Precise calcium-to-spike inference using biophysical generative models

Gerard Joey Broussard[1], Giovanni Diana[2], Francisco J. Urra Quiroz[2], B. Semihcan Sermet[2], Nelson Rebola[3], Laura A. Lynch[1], David A. DiGregorio[2,3,*], Samuel S.-H. Wang[1,*]

[1]Neuroscience Institute, Washington Road, Princeton University, Princeton, New Jersey USA 08540.
[2]Institut Pasteur, Université Paris Cité, CNRS UMR 3571, Synapse and Circuit Dynamics Laboratory, Paris, France.
[3]L'Institut du Cerveau et de la Moelle épinière, CNRS UMR 7225, Paris, France
[4]Department of Physiology and Biophysics, University of Colorado School of Medicine, Aurora, Colorado USA 80045.

*co-senior authors

Correspondence: sswang@princeton.edu; david.digregorio@cuasnchutz.edu

Acknowledgements: The authors thank Marissa Applegate, Kiri Couchman, and Minglu Wang for collaboration and technical assistance in early stages of this project, and Yan Zhang and Loren Looger for discussion and for providing a jGCaMP8f construct prior to publication. S.W. was supported by NIH R01 NS045193 and U19 NS104648. S.W. and D.D. were supported by NIH R21 EY026434. G.J.B. was supported by the Warren Alpert Foundation and the Brain & Behavior Research Foundation. The laboratory of D.D. was also supported by the Fondation pour la Recherche Médicale (FRM EQU202003010555), Fondation pour l'Audition (FPA-RD-2018-8), and the Agence Nationale de la Recherche (ANR-17-CE16-0019, and ANR-18-CE16-0018, ANR-19-CE16 0019-02, ANR-21-CE16-0036-01).

Abstract:

The intramolecular dynamics of fluorescent indicators of neural activity can distort the accurate estimate of action potential ("spike") times. In order to develop a more accurate spike inference algorithm we characterized the kinetic responses to calcium of three popular indicator proteins, GCaMP6f, jGCaMP7f, and jGCaMP8f, using in vitro stopped-flow and brain slice recordings. jGCaMP8f showed a use-dependent slowing of fluorescence responses that caused existing inference methods to generate numerous false positives. From these data we developed a multistate model of GCaMP and used it to create Bayesian Sequential Monte Carlo ($Biophys_{SMC}$) and machine learning ($Biophys_{ML}$) inference methods that reduced false positives substantially. This biophysical method dramatically improved spike time accuracy, detecting individual spikes with a median uncertainty of 4 milliseconds, a performance level that reached the theoretical limit and is twice as accurate as any previous method. Our framework thus highlights advantages of physical model-based approaches over model-free algorithms.

Introduction

Genetically encoded calcium indicators (GECIs), as exemplified by the GCaMP family of sensors (Tian et al., 2009; Akerboom et al., 2012; Chen et al., 2013; Dana et al., 2019; Zhang et al., 2023), have transformed the study of neural populations by enabling simultaneous recordings from large ensembles of neurons (Demas et al., 2021). These sensors convert rapid changes in intracellular calcium concentration, which accompany neuronal action potentials, into measurable fluorescence signals. However, the temporal resolution of GECIs is limited by their slower kinetics compared to the rapid voltage changes associated with neuronal spikes. This mismatch limits the temporal precision with which the neural dynamics underlying behavior can be accurately estimated (Xiao et al., 2017; Li et al., 2019; Wei et al., 2020).

A variety of computational strategies have been developed to address the challenge of inferring spikes from fluorescence signals. These methods can be broadly categorized as either supervised or unsupervised. Supervised approaches leverage training data where both fluorescence and spike times are known to train machine learning models to recognize statistical regularities mapping observed signals to underlying spiking activity (Rupprecht et al., 2021; Zhou et al., 2023). Unsupervised approaches, by contrast, do not require labeled data and instead use principled assumptions about how spikes generate fluorescence based on the statistical properties of fluorescence time series to infer the underlying “latent” spike trains (e.g. Vogelstein et al., 2010; Pnevmatikakis et al., 2013; Deneux et al., 2016; Friedrich et al., 2017; Sebastian et al., 2017; Pachitariu et al., 2018). The performance of both categories is enhanced if the relationship between spike occurrences and the resultant fluorescence responses is fundamentally linear. When this linearity assumption holds, unsupervised methods can utilize simple, computationally efficient models (in particular linear deconvolution; Friedrich et al., 2017), while supervised methods benefit from stable, direct mappings that allow for robust training and generalization.

However, the spike-to-fluorescence relationship often deviates from the linear ideal. Classic examples include saturation of indicator fluorescence at high calcium concentrations (Regehr and Atluri, 1995) and slowing of calcium dynamics due to calcium buffering by high-affinity indicators (Sabatini and Regehr, 1998). Thirdly, GECI responses to free calcium changes can be supralinear (Rose et al., 2014). Such nonlinearities pose potential challenges to both unsupervised and supervised methods. Popular unsupervised methods relying on low-complexity models such as linear convolution or generative frameworks reliant upon a static decay process would be expected to exhibit systematic errors dictated by the nature of the nonlinearity. Supervised methods, meanwhile, may fail to capture the complexity of novel, nonlinear regimes that differ substantially from the conditions under which they were trained (Seroussi and Zeitouni, 2022; Sariyildiz et al., 2023). Consequently, nonlinearities in the spike-to-fluorescence transformation have the potential to degrade the accuracy and reliability of inferred spike statistics.

The recently developed jGCaMP8 sensors represent a promising advance due to their faster rise times and improved single-spike signal-to-noise ratios (SNR) (Zhang et al., 2023; Diana et al., 2024), but little is known about what new non-linearities they introduce. Past generation indicators have been reported to exhibit supralinear responses arising from closely occurring spikes (Chen et al., 2013; Éltes et al., 2019; Wei et al., 2020; Huang et al., 2021;

Schoenfeld et al., 2021). Neither models designed to capture this supralinearity (Deneux et al., 2016; Wei et al., 2020) nor models that assume linear characteristics of the newer probes (Zhang et al., 2023) are able to extract more accurate spiking dynamics using jGCaMP8 than they do from previous-generation sensors.

To address these gaps, we first comprehensively characterized the dynamics of GCaMP6f, jGCaMP7f, and jGCaMP8f responses. We identified novel nonlinear properties of these responses, including a pronounced use-dependent slowing by jGCaMP8f during periods of heightened activity. We developed a biophysical model of all three GCaMPs, which we used as the core of two new spike inference approaches: a sequential Monte Carlo (SMC) method to model latent spike-time sequences, and a machine learning-based strategy trained on synthetic data generated by the SMC method. Both approaches identify spike times with higher accuracy and temporal precision than any currently used technique.

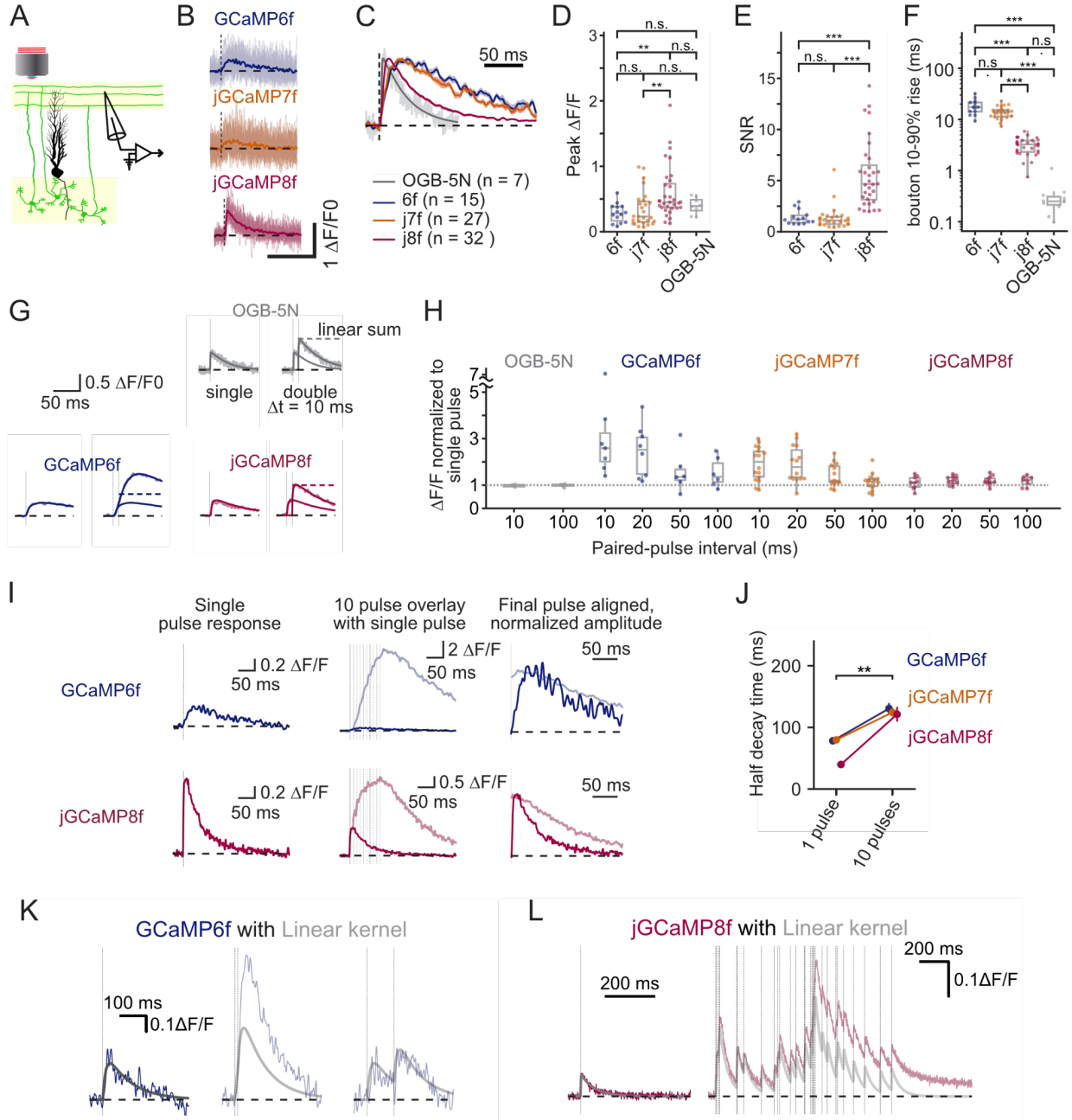
Results

Biophysical characterization of GCaMPs

jGCaMP8 variants have been reported to possess improved kinetic properties and linear responses to action potentials compared to earlier GCaMP versions (Zhang et al., 2023). To precisely quantify these improvements, we expressed GCaMP6f, jGCaMP7f, or jGCaMP8f in cerebellar granule cells using adeno-associated viral vectors. We performed high-speed (0.5-1 kHz) two-photon linescans to measure fluorescence transients in single boutons (parallel fibers, PFs; **Figure 1A-C**) in response to extracellularly-evoked action potentials. jGCaMP8f exhibited the fastest rise and decay times (**Figure 1C**), and a fractional per-spike fluorescence change of 0.59 ± 0.07 (n = 32 boutons, **Figure 1D**), larger than the other probes tested (GCaMP6f, 0.29 ± 0.04, n = 14 boutons; jGCaMP7f, 0.34 ± 0.05, n = 27; the small-molecule indicator Oregon Green BAPTA-5N, OGB-5N, 0.39 ± 0.04, n = 11; mean ± SEM; ANOVA, p = 0.0041). jGCaMP8f showed a signal-to-noise ratio of 5.6 ± 0.1 (**Figure 1E**), again higher than the other probes (GCaMP6f, 1.4 ± 0.2; jGCaMP7f, 1.3 ± 0.2; jGCaMP8f, 5.6 ± 0.1; mean ± SEM; ANOVA, p = 6.7e-10).

The rise time of jGCaMP8f was 3.1 ± 0.2 ms, approaching that of OGB-5N (0.3 ± 0.1 ms), which has near diffusion-limited calcium on-binding kinetics (DiGregorio et al., 1999) (**Figure 1F**) and was faster than all other probes (GCaMP6f, 18.1 ± 1.6 ms; jGCaMP7f, 13.7 ± 0.7 ms;; mean ± SEM; ANOVA, p = 6.7e-29; significant Tukey HSD post hoc comparisons: p = 1.1e-21, GCaMP6f vs. OGB5N; p = 5.4e-4, jGCaMP6f vs. jGCaMP7f; p = 3.9e-23, jGCaMP6f vs. jGCaMP8f; p = 8.7e-18, jGCaMP7f vs. OGB-5N; p = 1.5e-19, jGCaMP7f vs. jGCaMP8f).

Figure 1: Nonlinear response characteristics of fast GCaMPs in a controlled, ex vivo setting.



(**A**) Schematic of experimental approach showing sites of extracellular stimulation and imaged regions (highlighted in yellow). (**B**) Representative traces obtained across sensors for single-spike responses, with mean overlaid. (**C**) Traces from **B** averaged and normalized to compare kinetic properties of GCaMPs with the low affinity synthetic indicator, OGB-5N. (**D-F**) Peak response amplitude, signal-to-noise ratio, and 10-90% rise time for individual boutons. Each symbol represents an average across trials in one bouton. (**G**) averaged responses to one stimulus or two stimuli spaced 10 ms apart. (**H**) Relative amplitude of responses to a second stimulus spaced at various time intervals from the first stimulus. Each point represents the

average response for one bouton. (**I**) Responses to 1 stimulus (left), 10 stimuli at 100 Hz overlaid to the single stimulus response (center), and normalized to the amplitude of a single-stimulus response to demonstrate prolonged decay (right). (**J**) Half-decay times as a function of stimulus number. (**K-L**) Convolution of a single-stimulus response with the times of stimuli to demonstrate effect of short-term amplitude non-linearity in GCaMP6f responses in **K** and cumulative signal arising from many stimuli for jGCaMP8f in **L**. Stars indicate statistical significance as: *p<0.05, **p<0.01, ***p<0.001.

Recent reports of GCaMP6 indicators suggest a nonlinear relationship between the number of APs and the corresponding fluorescence response (Wei et al., 2020; Huang et al., 2021; Zhang et al., 2023). To quantify the time-dependence of this nonlinearity across fast indicator variants, we explored signals evoked by pairs of spikes delivered at defined time intervals. GCaMP6f and GCaMP7f showed short-term supralinearity that decreased to a linear summation for intervals greater than 100 ms (**Figure 1G, H**) (GCaMP6f, 3.1 ± 0.7 at 10 ms and 1.5 ± 0.3 at 100 ms fold increase, n = 7 boutons; jGCaMP7f, 1.9 ± 0.2 at 10 ms and 1.1 ± 0.1 at 100 ms, n = 14 boutons; mean ± SEM; two-way mixed ANOVA, $F(3,60) = 7.2$, p = 3.3e-4 main effect of indicator, with significant one-way ANOVA at 10 ms, $F(3,32) = 6.4$, $p = 0.0016$; significant Tukey HSD post hoc comparisons: $p = 0.01$, GCaMP6f vs. OGB-5N). In contrast, the jGCaMP8f responses were near-linear in their response amplitude to spikes at all tested intervals (**Figure 1G, H**) (testing against a null distribution centered on 1; one-sample t-test, p = 0.32).

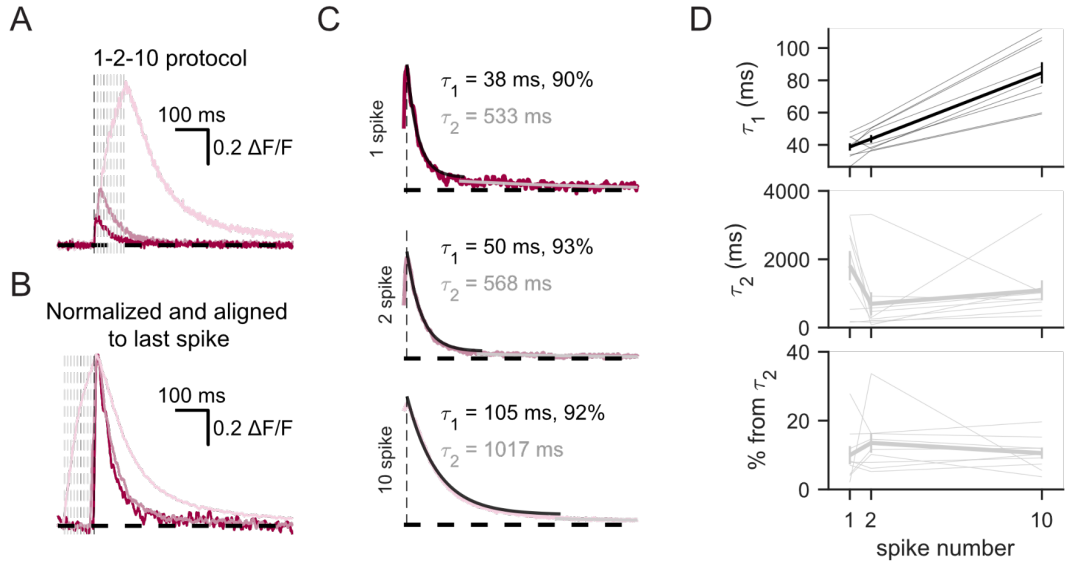
To examine responses to bursts of activity, we measured bouton fluorescence in response to 1, 2, and 10 AP stimuli delivered at 100 Hz. GCaMP6f responses showed a strong supralinearity, in which a 10-pulse train evoked signals approximately 20-fold larger than a single pulse (**Figure 1G**, top), consistent with positive cooperativity. In contrast, jGCaMP8f responses under the same comparison resulted in only a 3-fold larger response (**Figure 1F**, bottom). These findings were broadly consistent with indicator saturation and low apparent cooperativity (**Figure 1F**, bottom).

Unexpectedly, all three tested GCaMPs showed a slowing of the decay of fluorescence transients driven by ten spikes compared with a single spike (**Figure 1I**) (GCaMP6f, 78 ± 5 ms 1 pulse vs. 130 ± 9 ms 10 pulse, n = 6 boutons; jGCaMP7f, 77 ± 5 ms 1 pulse vs. 126 ± 7 ms 10 pulse, n = 22 boutons; jGCaMP8f, 41 ± 3 ms 1 pulse vs. 123 ± 12 ms 10 pulse, n = 14 boutons; mean ± SEM; two-way mixed ANOVA, $F(1,78) = 72.2$, p = 1.0e-12 main effect of time). This use-dependent effect was particularly pronounced for jGCaMP8f (~3-fold slowing) when compared to the older fast sensor variants (~1.5-fold slowing) (**Figure 1J, K**). Further analysis revealed that a two-component exponential was required to achieve acceptable fits to the decay even for single-spike responses. Both kinetic components tended to increase in duration with the number of spikes (**Supplementary Figure 1**).

To explore the impact of nonlinear properties of the old and new GCaMP variants, we compared experimentally-measured spike responses to a linear model. The model was calculated by convolving a kernel fitted to a single spike response with the spike train used to generate the observed data. In the paired-spike protocol, the GCaMP6f kernel underestimated the response to a 10 ms inter-spike interval. The estimate improved as the intervals lengthen and the nonlinearity decreased (**Figure 1J**). When jGCaMP8f boutons were exposed to a 10 Hz

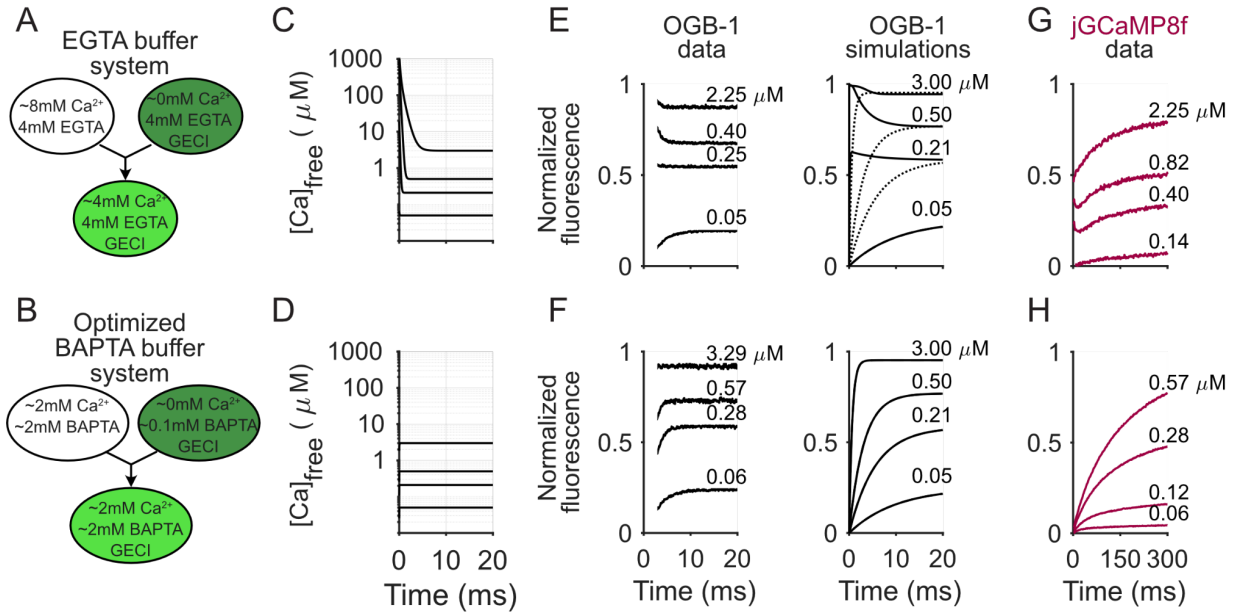
Poisson stimulus train, the kernel estimate diverged increasingly from the real data (**Figure 1K**). Overall, these analyses reveal that the GCaMP6f amplitude nonlinearity is dependent on spike timing, while the use-dependent slowing of jGCaMP8f signals is dependent on spike rate. These findings indicate activity regimes where linear methods for deconvolution are expected to be unsuccessful: short inter-spike intervals for older GCaMPs, and high spike rates for jGCaMP8f.

Supplementary Figure 1: jGCaMP8f response decay is captured by two kinetic components.



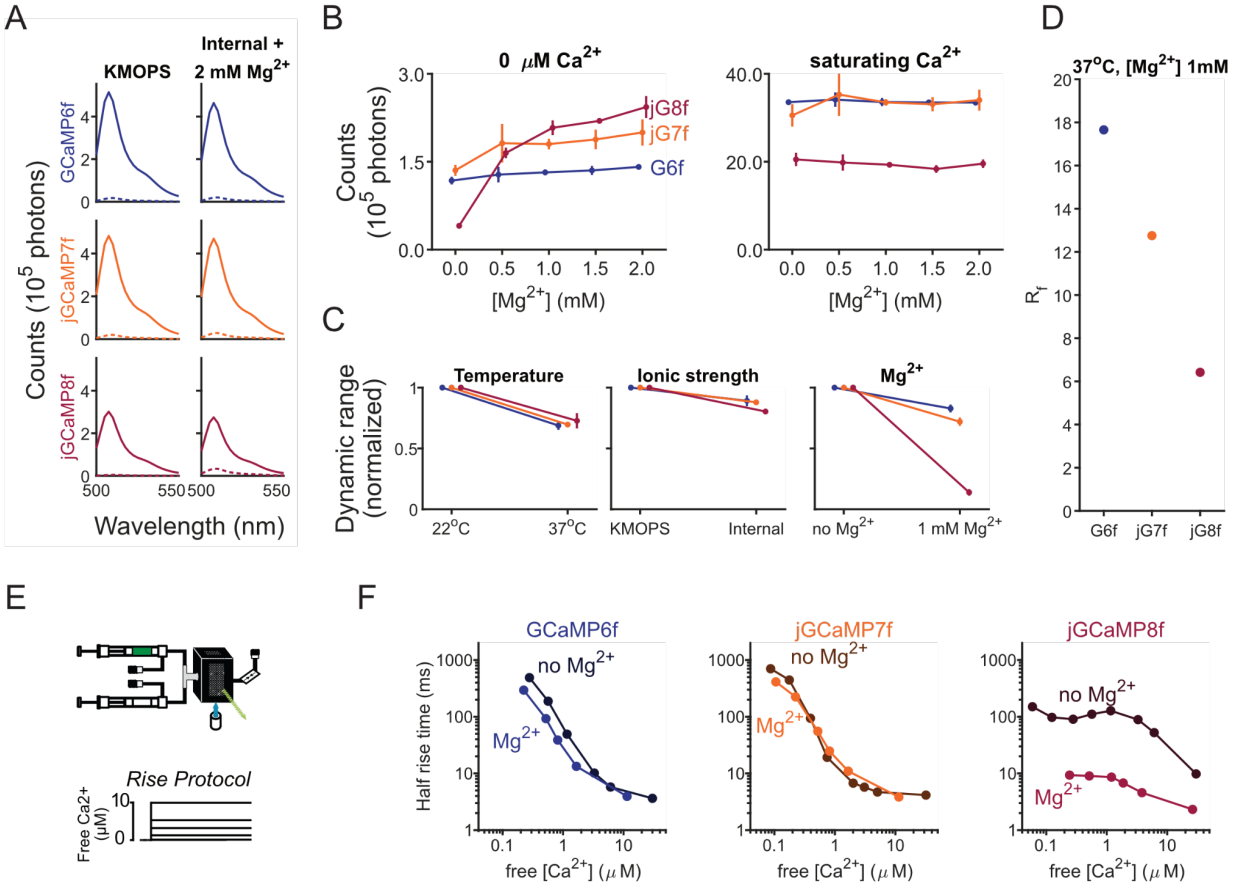
(**A**) Representative single-bouton responses to 1, 2, and 10 stimuli overlaid and aligned by first stimulus time. (**B**) Responses normalized and aligned to last spike. (**C**) Example fits to dual sums of exponentials showing fast and slow components of decay. (**D**) Dependence of fast (top) and slow (middle) time constants on the number of stimuli, and relative contribution of the slow component (τ_2) to the overall time course (bottom). Individual bouton values shown as low linewidth traces with group means shown as thicker line and SEM as error bars, *n* = 6 boutons and 3 granule cell somata.

Supplementary Figure 2: The impact of calcium clamp design on observed stopped-flow dynamics.



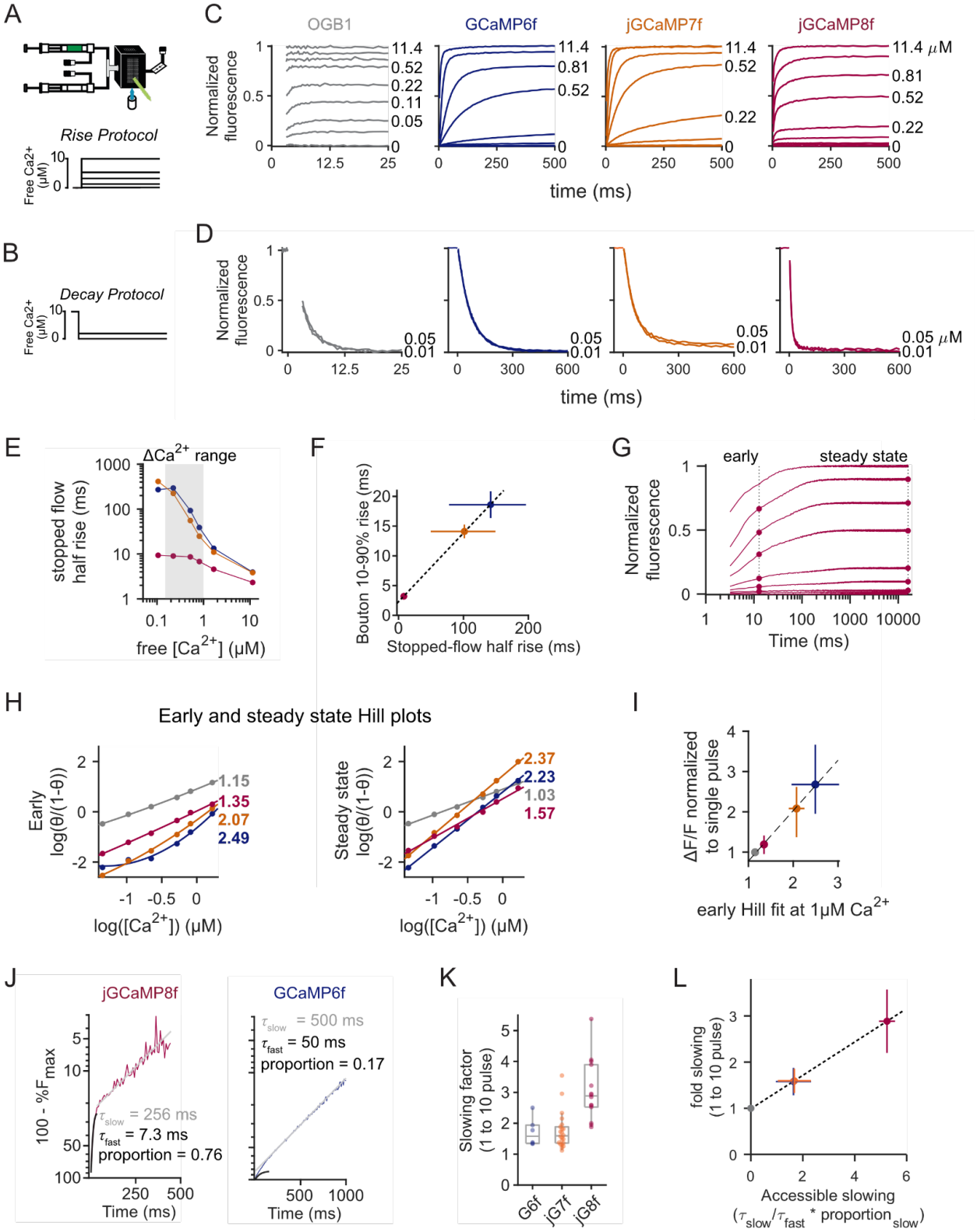
(**A-B**) Schematic of stopped-flow buffer systems using (**A**) a slow buffer (EGTA) and containing unbuffered calcium in the calcium-containing bolus (**A**) and (**B**) a system with a fast buffer (BAPTA) and fully buffered calcium in the calcium-containing bolus. (**C**) Simulations of the EGTA-based system presented in (**A**) show uncontrolled free calcium concentrations in the first few milliseconds of large calcium steps, but (**D**) BAPTA-based experiments do not. (**E**) Measurements of OGB-1 responses (left) reflect uncontrolled responses that are matched by simulations of the buffer system (right). (**F**) OGB-1 data and simulations in BAPTA-based buffers. (**G**) Artifactual fast on-response to jGCaMP8f in EGTA-based buffer system. (**H**) True jGCaMP8f response in BAPTA-based buffer system.

Supplementary Figure 3: Magnesium affects the properties of jGCaMP8f much more than older GCaMPs.



(**A**) Emission spectra of GCaMP variants in standard protein purification buffer (KMOPS, left) and under intracellular conditions that include magnesium (2 mM Mg^{2+}). (**B**) Fluorimeter counts at 490 nm under calcium-free (left) or saturating calcium (right) conditions. jGCaMP8f, but not other variants, showed strong magnesium-dependence of fluorescence emission under low-calcium conditions. No indicator showed this effect at saturating calcium. (**C**) Quantification of the weak dependence of GECI dynamic range R_f (saturating calcium divided by calcium-free fluorescence output) on temperature (left) or ionic strength (center) compared to the effect seen with magnesium (right). (**D**) Dynamic range of GCaMP variants at mammalian near-physiological conditions. (**E**) Schematic of stopped-flow experiment to measure responses to rising steps of calcium concentration. (**F**) Half-rise times to calcium steps were accelerated most by magnesium for jGCaMP8f, less so for GCaMP6f, and not at all for jGCaMP7f.

Figure 2: Stopped-flow fluorimetry of GCaMP dynamic responses.

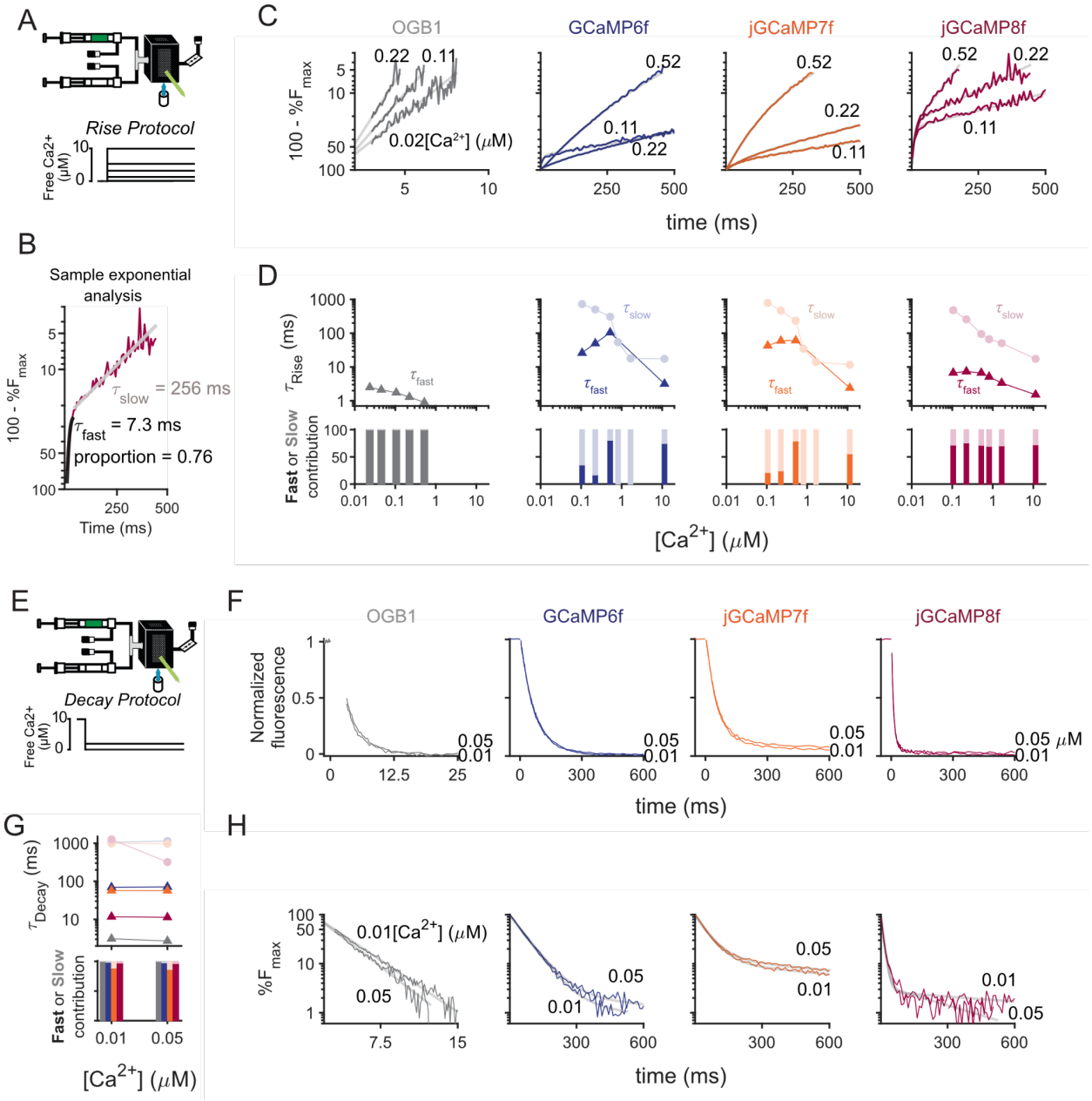


(**A**) Schematic of stopped-flow experiment and rising steps of calcium concentration. (**B**) Step-down protocols for decreasing calcium. (**C**) Responses to steps up from 0 μM free calcium. (**D**)

222 Responses to steps down from 10 μM free calcium. (**E**) Half-rise times for GCaMP6f,
223 jGCaMP7f, and jGCaMP8f as a function of calcium step size. Shaded gray region indicates the
224 physiological range of intracellular signals. (**F**) Comparison of brain slice presynaptic bouton
225 signals with stopped-flow rise times for the three indicator variants. (**G**) Rising phase responses
226 on logarithmic time scales. Symbols indicate the time points plotted in (**H**). (**H**) Hill plots of rising
227 and steady-state responses to steps in calcium. Colored numbers indicate Hill coefficients. (**I**)
228 Comparison of brain slice amplitude nonlinearity at 10 ms inter-stimulus response with respect
229 to Hill coefficients. (**J**) Representative stopped-flow traces with y-axis values presented on a log
230 scale for jGCaMP8f (left) and GCaMP6f (right). Sums of exponentials for each indicator
231 demonstrate the biphasic response and the greater contrast in speed of kinetic components
232 seen in jGCaMP8f. (**K**) Quantification of the factor of slowing seen in GCaMP fluorescence
233 decay in response to 1 and 10 stimuli in brain slices. (**L**) Comparison of brain slice slowing to
234 accessible slowing (see main text and methods). Error bars in all panels represent the
235 interquartile range of the distribution.
236
237 To gain insight into the mechanistic basis for GCaMP's use-dependent nonlinearities, we
238 measured GCaMP fluorescence responses to various calcium concentration steps directly using
239 stopped-flow fluorimetry on purified protein (**Figure 2A, B)**. We first optimized stopped-flow
240 buffer conditions to closely match intracellular ionic composition, including magnesium, which
241 we found to be necessary to reproduce the kinetics and dynamic range of the GCaMP
242 responses observed in living cells (**Supplementary Figure 2, 3**). The rise and decay kinetics of
243 jGCaMP8f were faster than 6f and 7f for all calcium concentration steps in the range of 0-10 μM
244 (**Figure 2C, D**). Within the range of calcium increase expected in response to a small number of
245 spikes (shaded region, **Figure 2E**), the rise time of spike-evoked calcium transients for each
246 GECI variant correlated with stopped-flow rise times (**Figure 2E, F**). Similarly, the paired-pulse
247 facilitation of a GCaMP's fluorescent response in neurons closely tracked cooperativity
248 measures derived from analysis of the early stopped-flow response (**Figure 2G-I**). These
249 experiments show that our stopped-flow measurements capture key features of action potential-
250 evoked calcium responses in neurons.
251 Previous studies showed that activation of GCaMPs by calcium occurs via two
252 pathways, one fast and one slow (Sun et al., 2013; Badura et al., 2014). Closer inspection of the
253 binding and unbinding kinetics across GCaMP sensors revealed fast and slow phases of
254 fluorescence increases. Responses of GCaMP6f and jGCaMP7f arose predominantly from a
255 slow component whose speed increased with the size of the calcium step. In contrast,
256 jGCaMP8f's fluorescence developed predominantly through a fast component. While both
257 kinetic components contributed to fluorescence changes in all GECIs, jGCaMP8f was
258 distinguished by having a faster and larger contribution from the fast component at all
259 concentrations tested (**Figure 2J**, **Supplementary Figure 4**).
260 The presence of two time constants in the rising phase of the fluorescence response is
261 consistent with entrance into at least two distinct fluorescence states. The presence of multiple
262 fluorescence states could account for fast and slow decay rates. We compared the time
263 constants governing this fast and slow increase to the use-dependent slowing effect observed in
264 our ex vivo (i.e. brain slice) recordings. We found a linear relationship between our ex vivo
265 observations and a metric we term accessible slowing which represents a tradeoff between the

speed of entrance into the fast versus slow state and stability of the slow state (**Figure 2K, L**). These observations support the conclusion that large calcium increases push GCaMPs into a stable fluorescent state from which they are slow to exit.

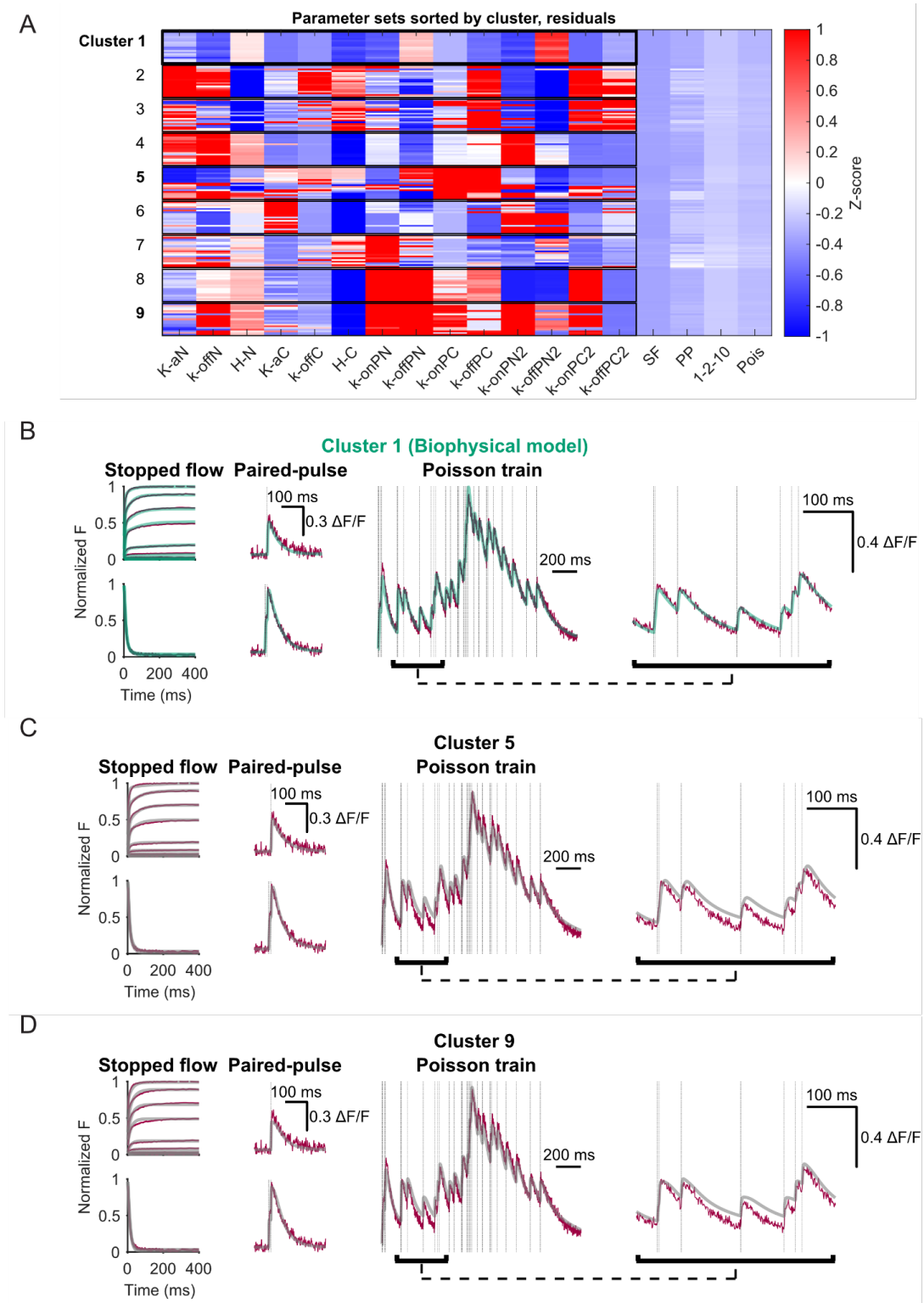
Supplementary Figure 4: Biphasic responses of GCaMPs to increasing calcium steps.



(**A**) Schematic of stopped-flow experiment and rising calcium-step protocol. (**B**) Example analysis of fit to a sum of two exponentials. (**C**) Monoexponential rising kinetics for Oregon Green BAPTA-1 (OGB-1) and biexponential kinetics for GCaMP variants. (**D**) Dependence of fast and slow components on amplitude of calcium step. Upper plots show rate constants represented in the response at the indicated step. In the lower plots, vertical bars show relative amplitude of fast (dark) and slow (light) components. (**E**) Schematic of decreasing-calcium step

protocol. (**F**) Fluorescence responses to decreasing calcium steps. (**G**) Breakdown of kinetic components contributing to the decay of fluorescence as in (**D**). Due to the low number of steps considered, analysis for all variants is included in this single plot. (**H**) Monoexponential decay for OGB-1 and more complex responses for GCaMP variants.

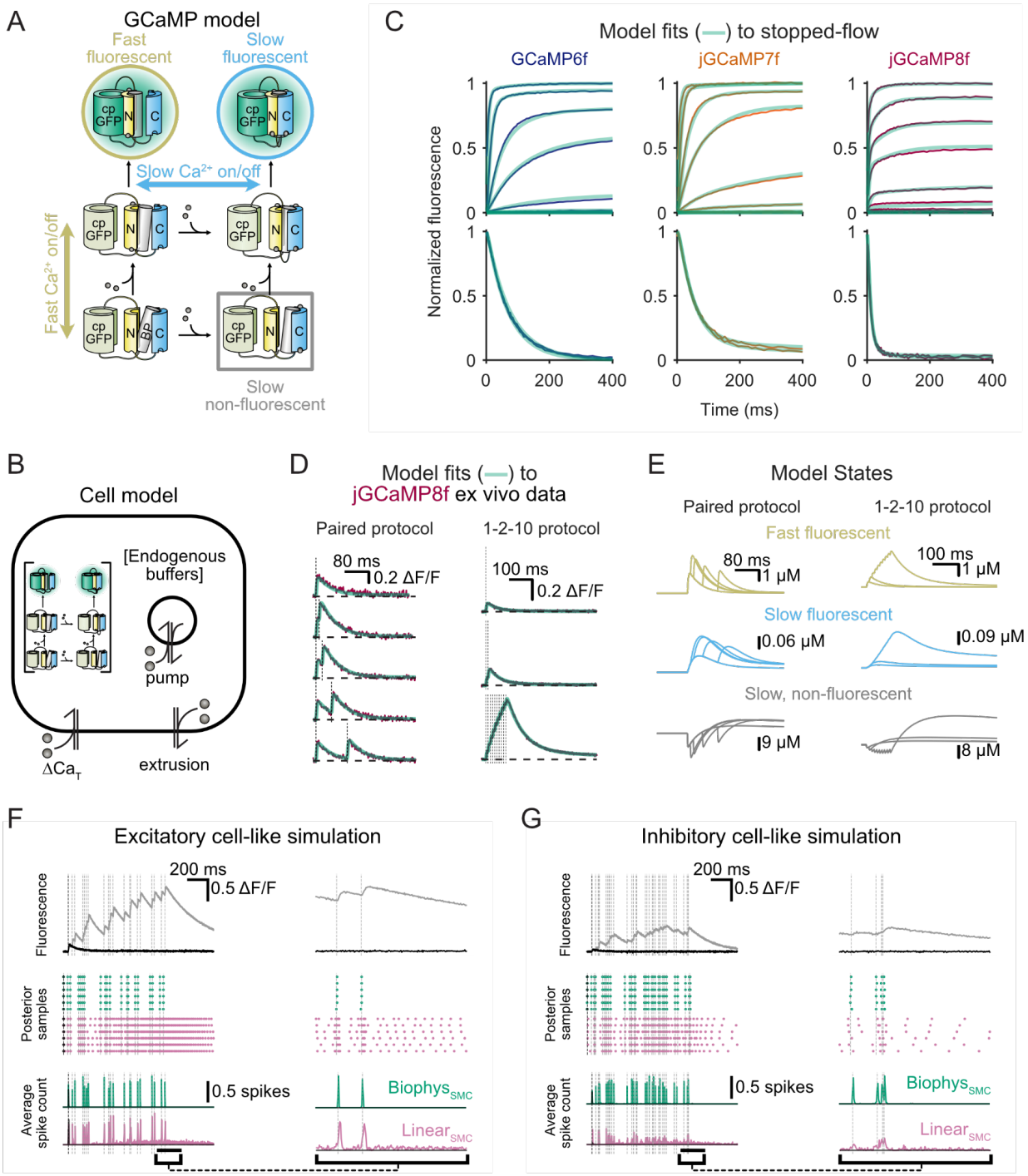
Supplementary Figure 5: Selection of GCaMP biophysical parameters using in vitro and ex vivo datasets.



(**A**) Distributions of parameters on fits from randomized seeds with GCaMP model parameters at figure left and residuals from best fits to stopped-flow and ex vivo experiments (paired pulse,

1-2-10 pulse, and Poisson train) based on optimizing cell parameters. GCaMP parameter sets were sorted by hierarchical clustering and then by lowest summed residuals within each cluster. This procedure allowed qualitative assessment of parameters that lay in different local minima of the parameter space. (**B**) Fitted data examples for cluster 1, which was used for all further modeling. (**C**) Fitted data for cluster 5. (**D**) Fitted data for cluster 9. In paired-pulse and Poisson train data, vertical lines indicate the times of action potentials. Expanded timescale of Poisson responses at right. Neither cluster 5 or 9 captured the differential slowing across the Poisson responses.

Figure 3: A biophysical model for GCaMP response.



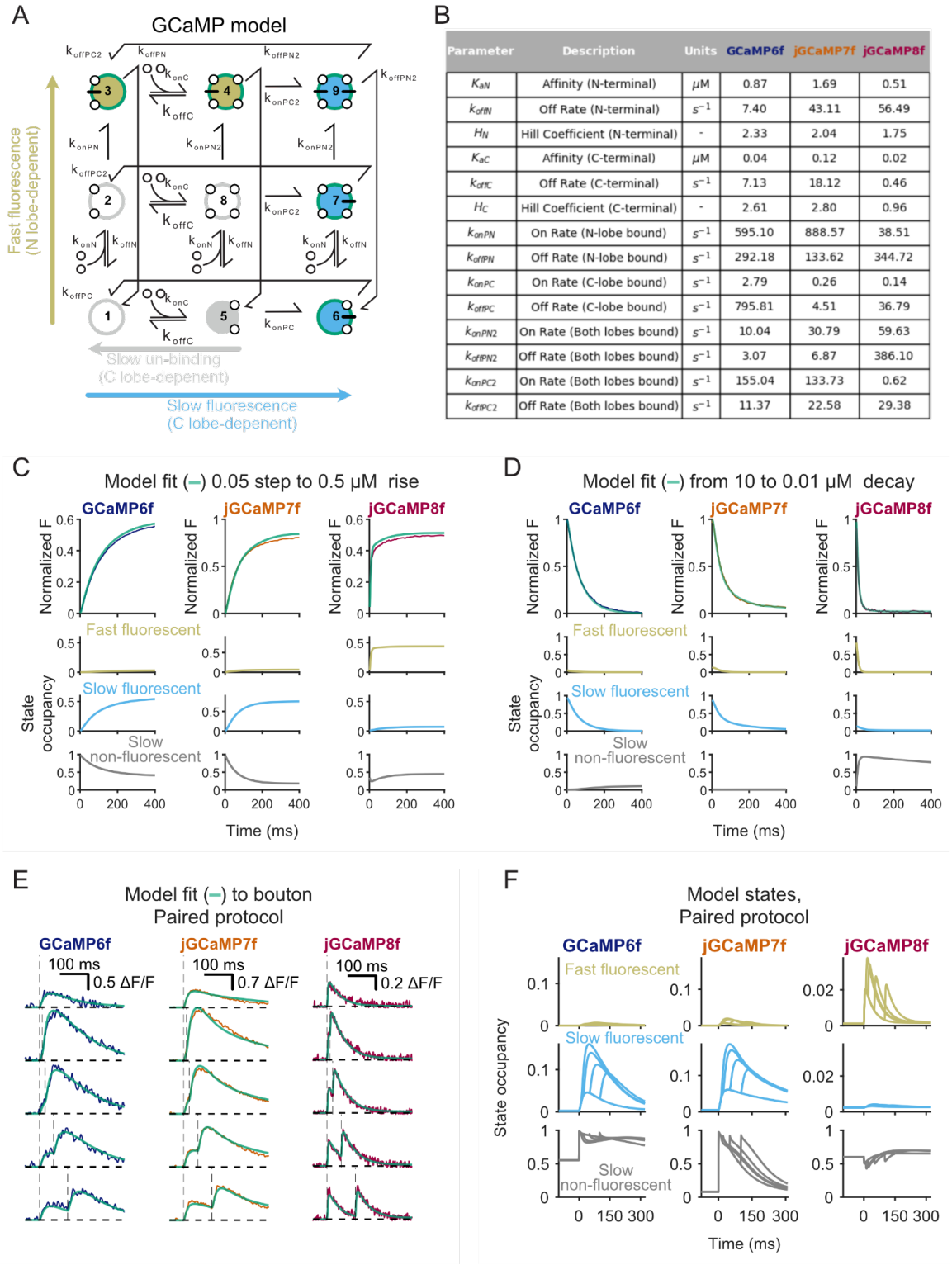
(**A**) Schematic GCaMP model depicting probe components, including circularly permuted green fluorescent protein (cp-GFP), the N- and C-lobes of the calcium binding-calmodulin, and the binding peptide (BP). Black arrows indicate key pathways to fluorescence with states highlighted for their role in model function. (**B**) Schematic diagram of physiological response of a cellular structure containing endogenous calcium buffer, GCaMP, and intracellular and plasma membrane calcium extrusion (pump) mechanisms. (**C**) Fits of the GCaMP model to the three

305 GCaMP variants with different parameters for each variant. (**D**) Fits of the cell model to
306 physiological data with fixed GCaMP model parameters. Vertical lines indicate action potential
307 (spike) times. (**E**) Occupancy of the fast-fluorescent, slow-fluorescent, and slow non-fluorescent
308 states of GCaMP indicating a mechanism for use-dependent slowing of fluorescence decay,
309 becoming most apparent after 10 action potentials. (**F**) Predictions on simulated ground-truth
310 data for an excitatory neuron. The data consists of a single spike (black) or a 10 Hz Poisson
311 train of spikes (gray) (top panel). Lower panels show the average posterior spike probability
312 (solid lines) and spikes times from a single trajectory (triangles) derived by our SMC method
313 employing either the biophysical model ($\text{Biophys}_{\text{SMC}}$, green) or a linear model ($\text{Linear}_{\text{SMC}}$, pink)
314 as its generative engine. Vertical lines indicate action potential (spike) times. (**G**) As in (**F**) with
315 simulated ground truth for an inhibitory cell (30 Hz Poisson spike rate, calcium increase per
316 spike reduced by 2, dynamic range reduced by a factor of 2).
317

318 **Development of a biophysical model of GCaMP**

319 We built a biophysical model of GCaMP that included parallel fast and slow kinetic
320 pathways to multiple fluorescence states. Multiple kinetic pathways are known to coexist in
321 calmodulin, the calcium-sensing domain of GCaMP, which semi-independently binds calcium
322 via its N-lobe (fast-binding) and C-lobe (slow-binding) (Linse et al., 1991; Faas et al., 2011),
323 each subsequently capable of interacting with a target peptide to induce fluorescence (Johnson
324 et al., 1996; Brown et al., 1997; Sun et al., 2013; Lai et al., 2015). We reasoned that such a
325 model structure might phenomenologically account for both the *in vitro* and *ex vivo* observations
326 (**Figure 3A**). In order to estimate the kinetic rate constants of the GCaMP model we fit it to
327 experimental data. We took advantage of a neuron type (cerebellar granule cells) with well-
328 characterized calcium buffering and handling (Rebola et al., 2019). The cell model used here
329 included endogenous buffering, calcium extrusion mechanisms, and calcium exchange from
330 intracellular stores (**Figure 3B**) and could reproduce the fast dynamics of previous buffer
331 models (Rebola et al., 2019), but we also added a second compartment to account for a slow
332 decay that was not used to constrain previous models. By jointly fitting the stopped-flow data
333 and *ex vivo* responses, we derived the final GCaMP model parameters for each variant, which
334 were then fixed for all subsequent analyses (**Supplementary Figure 5; Supplementary Figure**
335 **6A, B**). This approach allowed us to reproduce observed dynamics from all experimental
336 systems and GCaMP variants, despite the idiosyncrasies observed across the sensors (**Figure**
337 **3C, D; Supplementary Figure 6C-F**).
338 Further analyses of the GCaMP model states (**Figure 3A; Supplementary Figure 6A**)
339 revealed additional insights into how parameterization led to different operational modes. When
340 separating the fluorescent GCaMP model states into those that depended on C-lobe binding
341 (slow fluorescent) and those that did not (fast fluorescent), we observed that jGCaMP8f alone
342 showed stable occupancy of the fast fluorescent states in both stopped-flow and ex vivo fits
343 (yellow traces **Supplementary Figure 6C-F**). The slowing of jGCaMP8f responses following
344 spike bursts, in contrast, was driven by increased occupancy of a slow, fluorescent and non-
345 fluorescent state (**Figure 3E**, blue and gray, respectively).
346
347

Supplementary Figure 6: A full biophysical state analysis of GCaMP response.

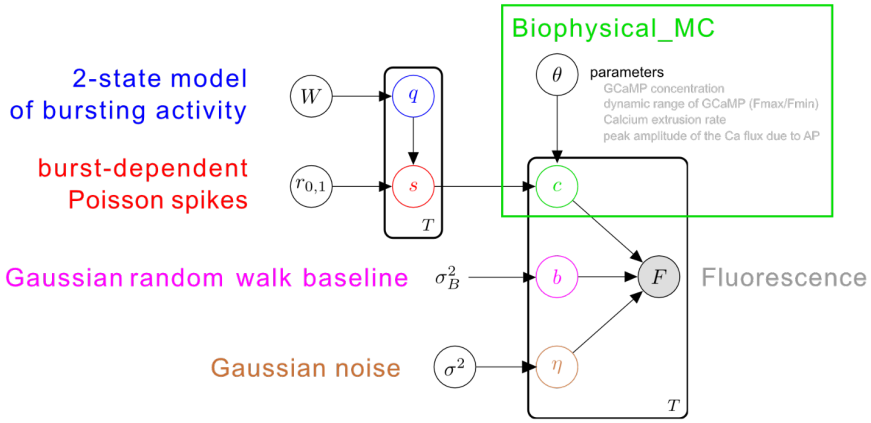


Parameter	Description	Units	GCaMP6f	jGCaMP7f	jGCaMP8f
K_{aN}	Affinity (N-terminal)	μM	0.87	1.69	0.51
k_{offN}	Off Rate (N-terminal)	s^{-1}	7.40	43.11	56.49
H_N	Hill Coefficient (N-terminal)	-	2.33	2.04	1.75
K_{aC}	Affinity (C-terminal)	μM	0.04	0.12	0.02
k_{offC}	Off Rate (C-terminal)	s^{-1}	7.13	18.12	0.46
H_C	Hill Coefficient (C-terminal)	-	2.61	2.80	0.96
k_{onPN}	On Rate (N-lobe bound)	s^{-1}	595.10	888.57	38.51
k_{offPN}	Off Rate (N-lobe bound)	s^{-1}	292.18	133.62	344.72
k_{onPC}	On Rate (C-lobe bound)	s^{-1}	2.79	0.26	0.14
k_{offPC}	Off Rate (C-lobe bound)	s^{-1}	795.81	4.51	36.79
k_{onPN2}	On Rate (Both lobes bound)	s^{-1}	10.04	30.79	59.63
k_{offPN2}	Off Rate (Both lobes bound)	s^{-1}	3.07	6.87	386.10
k_{onPC2}	On Rate (Both lobes bound)	s^{-1}	155.04	133.73	0.62
k_{offPC2}	Off Rate (Both lobes bound)	s^{-1}	11.37	22.58	29.38

(**A**) Full biophysical state model of GCaMP response. Specific states referred to as “fast fluorescent”, “slow fluorescent”, and “slow, non-fluorescent” have centers color-coded as in model state traces below. (**B**) Best-fit parameter values for GCaMP6f, jGCaMP7f, and jGCaMP8f. (**C**) Model fits to 0.5 μM increasing step of calcium (top), with model state broken

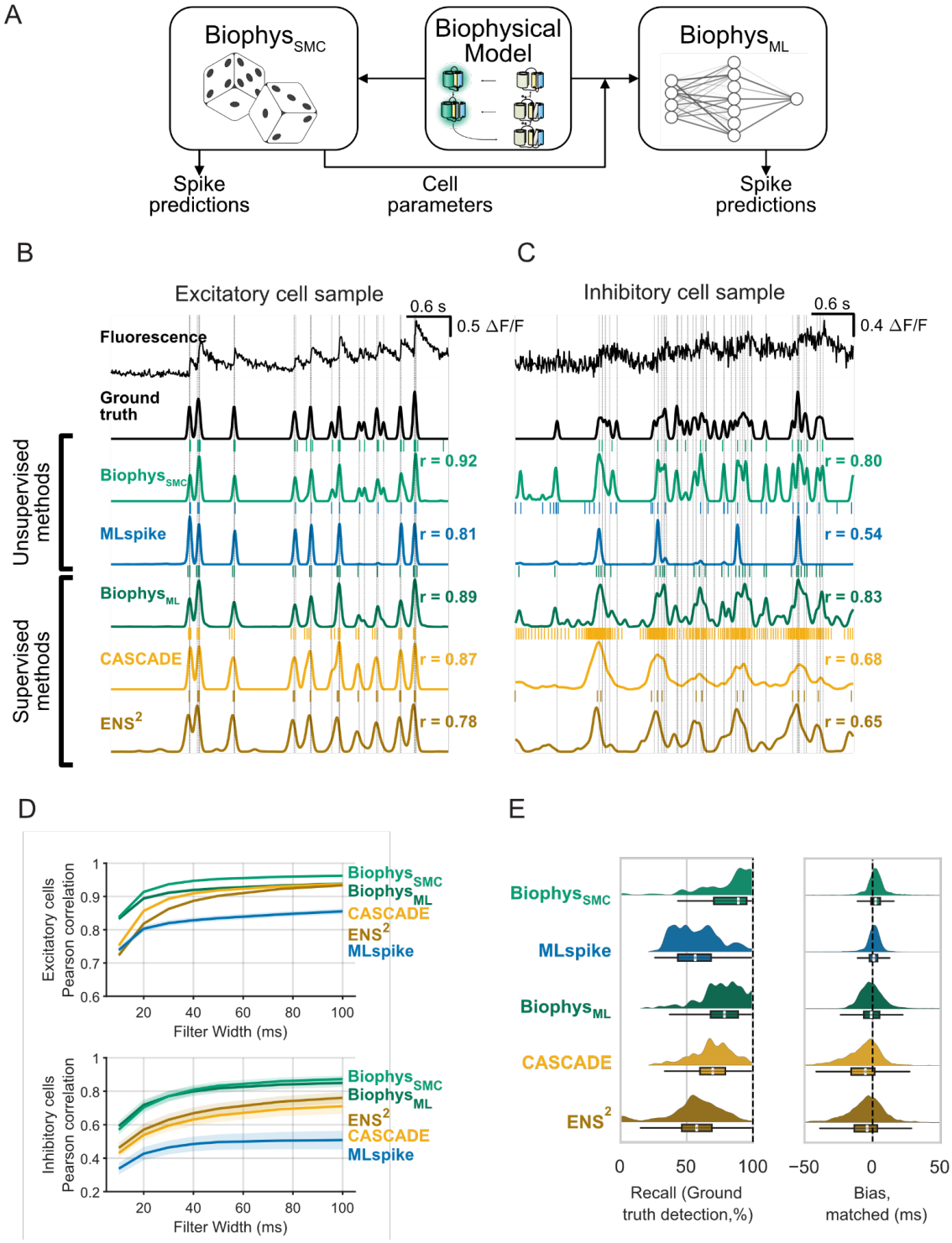
down by occupancy. Note that jGCaMP8f is distinguished by its near-exclusive use of the fast fluorescent states. (**D**) As in (**C**) for decreasing step. jGCaMP8f is distinguished by fast, persistent entry into the slow, non-fluorescent state. (**E**) Model fits to fluorescence responses in presynaptic boutons of granule cells. (**F**) State occupancy for the fluorescence responses shown in (**E**). Fast responses arise in part from high occupancy of a slow, non-fluorescent state at basal calcium for jGCaMP8f which efficiently transitions to fluorescence on elevation of calcium.

Supplementary Figure 7: Generative cell-based model of fluorescence time series



Graphical representation of the generative SMC model described in the main text. White circles denote unknown variables, gray circles denote measurements and bare variables are fixed prior hyperparameters. Plates denote groups of variables. Note that GCaMP model parameters are included, but fixed for the SMC approach.

Figure 4: Improved action potential inference from generative model-based methods.



(**A**) A generative model for imputing action potential times based on iterative fitting. The biophysical GCaMP model was held fixed while cell parameters were varied. Responses to action potential sequences were generated to fit cell parameters (Biophys$_{SMC}$). After fitting was complete, the Biophys$_{SMC}$ model was then used to train a machine learning based model. (**B, C**)

375 Imputed spike probabilities using a variety of unsupervised and supervised models. Values of r
376 indicate Pearson correlation between the time course of ground truth and that of the imputed
377 probabilities for data filtered at 20 ms. Vertical lines indicate true spike times and tick marks
378 indicate imputed spikes. (**B**) shows data from an excitatory cell while (**C**) shows data from an
379 inhibitory cell. (**D**) Pearson correlations for different filtering time constants. (**E**) Discrete spike
380 estimate statistics. Left panels shows the average time between the nearest imputed and
381 ground truth spikes while the right shows the normalized number of spikes that were over or
382 underestimated. Individual points are calculated over 8-second epochs covering 1000 data
383 points randomly selected from the full dataset.(**E**) Timing and detection accuracy of the different
384 imputation models. Left, distribution of per-spike timing error for the various methods. Smoothed
385 histograms and box/whisker plots summarize all errors. Right, epoch-wise recall of the various
386 methods indicating the percent of ground truth spikes recovered.
387

388 **Model-based insights enable improved methods of spike detection**

389 We next leveraged the GCaMP model to implement a generative spike inference
390 algorithm. Inspired by recent successes of a Bayesian autoregressive-model based approach
391 (Diana et al. 2024), we adopted the same Sequential Monte Carlo (SMC) framework in which
392 time-invariant parameters (such as those governing cell calcium handling) and latent states
393 (such as spike times) are jointly estimated in an iterative manner to estimate spike times
394 (**Supplementary Figure 7**).
395 As an initial test of this approach, we generated ground truth simulations using the
396 GCaMP model and attempted inference using our SMC framework with either the GCaMP
397 model ($\text{Biophys}_{\text{SMC}}$ hereafter) or a linear autoregressive model ($\text{Linear}_{\text{SMC}}$) as the generative
398 model (**Figure 3F, G**).Using this approach, $\text{Linear}_{\text{SMC}}$ produced a progressive increase in the
399 number of false spike predictions during periods of elevated spiking for both cell types,
400 particularly in the decay phase of the fluorescence. This is likely due to the use-dependent
401 slowing of jGCaMP8f fluorescence. In contrast, $\text{Biophys}_{\text{SMC}}$, which accommodated all
402 nonlinearities, made consistent, accurate predictions of spike times across the simulation
403 epoch. The $\text{Biophys}_{\text{SMC}}$ was able to reproduce the input spike train with higher accuracy and
404 precision in the context of simulations designed to mimic different cell types, including both
405 excitatory and inhibitory cells.
406 The $\text{Biophys}_{\text{SMC}}$ approach to spike inference is computationally time-consuming due to
407 the SMC sampling. Therefore we also used the GCaMP model to create a synthetic training
408 dataset to train a machine learning algorithm (Rupprecht 2021). Specifically, we generated
409 synthetic fluorescence data for training using cell parameters derived from the $\text{Biophys}_{\text{SMC}}$ along
410 with randomly generated spikes, noise, and baseline drift designed to mimic the statistics of the
411 dataset targeted for inference (**Figure 4A**). In this way, we obtained two methods for spike
412 inference: an unsupervised, generative Bayesian approach ($\text{Biophys}_{\text{SMC}}$), and a computationally
413 faster machine-learning approach ($\text{Biophys}_{\text{ML}}$) which was generated by using the outputs of the
414 unsupervised approach to train a supervised-learning neural network.
415 We benchmarked our spike inference methods against state-of-the-art supervised and
416 unsupervised approaches for spike inference, using excitatory and inhibitory cell responses
417 from the jGCaMP8f ground truth dataset (Zhang et al., 2023). We used 154 18-second epochs
418 of firing in 37 excitatory neurons and 3 inhibitory neurons, containing 8,161 ground truth action

potentials, with a median of 34 (IQR, 16-69) spikes per epoch. For a generative method comparison, we used MLspike, an established algorithm based on an alternative generative model of calcium-fluorescence dynamics (Deneux et al., 2016). For supervised-method comparisons, we used the CASCADE (Rupprecht et al., 2021) and ENS^2 (Zhou et al., 2023) global-trained networks, which have been reported to outperform other methods on a variety of benchmark tests.

We first examined the ability of our methods to estimate continuous changes in activity. Imputed spike predictions and ground truth spikes were smoothed with Gaussian filters of various widths. $Biophys_{SMC}$ and $Biophys_{ML}$ model predictions were tightly locked to true spike patterns for both excitatory and inhibitory classes (**Figure 4B, C**), as indicated by the higher correlations between our method predictions and ground truth at all filter widths, particularly at ten and twenty milliseconds. Results with CASCADE were not improved by training the network on the jGCaMP8f dataset itself (**Supplementary Figure 8**).

We quantified the ability of various methods to track changes in firing using the Pearson correlation between continuous spike probability, filtered on various timescales, and the smoothed time course of ground truth (**Figure 4D**). For excitatory neurons (n = 142 epochs), a mixed effects model shows that correlation depends on the method used (method and filter width as fixed effects, $F(4,4935) = 171.1$, p = 2e-137). Method-wise comparisons demonstrate that $Biophys_{SMC}$ outperforms MLspike (p = 6.5e-62, d = 1.1), CASCADE (p = 9.0e-47, d = 0.8), and ENS^2 (p = 2.3e-82, d = 1.0). Likewise, $Biophys_{ML}$ outperformed MLspike (p = 3.9e-56, d = 1.1), CASCADE (p = 9.8e-42, d = 0.3), and ENS^2 (p = 9.4e-76, d = 0.6). $Biophys_{ML}$ outperformed CASCADE at filter widths up to 30 milliseconds, ENS^2 up to 76 milliseconds, and MLspike up to 100 milliseconds, the widest filter tested.

For the inhibitory cell dataset (n = 12 epochs), a similarly-structured mixed effect model again showed that correlation depends on method ($F(4,385) = 14.7$, p = 3.6e-11) with improved performance with $Biphys_{SMC}$ (vs. MLspike p = 4.1e-9, d = 1.9; vs. CASCADE p = 3.3e-4, d = 1.0; vs. ENS^2 p = 0.005, d = 1.1) and $Biophys_{ML}$ (vs. MLspike p = 1.4e-10, d = 2.0; vs. CASCADE p = 3.3e-5, d = 1.2; vs. ENS^2 p = 8.2e-4, d = 1.4). Together, these comparisons demonstrate the superior ability of $Biophys_{SMC}$ and BiophysML to estimate continuous spike probability on a variety of timescales and in different cell types.

Next, we calculated how well the different methods were able to capture discrete spike times. First, we matched imputed spikes to the nearest ground-truth spike within a 10 ms interval to avoid double-counting of spikes and penalize imprecise imputed spikes (**Supplementary Figure 9**). Our biophysical methods consistently found ground truth events at a higher rate than other methods as assessed by sensitivity (also called recall, defined as true positives as a fraction of all ground truth events) (**Figure 4E** left, **Table 1**; repeated measures ANOVA, $Biophys_{SMC}$ vs. MLspike p = 4.0e-9, d = 1.1; vs. CASCADE p = 4.0e-9, d = 0.6; vs. ENS^2 p = 4.0e-9, d = 1.0; $Biophys_{ML}$ vs. MLspike p = 4.0e-9, d = 1.1; vs. CASCADE p = 4.0e-9, d = 0.7; vs. ENS^2 p = 4.0e-9, d = 1.0).

We further tested the methods using a standard measure of predictive performance, the F-score. The F-score is the harmonic mean of precision (true positives as a fraction of all imputed events) and recall and has a maximum value of 1 thus allowing assessment of performance considering both false negatives as well as false positives. Taking this broader assessment into account, we found that the biophysical methods again outperform the other

methods (**Table 1**; repeated measures ANOVA, $Biophys_{SMC}$ vs. MLspike p = 3.8e-9, d = 1.1; vs. CASCADE p = 3.8e-9, d = 0.8; vs. ENS^2 p = 3.8e-9, d = 0.9; $Biophys_{ML}$ vs. MLspike p = 3.8e-9, d = 0.9; vs. CASCADE p = 3.8e-9, d = 0.6; vs. ENS^2 p = 3.8e-9, d = 0.6). Thus, our biophysically-based methods provide unprecedented temporal accuracy and precision compared to current state-of-the-art approaches to spike predictions based on jGCaMP8f imaging datasets.

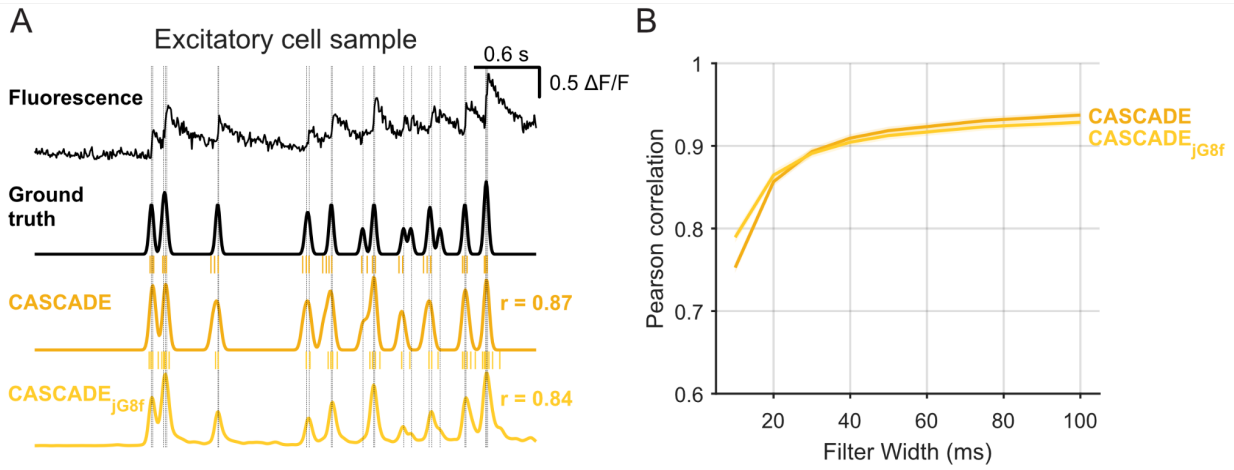
Finally, we tested the ability of our methods to estimate individual spike times. We first calculated the time between imputed spikes and the nearest matched ground-truth spike in a 100 ms interval to assess bias and uncertainty in spike time predictions (**Figure 4E**, right, **Supplementary Figure 10**). On an epoch-by-epoch basis, both of the supervised methods predicted spike times that tended to precede ground-truth times (CASCADE, 7.45 ms; ENS^2, 7.67 ms); MLspike showed a smaller bias (4.93 ms). In contrast, both biophysical methods removed this bias (**Table 1**; repeated measures ANOVA, $Biophys_{SMC}$ vs. MLspike p = 6.0e-8, d = 0.5; vs. CASCADE p = 6.0e-8, d = 0.8; vs. ENS^2 p = 6.0e-8, d = 0.6; $Biophys_{ML}$ vs. MLspike p = 1.8e-6, d = 0.4; vs. CASCADE p = 6.0e-8, d = 0.9; vs. ENS^2 p = 1.8e-7, d = 0.5).

The biophysical methods also reported spike times with lower uncertainty ($Biophys_{SMC}$ median 3.8, IQR 1.8-7.3 ms; $Biophys_{ML}$ median 5.8, IQR 2.7-10.3 ms) than the other supervised methods (CASCADE median 8.2, IQR 3.5-18.2 ms; ENS^2 median 8.3, IQR 3.8-16.6 ms) and superior epoch-wise statistics than all other methods (repeated measures ANOVA, $Biophys_{SMC}$ vs. MLspike p = 0.01, d = 0.3; vs. CASCADE p = 6.0e-8, d = 0.5; vs. ENS^2 p = 1.2e-7, d = 0.5; $Biophys_{ML}$ vs. MLspike p = 0.4, d = 0.1; vs. CASCADE p = 6.0e-8, d = 0.5; vs. ENS^2 p = 1.5e-5, d = 0.4). In summary, the generative biophysical model-based methods outperformed existing in-class and out-of-class alternatives, exceeding previous limits in the accuracy of GCaMP-based spike decoding.

Table 1. Single-spike classification metrics for calcium-based decoding methods. Median performance across epochs for a database of 37 excitatory cells with ground-truth electrical recordings and jGCaMP8f fluorescence signals. Values reported as median (IQR).

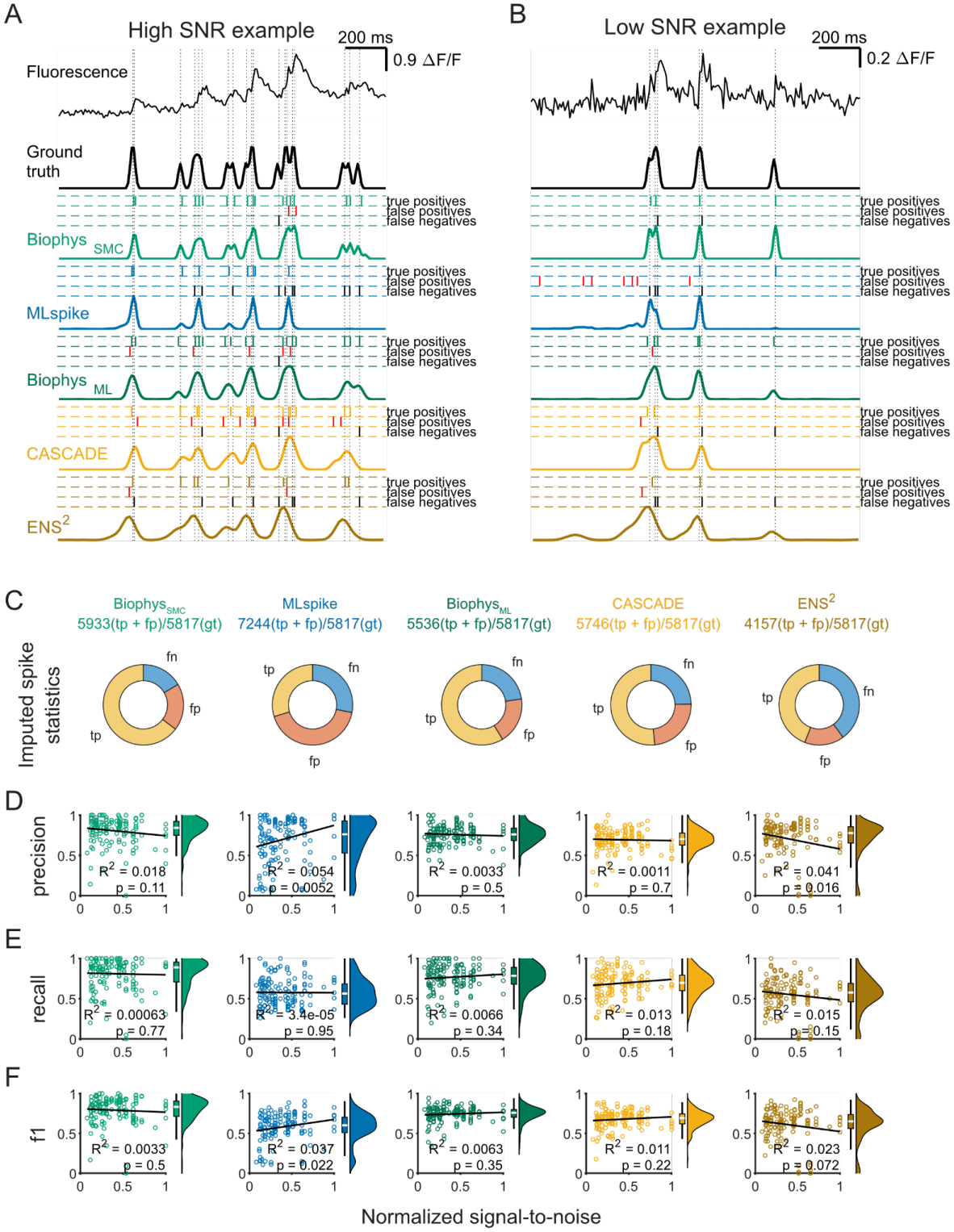
	Precision	Recall	F-Score	Accuracy	Mean bias (ms)	Uncertainty (ms)
$\text{Biophys}_{\text{SMC}}$	0.84 (0.75, 0.92)	0.88 (0.71, 0.95)	0.83 (0.71, 0.91)	0.71 (0.55, 0.83)	1.9 (-2.9, 3.0)	3.8 (1.8, 7.3)
MLspike	0.70 (0.53, 0.93)	0.56 (0.43, 0.68)	0.60 (0.51, 0.70)	0.43 (0.34, 0.54)	-2.0 (-0.6, 0.9)	3.2 (1.5, 6.2)
$\text{Biophys}_{\text{ML}}$	0.76 (0.68, 0.84)	0.78 (0.68, 0.89)	0.75 (0.70, 0.80)	0.61 (0.54, 0.67)	0.0 (-2.9, 3.0)	5.8 (2.7, 10.3)
CASCADE	0.69 (0.63, 0.78)	0.70 (0.60, 0.79)	0.69 (0.61, 0.74)	0.52 (0.45, 0.59)	-7.4 (-11.2, -3.2)	8.2 (3.5, 18.2)
ENS^2	0.71 (0.65, 0.86)	0.57 (0.46, 0.69)	0.65 (0.56, 0.74)	0.48 (0.39, 0.58)	-6.6 (-12.3, -2.3)	8.3 (3.8, 16.5)

Supplementary Figure 8: Training on jGCaMP8f data does not improve CASCADE performance.



(**A**) Imputed spike probabilities using the CASCADE Global Exc model or one trained on the jGCaMP8f dataset itself. Values of r indicate Pearson correlation between the time course of ground truth and that of the imputed probabilities for data filtered at 20 ms. Vertical lines indicate true spike times and tick marks indicate imputed spikes. (**B**) Pearson correlations for different filtering time constants.

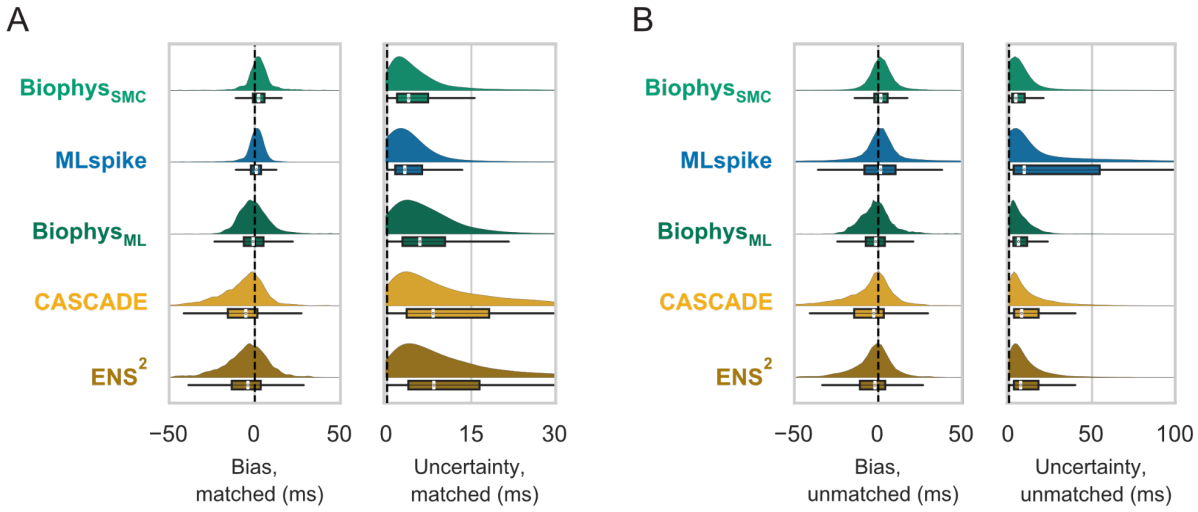
Supplementary Figure 9: Statistics of discrete spike estimate comparisons to ground truth spike times



(**A, B**) Continuous and discrete spike predictions across methods. The top two traces (black) show raw fluorescence and Gaussian-convolved (10 ms) ground truth spike times. Ground truth

spike times are indicated as vertical tick marks. Discrete imputed spikes for each method are shown above the corresponding continuous prediction trace, sorted into true positives, false positives, and false negatives. An epoch was selected from a representative cell which produced high signal-to-noise ratio (SNR) fluorescence changes in response to an action potential (**A**) or low SNR (**B**). (**C**) Columns show the proportions of imputed spikes that were scored as true positive (tp), false positive (fp), or false negative (fn) for each method with a comparison of all imputed spikes produced by each method (tp + fp) compared to total number of ground truth (gt) spikes from the 37 putative excitatory cells in the Janelia dataset. (**D-F**) Precision (tp/(tp + fn), **D**), Recall (tp/(tp + fp), **E**), and F-score (harmonic mean of recall and precision, **F**) as a function of the SNR of each recorded epoch. The black line in each plot indicates a regression fit to the data with r^2 indicating the amount of variation explained by SNR and p indicating significance of the t-statistic comparing a linear to a constant model. Of all methods, only MLspike shows reduced performance at the SNR levels tested across this dataset. In particular, the precision (propensity to find false negatives) was increased in low SNR conditions.

Supplementary Figure 10: Comparison of matched and unmatched temporal differences across methods.



(**A**) Bias and uncertainty between imputed and ground truth spikes matched with a 100 ms. Each plot is a kernel density estimate over a boxplot of indicated metric for each imputed spike. (**B**) As **A** for all imputed spikes compared to the nearest ground truth spike. Note that the supervised methods show increased bias and uncertainty when removing double-counted imputed spikes by using matched statistics.

Discussion

In this study, we identified and characterized nonlinear distortions arising from spike activity patterns in fluorescence signals from three fast GCaMP variants, GCaMP6f, jGCaMP7f, and jGCaMP8f. We discovered a novel use-dependent slowing of the fluorescence decay during

536 periods of sustained activity. Using these empirical observations, we constructed a biophysical
537 model which served as the foundation of an unsupervised ($Biophys_{SMC}$) and a supervised
538 ($Biophys_{ML}$) spike inference method. These biophysically-based inference methods extracted
539 spikes with greater precision than existing methods. The advantage of biophysical model-based
540 approaches was most apparent under high-bandwidth conditions.
541 Our spike-decoding framework was composed of three basic parts. First, we made
542 recordings in a minimal tissue-based system (acute slices) where paired fluorescent/spike
543 recordings were as easy as possible to make. Second, we used in vitro recordings where all
544 time-varying parameters of experiments were known precisely, save the GCaMP response
545 itself. Finally, we used these recordings to jointly constrain models of the GCaMP response per
546 se and to suggest an appropriate generative model of the cellular response.
547 Our study clarifies how response nonlinearities can differ between indicators. Earlier
548 GCaMPs (6f and 7f) showed a facilitation-like nonlinearity in their fluorescence response lasting
549 tens of milliseconds, amplifying signals from closely timed spikes as previously noted (Éltes et
550 al., 2019; Wei et al., 2020; Huang et al., 2021; Schoenfeld et al., 2021). We found an additional,
551 novel nonlinearity: a persistent use-dependence that builds progressively during periods of
552 heightened activity. This second nonlinearity was particularly apparent for jGCaMP8f, whose
553 half-decay responses were slowed by up to three-fold. We showed that this effect is problematic
554 for unsupervised methods that assume a linear mapping of spikes to fluorescence. In the linear
555 SMC approach we used as a benchmark (PGBAR, (Diana et al., 2024)), elongated decay
556 dynamics drove spurious predictions as the hysteretic effect forced the latent states to match
557 the fluorescence dynamics (**Figure 3G, H**). This linear SMC, as well as MLspike (Deneux et al.,
558 2016), were particularly poor spike predictors in the context of low SNR, high spike-rate data
559 from inhibitory cells (**Figure 3H**; **Figure 4C**).
560 By leveraging a biophysically-based understanding of sensor behavior, we developed
561 two complementary, high-performance spike inference methods. Initially, we applied the
562 generative model directly within an unsupervised SMC framework to substantially improve
563 accuracy, albeit with a computational cost inherent to Monte Carlo-based approaches
564 (Vogelstein et al., 2010; Speiser et al., 2017). To address this limitation, we then used our
565 biophysical model to produce large, synthetic datasets to train a supervised neural network that
566 performed more quickly.
567 To our knowledge, our study is the first to report improvements in a machine learning
568 training using a generative physical model approach. Because the biophysical model captures
569 important features of GCaMP responses, we were able to use it as the engine for exploring how
570 GCaMPs would react under different cellular and activity contexts. We achieved substantial
571 improvements in spike time estimates by training our model on simulated spike patterns that
572 explored a wide range of timing and patterns. The size and diversity of our training may explain
573 why $Biophys_{ML}$ outperformed CASCADE trained on the jGCaMP8f data itself. While large, the
574 jGCaMP8f excitatory cell dataset still falls within a range over which increasing training set size
575 would be expected to improve performance (Rupprecht et al., 2021; Zhou et al., 2023). Indeed,
576 limited training data is a fundamental hurdle for any machine learning approach, a problem that
577 is particularly acute in the context of labor-intensive datasets such as paired
578 optical/electrophysiology considered here.

However, training of supervised neural networks, even on extensive, heterogeneous datasets, was insufficient to guarantee accurate inference for sensors with previously unseen properties. Both ENS^2 and CASCADE, which rank highly on established benchmarks, exhibited temporal biases and overestimated rise times for jGCaMP8f (**Figure 4**). Notably, in contrast to a published report (Zhou et al., 2023), CASCADE outperformed ENS^2 on all aspects of inference in the jGCaMP8f dataset. This result highlights that observations made with neural networks trained on broad datasets do not generalize in instances where indicator properties differ substantially from those in the training dataset. Taken together, these results suggest that for both current and future calcium indicator proteins, building an effective supervised inference approach requires a thorough exploration of relevant dynamic regimes using either labor-intensive ground truth or the generative approach we describe here.

In summary, this study elucidates the biophysical mechanisms underlying GECI fluorescence responses and offers practical solutions for more accurate and reliable neural activity detection. As a final note, our approach may serve as a generalizable roadmap for linking fluorescence measurements to underlying neural activity. For example, the steps we introduce here are likely to generalize to other GECI variants—including red-shifted (Inoue et al., 2015, 2019; Dana et al., 2016) and chemigenetic sensors (Deo et al., 2021; Farrants et al., 2024). Appropriate state and sensor models may allow extension to other genetically encoded sensors including those designed to detect neuromodulator release (e.g., Patriarchi et al., 2018; Sun et al., 2018) and voltage change (e.g., Evans et al., 2023; Platisa et al., 2023).

Methods

Animals.

The following strains were used for physiological experiments: C57Bl/6J (Jackson Laboratories stock 000664); CB6F1 (BalbC and C57Bl/6J F1); and Gabra6 (B6;129P2-Gabra6tm2(cre)Wwis/Mmucd). Both males and females were used and were between 40 and 130 days old. Animals were used in accordance with Princeton University's Institutional Animal Care and Use Committee and Institut Pasteur's CEEA-Paris1 approved protocols.

Protein purification.

GCaMP variants were cloned into the pET28b protein expression vector using GCaMP6f, 7f, and 7s backbones generously provided by Douglas Kim (Janelia) and Addgene (104488). Plasmids were transformed into IPTG-inducible BL21(DE3) competent E. coli (MilliporeSigma 694504). Starter cultures of transformed cells were grown in 5 ml LB medium with 50 mg/L kanamycin at 37°C for 4 hours. Cultures were then added to 1 L LB medium with 50 mg/L kanamycin and grown at 37°C until the OD at 600 nm reached 1.0. 1 mM IPTG was added and the incubation temperature dropped to 25°C overnight for optimal protein expression.

Liquid cultures were centrifuged at 6,000 relative centrifugal force (rcf) for 10 minutes. Pellets were resuspended in 25 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, pH 8, with 1 mM PMSF for protease inhibition. Cells were mechanically lysed through an Emulsiflex homogenizer (ATA Scientific). Cell debris was removed through centrifugation at 13,000 rcf for 45 minutes at 4°C and the supernatant bound to nickel-NTA agarose (ThermoScientific 25215). Protein was eluted with 25 mM Tris-HCl, 500 mM NaCl, 500 mM imidazole, pH 8. Proteins were further

621 concentrated using 10 kDa MWCO centrifugal filters (MilliporeSigma UFC901008) and desalted
622 (GE Healthcare 17085101) into 130 mM KMOPS (in mM:100 KCl, 30 MOPS, pH 7.2), or a
623 buffer mimicking intracellular conditions (in mM: 125 KCl, 14.5 KOH, 30 NaCl, 0.7 Mg2Cl, 10
624 HEPES) termed internal buffer as modified from (Tran et al., 2018). Protein concentrations were
625 determined by absorption spectrophotometry using alkali denaturation in 0.1M NaOH. Under
626 these conditions, mature GCaMP chromophore exhibits a standard molar absorption coefficient
627 of 44,000M-1cm-1 at 447 nm, allowing direct recovery of GCaMP concentration using Beer's
628 law (Ward, 2005; Barnett et al., 2017).
629

630 **Protein steady-state and kinetic characterization.**

631 Excitation and emission spectra were measured on a PTI Quantamaster 800 spectro-
632 fluorometer (Horiba Jobin Yvon Inc., Edison, NJ, USA) with data collection under control of Felix
633 GX software. For calcium-dependent fluorescence, steady-state measurements were made at
634 20°C using 0.15 μM of purified protein suspended in zero-Ca2+ buffer or high-Ca2+ buffer from
635 a calcium calibration kit (ThermoFisher C3008MP). The zero-Ca2+ buffer was reciprocally
636 diluted with the high-Ca2+ buffer to reach free Ca2+ concentrations between 0.01 and 10 μM.
637 Hill coefficient and Kd were estimated by fitting steady state data to the Hill equation.
638 The fluorometer was also used to prepare reagents for stopped-flow kinetic
639 measurements. Nominally Ca2+-free buffers included KMOPS or internal buffer supplemented
640 by 2 mM BAPTA (Invitrogen B1204). Increasing amounts of 1 M CaCl2 were added to create
641 free calcium steps of 0 nM, 50 nM, and increasing by approximately 2x free calcium in ten steps
642 to 10 μM free calcium. Indicators were diluted to 175 nM in 0.1mM solution buffered by 0.1 mM
643 BAPTA (see design considerations in **Supplementary Figure 2**).
644 Off-response kinetics buffers were designed to have a starting free calcium
645 concentration of 10 μM in 100 μM BAPTA and a final nominal concentration of 0 nM or 50 nM
646 free calcium when mixed to Ca2+-free buffer. Proteins were diluted to 35 nM in the 10 μM free
647 calcium buffer. In all cases, fura-2 (Molecular Probes F6799) and fura-4F (Invitrogen F14174)
648 were used to calibrate free calcium concentrations.
649 On-response and off-response kinetics were measured using an Applied Photophysics
650 SX20 stopped-flow spectrometer. Excitation light was provided by a xenon arc lamp directed
651 through a monochromator set to a 4 nm passband centered on 480 nm, while emissions were
652 collected through a 550 nm longpass filter (Newport FR-OG550). Dead time in the range of 0.6-
653 1.5 ms was estimated by extrapolation from measurements using OGB-1 (Life Technologies,
654 O6806). Measurements were made at 37°C with 60 μl per reaction chamber.
655

656 **Adeno-associated virus constructs.**

657 Adeno-associated viruses (AAVs) were prepared by cloning the GCaMP variants into
658 pAAV backbones (GenScript, Piscataway, NJ). Plasmids were incorporated into AAV serotype 1
659 (AAV2/1) and serotype DJ (AAV DJ) by the Princeton Neuroscience Institute Viral
660 Neuroengineering Laboratory.

661 **Stereotactic viral Injection.**

662 Viral vectors were injected stereotactically into the cerebellar vermis. Mice were deeply
663 anesthetized by intraperitoneal injections of 1.5% ketamine (Mérial) and 0.05% xylazine (Bayer).
664 2% Xylocaine (Newpharma) was applied to the cranial incision. 100 nl vermal (6.5 mm caudal to

bregma, lateral 0.2 mm, ventral 3.6 mm and 3.4 mm) and 100 nl Crus I (6 mm caudal to bregma, lateral 3 mm, ventral 0.5 mm) injections were infused over one minute. To allow for optimal transgene expression, mice were imaged 2 to 4 weeks after injection.

Slice preparation.

Animals were euthanized by rapid decapitation. The brains were quickly removed and placed in an ice-cold solution containing (in mM) 2.5 KCl, 0.5 CaCl2, 4 MgCl2, 1.25 NaH2PO4, 24 NaHCO3, 25 glucose, 230 sucrose, and 0.5 ascorbic acid. The buffer was bubbled with 95% O2 and 5% CO2. A Leica VT1200S vibratome was used to prepare coronal slices (200 µm), which were incubated at 32 °C for 30 min in (in mM): 85 NaCl, 2.5 KCl, 0.5 CaCl2, 4 MgCl2, 1.25 NaH2PO4, 24 NaHCO3, 25 glucose, 75 sucrose, and 0.5 ascorbic acid. The external recording solution contained (in mM) 125 NaCl, 2.5 KCl, 1.5 or 2.0 CaCl2, 1.5 or 1.0 MgCl2, 1.25 NaH2PO4, 25 NaHCO3, 25 glucose, and 0.5 ascorbic acid. Slices were maintained at room temperature for up to 6 h. All slice recordings were performed at 36–38°C.

Imaging in cerebellar brain slices.

Parallel fiber boutons expressing fluorescent virus were imaged with an Ultima two-photon scanning scanhead (Bruker Nano Surfaces Division, Middleton, WI, USA) that was mounted on an Olympus BX61W1 microscope equipped with a water-immersion objective (60×/1.1-NA; Olympus Optical, Tokyo, Japan) and infrared Dodt-gradient contrast. Two-photon excitation was performed with a pulsed Ti:Sapphire laser (DeepSee, Spectra-Physics, France) tuned to 920 nm. Extracellular parallel fiber axonal stimulation was performed with a constant voltage stimulator (Digitimer Ltd, Letchworth Garden City, UK) and a patch pipette (typically with a tip resistance of 4–6 MΩ) filled with artificial cerebrospinal fluid and placed in the molecular layer adjacent to labeled parallel fibers. The fluorescence of single cerebellar granule cells was monitored using circular line scans along the cytoplasm of the GC somata or straight line scans across individual parallel fiber boutons. Typical line rates were ~1 kHz. To determine threshold fluorescence, extracellular stimulation intensity was increased until clear fluorescence responses were observed after a train of 20–60 60-µs pulses at 100 or 300 Hz. To ensure action potential initiation for each stimulus pulse, the stimulus voltage was increased by 5 V above threshold.

Biophysical and cell models.

We modeled GCaMP fluorescence as arising from semi-independent ("hemi-independent") (Lai et al., 2015) calcium binding at two lobes, each of which can subsequently bind a target peptide to become fluorescent. Formally, each lobe transitions between an unbound and a calcium-bound state, which is governed by first-order binding and unbinding reactions:

$$\frac{d(\mathrm{bound_{lobe}})}{dt} = k_{\mathrm{on,lobe}}\,\mathrm{unbound_{lobe}} - k_{\mathrm{off,lobe}}\,\mathrm{bound_{lobe}} \qquad \text{eqn. 1}$$

where $k_{\mathrm{on,lobe}}$ and $k_{\mathrm{off,lobe}}$ are rate constants for binding and unbinding at each lobe. We defined:

$$k_{\text{on,lobe}} = \frac{k_{\text{off,lobe}}}{\left(K_{a,\text{lobe}}\right)^{H_{\text{lobe}}}} \qquad \text{eqn. 2}$$

where $K_{a,\text{lobe}}$ is an apparent dissociation constant and H_{lobe} is a Hill coefficient describing cooperative binding. Once a lobe is bound with calcium, it can transition to a fluorescent state by binding the target peptide. This step similarly follows first-order kinetics:

$$\frac{d(\text{fluorescent})}{dt} = k_{\text{on,peptide,lobe}}\,\text{bound}_{\text{lobe}} - k_{\text{off,peptide,lobe}}\,\text{fluorescent} \qquad \text{eqn. 3}$$

All fluorescent states (i.e., those in which a lobe is calcium-bound and has the peptide bound) contribute equally to total fluorescence. For stopped-flow simulations, model fluorescence was normalized between 0 (occupancy of fluorescent states at zero calcium) and 1 (occupancy at saturating calcium). For cell-based simulations, fluorescence was scaled to a dimensionless quantity F by setting to 1 at 0 calcium and the parameter R_f – the dynamic range of the indicator – at saturating calcium level. $\frac{\Delta F}{F}$ was then calculated as:

$$\frac{\Delta F}{F} = \frac{F\left(\text{Ca}_{\text{current}}\right) - F\left(\text{Ca}_{\text{basal}}\right)}{F\left(\text{Ca}_{\text{basal}}\right)} \qquad \text{eqn. 4}$$

where $F(\text{Ca}_{\text{current}})$ is the instantaneous normalized fluorescence at the current free-calcium level, and $F(\text{Ca}_{\text{basal}})$ is the steady-state fluorescence at basal calcium.

To simulate intracellular calcium transients (e.g., after action potentials), we extended the model from Rebola et al. 2019 (Rebola et al., 2019) by including an internal calcium reservoir with influx and efflux governed by Michaelis–Menten dynamics. Because cerebellar granule cells have high concentrations of fast, low affinity endogenous buffers, we linearized the buffer's binding capacity κ_B (Neher and Augustine, 1992) to reduce computational cost. Specifically, each small increment of free calcium Ca_{in} was instantly equilibrated with endogenous buffer:

$$\frac{d(\text{BCa})}{dt} = \frac{\text{Ca}_{\text{in}}\,\kappa_B}{\kappa_B + 1} \qquad \text{eqn. 5}$$

where BCa is the portion of endogenous buffer bound to calcium, a ratio of approximately 40 for our standard granule cell model. This linearization preserves the overall calcium balance while allowing faster numerical integration of high buffer concentration simulations.

For obtaining full GCaMP model parameterization we globally fit all traces obtained from stopped flow for each variant by gradient descent from 1500 quasi-randomly selected initial points in parameter space. GCaMP parameters from each run were then fixed and best fits were obtained across our ex vivo datasets by gradient descent over cell-specific parameters.

The GCaMP model parameters which minimized errors across experimental contexts were then selected and fixed for each GCaMP variant.

Bayesian and machine-learning spike inference algorithms.

In order to infer spike times from fluorescence time series using the biophysical model of jGCaMP8f described in the main text we adopted the Bayesian framework developed by Diana et al, 2024 based on sequential Monte Carlo methods (particle Gibbs with ancestor sampling, PGAS; (Lindsten et al., 2014; Diana et al., 2024) which enabled us to estimate simultaneously the posterior distributions of spike times and the parameters of the cell model, including the GCaMP concentration, calcium extrusion rates, peak amplitude of the calcium flux due to single action potentials and the dynamic range of the probe. Single point estimates of model parameters and spike times were obtained by using maximum-a-posteriori estimates.

We compared the performance of our method with MLSpike (Deneux et al., 2016) which employs a different model for the calcium indicator dynamics and baseline fluorescence. Spike inference in MLspike is carried out by discretizing all continuous dynamical variables in the model and applying the Viterbi algorithm to extract maximum-a-posteriori estimates of spike times. One of the weaknesses of this approach is the difficulty of estimating the model parameters, for which Deneux et al provide ad hoc calibration schemes. We directly addressed this issue using the PGAS sampling algorithm.

We compared our method with state-of-the-art supervised machine-learning models: CASCADE (Rupprecht et al., 2021) and ENS^2 (Zhou et al., 2023). The two approaches use different network architectures. CASCADE returns the spike probability at each time point by taking as input a local time window by using a trained convolutional network while ENS^2 uses the U-net applied on fixed-length segments of the original calcium trace.

Analysis of discrete action potential imputation.

Continuous prediction traces were converted to discrete imputed spike times using the approach indicated by each published method. For $Biophys_{SMC}$, spike predictions were calculated using the maximum a-posteriori of the posterior distribution of the spike state over time. For $Biophys_{ML}$, spike predictions were calculated by optimally fitting a set of Gaussian kernels with prior width and height determined by approximating the smoothing filter applied to ground truth for model training as in (Rupprecht et al., 2021).

To avoid double-counting of imputed spikes for a given ground truth spike, we employed the linear assignment algorithm to identify unique imputed/ground truth pairs. We used the matlab function matchpairs at a cost of unassignment as specified in the main text to constrain time interval between such pairs as specified in the main text.True positives were then defined as imputed spikes with unique ground truth matches, with unmatched ground truth spikes giving the false negatives, and unmatched imputed spikes being false positives. Total true positives, false positives, and ground-truth spikes within epochs for each cell to form the distributions for F-score, precision, and recall/specificity.

For each epoch of ground-truth recording, the bias in each imputed spike time was calculated as the interval to the nearest ground-truth action potential while uncertainty was defined as the absolute value of this quantity. The measures were taken to either the nearest

ground truth spike using a nearest neighbors search (unmatched) or to the nearest matched spike.

Statistical presentation.

Box and whisker plots comprise a box indicating the median and 25-75th percentile range and whiskers indicating the shorter of 1.5 times the interquartile range or the extreme data point. Descriptive statistics were calculated using median-based methods, and the estimated standard deviation was defined as the median absolute deviation divided by 0.6745.

References

Akerboom J et al. (2012) Optimization of a GCaMP calcium indicator for neural activity imaging. J Neurosci Off J Soc Neurosci 32:13819–13840.

Badura A, Sun XR, Giovannucci A, Lynch LA, Wang SS-H (2014) Fast calcium sensor proteins for monitoring neural activity. Neurophotonics 1:025008.

Barnett LM, Hughes TE, Drobizhev M (2017) Deciphering the molecular mechanism responsible for GCaMP6m's Ca2+-dependent change in fluorescence. PLOS ONE 12:e0170934.

Brown SE, Martin SR, Bayley PM (1997) Kinetic Control of the Dissociation Pathway of Calmodulin-Peptide Complexes*. J Biol Chem 272:3389–3397.

Chen T-W, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS (2013) Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499:295–300.

Dana H, Mohar B, Sun Y, Narayan S, Gordus A, Hasseman JP, Tsegaye G, Holt GT, Hu A, Walpita D, Patel R, Macklin JJ, Bargmann CI, Ahrens MB, Schreiter ER, Jayaraman V, Looger LL, Svoboda K, Kim DS (2016) Sensitive red protein calcium indicators for imaging neural activity. eLife 5:e12727.

Dana H, Sun Y, Mohar B, Hulse BK, Kerlin AM, Hasseman JP, Tsegaye G, Tsang A, Wong A, Patel R, Macklin JJ, Chen Y, Konnerth A, Jayaraman V, Looger LL, Schreiter ER, Svoboda K, Kim DS (2019) High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. Nat Methods 16:649–657.

Demas J, Manley J, Tejera F, Kim H, Traub FM, Chen B, Vaziri A (2021) High-Speed, Cortex-Wide Volumetric Recording of Neuroactivity at Cellular Resolution using Light Beads Microscopy. bioRxiv:2021.02.21.432164.

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

Deo C, Abdelfattah AS, Bhargava HK, Berro AJ, Falco N, Farrants H, Moeyaert B, Chupanova M, Lavis LD, Schreiter ER (2021) The HaloTag as a general scaffold for far-red tunable chemigenetic indicators. Nat Chem Biol 17:718–723.

Diana G, Sermet BS, DiGregorio DA (2024) High frequency spike inference with particle Gibbs sampling. eLife 13.

DiGregorio DA, Peskoff A, Vergara JL (1999) Measurement of Action Potential-Induced Presynaptic Calcium Domains at a Cultured Neuromuscular Junction. J Neurosci 19:7846–7859.

Éltes T, Szoboszlay M, Kerti-Szigeti K, Nusser Z (2019) Improved spike inference accuracy by estimating the peak amplitude of unitary [Ca2+] transients in weakly GCaMP6f-expressing hippocampal pyramidal cells. J Physiol 597:2925–2947.

Evans SW et al. (2023) A positively tuned voltage indicator for extended electrical recordings in

the brain. Nat Methods 20:1104–1113.
Faas GC, Raghavachari S, Lisman JE, Mody I (2011) Calmodulin as a direct detector of Ca2+ signals. Nat Neurosci 14:301–304.
Farrants H et al. (2024) A modular chemigenetic calcium indicator for multiplexed in vivo functional imaging. Nat Methods 21:1916–1925.
Friedrich J, Zhou P, Paninski L (2017) Fast online deconvolution of calcium imaging data. PLOS Comput Biol 13:e1005423.
Huang L, Ledochowitsch P, Knoblich U, Lecoq J, Murphy GJ, Reid RC, de Vries SE, Koch C, Zeng H, Buice MA, Waters J, Li L (2021) Relationship between simultaneously recorded spiking activity and fluorescence signal in GCaMP6 transgenic mice Westbrook GL, Svoboda K, Higley M, Sabatini BL, eds. eLife 10:e51675.
Inoue M et al. (2019) Rational Engineering of XCaMPs, a Multicolor GECI Suite for In Vivo Imaging of Complex Brain Circuit Dynamics. Cell 177:1346-1360.e24.
Inoue M, Takeuchi A, Horigane S, Ohkura M, Gengyo-Ando K, Fujii H, Kamijo S, Takemoto-Kimura S, Kano M, Nakai J, Kitamura K, Bito H (2015) Rational design of a high-affinity, fast, red calcium indicator R-CaMP2. Nat Methods 12:64–70.
Johnson JD, Snyder C, Walsh M, Flynn M (1996) Effects of Myosin Light Chain Kinase and Peptides on Ca2+ Exchange with the N- and C-terminal Ca2+ Binding Sites of Calmodulin (∗). J Biol Chem 271:761–767.
Lai M, Brun D, Edelstein SJ, Novère NL (2015) Modulation of Calmodulin Lobes by Different Targets: An Allosteric Model with Hemiconcerted Conformational Transitions. PLOS Comput Biol 11:e1004063.
Li P, Geng X, Jiang H, Caccavano A, Vicini S, Wu J (2019) Measuring Sharp Waves and Oscillatory Population Activity With the Genetically Encoded Calcium Indicator GCaMP6f. Front Cell Neurosci 13.
Lindsten F, Jordan MI, Schön TB (2014) Particle Gibbs with Ancestor Sampling. J Mach Learn Res 15:2145–2184.
Linse S, Helmersson A, Forsén S (1991) Calcium binding to calmodulin and its globular domains. J Biol Chem 266:8050–8054.
Pachitariu M, Stringer C, Harris KD (2018) Robustness of Spike Deconvolution for Neuronal Calcium Imaging. J Neurosci 38:7976–7985.
Patriarchi T, Cho JR, Merten K, Howe MW, Marley A, Xiong W-H, Folk RW, Broussard GJ, Liang R, Jang MJ, Zhong H, Dombeck D, Zastrow M von, Nimmerjahn A, Gradinaru V, Williams JT, Tian L (2018) Ultrafast neuronal imaging of dopamine dynamics with designed genetically encoded sensors. Science 360.
Platisa J, Ye X, Ahrens AM, Liu C, Chen IA, Davison IG, Tian L, Pieribone VA, Chen JL (2023) High-speed low-light in vivo two-photon voltage imaging of large neuronal populations. Nat Methods 20:1095.
Pnevmatikakis EA, Merel J, Pakman A, Paninski L (2013) Bayesian spike inference from calcium imaging data.
Rebola N, Reva M, Kirizs T, Szoboszlay M, Lőrincz A, Moneron G, Nusser Z, DiGregorio DA (2019) Distinct Nanoscale Calcium Channel and Synaptic Vesicle Topographies Contribute to the Diversity of Synaptic Function. Neuron 104:693-710.e9.
Regehr WG, Atluri PP (1995) Calcium transients in cerebellar granule cell presynaptic terminals. Biophys J 68:2156–2170.
Rose T, Goltstein PM, Portugues R, Griesbeck O (2014) Putting a finishing touch on GECIs. Front Mol Neurosci 7.
Rupprecht P, Carta S, Hoffmann A, Echizen M, Blot A, Kwan AC, Dan Y, Hofer SB, Kitamura K, Helmchen F, Friedrich RW (2021) A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging. Nat Neurosci 24:1324–1337.

Sabatini BL, Regehr WG (1998) Optical measurement of presynaptic calcium currents. Biophys J 74:1549–1563.

Sariyildiz MB, Kalantidis Y, Alahari K, Larlus D (2023) No Reason for No Supervision: Improved Generalization in Supervised Models.

Schoenfeld G, Carta S, Rupprecht P, Ayaz A, Helmchen F (2021) In Vivo Calcium Imaging of CA3 Pyramidal Neuron Populations in Adult Mouse Hippocampus. eNeuro 8.

Sebastian J, Mari Ganesh Kumar M, Sreekar YS, Rikhye RV, Sur M, Murthy HA (2017) GDspike: An accurate spike estimation algorithm from noisy calcium fluorescence signals. 2017 IEEE Int Conf Acoust Speech Signal Process ICASSP:1043–1047.

Seroussi I, Zeitouni O (2022) Lower Bounds on the Generalization Error of Nonlinear Learning Models.

Speiser A, Yan J, Archer EW, Buesing L, Turaga SC, Macke JH (2017) Fast amortized inference of neural activity from calcium imaging data with variational autoencoders. In: Advances in Neural Information Processing Systems. Curran Associates, Inc.

Sun F et al. (2018) A Genetically Encoded Fluorescent Sensor Enables Rapid and Specific Detection of Dopamine in Flies, Fish, and Mice. Cell 174:481-496.e19.

Sun XR, Badura A, Pacheco DA, Lynch LA, Schneider ER, Taylor MP, Hogue IB, Enquist LW, Murthy M, Wang SS-H (2013) Fast GCaMPs for improved tracking of neuronal activity. Nat Commun 4:2170.

Tian L, Hires SA, Mao T, Huber D, Chiappe ME, Chalasani SH, Petreanu L, Akerboom J, McKinney SA, Schreiter ER, Bargmann CI, Jayaraman V, Svoboda K, Looger LL (2009) Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. Nat Methods 6:875.

Tran V, Park MCH, Stricker C (2018) An improved measurement of the Ca2+-binding affinity of fluorescent Ca2+ indicators. Cell Calcium 71:86–94.

Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, Yuste R, Paninski L (2010) Fast Nonnegative Deconvolution for Spike Train Inference From Population Calcium Imaging. J Neurophysiol 104:3691–3704.

Ward WW (2005) Biochemical and Physical Properties of Green Fluorescent Protein. In: Green Fluorescent Protein, pp 39–65. John Wiley & Sons, Ltd.

Wei Z, Lin B-J, Chen T-W, Daie K, Svoboda K, Druckmann S (2020) A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. PLOS Comput Biol 16:e1008198.

Xiao D, Vanni MP, Mitelut CC, Chan AW, LeDue JM, Xie Y, Chen AC, Swindale NV, Murphy TH (2017) Mapping cortical mesoscopic networks of single spiking cortical or sub-cortical neurons Kleinfeld D, ed. eLife 6:e19976.

Zhang Y et al. (2023) Fast and sensitive GCaMP calcium indicators for imaging neural populations. Nature 615:884–891.

Zhou Z, Yip HM, Tsimring K, Sur M, Ip JPK, Tin C (2023) Effective and efficient neural networks for spike inference from in vivo calcium imaging. Cell Rep Methods 3:100462.