Spike inference from calcium imaging data with GCaMP8

Peter Rupprecht1,2,\*, Fritjof Helmchen1,2,3

1 Laboratory of Neural Circuit Dynamics, Brain Research Institute, University of Zurich, Switzerland  
2 Neuroscience Center Zurich, University of Zurich, Switzerland

3 University Research Priority Program (URPP), Adaptive Brain Circuits in Development and Learning, University of Zurich, Switzerland  
  
\* Corresponding author: [rupprecht@hifo.uzh.ch](mailto:rupprecht@hifo.uzh.ch)

Target journals: Journal of Neuroscience; Journal of Neuroscience Methods; eNeuro; J Physiology; eLife maybe; or even Nature Methods (“Analysis” article type), or Nature Communications

**Abstract**

Reproducible neuroscience relies on a set of widely used tools. However, to accurately interpret scientific results, it is necessary to thoroughly understand the features and limitations of these tools. Calcium imaging is such a widely used method to record neuronal activity, but the recorded signals are only indirect readouts of neuronal activity and therefore difficult to interpret. In this study, we evaluate the recently developed calcium indicator GCaMP8 and investigate how existing spike inference methods (CASCADE, OASIS, MLSpike) should be adapted for optimal performance with this new tool. We demonstrate that algorithms require fine-tuning, either through supervised training or parameter optimization, to obtain optimal results with GCaMP8 data, and we provide such pretrained models for supervised approaches (CASCADE) and the best-fit parameters for model-based approaches (OASIS, MLSpike). Supervised algorithms not specifically adapted for GCaMP8 resulted in degraded performance due to the introduction of undesired non-linearities and the biased processing of complex spiking events. Systematic benchmarking revealed that the slower variants of GCaMP8, as opposed to GCaMP6, GCaMP7f and GCaMP8f, were able to reliably detect isolated action potentials, as long as recording noise levels remained low. Furthermore, we show that, due to its fast rise times, GCaMP8 enables shorter closed-loop latencies for real-time detection of neuronal activity. Finally, we train spike inference algorithms specifically on interneuron ground truth data and find XXX. Together, our study offers critical insights into the toolset of calcium indicators, providing guidance for the design of future experiments and for the interpretation of existing datasets.

**Introduction**

During the last decades, systems neuroscience has advanced at a rapid pace. This progress has been driven by the astonishingly fast development of new tools for circuit neuroscience, such as large-scale recording techniques1 or optogenetics2. Calcium imaging is one of this toolset’s most powerful examples, ideally suited to study neural activity in living tissues at high spatial and temporal resolution3. Calcium imaging is based on a small molecular tool, a fluorescent calcium indicator. The last decades have seen first a shift from organic dyes to genetically encoded calcium indicators4, followed by a continuous refinement of these genetically encoded indicators to more closely reflect neuronal activity through fluorescence signals 5–8. GCaMP8, the latest iteration of the jGCaMP indicator series, was designed to significantly improve sensitivity and provide faster kinetics compared to its predecessors9.

However, it is not intuitively clear whether calcium imaging results obtained with GCaMP8 should be processed and interpreted differently compared to results obtained with previous indicators. For example, since recorded fluorescent signals are only an indirect proxy of neuronal activity, spike inference is an essential processing step that estimates the true spike rate from the these calcium signals10–12 and simultaneously helps to denoise recording13,14. Is it possible to use already established methods for spike inference that were optimized and benchmarked for previous calcium indicators, or do they need to be adapted? If so, how?

Furthermore, it was demonstrated that GCaMP8 sensors exhibit an increased linearity and faster rise times compared to previous indicator generations9. However, what is the consequence of these properties in practice? How do signals need to be interpreted differently from signals obtained with older generation sensors? For example, does the smallest detectable calcium transient for GCaMP8 reflect the same spike pattern as the smallest detectable calcium transient for GCaMP6? Forthermore, for which applications are there particular benefits from these specific properties of GCaMP8? Only with answers to such questions, results obtained with GCaMP8 can be interpreted accurately and with confidence.

In this study, we systematically address these questions. We investigate how spike inference generalizes from previous calcium indicators to GCaMP8. We explore the limitations and potential benefits of using GCaMP8 together with spike inference as a precise readout of neuronal activity from the point of view of the experimenter. We perform analyses across recordings conditions, covering excitatory and inhibitory neurons, and spanning a range of recording noise levels that are typical for population calcium imaging.

**Results**

**Generalization of spike inference algorithms to GCaMP8**

First, we wanted to compare and evaluate existing algorithms for spike inference when applied to GCaMP8. To enable this evaluation, we used the simultaneous calcium imaging and electrophysiological recordings from the same cells (“ground truth recordings”) that were provided with the original GCaMP8 study9. This data set (Fig. 1a-b) consists of a large number of ground truth recordings obtained from layer 2/3 pyramidal neurons of visual cortex in anesthetized animals during spontaneous neuronal activity or during visual stimulation (36 neurons for GCaMP8f, 42 neurons for GCaMP8m, 39 neurons for GCaMP8s, plus 22 neurons for GCaMP7f, after quality control, see Methods). We complemented this ground truth dataset focused on GCaMP8 with an existing compound dataset based on GCaMP6 (Fig. 1b, recordings from 125 neurons, described in ref. 14). The simultaneous recording of calcium and ground truth spikes in these datasets enables us to systematically evaluate how well an algorithm for spike inference performs (Fig. 1c). To render evaluations comparable across datasets, we resampled all ground truth datasets to a common frame rate (30 Hz unless otherwise indicated) and added Gaussian noise to achieve the same level of standardized noise level (see Methods for details). This procedure additionally enables us to adjust noise levels and generate low-noise recordings (standardized noise levels ~2) or high-noise recordings (noise levels ~8) from the same ground truth data.

We evaluated the performance of several standard approaches that are used as a proxy for spiking activity. First, we used the raw ∆F/F as a baseline for the benchmark. Second, we applied the OASIS algorithm, a simple and popular method implemented in Suite2p15 and CaImAn16 as the default method. Third, we used MLSpike17, which has been repeatedly shown to be the state-of-the-art model-based method for spike inference14,18. For the model-based methods (OASIS and MLSpike), we provide benchmarking results for the default parameters (see Methods), but in addition also analyze how to optimize model parameters in order to optimize spike inference with GCaMP8 data. Fourth, we used CASCADE14, a supervised method based on deep learning that was trained on a large and diverse set of calcium indicators, but with a focus on GCaMP6. This method has been shown to outperform other existing approaches and is hence called *Default CASCADE*. For the supervised method CASCADE, we take advantage of the GCaMP8 ground truth data to train specific models for GCaMP8 in general (hence called *GC8-tuned CASCADE*) or for each indicator separately (*Finetuned CASCADE* or, e.g., *GC8s-tuned CASCADE*). Using these default and finetuned models, we compare the performance to GCaMP8 calcium imaging data (Fig. 1d-e).

In general, all models performed well above the baseline set by raw ∆F/F both for high noise level of the calcium recordings (Fig. 1) and for lower noise levels (Fig. Suppl. 1-1). OASIS and MLSpike performed reasonably well with default parameters, but their performance could be improved by finetuning (Fig. 1e). This improvement was more modest for OASIS (from 0.72 ± 0.16 to 0.74 ± 0.16, median correlation with ground truth ± s.d. across 105 neurons), and more prominent for MLSpike (from 0.71 ± 0.19 to 0.81 ± 0.10), where more parameters can be optimized. The optimal performance was not highly sensitive on parameter choices for neither OASIS (Fig. Suppl. 1-2) nor MLSpike (Fig. Suppl. 1-3). Interestingly, the optimal decay time found for GCaMP8 (~0.5 s for OASIS and 0.3-0.8 s for MLSpike) were clearly larger than the measured decay time constant of the indicator (0.05-0.02 s, ref. 9). These best-fit values therefore do not directly reflect physiological properties and have to be treated as heuristic parameters.

Default CASCADE as supervised method (0.75 ± 0.08) performed better than the default version of either OASIS and MLSpike (p = 4.2e-5 and 9.8e-4, Wilcoxon signed-rank test), but was surpassed by the finetuned version of MLSpike (2.1e-15). However, a GC8-tuned CASCADE performed better than almost all previous models (0.79 ± 0.10, p = 0.06 for finetuned MLSpike, p < 1e-19 for all other comparisons). A dataset-specific finetuning of the model (Finetuned CASCADE; 0.84 ± 0.11) improved the performance significantly compared to finetuned MLSpike or GC8-tuned CASCADE (p = 7.2e-4 and 0.03) but only with small average effect sizes (Fig. 1e). A transfer-learning approach to better benefit from the small amount of ground truth of each GCaMP8 variant (see Methods) performed reasonably well (0.79 ± 0.09) but did not outperform other finetuning approaches (Fig. 1e).

The most striking finding is however that the already relative good performance of Default CASCADE improved significantly for finetuned MLSpike or for tuned CASCADE (Fig. 1e). However, these models trained with GC8 ground truth are not simply better for all use cases: they systematically performed better for GCaMP8 data but worse for pre-GCaMP8 indicator datasets (Fig. Suppl. 1-1c). These results demonstrate that spike inference for calcium imaging data with GCaMP8 benefits from optimized model parameters or specifically trained supervised models.

The quality of spike inference from two-photon calcium imaging data is limited not only by the indicator and the inference model but also by the recording quality and in particular the shot noise level (refs). To assess this depdencen, we performed all model evaluations both for low and high noise levels. As a result, we did not notice a major difference for the relative evaluation of the spike inference models (Fig. 1e, Fig. Suppl. 1-1b). However, we observed that isolated spikes that were rarely detected by CASCADE for high-noise recordings were more reliably detected for low-noise data (Fig. 1f), as one would expect. Furthermore, we noticed that spike inference resulted in a more visible improvement for noisier calcium imaging data. This improvement was reflected by an increase of the performance compared to the baseline performance (∆F/F), which was more pronounced for high noise levels (improvement ∆c = 0.24 ± 0.09 across neurons for low noise, 0.39 ± 0.11 for high noise) (Fig. 1e,g). This finding expands on the well-established observation that spike inference results in a reduction of noise13,14.

Altogether, we found that spike inference generalized well from pre-GCaMP8 to GCaMP8, but specific finetuning of these methods to GCaMP8 further improved the performance. To disseminate use of optimized algorithms, we provide the best-fit parameters for each GCaMP8 variant in the Methods section both for OASIS and MLSpike, and pretrained models for CASCADE online (<https://github.com/HelmchenLabSoftware/Cascade>).

**Recovery of absolute spike rates across firing frequencies**

In an ideal world, spike inference would be able to recover not only an estimate proportional to spike rate but also the correctly scaled absolute spike rate. Existing algorithms for spike inference typically struggle to do so when they are applied to datasets that are not part of the training data14 or do not even attempt to recover absolute spike rates and instead rely on “correlation with ground truth” as a scaling-invariant metric to measure algorithm performance18. CASCADE was designed to recover absolute spike rates for unseen datasets and, as a state-of-the-art algorithm, managed to recover absolute spikes with an error factor of ~2 for indicators prior to GCaMP8. That is, estimated spike rates were typically within 50% and 200% of the true spike rate (ref. 14). In this section, we investigate how well absolute spike rates can be recovered from GCaMP8 data across different neuronal firing rates.

To visualize predicted vs. true spike rates, we smoothed both ground truth and predicted spike rates with a temporal Gaussian filter (standard deviation, 1 s) and binned the resulting ground truth in spike rate bins with increments of 0.6 Hz. This enabled us to separately plot the true spike rate vs. ∆F/F (Fig. 2a) or predicted spikes (Fig. 2b-d) for each neuron, hence called the “transfer function” associated with this particular neuron and spike inference approach. A striking feature of these transfer functions, shown as an example for the GCaMP8s dataset, is the sigmoidal nonlinearity for Default CASCADE (Fig. 2b). This is a property that has already been observed for CASCADE when trained with GCaMP6f (Figure 4g in ref.14). The shape of this nonlinear transfer function is likely to reflect a trade-off to deal with the nonlinear properties of GCaMP6, although this explanation remains speculative at this point. Irrespective of the underlying reasons, it is clear that GCaMP8 does not require this trade-off since CASCADE trained with GCaMP8 ground truth is capable of recovering a more linear transfer function (Fig. 2c-d). This example demonstrates that CASCADE trained with GCaMP8-data is capable of remediating a distortion of the model’s transfer function across spike frequencies that could not be prevented for GCaMP6.

To systematcially quantify this property across datasets, we computed for each neuron and algorithm the “deviation from linearity” (Fig. 2e), which measures the relative amount of variation that is not covered by a linear fit and which therefore contributes, on average, to the nonlinearity as illustrated in Fig. 2b (see Methods for a strict definition). We then quantified this metric across models, enabling the comparison of the transfer curve linearity for all approaches. We found that OASIS, which features a linear model of calcium dynamics19, performed well with respect to this metric and did not increase the deviation from linearity compared to raw ∆F/F across most datasets (Fig. 2e; p > 0.05 for GCaMP6, GCaMP8f, GCaMP8m; reduction for GCaMP8s with p = 3.5e-7; Wilcoxon paired signed-rank test across; number of neurons n = 125, 36, 42, and 39 for the four datasets). MLSpike, on the other hand, resulted in a difficult-to-predict behavior, with increased nonlinearity compared to ∆F/F for GCaMP6 and GCaMP8f (p < 1.8e-12, p = 0.0087), but no significant increase for GCaMP8m (p = 0.73) and a decrease for GCaMP8s (Fig. 2e; p = 4.4e-6). Default CASCADE performed well for GCaMP6, for which it was designed (decreased nonlinearity, p < 1.0e-6), and did not significantly increase or decrease the nonlinearity for GCaMP8f (p = 0.54) but, consistent with Fig. 2b, introduced a striking nonlinearity to the transfer function for the other GCaMP8 variants (Fig. 2e; p = 4.9e-6 for GCaMP8m and p = 7.2e-8 for GCaMP8s). When retrained with GCaMP8, however, either generally with GCaMP8, or for the dataset’s specific GCaMP8 variant, CASCADE exhibited a substantial and consistent decrease of nonlinearity compared to ∆F/F (Fig. 2e; p < 1e-4 for all comparisons). Together, these results demonstrate that Default CASCADE, trained for GCaMP6, while making very good predictions about relative spike rates (Fig. 1), nonlinearly distorts the transfer curve. A CASCADE model retrained with GCaMP8 data was able to correct for this undesired behavior.

Next, we observed that transfer functions of individual neurons not only deviated from a linear relationship but also from the identity relationship (Fig. 2f), with the deviation from identity calculated as the difference of the slope of the linear fit, , from unity (see Methods; Fig. 2f). A positive value would indicate an overestimated spike rate, while a negative value would indicate, on average, an underestimated spike rate. For this quantification, we did not consider raw ∆F/F and the OASIS algorithm, both of which cannot be used to estimate absolute spike rates for GCaMP indicators. The results highlight a key difference between CASCADE models that were trained with all GCaMP8 datasets and CASCADE trained only with a specific GCaMP8 dataset, e.g. with GCaMP8s (Fig. 2f). The models trained with specific ground truth made relatively accurate predictions of spike rates across the firing frequencies (∆m ± s.e.m across neurons: -0.12 ± 0.04, -0.01 ± 0.04 and +0.12 ± 0.07 for the “f”, “m” and “s” variants). In contrast, models trained on all GCaMP8 ground truth underestimated spike rates for GCaMP8f and GCaMP8m (∆m = -0.39 ± 0.05 and -0.30 ± 0.03) and overestimated spike rates for GCaMP8s (+0.71 ± 0.10; Fig. 2f). Such a bias is not necessarily prohibitive since a spike rate estimate within a factor of <2 is still very useful. For example, a ∆m of +0.71 indicates a systematic bias to overestimate spike rates by ~71%, which seems acceptable in most applications. This bias for the general GCaMP8-based model might be compensated for by the increased robustness that has been shown to come with a larger and more diverse ground truth dataset for training an algorithm14. However, if the goal of the analysis is to recover absolute spike rates as precisely as possible, our analyses clearly suggest to use a CASCADE model trained on this specific GCaMP8 variant.

**Spike rate estimation for complex events with GCaMP8**

Next, we investigated how specific spike patterns were recovered by GCaMP8 together with spike inference, in order to understand the differences seen by different versions of CASCADE (Fig. 1). To this end, we visually screened inferred spike rates together with ground truth in order to systematically detect conditions under which Default CASCADE performed worse than GC8-tuned CASCADE. We specifically noticed such cases during complex events, i.e., when multiple spikes and bursts were fired within a short time window. For illustration, we selected such examples from neurons of each GCaMP8 variant (Fig. 3a). For these complex events, Default CASCADE tended to overestimate the spike rate for the initial spikes and to underestimate the spike rate for subsequent spike patterns. These examples suggest that spike inference makes systematic mistakes when a model trained on pre-GCaMP8 data is applied to data recorded with the new GCaMP8 sensor variants. In line with that, these systematic mistakes were largely corrected when using GC8-tuned CASCADE (Fig. 3a).

To confirm this initial observation more systematically, we detected complex events with high spike rates based on a simple threshold (instantaneous smoothed ground truth spike rate >6 Hz). Then, we extracted the raw ∆F/F trace and inferred spike rates in a time window centered around the maximum true spike rate. We performed this analysis both for data based on GCaMP8 and GCaMP6 from our joint ground truth database (Fig. 1b). To quantify a potential bias towards the initial vs. late phase of such complex events, we computed the difference between normalized true and inferred spike rate. For example, for the GCaMP8s dataset, application of Default CASCADE resulted in a clear spike rate overestimation of the initial phase and an underestimation of the later phase of the complex event (Fig. 3b, based on 1218 events across 37 neurons), and these biases were eliminated when applying GC8-tuned CASCADE (Fig. 3b). Interestingly, the opposite observation could be made for GCaMP6-based datasets, where Default CASCADE resulted in unbiased recovery of complex event spike rates, while GC8-tuned CASCADE resulted in an overestimation of late phase-spike rates (Fig. 3c, based on 329 events across 23 neurons). To summarize this finding across datasets, we computed for each difference curve the area under the curve after () and before () the zero time and computed the normalized difference as a skewness metric, . Default CASCADE resulted in a negative bias for all GCaMP8 datasets (-0.40 ± 0.03, mean ± s.d. across the 3 datasets) and no visible bias for GCaMP6 data (-0.05 ± 0.06 across the 5 datasets; Fig. 3d). Conversely, GC8-tuned CASCADE resulted in no visible bias for GCaMP8 (-0.05 ± 0.11) but in a positive bias for all GCaMP6 datasets (0.34 ± 0.11; Fig. 3e). Interestingly, the calcium indicator GCaMP7f positioned itself somewhat between GCaMP6 and GCaMP8 (Fig. 3d-e). These systematic analyses show that the distinct properties of GCaMP6 vs. GCaMP8 specifically affect the analysis of large and complex events.

What could be the mechanism that underlies the distinct behavior of these two groups of calcium indicators? To address this question, it is essential to understand the effects of the supralinear nonlinearity displayed by calcium indicators such as GCaMP64. This supralinearity can result in a history-dependent amplification, where initial spikes might evoke only a small fluorescence signal but subsequent spikes a much higher one, boosted by the already bound calcium from the first spike (Fig. 3f). To provide an intuition of this suprelinearity, we demonstrate the effect in sigmoidal20 and biophysical21 models of GCaMP nonlinearities (Fig. 3g; see Methods). Default CASCADE compensates for this nonlinearity by reducing the expected spike rate underlying such supralinearly amplified events. This compensation results in­ a spike rate around complex events that is unbiased despite the sensor’s supralinearity (Fig. 3c). The GCaMP8 sensor variants, however, were designed to exhibit a lower degree of nonlinearity9, and a spike inference algorithm optimized for GCaMP6 will instead introduce a *sub*linear behavior. Our analyses therefore demonstrate the need for a GCaMP8-specific optimization of spike inference in order to prevent the biased estimation of neuronal activity during complex events.

**Single-spike detection with GCaMP8**

A question that has received wide attention is whether a calcium indicator or an algorithm can recover single action potentials. It was often claimed that single action potentials *can* be resolved using a specific algorithm or a specific indicator6,7,10. But it often turned out that either the conditions under which single action potentials were recovered could not be achieved with typical experiments22–24, and that large and detectable action potential-triggered calcium transients were observed in only a small and not necessarily representative subset of neurons14,25,26. The potential detection of single action potentials is important because these responses are an intuitive benchmark that complements other metrics (cf. Fig. 1). In addition, it is important for the interpretation of neuronal activity whether the underlying electrical signals are in reality bursts or isolated spikes27. Resolving this issue would make neuronal calcium imaging more interpretable and would help to tackle fundamental questions about the nature of the neuronal code. We therefore addressed the question in a systematic and comparative approach: can we detect single action potentials for GCaMP8?

Detection of single spikes is always a trade-off between the amount of false positive vs. false negative detections. For cortical recordings, neuronal silence is more common than activity, which can lead to very frequent false positive detections if the algorithm is not balanced. In addition, it is not fully clear how the arbitrary trade-off between false positive and false negative detection rates from traditional approaches22,24 will translate into practical applications. To overcome this trade-off, we used CASCADE which has been trained in a supervised manner to minimize the mean squared deviation from ground truth for the entire recordings, therefore taking into account both silent and active periods and their respective frequency. The single-spike detection ability of such an algorithm, which is not specifically optimized for single-spike detection, is therefore a benchmark that provides realistic performance metrics.

To assess detection of isolated action potentials from the existing ground truth datasets (Fig. 1b), we detected isolated single action potentials and the surrounding calcium traces (see Methods), added a variable amount of Gaussian recording noise to the calcium traces and inferred spike rates using a CASCADE model trained for this dataset (see Methods). Strikingly, we found that single spikes were almost never detectable with GCaMP6f, rarely with GCaMP6s, GCaMP7f, GCaMP8f, but relatively well for GCaMP8m and GCaMP8s as long as noise levels were low enough (Fig. 4a-c; Fig. 4-1). As expected noise levels increased, single action potentials were detected less reliably also for GCaMP8m and GCaMP8s (Fig. 4a-c; Fig. 4-1).

Next, to quantify these initial observations, we quantified the inferred spike rate for each single-action potential event and used a kernel density estimate to obtain the distribution of inferred spike rates for a single true spike. Importantly, we found that this distribution was clearly centered around zero for GCaMP6f, GCaMP6s, GCaMP7f and GCaMP8f, and this tendency was increased for high noise levels (Fig. 4d), reflecting the inability to detect single isolated spikes. Notably, the datasets exhibited slightly different abilities to pick up single action potentials even for the same calcium indicator, most clearly seen for GCaMP6s (Fig. 4-2). In any case, our results demonstrate that single spike-evoked calcium transients cannot be distinguished from noise under these conditions. As a consquence, the supervised algorithm decides to make the conservative estimate of “0” spikes. Interestingly, this situation changes for GCaMP8m and GCaMP8s, where the distribution of inferred spikes for an isolated single true spike shifted towards “1” spike (Fig. 4d). For GCaMP8m, the distribution returned to the “0” bin for intermediate noise levels, while this shift to the “0” bin occurred only for the highest noise levels for GCaMP8s.

To summarize these results, we quantified the fraction of isolated APs that can be detected as a single AP (inferred spike rate > 0.5 APs and < 1.5 APs), and the fraction of APs than can be detected at all as an event (inferred spike rate > 0.5 APs) across different noise levels. As expected from the distributions shown in Fig. 4d, GCaMP8s was the best available indicator, with a fraction of correct detections of > 80% and a fraction of detected events of close to 100% for the lowest noise levels (Fig. 4e,f). GCaMP8m also exhibited high detection rates that were, however, clearly lower than for GCaMP8s, followed with some distance by GCaMP8f, GCaMP7f, GCaMP6s and, with a detection fraction of close to 0% across all noise conditions, GCaMP6f. The calcium indicator XCaMP-Gf deserves a special mention. Although the available dataset is relatively small especially for low noise levels (see Suppl. Table 1), single APs could be detected with much better reliability than by GCaMP6 variants or GCaMP7f, probably reflecting the more linear behavior of the XCaMP indicator family8.

So far, all single AP-detection was quantified across isolated action potentials that had been pooled across all neurons. To exclude a bias introduced by a large number of isolated APs from a single neuron, we averaged effects of isolated APs from the same neuron and then visualized the performance and variability across neurons (Fig. 4g,h). These quantifications confirm the outstanding performance of GCaMP8s for the detection of isolated action potentials, and the relatively good performance of GCaMP8m and XCaMP-Gf. The analysis also demonstrates that data obtained with GCaMP6f were largely unable to recover isolated single action potentials. Together, these analyses dissect and compare the abilities of calcium indicators to detect single action potentials and therefore provide important guidance for the interpretation of calcium signals from these indicators.

**Online spike inference with fast rise times**

A striking feature of the GCaMP8 indicator family is that it reduced the rise time kinetics by a large margin compared to previous calcium indicators9. Here, we systematically study the relevance of this effect and analyze its potential benefits for calcium imaging and spike inference. First, we used linear Lucy-Richardson deconvolution to extract the fluorescence kernel evoked by the average action potential (Fig. 5a). This linear kernel is not the fluorescence triggered by isolated action potentials but of all action potentials from both complex and isolated events (see Methods). From this extracted kernel, we determined the rise time (half-maximum rise time) for each neuron across datasets. As expected, we observed low-latency rise times for GCaMP8 variants (6.4 ± 2.0 ms for GCaMP8f, 5.8 ± 1.4 ms for GCaMP8m, 7.4 ± 2.6 for GCaMP8s; median ± s.d. across neurons) and a slightly higher latency for GCaMP7f (17 ± 3 ms) (Fig. 5b). For GCaMP6, however, the results need to be unpacked more carefully. In agreement with previous analyses9, the AAV-mediated GCaMP6 variants had relatively high rise times (56 ± 9 ms for GCaMP6f and 75 ± 9 ms for GCaMP6s). However, datasets based on transgenic expression of GCaMP6f and, to a lesser extent, also GCaMP6s, resulted in surprisingly low rise times (20 ± 5 and 17 ± 3 ms for GCaMP6f datasets, 37 ± 10 ms for GCaMP6s). It seems likely that this obsevation can be explained by the lower expression levels for transgenic as opposed to AAV-based strategies, resulting in less calcium buffering and therefore faster calcium kinetics (ref). Interestingly, another dataset based on transgenic expression of GCaMP6f in zebrafish resulted in relatively long rise times (90 ± 19 ms), potentially due to the different temperature regimes for these two recordings (room temperature for zebrafish vs. body temperature for mice). These analyses show that rise times are significantly lower for GCaMP8 but are also affected by other factors like expression strategies, calcium buffering and temperature.

We reasoned that the fast rise times of GCaMP8 could be particularly advantageous for low-latency closed-loop processing. In such experiments, calcium signals are recorded and processed in real-time, enabling to use feedback signals that from the calcium traces of a defined set of neurons. Such paradigms have been applied to study learning and neuronal plasticity, but are also key for the development of brain-machine-interfaces (refs). A primary limitation of these feedback loops is the latency from the action potential to the feedback signal derived from this action potential. This latency involves the spike-induced calcium influx, the binding to the fluorescent calcium sensor, fluorescence imaging with a microscope, extraction of the fluorescence trace from the recorded data, and an algorithm to detect spiking from the extracted trace, for example through spike inference. Ideally, such a feedback loop should have a latency of <50 ms. GCaMP8 could reduce this latency because its fast rise time might enable a detection of spike events with lower latency. We therefore systematically analyzed calcium indicators and compared to which extent the short rise times of GCaMP8 will be beneficial for online spike inference.

For each indicator-specific dataset (Fig. 1b), we trained multiple CASCADE models that had access to a variable amount of time points after the current time point of interest, which we term “integration time” (see Methods; Fig. 5c). These CASCADE models were trained for a relatively high sampling rate of 60 Hz to enable comparisons at a high temporal resolution. Then, we quantified the performance of these models (correlation with ground truth), which resulted in a sigmoidal curve as a function of integration time, with performance increasing with integration time (Fig. 5c). We quantified the integration time necessary to recover 90% of the optimal performance (Fig. 5c). For low-noise calcium imaging data (standardized noise level of “2”), GCaMP8 data exhibited the lowest 90% integration times (36 ± 9 ms for GCaMP8f, 32 ± 6 ms for GCaMP8m, 37 ± 9 ms for GCaMP8s; median ± s.d. across neurons), closely followed by GCaMP7f (50 ± 7 ms) and transgenically expressed GCaMP6 (61 ± 10 ms and 84 ± 46 ms for the GCaMP6f datasets, 90 ± 68 for GCaMP6s), and, finally, by AAV-induced GCaMP6 (126 ± 45 ms for GCaMP6f, 188 ± 97 ms for GCaMP6s). These 90% integration times reflect the ranking of indicator rise times (Fig. 5d, left panel; cf. Fig. 5b). Interestingly, no clear advantage of GCaMP8f compared to the slower variants of GCaMP8 can be seen. For higher noise levels of calcium recordings, the required integration times increased and diminuished the advantage of GCaMP8 compared to the faster variants of GCaMP7f and GCaMP6f (Fig. 5d, middle and right panels). For example, at a standardized noise level of “8”, which is more typical of large-field of view recordings with fewer pixels per neuron, integration times around 70 ms were necessary for the GCaMP8 variants as well as for GCaMP7f (Fig. 5d, right panel). It seems reasonable to assume that the 90% integration time for GCaMP8 can, in principle, be reduced to even lower values than shown in Fig. 5d (left panel) when calcium recordings are performed with lowest noise. Altogether, our results show that online spike inference will directly benefit from the short rise times of GCaMP8, but also can take advantage of the faster rise times of the lower level of calcium indicators obtained via transgenic expression.

**Spike inference for interneurons**

*To be done.*

**Discussion**

The usefulness of new methods and tools depends not only on their quality but also on how well they are validated and understood. In this study, we investigated the practical use of the calcium indicator GCaMP8 and analyzed how algorithms for spike inference optimized for previous generations of calcium indicators should be adapted for use with GCaMP8. Together, this study provides an in-depth and comparative characerization of GCaMP8. As a result, we share guidelines, caveats, and pretrained models for spike inference with GCaMP8.

*The role of sensitivity*

*Comparison across indicators: limitations of available datasets and their respective recording conditions.*

*The role of non-linearity*

*The role of neuron-to-neuron variability. Autocalibration approaches.*

*Availability of models and parameter sets.*

(Snippet to be used for Discussion:)

As any tool, the calcium sensor GCaMP8 is only as powerful as its user. Only when we know the precision, strengths and weaknesses of a method, then we can use it with confidence and accurately interpret the results.

**Methods**

**Ground truth data sets and quality control**

Ground truth was extracted from publicly available datasets and quality-controlled for each neuron as described previously14. Specifically, ground truth for virally induced GCaMP6 expression was obtained from the primary publication related to the introduction of GCaMP6 (ref. 6). Ground truth for transgenically expressed GCaMP6 was obtained for mice from a dedicated publication from the Allen Institute24 and from the CASCADE publication for zebrafish14. All these datasets and associated quality controls have been described previously14.

For ground truth recorded with the three GCaMP8 variants, as well as ground truth recorded with GCaMP7f and XCaMP-Gf, we used the openly accessible data published with the primary publication related to the introduction of GCaMP8 (ref. 9). We performed additional quality controls as we did for previous similar analyses14 and excluded ground truth recordings from neurons that exhibited excessive motion artifacts, too high spike rates (indicative of unidentified interneuron cell types), or that did not allow the accurate isolation of the electrophysiological spikes from the recorded neuron. These cleaned-up datasets for XCaMP-Gf, GCaMP7f, GCaMP8f, GCaMP8m and GCaMP8s are available in addition to previously uploaded datasets in a format that can be accessed both in MATLAB and Python (datasets DS#28-#32 on <https://github.com/HelmchenLabSoftware/Cascade>).

**Standardized noise levels and resampling of ground truth**

Standardized noise levels were obtained exactly as previously described14,25. Briefly, the median absolute fluctuation of *ΔF/F* between adjacent timepoints was computed, normalized by the square root of the imaging frame rate *fr*:

When computed for ΔF/F data, *ν* is quantitatively comparable across datasets, even when frame rates differ, hence the name “standardized noise levels”. The units for *ν* are %·Hz−1/2, which we omit in the main text for readability.

To obtain ground truth for well-defined and comparable conditions, we used the same procedures as previously described14. Briefly, ground truth data were first temporally resampled to the desired sampling rate using the *signal.resample()* function from the Python package SciPy28. Then, to obtain a target noise level *ν*, Gaussian noise was added until the noise metric *ν* yielded the desired level. We have previously shown that such addition of Gaussian noise is, for this purpose, a sufficient approximation of typical Poisson-distributed shot noise14.

**OASIS for spike inference**

For spike inference with Default OASIS, the Python implementation of the algorithm in Suite2p15 was downloaded from <https://github.com/MouseLand/suite2p> and used in Python 3.7 with default parameters. For Finetuned OASIS, grid search was performed for the time constant parameter (Fig. Suppl. 1-3) for each dataset. The value for the time constant parameter that maximized the correlation with ground truth spike rates (median across neurons) was used for all neurons from the respective dataset.

**MLSpike for spike inference**

The MLSpike algorithm was downloaded from <https://github.com/MLspike/spikes> and used within Matlab 2022b (ref. 29). To optimize the performance for Default MLSpike (that is, a version of MLSpike that does not explicitly require ground truth), the inverse frame rate *dt* as a parameter of the algorithm was fixed to the value constrained by the ground truth. The *drift* parameter was set to 0.1. The nonlinearity parameter was set to 0.1. The hill coefficient was set to 1.84 for GCaMP6s, 2.99 for GCaMP6f, 3.1 for GCaMP7f, and 2.08/1.92/2.2 for GCaMP8f/m/s, as reported in previous publications. The parameters *amplitude* (amplitude of a single action potential in ∆F/F, in %), *tau2* (rise time constant, in milliseconds) and *tau* (decay time constant, in milliseconds) were set to 0.113, 0.034, 0.121, 0.294, 0.511 and 0.576 for *amplitude*, 70.2, 15.6, 15.3, 2.96, 3.31 and 4.72 for *tau2*, and 1870, 760, 129, 57, 107, and 267 for *tau* (for GCaMP6s, GCaMP6f, GCaMP7f, GCaMP8f, GCaMP8m and GCaMP8s). This version of MLSpike is considered “Default” because it takes the physiological values from the relevant papers for GCaMP6 (ref. 29) and GCaMP7-8 (ref. 9).

For Finetuned MLSpike, the same initial parameters were used, but a 2D grid search was performed to obtain the optimal parameters for *amplitude* and *tau* (Fig. Suppl. 1-2). Grid search was performed for each dataset and the best performance-parameters as quantified by correlation with ground truth were used for all neurons of this dataset. Due to long processing times with MLSpike, no out-of-dataset generalization, as it was done for CASCADE, was performed with MLSpike. Furthermore, a parameter search in higher-dimensional space, for example by including rise times or nonlinearity parameters, might result in even better performance for MLSpike but are prohibitive due to the slow processing speed of MLSpike.

**Spike inference with CASCADE**

For spike inference with Default CASCADE, pretrained models together with the algorithm were downloaded from <https://github.com/HelmchenLabSoftware/Cascade>. These models had been trained on a large database of excitatory neurons across different brain areas with a focus on cortical recordings using the GCaMP6 indicator (named “global CASCADE models” in ref. 14). Here, these models are called Default CASCADE because they were not retrained using GCaMP8 ground truth.

To optimize CASCADE for GCaMP8 ground truth, several approaches were tested. First, CASCADE was retrained from scratch on all GCaMP8 data (covering all variants, “f”, “m” and “s”). These models are called GC8-tuned CASCADE in this study. Second, CASCADE was retrained from scratch on a specific dataset, e.g., GCaMP8f only. These models are called, e.g., GC8f-tuned CASCADE or Finetuned CASCADE (akin to Finetuned MLSpike mentioned above). Third, CASCADE was initially trained from scratch on the same datasets as the Default CASCADE models, but then the model was retrained with a transfer-learning approach30 with dataset-specific ground truth while freezing the weights for all but the last layer. This procedure benefits from convolutional filter weights obtained by fitting a large and diverse ground truth, and fine-tuning the weighting of these filters for the specific ground truth dataset. Similar transfer learning procedures have been used successfully in neuroscience applications to reduce the amount of required training data31,32.

In all instances of retraining, CASCADE networks consist of a standard convolutional network with six hidden layers, including three convolutional layers. The input consists of a window of 64 time points (32 time points for frame rates <15 Hz), symmetric around the time point for which the inference was made. The three convolutional layers have relatively large filter sizes (31, 19 and 5 time points; 17, 9 and 3 time points for frame rates <15 Hz), with an increasing number of features (20, 30 and 40 filters per layer), with max pooling layers after the second and third layer, and a densely connected hidden layer consisting of ten neurons as the final layer. To avoid any effect of overfitting on our results, the same neuron was never used both for training and testing of a model. For example, if a model was tested for GCaMP8f data, separate models were trained for each test neuron of the GCaMP8f data while excluding during training the neuron tested for (“leave-one-out” strategy). This cross-validation strategy prevents fitting of test data and enables us to test the generalization of the algorithm to unseen data.

Models retrained with GCaM8 in general or for specific GCaMP8 datasets at various sampling rates and for a broad range of noise levels are already integrated into CASCADE (<https://github.com/HelmchenLabSoftware/Cascade>); further models tailored towards special use cases can be readily requested as described on the FAQ of the Github page.

**CASCADE for online spike inference**

For the investigation of online spike inference, specific new CASCADE models were trained for each dataset and each “integration time”. Integration time is here defined as the number of data points in the future from the current time point of inference that can be seen by the spike inference algorithm (Fig. 5c). For example, standard spike inference with CASCADE uses a symmetrical window of 64 time points, with 32 before and 32 after the time point of inference. For online spike inference, the number of time points after was decreased and the number of time points before was increased in order to maintain a 64-point window.

**Quantification of spike inference performance**

Performance of spike inference was quantified as described previously14,25. First, ground truth spike rates used for training and evaluation were generated from discrete ground truth spikes by convolution with a Gaussian smoothing kernel. The precision of the ground truth was adjusted by tuning the standard deviation of the Gaussian smoothing to the temporal sampling rate (σ = 0.05 s for 30 Hz recordings and σ = 0.025 s for 60 Hz recordings). Next, this smoothed ground truth spike rate was compared to the inferred spike rate using Pearson’s correlation. Additional metrics were used for specific analyses as described in the main text.

**Analysis of deviations from linearity and from identity**

To compute the deviation from linearity (Fig. 2e), the mean normalized squared deviations of the inferred spike rate, , from the linear fit to the inferred spike rate, , were computed. These deviations were averaged across spike rate bins of the transfer function (bins *k* from 1 to N in the equation below) and then converted to regular normalized units by taking the square root. This metric corresponds to the mean square root error but is normalized to obtain relative deviations from linearity.

To compute the deviation from identity for the transfer function of true spike rate to inferred spike rate (Fig. 2), a linear fit was computed, with the true spike rate , the inferred spike rate , and the slope of the linear transfer function fit, . The deviation from identity was defined as:

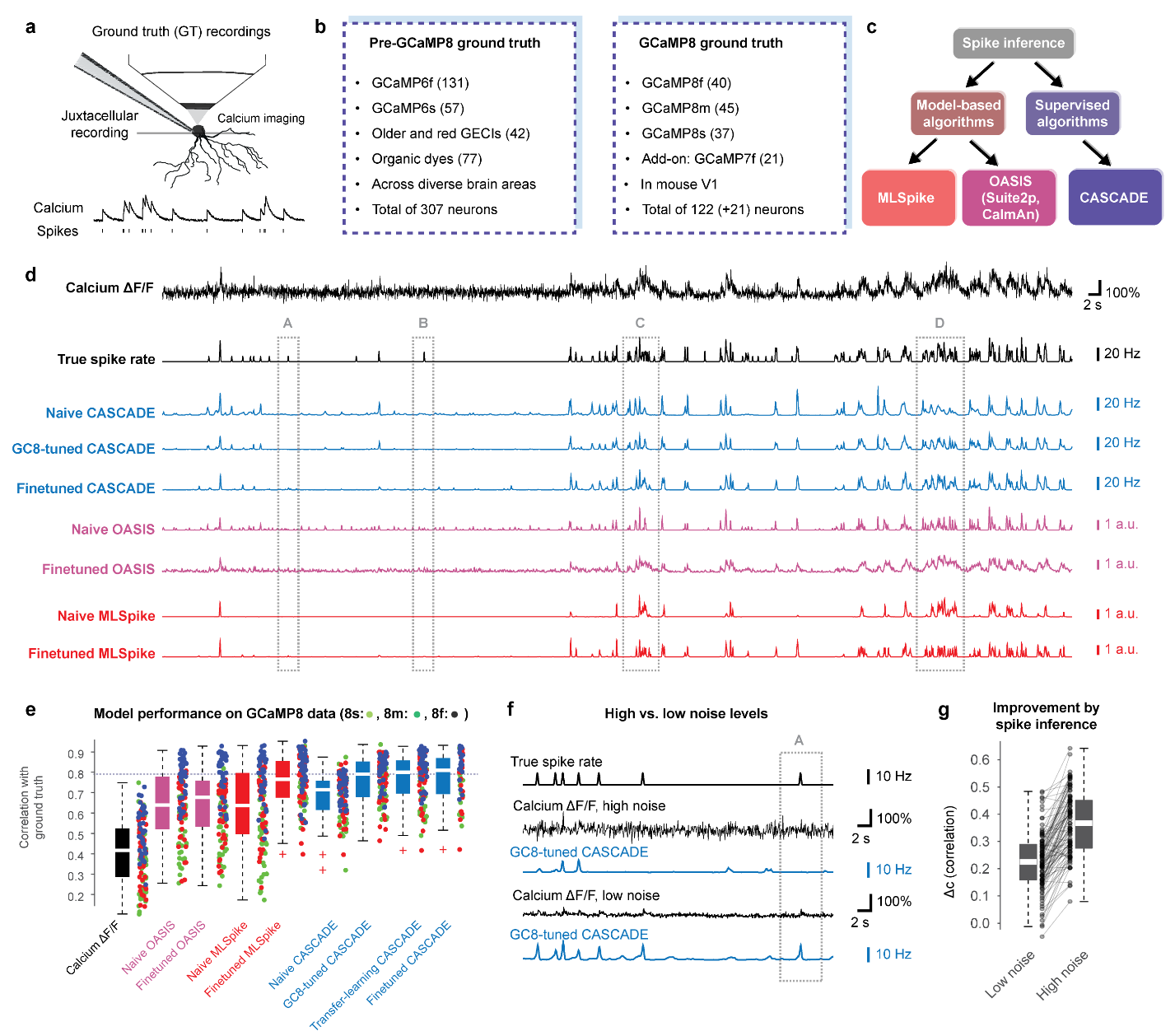
**Analysis of complex spike events**

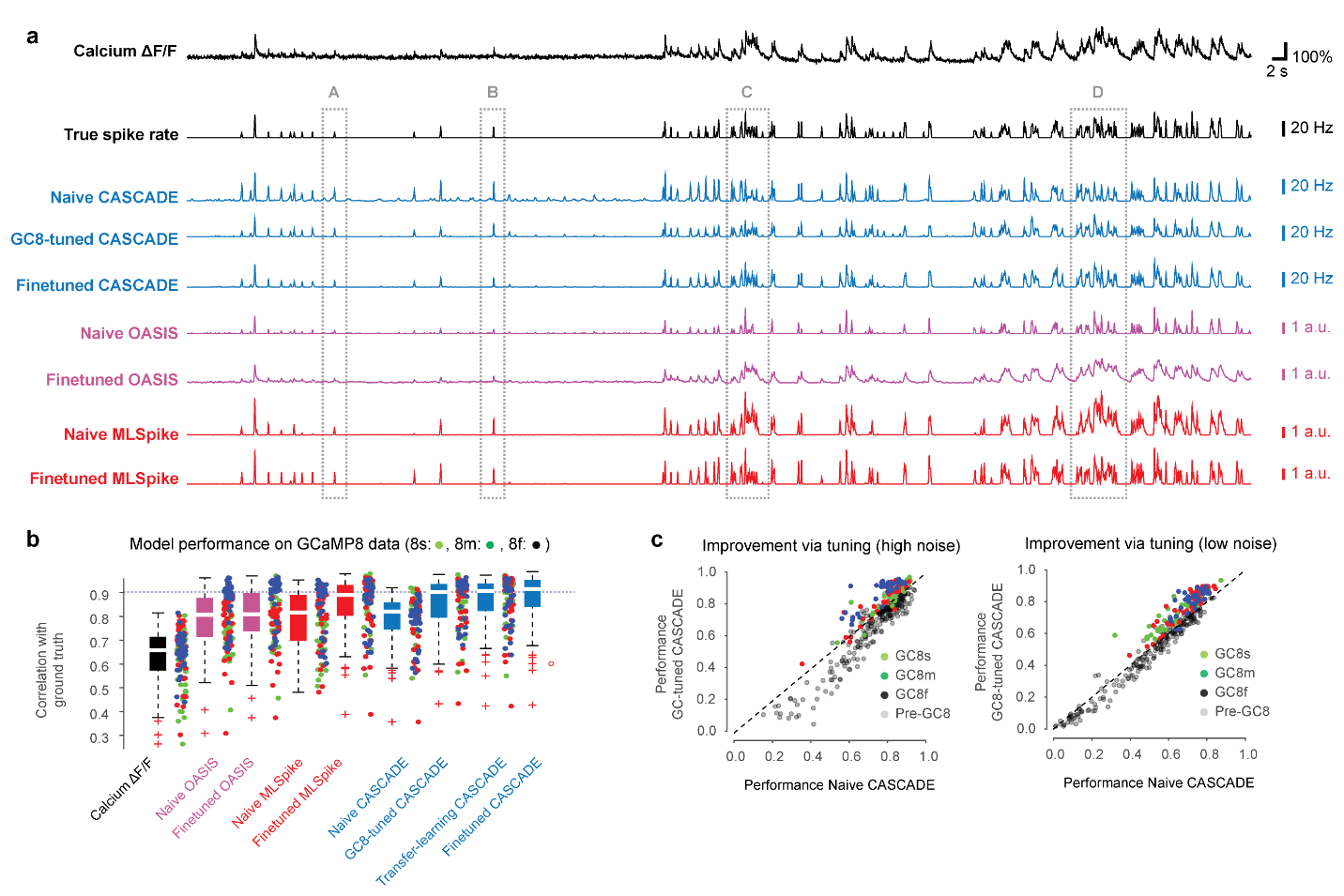
For detection of complex spike events, ground truth spiking binned to 30 Hz was smoothed with a 35-point running box filter. Then, all contiguous time windows with the smoothed ground truth spike rate >6 Hz were labeled using the Matlab function *regionprops()* and defined as events. For such events, the central time bin of each supra-threshold time window was retrieved. 50 time points before and 50 points after the central point were extracted as a local excerpt for this event. The goal of this procedure was to define events where a relatively large number of spikes occurred within a short time window.

The number of such extracted events across neurons were, with dataset ID in brackets: 92 (DS#9, GC6f, virally induced), 329 (DS#10, GC6f, transgenic), 281 (DS#11, GC6f, transgenic), 174 (DS#13, GC6s, transgenic), 36 (DS#14, GC6s, virally induced), 265 (DS#29, GC7f), 712 (DS#30, GC8f), 1070 (DS#31, GC8m) and 1218 (DS#32, GC8s).

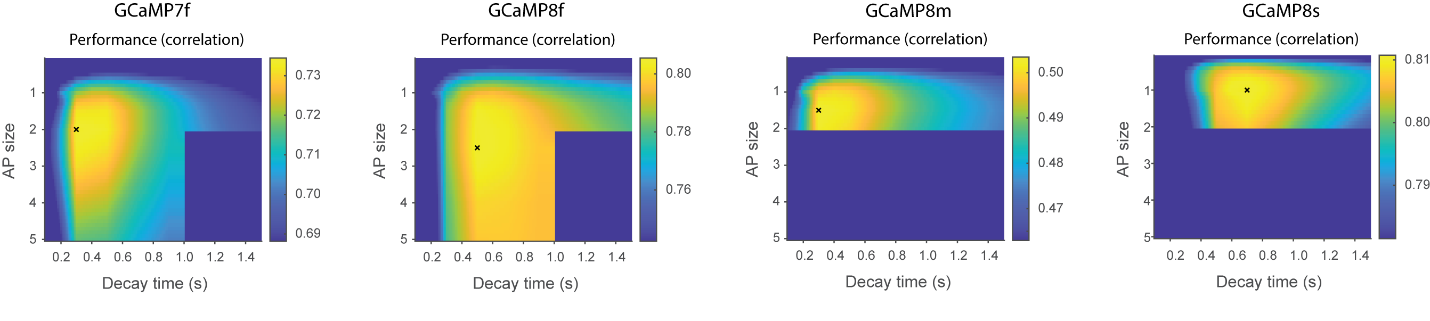
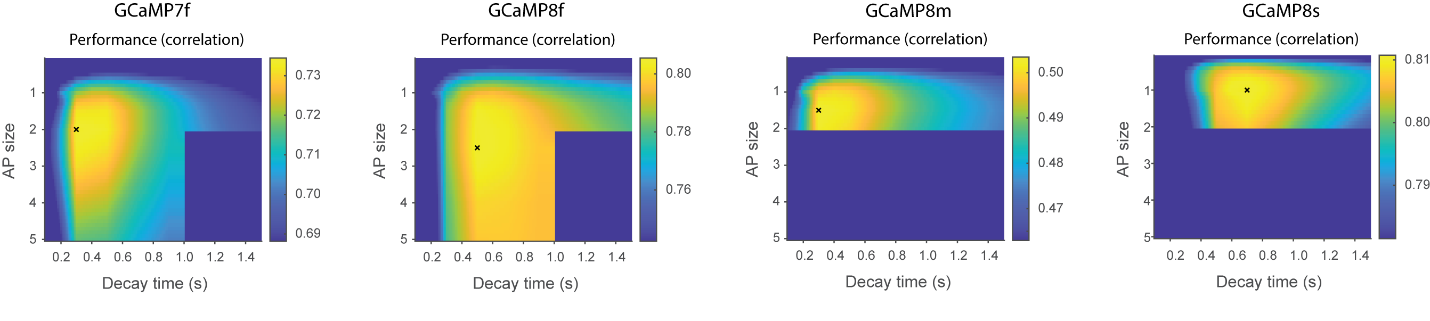
**Statistical analysis and box plots**

Statistical analyses were performed in MATLAB 2022b. Only non-parametric tests were used. The Mann–Whitney rank-sum test was used for non-paired samples, and the Wilcoxon signed-rank test was used for paired samples. Two-sided tests were applied unless otherwise stated. Box plots used standard settings in MATLAB, with the central line at the median of the distribution, the box at the 25th and 75th percentiles and the whiskers at extreme values excluding outliers (outliers defined as data points that are more than 1.5·D away from the 25th or 75th percentile value, with D being the distance between the 25th and 75th percentiles).

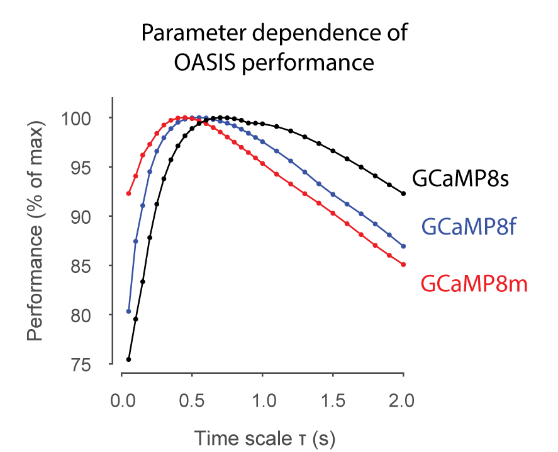
**Figure 1. Benchmarking of spike inference algorithms for GCaMP8**. **a,** Scheme of ground truth recordings, obtained by simultaneous juxtacellular electrophysiology and two-photon calcium imaging. **b,** Overview of calcium indicator datasets prior to GCaMP8 (left, ref. 14) and datasets from the GCaMP8 study (right, ref. 9). **c,** Overview of three commonly used algorithms for spike inference. **d,** Example of spike inference from calcium imaging data (top trace, resampled at 30 Hz, at a standardized noise level of “8”, GCaMP8m), with ground truth spike rate (black) as well as spike rates inferred by different variants of the CASCADE algorithm (blue), OASIS (pink) and MLSpike (red). Spike inference for the same recording but resampled at lower noise levels is shown in Fig. Suppl. 1-1. **e,** Quantification of model performance across all GCaMP8 ground truth data resampled at 30 Hz and with standardized noise level of “8”. **f,** Example of spike inference from the same ground truth, resampled at high vs. low noise levels. **g,** Improvement of spike inference (increase of correlation, ∆c, compared to using ∆F/F as a proxy for neuronal spike rate) for low and high noise levels.



**Figure 1 Supplement 1. Benchmarking of spike inference algorithms for GCaMP8 for low noise levels. a,** Example of spike inference from calcium imaging data (top trace, resampled at 30 Hz, at a standardized noise level of “2”, GCaMP8m), with ground truth spike rate (black) as well as spike rates inferred by different variants of the CASCADE algorithm (blue), OASIS (pink) and MLSpike (red). Spike inference for the same recording but resampled at lower noise levels is shown in Fig. 1d. **b,** Quantification of model performance across all GCaMP8 ground truth data resampled at 30 Hz and with standardized noise level of “2”. **c,** Performance of spike inference (correlation with ground truth) for Default CASCADE (trained with pre-GCaMP8 data) vs. GCaMP8-tuned CASCADE. Spike inference for GCaMP8 data is improved, performance for pre-GCaMP8 data is decreased, for both high and low noise levels.



**Figure 1 Supplement 2. Grid search for optimization of MLSpike parameters for spike inference with GCaMP8.** Search was performed in a 2D parameter space spanned by the two parameters *AP size* and *decay time*. Please note the different scaling of the four color maps; further, notice that the range of values is relatively narrow for all color maps and does not go to zero performance (typically spanning a range of <10% of the average value). Due to the high computational cost of running MLSpike, parameter search was stopped when the most likely global maximum was found, resulting in empty values of the grid search (dark blue).



**Figure 1 Supplement 3. Grid search for optimization of OASIS for spike inference with GCaMP8.** Search was performed in a 1D parameter space (*time scale*). Please notice that the range of values is relatively narrow and does not go to zero performance.

A collage of graphs and charts

Description automatically generated

**Figure 2. Inference of absolute spike rates: Deviations from linearity and identity. a,** Average ∆F/F as a function of true spike rate for the GCaMP8s dataset. Each connected line represents the values from a single neuron. **b,** Average spike rate as obtained with Default CASCADE as a function of true spike rate for the GCaMP8s dataset. The dashed line represents the unity relationship, here and in subsequent panels. **c,** Average spike rate as obtained with GCaMP8-tuned CASCADE as a function of true spike rate for the GCaMP8s dataset. **d,** Average spike rate as obtained with specifically GCaMP8s-tuned CASCADE as a function of true spike rate for the GCaMP8s dataset. **e,** Deviation from linearity. Measurements as in Fig. 2a-d are compared to the linear fit obtained for each neuron (dotted line). The normalized deviations (red double-arrows) are averaged and displayed across datasets and algorithms. **f,** Deviation from identity. For each neuron, measurements are compared to the identity relationship (dashed line). The normalized deviations (red double-arrows) are averaged and displayed across datasets and algorithms.

A graph of different colors and lines

Description automatically generated with medium confidence

**Figure 2 Supplement 1. Inference of absolute spike rates: Deviations from linearity and identity for MLSpike and transfer-learning GC8s-CASCADE.** Average inferred spike rate as a function of true spike rate for the GCaMP8s dataset. Each connected line represents the values from a single neuron. The dashed lines represent the unity relationship.

**A white paper with text and graphics

Description automatically generated with medium confidenceFigure 3. Improved spike inference during complex spike events for finetuned CASCADE models. a,** Examples of complex events that show a biased recovery of spike rates by Default CASCADE but not GC8-tuned CASCADE. The earlier time points of a complex spike event are exaggerated by Default CASCADE (light gray arrow) while later time points of a complex spike event are suppressed (dark gray arrow). The gray shading highlights cases where Default CASCADE underestimates the spike rate for trailing edges of complex events. **b,** Average complex event traces for ∆F/F; spike rates obtained with Default CASCADE and the difference compared to true spike rates; spike rates obtained with GC8-tuned CASCADE and the difference compared to true spike rates. The difference plot illustrates the bias induced by Default CASCADE but not GC8-tuned CASCADE. Data from the GCaMP8s ground truth dataset. **c,** Same as in (b) but for the GCaMP6f dataset. **d,** Temporal bias as shown in (b) and (c) quantified for each dataset when applying Default CASCADE. A box plot centered around zero indicates low or no bias. **e,** Same as in (d) but when applying GC8-tuned CASCADE. **f,** Scheme: fluorescence change ∆F as a function of calcium concentration change ∆C for a non-linear indicator such as GCaMP6. For such a non-linear relationship, the first action potential (∆C1) might go unnoticed (∆F1), while subsequent action potentials will trigger a much larger response (∆F2 or∆F3). **g,** Illustration of the effects of a non-linearity as shown in (f) on complex events, resulting in an exaggeration of later as opposed to early spikes. This property is able to explain the problems when applying Default CASCADE to GCaMP8 data due to the higher linearity of GCaMP8.

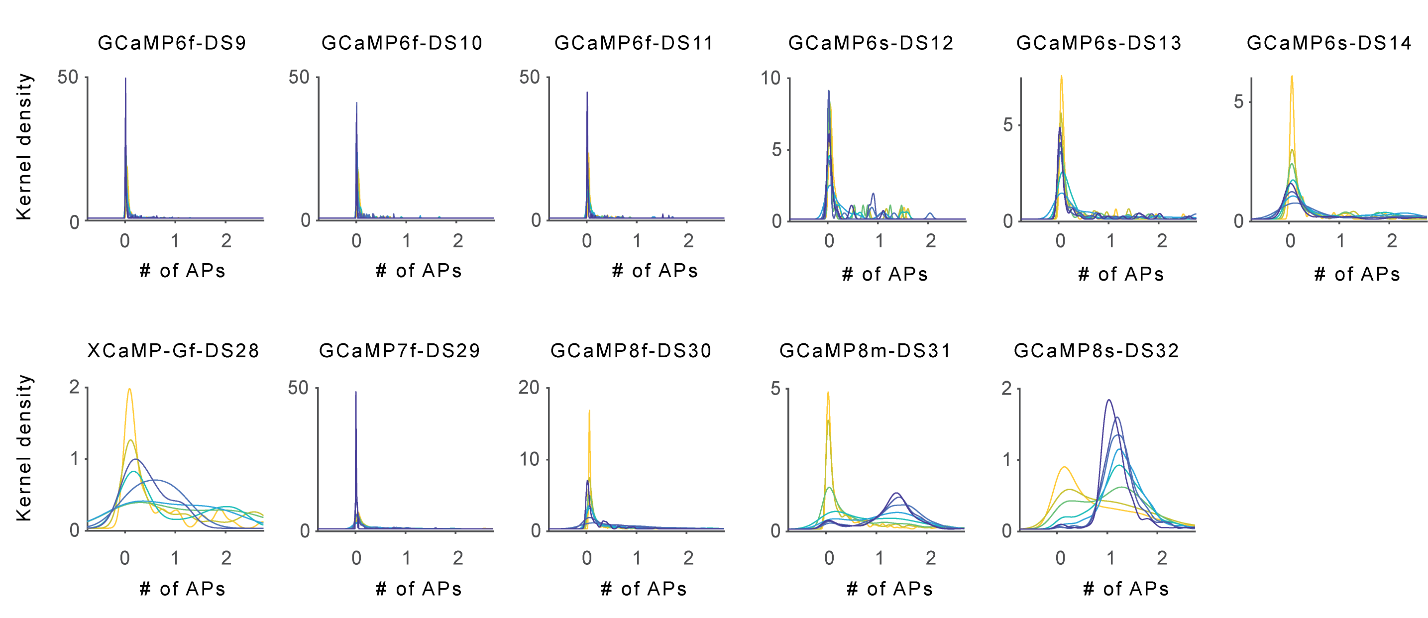
**A collage of graphs and diagrams

Description automatically generatedFigure 4. Detection of isolated action potentials from GCaMP8 with spike inference. a,** Examples of isolated action potentials and the corresponding ∆F/F trace (green) and the inferred spike rate (black) for an increasing standardized noise level ν. A low-noise population recording corresponds to ν ≈ 2, a higher-noise population recording to ν ≈ 8. Two example action potentials for the GCaMP8f dataset. **b-c,** Same as in (b) but for GCaMP8m and GCaMP8s datasets. **d,** Number of detected action potentials (APs) for a true isolated single AP, plotted as a distribution (kernel density estimate). Colors indicate the different noise levels (lowest noise level, blue; highest noise level, yellow). **e,** Fraction of APs correctly detected as a single AP (spike count >0.5 and <1.5) across datasets and noise levels. **f,** Fraction of APs correctly detected from noise (spike count >0.5) across datasets and noise levels. g, Percent of APs correctly detected as a single AP as defined in (e), averaged across noise levels ν = 2-4 for robustness. h, Percent of APs detected as defined in (f), averaged across noise levels ν = 2-4.

A screenshot of a graph

Description automatically generated

**Figure 4 Supplement 1. Examples of calcium signals triggered by single action potentials.** Examples of isolated action potentials and the corresponding ∆F/F trace (green) and the inferred spike rate (black) for an increasing standardized noise level ν. Examples shown for different datasets for GCaMP6 variants, and GCaMP7f under variable conditions (see Methods for details about the datasets).



**Figure 4 Supplement 2. Distributions of inferred spike numbers for single isolated action potentials.** Extension of Fig. 4d. Number of detected action potentials (APs) for a true isolated single AP, plotted as a distribution (kernel density estimate). Colors indicate the different noise levels (lowest noise level, blue; highest noise level, yellow).

A collage of graphs and diagrams

Description automatically generated

**Figure 5. Improved real-time spike inference for GCaMP8 due to faster onset kinetics.** **a,** Calcium transient evoked by the average action potential for a selected subset of datasets to illustrate differences of overall kinetics (left) and of rise times (zoom-in, right). **b,** Rise times to half of maximum across indicators. Box plots represent distributions across neurons for each dataset. **c,** Illustration of online spike inference, with a limited number of imaging frames available from the time point after the occurrence of a spike. The resulting performance (% of maximal performance for this dataset) increases sigmoidally as a function of time after the action potential used for inference. For each condition (2 frames, 5 frames, 16 frames, etc.), a new CASCADE model is trained to optimally take advantage of the limited amount of information. Sigmoidal performance dependences are shown for three example datasets. **d,** Time to reach 90% of the maximal performance, distribution across neurons shown for each dataset. **e-f,** Same as (d) but for higher noise levels.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Noise level | GCaMP6f | GCaMP6s | GCaMP7f | XGCaMP-Gf | GCaMP8f | GCaMP8m | GCaMP8s |
| 1.0 | 336 (45) | 167 (33) | 26 (2) | 0 (0) | 107 (10) | 229 (22) | 241 (19) |
| 1.5 | 342 (47) | 167 (33) | 56 (9) | 10 (2) | 252 (27) | 502 (35) | 508 (30) |
| 2.0 | 342 (47) | 167 (33) | 78 (12) | 10 (2) | 329 (30) | 551 (39) | 578 (34) |
| 3.0 | 342 (47) | 167 (33) | 113 (18) | 27 (4) | 362 (34) | 599 (42) | 589 (36) |
| 4.0 | 342 (47) | 167 (33) | 141 (19) | 34 (5) | 390 (36) | 611 (42) | 594 (37) |
| 6.0 | 342 (47) | 167 (33) | 146 (20) | 42 (7) | 392 (36) | 611 (42) | 597 (38) |
| 8.0 | 342 (47) | 167 (33) | 155 (21) | 44 (7) | 396 (36) | 611 (42) | 597 (38) |
| 10 | 342 (47) | 167 (33) | 155 (21) | 44 (7) | 396 (36) | 611 (42) | 600 (38) |

**Suppl. Table 1. Number of isolated action potentials detected.** Related to Fig. 4. For the analyses in Fig. 4, Gaussian noise was added to ground truth recordings to obtain data with a defined noise level (left-most column). Since some ground truth recordings were performed at a noise level higher than the lowest analyzed condition, they could not be included for low-noise condition analyses. Therefore, the number of detected isolated action potentials (values indicated in the dataset-specific columns) and the number of the underlying recorded neurons (values indicated in brackets) increase with higher noise levels.

**References**

1. Abdelfattah, A. *et al.* Neurophotonic Tools for Microscopic Measurements and Manipulation: Status Report. *Neurophotonics* **9**, (2022).

2. Emiliani, V. *et al.* Optogenetics for light control of biological systems. *Nat. Rev. Methods Primer* **2**, 55 (2022).

3. Grienberger, C., Giovannucci, A., Zeiger, W. & Portera-Cailliau, C. Two-photon calcium imaging of neuronal activity. *Nat. Rev. Methods Primer* **2**, 67 (2022).

4. Rose, T., Goltstein, P. M., Portugues, R. & Griesbeck, O. Putting a finishing touch on GECIs. *Front. Mol. Neurosci.* **7**, (2014).

5. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat. Methods* **6**, 875–881 (2009).

6. Chen, T.-W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).

7. Dana, H. *et al.* High-performance calcium sensors for imaging activity in neuronal populations and microcompartments. *Nat. Methods* **16**, 649–657 (2019).

8. Inoue, M. *et al.* Rational Engineering of XCaMPs, a Multicolor GECI Suite for In Vivo Imaging of Complex Brain Circuit Dynamics. *Cell* **177**, 1346-1360.e24 (2019).

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

10. Kerr, J. N. D., Greenberg, D. & Helmchen, F. Imaging input and output of neocortical networks in vivo. *Proc. Natl. Acad. Sci.* **102**, 14063–14068 (2005).

11. Yaksi, E. & Friedrich, R. W. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca2+ imaging. *Nat. Methods* **3**, 377–383 (2006).

12. Lütcke, H. *et al.* Optical recording of neuronal activity with a genetically-encoded calcium indicator in anesthetized and freely moving mice. *Front. Neural Circuits* **4**, (2010).

13. Pachitariu, M., Stringer, C. & Harris, K. D. Robustness of Spike Deconvolution for Neuronal Calcium Imaging. *J. Neurosci.* **38**, 7976–7985 (2018).

14. Rupprecht, P. *et al.* A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging. *Nat. Neurosci.* **24**, 1324–1337 (2021).

15. Pachitariu, M. *et al.* Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at www.biorxiv.org/content/10.1101/061507v2 (2017). (2019).

16. Giovannucci, A. *et al.* CaImAn an open source tool for scalable calcium imaging data analysis. *eLife* **8**, e38173 (2019).

17. Deneux, T. *et al.* Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nat. Commun.* **7**, 1–17 (2016).

18. Berens, P. *et al.* Community-based benchmarking improves spike rate inference from two-photon calcium imaging data. *PLoS Comput. Biol.* **14**, (2018).

19. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLOS Comput. Biol.* **13**, e1005423 (2017).

20. Wei, Z. *et al.* A comparison of neuronal population dynamics measured with calcium imaging and electrophysiology. *PLOS Comput. Biol.* **16**, e1008198 (2020).

21. Greenberg, D. S. *et al.* Accurate action potential inference from a calcium sensor protein through biophysical modeling. Preprint at www.biorxiv.org/content/10.1101/479055v1 (2018). (2020).

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

23. Beaulieu-Laroche, L., Toloza, E. H. S., Brown, N. J. & Harnett, M. T. Widespread and Highly Correlated Somato-dendritic Activity in Cortical Layer 5 Neurons. *Neuron* **103**, 235-241.e4 (2019).

24. Huang, L. *et al.* Relationship between spiking activity and simultaneously recorded fluorescence signals in transgenic mice expressing GCaMP6. Preprint at www.biorxiv.org/content/10.1101/788802v1 (2019). (2020).

25. Rupprecht, P., Fan, W., Sullivan, S. J., Helmchen, F. & Sdrulla, A. D. Spike inference from mouse spinal cord calcium imaging data. Preprint at https://doi.org/10.1101/2024.07.17.603957 (2024).

26. Éltes, T., Szoboszlay, M., Kerti-Szigeti, K. & Nusser, Z. 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 (2019).

27. Friedenberger, Z., Harkin, E., Tóth, K. & Naud, R. Silences, spikes and bursts: Three‐part knot of the neural code. *J. Physiol.* **601**, 5165–5193 (2023).

28. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

29. Deneux, T. *et al.* Accurate spike estimation from noisy calcium signals for ultrafast three-dimensional imaging of large neuronal populations in vivo. *Nat. Commun.* **7**, 1–17 (2016).

30. Yosinski, J., Clune, J., Bengio, Y. & Lipson, H. How transferable are features in deep neural networks? (2014) doi:10.48550/ARXIV.1411.1792.

31. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).

32. O’Neill, P. S. *et al.* A deep learning framework for automated and generalized synaptic event analysis. Preprint at https://doi.org/10.1101/2023.11.02.565316 (2023).