\mathbf{I} + \mathbf{\Lambda}(\omega)]\mathbf{N}(\omega), \tag{4}$$

where ω is the angular frequency, $\mathbf{I}$ is the unit matrix, the transfer matrix $\mathbf{T}$ is defined by comparison of (2) and (3), and the teCM $\mathbf{\Lambda}$ is defined by comparison of (2) and (4) [19].

If we define the spatial fCM to be the covariance matrix of the activity, it can be written

$$\mathcal{C} = \langle \mathbf{\Phi}(t)\mathbf{\Phi}^T(t)\rangle, \tag{5}$$

where the angle brackets indicate an average over t. We Fourier transform the quantities on the right of (5) and note that the average is the zero-frequency component of this expression. Then, if the average is done first, we find [19]

$$\mathcal{C} = \int \mathbf{C}(\omega)\frac{d\omega}{2\pi}, \tag{6}$$

$$\approx \mathbf{C}(\omega = 0), \tag{7}$$

where $\mathbf{C}$ is the normalized correlation matrix whose equal-time value in coordinate space is $\mathcal{C}$, as given by the integral in (6). The approximation (7) is valid for the very low frequencies of fMRI, which responds significantly only to frequencies $\omega \lesssim 1$ s^{-1} [7,29]; as an approximation to (6) it could be replaced without loss of generality by a weighted average over bandwidth.

Using (3) we can write the right side of (7) in terms of the transfer function, which leads to

$$\mathcal{C} = \mathbf{T}(\omega = 0)\mathbf{T}^{\dagger}(\omega = 0), \tag{8}$$

where the dagger in (8) indicates a Hermitian conjugate and (8) assumes uncorrelated inputs with

$$\langle \mathbf{N}(t)\mathbf{N}^{T}(t')\rangle = \mathbf{I}\delta(t - t'). \tag{9}$$

From this point on, all matrices assume very low ω and the argument ω is omitted from Fourier representations.

Standard matrix theory implies that $\mathbf{\Lambda}^{(0)}$ can be written in the form

$$\mathbf{\Lambda}^{(0)} = \mathbf{U}\mathbf{L}^{(0)}\mathbf{U}^{\dagger}, \tag{10}$$

where $\mathbf{U}$ is a unitary matrix whose columns are the eigenvectors of $\mathbf{\Lambda}^{(0)}$, $\mathbf{U}^{\dagger} = \mathbf{U}^{-1}$, and $\mathbf{L}^{(0)}$ is a diagonal matrix of the eigenvalues $\lambda_j^{(0)}$ of $\mathbf{\Lambda}^{(0)}$, written

$$\mathbf{L}^{(0)} = \mathrm{diag}\big(\lambda_j^{(0)}\big). \tag{11}$$

The matrix $\mathbf{\Lambda}^{(0)}$ represents direct connections between points, while its powers represent successively higher-order polysynaptic paths [19]. The sum of these contributions is the teCM $\mathbf{\Lambda}$, with

$$\mathbf{\Lambda} = \sum_{m=1}^{\infty}[\mathbf{\Lambda}^{(0)}]^m, \tag{12}$$

$$= \sum_{m=1}^{\infty}[\mathbf{U}\mathbf{L}^{(0)}\mathbf{U}^{\dagger}]^m, \tag{13}$$

$$= \sum_{m=1}^{\infty}\mathbf{U}[\mathbf{L}^{(0)}]^m\mathbf{U}^{\dagger}, \tag{14}$$

since $\mathbf{U}^{\dagger}\mathbf{U} = \mathbf{I}$. The sum in (14) commutes with $\mathbf{U}$, so

$$\mathbf{\Lambda} = \mathbf{U}\mathbf{L}\mathbf{U}^{\dagger}, \tag{15}$$

with

$$\mathbf{L} = \mathrm{diag}(\lambda_j), \tag{16}$$

$$= \mathrm{diag}\left(\frac{\lambda_j^{(0)}}{1 - \lambda_j^{(0)}}\right), \tag{17}$$

where (16) and (17) define λ_j in terms of $\lambda_j^{(0)}$. For convergence of the series in (12) and stability of the system, one requires $|\lambda_j^{(0)}| < 1$ for all j [19], with criticality at the boundary of this circle in the complex plane.

Following similar steps to the derivation of (11)–(17), Eqs. (3), (4), and (17) imply

$$\mathbf{T} = \mathbf{U}\Theta\mathbf{U}^{\dagger}, \tag{18}$$

with

$$\Theta = \mathbf{I} + \mathbf{L}, \tag{19}$$

$$= \mathrm{diag}(1 + \lambda_j), \tag{20}$$

$$= \mathrm{diag}\big(\big[1 - \lambda_j^{(0)}\big]^{-1}\big). \tag{21}$$

Similarly, from (8), (18), and (21), we find

$$\mathcal{C} = \mathbf{U}\mathbf{K}\mathbf{U}^{\dagger}, \tag{22}$$

with

$$\mathbf{K} = \mathrm{diag}(\kappa_j), \tag{23}$$

$$= \mathrm{diag}\big(\big|1 - \lambda_j^{(0)}\big|^{-2}\big), \tag{24}$$

where (23) and (24) define κ_j in terms of $\lambda_j^{(0)}$. A previous analysis of brain network dynamics with a purely spatial deCM followed a different approach to find a result with some similarity to (24) [16]. However, that result erred by setting to zero averages of quantities like $\mathbf{\Phi}(\omega)\mathbf{N}^{\dagger}(\omega)$ in the present notation, whereas the input $\mathbf{N}$ and the response $\mathbf{\Phi}$ that it generates are actually correlated. Another analysis restricted to networks of nonindentical oscillators governed by first-order ordinary differential equations with a specialized aCM whose row sums must be zero also found a related result [35].

The meaning of the result (18) is best seen when one retains the argument ω (which leaves these results otherwise unchanged). Then (18) is seen to be the standard quadratic-form expansion of the Green function $\mathbf{T}$ in terms of its eigenfunctions [36], in which poles due to zeros of the denominator of (21) give resonances in the propagation between points via excitation of the eigenmodes. When the system is spatially uniform, a spatial Fourier transform yields poles that correspond to the dispersion relations of the eigenmodes.

III. RESULTS

In this section we first outline some theoretical consequences of the analysis in Sec. II. We then demonstrate that the results obtained for the data in Fig. 1 enable the proximity of the system to criticality to be measured and are consistent with estimates obtained from electroencephalographic (EEG) analyses. We then show that the deCM obtained by inverting the fCM in Fig. 1(d) contains significant interhemispheric connections that are missing from the aCM in Fig. 1(a)—as is expected since the imaging technique used to obtain Fig. 1(a) is known to miss many such connections [3,4,17]. This demonstrates that the fCM can potentially be used to infer the deCM, rather than relying on estimates from the aCM [which is known to omit many long-range connections and hence doesn't generate the correct fCM, as seen in Fig.1(b)].

Several results follow from the above analysis.

(i) Given the fCM $\mathbf{C}$, one can immediately invert (24) to find

$$\lambda_j^{(0)} = 1 - \kappa_j^{-1/2}e^{i\psi_j}, \tag{25}$$

where

$$|\psi_j| < \cos^{-1}\left(\tfrac{1}{2}\kappa_j^{-1/2}\right) \tag{26}$$

places $\lambda_j^{(0)}$ in the unit circle required for stability. (Stability has previously been exploited to constrain strengths of connection in deCMs [10,11].) In the symmetric case, Eq. (23) then allows us to write the explicit result

$$\mathbf{\Lambda}^{(0)} = \mathbf{U}[\mathbf{I} - \mathbf{K}^{-1/2}]\mathbf{U}^{\dagger}, \tag{27}$$

$$\mathbf{K}^{-1/2} = \mathrm{diag}\big(\kappa_j^{-1/2}\big). \tag{28}$$

If $\mathbf{\Lambda}^{(0)}$ is symmetric (e.g., as measured by diffusion imaging), $\lambda_j^{(0)}$ is real and $\psi_j = 0$ in (25); otherwise, any complex eigenvalues come in complex conjugate pairs that

correspond to oscillatory solutions. fMRI is only sensitive to $\omega \lesssim 1\ \mathrm{s}^{-1}$ [7,29], so the only eigenvalues that can be reliably determined are real, or very nearly so [since $\mathrm{Im}\lambda_j^{(0)} = \mathrm{Re}\omega$]. This corresponds to the symmetric part of the connectivity, and $\mathbf{C}$ is inherently symmetric in any case. We also note that large amplitude oscillatory solutions likely correspond to pathologies such as epileptic seizures, not relevant to the normal brain [9,26,28].

(ii) Equation (24) implies that the eigenvalues κ_j of any covariance matrix must be non-negative. Moreover, the stability condition $|\lambda_j^{(0)}| < 1$ implies $\kappa_j > \frac{1}{4}$.

(iii) Values $\kappa_j < 1$ correspond to $\lambda_j^{(0)} < 0$; i.e., to inhibitory interactions.

(iv) The largest eigenvalues of $\mathbf{T}$ will dominate the fCM, the response to stimuli, the deCM, and the teCM, because these correspond to the least damped modes. This explains widely observed qualitative similarities between spontaneous activity patterns, evoked responses, deCMs, and fCMs [2–4], because the eigenfunctions are the same for all CMs and activity is dominated by the least-damped modes. We note that the system must be near criticality ($\max|\lambda_j^{(0)}| \approx 1$) for this similarity to apply, as previously inferred from EEG spectra [13,21–24,27], whose results imply $\max\lambda^{(0)}$ to have an average value of 0.85 ± 0.07 [22], 0.84 ± 0.07 [23], and 0.84 ± 0.05 [24], based on data from 100 to 1500 subjects. This implies that the brain is linearly stable and justifies the use of linear analysis.

We now demonstrate the above results on the data in Fig. 1 [17]. We begin by using the experimental aCM in Fig. 1(a) as an *approximation* to the deCM in the relevant brain state and use this to calculate the corresponding fCM. (This approximation will be poor if the aCM misses connections, or if some of the connections found in the aCM have greater or lesser activity than their anatomical weight would imply.) We then carry out the inverse process, starting from the experimental fCM in Fig. 1(d) to compute the corresponding deCM to see especially whether it reveals missing connections. One complication in this procedure is that the experimental fCM has had the global average signal removed, which prevents identification of this mode and induces substantial negative correlations [37,38] in Fig. 1(d), but should not affect other eigenvalues and eigenvectors. Hence this should be seen as a proof-of-principle illustration that we hope will stimulate experimental tests. Using Eqs. (10), (17), and (23), the aCM in Fig. 1(a) yields the fCM in Fig. 1(b) when we minimize the norm of the difference from the observed fCM in Fig. 1(d) to estimate the (unknown *a priori*) normalization of the aCM. This yields $\max\lambda_j^{(0)} = 0.87$, in excellent agreement with the EEG estimates, placing the network near criticality.

The calculated fCM 1(b) shows strong connections near the main diagonal, as in Fig. 1(d), but with weaker block-diagonal structure. Some functional connections are not reproduced, especially interhemispheric connections between homologous regions (the prominent secondary diagonals). This is not surprising because the imaging methods used in Fig. 1(a) have difficulty in identifying these known interhemispheric anatomical links that would otherwise appear in Fig. 1(a) [3,4,17] and the connection strengths inferred from Fig. 1(a) are only approximate estimates derived from numbers of fibers connecting regions, taking no account of whether or not these are active.

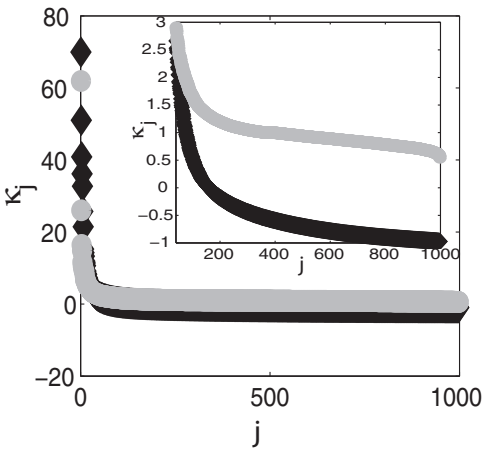

FIG. 2. CM eigenvalues κ_j vs mode number $j = 1$–998, ordered by size for the fCMs in Fig. 1(b) (gray) and Fig. 1(d) (black). Inset shows vertical zoom.

The fCM in Fig. 1(d) shows covariances of activity between different points. This activity is supported by all active connections, whether or not their strengths are inferred correctly by the measurements used to derive the aCM in Fig. 1(a), so the analysis in Sec. II should enable the deCM to be determined. Hence we next invert the fCM in Fig. 1(d) to obtain the deCM in Fig. 1(c), which we discuss below. The fCM's leading eigenvalues satisfy the stability condition $\kappa_j > \frac{1}{4}$, as seen in the plot in Fig. 2. However, the smaller eigenvalues do not all fulfill this requirement, and some are even negative (the most negative is -0.97). These discrepancies likely result from global signal removal, measurement noise, and the necessarily nonzero bandwidth employed. Certainly, we find that if we artificially add a small random component to the analytic fCM in Fig. 1(b) it induces similar negative eigenvalues, while waves of activity have previously been found to induce negative correlations between locations separated by around half a wavelength [19]. We conclude that such effects in Fig. 1(d) can easily produce eigenvalues $\kappa_j < \frac{1}{4}$ in measured CMs.

Unfortunately, the data in Fig. 1 are too noisy to resolve issues of negative eigenvalues. This will require further analysis and experimental verification with low-noise deCMs and fCMs for the same subjects and retention of the global signal in the data. In the present work, we conclude that eigenvalues are only reliable to within an amount of order ± 1, given that values of -0.97 are seen when only $\kappa_j > 1/4$ would be possible for a precisely measured fCM. We then proceed by retaining only the first 76 eigenvalues ($\kappa_j > 1.0$) and corresponding eigenvectors. Omission of activity in the remaining eigenvectors is found to have only a slight effect on the fCM, with its norm changing by less than 1.6%. The largest eigenvalue is 69, whence (22) implies $\max\lambda_j^{(0)} \approx 0.88$, in excellent agreement with EEG-based results and again implying near criticality. Notably, this implies that 88% of activity in the dominant mode is generated internally with only 12% from external sources [12,21]. Using this method, interrelating the deCM and fCM does not require us to minimize the norm of the difference between theoretical and

experimental fCMs, and the results determine the strengths of the deCM entries objectively.

Figure 1(c) captures the existence of interhemispheric connections, mostly between homologous regions, identifying these via the eigenfunctions of the fCM, which include their effects. The changes near the main diagonal are mostly due to an overall rescaling to accommodate the effects of the additional interhemispheric connections identified; however, some connections have different relative strengths than in Fig. 1(a) (which is an aCM, not a deCM, so is only approximately comparable in any case). Some deCM strengths are found to be slightly negative in Fig. 1(c), but this is due to the removal of the global signal from Fig. 1(d)—certainly, we find a similar result if Fig. 1(a) is modified by removing the contribution from the global mode.

Further experimental and anatomical testing using low-noise anatomical, effective, and functional data sets obtained for the same subjects is needed to explore these findings in detail and refine the overall criticality measure to improve global mode removal. However, we stress that the present proof-of-principle results demonstrate the potential to infer deCMs from fCMs in situations where the observed aCM (viewed as an approximation to the deCM) does not yield an accurate prediction of the fCM, either because connections are missed or their relative strengths are not estimated correctly. We further emphasize that there was no *a priori* reason why the additional connections in Fig. 1(c) would be the very ones that are known to be underestimated by the technique used to obtain Fig. 1(a) so their detection further underlines the potential of the method.

As a final result, we demonstrate that the incomplete "patchiness" of the secondary diagonals in the aCM in Fig. 1(a) is responsible for the strong square features in Fig. 1(b) when the aCM is used to approximate the deCM. Conversely, we show that the strong similarity of the interhemispheric submatrices in Fig. 1(d) (top right and bottom left quadrants) to the intrahemispheric ones (top left and bottom right quadrants) is linked directly to the existence of the approximately uniform one-to-one interhemispheric connections between homologous regions seen in the deCM in Fig. 1(c).

We begin by approximating the deCM as

$$\mathbf{\Lambda}^{(0)} = \begin{pmatrix} \mathbf{A} & \mathbf{D} \\ \mathbf{D} & \mathbf{A} \end{pmatrix}, \tag{29}$$

where $\mathbf{A}$ is the intrahemispheric deCM (approximated as being the same for both hemispheres) and $\mathbf{D}$ is a diagonal matrix that approximates the interhemispheric connectivity as being one to one but not necessarily the same for all pairs of homologous points. If we then write the transfer matrix as

$$\mathbf{T} = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix}, \tag{30}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are symmetric matrices, we find

$$\mathbf{X} = [\mathbf{I} - \mathbf{A} - \mathbf{D}(\mathbf{I} - \mathbf{A})^{-1}]^{-1}, \tag{31}$$

$$\mathbf{Y} = \mathbf{X}\mathbf{D}(\mathbf{I} - \mathbf{A})^{-1}, \tag{32}$$

where $\mathbf{I}$ is a unit matrix of the same size as $\mathbf{A}$.

Rather than proceed in full generality, we now assume that the norm of $\mathbf{D}$ is small compared to that of $\mathbf{A}$, which suffices to establish the key results. In this case, Eq. (8) implies

$$\mathcal{C} \approx \begin{pmatrix} \mathbf{X^2} & \mathbf{Z} \\ \mathbf{Z} & \mathbf{X}^2 \end{pmatrix}, \tag{33}$$

$$\mathbf{Z} \approx \mathbf{X}(\mathbf{X}\mathbf{D} + \mathbf{D}\mathbf{X})\mathbf{X}, \tag{34}$$

to first order in $\mathbf{D}$.

If interhemispheric connections are uniform and one to one, then $\mathbf{D} = a\mathbf{I}$, where we assume a is a small constant. Then, $\mathbf{Z} = 2a\mathbf{X}^3$. Since $\mathbf{X}$ is itself dominated by entries that fall off rapidly in magnitude with distance from the principal diagonal, the structure of $\mathbf{Z}$ is very similar to that of $\mathbf{X}^2$, which explains the qualitative form of the fCM in Fig. 1(d). Comparison of the norms of the submatrices implies $a \approx 0.4$, thereby confirming that interhemispheric connections are non-negligible with respect to intrahemispheric ones. [This value of a implies that second-order terms are of order 0.2, so a full analysis, starting from (31) and (32), is needed for detailed quantitative results.]

If interhemispheric connections are one to one, but do not exist between all homologous points, then $\mathbf{D}$ will be patchy, as in Fig. 1(a) where it has two regions of particularly strong entries. Each strong entry (or group) gives rise to a strong row (or group of adjacent rows) in the term $\mathbf{DX}$ in (34) and to a strong column (or group of adjacent columns) in the term $\mathbf{XD}$. In the case of Fig. 1(a), this produces a structure in $\mathbf{Z}$ that is dominated by the intersections of two groups of strong neighboring rows and the corresponding two groups of columns, leading to the four dominant regions seen in this submatrix in Fig. 1(b). This explains how incomplete observation of interhemispheric connectivities leads to strong artifacts in the fCM. The appearance of the same artifact in the intrahemispheric submatrices of the fCM arises from the higher-order terms in $\mathbf{D}$ that are retained if we analyze (31) and (32) exactly.

IV. SUMMARY AND DISCUSSION

We have shown that the fCM $\mathcal{C}$ can be used to infer the transfer matrix $\mathbf{T}$, the teCM $\mathbf{\Lambda}$, and the deCM $\mathbf{\Lambda}^{(0)}$ in the symmetric case. This avoids forward modeling (indeed, we do not assume any particular neural model) and fitting predicted to observed fCMs. It identifies the correctly normalized deCM, including interhemispheric connections and strengths that differ from ones estimated by diffusion imaging or from approximation of the deCM by the aCM (an approximation that does not allow for connections that may be physically present but inactive, or for ones that may have been omitted from the aCM). Inferences of interhemispheric connections, for example, are predictions that can be experimentally tested as better imaging techniques become available for tracking long-range axons.

Our results imply that the brain is in a near-critical state, with criticality measures very close to values obtained from EEG spectra, thereby explaining similarities between various CMs and brain activity patterns. We suggest that imposition of joint deCM-fCM constraints on stability and sign of connection strengths might further assist in distinguishing global fMRI signal from artifact [37,38], e.g., by restricting global-mode subtraction to levels that do not induce negative deCM entries if it can be reliably argued that these do not occur at the relevant scales. Generalization of the present work

will be needed to probe asymmetric, rapidly time varying, and oscillatory cases and to deal with measurement effects systematically. Rigorous experimental verification of inferred connections in individual subjects will also be required to fully validate the method—data that are not currently available to us with corresponding fMRI. Notably, a paper submitted and published since submission of the present work has found similar omission of interhemispheric connections in aCMs by simulating activity on the aCM, using a specific neural model, then progressively adjusting connection strengths until the best match with the resulting numerically calculated fCM is achieved [39]. That work also concluded that near criticality is required for a good match, in agreement with the present work and Ref. [19].

The present method has significant advantages for probing and mapping brain activity, structure (i.e., strengths of connections), and structure-function relationships. It differs from techniques such as dynamic causal modeling (DCM) [40,41], which requires a specific dynamic model to be assumed at each node, and which cannot handle networks with large numbers of nodes. In contrast, our method assumes no specific dynamics aside from approximate linearity (and justifies this for the brain via the resulting experimental stability measure), and can easily deal with networks such as the 998-node one in Fig. 1. Indeed, the reasons that DCM can work for sparse networks of regions of interest were recently elucidated by application of Eq. (1) [42].

Aside from brain applications, the method described here is also relevant to other complex networks discussed in the Introduction, whose structure is difficult to measure directly and more readily probed via activity supported—e.g., econophysical, genetic, economic, ecological, and social networks.

ACKNOWLEDGMENTS

The Australian Research Council, Westmead Millennium Institute, and Brain Resource Ltd. supported this work.

[1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[2] K. J. Friston, Brain Conn. **1**, 13 (2011).
[3] E. Bullmore and O. Sporns, Nat. Rev. Neurosci. **10**, 186 (2009).
[4] O. Sporns, *Networks of the Brain* (MIT, Cambridge, MA, 2011).
[5] R. Albert and A. L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002).
[6] M. Barthélemy, Phys. Rep. **499**, 1 (2011).
[7] *Functional MRI: An Introduction to Methods*, edited by P. Jezzard, P. M. Matthews, and S. M. Smith (Oxford University Press, Oxford, 2001).
[8] O. Sporns, G. Tononi, and G. M. Edelman, Cereb. Cortex **10**, 127 (2000).
[9] C. J. Honey, R. Kötter, M. Breakspear, and O. Sporns, Proc. Natl. Acad. Sci. USA **104**, 10240 (2007).
[10] P. A. Robinson, J. A. Henderson, E. Matar, P. Riley, and R. T. Gray, Phys. Rev. Lett. **103**, 108104 (2009).
[11] J. A. Henderson and P. A. Robinson, Phys. Rev. Lett. **107**, 018102 (2011).
[12] M. Rubinov, O. Sporns, J.-P. Thivierge, and M. Breakspear, PLoS Comput. Biol. **7**, e1002038 (2011).
[13] M. G. Kitzbichler, M. L. Smith, S. R. Christensen, and E. Bullmore, PLoS Comput. Biol. **5**, e1000314 (2009).
[14] S. A. Knock, A. R. McIntosh, O. Sporns, R. Kötter, P. Hagmann, and V. K. Jirsa, J. Neurosci. Meth. **183**, 86 (2009).
[15] M. D. Grecius, K. Supekar, V. Menon, and R. F. Dougherty, Cereb. Cortex **19**, 72 (2009).
[16] R. F. Galán, PLoS ONE **3**, e2148 (2008).
[17] C. J. Honey, J.-P. Thivierge, and O. Sporns, NeuroImage **52**, 766 (2010).
[18] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann, Proc. Natl. Acad. Sci. USA **106**, 2035 (2009).
[19] P. A. Robinson, Phys. Rev. E **85**, 011912 (2012).
[20] J. M. Beggs and D. Plenz, J. Neurosci. **23**, 11167 (2003).
[21] P. A. Robinson, C. J. Rennie, and J. J. Wright, Phys. Rev. E **56**, 826 (1997).
[22] D. L. Rowe, P. A. Robinson, and C. J. Rennie, J. Theor. Biol. **231**, 413 (2004).
[23] P. A. Robinson, C. J. Rennie, D. L. Rowe, and S. C. O'Connor, Hum. Brain Mapp. **23**, 53 (2004).
[24] S. J. van Albada, C. C. Kerr, C. J. Rennie, and P. A. Robinson, Clin. Neurophysiol. **121**, 21 (2010).
[25] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Rev. Mod. Phys. **80**, 1275 (2008).
[26] P. A. Robinson, C. J. Rennie, and D. L. Rowe, Phys. Rev. E **65**, 041924 (2002).
[27] P. A. Robinson, C. J. Rennie, J. J. Wright, H. Bahramali, E. Gordon, and D. L. Rowe, Phys. Rev. E **63**, 021903 (2001).
[28] M. Breakspear, J. A. Roberts, J. R. Terry, S. Rodrigues, N. Mahant, and P. A. Robinson, Cereb. Cortex **16**, 1296 (2006).
[29] K. M. Aquino, M. M. Schira, P. A. Robinson, P. M. Drysdale, and M. Breakspear, PLoS Comput. Biol. **8**, e1002435 (2012).
[30] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston, PLoS Comput. Biol. **4**, e1000092 (2008).
[31] G. Deco, V. K. Jirsa, and A. R. McIntosh, Nat. Rev. Neurosci. **12**, 43 (2011).
[32] G. Deco and V. K. Jirsa, J. Neurosci. **32**, 3366 (2012).
[33] M. E. Raichle, A. M. MacLeod, A. Z. Snyder, W. J. Powers, D. A. Gusnard, and G. L. Shulman, Proc. Natl. Acad. Sci. USA **98**, 676 (2001).
[34] D. Attwell and S. B. Laughlin, J. Cereb. Blood Flow Metab. **21**, 1133 (2001).
[35] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, Phys. Rev. Lett. **104**, 058701 (2010).
[36] W. Greiner, *Relativistic Quantum Mechanics, Wave Equations* (Springer, Berlin, 1990).
[37] K. Murphy, R. M. Birn, D. A. Handweerker, T. B. Jones, and P. A. Bandettini, NeuroImage **44**, 893 (2009).
[38] H. He and T. T. Liu, NeuroImage **59**, 2339 (2012).
[39] G. Deco, A. R. McIntosh, K. Shen, R. M. Hutchison, R. S. Menon, S. Everling, P. Hagmann, and V. K. Jirsa, J. Neurosci. **34**, 7910 (2014).
[40] K. J. Friston, L. Harrison, and W. Penny, NeuroImage **19**, 1273 (2003).
[41] M. L. Seghier and K. J. Friston, NeuroImage **68**, 181 (2013).
[42] P. A. Robinson, Phys. Rev. E **88**, 054702 (2013).