1
2 **A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology**

3 Ziqiang Wei[1,2], Bei-Jung Lin[1,3], Tsai-Wen Chen[1,3], Kayvon Daie[1], Karel Svoboda[1,*], Shaul Druckmann[1,4,*]

4 [1] Janelia Research Campus, HHMI, Ashburn, VA 20147

5 [2] The Solomon H. Snyder Department of Neuroscience, Johns Hopkins University School of Medicine,
6 Baltimore, MD 21205

7 [3] Institute of Neuroscience, National Yang-Ming University, Taipei, Taiwan

8 [4] Department of Neurobiology, Stanford University

9 [*] Correspondence: svobodak@janelia.hhmi.org (K.S.), shauld@stanford.edu (S.D.)

10 **Summary (150 words)**

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

22   **Introduction**

23   Electrophysiological recordings ('ephys') and calcium imaging offer distinct tradeoffs for interrogating
24   activity in neural populations. Ephys directly reports spiking of neurons with high signal-to-noise ratio,
25   temporal fidelity, and dynamic range, but typically offers access only to a sparse subset of relatively active
26   neurons (Buzsaki, 2004). In addition, the ability to track the same population of neurons across time,
27   important for understanding the neural basis of learning, remains challenging (Dhawale et al., 2017;
28   Ganguly and Carmena, 2009; Tolias et al., 2007).

29   Calcium imaging provides access to large numbers of neurons simultaneously (Huber et al., 2012; Peron
30   et al., 2015b; Sofroniew et al., 2016; Stringer et al., 2019), potentially with cell type specificity (Fu et al.,
31   2014; Peron et al., 2015a). Moreover, calcium imaging can track the activity of the same neuronal
32   populations over time (Huber et al., 2012; Peters et al., 2014). Indeed, with the development of sensitive
33   fluorescent protein-based indicators (Akerboom et al., 2012; Chen et al., 2013; Dana et al., 2016; Dana et
34   al., 2018; Inoue et al., 2015; Ohkura et al., 2012; Tian et al., 2009) and powerful new imaging methods
35   (Hamel et al., 2015; Sofroniew et al., 2016) calcium imaging has been rapidly adopted for measurements
36   of neural population activity.

37   However, calcium imaging reports spikes only indirectly (Grienberger and Konnerth, 2012; Peron et al.,
38   2015a). The transformation from spikes to calcium is inherently non-linear due to the dynamics of
39   intracellular calcium concentrations (Scheuss et al., 2006). Additional nonlinearities are imposed by the
40   protein-based indicators of calcium (Akerboom et al., 2012; Chen et al., 2013; Pologruto et al., 2004; Tian
41   et al., 2009). Together these produce a low-pass filtered, delayed, and transformed version of neural activity,
42   which complicates relating neural activity to behavior. Calcium imaging also has lower signal-to-noise ratio
43   for detecting spikes and limited dynamic range (Peron et al., 2015a). In addition, during animal behavior,
44   spike rates can vary by orders of magnitude across behavioral epochs and across neurons, even neurons
45   of the same type (Hromadka et al., 2008; Li et al., 2015; O'Connor et al., 2010) and spike rates change
46   over times of milliseconds to seconds (Brody et al., 2003; Li et al., 2015; Li et al., 2016). Finally, the coupling
47   between spikes and calcium-dependent fluorescence likely differs across different neuron types and even
48   individual neurons within a type (Chen et al., 2013; Maravall et al., 2000).

49   The complexities in the relation between spiking activity and calcium imaging at the level of single neurons
50   have been long appreciated (Akerboom et al., 2012; Chen et al., 2013; Greenberg et al., 2018; Pologruto
51   et al., 2004; Scheuss et al., 2006; Tian et al., 2009). However, the effect of these factors on analyses of
52   population activity are not yet fully known (Cunningham and Yu, 2014). Ideally a detailed understanding of
53   the transformation from spikes-to-calcium-dependent fluorescence would allow inversion of this
54   transformation and the reliable extraction of spikes. Calcium indicators with high sensitivity allow reliable
55   detection of action potentials, at least under conditions when single spikes or burst of spikes are separated
56   in time (Chen et al., 2013; Dana et al., 2014; Theis et al., 2016). However, under behaviorally relevant
57   conditions neurons operate with a large range of spike rates, and spiking responses are typically
58   superposed on a substantial background spike rate, which varies across the population (Li et al., 2015;
59   O'Connor et al., 2010). Moreover, neuron-to-neuron variability in calcium dynamics, calcium indicator
60   dynamics and patterns of firing rate could conspire to make this inversion challenging. These issues are
61   compounded by the paucity of simultaneously recorded spikes and fluorescence data. It is therefore unclear
62   if spike inference can invert the fluorescence data accurately to eliminate potential discrepancies between
63   analyses performed on ephys and calcium imaging data.

64   Here, we explore these issues empirically in the context of a decision-making task, where the dynamics of
65   the neural circuit are rich and variable across neurons. In particular, neurons in frontal cortex show a wide
66   range of spike rates and exhibit diverse temporal dynamics and selectivity (Brody et al., 2003; Li et al.,
67   2015; Wei et al., 2019). We analyzed ephys and calcium imaging measured in matched neuronal
68   populations in the same delayed response task and directly compared the results of standard
69   measurements of selectivity and population dynamics. We find qualitative discrepancies at both the level
70   of single cells and neural populations. Spike inference algorithms were limited in resolving these differences.
71   However, a phenomenological model of the spike-to-fluorescence transformation, based on simultaneous

72   imaging and electrophysiology data (Akerboom et al., 2012; Chen et al., 2013), explains many differences
73   across the data sets. Finally, we developed a web-based platform, im-phys.org, that allows quantification
74   of the effects of various transformations from electrophysiology to imaging.

Results

76   We measured neural activity using electrophysiology ('ephys') and calcium imaging under identical
77   behavioral conditions and in matched neural populations, but in separate experiments. Mice performed a
78   tactile delayed response task (Guo et al., 2014a; Guo et al., 2014b; Li et al., 2015) (**Figure 1A**). In each
79   trial, mice judged the location of an object with their whiskers. During the subsequent delay epoch
80   (approximately 1.3 seconds), mice planned an upcoming response. Following an auditory 'go' cue, mice
81   reported object location with directional licking (lick-left or lick-right).

82   Two-photon calcium imaging and ephys were performed in left anterolateral motor cortex (ALM; **Figure
83   1BDF**). We report the results of three variants of calcium indicators in this study: GCaMP6s delivered by
84   viral gene transfer, and GCaMP6s and GCaMP6f expressed in Thy-1 transgenic mice. In the first series of
85   imaging experiments, neurons were transduced with adeno-associated virus expressing GCaMP6s (6s-
86   AAV), a widely-used method (Chen et al., 2013; Peron et al., 2015a) (data from (Li et al., 2015), 1493
87   neurons, 4 mice). In the second, neural activity was recorded by imaging transgenic mice expressing
88   GCaMP6s in cortical pyramidal neurons (6s-TG, data from (Chen et al., 2017), 2293 neurons, 1 mouse).
89   We treated these datasets separately since the mode of delivery of GCaMP can affect its properties.
90   Specifically transgenic GCaMP typically results in neurons that have lower GCaMP6 expression levels and
91   faster fluorescence dynamics compared to neurons transduced with AAV (Dana et al., 2014). Finally, we
92   collected a dataset obtained with a faster, but less sensitive indicator, GCaMP6f (6f-TG, 2672 neurons, 2
93   mice). We refer to these three datasets as 6s-AAV, 6s-Tg and 6f-Tg, respectively. We compared this data
94   to ephys data acquired with silicon probes that record multiple neurons simultaneously (720 neurons, 19
95   mice (Li et al., 2015)) (**Figure 1CEG**). Ephys recordings were subsampled so that their recording depths
96   matched the generally more superficial calcium imaging experiments. Neurons were recorded by 6s-AAV
97   and 6s-Tg at $120 – 740$ μm. The matched ephys subset was taken at $100 – 800$ μm leaving n = 720 neurons.
98   Neurons were recorded by 6f-Tg at $140 – 470$ μm. The matched ephys subset was taken at $100 – 470$ μm,
99   leaving n = 225 neurons.

Filtering of selectivity by calcium imaging

101   Individual ALM neurons exhibit diverse temporal dynamics, including changes in selectivity over time
102   (**Figure 2B**) (Guo et al., 2014b; Li et al., 2015; Wei et al., 2019). We classified dynamics into three
103   categories: 'monophasic' neurons showed consistent selectivity across the trial (**Figure 2A**); 'multiphasic'
104   neurons changed selectivity over time (defined as having consistent selectivity for at least 335 ms which
105   then changes and remains stable for at least 335 ms more) (**Figure 2B**); 'non-selective' neurons responded
106   similarly across trial types but were still modulated during the task (**Figure 2C**). The proportion of
107   monophasic selective neurons was similar between the datasets (58% ephys; 66% 6s-AAV; 50% 6s-Tg;
108   45% 6f-Tg). However, the ephys data set contained a substantial proportion of multiphasic neurons
109   (220/720; 31%), much larger than the imaging datasets (6s-AAV: 76/1493, 5%; 6s-Tg, 98/2293, 4%;
110   compare to matched ephys, 220/720, 31%; 6f-Tg, 69/2672, 3%; compare to matched ephys, 52/225, 20%;
111   $p < .001$, χ^2 test; **Figure 2D-F**). As neural response properties can change across cortical layers, we
112   performed a more detailed analysis of the effect of recording depth on single neuron selectivity and find
113   that selectivity was reduced in imaging compared to ephys across depths (**Figure S1A-D**).

114   What could account for this difference? Ephys records a sparse subset of neurons blindly, which could
115   introduce biases, for example a bias towards neurons with higher spike rates (**Figure S1E-I**). In contrast,
116   in our imaging experiments all visualized neurons were analyzed. In addition, the spike sorting procedure
117   used to identify units from raw electrode potentials can introduce artifacts, including erroneous merging of
118   neurons (**Figure S1J**). However, we found that these factors were unlikely to explain our results for two
119   reasons. First, we considered intracellular recordings for which spike sorting is not required (Guo et al.,
120   2017). The fraction of multiphasic neurons was 25.7% (n = 9/35), similar to the extracellular ephys data (p

121   = .67, χ^2 test; **Figure 2G**) but significantly different from the imaging data (p < .001, χ^2 test with both 6s-
122   AAV and 6s-Tg). Second, we tested the question of spike-sorting induced biases by considering synthetic
123   data in which we deliberately introduced merges at different probabilities. Merging neurons generated more
124   multi-selective neurons when two neurons with different temporal selectivity profiles were merged, but the
125   ratio of accidental merging had to be unrealistically high to explain the difference between datasets (**Figure
126   S1J**).

127   Another source of difference could be that our comparisons so far were performed on the imaging data, not
128   on spike rates inferred from the imaging data. Spike inference algorithms attempt to undo the transformation
129   from spikes to claicum-dependent fluorescence, thereby recovering spike times (or spike rates) from
130   imaging data (Berens et al., 2018; Deneux et al., 2016; Pnevmatikakis et al., 2013; Pnevmatikakis et al.,
131   2016; Theis et al., 2016; Vogelstein et al., 2010; Vogelstein et al., 2009). We tested two published methods:
132   MLSpike (Deneux et al., 2016) and MCMC (Pnevmatikakis et al., 2013). In our hands, spike inference only
133   partially corrected the differences between the datasets and in some cases actually pushed the data even
134   further apart (**Figure 2H**). For instance, MLSpike produced even lower proportions of multiphasic neurons.
135   MCMC was more accurate, increasing the proportions of multiphasic neurons, but still far short of the actual
136   proportion in the ephys dataset (and for 6s-Tg and 6f-Tg decreased instead of increased the proportion of
137   monophasic neurons). Deconvolution at best recovered about half of the missing multiphasic selectivity
138   (6s-AAV, 18%; 6s-TG, 17%, compared to 31% in matched ephys; 6f-TG, 8%, compared to 20% in matched
139   ephys; **Figure 2H**).

140   Differences between calcium and ephys were not limited to the temporal nature of responses but were also
141   present in trial-type selectivity. Namely we found that in the ephys dataset right-preferring neurons (i.e.,
142   neurons whose firing rate before right licks was higher than before left licks) were as common as left-
143   preferring neurons (**Figure 2I**, left; p = .118, χ^2 test) (Guo et al., 2014b; Li et al., 2015). The same was true
144   for imaging with a fast calcium indicator (**Figure 2I**, right; p = .102, χ^2 test), but not for imaging with slow
145   indicators (**Figure 2I**, left; p < .001, χ^2 test; **Figure S4B**, spike-inference measure). What could be the
146   cause of these differences? Spike rates in individual ALM neurons often increase or decrease during a trial
147   in ramp-like patterns (Guo et al., 2014b; Inagaki et al., 2018; Li et al., 2015). Right-preferring selectivity was
148   more often associated with neurons ramping up on right trials, whereas left-preferring selectivity included
149   many neurons with firing rates ramping down in the non-preferred ('right') trial (**Figure 2J).** The large
150   difference between the rise and decay times of calcium indicators could lead to differences in how the
151   selectivity of neurons that ramp up or ramp down gets transformed by the indicator. To test to what degree
152   such explanations explain the data we developed a model of the spike-to- fluorescence transformation.

Simultaneous loose-seal electrophysiology and calcium imaging

154   Modeling the spike-to-fluorescence transformation requires simultaneously electrophysiology and calcium
155   imaging at the level of individual neurons. Since this data is not available for the transgenic mice used here
156   we performed loose-seal recordings and calcium imaging in individual neurons (**Figure 3**). The dataset
157   consists of gCAMP6f- and gCAMP6s-expressing L2/3 neurons in transgenic mice (6s-TG, 22 cells; 6f-TG,
158   18 cells; **Table S1**). This new data more than doubles the number of previously available simultaneously
159   recorded neurons (Theis et al., 2016). In addition we used published data with AAV-based gene
160   transduction (Chen et al., 2013)(http://dx.doi.org/10.6080/K02R3PMN). Bursts of spikes produced
161   fluorescence transients in the imaged neurons (**Figure 3A-D**). The ability to detect single spikes varied
162   considerably between neurons (**Figures 3E-G, S2**). The detection of single spikes was lower in transgenic
163   mice than with AAV-based gene transduction, likely reflecting the lower expression level in the transgenic
164   mice.

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

166   Using the newly recorded data we developed a spike-to-fluorescence (S2F) forward model to generate a
167   synthetic calcium imaging data based on a neuron's spike train (**Figure 4A**) (Akerboom et al., 2012; Chen
168   et al., 2013; Lütcke et al., 2013; Yasuda et al., 2004). In brief, spike times were first converted to a latent

169   variable, c(t), by convolution with a double-exponential kernel, with parameters rise-time (τ_r) and decay-
170   time (τ_d). This latent variable was pushed through a non-linearity, F(c), with a non-linearity sharpness
171   parameter (k), a half-activation parameter ($c_{1/2}$, corresponding to the half-rise point of the nonlinearity) and
172   a maximum fluorescence change (F_m) (**Materials and methods**). The neurons were well-fit by the model
173   (**Figure S3B**; variance explained, 6s-AAV, .87 $\pm$ .17, mean $\pm$ std.; 6s-TG, .80 $\pm$ .20; 6f-AAV, .82 $\pm$ .27, mean
174    $\pm$ std.; 6f-TG, .66 $\pm$ .23). The inferred parameter values reflected known indicator kinetics. For instance, the
175   decay times measured for neurons expressing GCaMP6s were longer than those expressing GCaMP6f
176   (**Figure 4B**). However, there was substantial variability between the parameter values inferred across
177   neurons (**Figures 4B, S3**). This variability is one factor that could explain the difficulty of the inversion of
178   calcium responses which is central to spike inference approaches. We refer to simulations of calcium-
179   dependent fluorescence based on application of the S2F model to spiking activity as '$\Delta F/F_{Synth}$'.

180   We applied the model to ramp-up and ramp-down neurons. For ramp-up cells the separation of activity
181   across trial types was retained in $\Delta F/F_{Synth}$, albeit with slower dynamics (**Figure 4C**). In contrast, for many
182   ramp-down cells $\Delta F/F_{Synth}$ became non-selective (**Figure 4D**). Overall, selectivity was conserved more
183   frequently for ramp-up cells than for ramp-down cells. Since right-preferring cells were more often
184   associated with ramp-up dynamics, and calcium imaging is more likely to capture ramp-up selectivity than
185   ramp-down selectivity, the model explains the greater fraction of right-preferring neurons in the calcium
186   imaging data (**Figure 4E).** This was true whether a neuron happened to be a right- or left-preferring neuron,
187   i.e., there were no significant differences in the fraction of detectability in the synthetic data once the data
188   was broken down into two categories, ramp-up and ramp-down (p > .05, χ^2 test for all imaging conditions;
189   **Figure S4A**). Consistent with the difference being produced by the slow decay kinetics of GCaMP6s, there
190   was little difference between the fraction of right- and left-preferring neurons in the 6f-TG data (p > .05 for
191   both cell types). In line with these results, we found that the forward model accounted for the drop in
192   multiphasic neurons presented in the previous section (**Figure S4C**). These data show that the spike-to-
193   fluorescence transformation introduces systematic discrepancies in comparing the same analysis
194   performed on ephys or imaging data.

Dimensionality reduction emphasizes different sources of variance in ephys and imaging

196   Large-scale recording methods are often used in combination with dimensionality reduction techniques to
197   provide a compact description of the data (Cunningham and Yu, 2014). For example, principal component
198   analysis (PCA) finds modes of population activity that capture the largest amount of variance in neural
199   activity (Cunningham and Yu, 2014). Data visualization and analysis are often performed after truncating
200   the decomposition after a few components. We found that the contribution of different sources of variance
201   to the first principal components diverges between ephys and imaging. Accordingly, truncation of PCA in
202   the first few principal components can lead to a qualitatively different PCA decomposition of neural activity
203   between ephys and imaging.

204   We found substantial differences in performing PCA on ephys and imaging datasets. First, the content of
205   the first PCs was remarkably different between ephys and imaging. In the ephys data, variance in the first
206   PC was mostly due to temporal dynamics (98.71 $\pm$ 0.06%, mean $\pm$ std., bootstrap analysis). In contrast, for
207   GCaMP6s imaging trial-type selectivity was the dominant source of variance in the first PC (6s-AAV: 60.39
208    $\pm$ 0.29%; 6s-TG: 44.51 $\pm$ 0.65%) (**Figure 5A**). This difference was consistent with the temporal smoothing
209   imposed by slower indicators, and as expected temporal dynamics were predominant in the first PC of
210   GCaMP6f, closer to the values found in ephys, (6f-TG: 64.87 $\pm$ 2.47%; depth matched ephys: 91.02 $\pm$
211   0.14%). Second, in the ephys data, a relatively large number of PCs (> 10) contribute substantially to the
212   variance, whereas in imaging and synthetic imaging most variance was explained by the first few PCs (test
213   for number of PCs required to explain 90% of the variance, p < .001; t test, bootstrap).

214   These differences in the sources of explained variance can be seen in the profiles of the PC scores (**Figure
215   5B**) as well as in the profiles obtained by a standard exploratory visualization, depicting the evolution of
216   activity over time as a trajectory in the space of the first two PCs (**Figure 5C**). Spike inference algorithms
217   correctly reduced the amount of trial-type variance in the first principal components (although not fully), but

218 with the caveat that the fraction of variance in the first two principal components was reduced too much
219 (**Figure S5**).The spike-to-fluorescence model captured the qualitative differences between ephys and
220 imaging, but overestimated the increase in variance in the first two principal components (**Figure 5A-C**).

221 **Population activity history affects instantaneous decoding differently in ephys and imaging**

222 Decoding analysis relating population activity to behavioral variables is widely used in systems
223 neuroscience (Ganguly and Carmena, 2009; Harvey et al., 2012; Huber et al., 2012; Wei et al., 2019). Such
224 analyses typically relate the state of population activity at a given time point to a behavioral variable of
225 interest, such as behavioral choice. We performed decoding analysis to predict either trial type or the
226 current behavioral epoch from population activity. Decodability of trial type in ephys increased earlier (one-
227 tail t-test, p < .001), but saturated at a lower level (one-tail t-test, p < .001) than in calcium imaging (**Figure
228 6A**). Spike inference models, the MCMC framework in particular, partially reduced the delay of the rise of
229 decodability but overestimated the decrease in decodability yielding lower performance in delay-response
230 epoch than the ephys data (**Figure S6**). Both observations were recapitulated by the S2F model (delay:
231 one-tail rank sum test, 6s-AAV, p < .001, 6s-TG, p < .001, 6f-TG, p < .001; enhancement: 6s-AAV, p < .001,
232 6s-TG, p < .001, 6f-TG, p < .001; **Figure 6B-C**). The counterintuitive result of higher decoding accuracy in
233 imaging for matched population size is explained by the long decay time of slow calcium imaging. The long
234 integration in calcium imaging causes instantaneous decoding on imaging to be equivalent not to
235 instantaneous decoding on spiking data, but to decoding on a more time averaged variable. Such a choice
236 is advantageous when a larger proportion of the selectivity is stable, as was the case in ALM sample and
237 delay selectivity (Li et al., 2016). Consistently, decoders built on ephys that incorporated a one second
238 integration time were more accurate than instantaneous ephys decoders and as accurate as slow indicators
239 (**Figure 6D**). The delayed increase of decodability was also explained by the forward model. GCaMP6f,
240 with its reduced signal to noise, yielded less accurate population decoders (**Figure 6E**; spike inference
241 measure, **Figure S6**).

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

254 **Population dynamics is temporally dispersed in calcium imaging**

255 Neurons show temporally complex responses, even in simple trial-based behaviors (Brody et al., 2003; Li
256 et al., 2015). These spike rate changes are critical for an understanding of neural circuit models of neural
257 computation. Our analysis revealed a qualitative difference in the dynamics between populations recorded
258 by ephys or imaging: a dispersion of the apparent dynamics. That is, the spike rates recorded in ALM
259 peaked at transitions between behavioral epochs (**Figure 7A**) (Li et al., 2015). In contrast, in the calcium
260 imaging data, peaks of fluorescence were delayed and jittered, causing a more sequence-like appearance
261 (**Figure 7B**) (Harvey et al., 2012; Scott et al., 2017).

262 To quantify this effect we computed a measure of the 'peakiness' of the distribution of neuronal activity ('s')
263 across recording modalities as the difference between observed neural activity and temporally uniformly
264 distributed neural activity ($P = \frac{1}{2T}$):

265
$$s = \frac{1}{P} \sqrt{\frac{1}{2T} \sum_{i=A,B} \int_0^T dt (P_i(t) - P)^2}$$

266 s was much larger for the ephys dataset (1.27 ± 0.23) compared to the 6s-AAV (0.49 ± 0.03; one-tail rank
267 sum test, p < .001), 6s-TG (0.38 ± 0.04; one-tail rank sum test, p < .001), and 6f-TG (0.58 ± 0.07; one-tail
268 rank sum test, p < .001) imaging data (**Figure 7C**). The forward model was able to recapitulate the
269 differences between ephys and imaging (s = 0.39 ± 0.04, 6s-AAV $\Delta F/F_{Synth}$; s = 0.37 ± 0.03, 6s-TG $\Delta F/F_{Synth}$;
270 s = 0.71 ± 0.07, 6f-TG $\Delta F/F_{Synth}$; **Figure 7DE**). Using the forward model we found that the degree of delay
271 in the peak response is dependent on interactions between multiple factors including the assumed temporal
272 and non-linear parameters of the indicator, as well as the absolute value of the underlying firing rate (**Figure
273 7FG**). Here, spike inference algorithms were able to partially undo the difference between imaging and
274 ephys, yielding a reduction in the temporal dispersal (**Figure 7H**). Similar overall results were obtained with
275 different metrics for the sharpness of the maximum-activity-time distribution relative to a uniform distribution,
276 such as the Kullback-Leibler divergence.

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

Discussion

280 Calcium imaging using fluorescent protein sensors is a powerful method for recording activity in large
281 neuronal populations (Peron et al., 2015a). In systems neuroscience, cellular calcium imaging fills a
282 complementary role to extracellular electrophysiology. Imaging can sample neural activity densely (Peron
283 et al., 2015a; Peron et al., 2015b) and reveal spatial relationships between neurons with related activity
284 patterns (Kerlin et al., 2010; Ohki et al., 2005). Imaging can be used in a cell-type specific mode to sample
285 rare neuronal populations that are difficult to target using electrophysiology (Fu et al., 2014). Imaging can
286 be combined with post-experiment molecular analysis (Kerlin et al., 2010; Lovett-Barron et al., 2017;
287 O'Connor et al., 2010) or even serial electron microscopy reconstruction (Bock et al., 2011; Briggman et al.,
288 2011). Imaging can track the activity of individual neurons over long time scales to explore the circuit basis
289 of learning (Huber et al., 2012; Komiyama et al., 2010). Finally, imaging allows recording activity in neuronal
290 microcompartments that are not accessible to electrophysiology (Chen et al., 2013; Jia et al., 2010;
291 Petreanu et al., 2012; Xu et al., 2012). Electrophysiological recordings report neural activity with high
292 temporal precision but have limitations of their own. Ephys recordings have a bias towards large neurons
293 with high spike-rates. In addition, the process of transforming raw recordings into spike times associated
294 with individual isolated units, i.e., spike sorting, can introduce artifacts such as merging spikes from different
295 neurons.

296 Although calcium imaging and ephys are often used almost interchangeably, the quantitative effects of the
297 differences between ephys and imaging on measures typically used in system neuroscience have not been
298 evaluated in a systematic manner. By comparing activity recorded with electrophysiology or imaging from
299 the same circuit during the same behavioral task we showed that the different recording methods can lead
300 to diverging results. On the level of single neurons, the proportion of neurons with specific response
301 properties and different dynamics of selectivity differs between calcium imaging and ehpys. At the level of
302 neuronal populations, we find diverging results for the content of population activity variance (trial condition
303 differences being the main source of variance in imaging while temporal dynamics are the main source of
304 variance in ephys), the relation of population activity to behavior, and the overall pattern of population
305 dynamics. Spike inference algorithms only partially recovered the difference between ephys and imaging
306 across the multiple metrics considered in this study (**Figure S8B**). Notably, we find large neuron-to-neuron
307 variability in the inferred parameters of a forward, spike-to-fluorescence model. Such variability coupled
308 with the large heterogeneity in firing rates and temporal patterns makes correctly solving the inverse
309 problem difficult, which potentially explains our results. At the same time, most of the differences we found
310 between ephys and imaging were explainable by a forward-model that generates a synthetic imaging
311 experiment counterpart of a neuron's ephys responses. Such a model takes into account the specific
312 heterogeneity found in ephys recordings and can take into account neuron-to-neuron variability in calcium

313   imaging properties by sampling randomly from the varying parameters of the spike-to-fluorescence
314   transformation.

Im-phys.org – a website for more detailed comparison ephys and imaging

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

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

333   Differences in metrics of population activity between calcium imaging and ephys not only complicate the
334   research literature but can result in the divergence of models used to understand the underlying data. Most
335   population models, whether models in which the single units are modeled in more biophysical detail or more
336   abstractly, are still highly reduced in the way they treat population heterogeneity. As such they often rely
337   on dimensionality reduction of the recorded data to define the aspects of population activity the model is
338   meant to capture. We found substantial differences between ephys and imaging data in application of PCA,
339   and the truncation of the data after a few important data components can further amplify differences. In
340   extreme cases one may be left with subsets that differ dramatically across imaging and electrophysiology.
341   The amplification of difference by dimensionality reduction is relevant not just for modeling of the data, but
342   more generally when generic forms of dimensionality reduction, such as PCA, are used early in the analysis
343   pipeline to improve signal-to-noise ratio (which is important given the limited duration of typical behavioral
344   experiments) for subsequent analysis, such as population decoding. Dimensionality reduction can be hard
345   to avoid when analyzing large datasets (Cunningham and Yu, 2014), but can be modified to be less
346   sensitive to known issues.

Going forward

348   Going forward, the discrepancies between ephys and calcium imaging can be reduced by improvements in
349   calcium indicators and adjustments to experimental design. Calcium indicators could be improved on
350   multiple fronts. They could be made faster and less nonlinear (Dana et al., 2018; Inoue et al., 2019). In
351   addition, more uniform expression across cells can allow for more aggressive modeling of the nonlinearities
352   that cannot be reduced, especially when coupled with priors on activity profiles derived from large scale
353   electrophysiology. Faster indicators will result in the effect of previous activity history washing away faster,
354   thus reducing effects that are history dependent. Imaging with multiple types of indicators in different
355   experiments might produce additional constraints and help reduce biases. Voltage imaging holds great
356   potential for fast accurate measurement of spiking activity, at least in sparsely labeled neuronal populations
357   (Abdelfattah et al., 2019; Adam et al., 2019). At the level of experimental design, when population activity
358   in a given behavioral epoch involves fixed dynamics, such as settling to a steady state or consistent ramping,
359   longer trial epochs will allow the effect of the previous dynamical state to decay away. Indeed, we found a
360   smaller discrepancy between the number of multiphasic neurons in ephys and 6s-TG data when the
361   behavioral paradigm was adapted to use longer delay epochs. If priors on the profile of activity are known

362 from electrophysiological recordings, an effort that should become easier as neurophysiology probes
363 become more powerful (Jun et al., 2017), data sharing more common, and preprocessing more
364 standardized, fluorescence-to-spike models can assist in evaluating experimental design and its effect on
365 dynamics.

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

374 **Authorship**

375 Z.W, K.S and S.D conceived the project. B.J.L performed simultaneous ephys-imaging recordings. T.W.C
376 and K.D. performed 6s-TG and 6f-TG imaging experiments. Z.W, K.S. and S.D performed model
377 simulations and analyzed the data. Z.W, K.S and S.D wrote the paper with comments from the other authors.

386 **Competing interests**

387 All authors have no competing interests declared.

388   **Materials and methods**

389   **Electrophysiological and imaging population activity recordings**

390   Electrophysiological ('ephys') or calcium imaging recordings were performed in separate experiments (Li et
391   al., 2015). Mice were trained to perform a delayed version of a tactile discrimination task. Mice reported the
392   position of a pole (anterior or posterior) by directional licking (lick-left or lick-right) after a delay period. The
393   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
394   s. Trials with early licking were excluded from analysis. Neuronal depths were 100 to 800 um (ephys), 150
395   – 740 um (6s-AAV), 120 – 640 um (Thy1-GP4.3 mice, 6s-TG), and 140 – 470 um (Thy1-GP5.17 mice, 6f-
396   TG). Only sessions with more than 20 trials for each type (right-trial and left-trial) were included. For imaging
397   data, we performed a post-hoc detection of outliers and removed trials where more than 30% of the time
398   points contain a signal with 3 standard deviations away from median (these outliers relate to baseline
399   fluctuations across trials, and removing them was necessary for variance-based analysis). Neurons were
400   limited to putative pyramidal neurons. These reduced the total number of neurons with sufficient number of
401   trials, yielding 1493, 2293, and 2672 units for 6s-AAV, 6s-TG and 6f-TG imaging, respectively.

402   We used two sets of data from loose-seal electrophysiological recordings and imaging from GCaMP6-
403   expressing neurons in primary visual cortex. In one set neurons were transduced with 6s-AAV and 6f-AAV
404   (data from (Chen et al., 2013)). In the other set we used 6s-TG and 6f-TG mice (Dana et al., 2014; Lin et
405   al., 2016). More details of all datasets are described at http://im-phys.org/data.

406   Simultaneous loose-seal recordings and imaging (Figure 3 and Figure S2) was performed as described
407   previously(Chen et al., 2013) (more details at http://im-phys.org/data). GP4.3 and GP5.17 mice (Dana et
408   al., 2014) were lightly anesthetized (0.5% isoflurane). Drifting grating visual stimuli were used to drive
409   activity in the visual cortex. Loose-seal recordings were made through a craniotomy windows over the
410   primary visual cortex. Two-photon imaging and loose-seal, cell-attached recordings were performed
411   simultaneously. We acquired images in both low (284 x 284 um^2) and high (38 x 38 um^2) zoom
412   configurations. Extraction of fluorescence transients was as described (Chen et al., 2013).

413   To analyze the spike-triggered fluorescence changes, we created 1.2-s snippets around action potentials
414   (APs), where a few APs only happened from 200 ms to 400 ms from the onset of each snippet. We
415   computed baseline fluorescence using the snippets without AP in the entire time series. For snippet with
416   APs, we required the fluorescence changes within the first 200 ms (before APs) was around baseline level
417   (**Figure 3C, S2A**). We computed ROC curve as the probability if the peak fluorescence changes after 1
418   (insert panels) or many APs (**Figure 3E, S2C**) can be detected at certain threshold comparing to baseline
419   fluorescence fluctuations. D-prime is computed as $\frac{<\max(\Delta F/F)_{1AP}> - <\max(\Delta F/F)_{no\ AP}>}{\mathrm{std}(\Delta F/F)_{no\ AP}}$.

420   **Spike-to-fluorescence model**

421   We developed a phenomenological model that converts spike times to synthetic fluorescence time series
422   (Akerboom et al., 2012; Chen et al., 2013; Li et al., 2015; Yasuda et al., 2004). This 'spike-to-fluorescence'
423   (S2F) model consists of two steps. First, spikes at times $\{t_k\}$ are converted to a latent variable, c(t), by
424   convolution with a double-exponential kernel:

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

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

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

430 k is a non-linearity sharpness parameter, $c_{1/2}$ is a half-activation parameter, F_m is the maximum possible
431 fluorescence change. $n_e(t) \sim N(0, \sigma_e^2)$ is Gaussian external noise (Maravall et al., 2000; Tsien, 1989;
432 Yasuda et al., 2004).

433 We estimated the model parameters for each imaging condition using the simultaneous ephys and imaging
434 experiments (**Figure S3ABC**). We then applied the S2F model to ephys data using parameters randomly
435 sampled from the parameter distributions except for the parameters directly related to the nonlinearity.
436 Since ALM spike rates in ephys vary over a larger range than the spike rates in the primary visual cortex
437 these parameters may be underconstrained. Accordingly, we followed an alternative strategy to choose
438 these parameters for a given neuron. For each neuron, after assigning the rest of the parameters, we
439 transformed the spike trains to calculate the phenomenological calcium variable $c(t)$. We then estimated
440 the nonlinear parameters for that neuron by calculating the values that would best transform $c(t)$ to the
441 fluorescence dynamics of any neuron in the imaging dataset. For all neurons we were able to find matches
442 with Spearman correlation higher than 0.7 between mean dF/F and mean synthetic dF/F. The parameters
443 inferred in this process recapitulated the correlation structure of $c_{1/2}$ and k found in the data (**Figure S3D**).

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

447 To relate this model to previously studied models, **Equation 2** can be generalized as $\Delta F/F_{Synth}(t) =$
448 $f(c(t)) + n_e(t)$, where $f(\cdot)$ and $n_e(t) \sim N(0, \sigma_e^2)$ is Gaussian external noise (Maravall et al., 2000; Tsien,
449 1989; Yasuda et al., 2004). We considered two alternative S2F models used by previous studies (note
450 though that both of these models did not contain internal noise in **Equation 1**):

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

453 **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
454 the nonlinearity; K_d is a half-activation parameter. Model performance is summarized in the supplementary
455 material and reported in http://im-phys.org/analyses for each single cell (**Figure S3G,** right).

456 Model parameter sensitivity (**Figure S3C**) was defined as the decrease of the fraction of explained variance,
457 as a function of the deviation of the parameter value from the estimated solution: $g = \frac{\Delta EV/EV}{\Delta P/P}$, where $P \in$
458 $\{\tau_r, \tau_d, k, c_{1/2}, F_m\}$.

459 **Calcium imaging to spikes for non-simultaneous ephys-imaging recordings**

460 We performed fluorescence-to-spike (F2S) inference using two published models (Deneux et al., 2016;
461 Pnevmatikakis et al., 2013). These were two state-of-the-art generative models based on inference
462 techniques developed by (Deneux et al., 2016; Pnevmatikakis et al., 2013). We used the authors' code
463 available on GitHub.

464 **Single neuron analyses**

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

470   Selective neurons (mono- and multiphasic) were classified into left- and right-preferring cells according to
471   the condition in which their activity was higher (**Figure 2IJ**). Ramp-down (ramp-up) were defined as neurons
472   that have activity that is greater (less) in the baseline epoch compared to the delay epoch (paired t-test, p
473   < .05 across trials). Note that ramp-down cells were excluded from the analysis of peakiness (**Figure 7**).

Principal component analysis

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

$$r(s,t) = Cx(s,t) + <r>_{s,t}$$ (Equation 3)

478   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
479   points for each trial type. $<r>_{s,t}$ is a vector of the mean activity of each neuron across time and trial type.
480   $x(s,t)$ is an $n \times 2T$ PC score matrix, where the ith row corresponds to the ith PC score. We estimated the
481   relative contribution to each PC of the different forms of variance: temporal dynamics, trial-type selectivity
482   and other. Explained variance (EV) of temporal dynamics $EV_i(t)$ and trial-type selectivity $EV_i(s)$ for the ith
483   principal component (PC) were computed as:

$$EV_i(t) = <<x_i(s,t)>_s{}^2>_t / <x_i(s,t)^2>_{t,s}$$ (Equation 4)

$$EV_i(s) = <<x_i(s,t)>_t{}^2>_s / <x_i(s,t)^2>_{t,s}$$ (Equation 5)

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

Population decoding

488   We applied regularized linear discriminant analysis (LDA) on neural dynamics grouped into bins
489   corresponding to single imaging frames (67 ms) to compute the instantaneous decodability of trial type.
490   Regularization was performed by sparsity-regularized LDA (Guo et al., 2007; Wei et al., 2019). The optimal
491   LDA decoder was computed separately for each time bin using correct trials only. We estimated
492   performance for the instantaneous LDA decoder by sampling subsets of units and averaging 100
493   subsamples. We separated the trials of each neuron into non-overlapping training (70%) and testing (30%)
494   sets. The instantaneous decoder of trial type was computed from training set and its performance was
495   evaluated on the testing set.

496   We tested the ability of neuronal population activity at different times to discriminate the behavioral epoch
497   by using a four-class LDA (**Figure 6F-H**). We defined the latency of neuronal response to behavioral epoch
498   by the first time at which decoding reached a 0.7 accuracy threshold (arrows on **Figure 6F-H**).
499   Regularization was performed by sparsity-regularized LDA (Guo et al., 2007; Wei et al., 2019).

Sensitivity analysis of peakiness

501   We used as a reference value an artificial, synthetic ephys dataset with 50 neurons whose firing rates were
502   manually set to be non-zero only at the time corresponding to one imaging frame. From left to right in **Figure**
503   **7G**, S2F model was configured (1) using the same parameters for all cells, except that the internal noise
504   and external noise were randomly generated (at the same amplitudes); (2) using the same parameters for
505   all cells, except that the spike times were jittered within the time length of the frame (i.e., all spikes were
506   kept in the same image frame); (3) using the same parameters for all cells, except that the spike rates in
507   the original frame varied from 0.1 Hz to 5 Hz (spike trains generated using Poisson process); (4) using the
508   same parameters for all cells, except that the decay time constant of calcium indicator was randomly
509   sampled from its distribution; (5) using the same parameters for all cells, except that nonlinearity of calcium
510   indicator was randomly sampled from its distribution; (6) both decay time constant and nonlinearity of

511 calcium indicator were randomly sampled; (7) the same as (6) except that the spike rates in the original
512 frame vary from 0.1 Hz to 5 Hz (spike trains generated using Poisson process).

513 **Distributions of measures**

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

518 $$P(\psi) = \int P(\psi, \Delta F/F_{Synth}(t), \{t_{spike}\}, \Theta) \qquad \text{(Equation 6)}$$

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

520 $$P(\psi, \Delta F/F_{Synth}(t), \{t_{spike}\}, \Theta) = P\left(\psi\big|\Delta F/F_{Synth}(t)\right) P\left(\Delta F/F_{Synth}(t)|\{t_{spike}\}, \Theta\right) P(\{t_{spike}\}) P(\Theta) \quad \text{(Equation 7)}$$

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

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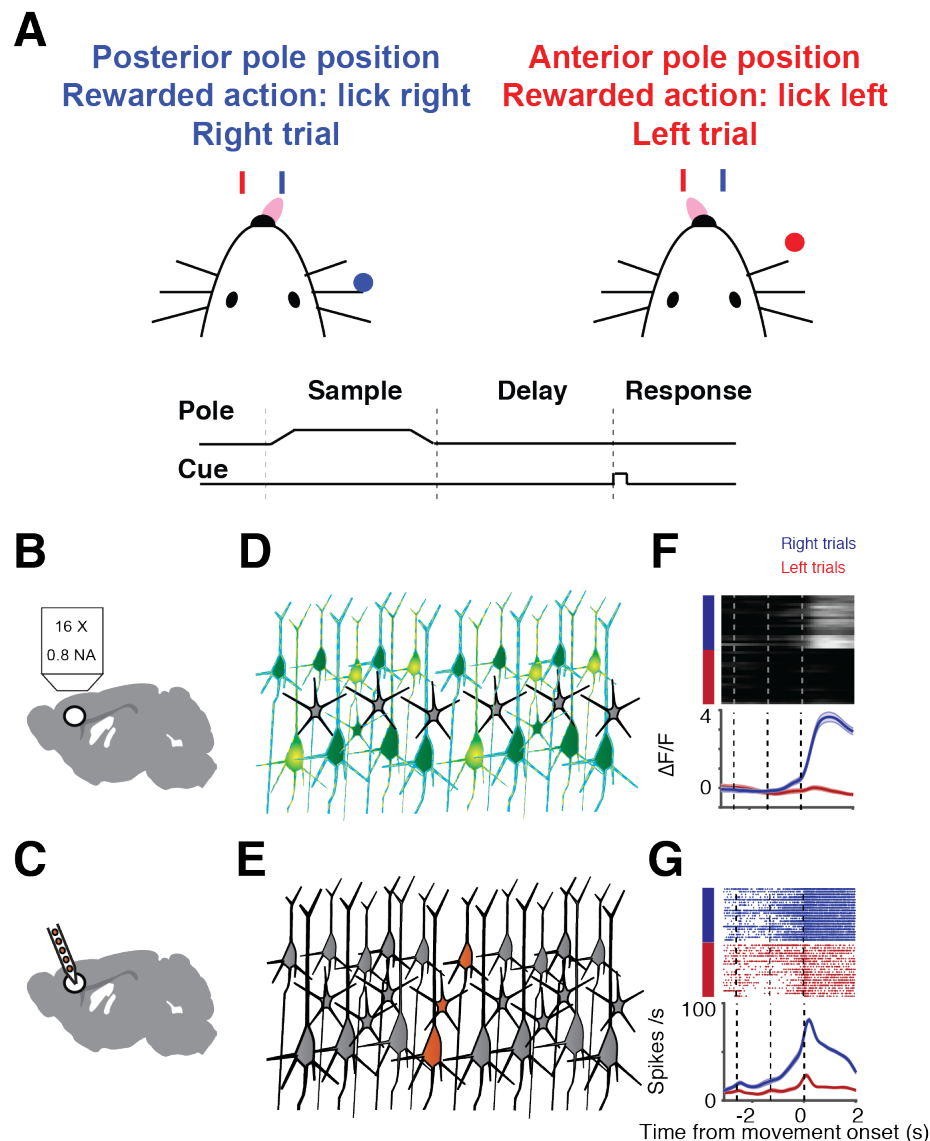525 of deconvolved synthetic ephys data to estimate distribution of the parameters.

526

Figure 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).
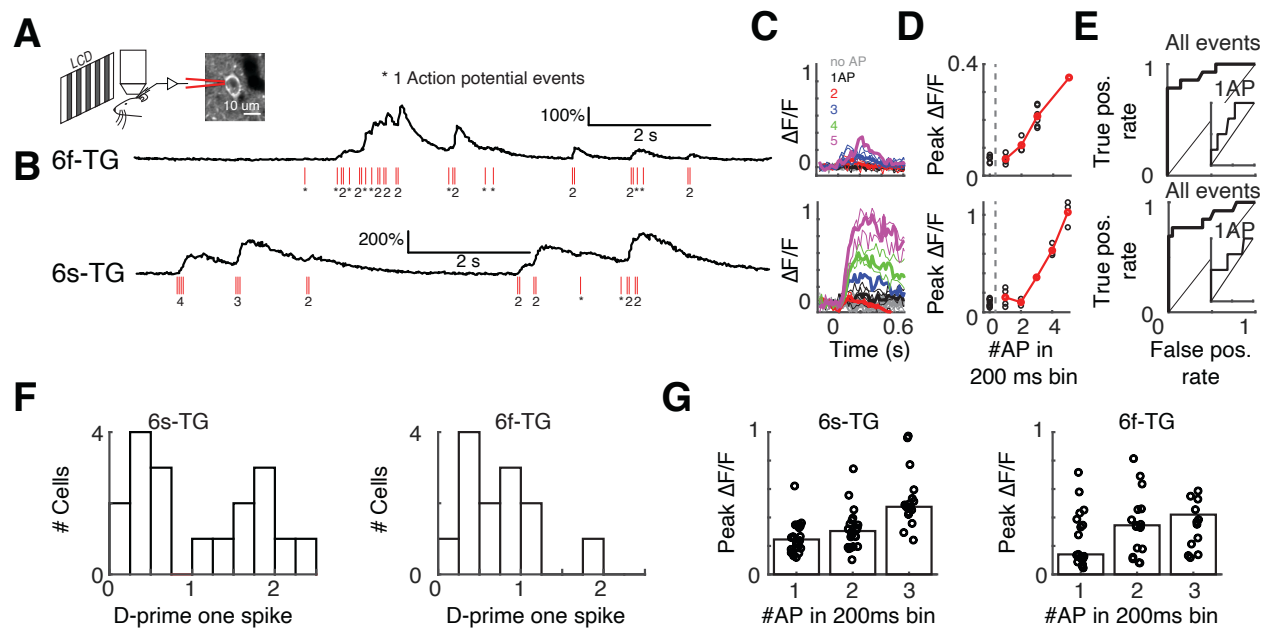
536

Figure 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 (Pnevmatikakis et al., 2013) are in cyan and for MLSpike (Deneux et al., 2016) 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 MLSpike). **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).

Figure 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, 3APs; green, 4APs; magenta, 5APs. 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.

567
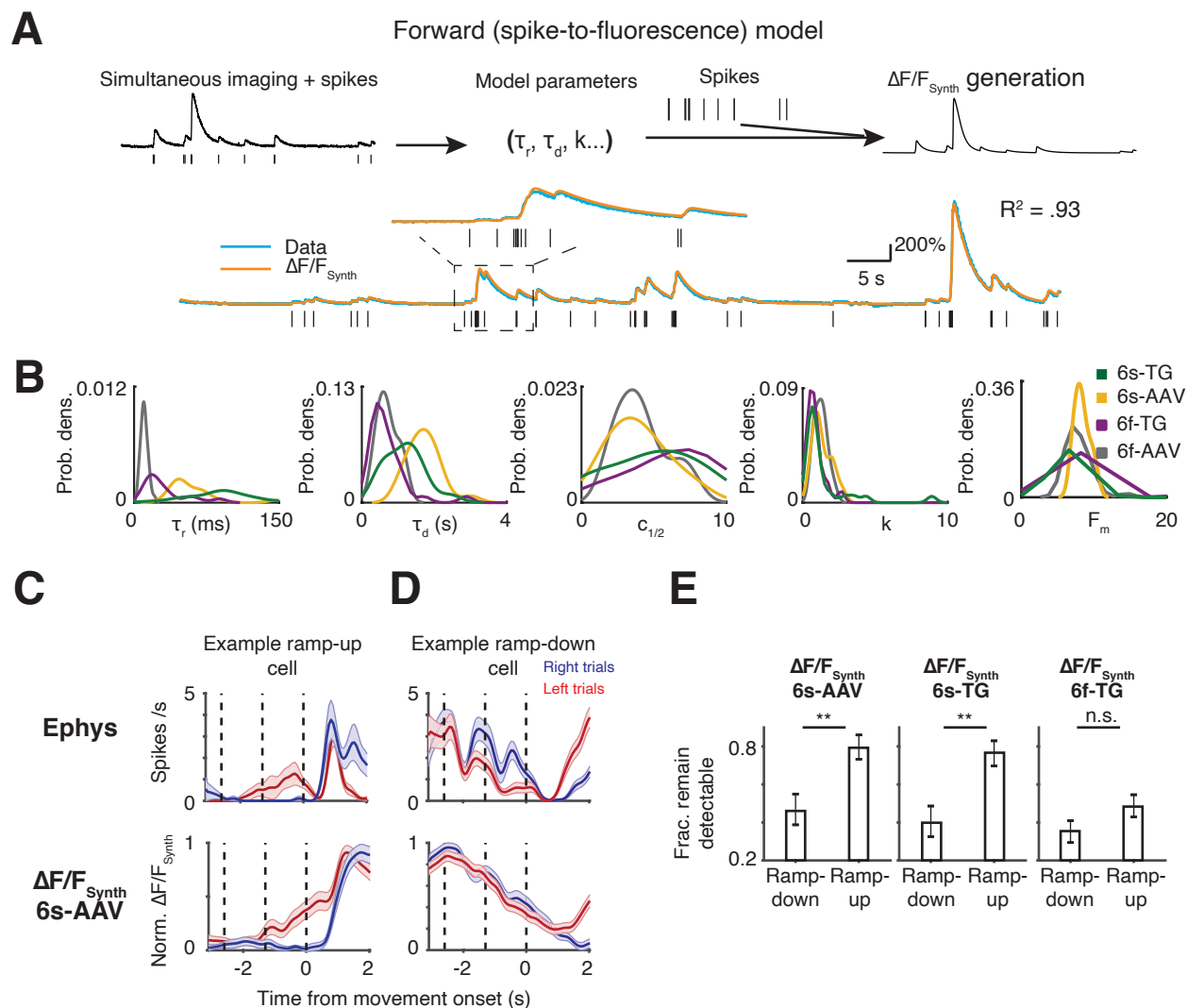
Figure 4. Forward modeling of the spike-to-fluorescence transformation largely explains difference in selectivity patterns

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. 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 cell (top, ephys; bottom, 6s-AAV synthetic); selectivity remains detectable in synthetic imaging data. **D.** An example ramp-down cell (top, ephys; bottom, 6s-AAV synthetic); 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.
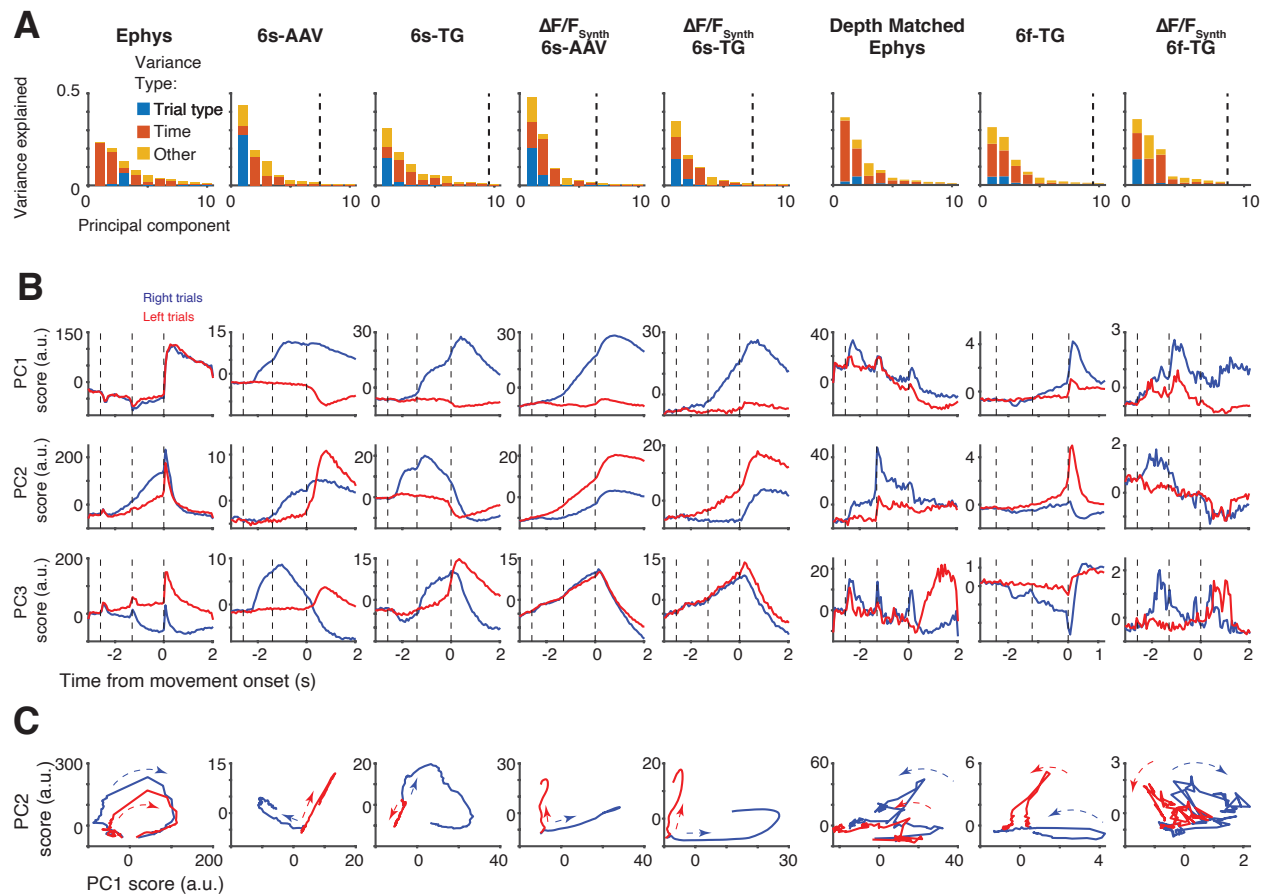
Figure 5. Different sources of variability extracted in dimensionality reduction on imaging and ephys

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.

591
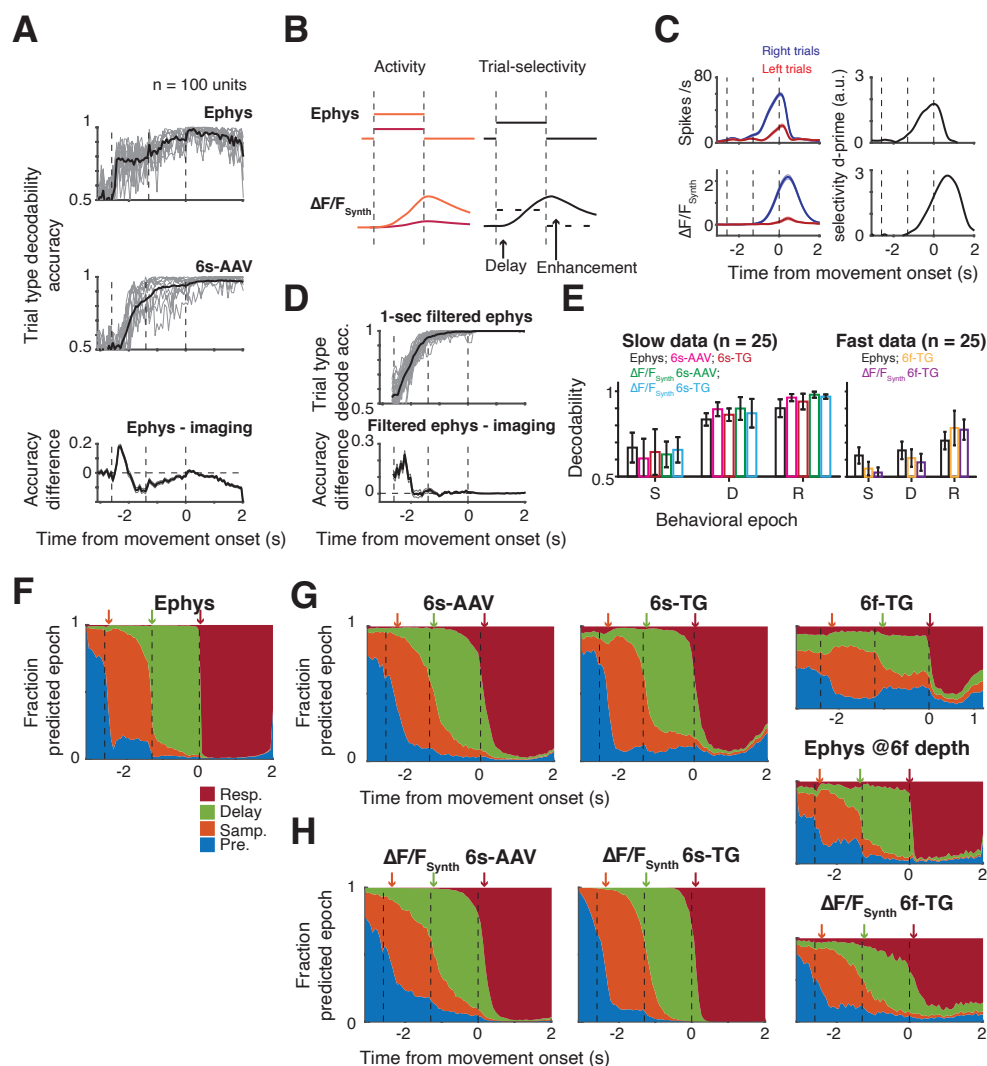
Figure 6. Population decoding differs in sensitivity and temporal profile between imaging and ephys

A. Performance of instantaneous regularized linear-discriminant-analysis (LDA) trail-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 think 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-G.** Performance of behavioral-epoch LDA decoders. **F.** 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. **G.** Same plot format as F, but for imaging. **H.** Sample plot format as F, but for synthetic imaging.
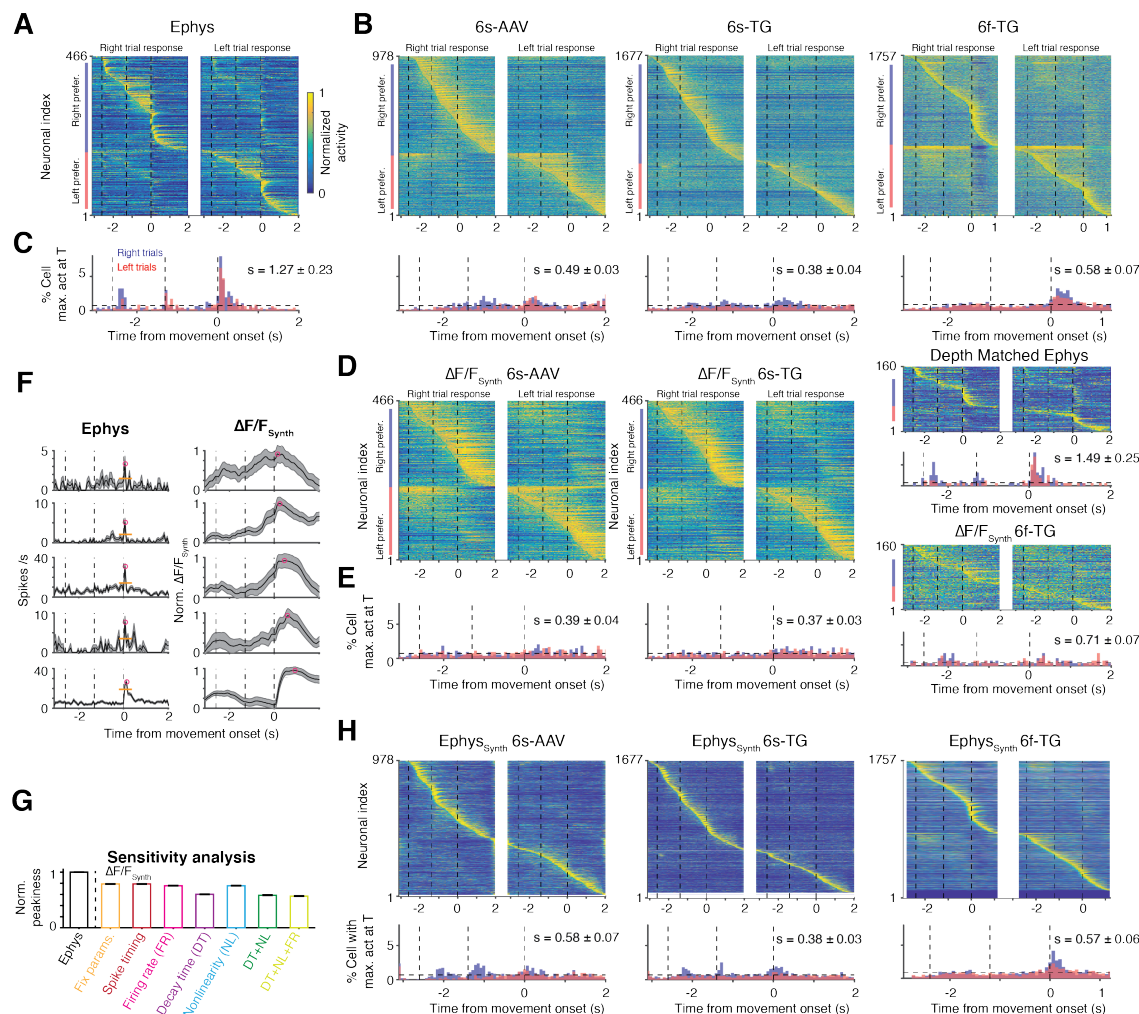
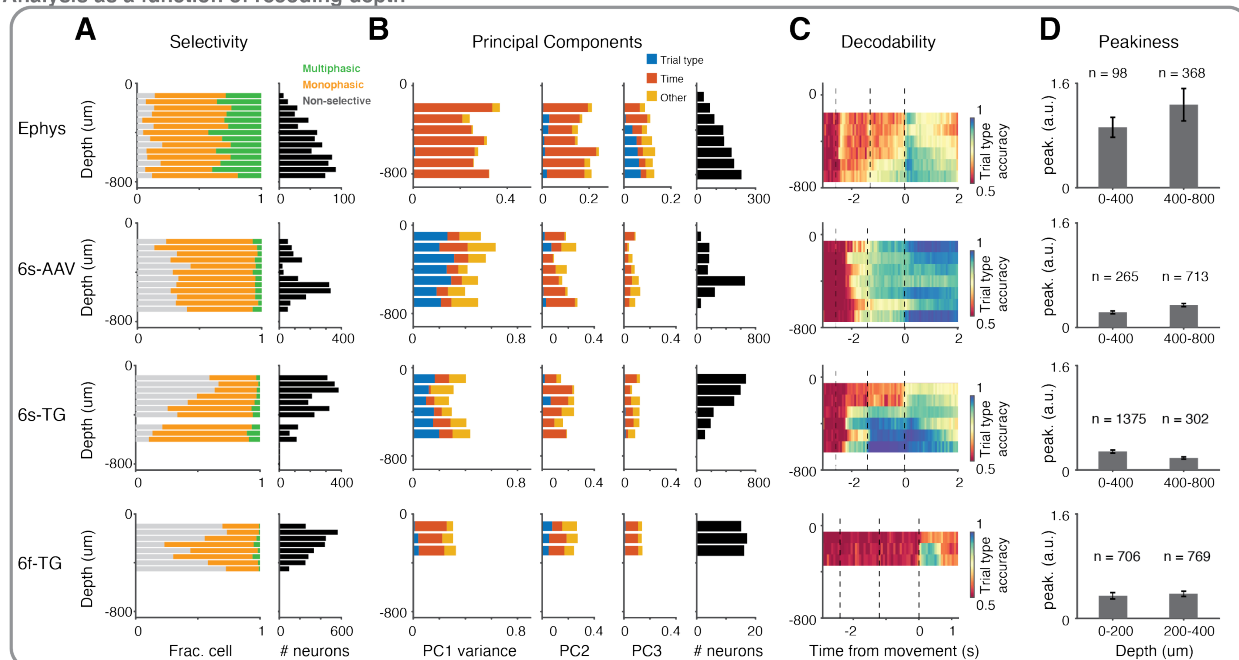**Figure 7. Temporal dispersion of population dynamics differs between imaging and ephys**

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.

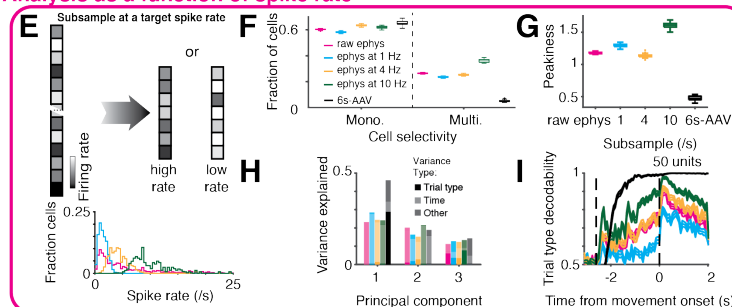632   **Supplementary materials**

633   **A comparison of neuronal population dynamics measured with calcium imaging and**
634   **electrophysiology**

635   Ziqiang Wei, Bei-Jung Lin, Tsai-Wen Chen, Kayvon Daie, Karel Svoboda, Shaul Druckmann
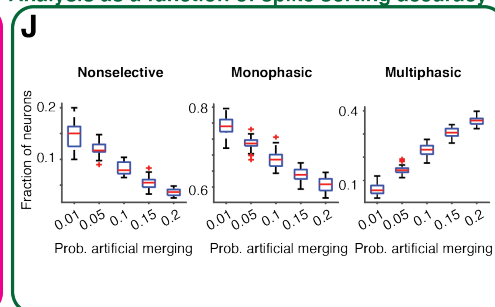
Figure S1. Effect of recording depth and firing rate in ephys

A-D. Analysis as a function of recoding depth. **A.** Single neuron selectivity-type analyses. Left: horizontal bar plots show breakdown of 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 roughly identical across depths (χ^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 neuron and more multiphasic selective neuron than imaging (χ^2-test, p < .001 for all). **B.** Percentage of variance of neural activity explained by each principal component (**Figure 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 1st PC show more temporal dynamics content in ephys and 6f-TG (χ^2-test, p < .001 for all), while that show more trial-type content in 6s-AAV and 6s-TG (χ^2-test, p < .001 for all). **C.** Decodability of trial type (**Figure 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 (**Figure 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

657   sum test, p < .001 for all, 1000 bootstrap). **E-I.** Analysis as a function of spike rates. **E.** Schematic of
658   resampling procedure to target firing rate and distribution of firing rates after subsampling to different
659   average spike rates (magenta: original data; cyan: ephys subsampled to 1 Hz average; yellow: ephys
660   subsampled to 4 Hz average; green: ephys subsampled to 10 Hz average. **F.** Effect of target firing rate
661   subsampling on fraction of monophasic (left) and multiphasic neurons (right) **G.** Values of peakiness are
662   shown with the same color code as F. **H.** Fraction of variance in the first principal components are shown
663   by length of bar with same color code as F. Saturation of bar shows the breakdown into different
664   components of variance (trial-type, time, other). **I.** Trial-type decodability over time shown with the same
665   color code as F and with 6s-AAV added as a reference. **J.** Analysis as a function of spike sorting accuracy
666   -- possible effects of merging. Increased fraction of multiphasic neurons is unlikely to have stemmed
667   exclusively from failures of spike-sorting. Box plots indicate fraction of neurons in each selectivity class (left:
668   non-selective, middle: monophasic, right: multiphasic) as a function of increased probability of artificially
669   induced merging between two neurons. Dashed line indicates fraction of selectivity type found in the ephys
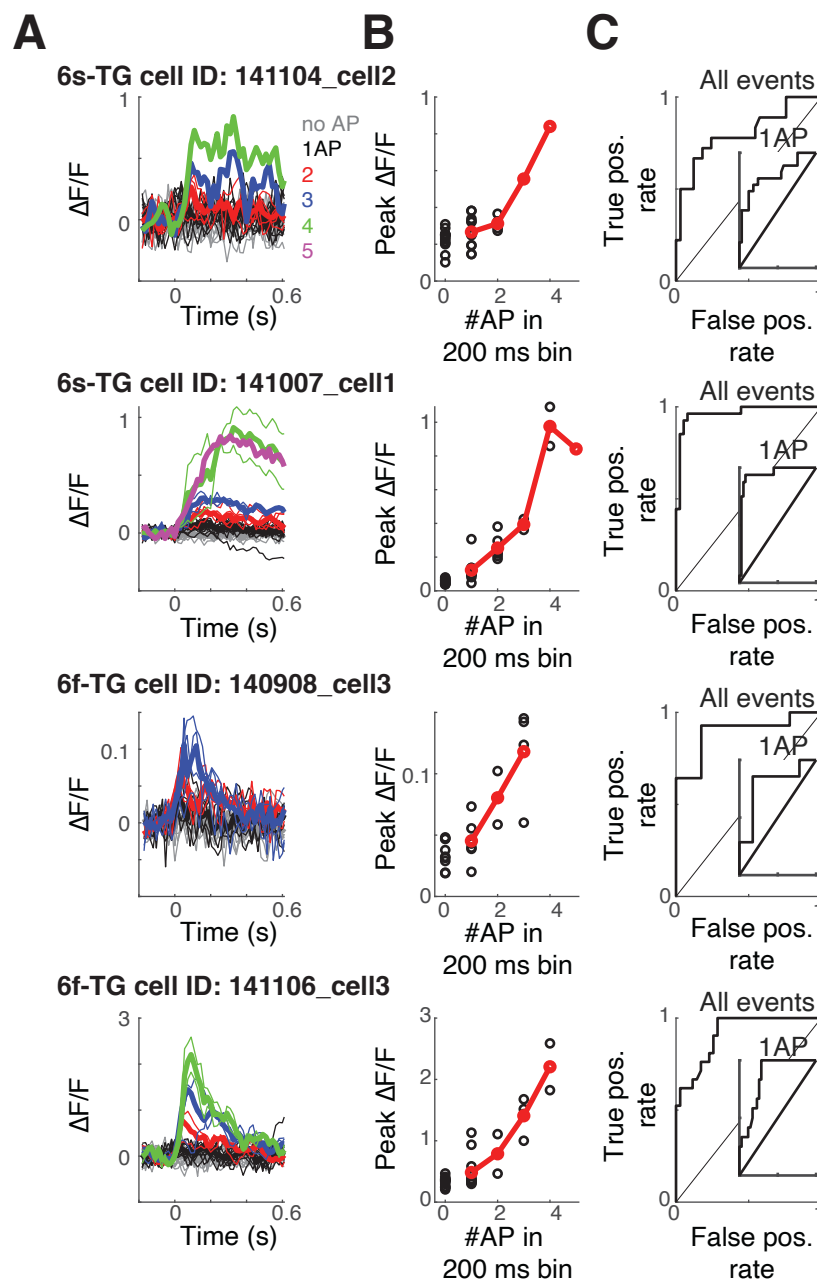670   dataset.

671

Figure S2. 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 Figure 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 as a function of the number of spikes (same plots as Figure 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 Figure 3E for additional examples).
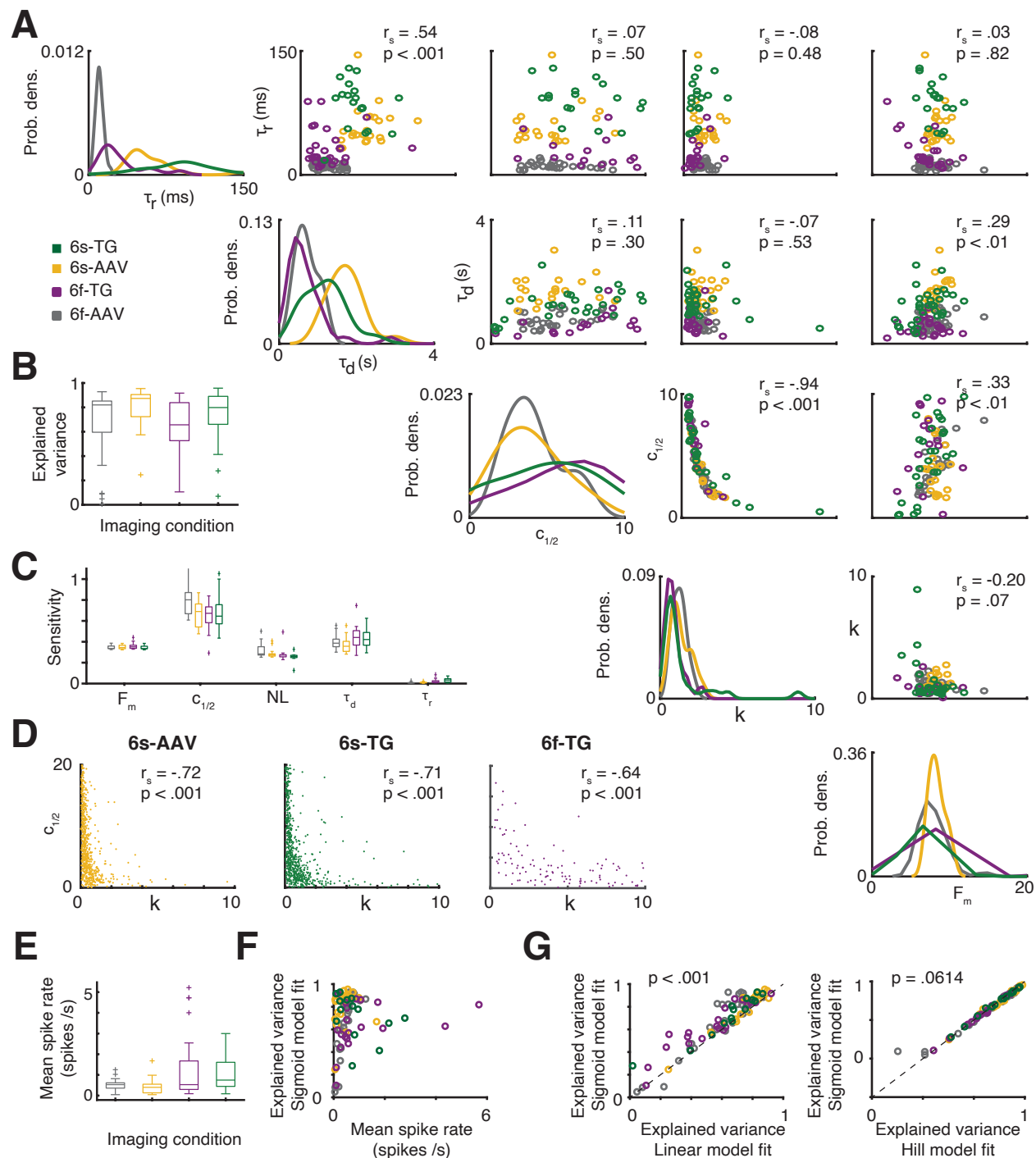
680   Figure S3. Detailed values of model parameters for simultaneously recorded neurons

681   **A.** Pairwise correlation plots for each of the spike-to-fluorescence parameters. Panels along the diagonal
682   describe the distribution of each parameter (these are identical to Figure 4B but reproduced to facilitate
683   comparisons). Off-diagonal panels depict the correlation between two parameters. Spearman's rank
684   correlation of parameters across cells (regardless of recording method) and associated p-value are
685   provided in each off-diagonal panel. Each circle corresponds to a response set. Data from the different
686   indicator conditions is overlaid and marked by color. (gray: 6f-AAV, 11 neurons, 37 response sets; yellow:

687   6s-AAV, 9 neurons, 21 response sets; purple: 6f-TG, 18 cells, 32 response sets; green: 6s-TG, 22 neurons,
688   33 recording periods). **B.** Boxplots of explained variance of S2F on validation data for simultaneously
689   recorded neurons (color follows the same convention as in A). **C.** Boxplot of distribution of parameter
690   sensitivity values. **D.** Pairwise correlation of re-estimation of k and $c_{1/2}$ using ALM imaging dynamics
691   (Materials and methods). The re-estimated parameter values are shown as a scatter plot. Each dot
692   corresponds to a neuron (n = 720 for 6s-AAV and 6s-TG; n = 225 for 6f-TG in matched depths). The
693   distribution of the re-estimated parameter values strongly overlapped with those obtained in simultaneous
694   imaging-ephys recordings. $c_{1/2}$ and k had a strong inverse correlation as in the simultaneously recorded
695   data ($r_s < -.64$, $p < .001$). **E.** Boxplots of firing rates of neurons in each recording sessions (6f-AAV, gray,
696   0.51 ± 0.25 Hz, mean $\pm$ std., range $0.05 - 1.25$ Hz; 6s-AAV, yellow, 0.43 ± 0.38 Hz, range $0.05 - 1.68$ Hz;
697   6f-TG, purple, 1.25 ± 1.48 Hz, range $0.09 - 5.22$ Hz; 6s-TG, green, 1.08 ± 0.85 Hz, range $0.09 - 3.00$ Hz).
698   **F**. Scatter of simultaneous ephys-imaging data model fit and the dynamical range of the data (expressed
699   as mean spike rate). **G.** Scatter of simultaneous ephys-imaging data model fit quality between different S2F
700   models (Materials and methods). Left: comparison between S2F linear model (x-axis) and S2F sigmoid
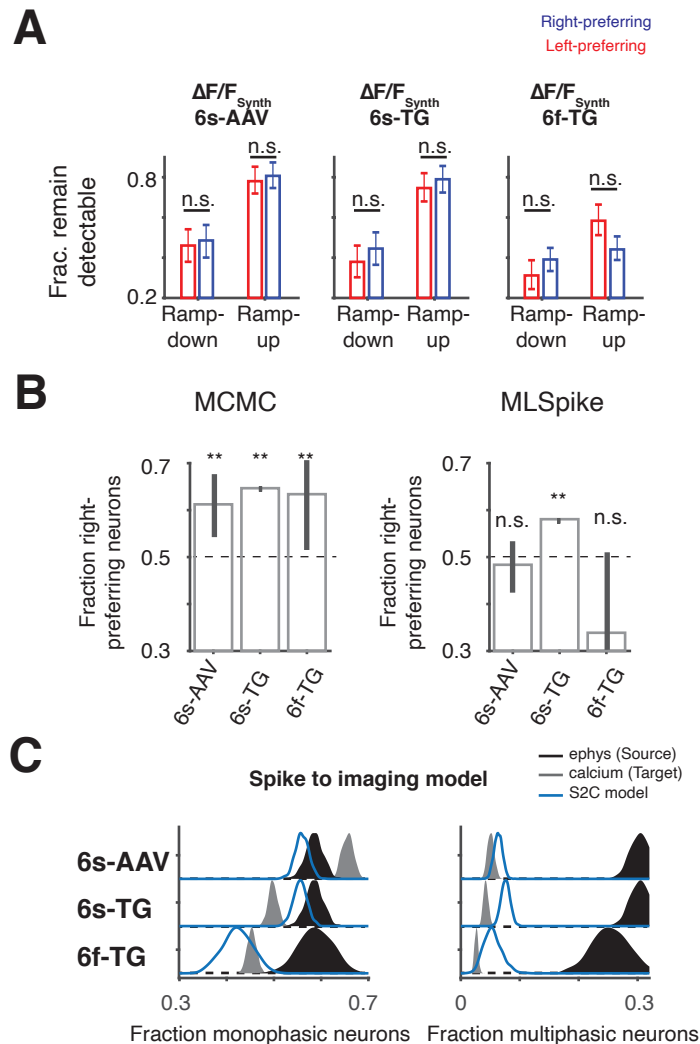701   model (y-axis); right, comparison between S2F hill model (x-axis) and S2F sigmoid model (y-axis).

Figure S4. 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 Figure 2I, but performed on inferred spiking data obtained via the MCMC framework; right: the same analyses as that in Figure 2I, but performed on inferred spiking data obtained via the MLSpike 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.
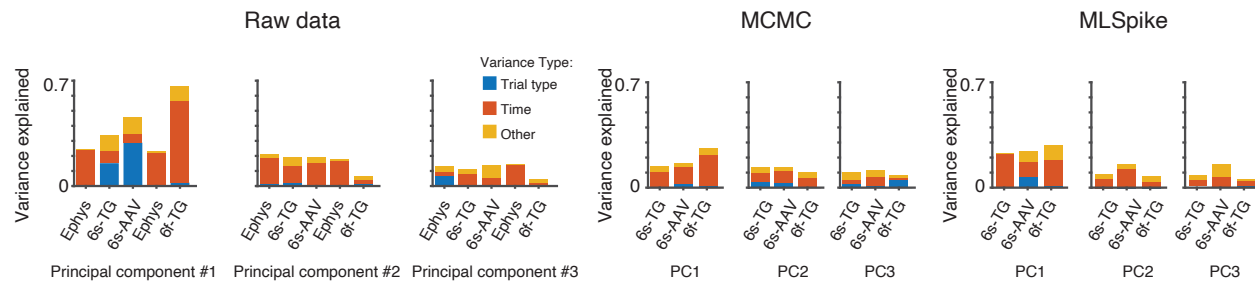
Figure S5. Spike inference results for principal component analysis

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 MLSpike framework.
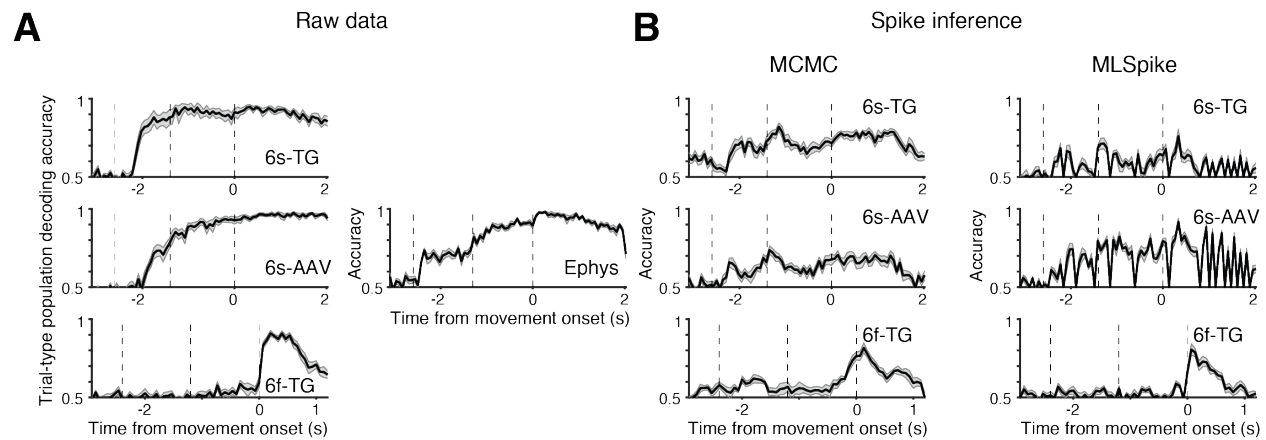
724

Figure S6. Spike inference results for trial-type population decoding

A. Accuracy of trial-type population decoding over time for different datasets. Left column top to bottom: 6s-TG, 6s-AAV, 6f-TG. Right column: ephys. **B.** 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: MLSpike framework.
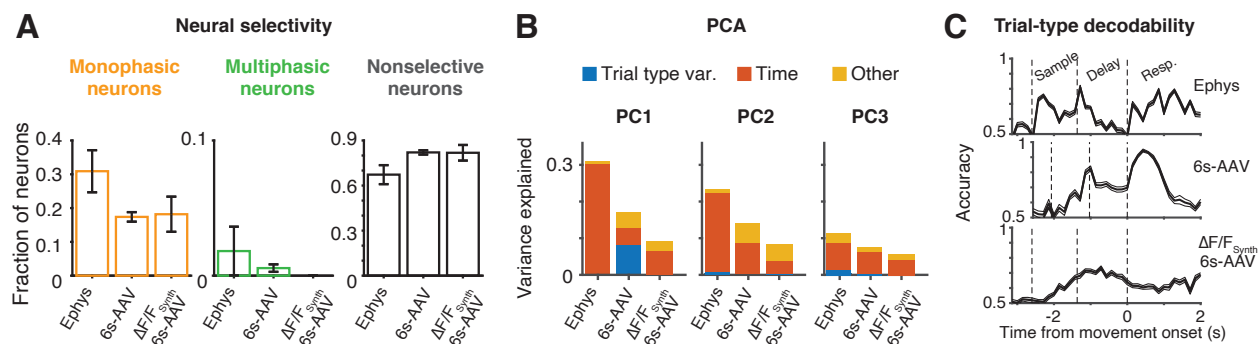
S1 dataset analyses

730

Figure S7. 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, χ^2 test) and there was no significant difference between the fraction of multiphasic neurons observed in ephys and imaging (p = .801, χ^2 test). Our forward model correctly predicted this lack of change (p = .674, χ^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.
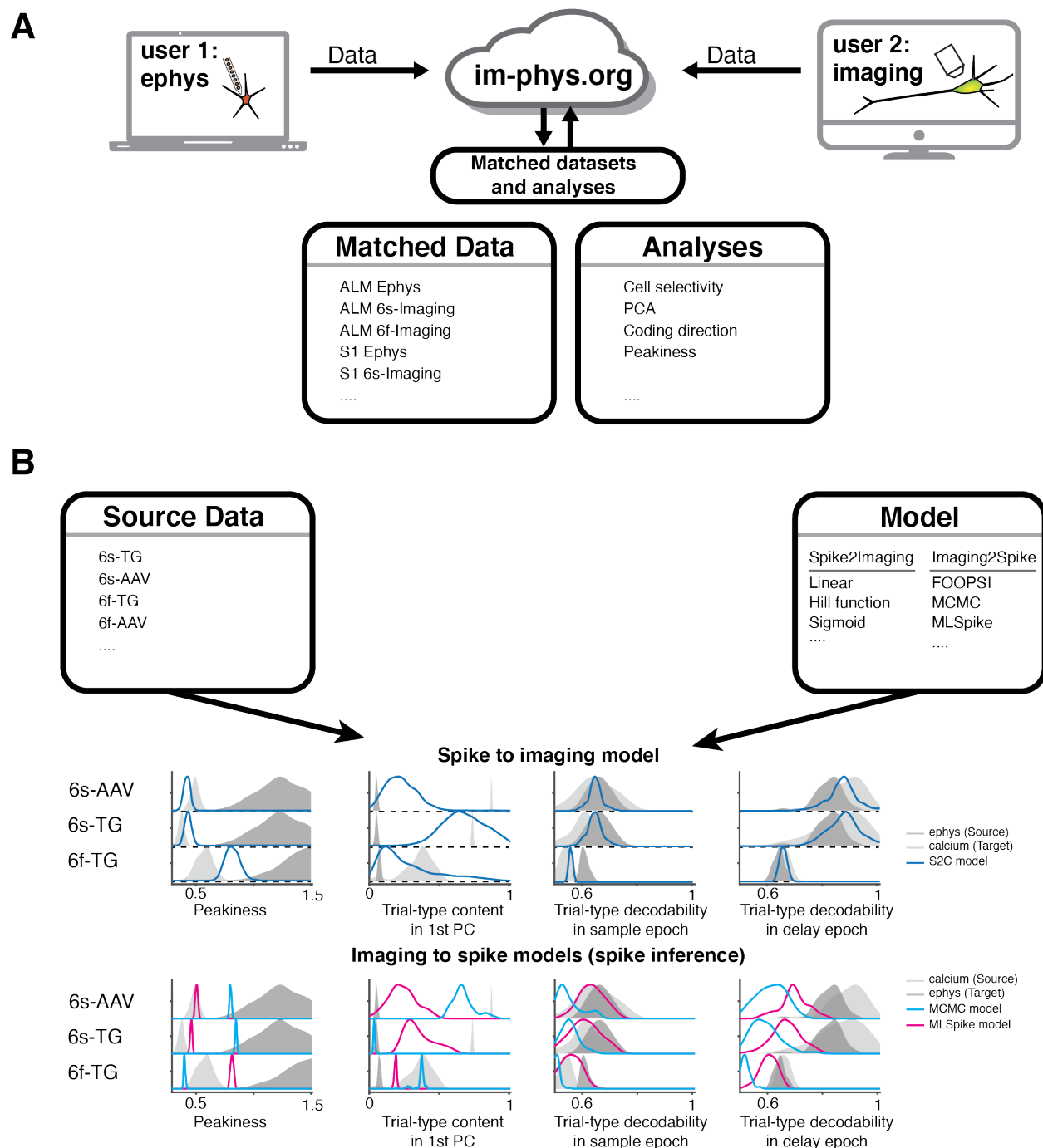
Figure S8. A community based online resource, im-phys.org, for determining quantitative effects of measuring population activity by imaging or ephys

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. **B.** Schematic of using 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.

Summary of recordings in GCaMP6s-TG mice

Cell ID	Recording duration (s)	Mean spike rate (Hz)	Mean spike rate during stimulus (4 s) (Hz)	Max spike rate in 100-ms bins (Hz)	Decay time (sec)
141001_cell1	292	0.26	4	40	2.55
141001_cell2	1067	0.92	9.5	100	1.13
141002_cell1	312	2.2	16.25	90	1.58
141002_cell2	156	0.75	3	20	1.37
141002_cell3	156	2.23	7.25	70	1.17
141002_cell4	412	1.76	9	60	2.02
141006_cell1	186	2.27	7.5	60	0.53
141006_cell2	132	0.4	4.75	90	0.40
141007_cell1	526	0.76	6.5	50	1.33
141007_cell3	1315	0.15	3.25	40	1.11
141007_cell4	443	3.39	10	40	0.55
141010_cell1	526	0.34	5	70	1.73
141010_cell2	165	0.29	2	50	1.56
141010_cell3	526	0.33	1.5	40	1.45
141010_cell5	611	3.19	13	60	0.79
141028_cell1	160	1.68	12.5	70	0.93
141029_cell1	363	2.67	6.25	50	0.56
141029_cell2	100	0.84	4.5	70	0.62
141029_cell4	526	1.89	8.5	60	1.90
141104_cell1	263	0.47	5.25	50	1.23
141104_cell2	766	2.31	14.25	70	0.90
141104_cell3	761	0.96	7.25	70	0.78
N=22	444±316	1.37±1.02	7.32±4.01	60±19	1.19±0.55

Summary of recordings in GCaMP6f-TG mice

Column1	Recording duration (s)	Mean spike rate (Hz)	Mean spike rate during stimulus (4 s) (Hz)	Max spike rate 100-ms bins (Hz)	Decay time (sec)
140822_cell1	240	0.29	3	100	0.39
140822_cell2	240	4.24	15	70	0.89
140826_cell1	100	0.7	3.25	40	0.23
140828_cell1	248	0.45	11.5	80	0.35
140904_cell2	744	0.59	8.5	220	0.41
140904_cell3	50	2.46	5	60	0.31
140908_cell1	744	2.19	11.75	70	0.81
140908_cell3	120	0.9	7.75	50	0.53
140911_cell1	403	1.11	5.5	40	0.69
141110_cell1	526	0.4	3	60	0.41
141110_cell2	789	0.95	8.25	50	0.85
141110_cell3	263	1.46	11.75	110	0.70
141113_cell1	263	1.89	12	70	0.91
141113_cell2	526	0.7	6	70	1.17
141113_cell3	263	1.76	8.5	50	0.39
141113_cell4	2630	0.54	8.5	70	1.04
141114_cell1	1152	4.35	18.75	80	0.67
141114_cell2	263	2.74	11.75	60	0.42
N=18	531±598	1.54±1.25	8.88±4.30	75±41	0.62±0.28

Table S1

Reference:

Abdelfattah, A.S., Kawashima, T., Singh, A., Novak, O., Liu, H., Shuai, Y., Huang, Y.-C., Campagnola, L., Seeman, S.C., Yu, J*., et al.* (2019). Bright and photostable chemigenetic indicators for extended in vivo voltage imaging. Science, eaav6416.

Adam, Y., Kim, J.J., Lou, S., Zhao, Y., Xie, M.E., Brinks, D., Wu, H., Mostajo-Radji, M.A., Kheifets, S., Parot, V*., et al.* (2019). Voltage imaging and optogenetics reveal behaviour-dependent changes in hippocampal dynamics. Nature *569*, 413-417.

Akerboom, J., Chen, T.W., Wardill, T.J., Tian, L., Marvin, J.S., Mutlu, S., Calderon, N.C., Esposti, F., Borghuis, B.G., Sun, X.R*., et al.* (2012). Optimization of a GCaMP Calcium Indicator for Neural Activity Imaging. The Journal of neuroscience : the official journal of the Society for Neuroscience *32*, 13819-13840.

Berens, P., Freeman, J., Deneux, T., Chenkov, N., McColgan, T., Speiser, A., Macke, J.H., Turaga, S.C., Mineault, P., Rupprecht, P*., et al.* (2018). Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. PLOS Computational Biology *14*, e1006157.

Bock, D.D., Lee, W.C., Kerlin, A.M., Andermann, M.L., Hood, G., Wetzel, A.W., Yurgenson, S., Soucy, E.R., Kim, H.S., and Reid, R.C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. Nature *471*, 177-182.

Briggman, K.L., Helmstaedter, M., and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. Nature *471*, 183-188.

Brody, C.D., Romo, R., and Kepecs, A. (2003). Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. Curr Opin Neurobiol *13*, 204-211.

Buzsaki, G. (2004). Large-scale recording of neuronal ensembles. Nat Neurosci *7*, 446-451.

Chen, T.-W., Li, N., Daie, K., and Svoboda, K. (2017). A Map of Anticipatory Activity in Mouse Motor Cortex. Neuron *94*, 866-879.e864.

Chen, T.W., Wardill, T.J., Sun, Y., Pulver, S.R., Renninger, S.L., Baohan, A., Schreiter, E.R., Kerr, R.A., Orger, M.B., Jayaraman, V*., et al.* (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature *499*, 295-300.

Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nature neuroscience.

Dana, H., Chen, T.-W., Hu, A., Shields, B.C., Guo, C., Looger, L., Kim, D., and Svoboda, K. (2014). Thy1-GCaMP6 Transgenic Mice for Neuronal Population Imaging In Vivo. PloS ONE *9*, e108697.

Dana, H., Mohar, B., Sun, Y., Narayan, S., Gordus, A., Hasseman, J.P., Tsegaye, G., Holt, G.T., Hu, A., Walpita, D*., et al.* (2016). Sensitive red protein calcium indicators for imaging neural activity. Elife *5*.

Dana, H., Sun, Y., Mohar, B., Hulse, B., Hasseman, J.P., Tsegaye, G., Tsang, A., Wong, A., Patel, R., Macklin, J.J*., et al.* (2018). High-performance GFP-based calcium indicators for imaging activity in neuronal populations and microcompartments. bioRxiv, 434589.

795   Deneux, T., Kaszas, A., Szalay, G., Katona, G., Lakner, T., Grinvald, A., Rózsa, B., and Vanzetta, I.
796   (2016). Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of
797   large neuronal populations in vivo. Nature Communications 7, 12190.

798   Dhawale, A.K., Poddar, R., Wolff, S.B.E., Normand, V.A., Kopelowitz, E., and Ölveczky, B.P. (2017).
799   Automated long-term recording and analysis of neural activity in behaving animals. eLife 6, e27702.

800   Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., and Stryker,
801   M.P. (2014). A cortical circuit for gain control by behavioral state. Cell 156, 1139-1152.

802   Ganguly, K., and Carmena, J.M. (2009). Emergence of a stable cortical map for neuroprosthetic control.
803   PLoS biology 7, e1000153.

804   Greenberg, D.S., Wallace, D.J., Voit, K.-M., Wuertenberger, S., Czubayko, U., Monsees, A., Handa, T.,
805   Vogelstein, J.T., Seifert, R., Groemping, Y., et al. (2018). Accurate action potential inference from a
806   calcium sensor protein through biophysical modeling. bioRxiv, 479055.

807   Grienberger, C., and Konnerth, A. (2012). Imaging calcium in neurons. Neuron 73, 862-885.

808   Guo, Y., Hastie, T., and Tibshirani, R. (2007). Regularized linear discriminant analysis and its application
809   in microarrays. Biostatistics 8, 86-100.

810   Guo, Z.V., Hires, S.A., Li, N., O'Connor, D.H., Komiyama, T., Ophir, E., Huber, D., Bonardi, C., Morandell,
811   K., Gutnisky, D., et al. (2014a). Procedures for behavioral experiments in head-fixed mice. PloS one 9,
812   e88678.

813   Guo, Z.V., Inagaki, H.K., Daie, K., Druckmann, S., Gerfen, C.R., and Svoboda, K. (2017). Maintenance of
814   persistent activity in a frontal thalamocortical loop. Nature 545, 181.

815   Guo, Z.V., Li, N., Huber, D., Ophir, E., Gutnisky , D.A., Ting, J.T., Feng, G., and Svoboda, K. (2014b).
816   Flow of cortical activity underlying a tactile decision in mice. Neuron 81, 179-194.

817   Hamel, E.J., Grewe, B.F., Parker, J.G., and Schnitzer, M.J. (2015). Cellular level brain imaging in
818   behaving mammals: an engineering approach. Neuron 86, 140-159.

819   Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a
820   virtual-navigation decision task. Nature 484, 62-68.

821   Hromadka, T., Deweese, M.R., and Zador, A.M. (2008). Sparse representation of sounds in the
822   unanesthetized auditory cortex. PLoS Biol 6, e16.

823   Huber, D., Gutnisky, D.A., Peron, S., O'Connor, D.H., Wiegert, J.S., Tian, L., Oertner, T.G., Looger, L.L.,
824   and Svoboda, K. (2012). Multiple dynamic representations in the motor cortex during sensorimotor
825   learning. Nature 484, 473-478.

826   Inagaki, H.K., Inagaki, M., Romani, S., and Svoboda, K. (2018). Low-Dimensional and Monotonic
827   Preparatory Activity in Mouse Anterior Lateral Motor Cortex. The Journal of Neuroscience 38, 4163.

828   Inoue, M., Takeuchi, A., Horigane, S., Ohkura, M., Gengyo-Ando, K., Fujii, H., Kamijo, S., Takemoto-
829   Kimura, S., Kano, M., Nakai, J., et al. (2015). Rational design of a high-affinity, fast, red calcium indicator
830   R-CaMP2. Nature methods 12, 64-70.

831   Inoue, M., Takeuchi, A., Manita, S., Horigane, S.-i., Sakamoto, M., Kawakami, R., Yamaguchi, K., Otomo,
832   K., Yokoyama, H., Kim, R., *et al.* (2019). Rational Engineering of XCaMPs, a Multicolor GECI Suite for
833   In Vivo Imaging of Complex Brain Circuit Dynamics. Cell *177*, 1346-1360.e1324.

834   Jia, H., Rochefort, N.L., Chen, X., and Konnerth, A. (2010). Dendritic organization of sensory input to
835   cortical neurons in vivo. Nature *464*, 1307.

836   Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou,
837   C.A., Andrei, A., Aydın, Ç., *et al.* (2017). Fully integrated silicon probes for high-density recording of
838   neural activity. Nature *551*, 232.

839   Kerlin, A.M., Andermann, M.L., Berezovskii, V.K., and Reid, R.C. (2010). Broadly tuned response
840   properties of diverse inhibitory neuron subtypes in mouse visual cortex. Neuron *67*, 858-871.

841   Komiyama, T., Sato, T.R., O'Connor, D.H., Zhang, Y.X., Huber, D., Hooks, B.M., Gabitto, M., and
842   Svoboda, K. (2010). Learning-related fine-scale specificity imaged in motor cortex circuits of behaving
843   mice. Nature *464*, 1182-1186.

844   Li, N., Chen, T.W., Guo, Z.V., Gerfen, C.R., and Svoboda, K. (2015). A motor cortex circuit for motor
845   planning and movement. Nature *519*, 51-56.

846   Li, N., Daie, K., Svoboda, K., and Druckmann, S. (2016). Robust neuronal dynamics in premotor cortex
847   during motor planning. Nature.

848   Lin, B.-J., Chen, T.-W., Kim, D.S., and Svoboda, K. (2016). Simultaneous calcium imaging using GCaMP
849   sensors and electrophysiology in L2/3 pyramidal neurons of the visual cortex in thy1 transgenic mice.

850   Lovett-Barron, M., Andalman, A.S., Allen, W.E., Vesuna, S., Kauvar, I., Burns, V.M., and Deisseroth, K.
851   (2017). Ancestral Circuits for the Coordinated Modulation of Brain State. Cell *171*, 1411-1423.e1417.

852   Lütcke, H., Gerhard, F., Zenke, F., Gerstner, W., and Helmchen, F. (2013). Inference of neuronal network
853   spike dynamics and topology from calcium imaging data. Frontiers in Neural Circuits *7*, 201.

854   Maravall, M., Mainen, Z.M., Sabatini, B.L., and Svoboda, K. (2000). Estimating intracellular calcium
855   concentrations and buffering without wavelength ratioing. Biophys J *78*, 2655-2667.

856   O'Connor, D.H., Peron, S.P., Huber, D., and Svoboda, K. (2010). Neural activity in barrel cortex
857   underlying vibrissa-based object localization in mice. Neuron *67*, 1048-1061.

858   Ohki, K., Chung, S., Ch'ng, Y.H., Kara, P., and Reid, R.C. (2005). Functional imaging with cellular
859   resolution reveals precise micro-architecture in visual cortex. Nature *433*, 597-603.

860   Ohkura, M., Sasaki, T., Sadakari, J., Gengyo-Ando, K., Kagawa-Nagamura, Y., Kobayashi, C., Ikegaya,
861   Y., and Nakai, J. (2012). Genetically Encoded Green Fluorescent Ca2+ Indicators with Improved
862   Detectability for Neuronal Ca2+ Signals. PLoS ONE *7*, e51286.

863   Peron, S., Chen, T.W., and Svoboda, K. (2015a). Comprehensive imaging of cortical networks. Curr Opin
864   Neurobiol *32*, 115-123.

865   Peron, S.P., Freeman, J., Iyer, V., Guo, C., and Svoboda, K. (2015b). A Cellular Resolution Map of Barrel
866   Cortex Activity during Tactile Behavior. Neuron *86*, 783-799.

867   Peters, A.J., Chen, S.X., and Komiyama, T. (2014). Emergence of reproducible spatiotemporal activity
868   during motor learning. Nature *510*, 263.

869   Petreanu, L., Gutnisky, D.A., Huber, D., Xu, N.L., O'Connor, D.H., Tian, L., Looger, L., and Svoboda, K.
870   (2012). Activity in motor-sensory projections reveals distributed coding in somatosensation. Nature *489*,
871   299-303.

872   Pnevmatikakis, E.A., Merel, J., Pakman, A., and Paninski, L. (2013). Bayesian spike inference from
873   calcium imaging data. Paper presented at: Signals, Systems and Computers, 2013 Asilomar Conference
874   on (IEEE).

875   Pnevmatikakis, E.A., Soudry, D., Gao, Y., Machado, T.A., Merel, J., Pfau, D., Reardon, T., Mu, Y.,
876   Lacefield, C., Yang, W*., et al.* (2016). Simultaneous Denoising, Deconvolution, and Demixing of Calcium
877   Imaging Data. Neuron *89*, 285-299.

878   Pologruto, T.A., Yasuda, R., and Svoboda, K. (2004). Monitoring neural activity and [Ca2+] with
879   genetically encoded Ca2+ indicators. J Neurosci *24*, 9572-9579.

880   Scheuss, V., Yasuda, R., Sobczyk, A., and Svoboda, K. (2006). Nonlinear [Ca2+] signaling in dendrites
881   and spines caused by activity-dependent depression of Ca2+ extrusion. J Neurosci *26*, 8183-8194.

882   Scott, B.B., Constantinople, C.M., Akrami, A., Hanks, T.D., Brody, C.D., and Tank, D.W. (2017). Fronto-
883   parietal Cortical Circuits Encode Accumulated Evidence with a Diversity of Timescales. Neuron *95*, 385-
884   398.e385.

885   Sofroniew, N.J., Flickinger, D., King, J., and Svoboda, K. (2016). A large field of view two-photon
886   mesoscope with subcellular resolution for in vivo imaging. Elife *5*.

887   Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K.D. (2019). High-dimensional
888   geometry of population responses in visual cortex. Nature *571*, 361-365.

889   Theis, L., Berens, P., Froudarakis, E., Reimer, J., Roman Roson, M., Baden, T., Euler, T., Tolias, A.S.,
890   and Bethge, M. (2016). Benchmarking Spike Rate Inference in Population Calcium Imaging. Neuron *90*,
891   471-482.

892   Tian, L., Hires, S.A., Mao, T., Huber, D., Chiappe, M.E., Chalasani, S.H., Petreanu, L., Akerboom, J.,
893   McKinney, S.A., Schreiter, E.R*., et al.* (2009). Imaging neural activity in worms, flies and mice with
894   improved GCaMP calcium indicators. Nat Methods *6*, 875-881.

895   Tolias, A.S., Ecker, A.S., Siapas, A.G., Hoenselaar, A., Keliris, G.A., and Logothetis, N.K. (2007).
896   Recording chronically from the same neurons in awake, behaving primates. J Neurophysiol *98*, 3780-
897   3790.

898   Tsien, R.Y. (1989). Fluorescent probes of cell signaling. Annu Rev Neurosci *12*, 227-253.

899   Vogelstein, J.T., Packer, A.M., Machado, T.A., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010).
900   Fast nonnegative deconvolution for spike train inference from population calcium imaging. Journal of
901   neurophysiology *104*, 3691-3704.

902   Vogelstein, J.T., Watson, B.O., Packer, A.M., Yuste, R., Jedynak, B., and Paninski, L. (2009). Spike
903   inference from calcium imaging using sequential Monte Carlo methods. Biophysical journal *97*, 636-655.

904   Wei, Z., Inagaki, H., Li, N., Svoboda, K., and Druckmann, S. (2019). An orderly single-trial organization of
905   population dynamics in premotor cortex predicts behavioral variability. Nature Communications *10*, 216.

906   Xu, N.L., Harnett, M.T., Williams, S.R., Huber, D., O'Connor, D.H., Svoboda, K., and Magee, J.C. (2012).
907   Nonlinear dendritic integration of sensory and motor input during an active sensing task. Nature *492*, 247-
908   251.

909   Yasuda, R., Nimchinsky, E.A., Scheuss, V., Pologruto, T.A., Oertner, T.G., Sabatini, B.L., and Svoboda,
910   K. (2004). Imaging calcium concentration dynamics in small neuronal compartments. Sci STKE *2004*, pl5.

911