

## RESEARCH ARTICLE

# A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology

Ziqiang Wei<sup>1,2</sup>, Bei-Jung Lin<sup>1,3</sup>, Tsai-Wen Chen<sup>1,3</sup>, Kayvon Daie<sup>1</sup>, Karel Svoboda<sup>1\*</sup>, Shaul Druckmann<sup>1,4\*</sup>

**1** Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia, the United States of America, **2** The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland, the United States of America, **3** Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan, **4** Department of Neurobiology, Stanford University, Stanford, California, the United States of America

\* [svobodak@janelia.hhmi.org](mailto:svobodak@janelia.hhmi.org) (K.S.); [shaul@stanford.edu](mailto:shaul@stanford.edu) (S.D.)



## OPEN ACCESS

**Citation:** Wei Z, Lin B-J, Chen T-W, Daie K, Svoboda K, Druckmann S (2020) A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. *PLoS Comput Biol* 16(9): e1008198. <https://doi.org/10.1371/journal.pcbi.1008198>

**Editor:** Boris S. Gutkin, École Normale Supérieure, Collège de France, CNRS, FRANCE

**Received:** April 6, 2020

**Accepted:** July 27, 2020

**Published:** September 15, 2020

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1008198>

**Copyright:** © 2020 Wei et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All data are available at both figshare and <http://im-phys.org/data>. Precompiled data used in the paper can be download at <https://doi.org/10.6084/m9.figshare>.

## Abstract

Calcium imaging with fluorescent protein sensors is widely used to record activity in neuronal populations. The transform between neural activity and calcium-related fluorescence involves nonlinearities and low-pass filtering, but the effects of the transformation on analyses of neural populations are not well understood. We compared neuronal spikes and fluorescence in matched neural populations in behaving mice. We report multiple discrepancies between analyses performed on the two types of data, including changes in single-neuron selectivity and population decoding. These were only partially resolved by spike inference algorithms applied to fluorescence. To model the relation between spiking and fluorescence we simultaneously recorded spikes and fluorescence from individual neurons. Using these recordings we developed a model transforming spike trains to synthetic-imaging data. The model recapitulated the differences in analyses. Our analysis highlights challenges in relating electrophysiology and imaging data, and suggests forward modeling as an effective way to understand differences between these data.

## Author summary

Many studies in neuroscience revolve around understanding the patterns of activity of neurons and their relation to behavior. To be able to address such questions one must first record the activity of neurons. Broadly speaking, two different approaches are commonly used, each with its own advantages and disadvantages. Imaging can sample neural activity of hundreds of neurons in a local area and can be targeted to specific cell-types. But it does not record activity directly, reporting it rather through a transformation from intracellular calcium. Electrophysiological recordings report neural activity directly with high temporal precision but have limitations of their own such as being less likely to accurately pickup neurons with low activity. We compared neuronal spikes and fluorescence recorded in matched neural populations in behaving mice performing the same task. We

12786296.v1. The new raw data in precompiled data include (1) simultaneously ephys-imaging data in TG mice in primary visual cortex (passive viewing task) that were recorded by B.J.L. and (2) imaging data in 6f-TG mice in anterior lateral motor cortex (delayed discrimination task) that were recorded by K.D. and roi-extracted by Z.W. manually; both of them can be downloaded at <https://doi.org/10.6084/m9.figshare.12792587>. All codes for model benchmarks and comparison metrics are recompiled and packed with data through im-phys-API (<https://github.com/zqwei/lm-phys-API>), which can be available at <http://im-phys.org/codes>. The API will come with a user-friendly interface in which one can reproduce all results in our paper and extensive results on <http://im-phys.org>. We also provide repos for benchmarks of S2F and F2S models at <https://github.com/zqwei/Ca-Imaging-Deconv-List> (DOI: 10.5281/zenodo.3960635) and comparison metrics at <https://github.com/zqwei/Neural-Recording-Methodology-Comparison> (DOI: 10.5281/zenodo.3979786) and website interface at <https://github.com/zqwei/lm-phys-org>.

**Funding:** T.W.C is supported by a career development grant (NHRI-ex-105-10509NC) from the Taiwan National Health Research Institute. K.D, K.S and S.D are funded by the Simons foundation collaboration on the global brain SCGB 542969SPI; Z.W is funded by SCGB 542943SPI; S.D is funded by NIH NS104781. K.S. is funded by HHMI, K.S. and S.D. funded by the visitor program at Janelia Research Campus. 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.

report multiple discrepancies between analyses performed on the two types of data at the single neuron and population level. We developed a model transforming spike trains to synthetic-imaging data which recapitulated many of the differences in analyses. Our analysis highlights challenges in relating electrophysiology and imaging data, and suggests forward modeling as an effective way to predict and understand differences between them.

## Introduction

Electrophysiological recordings ('ephys') and calcium imaging offer distinct tradeoffs for interrogating activity in neural populations. Ephys directly reports spiking of neurons with high signal-to-noise ratio, temporal fidelity, and dynamic range, but typically offers access only to a sparse subset of relatively active neurons[1]. In addition, the ability to track the same population of neurons across time, important for understanding the neural basis of learning, remains challenging[2–4].

Calcium imaging provides access to large numbers of neurons simultaneously[5–8], potentially with cell type specificity[9,10]. Moreover, calcium imaging can track the activity of the same neuronal populations over time[6,11]. Indeed, with the development of sensitive fluorescent protein-based indicators[12–18] and powerful new imaging methods[7,19] calcium imaging has been rapidly adopted for measurements of neural population activity.

However, calcium imaging reports spikes only indirectly[10,20]. The transformation from spikes to calcium is inherently non-linear due to the dynamics of intracellular calcium concentrations[21]. Additional nonlinearities are imposed by the protein-based indicators of calcium [12–14,22]. Together these produce a low-pass filtered, delayed, and transformed version of neural activity, which complicates relating neural activity to behavior. Calcium imaging also has lower signal-to-noise ratio for detecting spikes and limited dynamic range[10]. In addition, during animal behavior, spike rates can vary by orders of magnitude across behavioral epochs and across neurons, even neurons of the same type[23–25] and spike rates change over times of milliseconds to seconds[25–27]. Finally, the coupling between spikes and calcium-dependent fluorescence likely differs across different neuron types and even individual neurons within a type[13,28].

The complexities in the relation between spiking activity and calcium imaging at the level of single neurons have been long appreciated[12–14,21,22,29]. However, the effect of these factors on analyses of population activity are not fully known[30]. Ideally a detailed understanding of the transformation from spikes-to-calcium-dependent fluorescence would allow inversion of this transformation and the reliable extraction of spikes. Calcium indicators with high sensitivity allow reliable detection of action potentials, at least under conditions when single spikes or burst of spikes are separated in time[13,31,32]. However, under behaviorally relevant conditions neurons operate with a large range of spike rates, and spiking responses are typically superposed on a substantial background spike rate, which varies across the population[23,25]. Moreover, neuron-to-neuron variability in calcium dynamics, calcium indicator dynamics and patterns of firing rate could conspire to make this inversion challenging. These issues are compounded by the paucity of simultaneously recorded spikes and fluorescence data. It is therefore unclear if spike inference can invert the fluorescence data accurately to eliminate potential discrepancies between analyses performed on ephys and calcium imaging data.

Here, we explore these issues empirically in data collected in a decision-making task, where the dynamics of the neural circuit are rich and variable across neurons. In particular, neurons

in frontal cortex show a wide range of spike rates and exhibit diverse temporal dynamics and selectivity[25,26,33]. We first analyzed ephys and calcium imaging data, recorded separately but measured in matched neuronal populations in the same delayed response task, and directly compared the results of standard measurements of selectivity and population dynamics. We find qualitative discrepancies at both the level of single cells and neural populations. Spike inference algorithms were limited in resolving these differences. However, a phenomenological model of the spike-to-fluorescence transformation, based on a new set of simultaneous imaging and electrophysiology data[13,14], explains many differences across the data sets. Finally, we developed a web-based platform, [im-phys.org](http://im-phys.org), that allows quantification of the effects of various transformations from electrophysiology to imaging.

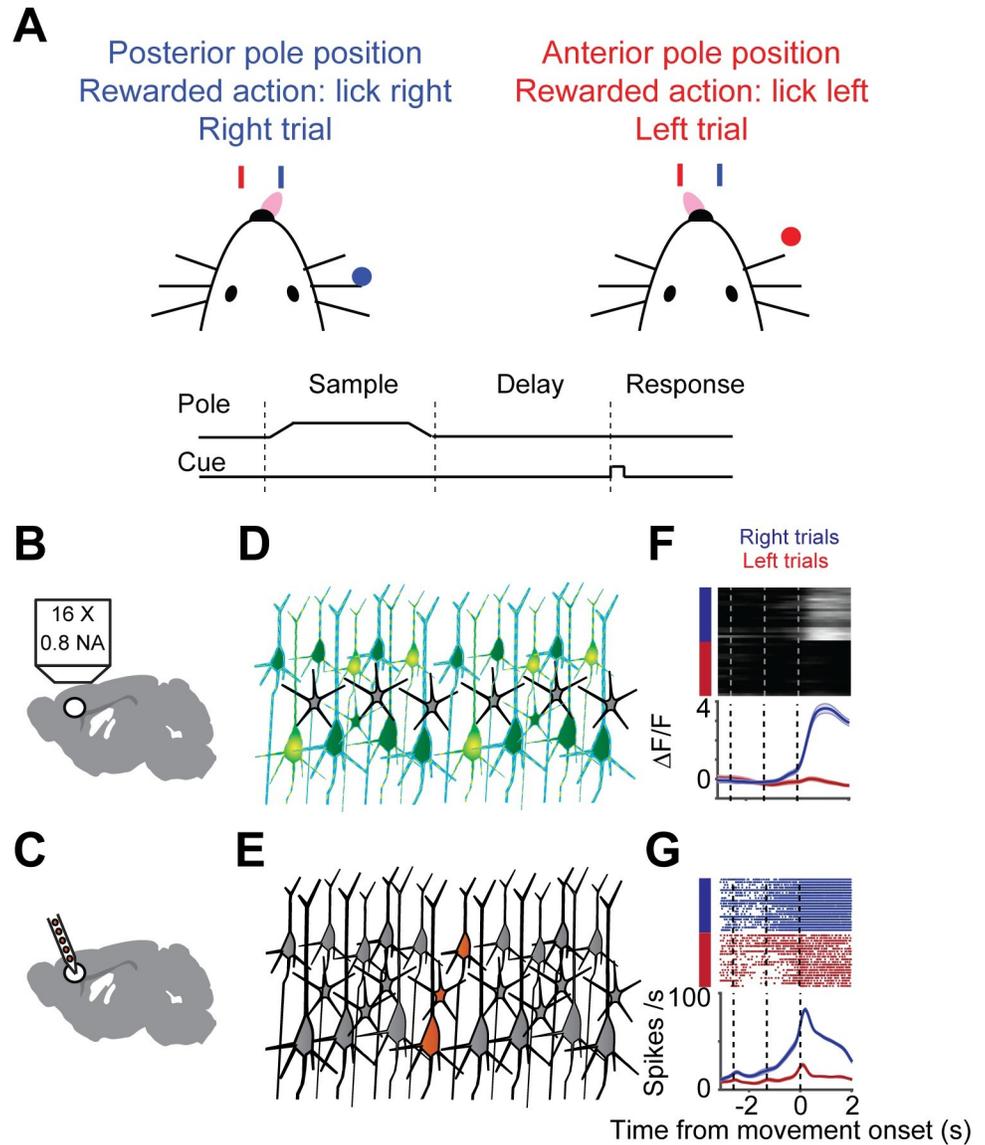
## Results

We measured neural activity using electrophysiology ('ephs') and calcium imaging under identical behavioral conditions and in matched neural populations, but in separate experiments. Mice performed a tactile delayed response task[25,34,35] (Fig 1A). In each trial, mice judged the location of an object with their whiskers. During the subsequent delay epoch (approximately 1.3 seconds), mice planned an upcoming response. Following an auditory 'go' cue, mice reported object location with directional licking (lick-left or lick-right).

Two-photon calcium imaging and ephys were performed in left anterolateral motor cortex (ALM; Fig 1B, 1D and 1F). We report the results of three variants of calcium indicators in this study: GCaMP6s delivered by viral gene transfer, and GCaMP6s and GCaMP6f expressed in Thy-1 transgenic mice. In the first series of imaging experiments, neurons were transduced with adeno-associated virus expressing GCaMP6s (6s-AAV), a widely-used method[10,13] (data from[25], 1493 neurons, 4 mice). In the second, neural activity was recorded by imaging transgenic mice expressing GCaMP6s in cortical pyramidal neurons (6s-TG, data from[36], 2293 neurons, 1 mouse). We treated these datasets separately since the mode of delivery of GCaMP can affect its properties. Specifically transgenic GCaMP typically results in neurons that have lower GCaMP expression levels and faster fluorescence dynamics compared to neurons transduced with AAV[31]. Finally, we collected a dataset obtained with a faster, but less sensitive indicator, GCaMP6f (6f-TG, 2672 neurons, 2 mice). We refer to these three datasets as 6s-AAV, 6s-Tg and 6f-Tg, respectively. The 6s-TG data, though containing a large number of neurons across multiple behavioral sessions, came from a single animal. We compared this data to ephys data acquired with silicon probes that record multiple neurons simultaneously (720 neurons, 19 mice[25]) (Fig 1C, 1E and 1G). Ephys recordings were subsampled so that their recording depths matched the generally more superficial calcium imaging experiments. Neurons were recorded by 6s-AAV and 6s-Tg at 120–740  $\mu\text{m}$ . The matched ephys subset was taken at 100–800  $\mu\text{m}$  leaving 720 neurons. Neurons were recorded by 6f-Tg at 140–470  $\mu\text{m}$ . The matched ephys subset was taken at 100–470  $\mu\text{m}$ , leaving 225 neurons (S1 Table).

### Filtering of selectivity by calcium imaging

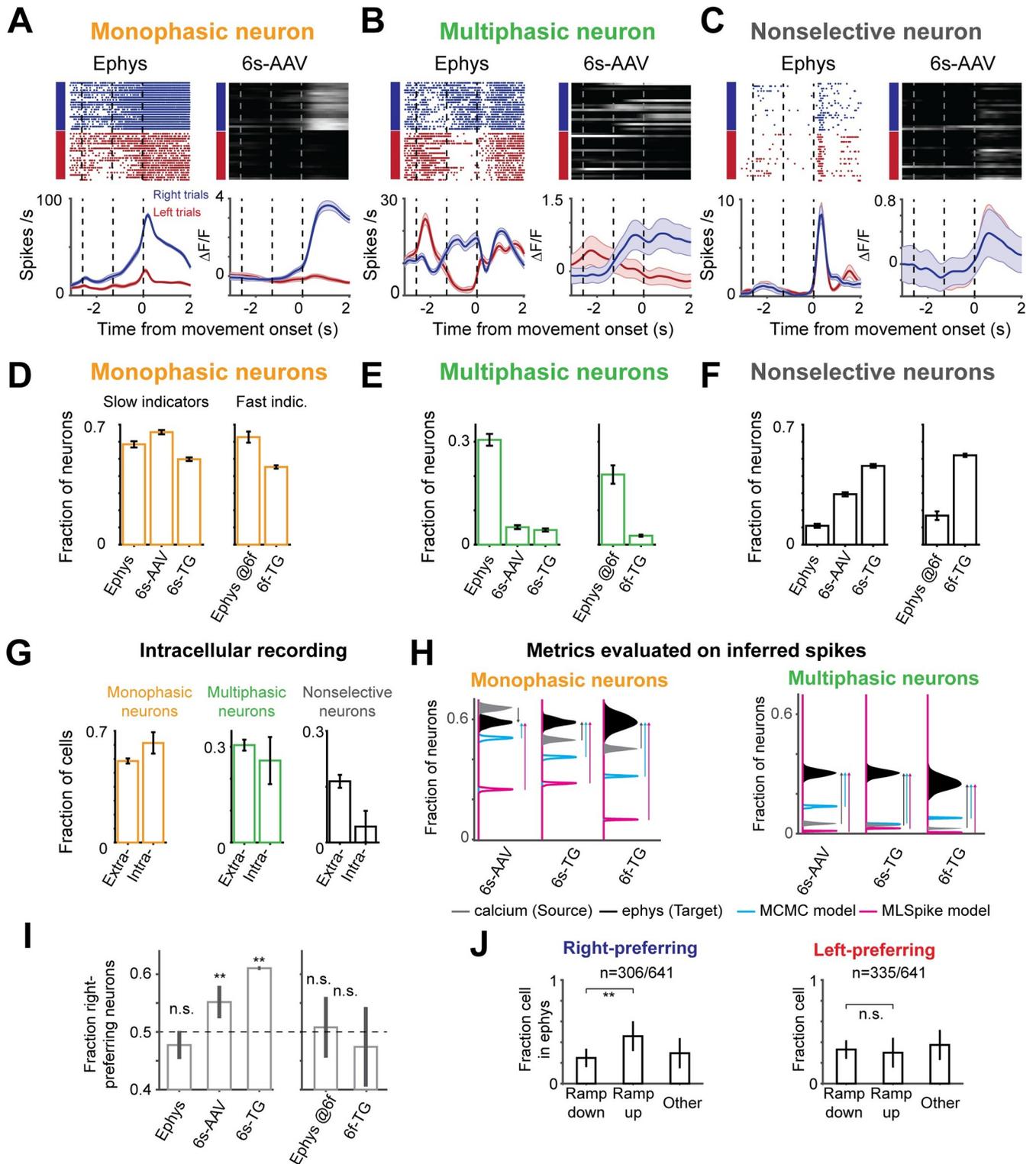
Individual ALM neurons exhibit diverse temporal dynamics, including changes in selectivity over time (Fig 2B)[25,33,34]. We classified dynamics into three categories: 'monophasic' neurons showed consistent selectivity across the trial (Fig 2A); 'multiphasic' neurons changed selectivity over time (defined as having consistent selectivity for at least 335 ms which then changes and remains stable for at least 335 ms more) (Fig 2B); 'non-selective' neurons responded similarly across trial types but were still modulated during the task (Fig 2C). The proportion of monophasic selective neurons was similar between the datasets (58% ephys; 66% 6s-AAV; 50% 6s-Tg; 45% 6f-Tg). However, the ephys data set contained a substantial



**Fig 1. Illustration of sampling population activity in anterior lateral motor cortex using imaging and electrophysiology.** **A.** Delayed-response, two alternative forced-choice task. Mice discriminated a pole position (anterior or posterior) and reported it by directional licking (lick right, blue; lick left, red) after a delay period. End of delay period was signaled by an auditory cue **B.** Schematic of imaging setup. **C.** Schematic of electrophysiological setup. **D.** Schematic of neurons sampled by imaging (green). **E.** Schematic of sampled neurons by electrophysiology (orange). **F.** Example neuron, imaging. Top, individual trials (blue, right trial; red, left trial). Bottom, mean activity (mean, thick line; sem., shaded area). **G.** Example neuron, electrophysiology. Top, raster plot. Bottom, peri-stimulus time histogram (PSTH).

<https://doi.org/10.1371/journal.pcbi.1008198.g001>

proportion of multiphasic neurons (220/720; 31%), much larger than the imaging datasets (6s-AAV: 76/1493, 5%; 6s-Tg, 98/2293, 4%; compare to matched ephys, 220/720, 31%; 6f-Tg, 69/2672, 3%; compare to matched ephys, 52/225, 20%;  $p < .001$ ,  $\chi^2$  test; **Fig 2D–2F**). As neural response properties can change across cortical layers, we performed a more detailed analysis of the effect of recording depth on single neuron selectivity and find that selectivity was reduced in imaging compared to ephys across depths (**S1A–S1D Fig**).



**Fig 2. Single neuron trial-type selectivity differs between imaging and ephys.** A. Example neurons with monophasic selectivity. Left, ephys; right, imaging. B. Same as A for multiphasic neurons. C. Same as A for non-selective neurons. D-F. Fraction of selective neurons in depth-matched ephys (“ephys @6f” indicates depth matched to the more superficial 6f-TG recordings) and when imaged with 6s-AAV, 6s-TG, or 6f-TG. D. Fractions of monophasic neurons. E. Fraction of multiphasic neurons. F. Fraction of nonselective. G. Proportion of multiphasic neurons in intracellular recordings is similar to that in extracellular recordings. Bar shows fraction of neurons in each of the categories for extracellular (left) and intracellular (right) ephys. H. Effect of spike inference on

estimates of fractions of monophasic (left) and multiphasic (right) neurons. The distribution of fraction of neurons for imaging data (source data), is given in gray for 6s-AAV (top), 6s-TG (middle) and 6f-TG (bottom). The distribution for ephys (target data) is in black. Distributions from inferred spike rates from MCMC (40) are in cyan and for MLSpoke (42) are in magenta. Arrows denote the difference between the imaging data and ephys data (gray arrow) or inferred ephys and ephys data (cyan arrow for MCMC and magenta arrow for MLSpoke). I. Fraction of right-preferring neurons in the different datasets divided into slow indicators (left) and fast indicators (right). J. Bar plot of fractions of ramp-down, ramp-up and 'other' cells in ephys for right-preferring (left) and left-preferring neurons (right).

<https://doi.org/10.1371/journal.pcbi.1008198.g002>

What could account for this difference? Ephys records a sparse subset of neurons blindly, which could introduce biases, for example a bias towards neurons with higher spike rates (S1E–S1I Fig). In contrast, in our imaging experiments all visualized neurons were analyzed. In addition, the spike sorting procedure used to identify units from raw electrode potentials can introduce artifacts, including erroneous merging of neurons (S1J Fig). However, we found that these factors were unlikely to explain our results for two reasons. First, we considered intracellular recordings for which spike sorting is not required[37] (we used only trials without photostimulation and excluded one cell due to the small number of trials; see [Materials and Methods](#)). The fraction of multiphasic neurons was 25.7% ( $n = 9/35$ ), similar to the extracellular ephys data ( $p = .67$ ,  $\chi^2$  test; Fig 2G) but significantly different from the imaging data ( $p < .001$ ,  $\chi^2$  test with both 6s-AAV and 6s-Tg). Second, we tested the question of spike-sorting induced biases by considering synthetic data in which we deliberately introduced merges at different probabilities. Merging neurons generated more multi-selective neurons when two neurons with different temporal selectivity profiles were merged, but the ratio of accidental merging had to be high (i.e., more than 10% of the neurons need to be completely conflated with another neuron) to explain the difference between datasets (S1J Fig). This suggests that the above differences are not driven by ephys being biased towards different populations of neurons than imaging.

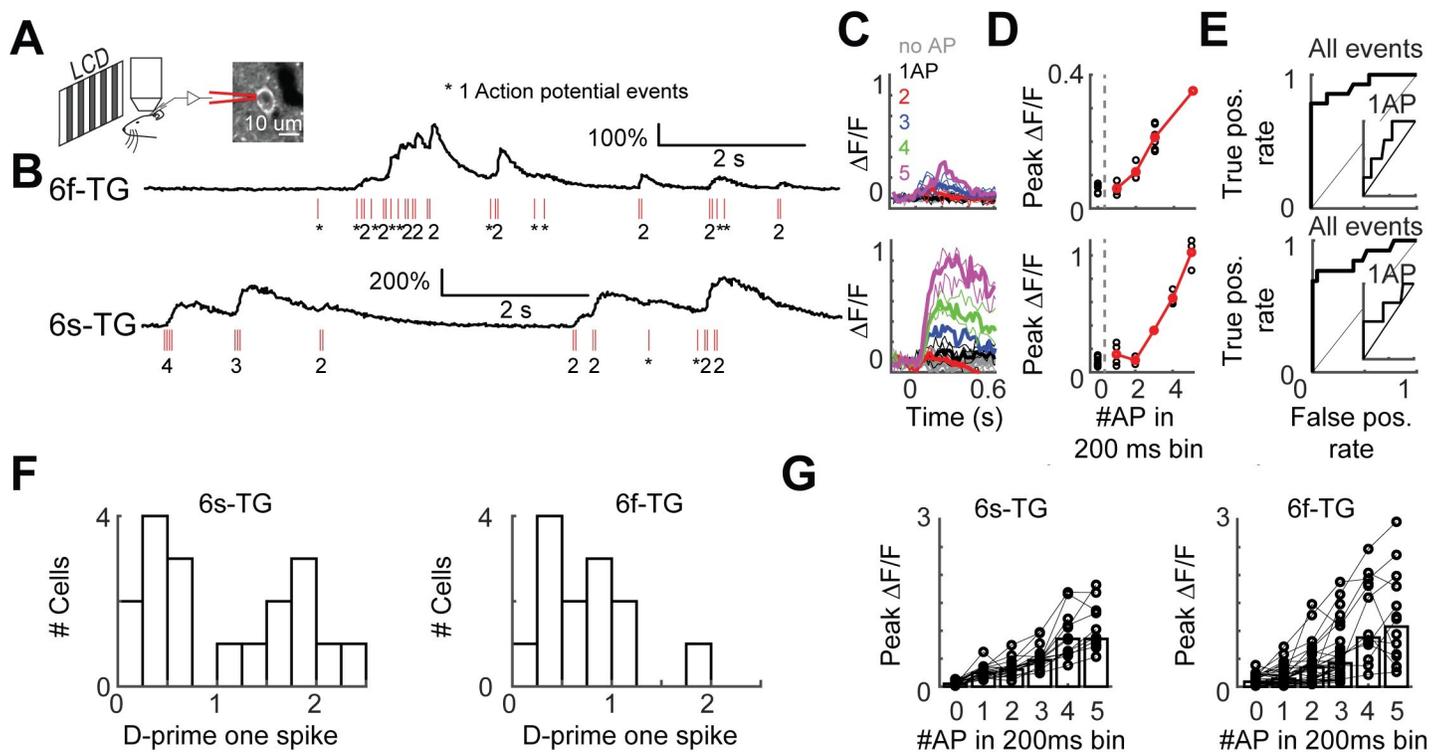
Another source of difference could be that our comparisons so far were performed on the imaging data, not on spike rates inferred from the imaging data. Spike inference algorithms attempt to undo the transformation from spikes to calcium-dependent fluorescence, thereby recovering spike times (or spike rates) from imaging data[32,38–43]. We tested two high-performing published methods: MLSpoke[42] and MCMC[40]. We also tested seven additional models (<http://im-phys.org/analyses>, for additional comparisons between inference techniques see[44]). In our hands, spike inference only partially corrected the differences between the datasets and in some cases actually pushed the data even further apart (Fig 2H). For instance, MLSpoke produced even lower proportions of multiphasic neurons. MCMC was more accurate, increasing the proportions of multiphasic neurons, but still far short of the actual proportion in the ephys dataset (and for 6s-Tg and 6f-Tg decreased instead of increased the proportion of monophasic neurons). Deconvolution at best recovered about half of the missing multiphasic selectivity (6s-AAV, 18%; 6s-TG, 17%, compared to 31% in matched ephys; 6f-TG, 8%, compared to 20% in matched ephys; Fig 2H).

Differences between calcium and ephys were not limited to the temporal nature of responses but were also present in trial-type selectivity. In the ephys dataset, right-preferring neurons (i.e., neurons whose firing rate before right licks was higher than before left licks) were as common as left-preferring neurons (Fig 2I, left;  $p = .118$ ,  $\chi^2$  test)[25,34]. The same was true for imaging with a fast calcium indicator (Fig 2I, right;  $p = .102$ ,  $\chi^2$  test), but not for imaging with slow indicators (Fig 2I, left;  $p < .001$ ,  $\chi^2$  test; S4B Fig, spike-inference measure). What could be the cause of these differences? Spike rates in individual ALM neurons often increase or decrease during a trial in ramp-like patterns[25,34,45]. Right-preferring selectivity was more often associated with neurons ramping up on right trials, whereas left-preferring selectivity included many neurons with firing rates ramping down in the non-preferred

(‘right’) trial (Fig 2J). The large difference between the rise and decay times of calcium indicators could lead to differences in how the selectivity of neurons that ramp up or ramp down gets transformed by the indicator. To test to what degree such explanations explain the data we developed a model of the spike-to-fluorescence transformation.

### Simultaneous loose-seal electrophysiology and calcium imaging

Modeling the spike-to-fluorescence transformation requires simultaneous electrophysiology and calcium imaging at the level of individual neurons. Since this data was not available for the transgenic calcium indicators used here we performed loose-seal recordings and calcium imaging in individual neurons (Fig 3). The dataset consists of GCaMP6f- and GCaMP6s-expressing L2/3 neurons in transgenic mice (6s-TG, 22 cells; 6f-TG, 18 cells; S2 Table). This new data, which we make publicly available, more than doubles the number of currently publicly available simultaneously recorded neurons[32]. In addition we used published data with AAV-based gene transduction[13] (<http://dx.doi.org/10.6080/K02R3PMN>). Bursts of spikes produced fluorescence transients in the imaged neurons (Fig 3A–3D; top, 6f-TG cell; bottom, 6s-TG cell). Peak fluorescence responses increased monotonically with the number of spikes. The ability to detect single spikes varied considerably between neurons (Figs 3E–3G and S2).



**Fig 3. Simultaneous loose-seal recordings and calcium imaging of layer 2/3 pyramidal neurons *in vivo*.** A. Illustration of the recording setup. Transgenic mice expressing GCaMP6s (GP4.3) or GCaMP6f (GP5.17) were lightly anesthetized and viewed drifting grating visual stimuli. GCaMP-expressing L2/3 neurons were recorded in the loose-seal mode during calcium imaging. B. Example recordings from neurons expressing GCaMP6f (top, 6f-TG) and GCaMP6s (bottom, 6s-TG). Red ticks, spikes. C. Traces of fluorescence dynamics following different numbers of action potentials (APs) for example neurons. Top, 6f-TG; bottom, 6s-TG. Gray, no AP; black, a single AP; red, 2 APs; blue, 3 APs; green, 4 APs; magenta, 5 APs. Thin lines, single trials; thick lines, average. D. Peak fluorescence increases as a function of the number of spikes in 200 ms bins. Black, single trials; red, trial average. E. ROC curve of all spike events. Inner panel, ROC curve for single AP events. F. Distribution of d-prime for single spikes across cells. Left, 6s-TG; right, 6f-TG. G. Mean peak fluorescence changes as a function of number of spikes in 200 ms time intervals across cells. Left, 6s-TG; right, 6f-TG. Each circle corresponds to a recorded neuron. Bars indicate average.

<https://doi.org/10.1371/journal.pcbi.1008198.g003>

The detection of single spikes was lower in transgenic mice than with AAV-based gene transduction, likely reflecting the lower expression level in the transgenic mice. Detection of spikes varied across neurons and, as expected, was better for 6s-TG than 6f-TG (Fig 3F). We used these recordings to build models of the spike to fluorescence transformation.

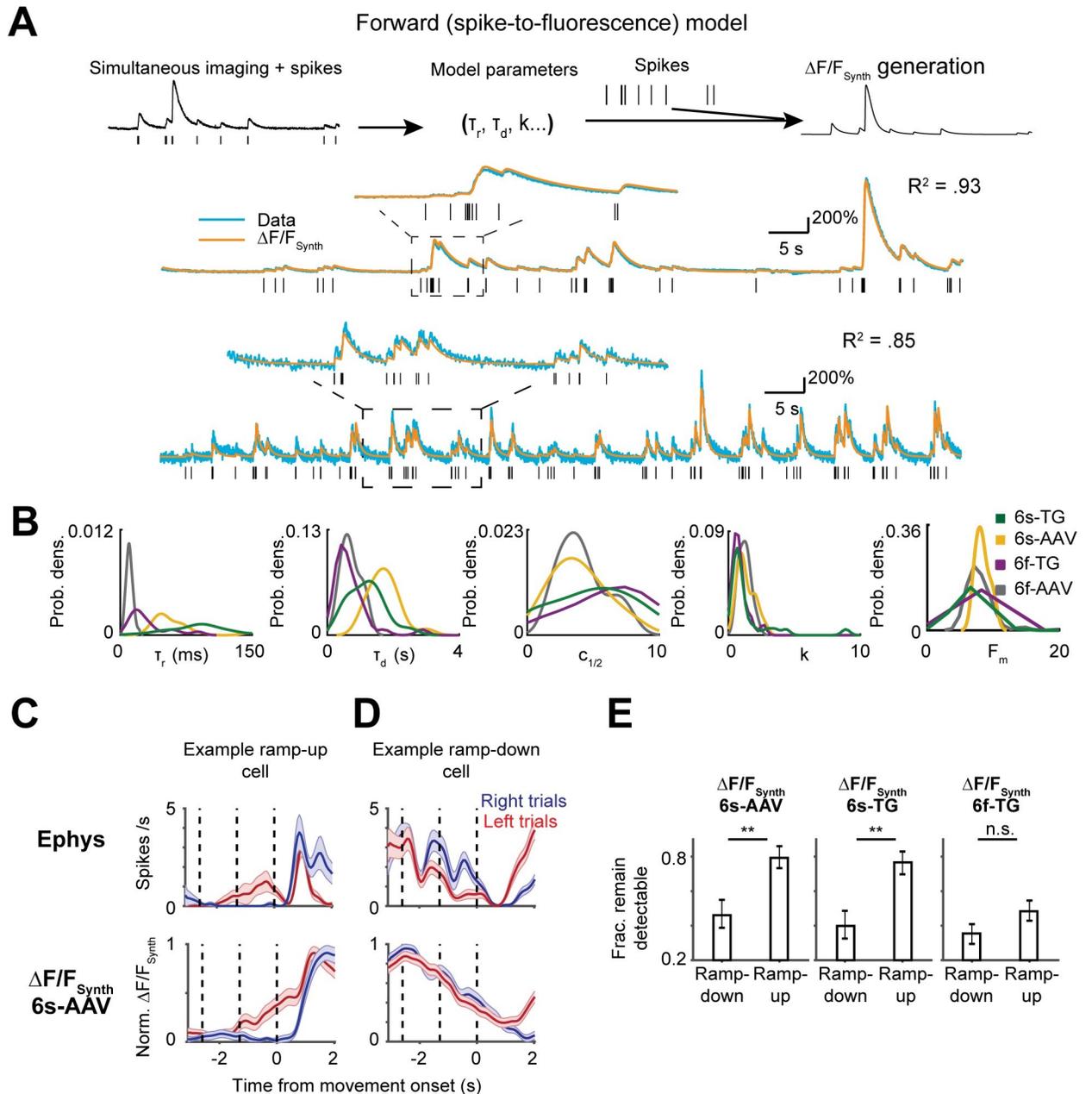
### Spike-to-fluorescence transformations explain differences in single neuron selectivity

Using the newly recorded data we developed a spike-to-fluorescence (S2F) forward model to generate a synthetic calcium imaging data based on a neuron's spike train (Fig 4A) [13,14,46,47]. In brief, spike times were first converted to a latent variable,  $c(t)$ , by convolution with a double-exponential kernel, with parameters rise-time ( $\tau_r$ ) and decay-time ( $\tau_d$ ). This latent variable was pushed through a non-linearity,  $F(c)$ , with a non-linearity sharpness parameter ( $k$ ), a half-activation parameter ( $c_{1/2}$ , corresponding to the half-rise point of the nonlinearity) and a maximum fluorescence change ( $F_m$ ) (Materials and Methods). The neurons were well-fit by the model (S3B Fig; variance explained, 6s-AAV,  $.87 \pm .17$ , mean  $\pm$  std.; 6s-TG,  $.80 \pm .20$ ; 6f-AAV,  $.82 \pm .27$ , mean  $\pm$  std.; 6f-TG,  $.66 \pm .23$ ). The inferred parameter values reflected known indicator kinetics. For instance, the decay times measured for neurons expressing GCaMP6s were longer than those expressing GCaMP6f (Fig 4B). However, there was substantial variability between the parameter values inferred across neurons (Figs 4B and S3). This heterogeneity likely reflects differences in calcium indicator expression and differences in calcium influx and calcium extrusion rates [28,48,49]. This variability is one factor that could explain the difficulty of the inversion of calcium responses which is central to spike inference approaches. We refer to simulations of calcium-dependent fluorescence based on application of the S2F model to spiking activity as ' $\Delta F/F_{\text{Synth}}$ '.

We applied the model to ramp-up and ramp-down neurons. For ramp-up cells the separation of activity across trial types was retained in  $\Delta F/F_{\text{Synth}}$ , albeit with slower dynamics (Fig 4C; top, an example neuron; bottom, synthetic 6s-AAV imaging of that neuron). In contrast, for many ramp-down cells  $\Delta F/F_{\text{Synth}}$  became non-selective (Fig 4D). Overall, selectivity was conserved more frequently for ramp-up cells than for ramp-down cells. Since right-preferring cells were more often associated with ramp-up dynamics, and calcium imaging is more likely to capture ramp-up selectivity than ramp-down selectivity, the model explains the greater fraction of right-preferring neurons in the calcium imaging data (Fig 4E). This was true whether a neuron happened to be a right- or left-preferring neuron, i.e., there were no significant differences in the fraction of detectability in the synthetic data once the data was broken down into two categories, ramp-up and ramp-down ( $p > .05$ ,  $\chi^2$  test for all imaging conditions; S4A Fig). Consistent with the difference being produced by the slow decay kinetics of GCaMP6s, there was little difference between the fraction of right- and left-preferring neurons in the 6f-TG data ( $p > .05$  for both cell types). In line with these results, we found that the forward model accounted for the drop in multiphasic neurons presented in the previous section (S4C Fig). We further confirmed that these (and previous differences) between ephys and imaging were not driven by low SNR manually-identified neurons (S5 Fig). These data show that the spike-to-fluorescence transformation introduces systematic discrepancies in comparing the same analysis performed on ephys or imaging data.

### Dimensionality reduction emphasizes different sources of variance in ephys and imaging

Large-scale recording methods are often used in combination with dimensionality reduction techniques to provide a compact description of the data[30]. For example, principal



**Fig 4. Forward modeling of the spike-to-fluorescence transformation largely explains difference in selectivity patterns.** A. spike-to-fluorescence model. Top: schematic plot of the spike-to-fluorescence (S2F) forward model that generates a synthetic fluorescence trace ( $\Delta F/F_{Synth}$ ) from an input spike train. Middle: example fit and data of two cells. Experimental, measured  $\Delta F/F$  (blue) is overlaid with the simulated  $\Delta F/F_{Synth}$  (orange) from the S2F model. The input to the model, the simultaneously recorded spikes (black), is shown below the traces. B. Distributions of the inferred model parameters for different indicators (yellow: 6s-AAV; green: 6s-TG; Purple: 6f-TG; gray: 6f-AAV). C. An example ramp-up neuron (top, ephys; bottom, 6s-AAV synthetic of that neuron); selectivity remains detectable in synthetic imaging data. D. An example ramp-down neuron (top, ephys; bottom, 6s-AAV synthetic of that neuron); selectivity becomes undetectable in synthetic imaging. E. S2F model predicts that selectivity of ramp-down neurons but not ramp-up neurons, would be often obscured in imaging datasets. Bar plot shows fraction of cells that remain detectably selective in synthetic imaging (6s-AAV synthetic, left; 6s-TG synthetic, middle; 6f-TG synthetic, right) plotted separately for ramp-down and ramp-up cells.

<https://doi.org/10.1371/journal.pcbi.1008198.g004>

component analysis (PCA) finds modes of population activity that capture the largest amount of variance in neural activity[30]. Data visualization and analysis are often performed after

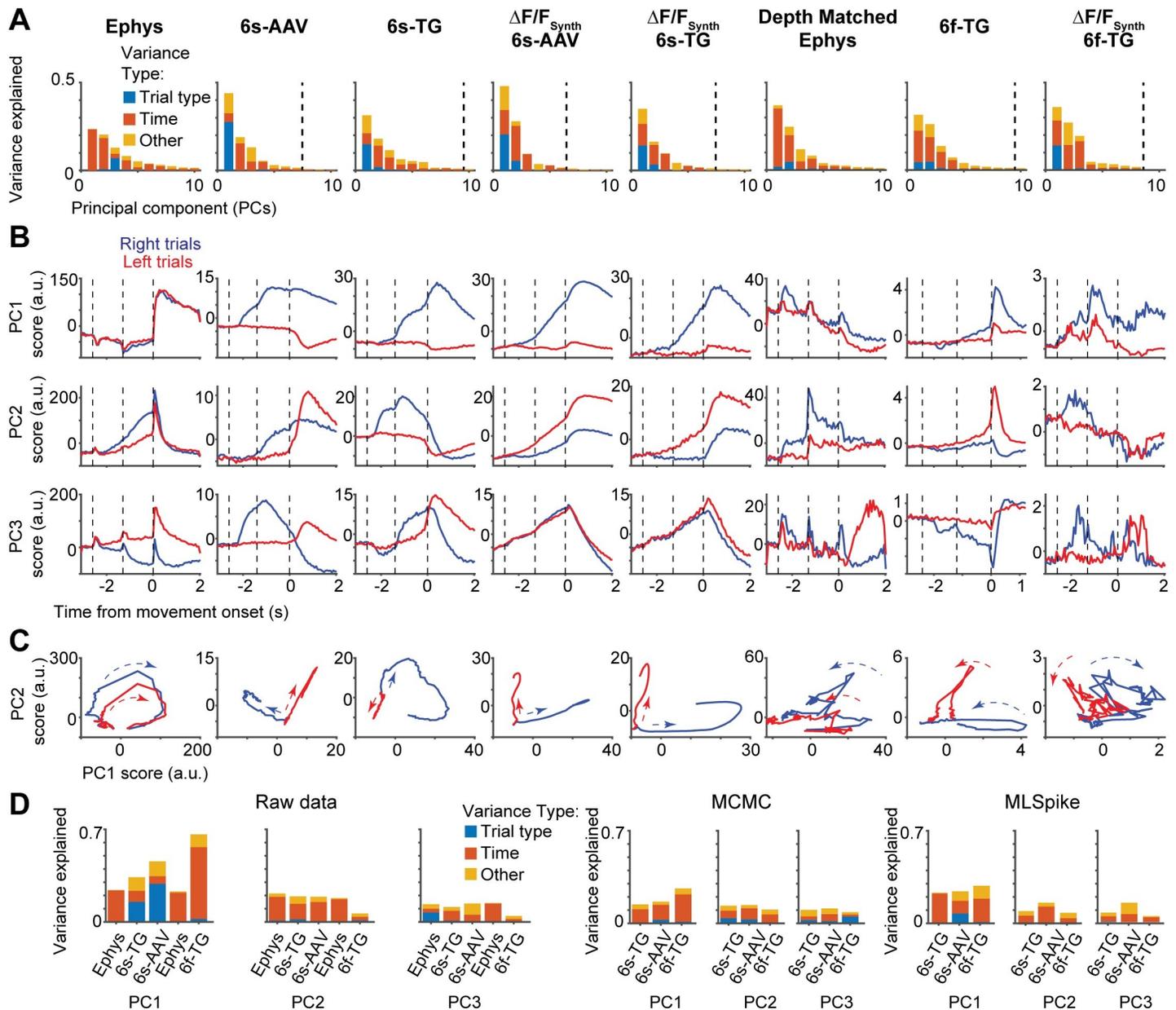
truncating the decomposition after a few components. Moreover, regression analyses are typically performed following dimensionality reduction to avoid having the number of variables (neurons) be close to the number of samples (trials). We found that the contribution of different sources of variance to the first principal components diverges between ephys and imaging. Accordingly, truncation of PCA in the first few principal components can lead to a qualitatively different PCA decomposition of neural activity between ephys and imaging.

We found substantial differences in performing PCA on ephys and imaging datasets. First, the content of the first PCs was remarkably different between ephys and imaging. In the ephys data, variance in the first PC was mostly due to temporal dynamics ( $98.71 \pm 0.06\%$ , mean  $\pm$  std., bootstrap analysis). In contrast, for GCaMP6s imaging trial-type selectivity was the dominant source of variance in the first PC (6s-AAV:  $60.39 \pm 0.29\%$ ; 6s-TG:  $44.51 \pm 0.65\%$ ) (Fig 5A). This difference was consistent with the temporal smoothing imposed by slower indicators, and as expected temporal dynamics were predominant in the first PC of GCaMP6f, closer to the values found in ephys, (6f-TG:  $64.87 \pm 2.47\%$ ; depth matched ephys:  $91.02 \pm 0.14\%$ ). Second, in the ephys data, a relatively large number of PCs ( $> 10$ ) contribute substantially to the variance, whereas in imaging and synthetic imaging most variance was explained by the first few PCs (test for number of PCs required to explain 90% of the variance,  $p < .001$ ; t test, bootstrap) (a difference in the explained variance per component has been previously reported[50]; Fig S9 there).

These differences in the sources of explained variance can be seen in the profiles of the PC scores (Fig 5B) as well as in the profiles obtained by a standard exploratory visualization, depicting the evolution of activity over time as a trajectory in the space of the first two PCs (Fig 5C). Spike inference algorithms correctly reduced the amount of trial-type variance in the first principal components (although not fully), but with the caveat that the fraction of variance in the first two principal components was reduced too much (Fig 5D). The spike-to-fluorescence model captured the qualitative differences between ephys and imaging, but overestimated the increase in variance in the first two principal components (Fig 5A–5C).

## Population activity history affects instantaneous decoding differently in ephys and imaging

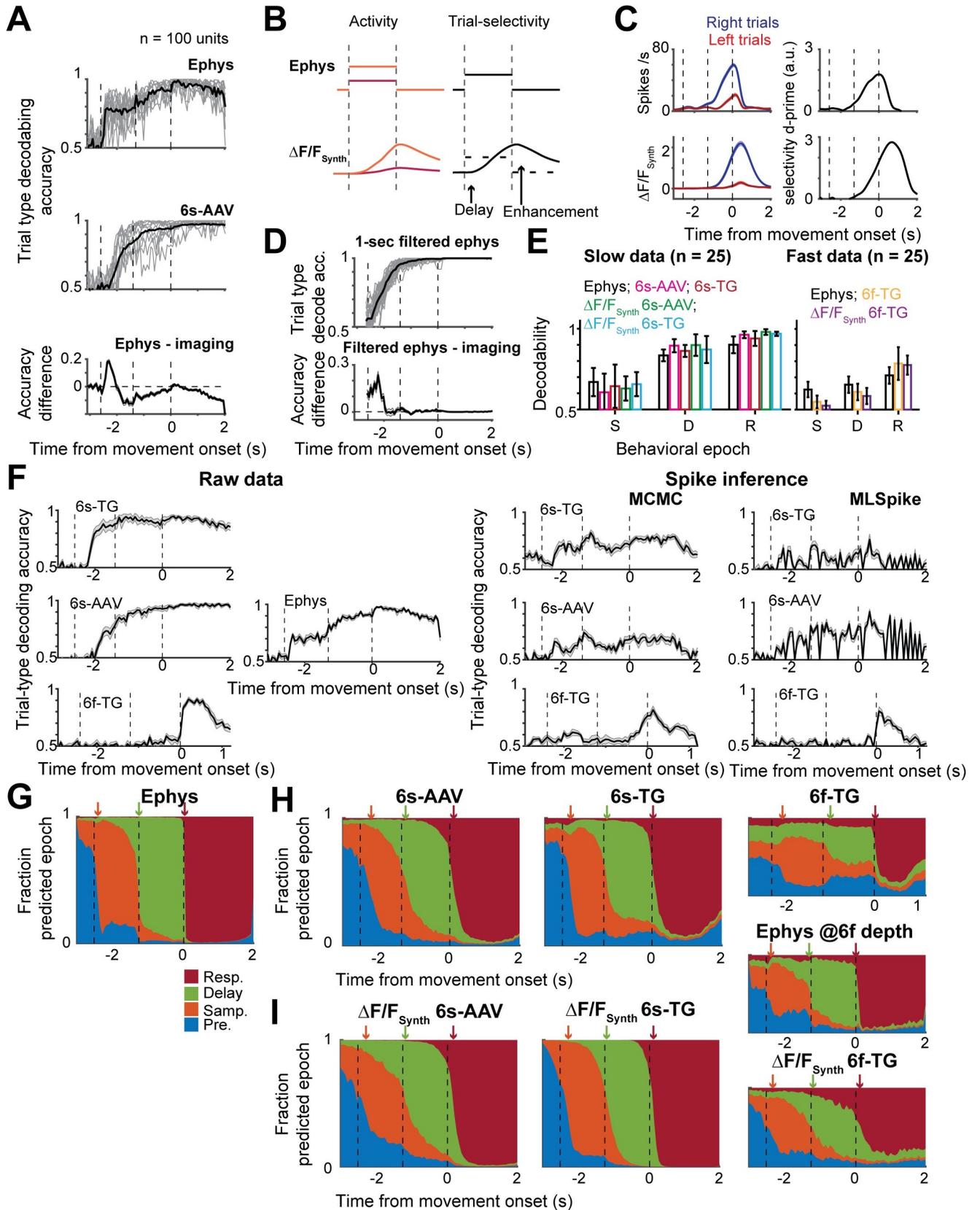
Decoding analysis relating population activity to behavioral variables is widely used in systems neuroscience[3,6,33,51]. Such analyses typically relate the state of population activity at a given time point to a behavioral variable of interest, such as behavioral choice. They are one of the most common analyses as they are a straightforward approach to addressing the question of what information does a population of neurons contain. We performed decoding analysis to predict either trial type or the current behavioral epoch from population activity (Eqs 3–5, Materials and Methods). Decodability of trial type in ephys increased earlier (one-tail t-test,  $p < .001$ ), but saturated at a lower level (one-tail t-test,  $p < .001$ ) than in calcium imaging (Fig 6A). Spike inference models, the MCMC framework in particular, partially reduced the delay of the rise of decodability but overestimated the decrease in decodability yielding lower performance in delay-response epoch than the ephys data (S6 Fig). Both observations were recapitulated by the S2F model (delay: one-tail rank sum test, 6s-AAV,  $p < .001$ , 6s-TG,  $p < .001$ , 6f-TG,  $p < .001$ ; enhancement: 6s-AAV,  $p < .001$ , 6s-TG,  $p < .001$ , 6f-TG,  $p < .001$ ; Fig 6B and 6C). The counterintuitive result of higher decoding accuracy in imaging for matched population size is explained by the long decay time of slow calcium imaging. The long integration in calcium imaging causes instantaneous decoding on imaging to be equivalent not to instantaneous decoding on spiking data, but to decoding on a more time averaged variable. Such a choice is advantageous when a larger proportion of the selectivity is stable, as was the case in



**Fig 5. Different sources of variability extracted in dimensionality reduction on imaging and ephys.** **A.** Fraction of variance of neural activity explained by principal components 1–10 divided into different sources of variability: red: temporal dynamics; blue: trial type; yellow: other (interaction term). From left to right: ephys, 6s-AAV, 6s-TG, 6s-AAV synthetic, 6s-TG synthetic; ephys depth-matched to 6f-TG recordings, 6f-TG, 6f-TG synthetic. Vertical dashed line indicates the PC index at which the remaining components capture <1% of total variance. **B.** Trial-averaged scores of first three PCs over time (from top to bottom), averaged separately for the two trial types (right trial, blue; left trial, red). Same order from left to right as in **A**. **C.** Trial dynamics in the first two-PC subspace for the two trial types (right trial, blue; left trial, red). Same order from left to right as in **A**. **D.** Left: fraction of variance explained by principal components 1–3 for each of the datasets, and its division into different sources of variability: red: temporal dynamics; blue: trial type; yellow: other (interaction term). Bars from left to right: ephys, 6s-TG, 6s-AAV; ephys depth-matched to 6f-TG recordings, 6f-TG. Middle: equivalent results for principal component analysis performed on inferred spiking data obtained via the MCMC framework. Right: equivalent results for principal component analysis performed on inferred spiking data obtained via the MLSpoke framework.

<https://doi.org/10.1371/journal.pcbi.1008198.g005>

ALM sample and delay selectivity(27). Consistently, decoders built on ephys that incorporated a one second integration time were more accurate than instantaneous ephys decoders and as accurate as slow indicators (Fig 6D). The delayed increase of decodability was also explained



**Fig 6. Population decoding differs in sensitivity and temporal profile between imaging and ephys.** **A.** Performance of instantaneous regularized linear-discriminant-analysis (LDA) trial-type decoder for 100-unit subpopulations. Vertical dotted lines indicate behavioral epochs, from left to right: presample, sample, delay, response. Top, decoders trained on ephys; middle, decoders trained on 6s-AAV; bottom, difference between the two. For top and bottom plots: individual gray lines show single subsample performance and black thick line shows average. In bottom plot mean is indicated by thick line and shaded area corresponds to standard deviation. **B.** Toy model demonstrating observed delayed but enhanced decodability in imaging data. Schematic of relation between activity (left) and decodability (right) when the model has two constant levels of activation for the two trial types (orange and red). **C.** Example cell showing similar behavior to the toy model. **D.** Comparison of decodability from imaging to decodability from 1-second filtered ephys. Top, 1-second filtered ephys; bottom, difference between filtered ephys and imaging. **E.** Comparison of decodability of trial type per behavioral epoch. Decodability for all datasets separated into slow indicators (left) and fast indicators (right). Bars color coded according to dataset. Left: black, ephys; magenta, 6s-AAV; red, 6s-TG; green, 6s-AAV synthetic; cyan, 6s-TG synthetic. Right: black, ephys (depth matched to 6f-TG); orange, 6f-TG; purple, 6f-TG synthetic. **F.** Accuracy of trial-type population decoding over time for different datasets. Left, top to bottom: 6s-TG, 6s-AAV, 6f-TG. Middle: ephys. Right, accuracy of trial-type population decoding over time of datasets comprised of inferred ephys from the different imaging datasets. top to bottom: 6s-TG, 6s-AAV, 6f-TG. Left column: MCMC framework, right column: MLSpikes framework. **G-I.** Performance of behavioral-epoch LDA decoders. **G.** Probability of decoder based on ephys to assign population activity to each of the different epochs shown in the following color scheme: pre-sample (blue), sample (orange), delay (green), and response (red) epoch; arrows indicate the inferred transition times of epochs from neural codes. **H.** Same plot format as G, but for imaging. **I.** Sample plot format as G, but for synthetic imaging.

<https://doi.org/10.1371/journal.pcbi.1008198.g006>

by the forward model. GCaMP6f, with its reduced signal to noise, yielded less accurate population decoders (Fig 6E; spike inference measure, Fig 6F).

For different decoding analyses such averaging can reduce accuracy. For instance, the neurons that can be used to decode trial-type change substantially between the delay and response period, i.e., the patterns of population selectivity are typically dynamical themselves. To test the interaction of these dynamics with calcium indicators, we trained decoders to distinguish the current epoch in the task from the pattern of neural activity. In ephys (Fig 6G) we observed a rapid decrease of the probability of activity to belong to the previous epoch following a change in behavioral epoch, along with a sharp increase in the probability of belonging to the current epoch. In contrast, in the calcium imaging data such changes tended to be delayed and gradual, even for the fast calcium indicator (Fig 6H). This effect was also recapitulated in the synthetic calcium data from the S2F model (Fig 6I). In other words, at the change of a behavioral epoch, the asymmetry of fast rise times and long decay times in calcium indicators yields calcium imaging signals that are a mix of the decaying profile of activity in the previous epoch and the newly activated profile of activity elicited by the response epoch.

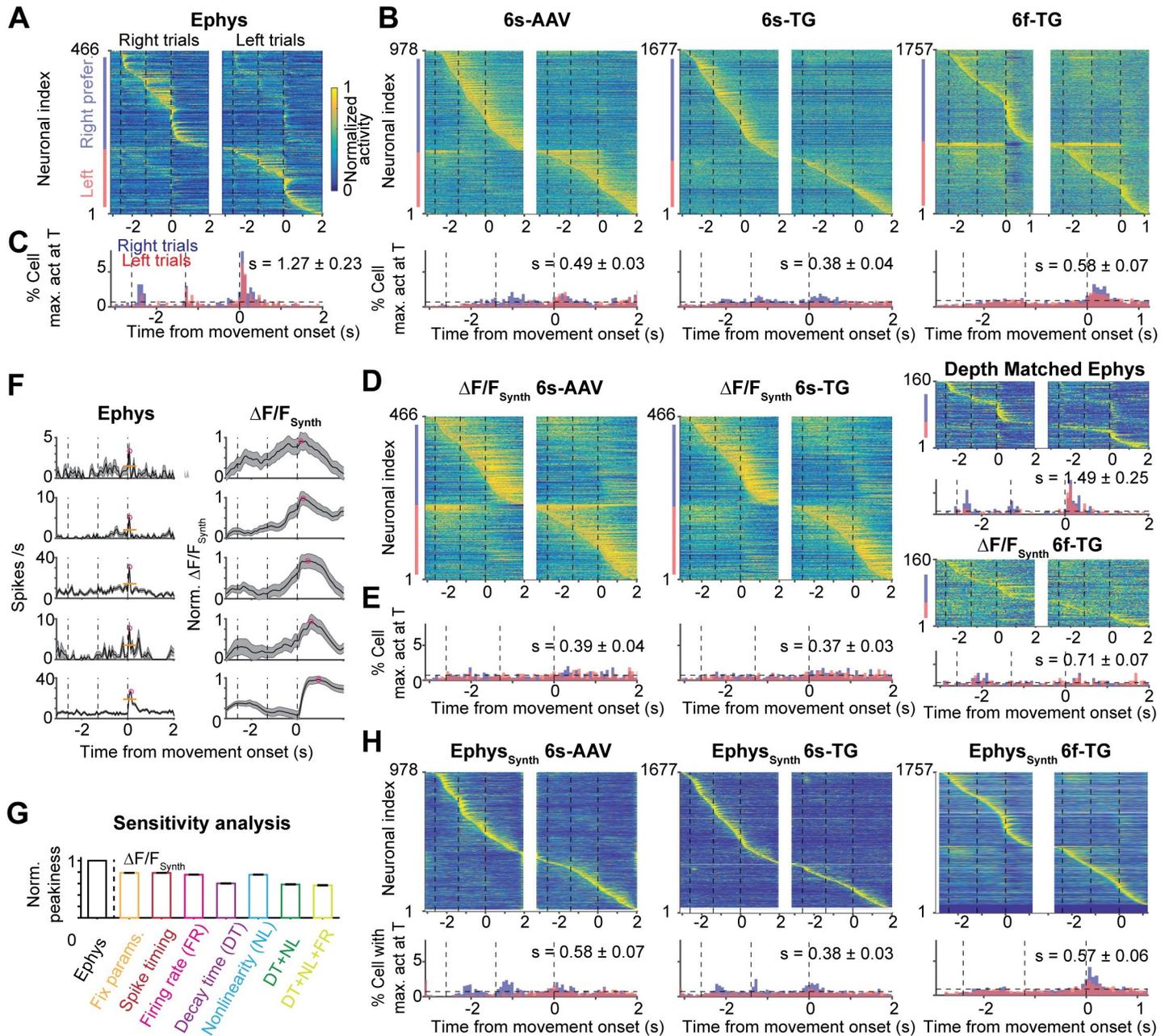
### Population dynamics is temporally dispersed in calcium imaging

Neurons show temporally complex responses, even in simple trial-based behaviors [25,26]. These spike rate changes are critical for an understanding of neural circuit models of neural computation. For instance, relative timing can be analyzed for propagation of information through neural circuits [52]. Our analysis revealed a qualitative difference in the dynamics between populations recorded by ephys or imaging: a dispersion of the apparent dynamics. That is, the spike rates recorded in ALM peaked at transitions between behavioral epochs (Fig 7A) [25]. In contrast, in the calcium imaging data, peaks of fluorescence were delayed and spread out over time, producing a more sequence-like appearance (Fig 7B) [51,53].

To quantify this effect we computed a measure of the ‘peakiness’ of the distribution of neuronal activity ( $s$ ) across recording modalities as the difference between observed neural activity and temporally uniformly distributed neural activity ( $P = \frac{1}{2T}$ ), summed over lick-left and lick-right trials:

$$s = \frac{1}{P} \sqrt{\frac{1}{2T} \sum_{i=\text{left, right}} \int_0^T dt (P_i(t) - P)^2}$$

$s$  was much larger for the ephys dataset ( $1.27 \pm 0.23$ ) compared to the 6s-AAV ( $0.49 \pm 0.03$ ; one-tail rank sum test,  $p < .001$ ), 6s-TG ( $0.38 \pm 0.04$ ; one-tail rank sum test,  $p < .001$ ), and 6f-



**Fig 7. Temporal dispersion of population dynamics differs between imaging and ephys.** **A.** Heatmap of normalized trial-averaged firing rates for right trials (left) and left trials (right) for ephys data. Firing rates were normalized to maximum of activity across both conditions. Neurons were first divided into two groups by their preferred trial type then sorted by latency of peak activity. **B.** Same plots as **A** but for 6s-AAV (left), 6s-TG (middle) and 6f-TG (right). Below the 6f-TG are neurons from ephys depth matched to 6f-TG. **C.** Fraction of neurons with a peak at given time point over time. Distribution in time plotted simultaneously for both trial types (red: right trials, blue: left trials, black horizontal line: uniform distribution). Datasets shown left to right (from left: ephys, 6s-AAV, 6s-TG, and 6f-TG respectively). **D-E.** The same plots as **B-C** for synthetic imaging (6s-AAV synthetic, left; 6s-TG synthetic, middle; 6f-TG synthetic, right). **F.** Example cells with peaks at a similar time in ephys (left; mean activity, thick black line; sem, shaded area; peak, magenta circle; baseline, orange thin line) along with the corresponding synthetic data (right). Neurons are sorted according to their peak times in synthetic imaging (early to late, from top to bottom). **G.** Sensitivity analysis of peakiness by synthetic, artificial data (**Materials and Methods**). Bars show normalized peakiness for the different model variants: (1) identical S2F parameters and identical spike times; (2) identical S2F parameters, jittered spike times (3) identical S2F parameters, variable firing rate (4) identical S2F parameters except for the decay time constant of the calcium indicator that was randomly sampled from its distribution; (5) identical S2F parameters, except for the nonlinearity of the calcium indicator that was randomly sampled from its distribution; (6) both decay time constant and nonlinearity of calcium indicator randomly sampled; (7) variable decay time constant, non-linearity and firing rates. **H.** Same plots as **B-C** for inferred firing rates from imaging, i.e., synthetic ephys.

<https://doi.org/10.1371/journal.pcbi.1008198.g007>

TG ( $0.58 \pm 0.07$ ; one-tail rank sum test,  $p < .001$ ) imaging data (Fig 7C). The forward model was able to recapitulate the differences between ephys and imaging ( $s = 0.39 \pm 0.04$ , 6s-AAV  $\Delta F/F_{\text{Synth}}$ ;  $s = 0.37 \pm 0.03$ , 6s-TG  $\Delta F/F_{\text{Synth}}$ ;  $s = 0.71 \pm 0.07$ , 6f-TG  $\Delta F/F_{\text{Synth}}$ ; Fig 7DE). Using the forward model we found that the degree of delay in the peak response is dependent on interactions between multiple factors including the assumed temporal and non-linear parameters of the indicator, as well as the absolute value of the underlying firing rate (Fig 7FG). In addition, slow changes in spike rate can interact with transient dynamics. For example, small fractional changes in spike rate can appear as prominent ramping in  $\Delta F/F$ , whereas subsequent brief increases in spike rate can appear blunted in  $\Delta F/F$  [54] (S6 Fig). Here, spike inference algorithms were able to partially undo the difference between imaging and ephys, yielding a reduction in the temporal dispersal (Fig 7H). Similar overall results were obtained with different metrics for the sharpness of the maximum-activity-time distribution relative to a uniform distribution, such as the Kullback-Leibler divergence.

Similar analyses on single neuron and population activity properties were performed on ephys and imaging data from the primary somatosensory cortex with qualitatively similar results (S7 Fig).

## Discussion

Calcium imaging using fluorescent protein sensors is a powerful method for recording activity in large neuronal populations [5,8]. In systems neuroscience, cellular calcium imaging fills a complementary role to extracellular electrophysiology. Imaging can sample neural activity densely [5,10] and reveal spatial relationships between neurons with related activity patterns [55,56]. Imaging can be used in a cell-type specific mode to sample rare neuronal populations that are difficult to target using electrophysiology [9]. Imaging can be combined with post-experiment molecular analysis [23,55,57] or serial electron microscopy reconstruction [58,59]. Imaging can track the activity of individual neurons over long time scales to explore the circuit basis of learning [6,60]. Finally, imaging allows recording activity in neuronal microcompartments that are not accessible to electrophysiology [13,61–63]. Electrophysiological recordings report neural activity with high temporal precision but have limitations of their own. Ephys recordings have a bias towards large neurons with high spike-rates. In addition, the process of transforming raw recordings into spike times associated with individual isolated units, i.e., spike sorting, can introduce artifacts such as merging spikes from different neurons.

Calcium imaging and ephys are often used almost interchangeably. A few studies have attempted to compare calcium imaging and electrophysiology and generally found qualitative agreement [8,13], but only using static and relatively coarse measures. More refined measurements reveal clear differences between the methods. For example, under standard recording conditions the detection efficiency for individual spikes is low for imaging and high for ephys (Figs 3E and S2C) [64]. Here we explored the effects of differences between ephys and imaging on measures typically used in system neuroscience. By comparing activity recorded with electrophysiology or imaging from matched neuronal populations during the same behavioral task we showed that the different recording methods can lead to diverging results. On the level of single neurons, the proportion of neurons with specific response properties and different dynamics of selectivity differs between calcium imaging and ephys. At the level of neuronal populations, we find diverging results for the content of population activity variance (trial condition differences being the main source of variance in imaging while temporal dynamics are the main source of variance in ephys), the relation of population activity to behavior, and the overall pattern of population dynamics. Spike inference algorithms only partially recovered the difference between ephys and imaging across the multiple metrics considered in this study

(S8B Fig). Notably, we find large neuron-to-neuron variability in the inferred parameters of a forward, spike-to-fluorescence model. Analytical approaches that ignore this heterogeneity, as most do, will likely infer an incorrect average inverse solution which will be a poor match for individual neurons. Indeed, such variability coupled with the large heterogeneity in firing rates and temporal patterns makes correctly solving the inverse problem difficult, which potentially explains our results. At the same time, most of the differences we found between ephys and imaging were explainable by a forward-model that generates a synthetic imaging experiment counterpart of a neuron's ephys responses. Such a model takes into account the specific heterogeneity found in ephys recordings and can take into account neuron-to-neuron variability in calcium imaging properties by sampling randomly from the varying parameters of the spike-to-fluorescence transformation. Lastly, baseline subtraction can distort inference of modulation of spiking activity when the underlying baseline spike rate is unknown. For example, a small gradual change in baseline spike rate can be amplified compared to a large phasic response (S6 Fig). This poses a challenge especially for the interpretation of photometry[54], where averaging is performed over neurons.

### **Im-phys.org—A website for more detailed comparison ephys and imaging**

We presented an extensive dataset with three calcium indicators, extracellular and intracellular electrophysiology and multiple models. However, a single research paper still represents a small distillation of all possible analyses. We developed an online resource, [im-phys.org](http://im-phys.org) (<http://im-phys.org/> S8 Fig) with three goals. First, the website allows analysis of all combinations of dataset and model, to evaluate the scenario that is most relevant to particular experiments. [Im-phys.org](http://im-phys.org) allows spike inference algorithms to be systematically tested in real use case scenarios, i.e., not just testing recovery of any aspect of the patterns of spike rates but rather testing the impact of performing spike inference on undoing differences in specific metrics extracted from ephys and imaging (S8B Fig). Second, we hope that other groups will share data, models and analyses to allow more general comparison of ephys and imaging data. [Im-phys.org](http://im-phys.org) allows submission of data that can be incorporated into various comparisons that are displayed on the website, controlled through UIs. Though few labs have matched ephys and imaging datasets, many labs have one or the other. Our resource can serve to aggregate and combine these datasets, as well as find a best match from an imaging to ephys dataset (S8A Fig). Third, [im-phys.org](http://im-phys.org) is linked to a github repository containing the analyses code, models (S2F and F2S), and related data. These allow the application of analyses and models on data without sharing it through [im-phys.org](http://im-phys.org).

### **Differences between interrogating population activity by ephys and imaging affect data-driven models**

Differences in metrics of population activity between calcium imaging and ephys not only complicate the research literature but can result in the divergence of models used to understand the underlying data. Most population models, whether models in which the single units are modeled in more biophysical detail or more abstractly, are still highly reduced in the way they treat population heterogeneity. As such they often rely on dimensionality reduction of the recorded data to define the aspects of population activity the model is meant to capture. We found substantial differences between ephys and imaging data in application of PCA, and the truncation of the data after a few important data components can further amplify differences. In extreme cases one may be left with subsets that differ dramatically across imaging and electrophysiology. The amplification of difference by dimensionality reduction is relevant not just for modeling of the data, but more generally when generic forms of dimensionality

reduction, such as PCA, are used early in the analysis pipeline to improve signal-to-noise ratio (which is important given the limited duration of typical behavioral experiments) for subsequent analysis, such as population decoding. Dimensionality reduction can be hard to avoid when analyzing large datasets[30], but can be modified to be less sensitive to known issues.

## Going forward

Going forward, the discrepancies between ephys and calcium imaging can be reduced by improvements in calcium indicators, adjustments to experimental design and use of forward models to identify the sensitivity of metrics of interest to the transformation in calcium dynamics. Calcium indicators could be improved on multiple fronts. They could be made faster and less nonlinear[18,65]. In addition, more uniform expression across cells can allow for more aggressive modeling of the nonlinearities that cannot be reduced, especially when coupled with priors on activity profiles derived from large scale electrophysiology. Faster indicators will result in the effect of previous activity history washing away faster, thus reducing effects that are history dependent. Imaging with multiple types of indicators in different experiments might produce additional constraints and help reduce biases. Voltage imaging holds great potential for fast accurate measurement of spiking activity, at least in sparsely labeled neuronal populations[66,67]. At the level of experimental design, when population activity in a given behavioral epoch involves fixed dynamics, such as settling to a steady state or consistent ramping, longer trial epochs will allow the effect of the previous dynamical state to decay away. Indeed, we found a smaller discrepancy between the number of multiphasic neurons in ephys and 6s-TG data when the behavioral paradigm was adapted to use longer delay epochs.

Finally and most importantly, the sensitivity to the specific properties of population activity that are of interest to a particular hypothesis can be evaluated by forward models, as we performed here. For example, imaging studies could use forward models on published ephys data to evaluate the potential effect of the spike to calcium transformation on the metrics of interest. Then differences from the expected value given the transformation can be analyzed and metrics that are shown to be more variable given heterogeneity in transformation parameters can be flagged. This effort will become easier as neurophysiology probes become more powerful [68], data sharing more common, and preprocessing more standardized.

Overall our results highlight the importance of a deeper understanding of the transformation imposed by calcium imaging. The fact that our model was able to reproduce differences between the recording methods suggests that additional data and associated analysis methodology developments could potentially better address quantitative comparisons between analyses of population activity performed from imaging or ephys data. The online resource we built allows researchers to better understand how the discrepancies we observed would be relevant for the circuit and recording method of interest. More quantitative interpretation of calcium imaging and full utilization of all its advantages will require investment in ground-truth data sets and new statistical approaches. We hope this study and our online resource will catalyze this crucial effort.

## Materials and methods

### Electrophysiological and imaging population activity recordings

Electrophysiological ('ephs')[25] or calcium imaging[25,36] recordings were performed in separate experiments and described in detail in the original publications (S1 Table;<http://imphys.org/data>). Mice were trained to perform a delayed version of a tactile discrimination task. Mice reported the position of a pole (anterior or posterior) by directional licking (lick-left or lick-right) after a delay period. The duration of sample and delay epoch was 2.6 s. In ephys, the

delay epoch was 1.3 s; in imaging, it was 1.4 s. Trials with early licking were excluded from analysis. Neuronal depths were 100 to 800  $\mu\text{m}$  (ephys), 150–740  $\mu\text{m}$  (6s-AAV), 120–640  $\mu\text{m}$  (Thy1-GP4.3 mice, 6s-TG), and 140–470  $\mu\text{m}$  (Thy1-GP5.17 mice, 6f-TG). Only sessions with more than 20 trials for each type (right-trial and left-trial) were included. For imaging data, we performed a post-hoc detection of outliers and removed trials where more than 30% of the time points contain a signal with 3 standard deviations away from median (these outliers relate to baseline fluctuations across trials, and removing them was necessary for variance-based analysis). Neurons were limited to putative pyramidal neurons. These reduced the total number of neurons with sufficient number of trials, yielding 1493, 2293, and 2672 units for 6s-AAV, 6s-TG and 6f-TG imaging, respectively. We note that despite the 6s-TG data containing many neurons across multiple sessions all data came from a single animal. Though we find this data to be consistent with other imaging data, the single animal source could be a potential issue in that differences in an individual animal's behavior may cause differences in neural encoding.

We used two sets of data from loose-seal electrophysiological recordings and imaging from GCaMP6-expressing neurons in primary visual cortex. In one set neurons were transduced with 6s-AAV and 6f-AAV (data from [13]). For 6s-AAV data, imaging was performed after 2–4 weeks of expression. In the other set we used 6s-TG and 6f-TG mice [31,69]. More details of all datasets are described at <http://im-phys.org/data>.

In the imaging data, individual neurons were visually identified based on average fluorescence images as well as “neighborhood correlation maps” (where the brightness of each pixel encodes the correlation of its fluorescent time course to that of its neighbors), which highlights active cells. Each ROI was inspected to correspond to a morphological neuron and have a generally donut-like shape (since the nucleus does not express the indicator). The fluorescence time course of each cell was measured by averaging all pixels within the ROI, with a correction for neuropil contamination. The fluorescence signal of a cell body was estimated as  $F_{\text{cell}}(t) = F_{\text{roi}}(t) - r * F_{\text{neuropil}}(t)$ , with  $r = 0.7$ . The neuropil signal  $F_{\text{neuropil}}(t)$  surrounding each cell was measured by averaging the signal of all pixels within a 40  $\mu\text{m}$  radius from the cell center (excluding all selected cells). For each imaging plane, the number of ROIs was about 30+/-19 cells in the 6s-AAV imaging conditions, and 82+/-48 cells per recording plane in 6s-TG, and 122+/-32 cells per recording plane in 6f-TG.

Whole-cell recordings were made using pulled borosilicate glass (Sutter instrument). A small craniotomy (100–300  $\mu\text{m}$  diameter) was created over the ALM (bregma AP 0.0 mm, ML 2.0 mm) under isoflurane anaesthesia and covered with cortex buffer during recording. Whole-cell patch pipettes (7–9 M $\Omega$ ) were filled with internal solution (in mM): 135 K-glucuronate, 4 KCl, 10 HEPES, 0.5 EGTA, 10 Na<sub>2</sub>-phosphocreatine, 4 Mg-ATP, 0.4 Na<sub>2</sub>-GTP and 0.3% Biocytin (293–303 mOsm, pH 7.3). The membrane potential,  $V_m$ , was amplified (Multi-clamp 700B, Molecular Devices) and sampled at 20 kHz using WaveSurfer (<http://wavesurfer.janelia.org/>).  $V_m$  were not corrected for liquid junction potential. After the recording the craniotomy was covered with Kwik-Cast (World Precision Instruments). Each animal was used for 2–3 recording sessions. Recordings were made from 350 to 850  $\mu\text{m}$  below the pia.

Simultaneous loose-seal recordings and imaging (Figs 3 and S2; S2 Table) was performed as described previously [13] (more details at <http://im-phys.org/data>). GP4.3 and GP5.17 mice [31] were lightly anesthetized (0.5% isoflurane). Drifting grating visual stimuli were used to drive activity in the visual cortex. Loose-seal recordings were made through a craniotomy windows over the primary visual cortex. Two-photon imaging and loose-seal, cell-attached recordings were performed simultaneously. We acquired images in both low (284 x 284  $\mu\text{m}^2$ ) and high (38 x 38  $\mu\text{m}^2$ ) zoom configurations. Extraction of fluorescence transients was as

described [13]. All procedures in mice experiments were performed in compliance with the Janelia Research Campus Institutional Animal Care and Use Committee.

To analyze the spike-triggered fluorescence changes, we created 1.2-s snippets around action potentials (APs), where a few APs only happened from 200 ms to 400 ms from the onset of each snippet. We computed baseline fluorescence using the snippets without AP in the entire time series. For snippet with APs, we required the fluorescence changes within the first 200 ms (before APs) was around baseline level (Fig 3C and S2A). We computed ROC curves for detecting one (Fig 3E, inset panels) or many APs (Figs 3E and S2C) compared to baseline fluorescence fluctuations. D-prime was computed as  $\frac{\langle \max(\Delta F/F)_{1AP} \rangle - \langle \max(\Delta F/F)_{no\_AP} \rangle}{\text{std}(\Delta F/F)_{no\_AP}}$ .

### Spike-to-fluorescence model

We developed a phenomenological model that converts spike times to synthetic fluorescence time series [13,14,25,46]. This ‘spike-to-fluorescence’ (S2F) model consists of two steps. First, spikes at times  $\{t_k\}$  are converted to a latent variable,  $c(t)$ , by convolution with a double-exponential kernel:

$$c(t) = \sum_{t > t_k} \exp\left(-\frac{t-t_k}{\tau_d}\right) \left[1 - \exp\left(-\frac{t-t_k}{\tau_r}\right)\right] + n_i(t) \quad (\text{Eq 1})$$

$\tau_r$  and  $\tau_d$  are the rise and decay times, respectively.  $n_i(t) \sim N(0, \sigma_i^2)$  is Gaussian distributed ‘internal’ noise.  $c(t)$  was truncated at zero if noise drove it to negative values. Second,  $c(t)$  was converted to a synthetic fluorescence signal through a sigmoidal function:

$$\Delta F/F_{\text{Synth}}(t) = \frac{F_m}{1 + \exp[-k(c(t) - c_{1/2})]} + n_e(t) \quad (\text{Eq 2})$$

$k$  is a non-linearity sharpness parameter,  $c_{1/2}$  is a half-activation parameter,  $F_m$  is the maximum possible fluorescence change.  $n_e(t) \sim N(0, \sigma_e^2)$  is Gaussian external noise [28,46,70].

We estimated the model parameters for each imaging condition using the simultaneous ephys and imaging experiments (S3A, S3B and S3C Fig). We then applied the S2F model to ephys data using parameters randomly sampled from the parameter distributions except for the parameters directly related to the nonlinearity. Since ALM spike rates in ephys vary over a larger range than the spike rates in the primary visual cortex these parameters may be underconstrained. Accordingly, we followed an alternative strategy to choose these parameters for a given neuron. For each neuron, after assigning the rest of the parameters, we transformed the spike trains to calculate the phenomenological calcium variable  $c(t)$ . We then estimated the nonlinear parameters for that neuron by calculating the values that would best transform  $c(t)$  to the fluorescence dynamics of any neuron in the imaging dataset. For all neurons we were able to find matches with Spearman correlation higher than 0.7 between mean  $dF/F$  and mean synthetic  $dF/F$ . The parameters inferred in this process recapitulated the correlation structure of  $c_{1/2}$  and  $k$  found in the data (S3D Fig).

Given the short timeframe over which baseline activity was recorded before each trial started, we extended the pre-trial period by simulating a Poisson spike train for the unrecorded time between trials with a constant rate equal to the baseline mean activity.

To relate this model to previously studied models, Eq 2 can be generalized as  $\Delta F/F_{\text{Synth}}(t) = f(c(t)) + n_e(t)$ , where  $f(\cdot)$  and  $n_e(t) \sim N(0, \sigma_e^2)$  is Gaussian external noise [28,46,70]. We considered two alternative S2F models used by previous studies (note though that both of these models did not contain internal noise in Eq 1):

**S2F Linear model:**  $f(c(t)) = F_{\text{max}}c(t) + F_0$ , where  $F_{\text{max}}$  is a scaling parameter (we kept the naming as max to clarify the relationship to other models);  $F_0$  is the baseline (S3G Fig, left).

**S2F Hill model:**  $f(c(t)) = F_{max} \frac{c(t)^n}{c(t)^n + K_d}$ , where  $F_{max}$  is the maximum possible fluorescence change;  $n$  is the nonlinearity;  $K_d$  is a half-activation parameter. Model performance is summarized in the supplementary material and reported in <http://im-phys.org/analyses> for each single cell (S3G Fig, right).

Model parameter sensitivity (S3C Fig) was defined as the decrease of the fraction of explained variance, as a function of the deviation of the parameter value from the estimated solution:  $g = \frac{\Delta EV/EV}{\Delta P/P}$ , where  $P \in \{\tau_r, \tau_d, k, c_{1/2}, F_m\}$ .

## Calcium imaging to spikes for non-simultaneous ephys-imaging recordings

We performed fluorescence-to-spike (F2S) inference using two published models[40,42] and code available on GitHub. Specifically, the default model in CaImAn was used to solve the FOOPSI problem to infer firing rates (using OASIS) and then MCMC was used to infer spike times. When performing algorithm comparison some published approaches may rely on parameters that are difficult to tune. In order to avoid mistuning the hyperparameters, we intentionally selected F2S models that do not require manual setting of parameters or hyperparameters. Instead, both toolboxes autonomously fit the parameters they required through optimization processes provided in the shared code. Moreover, we used the simultaneous recorded loose patch and imaging data (where the loose patch provides ground truth) to ensure that fluorescence-to-spike models were implemented correctly and return reasonable results.

## Single neuron analyses

Neural selectivity for left- or right-trials was determined using two-sample t-tests, with neural activity binned over 67 ms, which corresponds to one imaging frame. A neuron was selective if it showed selectivity ( $p < .05$ ) for  $>335$  ms (5 continuous frames). A selective neuron was multiphasic if the polarity of selectivity switched, with continuous periods of selectivity lasting at least 335 ms long. Selective neurons that were not classified as multiphasic according to this criterion were classified as monophasic.

Selective neurons (mono- and multiphasic) were classified into left- and right-preferring cells according to the condition in which their activity was higher (Fig 2IJ). Ramp-down (ramp-up) were defined as neurons that have activity that is greater (less) in the baseline epoch compared to the delay epoch (paired t-test,  $p < .05$  across trials). Note that ramp-down cells were excluded from the analysis of peakiness (Fig 7).

## Principal component analysis

Principal Component Analysis (PCA) was performed on the activity of neurons averaged across trial type ( $s \in \{left, right\}$ ):

$$\mathbf{r}(s, t) = \mathbf{C}\mathbf{x}(s, t) + \langle \mathbf{r} \rangle_{s,t} \quad (\text{Eq 3})$$

$\mathbf{r}$  is a  $n \times 2T$  matrix, where  $n$  is the number of recorded units in each dataset and  $T$  is the number of time points for each trial type.  $\langle \mathbf{r} \rangle_{s,t}$  is a vector of the mean activity of each neuron across time and trial type.  $\mathbf{x}(s, t)$  is an  $n \times 2T$  PC score matrix, where the  $i$ th row corresponds to the  $i$ th PC score. We estimated the relative contribution to each PC of the different forms of variance: temporal dynamics, trial-type selectivity and other. Explained variance (EV) of temporal dynamics  $EV_i(t)$  and trial-type selectivity  $EV_i(s)$  for the  $i$ th principal component (PC)

were computed as:

$$EV_i(t) = \langle \langle x_i(s, t) \rangle_s^2 \rangle_t / \langle x_i(s, t) \rangle_{t,s}^2 \quad (\text{Eq 4})$$

$$EV_i(s) = \langle \langle x_i(s, t) \rangle_t^2 \rangle_s / \langle x_i(s, t) \rangle_{t,s}^2 \quad (\text{Eq 5})$$

respectively. 6f-Tg related population analyses were only applied to cells with ROC > 0.7.

## Population decoding

We applied regularized linear discriminant analysis (LDA) on neural dynamics grouped into bins corresponding to single imaging frames (67 ms) to compute the instantaneous decodability of trial type. Regularization was performed by sparsity-regularized LDA [33,71]. The optimal LDA decoder was computed separately for each time bin using correct trials only. We estimated performance for the instantaneous LDA decoder by sampling subsets of units and averaging 100 subsamples. We separated the trials of each neuron into non-overlapping training (70%) and testing (30%) sets. The instantaneous decoder of trial type was computed from training set and its performance was evaluated on the testing set.

We tested the ability of neuronal population activity at different times to discriminate the behavioral epoch by using a four-class LDA (Fig 6F–6H). We defined the latency of neuronal response to behavioral epoch by the first time at which decoding reached a 0.7 accuracy threshold (arrows on Fig 6F–6H). Regularization was performed by sparsity-regularized LDA [33,71].

## Sensitivity analysis of peakiness

We used as a reference value an artificial, synthetic ephys dataset with 50 neurons whose firing rates were manually set to be non-zero only at the time corresponding to one imaging frame. From left to right in Fig 7G, S2F model was configured (1) using the same parameters for all cells, except that the internal noise and external noise were randomly generated (at the same amplitudes); (2) using the same parameters for all cells, except that the spike times were jittered within the time length of the frame (i.e., all spikes were kept in the same image frame); (3) using the same parameters for all cells, except that the spike rates in the original frame varied from 0.1 Hz to 5 Hz (spike trains generated using Poisson process); (4) using the same parameters for all cells, except that the decay time constant of calcium indicator was randomly sampled from its distribution; (5) using the same parameters for all cells, except that nonlinearity of calcium indicator was randomly sampled from its distribution; (6) both decay time constant and nonlinearity of calcium indicator were randomly sampled; (7) the same as (6) except that the spike rates in the original frame vary from 0.1 Hz to 5 Hz (spike trains generated using Poisson process).

## Distributions of measures

For S2F model, one can randomly sample all the parameters from the distributions measured using simultaneous ephys-imaging recordings and all possible noise levels. The distribution of a measure  $\psi$  (e.g. fraction of mono-selective neurons, peakiness etc.) can then be computed through synthetic data using randomly sampled S2F models. Specifically:

$$P(\psi) = \int P(\psi, \Delta F/F_{\text{Synth}}(t), \{t_{\text{spike}}\}, \Theta) \quad (\text{Eq 6})$$

where the joint distribution can be formulated through a chain rule:

$$P(\psi, \Delta F/F_{Synth}(t), \{t_{spike}\}, \Theta) \\ = P(\psi|\Delta F/F_{Synth}(t))P(\Delta F/F_{Synth}(t)|\{t_{spike}\}, \Theta)P(\{t_{spike}\})P(\Theta) \quad (\text{Eq 7})$$

where  $P(\Delta F/F_{Synth}(t)|\{t_{spike}\}, \Theta)$  is derived from **Eqs 1, 2**, and  $P(\psi|\Delta F/F_{Synth}(t))$  describes probability of measure  $\psi$  at a given value for dynamics  $\Delta F/F_{Synth}(t)$ ,  $P(\{t_{spike}\})$  is the empirical distribution of spike events in ground truth ephys and  $P(\Theta)$  is the distributions of S2F parameters.

For unsupervised-learning-based F2S models (i.e. MCMC and MLSpikes), we performed 100 subsamples of deconvolved synthetic ephys data to estimate distribution of the parameters.

## Computer code

All codes for model benchmarks and comparison metrics are recompiled and packed with data through im-phys-API (<https://github.com/zqwei/Im-phys-API>), which can be available at [im-phys.org/codes](http://im-phys.org/codes). The API will come with a user-friendly interface in which one can reproduce all results in our paper and extensive results on [im-phys.org](http://im-phys.org).

We also provide repos for benchmarks of S2F and F2S models at <https://github.com/zqwei/Ca-Imaging-Deconv-List> (DOI: [10.5281/zenodo.3960635](https://doi.org/10.5281/zenodo.3960635)) and comparison metrics at <https://github.com/zqwei/Neural-Recording-Methodology-Comparison> (DOI: [10.5281/zenodo.3979786](https://doi.org/10.5281/zenodo.3979786)) and website interface at <https://github.com/zqwei/Im-phys-org>.

## Supporting information

**S1 Fig. Effect of recording depth and firing rate in ephys. A-D.** Analysis as a function of recording depth. **A.** Single neuron selectivity-type analyses. Left: horizontal bar plots show breakdown of the population into selectivity types (gray: non-selective neurons, orange: monophasic-selective neurons, green, multiphasic-selective neuron). Right: horizontal bar plot shows number of neurons at each depth. The ratio of monophasic- to multiphasic selective neuron was similar across depths ( $\chi^2$ -test to depths with  $n > 50$  cells, ephys:  $p = .19$ ; 6s-AAV:  $p = .73$ ; 6s-TG:  $p = .97$ ; 6f-TG:  $p = .43$ ). For the same depth, ephys has more selective neurons and more multiphasic selective neurons than imaging ( $\chi^2$ -test,  $p < .001$  for all). **B.** Percentage of variance of neural activity explained by each principal component (**Fig 5**). Left: length of horizontal bar shows fraction of variance in each principal component. Colors show breakdown into different types of variance (blue: trial-type, red: time, orange: other). Right: horizontal bar shows number of neurons in each depth. For the same depth, the 1<sup>st</sup> PC show more temporal dynamics content in ephys and 6f-TG ( $\chi^2$ -test,  $p < .001$  for all), while that show more trial-type content in 6s-AAV and 6s-TG ( $\chi^2$ -test,  $p < .001$  for all). **C.** Decodability of trial type (**Fig 6**). The number of cells at each depth is identical to that in PCA analyses. The decodability differs across depths, where the neurons in superficial layers show weak decodability of trial type in sample-delay epoch (multivariate ANOVA test on time-series to depths with  $n > 50$  cells in ephys, 6s-AAV and 6s-TG; that to depth with  $n > 10$  cells at ROC  $> 0.7$  in 6f-TG;  $p < .001$ , 1000 bootstrap). For the same depth, the average decodability of trial type is higher in late delay to early response in imaging than that in ephys (rank sum test,  $p < .001$  for all, 1000 bootstrap). **D.** Peakiness (**Fig 7**). The peakiness differs across depths (rank sum test,  $p < .001$ , 1000 bootstrap). For the same depth, peakiness is higher in ephys than imaging (rank sum test,  $p < .001$  for all, 1000 bootstrap). **E-I.** Analysis as a function of spike rates. **E.** Schematic of resampling procedure to target firing rate and distribution of firing rates after subsampling to different average spike rates (magenta: original data; cyan: ephys subsampled to 1

Hz average; yellow: ephys subsampled to 4 Hz average; green: ephys subsampled to 10 Hz average. **F.** Effect of target firing rate subsampling on fraction of monophasic (left) and multiphasic neurons (right) **G.** Values of peakiness are shown with the same color code as **F.** **H.** Fraction of variance in the first principal components are shown by length of bar with same color code as **F.** Saturation of bar shows the breakdown into different components of variance (trial-type, time, other). **I.** Trial-type decodability over time shown with the same color code as **F.** and with 6s-AAV added as a reference. **J.** Analysis as a function of spike sorting accuracy—possible effects of merging. Increased fraction of multiphasic neurons is unlikely to have stemmed exclusively from failures of spike-sorting. Box plots indicate fraction of neurons in each selectivity class (left: non-selective, middle: monophasic, right: multiphasic) as a function of increased probability of artificially induced merging between two neurons. Dashed line indicates fraction of selectivity type found in the ephys dataset.

(TIF)

**S2 Fig. Single- and few-AP responses of neurons in transgenic GCaMP6s and 6f mice. A.**

Traces of fluorescence dynamics following different numbers of action potentials (APs) for example neurons (same plots as Fig 3C for additional examples). Gray, no AP; black, a single AP; red, 2 APs; blue, 3APs; green, 4APs; magenta, 5APs. Thin lines, single trials; thick lines, average. **B.** Peak fluorescence change as a function of the number of spikes (same plots as Fig 3D for additional examples). Black, single trials; red, trial average. **C.** ROC curve of all spike events. Inner panel, ROC curve for single AP events (same plots as Fig 3E for additional examples).

(TIF)

**S3 Fig. Detailed values of model parameters for simultaneously recorded neurons. A.**

Pairwise correlation plots for each of the spike-to-fluorescence parameters. Panels along the diagonal describe the distribution of each parameter (these are identical to Fig 4B but reproduced to facilitate comparisons). Off-diagonal panels depict the correlation between two parameters. Spearman's rank correlation of parameters across cells (regardless of recording method) and associated p-value are provided in each off-diagonal panel. Each circle corresponds to a response set. Data from the different indicator conditions is overlaid and marked by color. (gray: 6f-AAV, 11 neurons, 37 response sets; yellow: 6s-AAV, 9 neurons, 21 response sets; purple: 6f-TG, 18 cells, 32 response sets; green: 6s-TG, 22 neurons, 33 recording periods). **B.** Boxplots of explained variance of S2F on validation data for simultaneously recorded neurons (color follows the same convention as in A). **C.** Boxplot of distribution of parameter sensitivity values. **D.** Pairwise correlation of re-estimation of  $k$  and  $c_{1/2}$  using ALM imaging dynamics (Materials and methods). The re-estimated parameter values are shown as a scatter plot. Each dot corresponds to a neuron ( $n = 720$  for 6s-AAV and 6s-TG;  $n = 225$  for 6f-TG in matched depths). The distribution of the re-estimated parameter values strongly overlapped with those obtained in simultaneous imaging-ephyphys recordings.  $c_{1/2}$  and  $k$  had a strong inverse correlation as in the simultaneously recorded data ( $r_s < -.64$ ,  $p < .001$ ). **E.** Boxplots of firing rates of neurons in each recording sessions (6f-AAV, gray,  $0.51 \pm 0.25$  Hz, mean  $\pm$  std., range 0.05–1.25 Hz; 6s-AAV, yellow,  $0.43 \pm 0.38$  Hz, range 0.05–1.68 Hz; 6f-TG, purple,  $1.25 \pm 1.48$  Hz, range 0.09–5.22 Hz; 6s-TG, green,  $1.08 \pm 0.85$  Hz, range 0.09–3.00 Hz). **F.** Scatter of simultaneous ephys-imaging data model fit and the dynamical range of the data (expressed as mean spike rate). **G.** Scatter of simultaneous ephys-imaging data model fit quality between different S2F models (Materials and methods). Left: comparison between S2F linear model (x-axis) and S2F sigmoid model (y-axis); right, comparison between S2F hill model (x-axis) and S2F sigmoid model (y-axis).

(TIF)

**S4 Fig. Forward model explains differences in neuronal selectivity between imaging and ephys.** **A.** Fraction of cells that remain selective in synthetic imaging plotted separately for ramp-down and ramp-up cells (left: 6s-AAV synthetic, middle: 6s-TG synthetic, right: 6f-TG synthetic), which is further broken down into right- (blue) and left-preferring (red) trials. **B.** Fraction of right-preferring neurons in imaging after spike inference models. Left: the same analyses as that in Fig 2I, but performed on inferred spiking data obtained via the MCMC framework; right: the same analyses as that in Fig 2I, but performed on inferred spiking data obtained via the MLSpikes framework. **C.** Estimation of the fraction of monophasic and multiphasic neurons that would be discovered by an imaging experiment through use of the S2F forward model. Plots show the estimates for monophasic (left) and multiphasic (right) neurons. The proportion of the source data, ephys, is in black. The experimentally measured proportions in imaging are in gray. Blue color shows the distribution of selectivity type proportion for different repetitions of each algorithm on subsamples of the dataset for synthetic imaging using 6s-AAV (top), 6s-TG (middle) and 6f-TG (bottom) parameters. **D.** Example neurons that change their selectivity after F2S models. Top, a mono-phasic neuron becomes nonselective after F2S model; bottom, a mono-selective neuron becomes multi-phasic after F2S model. **E.** Fraction of selectivity change per selective group after F2S models. Left, MCMC F2S model; right, MLSpikes F2S model. Top, 6s-TG imaging; middle, 6s-AAV; bottom, 6f-TG. First column, non-selective neurons before F2S model; second column, mono-phasic; third column, multi-phasic. First bar, non-selective neurons after F2S model; second bar, mono-phasic; third bar, multi-phasic.

(TIF)

**S5 Fig. Fraction of selective neurons as a function of the imaging signal-to-noise ratio.** We estimated SNR using the procedure in the widely used CaImAn package, in which the noise level is estimated as the exponential of the mean of the logarithm of power spectral density. We then generated datasets including only a subset of neurons by moving the threshold up from its zeroth percentile to its 100th percentile. **A.** non-selective (yellow), mono-selective (blue) and multi-selective neurons (red). Top, 6s-AAV imaging; middle, 6f-TG; right, 6s-TG. **B.** Contra-selective (blue). The remaining was ipsi-selective neurons. **C.** Ramp-up (blue) and ramp-down (red). The remaining was other neurons.

(TIF)

**S6 Fig. Simulation of the effect of slow baseline spike dynamics on fluorescence readout.**

An important assumption of F2S models is that baseline fluorescence reflects zero spikes. This assumption is rarely met. For example, in our study, ALM neurons fire at about 6 Hz in the pre-sample period. This background firing rate, which can vary across time and from neuron to neuron, can distort measures of neural dynamics based on imaging. We explored this effect using computer simulations. The firing rate of a simulated neuron (baseline at 3 Hz) was gradually increased by 2 Hz over four seconds followed by a brief phasic response (1 to 5 spikes were evoked in 70 ms; Fig S6A). We computed the peak over ramp ratio (i.e. ratio of the maximum firing rate during phasic firing to the maximum firing rate before phasic firing) as the measure of the detectability of the phasic activity from tonic activity. We found that the small change of the tonic activity became prominent while detectability of phasic activity was reduced by a factor of  $>10$  in calcium imaging (Fig S6BC). This stems from the integration in calcium dynamics. Although the ramping activity was weak, it was integrated over seconds; although the phasic activity was strong, it was only integrated over 100 ms. The degree to which the detectability was reduced in imaging (comparing to ephys) increased with the level of baseline spike rate (Fig S6D). Therefore, baseline subtraction can be problematic for

inference when the underlying baseline spike rate is unknown. **A.** Simulation (200 trials) of a single neuron, whose firing rate slowly increased from 3 Hz to 5 Hz over ~4 seconds and was then followed by a transient increase (phasic firing) to 15Hz or 30 Hz in 70 ms, and then reset to 3 Hz (baseline). Black dots, spike events; gray dash line, onset time of transient increased spike events. **B.** Spike rate. Black line, phasic firing at 30 Hz; gray, phasic firing at 15 Hz. **C.** Mean  $\Delta F/F_{\text{synth}}$  from S2F model. The change of fluorescence came more from the small change of the baseline firing, and little from the strong phasic firing, which results in difficulty to detect transient modulations of spikes in calcium. **D.**  $\Delta F/F_{\text{synth}}$  peak/ramp ratio is less than that in spike, and such effect increases with baseline spike rates. Colors of circles correspond to baseline spike rates; size corresponds to number of spikes in phasic firing. Dash line corresponds to equal ratio between x and y axes.

(TIF)

**S7 Fig. Analysis of matched datasets from a primary somatosensory area.** Differences between ephys and imaging are likely to depend not only on the analysis and indicator, but also on the underlying dynamics which change from one brain area to the other. We analyzed a second group of matched population recordings, obtained from primary somatosensory area (S1) rather than ALM. We find that differences in some analyses were no longer present, but others remained. We find that the fraction of multiphasic neurons in S1 was far smaller than that in ALM ( $n = 1/55$ , ephys;  $n = 4/719$ , 6s-AAV;  $p < .001$ ,  $\chi^2$  test) and there was no significant difference between the fraction of multiphasic neurons observed in ephys and imaging ( $p = .801$ ,  $\chi^2$  test). Our forward model correctly predicted this lack of change ( $p = .674$ ,  $\chi^2$  test between imaging data and synthetic imaging data). Similarly to ALM data, trial type variance dominated the first principal component in imaging but not in ephys and population decoding was substantially delayed in imaging relative to ephys. **A.** Single neuron selectivity type. Bar plots show fraction of neurons found in each of the three selectivity types (left: monophasic, middle: multiphasic, right: nonselective) for the different recording methods (left: ephys, middle: 6s-AAV, right: 6s-AAV synthetic). **B.** principal component variance content. Bar plots show fraction of variance contained in the first three principal components (from left to right: PC1, PC2, PC3). Each bar is broken into the contribution from trial-type variance (blue), time variance (red) and other (yellow). **C.** Population trial-type decodability. Plot shows mean decodability over time for ephys: top, 6s-AAV: middle and synthetic 6s-AAV: bottom. Dashed lines designate different trial periods (sample, delay response). Note that the experiments with 6s-AAV had a slightly shorter delay period, hence the difference in location of dashed lines. Since 6s-AAV synthetic is derived from ephys it has the same trial structure as ephys.

(TIF)

**S8 Fig. A community based online resource, [im-phys.org](https://im-phys.org), for determining quantitative effects of measuring population activity by imaging or ephys.** **A.** Top, schematic of our community resource that can allow datasets acquired by different labs to be found in one location and matched in analyses. Bottom, schematic of combining different analyses with different datasets on [im-phys.org](https://im-phys.org). **B.** Schematic of using [im-phys.org](https://im-phys.org) to predict values (metric distributions) expected for different population analyses from datasets acquired by different techniques through use of a variety of forward and inverse models.

(TIF)

**S1 Table. Summary of large-scale ephys and imaging recording, more data can be found at [im-phys.org/data](https://im-phys.org/data).** List of datasets. Includes type of dataset, number of neurons, link to dataset, figures in manuscript and citation for data.

(DOCX)

**S2 Table. Summary of simultaneous ephys-imaging recording of single cells in GCaMP6-TG mice.** List of single neurons recorded simultaneously by ephys and imaging. Includes duration of recording, spike rate properties and inferred decay time constant of calcium imaging.  
(DOCX)

## Acknowledgments

We thank Arseny Finkelstein, Christopher Harvey, Daniel Huber, Aaron Kerlin, Daniel O'Connor and Louis K. Scheffer for comments on the manuscript and Nuo Li for many useful discussions. Simultaneous recordings and imaging experiments were performed with support from the Genetically Encoded Neural Indicator and Effector (GENIE) project.

## Author Contributions

**Conceptualization:** Ziqiang Wei, Karel Svoboda, Shaul Druckmann.

**Data curation:** Ziqiang Wei, Bei-Jung Lin, Kayvon Daie, Karel Svoboda.

**Formal analysis:** Ziqiang Wei, Karel Svoboda, Shaul Druckmann.

**Investigation:** Ziqiang Wei, Karel Svoboda, Shaul Druckmann.

**Methodology:** Ziqiang Wei, Bei-Jung Lin, Tsai-Wen Chen, Karel Svoboda, Shaul Druckmann.

**Resources:** Ziqiang Wei, Bei-Jung Lin, Karel Svoboda, Shaul Druckmann.

**Software:** Ziqiang Wei.

**Supervision:** Karel Svoboda, Shaul Druckmann.

**Validation:** Ziqiang Wei, Karel Svoboda, Shaul Druckmann.

**Visualization:** Ziqiang Wei.

**Writing – original draft:** Ziqiang Wei, Karel Svoboda, Shaul Druckmann.

**Writing – review & editing:** Ziqiang Wei, Bei-Jung Lin, Tsai-Wen Chen, Kayvon Daie, Karel Svoboda, Shaul Druckmann.

## References

1. Buzsaki G. Large-scale recording of neuronal ensembles. *Nat Neurosci.* 2004; 7(5):446–51. <https://doi.org/10.1038/nn1233> PMID: 15114356
2. Dhawale AK, Poddar R, Wolff SBE, Normand VA, Kopelowitz E, Ölveczky BP. Automated long-term recording and analysis of neural activity in behaving animals. *eLife.* 2017; 6:e27702. <https://doi.org/10.7554/eLife.27702> PMID: 28885141
3. Ganguly K, Carmena JM. Emergence of a stable cortical map for neuroprosthetic control. *PLoS biology.* 2009; 7(7):e1000153. <https://doi.org/10.1371/journal.pbio.1000153> PMID: 19621062
4. Tolias AS, Ecker AS, Siapas AG, Hoenselaar A, Keliris GA, Logothetis NK. Recording chronically from the same neurons in awake, behaving primates. *J Neurophysiol.* 2007; 98(6):3780–90. <https://doi.org/10.1152/jn.00260.2007> PMID: 17942615
5. Peron SP, Freeman J, Iyer V, Guo C, Svoboda K. A Cellular Resolution Map of Barrel Cortex Activity during Tactile Behavior. *Neuron.* 2015; 86(3):783–99. <https://doi.org/10.1016/j.neuron.2015.03.027> PMID: 25913859
6. Huber D, Gutnisky DA, Peron S, O'Connor DH, Wiegert JS, Tian L, et al. Multiple dynamic representations in the motor cortex during sensorimotor learning. *Nature.* 2012; 484(7395):473–8. <https://doi.org/10.1038/nature11039> PMID: 22538608
7. Sofroniew NJ, Flickinger D, King J, Svoboda K. A large field of view two-photon mesoscope with subcellular resolution for in vivo imaging. *Elife.* 2016;5.

8. Stringer C, Pachitariu M, Steinmetz N, Carandini M, Harris KD. High-dimensional geometry of population responses in visual cortex. *Nature*. 2019; 571(7765):361–5. <https://doi.org/10.1038/s41586-019-1346-5> PMID: 31243367
9. Fu Y, Tucciarone JM, Espinosa JS, Sheng N, Darcy DP, Nicoll RA, et al. A cortical circuit for gain control by behavioral state. *Cell*. 2014; 156(6):1139–52. <https://doi.org/10.1016/j.cell.2014.01.050> PMID: 24630718
10. Peron S, Chen TW, Svoboda K. Comprehensive imaging of cortical networks. *Curr Opin Neurobiol*. 2015; 32:115–23. <https://doi.org/10.1016/j.conb.2015.03.016> PMID: 25880117
11. Peters AJ, Chen SX, Komiyama T. Emergence of reproducible spatiotemporal activity during motor learning. *Nature*. 2014; 510:263. <https://doi.org/10.1038/nature13235> PMID: 24805237
12. Tian L, Hires SA, Mao T, Huber D, Chiappe ME, Chalasani SH, et al. Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat Methods*. 2009; 6(12):875–81. <https://doi.org/10.1038/nmeth.1398> PMID: 19898485
13. Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*. 2013; 499(7458):295–300. <https://doi.org/10.1038/nature12354> PMID: 23868258
14. Akerboom J, Chen TW, Wardill TJ, Tian L, Marvin JS, Mutlu S, et al. Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. *The Journal of neuroscience: the official journal of the Society for Neuroscience*. 2012; 32(40):13819–40.
15. Ohkura M, Sasaki T, Sadakari J, Gengyo-Ando K, Kagawa-Nagamura Y, Kobayashi C, et al. Genetically Encoded Green Fluorescent Ca<sup>2+</sup> Indicators with Improved Detectability for Neuronal Ca<sup>2+</sup> Signals. *PLoS ONE*. 2012; 7(12):e51286. <https://doi.org/10.1371/journal.pone.0051286> PMID: 23240011
16. Inoue M, Takeuchi A, Horigane S, Ohkura M, Gengyo-Ando K, Fujii H, et al. Rational design of a high-affinity, fast, red calcium indicator R-CaMP2. *Nature methods*. 2015; 12(1):64–70. <https://doi.org/10.1038/nmeth.3185> PMID: 25419959
17. Dana H, Mohar B, Sun Y, Narayan S, Gordus A, Hasseman JP, et al. Sensitive red protein calcium indicators for imaging neural activity. *Elife*. 2016; 5.
18. Dana H, Sun Y, Mohar B, Hulse B, Hasseman JP, Tsegaye G, et al. High-performance GFP-based calcium indicators for imaging activity in neuronal populations and microcompartments. *bioRxiv*. 2018:434589.
19. Hamel EJ, Grewe BF, Parker JG, Schnitzer MJ. Cellular level brain imaging in behaving mammals: an engineering approach. *Neuron*. 2015; 86(1):140–59. <https://doi.org/10.1016/j.neuron.2015.03.055> PMID: 25856491
20. Grienberger C, Konnerth A. Imaging calcium in neurons. *Neuron*. 2012; 73(5):862–85. <https://doi.org/10.1016/j.neuron.2012.02.011> PMID: 22405199
21. Scheuss V, Yasuda R, Sobczyk A, Svoboda K. Nonlinear [Ca<sup>2+</sup>] signaling in dendrites and spines caused by activity-dependent depression of Ca<sup>2+</sup> extrusion. *J Neurosci*. 2006; 26(31):8183–94. <https://doi.org/10.1523/JNEUROSCI.1962-06.2006> PMID: 16885232
22. Pologruto TA, Yasuda R, Svoboda K. Monitoring neural activity and [Ca<sup>2+</sup>] with genetically encoded Ca<sup>2+</sup> indicators. *J Neurosci*. 2004; 24(43):9572–9. <https://doi.org/10.1523/JNEUROSCI.2854-04.2004> PMID: 15509744
23. O'Connor DH, Peron SP, Huber D, Svoboda K. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron*. 2010; 67(6):1048–61. <https://doi.org/10.1016/j.neuron.2010.08.026> PMID: 20869600
24. Hromadka T, Deweese MR, Zador AM. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol*. 2008; 6(1):e16. <https://doi.org/10.1371/journal.pbio.0060016> PMID: 18232737
25. Li N, Chen TW, Guo ZV, Gerfen CR, Svoboda K. A motor cortex circuit for motor planning and movement. *Nature*. 2015; 519(7541):51–6. <https://doi.org/10.1038/nature14178> PMID: 25731172
26. Brody CD, Romo R, Kepecs A. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Curr Opin Neurobiol*. 2003; 13(2):204–11. [https://doi.org/10.1016/s0959-4388\(03\)00050-3](https://doi.org/10.1016/s0959-4388(03)00050-3) PMID: 12744975
27. Li N, Daie K, Svoboda K, Druckmann S. Robust neuronal dynamics in premotor cortex during motor planning. *Nature*. 2016.
28. Maravall M, Mainen ZM, Sabatini BL, Svoboda K. Estimating intracellular calcium concentrations and buffering without wavelength ratioing. *Biophys J*. 2000; 78:2655–67. [https://doi.org/10.1016/S0006-3495\(00\)76809-3](https://doi.org/10.1016/S0006-3495(00)76809-3) PMID: 10777761
29. Greenberg DS, Wallace DJ, Voit K-M, Wuertenberger S, Czubayko U, Monsees A, et al. Accurate action potential inference from a calcium sensor protein through biophysical modeling. *bioRxiv*. 2018:479055.

30. Cunningham JP, Yu BM. Dimensionality reduction for large-scale neural recordings. *Nature neuroscience*. 2014.
31. Dana H, Chen T-W, Hu A, Shields BC, Guo C, Looger L, et al. Thy1-GCaMP6 Transgenic Mice for Neuronal Population Imaging In Vivo. *PloS ONE*. 2014; 9(9):e108697. <https://doi.org/10.1371/journal.pone.0108697> PMID: 25250714
32. Theis L, Berens P, Froudarakis E, Reimer J, Roman Roson M, Baden T, et al. Benchmarking Spike Rate Inference in Population Calcium Imaging. *Neuron*. 2016; 90(3):471–82. <https://doi.org/10.1016/j.neuron.2016.04.014> PMID: 27151639
33. Wei Z, Inagaki H, Li N, Svoboda K, Druckmann S. An orderly single-trial organization of population dynamics in premotor cortex predicts behavioral variability. *Nature Communications*. 2019; 10(1):216. <https://doi.org/10.1038/s41467-018-08141-6> PMID: 30644387
34. Guo ZV, Li N, Huber D, Ophir E, Gutnisky DA, Ting JT, et al. Flow of cortical activity underlying a tactile decision in mice. *Neuron*. 2014; 81(1):179–94. <https://doi.org/10.1016/j.neuron.2013.10.020> PMID: 24361077
35. Guo ZV, Hires SA, Li N, O'Connor DH, Komiyama T, Ophir E, et al. Procedures for behavioral experiments in head-fixed mice. *PloS one*. 2014; 9(2):e88678. <https://doi.org/10.1371/journal.pone.0088678> PMID: 24520413
36. Chen T-W, Li N, Daie K, Svoboda K. A Map of Anticipatory Activity in Mouse Motor Cortex. *Neuron*. 2017; 94(4):866–79.e4. <https://doi.org/10.1016/j.neuron.2017.05.005> PMID: 28521137
37. Guo ZV, Inagaki HK, Daie K, Druckmann S, Gerfen CR, Svoboda K. Maintenance of persistent activity in a frontal thalamocortical loop. *Nature*. 2017; 545:181. <https://doi.org/10.1038/nature22324> PMID: 28467817
38. Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, Yuste R, et al. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of neurophysiology*. 2010; 104(6):3691–704. <https://doi.org/10.1152/jn.01073.2009> PMID: 20554834
39. Vogelstein JT, Watson BO, Packer AM, Yuste R, Jedynak B, Paninski L. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophysical journal*. 2009; 97(2):636–55. <https://doi.org/10.1016/j.bpj.2008.08.005> PMID: 19619479
40. Pnevmatikakis EA, Merel J, Pakman A, Paninski L, editors. Bayesian spike inference from calcium imaging data. *Signals, Systems and Computers, 2013 Asilomar Conference on*; 2013: IEEE.
41. Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, et al. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron*. 2016; 89(2):285–99. <https://doi.org/10.1016/j.neuron.2015.11.037> PMID: 26774160
42. Deneux T, Kaszas A, Szalay G, Katona G, Lakner T, Grinvald A, et al. Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nature Communications*. 2016; 7:12190. <https://doi.org/10.1038/ncomms12190> PMID: 27432255
43. Berens P, Freeman J, Deneux T, Chenkov N, McColgan T, Speiser A, et al. Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. *PLOS Computational Biology*. 2018; 14(5):e1006157. <https://doi.org/10.1371/journal.pcbi.1006157> PMID: 29782491
44. Pachitariu M, Stringer C, Harris KD. Robustness of Spike Deconvolution for Neuronal Calcium Imaging. *The Journal of Neuroscience*. 2018; 38(37):7976. <https://doi.org/10.1523/JNEUROSCI.3339-17.2018> PMID: 30082416
45. Inagaki HK, Inagaki M, Romani S, Svoboda K. Low-Dimensional and Monotonic Preparatory Activity in Mouse Anterior Lateral Motor Cortex. *The Journal of Neuroscience*. 2018; 38(17):4163. <https://doi.org/10.1523/JNEUROSCI.3152-17.2018> PMID: 29593054
46. Yasuda R, Nimchinsky EA, Scheuss V, Pologruto TA, Oertner TG, Sabatini BL, et al. Imaging calcium concentration dynamics in small neuronal compartments. *Sci STKE*. 2004; 2004(219):pl5. <https://doi.org/10.1126/stke.2192004pl5> PMID: 14872098
47. Lütcke H, Gerhard F, Zenke F, Gerstner W, Helmchen F. Inference of neuronal network spike dynamics and topology from calcium imaging data. *Frontiers in Neural Circuits*. 2013; 7:201. <https://doi.org/10.3389/fncir.2013.00201> PMID: 24399936
48. Helmchen F, Imoto K, Sakmann B. Ca<sup>2+</sup> buffering and action potential-evoked Ca<sup>2+</sup> signaling in dendrites of pyramidal neurons. *Biophys J*. 1996; 70(2):1069–81. [https://doi.org/10.1016/S0006-3495\(96\)79653-4](https://doi.org/10.1016/S0006-3495(96)79653-4) PMID: 8789126
49. Zariwala HA, Borghuis BG, Hoogland TM, Madisen L, Tian L, De Zeeuw CI, et al. A Cre-Dependent GCaMP3 Reporter Mouse for Neuronal Imaging & In Vivo. *The Journal of Neuroscience*. 2012; 32(9):3131. <https://doi.org/10.1523/JNEUROSCI.4469-11.2012> PMID: 22378886
50. Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. Spontaneous behaviors drive multidimensional, brainwide activity. *Science*. 2019; 364(6437):eaav7893.

51. Harvey CD, Coen P, Tank DW. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*. 2012; 484(7392):62–8. <https://doi.org/10.1038/nature10918> PMID: 22419153
52. Siegel M, Buschman TJ, Miller EK. Cortical information flow during flexible sensorimotor decisions. *Science*. 2015; 348(6241):1352. <https://doi.org/10.1126/science.aab0551> PMID: 26089513
53. Scott BB, Constantinople CM, Akrami A, Hanks TD, Brody CD, Tank DW. Fronto-parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. *Neuron*. 2017; 95(2):385–98.e5. <https://doi.org/10.1016/j.neuron.2017.06.013> PMID: 28669543
54. Kim HR, Malik AN, Mikhael JG, Bech P, Tsutsui-Kimura I, Sun F, et al. A unified framework for dopamine signals across timescales. *bioRxiv*. 2019:803437.
55. Kerlin AM, Andermann ML, Berezovskii VK, Reid RC. Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron*. 2010; 67(5):858–71. <https://doi.org/10.1016/j.neuron.2010.08.002> PMID: 20826316
56. Ohki K, Chung S, Ch'ng YH, Kara P, Reid RC. Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. *Nature*. 2005; 433(7026):597–603. <https://doi.org/10.1038/nature03274> PMID: 15660108
57. Lovett-Barron M, Andalman AS, Allen WE, Vesuna S, Kauvar I, Burns VM, et al. Ancestral Circuits for the Coordinated Modulation of Brain State. *Cell*. 2017; 171(6):1411–23.e17. <https://doi.org/10.1016/j.cell.2017.10.021> PMID: 29103613
58. Briggman KL, Helmstaedter M, Denk W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature*. 2011; 471(7337):183–8. <https://doi.org/10.1038/nature09818> PMID: 21390125
59. Bock DD, Lee WC, Kerlin AM, Andermann ML, Hood G, Wetzel AW, et al. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*. 2011; 471(7337):177–82. <https://doi.org/10.1038/nature09802> PMID: 21390124
60. Komiyama T, Sato TR, O'Connor DH, Zhang YX, Huber D, Hooks BM, et al. Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. *Nature*. 2010; 464(7292):1182–6. <https://doi.org/10.1038/nature08897> PMID: 20376005
61. Jia H, Rochefort NL, Chen X, Konnerth A. Dendritic organization of sensory input to cortical neurons in vivo. *Nature*. 2010; 464:1307. <https://doi.org/10.1038/nature08947> PMID: 20428163
62. Petreanu L, Gutnisky DA, Huber D, Xu NL, O'Connor DH, Tian L, et al. Activity in motor-sensory projections reveals distributed coding in somatosensation. *Nature*. 2012; 489(7415):299–303. <https://doi.org/10.1038/nature11321> PMID: 22922646
63. Xu NL, Harnett MT, Williams SR, Huber D, O'Connor DH, Svoboda K, et al. Nonlinear dendritic integration of sensory and motor input during an active sensing task. *Nature*. 2012; 492(7428):247–51. <https://doi.org/10.1038/nature11601> PMID: 23143335
64. Huang L, Knoblich U, Ledochowitsch P, Lecoq J, Reid RC, de Vries SEJ, et al. Relationship between spiking activity and simultaneously recorded fluorescence signals in transgenic mice expressing GCaMP6. *bioRxiv*. 2019:788802.
65. Inoue M, Takeuchi A, Manita S, Horigane S-i, Sakamoto M, Kawakami R, et al. Rational Engineering of XCaMPs, a Multicolor GECI Suite for In Vivo Imaging of Complex Brain Circuit Dynamics. *Cell*. 2019; 177(5):1346–60.e24. <https://doi.org/10.1016/j.cell.2019.04.007> PMID: 31080068
66. Abdelfattah AS, Kawashima T, Singh A, Novak O, Liu H, Shuai Y, et al. Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. *Science*. 2019:eaav6416.
67. Adam Y, Kim JJ, Lou S, Zhao Y, Xie ME, Brinks D, et al. Voltage imaging and optogenetics reveal behaviour-dependent changes in hippocampal dynamics. *Nature*. 2019; 569(7756):413–7. <https://doi.org/10.1038/s41586-019-1166-7> PMID: 31043747
68. Jun JJ, Steinmetz NA, Siegle JH, Denman DJ, Bauza M, Barbarits B, et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature*. 2017; 551:232. <https://doi.org/10.1038/nature24636> PMID: 29120427
69. Lin B-J, Chen T-W, Kim DS, Svoboda K. Simultaneous calcium imaging using GCaMP sensors and electrophysiology in L2/3 pyramidal neurons of the visual cortex in thy1 transgenic mice. 2016.
70. Tsien RY. Fluorescent probes of cell signaling. *Annu Rev Neurosci*. 1989; 12:227–53. <https://doi.org/10.1146/annurev.ne.12.030189.001303> PMID: 2648950
71. Guo Y, Hastie T, Tibshirani R. Regularized linear discriminant analysis and its application in microarrays. *Biostatistics*. 2007; 8(1):86–100. <https://doi.org/10.1093/biostatistics/kxj035> PMID: 16603682