

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE**  
**INSTITUTO DO CÉREBRO**

# **CONTRIBUIÇÕES PARA O ESTUDO DO CÓDIGO NEURAL**

**Tese apresentada ao Curso de Pós-Graduação  
em Neurociências, Instituto do Cérebro,  
Universidade Federal do Rio Grande do Norte,  
como requisito parcial para a obtenção do título  
de Doutor em Neurociências.**

Orientador: **Sidarta TG Ribeiro**  
Co-orientador: **Adriano BL Tort**  
Co-orientador no exterior: **Rodrigo Quian Quiroga**

**Natal**

**2015**

**Vítor Lopes dos Santos**

#### FICHA CATALOGRÁFICA

Lopes dos Santos, Vítor

Contribuições para o estudo do código neural – Natal, 2015.

85 páginas

Área de concentração: Neurofisiologia.

Orientador: Prof. Dr. Sidarta Tollendal Gomes Ribeiro.

## *A IDÉIA*

*De onde ela vem?! De que matéria bruta  
Vem essa luz que sobre as nebulosas  
Cai de incógnitas criptas misteriosas  
Como as estalactites numa gruta?!  
Vem da psicogenética e alta luta  
Do feixe de moléculas nervosas,  
Que, em desintegrações maravilhosas,  
Delibera, e depois, quer e executa!*

*Vem do encéfalo absconso que a constringe,  
Chega em seguida às cordas do laringe,  
Tísica, tênue, mínima, raquítica ...*

*Quebra a força centrípeta que a amarra,  
Mas, de repente, e quase morta, esbarra  
No mulambo da língua parálitica.*

**Augusto dos Anjos**

# SUMÁRIO

<i>Resumo</i>	- 5 -
<i>Introdução</i>	- 6 -
<i>Capítulo 1: Detectando assembleias em populações neuronais</i>	- 11 -
<i>Capítulo 2: Extraíndo informação de padrões temporais de disparo</i>	- 30 -
<i>Discussão</i>	- 50 -
<i>Anexo I: Detectando assembleias com análise de componentes principais</i>	- 62 -
<i>Anexo II: On the neurophysiology of serotonergic hallucinogens</i>	- 79 -
<i>Referências</i>	- 90 -

---

## RESUMO

Os recentes avanços técnicos das duas últimas décadas para o registro de sinais neuroeletrofisiológicos foram essenciais para que se testassem hipóteses há muito propostas acerca de como células nervosas processam e armazenam informação. No entanto, ao permitir maior detalhamento dos dados coletados, as novas tecnologias levam inevitavelmente ao aumento de sua complexidade estatística e, conseqüentemente, à necessidade de novas ferramentas matemático-computacionais para sua análise.

Nesta tese, apresentamos novos métodos para a análise de dois componentes fundamentais nas atuais teorias da codificação neural: (1) assembleias celulares, definidas pela co-ativação de subgrupos neuronais; e (2) o padrão temporal de atividade de neurônios individuais. Em relação a (1), desenvolvemos um método baseado em análise de componentes independentes para identificar e rastrear padrões de co-ativação significativos com alta resolução temporal. Superamos limitações de métodos anteriores, ao efetivamente isolar assembleias e abrir a possibilidade de analisar simultaneamente grandes populações neuronais. Em relação a (2), apresentamos uma nova técnica para a extração de padrões de atividade em trens de disparo baseada na decomposição *wavelet*. Demonstramos, por meio de simulações e de aplicação a dados reais, que nossa ferramenta supera as mais utilizadas atualmente para decodificar respostas de neurônios e estimar a informação de Shannon entre trens de disparos e estímulos externos.

---

## INTRODUÇÃO

O trabalho seminal de Adrian e Zotterman (1926) foi o primeiro a demonstrar que células nervosas poderiam transmitir informação, disparando pulsos elétricos (potenciais de ação) em resposta a estímulos externos. Hoje, depois de décadas de pesquisa, acredita-se, axiomaticamente, que a atividade elétrica dos neurônios é o elemento mais fundamental para o processamento de informação no sistema nervoso. O estudo do código neural a que me refiro no título desta tese é, portanto, a investigação científica acerca de como os bilhões de neurônios espalhados nas diversas áreas do encéfalo codificam informação por meio da emissão de potenciais de ação.

Nesses já quase cem anos transcorridos desde os resultados de Adrian e Zotterman, diversas tentativas de compreensão dos mecanismos de codificação têm se sucedido (Abeles, 2009; Buzsáki, 2010; Harris, 2005; Hebb, 1949; Izhikevich, 2006; Marr, 1970; Singer, 1993), produzindo avanços científicos, e compartilhando uma característica fundamental. Todas elas podem ser definidas com base em dois *eixos* ortogonais: o *populacional* e o *temporal*.

O que designo como *eixo populacional* de uma teoria do código neural é o modo de interação entre células que ela propõe. Em um extremo do eixo populacional, situam-se as hipóteses anedóticas de *células da avó* (Gross, 2002); as teorias de codificação *localista*. Para elas, cada *objeto* é codificado por poucas células e cada célula codifica no máximo alguns poucos objetos. Percorrendo esse eixo populacional, encontramos teorias de codificação menos esparsas, até chegarmos às teorias de codificação *densa* ou *populacional*. Nesse outro extremo encontramos a hipótese de *equipotencialidade* apresentada por Karl Lashley (1950), Segundo a qual a representação de qualquer objeto está completamente distribuída no sistema nervoso e o conhecimento do estado de uma célula específica não possui nenhum significado.

Dado o contexto histórico de sua época, uma das maiores contribuições de Donald Hebb no clássico *The Organization of Behavior* (1949) foi apresentar uma teoria posicionada no centro do eixo. A teoria hebbiana das assembleias neurais propõe que a sincronia entre grupos de neurônios seja a fundamentação do processamento de informação no cérebro. A co-ativação de um grupo de células - ou seja, a ativação de uma assembleia - seria a unidade elementar de informação. Dessa forma, neurônios só seriam equipotenciais quando forem parte da mesma assembleia. Do ponto de vista de Lashley, Hebb era localista, já que ele acreditava que a unidade elementar da codificação era um grupo de células com *fronteiras*,

as assembleias. Visto do outro extremo, no entanto, Hebb propõe um código distribuído, pois a representação de objetos estaria espalhada em diversas células.

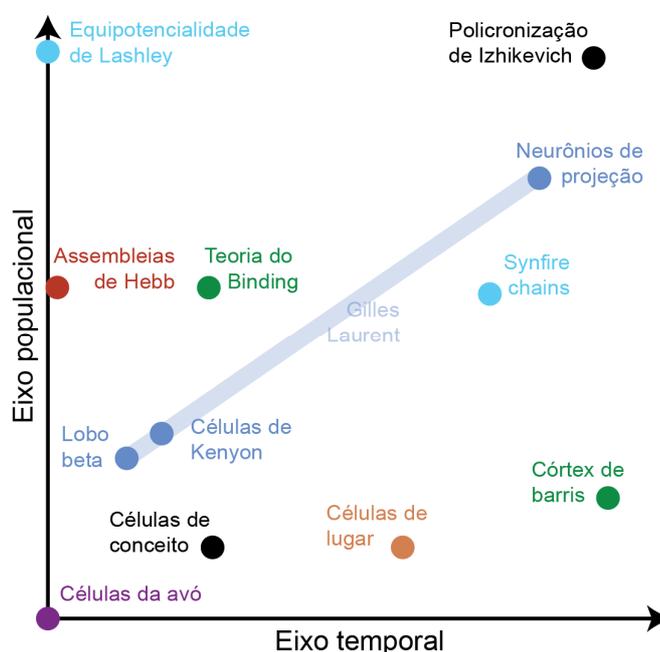
No que concerne ao eixo *temporal*, o posicionamento de uma teoria nessa dimensão é dado pelo modo como ela incorpora a importância dos padrões temporais de disparo dos neurônios para a codificação. Em um extremo do eixo temporal, estariam as hipóteses de codificação por taxa de disparo (*rate coding*), as quais afirmam que células neurais respondem a estímulos modulando a quantidade de potenciais de ação emitidos em uma longa janela de integração temporal.

Percorrendo esse eixo, deparamo-nos com teorias para as quais os padrões temporais de disparo possuem informação além da taxa de disparos. Esse esquema é chamado de codificação temporal (*temporal coding*). O leitor poderia argumentar que a teoria *hebbiana*, assim como outras hipóteses populacionais, teria necessariamente que adotar a codificação temporal, pois a integração de uma população (como uma assembleia) seria feita pela coordenação de disparos no domínio do tempo. No entanto, de acordo com a definição que emprego nesse texto, a ideia de padrão temporal de ativação é essencialmente ortogonal à de sincronia de assembleias. Por exemplo, Hebb apenas exige que uma célula, ao co-ativar-se com membros de uma assembleia, aumente sua taxa de disparo. Em cada ativação de uma assembleia, seus membros podem disparar de forma não consistente, contanto que aumentem sua taxa de disparo. Hebb ainda propôs o encadeamento temporal de assembleias (*phase sequences*) como o mecanismo para conectar diferentes compartimentos de informação em processos mais complexos. No entanto, Hebb ainda nesse ponto não se preocupa com padrões temporais, apenas com sequências de ativação. A teoria do binding (*binding theory*), ao reescrever a teoria *hebbiana*, adiciona um elemento relevante que move a teoria no eixo temporal: a sincronização de membros de assembleias por oscilações (Singer, 2007). Por sua vez, Moshe Abeles (2009) vai além no eixo temporal com a hipótese que a sucessão de assembleias é realizada em uma sequência estereotipada, as *synfire chains*. Nesse caso, o intervalo entre a ativação de assembleias deve ser consistente e com precisão da ordem de poucos milissegundos.

A Figura 1.1 ilustra esse espaço de acordo como eu entendo diversos modelos de codificação.

Atualmente, não há consenso para uma região limite no espaço formado por esses eixos para posicionar uma teoria contemporânea. Talvez, o debate a ser feito seja se uma teoria pode ser estática nesse espaço. Diversos estudos sugerem que códigos temporais são empregados em áreas sensoriais primárias (e.g., córtex de barris; Panzeri et al., 2001, mas não em áreas multimodais no lobo temporal medial (e.g., células de lugar - *place cells*; O'Keefe

e Dostrovsky, 1971). Paralelamente, o código neural parece tornar-se cada vez mais esparsos conforme subimos a *ordem* da área analisada. Por exemplo, o código é densamente distribuído no início da via ventral (*ventral stream*) do sistema visual, mas torna-se cada vez mais esparsos até chegar ao córtex temporal inferior onde respostas são extremamente seletivas e invariantes (e.g., células de conceito - *concept cells*; Quian Quiroga, 2012). Esse percurso no espaço dimensional é bem ilustrado pelo trabalho no sistema nervoso de insetos de Gilles Laurent (revisado em Laurent, 2002). A teoria que considero mais longe da origem dos eixos é a policronização de Izhikevich (2006). Voltarei a ela no Capítulo 4.



**Figura 1.1. Representação dos modelos teóricos de codificação neural.**

De qualquer forma, ninguém pode afirmar que a comunidade científica está de acordo em relação à natureza do código neural. Provavelmente, o principal motivo dessa falta de clareza está nas dificuldades técnicas para aquisição de dados, dificuldades que começaram a ser superadas apenas nas últimas três décadas. Porém, com o desenvolvimento técnico que possibilitou o registro simultâneo de grandes populações neurais com alta precisão (Stevenson and Kording, 2011), ficou clara a carência de métodos estatísticos que pudessem analisar dados de alta complexidade (Brown et al., 2004). Se dados são obtidos com mais detalhes, é natural que as técnicas de análise anteriormente usadas tornem-se obsoletas. Portanto, tão essencial quanto o desenvolvimento tecnológico de registro é o desenvolvimento de ferramentas de análise que possam extrair significados dos dados.

Nesta tese, apresentamos dois novos métodos computacionais (Lopes-dos-Santos et al., 2013; Lopes-Dos-Santos et al., 2015) aplicáveis à investigação da codificação neural, cada um deles para o estudo de um dos eixos definidos acima.

O primeiro deles, apresentado no Capítulo 2 desta tese, trabalha no eixo populacional. Como será revisado, por muitos anos o estudo de interações entre neurônios se limitou a observações de células individuais ou a análises par-a-par. Apenas mais recentemente, ferramentas começaram a ser apresentadas com a intenção de analisar uma população neural como um todo (ver Capítulo 2 para detalhes). Nesse contexto, contribuímos com um novo método, baseado na análise de componentes independentes, para a detecção e rastreamento de assembleias neurais. De forma não-supervisionada, nosso método é capaz de extrair padrões de co-ativação em trens de disparo imersos em uma grande população neuronal registrada simultaneamente. Demonstramos que essa ferramenta supera dificuldades de outros métodos publicados na detecção de significância estatística de assembleias e no isolamento entre padrões de co-ativação independentes. Um trabalho prévio a este, elaborado durante meu mestrado e publicado no início de meu doutorado (Lopes-dos-Santos et al., 2011), é incluído nesta tese no Anexo I.

No Capítulo 3 descrevo outro método que se aplica ao eixo temporal. A investigação de um potencial código temporal foi impossibilitada por décadas, dadas as limitações experimentais que impedem uma prolongada sessão de apresentação de estímulos a um animal. Considerar padrões temporais na atividade de uma célula, e não só o número total de disparos, leva inevitavelmente ao aumento da dimensionalidade das respostas. Várias técnicas foram propostas para reduzir o viés estatístico dessas análises nos últimos anos. Aqui, usamos decomposição *wavelet* para extrair componentes em trens de disparo que representem de maneira eficiente seus padrões temporais. Demonstramos que esse método supera os métodos mais usados em vários aspectos.

É relevante uma observação sobre a aplicabilidade dessas ferramentas. Métodos computacionais dão a oportunidade para que dados possam ser entendidos e processados contribuindo de maneira imprescindível para o estudo de uma hipótese. Porém, seria ingênuo concluir que elas são suficientes para tal. Métodos de análise devem ser usados complementarmente a paradigmas experimentais adequados. Por exemplo, no próximo capítulo demonstramos que assembleias podem ser detectadas no hipocampo e sua dinâmica pode ser caracterizada por nossa ferramenta. Porém, ainda falta ser demonstrado que as assembleias *per se* impactam o comportamento do animal.

No Capítulo 4 retomo a discussão acerca das teorias propostas para os mecanismos de processamento de informação no cérebro, e aponto direções em que as ferramentas

elaboradas na presente tese podem vir a ser úteis, especialmente para gerar novos modelos de codificação e decodificação neural.

Por fim, em um trabalho preliminar anexo a esta tese (Anexo II), apresento um ensaio sobre a neurofisiologia dos alucinogênicos serotoninérgicos. Acredito que a eletrofisiologia de sistemas é crucial para entendermos os mecanismos responsáveis pelos efeitos psicoativos dessas substâncias. Considero essa linha de pesquisa por si só um programa científico interessantíssimo e abandonado. Além disso, acredito também que esses alucinogênicos podem ser uma ferramenta poderosa para o estudo do cérebro. No Anexo II apresento meus argumentos.

---

# CAPÍTULO 1: DETECTANDO ASSEMBLEIAS EM POPULAÇÕES NEURONAIS



## Computational Neuroscience

## Detecting cell assemblies in large neuronal populations

Vítor Lopes-dos-Santos\*, Sidarta Ribeiro, Adriano B.L. Tort

Brain Institute, Federal University of Rio Grande do Norte, Brazil

## ARTICLE INFO

## Article history:

Received 6 January 2013  
 Received in revised form 11 April 2013  
 Accepted 17 April 2013

## Keywords:

Cell assemblies  
 Principal component analysis  
 Independent component analysis  
 Assembly vectors

## ABSTRACT

Recent progress in the technology for single unit recordings has given the neuroscientific community the opportunity to record the spiking activity of large neuronal populations. At the same pace, statistical and mathematical tools were developed to deal with high-dimensional datasets typical of such recordings. A major line of research investigates the functional role of subsets of neurons with significant co-firing behavior: the Hebbian cell assemblies. Here we review three linear methods for the detection of cell assemblies in large neuronal populations that rely on principal and independent component analysis. Based on their performance in spike train simulations, we propose a modified framework that incorporates multiple features of these previous methods. We apply the new framework to actual single unit recordings and show the existence of cell assemblies in the rat hippocampus, which typically oscillate at theta frequencies and couple to different phases of the underlying field rhythm.

© 2013 Elsevier B.V. All rights reserved.

## Contents

1. Introduction .....	149
2. Review of the methods .....	150
2.1. Construction of the spike matrix .....	150
2.2. Determination of the number of cell assemblies .....	150
2.3. Extraction of cell assembly patterns and estimation of cell assembly activity .....	153
2.3.1. Principal component analysis .....	153
2.3.2. Assembly vector estimation .....	156
2.3.3. Independent component analysis .....	158
2.3.4. Other examples .....	158
3. Real data applications .....	162
4. Discussion .....	164
4.1. Limitations .....	165
5. Conclusion .....	165
Acknowledgements .....	165
Appendix A. Supplementary data .....	165
References .....	165

## 1. Introduction

A main concern in systems neuroscience is to understand how single neurons form functional neuronal circuits ultimately giving rise to complex information processing and behavior. Much of the current knowledge has been derived from the analysis of

the firing rate of single units (Adrian and Zotterman, 1926; Hubel and Wiesel, 1959; Okeefe and Dostrovsk, 1971; Perrett et al., 1982), pairwise correlations (Wilson and McNaughton, 1994), and synchrony between single cells and local field potentials (Siapas et al., 2005). Recent technological advances have opened the possibility of recording large populations of neurons simultaneously (Buzsaki, 2004). These recordings created a new demand for mathematical and statistical tools to analyze the activity of neuronal ensembles as a whole, instead of each unit at a time. New methods with different strategies have been employed, such as template matching (Lee and Wilson, 2002; Louie and Wilson, 2001; Ribeiro et al., 2004), principal component analysis (Chapin and Nicolelis, 1999;

\* Corresponding author at: Brain Institute, Federal University of Rio Grande do Norte, Rua Nascimento de Castro, 2155 – Lagoa Nova, Natal, RN 59056-450, Brazil. Tel.: +55 84 32152709.

E-mail address: [vitor@neuro.ufrn.br](mailto:vitor@neuro.ufrn.br) (V. Lopes-dos-Santos).

Lopes-dos-Santos et al., 2011; Peyrache et al., 2010), independent component analysis (Laubach et al., 1999), shuffling methods for detecting repeated firing sequences (Abeles and Gat, 2001; Abeles and Gerstein, 1988; Berger et al., 2010; Gansel and Singer, 2012), and methods based on information theory (Arabzadeh et al., 2004; Quiroga and Panzeri, 2009), among others.

In this review we focus on linear methods for detecting and tracking the activity of cell assemblies embedded in large neuronal populations. Here we define cell assemblies as subsets of neurons with significant co-activation behavior, as proposed by Hebb (1949). We compare three methods that rely on principal component analysis (PCA). Based on their efficiency in simulated neuronal networks, we propose a modified framework that incorporates multiple features of these methods. We aimed to provide an intuitive view of the methods and their possible applications. Rigorous mathematical derivations can be found in the references cited along the text. MATLAB codes and a tutorial for running the methods can be obtained from the corresponding author upon request.

## 2. Review of the methods

The general procedure can be structured in three main steps: (1) Construction of the spike matrix, where spike trains are binned and normalized; (2) Determination of the number of cell assemblies, where a null hypothesis distribution for cell assembly activity is generated; and (3) Extraction of cell assembly patterns and estimation of cell assembly activity, where co-activation patterns are found and used to track the activity of cell assemblies with single-bin resolution. Step 1 is standard for all methods reviewed here, while steps 2 and 3 may differ. Regarding step 2, here we compare surrogate methods with an analytical threshold recently introduced in Peyrache et al. (2009), and, regarding step 3, we compare previously published methods (Chapin and Nicolelis, 1999; Laubach et al., 1999; Lopes-dos-Santos et al., 2011; Peyrache et al., 2010) and suggest improvements.

### 2.1. Construction of the spike matrix

All methods studied here use matrix representations of spike trains. In this section we show how to construct the spike matrix.

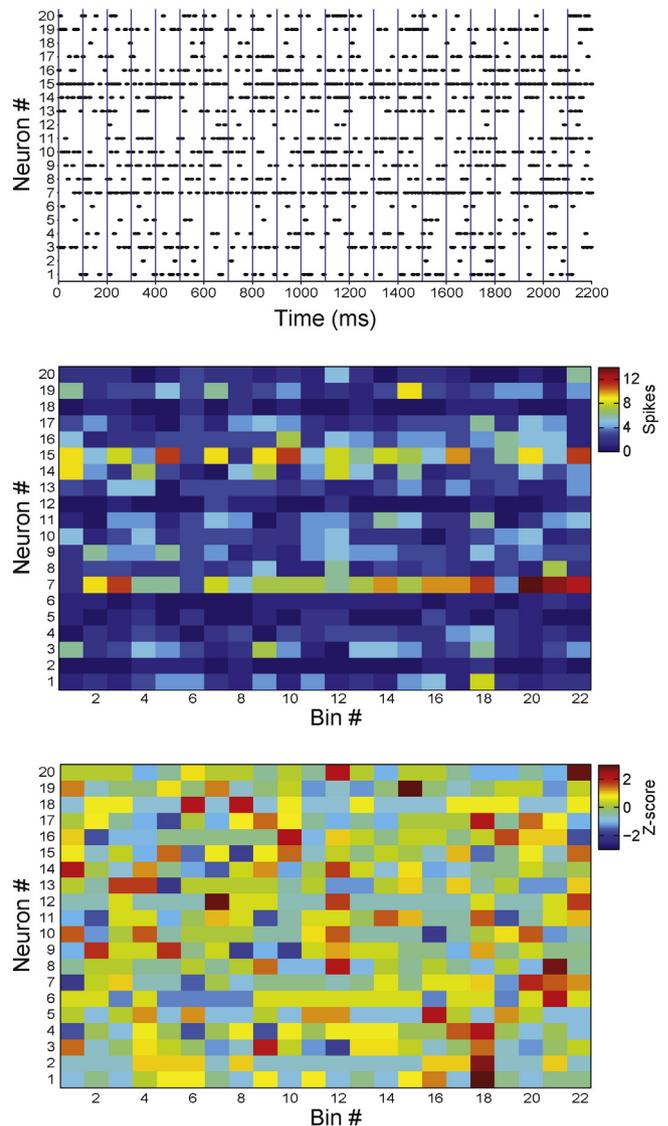
Top panel of Fig. 1 shows the activity of 20 simulated neurons by means of a spike raster plot, in which each black mark denotes an action potential of a given neuron (vertical axis) in a given time (horizontal axis). The spike matrix is constructed by binning the raster plot and counting the number of spikes elicited by each neuron within each bin (Fig. 1 middle); At this point each matrix entry denotes the number of spikes of a given neuron (rows) in a given time bin (columns). In this example we employ a bin size of 100 milliseconds. Next, the spike count of each neuron (i.e., each row of the matrix) is normalized by z-score transformation (Fig. 1 bottom):

$$z_{ib} = \frac{s_{ib} - \langle s_i \rangle}{\sigma_{s_i}}$$

where  $z_{ib}$  is the z-scored spike count of neuron  $i$  in time bin  $b$ ,  $s_{ib}$  is the number of spikes of neuron  $i$  in bin  $b$ ,  $\langle s_i \rangle$  is the mean spike count of neuron  $i$  over all time bins, and  $\sigma_{s_i}$  is the standard deviation of the spike counts of neuron  $i$  over bins. Thus, in the z-scored spike matrix each neuron is set to have null mean and unitary variance.

### 2.2. Determination of the number of cell assemblies

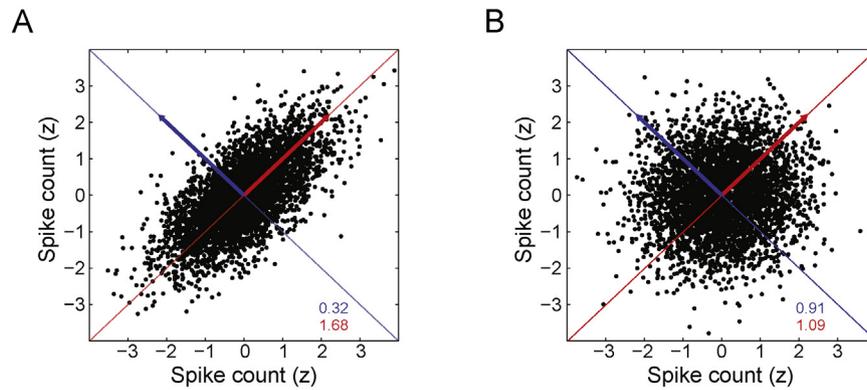
Before extracting assembly patterns it is important to know how many cell assemblies are there to be found. Peyrache et al. (2010, 2009) introduced the use of eigenvalue analysis for determining the statistical significance of assembly patterns. We provide some toy examples below to illustrate this procedure.



**Fig. 1.** Pre-processing of spiking activity data. Top panel: spike raster plot of 20 simulated neurons. Dots represent spike times and blue lines denote the boundaries of 100-ms time bins. Middle panel: Non-normalized spike matrix. Each element of the matrix is the number of spikes of a given neuron (row) within a given time bin (columns). Bottom panel: Normalized spike matrix. The spiking activity of each neuron is z-scored. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 2 shows two scatterplots in which each point represents the normalized spike count of a pair of neurons at the same time bin. The spiking activity of the neuron pair displayed in Fig. 2A is correlated, while the activity of the pair in Fig. 2B is not correlated. Notice that the variance of the data concentrates in a given direction when variables are correlated (Fig. 2A), while the variance is homogeneously distributed across all directions when variables are not correlated (Fig. 2B).

The directions of largest variances of multidimensional data can be found by PCA. By definition, the first principal component (PC) gives the direction of largest variance (that is, when data points are projected onto the direction defined by the first PC, the variance is larger than when projected onto any other direction); the second PC points the direction of largest variance orthogonal to the first PC, and so on. PCs are the eigenvectors of the covariance matrix of the spike matrix. Since spike counts are normalized to have null mean



**Fig. 2.** Principal components and spike count correlations. (A,B) Panels show scatterplots of the spiking activity of two neurons. Each point denotes the z-scored spiking activity of a pair of simulated neurons (100-ms time bins were used; each neuron is represented by an axis). Red and blue arrows denote first and second principal components, respectively. Colored numbers inform the variance of the data in the axes spanned by the principal components (colored lines; same color convention). Panel A shows a correlated pair of neurons and panel B shows an uncorrelated pair. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

and unitary variance, in our case the covariance matrix is equal to the correlation matrix, and can be calculated as:

$$C = \frac{ZZ^T}{N_{\text{columns}}}$$

where  $Z$  is the (z-scored) spike matrix,  $T$  the transpose operator, and  $N_{\text{columns}}$  is the number of time bins of  $Z$ . Thus, the element at the  $i$ -th column and  $j$ -th row of  $C$  is the linear correlation between neurons  $i$  and  $j$ . Since  $C$  is necessarily real and symmetric, it follows from the spectral theorem that it can be decomposed as:

$$C = \sum \lambda_i x_i x_i^T$$

where  $x_i$  is the  $i$ -th eigenvector of  $C$  (that is, the  $i$ -th PC of  $Z$ ) and  $\lambda_i$  its corresponding eigenvalue. Moreover, the outer product  $x_i x_i^T$  is the projection matrix onto the direction of  $x_i$ , and  $\lambda_i$  is the variance of the data along the same axis.

In Fig. 2 the red and blue lines represent the first and second PCs, respectively, and their eigenvalues (i.e., variances) are shown with same color labels. As mentioned above, when spike trains are uncorrelated, as in Fig. 2B, the variance of the data projected onto any direction is roughly the same. In fact, they would be exactly equal if infinite samples were analyzed. Random fluctuations due to undersampling produce a direction with slightly larger variance that is detected as the first PC. In the case of correlated activity (Fig. 2A), the variance is significantly more concentrated in the direction of the first PC. These observations can be generalized for higher dimensional signals: data variance will be significantly larger in a given direction if there is a linearly correlated group of cells. Following this reasoning, a possible strategy to estimate the number of cell assemblies (subsets of neurons with correlated activity) is to find the number of PCs of the spike matrix with significantly large eigenvalues.

Peyrache et al. (2009, 2010) proposed the use of the Marčenko–Pastur distribution as a null hypothesis for the existence of cell assemblies. Marčenko and Pastur (1967) demonstrated that the eigenvalues of the correlation matrix of a normal random matrix  $M$  with statistically independent rows follow a probability function described by:

$$p(\lambda) = \frac{q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{\max} - \lambda)(\lambda - \lambda_{\min})}}{\lambda},$$

with  $q = N_{\text{columns}}/N_{\text{rows}} \geq 1$ , where  $\sigma^2$  is the variance of the elements of  $M$  (in our case  $\sigma^2 = 1$  due to z-score normalization),  $N_{\text{columns}}$  is the number of columns and  $N_{\text{rows}}$  the number of rows.  $\lambda_{\max}$  and

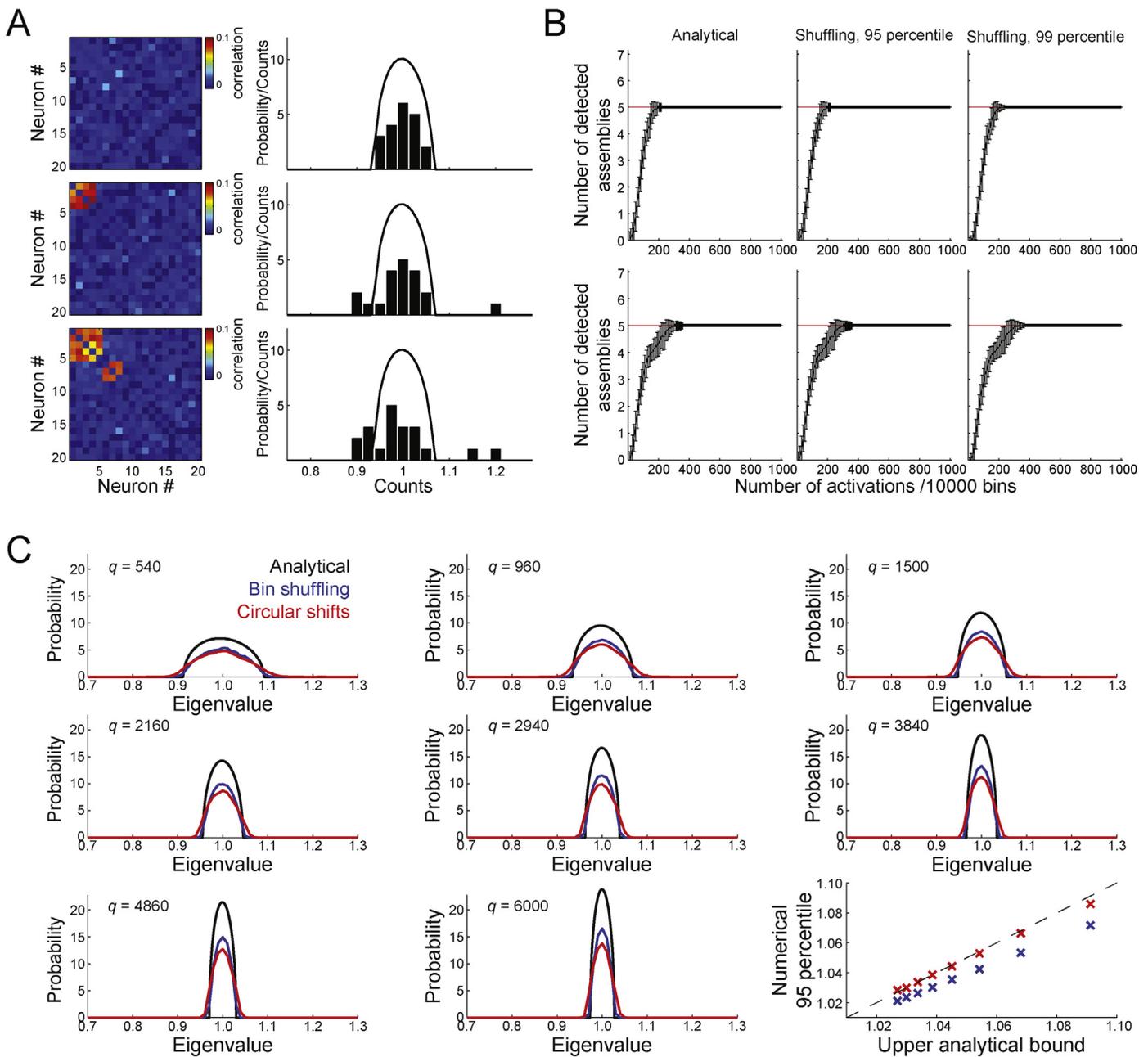
$\lambda_{\min}$  are the maximum and minimum bounds, respectively, and are calculated as:

$$\lambda_{\min}^{\max} = \sigma^2(1 \pm \sqrt{1/q})^2$$

This probability function has finite support given by the interval  $\lambda_{\min} \leq \lambda \leq \lambda_{\max}$ . Thus, if the rows of  $M$  are statistically independent, the probability of finding an eigenvalue outside these bounds is zero. In other words, the variance of the data in any axis cannot be larger than  $\lambda_{\max}$  when neurons are uncorrelated. Therefore,  $\lambda_{\max}$  can be used as a statistical threshold for detecting cell assembly activity (Peyrache et al., 2010). That is, the number of eigenvalues above  $\lambda_{\max}$  can be used to estimate the number of cell assemblies in the spike matrix.

Although this theoretical bound was derived for large matrices, it is also a robust threshold for smaller matrices (Lopes-dos-Santos et al., 2011; Plerou et al., 2002). Nevertheless, Peyrache et al. (2010) proposed the use of a finite sample bias correction based on the Tracy–Widom distribution (Tracy and Widom, 1994). In addition, it should be noted that although the Marčenko–Pastur distribution was proved for random matrices whose entries are derived from Gaussian distributions, empirical simulations show that this distribution also provides a good bound for eigenvalues of matrices composed by independent rows (in our case, uncorrelated neurons) originated from other random processes (Biroli et al., 2007; Lopes-dos-Santos et al., 2011; Seba, 2003). Moreover, we never observed “false positive” eigenvalues in simulations employing Poisson neurons, i.e., the number of eigenvalues significantly larger than chance was always equal to or less than the real number of assemblies.

Fig. 3A displays examples of correlation matrices computed from three simulated spike matrices (the main diagonal was set to zero for clearer visualization); the empirical eigenvalue distributions are plotted along with the Marčenko–Pastur distribution (black line). Neurons were modeled as Poisson processes with mean rate of 1 spike/bin. In the example shown on top, all neurons fired independently of each other. Note that all eigenvalues lie within the bounds of the analytical distribution. In the second example, a subset of four neurons is correlated (Neurons # 1–4), as can be seen from the correlation matrix. In this case, one eigenvalue lies above the analytical distribution. The last example has two subsets of correlated neurons (Neurons # 1–5 and 6–8), and two eigenvalues are larger than the theoretical threshold. Thus, the upper bound of the Marčenko–Pastur distribution represents a statistical threshold for assembly activity, i.e., the number of cell assemblies in the network is estimated by the number of eigenvalues above this threshold (Lopes-dos-Santos et al., 2011).



**Fig. 3.** Estimating the number of cell assemblies. (A) An analytical threshold for assembly detection. Left panels show correlation matrices of three simulated neuronal populations (20 neurons were modeled as Poisson processes with mean rate of 1 spike/bin for 8000 time bins). Main diagonals set to zero for clearer visualization. Right panels show the theoretical Marčenko–Pastur distribution (black line) along with the empirical distribution of eigenvalues of the correlation matrices. The upper bound of the analytical distribution can be used as statistical threshold for assembly detection. In the first simulation all neurons fired independently (top left panel), and all eigenvalues lie within the analytical distribution (top right panel). In the second simulation four neurons (neurons #1, #2, #3 and #4) were correlated (middle left panel). Note that one eigenvalue lies above the upper bound of the analytical distribution (middle right panel). In the third simulation two subgroups presented significant correlation (bottom left panel), and two eigenvalues lie above the upper bound of the analytical distribution (bottom right panel). (B) Top panels show the mean number of assemblies detected in simulated spike matrices as a function of assembly activation rate (150 simulations were run for each activation rate; error bars denote standard deviation). Each spike matrix comprised 32 neurons modeled as Poisson processes; five cell assemblies composed by five non-overlapping neurons were programmed. Left panel displays results when the upper bound of the analytical distribution is used as threshold for detecting assembly activity. Middle and right panels display results when the threshold was determined by a surrogate method, in which the bins of each neuron are shuffled independently and the maximum eigenvalue of the correlation matrix is calculated. This procedure is repeated 3200 times in order to generate a null hypothesis distribution for maximal eigenvalues. A statistical threshold is defined as a percentile of this distribution (95% for middle panel and 99% for right panel). Bottom panels show the same as above but for overlapping assembly membership. See text for further details. (C) Eigenvalue distribution for independent activity. The activity of an actual hippocampal neuron was binned into 10-ms windows and spike matrices were constructed by two methods: in the first method, neurons were created by randomly shuffling bins; in the second method, neurons were created by random circular time shifts of the original spike train, thus preserving autocorrelation and ISI counts. No cell assembly was programmed (but see Fig. 10). For each method and values of  $q$  (number of time bins/number of neurons, see text), eigenvalue distribution was estimated by analyzing 1000 random spike matrices composed by 20 neurons. The analytical Marčenko–Pastur distribution is also displayed. Rightmost panel in the bottom row shows correspondence between the 95th percentile of the numerical distributions ( $y$ -axis, color coded) and the upper bound of the analytical distribution ( $x$ -axis). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

An alternative to the use of the Marčenko–Pastur distribution is to define a statistical threshold based on surrogate data. This procedure entails the shuffling of time bins for each neuron independently in order to destroy their temporal relations while maintaining the distribution of spike counts unchanged. The eigenvalues of correlation matrices obtained from shuffled spike matrices can be used to construct a null distribution; for example, the eigenvalues of the original spike matrix that are larger than a certain percentile (e.g., 95% or 99%) of the distribution of the maximal eigenvalues computed from shuffled data are regarded as significant (that is, in this framework each surrogate matrix contributes its maximal eigenvalue to the chance distribution, which is a more conservative approach than considering all surrogate eigenvalues).

In order to compare the analytical and shuffling methods, we simulated spike matrices with 32 neurons and 10,000 bins. Neurons were modeled as Poisson processes with unitary mean. Additionally,  $\eta$  bins of each neuron were randomly chosen as *activation bins*, in which the spike count was derived from a uniform random distribution from 0 to 6 spikes, thus likely above the overall mean. In order to simulate assemblies, we set five subsets of neurons (i.e., five cell assemblies) to have coincident activation bins (*assembly activations*). Note that although assembly members increase firing together, each assembly activation is unique since the firing of assembly members is stochastic. The top panels of Fig. 3B display results for spike matrices with the following assembly composition: assembly 1 neurons: #1, #2, #3, #4; assembly 2: #5, #6, #7, #8; assembly 3: #9, #10, #11, #12; assembly 4: #13, #14, #15, #16; assembly 5: #17, #18, #19, #20. Lines represent the mean over 150 simulations and bars indicate standard deviation. This example shows that the analytical and the shuffling methods converge to the correct number of assemblies after a certain number of activations have occurred. The bottom panels of Fig. 3B show similar results but for a set of assemblies in which each assembly shares neurons with at least two other assemblies (assembly 1 neurons: #1, #2, #3, #4, #5; assembly 2: #5, #6, #7, #8, #9; assembly 3: #9, #10, #11, #12, #13; assembly 4: #12, #13, #14, #15; assembly 5: #15, #16, #17, #1). The analytical and the shuffling methods also converge to the correct number of assemblies in this case, although more assembly activations are necessary than in the case of non-overlapping assemblies.

Next, we compare analytical and numerical distributions of random eigenvalues for data with a more realistic structure. Instead of using Poisson neurons, we created 20 neurons based on the spike trains of a neuron recorded from the rat hippocampus. Spikes were binned into 10-ms windows (120,000 bins, or 20 min, in total) and new spike trains were created by two surrogate methods. The first of them is the same bin shuffling method used for Fig. 3B. The second is achieved by means of a circular random shift. That is, for each neuron a time point is randomly chosen and used to divide the spike train in two non-overlapping segments which switch position. This method creates surrogate neurons with same rate and variance of the original neuron, and virtually the same autocorrelation and inter-spike-interval (ISI) distribution.

We estimated the distribution of eigenvalues for independent activity by computing the average eigenvalue histogram across 1000 simulations for each surrogate method. Since the distribution is affected by the ratio  $q = N_{\text{columns}}/N_{\text{rows}}$ , we repeated the procedure for matrices with different number of bins ( $N_{\text{columns}}$ ) while keeping the number of neurons fixed. Fig. 3C displays the eigenvalue distribution estimated by each surrogate method along with the Marčenko–Pastur distribution for each  $q$ . Finally, we compare the 95 percentile drawn from the numerical distributions with the upper bound of the analytical distribution in the bottom right panel of Fig. 3C. Surprisingly, the 95 percentile of the circular shift method matches very closely the upper bound of the analytical distribution.

The 95 percentile of the bin shuffling procedure was lower than the upper analytical bound for all values of  $q$  studied. Altogether, the results indicate that the theoretical bound derived from the Marčenko–Pastur distribution is well suited for tracking the number of cell assemblies, since surrogate methods give similar results but are computationally demanding. More detailed parametrical studies on the Marčenko–Pastur distribution can be found in Lopes-dos-Santos et al. (2011). It is nevertheless advisable to compare the analytical and shuffling methods when working with actual spike recordings, since the specific statistical properties of a given dataset may be different from the simulated examples employed here. In this regard, a valuable feature of the surrogate method is that surrogate spike matrices can be generated in such a way that neurons preserve their ISI distribution and autocorrelation, thus preserving more statistical properties of the original data.

### 2.3. Extraction of cell assembly patterns and estimation of cell assembly activity

In linear models, the activity of a cell assembly is assumed to be a linear combination (a weighted sum) of the spike activity of all neurons. Thus, each co-activation pattern (also called assembly pattern) is a vector that attributes weights to each neuron. For each pattern, cell assembly neurons can be identified as those with the largest weights.

In mathematical terms, the activity of a cell assembly in a given time bin  $b$  can be defined as:

$$R_b = \sum_{i=N_{\text{neurons}}} w_i z_{ib} = w^T Z_b,$$

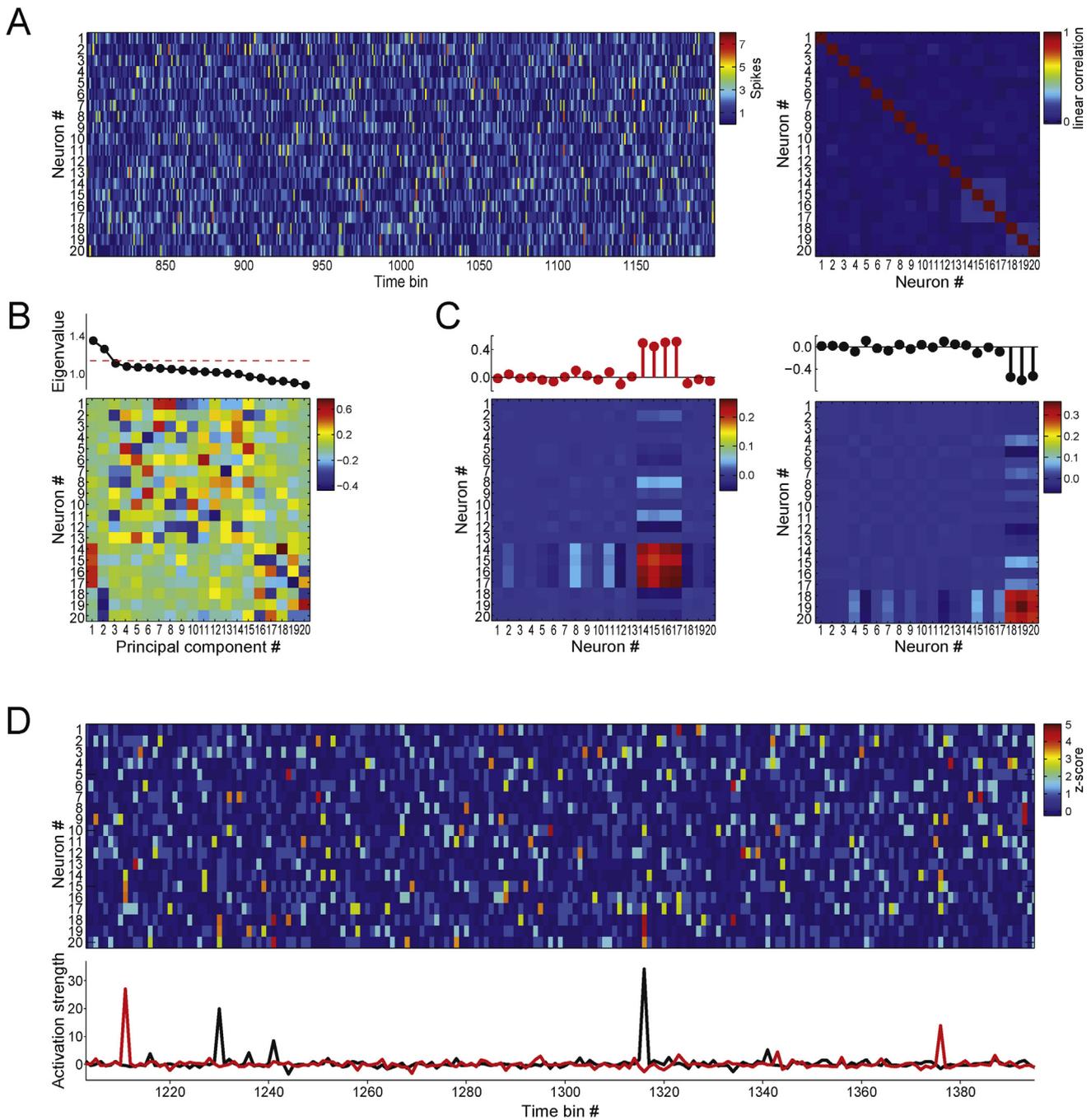
where  $N_{\text{neurons}}$  is the number of neurons,  $z_{ib}$  is the z-scored activity of neuron  $i$  at the time bin  $b$ ,  $w_i$  is the weight of the corresponding neuron in the assembly,  $w$  is a column vector containing the weights of all neurons in the assembly and  $Z_b$  is the  $b$ -th column of the spike matrix  $Z$ .

This section deals with the problem of finding co-activation patterns  $w$  for each cell assembly. We study three different methods for obtaining  $w$ , which are based on PCA (Chapin and Nicolelis, 1999; Peyrache et al., 2010), assembly vectors (AV) (Lopes-dos-Santos et al., 2011), and independent component analysis (ICA) (Laubach et al., 1999).

#### 2.3.1. Principal component analysis

Nicolelis et al. (1995) employed PCA to study the activity of thalamo-cortical assemblies during the occurrence of *mu* rhythm in rats, using the PCs themselves as co-activation patterns to track ensemble activity (Chapin and Nicolelis, 1999). Later, Peyrache et al. (2010) introduced a template-match protocol by which assembly patterns identified during learning epochs are used to assess assembly activity during pre- and post-learning sleep episodes. Employing this framework, they showed that subsets of medial prefrontal. Cortex neurons that were co-active during rewarded runs in a T-mask task also co-activated during subsequent slow-wave sleep, mainly during hippocampal ripple events (Peyrache et al., 2009). More recently, Benchenane et al. (2010) used the PCA approach to show that the co-firing of assembly neurons occurs at a preferred phase of the theta rhythm.

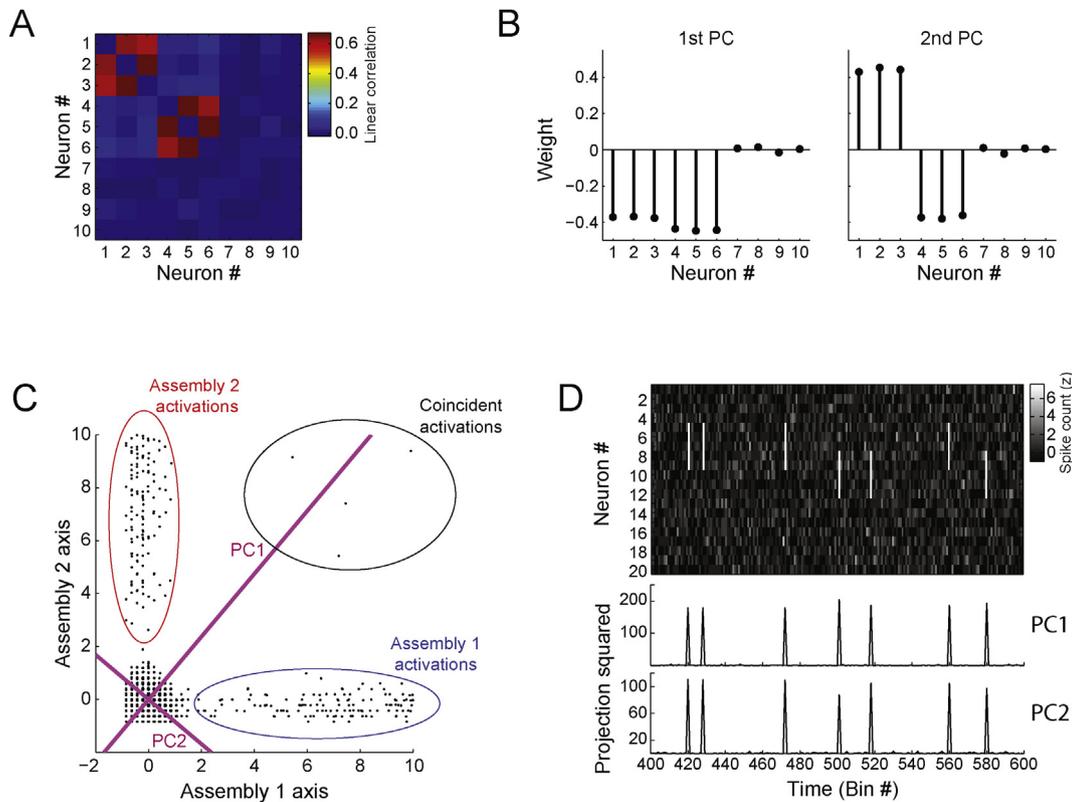
Roughly speaking, PCA extracts uncorrelated orthogonal components from the spike matrix, which are assumed to represent co-activation patterns. The framework is illustrated in Fig. 4. A time interval of a spike matrix and its correlation matrix are shown in Fig. 4A, with 20 neurons simulated as Poisson processes for 8000 bins with 1 spike/bin mean rate. Additionally, non-stationarities were introduced in the activity of each neuron by randomly choosing 250 bins as activation bins. In these bins, the firing rate



**Fig. 4.** Detecting cell assemblies and tracking their activity using principal component analysis. (A) Left panel: Non-normalized spike matrix composed by 20 simulated neurons (total simulation time: 8000 bins). Each neuron was modeled as a Poisson process with mean rate of 1 spike/bin. Additionally, 250 *activation bins* (around 3% of the total number of bins) were randomly chosen for each neuron: in these bins, the spike count was derived from a uniform distribution between 0 and 6 spikes. In order to simulate assemblies, subsets of neurons were set to have coincident activation bins. In this example, two assemblies are present: assembly 1 (neurons #14, #15, #16, #17) and assembly 2 (neurons #18, #19 and #20). Right panel: Correlation matrix. (B) Principal components and eigenvalues. Top panel shows the eigenvalues of the correlation matrix shown in A. Red dashed line denotes the analytical significance threshold. Two eigenvalues are significant in this case. Bottom panel shows principal components associated with the eigenvalues shown above. (C) Top panels display the principal components associated with the eigenvalues significantly above chance. Lower panels show the corresponding outer products of each principal component, which are projectors of the spike matrix used for computing cell assembly activity with single-bin resolution. (D) Top panel shows a representative time interval of the same spike matrix as in A after z-score normalization. Lower panel shows the activity of the detected assemblies estimated by the principal components shown in C. Note that the peaks of assembly activity correspond to co-activations of the corresponding assembly members. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

was uniformly distributed between 0 and 6 spikes. Two assemblies were programmed: assembly 1, composed by neurons #14, #15, #16, #17; and assembly 2, composed by neurons #18, #19 and #20. Fig. 4B displays the PCs along with their associated eigenvalues. The red dashed line represents the significance threshold provided by the Marčenko–Pastur distribution. In this example,

two eigenvalues are significant. Fig. 4C shows the significant PCs (top panels) along with their outer product (bottom panels). Note that these patterns isolate the activity of the cell assemblies, i.e., they attribute large weights to specific assembly members; in fact, in this example the first and second PCs represent the co-activation patterns of assemblies 1 and 2, respectively.



**Fig. 5.** Examples of limitations of assembly pattern estimation by principal components. (A) Correlation matrix of a simulated spike matrix with 8000 bins and 10 neurons. Two assemblies were programmed with the following composition: assembly 1, neurons #1, #2, #3; assembly 2, neurons #4, #5, #6. Neurons and assemblies were simulated as in Fig. 4. Each assembly was active for 200 bins, where the spike count of assembly members assumed a uniform distribution from 0 to 30. (B) Significant principal components of the correlation matrix shown in A. (C) Projection of the columns of the spike matrix onto the subspace spanned by the *ideal* assembly patterns. Ideal assembly patterns have same weights for assembly members and null weights for other neurons. Points representing activations of assembly 1 and 2 are circumscribed by blue and red ellipses, respectively. Bins in which both assemblies were active are circumscribed by a black ellipse. Axes spanned by the principal components are represented by magenta lines. (D) Top panel: A time interval of a simulated spike matrix with 20 neurons showing activations of two assemblies with overlapping membership (assembly 1, neurons #6, #7, #8, #9; assembly 2, neurons #8, #9, #10, #11). Bottom panels: assembly activities estimated when principal components are used to define assembly patterns. Adapted from Lopes-dos-Santos et al. (2011). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

The next step is to use the assembly patterns to compute the time course of each cell assembly activity with single-bin resolution. The activity of each assembly can be estimated by projecting the columns of the spike matrix onto the axis spanned by the corresponding assembly pattern (in this case, assumed to be the PCs). The projection of a column  $Z_b$  onto an axis can be calculated as:

$$proj = PZ_b,$$

where the projection matrix  $P$  is defined as:

$$P = w \otimes w = ww^T,$$

where  $\otimes$  is the outer product operator and  $w$  is a unitary vector that spans the axis. The vector  $w$  is the assembly pattern. At each time bin, the length of the projection is a measure of the similarity between the activity of the whole population and the assembly pattern. This length is calculated by taking the inner product between the assembly pattern  $w$  and  $Z_b$  (i.e., a weighted sum of the normalized spike counts). Alternatively, the *strength* of the projection was defined by Peyrache et al. (2010) as the square of the projection length, which can be calculated as:

$$R_b = Z_b^T P Z_b,$$

where  $P$  is the outer product of the assembly pattern. Hereafter we will adopt this definition of assembly activity. The projection matrices for the assembly patterns in the last example are shown in Fig. 4C bottom panels. Of note, since we are concerned with

co-activation patterns, the main diagonal of  $P$  is set to zero, which assures that isolated activations of assembly members do not contribute to  $R$  (Lopes-dos-Santos et al., 2011; Peyrache et al., 2010). Fig. 4D shows a time interval of the spike matrix  $Z$  along with the estimated time course of the activity of the two cell assemblies. Note that the peaks of the *black* time-course are coincident with the co-activation of neurons composing assembly 2, while the *red* time-course peaks with the activations of assembly 1.

Although the original PCA method can be efficient for extracting assembly patterns (Benchenane et al., 2010; Chapin and Nicolelis, 1999; Nicolelis et al., 1995; Peyrache et al., 2009, 2010), this framework has important limitations. As mentioned above, linear correlations among a subset of neurons lead to larger variance in a given axis that can be found by PCA. However, if two assemblies concentrate a similar amount of variance in their corresponding axes, the first PC will represent the average of the two assemblies and not an individual assembly. This issue is illustrated in Fig. 5. A spike matrix was simulated as before but with 10 neurons; each assembly was active in 200 bins, where the spike count of the assembly neurons assumed a uniform distribution from 0 to 30 spikes. Assembly 1 was composed by neurons #1, #2 and #3, and assembly 2 by neurons #4, #5 and #6. Since the two assemblies have the same number of neurons and activation bins, they have similar mean strength; consequently, the axes that describe these assemblies have similar variance. The correlation matrix for this simulation and its significant eigenvectors are shown in Fig. 5A,B. Notice that the first PC attributes large weights to neurons of assemblies 1 and 2, whereas

ideal assembly patterns should have large weights (with the same sign) only for the members of the corresponding assembly. Fig. 5C shows the projection of the data onto the space spanned by the ideal assembly patterns (a large value in a given axis means that the corresponding assembly was active at that time bin). The axes of the significant PCs are displayed as magenta lines. Note that length of the projections onto the axis defined by the first PC are similar for both assemblies; therefore, in this example the activity time course estimated by the outer product of the first PC peaks when either assembly 1 or 2 is active (not shown).

This problem gets more critical when assemblies share neurons (Lopes-dos-Santos et al., 2011). A shared neuron should ideally have large weight in both assembly patterns; however, this cannot be represented by PCs since PCs are orthogonal by definition. In order to illustrate this issue, a spike matrix was simulated with 20 neurons modeled as Poisson processes with unitary mean for 8000 time bins. For each assembly, 40 bins were randomly chosen as assembly activations, in which assembly members elicited 6 spikes. Two assemblies were included: assembly 1 composed by neurons #6, #7, #8 and #9, and assembly 2 composed by neurons #8, #9, #10, #11 and #12. Fig. 5D shows a time interval of the spike matrix along with assembly activities computed from the outer product of the significant PCs. Notice that both time courses peak whenever assembly 1 or 2 is active.

In an attempt to overcome these limitations, Lopes-dos-Santos et al. (2011) have proposed a method that searches for clusters of neurons in the subspace spanned by the PCs. We review this framework in the following section.

### 2.3.2. Assembly vector estimation

The PC weights (loadings) carry information about cell assembly membership, in the sense that large weighted neurons are part of an assembly while neurons with near zero weights are not. However, how large should the weight of a neuron be in order to make it part of an assembly? The assembly vector (AV) estimation method was developed to identify the membership of cell assemblies in addition to extracting assembly patterns (Lopes-dos-Santos et al., 2011). The framework is summarized in Fig. 6. A spike matrix composed of 32 neurons was simulated as in the example of Fig. 4. Two assemblies were programmed: assembly 1 with neurons #27, #28, #29, #30; and assembly 2 with neurons #29, #30, #31, #32. Fig. 6A shows the correlation matrix (left panel) along with its eigenvalues distribution (right panel). As in the previous method, the number of cell assemblies is determined by means of the Marčenko–Pastur distribution.

Note that even when the PCs mix assembly patterns, the threshold derived from the Marčenko–Pastur distribution still holds (Lopes-dos-Santos et al., 2011). This happens because activations of different assemblies cannot lie in the same axis (otherwise they would be a single assembly), and therefore variance increases in the same number of axes as the number of assemblies. Moreover, note that when two neurons are correlated, the variance concentrates in a given axis (the assembly axis) and decreases in the orthogonal axis (Fig. 2A), which leads to one eigenvalue significantly above chance distribution and another significantly below. This can be generalized for groups of neurons: for example, if four neurons participate in one assembly, the variance will increase in one axis and decrease in other three. Following this reasoning, Lopes-dos-Santos et al. (2011) have shown numerically that the number of eigenvalues outside the theoretical distribution can be used to estimate the number of assembly neurons. Thus, in the AV framework the number of neurons composing at least one assembly is estimated by counting the total number of eigenvalues outside the Marčenko–Pastur distribution, i.e., above or below the theoretical bounds (Lopes-dos-Santos et al., 2011). These bounds are indicated in the right panel of Fig. 6A by dashed red lines. In this example,

six eigenvalues lie outside the distribution (2 above and 4 below), which matches the number of neurons participating in cell assemblies.

In order to identify cell assembly neurons, the columns of the correlation matrix are projected onto the subspace spanned by the PCs associated with significantly large eigenvalues (i.e., above the upper limit of the Marčenko–Pastur distribution):

$$N_i = P_{AS} C_i,$$

where  $C_i$  is the  $i$ -th column of the correlation matrix and  $P_{AS}$  is the projection matrix of the assembly space, defined as:

$$P_{AS} = \sum_i P_{ci} P_{ci}^T = P_c P_c^T,$$

where  $P_{ci}$  is the  $i$ -th significant PC and  $P_c$  is a matrix containing all significant PCs (columns).

In this framework, this subspace is called *Assembly Space* (Fig. 6B), and the columns of the correlation matrix projected onto this subspace ( $N_i$ ) are called *neuron vectors*.

Since six eigenvalues lie outside the theoretical distribution, the six neurons with largest neuron vectors are regarded as *significant neurons*. Note that neurons that have similar correlation patterns will have neuron vectors clustered together in the Assembly Space. On the other hand, if neurons have orthogonal correlation patterns their neuron vectors will tend to be separated in this space. Following this reasoning, the inner products between all neuron vectors and those of the significant neurons is a measure of similarity between their co-activation patterns. The matrix containing these inner products is called *Interaction Matrix* (Fig. 6C), which can be formally defined as:

$$M_{i,j} = N_i^T N_j,$$

where  $M_{i,j}$  is its element at row  $i$  and column  $j$ ,  $N_i$  is the neuron vector of neuron  $i$  and  $N_j$  is the neuron vector of significant neuron  $j$ .

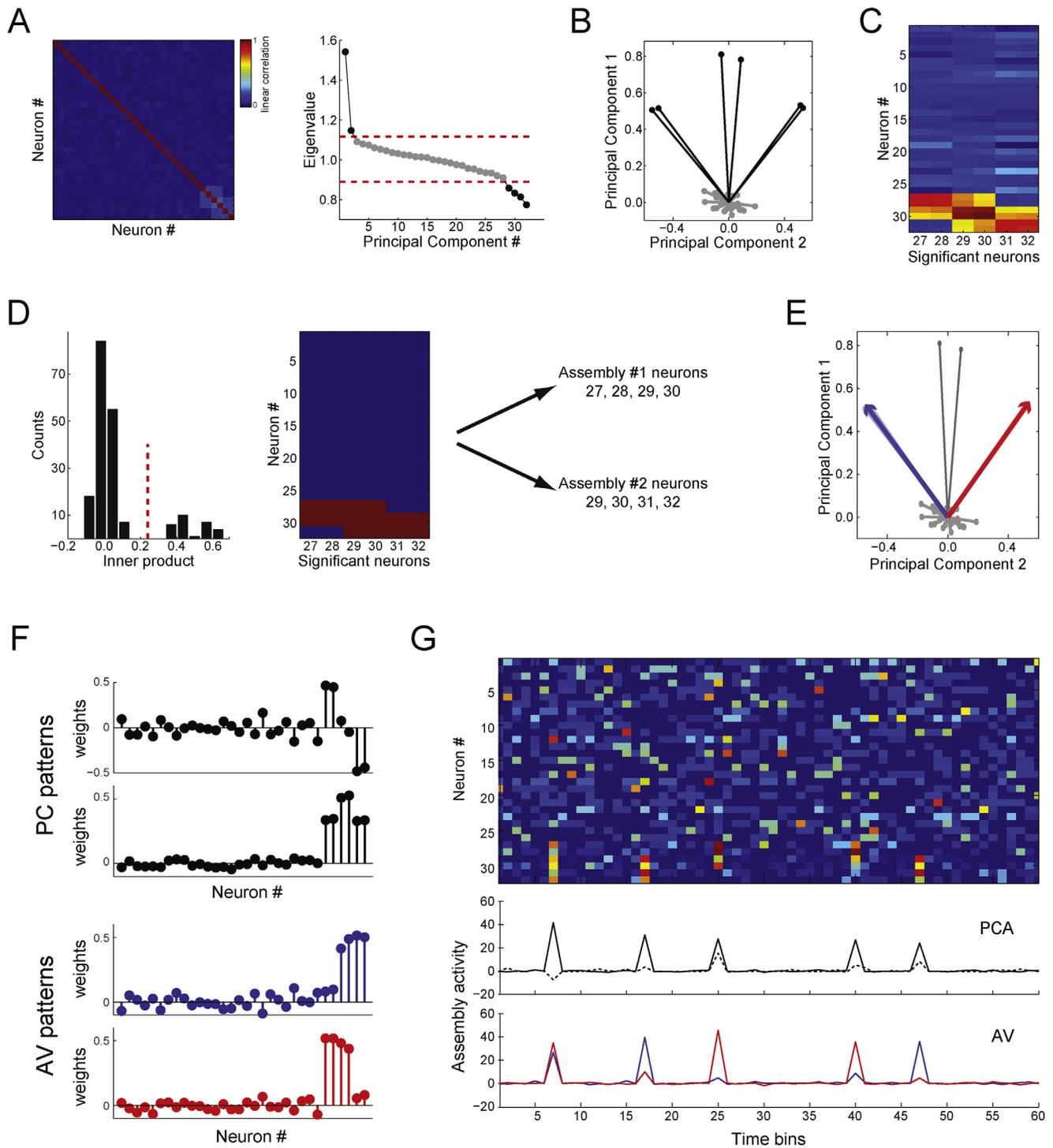
In principle the same could be done by the direct use of the columns of the correlation matrix. However, it is reasonable to assume that assembly patterns can be described by a linear combination of the significant PCs (as in Fig 5B,C). Thus, PCs are used to *filter* the data and the inner products are taken inside the Assembly Space.

Ideally, the distribution of inner products is bimodal, and there is a threshold that separates large and low values (Fig 6D, left) and can be used to digitalize the interaction matrix ('1' is assigned to values above the threshold and '0' to others). Automated algorithms, such as standard k-means, can find such a threshold. A clustering algorithm is then applied to the digitalized interaction matrix to identify assembly membership (Fig. 6D, right), under the constraint that there are two assemblies (which correspond to the number of eigenvalues above the Marčenko–Pastur distribution in this example). Finally, the AVs are defined as the mean of all neuron vectors exclusive to an assembly (Fig. 6E):

$$AV_a = \frac{1}{n_a} \sum_i N_i,$$

where  $N_i$  is the neuron vector of an exclusive neuron of the corresponding assembly and  $n_a$  is the number of exclusive neurons.

For example, the vector of assembly 1 is the mean between the vectors of neurons #27 and #28. Fig. 6F compares assembly patterns defined by PCA and by the AV method. While the PCA method fails to sort apart the assembly patterns, the AV method properly identifies the two patterns. Thus, the projectors of a spike matrix constructed from the outer product of AVs will correctly track the



**Fig. 6.** Detecting and tracking cell assembly activity using *assembly vectors*. (A) Left: Correlation matrix. For this example, 32 neurons were simulated (mean 1 spike/bin; total simulation time: 10,000 bins). Each neuron had 500 activation bins (0 to 6 spikes uniformly distributed). Right: Distribution of eigenvalues of the correlation matrix. The two significance bounds (upper and lower) of the analytical distribution are shown (red dashed lines). The number of eigenvalues above the upper bound denotes the number of cell assemblies, as in the previous method. Additionally, the number of eigenvalues outside the chance distribution (above and below) is used as a lower bound for the number of neurons participating in cell assemblies. (B) Projection of all correlation matrix columns in the subspace spanned by the significant principal components (*Assembly Space*). In this framework, these projections are referred to as neuron vectors. Neurons with largest projections in the Assembly Space are regarded as *significant neurons*. (C) Interaction matrix: The inner product of all neuron vectors with those of the significant neurons is computed. (D) Membership estimation. Left: histogram of the entries of the interaction matrix. The value that best separates the bimodal distribution is obtained by a one-dimensional k-means algorithm and used to threshold the interaction matrix (values above the threshold are set to '1', values below to '0'). Right: Interaction matrix shown in C after thresholding (*binary interaction matrix*). Assembly members are estimated by means of an iterative clustering algorithm applied to the binary interaction matrix (for details, see Lopes-dos-Santos et al., 2011). (E) Estimation of assembly vectors. Each assembly vector (blue and red arrows) is defined as the mean over all vectors in Assembly Space associated with neurons exclusively pertaining to an assembly. Similar to the previous method, the assembly vectors then form assembly patterns from which activity projectors (i.e., outer products) are constructed. (F) Assembly patterns estimated by principal components (PC) and assembly vectors (AV). Note in this example that the PC method mixes assembly members, while the AV method estimates assembly members more reliably. (G) Estimation of assembly activity. A time interval of the spike matrix is displayed (top) along with assembly activities estimated by PC (middle) and AV (bottom) methods. Note that the activities estimated by AVs better isolate activations of different assemblies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

activity of individual cell assemblies, which cannot be achieved by the original PCA method in this case (Fig. 6G).

Despite being an improvement over the original PCA framework, the AV method has limitations. First, the AV method by definition requires each assembly to have at least one exclusive member. Moreover, the separation between small and large inner products in the Assembly Space (Fig. 6C) is not always clear. This is critical because k-means splits the data into two groups irrespective of whether they are clearly separated or not, which may lead to spurious separation. Below we describe another method for extracting assembly patterns that also relies on first analyzing the spike matrix using PCA, and then proceeds with ICA to identify assembly patterns even when cell assemblies have no exclusive neuron.

### 2.3.3. Independent component analysis

Laubach et al. (1999, 2000) introduced ICA to quantify interactions among neurons. In this section we review basic ICA theory and show how this tool can be used to estimate assembly patterns. Roughly speaking, ICA is a method to extract statistically independent non-Gaussian components of a multivariate signal. Rigorous mathematical derivations can be found elsewhere (Comon, 1994; Hyvarinen and Oja, 1997; Jutten and Herault, 1991). Assuming that the time course of the activity of an assembly is described as:

$$R = w^T Z,$$

note that we only know  $Z$  (the spike matrix) and must estimate both  $w$  and  $R$ .

ICA relies on the Central Limit Theorem, which states that (under certain conditions) the mix of two independent random variables is more Gaussian than the original variables. In other words, any linear combination of independent assembly activities is expected to be more Gaussian than the individual assembly activities themselves. From this reasoning, ICA iteratively rotates the axes of an initial random vector  $w$  in order to minimize some Gaussianity measure of  $y = w^T Z$ . As noted by Hyvarinen and Oja (2000), there are several measures of Gaussianity (or non-Gaussianity). Here we employ the fastICA algorithm (Hyvarinen, 1999; Hyvarinen and Oja, 1997) which uses approximating negentropy, defined as:

$$J_{\text{appr}}(y) = \sum_i [E\{G_i(y)\} - E\{G_i(v)\}]^2,$$

where  $v$  is a Gaussian variable with zero mean and unitary variance, and  $G_i$  are the non-quadratic functions:

$$G_1(u) = \frac{1}{a} \log \cosh au, \quad G_2(u) = -\exp\left(-\frac{u^2}{2}\right),$$

where  $1 \leq a \leq 2$ . For further details on approximating negentropy and other Gaussianity measures see Hyvarinen and Oja (2000).

Using ICA, Laubach et al. (1999, 2000) studied how neuronal ensembles in the motor cortex of rats evolved as the animals learned to perform a reaction-time task. In their work, however, the authors assumed that the number of eigenvalues larger than 1 represented the number of significant independent components (and hence the number of assembly patterns considered for further analyses). This procedure considerably overestimates the number of cell assemblies in the network, which, as reviewed above, should be taken as the number of eigenvalues above the upper limit of the Marčenko–Pastur distribution. We observed through simulations that the incorrect determination of the number of assemblies may lead to spurious results (not shown). This becomes more critical if ICA is performed in the original data space, i.e., with no prior dimensionality reduction. In this case, assembly patterns are broken into two or more independent components. In addition, since ICA by design does not extract co-activation patterns, some independent components may represent the activity of single neurons

that do not fire following a Gaussian distribution. To avoid these issues, here we suggest a modification of the original ICA approach which is achieved by incorporating the use of the Marčenko–Pastur distribution in order to estimate the number of cell assemblies. This modified approach involves first reducing the dimensionality of the spike matrix  $Z$  by projecting  $Z$  onto the subspace spanned by the significant PCs, and then computing the independent components through the fastICA algorithm. In simple terms, the new framework first finds the significant PCs and then rotates them to match the ideal assembly patterns.

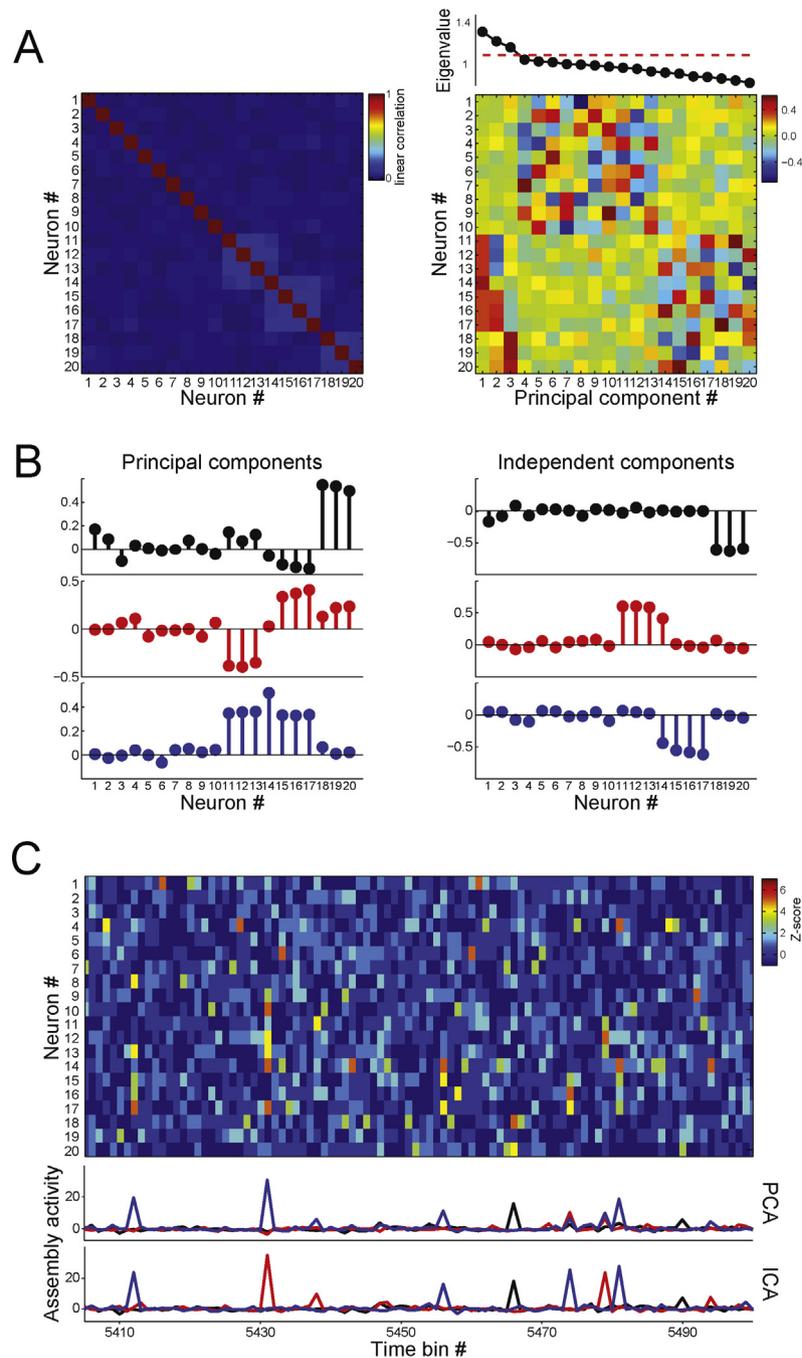
Fig. 7 illustrates the performance of PCA and modified ICA frameworks in extracting non-orthogonal assembly patterns. A spike matrix was simulated as before, but with 20 neurons. Assembly membership was: assembly 1, neurons #11, #12, #13, #14; assembly 2, neurons #14, #15, #16, #17; and assembly 3, neurons #18, #19, #20. Note that neuron #14 participates in assemblies 1 and 2. The correlation matrix is shown in Fig. 7A, along with its eigenvectors and eigenvalues. Since there are three cell assemblies, three eigenvalues lie above the theoretical threshold. The significant PCs and the independent components are shown in Fig. 7B. Note that while the independent components correctly attribute larger weights to the proper assembly members, the PCs fail to correctly segregate assembly patterns. Fig. 7C shows a time interval of the spike matrix along with the time course of assembly activity estimated by both methods. Note that the blue assembly computed by the PCA method peaks for the activations of both assemblies 1 and 2, while assembly activities computed from independent components correctly track individual assembly activations.

### 2.3.4. Other examples

For further comparisons, we applied the three methods described above to two other examples. In both cases, assemblies were highly distributed and thus difficult to segregate.

The first example is shown in Fig. 8. A spike matrix with 40 neurons (1 spike/bin mean rate) and 20,000 time bins was simulated; each neuron had 800 activation bins. As shown in Fig. 8A, the assembly identities were: assembly 1, neurons #1, #2, #3, #4, #5, #13, #24; assembly 2, neurons #4, #5, #6, #7, #8, #9, #18, #29; assembly 3, neurons #9, #10, #11, #12, #13, #14, #30; assembly 4, neurons #3, #14, #15, #16, #17, #18, #19, #20; assembly 5, neurons #18, #20, #21, #22, #23, #24, #25; assembly 6, neurons #2, #6, #10, #15, #21, #31. Fig. 8B shows the correlation matrix; the eigenvalue distribution is shown in Fig. 8C. Six eigenvalues lie above the upper bound and twenty-two below the lower threshold, which matches the number of simulated assemblies (6) and the number of assembly neurons (28).

Fig. 8D shows how each method estimated the assembly patterns. While PCA mixed the assembly patterns, the AV and ICA methods correctly attributed larger weights with same signs for assembly members within a pattern. For instance, independent component #1 and assembly vector #1 represented assembly 6 (green assembly in Fig. 8A). Since activation bins are independent among assemblies, assembly activities should ideally not correlate. Fig. 8E shows the histogram of linear correlation coefficients obtained for all possible pairs of assembly activities. Note that the PCA method estimates assembly activities that are more correlated than when estimated by the AV and ICA methods. The left panels in Fig. 8F show the time course of estimated assembly activities. Each black circle denotes actual assembly activations (some are stronger than others due to increased but random spike counts in activation bins). Note that the activities estimated by the AV and ICA methods peak at the actual assembly activations, while the time course of assembly activity estimated by PCA does not provide robust results for all assemblies. The right panels in Fig. 8F show matrices whose entries are the correlation coefficients for pairs of assembly activities estimated by the different methods (these values were used

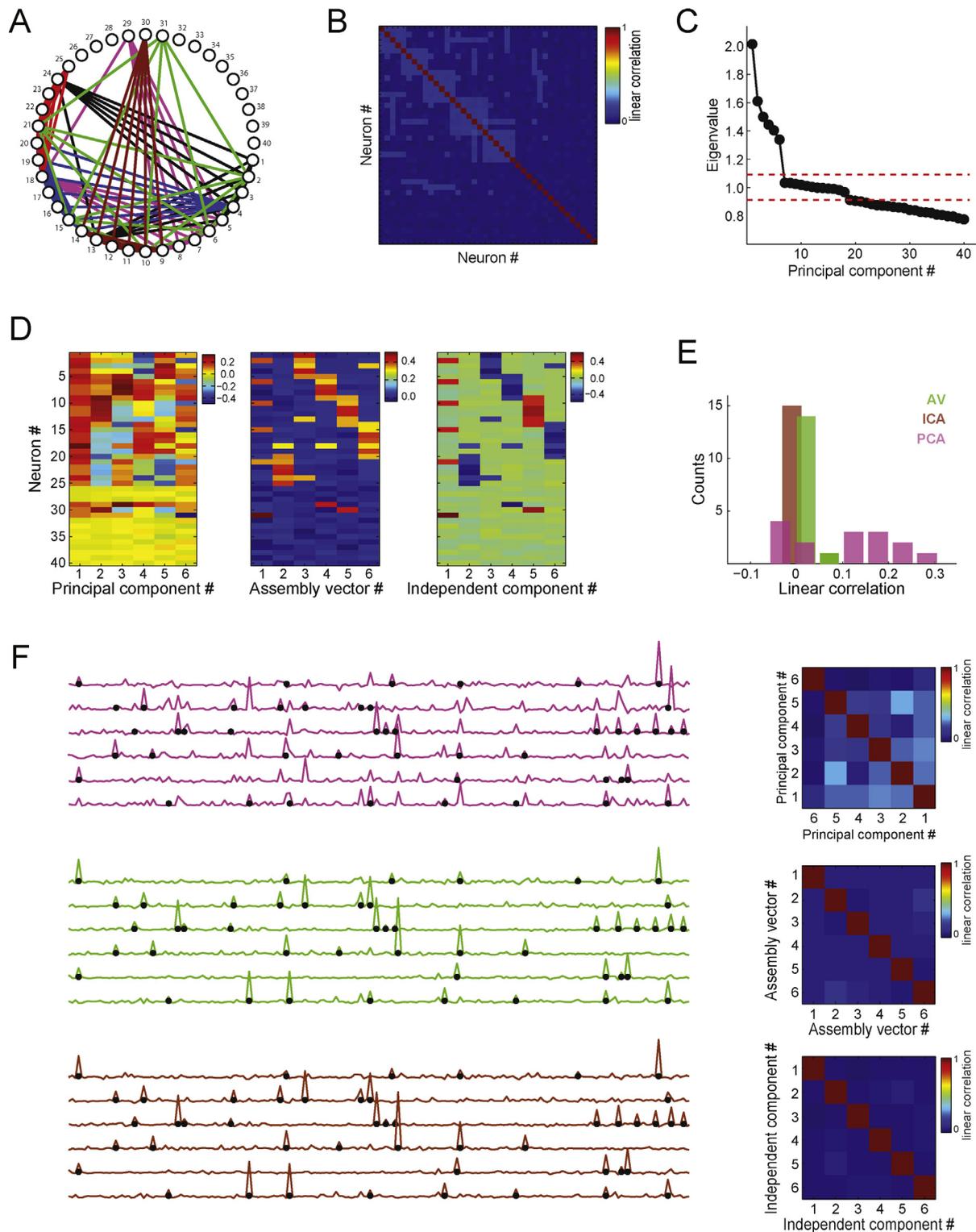


**Fig. 7.** Detecting and tracking cell assembly activity using independent component analysis. (A) Left: Correlation matrix of a spike matrix (not shown) with 20 neurons. In this example, 3 assemblies were present in the network (assembly 1: neurons #11, #12, #13, #14; assembly 2: neurons #14, #15, #16, #17; assembly 3: neurons #18, #19, #20). Right: Associated eigenvalues and principal components. Note that three eigenvalues are significant. (B) Assembly patterns estimated by principal components (left) and by independent components (right). Note that assembly patterns are better segregated in the latter case, i.e., each assembly pattern has large weights for its corresponding members. (C) Representative time interval of a spike matrix (top) and assembly activity time course estimated by principal component analysis (PCA) and independent component analysis (ICA) (bottom). Different colors represent projections obtained using different assembly patterns (as before, projectors are defined by the outer product of assembly patterns). Note that the blue trace in the PCA framework peaks for activations of both assemblies 1 and 2. On the other hand, assembly activity estimated by ICA properly segregates the three assemblies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

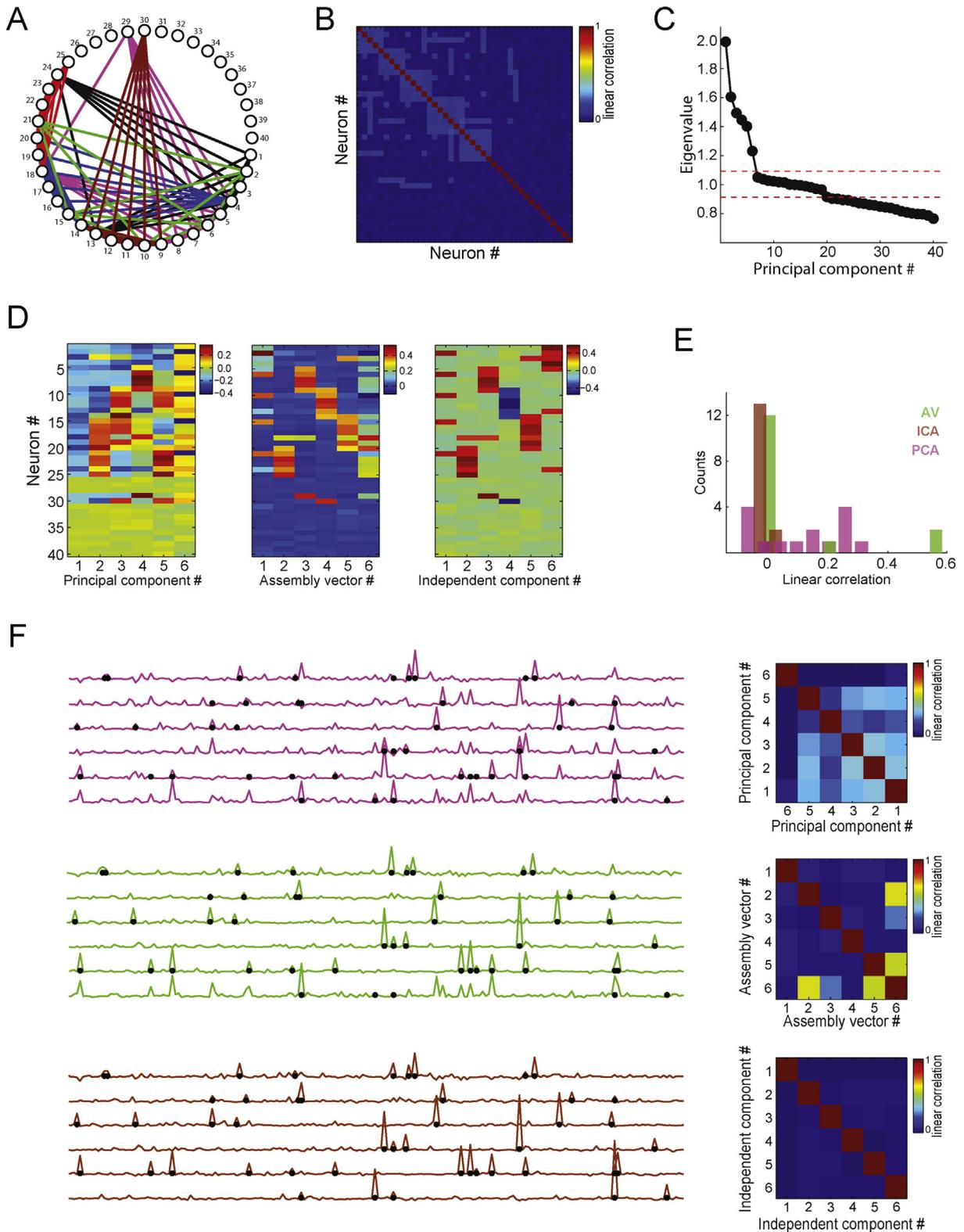
in Fig. 8E); again, notice lower correlations for the AV and ICA methods.

Fig. 9 shows a case in which the AV method fails. The spike matrix was simulated as in the last example but with a slight difference in assembly composition: neuron #31 was removed from assembly 6, and thus this assembly had no exclusive neuron (Fig. 9A). The correct number of assemblies is still detected by the analysis of the eigenvalues of the correlation matrix (Fig. 9B,C).

The assembly patterns estimated by all three methods are shown in Fig. 9D. The AV method fails to identify the pattern of assembly 6; in fact, the assembly vector #6 is a mixture of other assembly vectors. This is reflected in the histogram of the correlation coefficients between pairs of estimated assembly activities (Fig. 9E); in contrast to the previous example, now both the PCA and AV methods exhibit cases of high correlations. Fig. 9F shows the misleading estimation of the activity of assembly 6 by the AV method more clearly:



**Fig. 8.** PCA, ICA, and AV method performance for highly distributed assemblies. 40 neurons were simulated (mean 1 spike/bin) for 20,000 time bins. Each neuron had 800 activation bins (0 to 6 spikes uniformly distributed). (A) Simulated cell assemblies. Each circle represents a neuron. Neurons in the same assembly are connected by lines with same color (assembly 1: neurons #1, #2, #3, #4, #5, #13, #24; assembly 2: neurons #4, #5, #6, #7, #8, #9, #18, #29; assembly 3: neurons #9, #10, #11, #12, #13, #14, #30; assembly 4: neurons #3, #14, #15, #16, #17, #18, #19, #20; assembly 5: neurons #18, #20, #21, #22, #23, #24, #25; assembly 6: neurons #2, #6, #10, #15, #21, #31). (B) Correlation matrix. (C) Eigenvalue distribution. Six eigenvalues lie above the upper threshold and twenty-two lie below the lower threshold. (D) Assembly patterns estimated by the three methods. (E) Distribution of linear correlation coefficients among the time course of assembly activities when estimated by each of the three methods. Notice large correlations among cell assembly activities when estimated by the PCA framework. (F) Top panels: representative interval of the time course of assembly activities estimated from principal components (left). Black circles denote true (simulated) assembly activations. Note that assembly activations estimated by PCA do not perfectly match simulated activations. The right panel shows pairwise linear correlations between the time courses of different assemblies when estimated by PCA. Middle and bottom panels: Same as above but for ICA and AV methods, respectively. In these cases the peaks of the estimated cell assembly activities correspond well to the simulated activations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

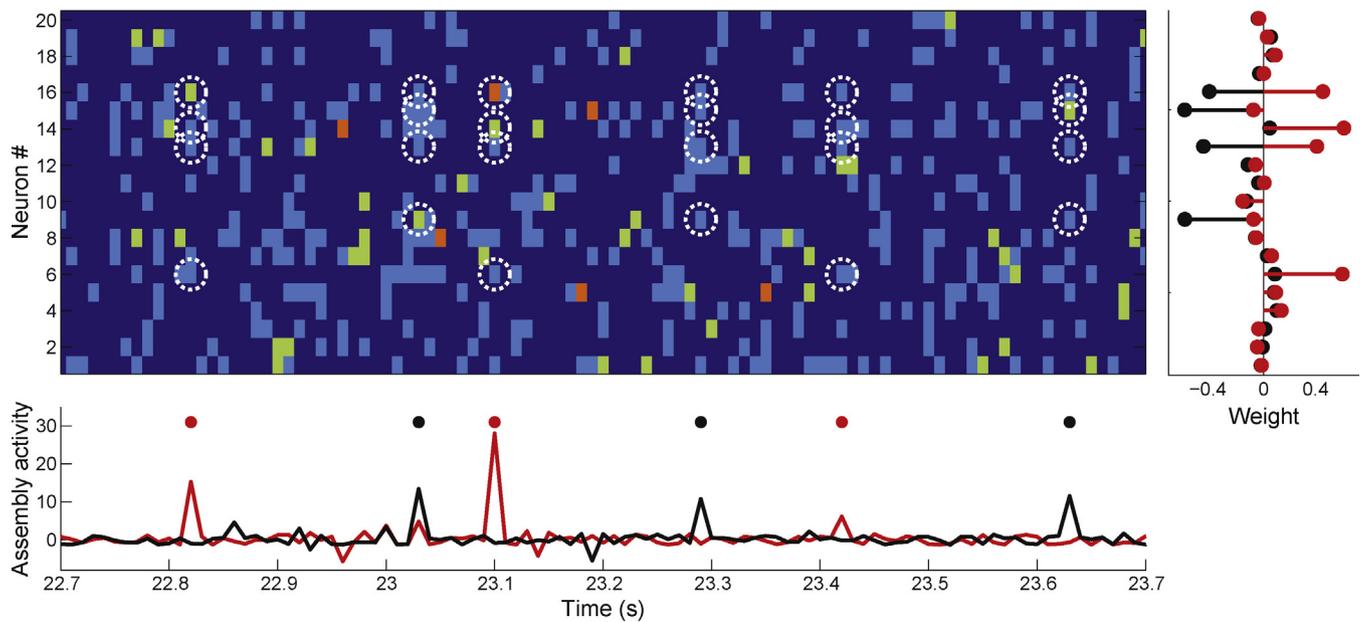


**Fig. 9.** PCA, ICA and AV method performance for cell assemblies with no exclusive neurons. 40 neurons were simulated as in Fig. 8. (A) Simulated cell assemblies. Same as Fig. 8, except that neuron #31 does not pertain to assembly 6, which consequently has no *exclusive* neuron (that is, all neurons in assembly 6 participate in at least one other assembly). (B) Correlation matrix. (C) Eigenvalue distribution. (D) Assembly patterns estimated by each method. Note that for this case the AV framework does not find assembly 6. (E) Same as Fig. 8E. (F) Same as Fig. 8F. Note that the AV method does not properly track the activity of assembly 6, which is estimated as a mixture of multiple assemblies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

note that while the activities of the first five assemblies are well estimated, the activity of the sixth assembly is mostly a mixture of the third and the fifth cell assemblies (left panel), as can also be seen from the correlation coefficients shown in the right panel.

Finally, notice that, as in Fig. 8, the ICA method still provides a good estimation for all cell assemblies.

We next tested the ICA method in a more realist condition. A spike matrix with 20 neurons was generated by circular shifting



**Fig. 10.** Application of the ICA-based framework to realistic simulations. Each neuron of the spike matrix was created from a circular shift of actual spike times of a hippocampal neuron (same procedure as in Fig. 3C). Two cell assemblies were simulated by including common spikes to assembly members, which were drawn from independent Poisson processes for each assembly (mean rate: 2 Hz). Assembly 1 was composed by neurons #9, #13, #15 and #16; assembly 2 by neurons #6, #13, #14 and #16 (notice 50% of overlap between assemblies). Top left panel shows a time interval of the spike matrix; circles mark co-firing of assembly members. Right panel shows assembly patterns extracted by the ICA-based framework. Bottom panel shows estimated time course of assembly activity computed from the assembly patterns. Events of the Poisson processes used to generate assembly activations are indicated by colored dots. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

actual spike trains as in Fig. 3C. In order to simulate a cell assembly, common spike times generated by a Poisson process with mean rate of 0.02 spikes/bin (2 Hz) were added to the spike trains of a group of neurons. We simulated two overlapping cell assemblies: assembly 1, composed by neurons #6, #13, #14 and #16; and assembly 2, composed by neurons #9, #13, #15 and #16. Accordingly, the ICA-based method extracted two assembly patterns, which corresponded to the programmed cell assemblies (Fig. 10 top right panel). We show a representative time interval of the spike matrix in the top left panel of Fig. 10. Dashed circles show co-activations of assembly members. The bottom panel shows estimated assembly activities along with colored dots that indicate events of the Poisson processes used to generate assembly activations. Note that the time courses of assembly activities estimated by the ICA method peak accordingly.

### 3. Real data applications

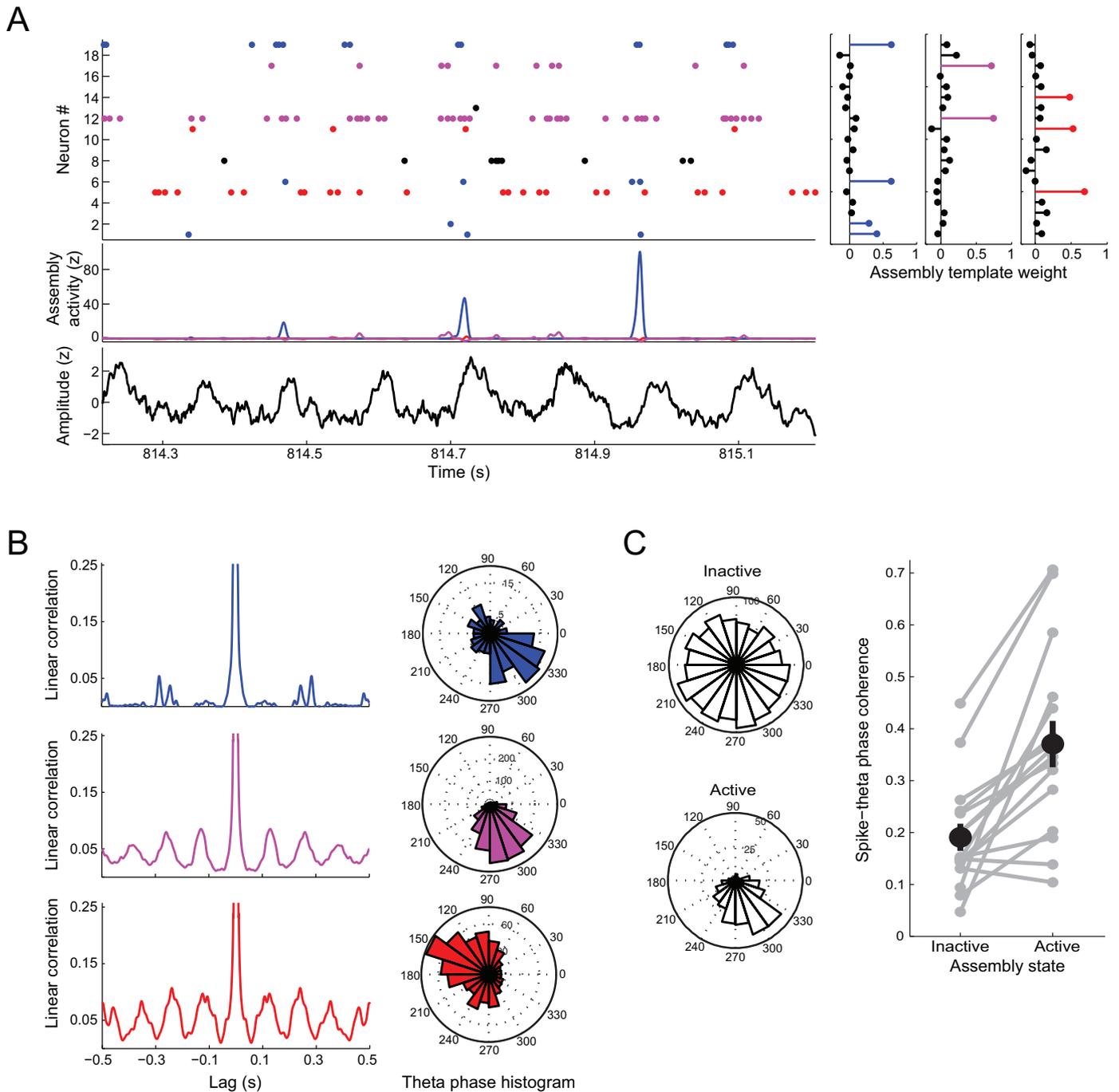
In this section we analyze spike trains of 19 single-units recorded from the CA1 region of the hippocampus of a freely moving rat exploring an open field. The data was downloaded from the Collaborative Research in Computational Neuroscience (CRCN) webpage (<http://crcns.org/>). Detailed information on recordings, behavior and surgery can be found in Mizuseki et al. (2009). The spike matrix was constructed using 10-ms bins. To estimate the number of cell assemblies, the eigenvalues of the correlation matrix were compared to the theoretical threshold derived from the Marčenko–Pastur distribution. Assembly patterns were extracted from the spike matrix using the modified ICA-based framework described above. We here further adapted this framework by increasing its time resolution. Since neuronal firing can occur in different bins even when spikes are separated by a few milliseconds, we estimated the time course of assembly activity by directly applying the projector operators to spike trains after smoothing them with a Gaussian kernel (maximum value = 1, standard deviation = 12 ms). In this framework, therefore, assembly activities can

be computed with the same time resolution as the local field potential (LFP), and coupling between these signals can be evaluated.

A time interval of the spike raster plot of the neuronal population is shown in the top left panel of Fig. 11A. The right panels show three assembly patterns extracted from the associated spike matrix. Neurons with large weights in a same pattern are displayed with same color. Estimated assembly activities are shown in the plot underneath the raster plot. Notice that only the blue assembly is active in the depicted period, and that assembly activity peaks when most of its neurons are co-active; in contrast, isolated firing of the same neurons does not affect assembly activity. The ongoing LFP is shown in the bottom panel. Examples from the same recording session in which the magenta and red assemblies are active are shown in Supplementary Fig. 1.

We used standard analysis techniques to study the activity of CA1 cell assemblies identified by this framework. Although it is usually assumed that neurons with higher firing probability at the same phase of an ongoing oscillation are part of the same cell assembly, non-overlapping assemblies with same phase preference has been previously shown by Harris et al. (2003). In accordance with this result, here we also observed that independent assemblies can have similar phase preferences. Fig. 11B shows the autocorrelation of the three assemblies highlighted in Fig. 11A along with their corresponding theta-phase distributions of assembly activations. Theta-phase distributions were obtained by (1) filtering the LFP into the theta range (6–10 Hz), (2) computing the instantaneous theta phases using the Hilbert transform, (3) localizing the phases associated with the peaks of assembly activity, and (4) expressing the phases by means of a circular histogram. Notice that the blue and magenta assemblies have similar preferred phase, while the red assembly was most active in a different theta phase. All assemblies were significantly coupled to ongoing theta oscillations (Rayleigh test,  $p < 10^{-6}$ ).

We next computed two theta-phase distributions for each cell assembly neuron. The first distribution was obtained by only taking into account spikes occurring during assembly activations (defined



**Fig. 11.** Real data applications. (A) Top-left panel shows a time interval of the spike raster plot of 19 single-units recorded from the hippocampal CA1 pyramidal layer. Top-right panels show three assembly patterns detected by the ICA-based method. The neurons with largest weights in each assembly were color-coded. Estimated assembly activities are shown below the raster plot. The bottom panel shows the ongoing local field potential (LFP). Assembly activities and LFP are z-scored. Note that the co-activation of the *blue* neurons is indicated by the peaks of the corresponding assembly activity. (B) Left panels show autocorrelations of the same cell assemblies as in A. Right panels show the theta-phase histogram of assembly activations. Note clear coupling between assembly activity and hippocampal theta oscillations. (C) Left panels show the theta phase distribution of spikes for an assembly neuron inside (bottom) and outside (top) assembly activations. Right panel shows mean spike-theta phase coherence of assembly neurons when the cell assembly was inactive or active (black circle; error bars denote standard deviation). Individual cases are shown in gray. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

as assembly activity above its mean + 2 standard deviations). The remaining spikes were used to construct the second distribution. A representative example is shown in Fig. 11C: the bottom left panel shows the distribution of spike phases during assembly activations, whereas the top left panel shows the phase distribution for spikes occurring when the cell assembly was inactive. Note that the neuron in this example is strongly coupled to the theta rhythm

when its assembly is active. To compute a group result, for each assembly neuron we calculated the spike-phase coherence inside and outside assembly activations. Spike-phase coherence was defined as the length of the mean phasor  $e^{i\varphi}$ , where  $\varphi$  is a theta-phase in which a spike occurred (Siapas et al., 2005). The right panel of Fig. 11C shows the spike-phase coherence for each assembly neuron in gray; the black circle denotes the means for

each case and bars indicate standard deviations. These results show that neurons couple more strongly to the theta rhythm when firing in synchrony with other assembly members ( $p=0.0039$ , Wilcoxon rank sum test).

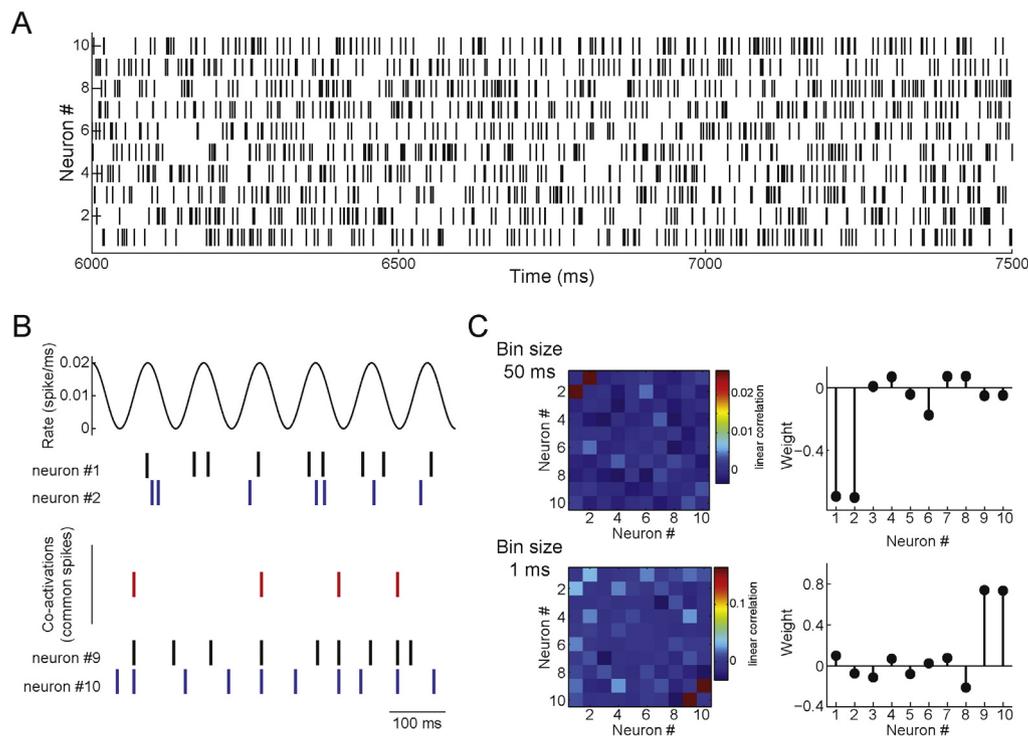
#### 4. Discussion

Recent advances have opened the possibility of testing influential theories on network functioning and its relation to behavior and memory. Here we reviewed three linear methods for computing cell assembly activity. We began by introducing the original PCA approach, which was the first employed to track cell assemblies (Nicolelis et al., 1995), and is still useful nowadays in generating insights about organized spiking activity (Benchenane et al., 2010; Peyrache et al., 2009). Next, we showed that assembly patterns estimated by PCs have important limitations that can be overcome the AV framework (Lopes-dos-Santos et al., 2011), which identifies cell assemblies based on correlation patterns within the subspace spanned by the significant PCs. Finally, we presented a modified version of a previously described ICA approach (Laubach et al., 1999), which incorporates the statistical threshold provided by the Marčenko–Pastur distribution. We showed that this method provides even more robust estimates of assembly patterns than the AV method.

The large interest in this field led to the development of analytical tools able to identify high-order correlations that can potentially replace pair-wise analyses (Stevenson and Kording, 2011). However, several methods can only be applied to small networks due to combinatorial explosion (Berger et al., 2010; Grun et al., 2002; Tetko

and Villa, 2001). In order to overcome this problem, some authors have proposed methodologies that detect high-order correlations ignoring how cells are involved in the coalitions (Louis et al., 2010; Staude et al., 2010a,b). The methods reviewed here can be applied to large neuronal populations, and neuronal activity is assessed as a whole through the use of eigenvalue analysis. Additionally, the Marčenko–Pastur distribution opens the possibility of using an analytical and reliable statistics instead of surrogate methods employed in previous frameworks (Abeles, 2009; Abeles and Gat, 2001; Abeles and Gerstein, 1988; Humphries, 2011; Maldonado et al., 2008; Shmiel et al., 2006). In addition to being computationally demanding, a problem inherent to the use of surrogates is the fact that there is no consensus about which statistical properties should be preserved in these control data (Berger et al., 2010; Grun, 2009). For example, we have shown here that surrogate methods that preserve ISI counts and auto-correlations provide different thresholds for statistical significance than less conservative shufflings (Fig. 3C).

We note that although ICA assumes a linear model for cell assemblies, it employs nonlinear equations in order to quantify Gaussianity, which can be greatly optimized by the fastICA algorithm (Hyvarinen and Oja, 1997). In fact, while the ICA-based framework is more computationally demanding than the PCA and AV methods, it is still faster than most nonlinear algorithms. In general, frameworks employing more complex mathematics require extensive data crunching that often yield results difficult to interpret by non-specialized researchers. The methods studied here are intuitive and generate results easy visualize in raw data, which should favor their comprehension and broad use by the scientific community.



**Fig. 12.** The influence of bin size on assembly detection. (A) Spike raster plot of ten simulated neurons. Neurons #1 and #2 are Poisson processes with a common varying mean rate following a 10-Hz sinusoid (minimum and maximum rate: 0 and 0.02 spike/ms). Neurons #3 to #10 are independent Poisson processes with 0.01 spike/ms mean rate. In addition, spikes drawn from a Poisson process with 0.3-Hz mean rate were added to neurons #9 and #10 in order to simulate a cell assembly. (B) Top panels show the varying mean rate of neurons #1 and #2 along with their spike times. Note that neurons are more active around the peak of the sinusoid. Bottom panels show the Poisson events used to create common spikes for neurons #9 and #10, along with their spike times. (C) Top panels show the correlation matrix when 50-ms bins are used to construct the spike matrix (left) and the assembly pattern extracted by the ICA-based method (right). Note that the pattern corresponds to neurons #1 and #2 that co-activate at a time scale defined by the sinusoid. Bottom panels show the same results but for when 1-ms bins are used to construct the spike matrix. Note that at this time resolution only co-activations of neurons #9 and #10 are detected. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

#### 4.1. Limitations

None of revised methods is optimal, and there is room for further improvements. For example, ICA cannot find assemblies if they follow a Gaussian distribution. Moreover, non-linear spike correlations may produce spurious results in assembly detection methods that depend on PCA. Roughly speaking, when the scatterplot of two random variables has a curved shape, the eigenvalue analysis may reveal two dimensions with large variance. By consequence, PCA-based methods would falsely identify two assemblies in the network, while only one pair of neurons is correlated. It should be noted that other methods, such as the one described in Humphries (2011), can be adapted to other measures of spike train similarity besides linear correlations.

It is also relevant to note that the bin size used to create the spike matrix (Fig. 1) defines the time scale of the detected assemblies, and thus different bin sizes can yield different results. We illustrate this possibility in Fig. 12. In this example, neurons in assembly 1 fire at a Poisson rate defined by a 10-Hz sinusoid (Fig. 12B top), while neurons in assembly 2 fire synchronous spikes based on the realizations of a Poisson process, despite also emitting independent random spikes outside assembly activations (Fig. 12B bottom). Using a 1-ms bin size to construct the spike matrix leads to the detection of assembly 2 (Fig. 12C bottom), which neurons spike in precise synchrony; assembly 1 is not detected because of the jitter in the spike times of its neurons. On the other hand, employing a 50-ms bin size leads to the exclusive detection of assembly 1 (Fig. 12C top), since at this time scale assembly 1 neurons are co-active and the random spikes of assembly 2 neurons occurring when the assembly is not active mask assembly 2 activations. Thus, the extracted co-activation patterns have a time scale defined by the bin size, which should be considered when interpreting results.

Finally, one should note that the reviewed methods do not disambiguate between stimulus-driven and internally generated spike correlations. For instance, spike correlations can be detected when neurons have similar tuning curves, irrespective of whether they are wired together or not. Thus, a pair of CA1 pyramidal cells with overlapping place fields can be potentially identified as composing an assembly depending on the time scale (bin size) employed in the analysis. However, we note that this feature is inherent to the definition of a cell assembly, which is usually taken as a group of cells that fire together (and collectively represent an object, a sensation, an action, etc.), independently of what causes the correlated firing. In our view, whether neurons presenting correlated activity due to common sensory inputs should or not be considered a cell assembly is a matter of definition.

## 5. Conclusion

Linear methods are computationally low demanding, and yet quite efficient for tracking cell assembly activity. Hebb's seminal work comprises one of the most influential theories in modern neuroscience. Yet, to date only few studies have addressed Hebb's ideas at the systems level. We hope that the methods reviewed here can lead to a proper estimation of co-activation patterns and help answering whether cell assemblies have any functional role, as originally envisioned by Hebb.

## Acknowledgements

Supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Apoio à Pesquisa do Estado do Rio Grande do Norte (FAPERN). We thank the Buzsáki lab for making in vivo CA1 recordings publicly available through

the Collaborative Research in Computational Neuroscience website (<http://crcns.org/>), a data sharing website funded by the National Science Foundation. We also thank Sergio Conde-Ocazonez, Enio Aguiar and Adrien Peyrache for helpful discussions. The authors declare no competing financial interests.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jneumeth.2013.04.010>.

## References

- Abeles M. Synfire Chains 2009, [http://www.scholarpedia.org/article/Synfire\\_chains](http://www.scholarpedia.org/article/Synfire_chains)
- Abeles M, Gat I. Detecting precise firing sequences in experimental data. *J Neurosci Methods* 2001;107:141–54.
- Abeles M, Gerstein GL. Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J Neurophysiol* 1988;60:909–24.
- Adrian ED, Zotterman Y. The impulses produced by sensory nerve-endings Part II. The response of a single end-organ. *J Physiol (Lond)* 1926;61:151–71.
- Arabzadeh E, Panzeri S, Diamond M. Whisker vibration information carried by rat barrel cortex neurons. *J Neurosci* 2004;24:6011–20.
- Benchenane K, Peyrache A, Khamassi M, Tierney PL, Gioanni Y, Battaglia FP, et al. Coherent Theta oscillations and reorganization of spike timing in the hippocampal-prefrontal network upon learning. *Neuron* 2010;66:921–36.
- Berger D, Borgelt C, Louis S, Morrison A, Grün S. Efficient identification of assembly neurons within massively parallel spike trains. *Comput Intel Neurosci* 2010;18.
- Biroli G, Bouchaud JP, Potters M. On the top eigenvalue of heavy-tailed random matrices. *Europhys Lett* 2007;78:5.
- Buzsáki G. Large-scale recording of neuronal ensembles. *Nat Neurosci* 2004;7:446–51.
- Chapin JK, Nicolelis MAL. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *J Neurosci Methods* 1999;94:121–40.
- Comon P. Independent component analysis, a new concept? *Signal Process* 1994;36:287–314.
- Gansel KS, Singer W. Detecting multineuronal temporal patterns in parallel spike trains. *Front Neuroinform* 2012;6.
- Grun S. Data-driven significance estimation for precise spike correlation. *J Neurophysiol* 2009;101:1126–40.
- Grun S, Diesmann M, Aertsen A. Unitary events in multiple single-neuron spiking activity: 1 detection and significance. *Neural Comput* 2002;14:43–80.
- Harris KD, Csicsvari J, Hirase H, Dragoi G, Buzsáki G. Organization of cell assemblies in the hippocampus. *Nature* 2003;424:552–6.
- Hebb DO. *The organization of behaviour: a neuropsychological theory*. New York: Psychology Press; 1949.
- Hubel DH, Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *J Physiol (Lond)* 1959;148:574–91.
- Humphries MD. Spike-train communities finding groups of similar spike trains. *J Neurosci* 2011.
- Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Networks* 1999;10:626–34.
- Hyvarinen A, Oja E. A fast fixed-point algorithm for independent component analysis. *Neural Comput* 1997;9:1483–92.
- Hyvarinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Networks* 2000;13:411–30.
- Jutten C, Herault J. Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Process* 1991;24:1–10.
- Laubach M, Shuler M, Nicolelis M. Independent component analyses for quantifying neuronal ensemble interactions. *J Neurosci Methods* 1999;94:141–54.
- Laubach M, Wessberg J, Nicolelis MAL. Cortical ensemble activity increasingly predicts behaviour outcomes during learning of a motor task. *Nature* 2000;405.
- Lee AK, Wilson MA. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* 2002;36:1183–94.
- Lopes-dos-Santos V, Conde-Ocazonez S, Nicolelis M, Ribeiro S, Tort A. Neuronal assembly detection and cell membership specification by principal component analysis. *PLoS One* 2011;6:e20996.
- Louie K, Wilson M. Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* 2001;29:145–56.
- Louis S, Borgelt C, Grun S. Complexity distribution as a measure for assembly size and temporal precision. *Neural Networks* 2010;23:705–12.
- Maldonado P, Babul C, Singer W, Rodriguez E, Berger D, Grun S. Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. *J Neurophysiol* 2008;100:1523–32.
- Marčenko VA, Pastur LA. Distribution of eigenvalues for some sets of random matrices. *Math USSR-Sbornik* 1967;1:457–83.
- Mizusaki K, Sirota A, Pastalkova E, Buzsáki G. Theta oscillations provide temporal windows for local circuit computation in the entorhinal-hippocampal loop. *Neuron* 2009;64.

- Nicolelis MAL, Baccala LA, Lin RCS, Chapin JK. Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science* 1995;268:1353–8.
- Okeefe J, Dostrov J. Hippocampus as a spatial map – preliminary evidence from unit activity in freely-moving rat. *Brain Res* 1971;34:171–5.
- Perrett D, Rolls E, Caan W. Visual neurones responsive to faces in the monkey temporal cortex. *Exp Brain Res* 1982;47:329–42.
- Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP. Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat Neurosci* 2009;12(919), U143.
- Peyrache A, Benchenane K, Khamassi M, Wiener SI, Battaglia FP. Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution. *J Comput Neurosci* 2010;29:309–25.
- Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, Stanley HE. Random matrix approach to cross correlations in financial data. *Phys Rev E* 2002;65.
- Quiroga RQ, Panzeri S. Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci* 2009;10:173–85.
- Ribeiro S, Gervasoni D, Soares ES, Zhou Y, Lin SC, Pantoja J, et al. Long-lasting novelty-induced neuronal reverberation during slow-wave sleep in multiple forebrain areas. *PLoS Biol* 2004;2, 126–37.
- Seba P. Random matrix analysis of human EEG data. *Phys Rev Lett* 2003;91:4.
- Shmuel T, Drori R, Shmuel O, Ben-Shaul Y, Nadasdy Z, Shemesh M, et al. Temporally precise cortical firing patterns are associated with distinct action segments. *J Neurophysiol* 2006;96, 2645–52.
- Siapas AG, Lubenov EV, Wilson MA. Prefrontal phase locking to hippocampal theta oscillations. *Neuron* 2005;46, 141–51.
- Staud B, Grun S, Rotter S. Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference. *Front Comput Neurosci* 2010a:4.
- Staud B, Rotter S, Grun S. CuBIC: cumulant based inference of higher-order correlations in massively parallel spike trains. *J Comput Neurosci* 2010b;29, 327–50.
- Stevenson IH, Kording KP. How advances in neural recording affect data analysis. *Nat Neurosci* 2011;14:139–42.
- Tetko IV, Villa AEP. A pattern grouping algorithm for analysis of spatiotemporal patterns in neuronal spike trains, detection of repeated patterns. *J Neurosci Methods* 2001;105:1–14.
- Tracy CA, Widom H. Level-spacing distributions and the airy kernel. *Commun Math Phys* 1994;159:151–74.
- Wilson MA, McNaughton BL. Reactivation of hippocampal ensemble memories during sleep. *Science* 1994;265:676–9.

---

## CAPÍTULO 2: EXTRAINDO INFORMAÇÃO DE PADRÕES TEMPORAIS DE DISPARO

# Extracting information in spike time patterns with wavelets and information theory

Vítor Lopes-dos-Santos,<sup>1,2</sup> Stefano Panzeri,<sup>3,4</sup> Christoph Kayser,<sup>5,6</sup> Mathew E. Diamond,<sup>7</sup> and Rodrigo Quian Quiroga<sup>2</sup>

<sup>1</sup>Brain Institute, Federal University of Rio Grande do Norte, Natal, Brazil; <sup>2</sup>Centre for Systems Neuroscience, University of Leicester, Leicester, United Kingdom; <sup>3</sup>Center for Neuroscience and Cognitive Systems @UniTn, Istituto Italiano di Tecnologia, Rovereto, Italy; <sup>4</sup>Max Planck Institute for Biological Cybernetics, Tübingen, Germany; <sup>5</sup>Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom; <sup>6</sup>Bernstein Center for Computational Neuroscience, Tübingen, Germany; and <sup>7</sup>Tactile Perception and Learning Laboratory, International School for Advanced Studies, Trieste, Italy

Submitted 23 May 2014; accepted in final form 12 November 2014

**Lopes-dos-Santos V, Panzeri S, Kayser C, Diamond ME, Quian Quiroga R.** Extracting information in spike time patterns with wavelets and information theory. *J Neurophysiol* 113: 1015–1033, 2015. First published November 12, 2014; doi:10.1152/jn.00380.2014.—We present a new method to assess the information carried by temporal patterns in spike trains. The method first performs a wavelet decomposition of the spike trains, then uses Shannon information to select a subset of coefficients carrying information, and finally assesses timing information in terms of decoding performance: the ability to identify the presented stimuli from spike train patterns. We show that the method allows: 1) a robust assessment of the information carried by spike time patterns even when this is distributed across multiple time scales and time points; 2) an effective denoising of the raster plots that improves the estimate of stimulus tuning of spike trains; and 3) an assessment of the information carried by temporally coordinated spikes across neurons. Using simulated data, we demonstrate that the Wavelet-Information (WI) method performs better and is more robust to spike time-jitter, background noise, and sample size than well-established approaches, such as principal component analysis, direct estimates of information from digitized spike trains, or a metric-based method. Furthermore, when applied to real spike trains from monkey auditory cortex and from rat barrel cortex, the WI method allows extracting larger amounts of spike timing information. Importantly, the fact that the WI method incorporates multiple time scales makes it robust to the choice of partly arbitrary parameters such as temporal resolution, response window length, number of response features considered, and the number of available trials. These results highlight the potential of the proposed method for accurate and objective assessments of how spike timing encodes information.

neuronal coding; temporal coding; wavelets; information theory; decoding

THE IMPORTANCE OF PRECISE spike timing in carrying meaningful information has attracted much attention (Quian Quiroga and Panzeri 2009; Rieke et al. 1999). Does the temporal structure of spike trains provide information beyond the total spike count, or does it merely reflect noise? According to the “rate coding” view, neurons represent stimuli solely by the rate of firing within an encoding time window (Adrian and Zotterman 1926; Shadlen and Newsome 1994). In contrast, according to the “temporal coding” view, the time structure of the responses

conveys additional information not provided by the total spike count (de Ruyter van Steveninck et al. 1997; Optican and Richmond 1987; Richmond and Optican 1987; Victor and Purpura 1996).

Experimental evidence accumulated over the last 3 decades has suggested that precise spike patterns, on the scale of milliseconds, do indeed convey information not available in rate codes (Arabzadeh et al. 2006; de Ruyter van Steveninck et al. 1997; Di Lorenzo et al. 2009; Eckhorn and Popel 1975; Foffani et al. 2009; Fontanini and Katz 2006; Kayser et al. 2010; Laurent et al. 1996; Montemurro et al. 2007; Panzeri et al. 2001, 2010; Quian Quiroga and Panzeri 2009; Richmond and Optican 1987; Victor 2000). For this, a straightforward way to assess the significance of spike timing has been to represent spike trains as sequences of “0s” and “1s,” denoting the absence or the presence of a spike in poststimulus time bins and then, using the formalism of information theory, evaluate whether the information about stimulus identity carried by such patterns is significantly larger than the information carried by spike counts alone (de Ruyter van Steveninck et al. 1997; Kayser et al. 2009; Panzeri et al. 2001; Strong et al. 1998). However, this approach leads to a combinatorial explosion (the “curse of dimensionality”) because the number of possible response patterns increases exponentially with the number of bins (Panzeri et al. 2007). Thus, for an experimentally feasible number of trials, this limits the precision of the temporal patterns to be studied (i.e., the size of the time bin) and the length of the response considered.

A solution to the combinatorial explosion problem is to reduce the dimensionality of the spike trains. To this end, a well-known approach is to compress the neural responses into a small number of features using principal component analysis (PCA). By using this method, Richmond and Optican (1987) showed that time patterns in responses from neurons in the macaque inferior temporal cortex could disambiguate visual stimuli that could not be distinguished by firing rate alone (Optican and Richmond 1987; Richmond and Optican 1987). Despite the value of this application, the PCA-based decomposition has two main caveats. First, PCA represents directions of maximum variance, which are not necessarily the directions with the largest information. Second, PCA coefficients are not localized in time and may not capture sources of information that are precisely localized at one or a few poststimulus time

Address for reprint requests and other correspondence: R. Quian Quiroga, Centre for Systems Neuroscience, Univ. of Leicester, Leicester, LE1 7QR, United Kingdom (e-mail: rqqg1@le.ac.uk).

points (Panzeri et al. 2001) or even encoded at multiple temporal scales (Fotowat et al. 2011; Harvey et al. 2013; Kayser et al. 2009; Panzeri et al. 2010). Here, we propose another type of dimensionality reduction that is able to capture time-localized information encoded at multiple time scales (Fig. 1). The method combines wavelet decomposition and information theory to identify first the features in the spike patterns carrying relevant information and then use these features to quantify the amount of sensory encoding carried by these responses using a decoding approach. We validate the method on simulated spike trains and compare its performance with that obtained with PCA, direct estimations of information from digitized neuronal responses, and a widely used metric-space (MS) method (Victor and Purpura 1996). Results on simulated data demonstrate that the Wavelet-Information (WI) method is more robust and extracts more spike timing information than previous methods for a wide range of background firing rates and intertrial jitters. The advantages of the WI method are confirmed by evaluating its performance with experimental data from the monkey auditory cortex and the rat somatosensory cortex. Additionally, we show that the same approach can be used: 1) to denoise spike trains, providing a more robust quantification of the stimulus selectivity; and 2) to assess and visualize the information carried by the synchronous firing of neurons.

**MATERIALS AND METHODS**

*Wavelet Decomposition*

The wavelet transform is the inner product of a signal with dilated and translated versions of a wavelet function (Mallat 2008; Strang and Nguyen 1996). Formally, given a signal  $x(t)$  and a wavelet function  $\psi_{a,b}(t)$ , the continuous wavelet transform (CWT) is defined as:

$$W_{\Psi}X(a, b) = \langle x, \Psi_{a,b} \rangle, \text{ with}$$

$$\Psi_{a,b} = |a|^{-1/2} \Psi\left(\frac{t-b}{a}\right),$$

where  $a, b \in \mathbb{R}$  are the scale and translation parameters, respectively. The translation parameter changes the location of the wavelet function, whereas the scale parameter dilates or compresses it. The correlation of the signal  $x(t)$  with the dilated (contracted) versions of the wavelet  $\psi_{a,b}(t)$  gives the low- (high-) frequency components. The CWT is redundant, and, without any loss of information, it is practical to define the wavelet transform only at discrete scales  $a_j = 2^j$  and times  $b_{j,k} = 2^j k$ , which is called the dyadic wavelet transform (DWT). The DWT is nonredundant in the sense that from  $N$  data points we obtain  $N$  wavelet coefficients, each of them representing the amount of activity of the original signal at a specific time and scale. Furthermore, patterns in the signal with different frequency and time localizations are represented by specific wavelet coefficients. The DWT can be computed using a hierarchical and very efficient algorithm called multiresolution decomposition (Mallat 1999). This algorithm successively divides the signal into coarse approximations and details at different scales. The end result is the decomposition of the original signal into a series of detail scales and a final approximation, corresponding to the time-localized activity in different frequency bands.

Starting from the binned spike trains, in this study we implemented a five-scale dyadic wavelet decomposition using Haar wavelets, which is a square function that is ideally suited to identify local contrasts at different scales. The spike trains were always binned with 1-ms windows unless stated otherwise.

*Selection of Wavelet Coefficients*

From the total set of wavelet coefficients, equal to the number of bins in the spike trains, we selected a subset of coefficients based on their mutual information with the stimuli, defined as (Shannon 1948):

$$I_{S,w_c} = \sum_{S,w_c} P(S, w_c) \log_2 \frac{P(S, w_c)}{P(S)P(w_c)},$$

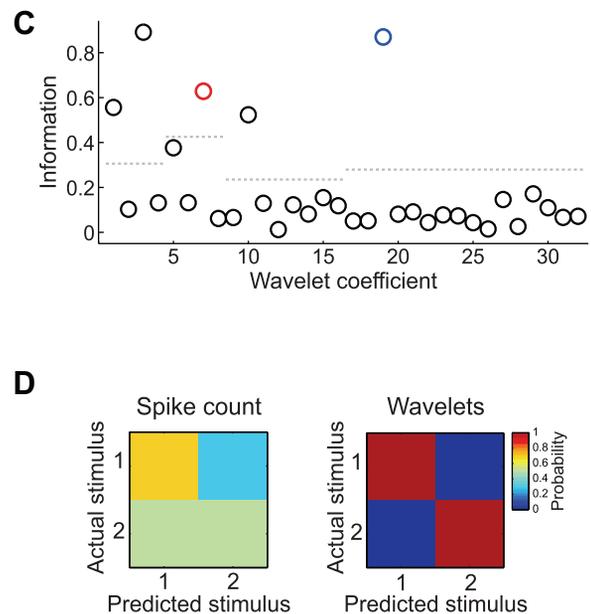
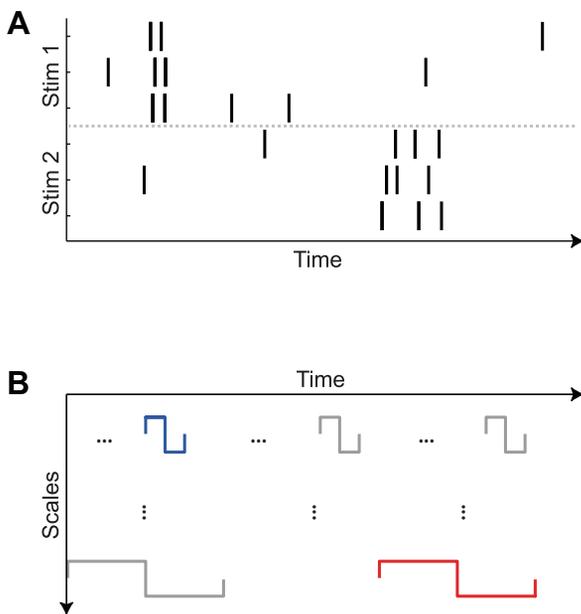


Fig. 1. Description of the method. *A*: simulated neural responses to 2 stimuli (Stim). *B*: responses are binned and convolved with Haar wavelets. Note that for the example in *A*, the time pattern associated with *stimulus 1* matches both scale and time localization of the wavelet in the 1st level displayed in blue, whereas the time pattern associated with *stimulus 2* has a larger jitter and fits with a wavelet in a higher scale (shown in red). *C*: wavelet coefficients with significant time-pattern information are identified. Dashed lines represent statistical threshold for information significance for each decomposition level (see MATERIALS AND METHODS). The red and blue circles denote the information values corresponding to the wavelet coefficients in *B*. *D*: decoding results in the form of confusion matrices using the total spike count (*left*) or the informative wavelet coefficients (*right*) as inputs to the decoder.

where  $S$  is the set of stimuli and  $w_c$  is the set of values of a wavelet coefficient  $c$ . The significance of the information given by each coefficient  $I_{S,w_c}$  was established based on surrogate testing: for each coefficient, we calculated a distribution of information values obtained by shuffling trials (i.e., randomizing trial-stimulus relations) 20 times. Surrogate distributions were calculated separately for each decomposition level (information values obtained from coefficients of the same levels were combined), and the 95th percentiles of each distribution were used as statistical thresholds (horizontal dashed lines in Fig. 1C). To avoid having too many features with significant information, if  $>25$  coefficients were significant, we used the 25 with the largest unbiased information (the same restriction was applied to the other methods). In this context, unbiased information was defined as the difference between the direct measure of information and its corresponding statistical threshold computed from shuffling the trials. Additionally, if none of the coefficients crossed the statistical threshold, we used the 2 with the largest information.

#### Stimulus Decoding and Information Estimation from Confusion Matrices

To estimate the information in a set of features, we used a cross-validated (leave-1-out) naïve Bayesian decoder to assign the response on each trial in the testing set to a given stimulus, which gives a lower bound of the information available in the spike trains (Quiroga and Panzeri 2009). Decoding performance was computed as the proportion of correct predictions. The decoder optimization and the selection of features were based solely on the training trials. For comparison, we also used linear discriminant analysis (Fisher 1936; Quiroga et al. 2007) and nearest neighbors classifiers (in which case, features were assigned  $z$ -scores to avoid scaling problems) and obtained virtually the same results.

In cases where the linear decoder introduced systematic errors (Fig. 7), we computed the mutual information between the actual and the predicted stimuli from the confusion matrices:

$$I(S, S^P) = \sum_{S, S^P} P(S, S^P) \log_2 \frac{P(S, S^P)}{P(S)P(S^P)},$$

where  $S$  is the set of actual stimuli presented to the decoder and  $S^P$  is the set of predicted stimuli by the decoder. To correct for the upward limited-sampling bias in the information estimate, we used the quadratic extrapolation procedure described elsewhere (Panzeri et al. 2007; Strong et al. 1998) and implemented in Magri et al. (2009).

#### MS Method

We also compared the WI method with the MS approach (Victor and Purpura 1996), which clusters responses based on a distance metric between spike trains. This distance is defined as the minimum “cost” of converting a spike train into another one by deleting, inserting, or moving spikes. The cost of deleting or inserting a spike is always set to 1, and the cost of moving a spike per unit of time is given by the free parameter  $q$  (expressed in units of 1/ms), which has to be optimized for each data set. Thus, when  $q$  is 0, moving a spike is free, and therefore only the spike count is taken into consideration. As  $q$  is increased, more weight is given to the precise timing of the spikes. Note that since moving a spike by  $1/q$  ms has the same cost as deleting it,  $1/q$  defines the temporal precision of the analysis. With the MS method, we classified trials using a nearest neighbor decoder; more specifically, we assigned each tested trial to the class of its nearest neighbor in the training set. We systematically varied  $q$  from 0.001 to 524/ms in half-octave intervals. Only representative  $q$ -values are reported. For computing the spike train distances, we used a MATLAB function available at <http://www-users.med.cornell.edu/~jdvicto/spkdm.html>.

#### Spike Train Denoising

To visualize spike patterns containing information, we adapted the WI method to denoise the spike trains by: 1) computing the wavelet decomposition of the mean peristimulus time histogram (PSTH) of each stimulus; 2) denoising the mean PSTHs by reconstructing them using only the wavelet coefficients with significant information; 3) setting to 0 the denoised PSTHs values below a threshold at 1 SD of the absolute values (taken from the denoised PSTHs of all stimuli); and 4) using the denoised and thresholded PSTHs as masks, preserving only spikes in bins passing this threshold. This effectively preserves spikes conveying stimulus information and deletes the others. We remark that although the selection of informative wavelet coefficients was the same for all stimuli, the masks were different for each stimulus.

#### Estimation of Information in Correlated Spike Patterns

To estimate the information given by the correlated firing of pairs of neurons ( $i, j$ ), we computed the wavelet decomposition for each trial  $n$  and calculated the normalized distance between the values of a corresponding wavelet coefficient (i.e., considering a specific time location and scale) in both neurons as:

$$D_{c,n}^{i,j} = \frac{|w_{c,n}^i - w_{c,n}^j|}{|w_{c,n}^i| + |w_{c,n}^j|},$$

where  $w_{c,n}^i$  is the value of wavelet coefficient  $c$  of neuron  $i$  at trial  $n$ . Analogous to the procedure described in Fig. 1C, we then selected the distances  $D_{c,n}^{i,j}$  that had significant information about the stimuli and used these distance values for decoding. In other words, we implemented the same procedure as before but using the distances  $D_{c,n}^{i,j}$  between the wavelet coefficients of each neuron instead of the value of the coefficients themselves.

#### Experimental Data

All procedures with monkeys reported here were approved by the local authorities (Regierungspräsidium Tübingen) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC). The animals were socially (group-) housed in an enriched environment under daily veterinary supervision. Experiments with rats were conducted in accordance with National Institutes of Health and international standards for the care and use of animals in research and were approved by the International School for Advanced Studies Ethics Committee. Procedures were supervised by a consulting veterinarian.

**Monkey A1 data.** As described in a previous work (Kayser et al. 2010), neural activity was recorded from caudal auditory cortex (mainly areas A1 and caudal belt) of three alert animals using multiple microelectrodes. The data were high-pass filtered (4 Hz), amplified (AlphaOmega, Nazareth, Israel), and digitalized at 20.83 kHz. Recordings were performed in a dark and anechoic booth while animals passively listened to acoustic stimuli. The sound stimulus consisted of a 40-s sequence of pseudorandom tones (“random chords”). This sequence was generated by presenting multiple tones (125-ms duration) in different sequences (12 fixed-frequency bins per octave) with each tone frequency appearing (independently of the others) with an exponentially distributed intertone interval (range 30–1,000 ms, median 250 ms). To estimate spectrograms of the acoustic stimulus, the signal was convolved with complex Morlet wavelets with central frequencies ranging from 20 to 1,600 Hz. Then, the instantaneous amplitude of each frequency was computed as the norm of the complex values. The  $z$ -scored instantaneous amplitudes were used for computing spike-triggered averages (STAs) for reverse correlation.

**Rat primary somatosensory cortex (S1) data.** As described in previous works (Lebedev et al. 2000; Panzeri et al. 2001), recordings

in the somatosensory cortex of adult Wistar rats were performed with an array of six tungsten microelectrodes. Neurons in barrel columns C1–3, D1–3, and E1–3 were recorded while their corresponding whiskers were stimulated individually. The stimulus was an up-down step function of 80- $\mu$ m amplitude and 100-ms duration delivered 48 times for each vibrissa with a 1-s interstimulus interval. Neuronal activity was amplified and band-pass filtered in the range 300–7,500 Hz. Spike waveforms were digitized at 25 kHz (Discovery; DataWave Technologies, Boulder, CO).

## RESULTS

### Outline of the WI Framework

The first step of the WI method is to convolve the spike train responses (to repeated presentations of a set of stimuli; Fig. 1A) with Haar wavelets (Fig. 1B; see MATERIALS AND METHODS). Thus each trial is decomposed into a set of wavelet coefficients representing local spike patterns at different time scales. To identify wavelets carrying meaningful information, we then compute the mutual information between each coefficient and the stimuli (Fig. 1C). Information values are compared with distributions constructed by stimulus label permutations (see MATERIALS AND METHODS). Then, the wavelet coefficients with significant information are used to represent the data. We used a decoding approach to quantify the performance of the WI method (and other methods for comparison) in extracting stimulus information. Figure 1D shows confusion matrices of naïve Bayesian decoders trained to classify responses with time patterns as in Fig. 1A with either spike counts or the selected wavelet coefficients. As expected, by construction, the performance with the WI method clearly outperformed the one obtained with spike counts.

### Performance with Simulated Data

We used simulated data to quantify the performance of the WI method and compared it with other approaches. The simulated data consisted of a set of 200-ms responses to four hypothetical stimuli created with a two-step procedure: 1) a specific spike time pattern (a sequence of predefined spike times) was assigned to each stimulus and inserted into the response of a given trial with a random shift within a window centered at their original time; and 2) background activity was generated independently for each trial following a Poisson process with a given mean rate and then added to these patterns.

The example in Fig. 2A illustrates the ability of the WI method to extract information at different time scales. A relatively precise spike timing distinguishes the first two stimuli (at the bottom; generated using a jitter of 0.5 ms), and a pattern at a coarser scale distinguishes the remaining two (at the top; generated using a jitter of 8 ms). The mean background rate was eight spikes per second. Figure 2B, right, shows the outcome of the WI method using a fine (1 ms) binning of the data. The mean decoding performance was close to perfect (0.975), thus indicating the ability of the method to capture time patterns at different scales. To understand this result further, in Fig. 2C we show the decoding outcomes obtained when considering coefficients of each wavelet scale separately. The coefficients from the coarser scales (scales 4 and 5 and last approximation) distinguished the coarse time patterns of stimuli 3 and 4, and the high-frequency coefficients (scales 1–3)

distinguished the more precise patterns of stimuli 1 and 2. For scale 1, no coefficient crossed the statistical threshold, and we therefore used the two with the largest information.

For comparison with another dimensionality reduction method, we applied the PCA-based developed by Richmond and Optican (1987). For this, we computed the principal components (PCs) from the spike trains binned with either 1- or 8-ms windows (vertical lines in Fig. 2A). As with the WI method, we estimated the time-pattern information by decoding stimuli based on the scores of the four PCs with the largest variance. As shown in Fig. 2B, neither the 1- nor 8-ms bins could capture the information at both time scales. In particular, with the 1-ms binning, the decoder could distinguish between stimulus 1 and 2, but not 3 and 4, given that the pattern of these 2 stimuli were scattered across several bins. Likewise, when using the 8-ms bins, the decoder could distinguish between stimulus 3 and 4, but this gridding was too coarse to distinguish between stimulus 1 and 2.

To test the WI method in scenarios mimicking different recording conditions, we generated three examples (Fig. 3) including patterns with different time localizations, precision, and complexity, and we systematically varied the background firing rate (2–64 Hz) and the time jitter (2–64 ms). Left panels show simulations with low background rate and jitter (4 Hz and 2 ms, respectively), whereas right panels show simulations of the same patterns but with larger baseline firing and jitter values (16 Hz and 8 ms).

We further implemented a similar PCA-based approach but selecting the PCs with the highest information. Also, we calculated performance using the whole binned responses (i.e., with no reduction of dimensionality). We then compared these time pattern strategies to a total spike count decoding, simply summing the total number of spikes of each trial. The selection of features was always performed on a set of trials used to train the decoder (training set), and then performance was evaluated in a different set of trials (test set).

Figure 4A displays results for different jitters and background rates for example 1 of Fig. 3 (using 15 trials per stimulus for training and 20 trials for testing). The results displayed are the averages of 20 simulations for each combination of parameters. As expected, there was an overall decay of performance when increasing the background firing rate and jitter due to the increasing difficulty in extracting time patterns. Still, the WI method provided the best decoding accuracy in nearly all cases except when very large jitters were used (of the order of or larger than the time patterns themselves). In this case, all information in the time patterns was destroyed, and only spike count carried information. Similar results were obtained for examples 2 and 3.

We then investigated how the number of trials used for training the decoder and for selecting the set of response features used for decoding affected performance. To do so, we repeated the analysis of Fig. 4A but systematically varied the training set size (from 5 to 65 trials per stimulus in 5-trial steps). We used 20 simulations for each method and each set of parameters (jitter and background rate). Figure 4B reports results for all 3 examples averaging across all baseline firing and jitter values of Fig. 4A. The performance of the WI, PC coefficients chosen based on information (PCinfo), and no-reduction method increased monotonically with the number of trials used for training. In contrast to the PCinfo and no-

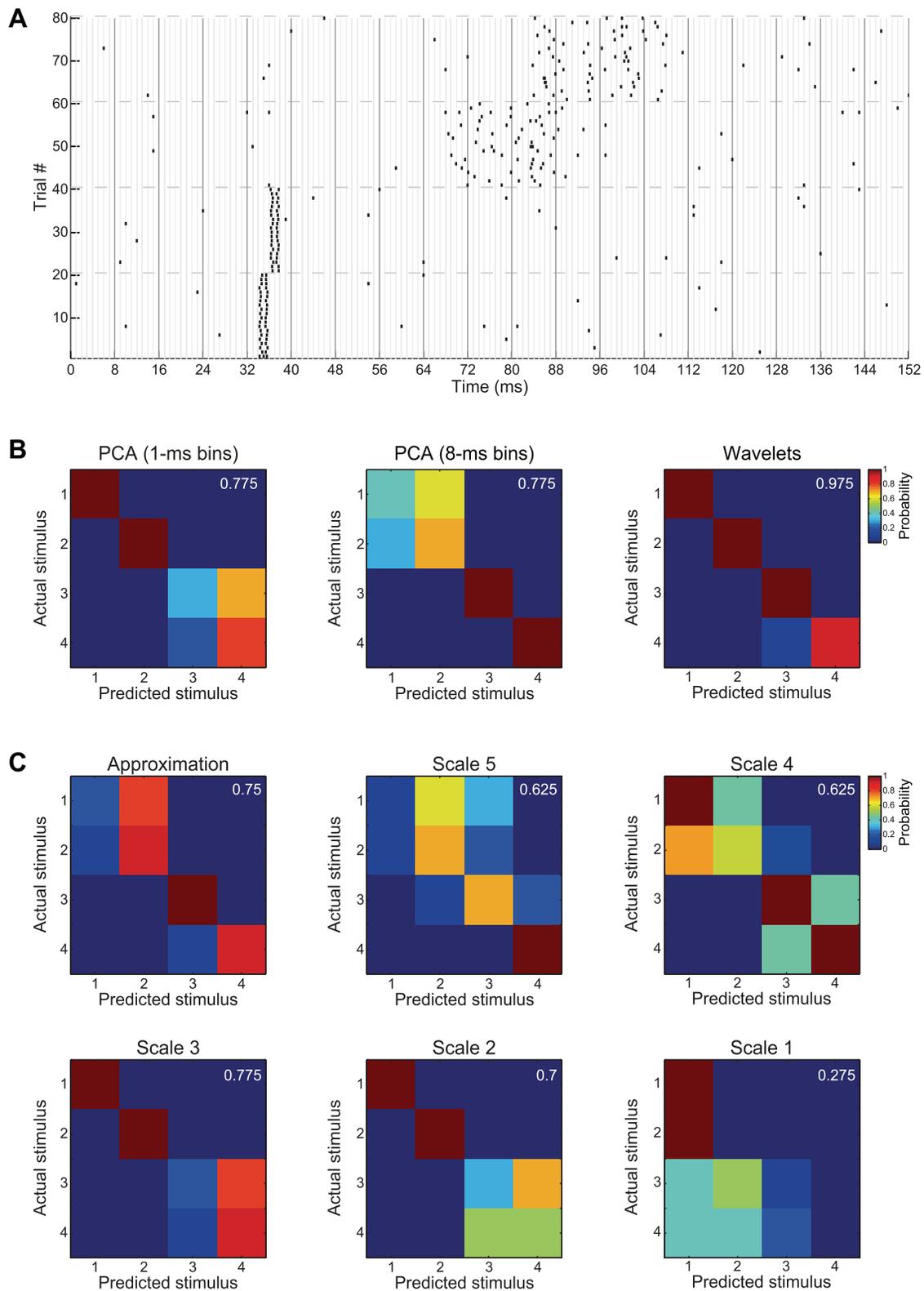
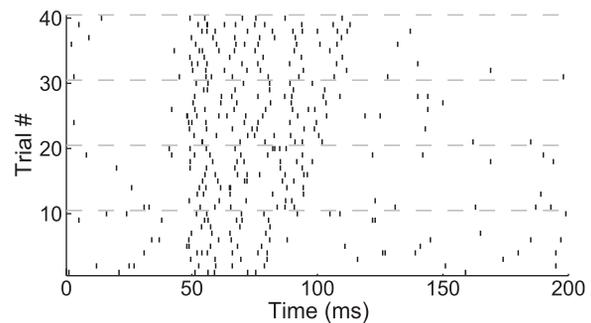
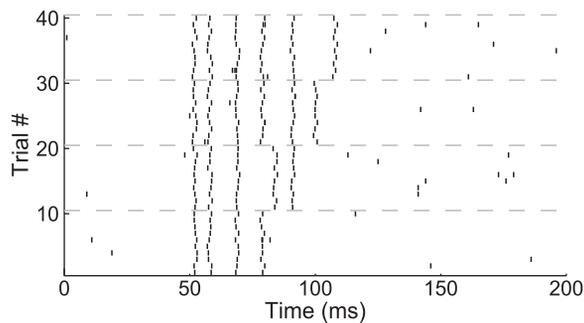
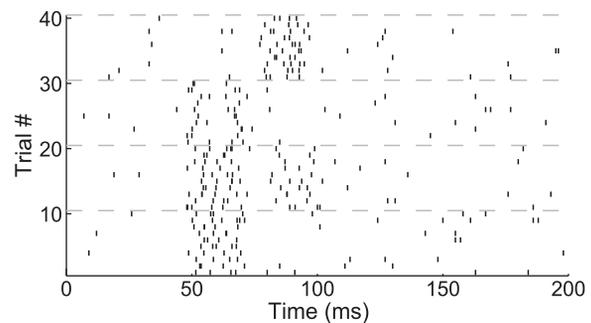
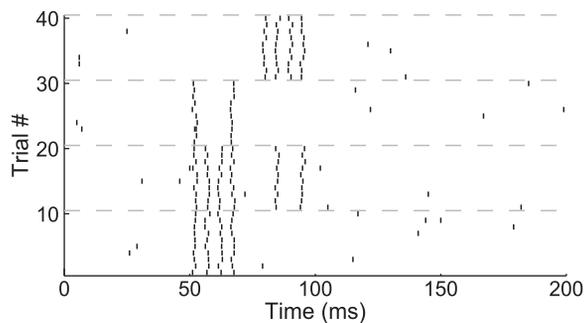


Fig. 2. Decoding time patterns at different time scales. *A*: simulated responses to 4 stimuli. Marks represent spikes at a given time (horizontal axis) in a given trial (vertical axis). Twenty trials per stimulus are displayed (dashed lines separate trials of different stimuli), and only the 1st 152 ms of each trial are displayed for clearer visualization. Vertical axes denote the size of the bins used (light gray: 1-ms; dark gray: 8-ms bins). *B*: confusion matrices of naïve Bayesian decoders trained with principal components (PCs) extracted from responses binned with 1- or 8-ms bins (*left* and *middle*, respectively) and based on informative wavelet coefficients (*right*). Decoding performances are shown in white. PCA, PC analysis. *C*: confusion matrices obtained when using wavelet coefficients from specific decomposition levels as labeled.

Example 1



Example 2



Example 3

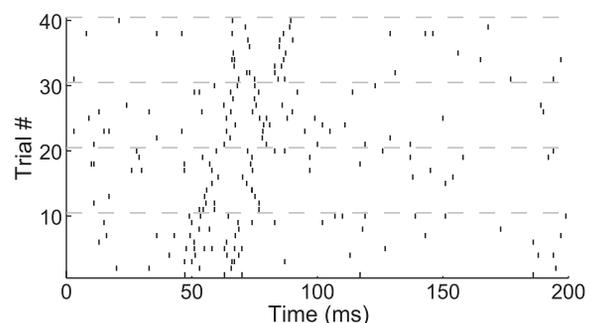
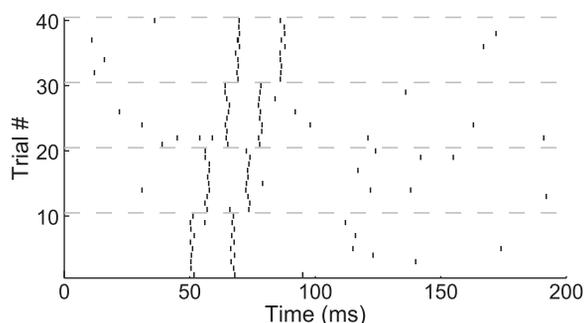


Fig. 3. Results for different examples, background firing, and jitter. The panels show 3 examples of simulated responses to 4 different stimuli. Ten trials per stimulus are displayed. The background firing was modeled as a Poisson process with mean rate of 4 (*left panels*) or 16 spikes per second (*right panels*). Stimulus-specific time patterns were added to the background firing with a 2- (*left panels*) or 8-ms (*right panels*) jitter.

reduction methods, the WI method reached a value close to its maximum performance within <20 trials, stressing its robustness to undersampling.

To evaluate the efficiency of each method in reducing the dimensionality of the responses without loss of information, we computed the performance of the naïve Bayesian classifiers as a function of the number of response features used for classification. For this, we ranked the features either by variance, in the case of PCs (PCAv<sub>ar</sub>), or by information, in the case of wavelets and PCinfo. Additionally, we performed a similar analysis by ranking the response time bins by their amount of information and then selecting only the

*n* most informative ones (referred hereafter as binned responses). Figure 5A reports the results for each of the 3 examples of Fig. 3 when using a training set of 15 trials per stimulus. Whereas the WI method needed 10 or fewer response features to reach maximal performance, all other methods needed a larger number of response features, which also varied substantially across examples. Thus the performance for these latter methods was very sensitive to the number of response features used. In sum, the WI method reduced the dimensionality of responses in a more efficient and robust way.

Interestingly, the performance using PCs with the largest variance (PCv<sub>ar</sub>) had a much steeper increase of information

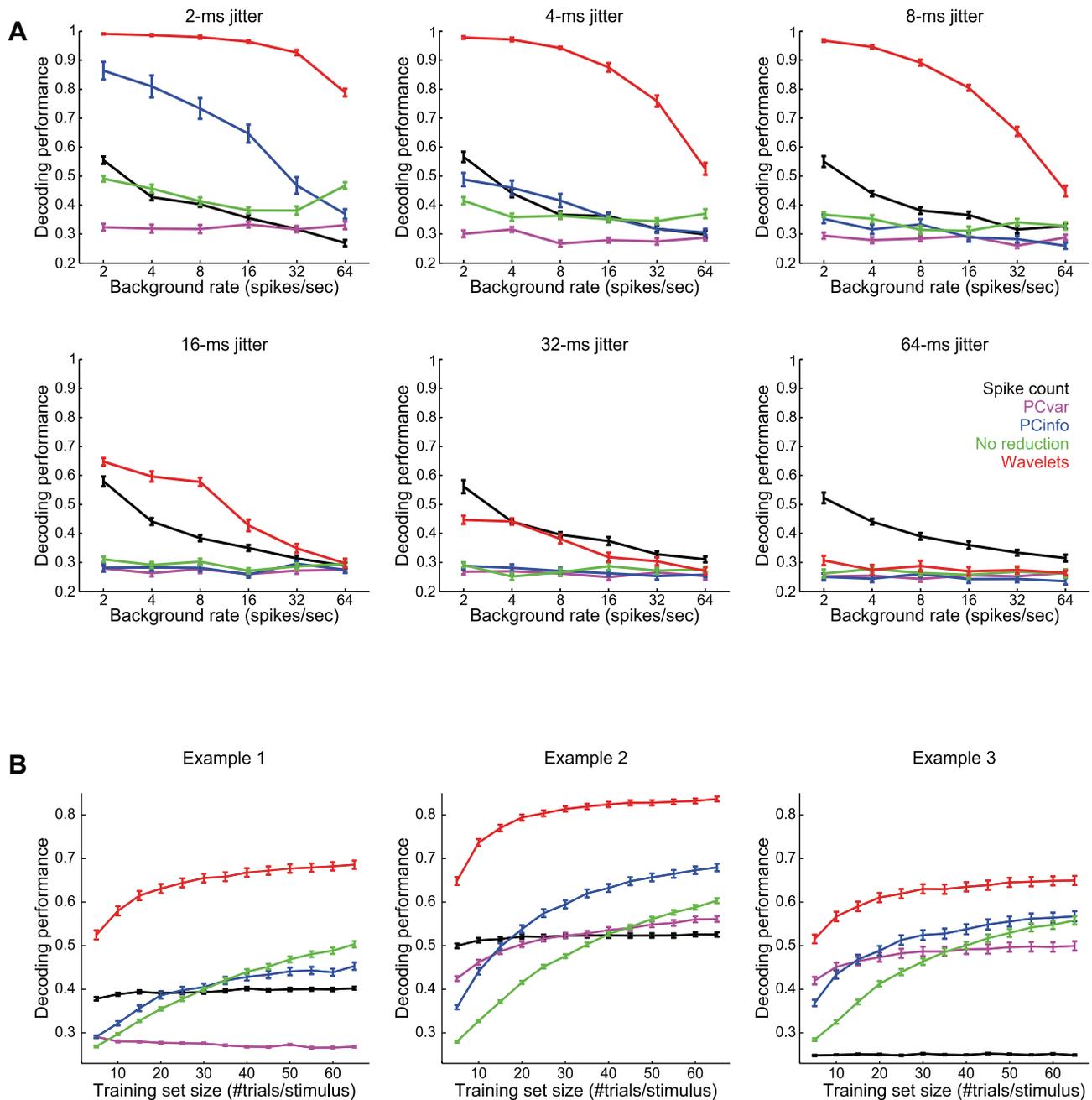


Fig. 4. Performance of the Wavelet-Information (WI) method. *A*: the different plots show the time-pattern information, quantified in terms of decoding performance using spike counts (black), PCs with the largest variance (purple), PCs with significant information (blue), no dimensionality reduction (green), and wavelets (red). Each data point represents the average of 20 simulations (error bars show SE) for *example 1* of Fig. 2 with varying jitters and firing rates as labeled. PCvar and PCinfo correspond to results with PCs with largest variance and information, respectively. *B*: relative decoding performance as a function of the training set size for the different methods and all examples of Fig. 2. Results are averages across all combinations of background firing rates and jitters. Note that in nearly all cases, the wavelet-based algorithm outperformed the other methods.

with the number of features compared with the performance obtained with PCinfo. This seemingly counterintuitive result can be attributed to the low number of trials (15 per stimulus) used for training in this case, which gave a relatively poor estimation of information carried by each PC. To verify this, we ran the same analysis as in Fig. 5A but using 50 trials per stimulus (Fig. 5B). As expected, with the larger training set, PCinfo showed a clear increase in performance and was much more efficient than PCvar for small numbers of features. Consistent with the results shown in Fig. 4B, increasing the

number of trials had little impact on the performance of the WI method, thus highlighting its robustness to sample size.

*Performance with Data from the Monkey Auditory Cortex*

Single neuron recordings were performed in primary auditory (A1) cortex in response to a 40-s-long sequence of pseudorandom tones (see MATERIALS AND METHODS for details). We divided the sequence into 500-ms time intervals and denoted each time interval as a different discrete stimulus. For this data set, we trained naïve Bayesian decoders to predict

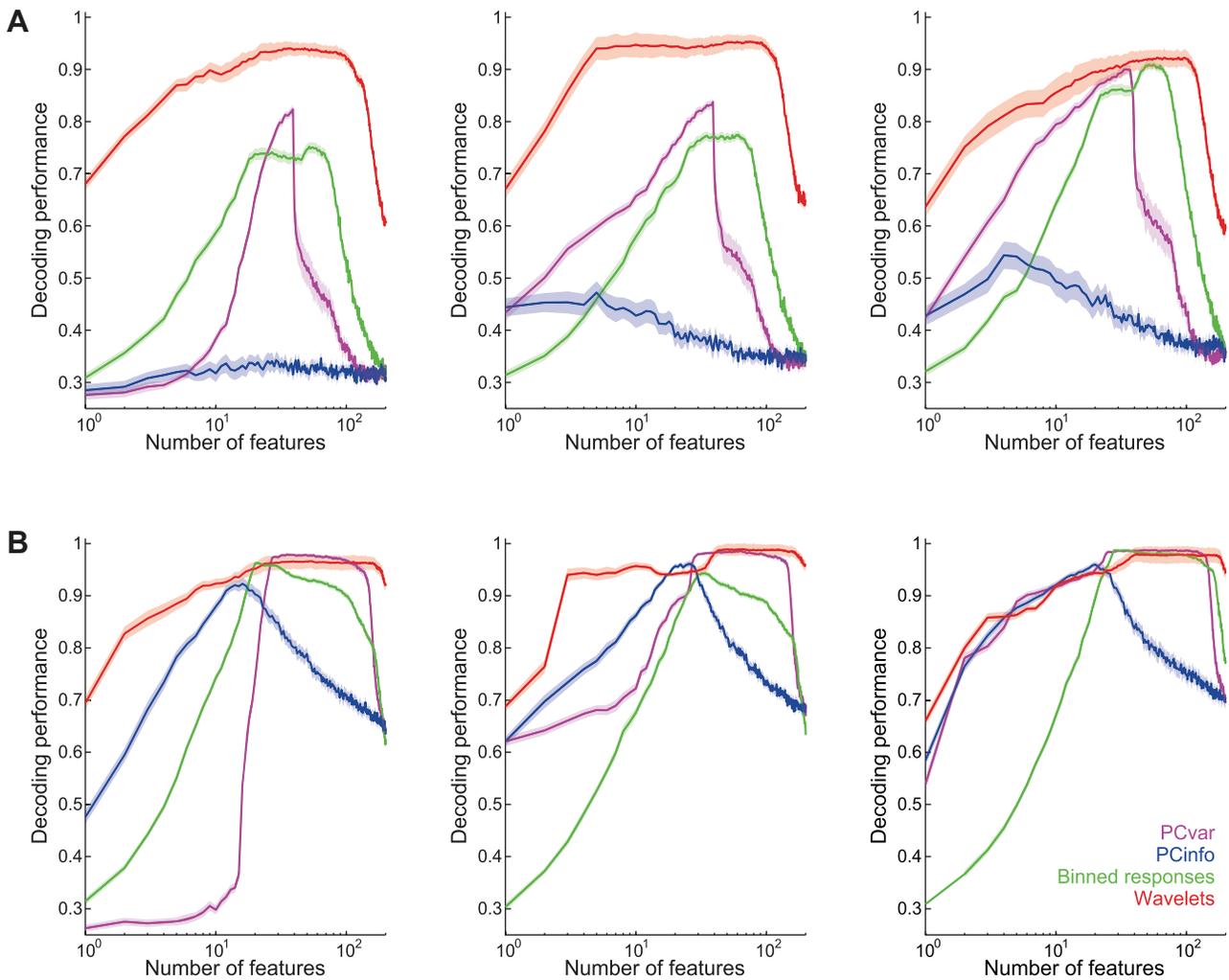


Fig. 5. Information with varying number of features. *A*: decoding performance for all 3 examples as function of number of features used for decoding (mean  $\pm$  SE; 40 simulations). For all of these simulations, we chose intermediate values of background firing rate (8-Hz) and jitter (4 ms). Features were ranked by information (wavelets, binned responses, and PCinfo) or by variance (PCvar). The training set size was of 15 simulations per stimulus. *B*: same as *A* but with a training set of 50 trials per stimulus.

which of different chunks of the time-varying stimulus was being presented. In total, 34 responsive neurons (with >1-Hz mean rate) recorded in 12 sessions were included in this analysis. Each session consisted of 50–60 presentations of the stimuli, which we separated into 2 nonoverlapping sets of training and test trials.

For each neuron, we 1st evaluated the performance of the various methods with a time resolution of 1 ms using 15 trials for training. Figure 6A shows the decoding performance of each neuron using the different approaches described above (y-axis) against the performance achieved with wavelets (x-axis). Note that the WI method outperformed the other methods for virtually all neurons.

Figure 6B shows decoding performance vs. training set size and reveals that the WI method performed significantly better than the other methods. Moreover, performance with wavelets decreased only slightly when decreasing the training set size and was close to optimal with as few as 10–15 trials. In contrast, the other methods showed a marked decrease when using few trials. Consistently, however, all methods revealed that spike timing contained more stimulus information than the total spike count.

We then quantified the impact of temporal precision used to quantify the neural responses. Figure 6C shows that decoding performance was maximal when using a bin size of about 5–10 ms (and more toward 10 ms for the PCA-based methods). This result is comparable with the optimal resolution previously reported for these data using a direct information estimate (Kayser et al. 2010). In all cases, the performance decreased as the bin size increased, meaning that larger bin sizes missed relevant information arising from the precise pattern of firing of these neurons. Moreover, the performance of all methods also decreased for bin sizes smaller than 5–10 ms. This arises because the outcome of increasing temporal precision is the trade-off of two opposing effects. On one hand, a finer resolution leads to potentially higher information content in the neural responses. Because of the data processing inequality (Quian Quiroga and Panzeri 2009), increasing resolution can only increase or leave invariant the information available in the responses. On the other hand, a finer resolution increases the dimensionality of the responses, thus making it more difficult for a decoder to extract the available information. A drop in decoding performance when increasing the resolution thus means that from that resolution onward the additional infor-

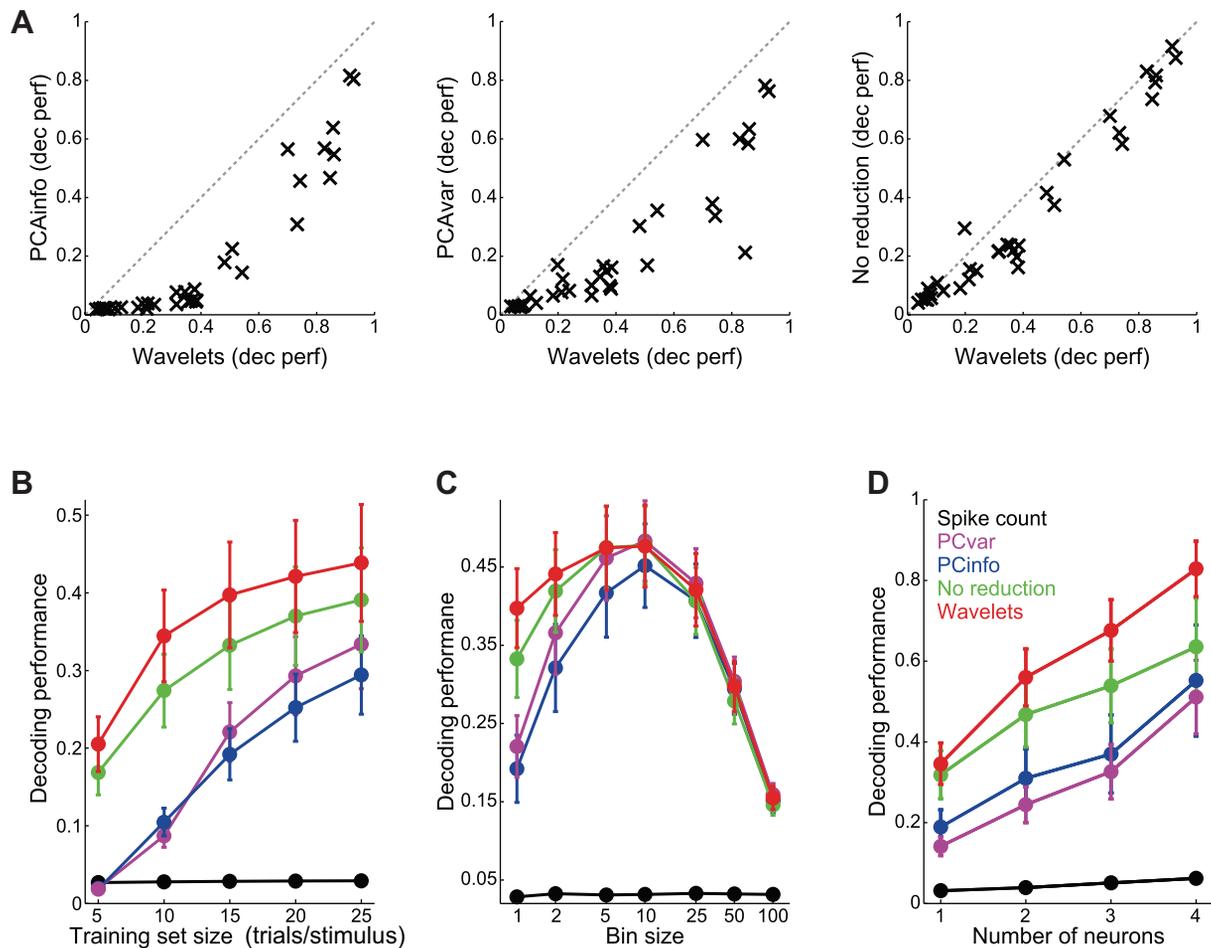


Fig. 6. Time patterns in responses from monkey A1 neurons. *A*: time-pattern information of 34 A1 neurons extracted using the PCA-based and no-reduction methods (y-axis) against the performance achieved with wavelet (x-axis). Most cases are below the diagonal (dashed lines), indicating a better performance of the wavelet-based method. The training set size was of 15 trials per stimulus. *B*: average decoding performance for different methods as a function of the training set size. Bars denote SE. The performance with wavelets was larger than with the other methods. *C*: average decoding performance for different methods as a function of the bin size (15 trials per stimulus used for training). *D*: average decoding performance for the different methods as a function of number of simultaneously recorded neurons used. Marks show averages across 10 sessions, and bars denote SE. As in *A*, 15 trials per stimulus were used for training. The performance with wavelets was significantly larger when considering >1 neuron.

information available at finer resolutions is not sufficient to overcome the added difficulty in decoding many extra and weakly informative dimensions. Thus the ability to extract more information at finer resolutions with WI compared with other methods is due to the optimal dimensionality reduction implemented in the WI method. For example, with a bin size of 1 ms (i.e., an order of magnitude increase of the dimensionality of the response space used for decoding), the performance with wavelet decreased only ~15% with respect to the 5- to 10-ms resolution, whereas decreases of ~30 and >50% were observed for the no-reduction and the PCA-based methods, respectively.

Next, we evaluated the ability of all methods to extract information from populations of simultaneously recorded cells. For this, we used the data of 10 (out of 12) sessions where 2 or more responsive neurons were recorded simultaneously. For each session, all possible combinations of a varying number of cells were used. In this case, we did not set a minimum number of wavelet coefficients or PCs for each neuron; i.e., if a neuron had no significantly informative features, then this neuron would provide no features to the decoder. Results (Fig. 6D) show that the decoding

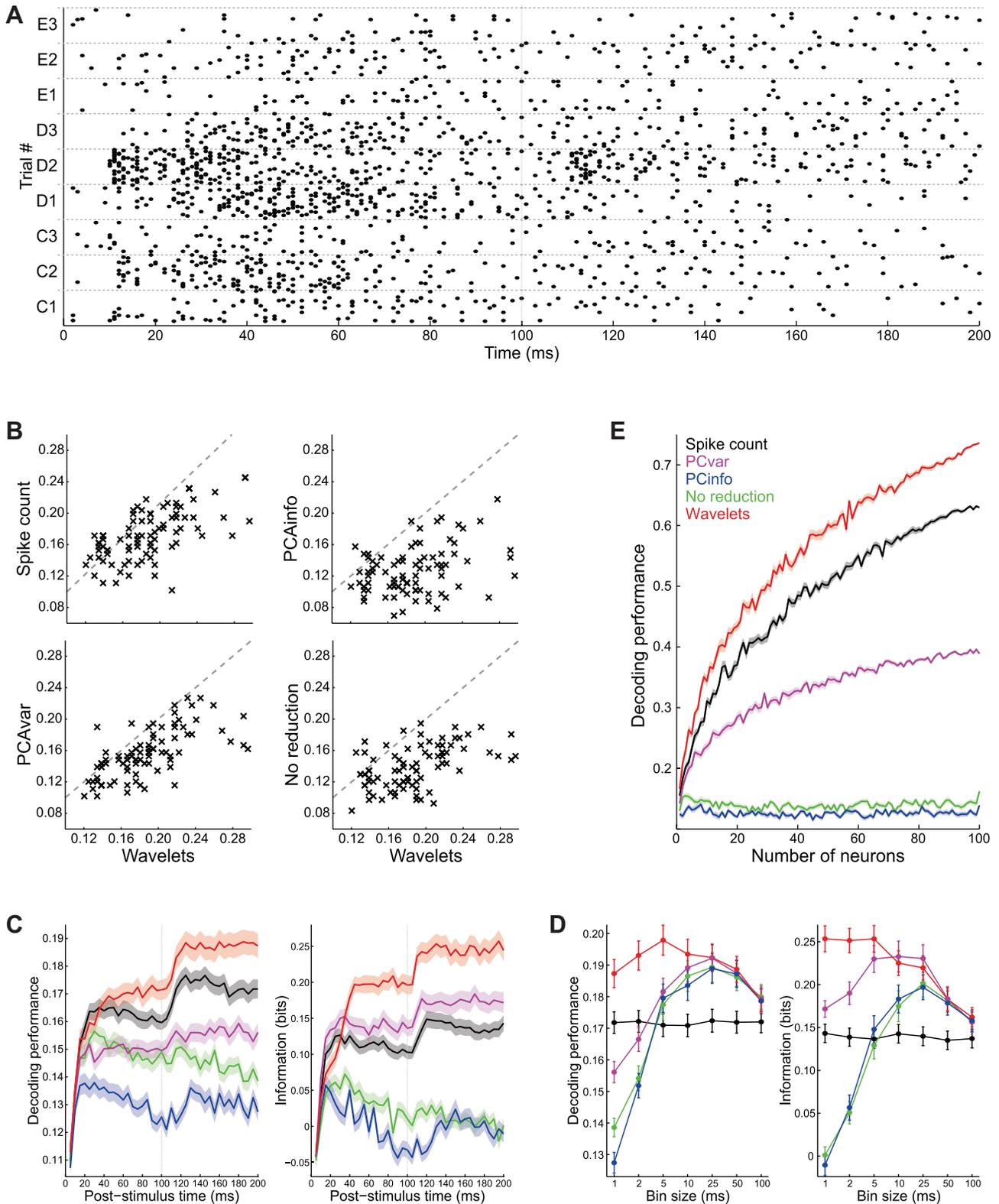
performance with wavelets was significantly larger than the one achieved with the other methods.

#### Performance with Data from the Rat Barrel Cortex

To evaluate the potential of the WI method to extract time-localized information optimally, we analyzed neuronal responses in the rat barrel cortex. In these data, a precise onset time is given by the time of whisker stimulation. We used a naïve Bayesian decoder, as before, to classify which vibrissa (from the set of C1–3, D1–3, and E1–3) was stimulated in each trial. Forty-eight trials were available for each vibrissa, and we randomly assigned half of these trials for training and the other half for testing. Figure 7A shows the responses of a representative neuron. Responses were binned in 1-ms windows, as before, and we considered a 200-ms response window starting at stimulus onset. Figure 7B compares the WI method with the other approaches for 84 responsive neurons (with >1-Hz mean firing rate). The WI method again outperformed all other methods for the large majority of the cells. Figure 7C, left, displays the decoding performance as a function of the poststimulus window (i.e.,

a window of 200 ms means taking the whole response shown in Fig. 7A). Consistent with the previous results, the WI method gave the best results for nearly all time windows considered. In contrast to the auditory data above (Fig. 6), in this case the no-reduction approach did not give a good performance because only a fraction of the bins provided

relevant information. Moreover, whereas for wavelets the performance kept increasing while increasing the poststimulus window, for the no-reduction and the PCA-based methods, the performance decayed or remained at the same level. This again indicates that these methods do not perform as well as wavelets when the dimensionality of the response



increases. We also observe that for these data the spike count decoding provided better results than both PCA-based (and the no-reduction) approaches. However, this was in part due to systematic errors in the decoder. For example, the PCAvar decoder classified D1 responses as D1, D2, or D3 (but still ruling out the possibility that other vibrissae were stimulated). To verify this, we computed the information between the predicted and the actual stimuli from the confusion matrices shown in Fig. 7C, right (see MATERIALS AND METHODS for details). In this case, PCAvar provided more information than the spike counts. As before, in nearly all cases, the wavelet-based decoding provided the best results.

Figure 7D shows both the decoding performance (left) and the information extracted from the confusion matrices (right) as a function of the bin size used. The peak of performance and information was at a bin size of  $\sim 25$  ms for all methods except wavelets. In particular, all methods showed decay in information for larger bin sizes because with larger bin sizes the information given by precise time patterns is lost. Interestingly, however, all methods except wavelets showed also decay in performance (and information) for bin sizes  $< 25$  ms. As with the monkey data, this is due to the increase in the response space dimensionality accompanying the increase in resolution. With these same data, a previous work (Panzeri et al. 2001) reported an optimal bin size  $< 25$  ms using a method analogous to the no-reduction shown here, i.e., calculating the mutual information from the binned responses. However, in that case, a high-dimensional space was avoided by considering a much smaller response window of 20 ms. In this regard, the advantage of wavelets is crucial whenever the optimal response window is not known a priori.

Finally, we studied the information carried by populations of neurons. For this, we assumed that all neurons were recorded simultaneously (an approach that does not take into consideration the effects of correlations) and repeated the procedure used in Fig. 6C. Since 100 neurons were available, we averaged across 30 randomly chosen combinations for each number of neurons. Additionally, to avoid an excessive number of features, we only allowed a maximum of 5 wavelet coefficients or PCs per neuron (instead of 25 as before). Results are shown in Fig. 7E. The performance achieved with wavelets surpassed the 1 achieved with the other methods. Except for the no-reduction and PCinfo cases (due to the above-mentioned limitation in dealing with high-dimensional response spaces), the performance increased monotonically with the number of neurons.

### Comparison with the MS Method

Next, we compared the WI method with the MS approach for different  $q$ -values (see MATERIALS AND METHODS) with both

the simulated and real data reported above. Figure 8A shows the results of such comparison with the same 20 simulations illustrated in Fig. 2A using different background firing rates. Note that WI is more robust to increases of noise levels for this example. Figure 8B shows the results for the simulations shown in Fig. 4A (to save space, results with 32- and 64-ms jitter are not shown since in these cases the performance with all methods was close to chance), and Fig. 8B the result for the 3 examples presented in Fig. 4B. Altogether, we observe an overall better performance with the WI method proposed here. Note also that the performance of the MS method is dependent on the choice of the parameter  $q$ . For instance, a  $q$ -value of 0.128 gave the best performance with a jitter of 16 ms but also the worst performance with a jitter of 2 ms. Conversely, a  $q$ -value of 0.362 gave the best performance for the 2-ms jitter case but a relatively poor performance for larger jitters. In this respect, the advantage of the WI method is that it does not require the tuning of any parameter and automatically gives a performance that in most cases surpassed the one obtained with the MS method even when choosing the optimal  $q$ -value.

Figure 8C displays results as function of training set size (as in Fig. 4B). For these examples, a  $q$ -value of 0.128 gave the best overall results for the MS method. These results were, however, not as good as the ones obtained with the WI method. We note that the MS method was remarkably robust to under-sampling as can be seen in the results of *example 2* where it outperformed wavelets when  $< 10$  trials per stimulus was used for training, likely due to the fact that this low number of trials was not sufficient for a good selection of wavelet coefficients.

Results of the comparison between the WI and MS method for the real data are shown in Fig. 9. In particular, Fig. 9A (left panels) displays the results obtained for the A1 monkey neurons (as in Fig. 6) using four representative  $q$ -values for the MS method. For this data set, the average performance obtained with the WI method was significantly better than that obtained with the MS method for all  $q$ -values (paired  $t$ -test,  $P < 0.05$  in all cases; see Fig. 9A, right). Figure 9B shows the results obtained for the barrel cortex neurons. As with the monkey data, performance with the WI method was significantly better than that obtained with the MS approach for all  $q$ -values (paired  $t$ -test,  $P < 10^{-23}$  in all cases; see Fig. 9B, right). The relatively poor performance of the MS method for the rat data set is likely due to the compact time localization of the informative spikes.

### Denoising Spike Patterns

We next investigated whether the WI method could be used to denoise single-trial spike trains. Denoising entails, in brief, identifying spikes correlated with the informative wavelet

Fig. 7. Time patterns in rat barrel cortex neurons. *A*: responses of a representative neuron to mechanical stimulation of different whiskers (C1–3, D1–3, and E1–3). Each dot denotes a spike in a given trial (vertical axis) at a given time (horizontal axis). Forty-eight trials were recorded for each whisker. Stimuli were delivered at 0 ms and lasted 100 ms, which corresponded to the OFF response. *B*: decoding performance for 84 responsive neurons with wavelets ( $x$ -axis) compared with spike count, PCAinfo, PCAvar, and no dimensionality reduction, respectively. Most entries are below the diagonal (dashed line), thus indicating a better performance with wavelets. *C*, left: mean decoding performance for different methods as a function of the poststimulus time window. Right: mean information for the different methods estimated from the confusion matrices (see main text for details). Solid lines denote averages, and shaded areas SE. *D*, left: decoding performance for different methods as function of bin size. Right: stimulus information for different methods estimated from confusion matrices as function of bin size. *E*: mean decoding performance for the different methods as a function of number of neurons used. Solid lines show averages across a set of 30 randomly chosen neurons (except for 100 neurons, where only 1 combination is possible), and shaded areas denote SE (barely visible due to the very small error). As before, note the better performance of wavelets for nearly all poststimulus time windows and any number of neurons considered.

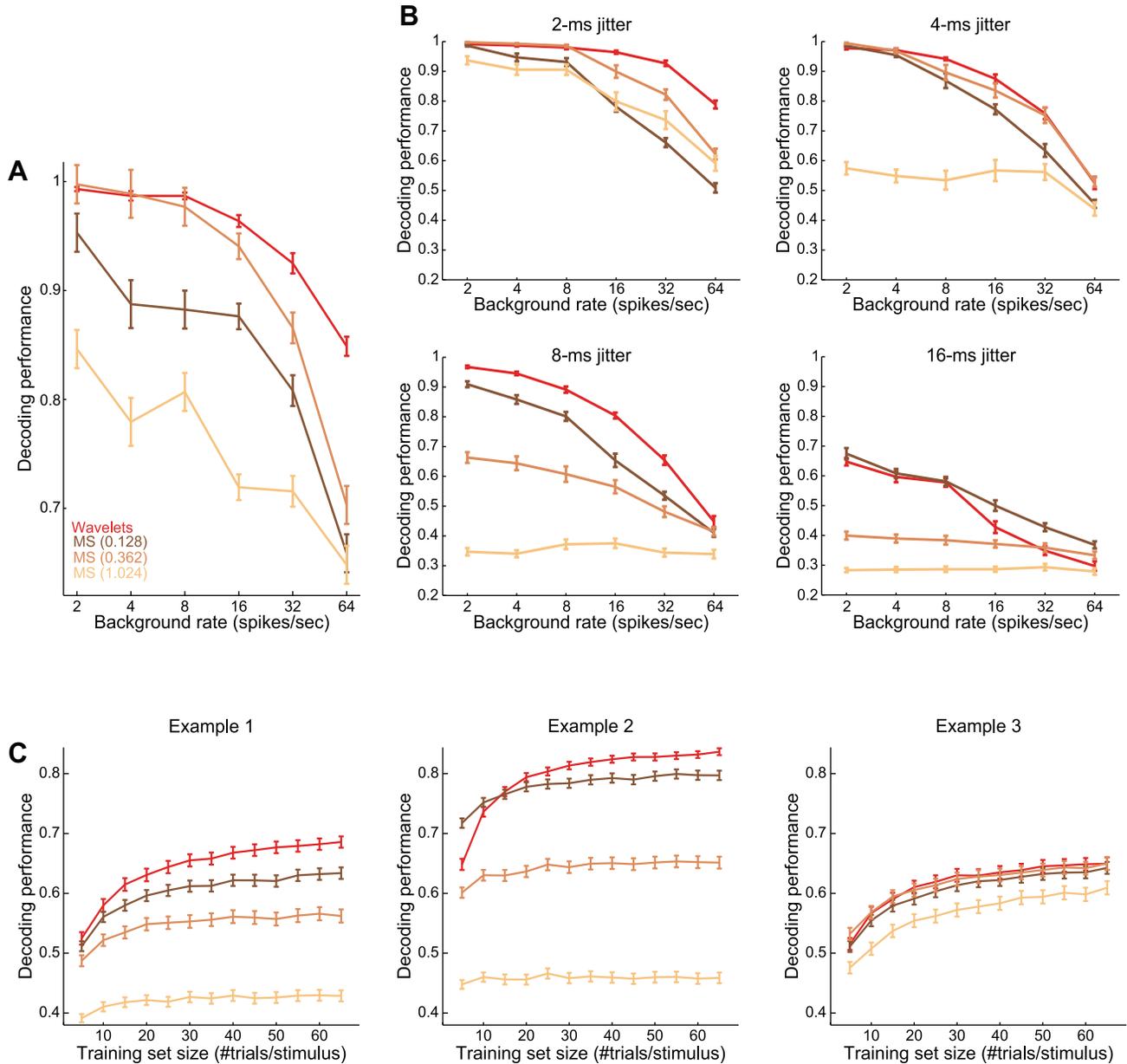


Fig. 8. Comparison of the WI and metric-space (MS) methods with simulated data. *A*: performance of WI and MS (for 3 different values of  $q$  as labeled) methods for multiscale responses of Fig. 2A with varying background rates. Shown are mean and SE across 20 simulations for each noise level. *B*: performance of WI and MS for the same examples in Fig. 4A. For space reasons, we do not show results with 32- and 64-ms jitters (performance was close to chance levels in all of these cases). *C*: same as Fig. 4B comparing the WI and MS approaches.

coefficients and discarding the remaining, noninformative ones (see MATERIALS AND METHODS). Figure 10A shows the denoising of the time pattern of *example 3* (Fig. 3) with a jitter of 4 ms and a background firing rate of 64 Hz. We observe that a large amount of the background noninformative spikes were removed, and the remaining spike rasters after denoising (Fig. 10A, *bottom left*) were very similar to the spike patterns embedded in the data (see Fig. 3). A similar reduction of background “noisy” activity is evident when comparing the PSTHs before and after denoising (Fig. 10A, *right panels*).

To quantify these observations, we repeated this procedure varying systematically the jitter and background firing rate (using 15 trials per stimulus). Figure 10B shows the number of errors obtained by the denoising procedure for different jitters as a function of background rate. Here, we defined errors as the

sum of false positives (not deleting a spike corresponding to background activity) and false negatives (incorrectly discarding a spike that belonged to the informative time pattern). For comparison, we also calculated the number of errors obtained when thresholding the original PSTHs but without a prior wavelet denoising. This was done to assess whether denoising could be achieved by a simple PSTH thresholding. Furthermore, to show that results were not just due to the smaller number of spikes obtained after denoising, we also calculated the number of errors obtained when randomly erasing the same number of spikes. In general, the wavelet-based denoising approach gave the lowest number of errors. Results obtained for the other examples of Fig. 3 were similar.

Next, we applied this denoising approach to the spikes of a representative neuron taken from the auditory data set. Figure 11A

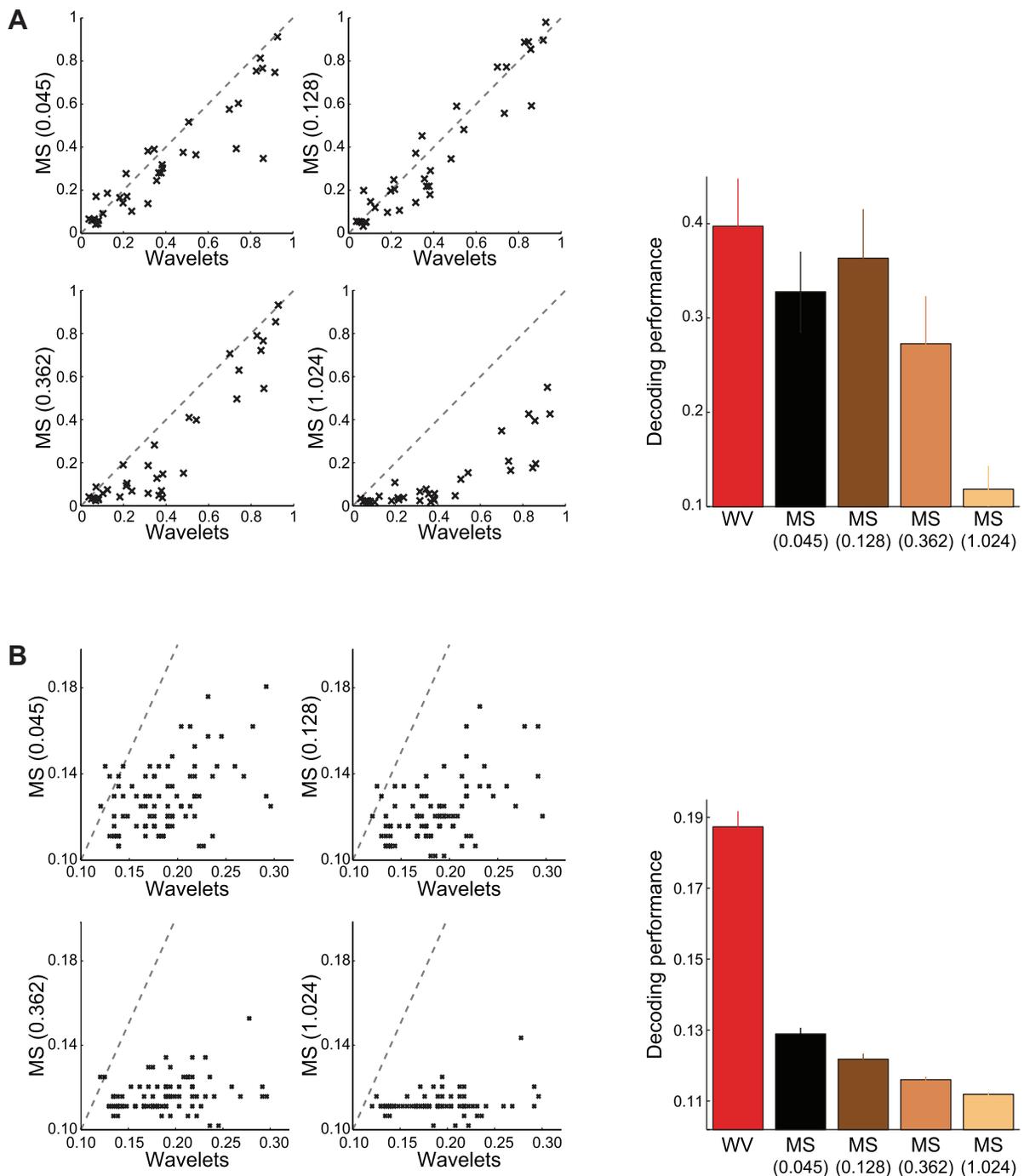


Fig. 9. Comparison of the MI and MS methods with real data. *A, left*: same as in Fig. 6A for the WI and the MS method with 4 representative  $q$ -values. On average, decoding performance achieved with wavelets (WV) was  $26 \pm 8$ ,  $24 \pm 8$ ,  $144 \pm 35$ , and  $449 \pm 73\%$  (means  $\pm$  SE) better than the 1 obtained with the MS method with  $q$ -values 0.045, 0.128, 0.362, and 1.024, respectively. *Right*: decoding performances for WI and MS methods (for the same  $q$ -values). Shown are means and SE. Wavelets gave better results for all  $q$ -values ( $n = 34$ ; paired  $t$ -test,  $P < 0.05$ ). *B, left*: comparison of WI and MS for the same data as in Fig. 7B. For this data set, decoding performance achieved with wavelets was  $46 \pm 5$ ,  $54 \pm 5$ ,  $61 \pm 5$ , and  $68 \pm 6\%$  (means  $\pm$  SE) superior to the 1 obtained with the MS method (same  $q$ -values as in A). *Right*: comparison of WI and MS performance for the rat data. Wavelets gave better results for all  $q$ -values ( $n = 84$ ; paired  $t$ -test,  $P < 10^{-23}$ ).

shows the raster plots before (*top*) and after (*bottom*) denoising, where it is clear that time patterns (spikes consistent across trials for each stimulus) are better visualized after denoising. Given this encouraging result, we then examined whether wavelet denoising could lead to obtaining cleaner and sharper STAs of the stimulus. STAs are commonly used representations to assess which stimulus features (out of the many in a

complex dynamic stimulus) drive the responses of the neurons (Dayan and Abbot 2005). For this, we used a time-frequency representation of the acoustic stimulus (see MATERIALS AND METHODS) and computed the STA in the form of the spectrotemporal receptive field, i.e., the average frequency spectrum of the stimulus around the time of spiking. To test the effectiveness of denoising in removing the detrimental effect of

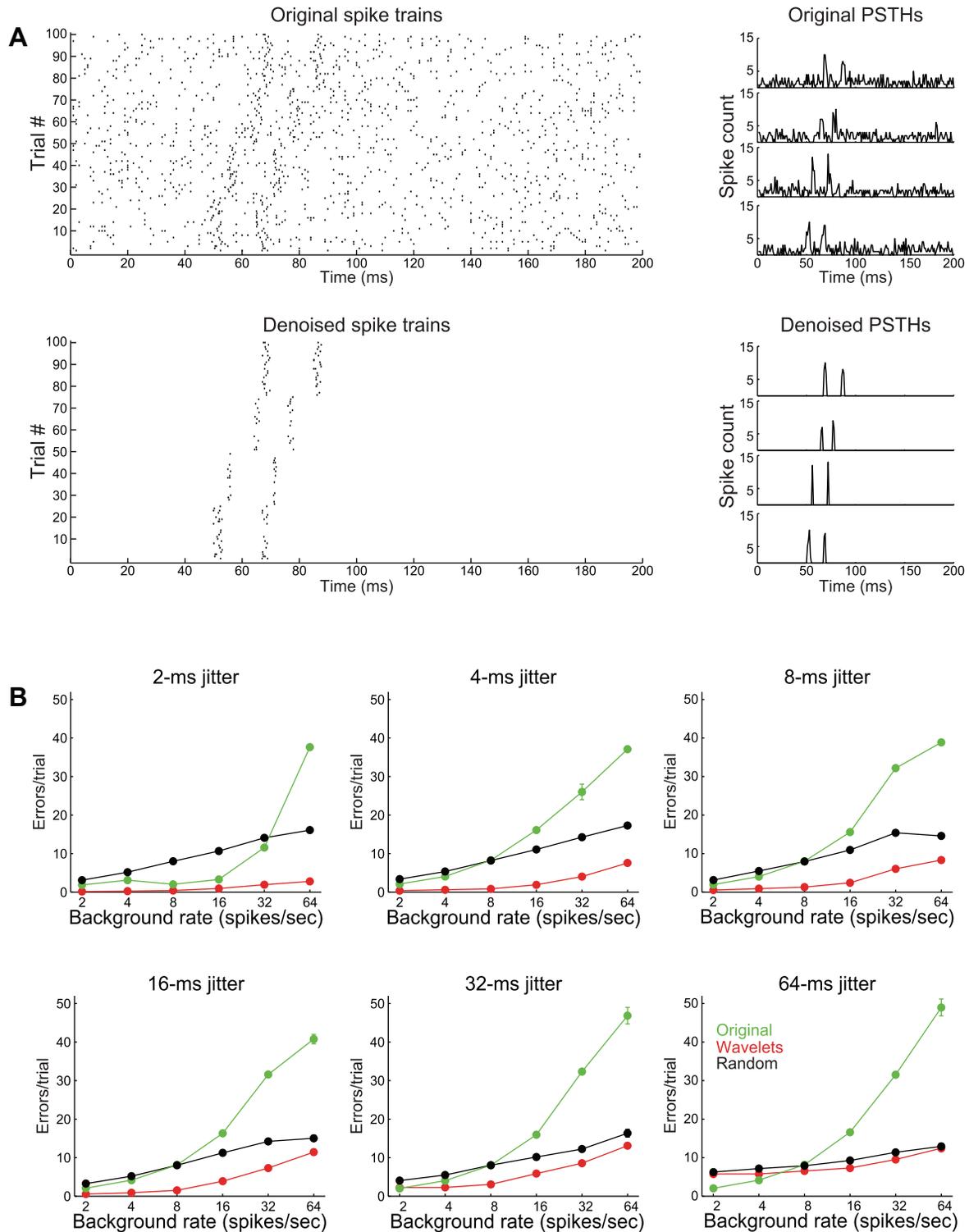


Fig. 10. Denoising time patterns with simulated data. *A, left*: a 100-trial realization (25 trials per stimulus) of *example 3* from Fig. 2 before (*top*) and after (*bottom*) denoising. *Right*: peristimulus time histograms (PSTHs) for each of the 4 stimuli before (*top*) and after (*bottom*) denoising. *B*: number of errors (see main text) after denoising (in red) obtained for *example 3* with different jitters (same values used in Fig. 3) as function of background firing rate. Results are averages across 40 simulations (error bars denote SE but in most cases are too small to be noticeable). For comparison, results obtained when thresholding the original PSTHs (without prior wavelet denoising; see main text for details) are shown in green. Results obtained by randomly erasing the same number of spikes as with wavelet denoising are displayed in black. In most cases, the lowest number of errors was obtained after wavelet denoising.

nonstimulus-driven spikes, we first computed the STA for the original response, then added background Poisson noise with a mean firing rate of 100 Hz, and computed the STA in this noisy condition. Finally, we denoised the data with the Poisson noise

and recomputed the STA. Results are shown in Fig. 11*B*. The original STA showed a clear tuning to stimuli with energy in the 200- to 300-Hz frequency range, ~50 ms preceding the spike occurrences. This stimulus selectivity was dramatically

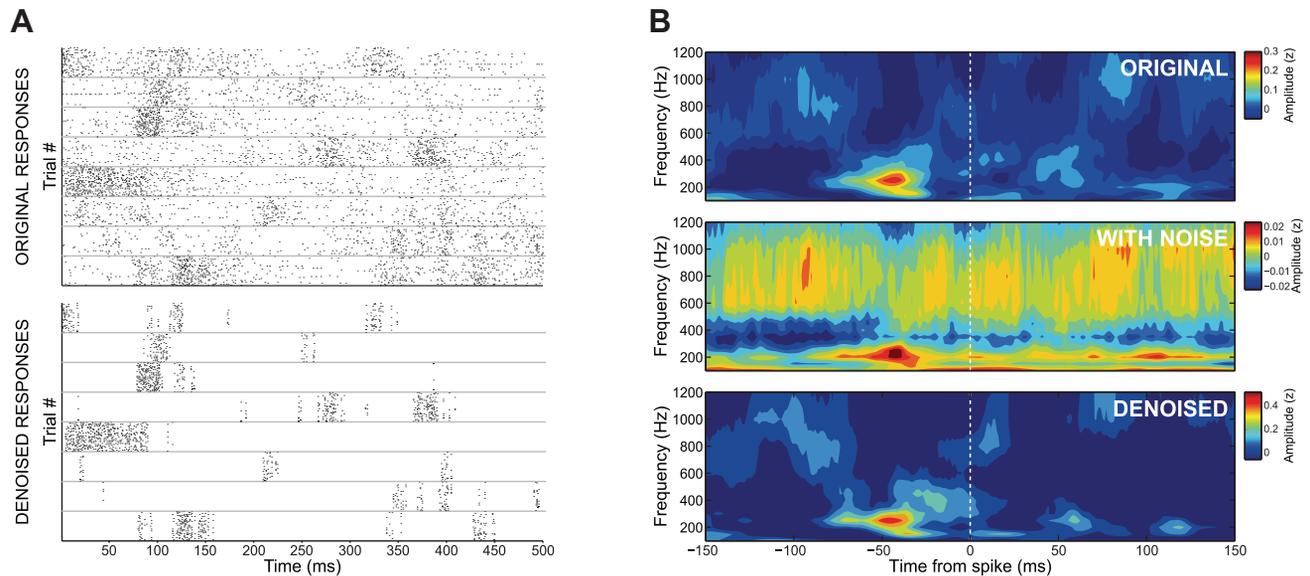


Fig. 11. Denoising time patterns with monkey A1 responses. *A*: responses to 8 stimuli of a representative A1 neuron before (*top*) and after (*bottom*) denoising. Gray lines separate trials corresponding to the different stimuli. *B*: spike-triggered averages of the time-frequency representation of the acoustic stimuli computed from the original spike trains (*top*), when adding a Poisson noise with 100-Hz mean firing rate (*middle*), and after denoising the spike train with the Poisson noise (*bottom*). Note that adding the Poisson noise dilutes the original time-frequency representation, which is recovered after denoising.

diluted after adding the background noise but was again recovered after denoising.

#### Information in the Synchronous Firing of Neurons

We next investigated whether the WI method could be extended to quantify the information conveyed by the synchronous firing of pairs of neurons by assuming that in this case the informative wavelet coefficients for both neurons should covary across trials (see MATERIALS AND METHODS). For comparison, we also evaluated the information carried by: 1) coincident spikes, i.e., spikes from different neurons fired within a short time window (Grun et al. 2002; although it should be noted that, in contrast to WI, with this method, a single, unique time scale that defines “coincidence” has to be defined a priori); and 2) an implementation of the MS approach calculated by computing the distance between the spike trains for the pair of neurons and then using a decoder similar to the one used for wavelets (using the MS distance instead of the wavelet distance). For this analysis, we used the three  $q$ -values shown in Fig. 8.

Performance was tested by simulating the activity of a pair of neurons with a correlated firing during a “synchrony window” (200–360 ms in the case of the example displayed in Fig. 12A, *left*) during the presentation of one of two stimuli (*stimulus 1*). In the rest of the response window, we added spikes generated independently following a Poisson distribution with a mean firing probability equal to the one of the synchronous spike pattern (40 Hz). Therefore, by construction, all of the information about the stimuli was only given by the transient synchronization.

We ran several simulations like the one illustrated in Fig. 12A, *left*, varying systematically the duration of the synchrony window and adding time jitters in the coactivation. We used 20 trials per stimulus for training the decoders and 20 for testing. Figure 12B shows the decoding performances for different jitters as function of the length of the synchrony duration. Clearly, the WI method outperformed the coincidence count

for nearly all window sizes and jitters. The overall decoding performance for all methods is shown in Fig. 12C.

We then extended the spike train denoising procedure described above to visualize better the informative patterns of synchronous firing. For this, we used the distances of the selected wavelets. The mask for a given stimulus (analogous to the denoised PSTHs in the previous case) was constructed by averaging the distances of each wavelet coefficient across trials and then adding together the mean distance of each coefficient multiplied by its time support (with a value of 1 within the time range spanned by the wavelet and 0 elsewhere). Figure 12A, *right* panels, displays the raster plots of the neurons shown on the *left* panels after denoising where it is clear that the coincident activations are highlighted and noninformative spikes deleted.

#### DISCUSSION

A key problem in assessing the contribution of precise time patterns to sensory coding is the high-dimensionality of the data sets. Although several statistical methods have been developed to correct for sampling biases arising in these cases (Montemurro et al. 2007; Nemenman et al. 2004; Paninski 2003; Panzeri et al. 2007; Panzeri and Treves 1996; Strong et al. 1998), these methods are still of limited value when dealing with long response spaces with high temporal precision. Other approaches to tackle this issue have been proposed, for example, based on simplifying the structure of interactions among possible information-carrying symbols with minimum information loss (Ganmor et al. 2011; Panzeri and Schultz 2001; Shew et al. 2011), developing binless estimations (Victor 2002) or defining spike train distances to quantify information (Victor and Purpura 1996).

In a classic work, Richmond and Optican (1987) proposed to reduce the dimensionality of the response space by using PCA. However, PCA offers no time resolution, thus being limited for characterizing time-resolved patterns, and relies on identifying directions of maximum variance of the data, which may or may

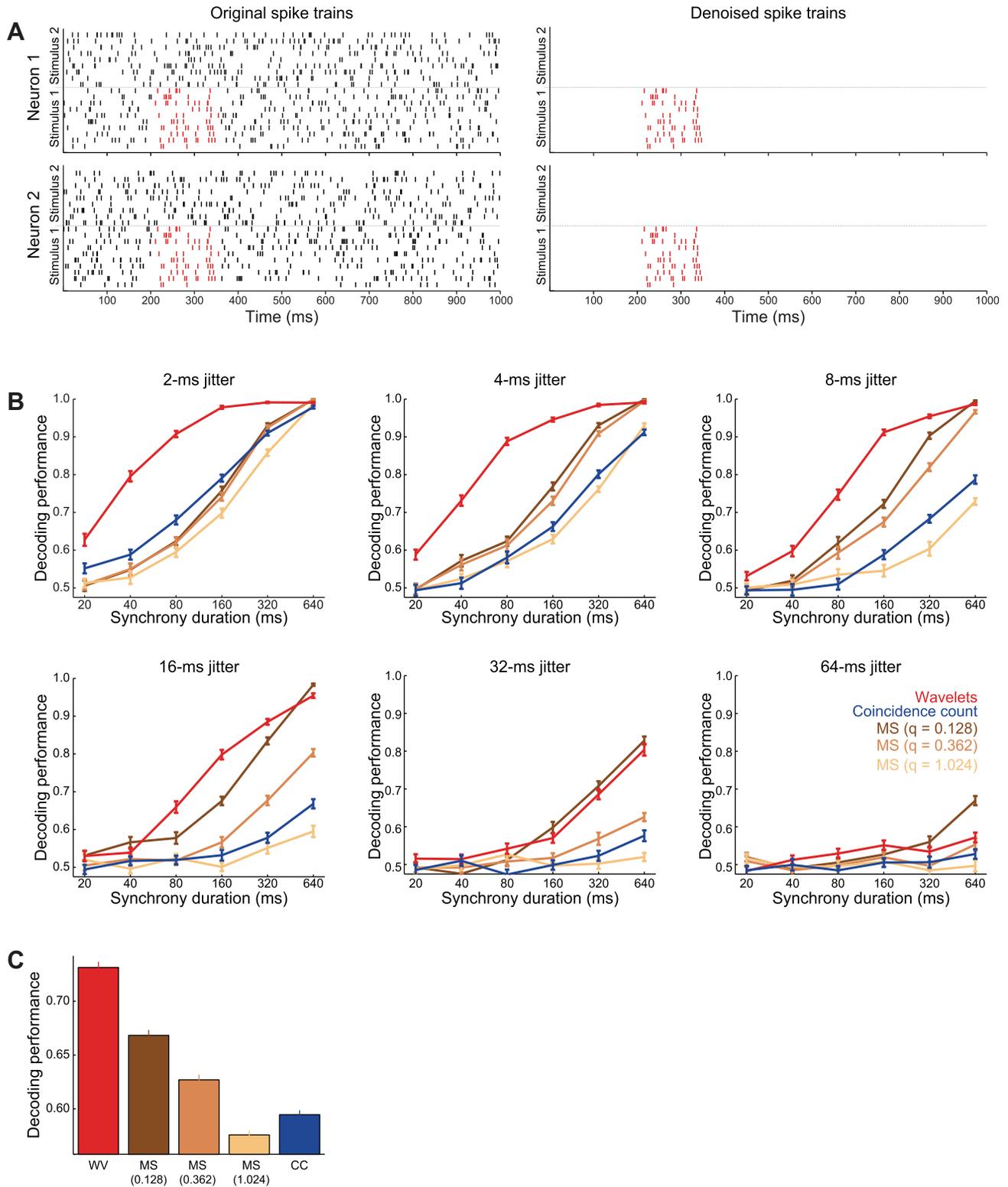


Fig. 12. Extracting information from spike synchrony. *A, left*: responses of 2 simulated neurons to 2 stimuli. Spikes between 200 and 360 ms were coincident in trials corresponding to *stimulus 1* (in red). This transitory synchrony was the only source of stimulus information in the responses. *Right*: denoised spike trains. *B*: performance obtained with the wavelet-based method with the MS approach and with the number of spike coincidences for different spike jitters and synchrony durations. *C*: average decoding performance for all methods across all jitter and background levels.

not match the dimensions with relevant information. Here, following the general idea of dealing with high-dimensional spaces by implementing a dimensionality reduction that captures relevant information, we proposed a new computational

approach to assess information carried by time patterns in single and multiple neurons. This approach is based on: 1) extracting features of the spike trains with the wavelet transform; 2) a dimensionality reduction by which a subset of

wavelet coefficients are selected using information theory; and 3) a quantification of time-pattern information by using a decoding approach.

With both simulated and real data, we demonstrated a robust performance of the WI method in capturing meaningful information in the spike trains without committing to specific assumptions about the time scales at which information is encoded and even capturing information in localized patterns at multiple time scales. This feature is of utmost importance considering that recent studies have shown that neural responses carry complementary information at a number of different time scales, ranging from millisecond-precise spike patterns to slow rate variations or slow network oscillations on the scale of hundreds of milliseconds (Bullock 1997; Fairhall et al. 2001; Lisman 2005; Nadasy 2009; Panzeri et al. 2010; Victor 2000). However, it has been challenging to characterize how these neural responses work together to represent information because most spike train analysis methods are committed from the beginning by the specific choice of an optimal time scale.

The performance of the WI method was minimally affected by increases in the dimensionality of the responses (obtained by increasing the time resolution, the length of the response considered, or the number of neurons) compared with other methods. In fact, compared with methods like PCA, using the whole (binned) response space (i.e., without any dimensionality reduction) or using spike counts, the WI method was able to extract more information from the spike trains, as quantified by decoding performance. In addition, the information obtained with wavelets was more robust to varying degrees of background activity and jitter in the precise timing of the spikes. For the cortical data analyzed here, high information values were found with PCA but only when the optimal time scale (i.e., response length and resolution bin) was considered. Crucially, information values obtained with wavelets were much less sensitive to the choice of the time scale used to study the neural responses. In our view, these advantages arise because the convolution with Haar functions implemented with the DWT identifies local contrasts at different time locations and at different time scales. In other words, the wavelet transform offers a time-resolved, multiscale representation that automatically and efficiently represents time patterns of different lengths and resolutions, appearing at different times. The advantages of wavelets for spike train analysis reported here are in line with the reported advantages of wavelets for spike sorting (Quiroga 2012; Quiroga et al. 2004) and for denoising evoked potentials (Ahmadi and Quiroga 2013).

We also compared the WI method with the MS method (Victor and Purpura 1996), a widely used metric-based approach to estimate information in spike trains. We found that overall the WI method performed better and more robustly than the MS method. In particular, MS results varied substantially depending on the choice of  $q$  (defining the weight given to precise timing vs. number of spikes), and no  $q$ -value gave good results in all conditions, i.e., for different firing rates and jitters. Therefore, it may not be possible to find a single  $q$ -value that is suitable for different neurons in a data set (with different firing rates, time pattern resolution, degrees of information, etc.). On the contrary, WI is parameter-free and performed well in all tested cases. With the real data, the performance of the MS method was lower than WI, likely because of the compact

time localization of the informative spikes, something that is captured by the WI method when doing the selection of informative wavelet coefficients. Another aspect of practical importance is that the MS method took significantly longer to compute than WI. For the monkey data, whereas the WI method required consistently between 100 and 150 s to compute the results for each neuron, the MS method processing time was strongly correlated with the mean rate of the responses: computations for neurons firing  $\sim 2.5$  Hz on average took  $\sim 200$  s or more per  $q$ -value, whereas computations for neurons firing  $\sim 5$  Hz took  $\sim 400$  s. Since, as in previous works (Roussin et al. 2012; Victor and Purpura 1996), results were calculated for dozens of different  $q$ -values, computations with the MS method took about two orders of magnitude longer compared with wavelets.

A key strength of the WI approach is its data robustness. In particular, the WI method required fewer trials to achieve optimal results and needed fewer features to represent the relevant information in the spike trains compared with other methods. This efficient compression of the responses mitigates the curse of dimensionality and allows the analysis of larger responses, the use of higher resolutions, and the possibility of population coding analyses where features of several neurons are considered together. This represents a significant advance compared with the other dimensionality reduction-based approaches examined in our study, which in some cases tended to underestimate the time resolution or the amount of information in the spike trains because actual increases in information were counterbalanced by the limitations of these methods to deal with higher-dimensional responses.

It is of interest to discuss briefly how the data robustness of the WI method (due to its highly efficient reduction of dimensionality) relates to the biases in extracting information from neural responses often discussed in the literature (Panzeri et al. 2007). First, the success of the WI method in effectively compressing the responses to a very small number of informative dimensions leads to a strong reduction in the downward bias in decoding performance, which is given by a limited number of data for training the decoder (Jacobs et al. 2009; Quiroga and Panzeri 2009). This is demonstrated by the success of the WI method to extract high information values even for very fine temporal resolutions in real data and the nearly optimal performance obtained with relatively few trials with the simulated data. As a rule of thumb, 10 or more trials per stimulus were found to be enough to avoid a major downward bias problem (see Fig. 4B). Second, the higher information values achieved with the WI method allow reducing (and correcting more efficiently) the upward bias in estimation of information from the confusion matrix due to the limited number of experimentally available test data. As shown in Panzeri and Treves (1996), this upward bias is roughly proportional to the number of different stimuli that are predicted by the decoder when a given stimulus is presented. As a rule of thumb, the corrections for the upward bias in the information calculation from the confusion matrix work well if the number of different stimuli that are predicted by the decoder is two to four times smaller than the number of trials per stimulus (Panzeri et al. 2007). The WI method gives less decoding errors and therefore a smaller upward bias in confusion matrix information calculations.

We stress that with the WI method we do not estimate directly the mutual information contained in the stimulus response probabilities in the form of either binary words, as in Strong et al. (1998), or PC scores, as in Optican and Richmond (1987). Instead, we limited ourselves to the calculation of cross-validated confusion matrices. These estimations give a lower bound to the information available in the data (Quián Quiroga and Panzeri 2009). However, it is in principle possible to extend the WI method to perform also direct calculations of information from neural activity (Strong et al. 1998). It is well-known that such direct computations of information from binary response words are in practice possible only for short response windows due to the curse of dimensionality (Kayser et al. 2009). In this respect, the WI method may be further developed to extend the applicability of direct calculations of information to longer windows by providing a relatively low-dimensional representation of the binary word response, thus providing more data-robust, direct calculations of information. Such implementations are beyond the scope of the present paper but are ripe for future work.

We have also shown that the WI approach can be adapted to denoise spike trains and to estimate and visualize correlations across neurons. In particular, we found an improved performance in estimating information in the correlated firing of neurons compared with a MS approach (Victor and Purpura 1996) or the standard technique of analyzing coincidence counts, what has been called “unitary events” (Grun et al. 2002). As in the case of other methods to assess time-pattern information, the caveat of the unitary event approach is that it is highly dependent on the time resolution used to bin the spike trains. For instance, coactivations can be missed if the time scale at which they occur does not match the window used for binning the spiking activity (Lopes-dos-Santos et al. 2013). Likewise, results with the MS approach rely on the choice of an optimal  $q$ -value. The advantage of wavelets in this respect is the fact that it allows evaluating correlated firing at different time scales and for specific time localizations.

Finally, the possibility of denoising spike trains allows a much clearer visualization of informative time pattern and a better characterization of the selectivity of the neuron (through reverse correlations) as we illustrated with data from monkey auditory cortex. The spike train denoising based on WI could in principle be used for a number of other novel applications. For example, it can be used to characterize better the relationships between spiking activity and local field potentials (particularly their phase at selected frequency bands) carrying out information about the stimuli. The better visualization of the stimulus-driven spikes can be also useful to study relationships between stimulus-driven and “internal state” components of neural activity, a topic of current interest in systems neuroscience (Harris and Thiele 2011). In general, the more accurate characterization of the tuning of the neurons offered by WI denoising will likely offer important practical advantages for the discovery of the “thesaurus” that relates sensory stimulus to neural responses and could further facilitate the understanding of what features encoded by the neurons do contribute to behavior.

A MATLAB implementation of the WI is available from <http://www.le.ac.uk/csn/WI>.

## GRANTS

V. Lopes-dos-Santos was funded by the Science Without Borders program from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) Foundation, Ministry of Education of Brazil. This work was supported by grants from the Engineering and Physical Sciences Research Council and the Human Frontier Science Program. S. Panzeri acknowledges support from the SI-CODE European Union Future Emerging Technology (FET)-Open FP7-284533 Project and from the Autonomous Province of Trento (Call “Grandi Progetti 2012,” Project “Characterizing and Improving Brain Mechanisms of Attention-ATTEND”).

## DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

## AUTHOR CONTRIBUTIONS

V.L.-d.S., S.P., and R.Q.Q. conception and design of research; R.Q.Q. performed preliminary developments and simulations; C.K. and M.E.D. performed monkey and rat experiments, respectively; V.L.-d.S. performed simulation experiments; V.L.-d.S. analyzed data; V.L.-d.S., S.P., C.K., M.E.D., and R.Q.Q. interpreted results of experiments; V.L.-d.S. prepared figures; V.L.-d.S. and R.Q.Q. drafted manuscript; V.L.-d.S., S.P., C.K., M.E.D., and R.Q.Q. edited and revised manuscript; V.L.-d.S., S.P., C.K., M.E.D., and R.Q.Q. approved final version of manuscript.

## REFERENCES

- Adrian ED, Zotterman Y.** The impulses produced by sensory nerve-endings: Part II. The response of a single end-organ. *J Physiol* 61: 151–171, 1926.
- Ahmadi M, Quián Quiroga R.** Automatic denoising of single-trial evoked potentials. *Neuroimage* 66: 672–680, 2013.
- Arabzadeh E, Panzeri S, Diamond ME.** Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway. *J Neurosci* 26: 9216–9226, 2006.
- Bullock TH.** Signals and signs in the nervous system: the dynamic anatomy of electrical activity is probably information-rich. *Proc Natl Acad Sci USA* 94: 1–6, 1997.
- Dayan P, Abbot L.** *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. Cambridge, MA: MIT Press, 2005.
- de Ruyter van Steveninck RR, Lewen GD, Strong SP, Koberle R, Bialek W.** Reproducibility and variability in neural spike trains. *Science* 275: 1805–1808, 1997.
- Di Lorenzo PM, Chen JY, Victor JD.** Quality time: representation of a multidimensional sensory domain through temporal coding. *J Neurosci* 29: 9227–9238, 2009.
- Eckhorn R, Popel B.** Rigorous and extended application of information theory to afferent visual system of the cat. II. Experimental results. *Biol Cybern* 17: 7–17, 1975.
- Fairhall AL, Lewen GD, Bialek W, van Steveninck RR.** Efficiency and ambiguity in an adaptive neural code. *Nature* 412: 787–792, 2001.
- Fisher RA.** The use of multiple measurements in taxonomic problems. *Ann Hum Genet* 7: 179–188, 1936.
- Foffani G, Morales-Botello ML, Aguilar J.** Spike timing, spike count, and temporal information for the discrimination of tactile stimuli in the rat ventrobasal complex. *J Neurosci* 29: 5964–5973, 2009.
- Fontanini A, Katz DB.** State-dependent modulation of time-varying gustatory responses. *J Neurophysiol* 96: 3183–3193, 2006.
- Fotowat H, Harrison RR, Gabbiani F.** Multiplexing of motor information in the discharge of a collision detecting neuron during escape behaviors. *Neuron* 69: 147–158, 2011.
- Ganmor E, Segev R, Schneidman E.** The architecture of functional interaction networks in the retina. *J Neurosci* 31: 3044–3054, 2011.
- Grun S, Diesmann M, Aertsen A.** Unitary events in multiple single-neuron spiking activity: 1. Detection and significance. *Neural Comput* 14: 43–80, 2002.
- Harris KD, Thiele A.** Cortical state and attention. *Nat Rev Neurosci* 12: 509–523, 2011.
- Harvey MA, Saal HP, Dammann JF 3rd, Bensmaia SJ.** Multiplexing stimulus information through rate and temporal codes in primate somatosensory cortex. *PLoS Biol* 11: e1001558, 2013.
- Jacobs AL, Fridman G, Douglas RM, Alam NM, Latham PE, Prusky GT, Nirenberg S.** Ruling out and ruling in neural codes. *Proc Natl Acad Sci USA* 106: 5936–5941, 2009.

- Kayser C, Logothetis NK, Panzeri S.** Millisecond encoding precision of auditory cortex neurons. *Proc Natl Acad Sci USA* 107: 16976–16981, 2010.
- Kayser C, Montemurro M, Logothetis N, Panzeri S.** Spike-phase coding boosts and stabilizes information carried by spatial and temporal spike patterns. *Neuron* 61: 597–608, 2009.
- Laurent G, Wehr M, Davidowitz H.** Temporal representations of odors in an olfactory network. *J Neurosci* 16: 3837–3847, 1996.
- Lebedev MA, Mirabella G, Erchova I, Diamond ME.** Experience-dependent plasticity of rat barrel cortex: redistribution of activity across barrel-columns. *Cereb Cortex* 10: 23–31, 2000.
- Lisman J.** The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus* 15: 913–922, 2005.
- Lopes-dos-Santos V, Ribeiro S, Tort AB.** Detecting cell assemblies in large neuronal populations. *J Neurosci Methods* 220: 149–166, 2013.
- Magri C, Whittingstall K, Singh V, Logothetis NK, Panzeri S.** A toolbox for the fast information analysis of multiple-site LFP, EEG and spike train recordings. *BMC Neurosci* 10: 81, 2009.
- Mallat S.** *A Wavelet Tour of Signal Processing: The Sparse Way*. San Diego, CA: Academic Press, 2008.
- Mallat SG.** *A Wavelet Tour of Signal Processing*. San Diego, CA; London: Academic Press, 1999.
- Montemurro MA, Panzeri S, Maravall M, Alenda A, Bale MR, Brambilla M, Petersen RS.** Role of precise spike timing in coding of dynamic vibrissa stimuli in somatosensory thalamus. *J Neurophysiol* 98: 1871–1882, 2007.
- Nadasdy Z.** Information encoding and reconstruction from the phase of action potentials. *Front Syst Neurosci* 3: 6, 2009.
- Nemenman I, Bialek W, de Ruyter van Steveninck R.** Entropy and information in neural spike trains: progress on the sampling problem. *Phys Rev E Stat Nonlin Soft Matter Phys* 69: 056111, 2004.
- Optican LM, Richmond BJ.** Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. Information theoretic analysis. *J Neurophysiol* 57: 162–178, 1987.
- Paninski L.** Estimation of entropy and mutual information. *Neural Comput* 15: 1191–1253, 2003.
- Panzeri S, Brunel N, Logothetis NK, Kayser C.** Sensory neural codes using multiplexed temporal scales. *Trends Neurosci* 33: 111–120, 2010.
- Panzeri S, Petersen RS, Schultz SR, Lebedev M, Diamond ME.** The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron* 29: 769–777, 2001.
- Panzeri S, Schultz SR.** A unified approach to the study of temporal, correlational, and rate coding. *Neural Comput* 13: 1311–1349, 2001.
- Panzeri S, Senatore R, Montemurro M, Petersen R.** Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol* 98: 1064–1072, 2007.
- Panzeri S, Treves A.** Analytical estimates of limited sampling biases in different information measures. *Network* 7: 87–107, 1996.
- Quiara Quiroga R, Panzeri S.** Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci* 10: 173–185, 2009.
- Quiroga RQ.** Spike sorting. *Curr Biol* 22: R45–R46, 2012.
- Quiroga RQ, Nadasdy Z, Ben-Shaul Y.** Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput* 16: 1661–1687, 2004.
- Quiroga RQ, Reddy L, Koch C, Fried I.** Decoding visual inputs from multiple neurons in the human temporal lobe. *J Neurophysiol* 98: 1997–2007, 2007.
- Richmond BJ, Optican LM.** Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *J Neurophysiol* 57: 147–161, 1987.
- Rieke F, Warland D, de Ruyter van Steveninck R, Bialek W.** *Spikes: Exploring the Neural Code*. Cambridge, MA: MIT Press, A Bradford Book, 1999, p. 416.
- Roussin AT, D'Agostino AE, Fooden AM, Victor JD, Di Lorenzo PM.** Taste coding in the nucleus of the solitary tract of the awake, freely licking rat. *J Neurosci* 32: 10494–10506, 2012.
- Shadlen MN, Newsome WT.** Noise, neural codes and cortical organization. *Curr Opin Neurobiol* 4: 569–579, 1994.
- Shannon CE.** A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423, 623–656, 1948.
- Shew WL, Yang H, Yu S, Roy R, Plenz D.** Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches. *J Neurosci* 31: 55–63, 2011.
- Strang G, Nguyen T.** *Wavelets and Filter Banks*. Cambridge, MA: Wellesley-Cambridge Press, 1996, p. 500.
- Strong SP, Koberle R, van Steveninck RR, Bialek W.** Entropy and information in neural spike trains. *Phys Rev Lett* 80: 197–200, 1998.
- Victor JD.** Binless strategies for estimation of information from neural data. *Phys Rev E Stat Nonlin Soft Matter Phys* 66: 051903, 2002.
- Victor JD.** How the brain uses time to represent and process visual information. *Brain Res* 886: 33–46, 2000.
- Victor JD, Purpura KP.** Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol* 76: 1310–1326, 1996.

Como discutido no primeiro capítulo desta tese, os métodos aqui apresentados são apenas ferramentas. Eles precisam de dados adquiridos em experimentos adequados para testar hipóteses. Ao desenvolvê-los, minha intenção foi a de criar ferramentas simples e intuitivas sem comprometer o rigor estatístico. Se cheguei a meu objetivo, seus resultados podem ser interpretados de maneira crítica e lúcida.

Acredito que não é controverso afirmar que há grupos de neurônios que se co-ativam com frequência ou que neurônios podem responder a certos estímulos com padrões temporais consistentes. Acredito que a discussão deve ser em torno da importância dessas propriedades para o funcionamento adequado do cérebro.

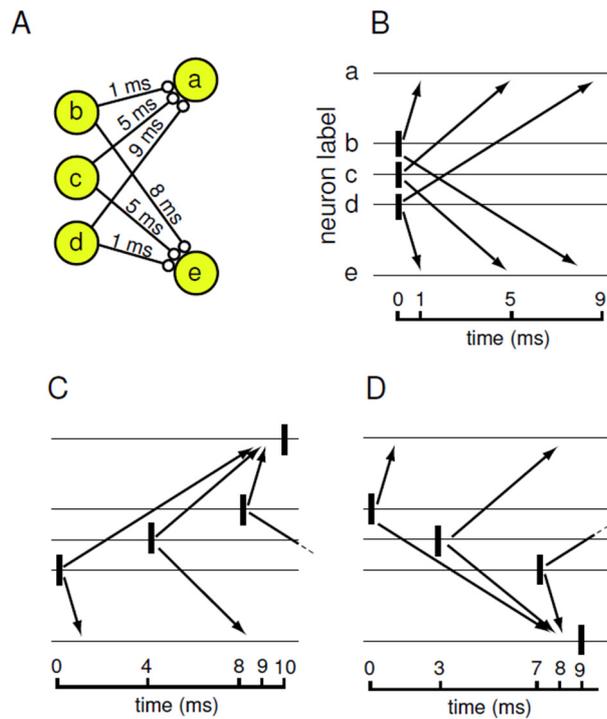
É possível perceber um *objeto* caso as células responsáveis por codificá-lo não disparem com um determinado padrão temporal? Seriam esses padrões temporais um epifenômeno ou a chave da codificação neural? Precisamos de experimentos que testem se a percepção de um objeto é comprometida quando as células que o codificam disparam de modo diferente desses padrões. O método apresentado no Capítulo 3 pode quantificar sua consistência, mas ele é inútil sem dados que provem que esses padrões temporais se relacionam com alguma variável comportamental. Da mesma forma, assembleias podem ser detectadas e rastreadas temporalmente com o método que descrevi no Capítulo 2, mas ainda precisamos provar que a dessincronização entre seus membros resulta em algum prejuízo cognitivo.

Minha percepção é que essas questões são muitas vezes tratadas na literatura científica como se houvesse um homúnculo em cada cérebro observando os disparos de bilhões de células. Quando alguém argumenta que o cérebro leva em consideração padrões de disparo ou co-ativações, pois assim tem sua capacidade de codificação aumentada, ele se esquece de que essas possíveis formas de representação necessitam de um mecanismo fisiológico. Ninguém lê o código neural. O código neural faz parte de um processo que ainda não entendemos.

### **MECANISMOS PARA O CÓDIGO TEMPORAL**

À primeira vista, a co-ativação entre células tem um mecanismo de leitura simples. A rede dendrítica de um neurônio *alvo* pode atuar como um detector de coincidências. Se os neurônios pré-sinápticos disparam juntos, seu efeito no leitor seria somado e, portanto, eles

seriam mais eficientes em transmitir informação. Um problema desse pensamento é desconsiderar as latências de propagação dos potenciais de ação. Sabemos que no neocórtex essas latências podem variar de poucos até dezenas de milissegundos (Ferraina et al., 2002; Miller, 1975; Swadlow, 1985, 1992). Portanto, a co-ativação de neurônios não garante que eles atinjam um leitor sincronamente.



**Figura 4.1. Mecanismo de leitura de padrões temporais de disparo.** Extraído de Izhikevich (2006).

Quando levamos em consideração a distribuição de latências de propagação, os padrões temporais de disparo fazem mais sentido. Izhikevich (2006) tratou essa questão de forma brilhante. Considere a rede ilustrada na Figura 4.1A. Os neurônios b, c e d conectam-se pré-sinápticamente às células a e e com diferentes tempos de condução. Quando b, c e d disparam de modo perfeitamente coincidente, seus disparos chegam descoordenados em a e e (Figura 4.1B). Para atingirem a ou e de forma eficiente, as células pré-sinápticas precisam disparar com uma sequência adequada (Figuras 4.1C-D). Cada padrão temporal de disparo tem sua própria consequência na propagação da atividade na rede. O modelo de Izhikevich não só resolve o problema dos tempos de propagação no córtex, mas mostra de maneira biologicamente plausível como a dimensão temporal pode aumentar a capacidade do código neural.

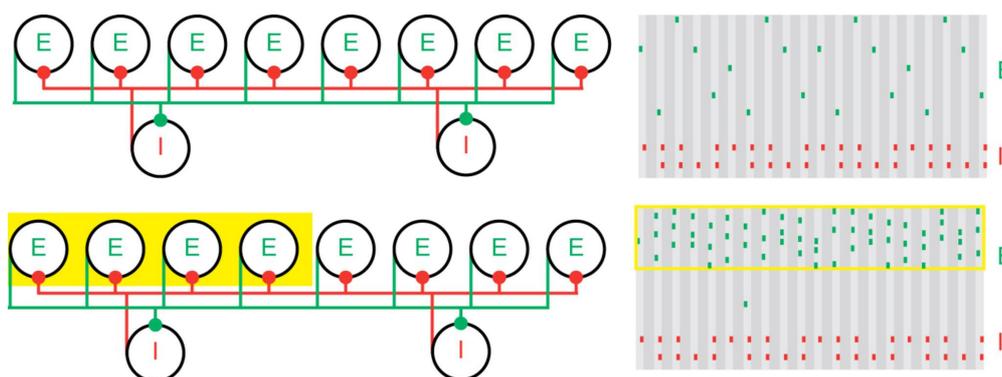
## MECANISMOS PARA O CÓDIGO POPULACIONAL

Acima mencionei alguns problemas para a implementação de um código que levasse em conta a co-ativação de neurônios dispersos. No entanto, como discutido nos Capítulos 1 e 2, existem diversas evidências experimentais para a existência de assembleias neurais. Nessa seção, discuto como esses problemas podem ser resolvidos.

Hebb (1949) acreditava que membros de uma assembleia eram anatomicamente conectados e, conseqüentemente, a ativação de parte deles era suficiente para recrutar todo o grupo. Segundo a teoria de Hebb, a ativação da assembleia pode ser entendida como uma reverberação, um *loop* de atividade elétrica que funciona como atrator.

Buzsáki (2010) propõe que uma assembleia é um grupo de neurônios que, quando ativados conjuntamente, provocam uma consequência perceptível em um *leitor*. Nessa visão mais moderna não há necessidade de *loops* nem de conexões anatômicas entre eles; apenas uma relação funcional.

Nesse ponto voltamos ao problema apresentado anteriormente: como neurônios desconectados anatomicamente e distribuídos podem coordenar suas atividades? Acredita-se que esse processo esteja relacionado às *oscilações* eletrofisiológicas medidas no meio extracelular em diversas áreas do cérebro (Freeman, 1978; Fries, 2005; Gray, 1994; Lisman, 2005; Singer, 1993). Oscilações refletem principalmente a troca de cargas elétricas entre os dendritos de milhares de neurônios que acontecem de forma regular, i.e., de ciclos de excitação e inibição em um volume (Buzsáki et al., 2012).



**Figura 4.2. Modelo PING.** *Painel superior esquerdo:* conexões entre células piramidais excitatórias (E) e inibitórias (I). *Painel superior direito:* Atividade da rede (traços verdes e vermelhos representam potenciais de ação de células E e I, respectivamente) para a rede mostrada no painel superior a esquerda. O eixo horizontal representa tempo de simulação. *Painel inferior esquerdo:* Mesmo que painel acima, porém quando um subgrupo da população E recebe uma excitação extra (subgrupo destacado em amarelo). *Painel inferior direito:* atividade da rede mostrada no painel à esquerda. Figura extraída de Koppel (2010).

O modelo de rede PING (do inglês *Pyramidal-Interneuron Network Gamma*, Koppel (2010)) é um exemplo de teoria que relaciona oscilações com a organização de assembleias (Figura 4.2).

A fundamentação desse mecanismo vem da interação entre células piramidais e interneurônios (Traub et al., 1997). As células piramidais são neurônios excitatórios (E) que, presumidamente, codificam a memória; ou seja, são as células que formariam as assembleias hebbianas. Os interneurônios são inibitórios (I) e, apesar de não participarem a codificação *per se*, participam do processamento de informação modulando a atividade da rede.

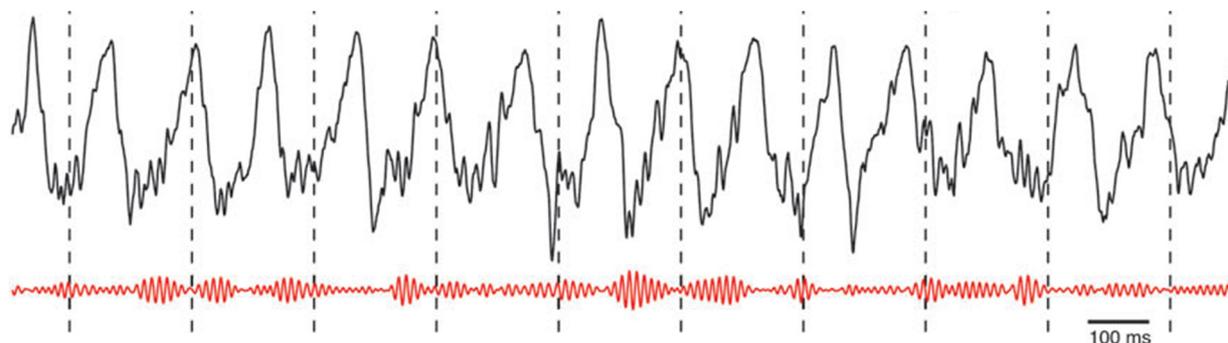
O painel superior esquerdo da Figura 4.2 mostra a arquitetura de uma rede que gera atividade PING. Quando uma entrada aleatória é aplicada nessa rede, os neurônios E que, por acaso, receberem maior excitação disparam e ativam as células I, que por sua vez inibem todos os neurônios da rede. Quando a inibição passa (normalmente o decaimento da inibição é considerado exponencial), células E ficam livres para dispararem novamente (Figura 4.2, painel superior direito). Em outras palavras, existem janelas de oportunidade de disparo que surgem ciclicamente (Börgers and Kopell, 2005). Como a constante de decaimento da inibição e a latência de conexão entre as células é considerada fixa, a atividade global dessa rede cicla em uma frequência definida por esses dois parâmetros. Para parâmetros biologicamente plausíveis, a rede pode oscilar estavelmente entre 30 e 100 Hz, gerando as denominadas *oscilações gama* ( $\gamma$ ). Considere agora que algum subgrupo de células E recebe uma maior excitação que as demais por um viés sensorial. Como consequência, essas células disparam antes das outras e indiretamente, pela ativação das células I, evitam os disparos das concorrentes (Figura 4.2, painéis inferiores). Nesse caso, apenas o subgrupo de células com maior excitação dispara nas janelas de oportunidade, caracterizando a atividade de assembleia.

Essa assembleia é formada por neurônios que disparam para diferentes componentes de um objeto (o viés sensorial). Quando disparam juntos dessa forma, sincronizados como resultado das conexões intrínsecas da rede, formam a coalizão que *representa* aquele objeto como uma unidade. Essa coordenação temporal pode promover um reforço de conexões fracas entre as células. O leitor da assembleia por sua vez, se estiver submetido a uma oscilação acoplada com a da assembleia, terá maior eficiência em *perceber* sua ativação. Em outras palavras, a oscilação que abre janelas de oportunidade para disparos da assembleia deve estar em fase com os ciclos de excitação de seu alvo.

Na próxima seção, uso a rede hipocampal como estudo de caso desse possível esquema de codificação e sua possível relação com o aprendizado.

## ASSEMBLEIAS HIPOCAMPAIS

O hipocampo recebeu atenção especial por ser essencial para a formação e evocação de memórias (Cohen et al., 1999; Eichenbaum, 2004; Scoville and Milner, 1957; Squire and Zola-Morgan, 1991; Wood et al., 1999). Como consequência, um grande corpo de conhecimento acerca de sua anatomia, arquitetura de rede, tipos celulares, propriedades biofísicas e padrões eletrofisiológicos foi formado (Andersen et al., 2006).

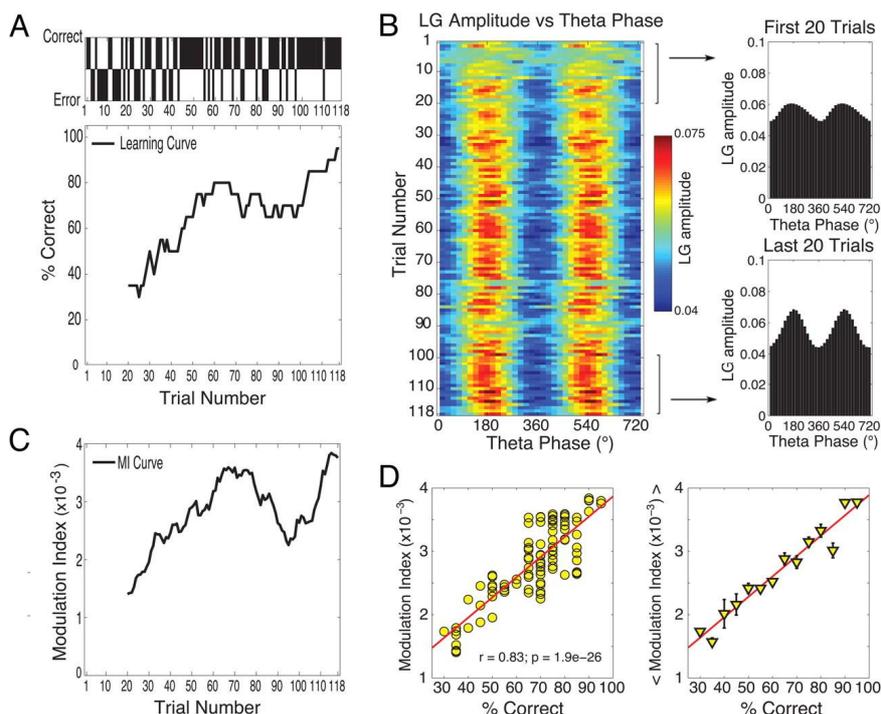


**Figura 4.3. Acoplamento entre oscilações teta e gama no hipocampo.** Traçado negro no topo mostra o potencial de campo medido na subregião CA1 do hipocampo de um rato. Traçado vermelho abaixo mostra esse sinal filtrado entre 60 e 100 Hz. Figura extraída de Scheffer-Teixeira (2012).

No hipocampo, as oscilações  $\gamma$  estão geralmente acopladas a oscilações *teta* ( $\theta$ , 4 - 12 Hz; (Buzsáki, 2002; Vanderwolf, 1969; Winson, 1972). Mais especificamente, a energia em  $\gamma$  aumenta durante fases específicas das ondas  $\theta$  (Bragin et al., 1995). Esse tipo de sincronia é chamado de *acoplamento fase-amplitude* (Jensen and Colgin, 2007). A Figura 4.3 mostra um exemplo dessa relação. O traçado negro mostra um trecho de um sinal registrado na subregião CA1 do hipocampo de um rato em livre movimento. Abaixo, em vermelho, destaca-se esse mesmo sinal filtrado em uma faixa de frequências correspondente a  $\gamma$  (nesse caso, 60 a 100 Hz). Note que maiores amplitudes de  $\gamma$  coincidem com vales de  $\theta$ .

Interessantemente, algumas evidências indicam que a força dessa modulação é importante para algumas demandas cognitivas (Demiralp et al., 2007; Jensen and Colgin, 2007; Kendrick et al., 2011; Lisman, 2010; Schack and Klimesch, 2002). A Figura 4.4A mostra a curva de aprendizado de um rato (painel inferior) em uma tarefa de associação entre itens e contexto (Tort et al., 2009). O painel esquerdo da Figura 4.4B mostra a amplitude média em  $\gamma$  (30 a 60 Hz) em função da fase de  $\theta$  para cada tentativa da tarefa. Note que a amplitude de  $\gamma$  aumenta quando a fase de  $\theta$  aproxima-se de 180º (nessa convenção, 180º representa o vale da oscilação). Painéis a direita da mesma figura mostram a distribuição média de amplitudes para as vinte primeiras e últimas tentativas (painéis superior e inferior,

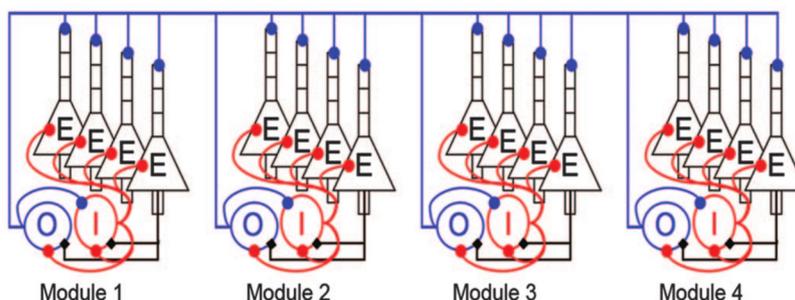
respectivamente). Claramente, nas últimas tentativas, as amplitudes de  $\gamma$  ficam mais concentradas; ou seja, a modulação amplitude-fase dessas oscilações fica mais forte. A Figura 4.4C mostra como essa força de acoplamento varia em função da tentativa. A Figura 4.4D mostra que existe uma correlação linear entre o índice de modulação e a performance na tarefa. Em outras palavras, o aprendizado do animal parece estar relacionado com quão organizadas são as amplitudes de gama em relação a fase do teta no hipocampo.



**Figura 4.4. Força do acoplamento entre oscilações teta e gama se correlaciona com aprendizado.** (A) Curva de aprendizado. *Painel superior:* perfil de performance de um rato em uma tarefa de contexto. *Painel inferior:* Eixo horizontal representa o número da tentativa, eixo vertical representa a proporção de tentativas corretas. (B) *Painel esquerdo:* Amplitude média de oscilações  $\gamma$  em função da fase de  $\theta$  (eixo horizontal) para 118 tentativas (eixo vertical). *Painéis direitos:* Média das 20 tentativas iniciais (painel superior) e das 20 tentativas finais (painel inferior). Note que nas últimas tentativas a concentração de energia de  $\gamma$  é mais concentrada, ou seja, o acoplamento é mais forte. (C) Índice de modulação  $\gamma$ - $\theta$  em função no número de tentativas. (D) *Painel esquerdo:* Índice de modulação em função da performance. *Painel direito:* Média do índice de modulação para cada performance. Figura extraída de Tort et al. (2009).

Alguns modelos teóricos propuseram que esse acoplamento emerge da interação entre as células E e I com outro tipo de neurônio inibitório: as células *oriens-lacunosum-moleculare* (Cutsuridis et al., 2010; Kopell et al., 2010; Kunec et al., 2005). As células *oriens-lacunosum-moleculare* (O) podem ser fundamentais na geração de oscilações  $\theta$  no hipocampo, tendo em vista suas particulares propriedades anatômicas e características

biofísicas (Gloveli et al., 2005) A Figura 4.4 mostra as conexões de células E, I e O (Tort et al., 2007). Simplificadamente, a interação entre células E e I geram oscilações PING, porém as células O inibem essa atividade em períodos regulares correspondentes ao ciclo  $\theta$ .



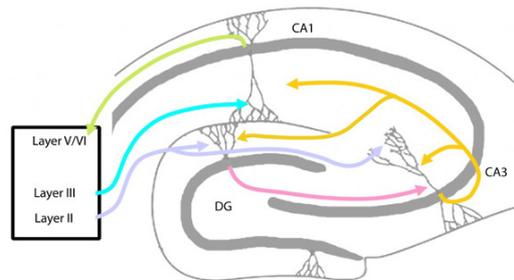
**Figura 4.5. Arquitetura de rede do hipocampo para geração do acoplamento tetra-gama observado experimentalmente.** Figura mostra conexões entre os três tipos de célula que estariam envolvidos. Figura extraída de (2007).

Portanto, de acordo com esses modelos, assembleias formadas por células E têm seus disparos organizados em oscilações  $\gamma$  por suas conexões intrínsecas no hipocampo (Figura 4.2); e células O, por sua vez, impõem na rede um ritmo  $\theta$ , que abre janelas de oportunidade para que essas oscilações  $\gamma$  aconteçam. Porém, como explicar mecanisticamente que essa interação entre oscilações  $\theta$  e  $\gamma$  (Figura 4.3) se relacione com aprendizado (Figura 4.4)?

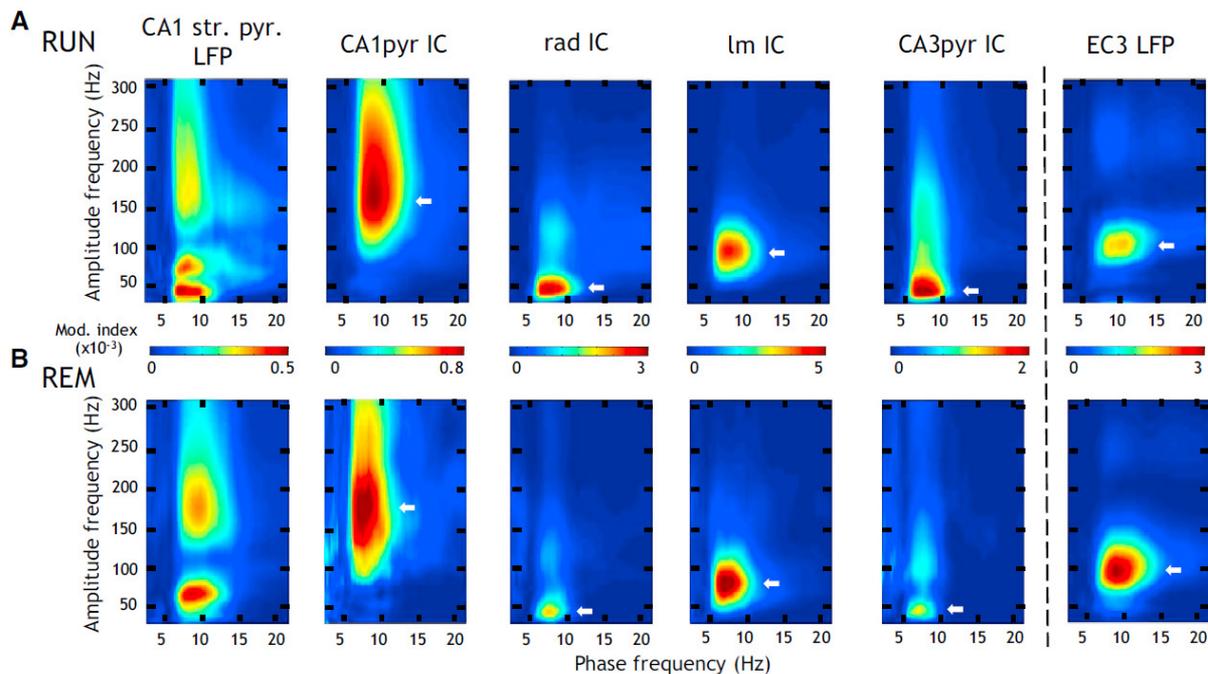
Podemos considerar que o desempenho em uma tarefa como a exemplificada na Figura 4.4 depende de dois processos: **1) a codificação**, que consiste no processamento e no armazenamento de novas informações; e **2) a evocação**, que é a recuperação de uma memória importante pela tarefa. Uma possibilidade é que o ciclo  $\theta$  sirva de referência temporal para segregar esses dois processos no hipocampo (Hasselmo et al., 2002), que seriam refletidos como diferentes oscilações  $\gamma$ , como se detalha a seguir.

A Figura 4.6 mostra como é o fluxo de informação entre o córtex entorrinal (CE, do inglês *entorhinal cortex*) e as três principais subáreas do hipocampo: giro dentado (GD, do inglês *dentate gyrus*), CA1 e CA3. Atualmente, acredita-se que, durante a codificação, as informações sensoriais pré-processadas no neocórtex são enviadas ao hipocampo via CE. A camada 3 do CE projeta para CA1, e a camada 2 projeta para o GD e CA3. Por possuir muitas recorrências, a rede de células E do CA3 é pensada como ideal para um armazenamento rápido e eficiente de novas assembleias (Lisman, 2005; Rennó-Costa et al., 2014; Treves and Rolls, 1994). Por sua vez, CA1 teria a função de integrar as informações sensoriais vindas do

CE e as evocadas em CA3 (Vinogradova, 2001). Essa integração seria importante para a consolidação da memória, pois CA1 é a porta de saída de informação do hipocampo e, portanto poderia ligar memórias dispersas no neocórtex (McClelland et al., 1995).

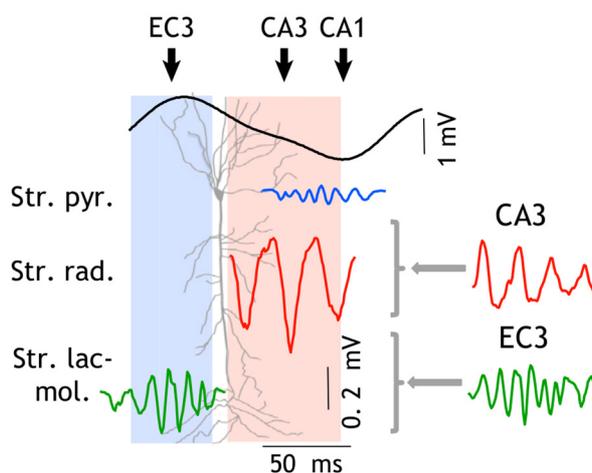


**Figura 4.6. Fluxo de informação no hipocampo.** Figura extraída de Petrantonakis and Poirazi (2014).



**Figura 4.7. Índices de acoplamento fase-amplitude no hipocampo e no córtex entorrinal enquanto ratos exploram um ambiente (A) ou durante o sono REM (B).** Cores representam índices de acoplamento fase-amplitude. Valores altos (cores quentes) indicam que a energia de certa frequência (eixo vertical) é modulada por uma oscilação de uma certa frequência central (eixo horizontal). *CA1 str pyr LFP* e *CA3 LFP* correspondem aos potenciais de campo da camada piramidal de CA1 e da terceira camada do CE, respectivamente. *CA1pyr IC*, *rad IC*, *lm IC*, *CA3pyr IC*, correspondem aos potenciais de campo das camadas piramidal, *radiatum*, *lacunosum-moleculare* de CA1 e da camada piramidal de CA3, respectivamente; corrigidas com análise de componentes independentes para evitar contaminação de condução de volume. Figura extraída de Schomburg (2014).

Colgin et al. (2009) mostraram que células em CA1 oscilam em diferentes frequências  $\gamma$ , dependendo se estão sincronizadas com o CA3 (25 a 50 Hz) ou com o EC (65 a 140 Hz). Eles chamaram essas subfaixas de *low*  $\gamma$  e *high*  $\gamma$  ( $\gamma$  lento e rápido goravante), e propuseram que elas representariam diferentes canais de comunicação entre essas áreas. Recentemente, Schomburg et al. (2014) mediram os índices de acoplamento fase-amplitude em diferentes camadas de CA1 e os compararam com os obtidos em CA3 e na camada 3 do córtex entorrinal (CE3; Figura 4.7).



**Figura 4.8. Desenho esquemático da teoria para a integração de informações em CA1.** Setas no topo indicam quando acontecem os disparos das células E da camada 3 do CE (EC3 na figura), CA3 e CA1. Um ciclo  $\theta$  é mostrado abaixo como referência temporal. Traçados vermelhos representam oscilações  $\gamma$ -lento na camada R (Str. rad. na figura, de *stratum radiatum*). Traçados verdes representam oscilações  $\gamma$ -rápido na camada LM (Str. lac-mol. na figura, de *stratum-lacunosum-moleculare*). O eixo horizontal indica quando acontecem as oscilações usando o ciclo  $\theta$ ; o eixo vertical indica onde as oscilações acontecem usando o neurônio E como referência espacial. Traçado azul representa as oscilações de mais alta frequência que são medidas na região (ver painéis CA1pyr IC na Figura 4.7), e que provavelmente refletem potenciais de ação das células da camada piramidal (Scheffer-Teixeira et al., 2013). Figura extraída de Schomburg (2014).

Seus resultados mostraram que no CE3 a energia de  $\gamma$ -rápido (por eles definido como a banda de 60 a 120 Hz) é fortemente modulado por  $\theta$ , mas o mesmo não acontece com  $\gamma$ -lento (30 a 80 Hz). Esse padrão é inverso em CA3. Paralelamente, em CA1, a camada *lacunosum-moleculare* (LM) apresentou um padrão muito semelhante àquele encontrado no CE, enquanto a camada *radiatum* (R) refletia o padrão encontrado em CA3. Essas observações são consistentes com a teoria de Colgin et al. (2009) já que CE3 e CA3 projetam para as camadas LM e R, respectivamente. A Figura 4.8 mostra um desenho esquemático resumindo esses resultados.

Embora o corpo de resultados tenha crescido substancialmente nos últimos anos, o entendimento da relação entre as observações descritivas revisadas acima e variáveis comportamentais continua elusiva e pouco discutida. A compreensão dos mecanismos por trás das correlações entre performance cognitiva e acoplamento de oscilações no hipocampo pode levar a importantes hipóteses de como o cérebro funciona.

Nessa linha, nós mostramos que assembleias formadas por neurônios da camada piramidal de CA1 no hipocampo apresentam atividade oscilatória e sincronizada com ritmos  $\theta$  presentes nos potenciais de campo medidos nessa mesma região (Capítulo 2). Interessantemente, as células que compunham as assembleias não apresentavam acoplamento com  $\theta$  quando não sincronizadas com o grupo, indicando que essas oscilações estão relacionadas à organização de atividade dessas células espacialmente dispersas.

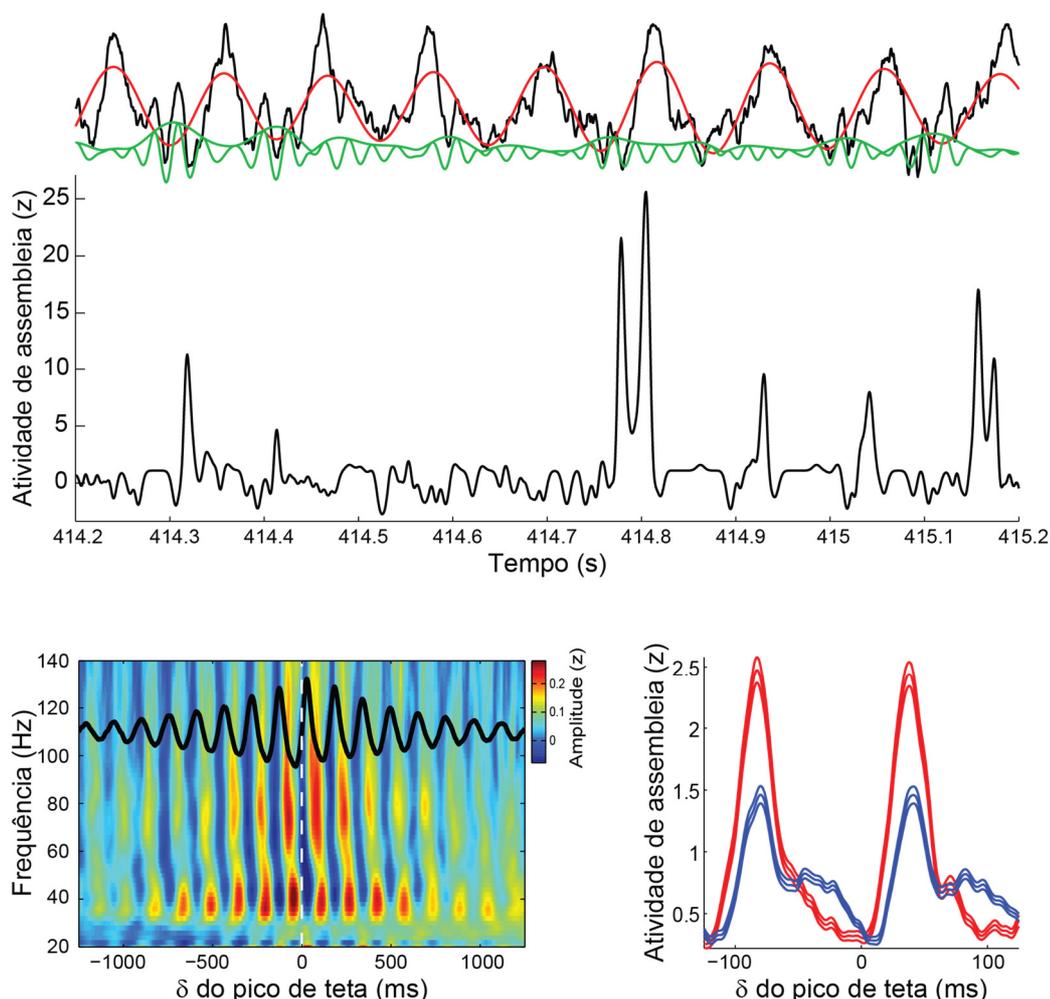
Durante essa mesma análise, algumas observações preliminares ainda não publicadas foram feitas em relação às oscilações  $\gamma$ , resumidas na Figura 4.9. O painel superior mostra um trecho representativo do potencial de campo medido em CA1. O traçado negro acima no topo do painel mostra o sinal não processado; em vermelho e em verde está o mesmo sinal filtrado em  $\theta$  (5 a 12 Hz) e  $\gamma$ -baixo (30 a 60 Hz), respectivamente. O sinal em verde está acompanhado por seu envelope (energia instantânea) calculado com a Transformada de Hilbert. O traçado negro abaixo desses sinais representa a atividade de uma assembleia formada por neurônios piramidais em CA1. Note-se que em alguns ciclos de  $\theta$  há uma clara aparição de oscilações  $\gamma$ -lento.

O painel inferior esquerdo mostra a energia instantânea (escore  $z$ ) média de várias frequências (eixo vertical) centradas nos picos de atividade da assembleia (eixo horizontal). Ou seja, cores vermelhas denotam que há, em média, um aumento de energia de certa frequência em algum tempo relativo às ativações da assembleia. O traçado negro na figura mostra a atividade média do potencial de campo não filtrado calculado da mesma forma. Em resumo, esses resultados indicam que dentro de um ciclo  $\theta$  existe uma cadeia de eventos: as ativações de assembleias em CA1 são precedidas por um aumento de  $\gamma$ -rápido e por um aumento de  $\gamma$ -lento (20 e 10 ms, respectivamente).

O painel inferior direito mostra que as ativações da assembleia nos ciclos  $\theta$  são mais fortes quando precedidas por aumento de  $\gamma$ -lento. Mais especificamente, os traçados vermelhos e azuis mostram as médias (com seus intervalos de 95% de confiança) da atividade de assembleias em ciclos  $\theta$  com e sem  $\gamma$ -lento, respectivamente.

Tendo em vista a literatura, esses resultados indicam que antes das assembleias de CA1 dispararem, essa população recebe informação vinda do CE ( $\gamma$ -rápido) e de CA3  $\gamma$ -lento em sequência. Minha hipótese é que essa integração seja computada nos dendritos da rede

e que dê o viés para que um subgrupo de células dispare no ciclo  $\theta$  correspondente, inibindo indiretamente os demais neurônios. Esse subgrupo de células, quando ativo conjuntamente de forma persistente, formaria uma assembleia que ligaria memórias já formadas anteriormente (informação vinda de CA3) com novos dados sensoriais (informação vinda do CE). Portanto, o ciclo  $\theta$  pode dar a referência temporal para que o fluxo sequencial de informação seja organizado e processado em CA1. A maior organização desse fluxo levaria a maior eficiência em formar novas assembleias e seria refletida como um acoplamento fase-amplitude similar ao mostrado na Figura 4.4.



**Figura 4.9. Assembleias em CA1 disparam em fases consistentes de  $\theta$  em ciclos com aumento de  $\gamma$ -baixo.**

*Painel superior:* Traçados acima mostram um trecho de um potencial de campo medido na camada piramidal de CA1 (negro), juntamente com duas filtragens suas: de 5 a 12 Hz (correspondente a  $\theta$ ) em vermelho e de 30 a 60 Hz (correspondente a  $\gamma$ -baixo) em verde. A componente  $\gamma$ -lento está plotada juntamente com seu envelope de amplitude (sua energia instantânea) computada com a Transformada de Hilbert. *Abaixo está a atividade de uma assembleia registrada na mesma região.* *Painel inferior esquerdo:* Cores representam o espectrograma médio (computado via Transformada *Wavelet* Contínua) do potencial de campo centrado em ativações da mesma assembleia. Ativações de

assembleias são definidas como valores três desvios padrão acima da média. Cores indicam a energia média (escore  $z$ ) de uma certa frequência (eixo vertical) em dado instante de tempo relativo à ativação da assembleia (eixo horizontal). Traçado negro mostra a média do potencial de campo não filtrado. *Painel inferior direito*: Traçados representam a atividade da assembleia (média e intervalo de 95% de confiança) centrada nos picos de  $\theta$  em ciclos com e sem  $\gamma$ -lento (vermelho e azul, respectivamente).

## CONSIDERAÇÕES FINAIS

O mito da elegância e da singularidade provavelmente nasceu nas mais antigas ciências, mas seus rastros podem ser encontrados atualmente nas neurociências. O cérebro não emergiu dos caminhos tortuosos da evolução como uma máquina engenhosamente construída. O fetiche silencioso de modelar todo o sistema neural com um único esquema simples pode ajudar a vender jornais e ganhar editais, mas não ajuda em seu entendimento científico.

Isso não significa que os modelos teóricos precisem ser máquinas de *Rube Goldberg*. Nossos modelos devem ser intuitivos para poderem ser interpretados, mas devem ser biologicamente plausíveis para levantar hipóteses experimentalmente falseáveis.

Nesta tese, minha intenção foi contribuir para o estudo do código neural desenvolvendo métodos de análise para estudar dados de alta dimensionalidade típicos dos registros eletrofisiológicos contemporâneos. Espero que esses métodos possam gerar resultados de fácil interpretação e, dessa forma, possam ajudar no processo de entendimento desses sinais e conseqüentemente na melhoria de nossos modelos do sistema neural.

---

# ANEXO I: DETECTANDO ASSEMBLEIAS COM ANÁLISE DE COMPONENTES PRINCIPAIS

# Neuronal Assembly Detection and Cell Membership Specification by Principal Component Analysis

Vítor Lopes-dos-Santos<sup>1,2,\*</sup>, Sergio Conde-Ocazonez<sup>1,2,3</sup>, Miguel A. L. Nicolelis<sup>2,3</sup>, Sidarta T. Ribeiro<sup>1,2</sup>, Adriano B. L. Tort<sup>1,2\*</sup>

**1** Brain Institute, Federal University of Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil, **2** Edmond and Lily Safra International Institute of Neuroscience of Natal, Natal, Rio Grande do Norte, Brazil, **3** Duke University Center for Neuroengineering and Department of Neurobiology, Duke University, Durham, North Carolina, United States of America

## Abstract

In 1949, Donald Hebb postulated that assemblies of synchronously activated neurons are the elementary units of information processing in the brain. Despite being one of the most influential theories in neuroscience, Hebb's cell assembly hypothesis only started to become testable in the past two decades due to technological advances. However, while the technology for the simultaneous recording of large neuronal populations undergoes fast development, there is still a paucity of analytical methods that can properly detect and track the activity of cell assemblies. Here we describe a principal component-based method that is able to (1) identify all cell assemblies present in the neuronal population investigated, (2) determine the number of neurons involved in ensemble activity, (3) specify the precise identity of the neurons pertaining to each cell assembly, and (4) unravel the time course of the individual activity of multiple assemblies. Application of the method to multielectrode recordings of awake and behaving rats revealed that assemblies detected in the cerebral cortex and hippocampus typically contain overlapping neurons. The results indicate that the PCA method presented here is able to properly detect, track and specify neuronal assemblies, irrespective of overlapping membership.

**Citation:** Lopes-dos-Santos V, Conde-Ocazonez S, Nicolelis MAL, Ribeiro ST, Tort ABL (2011) Neuronal Assembly Detection and Cell Membership Specification by Principal Component Analysis. PLoS ONE 6(6): e20996. doi:10.1371/journal.pone.0020996

**Editor:** Matjaz Perc, University of Maribor, Slovenia

**Received:** April 11, 2011; **Accepted:** May 17, 2011; **Published:** June 15, 2011

**Copyright:** © 2011 Lopes-dos-Santos et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Conselho Nacional de Desenvolvimento Científico e Tecnológico, and Associação Alberto Santos Dumont para Apoio à Pesquisa. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: vtlsantos@gmail.com (VL-d-S); tort@neuro.ufrn.br (ABL)

These authors contributed equally to this work.

## Introduction

Hebb's seminal work constitutes a landmark of modern neuroscience [1]. His theory proposes detailed neural mechanisms for the processing and learning of information, from the molecular, cellular and circuit levels to the emergence of complex cognitive functions. According to Hebb's hypothesis, the recurrent co-activation of a subset of neurons would increase the efficiency of their connections, leading to the formation of a cell assembly. Therefore, synchronization of spike times would play a critical role in the creation of new assemblies [2,3,4,5,6]. In this context, a cell assembly is defined as a group of neurons that fire together and wire together. Due to the increased strength of the connections linking members of the assembly, activation of some of its neurons would trigger the activation of the entire neuronal group, leading to pattern completion [7,8,9]. Hebb also postulated that the activation of a cell assembly can lead to the sequential activation of other assemblies, a phenomenon he termed as phase-sequences, and proposed to underlie complex brain computations (see also [10,11,12]). In line with this view, neocortical and hippocampal information has been shown to be widely distributed over neuronal populations, rather than encoded by the activity of highly specialized cells [13,14,15,16,17,18,19].

The actual investigation of Hebbian cell assemblies and their dynamics is only beginning to be possible, thanks to major technological advances that allow the simultaneous and chronic recording of large neuronal populations [20,21,22]. In parallel with these advances, mathematical methods have been developed to address Hebb's hypotheses in experimental data, such as template matching of neuronal population activity [23,24,25] and the detection of precise multi-neuron firing [26,27,28,29,30]. Powerful methods for the detection of neuronal co-activation based on Principal Component Analysis (PCA) were also described [31,32,33], which have recently been extended to incorporate strong statistical support [34]. The latter framework is able to reliably detect the presence of cell assemblies and to assess ensemble activation with high temporal resolution based on the projection of network activity on the principal components (PCs) of the neuronal correlation matrix (see next section for a definition).

Despite its successful initial applications [32,35,36], the PCA-based method presents some limitations. First, it does not identify which specific neurons compose the detected assemblies. In addition, as demonstrated in the present work, the use of individual PCs in order to represent assembly activity patterns is misleading when there are neurons shared by different assemblies. As a consequence, in these cases the projection of neuronal activity

based on PCs does not match the actual time course of individual assembly activation. Since it is currently believed that most, if not all, neocortical and hippocampal neurons take part in multiple assemblies (see Discussion), such limitation is an important one.

To address these gaps, we present here an exploration of some of the key properties of the PCA method for assembly detection, and propose critical modifications of the current framework. First, we show that the number of assemblies and assembly neurons can be computed from the analysis of the eigenvalues of the neuronal correlation matrix. We then show that the subspace spanned by the PCs can reveal which neurons compose the detected assemblies. We go on to show how the time course of the activity of individual assemblies can then be estimated, even when different cell assemblies have a subset of common neurons. Finally, we show that our method can properly detect, track and specify the neuronal membership of neocortical and hippocampal assemblies recorded from behaving rats.

## Results

First we briefly outline the general framework as proposed in [31,32,34,35]. Figure 1A shows an example of neuronal population activity represented by means of a standard spike rastergram plot, in which each mark denotes the firing of an action potential by a neuron (the y-axis indicates the neuron labels). The procedure begins by binning the spike rastergram into non-overlapping, short time windows (referred to as bins) and counting the number of spikes in each bin, as indicated in Figure 1B. In this way, the rows of the resulting matrix represent neuronal units, and the columns represent the time bins. More specifically, the element  $a_{ij}$  denotes the number of spikes of the  $i$ th neuron in the  $j$ th bin (Figure 1B inset). For the sake of generality, in this work we use the “bin number” (“bin #”) as our arbitrary unit of time. Next, the binned spike activity is z-scored in order to normalize the spike rate of each neuron (Figure 1C). Thus, the rows of the normalized matrix are vectors with zero mean and unit variance. The autocorrelation matrix of the normalized spike activity is then computed (Figure 1D); each entry  $ij$  of the autocorrelation matrix is the Pearson correlation coefficient ( $r$ ) between the rows  $i$  and  $j$  of the matrix shown in Figure 1C (i.e., a correlation between two spike rate vectors).

The next steps of the method involve the computation of the eigenvalues of the autocorrelation matrix (Figure 1E) and the associated eigenvectors (Figure 1F), which in this context are referred to as Principal Components (PCs). Finally, the PCs associated with significant eigenvalues (see below) are used to track the activity of cell assemblies in each time bin (Figure 2A).

An important question is to know when the correlation coefficient of the spike activity of two neurons can be considered statistically significant for a given dataset. To this end, a statistical threshold that separates non-significant correlations from values above chance is needed. Instead of using exhaustive surrogate methods [37,38,39,40,41,42], Peyrache et al. elegantly addressed this problem by analyzing the distribution of the eigenvalues of the autocorrelation matrix [34,35]. From random matrix theory, it can be demonstrated that the eigenvalues of an autocorrelation matrix computed from a matrix with statistically independent rows (in our case, neurons with independent activity) follow the so-called Marčenko-Pastur distribution [43]. Since the goal is to identify ensemble activity, i.e. groups of neurons with correlated firing, the theoretical upper limit provided by the Marčenko-Pastur distribution can be used as statistical threshold. Thus, if there are groups of significantly correlated neurons in the population recorded, some eigenvalues will lie above this statistical

threshold. Furthermore, the PCs associated with significant eigenvalues can be used to track assembly activity. This is accomplished by projecting the normalized spike activity matrix using projector operators computed from the PCs, resulting in a unidimensional signal representing the time series of ensemble activity.

Using simulated data, we show in Figure 2A that the activation time course computed as described above is able to represent the activity of specific cell assemblies in some cases. However, as shown in Figure 2B, this approach is unable to separate the activity of individual assemblies when the neuronal population is composed of assemblies with overlapping cells. Note in Figure 2A that the estimated time courses of the activation strength correspond to increases of firing rate of specific subsets of neurons, as desired. However, for the case depicted in Figure 2B, the projection of population activity using the PCs does not separate the activity of the two cell assemblies. This constitutes an important limitation since the existence of assemblies with shared neurons is expected (see Discussion).

In the following sections we explore in more detail the general characteristics of this method, and propose modifications to allow tracking the activity of individual assemblies even when they share neurons. We also show that it is possible to precisely identify the neurons participating in each cell assembly.

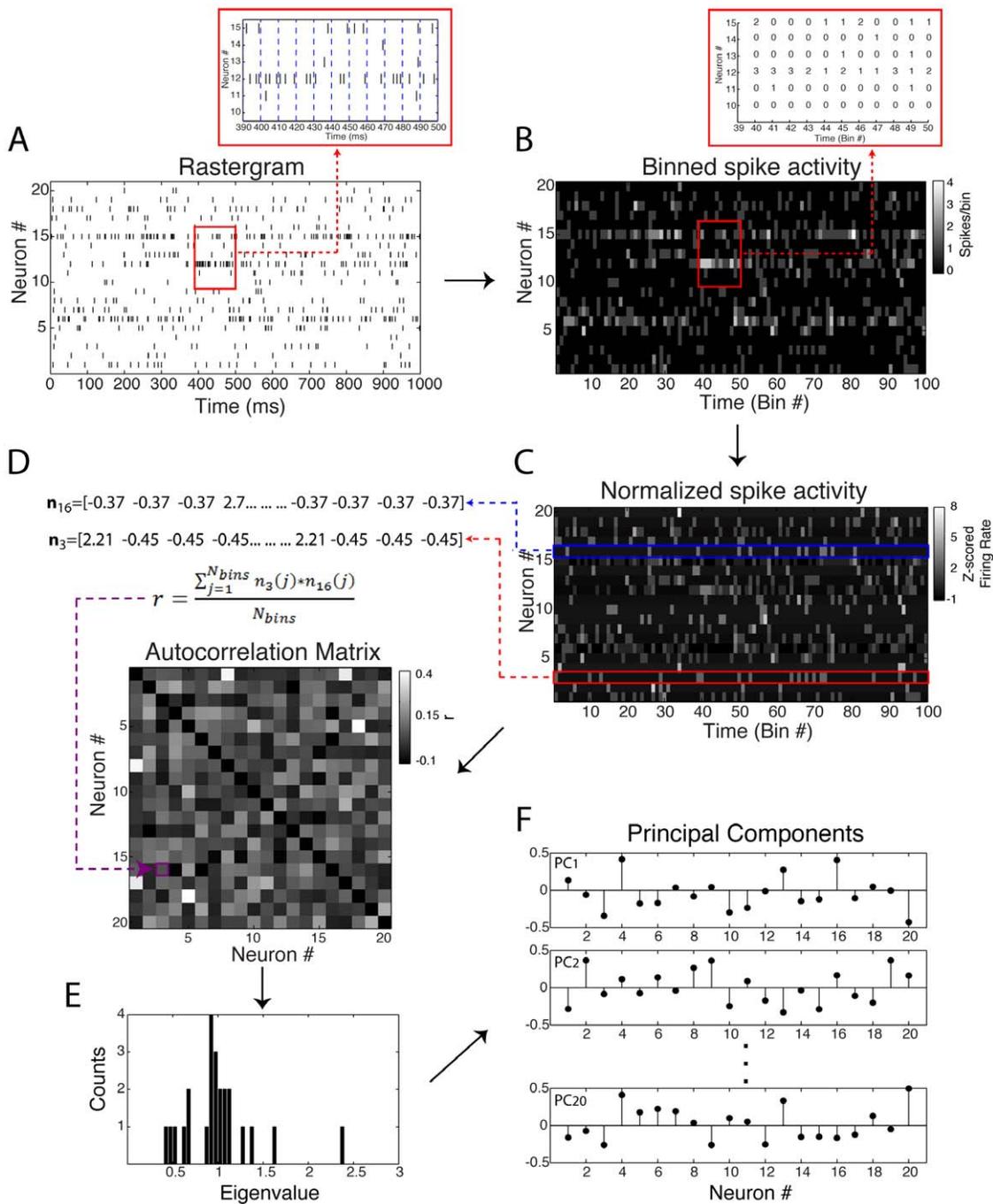
## Marčenko-Pastur distribution and the null hypothesis of independent neuronal activity

We start by exploring example cases of networks where no organized neuronal activity is present, that is, when there is no cell assembly in the network. As already introduced in the previous section, the eigenvalues of an autocorrelation matrix computed from a matrix with independent rows follow the Marčenko-Pastur distribution (see Methods for its formula). In order to illustrate this prediction, we show in Figure 3A–C three examples of random network activity differing in the number of neurons and time windows analyzed (i.e., the total number of bins). Each neuron is modeled as an independent Poissonian process (mean = 1 spike/bin). The predicted distribution of eigenvalues is shown below the corresponding network along with its empirical eigenvalues histogram. As expected, the actual eigenvalues follow the Marčenko-Pastur distribution. Note that the theoretical distribution has lower variance for greater values of the ratio  $q = \mathcal{N}_{\text{bins}}/\mathcal{N}_{\text{neurons}}$ .

We next performed a systematic parametric study of matrices with independent rows to investigate this property further. To this end, we defined “accuracy” as the percentage of eigenvalues that lie within theoretical bounds, that is, 100% accuracy means that all eigenvalues are within the limits predicted by the Marčenko-Pastur distribution. In other words, accuracy assesses the performance of the use of the theoretical bounds in determining the absence of cell assemblies in the network.

Figure 3D shows accuracy as a function of network size and total number of bins. Notice that, for a given network size, higher levels of accuracy are achieved with a higher number of time bins. In fact, as better seen in Figure 3E, accuracy is highly dependent on the condition  $q = \mathcal{N}_{\text{bins}}/\mathcal{N}_{\text{neurons}} > 1$ , i.e., the number of analyzed bins has to be greater than the number of neurons in the network. Figure 3F displays the results shown in Figure 3D for three specific network sizes. Similar results were obtained for different firing rates and also for the more realistic case in which the mean firing rate of each neuron differs from the mean rate of other neurons (data not shown). This latter result was expected since the firing rates are normalized.

Overall, we conclude that the theoretical limits predicted by the Marčenko-Pastur distribution can be used as the null hypothesis of

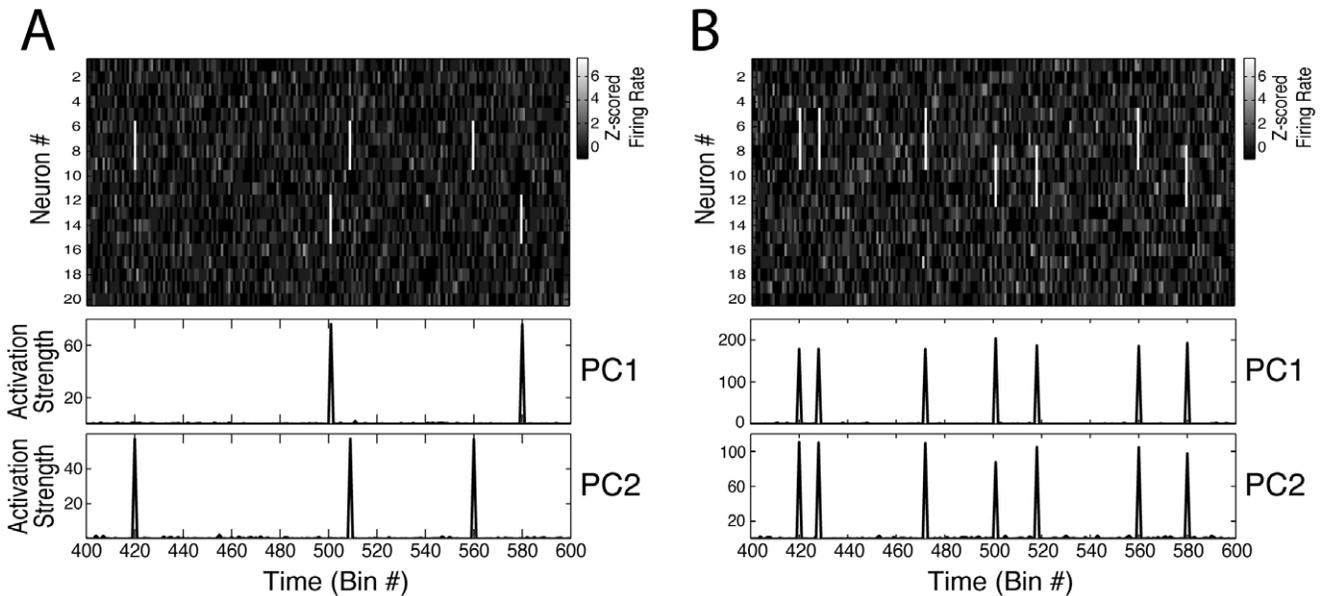


**Figure 1. Original method overview.** (A) Raster plot activity. Each row represents a neuron; marks denote an action potential and x-axis represents time. Panel inset shows the binning procedure into non-overlapping time windows. (B) Binned spike activity matrix obtained from raster plot in A. Each element is the count of the number of spikes in a given bin. (C) Z-scored binned spike activity matrix obtained by mean and variance normalization of the matrix in B. (D) Autocorrelation matrix (ACM) of the normalized binned spike activity in C. Each element denotes the linear correlation between two neurons. The main diagonal is set to zero for clearer visualization. (E) Eigenvalue histogram of the ACM shown in D. (F) Principal components (PCs) of the ACM, which are the eigenvectors associated with the eigenvalues shown in E. PCs are ordered in respect to their eigenvalues, i.e., the PC1 is associated with the highest eigenvalue and so on. doi:10.1371/journal.pone.0020996.g001

independent neuronal activity, as long as the number of bins analyzed is higher than the number of neurons in the network. In the next section, we show how we can also use this theoretical distribution to determine the precise number of cell assemblies in the network.

Eigenvalues outside theoretical bounds mark the number of cell assemblies and assembly neurons

We have shown above that eigenvalues of autocorrelation matrices computed from independent neuronal activity remain



**Figure 2. PCs do not always isolate the activity of different cell assemblies.** (A) Top panel shows a binned spike activity matrix with 20 neurons (modeled as Poissonian processes) and 8000 time bins. Two cell assemblies were simulated in the network, each having four neurons (Assembly 1 neurons: #6, #7, #8, #9; Assembly 2 neurons: #12, #13, #14, #15). Neurons in the same assembly were set to fire together six times above their mean firing rate at 0.5% of the bins. Bottom panels show the estimated time course of ensemble activity obtained by the projection of the binned spike activity using the projector operator defined as the outer product of the PCs (see Methods). Note that PC1 marks the activations of Assembly 2, and PC2 marks the activations of Assembly 1. (B) Same as A, but with assemblies sharing neurons (Assembly 1 neurons: #5, #6, #7, #8, #9; Assembly 2 neurons: #8, #9, #10, #11, #12). Note that for this example this framework fails to isolate the activity of individual assemblies. doi:10.1371/journal.pone.0020996.g002

within predicted limits as long as the condition  $q > 1$  is satisfied. Now we go further to show that the number of eigenvalues above the theoretical upper limit not only indicates the presence of ensemble activity, but it is also an accurate estimation of the number of cell assemblies in the network.

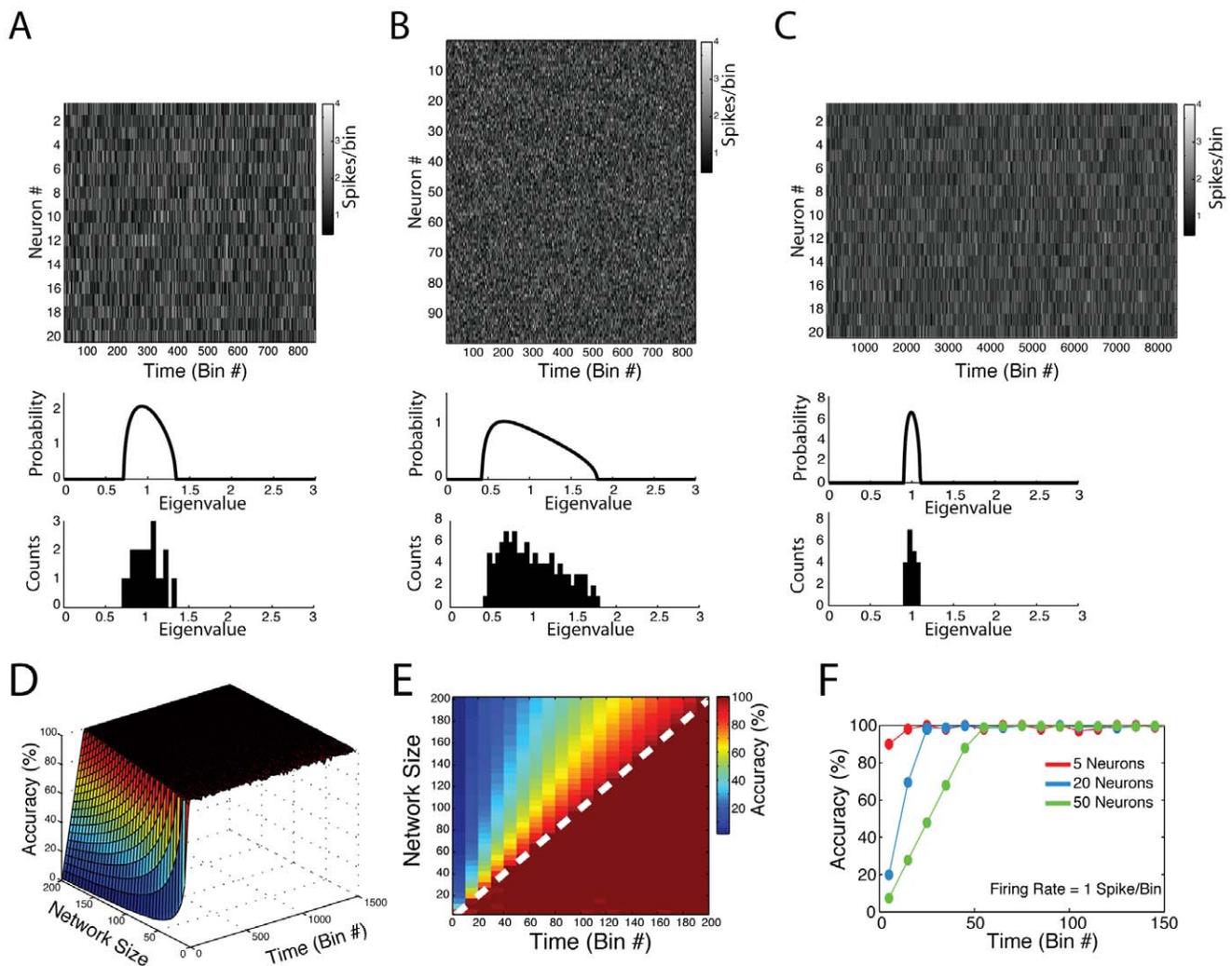
In Figure 4A two examples of neuronal network activity are shown. Neurons were modeled as Poissonian processes as in Figure 3, but, in addition, simulated assembly activity was added to the network. Assembly activations were modeled as an increase of the firing rates of a subset of neurons in specific bins. More specifically, in these “activation bins”, neurons were set to fire between 6 and 9 spikes, uniformly distributed. Both examples have 32 neurons and 8000 time bins. In each example, the mean firing rates (over all time bins) of cell assembly neurons were not necessarily higher than those of the other neurons in the network (Figure 4A, leftmost panels). In other words, the specific bins of assembly activation did not lead to a considerable net change in the average spike frequency of these neurons. In Figure 4A we depict a period of 150 bins in which assembly activations can be seen (second panels from left), along with the autocorrelation matrix of the simulated network (third panels from left); the theoretical eigenvalue distribution and the empirical eigenvalue histogram are also shown (top and bottom rightmost panels, respectively).

In the first example, a cell assembly with four neurons (neurons #7, #8, #9 and #10) is present in the network. Neurons have independent activity, with the exception of the cell assembly neurons that have higher firing rate in 0.5% of the bins randomly selected (i.e., the activation bins; cell assembly neurons have independent activity in the other bins). A simple visual inspection of the autocorrelation matrix already reveals higher correlations among cell assembly neurons. Importantly, notice that one eigenvalue of the empirical distribution lies above the upper limit predicted for independent neuronal activity in this example. In the

second example, three cell assemblies were added to the network. Notice that three eigenvalues fall above the theoretical upper limit in this case. These results therefore suggest that the number of eigenvalues above the Marcenko-Pastur distribution mark the number of cell assemblies in the network. We next performed a parametric analysis to investigate in more detail such property.

In Figure 4B, we analyze networks with different numbers of assemblies and different firing rates during activation bins (“activation firing rate”). We simulated networks with 40 neurons (mean spike rate = 1 spike/bin) and 8000 time bins; assemblies were composed by 4 neurons and set to be active in 0.5% of the bins. Each data point in Figure 4B corresponds to a network with a given level of activation firing rate (labeled by colors) and number of assemblies (varying from 1 to 10, as indicated in the x-axis). The number of eigenvalues above the theoretical upper limit is plotted as a function of the number of assemblies for different activation firing rates. Note that a perfect match between the number of eigenvalues above the upper limit and the number of assemblies in the network is indicated by  $\beta = 1$  in the linear fit  $y = \alpha + \beta x$ . We found that the number of eigenvalues above the upper limit underestimated the number of assemblies in the network ( $\beta < 1$ ) in cases in which assembly activations had a firing rate below 5; on the other hand, all cases with activation firing rate equal or above 5 presented a perfect match ( $\beta = 1$ ). Taken together, these results suggest that a minimal activation firing rate is required for the proper detection of the number of assemblies.

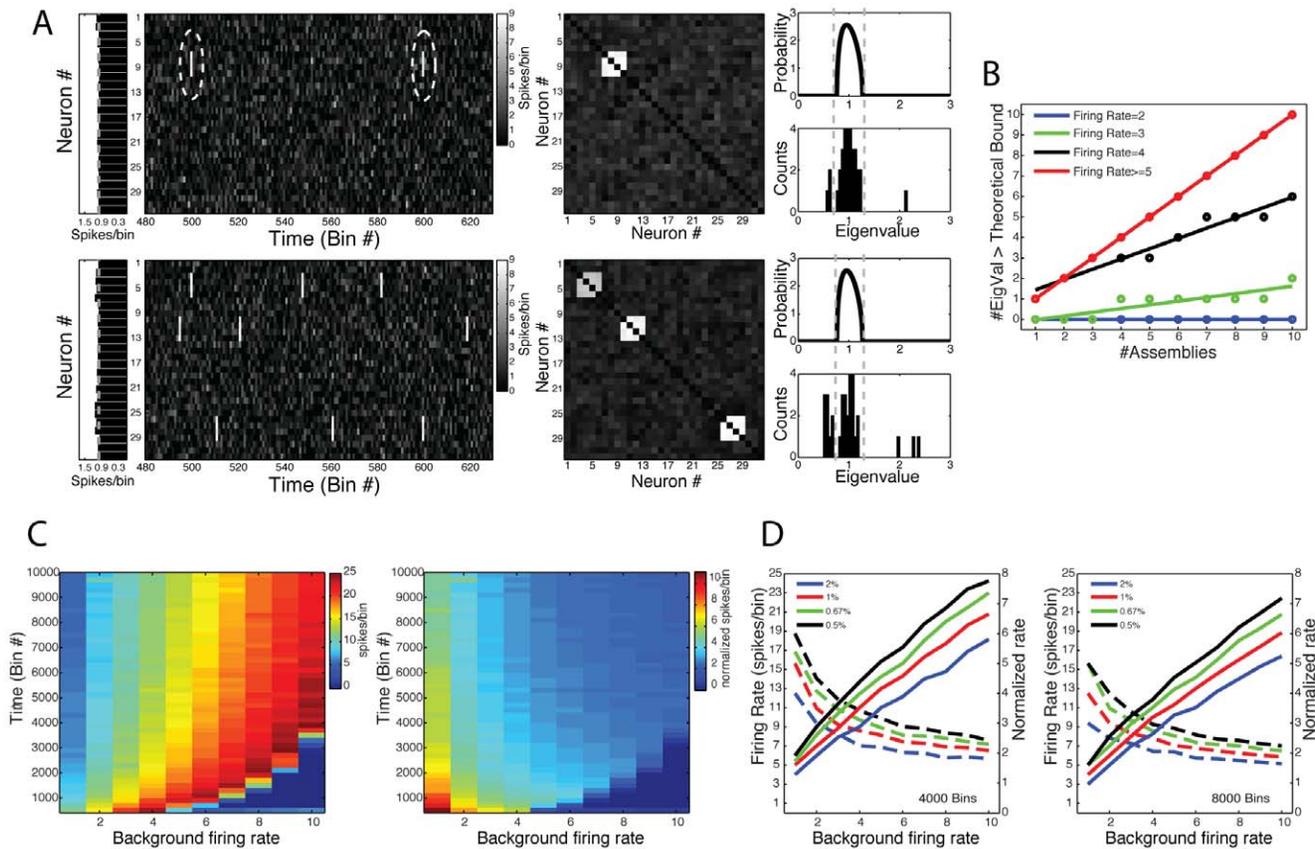
Our next step was to conduct exhaustive simulations to investigate the requirements for reaching the criterion  $\beta = 1$ . Figure 4C shows the minimal assembly activation firing rate required to achieve such criterion as a function of background firing rate and number of analyzed bins. In the left panel, the minimal activation firing rate is shown in absolute values, while in the right panel it is expressed as a ratio relative to the background firing rate.



**Figure 3. Eigenvalues of autocorrelation matrices derived from the activity of independent neurons fall within theoretical bounds.** (A) Top Panel: Binned spiking activity of 20 independent neurons. Each neuron was simulated as following a Poisson process (mean = 1 spike/bin). Middle Panel: Theoretical Marčenko-Pastur distribution. Bottom Panel: Histogram of eigenvalues obtained from the autocorrelation matrix computed from the neuronal activity shown in the top panel. (B,C) Similar panels as in A but for network activities presenting a greater number of neurons (B) or bins (C). Notice that the eigenvalues follow the Marčenko-Pastur distribution in all cases, and that the width of the predicted distribution is dependent on the ratio  $N_{neurons}/N_{bins}$ , where  $N_{neurons}$  = number of neurons and  $N_{bins}$  = number of bins. (D) Percentage of eigenvalues falling within Marčenko-Pastur theoretical bounds as a function of network size and number of time bins. For each parameter set, neurons were simulated as independent Poisson processes (mean = 1 spike/bin). Values represent the mean over 20 simulations. (E) Top-down view of the surface in D. Notice that virtually 100% accuracy occurs when  $N_{bins} > N_{neurons}$ . Dashed white line denotes  $N_{bins} = N_{neurons}$ . (F) Transactions of the surface in D obtained for three different network sizes.  
doi:10.1371/journal.pone.0020996.g003

Note that for a higher number of bins analyzed, a lower activation firing rate is required for a perfect match between the number of assemblies and of the eigenvalues above the upper limit. Figure 4D illustrates the dependence of  $\beta = 1$  on the number of assembly activation events. We studied network activities with 4000 (Figure 4D left panel) and 8000 (Figure 4D right panel) bins for four different “activation frequencies” (number of activation bins/number of time bins), and we show the minimal activation firing rate for  $\beta = 1$  as a function of background activity. Notice that, as the activation frequency gets higher, lower assembly activation firing rates are sufficient for  $\beta = 1$ . Overall, these simulations show that the number of eigenvalues above the theoretical bound is related to the number of assemblies present in the network. The efficiency of such estimation depends on how many bins the assembly neurons are correlated and how high this correlation is.

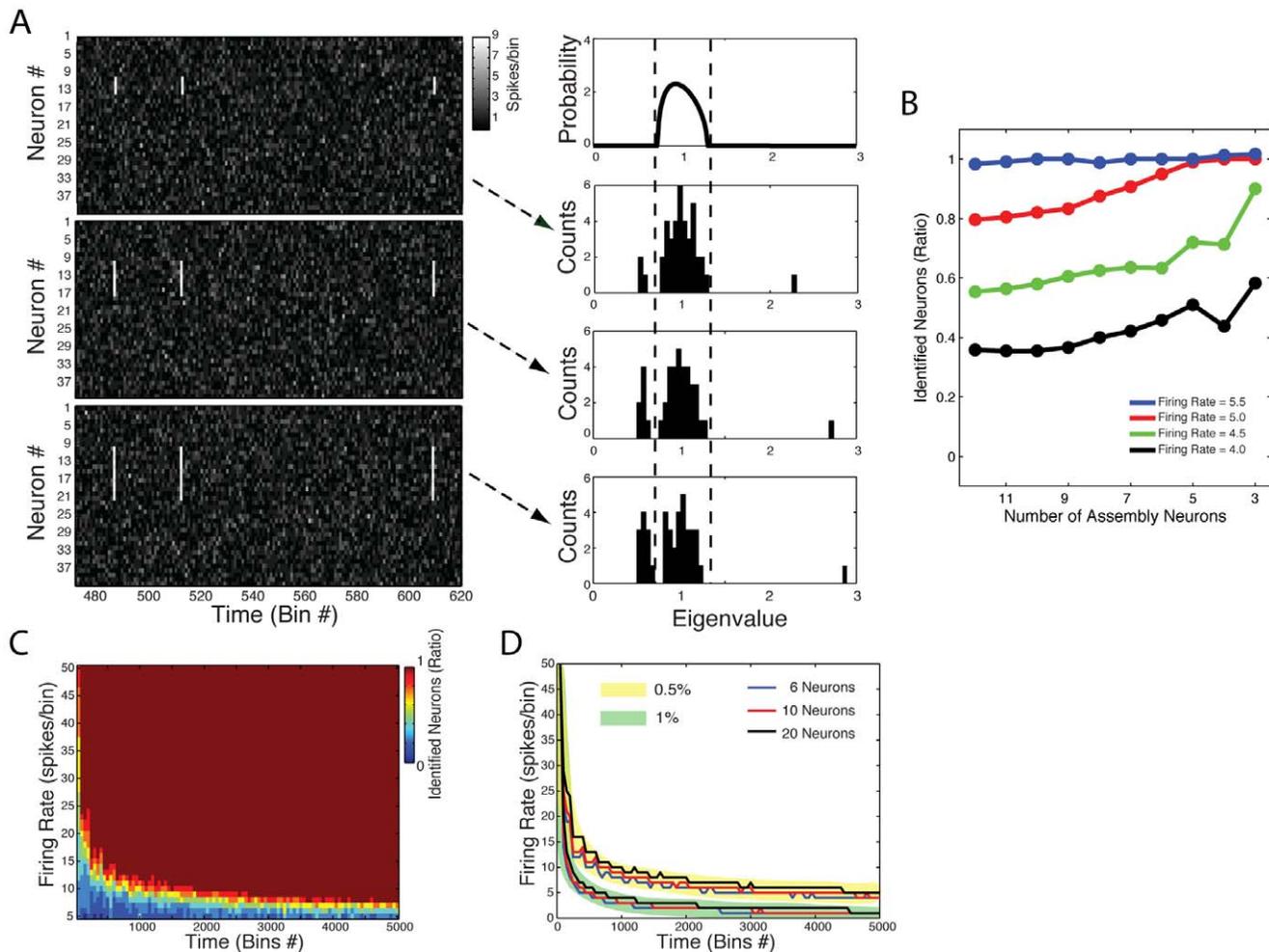
Next, we studied the eigenvalues that fall below the lower theoretical bound. Inspection of Figure 4A suggests that the number of eigenvalues below the predicted limit for independent activity increases when more assemblies are added to the network. In Figure 5, we show that, in fact, the total number of eigenvalues outside the theoretical distribution (below or above) is a good estimation of the total number of neurons involved in ensemble activity. More specifically, in Figure 5A we show three examples of networks with 40 simulated neurons and 8000 analyzed bins. A cell assembly was added to the network (active in 0.5% of the bins) and the number of neurons composing the ensemble was varied (4, 8 and 12 assembly neurons from top to bottom panels). The eigenvalue histograms shown in Figure 5A indicate that the number of eigenvalues below the predicted limit increases with increasing the number of cell assembly neurons; in fact, for the 3



**Figure 4. Eigenvalues above theoretical bound mark the number of cell assemblies.** (A) Top: (Left panels) Shown are binned spiking activity of a network composed of 32 neurons (second panel), along with the average firing rate of each neuron (first panel). Total simulation time was 8000 bins; neurons were modeled as possessing a Poissonian firing rate (mean = 1 spike/bin). In order to simulate a cell assembly, we set a group of neurons to activate simultaneously at 0.5% of the bins (firing rate within activation events = 6–9 spikes/bin). To facilitate visual inspection, neighbor neurons were chosen as composing the cell assembly (neurons #7, #8, #9, #10; dashed circle). (Middle Panel) Network correlation matrix. Notice that a cluster of correlated activity corresponding to the cell assembly. (Right Panels) Theoretical eigenvalues distribution for independent neuronal activity (top panel), and the eigenvalues histogram computed from the simulated network (bottom panel). Notice that 1 eigenvalue lies above the theoretical upper limit predicted for random activity. Bottom: Same as above, but for a network presenting three cells assemblies (Cell assembly 1: neurons #3, #4, #5, #6; Cell assembly 2: #10, #11, #12, #13; Cell assembly 3: neurons #26, #27, #28, #29). Notice that three eigenvalues lie above the theoretical bound. (B) Number of eigenvalues above the theoretical bound as a function of the number of cell assemblies in the network for different values of firing rate during cell assembly activation events. Networks were composed of 40 neurons; neurons were simulated as Poissonian processes (background mean = 1 spike/bin). Total simulation time was 8000 bins; assembly activation frequency was set to 0.5% of the bins. Each cell assembly was composed by 4 neurons (non-overlapping). Colored lines denote the linear fit  $y = \alpha + \beta x$  for each activation firing rate. Notice that the higher the firing rate within activation bins, the higher the slope coefficient ( $\beta$ ). If the firing rate is high enough,  $\beta$  equals 1, which characterizes the regimes in which the number of eigenvalues perfectly corresponds to the number of cell assemblies. Each data point represents a single simulation result. (C) Pseudocolors denote the minimal firing rate within activation bins leading to  $\beta$  equal to 1 as a function of the background mean firing rate (x-axis) and total number of time bins (y-axis). Results are expressed as absolute values (left) and as a ratio relative to the background firing rate (right). Assembly activation frequency was set to 0.5% of the bins. Networks were composed by 40 neurons, and each cell assembly was composed by 10 neurons. For each parameter set, values represent the mean over 20 simulations. (D) Left panel: Black line represents a transection of the result in **C** for network activities of 4000 time-bins. Other colored lines represent equivalent results obtained for different frequencies of cell assembly activation, as labeled. Notice that the higher the frequency of cell assembly activation, the lower the minimal firing rate leading to  $\beta$  equal to 1. Colored dashed lines represent the same result but as a ratio to the background firing rate. Right panel: Similar results as before, but for a network activity composed of 8000 time-bins. doi:10.1371/journal.pone.0020996.g004

examples, the total number of eigenvalues outside the theoretical distribution perfectly matched the number of cell assembly neurons. In Figure 5B we show that this property depends on the assembly activation firing rate. We again used exhaustive simulations in order to assess the robustness of this estimation. Figure 5C shows the number of eigenvalues outside predicted limits as a function of assembly activation firing rate and analyzed bins; the result is expressed as a ratio of the number of the neurons composing the assembly ( $\#$  outer eigenvalues/ $\#$  assembly neurons). Note that a virtually perfect estimation (ratio = 1) is

approached as the activation firing rate and the number of analyzed bins increase. Figure 5D shows the minimal activation firing rate for ratio = 1 as a function of the number of analyzed bins. We show this relation for different assembly activation frequencies and for different assembly sizes. While the estimation does not depend significantly on the number of neurons in the assembly, it is improved if the assembly is active in more bins. Similar findings were obtained in networks composed by multiple assemblies, even when some neurons were shared by two or more assemblies (simulations not shown, but see Figures 6 and 7).



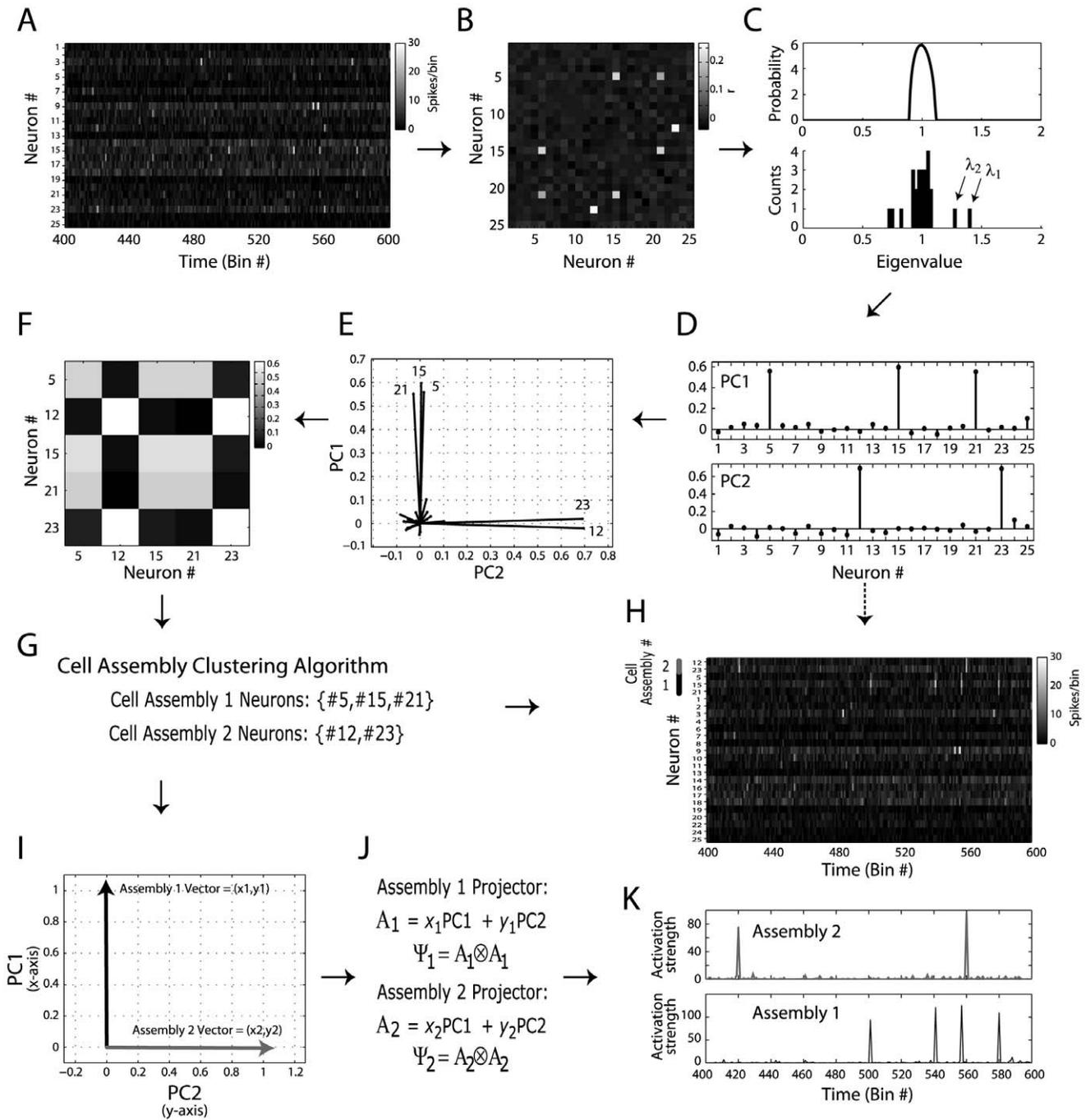
**Figure 5. The number of eigenvalues lying outside the theoretical distribution limits corresponds to the number of cell assembly neurons.** (A) Shown are the binned spiking activity matrices of networks composed of 40 neurons (left panels), along with the predicted eigenvalues distribution for independent neuronal activity (top right panel) and the actual eigenvalue histogram (bottom right panels). Total simulation time was 8000 bins; neurons were modeled as possessing a Poissonian firing rate (mean = 1 spike/bin). The 3 cases depicted differ in the number of neurons that compose the cell assembly. Notice that, for all cases, the number of eigenvalues outside the theoretical limits (dashed lines) matches the number of neurons in the cell assembly. (B) Ratio of the number of eigenvalues outside theoretical limits to the number of cell assembly neurons (ratio = 1 means that the number of significant eigenvalues perfectly corresponds to the number of cell assembly neurons). Different data points denote the mean over 20 simulations for different number of cell assembly neurons (x-axis) and activation firing rates (colored lines), as labeled. Networks were composed of 40 neurons; neurons were simulated as Poissonian processes (background mean = 1 spike/bin). Total simulation time was 8000 bins; assembly activation frequency was set to 1% of the bins. (C) Pseudocolors denote the ratio of the number of eigenvalues outside theoretical limits to the number of cell assembly neurons as a function of the activation firing rate and total time bins. Values represent the mean over 20 simulations. Networks were composed of 40 neurons; the cell assembly was made of 10 neurons set to activate at a frequency of 0.5% of the bins. Notice that for each activation firing rate, a perfect estimation of the number of cell assembly neurons (ratio = 1) is achieved if the number of bins analyzed is large enough. (D) Minimal activation firing rate required for a perfect match between the number of eigenvalues outside predicted limits and the number of cell assembly neurons as a function of the number of analyzed bins and in the frequency of cell assembly activation, as labeled. doi:10.1371/journal.pone.0020996.g005

In conclusion, we observed that the empirical distribution of eigenvalues not only indicates the presence of ensemble activity in the network, but can also be used to estimate the number of cell assemblies present in the network as well as the number of neurons involved in ensemble activity. In the next section we show how this information can be used to identify which neurons belong to each detected assembly.

### Identification of cell assemblies and time course of their activation

So far we have shown that eigenvalues of autocorrelation matrices that are higher than a well-established statistical threshold

have a strong relation with subsets of correlated neurons. Since these eigenvalues are by definition associated with PCs, it is reasonable to expect that these vectors also carry information about ensemble activity. In order to show how they can be used to identify assemblies in a network (in terms of which neurons compose them) we created a simulated network as an illustrative example. Neurons were again modeled as Poissonian processes, but with different mean spike rates (uniformly distributed between one and five spikes/bin). In addition, we set every neuron to fire 6 times above their mean rate at 0.5% of the bins. Two groups of neurons (cell assembly 1 neurons: #5, #15, #21; cell assembly 2 neurons: #12, #23) had these firing peaks at the same bins,



**Figure 6. Principal component-based analysis identifies cell assembly neurons and the time course of their activation.** (A) Binned spiking activity of a network composed of 25 neurons simulated for 8000 bins (200 bins shown). Neurons are modeled as Poissonian processes with random mean rate between 1 and 5 spikes/bin, uniformly distributed across the neurons. In addition, each neuron is set to fire at  $6 \times$  its mean rate at 0.5% of the bins randomly chosen (referred to as activation bins). In order to simulate cell assemblies, we set all activation bins to be independent, except for two groups of neurons which have simultaneous activation bins. (B) Autocorrelation matrix (ACM). (C) Theoretical eigenvalues distribution for independent neuronal activity (top), and the eigenvalues histogram computed from the simulated network (bottom). Note that 2 eigenvalues fall above the theoretical upper limit predicted for random activity, which correspond to the two cell assemblies present in the network. Notice further that three other eigenvalues fall below the lower limit; the number of eigenvalues outside the theoretical limits is therefore 5, which corresponds to the number of neurons participating in cell assemblies. (D) ACM eigenvectors associated with the two eigenvalues above the theoretical limit for random activity. These vectors are referred to as principal components (PCs). (E) Neuronal representations in the subspace spanned by the PCs (referred to as the Assembly Space). Since the PCs are the vectors which best describe strong correlated activity, neurons with larger projections on the Assembly Space are the neurons involved in cell assemblies (the label of these neurons are also shown). (F) Interaction Matrix; the entries of this matrix are measures of correlated activity of cell assembly neurons in the Assembly Space. Higher values denote neuron pairs pertaining to the same cell assembly, whereas lower values denote neurons whose activity is orthogonal. (G) From the Interaction Matrix, a simple clustering algorithm (described in Supplementary Information files) identifies the neurons of each cell assembly. (H) Same binned spiking activity as in A but rearranged in order to show cell assembly neurons on top, as labeled. (I, J) Assembly Vectors are defined as mean vectors in the Assembly Space (I); these vectors

are used to compute projector operators (**J**). (**K**) The projector operators are then applied to the binned spiking activity, revealing the time course of the activation strength of each cell assembly. Note that these results corroborate the activations seen by visual inspection of **H**. Since the cell assemblies were non-overlapping in this example, the identity of cell assembly neurons can be directly inferred by a simple analysis of the PCs (represented by the dashed line from **D** to **H**). However, such straight inference cannot be performed in cases where one or more neurons pertain to two or more assemblies (see Figure 7).  
doi:10.1371/journal.pone.0020996.g006

simulating assembly activations; non-assembly neurons had peak firing at independent (randomly chosen) bins.

Figure 6A shows a 200-bin interval of the simulated network; the associated autocorrelation matrix is shown in Figure 6B. Two eigenvalues of this matrix fall above the upper theoretical limit, whereas three eigenvalues lie below the lower bound (Figure 6C). This analysis therefore indicates that two assemblies and a total of five assembly neurons are present in the network, consistent with predefined simulation parameters.

Since eigenvalues above statistical threshold represent ensemble activity, we use the PCs associated with them (Figure 6D) to search for the identity of assembly neurons. The autocorrelation matrix can be seen as 25 vectors in a 25-dimensional space. In this case, PCA roughly means that the detected assembly activity is better described by the subspace spanned by the PCs; in the present work, we refer to this subspace as “Assembly space”. Removing the non-principal components of our analysis is equivalent to filtering the autocorrelation matrix in order to unravel assembly activity.

Figure 6E shows the neuron vectors on the Assembly space, which are obtained straight from the PC entries (see Methods). Note that some neurons present large vector length in this space, indicating that their spike activity is related to the detected assemblies. In fact, the five neurons with large vector length in the Assembly space (labeled in Figure 6E) correspond to the five units participating in assembly activity. Notice further that there are two clusters of neuron vectors in the Assembly space; these clusters are roughly orthogonal to each other, indicating independent activity. Indeed, notice that neurons orthogonal to each other pertain to different assemblies. Thus, we computed the length of the projection of each neuron vector onto the direction of the others and expressed these results in an “Interaction Matrix” (Figure 6F; see Methods). From the Interaction matrix, we used a simple clustering algorithm in order to determine which neurons were in each assembly (Figure 6G). Although the identification of assembly neurons was straightforward in this example from the visual inspection of Figure 6F, we noted that this was not always the case, making the use of a robust algorithm necessary (see Methods and Figures S1 and S2 for details about the algorithm). Figure 6H shows the same binned spike activity as in Figure 6A but with rows reordered with respect to the identified assemblies. Note that neurons within an assembly have firing peaks at the same bins.

The use of PCs was previously proposed in order to create projectors for computing ensemble activity with a single bin resolution [31,32,34,35]. An activity projector can be defined as the outer product of a PC with itself ([34]; see Methods for details). Since each PC represents an activity pattern, it is possible to compute the instantaneous strength of this pattern by multiplying the z-scored binned spike activity with the projector derived from the PC (see Methods). However, as shown in Figure 2, in some cases this method does not represent individual assemblies. To overcome this limitation, we propose another vector to construct the projectors. This vector, called “assembly vector”, is defined as the mean over all neuron vectors in the Assembly space that exclusively pertain to a given assembly (Figure 6I). Notice that the assembly vector is a linear combination of the PCs (Figure 6I,J), which allows obtaining this vector in the 25-dimensional space. By

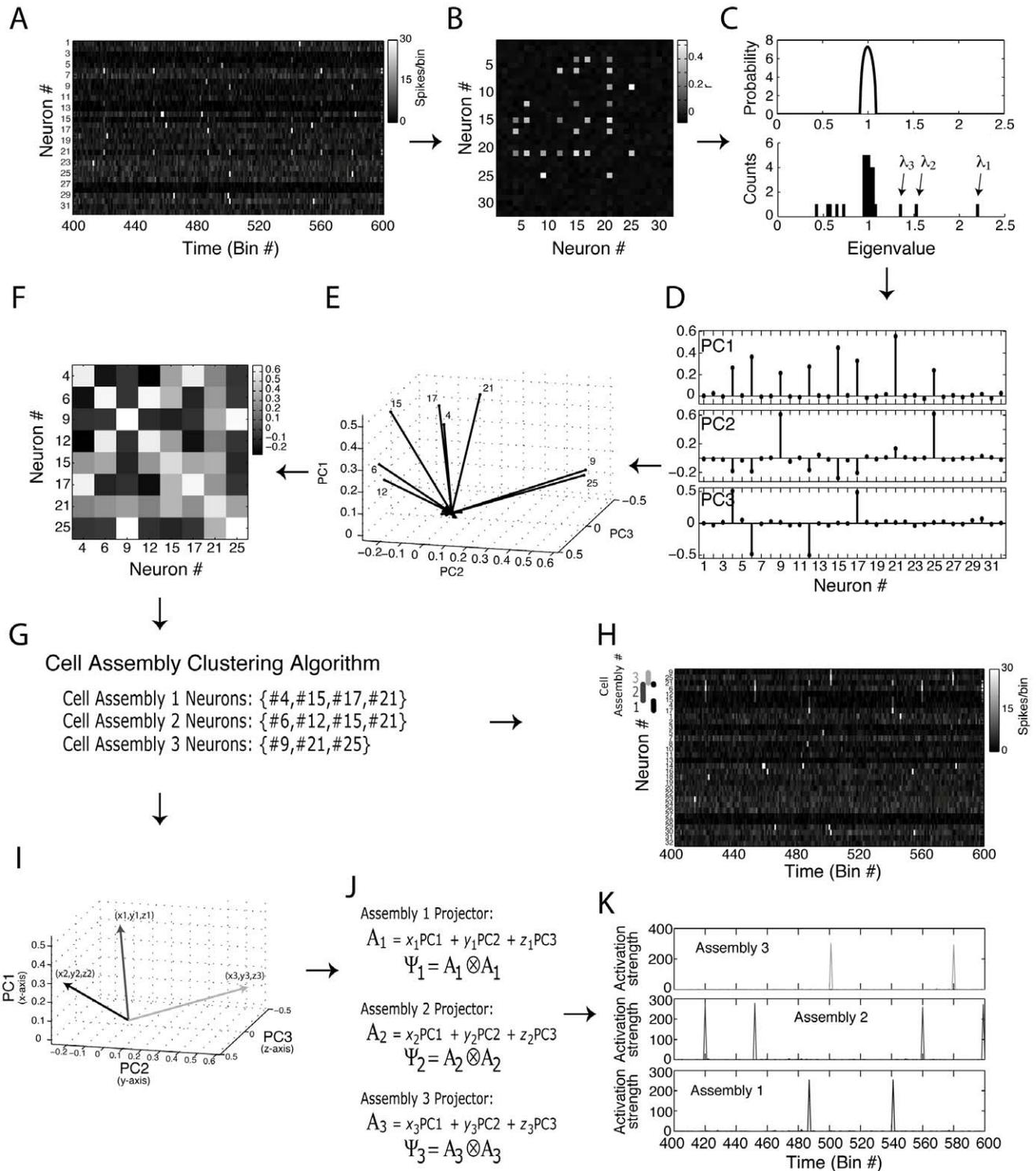
using this optimal assembly vector to construct the activity projector, we were then able to obtain the time course of the activity of the corresponding cell assembly. Figure 6K shows the results of such approach. For each assembly, the peaks of the time course matched perfectly the assembly activations seen in Figure 6H.

Note that in this example the PC weights directly reveal the neurons composing each assembly (Figure 6D). For instance, PC1 had higher values in dimensions 5, 15 and 21, which correspond to Cell Assembly 1 neurons; by the same token, the high values of PC2 denote Cell Assembly 2 neurons. Consequently, the estimated assembly optimal vectors in Figure 6I are very similar to the PCs and thus the activity projectors computed from the assembly vectors are virtually the same as the ones calculated from the PCs. As already mentioned (see Figure 2), the previous framework is able to track individual assembly activity when there are no overlapping neurons among the assemblies, as is the case of the example shown in Figure 6; therefore, our modified approach is equivalent to the original in these cases (see Figure S3).

In Figure 7 a more complex example is shown. The network activity was modeled as in Figure 6, but with three assemblies present in the network. Moreover, we simulated overlapping neurons between the assemblies (Assembly 1 neurons: #4, #15, #17, #21; Assembly 2 neurons: #6, #12, #15, #21; Assembly 3 neurons: #9, #21, #25). Figure 7C shows that 3 eigenvalues lie above the upper theoretical limit, denoting the three cell assemblies; moreover, the number of eigenvalues outside the theoretical limits matches the number of cell assembly neurons (8 in this example). Note in Figure 7D that it is no longer possible to identify the assemblies (in terms of which neurons compose them) by a visual inspection of the PC weights. Therefore, the estimation of the time course of assembly activity by computing the projectors from the PCs would be misleading in this case (see Figure S3).

As in the former example, projecting the neuron vectors on the Assembly space reveals the cell assembly neurons (Figure 7E). Note that the assembly neurons are not clearly clustered as in the example shown in Figure 6. While neurons that only pertain to the same assembly still tend to cluster together, neurons that participate in more than one assembly cannot be in two clusters simultaneously. For instance, projected neuron #15 is orthogonal to projected neurons #9 and #25. This is because neuron #15 does not compose the assembly in which neurons #9 and #25 participate. Conversely, neuron vector #15 is not orthogonal to any of the other neuron vectors, since they all participate in at least one assembly together with neuron #15. That is, overlapping neurons still have relatively large degree of collinearity with neurons that compose the same assemblies (Figure 7F). In this sense, since neuron #21 is in all assemblies, it is not orthogonal to any other assembly neuron.

As in the former example, pairwise relations between neurons in Assembly space can be inferred from the Interaction matrix. Notice however that in this case it is not straightforward to identify the cell assemblies by visual inspection of the Interaction Matrix (Figure 7F). Nevertheless, the clustering algorithm we developed (see Methods and Figures S1 and S2) was able to identify the precise composition of each assembly (Figure 7G). As before, after

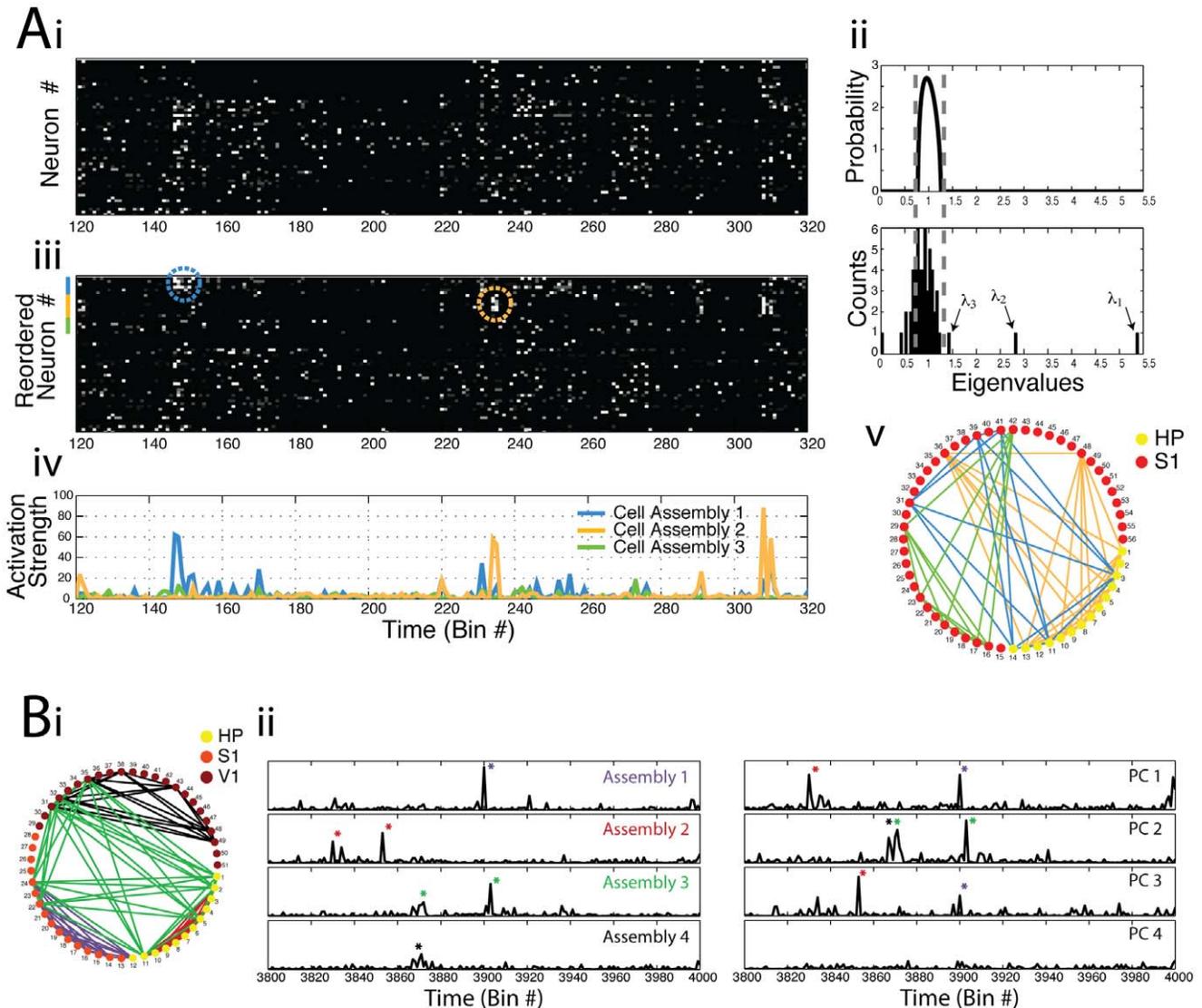


**Figure 7. Identification of cell assemblies with overlapping neurons.** (A–K) Same panels as in Figure 6, but for a network composed of three cell assemblies presenting common neurons. See text for further details. doi:10.1371/journal.pone.0020996.g007

identifying the assemblies we computed the optimal assembly vectors (Figure 7I,J) and used them to project the proper time course of assembly activations (Figure 7K; compare with Figure S3). This example therefore shows that the use of assembly vectors instead of PCs is better suited for computing assembly activity.

### Examples of applications to real data

So far we have used simulations to introduce a PCA-based method for cell assembly detection, providing details about how each step worked. In this section we apply the framework to real



**Figure 8. Example of cell assembly identification using principal components in an experimental data-set.** (A) Ai: Binned spiking activity for 14 hippocampal and 42 S1 neurons obtained from a rat during exploration of a novel object (see Ribeiro et al. [44]). Bin size = 30 ms; total time analyzed: 117.51 s. Aii: Theoretical eigenvalues distribution for independent neuronal activity (top) and the eigenvalues histogram computed from the actual network (bottom) exhibiting 3 eigenvalues above the theoretical upper limit predicted for random activity. Aiii: Same binned spiking activity as above, but with reordered rows such that neurons pertaining to cell assemblies are displayed in the top rows (color bars near the top of the y-axis mark cell assembly neurons; colored dashed circles highlight example periods of assembly activation). Aiv: Projection analysis yielding the activation time course for the three cell assemblies identified in this network (notice that cell assembly 3 does not activate in the period shown). Av: Graph diagram showing detected cell assemblies (connected neurons). Notice that inter-regional cell assemblies are revealed. (B) Bi: Graph diagram showing four assemblies detected in recordings from S1, V1 and hippocampus (HP) during slow-wave sleep (Bin size=30 ms; total time analyzed=124 s). Bii: Time course of ensemble activity as estimated by the original (right) and modified (left) framework.  
doi:10.1371/journal.pone.0020996.g008

data and further compare the modifications we propose with the original method.

We analyzed spike activity recorded from rats chronically implanted with multielectrode arrays (see Methods). In the first example (Figure 8A), neuronal activity was obtained from the hippocampus and primary somatosensory cortex (S1) during the exploration of novel objects [44]. In Figure 8Ai, we show a 200-bin period (bin size = 30 ms) of spike activity of this network; Figure 8Aii shows the Marčenko-Pastur distribution along with the empirical eigenvalue distribution computed from the associated autocorrelation matrix. Note the detection of three assemblies in this example. We then applied the framework described above,

and in Figure 8Aiii we plot the reordered spike activity with respect to the assemblies; dashed circles depict two examples of assembly activations occurring in the time period displayed. Figure 8Aiv shows the time course of activation of the detected assemblies; notice that the activation peaks match the activations seen in Figure 8Aiii. Finally, in Figure 8Av we show all neurons in a circular grid (hippocampal neurons: #1–14; S1 neurons: #15–56) and represent the assemblies by colored lines. Notice that our modified method allows us to infer that two assemblies have neurons in both brain regions.

In the second example (Figure 8B), we analyzed neurons recorded from the hippocampus, S1 and primary visual cortex

(V1) during slow-wave sleep (hippocampal neurons: #1–12; S1 neurons: #13–28; V1 neurons: #29–51). Analysis of the eigenvalues revealed that 4 cell assemblies were present in this network (not shown). We again applied our framework to get to the precise identity of the assembly neurons and depict the four assemblies in Figure 8Bi. Notice in this panel that one assembly was composed by neurons from the three brain areas, whereas three other assemblies were restricted to a single brain region. Notice further in this example that some neurons participate in two assemblies. We then compared the time course of ensemble activity when the PCs were used to build the projectors with projectors derived from assembly optimal vectors. The left and right panels in Figure 8Bii show the activity time course estimated by the assembly vector and by the direct use of the PCs, respectively. Note that the individual assembly activations estimated by the assembly vector approach appear mixed in different PC projections. For instance, the PC1 projection carries mixed activations of Assemblies 1 and 2, whereas PC2 carries information about Assemblies 3 and 4. Based on these results and the simulations presented above, we conclude that the use of assembly vectors to compute the activity time course is well suited for discriminating the activation of individual assemblies, even in the presence of overlapping neurons.

## Discussion

We have presented a mathematical method for the identification of cell assemblies and for computing their activity as a function of time (in units of time bins). The overall algorithm is based on PCA and can be divided in three major steps: **(1) Detection of the number of cell assemblies and assembly neurons;** **(2) Identification of cell assemblies;** and **(3) Computation of assembly activity as a function of time.** The algorithm presented here constitutes an extension of powerful methods introduced previously [31,32,34]. The adaptations and extensions we propose make our framework able to circumvent important limitations present in former methods.

With respect to step 1, Peyrache et al. introduced the use of the Marčenko-Pastur distribution as the null hypothesis to determine the presence of ensemble activity [34]. This is an important achievement in terms of computational cost because most of the previous methods relied on surrogate data analyses to determine statistical significance [26,30,45,46,47,48,49]. Moreover, other methods are only feasible for a small number of neurons [29,30,50,51] or only analyze pairwise correlations [45,52,53], making the analysis of large networks troublesome. The framework presented here inherits the computational advantages of the method envisioned by Peyrache et al. [34]. Additionally, it provides a clear interpretation for the eigenvalues derived from autocorrelation matrices of neuronal spike activity and their relation to the Marčenko-Pastur distribution: we showed that the number of eigenvalues significantly different from the random distribution contain useful information about the number of assemblies and the number of neurons participating in cell assemblies. This constitutes step 2 in our framework, which was not present in previous formulations.

Regarding step 3, the use of PCs in order to construct a time series of ensemble activity had already been introduced by Nicolelis et al. [31,32]. More recently, Peyrache et al. [34] proposed the use of projectors computed from the PC vectors associated with significant eigenvalues to extract *patterns of neural activity* from a defined *template* epoch to be later assessed in a *match* epoch. Peyrache and colleagues used this approach to obtain ensemble activation signatures from spike activity of medial prefrontal cortex (mPfc) neurons during a learning stage (*template*

epoch). Next they used these operators to measure instantaneous similarities (i.e., activations) of mPfc activity during a subsequent slow-wave sleep epoch (*match* epoch). It was found that (re)activations occurred preferentially during sharp wave/ripple complexes in post experience episodes, but not during previous sleep phases [35]. In another recent study, Benchenane et al. [36] reported that Pfc ensemble activations occur preferentially during periods of high theta coherence between the hippocampus and Pfc in a Y-maze task, which tended to occur during the decision point. These remarkable findings demonstrate that the use of PCA to estimate ensemble activity is a powerful tool to study network functioning. However, as illustrated in the present report, the framework applied in previous studies [32,33,35,36] possibly merges the activity of multiple cell assemblies into a single activity pattern. In this sense, the extension of the method now introduced allows for the isolation of the activity patterns of distinct groups of neurons. We believe that sorting out the individual activity of different assemblies will provide important insights in future studies.

While the studies mentioned above have focused on a template matching approach, the results shown in Figure 8 were obtained by first identifying all cell assemblies present in the network and subsequently assessing their activity time course in the same time period used to identify them; notice therefore that the method can be employed in different ways. One should however be cautious to avoid potential spurious results derived from circular analysis [54] when using the template-match approach. For example, it will likely happen that assembly activity during the template epoch (in which the assemblies are defined) is higher than that of any other epoch not used for computing the activity projectors; therefore, we believe one should not make quantitative inferences about assembly activity during the template epoch compared to other epochs.

It is important to emphasize that the PCA-based method is not sensitive to sequences of neuronal activity, such as *synfire chains* [27,55,56]. As pointed in Peyrache et al. [34], the statistical difficulties accompanying methods that look for firing sequences are overwhelming when one needs to analyze larger networks [47,49]. In fact, a common strategy to bypass combinatorial explosion (the number of possible temporal patterns is larger than the number of samples) is to detect ensemble activity disregarding the precise identity of the cell assemblies [57,58,59]. It is also important to note that only a tiny fraction of the neurons in the brain is observable, and therefore *synfire chains* are likely the effect of underlying sequences of cell assemblies, also known as Hebb's phase sequences [10]. The assessment of assembly sequences can be potentially achieved by the use of our method in combination with methods for detecting sequential activations [26,50,53].

It is also important to consider that the bin size used for the analyses can be critical for the interpretation of the results. As recently noted [12], bin sizes up to 30 ms are potentially well suited to analyze assembly activations. For instance, the typical membrane integration time in the waking cerebral cortex is estimated to vary between 10 and 30 ms [60,61]. Moreover, previous work has shown that neuronal members of a putative cell assembly tend to synchronize transiently in time windows of approximately 25 ms [14,62]. Interestingly, the time window for spike timing dependent plasticity is also consistent with this time-scale [63,64,65]. Finally, this time-scale corresponds to the period of gamma oscillations, which are believed to play a key role in binding representations coded by transiently active cell assemblies [66].

The novel framework described here allows the study of cell assemblies with shared neurons. The importance of this achievement is related to how information is processed and stored in the

brain. Some authors suggest that each neuron would only fire to a specific concept or stimulus (grandmother cells) [67]; therefore, cell assemblies encoding different “things” would not be expected to share neurons. However, a mounting body of work shows that neurons can be very selective (*sparse* coding), but are not grandmother cells [68,69]. The apparent grandmother cells in the human medial temporal lobe [70] may actually respond to between 50 and 150 distinct concepts [71]. Neurons participating in the representation of multiple concepts imply that the processing of information is distributed and occurs through a *multiplexed* code, in which concepts are represented by the activity of partially-overlapping groups of neurons, as postulated by Hebb.

Despite the worldwide acceptance of the cell assembly theory, there is still a paucity of evidence corroborating (or disproving) it. Hebb’s hypotheses not only deal with the formation of assemblies and phase sequences, but also constitute a complete theory describing how learning, fear, hunger, and other complex behaviors emerge from the brain [1]. Most of the difficulty in testing the theory resides in the fact that only a tiny fraction of neurons in the brain can be simultaneously recorded at any given time. However, techniques for massive neuronal recordings are being developed at accelerating rates [22], and while we still lack proper tools for analyzing large quantities of neurons [57,72], much progress is being made to circumvent this limitation. We hope the work presented here constitutes a useful step in this direction.

## Methods

Simulations and data analyses were programmed in MATLAB (The Mathworks, Inc); MATLAB codes for the computation of cell assemblies and their dynamics can be obtained from the authors upon request.

### Analytical formula of the Marčenko-Pastur distribution

The spectrum of eigenvalues of an autocorrelation matrix computed from a random matrix  $\mathbf{M}$  of  $N_{bins}$  columns and  $N_{neurons}$  rows follow the Marčenko-Pastur distribution, which in the limit of  $N_{bins} \rightarrow \infty$  and  $N_{neurons} \rightarrow \infty$ , with  $q = N_{bins}/N_{neurons} \geq 1$  constant, is given by

$$p(\lambda) = \frac{q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda_{min} - \lambda)}}{\lambda}$$

where  $\sigma$  is the standard deviation of the elements of  $\mathbf{M}$  (in our case, we have  $\sigma = 1$  since we apply the z-score normalization to the binned spike activity);  $\lambda_{max}$  and  $\lambda_{min}$  are the upper and lower limits of the Marčenko-Pastur distribution, and they are given by:

$$\lambda_{min}^{max} = \sigma^2(1 \pm \sqrt{1/q})^2$$

Notice that  $\lambda_{max}$  and  $\lambda_{min}$  converge to 1 when  $q \rightarrow \infty$  and in this limit the theoretical distribution becomes a Dirac delta function at  $\lambda = 1$ . Therefore, the predicted eigenvalues distribution for independent neuronal activity has lower variance when a greater number of time bins are analyzed for a given number of neurons (compare Figure 3A and 3C).

We note that even though the analytical formula for the Marčenko-Pastur distribution was derived in the limit case of large  $N_{bins}$  and  $N_{neurons}$ , this theoretical distribution also approximates the actual distribution in cases of finite matrices, as shown in Plerou et al. [73] and in the present work. Nevertheless, one can

also make use of the bias correction for finite size matrices suggested in [74]. The upper theoretical limit then becomes  $\lambda_{max} + N_{neurons}^{-2/3}$ . We found however that this correction did not influence the results shown in the present work.

### Outer product and the definition of the activity projector operators

The outer product of two vectors  $\mathbf{u}$  and  $\mathbf{v}$  of length  $N$  is defined as

$$\mathbf{u} \otimes \mathbf{v} = \begin{pmatrix} u_1 v_1 & \cdots & u_1 v_N \\ \vdots & \ddots & \vdots \\ u_N v_1 & \cdots & u_N v_N \end{pmatrix}$$

The outer product is used to construct the projectors of ensemble activity, as explained in the following. Let  $\mathbf{C}$  be the autocorrelation matrix of a z-scored binned network activity  $\mathbf{Z}$  of dimension  $N_{neurons} \times N_{bins}$ , and let  $\mathbf{p}_i$  ( $i = 1, 2, \dots, N_{neurons}$ ) denote the principal components of  $\mathbf{C}$ . The projector  $\mathbf{P}_i$  associated with  $\mathbf{p}_i$  is given by

$$\mathbf{P}_i = \mathbf{p}_i \otimes \mathbf{p}_i$$

If  $\lambda_i$  is the eigenvalue associated with the principal component  $\mathbf{p}_i$ ,  $\mathbf{C}$  can be decomposed as

$$\mathbf{C} = \lambda_1 \mathbf{P}_1 + \lambda_2 \mathbf{P}_2 + \dots + \lambda_{N_{neurons}} \mathbf{P}_{N_{neurons}} = \sum_{i=1}^{N_{neurons}} \lambda_i \mathbf{P}_i$$

Assuming that each principal component  $\mathbf{p}_i$  represents an ensemble co-activation pattern, the equation above shows that  $\mathbf{C}$  can be represented by a linear combination of the pattern representations encoded in the matrices  $\mathbf{P}_i$ .

### Assembly activity time-course

Peyrache et al. [34] has recently proposed the use of the principal components associated with significant eigenvalues for assessing ensemble activity with a single-bin resolution. The idea is to calculate the instantaneous similarity of the binned spike activity and the ensemble activity pattern as a function of time.

Let  $\mathbf{P}$  be outer product of a significant principal component with itself and  $\mathbf{Z}(b)$  be the  $b$ -th column of the z-scored binned spike activity (in other words, the number of spikes of all neurons in the  $b$ -th bin). The measure of instantaneous similarity of  $\mathbf{P}$  and  $\mathbf{Z}$  as a function of time is given by

$$R(b) = \mathbf{Z}(b)^T \mathbf{P} \mathbf{Z}(b)$$

This equation can be rewritten as

$$R(b) = \sum_{i,j} Z_{ib} P_{ij} Z_{jb}$$

where  $Z_{ib}$  is the normalized firing rate of neuron  $i$  in bin  $b$ , and  $P_{ij}$  is the entry in the  $i$ -th row and  $j$ -th column of  $\mathbf{P}$ . Note that when  $i = j$ , the corresponding term of the summation only takes into account the activity of a single neuron  $i$ . Since our goal is to measure ensemble activity more than single neuron activations, this term can be set to zero and the equation reduces to

$$R(b) = \sum_{i,j,i \neq j} Z_{ib} P_{ij} Z_{jb}$$

which is the equation for computing the time course of ensemble activity used in Peyrache et al. [34].

As we show in the present work (Figure 2 and S3), projectors computed as above are not appropriate to track the activity time course of individual cell assemblies if there are overlapping neurons among assemblies. To overcome this problem, we propose constructing the projectors using the optimal assembly vectors in Assembly space (Figures 6 and 7). This is achieved as follows: The Assembly space is defined as the metric subspace spanned by the principal components  $\mathbf{p}_i$  associated with eigenvalues  $\lambda_i$  that are significantly above chance. Let  $\mathbf{a}_k$  ( $k = 1, \dots, N_{neurons}$ ) denote the neuron vectors in the Assembly space; each  $\mathbf{a}_k$  is given by (see Figure 6E):

$$\mathbf{a}_k = (\mathbf{p}_1(k), \mathbf{p}_2(k), \dots, \mathbf{p}_n(k))$$

where  $n$  is the number of significant eigenvalues. As we show in the present work, the number of eigenvalues outside the theoretical distribution gives the total number of neurons participating in cell assemblies. Supposing there are  $N_{CAneurons}$  assembly neurons, they correspond to the  $N_{CAneurons}$  vectors with largest norm (vector length) in Assembly space (Figures 6E and 7E). Then, the projections of each neuron vector in the Assembly space onto the direction of the other vectors are computed and used to build the Interaction Matrix (Figures 6F and 7F). That is, given two neuron vectors  $\mathbf{a}_i$  and  $\mathbf{a}_j$ , the corresponding  $(i,j)$  entry of the Interaction Matrix is given by  $(\mathbf{a}_i \cdot \mathbf{a}_j) / (\|\mathbf{a}_i\| \|\mathbf{a}_j\|)$ . From the Interaction Matrix, it is possible to determine which neurons compose each assembly by means of a clustering algorithm (see next section). The estimated optimal assembly vector  $\bar{\mathbf{a}}$  is then defined for an assembly A as the mean over  $\mathbf{a}_i$ 's for all neurons  $i$  exclusive to A, normalized to have unitary norm:

$$\bar{\mathbf{a}} = \frac{\sum_i \mathbf{a}_i}{|\sum_i \mathbf{a}_i|}$$

Next,  $\bar{\mathbf{a}}$  is expressed as a linear combination of the significant principal components:

$$\alpha = \sum_{i=1}^n (\bar{\mathbf{a}} \cdot \mathbf{p}_i) \mathbf{p}_i$$

A projector  $\Psi$  is then calculated as the outer product of  $\alpha$  with itself ( $\Psi = \alpha \otimes \alpha$ ). Finally, we use  $\Psi$  to compute the activity time course of assembly A as follows:

$$R_A(b) = \sum_{i,j,i \neq j} Z_{ib} \Psi_{ij} Z_{ib}$$

### Binary Interaction Matrix and clustering algorithm

The algorithm identifies the neurons pertaining to each cell assembly based on the analysis of the Interaction Matrix. The entries of the Interaction Matrix are a measure of correlation between two neuron vectors in Assembly space (taking into account only cell assembly neurons). As we have shown in Figures 6 and 7, neurons that pertain to different assemblies are orthogonal to each other, while high collinearity levels indicate that neurons are correlated in the Assembly space. Therefore, it is expected that the distribution of Interaction Matrix entries is bimodal, having sets of low and high values (see Figure S1). We then apply a uni-dimensional version of the K-means clustering algorithm [75] in order to find a threshold that best separate these groups. We use this threshold to create a binary Interaction

Matrix; that is, we transform all matrix values in 0's (values below a threshold) and 1's (values above the threshold). This binary matrix is the input to the clustering algorithm which is then able to sort apart the neurons of different assemblies. In Figure S1 we provide an overview of the thresholding procedure and in Figure S2 we describe the clustering algorithm.

### Electrophysiological recordings

Male Long-Evans rats were chronically implanted with tungsten microelectrode arrays aimed at the hippocampus, primary visual cortex and primary somatosensory cortex. Data recorded from these animals were described in a previous study [44], in which a detailed description of surgery, data collection, behavior and histology can be found.

### Supporting Information

**Figure S1** Interaction Matrix thresholding. (A) Ai: Interaction Matrix of the example shown in Figure 7. Aii: Histogram of the entries of the Interaction Matrix shown in Ai. Dashed red line indicates the threshold found by a K-means algorithm. The threshold is the mean between the borders of the clusters. Aiii: Binary Interaction Matrix. Values lower and higher than the threshold are set to 0 and 1, respectively. This matrix is later used as input to the clustering algorithm described in Figure S2. (B) Same as (A) but for the real data shown in Figure 8B. Note that the threshold found separates the bimodal distribution. (TIF)

**Figure S2** Description of the assembly clustering algorithm. (A,B) Flux diagram representing the three main steps of the algorithm (A) and an example using simulated data of nine neurons (B). The algorithm receives as input a Binary Interaction Matrix (BIM; depicted in B top panel), which is obtained by thresholding the Interaction Matrix (see Figure S1), and provides as output the assembly label(s) for each neuron (B bottom panel). Step 1 involves re-organizing the BIM according to the number of interactions in each row and also removing repeated rows; we denote the resulting matrix as the OBIM (B second panel from top). Notice in B that row #8 does not appear in the OBIM since it was equal to row #5. In Step 2 assembly labels are created and assigned to the neurons. This is achieved based on sequentially examining each row of OBIM and identifying for each neuron (row) all other neurons that interact with it; a common assembly label is ascribed to all interacting neurons. New assembly labels are created whenever the neuron (row) being processed has not been previously assigned to any of the existing assembly labels. This step generates the Assembly Label Matrix (ALM), which entry  $(i,j)$  informs the assemblies shared by neurons  $\#i$  and  $\#j$ . Notice that neuron #8 automatically appears in ALM under this procedure (B third panel from top). Finally, in Step 3 the assembly labels in the diagonal of ALM are extracted; they indicate the assemblies in which each neuron participates. (TIF)

**Figure S3** Estimation of time course of cell assembly activity based on individual PCs for the examples shown in Figures 6 (A) and 7 (B). The estimation of assembly activity based on assembly vectors is also reproduced from Figures 6 and 7 for comparison. (TIF)

### Author Contributions

Conceived and designed the experiments: ABLT. Performed the experiments: VL-d-S SC-O. Analyzed the data: VL-d-S SC-O ABLT STR MALN. Wrote the paper: ABLT VL-d-S STR.

## References

- Hebb DO (1949) *The Organization of Behavior*. New York: John Wiley & Sons.
- Wang Q, Chen G, Perc M (2011) Synchronous bursts on scale-free neuronal networks with attractive and repulsive coupling. *PLoS One* 6: e15851.
- Wang Q, Perc M, Duan Z, Chen G (2009) Synchronization transitions on scale-free neuronal networks due to finite information transmission delays. *Phys Rev E Stat Nonlin Soft Matter Phys* 80: 026206.
- Pikovsky A, Rosenblum M, Kurths J (2001) *Synchronization, A Universal Concept in Nonlinear Sciences* Cambridge University Press.
- Singer W (1993) Synchronization of cortical activity and its putative role in information processing and learning. *Annu Rev Physiol* 55: 349–374.
- Freiwald WA, Kreiter AK, Singer W (2001) Synchronization and assembly formation in the visual cortex. *Prog Brain Res* 130: 111–140.
- Volman V, Perc M (2010) Fast random rewiring and strong connectivity impair subthreshold signal detection in excitable networks. *New J Phys* 12.
- Marr D (1971) Simple memory: a theory for archicortex. *Philos Trans R Soc Lond B Biol Sci* 262: 23–81.
- Wallenstein GV, Eichenbaum H, Hasselmo ME (1998) The hippocampus as an associator of discontinuous events. *Trends Neurosci* 21: 317–323.
- Harris KD (2005) Neural signatures of cell assembly organization. *Nature Reviews Neuroscience* 6: 399–407.
- Pastalkova E, Itskov V, Amarasingham A, Buzsaki G (2008) Internally generated cell assembly sequences in the rat hippocampus. *Science* 321: 1322–1327.
- Buzsaki G (2010) Neural syntax: cell assemblies, synapse ensembles, and readers. *Neuron* 68: 362–385.
- Wilson MA, McNaughton BL (1993) Dynamics of the hippocampal ensemble code for space. *Science* 261: 1055–1058.
- Harris KD, Csicsvari J, Hirase H, Dragoi G, Buzsaki G (2003) Organization of cell assemblies in the hippocampus. *Nature* 424: 552–556.
- Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, et al. (2003) Learning to control a brain-machine interface for reaching and grasping by primates. *Plos Biology* 1: 193–208.
- Hung CP, Kreiman G, Poggio T, DiCarlo JJ (2005) Fast readout of object identity from macaque inferior temporal cortex. *Science* 310: 863–866.
- Nicoletis MA, Ghazanfar AA, Stambaugh CR, Oliveira LM, Laubach M, et al. (1998) Simultaneous encoding of tactile information by three primate cortical areas. *Nat Neurosci* 1: 621–630.
- Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, et al. (2000) Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* 408: 361–365.
- Nicoletis MA, Fanselow EE, Ghazanfar AA (1997) Hebb's dream: the resurgence of cell assemblies. *Neuron* 19: 219–221.
- Nicoletis MAL, De Schutter E (1999) Methods for recording and analyzing neuronal ensemble activity. *Journal of Neuroscience Methods* 94: 3–4.
- Buzsaki G (2004) Large-scale recording of neuronal ensembles. *Nature Neuroscience* 7: 446–451.
- Stevenson IH, Kording KP (2011) How advances in neural recording affect data analysis. *Nat Neurosci* 14: 139–142.
- Lee AK, Wilson MA (2002) Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* 36: 1183–1194.
- Ribeiro S, Gervasoni D, Soares ES, Zhou Y, Lin SC, et al. (2004) Long-lasting novelty-induced neuronal reverberation during slow-wave sleep in multiple forebrain areas. *Plos Biology* 2: 126–137.
- Louie K, Wilson MA (2001) Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron* 29: 145–156.
- Abeles M, Gat I (2001) Detecting precise firing sequences in experimental data. *Journal of Neuroscience Methods* 107: 141–154.
- Ikegaya Y, Aaron G, Cossart R, Aronov D, Lampl I, et al. (2004) Synfire chains and cortical songs: Temporal modules of cortical activity. *Science* 304: 559–564.
- Maldonado P, Babul C, Singer W, Rodriguez E, Berger D, et al. (2008) Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. *J Neurophysiol* 100: 1523–1532.
- Grun S, Diesmann M, Aertsen A (2002) Unitary events in multiple single-neuron spiking activity: 1. Detection and significance. *Neural Computation* 14: 43–80.
- Pipa G, Wheeler DW, Singer W, Nikolic D (2008) NeuroXidence: reliable and efficient analysis of an excess or deficiency of joint-spike events. *Journal of Computational Neuroscience* 25: 64–88.
- Chapin JK, Nicoletis MA (1999) Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *J Neurosci Methods* 94: 121–140.
- Nicoletis MAL, Baccala LA, Lin RCS, Chapin JK (1995) Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science* 268: 1353–1358.
- Humphries MD (2011) Spike-train communities: finding groups of similar spike trains. *J Neurosci* 31: 2321–2336.
- Peyrache A, Benchenane K, Khamassi M, Wiener SI, Battaglia FP (2010) Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution. *Journal of Computational Neuroscience* 29: 309–325.
- Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP (2009) Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nature Neuroscience* 12: 919–U143.
- Benchenane K, Peyrache A, Khamassi M, Tierney PL, Gioanni Y, et al. (2010) Coherent Theta Oscillations and Reorganization of Spike Timing in the Hippocampal-Prefrontal Network upon Learning. *Neuron* 66: 921–936.
- Theiler J, Eubank S, Longtin A, Galdrikian B, Doynne Farmer J (1992) Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena* 58: 77–94.
- Schreiber T, Schmitz A (1996) Improved Surrogate Data for Nonlinearity Tests. *Physical Review Letters* 77: 635.
- Perc M, Green AK, Dixon CJ, Marhl M (2008) Establishing the stochastic nature of intracellular calcium oscillations from experimental data. *Biophys Chem* 132: 33–38.
- Pradhan N, Sadasivan PK (1996) Relevance of surrogate-data testing in electroencephalogram analysis. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 53: 2684–2692.
- Khoa TQD, Thang HM, Nakagawa M (2008) Testing for Nonlinearity in Functional Near-Infrared Spectroscopy of Brain Activities by Surrogate Data Methods. *The Journal of Physiological Sciences* 58: 47–52.
- Schreiber T (1999) Interdisciplinary application of nonlinear time series methods. *Physics Reports* 308: 1–64.
- Marčenko VA, Pastur LA (1967) Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* 1: 457–483.
- Ribeiro S, Shi X, Engelhard M, Zhou Y, Zhang H, et al. (2007) Novel experience induces persistent sleep-dependent plasticity in the cortex but not in the hippocampus. *Front Neurosci* 1: 43–55.
- Schneidman E, Berry MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440: 1007–1012.
- Berger D, Warren D, Normann R, Arieli A, Grun S (2007) Spatially organized spike correlation in cat visual cortex. *Neurocomputing* 70: 2112–2116.
- Mokeichev A, Okun M, Barak O, Katz Y, Ben-Shahar O, et al. (2007) Stochastic emergence of repeating cortical motifs in spontaneous membrane potential fluctuations in vivo. *Neuron* 53: 413–425.
- Berger D, Borgelt C, Louis S, Morrison A, Grun S (2010) Efficient identification of assembly neurons within massively parallel spike trains. *Comput Intell Neurosci*: 439648.
- Grun S (2009) Data-Driven Significance Estimation for Precise Spike Correlation. *Journal of Neurophysiology* 101: 1126–1140.
- Tetko IV, Villa AEP (2001) A pattern grouping algorithm for analysis of spatiotemporal patterns in neuronal spike trains. 1. Detection of repeated patterns. *Journal of Neuroscience Methods* 105: 1–14.
- Abeles M, Gerstein GL (1988) Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *Journal of Neurophysiology* 60: 909–924.
- Wilson MA, McNaughton BL (1994) Reactivation of hippocampal ensemble memories during sleep. *Science* 265: 676–679.
- Shmiel T, Drori R, Shmiel O, Ben-Shaul Y, Nadasdy Z, et al. (2006) Temporally precise cortical firing patterns are associated with distinct action segments. *Journal of Neurophysiology* 96: 2645–2652.
- Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience* 12: 535–540.
- Prut Y, Vaadia E, Bergman H, Haalman I, Slovlin H, et al. (1998) Spatiotemporal structure of cortical activity: Properties and behavioral relevance. *Journal of Neurophysiology* 79: 2857–2874.
- Abeles M (1991) *Corticonics: Neural Circuits of the Cerebral Cortex*: Cambridge University Press, New-York.
- Louis S, Borgelt C, Grun S (2010) Complexity distribution as a measure for assembly size and temporal precision. *Neural Networks* 23: 705–712.
- Staud B, Rotter S, Grun S (2010) CuBIC: cumulant based inference of higher-order correlations in massively parallel spike trains. *Journal of Computational Neuroscience* 29: 327–350.
- Staud B, Grun S, Rotter S (2010) Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference. *Frontiers in Computational Neuroscience* 4.
- Koch C, Rapp M, Segev I (1996) A brief history of time (constants). *Cerebral Cortex* 6: 93–101.
- Leger JF, Stern EA, Aertsen A, Heck D (2005) Synaptic integration in rat frontal cortex shaped by network activity. *Journal of Neurophysiology* 93: 281–293.
- Kelemen E, Fenton AA (2010) Dynamic Grouping of Hippocampal Neural Activity During Cognitive Control of Two Spatial Frames. *Plos Biology* 8.
- Bi GQ, Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* 18: 10464–10472.
- Markram H, Lubke J, Frotscher M, Sakmann B (1997) Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* 275: 213–215.
- Levy WB, Steward O (1983) Temporal contiguity requirements for long-term associative potentiation depression in the hippocampus. *Neuroscience* 8: 791–797.
- Fries P, Nikolic D, Singer W (2007) The gamma cycle. *Trends Neurosci* 30: 309–316.

67. Bowers JS (2009) On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychol Rev* 116: 220–251.
68. Quiroga RQ, Kreiman G, Koch C, Fried I (2008) Sparse but not 'Grandmother-cell' coding in the medial temporal lobe. *Trends in Cognitive Sciences* 12: 87–91.
69. Gelbard-Sagiv H, Mukamel R, Harel M, Malach R, Fried I (2008) Internally generated reactivation of single neurons in human hippocampus during free recall. *Science* 322: 96–101.
70. Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435: 1102–1107.
71. Waydo S, Kraskov A, Quiroga RQ, Fried I, Koch C (2006) Sparse representation in the human medial temporal lobe. *Journal of Neuroscience* 26: 10232–10234.
72. Brown EN, Kass RE, Mitra PP (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nature Neuroscience* 7: 456–461.
73. Plerou V, Gopikrishnan P, Rosenow B, Amaral LAN, Guhr T, et al. (2002) Random matrix approach to cross correlations in financial data. *Physical Review E* 65: 066126.
74. Tracy CA, Widom H (1994) Level-spacing distributions and the airy kernel. *Communications in Mathematical Physics* 159: 151–174.
75. Lloyd SP (1982) Least-squares quantization in PCM. *IEEE Transactions on Information Theory* 28: 129–137.

---

## ANEXO II: ON THE NEUROPHYSIOLOGY OF SEROTONERGIC HALLUCINOGENS

For thousands of years, in their struggle for survival, ancient hominids explored nature searching for sources of nourishment. Eventually, driven by their needs, curiosity and courage, women and men accidentally discovered plants and fungi that had remarkable properties. In all continents of the globe, hallucinogenic plants and fungi were regarded as sacred gifts from the Gods and their use was perpetuated in religious ceremonies. The *Homo* brain evolved for millions of years and eventually became able to plan, feel and create the sense of *self*. How fascinating is the fact that psychoactive substances can hijack elements of this machinery to create such unexpected and profound subjective experiences? Experiences that could only be called divine by our ancestors.

Actually, ancient men were not the only ones to be astonished by psychedelic experiences. Freedman (1968) provides an outstanding description of the psychoactive effects induced by Lysergic acid diethylamide (LSD): “*It is my impression that one basic dimension of behavior latently operative at any level of function and compellingly revealed in LSD states is ‘portentiousness’ - the capacity of the mind to see more than it can tell, to experience more than it can explicate, to believe in and be impressed with more than it can rationally justify, to experience boundlessness and ‘boundaryless’ events, from the banal to the profound.*”. It is certainly amazing that these substances, some of them with simple molecular structures, can cause such profound alterations in consciousness when systemically administered in humans.

Although the action of hallucinogens is often outlined as simply derived from their binding properties to certain receptors in neuronal cell membranes, it is fair to say that only the understanding of their effects in highly distributed and interconnected brain circuits will unravel the mechanisms of how they alter consciousness. How can such simple molecules provoke such dramatic effects in the functioning of neural systems, at such diminutive doses? In pursuit of answers to this question, the path may also lead to insights into the functioning of the neural system and the mechanisms of consciousness.

In this chapter I focus on the classic hallucinogens: the *serotonergic agonists*. I aim to review the current knowledge on their molecular mechanisms and then I will try to develop a working hypothesis for their systemic action. Finally, I will present and discuss some preliminary results from hippocampal recordings.

## WHAT IS A HALLUCINOGEN?

If the term *hallucinogen* is taken literally, it implies that these are substances that induce hallucinations. As noted by Oliver Sacks (2013), the word *hallucination* itself does not have a straightforward definition. William James (1890) stated that “*An hallucination is a strictly sensational form of consciousness, as good and true a sensation as if there were a real object there. The object happens to be not here, that is all*”. In the same line of thought, Oliver Sacks (2013) emphasizes that when one voluntarily imagines an *object*, such as the face of a relative, this mental image is not a *true* hallucination. In contrast, true hallucinations happen spontaneously and are projected into the external world without the presence of an external stimulus. Thus, they are liable to be interpreted as real. If we take these definitions rigorously, drugs like LSD and psilocybin are rarely hallucinogens (Bakalar and Grinspoon, 1997; Nichols, 2004).

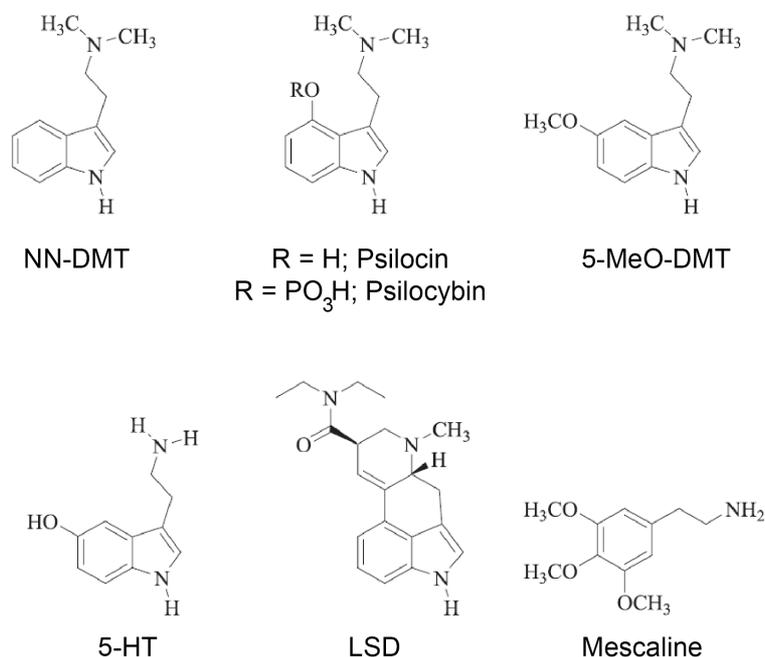
In a recently published meta-analysis, Studerus et al. (2011) investigated subjective effects reported by 110 healthy volunteers that received different doses of psilocybin. Although visual effects, such as facilitated imagination, synesthesia and non-voluntary imagery with eyes closed were typical, vivid true hallucinations were not reported. Hence, the authors argued that the visual experiences induced by psilocybin (in these subjects) were actually *pseudo-hallucinations*, a term used to refer to visions that happen inside the subjective space of the mind (Bakalar and Grinspoon, 1997; Sacks, 2013). In another study, volunteers under the effect of the psychoactive drink *ayahuasca* reported visual distortions and misperceptions such as “...*increases in an object's brightness and sharpness, or as vibrations in the visual field...*” (Riba et al., 2001). Unfortunately, scientific reports generally disregard the distinctions between hallucinations, pseudo-hallucinations, misperceptions and distortions when quantifying their observations. In any case, it seems reasonable to conclude that *true* hallucinations are much rarer to be induced by hallucinogens than what is anecdotally announced to the lay public. Accordingly, the terms *pseudohallucinogenic* and *illusionogenic* have been proposed in the past as being more adequate (Bakalar and Grinspoon, 1997). Yet, these cumbersome terminologies do not help to properly classify these substances. In my opinion, their difficulties can be summarized in two issues: (1) they do not contemplate the complex and wide-ranging actions of these drugs that certainly go beyond distortions in sensory perceptions and (pseudo-)hallucinations; and (2) they are far from accurate, since if we classify every drug that may cause pseudo-hallucinations in a single category, we would pool together substances with completely different pharmacological properties and with subjective effects that are easily distinguishable.

The term *psychotomimetic* was proposed due to supposed similarities between the acute effects of these substances and symptoms of psychosis. Nevertheless, the current scientific knowledge does not convincingly support any link between the physiological basis of real psychosis and of the acute actions of hallucinogens. Thus, I argue that this term has a detractory connotation since it leads to the erroneous interpretation that these drugs are primarily characterized as triggers to psychotic disorders. This agrees with Humphry Osmond (1957) dissatisfaction exposed over 50 years ago: “*Why are we always preoccupied with the pathological, the negative?*”. Throughout, in my opinion, this term is more misleading than informative.

As mentioned previously, drugs with completely distinct pharmacology are sometimes referred to as hallucinogens. For instance, Schultes, Hofmann and Ratsch list marijuana and *salvia divinorum* as hallucinogens in the seminal book *Plants of the Gods* (2001), although the systems that the psychoactive components of these plants affect are fairly distinct: the former being a well-known cannabinoid (Ashton, 2001) and the latter primarily a  $\kappa$ -opioid agonist (Prisinzano, 2005). Other examples range from N-methyl-D-aspartate (NMDA) antagonists, like ketamine (Kohrs and Durieux, 1998); to drugs with more complex pharmacodynamics (Morton, 2005), like 3,4-methylenedioxymethamphetamine (MDMA, popularly known as *ecstasy*).

Here I am concerned with the *serotonergic* hallucinogens. These are the substances Osmond (1957) proposed to be referred to as *psychedelics* (derived from greek words for ‘mind’ and ‘manifest’): “...*a name that will include the concepts of enriching the mind and enlarging the vision.*”. As I will discuss in more details along this manuscript, the *serotonergic* hallucinogens, as known as the *classic* hallucinogens, compose a well-defined class of drugs. Here, I employ a formal definition based on Glennon’s criteria (1999): A drug (or a chemical) is a serotonergic (or a classic) hallucinogen if (1) it meets Hollister’s standards (see below), (2) its psychoactive action is dependent on the activation of 5-HT<sub>2A</sub> receptors, and (3) it is perceived as 1-(2,5-dimethoxy-4-methylphenyl)-2-aminopropane (DOM) by rats performing a discriminative task. Regarding 1, Hollister (1972) defines hallucinogens as substances that produce thought, mood and perceptual changes; **without** causing physical addiction, craving, major physical disturbances, cognitive impairments, amnesia or delirium. Criteria 2 and 3 will be detailed in the next section. Important examples of serotonergic hallucinogens would include the tryptamines NN-DMT, 5-MeO-DMT, psilocin (4-HO-DMT) and LSD; and the phenethylamines mescaline, 2,5-dimethoxy-4-iodoamphetamine (DOI) and DOM. Below we review their binding properties to 5-HT receptors.

## THE ROLE OF 5-HT<sub>2A</sub> RECEPTORS



**Figure A2.1. Chemical structure of serotonergic hallucinogens.** Adapted from Nichols (2004).

The presence of serotonin in the brain was first reported by Gaddum (1953) ten years after the discovery of LSD by Albert Hoffman. The resemblance of their chemical structure was soon noted by the scientific community. In fact, diverse studies have shown that these substances were ligands of 5-HT receptors, in particular to 5-HT<sub>2A</sub> receptors (Lyon et al., 1988; McKenna et al., 1990; Sadzot et al., 1989; Shannon et al., 1984). Figure A2.1 displays the chemical structure of important serotonergic hallucinogens and the neurotransmitter serotonin (5-hydroxytryptamine; 5-HT).

Important evidence for the agonist action of hallucinogens on 5-HT<sub>2A</sub> receptors (5-HT<sub>2A</sub>Rs) came from the Drug Discrimination Paradigm (DDP) (Glennon et al., 1983a). In these experiments, rats are trained to emit specific responses to different conditions. For example, they are rewarded when they press lever A if they are injected with saline, whereas they should press lever B if they receive DOM. Thus, it is like the rat can *tell* the experimenter that he noticed that DOM was given to him. Remarkably, serotonergic hallucinogens present cross-generalization in these experiments (Glennon et al., 1984). For instance, rats trained to recognize DOM would respond to LSD, psilocybin or DMTs as if they had been given DOM (Glennon et al., 1979, 1983b). Of note, the hallucinogenic doses of these substances in humans were extremely correlated with the minimum doses recognized by rats (Glennon et al., 1983b). Furthermore, it has been shown that pretreatment with 5-HT<sub>2A</sub>R antagonists could

prevent animals to detect serotonergic hallucinogens in general (Appel and Callahan, 1989; Glennon et al., 1983c; May et al., 2009; Schreiber et al., 1994; Smith et al., 2003; Vollenweider et al., 1998; Winter et al., 2007). These evidences plus the fact that the hallucinogenic potency of these drugs is extremely correlated with their affinities for these receptors (Glennon et al., 1984) strongly suggest that 5-HT<sub>2A</sub>R activation is the basis for hallucinogenic psychoactive effects.

5-HT<sub>2A</sub>Rs are G protein-coupled receptors that when activated induce an excitatory effect (Aghajanian and Marek, 1997). More specifically, they are mainly located at apical dendrites (Puig and Gullledge, 2011), thus they act amplifying input currents (Aghajanian and Marek, 1999). They are expressed within the neocortex in about 60% of the pyramidal cells (Amargós-Bosch et al., 2004; Santana et al., 2004), and a smaller proportion of inhibitory cells in the same region (Puig et al., 2010), mainly in deeper layers (Puig et al., 2010; Xu and Pandey, 2000).

Accordingly, systemic administration of DOI increases the firing rate of pyramidal cells in the deep layers of the rat prefrontal cortex (Puig et al., 2003). In addition, DOI disrupts neocortical slow LFP oscillations related to the anesthesia (Celada et al., 2008), indicating that the activation of 5-HT<sub>2A</sub>R induces desynchronization of neocortical activity. Conversely, Wood et al (2012) reported that although DOI induced an increase in firing rate in some neocortical neurons, it had an inhibitory effect in most of them. Replacing DOI by 5-MeO-DMT in similar experiments, Riga et al (2014) showed that most of the cells in the rat frontal cortex increased firing rate and that, in fact, slow oscillations were eliminated. Thus, it seems these recent studies of the electrophysiological effects of activating 5-HT<sub>2A</sub>Rs by hallucinogens present results that are only partially consistent.

In parallel, Carhart-Harris et al (2012) have shown that *Blood-oxygen-level dependent* (BOLD) signals measured by functional magnetic resonance from the frontal cortex are reduced by the intravenous administration of psilocin in humans. Last year, the same group reported that the electrical activity in the same area (measured by magnetoencephalography - MEG) was also disrupted by psilocin (Muthukumaraswamy et al., 2013). The authors proposed that the excitatory stimulation of pyramidal cells in layer 5 would induce an overall desynchronization of the network.

Note that in fact Wood et al (2012) and Riga et al (2014) have not related electrophysiological effects with anatomical information, and therefore, it can be expected that the cells these authors reported to be inhibited by 5-HT<sub>2A</sub>R activation were located in superficial layers whereas the excited cells were from deep ones.

I prefer to be more careful with regard to activity synchronization since the elimination of slow oscillations was observed in anesthetized animals. It seems to me that when Wood et al (2012) claimed that DOI *disrupts* activity in the neocortex and, similarly, when Riga et al (2014) stated that 5-MeO-DMT *disrupts* cortical function in the title of their work, they did not consider that the slow oscillations eliminated by these substances were anesthesia effects and have very little relation to normal cortical function.

In the next section I try to present a hypothesis for hallucinogenic mechanisms based on a simple model for the neocortex network and the observations mentioned above.

### AN INITIAL HYPOTHESIS FOR EFFECTS ON NEURONAL NETWORKS

I start this section by reviewing the canonical microcircuit model for the neocortex (Figure A2.2). By microcircuit I mean a processing module, such as a cortical column, for example. Although this is just a toy model, I consider this scheme a useful starting point to develop a network hypothesis for the action of serotonergic hallucinogens in the brain.

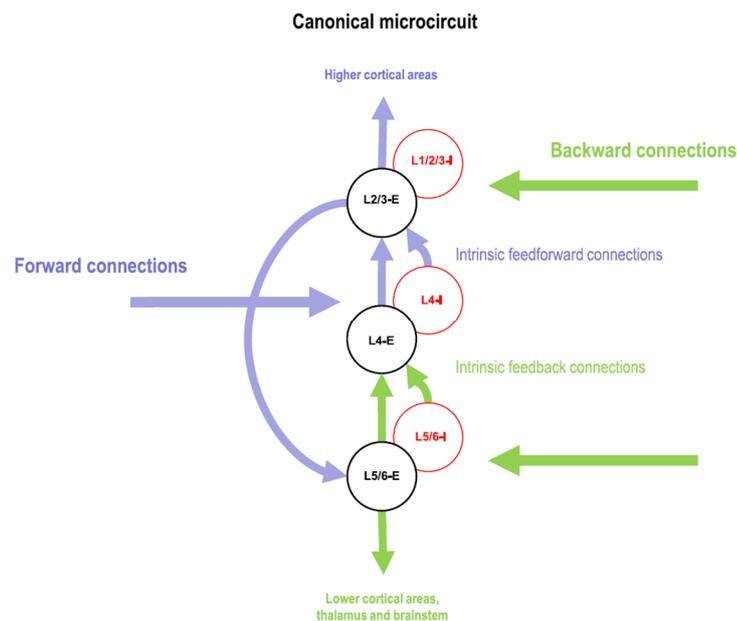


Figure A2.2. Canonical microcircuit model. From Bastos et al (2012).

The canonical microcircuit model outlined by Bastos et al (2012) has the fundamental assumption that the neocortex is organized hierarchically. This means that neural networks are organized in information processing levels. Note that this model does not assume that the information flux is unidirectional (feedforward) since higher areas can modulate lower ones through backward connections.

As displayed in Figure A2.2, layer 4 receives inputs from the lower areas whereas layers 2/3 and 5/6 receive information from higher areas. Moreover, layers 2/3 transmit information to higher areas, while deeper layers send information downwards in hierarchy. Thus, we assume that the inputs and outputs of the canonical microcircuit are anatomically organized.

Now the question that follows is: what are the consequences of the excitation of the deep layers by hallucinogens in light of the canonical microcircuit model of the neocortex? My initial working hypothesis for the systemic mechanisms of hallucinogens is that the dendritic amplification of excitatory currents in neocortical layer 5 biases the information flow downwards in hierarchy. Thus, under increasing doses of hallucinogens the inner information of the brain would increasingly rule sensory information, eventually trumping even the most salient signals from the external environment. Introspectively, this could be perceived as sensory distortion, facilitation of vivid imagery and recall of deep memories. This would be in accordance with the observation that the activity in the primary visual areas of subjects under the effects of ayahuasca, when performing a closed-eyes imagery task, is as strong as if they were actually seeing pictures (de Araujo et al., 2012). Electrophysiological recordings in animals evaluating laminar profiles of neocortical areas and interareal coupling analysis must therefore be done in order to test this hypothesis.

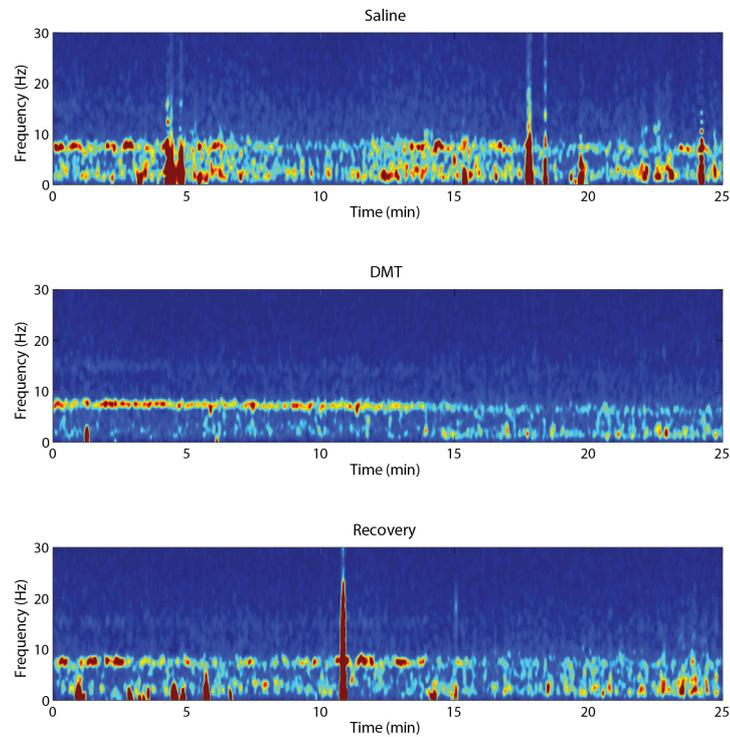
## **PRELIMINARY OBSERVATIONS IN HIPPOCAMPAL RECORDINGS**

Some pilot experiments were performed in order to evaluate the effects of 5-MeO-DMT in the hippocampal electrophysiology. The protocol consisted of four epochs: 1) basal; during which the animal behaves freely in an open field (30 minutes); 2) saline; which starts with an intraperitoneal injection of saline (120 minutes); 3) DMT; which starts with an intraperitoneal injection of (10mg/kg) 5-MeO-DMT (240 minutes); 4) Recovery; which happens in the following day and starts with another intraperitoneal injection of saline (240 minutes).

Figure A2.3 displays the spectrograms (1-sec windows) of the hippocampal local field potentials during the initial 25 minutes of epochs 2, 3 and 4. Remarkably, after the injection of 5-MeO-DMT, theta oscillations dominated the signal for around 12 minutes.

Visual evaluation of the video during this period showed that the animal was in an alert-like state. On the other hand, theta oscillations are well known correlates of locomotion (Buzsáki, 2002; Vanderwolf, 1969). Thus, we divided the epochs in five behavioral states for further analysis: REM sleep, slow-wave sleep (SWS), static (quiet waking), slow movements and exploration. These states were classified based on the instantaneous locomotion speed of the animal (tracking performed with the software IDtracker, developed by Pérez-Escudero

et al (2014)), and by visual inspection of the video. REM periods were detected by no locomotion, sleep behavior on the video and presence of theta oscillations in the local field potentials.

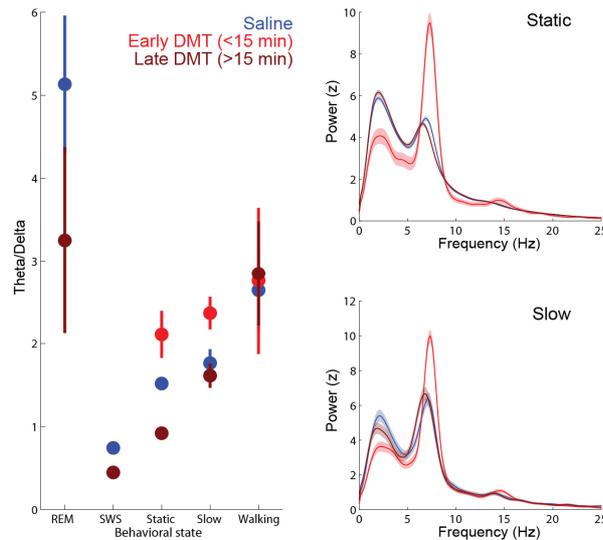


**Figure A2.3. Hippocampal local field potentials spectrograms for saline, DMT and recovery epochs.**

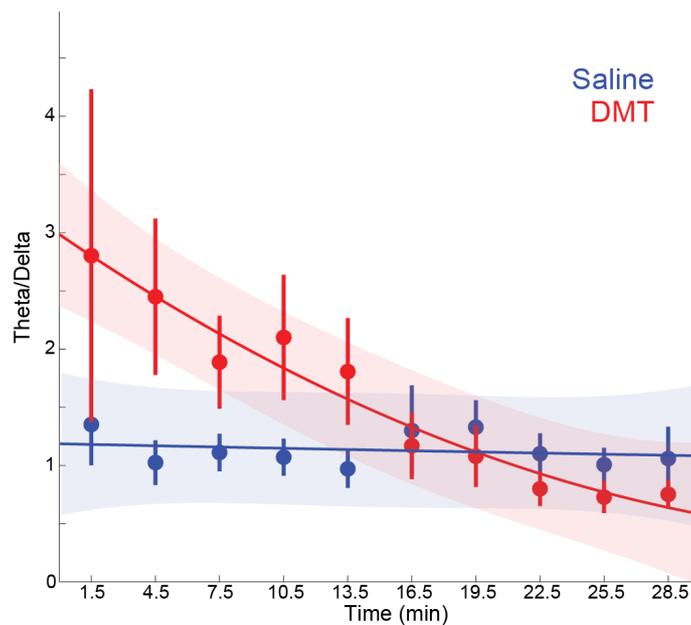
For each window of the spectrograms displayed in Figure A2.3, we computed the ratio between the power within the delta range (0 to 5 Hz) and the power within the theta range (5 to 12 Hz). The left panel of Figure A2.4 shows these ratios for each behavioral state for different epochs, as labeled. Further, the DMT epoch was divided in *early DMT*, which comprises only the initial fifteen minutes of this epoch; and *late DMT*, which contains the remaining epoch. Note that theta oscillations were stronger during the DMT effects in *static* and *slow* periods.

In order to show that these effects were not due the Theta/Delta normalization, we further computed the average z-scored power spectra for different epochs in *static* and *slow* states. These are shown in the right panels of Figure A2.4. Figure A2.5 shows the time course of the Theta/Delta ratio only considering *static* periods. Note that the ratio following DMT equalizes to saline levels after 12 to 15 minutes.

Thus, prominent theta oscillations could not be explained by locomotor behavior. In sum, the results above indicate that the rat was in an alert-like still state following the injection of DMT for at least 12 minutes.

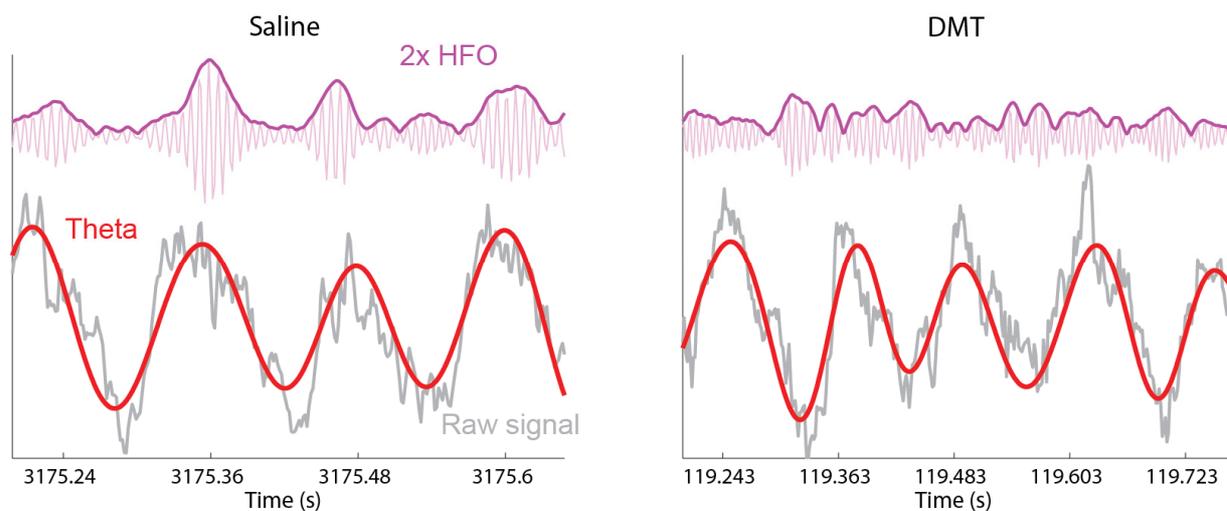


**Figure A2.4. Power of theta oscillations during different behavioral states.** All results are means and confidence intervals ( $\alpha = 0.05$ )



**Figure A2.5. Time course of Theta/Delta ratios in *static* periods for DMT and saline conditions.** All results are shown as means and confidence interval ( $\alpha = 0.05$ ). Lines are 2nd order polynomials fitted to data.

Next, we investigated whether the DMT-induced theta oscillations had common features with regular theta oscillations recorded in the same region. Visual inspection of the components of the signals during the *basal* epoch revealed the presence of high frequency oscillations (HFO; 120 to 160 Hz) increasing in energy at the peaks of theta oscillations (Figure A2.6, left panel). This coupling is typically detected in the superficial layers of the hippocampus (Scheffer-Teixeira et al., 2012). Remarkably, the total absence of such coupling was noted in the recordings of the *early DMT* epoch in the same electrode (Figure A2.6, right panel).



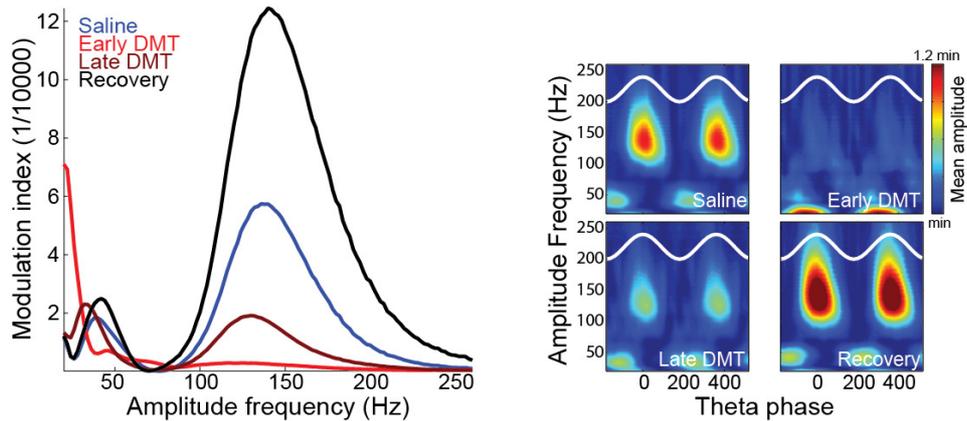
**Figure A2.6. Representative examples of local field potentials in the Saline and DMT epochs.** Raw signals are displayed in gray. Theta (4 to 12 Hz) and HFO (120 to 160 Hz) components are displayed in red and magenta, respectively. HFO amplitude envelopes computed with the Hilbert transform are also shown. Note that in the Saline epoch, the instantaneous amplitude of HFOs coincide with theta peaks; whereas DMT HFO amplitudes seem to be completely uncoupled.

In order to confirm these observations, we concatenated windows corresponding to  $\text{Theta}/\text{Delta} > 1$  and computed the modulation indexes (Tort et al., 2010) between the amplitude at several frequencies (20 to 260 Hz) and theta phase. Figure A2.7 left panel displays these results. Panels on the right of the same figure display the mean amplitudes of the frequencies studied as a function of theta phase. Note that during the *early DMT* epoch, coupling is absent.

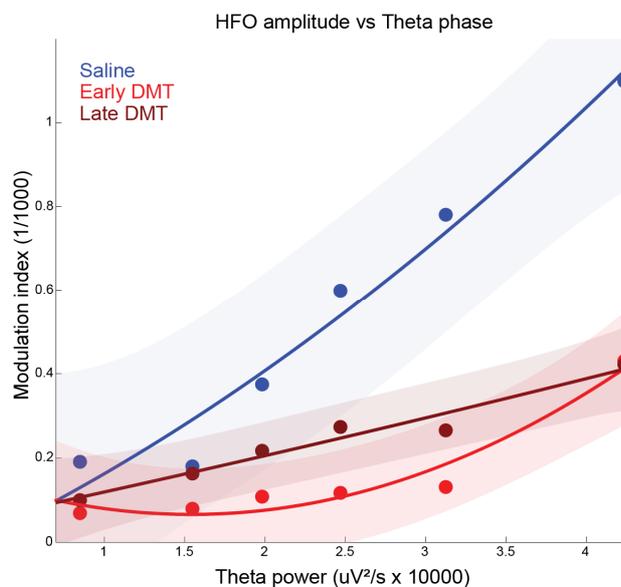
Since this cross-frequency coupling magnitude is correlated with the power of theta oscillations (Scheffer-Teixeira et al., 2012), we performed a final control in order to evaluate whether this could explain the differences between epochs. For that, we grouped the 1-sec time windows according to their theta power. Figure A2.8 shows the modulation indexes of HFO amplitude and theta phase for different values of theta power. Note that for the same values of theta power, the DMT epoch still presents the weaker cross-frequency coupling between these oscillations.

Taken together, these preliminary results indicate that the coupling between HFOs and hippocampal theta waves is eliminated by systemic 5-MeO-DMT. The functions and origins of HFOs are largely unknown. Since HFOs can be recorded in neocortical areas (Scheffzük et al., 2011; Sirota et al., 2008), it is possible that the effects observed derive from an interruption of the information flux from the entorhinal cortex to the hippocampus. In fact, 5-MeO-DMT decreased the coupling between HFO and theta oscillations without affecting their individual power, which is compatible with a transmission disruption. Although some data suggest that

HFOs can also be generated within the hippocampus (Tort et al., 2013), it is possible that HFOs need a neocortical trigger. Simultaneous recordings in vivo from neocortical and hippocampal areas are needed to clarify these questions. Overall, these preliminary results demonstrate that system electrophysiology in behaving animals can provide key insights into the understanding of hallucinogen mechanisms. In particular, this investigation with 5-MeO-DMT may unravel the mechanisms of a largely unknown oscillation.



**Figure A2.7. Coupling of several frequencies and theta oscillations.** *Left:* modulation indexes between theta phase and amplitude of several frequencies (horizontal axis) for different experimental epochs (as labeled). *Right:* Mean normalized amplitudes of different frequencies (vertical axis) for different theta phases (horizontal axis). Mean amplitudes were divided by their minimum. Theta oscillations are shown in white as a reference.



**Figure A2.8. Modulation index between HFO amplitude and theta phase for different theta power values.**

---

## REFERÊNCIAS

- Abeles, M. (2009). Synfire chains. *Scholarpedia* 4, 1441.
- Adrian, E.D., and Zotterman, Y. (1926). The impulses produced by sensory nerve endings. *J. Physiol.* 61, 465–483.
- Aghajanian, G.K., and Marek, G.J. (1997). Serotonin induces excitatory postsynaptic potentials in apical dendrites of neocortical pyramidal cells. *Neuropharmacology* 36, 589–599.
- Aghajanian, G.K., and Marek, G.J. (1999). Serotonin, via 5-HT<sub>2A</sub> receptors, increases EPSCs in layer V pyramidal cells of prefrontal cortex by an asynchronous mode of glutamate release. *Brain Res.* 825, 161–171.
- Amargós-Bosch, M., Bortolozzi, A., Puig, M.V., Serrats, J., Adell, A., Celada, P., Toth, M., Mengod, G., and Artigas, F. (2004). Co-expression and in vivo interaction of serotonin<sub>1A</sub> and serotonin<sub>2A</sub> receptors in pyramidal neurons of prefrontal cortex. *Cereb. Cortex N. Y. N* 1991 14, 281–299.
- Andersen, P., Morris, R., Amaral, D., Bliss, T., and O’Keefe, J. (2006). *The Hippocampus Book* (Oxford University Press).
- Appel, J.B., and Callahan, P.M. (1989). Involvement of 5-HT receptor subtypes in the discriminative stimulus properties of mescaline. *Eur. J. Pharmacol.* 159, 41–46.
- De Araujo, D.B., Ribeiro, S., Cecchi, G.A., Carvalho, F.M., Sanchez, T.A., Pinto, J.P., de Martinis, B.S., Crippa, J.A., Hallak, J.E.C., and Santos, A.C. (2012). Seeing with the eyes shut: neural basis of enhanced imagery following Ayahuasca ingestion. *Hum. Brain Mapp.* 33, 2550–2560.
- Ashton, C.H. (2001). Pharmacology and effects of cannabis: a brief review. *Br. J. Psychiatry J. Ment. Sci.* 178, 101–106.
- Bakalar, J.B., and Grinspoon, L. (1997). *Psychedelic Drugs Reconsidered* (New York, NY: The Lindesmith Center).
- Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711.
- Börgers, C., and Kopell, N. (2005). Effects of Noisy Drive on Rhythms in Networks of Excitatory and Inhibitory Neurons. *Neural Comput.* 17, 557–608.
- Bragin, A., Jando, G., Nadasdy, Z., Hetke, J., Wise, K., and Buzsáki, G. (1995). Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.* 15, 47–60.
- Brown, E.N., Kass, R.E., and Mitra, P.P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.* 7, 456–461.
- Buzsáki, G. (2002). Theta Oscillations in the Hippocampus. *Neuron* 33, 325–340.
- Buzsáki, G. (2010). Neural syntax: cell assemblies, synapse ensembles, and readers. *Neuron* 68, 362–385.
- Buzsáki, G., Anastassiou, C.A., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420.

- Carhart-Harris, R.L., Erritzoe, D., Williams, T., Stone, J.M., Reed, L.J., Colasanti, A., Tyacke, R.J., Leech, R., Malizia, A.L., Murphy, K., et al. (2012). Neural correlates of the psychedelic state as determined by fMRI studies with psilocybin. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 2138–2143.
- Celada, P., Puig, M.V., Díaz-Mataix, L., and Artigas, F. (2008). The hallucinogen DOI reduces low-frequency oscillations in rat prefrontal cortex: reversal by antipsychotic drugs. *Biol. Psychiatry* *64*, 392–400.
- Cohen, N.J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., and Nash, C. (1999). Hippocampal system and declarative (relational) memory: Summarizing the data from functional neuroimaging studies. *Hippocampus* *9*, 83–98.
- Colgin, L.L., Denninger, T., Fyhn, M., Hafting, T., Bonnevie, T., Jensen, O., Moser, M.-B., and Moser, E.I. (2009). Frequency of gamma oscillations routes flow of information in the hippocampus. *Nature* *462*, 353–357.
- Cutsuridis, V., Cobb, S., and Graham, B.P. (2010). Encoding and retrieval in a model of the hippocampal CA1 microcircuit. *Hippocampus* *20*, 423–446.
- Demiralp, T., Bayraktaroglu, Z., Lenz, D., Junge, S., Busch, N.A., Maess, B., Ergen, M., and Herrmann, C.S. (2007). Gamma amplitudes are coupled to theta phase in human EEG during visual perception. *Int. J. Psychophysiol.* *64*, 24–30.
- Eichenbaum, H. (2004). Hippocampus: Cognitive Processes and Neural Representations that Underlie Declarative Memory. *Neuron* *44*, 109–120.
- Ferraina, S., Paré, M., and Wurtz, R.H. (2002). Comparison of cortico-cortical and cortico-collicular signals for the generation of saccadic eye movements. *J. Neurophysiol.* *87*, 845–858.
- Freedman, D.X. (1968). On the use and abuse of LSD. *Arch. Gen. Psychiatry* *18*, 330–347.
- Freeman, W.J. (1978). Spatial properties of an EEG event in the olfactory bulb and cortex. *Electroencephalogr. Clin. Neurophysiol.* *44*, 586–605.
- Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* *9*, 474–480.
- Glennon, R.A. (1999). Arylalkylamine drugs of abuse: an overview of drug discrimination studies. *Pharmacol. Biochem. Behav.* *64*, 251–256.
- Glennon, R.A., Rosecrans, J.A., Young, R., and Gaines, J. (1979). Hallucinogens as a discriminative stimuli: generalization of DOM to a 5-methoxy-N, N-dimethyltryptamine stimulus. *Life Sci.* *24*, 993–997.
- Glennon, R.A., Rosecrans, J.A., and Young, R. (1983a). Drug-induced discrimination: a description of the paradigm and a review of its specific application to the study of hallucinogenic agents. *Med. Res. Rev.* *3*, 289–340.
- Glennon, R.A., Young, R., Jacyno, J.M., Slusher, M., and Rosecrans, J.A. (1983b). DOM-stimulus generalization to LSD and other hallucinogenic indolealkylamines. *Eur. J. Pharmacol.* *86*, 453–459.
- Glennon, R.A., Young, R., and Rosecrans, J.A. (1983c). Antagonism of the effects of the hallucinogen DOM and the purported 5-HT agonist quipazine by 5-HT2 antagonists. *Eur. J. Pharmacol.* *91*, 189–196.
- Glennon, R.A., Titeler, M., and McKenney, J.D. (1984). Evidence for 5-HT2 involvement in the mechanism of action of hallucinogenic agents. *Life Sci.* *35*, 2505–2511.
- Gloveli, T., Dugladze, T., Saha, S., Monyer, H., Heinemann, U., Traub, R.D., Whittington, M.A., and Buhl, E.H. (2005). Differential involvement of oriens/pyramidal interneurons in hippocampal network oscillations in vitro. *J. Physiol.* *562*, 131–147.

- Gray, C.M. (1994). Synchronous oscillations in neuronal systems: Mechanisms and functions. *J. Comput. Neurosci.* *1*, 11–38.
- Gross, C.G. (2002). Genealogy of the “grandmother cell.” *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* *8*, 512–518.
- Harris, K.D. (2005). Neural signatures of cell assembly organization. *Nat. Rev. Neurosci.* *6*, 399–407.
- Hasselmo, M., Bodelón, C., and Wyble, B. (2002). A Proposed Function for Hippocampal Theta Rhythm: Separate Phases of Encoding and Retrieval Enhance Reversal of Prior Learning. *Neural Comput.* *14*, 793–817.
- Hebb, D.O. (1949). *The Organization of Behavior: A Neuropsychological Theory* (Taylor & Francis).
- Hollister, L.E. (1972). *Chemical Psychoses: Lsd and Related Drugs* (Springfield, Ill.: Charles C Thomas Pub Ltd).
- Izhikevich, E.M. (2006). Polychronization: computation with spikes. *Neural Comput.* *18*, 245–282.
- James, W. (1950). *The Principles of Psychology, Vol. 1* (New York: Dover Publications).
- Jensen, O., and Colgin, L.L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends Cogn. Sci.* *11*, 267–269.
- Kendrick, K.M., Zhan, Y., Fischer, H., Nicol, A.U., Zhang, X., and Feng, J. (2011). Learning alters theta amplitude, theta-gamma coupling and neuronal synchronization in inferotemporal cortex. *BMC Neurosci.* *12*, 55.
- Kohrs, R., and Durieux, M.E. (1998). Ketamine: teaching an old drug new tricks. *Anesth. Analg.* *87*, 1186–1193.
- Kopell, N., Börgers, C., Pervouchine, D., Malerba, P., and Tort, A. (2010). Gamma and Theta Rhythms in Biophysical Models of Hippocampal Circuits. In *Hippocampal Microcircuits*, V. Cutsuridis, B. Graham, S. Cobb, and I. Vida, eds. (Springer New York), pp. 423–457.
- Kunec, S., Hasselmo, M.E., and Kopell, N. (2005). Encoding and Retrieval in the CA3 Region of the Hippocampus: A Model of Theta-Phase Separation. *J. Neurophysiol.* *94*, 70–82.
- Lashley, K. (1950). In search of the engram. In *Physiological Mechanisms in Animal Behavior. (Society’s Symposium IV.)*, (Oxford, England: Academic Press), pp. 454–482.
- Laurent, G. (2002). Olfactory network dynamics and the coding of multidimensional signals. *Nat. Rev. Neurosci.* *3*, 884–895.
- Lisman, J. (2005). The theta/gamma discrete phase code occurring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus* *15*, 913–922.
- Lisman, J. (2010). Working Memory: The Importance of Theta and Gamma Oscillations. *Curr. Biol.* *20*, R490–R492.
- Lopes-dos-Santos, V., Conde-Ocazonez, S., Nicolelis, M.A.L., Ribeiro, S.T., and Tort, A.B.L. (2011). Neuronal Assembly Detection and Cell Membership Specification by Principal Component Analysis. *PLoS ONE* *6*, e20996.
- Lopes-dos-Santos, V., Ribeiro, S., and Tort, A.B.L. (2013). Detecting cell assemblies in large neuronal populations. *J. Neurosci. Methods* *220*, 149–166.
- Lopes-Dos-Santos, V., Panzeri, S., Kayser, C., Diamond, M.E., and Quiñero, R. (2015). Extracting information in spike time patterns with wavelets and information theory. *J. Neurophysiol.* *113*, 1015–1033.

- Lyon, R.A., Titeler, M., Seggel, M.R., and Glennon, R.A. (1988). Indolealkylamine analogs share 5-HT<sub>2</sub> binding characteristics with phenylalkylamine hallucinogens. *Eur. J. Pharmacol.* *145*, 291–297.
- Marr, D. (1970). A Theory for Cerebral Neocortex. *Proc. R. Soc. Lond. B Biol. Sci.* *176*, 161–234.
- May, J.A., Sharif, N.A., Chen, H.-H., Liao, J.C., Kelly, C.R., Glennon, R.A., Young, R., Li, J.-X., Rice, K.C., and France, C.P. (2009). Pharmacological properties and discriminative stimulus effects of a novel and selective 5-HT<sub>2</sub> receptor agonist AL-38022A [(S)-2-(8,9-dihydro-7H-pyrano[2,3-g]indazol-1-yl)-1-methylethylamine]. *Pharmacol. Biochem. Behav.* *91*, 307–314.
- McClelland, J.L., McNaughton, B.L., and O'Reilly, R.C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* *102*, 419–457.
- McKenna, D.J., Repke, D.B., Lo, L., and Peroutka, S.J. (1990). Differential interactions of indolealkylamines with 5-hydroxytryptamine receptor subtypes. *Neuropharmacology* *29*, 193–198.
- Miller, R. (1975). Distribution and properties of commissural and other neurons in cat sensorimotor cortex. *J. Comp. Neurol.* *164*, 361–373.
- Morton, J. (2005). Ecstasy: pharmacology and neurotoxicity. *Curr. Opin. Pharmacol.* *5*, 79–86.
- Muthukumaraswamy, S.D., Carhart-Harris, R.L., Moran, R.J., Brookes, M.J., Williams, T.M., Erntizoe, D., Sessa, B., Papadopoulos, A., Bolstridge, M., Singh, K.D., et al. (2013). Broadband cortical desynchronization underlies the human psychedelic state. *J. Neurosci. Off. J. Soc. Neurosci.* *33*, 15171–15183.
- Nichols, D.E. (2004). Hallucinogens. *Pharmacol. Ther.* *101*, 131–181.
- O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* *34*, 171–175.
- Osmond, H. (1957). A review of the clinical effects of psychotomimetic agents. *Ann. N. Y. Acad. Sci.* *66*, 418–434.
- Panzeri, S., Petersen, R.S., Schultz, S.R., Lebedev, M., and Diamond, M.E. (2001). The Role of Spike Timing in the Coding of Stimulus Location in Rat Somatosensory Cortex. *Neuron* *29*, 769–777.
- Pérez-Escudero, A., Vicente-Page, J., Hinz, R.C., Arganda, S., and de Polavieja, G.G. (2014). idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nat. Methods* *11*, 743–748.
- Petrantonakis, P.C., and Poirazi, P. (2014). A compressed sensing perspective of hippocampal function. *Front. Syst. Neurosci.* *8*.
- Prisinzano, T.E. (2005). Psychopharmacology of the hallucinogenic sage *Salvia divinorum*. *Life Sci.* *78*, 527–531.
- Puig, M.V., and Gullledge, A.T. (2011). Serotonin and prefrontal cortex function: neurons, networks, and circuits. *Mol. Neurobiol.* *44*, 449–464.
- Puig, M.V., Celada, P., Díaz-Mataix, L., and Artigas, F. (2003). In vivo modulation of the activity of pyramidal neurons in the rat medial prefrontal cortex by 5-HT<sub>2A</sub> receptors: relationship to thalamocortical afferents. *Cereb. Cortex N. Y. N 1991* *13*, 870–882.
- Puig, M.V., Watakabe, A., Ushimaru, M., Yamamori, T., and Kawaguchi, Y. (2010). Serotonin modulates fast-spiking interneuron and synchronous activity in the rat prefrontal cortex through 5-HT<sub>1A</sub> and 5-HT<sub>2A</sub> receptors. *J. Neurosci. Off. J. Soc. Neurosci.* *30*, 2211–2222.

- Quiroga, R.Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nat. Rev. Neurosci.* *13*, 587–597.
- Rennó-Costa, C., Lisman, J.E., and Verschure, P.F.M.J. (2014). A signature of attractor dynamics in the CA3 region of the hippocampus. *PLoS Comput. Biol.* *10*, e1003641.
- Riba, J., Rodríguez-Fornells, A., Urbano, G., Morte, A., Antonijóan, R., Montero, M., Callaway, J.C., and Barbanoj, M.J. (2001). Subjective effects and tolerability of the South American psychoactive beverage Ayahuasca in healthy volunteers. *Psychopharmacology (Berl.)* *154*, 85–95.
- Riga, M.S., Soria, G., Tudela, R., Artigas, F., and Celada, P. (2014). The natural hallucinogen 5-MeO-DMT, component of Ayahuasca, disrupts cortical function in rats: reversal by antipsychotic drugs. *Int. J. Neuropsychopharmacol. Off. Sci. J. Coll. Int. Neuropsychopharmacol. CINP* *17*, 1269–1282.
- Sacks, O. (2013). *Hallucinations* (Waterville, Maine: Vintage).
- Sadzot, B., Baraban, J.M., Glennon, R.A., Lyon, R.A., Leonhardt, S., Jan, C.R., and Titeler, M. (1989). Hallucinogenic drug interactions at human brain 5-HT<sub>2</sub> receptors: implications for treating LSD-induced hallucinogenesis. *Psychopharmacology (Berl.)* *98*, 495–499.
- Santana, N., Bortolozzi, A., Serrats, J., Mengod, G., and Artigas, F. (2004). Expression of serotonin<sub>1A</sub> and serotonin<sub>2A</sub> receptors in pyramidal and GABAergic neurons of the rat prefrontal cortex. *Cereb. Cortex N. Y. N* *1991* *14*, 1100–1109.
- Schack, B., and Klimesch, W. (2002). Frequency characteristics of evoked and oscillatory electroencephalic activity in a human memory scanning task. *Neurosci. Lett.* *331*, 107–110.
- Scheffer-Teixeira, R., Belchior, H., Caixeta, F.V., Souza, B.C., Ribeiro, S., and Tort, A.B.L. (2012). Theta phase modulates multiple layer-specific oscillations in the CA1 region. *Cereb. Cortex N. Y. N* *1991* *22*, 2404–2414.
- Scheffer-Teixeira, R., Belchior, H., Leão, R.N., Ribeiro, S., and Tort, A.B.L. (2013). On high-frequency field oscillations (>100 Hz) and the spectral leakage of spiking activity. *J. Neurosci. Off. J. Soc. Neurosci.* *33*, 1535–1539.
- Scheffzük, C., Kukushka, V.I., Vyssotski, A.L., Draguhn, A., Tort, A.B.L., and Brankač, J. (2011). Selective coupling between theta phase and neocortical fast gamma oscillations during REM-sleep in mice. *PLoS One* *6*, e28489.
- Schomburg, E.W., Fernández-Ruiz, A., Mizuseki, K., Berényi, A., Anastassiou, C.A., Koch, C., and Buzsáki, G. (2014). Theta Phase Segregation of Input-Specific Gamma Patterns in Entorhinal-Hippocampal Networks. *Neuron* *84*, 470–485.
- Schreiber, R., Brocco, M., and Millan, M.J. (1994). Blockade of the discriminative stimulus effects of DOI by MDL 100,907 and the “atypical” antipsychotics, clozapine and risperidone. *Eur. J. Pharmacol.* *264*, 99–102.
- Schultes, R.E., Hofmann, A., and Rätsch, C. (2001). *Plants of the Gods: Their Sacred, Healing, and Hallucinogenic Powers* (Rochester, Vt: Healing Arts Press).
- Scoville, W.B., and Milner, B. (1957). LOSS OF RECENT MEMORY AFTER BILATERAL HIPPOCAMPAL LESIONS. *J. Neurol. Neurosurg. Psychiatry* *20*, 11–21.
- Shannon, M., Battaglia, G., Glennon, R.A., and Titeler, M. (1984). 5-HT<sub>1</sub> and 5-HT<sub>2</sub> binding properties of derivatives of the hallucinogen 1-(2,5-dimethoxyphenyl)-2-aminopropane (2,5-DMA). *Eur. J. Pharmacol.* *102*, 23–29.

- Singer, W. (1993). Synchronization of Cortical Activity and its Putative Role in Information Processing and Learning. *Annu. Rev. Physiol.* *55*, 349–374.
- Singer, W. (2007). Binding by synchrony. *Scholarpedia* *2*, 1657.
- Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., and Buzsáki, G. (2008). Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. *Neuron* *60*, 683–697.
- Smith, R.L., Barrett, R.J., and Sanders-Bush, E. (2003). Discriminative stimulus properties of 1-(2,5-dimethoxy-4-iodophenyl)-2-aminopropane [(+/-)DOI] in C57BL/6J mice. *Psychopharmacology (Berl.)* *166*, 61–68.
- Squire, L.R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science* *253*, 1380–1386.
- Stevenson, I.H., and Kording, K.P. (2011). How advances in neural recording affect data analysis. *Nat. Neurosci.* *14*, 139–142.
- Studerus, E., Kometer, M., Hasler, F., and Vollenweider, F.X. (2011). Acute, subacute and long-term subjective effects of psilocybin in healthy humans: a pooled analysis of experimental studies. *J. Psychopharmacol. Oxf. Engl.* *25*, 1434–1452.
- Swadlow, H.A. (1985). Physiological properties of individual cerebral axons studied in vivo for as long as one year. *J. Neurophysiol.* *54*, 1346–1362.
- Swadlow, H.A. (1992). Monitoring the excitability of neocortical efferent neurons to direct activation by extracellular current pulses. *J. Neurophysiol.* *68*, 605–619.
- Tort, A.B.L., Rotstein, H.G., Dugladze, T., Gloveli, T., and Kopell, N.J. (2007). On the formation of gamma-coherent cell assemblies by oriens lacunosum-moleculare interneurons in the hippocampus. *Proc. Natl. Acad. Sci.* *104*, 13490–13495.
- Tort, A.B.L., Komorowski, R.W., Manns, J.R., Kopell, N.J., and Eichenbaum, H. (2009). Theta–gamma coupling increases during the learning of item–context associations. *Proc. Natl. Acad. Sci.* *106*, 20942–20947.
- Tort, A.B.L., Komorowski, R., Eichenbaum, H., and Kopell, N. (2010). Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *J. Neurophysiol.* *104*, 1195–1210.
- Tort, A.B.L., Scheffer-Teixeira, R., Souza, B.C., Draguhn, A., and Brankač, J. (2013). Theta-associated high-frequency oscillations (110-160Hz) in the hippocampus and neocortex. *Prog. Neurobiol.* *100*, 1–14.
- Traub, R.D., Jefferys, J.G.R., and Whittington, M.A. (1997). Simulation of Gamma Rhythms in Networks of Interneurons and Pyramidal Cells. *J. Comput. Neurosci.* *4*, 141–150.
- Treves, A., and Rolls, E.T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus* *4*, 374–391.
- Vanderwolf, C.H. (1969). Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr. Clin. Neurophysiol.* *26*, 407–418.
- Vinogradova, O. s. (2001). Hippocampus as comparator: Role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus* *11*, 578–598.
- Vollenweider, F.X., Vollenweider-Scherpenhuyzen, M.F., Bäbler, A., Vogel, H., and Hell, D. (1998). Psilocybin induces schizophrenia-like psychosis in humans via a serotonin-2 agonist action. *Neuroreport* *9*, 3897–3902.
- Winston, J. (1972). Interspecies differences in the occurrence of theta. *Behav. Biol.* *7*, 479–487.

Winter, J.C., Rice, K.C., Amorosi, D.J., and Rabin, R.A. (2007). Psilocybin-induced stimulus control in the rat. *Pharmacol. Biochem. Behav.* *87*, 472–480.

Wood, E.R., Dudchenko, P.A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature* *397*, 613–616.

Wood, J., Kim, Y., and Moghaddam, B. (2012). Disruption of prefrontal cortex large scale neuronal activity by different classes of psychotomimetic drugs. *J. Neurosci. Off. J. Soc. Neurosci.* *32*, 3022–3031.

Xu, T., and Pandey, S.C. (2000). Cellular localization of serotonin(2A) (5HT(2A)) receptors in the rat brain. *Brain Res. Bull.* *51*, 499–505.