PHYSICAL REVIEW E **88**, 054702 (2013)

Discrete-network versus modal representations of brain activity: Why a sparse regions-of-interest approach can work for analysis of continuous dynamics

P. A. Robinson
School of Physics, University of Sydney, New South Wales 2006, Australia
Brain Dynamics Center, Sydney Medical School–Western, University of Sydney, Westmead, New South Wales 2145, Australia
and Center for Integrated Research and Understanding of Sleep, 431 Glebe Point Road, Glebe, NSW 2037, Australia
(Received 22 August 2013; published 26 November 2013)

The efficacy of the common practice of tracking brain dynamics using a few key regions of interest is explained via the fact that these regions are sensitive to underlying extended modes of activity, not just local dynamics. This underlines the inseparable interplay between modes and regions and reflects the reality that brain functions range from highly localized to highly extended.

DOI: 10.1103/PhysRevE.88.054702 PACS number(s): 87.10.−e, 87.18.Sn, 87.61.Qr, 89.75.Hc

I. INTRODUCTION

It has long been debated whether brain functions are localized or extended, discrete or continuous, and in what terms they are best described. Much evidence for localization comes from studies of normal function, lesions, and brain injuries, which show that specific functions and deficits are often associated with to localized regions [1,2], such as parts of the visual cortex. Conversely, integration of responses to produce unified percepts from disparate sensory stimuli appears to be widely dispersed even when specific processing is localized [2,3]. The full situation is thus a mixture of the two limiting cases.

Much data has been collected in recent years on the networks that underpin the flow of information between different areas of the brain [3,4]. On one hand, individual neurons are often highly specific in their projections from one region to another; conversely, there is sufficient divergence in the projections from a single brain area that all parts of the brain are connected to all others within only a few steps [2,3], so networks are generally not highly compartmentalized. Moreover, neural field theory (NFT) approaches that approximate couplings as spatially continuous have had much success in explaining a wide variety of brain phenomena [5–10], often via extended modes of activity. Such extended modes are also widely reported in studies of spontaneous and task-evoked brain activity [3,4,11–15]. These activity patterns are sometimes pictured as discrete structural networks that either exist continuously or are formed temporarily to process particular inputs [2,4,11,13–16], and analysis is often conducted in terms of the dynamics of just a few nodes or regions of interest (ROIs) that are specific to the brain state or task of interest. An alternative viewpoint is that the resting state networks (or modes) are simply the lowest (principal) eigenmodes of the brain structure [17] and that these are likely to be preferentially excited during task execution. All these points imply that there are roles for both localized and delocalized activity in brain dynamics and that analysis methods based on these two pictures need to be reconciled.

An approach based on a limited number of ROIs is often successful in distinguishing between different dynamics in task conditions, for example [13,14]. ROIs are the basis for dynamical causal modeling (DCM), which attempts to infer causal links between ROIs by using a generic model of their dynamics and Bayesian inference [13,14]. Recently, it was noted that even when one seeks a description in terms of dynamics at these ROIs only, these links must be mediated by activity in modes of the system [13], whether these are viewed as modes of a discrete network or of a continuous medium.

This Brief Report shows how the dynamics of a relatively small number of localized ROIs can accurately capture even distributed brain dynamics via their behavior being representative of the dominant underlying modes. The properties that distinguish "good" ROIs (i.e., ones that capture the dynamics successfully) from other choices are discussed, and the number of ROIs required is estimated. In Sec. II we introduce a class of equations widely used to describe brain dynamics and express them in both network and field forms. We then carry out their systematic decimation to a restricted network of ROIs and use underlying modal dynamics to explain why this decimated ROI representation works.

II. THEORY

In this section we introduce equations for neural dynamics in a form compatible with a wide range of analyses in the literature and show how their dynamics can be systematically approximated via decimation to a set of discrete ROIs.

A. Dynamical equations

Because time delays are not critical to the issues addressed here, we assume that we are dealing with low-frequency signals and neglect time delays for neural signals to propagate between different parts of the brain, which is a reasonable approximation under these circumstances [13,14]; generalization to include time delays is possible but not needed to address the issues considered here. We write the following equation for the neural firing rate $Q(\mathbf{r},t)$ at position $\mathbf{r}$ and time t of the dominant population (such a form also results when multiple populations are included):

$$DQ(\mathbf{r},t) = \int \Lambda^{(0)}(\mathbf{r},\mathbf{r}')Q(\mathbf{r}',t)d\mathbf{r}' + N(\mathbf{r},t). \qquad (1)$$

Here, $\Lambda^{(0)}$ is the propagator for activity to travel to $\mathbf{r}$ from $\mathbf{r}'$ (a special case is the direct effective connection matrix, or gain matrix, which measures the instantaneous strengths of anatomical connections between points [3,4,14,16,17]), D is a purely temporal differential operator, N embodies any external inputs, and the integral extends over all brain coordinates $\mathbf{r}'$. If the brain is discretized into a fine network of ROIs, or if a coarse-grained subset of these ROIs is used, the integral in Eq. (1) is replaced by a sum, and we can write it in matrix notation as

$$DQ_j(t) = \sum_k \Lambda^{(0)}_{jk} Q_k(t) + N_j(t), \tag{2}$$

where j and k denote locations. Equation (2) makes the reason for identification of $\Lambda^{(0)}$ as a connection matrix [17] more obvious. This equation is of the form used in DCM, except that DCM assumes $\Lambda^{(0)}$ to be a function of the activity.

For present purposes, Dirac bra-ket notation is most useful and compact [18,19], with Eqs. (1) and (2) becoming

$$\langle \mathbf{r}|D|Q(t)\rangle = \langle \mathbf{r}|\Lambda^{(0)}|\mathbf{r}'\rangle\langle \mathbf{r}'|Q(t)\rangle + \langle \mathbf{r}|N(t)\rangle. \tag{3}$$

Here, the eigenstates $|\mathbf{r}\rangle$ and their adjoints $\langle \mathbf{r}|$ are orthonormal δ-function eigenstates of the position operator and $\Lambda^{(0)}$ is interpreted as an operator. In Eq. (3) and henceforth we adopt the Einstein summation convention in which any repeated index is summed over and this is interpreted as an integral wherever the index is continuous.

B. Decimation

We wish to represent the above system via its dynamics at a less numerous, decimated, set of spatial points $\mathbf{R}$, which we denote by upper case letters [there should be no confusion with the state $Q(t)$]. We discuss the choice of these points below. We also introduce eigenfunctions $|\phi\rangle$ of the propagator $\Lambda^{(0)}$, which we denote by Greek letters. We thus write

$$\langle \mathbf{r}|D|Q(t)\rangle = \langle \mathbf{r}|\phi\rangle\langle\phi|\Lambda^{(0)}|\psi\rangle\langle\psi|\mathbf{r}'\rangle\langle \mathbf{r}'|Q(t)\rangle + \langle \mathbf{r}|N(t)\rangle, \tag{4}$$

$$= \langle \mathbf{r}|\phi\rangle\langle\phi|\Lambda^{(0)}|\psi\rangle\langle\psi|Q(t)\rangle + \langle \mathbf{r}|N(t)\rangle, \tag{5}$$

where $\langle\psi|$ and $\langle\phi|$ are adjoints of $|\psi\rangle$ and $|\phi\rangle$.

We note that if

$$\Lambda^{(0)}|\phi\rangle = \lambda_\phi|\phi\rangle, \tag{6}$$

where λ_ϕ is the eigenvalue corresponding to the eigenstate $|\phi\rangle$, then [18,19]

$$[\Lambda^{(0)}]^{-1} = \sum_\phi \frac{|\phi\rangle\langle\phi|}{\lambda_\phi}. \tag{7}$$

We further note that only the lowest eigenmodes of brain activity are weakly damped [6,8,10], and hence that activity is dominated by relatively few modes, which we denote by upper-case Greek letters. We thus obtain

$$\langle \mathbf{R}|D|Q(t)\rangle \approx \langle \mathbf{R}|\Phi\rangle\langle\Phi|\Lambda^{(0)}|\Psi\rangle\langle\Psi|Q(t)\rangle + \langle \mathbf{R}|N(t)\rangle, \tag{8}$$

where we have restricted attention to the locations $\mathbf{R}$ without loss of generality; the only approximation in going from Eq. (5) to Eq. (8) is the restriction to a subset of (say) M propagator eigenstates.

Equation (8) involves the evaluation of inner products of each of the M retained propagator eigenstates with $|Q(t)\rangle$. As in the general theory of integration using orthogonal functions, one can evaluate the relevant integrals exactly using only $\sim M$ appropriately chosen points $\mathbf{R}'$ and associated weights $w_{\Psi\mathbf{R}'}$ [20]. Then the approximation in going from Eq. (5) to Eq. (8) involves only the neglect of higher modes, not errors in evaluation of the terms retained. We can then write

$$\langle \mathbf{R}|D|Q(t)\rangle \approx \langle \mathbf{R}|\Phi\rangle\langle\Phi|\Lambda^{(0)}|\Psi\rangle \sum_{\mathbf{R}'} w_{\Psi\mathbf{R}'}\langle\Psi|\mathbf{R}'\rangle\langle\mathbf{R}'|Q(t)\rangle + \langle \mathbf{R}|N(t)\rangle, \tag{9}$$

$$= \sum_{\mathbf{R}'}[\langle \mathbf{R}|\Phi\rangle\langle\Phi|\Lambda^{(0)}|\Psi\rangle w_{\Psi\mathbf{R}'}\langle\Psi|\mathbf{R}'\rangle]\langle\mathbf{R}'|Q(t)\rangle + \langle \mathbf{R}|N(t)\rangle, \tag{10}$$

$$= \langle \mathbf{R}|\Lambda^{(0)}_{\mathrm{eff}}|\mathbf{R}'\rangle\langle\mathbf{R}'|Q(t)\rangle + \langle \mathbf{R}|N(t)\rangle, \tag{11}$$

where Eq. (11) defines $\Lambda^{(0)}_{\mathrm{eff}}$, which is not equal to $\Lambda^{(0)}$ unless M equals the total number of points required to fully resolve the dynamics. This is an important point, since observed brain networks are already coarse-grained representations of brain tissue.

The above analysis implies that ROIs must be at locations appropriate to approximate the inner products in Eq. (8) well and should thus predominantly lie in places where $|\Psi\rangle$ and $|\Phi\rangle$ do not have zeros. Most importantly, the result Eq. (11) is not a trivial approximation of Eq. (3). It is justified only by the relative unimportance of high-order modes of the system. The contribution of these modes can be estimated from the spectrum of the system activity in the relevant resting-state or task-evoked conditions and can be more formally estimated by Bayesian means, for example [4,13,14].

C. Transfer function and evoked activity

Activity is often evoked transiently by stimuli, including ones applied by experimenters. To analyze such situations we rewrite Eq. (3) as

$$\langle \mathbf{r}|D - \Lambda^{(0)}|\mathbf{r}'\rangle\langle\mathbf{r}'|Q(t)\rangle = \langle \mathbf{r}|N(t)\rangle. \tag{12}$$

Equation (12) can be inverted to yield

$$\langle \mathbf{r}|Q(t)\rangle = \langle \mathbf{r}|T|\mathbf{r}'\rangle\langle\mathbf{r}'|N(t)\rangle, \tag{13}$$

$$T = [D - \Lambda^{(0)}]^{-1}, \tag{14}$$

where T is the linear transfer function, and we have swapped the primed and unprimed coordinates in going from Eq. (12) to Eq. (13) without loss of generality. Recent work identified $\Lambda = T - I$ as the total effective connection matrix that includes both direct and indirect influences of one point on another [17] and I is the unit matrix. The expansion in [17] of Λ in powers of $\Lambda^{(0)}$ implies that T can also be written in the form of Eq. (7), with the same eigenfunctions but different eigenvalues. Equations (13) and (14) recast the present cases of these recent results in Dirac notation.

The decimation in Sec. IIB can be carried out in a similar way to obtain the analog of Eq. (13) involving an effective transfer function T_{eff}, giving

$$\langle \mathbf{R}|Q(t)\rangle = \langle \mathbf{R}|T_{\mathrm{eff}}|\mathbf{R}'\rangle\langle \mathbf{R}'|N(t)\rangle, \tag{15}$$

$$T_{\mathrm{eff}} = \left[D - \Lambda_{\mathrm{eff}}^{(0)}\right]^{-1}. \tag{16}$$

ROIs are often chosen to include the site $\mathbf{R}_0$ of a localized incoming stimulus, in which case, one can write

$$|N(t)\rangle = N(t)|\mathbf{R}_0\rangle, \tag{17}$$

$$\langle \mathbf{R}|Q(t)\rangle = N(t)\langle \mathbf{R}|T_{\mathrm{eff}}|\mathbf{R}_0\rangle. \tag{18}$$

Equation (18) involves only one row of the $M \times M$ matrix T_{eff}, and it may be that only a subset of entries $\langle \mathbf{R}|T|\mathbf{R}_0\rangle$ in this row of the matrix are significant for a particular stimulus. This further explains how it is often possible for just a few ROIs to capture much of the relevant dynamics, as is often found in DCM, for example [2–4,13–15].

III. SUMMARY AND DISCUSSION

This Brief Report has used a modal decomposition to show how to capture brain dynamics via the behavior of a decimated set of $\sim M$ pointlike regions of interest, so long as the relevant activity is confined to a set of M dominant modes. However, the modes themselves do not need to be tracked because the point dynamics of the decimated ROI network are equivalent. The number of ROIs required can be very small in cases where only a few entries in the transfer matrix couple significantly to the location of an applied stimulus, as is the case when specialized pathways are activated.

These results explain why the ROI approach often works, highlight the need to interpret the inferred connectivities $\Lambda_{\mathrm{eff}}^{(0)}$ carefully, and emphasize that these effective interactions include couplings via extended modes, not just discrete direct connections. They thus elucidate the combination of local and extended dynamics inherent in brain activity. Moreover, they imply that many researchers have selected ROIs that sample dominant modes effectively, but that this selection could be done on a more systematic first-principles basis in future.

ACKNOWLEDGMENTS

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

[1] E. R. Kandel, J. H. Schwartz, and T. M. Jessell, *Principles of Neural Science*, 4th ed. (McGraw-Hill, New York, 2000), pp. 5ff.

[2] O. Sporns, *Networks of the Brain* (MIT, Cambridge, 2011), p. 184.

[3] E. Bullmore and O. Sporns, Nat. Rev. Neurosci. **10**, 186 (2009).

[4] K. J. Friston, Brain Conn. **1**, 13 (2011).

[5] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston, Pub. Lib. Sci. Comp. Biol. **4**, e1000092 (2008).

[6] P. L. Nunez, *Neocortical Dynamics and Human EEG Rhythms* (Oxford University Press, Oxford, 1995).

[7] V. K. Jirsa and H. Haken, Phys. Rev. Lett. **77**, 960 (1996).

[8] P. A. Robinson, C. J. Rennie, and J. J. Wright, Phys. Rev. E **56**, 826 (1997).

[9] R. L. Beurle, Phil. Trans. Roy. Soc. (Lond.) Ser. B, Biol. Sci. **240**, 55 (1956).

[10] P. A. Robinson, P. N. Loxley, S. C. O'Connor, and C. J. Rennie, Phys. Rev. E **63**, 041909 (2001).

[11] M. D. Grecius, K. Supekar, V. Menon, and R. F. Dougherty, Cereb. Cortex **19**, 72 (2009).

[12] 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).

[13] M. L. Seghier and K. J. Friston, NeuroImage **68**, 181 (2013).

[14] K. J. Friston, L. Harrison, and W. Penny, NeuroImage **19**, 1273 (2003).

[15] D. S. Bassett, N. F. Wymbs, M. A. Porter, J. M. Carlson, and S. T. Grafton, Proc. Natl. Acad. Sci. USA **108**, 7641 (2011).

[16] P. A. Robinson, J. A. Henderson, E. Matar, P. Riley, and R. T. Gray, Phys. Rev. Lett. **103**, 108104 (2009).

[17] P. A. Robinson, Phys. Rev. E **85**, 011912 (2012).

[18] L. I. Schiff, *Quantum Mechanics*, 3rd ed. (McGraw-Hill Kogakusha, Tokyo), pp. 164ff.

[19] W. Greiner, *Relativistic Quantum Mechanics, Wave Equations* (Springer, Berlin, 1990), p. 64.

[20] F. W. Olver, D. W. Lozier, D. F. Boisvert, and C. W. Clark (eds.), *NIST Handbook of Mathematical Functions* (Cambridge, Cambridge, 2010), Chaps. 3 and 18.