



Computational Neuroscience

Extracting spatial–temporal coherent patterns in large-scale neural recordings using dynamic mode decomposition



Bingni W. Brunton ^{a,b,c,*}, Lise A. Johnson ^{d,e}, Jeffrey G. Ojemann ^e, J. Nathan Kutz ^b

^a Department of Biology, University of Washington, Seattle, WA 98195, USA

^b Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA

^c Institute for Neuroengineering, University of Washington, Seattle, WA 98195, USA

^d Center for Sensorimotor Neural Engineering, University of Washington, Seattle, WA 98195, USA

^e Department of Neurological Surgery, University of Washington, Seattle, WA 98195, USA

HIGHLIGHTS

- Dynamic mode decomposition (DMD) extracts dynamically coherent patterns from large-scale neuronal recordings.
- Multiple, distinct sleep spindle networks are identified by DMD as measured in subdural array recordings.
- Sleep spindle networks are characterized by different cortical distribution patterns, carrying frequencies and durations.

ARTICLE INFO

Article history:

Received 7 April 2015

Received in revised form

17 September 2015

Accepted 20 October 2015

Available online 31 October 2015

Keywords:

Dynamic mode decomposition

Spatiotemporal modes

Electrocorticography

Feature extraction

Sleep spindles

ABSTRACT

Background: There is a broad need in neuroscience to understand and visualize large-scale recordings of neural activity, big data acquired by tens or hundreds of electrodes recording dynamic brain activity over minutes to hours. Such datasets are characterized by coherent patterns across both space and time, yet existing computational methods are typically restricted to analysis either in space or in time separately. **New method:** Here we report the adaptation of dynamic mode decomposition (DMD), an algorithm originally developed for studying fluid physics, to large-scale neural recordings. DMD is a modal decomposition algorithm that describes high-dimensional dynamic data using coupled spatial–temporal modes. The algorithm is robust to variations in noise and subsampling rate; it scales easily to very large numbers of simultaneously acquired measurements.

Results: We first validate the DMD approach on sub-dural electrode array recordings from human subjects performing a known motor task. Next, we combine DMD with unsupervised clustering, developing a novel method to extract spindle networks during sleep. We uncovered several distinct sleep spindle networks identifiable by their stereotypical cortical distribution patterns, frequency, and duration.

Comparison with existing methods: DMD is closely related to principal components analysis (PCA) and discrete Fourier transform (DFT). We may think of DMD as a rotation of the low-dimensional PCA space such that each basis vector has coherent dynamics.

Conclusions: The resulting analysis combines key features of performing PCA in space and power spectral analysis in time, making it particularly suitable for analyzing large-scale neural recordings.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Advances in technology and infrastructure are delivering the capacity to record signals from brain cells in much greater numbers and at even faster speeds. This deluge of data is central to answering

many critical open questions in neuroscience and motivates the continued development of computational approaches to analyze, visualize, and understand large-scale recordings of neural activity. Fortunately, the activity of complex networks of neurons can often be described by relatively few distinct patterns (for instance, Broome et al. (2006), Byron et al. (2009), Churchland et al. (2012) and Machens et al. (2010)). Identifying these spatial–temporal patterns enables the reduction of complex measurements through projection onto coherent structures, where it is tractable to build dynamical models and apply machine learning tools for pattern

* Corresponding author at: Department of Biology, University of Washington, Box 351800, Seattle, WA 98195, USA. Tel.: +1 2062219330.

E-mail address: bbrunton@uw.edu (B.W. Brunton).

analysis. Here we introduce dynamic mode decomposition (DMD) as a novel approach to explore spatial-temporal patterns in large-scale neural recordings. The method combines well-characterized advantages from two of the most powerful data analytic tools in use today: power spectral analysis in time and principal components analysis (PCA) in space.

Measurements of neural activity from tens to hundreds of simultaneously recorded channels are traces in time that probe a network with complex dynamics; one principled way to make sense of such dynamic networks is with modal decomposition (Holmes et al., 1998). Modal decomposition has been successfully applied in almost every discipline of science and engineering because it makes tractable the analysis of very high-dimensional data, reducing them to combinations of relatively few distinct patterns, or *modes*. A particularly popular modal decomposition tool is PCA, which derives modes ordered by their ability to account for energy or variance in the data (Jolliffe, 2005). PCA has already been widely applied in the study of high-dimensional biological systems; however, it suffers from a few well known drawbacks. In particular, PCA is a static technique and does not model temporal dynamics of time-series data explicitly, so it often performs poorly in reproducing dynamic data, such as recordings of neural activity.

Neural dynamics are well known to be characterized by dynamic oscillations at many frequency bands, which are implicated in a variety of neural functions (Buzsáki and Draguhn, 2004; Fries, 2005; Raghavachari et al., 2001; Uhlhaas and Singer, 2010). Most tools analyzing the frequency content of a signal are related to the Fourier transform, which transforms time-varying signals into a spectrum in the frequency domain. Importantly, the power spectrum can be computed efficiently using the fast Fourier transform (FFT) algorithm (Welch, 1967), whose efficient implementation has contributed to its ubiquitous use. One example of a modal decomposition in time that goes beyond the Fourier transform is empirical mode decomposition (EMD), which computes intrinsic oscillatory modes from time-varying data (Huang et al., 1998). EMD has been used to analyze neural data, including cortical local field potential (Liang et al., 2005) and EEG (Sweeney-Reed and Nasuto, 2007). There are several extension of frequency-domain analyses that also support spatial structures (Rehman and Mandic, 2009; Rudrauf et al., 2006).

A relatively new modal decomposition method is DMD (Rowley et al., 2009; Schmid and Sesterhenn, 2008; Schmid, 2010). DMD was developed initially to study experiments and simulations in the fluid mechanics community, where it was introduced to reduce very high-dimensional dynamic data into relatively few coupled spatial-temporal modes. Importantly, it has been shown that DMD is related to Koopman spectral analysis, motivating its usefulness in characterizing dynamics of nonlinear systems (Budišić et al., 2012; Rowley et al., 2009). Beyond fluid mechanics, DMD has recently been applied to the fields of robotics (Berger et al., 2015) and disease modeling (Proctor and Eckhoff, 2015). In the context of analyzing neural recordings, DMD modes can be thought of as coherent structures in the neural activity.

1.1. Summary of computational developments

In Section 2, we describe a set of adaptations of the DMD that make it useful in the extraction of spatial-temporal patterns from neural recordings. The base DMD algorithm is given in Section 2.1, where we also note its relationship to more established methods and compare DMD modes to PCA modes for an illustrative synthetic dataset. Notably, in contrast to experiments and simulations in fluid mechanics, neural recordings often have fewer measurements (channels of electrodes) than time snapshots, so in Section 2.2 we describe the construction of an augmented data matrix. We give intuition and recommendations for how to choose a set of

parameters such that the extracted DMD modes are interpretable. For instance, DMD modes are useful as features in machine learning algorithms that undercover stereotyped patterns in the data, an attribute we leverage for a specific example described in Section 3.3. Examples of ECoG data decomposed by DMD are given in Section 2.3. Section 2.4 describes the DMD spectrum, which has units easily interpretable in comparison with traditional power spectral analyses. Next in Section 2.5, we characterize to what extent spatial-temporal modes extracted by DMD from human subdural recordings are robust to noise and subsampling.

For very large datasets whose dimensionality strains typical computing resources, DMD may be readily implemented using standard linear algebra routines to take advantage of cluster computing (for example, see Freeman et al. (2014)). We suggest that DMD may be useful in understanding spatial-temporal coherent patterns in data of escalating scale in neuroscience, including non-invasive and invasive measurements such as functional MRI, MEG, neurophysiological recordings with electrode arrays, and optical imaging of neural activity.

1.2. Summary of experimental demonstrations

To demonstrate DMD's applicability to large-scale neural recordings, we analyzed sub-dural electrode array recordings from human subjects in two different contexts.

1.2.1. Sensorimotor maps

First, we validated the DMD approach to derive sensorimotor maps based on a simple movement task. Our sensorimotor maps show statistically significant changes in activation over the sensorimotor cortex in two frequency ranges. These changes are distinct for movements of the hand and tongue, and they are consistent with results previously described by Miller et al. (2007).

1.2.2. Sleep spindle networks

Next, we leveraged DMD in combination with unsupervised clustering techniques to detect and characterize spindle networks present during sleep; a method to automatically extract these networks had not been described previously in the literature. Sleep spindles are distinctive, transient oscillations around 14 Hz that are characteristic of non-rapid eye movement (NREM) sleep, and their presence is commonly used to classify sleep stages (De Gennaro and Ferrara, 2003). Spindles have been the subject of scientific investigation since the early 1930s and their mechanisms of generation are now quite well understood (Steriade et al., 1993). In brief, sleep spindles oscillations are generated in the thalamus and their electrographic signature arises from thalamocortical connections. Even so, the role these transient oscillatory events play in brain function remains unclear. A line of evidence suggests that sleep spindles facilitate the consolidation of recently acquired memories (Clemens et al., 2005; Eschenko et al., 2006; Gais et al., 2002; Johnson et al., 2012). This hypothesis is supported by recent work demonstrating that sleep spindles can be locally, rather than globally, synchronous events (Johnson et al., 2012; Nir et al., 2011).

Historically, sleep spindles have been scored by experts on scalp EEG data. Spindles vary in amplitude, duration, central frequency, and often concur with other regularly observed sleep features. Automated detection algorithms typically rely on band-pass filtering the signal followed by an amplitude threshold on some moving average window (for instance, Ray et al. (2010) and Schimicek et al. (1994)). Recently, a number of these algorithms were evaluated against experts and crowd-sourced spindle detectors (Warby et al., 2014). It is important to point out that all of these approaches only address spindle detection. The reliable identification and characterization of spatial networks of electrodes showing synchronous spindle activity has remained a challenge. The structure of sleep

spindle networks may reveal insights about thalamocortical connections and other neural circuitry, which motivates development of more sophisticated methods.

We show that sleep spindles tended to occur coincidentally in different groups of electrodes at different times. These patterns may reflect some underlying anatomical or functional connectivity in the brain, and studying relatively local spindle networks may be a novel way to understand the organization of the brain. A companion manuscript describes our detailed analysis of the significance of these patterns (Johnson et al., 2015). Our spindle network detection algorithm enables the exploration of sleep spindle networks in a large number of experimental subjects, which will shed light on thalamocortical connections and local cortical networks.

2. Computing the dynamic mode decomposition (DMD)

In Section 2.1, we briefly summarize the DMD algorithm (Rowley et al., 2009; Schmid, 2010; Tu et al., 2013). We expand on connections between DMD and related methods, including its relationship with the Koopman operator, in Section 2.1.3. To build some intuition, DMD is compared to principal components analysis (PCA) and independent components analysis (ICA) on a synthetic noisy timeseries dataset. We then detail a few adaptations useful for its application to large-scale recordings of neural activity, including instructions for how to choose parameters for use in the computations. These adaptations are DMD with an augmented data matrix (Section 2.2) and the DMD “power” spectrum (Section 2.4). Section 2.5 explores the robustness and accuracy of DMD, describing to what extent the decomposition is reliable in the presence of noise and with data subsampling.

2.1. Exact DMD

Consider measurements taken from n observable locations at times $k\Delta t$, where we arrange measurements at snapshot k to make a column vector \mathbf{x}_k . For instance, in the case of electrophysiological recordings, these measurements may be voltages from n channels of an electrode array sampled every Δt .

Gathering measurements from m snapshots in time, we may construct two $n \times (m - 1)$ raw data matrices:

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix}, \\ \mathbf{X}' &= \begin{bmatrix} | & | & & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_m \\ | & | & & | \end{bmatrix}. \end{aligned} \quad (1)$$

Note that \mathbf{X} and \mathbf{X}' contain largely overlapping data, differing in that columns of \mathbf{X}' are shifted one Δt from those in \mathbf{X} .

Let us suppose that there is an unknown linear operator \mathbf{A} such that

$$\mathbf{X}' = \mathbf{A}\mathbf{X}. \quad (2)$$

The *dynamic mode decomposition* of the data matrix pair \mathbf{X} and \mathbf{X}' is given by the *eigendecomposition* of \mathbf{A} . We may think of \mathbf{A} as describing a high-dimensional linear regression of the nonlinear dynamics that relate \mathbf{X} to \mathbf{X}' .

To obtain an approximation of \mathbf{A} , one possible approach is to use the singular value decomposition (SVD) of the data matrix $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^*$ to compute its pseudoinverse:

$$\mathbf{A} \approx \mathbf{X}'\mathbf{X}^\dagger \triangleq \mathbf{X}'\mathbf{V}\Sigma^{-1}\mathbf{U}^*. \quad (3)$$

However, if n is large, computing the eigendecomposition of the $n \times n$ matrix \mathbf{A} may be prohibitively expensive, so that this approach is not readily scalable.

In addition, we assume that there exists some low-dimensional spatial structure in \mathbf{X} . Under this assumption, the following procedure allows the DMD modes and eigenvalues to be calculated without direct computation of \mathbf{A} .

DMD algorithm (Tu et al., 2013)

1. Compute the SVD of our first data matrix, $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^*$. We may now make the substitution into Eq. (2) and write $\mathbf{X}' = \mathbf{A}\mathbf{U}\Sigma\mathbf{V}^*$.
 2. Define $\mathbf{A} \triangleq \mathbf{U} * \mathbf{A}\mathbf{U} = \mathbf{U} * \mathbf{X}'\mathbf{V}\Sigma^{-1}$.
 3. Compute the eigendecomposition of \mathbf{A} , $\mathbf{A}\mathbf{W} = \mathbf{W}\Lambda$, where \mathbf{W} is the matrix of eigenvectors, and Λ is the diagonal matrix of eigenvalues λ_i . Each eigenvalue λ_i is a DMD eigenvalue.
 4. Compute the DMD modes, $\Phi \triangleq \mathbf{X}'\mathbf{V}\Sigma^{-1}\mathbf{W}$. Each column of Φ is a DMD mode ϕ_i corresponding to eigenvalue λ_i .
-

Finally, we may write an approximation of the observed data as a simple dynamic model $\hat{\mathbf{X}}(t)$,

$$\hat{\mathbf{X}}(t) = \Phi \exp(\Omega t)\mathbf{z}, \quad (5)$$

where $\Omega = \log(\Lambda)/\Delta t$, t is time, and \mathbf{z} is set of weights to match the first time point measured such that $\mathbf{x}_1 = \Phi\mathbf{z}$ (this equation can be solved by using the pseudoinverse of Φ).

The key feature of the above algorithm is that it decomposes data, arranged as in Eq. (1), into a set of *coupled spatial-temporal modes*. Note that Φ and Λ are both complex valued, and that $\hat{\mathbf{X}}$ has, in general, non-zero imaginary components; if the raw data \mathbf{X} is strictly real valued, then we may consider only the real component of $\hat{\mathbf{X}}$.

2.1.1. Data reconstruction from DMD modes

To see that the dynamic model reconstruction in Eq. (5) follows from the DMD algorithm above, we make use of the theorem that each pair of DMD modes and eigenvalues ϕ_i and λ_i are eigenvector/eigenvalues of \mathbf{A} (Theorem 1 in Tu et al. (2013)), satisfying

$$\mathbf{A}\phi_i = \lambda_i\phi_i.$$

We may expand each snapshot of data \mathbf{x}_k as a linear combination of the DMD modes ϕ_i ,

$$\mathbf{x}_k = \sum_i c_{ik}\phi_i,$$

for some constants c_{ik} . Since we have $\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$,

$$\mathbf{x}_{k+1} = \sum_i c_{ik}\mathbf{A}\phi_i = \sum_i c_{ik}\lambda_i\phi_i.$$

Further, $\mathbf{x}_k = \mathbf{A}^{k-1}\mathbf{x}_1$, so we may write

$$\mathbf{x}_k = \sum_i \mathbf{A}^{k-1}\phi_i z_i = \sum_i \lambda_i^{k-1}\phi_i z_i = \Phi\Lambda^{k-1}\mathbf{z},$$

where \mathbf{z} is a set of constants satisfying the expansion of the first data snapshot \mathbf{x}_1 such that $\mathbf{x}_1 = \Phi\mathbf{z}$.

Let us now make a change of units from data snapshots k (observed at every Δt) to units of time t , and also define $\Omega = \log(\Lambda)/\Delta t$, so that we have

$$\mathbf{x}(t) = \Phi \exp(\Omega t)\mathbf{z}.$$

This last expression is Eq. (5).

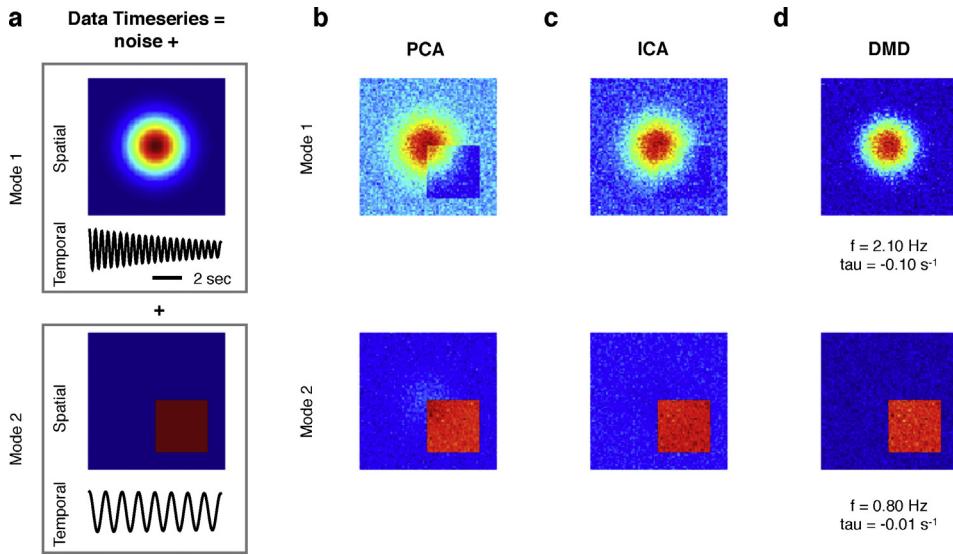


Fig. 1. A comparison of DMD with common modal decomposition algorithms on a synthetic timeseries dataset. (a) The dataset is a movie with 6400 pixels in each frame, and this noisy, high-dimensional time series dataset has two underlying, overlapping patterns, a Gaussian oval and a square. Each mode also has a distinct temporal evolution that includes both growth/decay and oscillation. The magnitude of the noise is $0.75 \times$ the magnitude of the signals. (b) PCA derives modes that mix the underlying modes. (c) ICA modes more closely resemble the generative modes, but the two underlying modes are still mixed. (d) DMD extracts the spatial-temporal coherent modes in the movie. These DMD modes closely resemble the underlying spatial modes and provide an estimate of the temporal evolution of these patterns.

2.1.2. DMD, PCA and ICA on a synthetic dataset

To build some intuition of DMD modes, Fig. 1 compares results of modal decomposition by PCA, independence components analysis (ICA, Hyvärinen and Oja (2000)), and DMD on a synthetic timeseries dataset. The timeseries dataset is a movie 10 s in duration sampled at 50 frames per second, where each frame is a $80 \times 80 = 6400$ pixel image (so $n=6400$ and $m=500$). The dataset is constructed to be the sum of two generative modes, each with a spatial pattern that evolves according to some coherent temporal dynamics (Fig. 1a). Mode 1 is a Gaussian oval in space that oscillates and decays in time; Mode 2 is a square that oscillates at a lower frequency than the oval does. The two modes are spatially overlapping. Each mode's magnitude is of range $[-1, 1]$, and independent noise drawn from a Gaussian distribution $\mathcal{N}(0, 0.75)$ was added at each pixel.

Fig. 1b-d shows the first two modes computed by each method. As shown in Fig. 1b, PCA derives modes that mix the two generative modes in Fig. 1a. These PCA modes are vectors in \mathbb{R}^n , ordered by their ability to explain the greatest fraction of variance in the data; PCA assumes the data is distributed as a multi-dimensional Gaussian. Fig. 1c shows that ICA can potentially do better than PCA, but the two generative modes are still mixed. ICA mode 1 (top of Fig. 1c) contains the Gaussian oval with a shadow of the square. Unlike PCA, ICA modes are computed assuming the underlying signals are non-Gaussian and statistically independent.

In contrast, DMD is an explicitly temporal decomposition and takes the sequences of snapshots into account, deriving spatial-temporal coherent patterns in the movie. DMD modes are closely related to PCA modes and also assumes variance in the data is Gaussian. The two largest DMD modes not only closely resemble the two generative modes, but they also contain an estimate of the temporal dynamics of the two modes, including an estimate of their frequencies of oscillation and time constant of exponential growth/decay. These temporal parameters are computed from the DMD eigenvalues by Eq. (6) as explained in Section 2.4. Further, the computational complexity of DMD is within the same order of magnitude as that of PCA.

2.1.3. Connections to related methods

DMD has deep mathematical connections to Koopman spectral analysis. The Koopman operator is an infinite-dimensional, linear

operator that represents finite-dimensional, nonlinear dynamics. The eigenvalues and modes of the Koopman operator capture the evolution of data measuring the nonlinear dynamical system (Budišić et al., 2012; Mezić, 2005). DMD is an approximation of Koopman spectral analysis (Rowley et al., 2009), so that DMD modes are able to describe even nonlinear systems.

As an algorithm, it is convenient to think of spatial-temporal decomposition by DMD as a hybrid of static mode extraction by principal components analysis (PCA) in the spatial domain and discrete Fourier transform (DFT) in the time domain. In fact, DMD modes are a rotation of PCA space such that each basis vector has coherent dynamics. The DMD algorithm in Section 2.1 starts with a SVD of the data matrix $\mathbf{X} = \mathbf{U}\Sigma\mathbf{V}^*$ as the first step, where \mathbf{U} are identical to PCA modes. DMD modes are eigenvectors of $\mathbf{A} = \mathbf{U}^*\mathbf{A}\mathbf{U}$, so that we can think of \mathbf{A} as the correlation between PCA modes \mathbf{U} and PCA modes in one time step \mathbf{AU} . Like the DFT, DMD extracts frequencies of oscillations observed in the measurements. In addition, DMD goes beyond DFT to also estimate rates of growth/decay, where the DFT eigenvalues always have magnitudes of exactly one.

The general formulation of the high-dimensional timeseries problem is related to several methods in the statistics literature, including vector autoregression (VAR, Charemza and Deadman (1992)). DMD differs from VAR in that the \mathbf{A} matrix in Eq. (2) is never explicitly estimated, but rather we seek its eigendecomposition by computing \mathbf{A} in step 2 of the DMD algorithm. The resultant modes are interpreted as a low-rank dynamical system expressed in Eq. (5). Further, these modes represent separable spatiotemporal features of the data. Interestingly, this approach of computing \mathbf{A} is mathematically related to Principal Components Regression (PCR, Jolliffe (2005)).

2.1.4. Additional properties and practical limitations

A few general properties of DMD are interesting to note. The data \mathbf{X} may be real or complex valued; in the case of recordings from electrode arrays, we will proceed assuming \mathbf{X} are real valued measurements of voltage. Further, the decomposition is unique (Chen et al., 2012), and it is also possible to compute the DMD of non-uniformly sampled data (Tu et al., 2013).

The relationship of DMD to PCA and DFT points to a few limitations of the technique that guide its application. DMD spatial

modes are based on PCA modes; therefore, relatively local spatial correlations (affecting only a very small number of measurement locations) that do not contain a lot of energy may be less likely to emerge as coherent modes in both PCA and DMD. Since the formulation DMD contain no explicit spatial relationship between neighboring measurements in \mathbf{X} , it is difficult to capture traveling waves with a few coherent DMD modes without transformation to a different spatial basis.

In the temporal domain, DMD converts timeseries information into a sum of complex sinusoids. It follows that nonlinear temporal dynamics that are not well approximated by $e^{\omega t}$, where ω is a complex number and t is time, will not be well approximated by DMD. The possible evolution of $e^{\omega t}$ includes functions of exponential growth or decay with some sinusoidal oscillation (for instance, see the temporal evolution of Modes 1 and 2 in Fig. 1a), so DMD is ill-suited to capture transient dynamics (even so, see a multi-resolution extension of DMD in Kutz et al. (2015)). In practice, DMD is used as a windowed technique, and we use caution in choosing the sizes of the windows such that the $e^{\omega t}$ approximation is useful. Interestingly, while the temporal window size constrains the DFT in the lowest frequency it is able to capture, DMD is not subject to the same constraint on the lower bound. Still, the sampling rate of the data does constrain the highest frequency DMD is able to capture, as dictated by Nyquist sampling.

2.2. Implementing DMD with an augmented data matrix

DMD was originally used in the study of large fluid flow fields, where typically $n \gg m$. In contrast, in neuroscience we are often interested in electrode arrays that have tens of channels sampled at hundreds of samples per second, so for a data matrix \mathbf{X} over a 1-s window of data, $n < m$. The SVD of \mathbf{X} produces v non-zero singular values, where v is the smaller of n and $m - 1$. This property restricts the maximum number of DMD modes and eigenvalues to n , which is often too few to fully capture the dynamics over m snapshots in time.

The solution to this rank mismatch is to construct augmented versions of the data matrices, appending to the snapshot measurements with $h - 1$ time-shifted versions of themselves, thus augmenting the number of measurements to be hn .

Specifically, we construct a new augmented data matrix \mathbf{X}_{aug} ,

$$\mathbf{X}_{\text{aug}} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_{m-h} \\ | & | & & | \\ | & | & & | \\ \mathbf{x}_2 & \mathbf{x}_3 & \cdots & \mathbf{x}_{m-h+1} \\ | & | & & | \\ & \vdots & & \\ | & | & & | \\ \mathbf{x}_h & \mathbf{x}_{h+1} & \cdots & \mathbf{x}_{m-1} \\ | & | & & | \end{bmatrix};$$

and similarly for \mathbf{X}'_{aug} . It is important to note that in shift-stacking the snapshots, we preserve the $\mathbf{X}' = \mathbf{A}\mathbf{X}$ relationship in Eq. (2) for all rows (measurements) of the new augmented data matrix. The data matrices augmented by shift-stacking is inspired by the Hankel matrix as constructed in the Eigenvalue Realization Algorithm (ERA, Juang and Pappa (1985)). A Hankel matrix is constant across its skew diagonal; similarly, \mathbf{X}_{aug} and \mathbf{X}'_{aug} have shifted repeats of \mathbf{x}_k in blocks across its skew diagonal direction.

DMD is then applied as described in Section 2.1 using \mathbf{X}_{aug} and \mathbf{X}'_{aug} instead of \mathbf{X} and \mathbf{X}' . Thus computed, the spatial DMD modes

Φ are now $hn \times m$ matrices; even so, because the augmented data matrices are h time-shifted stacks of the same raw data, the computed DMD modes are also stacks of h repeats.

2.3. DMD modes of ECoG data

Fig. 2 illustrates the separation of spatial modes Φ from their temporal dynamics Λ for an example window of ECoG array recordings. In this example, the 44-channel dataset with 100 time points (0.5 s sampled at 200 Hz) was stacked $h = 5$ times to achieve an accurate decomposition of the data. Comparing Fig. 2a and b, we see that the reconstruction $\hat{\mathbf{X}}$ as computed by Eq. (5) captures most spatial and dynamic features of the raw data; Fig. 2c shows the error of the reconstruction.

Fig. 2d represents DMD reconstruction of the data as the multiplication of complex-valued matrices, following Eq. (5). Here complex-valued $\Phi \in \mathbb{R}^{hn \times m}$ and $\Lambda \in \mathbb{R}^{m \times m}$ are visualized in their real-valued and imaginary parts using separate colormaps. Each column of Φ corresponds to one element on the diagonal of Λ . The magnitude (absolute value) and phase of each element of ϕ_i represents spatially coherent activation and the relative phase of this activation across the n electrodes, respectively. The magnitude and phase of each eigenvalue λ_i contain information about time dynamics; their interpretation and use are discussed in Section 2.4. Since a stacking depth of $h = 5$ is used in this example, the dashed horizontal lines break Φ into 5 stacks, emphasizing that the modes are likewise 5 stacks of essentially identical repeats.

To decide the appropriate parameters for number of stacks h , Fig. 3a shows an examination of the DMD approximation error as a function of h ; this plot uses the same example window of data as in Fig. 2a. The approximation error E was defined as

$$E = \frac{|\mathbf{X} - \hat{\mathbf{X}}|_F}{|\mathbf{X}|_F},$$

where $|M|_F = \sqrt{\sum_j \sum_i |m_{ij}|^2}$ is the Frobenius norm. The error decreased as the number of modes included in the approximation r increased. This decrease reached a plateau for all $h \geq 5$, so that further stacking did not significantly increase the accuracy of the DMD approximation.

Fig. 3b shows mean and variation of DMD approximation error over 100 random windows of data for which $n = 44$ and $m = 100$. Fig. 3c and d quantifies the mean and standard deviation of approximation error for different sizes of data matrices, using 100 random windows of data each. As the degree of stacking h increases, the error decreases.

2.3.1. Recommended parameter values

In general, to determine the stacking depth parameter h , we recommend choosing the smallest integer h such that $hn > 2m$, where n is the number of channels in the recording and m is the number of snapshots in the windowed data matrix. This value of h is shown as dashed lines in Fig. 3b-d. We observed that this degree of stacking and truncation provided enough DMD modes to capture the observed dynamics, as quantified by the reconstruction error (Fig. 3) reaching an asymptotic (or nearly asymptotic) level. The resultant augmented DMD modes ϕ_i are now vectors of hn elements; Fig. 2d shows that the computed DMD modes are also stacks of h repeats. Generally, the first n elements of each mode are used in subsequent analyses.

2.4. The DMD power spectrum

Each spatial DMD mode ϕ_i has a corresponding eigenvalue λ_i that describes its temporal dynamics. Specifically, the rate of

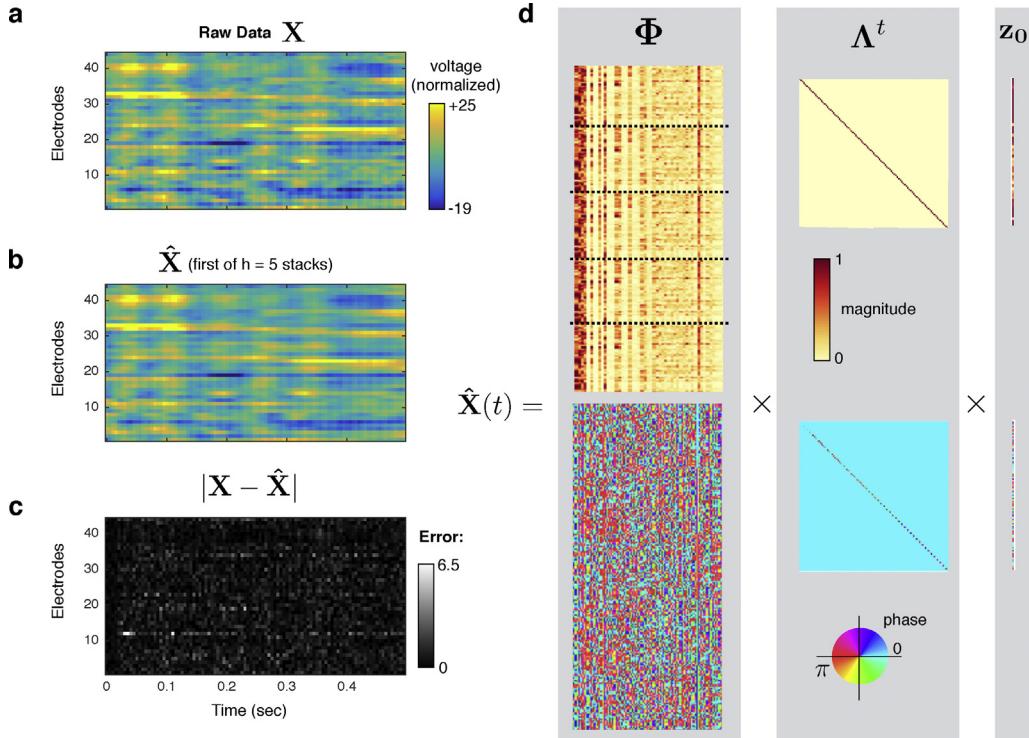


Fig. 2. Coupled spatial–temporal decomposition of ECoG data computed by DMD. The data used in this example, shown in (a) as a heat map where the colors represent (normalized) voltage, has $n = 44$ channels and spans 0.5 s of recordings acquired at 200 Hz (so $m = 100$ snapshots). To compute the DMD of \mathbf{X} , we used $h = 5$ stacks. Comparing (a) to the DMD approximation in (b) shows the reconstruction error (c) is relatively small. Note that (b) visualizes only the real part of \mathbf{X} , since our original data is strictly real valued. The spatial modes $\Phi \in \mathbb{R}^{hn \times m}$ and temporal components $\Lambda \in \mathbb{R}^{m \times m}$ are visualized in (d). Since these modes are all complex valued, their magnitude and phase information are visualized separately. The dashed lines break Φ into $h = 5$ stacks, emphasizing that modes are 5 stacks of the same repeats.

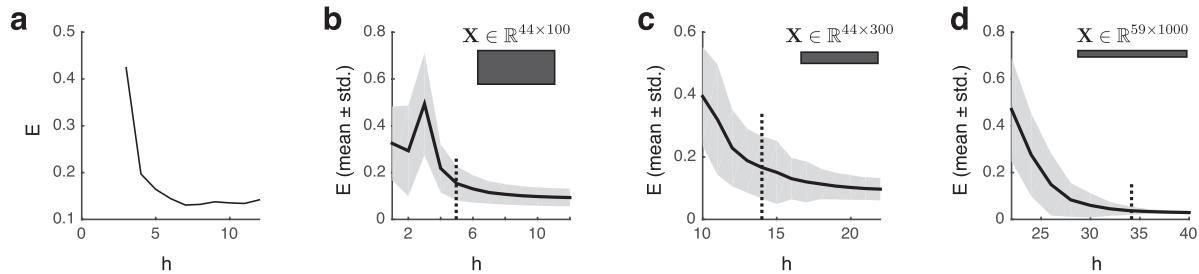


Fig. 3. The accuracy of the DMD approximation improves as the degree of stacking in the augmented data matrix h . (a) The error of the approximation E as defined in Section 2.3 is plotted as a function of h for the example raw data shown in Fig. 2a. The error of DMD approximation is plotted for raw data of several example dimensions $\mathbf{X} \in \mathbb{R}^{n \times m}$ in (b–d); the gray boxes illustrate dimensions of \mathbf{X} and have aspect ratios $n \times m$. In each figure, the mean and standard deviation of E is shown over 100 random time points of ECoG recordings. The dashed lines indicate the recommended value of h as described in Section 2.3.1.

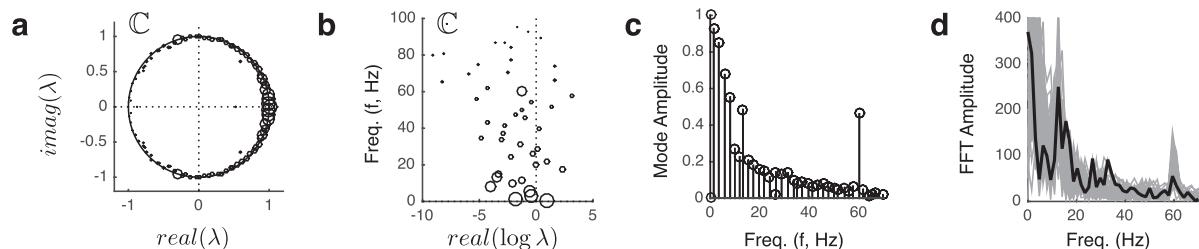


Fig. 4. The DMD spectrum recovers dynamics of the data, revealing the growth/decay and oscillations of each mode. This figure shows the spectrum computed for the same data as in Fig. 2. (a) and (b) Eigenvalues λ and $\log \lambda$ are visualized as complex values, where each open circle is an eigenvalue and its diameter is proportional to the amplitude of the corresponding spatial mode (Eq. (7)). The vertical axis of (b) shows frequency f as defined by Eq. (6), which has units of cycles/sec (Hz). Another way to show the same spectrum is in (c), which plots mode amplitude as a function of f . This view is the most easily comparable to (d), the amplitude of oscillations at a range of frequencies as computed by the Fourier transform. Each gray line in (d) shows the FFT computed separately for each channel of the recording and the solid black line is their mean.

growth/decay and frequency of oscillation are reflected in the magnitude and phase components of λ_i , respectively (Fig. 2d). The distribution of DMD eigenvalues λ may be visualized relative to the unit circle on the complex plane, as in Fig. 4a.

As we can see from Eq. (5), the dynamical equation derived from DMD modes and eigenvalues has a $\exp(\omega_i t)$ term, where $\omega_i = \log(\lambda_i)/\Delta t$ and t is time. The sign of the real component of ω determines if the dynamics of the corresponding mode are growing, decaying, or stable. The imaginary component of ω determines the frequency of oscillations. To make the phase of each eigenvalue more interpretable in the context of brain oscillations, we can make a change of units to cycles per second (Hz). Specifically,

$$f_i = \left| \frac{\text{imag}(\omega_i)}{2\pi} \right| \quad (6)$$

gives the frequency of oscillation of mode ϕ_i in units of cycles per second (Fig. 4b and c). In the case of ECoG data, since the raw data is strictly real valued, the decomposition produces complex conjugate pairs of eigenvalues and modes.

The definition of DMD as the eigendecomposition of \mathbf{A} in Eq. (2) allows for arbitrary scaling of the DMD modes. The algorithm outlined in Section 2.1 produces modes with unit norm. To use the magnitude of these modes in a way similar to the power spectrum for selecting modes with high energy, we must choose an appropriate scaling.

One useful scaling is to multiply each mode by the corresponding singular value from the SVD. We modified step 3 of the DMD algorithm to scale the modes by energy as follows.

-
- 3†. Compute the eigendecomposition of $\hat{\mathbf{A}} = \Sigma^{-1/2} \mathbf{A} \Sigma^{1/2}$,
 $\hat{\mathbf{A}}\hat{\mathbf{W}} = \hat{\mathbf{W}}\Lambda$,
where $\hat{\mathbf{W}}$ is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues. The eigenvalues of $\hat{\mathbf{A}}$ are identical to the eigenvalues of \mathbf{A} , and
 $\mathbf{W} = \Sigma^{1/2}\hat{\mathbf{W}}$
are eigenvectors of \mathbf{A} that we will use in step 4. These eigenvectors are scaled by $\Sigma^{1/2}$, so they do not have unit norm.
-

To visualize the DMD spectrum, we plot the “power” of each mode ϕ_i against its frequency of oscillation f_i . The mode amplitude of ϕ_i is defined as the square of its vector magnitude:

$$P_i = |\phi_i|_2^2. \quad (7)$$

In Fig. 4a and b, the sizes of the dots indicate this mode amplitude. Fig. 4c illustrates the intuition that the DMD spectrum qualitatively resembles the power spectrum computed by the FFT for a 0.5 s window of ECoG data, shown for comparison in Fig. 4d. Despite this resemblance, it is important to remember that the power spectrum is computed independently for every channel of the recording, whereas every point in the DMD spectrum corresponds to a specific spatial mode across all channels.

2.5. Robustness and accuracy of DMD

The DMD reliably extracts coupled spatial-temporal modes under subsampling of the raw data. As an illustrative example, Fig. 5 shows the effects of subsampling on one mode extracted from one window of data. Fig. 5a shows the DMD spectrum for a 1-s window of data ($n = 59$ channels) sampled at 1 kHz ($m = 1000$), and Fig. 5b–e shows the DMD spectra for the same window of data subsampled by factors of 2, 3, 4 and 5 (so that $m = 500, 333, 250$ and 200). To be explicit, subsampling by a factor of 2 consists of keeping every other time snapshot in the data matrix \mathbf{X} . The degree of stacking was determined as described in Section 2.3.1. The eigenvalues are not identical, but the spectra qualitatively resemble each other.

Next, we examined the spatial DMD mode corresponding to one particular eigenvalue, indicated by the red cross in every spectrum. This mode is chosen as the mode with the largest magnitude between 25 and 26 Hz. The magnitude and phase of this particular mode across all subsamples is shown in Fig. 5f–j and k–o, respectively. Each mode is plotted in space as a 8×8 spatial grid of electrodes, arranged as they were placed on the brain surface during the recording. Although these modes are shown as an abstract grid here for clarity, the raw data is the same as used in the sensorimotor map analysis in Section 4.1, and their positions are rendered relative to subject A’s brain structure as shown in Fig. 8. Note that there had been 64 electrodes in the grid, but only $n = 59$ channels with good recordings are included in this analysis.

Across all subsamples, mode magnitudes $|\Phi|$ and mode phases $\angle\Phi$ remain very similar to each other and are highly correlated, as quantified in Fig. 5p and q. All correlation coefficients are large and positive (>0.75) with highly significant p -values ($p < 10^{-6}$). The correlation coefficients for $\angle\Phi$ are computed considering that the values are circular.

2.5.1. Robustness of spatial modes to subsampling the data

Across a random ensemble of windowed ECoG data, correlations of mode amplitude and angle remain high across subsamples. Fig. 6a and b shows the correlation coefficients of $|\Phi|$ and $\angle\Phi$ as a function of sampling rate for four different example mode frequencies, averaged across 100 random windows of data. Each window comprised of 1 s of data for $n = 59$ channels and, before subsampling, was sampled at 1000 Hz. When subsampled by a factor of 1, 2, 3, 4 and 5, the resultant data matrix \mathbf{X} had $m = 1000, 500, 333, 250$ and 200 snapshots, respectively. For each window of data, the DMD spectrum is computed, similar to what is shown in Fig. 5a–e, and the largest amplitude mode whose frequency was within 1 Hz of 5, 15, 25 and 35 Hz was chosen and compared to the corresponding mode derived with no subsampling. In other words, for each window of data, we computed the correlation coefficients in the top rows of Fig. 5p and q. On average, these modes computed under subsampling remain highly correlated to modes computed with no subsampling.

Since the decomposition reliably extracts spatially coherent patterns, it follows that if we are interested in spatial modes of frequency f , then the minimum sampling frequency of the dataset obeys Nyquist sampling and should be larger than $2f$. In practice, a minimum sampling frequency of closer to $3f$ may be preferable, as we observe some decrease in correlation coefficients for larger mode frequencies.

2.5.2. Consistency of spatial modes to additive noise

We perform a similar analysis of spatial mode magnitude and phase correlation as a function of additive noise. We restrict our description here to the effects of additive noise to (already noisy) ECoG data. Fig. 6c and d shows correlation coefficients of $|\Phi|$ and $\angle\Phi$ for increasing noise magnitude for four different example mode frequencies, averaged across 100 random windows of data. Similar to the analysis described in Section 2.5.1, each window of data comprised of half a second of data across $n = 59$ channels and sampled at 100 Hz (so $m = 500$), the correlation coefficients were computed to the corresponding mode from the same window of data with no additive noise. The additive noise was drawn independently from Gaussian distributions of zero mean and standard deviation as specified on the horizontal axis. Since the raw data before adding noise has been normalized so that each channel has unit standard deviation, Fig. 6c and d shows that the correlation coefficients remain high for noise standard deviations up to at least 1/10 of the mean signal magnitude.

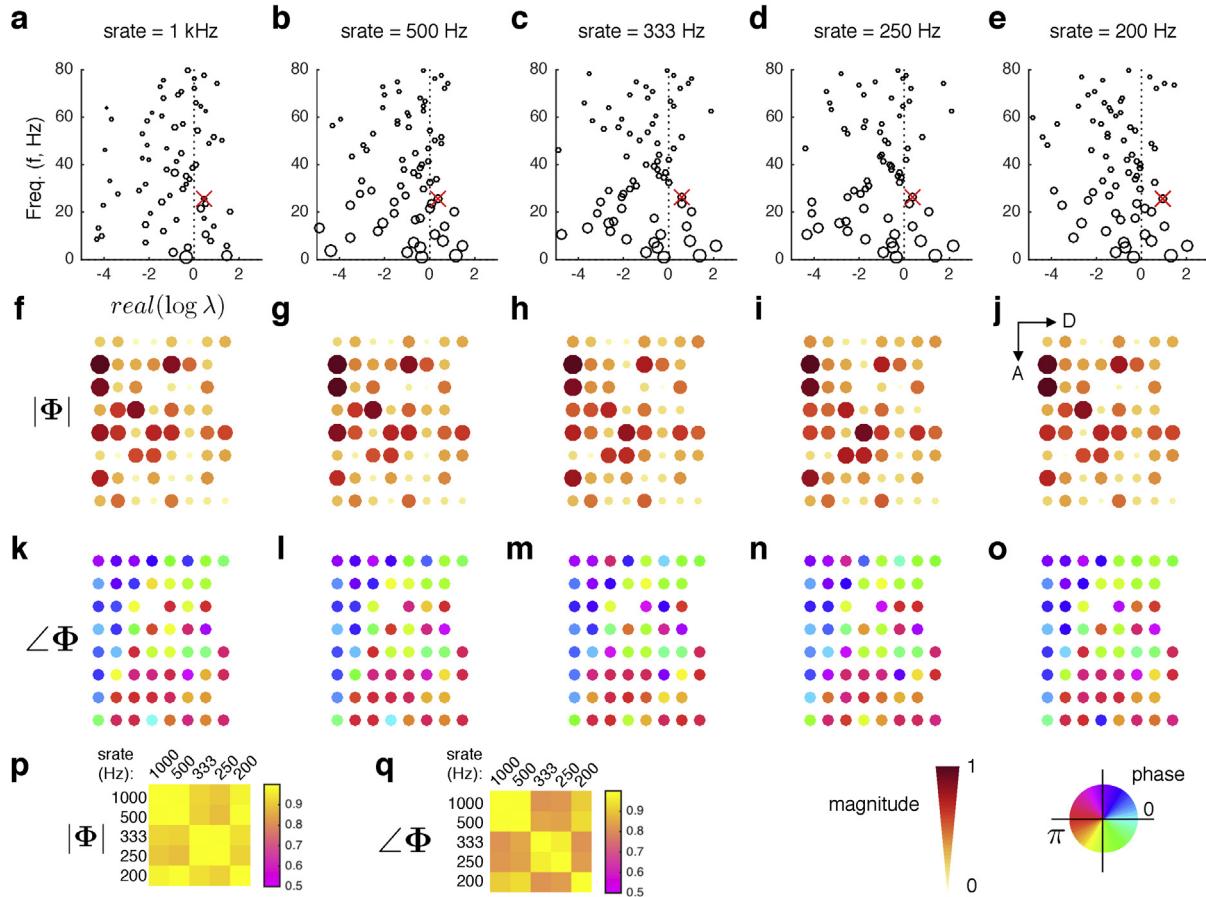


Fig. 5. Modes derived by DMD are robust to subsampling the raw data. DMD was computed for a window of ECoG data ($n=59$ channels) sampled at 1 kHz, and the same window of data subsampled by factors of 2, 3, 4 and 5. (a)–(e) show DMD spectra plotted the same way as in Fig. 4b. Next, we examine one particular spatial mode from each of spectra, marked by a red cross. The magnitude $|\Phi|$ (f)–(j) and phase $\angle\Phi$ (k)–(o) of this mode is consistent across subsamples. Modes are visualized as 8×8 grids of electrodes, and colors are coded as in the bottom right. Their similarity is quantified by pair-wise cross-correlations (p)–(q). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

3. Materials and methods

3.1. Collection and preprocessing of ECoG recordings

3.1.1. Subjects

All subjects were patients undergoing long-term electrocorticography (ECoG) monitoring in preparation for surgical treatment of intractable epilepsy. Data were collected from two subjects (female subjects A and B, ages 29 and 11 respectively) with subdural platinum electrode arrays (Ad-Tech, Racine, WI). Electrodes had a 2.3 mm exposed surface diameter and were spaced at 1 cm.

Subjects were patients at the University of Washington and gave informed consent according to the protocol approved by the Institutional Review Board of that institution.

3.1.2. Recording and data preprocessing

The ECoG signals were recorded by the XLTEK (Natus Medical Incorporated, San Carlos, California) clinical monitoring system at a sampling rate of 500 or 2000 Hz. The standard system parameters impose a high-pass filter at about 0.1 Hz. For the motor mapping, recordings were high pass filtered above 6 Hz and down sampled to 1000 samples/s. Further, each channel of recording was

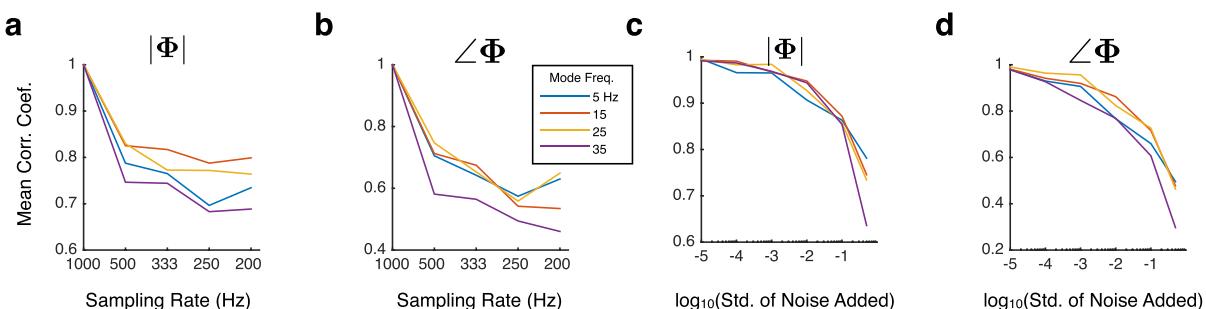


Fig. 6. Modes derived by DMD are robust to subsampling the raw data and additive noise. (a) and (b) The mean correlation coefficients of mode magnitude $|\Phi|$ and phase $\angle\Phi$ are computed over 100 random windows of data under subsampling. This comparison was repeated for modes at 4 different frequencies $f=5, 15, 25$ and 35 Hz. (c)–(e) A similar mode correlation analysis was repeated for additive Gaussian distributed noise.

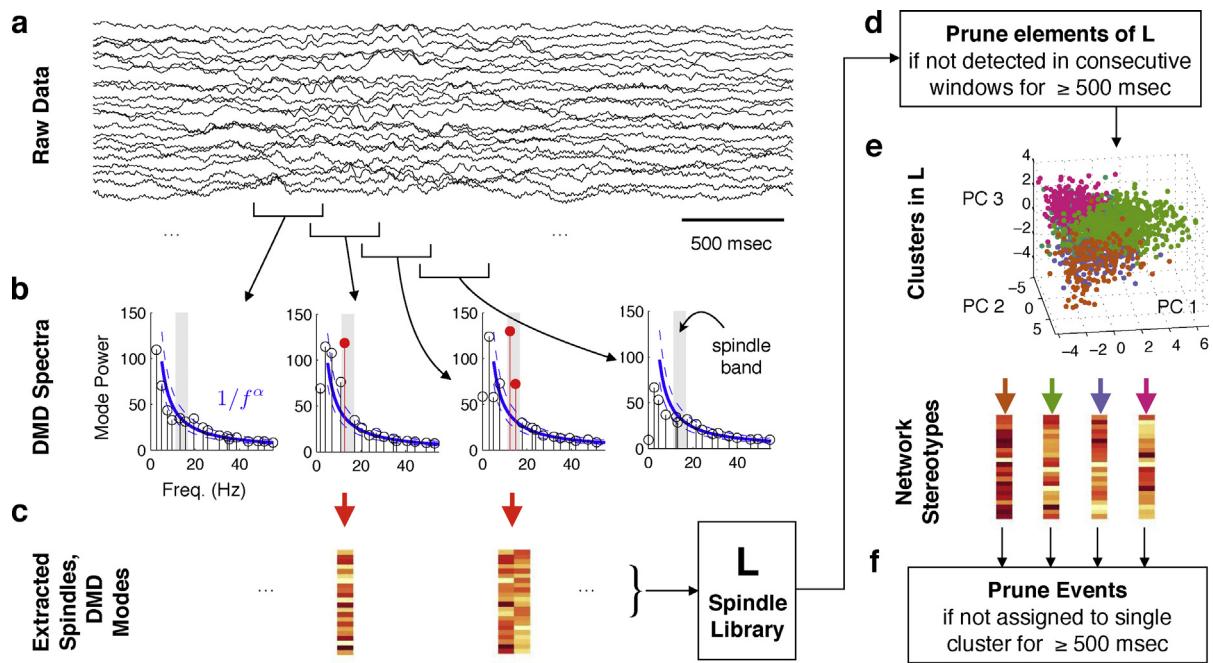


Fig. 7. Schematic of methods for spindle network detection and characterization. (a) ECoG data shown as voltage traces in time where different electrodes are stacked vertically. (b) A DMD spectrum is computed for each window of data, and modes in the spindle band (gray bar) with magnitudes larger than expected from a $1/f^{\alpha}$ fit of the ongoing spectrum (blue lines) are detected (modes highlighted in red). (c) These candidate spindle modes are gathered over the entire recording into a spindle library **L**, and elements in **L** are pruned if they not detected in consecutive windows for ≥ 500 ms (d). (e) The remaining elements in **L** are clustered using a mixture of Gaussians models, and the centroids of each cluster are consider the stereotypes of spindle networks. Each instance in **L** is assigned to a spindle network cluster and consolidated into spindle events. (f) Events shorter than 500 ms in duration are pruned and ignored from further analysis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

normalized to have unit standard deviation. The total length of the motor screen recording was approximately 4 min. For spindle network extraction, recordings were bandpass filtered between 6 and 80 Hz, and down sampled at 200 samples/s; the total length of recording of a sleep epoch from each subject was approximately 20–30 min. Subject A had an electrode array with 64 channels; Subject B had an electrode array with 48 channels.

3.2. Sensorimotor mapping

3.2.1. Motor screen task

ECoG data was collected while a simple motor screen task was performed as described previously in Miller et al. (2007). Briefly, the patient was instructed to not move (baseline), to move their tongue, or to move the hand contralateral to the implanted electrode array. Twenty (20) trials of each movement were repeated in pseudo-random order with instructions to not move in between; each instruction lasted three (3) seconds. The data used in this analysis was from Subject A.

3.2.2. Sensorimotor mapping by DMD

To map cortical areas whose activity is modified during movement of the tongue and hands in a frequency-dependent way, we divided the ECoG array recording into baseline (no instructed movement), tongue movement, and hand movement trials. The middle 2 s of each trial (between 500 and 2500 ms from the beginning of the instruction) where used in the sensorimotor mapping analysis. In each trial, the 2 s recording window over all channels was decomposed by DMD; the time series data were augmented by stacking as described in Section 2.2. The DMD spectrum was examined to extract spatial modes in two frequency ranges, a low frequency band (8–32 Hz) and a high frequency band (75–100 Hz). All the absolute value of DMD modes in each frequency range was averaged to extract the mean mode magnitude. Next, for both the

low and high frequency bands, we compared the mean mode across all trials for hand and tongue movements by subtraction of the mean baseline modes.

Fig. 8 shows the z-scores of relative changes in mean mode magnitude as assessed by bootstrapping. To estimate the variation in mode magnitude in frequency bands of interest unrelated to this task, the trial labels (no movement, tongue movement, or hand movement) were shuffled and the same movement minus baseline analysis was repeated for each set of shuffled data. This random shuffle was repeated 1000 times, and the standard deviation of movement minus baseline mode magnitude at each electrode was computed. Next, the z-score of change from baseline was computed at each electrode based on these bootstrapped estimates of standard deviation.

3.3. Spindle network extraction

To detect the presence of spindle networks, we first separated the ECoG recordings of sleep into 300 ms windows in time and decomposed the windowed data using DMD. Section 3.3.2 described how we used the resultant DMD spectrum to identify spectra characteristic of spindle-associated activity. We used overlapping windows, sliding 50 ms between consecutive windows. Next, Section 3.3.3 described how spatial DMD mode magnitudes corresponding to excess power in the spindle band were collected, and stereotypes of spindle networks were discovered by clustering using a Gaussian mixtures model.

3.3.1. Sleep identification

Non-REM sleep epochs were identified by inspection using the increased power in the delta band (1–6 Hz) and verified by the presence of K-complexes and spindles. For every sleep epoch, all electrode traces were normalized by z-scoring with respect to the amplitude in the 5–50 Hz range. This frequency range was chosen

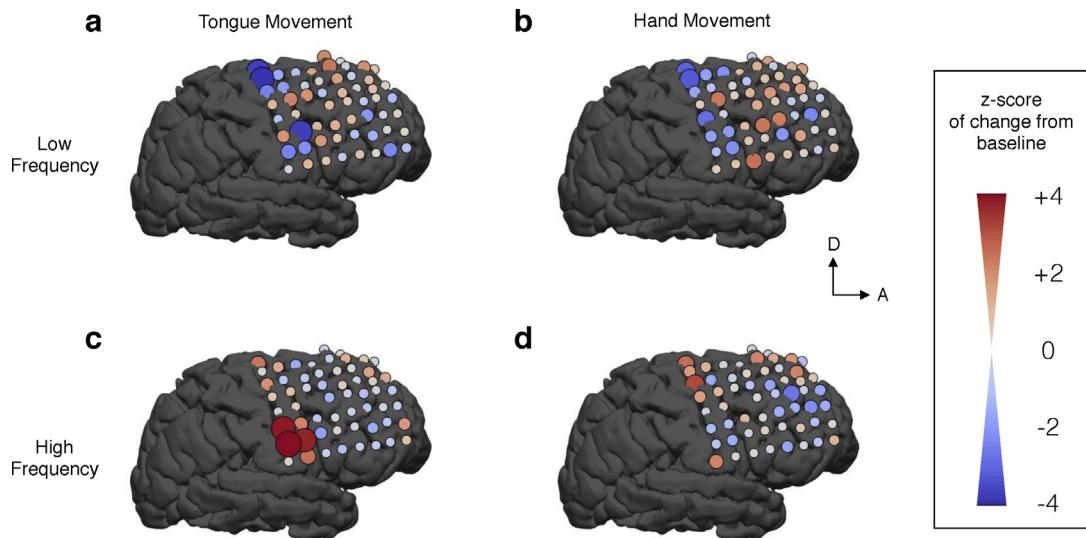


Fig. 8. Sensorimotor map as derived from DMD modes. Relative to baseline, during movement, sites over sensorimotor cortex (the most posterior columns of the electrode array) showed decreased DMD mode magnitudes in the low frequency range (8–32 Hz, a and b), and increased magnitudes in the high frequency range (76–100 Hz, c and d). Z-scores were computed by bootstrapping. Consistent with previous literature, the sites of increased high frequency power were separable for tongue (c) and hand (d) movements. Data from Subject A was used in this analysis.

to eliminate the variable amplitude effects of K-complexes, which have a maximum amplitude in medial-frontal regions (Wennberg, 2010).

3.3.2. Spindle detection criteria

The DMD power spectrum allows us to formulate a principled criterion to identify the presence of increased power in the spindle band (14 ± 5 Hz). It has been well described that the power spectral density of recordings from the brain follows a $1/f^\alpha$ distribution, where α is a positive scalar (Bedard et al., 2006; Miller et al., 2009). We used this observation to fit the α of the DMD spectrum, excluding power at frequencies below 5 Hz and above 57 Hz. This $1/f^\alpha$ distribution was fit over the cumulative DMD spectra of each epoch of a subject's recording (usually about 30 min) using robust regression (Holland and Welsch, 1977), which rejected outliers due to sporadic electrical noise in the recordings. Windows including presumptive epileptiform activity were rejected based on a generalized elevation in power over all frequencies.

For each window of data, a DMD mode ϕ was considered part of a spindle network if and only if its power significantly exceeded $2.5 \times$ standard deviations of the $1/f^\alpha$ fit (upper blue dashed lines in Fig. 7b); in the spectra shown in Fig. 7b, these modes are shown highlighted red. The magnitude of spatial DMD modes passing this detection criteria are gathered to form a library of putative spindle networks \mathbf{L} (Fig. 7c). Further, to reduce the incidence of spurious identification of spindle networks, significant power in the spindle band must be detected in consecutive, overlapping windows for ≥ 500 ms (Fig. 7d).

3.3.3. Clustering spindle networks

Elements of the library of spindle networks \mathbf{L} were clustered into distinct types of spindle networks. Specifically, we constructed a library of DMD modes \mathbf{L} ,

$$\mathbf{L} = \begin{bmatrix} | & | & | \\ \phi_1 & \phi_2 & \cdots & \phi_N \\ | & | & | \end{bmatrix}.$$

For purposes of clustering, we considered only the absolute value of DMD modes and each mode was normalized to unit length. Clusters were determined in r -dimensional principal components space,

using the projections of each column of $|\mathbf{L}|$ onto the first r principal components of $|\mathbf{L}|$:

$$|\mathbf{L}| = \mathbf{U}_L \Sigma_L \mathbf{V}_L^*,$$

$$\mathbf{a} = \mathbf{U}_r^T \mathbf{L},$$

where \mathbf{U}_r^T is the transpose of the first r columns of \mathbf{U}_L .

As illustrated schematically in Fig. 7e, we used a Gaussian mixtures model to group columns of \mathbf{a} as points in \mathbb{R}^r into k clusters (Murphy, 2012). Since this approach is unsupervised, we selected the appropriate value of k in the model using the Bayes Information Criterion (BIC), a statistical metric to guide comparison of models with different numbers of parameters that punishes overfitting. Briefly, a model with a smaller BIC is more parsimonious with the available data than a model with a larger BIC. We evaluated descriptions of $|\mathbf{L}|$ using a Gaussian mixtures model in r dimensions using k and systematically varied r from 2 to 10 and varied k from 2 to 10. The number of stereotyped spindle networks k was chosen to minimize BIC of this family of Gaussian mixtures model. For the two subjects we tested, the median k for which BIC was minimal was 4 clusters for both subjects. We chose to fit the Gaussian mixtures model with $r=5$ reduced dimensions, because increasing r beyond 5 did not significantly change the assignment of clusters.

After clustering, each element of \mathbf{L} is assigned to a single cluster (a stereotype of spindle network). These assignments are consolidated into spindle events, defined as a block of continuous windows of data assigned to the same cluster. A final step (Fig. 7f) prunes events shorter than 500 ms in duration.

3.4. Software availability

In the interest of reproducibility, a software implementation of DMD as described in this work is available as a set of Matlab scripts. The code can be found at github.com/bwbrunton/dmd-neuro/.

4. Results

4.1. Sensorimotor maps

We analyzed datasets recorded from sub-dural electrode arrays implanted in two patients monitored for seizure loci localization.

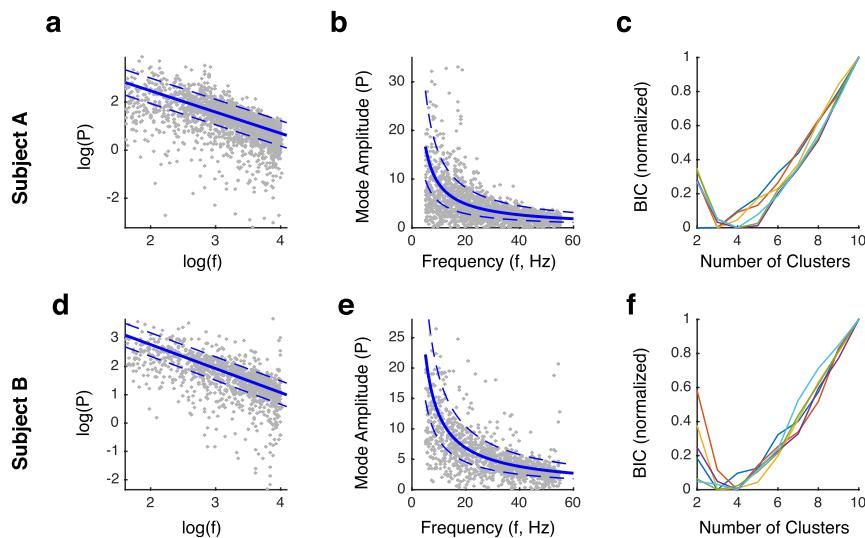


Fig. 9. Automated spindle detection threshold and clustering parameters for subjects A (top row) and B (bottom row). The spindle detection thresholds were computed by robust regression of a linear fit to $\log(P)$ and $\log(f)$ collected over all windowed DMD spectra of the sleep recordings. The results of the fit (blue solid lines) and the extent of the $\pm 2.5 \times$ standard deviation of the residuals (blue dashed lines) are shown in (a) and (b) and (d)–(e) for data from Subjects A and B, respectively. (c) and (f) Normalized BIC values computed as a function of number of clusters for various values r of the reduced dimensionality of \mathbf{L} , shown as differently colored lines. The number of clusters for which the BIC was minimal was consistently close to 4, which was chosen as the most parsimonious number of spindle stereotypes in the recording for both subjects. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

To validate the DMD approach to spatial–temporal pattern analysis, we derived sensorimotor maps based on a simple motor repetition task where the subjects were instructed to move either their tongue or the hand contralateral to the implanted electrode array. This task was previously analyzed for a large cohort of patients, and a spectral analysis of each channel revealed consistent sites in sensorimotor cortex whose power changed specifically for each movement Miller et al. (2007).

We performed a related analysis based on the DMD spectrum and derived motor maps of tongue and hand movements that are closely consistent with the previously described results of Miller et al. (2007). DMD spectra were computed for each epoch of the motor screen, and DMD modes within a low frequency band (8–32 Hz) and a high frequency band (76–100 Hz) were averaged to derive mean modes in each band. Next the baseline low frequency and high frequency mean modes were subtracted from the movement epochs to produce the frequency-dependent sensorimotor maps. Fig. 8 shows these frequency-dependent sensorimotor maps as z-scores of change from baseline, as described in Section 3.2.2.

4.1.1. Hand and tongue movements led to differentiable changes at low and high frequencies

In the low frequency band, there was a generalized decrease in correlation across sensorimotor cortex in both the tongue and hand movement epochs. These decreases are visualized by decreases in relative DMD mode magnitude across the most posterior columns of the electrode array (Fig. 8a and b). In the high frequency band, we found relatively local groups of electrodes within sensorimotor cortex where DMD mode amplitude increased selectively for tongue and hand movement. The foci for hand movement were more dorsal than the tongue movement sites (Fig. 8c and d), consistent with known organization of the human sensorimotor cortex.

4.2. Sleep spindle networks

In a novel application of DMD, we developed methods to extract and identify sleep spindle networks in human ECoG recordings. Our approach to spindle network extraction is based on the DMD spectra of windowed ECoG data (Fig. 7). DMD modes corresponding

to larger than expected power in the spindle band were collected in a library, whose elements were then clustered to determine stereotypes of spindle networks in the recording. To consider all spindle-type activity in an inclusive way, we defined the spindle band to be 14 ± 5 Hz. Each spindle network stereotype represents strength of spatial correlations between electrodes and may be visualized on a grid of electrode locations.

We applied our spindle network extraction algorithm to ECoG recorded during non-REM sleep from sub-dural electrodes implanted over the temporal lobe of two human subjects. Subject A had a 64-channel electrode grid implanted over the right hemisphere, and Subject B had a 48-channel electrode grid implanted over the left hemisphere.

4.2.1. Spindle detection thresholds and clustering parameters

As described in Section 3.3.2, detection of candidate spindle networks was based on larger than expected power of the windowed DMD spectrum in the spindle band. The expected power in this band was computed as a fit to all the DMD spectra in the duration of the recording (frequency f versus mode amplitude P). Fig. 9a and d shows these fits for Subjects A and B, respectively, where the data points in $\log(f)$ and $\log(P)$ was fit by a robust linear regression. Fig. 9b and e shows the same fits plotted as P as a function of f , where the distribution was proportional to $1/f^\alpha$ where α was 0.8838 and 0.8473 for Subjects A and B, respectively. Candidate spindle networks were required to have power above $2.5 \times$ the standard deviation of residuals in the regression, shown as the dashed blue lines in Fig. 9.

After gathering all putative spindle networks in the spindle library \mathbf{L} , we used a BIC analysis to determine the number of clusters (spindle networks stereotypes) most parsimonious with the data. Fig. 9c and f shows the normalized BIC values as a function of number of clusters for the two subjects' datasets. The different colored lines show different dimensionality of subspaces r on which the BIC analysis was computed (see Section 3.3.3). For each colored line, the range of BIC values were normalized to be visualized between 0 and 1. Fortunately, although using different values of r led to slightly different numbers of clusters that optimized the BIC, the results were closely consistent regardless of r . For Subject A,

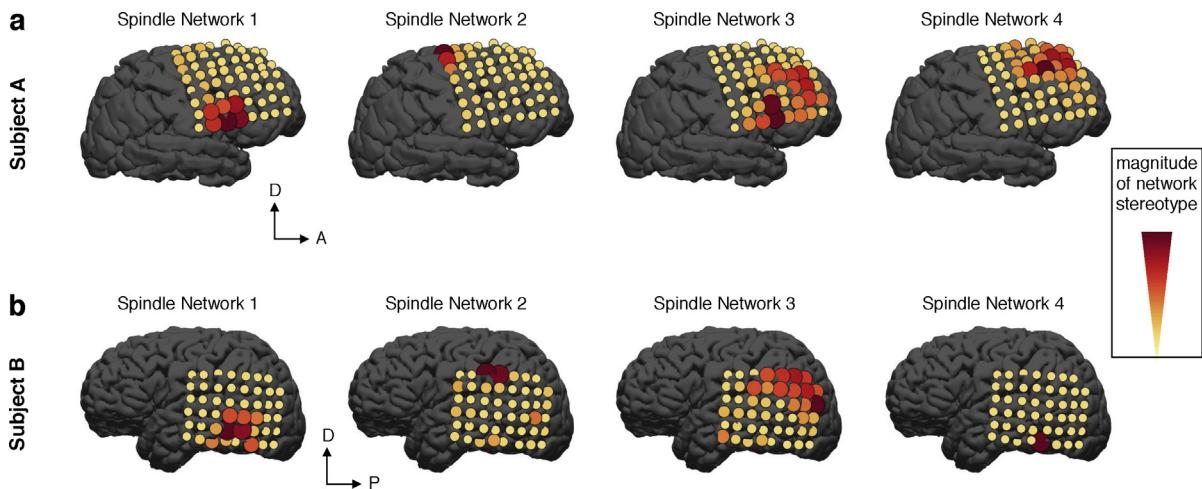


Fig. 10. Visualization of spindle network stereotypes for two subjects, computed from ECoG recordings during non-REM sleep. The spindle networks are shown on the ECoG recording grid, where both the size and color of dots at each electrode represent the strength of the spatial correlation between electrodes (see color bar).

the most parsimonious number of clusters was between 3 and 5, so we chose the median value of 4. Similarly for Subject B, the most parsimonious number of clusters was either 3 or 4, so we chose to round the median value of 3.5 up to 4 clusters.

4.2.2. Spindle networks have separable but overlapping spatial patterns

The spindle network detection and clustering approach found distinct stereotypes of spindle networks, whose spatial distributions are visualized for Subjects A and B in Fig. 10. These stereotypes are taken as the centroids of the clusters from the clustering analysis, reshaped based on the arrangement of electrodes within the arrays and visualized on top of a structural rendering of the brain. The spindle networks are not spatially exclusive; in other words, some electrodes participate in multiple spindle networks. For instance, compare spindle networks 3 and 4 for Subject A, which have many electrodes in common despite clearly separable distributions of mode weights. These observations quantify previous observations that spindle-band activity co-occurred at different electrodes in a number of distinct patterns, as noted by Johnson et al. (2012) and Nir et al. (2011). A much more detailed exploration of our findings about spindle networks is in a companion manuscript (Johnson et al. (2015)).

Two specific examples of spindle events classified as different stereotypes are shown in Fig. 11. The electrode array recordings around the spindle event are shown as normalized voltage recordings over time, where different electrodes are offset vertically for

clarity of visualization. Gray traces show non-spindle parts of the recording, and the spindle event is highlighted using the same colormap as in Fig. 10. During each event, many electrodes across the grid had sinusoidal-type oscillations that co-occurred. These oscillations were of different magnitudes, and the color of each voltage trace corresponds to the relative magnitudes as extracted by the DMD modes. In Fig. 11b, the spindle event is preceded by a K-complex; it is well known that spindle events often follow K-complexes. The K-complex was correctly rejected as a non-spindle event despite a generalized elevation in power overlapping with the spindle band.

4.2.3. Basic statistics of spindle events

In addition to separation of spindle activity to distinct locations on the cortex, spindle networks were also distinct from each other by their characteristic frequencies of oscillation. Importantly, these frequencies were not used in the unsupervised clustering. Fig. 12a and e shows the distribution of frequencies for spindle events of each network stereotype as box plots. Outliers were omitted from these plots. For Subject A, frequency distributions for all networks were pair-wise statistically significantly different from each other (except for networks 3 versus 4), as assessed by *t*-tests, with $p < 10^{-3}$. For Subject B, spindle network 4 had a frequency that was statistically significantly different from the other networks (*t*-test, $p < 10^{-3}$).

The different networks also differed in their occurrence and duration. For Subject A (Fig. 12b), network 1 occurred less often

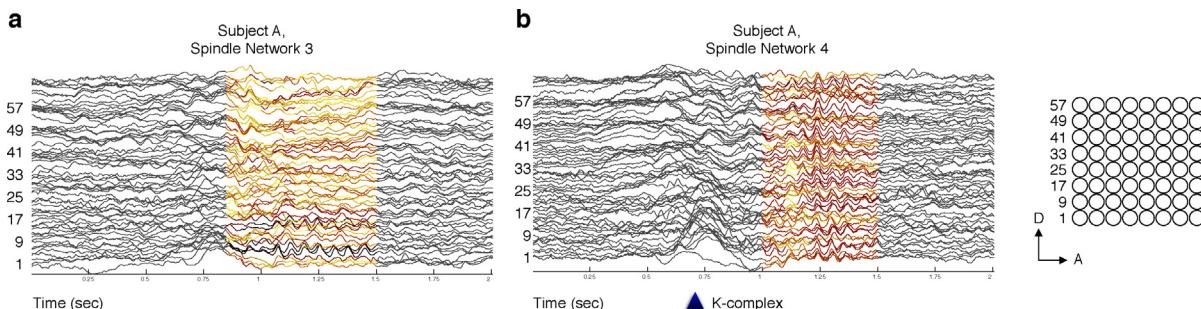


Fig. 11. Examples of spindle events from Subject A belonging to network 3 (a) and network 4 (b) in Fig. 10. The normalized recorded voltage at each electrode is shown as a function of time, where different electrodes are offset vertically for clarity. The gray traces are non-spindle periods, and the highlight regions are detected spindle events. The color of traces reflect the magnitude of the DMD mode in the spindle band, using the same colormap as in Fig. 10. The schematic at right gives the mapping of electrode numbers to their relative positions in the grid for subject A. Note in (b) that the spindle event followed a K-complex, which was correctly rejected as a non-spindle event. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

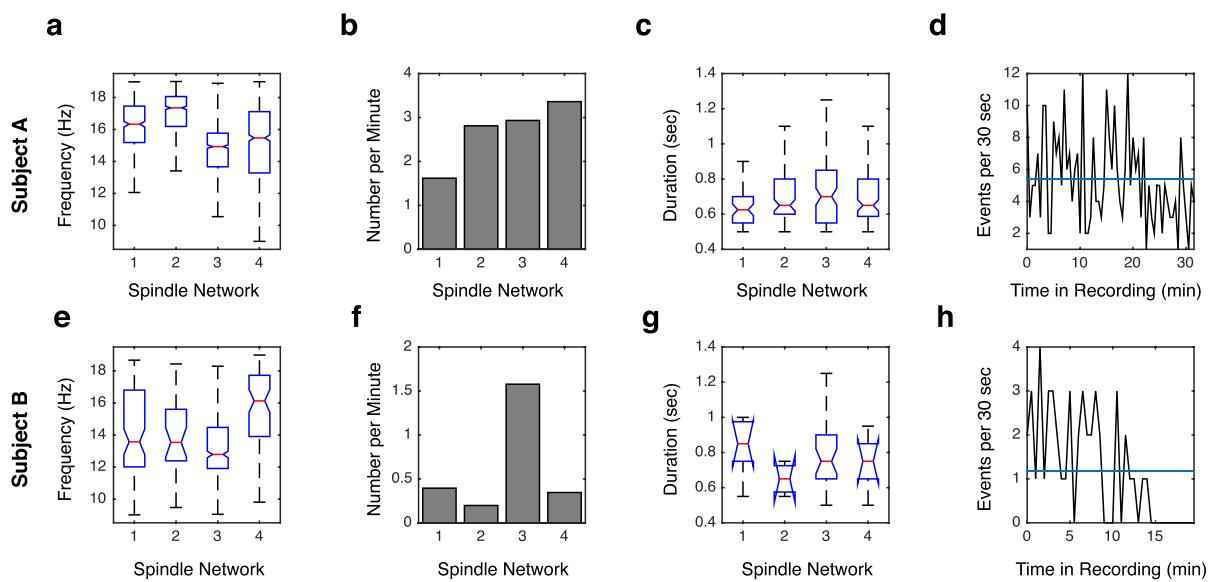


Fig. 12. Basic statistics of spindle events. For Subject A (top row) and Subject B (bottom row), spindle events belonging to each spindle network stereotype were quantified by their carrying frequency (a) and (e), occurrence per minute (b) and (f), and duration (c) and (g). In addition, the total number of events (of all network stereotypes) as a function of time in recording is shown in (d) and (h). The mean number of events per 30 s is shown as the blue horizontal line.

than the other networks by a factor of approximately half. For Subject B (Fig. 12f), network 3 occurred more often than the other networks by a factor of approximately three. Fig. 12c and g show the distribution of durations of events for each spindle network. Since we imposed a minimum duration of 500 ms (Section 3.3.3), this was the shortest possible event duration. Fig. 12d and h plot the total number of events (of all network stereotypes) as a function of time in the sleep recording, at a 30-s resolution. In both subjects, the number of events decreased towards the end of the recording; it is possible this decrease reflects a transition to a different state, including a non-REM sleep stage, REM sleep, or wakefulness.

The characteristic frequencies of networks from Subject A were generally higher than those from Subject B. Further, we detected more total spindle events in Subject A than in Subject B. These differences likely reflect both variations between individuals and electrode array placement.

5. Discussion

In this work, we presented the first application of a modal decomposition technique known as dynamic mode decomposition (DMD) to analyze and visualize large-scale neural recordings. Such neural recordings are often characterized by coherent patterns of activity in space (across many channels) at multiple temporal frequencies. We described how DMD can be used to extract these coherent patterns by decomposing the data into a low-dimensional representation in both space and time. This extraction was validated on a simple motor task where selective and separable regions of sensorimotor cortex were identified. Although DMD has previously been described in studying high-dimensional dynamical systems in a different discipline (Rowley et al., 2009; Schmid and Sesterhenn, 2008; Schmid, 2010), here we described in detail its adaption to use with ECoG recordings, providing rules-of-thumb for its computational implementation and intuition for interpretation of the extracted modes in the context of neural recordings.

Next, we presented a novel method to detect and analyze sleep spindle networks from large-scale human ECoG recordings; automated approaches to characterize these stereotyped spindle networks had not been previously reported, and the high-dimensionality of the dataset makes manual scoring by experts

an infeasible task. We showed that our method reliably detected spindle networks, which are correlated groups of electrodes that showed significantly power in the spindle band (14 ± 5 Hz). Further, we found multiple stereotypes of spindle networks for two different subjects. Spindle networks were distinct from each other in spatial localization as well as carrying frequency, frequency of occurrence, and duration. These sleep spindle networks are discovered independently of an explicit behavior and may represent functionally connected cortical areas. Although it has not been possible to make any general observations across these two subjects due to limited sample size and large variability in electrode location, our automated method enables these analyses to be performed across many subjects, and its computation scales favorably with increasing numbers of channels in the electrode array recordings.

It was our goal to develop a spindle network detection and characterization algorithm that minimizes fine-tuning of parameters while maximizing accuracy. Unfortunately, it was not possible to assess the accuracy of our algorithm against a “gold standard” labeled dataset, as such a dataset for sleep spindles in large-scale ECoG does not exist. Most experts who curate sleep data do not typically identify spindles in subdural grid recordings. Further, it would be extremely odious to identify spindles on individual electrodes and also group them into distinct networks. To the best of our knowledge, our algorithm reliably identifies most spindle-type events detected by the authors’ visual inspection of the data, and it correctly rejects non-spindle events, including ictal events and K-complexes. It remains possible that there are other types of epileptiform activity that we have not controlled for, although such activity would only impact our analysis if they have significant power in the spindle range. Our further explorations of these networks are described in a separate manuscript (Johnson et al., 2015).

Dimensionality reduction is a useful concept in building meaningful models based on dynamics of neuronal networks because there exists low-dimensional structures in large-scale data (Cunningham and Byron, 2014). Ideas from compressive sensing further suggest that such models may be accessible even from limited data, when the system is known to be sufficiently sparse (Ganguli and Sompolinsky, 2012). Extensions of DMD that exploit

underlying data sparsity (including Jovanović et al. (2014) and Brunton et al. (2013)) have potential to expand the usefulness of the framework to a larger class of incomplete measurements with improved robustness.

In contrast to other static modal decomposition techniques, DMD provides not only modes, but also a relatively low-dimensional, efficient linear model for how the most dynamically important modes evolve in time. In other words, the DMD approximation of the dataset (Eq. (5)) is also a prediction of the system's dynamic trajectory. Such dynamic models of high-dimensional dynamic data are *equation-free*, in the sense that they are entirely data-driven and do not rely on a set of governing equations. The theory of DMD is well grounded in dynamical systems, so that this modal decomposition technique can be used to assess system stability and provide a meaningful basis on which reduced order dynamic models of large-scale recordings can be made, for example in neural decoding tasks.

Since the pioneering work of Kalman, linear dynamical systems have underpinned many of our analytic techniques to understand noisy, dynamic data, including use of the Kalman Filter for optimal state estimation. Traditionally, these techniques have been developed for use in physical systems where the equations of motion are well known. DMD interfaces dynamical systems theory with modern “big-data” measurements and complex systems where it is infeasible or impossible to acquire true equations of motion; this perspective has been called the “dynamics of observables” in the review by Budišić et al. (2012). In particular, the dynamic model derived by DMD can serve the place of the traditional linear dynamical system in a Kalman Filter framework. Model predictions do not typically hold for long term; however, even short-term DMD predictions can inform design of feedback control in complex, non-linear systems (Potter et al., 2014).

We propose that DMD is generally useful for the analysis, understanding, and visualization of large-scale neural recordings. In particular, we are exploring DMD-based methods to characterize network-level responses across multiple brain areas to complex stimulus inputs. We envision this spatial-temporal modal decomposition may be applied to a variety of modalities of neural measurements, including ECoG, MEG, functional MRI, calcium imaging, LFP, and spike rates of a large collection of single units.

Author's contribution

B.W.B. and L.A.J. conceived of the project; L.A.J. collected the sleep data. B.W.B. designed the analyses and implemented the algorithms, with contributions from L.A.J. J.G.O. and J.N.K. supervised the project. B.W.B. wrote the manuscript, with input from all authors.

Funding

This research was supported by National Institutes of Health Grant R01 NS065186-01 and award number EEC-1028725 from the National Science Foundation.

Acknowledgements

We thank S.L. Brunton and J.L. Proctor for comments on the manuscript; J.P. Dyhr, T.L. Daniel, and R. Curtu for constructive discussions; D. Witten for pointing out the connection between DMD and PCR; and T. Blakely for collecting the dataset used for the sensorimotor mapping.

References

- Bedard C, Kroeger H, Destexhe A. Does the 1/f frequency scaling of brain signals reflect self-organized critical states? *Phys Rev Lett* 2006;97(11):118102.
- Berger E, Sastuba M, Vogt D, Jung B, Ben Amor H. Estimation of perturbations in robotic behavior using dynamic mode decomposition? *Adv Robot* 2015;29(5):331–43.
- Broome BM, Jayaraman V, Laurent G. Encoding and decoding of overlapping odor sequences? *Neuron* 2006;51(4):467–82.
- Brunton SL, Proctor JL, Kutz JN. Compressive sampling and dynamic mode decomposition; 2013 arXiv:1312.5186.
- Budišić M, Mohr R, Mezić I. Applied Koopmanism. *Chaos: Interdiscip J Nonlinear Sci* 2012;22(4):047510.
- Buzsáki G, Draguhn A. Neuronal oscillations in cortical networks? *Science* 2004;304(5679):1926–9.
- Byron MY, Cunningham JP, Santhanam G, Ryu SI, Shenoy KV, Sahani M. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *Adv Neural Inform Process Syst* 2009:1881–8.
- Charemza WW, Deadman DF. New directions in econometric practice: general to specific modelling, cointegration and vector autoregression. E. Elgar; 1992.
- Chen KK, Tu JH, Rowley CW. Variants of dynamic mode decomposition: boundary condition, Koopman, and Fourier analyses. *J Nonlinear Sci* 2012;22(6):887–915.
- Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, et al. Neural population dynamics during reaching. *Nature* 2012.
- Clemens Z, Fabo D, Halasz P. Overnight verbal memory retention correlates with the number of sleep spindles? *Neuroscience* 2005;132(2):529–35.
- Cunningham JP, Byron MY. Dimensionality reduction for large-scale neural recordings. *Nat Neurosci* 2014.
- De Gennaro L, Ferrara M. Sleep spindles: an overview? *Sleep Med Rev* 2003;7(5):423–40.
- Eschenko O, Mölle M, Born J, Sara SJ. Elevated sleep spindle density after learning or after retrieval in rats? *J Neurosci* 2006;26(50):12914–20.
- Freeman J, Vladimirov N, Kawashima T, Mu Y, Sofroniew NJ, Bennett DV, et al. Mapping brain activity at scale with cluster computing. *Nat Methods* 2014.
- Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence? *Trends Cogn Sci* 2005;9(10):474–80.
- Gais S, Mölle M, Helms K, Born J. Learning-dependent increases in sleep spindle density? *J Neurosci* 2002;22(15):6830–4.
- Ganguli S, Sompolinsky H. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. *Annu Rev Neurosci* 2012;35:485–508.
- Holland PW, Welsch RE. Robust regression using iteratively reweighted least-squares? *Commun Stat Theory Methods* 1977;6(9):813–27.
- Holmes P, Lumley JL, Berkooz G. Turbulence, coherent structures, dynamical systems and symmetry. Cambridge University Press; 1998.
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond Ser A: Math Phys Eng Sci* 1998;454(1971):903–95.
- Hyvärinen A, Oja E. Independent component analysis: algorithms and applications? *Neural Netw* 2000;13(4):411–30.
- Johnson LA, Blakely T, Hermes D, Hakimian S, Ramsey NF, Ojemann JG. Sleep spindles are locally modulated by training on a brain-computer interface? *Proc Natl Acad Sci U S A* 2012;109(45):18583–8.
- Johnson LA, Weaver KE, Brunton BW, Wunder JD, Hakimian S, Ojemann JG. Spatio-temporal sleep spindle networks reflect the intrinsic connectivity revealed by waking behavior and resting state fMRI; 2015 [in review].
- Jolliffe I. Principal component analysis. Wiley Online Library; 2005.
- Jovanović MR, Schmid PJ, Nichols JW. Sparsity-promoting dynamic mode decomposition. *Phys Fluids* 2014;26(2):024103.
- Juang J-N, Pappa RS. An eigensystem realization algorithm for modal parameter identification and model reduction. *J Guid Control Dyn* 1985;8(5):620–7.
- Kutz JN, Fu X, Brunton SL. Multi-resolution dynamic mode decomposition; 2015 arXiv:1506.00564.
- Liang H, Bressler SL, Desimone R, Fries P. Empirical mode decomposition: a method for analyzing neural data. *Neurocomputing* 2005;65:801–7.
- Machens CK, Romo R, Brody CD. Functional, but not anatomical, separation of “what” and “when” in prefrontal cortex. *J Neurosci* 2010;30(1):350–60.
- Mezić I. Spectral properties of dynamical systems, model reduction and decompositions. *Nonlinear Dyn* 2005;41(1–3):309–25.
- Miller KJ, Leuthardt EC, Schalk G, Rao RP, Anderson NR, Moran DW, et al. Spectral changes in cortical surface potentials during motor movement? *J Neurosci* 2007;27(9):2424–32.
- Miller KJ, Sorensen LB, Ojemann JG, Den Nijs M. Power-law scaling in the brain surface electric potential. *PLoS Comput Biol* 2009;5(12):e1000609.
- Murphy KP. Machine learning: a probabilistic perspective. MIT Press; 2012.
- Nir Y, Staba RJ, Andrillon T, Vyazovskiy VV, Cirelli C, Fried I, et al. Regional slow waves and spindles in human sleep? *Neuron* 2011;70(1):153–69.
- Potter SM, El Hady A, Fetzer EE. Closed-loop neuroscience and neuroengineering. *Front Neural Circuit* 2014;8:115.
- Proctor JL, Eckhoff PA. Discovering dynamic patterns from infectious disease data using dynamic mode decomposition? *Int Health* 2015;7(2):139–45.
- Raghavachari S, Kahana MJ, Rizzuto DS, Caplan JB, Kirschen MP, Bourgeois B, et al. Gating of human theta oscillations by a working memory task? *J Neurosci* 2001;21(9):3175–83.

- Ray LB, Fogel SM, Smith CT, Peters KR. **Validating an automated sleep spindle detection algorithm using an individualized approach?** *J Sleep Res* 2010;19(2):374–8.
- Rehman N, Mandic DP. **Multivariate empirical mode decomposition.** *Proc R Soc A: Math Phys Eng Sci* 2009; rspa20090502.
- Rowley CW, Mezić I, Bagheri S, Schlatter P, Henningson DS. **Spectral analysis of nonlinear flows.** *J Fluid Mech* 2009;641:115–27.
- Rudrauf D, Douiri A, Kovach C, Lachaux J-P, Cosmelli D, Chavez M, et al. **Frequency flows and the time-frequency dynamics of multivariate phase synchronization in brain signals.** *Neuroimage* 2006;31(1):209–27.
- Schimicek P, Zeithofer J, Anderer P, Saletu B. **Automatic sleep-spindle detection procedure: aspects of reliability and validity?** *Clin EEG Neurosci* 1994;25(1):26–9.
- Schmid P, Sesterhenn J. **Dynamic mode decomposition of numerical and experimental data.** *Bull Am Phys Soc* 2008.
- Schmid PJ. **Dynamic mode decomposition of numerical and experimental data.** *J Fluid Mech* 2010;656:5–28.
- Steriade M, McCormick DA, Sejnowski TJ. **Thalamocortical oscillations in the sleeping and aroused brain?** *Science* 1993;262(5134):679–85.
- Sweeney-Reed CM, Nasuto SJ. **A novel approach to the detection of synchronization in EEG based on empirical mode decomposition?** *J Comput Neurosci* 2007;23(1):79–111.
- Tu JH, Rowley CW, Luchtenburg DM, Brunton SL, Kutz JN. **On dynamic mode decomposition: theory and applications;** 2013 arXiv:1312.0041.
- Uhlhaas PJ, Singer W. **Abnormal neural oscillations and synchrony in schizophrenia?** *Nat Rev Neurosci* 2010;11(2):100–13.
- Warby SC, Wendt SL, Welinder P, Munk EG, Carrillo O, Sorensen HB, et al. **Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods.** *Nat Methods* 2014;11(4):385–92.
- Welch PD. **The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms.** *IEEE Trans Audio Electroacoust* 1967;15(2):70–3.
- Wennberg R. **Intracranial cortical localization of the human k-complex?** *Clin Neurophysiol* 2010;121(8):1176–86.