

---

# Sparse nonnegative deconvolution for compressive calcium imaging: algorithms and phase transitions

---

Eftychios A. Pnevmatikakis and Liam Paninski

Department of Statistics, Center for Theoretical Neuroscience  
Grossman Center for the Statistics of Mind, Columbia University, New York, NY  
{eftychios, liam}@stat.columbia.edu

## Abstract

We propose a compressed sensing (CS) calcium imaging framework for monitoring large neuronal populations, where we image randomized projections of the spatial calcium concentration at each timestep, instead of measuring the concentration at individual locations. We develop scalable nonnegative deconvolution methods for extracting the neuronal spike time series from such observations. We also address the problem of demixing the spatial locations of the neurons using rank-penalized matrix factorization methods. By exploiting the sparsity of neural spiking we demonstrate that the number of measurements needed per timestep is significantly smaller than the total number of neurons, a result that can potentially enable imaging of larger populations at considerably faster rates compared to traditional raster-scanning techniques. Unlike traditional CS setups, our problem involves a block-diagonal sensing matrix and a non-orthogonal sparse basis that spans multiple timesteps. We provide tight approximations to the number of measurements needed for perfect deconvolution for certain classes of spiking processes, and show that this number undergoes a “phase transition,” which we characterize using modern tools relating conic geometry to compressed sensing.

## 1 Introduction

Calcium imaging methods have revolutionized data acquisition in experimental neuroscience; we can now record from large neural populations to study the structure and function of neural circuits (see e.g. Ahrens et al. (2013)), or from multiple locations on a dendritic tree to examine the detailed computations performed at a subcellular level (see e.g. Branco et al. (2010)). Traditional calcium imaging techniques involve a raster-scanning protocol where at each cycle/timestep the microscope scans the image in a voxel-by-voxel fashion, or some other predetermined pattern, e.g. through random access multiphoton (RAMP) microscopy (Reddy et al., 2008), and thus the number of measurements per timestep is equal to the number of voxels of interest. Although this protocol produces “eye-interpretable” measurements, it introduces a tradeoff between the size of the imaged field and the imaging frame rate; very large neural populations can be imaged only with a relatively low temporal resolution.

This unfavorable situation can potentially be overcome by noticing that many acquired measurements are redundant; voxels can be “void” in the sense that no neurons are located there, and active voxels at nearby locations or timesteps will be highly correlated. Moreover, neural activity is typically sparse; most neurons do not spike at every timestep. During recent years, imaging practitioners have developed specialized techniques to leverage this redundancy. For example, Nikolenko et al. (2008) describe a microscope that uses a spatial light modulator and allows for the simultaneous imaging of different (predefined) image regions. More broadly, the advent of compressed sensing (CS) has found many applications in imaging such as MRI (Lustig et al., 2007), hyperspectral imaging (Gehm et al., 2007), sub-diffraction microscopy (Rust et al., 2006) and ghost imaging (Katz et al., 2009),

with available hardware implementations (see e.g. Duarte et al. (2008)). Recently, Studer et al. (2012) presented a fluorescence microscope based on the CS framework, where each measurement is obtained by projection of the whole image on a random pattern. This framework can lead to significant undersampling ratios for biological fluorescence imaging.

In this paper we propose the application of the imaging framework of Studer et al. (2012) to the case of neural population calcium imaging to address the problem of imaging large neural populations with high temporal resolution. The basic idea is to not measure the calcium at each location individually, but rather to take a smaller number of “mixed” measurements (based on randomized projections of the data). Then we use convex optimization methods that exploit the sparse structure in the data in order to simultaneously demix the information from the randomized projection observations and deconvolve the effect of the slow calcium indicator to recover the spikes. Our results indicate that the number of required randomized measurements scales merely with the number of expected spikes rather than the ambient dimension of the signal (number of voxels/neurons), allowing for the fast monitoring of large neural populations. We also address the problem of estimating the (potentially overlapping) spatial locations of the imaged neurons and demixing these locations using methods for nuclear norm minimization and nonnegative matrix factorization. Our methods scale linearly with the experiment length and are largely parallelizable, ensuring computational tractability. Our results indicate that calcium imaging can be potentially scaled up to considerably larger neuron populations and higher imaging rates by moving to compressive signal acquisition.

In the traditional static compressive imaging paradigm the sensing matrix is dense; every observation comes from the projection of all the image voxels to a random vector/matrix. Moreover, the underlying image can be either directly sparse (most of the voxels are zero) or sparse in some orthogonal basis (e.g. Fourier, or wavelet). In our case the sensing matrix has a block-diagonal form (we can only observe the activity at one specific time in each measurement) and the sparse basis (which corresponds to the inverse of the matrix implementing the convolution of the spikes from the calcium indicator) is non-orthogonal and spans multiple timelags. We analyze the effect of these distinctive features in Sec. 3 in a noiseless setting. We show that as the number of measurements increases, the probability of successful recovery undergoes a phase transition, and study the resulting phase transition curve (PTC), i.e., the number of measurements per timestep required for accurate deconvolution as a function of the number of spikes. Our analysis uses recent results that connect CS with conic geometry through the “statistical dimension” (SD) of descent cones (Amelunxen et al., 2013). We demonstrate that in many cases of interest, the SD provides a very good estimate of the PTC.

## 2 Model description and approximate maximum-a-posteriori inference

See e.g. Vogelstein et al. (2010) for background on statistical models for calcium imaging data. Here we assume that at every timestep an image or light field (either two- or three-dimensional) is observed for a duration of  $T$  timesteps. Each observed field contains a total number of  $d$  voxels and can be vectorized in a single column vector. Thus all the activity can be described by  $d \times T$  matrix  $F$ . Now assume that the field contains a total number of  $N$  neurons, where  $N$  is in general unknown. Each spike causes a rapid increase in the calcium concentration which then decays with a time constant that depends on the chemical properties of the calcium indicator. For each neuron  $i$  we assume that the “calcium activity”  $c_i$  can be described as a stable autoregressive process AR(1) process<sup>1</sup> that filters the neuron’s spikes  $s_i(t)$  according to the fast-rise slow-decay procedure described before:

$$c_i(t) = \gamma c_i(t-1) + s_i(t), \quad (1)$$

where  $\gamma$  is the discrete time constant which satisfies  $0 < \gamma < 1$  and can be approximated as  $\gamma = 1 - \exp(-\Delta t/\tau)$ , where  $\Delta t$  is the length of each timestep and  $\tau$  is the continuous time constant of the calcium indicator. In general we assume that each  $s_i(t)$  is binary due to the small length of the timestep in the proposed compressive imaging setting, and we use an i.i.d. prior for each neuron  $p(s_i(t) = 1) = \pi_i$ .<sup>2</sup> Moreover, let  $\mathbf{a}_i \in \mathbb{R}_+^d$  the (nonnegative) location vector for neuron  $i$ , and  $\mathbf{b} \in \mathbb{R}_+^d$  the (nonnegative) vector of baseline concentration for all the voxels. The spatial calcium concentration profile at time  $t$  can be described as

$$\mathbf{f}(t) = \sum_{i=1}^N \mathbf{a}_i c_i(t) + \mathbf{b}. \quad (2)$$

<sup>1</sup>Generalization to general AR( $p$ ) processes is straightforward, but we keep  $p = 1$  for simplicity.

<sup>2</sup>This choice is merely for simplicity; more general prior distributions can be incorporated in our framework.

In conventional raster-scanning experiments, at each timestep we observe a noisy version of the  $d$ -dimensional image  $\mathbf{f}(t)$ . Since  $d$  is typically large, the acquisition of this vector can take a significant amount of time, leading to a lengthy timestep  $\Delta t$  and low temporal resolution. Instead, we propose to observe the *projections* of  $\mathbf{f}(t)$  onto a random matrix  $B_t \in \mathbb{R}^{n \times d}$  (e.g. each entry of  $B_t$  could be chosen as 0 or 1 with probability 0.5):

$$\mathbf{y}(t) = B_t \mathbf{f}(t) + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma_t), \quad (3)$$

where  $\boldsymbol{\varepsilon}_t$  denotes measurement noise (Gaussian, with diagonal covariance  $\Sigma_t$ , for simplicity). If  $n = \dim(\mathbf{y}(t))$  satisfies  $n \ll d$ , then  $\mathbf{y}(t)$  represents a compression of  $\mathbf{f}(t)$  that can potentially be obtained more quickly than the full  $\mathbf{f}(t)$ . Now if we can use statistical methods to recover  $\mathbf{f}(t)$  (or equivalently the location  $\mathbf{a}_i$  and spikes  $s_i$  of each neuron) from the compressed measurements  $\mathbf{y}(t)$ , the total imaging throughput will be increased by a factor proportional to the undersampling ratio  $d/n$ . Our assumption here is that the random projection matrices  $B_t$  can be constructed quickly. Recent technological innovations have enabled this fast construction by using digital micromirror devices that enable spatial light modulation and can construct different excitation patterns with a high frequency (order of kHz). The total fluorescence can then be detected with a single photomultiplier tube. For more details we refer to Duarte et al. (2008); Nikolenko et al. (2008); Studer et al. (2012).

We discuss the statistical recovery problem next. For future reference, note that eqs. (1)-(3) can be written in matrix form as ( $\text{vec}(\cdot)$  denotes the vectorizing operator)

$$\begin{aligned} S &= CG^T \\ F &= AC + \mathbf{b}\mathbf{1}_T^T \quad \text{with } G = \begin{bmatrix} 1 & 0 & \dots & 0 \\ -\gamma & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -\gamma & 1 \end{bmatrix}, \quad B = \text{blkdiag}\{B_1, \dots, B_T\}. \quad (4) \\ \text{vec}(Y) &= B\text{vec}(F) + \boldsymbol{\varepsilon}, \end{aligned}$$

## 2.1 Approximate MAP inference with an interior point method

For now we assume that  $A$  is known. In general MAP inference of  $S$  is difficult due to the discrete nature of  $S$ . Following Vogelstein et al. (2010) we relax  $S$  to take continuous values in the interval  $[0, 1]$  (remember that we assume binary spikes), and appropriately modify the prior for  $s_i(t)$  to  $\log p(s_i(t)) \propto -(\lambda_i s_i(t)) \mathbf{1}(0 \leq s_i(t) \leq 1)$ , where  $\lambda_i$  is chosen such that the relaxed prior has the same mean  $\pi_i$ . To exploit the banded structure of  $G$  we seek the MAP estimate of  $C$  (instead of  $S$ ) by solving the following convex quadratic problem (we let  $\bar{\mathbf{y}}(t) = \mathbf{y}(t) - B_t \mathbf{b}$ )

$$\begin{aligned} \underset{C}{\text{minimize}} \quad & \sum_{t=1}^T \frac{1}{2} (\bar{\mathbf{y}}(t) - B_t A c(t))^T \Sigma_t^{-1} (\bar{\mathbf{y}}(t) - B_t A c(t)) - \log p(C) \\ \text{subject to} \quad & 0 \leq CG^T \leq 1, c(1) \geq 0, \end{aligned} \quad (\text{P-QP})$$

Using the prior on  $S$  and the relation  $S = CG^T$ , the log-prior of  $C$  can be written as  $\log p(C) \propto -\lambda^T CG^T \mathbf{1}_T$ . We can solve (P-QP) efficiently using an interior point method with a log-barrier (Vogelstein et al., 2010). The contribution of the likelihood term to the Hessian is a block-diagonal matrix, whereas the barrier-term will contribute a block-tridiagonal matrix where each non-zero block is diagonal. As a result the Newton search direction  $-H^{-1} \nabla$  can be computed efficiently in  $O(TN^3)$  time using a block version of standard forward-backward methods for tridiagonal systems of linear equations. We note that if  $N$  is large this can be inefficient. In this case we can use an augmented Lagrangian method (Boyd et al., 2011) to derive a fully parallelizable first order method, with  $O(TN)$  complexity per iteration. We refer to the supplementary material for additional details.

As a first example we consider a simple setup where all the parameters are assumed to be known. We consider  $N = 50$  neurons observed over  $T = 1000$  timesteps. We assume that  $A, \mathbf{b}$  are known, with  $A = I_N$  (corresponding to non-overlapping point neurons, with one neuron in each voxel) and  $\mathbf{b} = \mathbf{0}$ , respectively. This case of known point neurons can be thought as the compressive analog of RAMP microscopy where the neuron locations are predetermined and then imaged in a serial manner. (We treat the case of unknown and possibly overlapping neuron locations in section 2.2.) Each neuron was assumed to fire in an i.i.d. fashion with probability per timestep  $p = 0.04$ . Each measurement was obtained by projecting the spatial fluorescence vector at time  $t$ ,  $\mathbf{f}(t)$ , onto a random matrix  $B_t$ . Each row of  $B_t$  is taken as an i.i.d. normalized vector  $2\boldsymbol{\beta}/\sqrt{N}$ , where  $\boldsymbol{\beta}$  has i.i.d. entries following a fair Bernoulli distribution. For each set of measurements we assume that  $\Sigma_t = \sigma^2 I_n$ , and

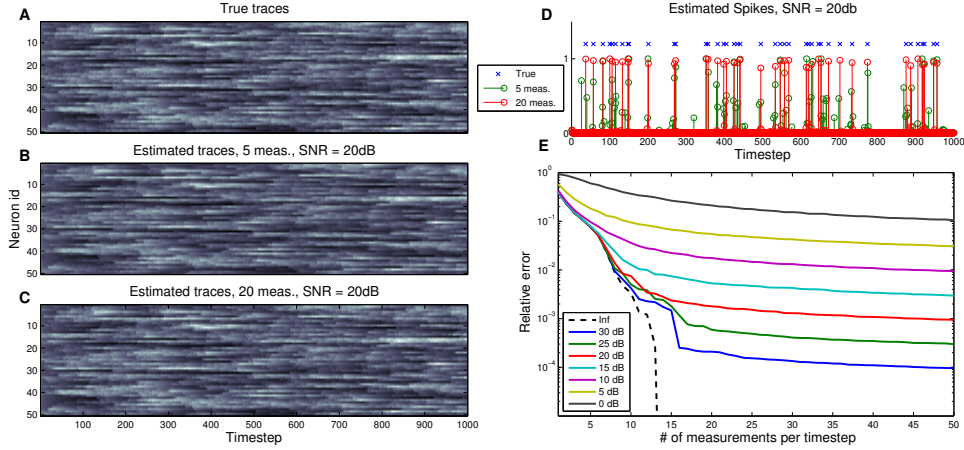


Figure 1: Performance of proposed algorithm under different noise levels. A: True traces, B: Estimated traces with  $n = 5$  ( $10\times$  undersampling), SNR = 20dB. C: Estimated traces with  $n = 20$  ( $2.5\times$  undersampling), SNR = 20dB. D: True and estimated spikes from the traces shown in panels B and C for a randomly selected neuron. E: Relative error between true and estimated traces for different number of measurements per timestep under different noise levels. The error decreases with the number of observations and the reconstruction is stable with respect to noise.

the signal-to-noise ratio (SNR) in dB is defined as  $\text{SNR} = 10 \log_{10}(\text{Var}[\beta^T \mathbf{f}(t)]/N\sigma^2)$ ; a quick calculation reveals that  $\text{SNR} = 10 \log_{10}(p(1-p)/(1-\gamma^2)\sigma^2)$ .

Fig. 1 examines the solution of (P-QP) when the number of measurements per timestep  $n$  varied from 1 to  $N$  and for 8 different SNR values 0, 5,  $\dots$ , 30 plus the noiseless case (SNR =  $\infty$ ). Fig. 1A shows the noiseless traces for all the neurons and panels B and C show the reconstructed traces for SNR = 20dB and  $n = 5, 20$  respectively. Fig. 1D shows the estimated spikes for these cases for a randomly picked neuron. For very small number of measurements ( $n = 5$ , i.e.,  $10\times$  undersampling) the inferred calcium traces (Fig. 1B) already closely resemble the true traces. However, the inferred MAP values of the spikes (computed by  $S = CG^T$ , essentially a differencing operation here) lie in the interior of  $[0, 1]$ , and the results are not directly interpretable at a high temporal resolution. As  $n$  increases ( $n = 20$ , red) the estimated spikes lie very close to  $\{0, 1\}$  and a simple thresholding procedure can recover the true spike times. In Fig. 1E the relative error between the estimated and true traces ( $\|C - \hat{C}\|_F / \|C\|_F$ , with  $\|\cdot\|_F$  denoting the the Frobenius norm) is plotted. In general the error decreases with the number of observations and the reconstruction is robust with noise. Finally, by observing the noiseless case (dashed curve) we see that when  $n \geq 13$  the error becomes practically zero indicating fully compressed acquisition of the calcium traces with a roughly  $4\times$  undersampling factor. We will see below that this undersampling factor is inversely proportional to the firing rate: we can recover highly sparse spike signals  $S$  using very few measurements  $n$ .

## 2.2 Estimation of the spatial matrix $A$

The above algorithm assumes that the underlying neurons have known locations, i.e., the matrix  $A$  is known. In some cases  $A$  can be estimated a-priori by running a conventional raster-scanning experiment at a high spatial resolution and locating the active voxels. However this approach is expensive and can still be challenging due to noise and possible spatial overlap between different neurons. To estimate  $A$  within the compressive framework we note that the baseline-subtracted spatiotemporal calcium matrix  $F$  (see eqs. (2) and (4)) can be written as  $\bar{F} = F - \mathbf{b}\mathbf{1}_T^T = AC$ ; thus  $\text{rank}(\bar{F}) \leq N$  where  $N$  is the number of underlying neurons, with typically  $N \ll d$ . Since  $N$  is also in general unknown we estimate  $\bar{F}$  by solving a nuclear norm penalized problem (Recht et al., 2010)

$$\begin{aligned}
 & \underset{\bar{F}}{\text{minimize}} && \sum_{t=1}^T \frac{1}{2} (\bar{\mathbf{y}}(t) - B_t \bar{\mathbf{f}}(t))^T \Sigma_t^{-1} (\bar{\mathbf{y}}(t) - B_t \bar{\mathbf{f}}(t)) - \log p(\bar{F}) + \lambda_{NN} \|\bar{F}\|_* \\
 & \text{subject to} && \bar{F} G^T \geq 0, \bar{\mathbf{f}}(1) \geq 0,
 \end{aligned} \tag{P-NN}$$

where  $\|\cdot\|_*$  denotes the nuclear norm (NN) of a matrix (i.e., the sum of its singular values), which is a convex approximation to the nonconvex rank function (Fazel, 2002). The prior of  $\bar{F}$  can be chosen in a similar fashion as  $\log p(C)$ , i.e.,  $\log p(\bar{F}) \propto -\lambda_F^T \bar{F} G^T \mathbf{1}_T$ , where  $\lambda_F \in \mathbb{R}^d$ . Although more complex than (P-QP), (P-NN) is again convex and can be solved efficiently using e.g. the ADMM method of Boyd et al. (2011). From the solution of (P-NN) we can estimate  $N$  by appropriately thresholding the singular values of the estimated  $\bar{F}$ .<sup>3</sup> Having  $N$  we can then use appropriately constrained nonnegative matrix factorization (NMF) methods to alternately estimate  $A$  and  $C$ . Note that during this NMF step the baseline vector  $\mathbf{b}$  can also be estimated jointly with  $A$ . Since NMF methods are nonconvex, and thus prone to local optima, informative initialization is important. We can use the solution of (P-NN) to initialize the spatial component  $A$  using clustering methods, similar to methods typically used in neuronal extracellular spike sorting (Lewicki, 1998). Details are given in the supplement (along with some discussion of the estimation of the other parameters in this problem); we refer to Pnevmatikakis et al. (2013) for full details.

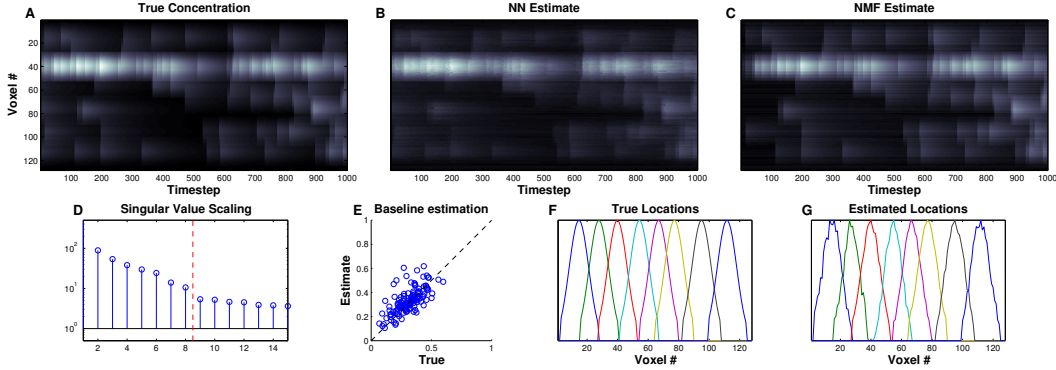


Figure 2: Estimating locations and calcium concentration from compressive calcium imaging measurements. A: True spatiotemporal concentration B: estimate by solving (P-NN) C: estimate by using NMF methods. D: Logarithmic plot of the first singular values of the solution of (P-NN), E: Estimation of baseline vector, F: true spatial locations G: estimated spatial locations. The NN-penalized method estimates the number of neurons and the NMF algorithm recovers the spatial and temporal components with high accuracy.

In Fig. 2 we present an application of this method to an example with  $N = 8$  spatially overlapping neurons. For simplicity we consider neurons in a one-dimensional field with total number of voxels  $d = 128$  and spatial positions shown in Fig. 2E. At each timestep we obtain just  $n = 5$  noisy measurements using random projections on binary masks. From the solution to the NN-penalized problem (P-NN) (Fig. 2B) we threshold the singular values (Fig. 2D) and estimate the number of underlying neurons (note the logarithmic gap between the 8th and 9th largest singular values that enables this separation). We then use the NMF approach to obtain final estimates of the spatial locations (Fig. 2G), the baseline vector (Fig. 2E), and the full spatiotemporal concentration (Fig. 2C). The estimates match well with the true values. Note that  $n < N \ll d$  showing that compressive imaging with significant undersampling factors is possible, even in the case of classical raster scanning protocol where the spatial locations are unknown.

### 3 Estimation of the phase transition curve in the noiseless case

The results presented above indicate that reconstruction of the spikes is possible even with significant undersampling. In this section we study this problem from a compressed sensing (CS) perspective in the idealized case where the measurements are noiseless. For simplicity, we also assume that  $A = I$  (similar to a RAMP setup). Unlike the traditional CS setup, where a sparse signal (in some basis) is sensed with a dense fully supported random matrix, in our case the sensing matrix  $B$  has a block-diagonal form. A standard justification of CS approaches proceeds by establishing that the sensing matrix satisfies the “restricted isometry property” (RIP) for certain classes of sparse signals

<sup>3</sup>To reduce potential shrinkage but promote low-rank solutions this “global” NN penalty can be replaced by a series of “local” NN penalties on spatially overlapping patches.

with high probability (w.h.p.); this property in turn guarantees the correct recovery of the parameters of interest (Candes and Tao, 2005). Yap et al. (2011) showed that for signals that are sparse in some orthogonal basis, the RIP holds for random block-diagonal matrices w.h.p. with a number of sufficient measurement that scales with the squared coherence between the sparse basis and the elementary (identity) basis. For non-orthogonal basis the RIP property has only been established for fully dense sensing matrices (Candes et al., 2011). For signals with sparse variations Ba et al. (2012) established perfect and stable recovery conditions under the assumption that the sensing matrix at each timestep satisfies certain RIPs, and the sparsity level at each timestep has known upper bounds.

While the RIP is a valuable tool for the study of convex relaxation approaches to compressed sensing problems, its estimates are usually up to a constant and can be relatively loose (Blanchard et al., 2011). An alternative viewpoint is offered from conic geometric arguments (Chandrasekaran et al., 2012; Amelunxen et al., 2013) that examine how many measurements are required such that the convex relaxed program will have a unique solution which coincides with the true sparse solution. We use this approach to study the theoretical properties of our proposed compressed calcium imaging framework in an idealized noiseless setting. When noise is absent, the quadratic program (P-QP) for the approximate MAP estimate converges to a linear program<sup>4</sup>:

$$\underset{C}{\text{minimize}} \quad f(C), \quad \text{subject to:} \quad B\text{vec}(C) = \text{vec}(Y) \quad (\text{P-LP})$$

$$\text{with} \quad f(C) = \begin{cases} (\mathbf{v} \otimes \mathbf{1}_N)^T \text{vec}(C), & (G \otimes I_d)\text{vec}(C) \geq 0 \\ \infty, & \text{otherwise} \end{cases}, \quad \text{and } \mathbf{v} = G^T \mathbf{1}_T.$$

Here  $\otimes$  denotes the Kronecker product and we used the identity  $\text{vec}(CG^T) = (G \otimes I_d)\text{vec}(C)$ . To examine the properties of (P-LP) we follow the approach of Amelunxen et al. (2013): For a fully dense sensing i.i.d. Gaussian (or random rotation) matrix  $B$ , the linear program (P-LP) will succeed w.h.p. to reconstruct the true solution  $C_0$ , if the total number of measurements  $nT$  satisfies

$$nT \geq \delta(\mathcal{D}(f, C_0)) + O(\sqrt{TN}). \quad (5)$$

$\mathcal{D}(f, C_0)$  is the descent cone of  $f$  at  $C_0$ , induced by the set of non-increasing directions from  $C_0$ , i.e.,

$$\mathcal{D}(f, C_0) = \cup_{\tau \geq 0} \{ \mathbf{y} \in \mathbb{R}^{N \times T} : f(C_0 + \tau \mathbf{y}) \leq f(C_0) \}, \quad (6)$$

and  $\delta(\mathcal{C})$  is the ‘‘statistical dimension’’ (SD) of a convex cone  $\mathcal{C} \subseteq \mathbb{R}^m$ , defined as the expected squared length of a standard normal Gaussian vector projected onto the cone

$$\delta(\mathcal{C}) = \mathbb{E}_{\mathbf{g}} \|\Pi_{\mathcal{C}}(\mathbf{g})\|^2, \quad \text{with } \mathbf{g} \sim \mathcal{N}(\mathbf{0}, I_m).$$

Eq. (5), and the analysis of Amelunxen et al. (2013), state that as  $TN \rightarrow \infty$ , the probability that (P-LP) will succeed to find the true solution undergoes a phase transition, and that the phase transition curve (PTC), i.e., the number of measurements required for perfect reconstruction normalized by the ambient dimension  $NT$  (Donoho and Tanner, 2009), coincides with the normalized SD. In our case  $B$  is a block-diagonal matrix (not a fully-dense Gaussian matrix), and the SD only provides an estimate of the PTC. However, as we show below, this estimate is tight in most cases of interest.

### 3.1 Computing the statistical dimension

Using a result from Amelunxen et al. (2013) the statistical dimension can also be expressed as the expected squared distance of a standard normal vector from the cone induced by the subdifferential (Rockafellar, 1970)  $\partial f$  of  $f$  at the true solution  $C_0$ :

$$\delta(\mathcal{D}(f, C_0)) = \mathbb{E}_{\mathbf{g}} \inf_{\tau > 0} \min_{\mathbf{u} \in \tau \partial f(C_0)} \|\mathbf{g} - \mathbf{u}\|^2, \quad \text{with } \mathbf{g} \sim \mathcal{N}(\mathbf{0}, I_{NT}). \quad (7)$$

Although in general (7) cannot be solved in closed form, it can be easily estimated numerically; in the supplementary material we show that the subdifferential  $\partial f(C_0)$  takes the form of a convex polytope, i.e., an intersection of linear half spaces. As a result, the distance of any vector  $\mathbf{g}$  from  $\partial f(C_0)$  can be found by solving a simple quadratic program, and the statistical dimension can be estimated with a simple Monte-Carlo simulation (details are presented in the supplement). The characterization

<sup>4</sup>To illustrate the generality of our approach we allow for arbitrary nonnegative spike values in this analysis, but we also discuss the binary case that is of direct interest to our compressive calcium framework.

of (7) also explains the effect of the sparsity pattern on the SD. In the case where the sparse basis is the identity then the cone induced by the subdifferential can be decomposed as the union of the respective subdifferential cones induced by each coordinate. It follows that the SD is invariant to coordinate permutations and depends only on the sparsity level, i.e., the number of nonzero elements. However, this result is in general not true for a nonorthogonal sparse basis, indicating that the precise location of the spikes (sparsity pattern) and not just their number has an effect on the SD. In our case the calcium signal is sparse in the non-orthogonal basis described by the matrix  $G$  from (4).

### 3.2 Relation with the phase transition curve

In this section we examine the relation of the SD with the PTC for our compressive calcium imaging problem. Let  $S$  denote the set of spikes,  $\Omega = \text{supp}(S)$ , and  $C$  the induced calcium traces  $C = SG^{-T}$ . As we argued, the statistical dimension of the descent cone  $\mathcal{D}(f, C)$  depends both on the cardinality of the spike set  $|\Omega|$  (sparsity level) and the location of the spikes (sparsity pattern). To examine the effects of the sparsity level and pattern we define the normalized expected statistical dimension (NESD) with respect to a certain distribution (e.g. Bernoulli)  $\chi$  from which the spikes  $S$  are drawn.

$$\tilde{\delta}(k/NT, \chi) = \mathbb{E}_{\Omega \sim \chi} [\delta(\mathcal{D}(f, C))/NT], \quad \text{with } \text{supp}(S) = \Omega, \text{ and } \mathbb{E}_{\Omega \sim \chi} |\Omega| = k.$$

In Fig. 3 we examine the relation of the NESD with the phase transition curve of the noiseless problem (P-LP). We consider a setup with  $N = 40$  point neurons ( $A = I_d, d = N$ ) observed over  $T = 50$  timesteps and chose discrete time constant  $\gamma = 0.99$ . The spike-times of each neuron came from the same distribution and we considered two different distributions: (i) Bernoulli spiking, i.e., each neuron fires i.i.d. spikes with the probability  $k/T$ , and (ii) desynchronized periodic spiking where each neuron fires deterministically spikes with discrete frequency  $k/T$  timesteps<sup>-1</sup>, and each neuron has a random phase. We considered two forms of spikes: (i) with nonnegative values ( $s_i(t) \geq 0$ ), and (ii) with binary values ( $s_i(t) = \{0, 1\}$ ), and we also considered two forms of sensing matrices: (i) with time-varying matrix  $B_t$ , and (ii) with constant, fully supported matrices  $B_1 = \dots = B_T$ . The entries of each  $B_t$  are again drawn from an i.i.d. fair Bernoulli distribution. For each of these 8 different conditions we varied the expected number of spikes per neuron  $k$  from 1 to  $T$  and the number of observed measurements  $n$  from 1 to  $N$ . Fig. 3 shows the empirical probability that the program (P-LP) will succeed in reconstructing the true solution averaged over a 100 repetitions. Success is declared when the reconstructed spike signal  $\hat{S}$  satisfies<sup>5</sup>  $\|\hat{S} - S\|_F / \|S\|_F < 10^{-3}$ . We also plot the empirical PTC (purple dashed line), i.e., the empirical 50% success probability line, and the NESD (solid blue line), approximated with a Monte Carlo simulation using 200 samples, for each of the four distinct cases (note that the SD does not depend on the structure of the sensing matrix  $B$ ).

The results indicate that in all cases, our problem undergoes a sharp phase transition as the number of measurements per timestep varies: in the white regions of Fig. 3,  $S$  is recovered essentially perfectly, with a sharp transition to a high probability of at least some errors in the black regions. Note that these phase transitions are defined as functions of the signal sparsity index  $k/T$ ; the signal sparsity sets the compressibility of this data.

In addition, in the case of time-varying  $B_t$ , the NESD provides a surprisingly good estimate of the PTC, especially in the binary case or when the spiking signal is actually sparse ( $k/T < 0.5$ ), a result that justifies our overall approach. Although using time-varying sensing matrices  $B_t$  leads to better results, compression is also possible with a constant  $B$ . This is an important result for implementation purposes where changing the sensing matrix might be a costly or slow operation. On a more technical side we also observe the following interesting properties:

- Periodic spiking requires more measurements for accurate deconvolution, a property again predicted by the SD. This comes from the fact that the sparse basis is not orthogonal and shows that for a fixed sparsity level  $k/T$  the sparsity pattern also affects the number of required measurements. This difference depends on the time constant  $\gamma$ . As  $\gamma \rightarrow 0$ ,  $G \rightarrow I$ ; the problem becomes equivalent to a standard nonnegative CS problem, where the spike pattern is irrelevant.
- In the binary case the results exhibit a symmetry around the axis  $k/T = 0.5$ . In fact this symmetry becomes exact as  $\gamma \rightarrow 1$ . In the supplement we prove that this result is predicted by the SD.

<sup>5</sup>Note that when calculating the reconstruction error we excluded the last 10 timesteps. As every spike is filtered by the AR process it has an effect for multiple timelags in the future and an optimal encoder (with fixed number of measurements per timestep) has to sense each spike over multiple timelags. The number of excluded timesteps depends only on  $\gamma$  and not on the length  $T$ , and thus this behavior becomes negligible as  $T \rightarrow \infty$ .

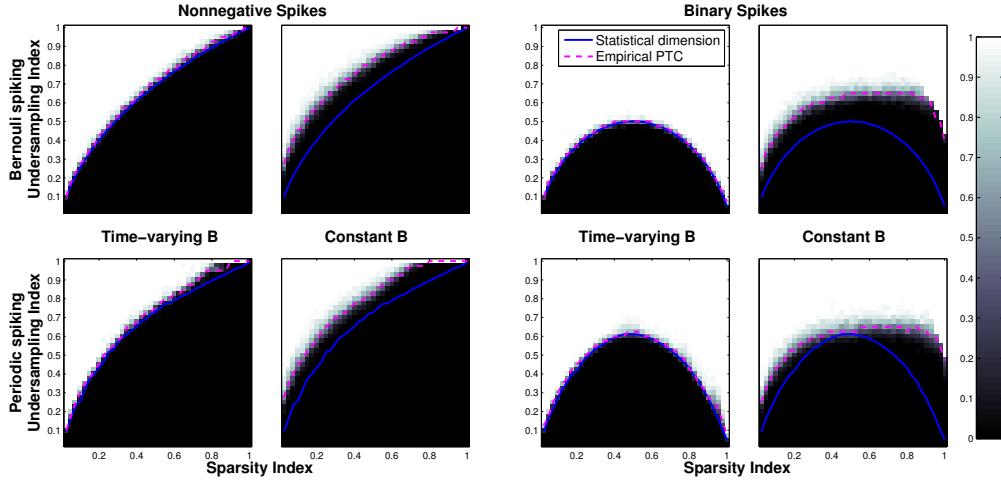


Figure 3: Relation of the statistical dimension with the phase transition curve for two different spiking distributions (Bernoulli, periodic), two different spike values (nonnegative, binary), and two classes of sensing matrices (time-varying, constant). For each panel: x-axis normalized sparsity  $k/T$ , y-axis undersampling index  $n/N$ . Each panel shows the empirical success probability for each pair  $(k/T, n/N)$ , the empirical 50%-success line (dashed purple line) and the SD (blue solid line). When  $B$  is time-varying the SD provides a good estimate of the empirical PTC.

- In the Bernoulli spiking nonnegative case, the SD is numerically very close to the PTC of the standard nonnegative CS problem (not shown here), adding to the growing body of evidence for universal behavior of convex relaxation approaches to CS (Donoho and Tanner, 2009).

As mentioned above, our analysis is only approximate since  $B$  is block-diagonal and not fully dense. However, this approximation is tight in the time-varying case. Still, it is possible to construct adversarial counterexamples where the SD approach fails to provide a good estimate of the PTC. For example, if all neurons fire in a completely synchronized manner then the required number of measurements grows at a rate that is not predicted by (5). We present such an example in the supplement and note that more research is needed to understand such extreme cases.

## 4 Conclusion

We proposed a framework for compressive calcium imaging. Using convex relaxation tools from compressed sensing and low rank matrix factorization, we developed an efficient method for extracting neurons' spatial locations and the temporal locations of their spikes from a limited number of measurements, enabling the imaging of large neural populations at potentially much higher imaging rates than currently available. We also studied a noiseless version of our problem from a compressed sensing point of view using newly introduced tools involving the statistical dimension of convex cones. Our analysis can in certain cases capture the number of measurements needed for perfect deconvolution, and helps explain the effects of different spike patterns on reconstruction performance.

Our approach suggests potential improvements over the standard raster scanning protocol (unknown locations) as well as the more efficient RAMP protocol (known locations). However our analysis is idealistic and neglects several issues that can arise in practice. The results of Fig. 1 suggest a tradeoff between effective compression and SNR level. In the compressive framework the cycle length can be relaxed more easily due to the parallel nature of the imaging (each location is targeted during the whole "cycle"). The summed activity is then collected by the photomultiplier tube that introduces the noise. While the nature of this addition has to be examined in practice, we expect that the observed SNR will allow for significant compression. Another important issue is motion correction for brain movement, especially in vivo conditions. While new approaches have to be derived for this problem, the novel approach of Cotton et al. (2013) could be adaptable to our setting. We hope that our work



will inspire experimentalists to leverage the proposed advanced signal processing methods to develop more efficient imaging protocols.

## A Parallel first order implementation using the ADMM method

The quadratic problem (P-QP) for finding the MAP estimate (repeated here for simplicity)

$$\begin{aligned} \underset{C}{\text{minimize}} \quad & \sum_{t=1}^T \frac{1}{2} (\bar{\mathbf{y}}(t) - B_t \mathbf{c}(t))^T \Sigma_t^{-1} (\bar{\mathbf{y}}(t) - B_t \mathbf{c}(t)) + \boldsymbol{\lambda}^T C G^T \mathbf{1}_T \\ \text{subject to} \quad & C G^T \geq 0, \mathbf{c}(1) \geq 0, \end{aligned} \quad (\text{P-QP})$$

can be also expressed as

$$\begin{aligned} \underset{C, Z}{\text{minimize}} \quad & \sum_{t=1}^T \frac{1}{2} (\bar{\mathbf{y}}(t) - B_t \mathbf{c}(t))^T \Sigma_t^{-1} (\bar{\mathbf{y}}(t) - B_t \mathbf{c}(t)) + \boldsymbol{\lambda}^T Z G^T \mathbf{1}_T \\ \text{subject to} \quad & C = Z, C Z^T \geq 0, \mathbf{z}(1) \geq 0. \end{aligned}$$

Using the alternate direction method of multipliers (ADMM) (Boyd et al., 2011) we solve the problem using the following iterative scheme until convergence:

$$\begin{aligned} C^{k+1} &= \arg \min_C \sum_{t=1}^T \frac{1}{2} (\bar{\mathbf{y}}(t) - B_t \mathbf{c}(t))^T \Sigma_t^{-1} (\bar{\mathbf{y}}(t) - B_t \mathbf{c}(t)) + (\rho/2) \|C - Z^k + U^k\|^2 \\ Z^{k+1} &= \arg \min_{Z: Z G^T \geq 0, \mathbf{z}(1) \geq 0} \boldsymbol{\lambda}^T Z G^T \mathbf{1}_T + (\rho/2) \|C^{k+1} - Z + U^k\|^2 \\ U^{k+1} &= U^k + C^{k+1} - Z^{k+1}, \end{aligned}$$

with  $\rho > 0$ . Now for  $C^{k+1}$  each  $\mathbf{c}(t)$  can be estimated in parallel by solving a simple unconstrained quadratic program. The Hessian of each of these programs is of the form  $H_t = \rho I_d + B_t^T \Sigma_t^{-1} B_t$  and therefore can be inverted in  $O(Nn^2)$  time via the application of the Woodbury matrix identity (remember that each  $B_t \in \mathbb{R}^{n \times d}$ ). Similarly, the estimation of  $Z^{k+1}$  can be split into  $N$  parallel programs each of which determines a row of  $Z^{k+1}$  with cost  $O(T)$  via a plain log-barrier method, similar to Vogelstein et al. (2010). Unlike the interior point method described in Sec. 2.1, the ADMM method is a first order method and typically requires more iterations for convergence. However, it may be a method of choice in the case where the spatial dimensionality is high and parallel computing resources are available.

## B Parameter estimation details

### B.1 Choice of number of neurons

To estimate the number of neurons we compute the vector  $\boldsymbol{\sigma}$  of the singular values of  $\hat{F}$ , where  $\hat{F}$  is the solution of the NN-penalized problem (P-NN). We then construct the vector of consecutive ratios  $\mathbf{a}$  with  $a(i) = \sigma(i)/\sigma(i+1)$  and find the local minima of  $\mathbf{a}$ . We pick  $N$  as the location of the first local minimum of  $\mathbf{a}$  such that the  $N$  chosen singular values capture a large fraction (e.g. 99%) of  $\|\hat{F}\|_F^2$ . We have observed that this method in general works well in practice.

### B.2 NMF step

We first note that the baseline vector  $\mathbf{b}$  can be jointly estimated with the matrix  $A$ , since it can be incorporated as an additional column of  $A$  that is multiplied with an additional row of  $C$  where each entry is equal to 1, i.e., the spatiotemporal calcium matrix can be written as

$$F = [A, \mathbf{b}] \begin{bmatrix} C \\ \mathbf{1}_T^T \end{bmatrix}.$$

Given  $A$  and  $\mathbf{b}$ , the calcium traces  $C$  for all the neurons can be estimated by solving the (P-QP) problem presented in Sec. 2. Given  $C$ , the log-likelihood of  $[A, \mathbf{b}]$  can be expressed as

$$\log(Y|A, \mathbf{b}; C) \propto -\frac{1}{2} \|\Sigma^{-1/2} B((C^T \otimes I_d) \text{vec}(A) + (\mathbf{1}_T \otimes I_d) \mathbf{b} - \text{vec}(Y))\|^2. \quad (8)$$

To jointly estimate  $A$  and  $\mathbf{b}$  we can maximize (8) subject to nonnegativity constraints  $A \geq 0, \mathbf{b} \geq \mathbf{0}$ . Additional regularizers can be introduced to  $A$  to penalize the size of each neuron (via an  $l_1$ -norm on  $A$ ) or to smooth the shape of each neuron (via a spatial derivative penalty on each row of  $A$ ). Note that since  $A \geq 0$  and  $\mathbf{b} \geq 0$ , the  $l_1$ -norm just becomes the sum of the elements and therefore the problem remains a nonnegative least squares problem that can be solved efficiently (e.g. with a log-barrier method).

To initialize the NMF procedure we start from the solution of the NN-penalized problem (P-NN) and initialize the spatial component as follows:

1. Extract the ‘‘voxel’’ spikes from the solution of problem (P-NN)

$$\hat{S} = \hat{F} G^T$$

2. Threshold the extracted spikes with a relatively high threshold (e.g. 90% quantile).
3. Perform  $k$ -means clustering of the spike vectors at each timestep, with  $N + 1$  clusters and discard the cluster with the centroid closest to zero.
4. Use the remaining clusters to initialize  $A^0$ .
5. The baseline vector can be initialized by computing the remainder between the solution  $\hat{F}$  and the extracted spikes and taking the mean for each voxel.

A full presentation of these methods will be pursued elsewhere (Pnevmatikakis et al., 2013).

## C Computing the statistical dimension

We first rewrite the function  $f$  for convenience:

$$f(C) = \begin{cases} (\mathbf{v} \otimes \mathbf{1}_N)^T \text{vec}(C), & (G \otimes I_d) \text{vec}(C) \geq 0 \\ \infty, & \text{otherwise} \end{cases}, \text{ and } \mathbf{v} = G^T \mathbf{1}_T.$$

We can express  $\delta(\mathcal{D}(f, C))$  in terms of the SDs of the descent cones induced by the rows of  $C$ . Let  $\mathbf{c}_i$  denote the  $i$ -th row of  $C$  (in column format) and define the function  $f_r(\mathbf{c}) = \mathbf{v}^T \mathbf{c}$ , for  $G\mathbf{c} \geq \mathbf{0}$  (and  $f_r(\mathbf{c}) = \infty$  otherwise). Note that  $f(C) = \sum_{i=1}^N f_r(\mathbf{c}_i)$ . Now if  $\mathcal{D}(f_r, \mathbf{c}_i)$  is the descent cone of  $f_r$  at  $\mathbf{c}_i$ , since the sparsifying basis does not multiplex the spikes of the different neurons we have

$$\mathcal{D}(f, C) = \mathcal{D}(f_r, \mathbf{c}_1) \times \mathcal{D}(f_r, \mathbf{c}_2) \times \dots \times \mathcal{D}(f_r, \mathbf{c}_N) \implies \delta(\mathcal{D}(f, C)) = \sum_{i=1}^N \delta(\mathcal{D}(f_r, \mathbf{c}_i)), \quad (9)$$

where the last equality follows from a property of the SD of direct product cones (Amelunxen et al., 2013). Using the expression of the statistical dimension in terms of the subdifferential (7), we have

$$\delta(\mathcal{D}(f_r, \mathbf{c})) = \mathbb{E}_{\mathbf{g}} \inf_{\tau > 0} \min_{\mathbf{u} \in \tau \partial f_r(\mathbf{c})} \|\mathbf{g} - \mathbf{u}\|^2.$$

To characterize the subdifferential consider the function  $z : \mathbb{R}^T \mapsto \mathbb{R}$  with  $z(\mathbf{x}) = \mathbf{1}_T^T \mathbf{x}$  if  $\mathbf{x} \geq \mathbf{0}$  (pointwise), and  $z(\mathbf{x}) = \infty$  otherwise. Then if  $\Omega$  is the set of entries where  $\mathbf{x}$  is nonzero, and  $\Omega^c$  its complement, we have

$$\partial z(\mathbf{x}) = \mathbf{w}, \quad \text{with } \begin{cases} w(j) = 1, & j \in \Omega, \\ w(j) \leq 1, & j \in \Omega^c \end{cases}.$$

For our case note that  $f_r(\mathbf{c}) = z(G\mathbf{c})$ . Let  $\Omega_i$  the set of spiketimes of neuron  $i$  and  $\Omega_i^c$  its complement in the set  $[1, \dots, T]$ . Using the relation for affine transformations of the subdifferential  $\partial f_r(\mathbf{c}) = G^T \partial z(G\mathbf{c})$ , the subdifferential at  $\mathbf{c}_i, \partial f_r(\mathbf{c}_i)$ , can be characterized as

$$\partial f_r(\mathbf{c}_i) = G^T \mathbf{w}, \quad \text{with } \begin{cases} w(j) = 1, & j \in \Omega_i \\ w(j) \leq 1, & j \in \Omega_i^c \end{cases} \quad (10)$$

Using (7) the SD equals the average value over all standard normal i.i.d. vectors  $\mathbf{g}$  of the quadratic programs

$$\underset{\mathbf{w}, \tau}{\text{minimize}} \|\mathbf{g} - G^T \mathbf{w}\|^2, \quad \text{subject to: } \tau \geq 0, \{w(j) = \tau, j \in \Omega_i\}, \{w(j) \leq \tau, j \in \Omega_i^c\}. \quad (\text{SD-QP})$$

The statistical dimension can be estimated with a simple Monte Carlo simulation by averaging the values of a series of quadratic programs. Since  $G$  is a bidiagonal matrix, each of the quadratic programs can be solved in  $O(T)$  time using a standard interior point method. Note that the sparsity pattern here matters as opposed to the function  $z(\mathbf{x})$  where the expected distance from the subdifferential cone is invariant under permutations.

## D Symmetry of the statistical dimension in the binary case

The above analysis assumes that the spikes signal  $S$  takes arbitrary nonnegative values. In the case where the spikes are restricted to take only binary  $\{0, 1\}$  values, a similar analysis can be carried: By inserting the additional condition  $G\mathbf{c} \leq \mathbf{1}$ , i.e.,  $f_r(\mathbf{c}_i) = \infty$  if  $G\mathbf{c}_i \not\leq \mathbf{1}$ , the subdifferential  $\partial f_r(\mathbf{c}_i)$  is now given by

$$\partial f_r(\mathbf{c}_i) = G^T \mathbf{w}, \quad \text{with} \quad \begin{cases} w(j) \geq 1, & j \in \Omega_i \\ w(j) \leq 1, & j \in \Omega_i^c \end{cases}. \quad (11)$$

Note that (11) is very similar to (10) with the difference that  $w(j) \geq 1$  for  $j \in \Omega_i$ .

When  $\gamma = 1$  the objective function  $f_r$  becomes a nonnegative total-variation norm. We now prove the statistical dimension with is asymptotically symmetric. Let two binary spiking signals  $\mathbf{s}_1, \mathbf{s}_2$  that satisfy  $\mathbf{s}_1 = \mathbf{1} - \mathbf{s}_2$  and also let  $\Omega_1, \Omega_2$  the corresponding sets of spiketimes, and  $\mathbf{c}_1, \mathbf{c}_2$  the corresponding calcium traces. It holds that  $\Omega_1 = [1, \dots, T] \setminus \Omega_2$ . Define  $\mathcal{C}_1, \mathcal{C}_2$  the set of conic constraints that these signals induce when computing the statistical dimension of the descent cones (eq. (SD-QP)):

$$\begin{aligned} \mathcal{C}_1 &= \{\mathbf{w}, \tau : (\tau \geq 0) \cap (w(j) \geq \tau, j \in \Omega_1) \cap (w(j) \leq \tau, j \in \Omega_2)\} \\ \mathcal{C}_2 &= \{\mathbf{w}, \tau : (\tau \geq 0) \cap (w(j) \geq \tau, j \in \Omega_2) \cap (w(j) \leq \tau, j \in \Omega_1)\} \end{aligned}$$

Now define the functions  $h_i : \mathbb{R}^T \mapsto \mathbb{R}_+$  ( $i = 1, 2$ ) as the value of the quadratic problems

$$h_i(\mathbf{g}) = \min_{\mathbf{w}, \tau \in \mathcal{C}_i} \|\mathbf{g} - G^T \mathbf{w}\|^2. \quad (12)$$

By making the change of variables  $\mathbf{w} \leftarrow \mathbf{w} - \tau$ , the set of constraints and the value functions become respectively

$$\begin{aligned} \tilde{\mathcal{C}}_1 &= \{\mathbf{w}, \tau : (\tau \geq 0) \cap (w(j) \geq 0, j \in \Omega_1) \cap (w(j) \leq 0, j \in \Omega_2)\} \\ \tilde{\mathcal{C}}_2 &= \{\mathbf{w}, \tau : (\tau \geq 0) \cap (w(j) \geq 0, j \in \Omega_2) \cap (w(j) \leq 0, j \in \Omega_1)\}, \end{aligned}$$

and the value functions of (12) can be expressed as

$$h_i(\mathbf{g}) = \min_{\mathbf{w}, \tau \in \tilde{\mathcal{C}}_i} \|\mathbf{g} - G^T \mathbf{w} + \tau G^T \mathbf{1}_T\|^2.$$

If  $\gamma = 1$  we have  $G^T \mathbf{1}_T = [0, 0, \dots, 0, 1]^T$  and therefore we argue that the contribution of the term  $\tau G^T \mathbf{1}_T$  in the value function is of order  $O(1)$  and that we have

$$h_i(\mathbf{g}) = \bar{h}_i(\mathbf{g}) + O(1),$$

where  $\bar{h}_i$  is defined as

$$\bar{h}_i(\mathbf{g}) = \min_{\mathbf{w} \in \tilde{\mathcal{C}}_i} \|\mathbf{g} - G^T \mathbf{w}\|^2,$$

with

$$\begin{aligned} \tilde{\mathcal{C}}_1 &= \{\mathbf{w} : (w(j) \geq 0, j \in \Omega_1) \cap (w(j) \leq 0, j \in \Omega_2)\} \\ \tilde{\mathcal{C}}_2 &= \{\mathbf{w} : (w(j) \geq 0, j \in \Omega_2) \cap (w(j) \leq 0, j \in \Omega_1)\} \end{aligned}$$

Now note that  $\tilde{\mathcal{C}}_1 = -\tilde{\mathcal{C}}_2$  and therefore  $\bar{h}_1(\mathbf{g}) = \bar{h}_2(-\mathbf{g})$ . Therefore

$$\delta(\mathcal{D}(f_r, \mathbf{c}_1)) = \mathbb{E}_{\mathbf{g}}(h_1(\mathbf{g})) = \mathbb{E}_{\mathbf{g}}(\bar{h}_1(\mathbf{g})) + O(1) = \mathbb{E}_{\mathbf{g}}(\bar{h}_2(\mathbf{g})) + O(1) = \delta(\mathcal{D}(f_r, \mathbf{c}_2)) + O(1)$$

and in the asymptotic regime

$$\lim_{T \rightarrow \infty} \frac{1}{T} \delta(\mathcal{D}(f_r, \mathbf{c}_1)) = \lim_{T \rightarrow \infty} \frac{1}{T} \delta(\mathcal{D}(f_r, \mathbf{c}_2)),$$

which establishes our symmetry argument.

## E Counterexamples

As the mentioned in the main text, the condition (5)

$$nT \geq \delta(\mathcal{D}(f, C_0)) + O(\sqrt{NT}), \quad (13)$$

is exact only when  $B$  is a fully dense i.i.d. random matrix. Here we construct a counterexample where the statistical dimension fails to provide a tight lower bound on the phase transition curve. Consider again  $N$  neurons. Each neuron has Bernoulli spiking with probability of spike per timestep  $k/T$ , but now assume that all the neurons are perfectly synchronized. From (9), the statistical dimension  $\delta(\mathcal{D}(f, C))$  can be decomposed as

$$\delta(\mathcal{D}(f, C)) = \sum_{i=1}^N \delta(\mathcal{D}(f_r, c_i)),$$

i.e., it does not depend on the spike correlations between the different neurons. However we expect that the synchrony between the neurons will make the reconstruction harder. For example, when the sensing matrix  $B$  is constant and the spikes have nonnegative values, then compressive acquisition is impossible since the dense calcium signal will always projected in the same lower-dimensional subspace spanned by the constant  $B$ , and thus it cannot be recovered.

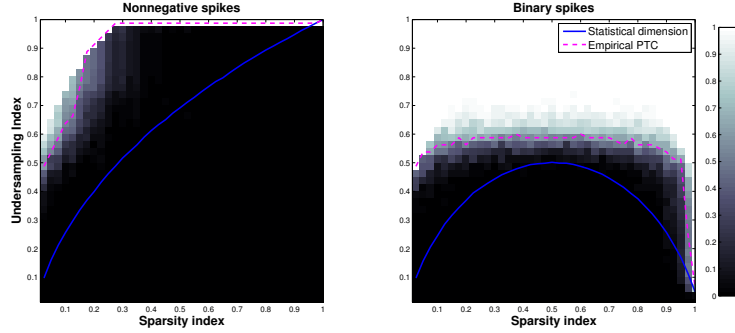


Figure 4: Relation of the statistical dimension with the phase transition curve in a synchronized firing case. For each panel: x-axis normalized sparsity  $k/T$ , y-axis undersampling index  $n/N$ . Left: Nonnegative spikes. Right: Binary spikes. The statistical dimension (solid blue line) does not provide a good estimate of the empirical PTC (purple dashed line) for this case.

In Fig. 4 we consider a time-varying sensing matrix  $B$  and examine the cases where the spikes take nonnegative values (left) and binary values (right). We again considered  $N = 40$  and  $T = 50$  and removed the last 10 timesteps from the calculation of the reconstruction error. We again varied the probability of spiking  $k/T$  from 0 to 1 and the number of measurements per timestep  $n$  from 1 to  $N$ . In both cases we see that the statistical dimension fails to provide a tight bound on the empirical PTC. However, note that compression is still possible for the nonnegative spikes case, if the signal is very sparse, a result which indicates that occasional synchronies can be tolerated within the proposed compressive framework. In the binary case, compression is always possible, a result that is expected from the relatively simple structure of binary signals, and is important for our compressive calcium imaging framework, where the spiking signal is expect to be binary (present or absent).

Note that the failure of the SD to capture the PTC should not be attributed to the decomposition property (9) which ignores the correlations between the different neurons, but rather on the block-diagonal structure of the sensing matrix  $B$ . If  $B$  is fully dense, then the SD coincides with the PTC as is predicted in general result of Amelunxen et al. (2013) (data not shown). More research is needed to understand the effects of restricting  $B$  to be a block-diagonal matrix.

## References

Ahrens, M. B., M. B. Orger, D. N. Robson, J. M. Li, and P. J. Keller (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods* 10(5), 413–420.

- Amelunxen, D., M. Lotz, M. B. McCoy, and J. A. Tropp (2013). Living on the edge: A geometric theory of phase transitions in convex optimization. *arXiv preprint arXiv:1303.6672*.
- Ba, D., B. Babadi, P. Purdon, and E. Brown (2012). Exact and stable recovery of sequences of signals with sparse increments via differential  $l_1$ -minimization. In *Advances in Neural Information Processing Systems* 25, pp. 2636–2644.
- Blanchard, J. D., C. Cartis, and J. Tanner (2011). Compressed sensing: How sharp is the restricted isometry property? *SIAM review* 53(1), 105–125.
- Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning* 3(1), 1–122.
- Branco, T., B. A. Clark, and M. Häusser (2010). Dendritic discrimination of temporal input sequences in cortical neurons. *Science* 329, 1671–1675.
- Candes, E. J., Y. C. Eldar, D. Needell, and P. Randall (2011). Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis* 31(1), 59–73.
- Candes, E. J. and T. Tao (2005). Decoding by linear programming. *Information Theory, IEEE Transactions on* 51(12), 4203–4215.
- Chandrasekaran, V., B. Recht, P. A. Parrilo, and A. S. Willsky (2012). The convex geometry of linear inverse problems. *Foundations of Computational Mathematics* 12(6), 805–849.
- Cotton, R. J., E. Froudarakis, P. Storer, P. Saggau, and A. S. Tolia (2013). Three-dimensional mapping of microcircuit correlation structure. *Frontiers in Neural Circuits* 7.
- Donoho, D. and J. Tanner (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367(1906), 4273–4293.
- Duarte, M. F., M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk (2008). Single-pixel imaging via compressive sampling. *Signal Processing Magazine, IEEE* 25(2), 83–91.
- Fazel, M. (2002). *Matrix rank minimization with applications*. Ph. D. thesis, Stanford University.
- Gehm, M., R. John, D. Brady, R. Willett, and T. Schulz (2007). Single-shot compressive spectral imaging with a dual-disperser architecture. *Opt. Express* 15(21), 14013–14027.
- Katz, O., Y. Bromberg, and Y. Silberberg (2009). Compressive ghost imaging. *Applied Physics Letters* 95(13).
- Lewicki, M. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems* 9, R53–R78.
- Lustig, M., D. Donoho, and J. M. Pauly (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic resonance in medicine* 58(6), 1182–1195.
- Nikolenko, V., B. Watson, R. Araya, A. Woodruff, D. Peterka, and R. Yuste (2008). SLM microscopy: Scanless two-photon imaging and photostimulation using spatial light modulators. *Frontiers in Neural Circuits* 2, 5.
- Pnevmatikakis, E., T. Machado, L. Grosenick, B. Poole, J. Vogelstein, and L. Paninski (2013). Rank-penalized nonnegative spatiotemporal deconvolution and demixing of calcium imaging data. In *Computational and Systems Neuroscience Meeting COSYNE*.
- Recht, B., M. Fazel, and P. Parrilo (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review* 52(3), 471–501.
- Reddy, G., K. Kelleher, R. Fink, and P. Saggau (2008). Three-dimensional random access multiphoton microscopy for functional imaging of neuronal activity. *Nature Neuroscience* 11(6), 713–720.
- Rockafellar, R. (1970). *Convex Analysis*. Princeton University Press.
- Rust, M. J., M. Bates, and X. Zhuang (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature methods* 3(10), 793–796.
- Studer, V., J. Bobin, M. Chahid, H. S. Mousavi, E. Candes, and M. Dahan (2012). Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proceedings of the National Academy of Sciences* 109(26), E1679–E1687.
- Vogelstein, J., A. Packer, T. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski (2010). Fast non-negative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology* 104(6), 3691–3704.
- Yap, H. L., A. Eftekhari, M. B. Wakin, and C. J. Rozell (2011). The restricted isometry property for block diagonal matrices. In *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*, pp. 1–6.