

TECHNIQUE

A machine-learning tool to identify bistable states from calcium imaging data

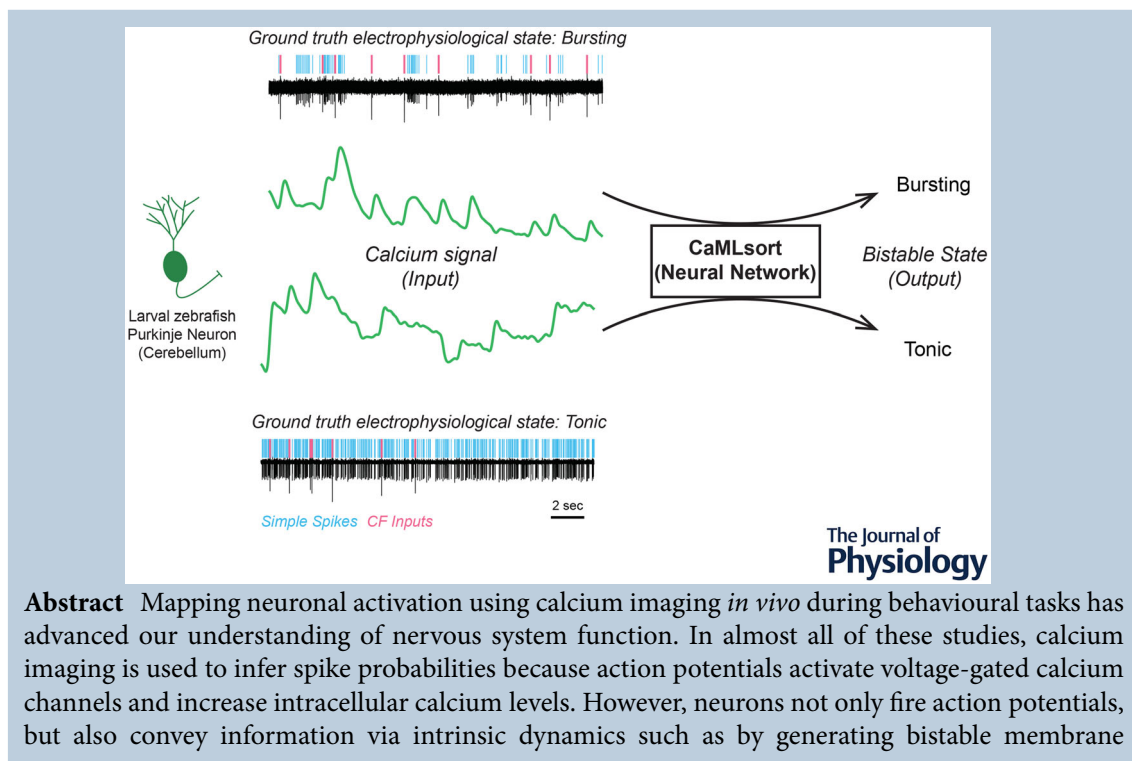
Aalok Varma¹ , Sathvik Udupa², Mohini Sengupta¹ , Prasanta Kumar Ghosh² and Vatsala Thirumalai¹ 

¹National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India

²Department of Electrical Engineering, Indian Institute of Science, Bangalore, India

Handling Editors: David Wyllie & Conny Kopp-Scheinflug

The peer review history is available in the Supporting Information section of this article (<https://doi.org/10.1113/JP284373#support-information-section>).



Abstract Mapping neuronal activation using calcium imaging *in vivo* during behavioural tasks has advanced our understanding of nervous system function. In almost all of these studies, calcium imaging is used to infer spike probabilities because action potentials activate voltage-gated calcium channels and increase intracellular calcium levels. However, neurons not only fire action potentials, but also convey information via intrinsic dynamics such as by generating bistable membrane

Aalok Varma is a graduate student at the National Centre for Biological Sciences (NCBS-TIFR), Bangalore, India. He obtained a bachelor's degree (BSc) in Biotechnology from Manipal University, India, graduating as the university gold medallist (valedictorian). His PhD studies, which include the work described in the present study, have explored what bistability in individual Purkinje neurons means for the ensemble of neurons in the cerebellum. Besides academic research, he advocates for systemic change in academia as a member of eLife's Early Career Advisory Group (ECAG) and enjoys communicating his science and sharing his experience by teaching.



This article was first published as a preprint. Varma A, Udupa S, Sengupta M, Kumar Ghosh P, Thirumalai V. 2022. A machine-learning tool to identify bistable states from calcium imaging data. bioRxiv. <https://doi.org/10.1101/2022.11.10.515941>

potential states. Although a number of tools for spike inference have been developed and are currently being used, no tool exists for converting calcium imaging signals to maps of cellular state in bistable neurons. Purkinje neurons in the larval zebrafish cerebellum exhibit membrane potential bistability, firing either tonically or in bursts. Several studies have implicated the role of a population code in cerebellar function, with bistability adding an extra layer of complexity to this code. In the present study, we develop a tool, CaMLSort, which uses convolutional recurrent neural networks to classify calcium imaging traces as arising from either tonic or bursting cells. We validate this classifier using a number of different methods and find that it performs well on simulated event rasters as well as real biological data that it had not previously seen. Moreover, we find that CaMLsort generalizes to other bistable neurons, such as dopaminergic neurons in the ventral tegmental area of mice. Thus, this tool offers a new way of analysing calcium imaging data from bistable neurons to understand how they participate in network computation and natural behaviours.

(Received 13 January 2023; accepted after revision 9 February 2024; first published online 14 March 2024)

Abstract figure legend Larval zebrafish Purkinje neurons (PNs) (left) are bistable: they can fire tonically (middle, bottom) or in bursts (middle, top). They also generate two types of electrical events: simple spikes (cyan in rasters) or climbing fibre (CF) EPSPs called CF events (pink in rasters). Calcium imaging of larval zebrafish PNs shows $\Delta F/F$ peaks for both types of electrical events during bursting or tonic firing (green traces). We developed CaMLsort, a machine learning based sorter that trained on ground truth datasets, to make accurate calls on bursting or tonic state of larval zebrafish PNs.

Key points

- Calcium imaging, comprising the gold standard of inferring neuronal activity, does not report cellular state in neurons that are bistable, such as Purkinje neurons in the cerebellum of larval zebrafish. We model the relationship between Purkinje neuron electrical activity and its corresponding calcium signal to compile a dataset of state-labelled simulated calcium signals.
- We apply machine-learning methods to this dataset to develop a tool that can classify the state of a Purkinje neuron using only its calcium signal, which works well on real data even though it was trained only on simulated data.
- CaMLsort (**C**alcium imaging and **M**achine Learning based tool to **sort** intracellular state) also generalizes well to bistable neurons in a different brain region (ventral tegmental area) in a different model organism (mouse).
- This tool can facilitate our understanding of how these neurons carry out their functions in a circuit.

Introduction

Calcium imaging enables us to sample the activity of populations of neurons and even whole brains during animal behaviour (Ahrens et al., 2013; Grewe & Helmchen, 2009; Grienberger & Konnerth, 2012), providing us a window into the functioning of the nervous system. When a neuron fires an action potential, somatic calcium levels increase dramatically as a result of the opening of voltage-dependent calcium channels. Simultaneous electrophysiology and calcium imaging recordings of neuronal activity demonstrate the correspondence of somatic calcium imaging signals with action potential firing (Chen et al., 2013; Kerr et al., 2005; Kwan & Dan, 2012; Tian et al., 2009). Nevertheless, the calcium imaging signal is at best a delayed, low-pass

filtered and non-linearly transformed proxy for neuronal activity. Non-linearities are introduced as a result of the dynamics of calcium release, and the kinetics of calcium binding and unbinding by the sensor molecules (Ali & Kwan, 2020). This technique would thus be most useful in cases where the neurons being studied have low baseline firing rate. However, this is not often the case and several neuronal classes, such as cerebellar Purkinje neurons (PNs) and cortical interneurons are known to fire at very high spike rates spontaneously (Armstrong & Edgley, 1984; Markram et al., 2004). Further, many neuronal types have additional intracellular dynamics apart from action potential firing that can strongly modulate calcium imaging signals (Lin et al., 2007; Moreaux & Laurent, 2007). Lastly, confounds can be introduced when more

than one type of electrical event leads to somatic calcium increase (Lev-Ram et al., 1992; Ramirez & Stell, 2016).

PNs show all of the features mentioned above that make the interpretation of calcium signals in these neurons especially difficult. PNs are principal neurons of the cerebellum, integrate thousands of synaptic inputs and are indispensable for cerebellar function (Narayanan & Thirumalai, 2019). Therefore, studying PN population activity during motor behaviours is critical to deciphering cerebellar computation. In mammals, PNs fire simple spikes and complex spikes, comprising climbing fibre (CF) input from the inferior olive leads to calcium entry within the dendrites and several spikelets that ride the depolarization wave (Kitamura & Häusser, 2011; Miyakawa et al., 1992). In fish, the olivary input causes an all-or-none AMPAR-mediated giant synaptic potential with concomitant calcium entry but without any spikelets (Sengupta & Thirumalai, 2015). Although the CF input causes a relatively large calcium signal, bursts of simple spikes can also generate calcium signals via the activation of somatic voltage-dependent calcium channels, albeit smaller in amplitude (Knogler et al., 2019; Sengupta, 2015). This means that optically recorded calcium $\Delta F/F$ signals cannot be uniquely assigned to simple spikes or CF inputs.

In addition to the above, larval zebrafish PNs exhibit membrane potential bistability. These neurons can exist in one of two stable membrane potentials. At a depolarized state, they generate action potentials at a more or less regular rate and are said to be tonic. When hyperpolarized, they generate bursts of action potentials (Sengupta & Thirumalai, 2015). Simple spikes in the bursting state are highly correlated with motor bouts, whereas those in the tonic state are not (Sengupta, 2015). This implies that bistability has important ramifications for how motor bouts are represented across the entire PN population. We considered whether calcium imaging can be used to deduce the state of PNs. Because both CF inputs and simple spikes contribute to calcium signals, assigning optically recorded $\Delta F/F$ signals is not straight-forward and deduction of spike rate using existing spike inference algorithms is not possible.

Here, we first performed simultaneous electrophysiological recordings and calcium imaging in larval zebrafish PNs to obtain ground truth data. Unsurprisingly, we found that none of the usually measured parameters such as peak amplitude, area under the curve, etc., of the calcium signal uniquely correlated with cellular state. We therefore built a sorter using convolutional recurrent neural networks to identify cellular state from $\Delta F/F$ data. We have tested our sorter with data that was not part of the training set and validated that it performs at above 90% accuracy and with F1 scores greater than 0.8 in most cases. Our method demonstrates that calcium imaging

datasets can be mined using such tools to glean greater insights into neuronal activity states.

Methods

Ethical approval

All experimental procedures were approved by the institutional animal ethics committee [IAEC; NCBS-IAE-2020/14 (R1E)] and institutional biosafety committee (IBSC; TFR/NCBS/14 IBSC/VT-1/2011) at the National Centre for Biological Sciences.

Experimental animals

Experiments were performed on Indian wild-type (Ind WT) zebrafish (*Danio rerio*). Embryos were obtained by setting up an in-cross between adult Ind WT zebrafish raised and housed in a ZebTec multi linking system (Tecniplast, Buguggiate, Italy) with a pH of 7.8 and conductivity of 1200 μS . Larvae were kept in a MIR-154 incubator (Sanyo, Osaka, Japan) under a 14:10 h light/dark photocycle maintained at 28°C in E3 medium (composition in mM: 5 NaCl, 0.17 KCl, 0.33 CaCl₂, and 0.33 MgSO₄, pH 7.8) in standard 90 mm Petri dishes (Tarsons, New Delhi, India). The medium was routinely replaced and, starting at 5 days post fertilization (dpf), larvae were fed Zeigler Larval diet AP100 (<100 μm) (Pentair AES, Cary, NC, USA). Larvae have not yet undergone sex specification by these stages, and so the experiments performed were agnostic to the sex of the animal.

Transient transgenesis by microinjection

Thin-walled borosilicate capillaries (1.0 mm outer diameter, 0.56 mm inner diameter; Warner Instruments, Holliston, MA, USA) were pulled using a Flaming/Brown P-97 pipette puller (Sutter Instruments, Novato, CA, USA) to produce long and thin pipettes for microinjection. Indian WT embryos at the one to two cell stage were injected with ~ 10 nL of a mixture of tagRFP-T:PC:GCaMP5G plasmid (Matsui et al., 2014) (50 ng μL^{-1}) and Tol2 transposase mRNA (50 ng μL^{-1}) (Urasaki et al., 2006), using a Picospritzer III (Parker Hannifin, Cleveland, OH, USA) pressure injection apparatus, and raised as described above. Larvae were screened at 5 dpf at an epifluorescence dissection microscope (SZX-16; Olympus, Tokyo, Japan) and those showing mosaic labelling of PNs were selected for experiments, which were performed at 7 dpf at room temperature (25–28°C).

Electrophysiology

Larvae were anaesthetized in 0.01% MS-222 (Sigma-Aldrich, St Louis, MO, USA) and transferred to a custom-made acrylic recording chamber. Fine tungsten wire (California Fine Wire, Grover Beach, CA, USA) was used to pin the larva onto a piece of Sylgard (Dow Corning, Midland, MI, USA) glued to the recording chamber. Two pins were put through the notochord at the level of the swim bladder and the tail, respectively. A third pin was then used to position the larval head dorsal-up. The MS222 was replaced by external solution (composition in mM: 134 NaCl, 2.9 KCl, 1.2 MgCl₂, 10 HEPES, 10 glucose, 2.1 CaCl₂, 0.01 D-tubocurarine; pH 7.8; 290 mOsm) and the skin on the dorsal surface of the larva was carefully removed using fine no. 5 forceps (Fine Science Tools, Heidelberg, Germany) to expose the brain.

The Flaming/Brown pipette puller was used to pull patch pipettes made of standard-wall borosilicate capillaries with filament (1.5 mm outer diameter; 0.86 mm inner diameter; Warner Instruments), pulled to a final tip diameter of 1–1.5 μm . For loose patch recordings, pipettes were backfilled with external solution, whereas, for whole-cell recordings, pipettes were backfilled with a K-gluconate-based internal solution (composition in mM: 115 K-gluconate, 15 KCl, 2 MgCl₂, 10 HEPES, 10 EGTA, 4 Mg-ATP; pH 7.2; 290 mOsm). Solutions were filtered using a 0.2 μm filter (Millipore, Burlington, MA, USA) before backfilling pipettes, which typically had resistances between 8 and 10 M Ω . Sulforhodamine (Sigma-Aldrich) was added to the patch internal solution at a final concentration of 5 $\mu\text{g mL}^{-1}$ to facilitate visualization of cellular morphology post-recording.

Recordings were acquired with a Multiclamp 700B amplifier, Digidata 1440A digitizer and pCLAMP 10 software (Molecular Devices, San Jose, CA, USA). Data were low pass filtered at 2 kHz and sampled at 20–50 kHz with a gain of 1 (whole-cell recordings) or between 20 and 1000 (loose patch recordings), aiming to optimally use the dynamic range of the digitizer.

Simultaneous calcium imaging and electrophysiology

Indian WT larvae microinjected with the Tol2-tagRFP-T:PC:GCaMP5G construct (gift from Dr Hideaki Matsui, Niigata University, Japan) and screened for sparse labelling were used for this experiment. PN8 expressing GCaMP5G were targeted for the recordings. Widefield imaging was performed using a 60 \times water-immersion objective (1.0 NA) on a BX61WI microscope (Olympus). Excitation light was provided using a pE300-ultra (CoolLED). Imaging was performed using an EVOLV EM-CCD camera (Photometrics, Tucson, AZ, USA) at the rig, with an exposure time of 30 ms and an on-chip gain of 170. A region of interest (ROI)

was drawn around the targeted cell, so that only an isolated cell was imaged. Imaging was synchronized with electrophysiology by interfacing with the Digidata 1440A.

Calcium imaging analysis

Image analysis was performed in Fiji/ImageJ (Schindelin et al., 2012). To identify cell boundaries, the averaged time series was used as a reference and a ROI was drawn around the cell. Bleach correction was performed using the exponential fitting method in the Bleach Correction plugin in Fiji. The average raw pixel intensity values within the ROI for each time point were then extracted. The baseline intensity, F_0 , was taken to be the 5th percentile of these values. $\Delta F/F$ was then calculated for each frame using the formula $(F - F_0)/F_0$, where F represents the fluorescence in that frame and F_0 is the baseline fluorescence. Following this calculation, a Gaussian filter was applied to the time series in MATLAB R2019b (Mathworks, Natick, MA, USA), using its function 'smoothdata' with a window size of 15.

To detect transients in the calcium signal, $\Delta F/F$ traces were first normalized to a range of [0, 1] using min-max feature scaling. Peaks in the normalized trace were then detected using the 'findpeaks' function in MATLAB, with a minimum peak height of 0.1, a minimum peak prominence of 0.05 and a minimum interval of 500 ms between consecutive peaks.

Spike detection and analysis

All electrophysiological analyses were performed using custom scripts written in MATLAB, using functions from a publicly available repository (<https://github.com/wagenadl/mbi-nsb-toolbox>) to import electrophysiology data. For extracellular recordings, spikes were identified as events that showed large fluctuations relative to a 25 ms rolling window baseline and the peak timing and amplitude of these were extracted. Events were sorted based on amplitude, with small amplitude events being classified as simple spikes and large amplitude events as climbing fibre (CF) inputs, as per (Sengupta & Thirumalai, 2015). For intracellular recordings, peak detection algorithms in-built in MATLAB were used to identify spikes. Once again, events were sorted as simple spikes or CF inputs based on their amplitude relative to baseline.

Calcium imaging reconstruction from electrophysiology data

Simultaneous calcium imaging and extracellular electrophysiology data was used to infer optimal GCaMP5G

kernels, which are of the general form:

$$A1 * e^{-\frac{t}{\tau_{\text{decay}}}} - A2 * e^{-\frac{t}{\tau_{\text{rise}}}}$$

for simple spikes and CF inputs of zebrafish PNs. Simple spike and CF input timings were first extracted from the electrophysiology data for each cell. Simple spike timings were convolved with a simple spike kernel, with amplitudes A1s and A2s and CF input timings were convolved with a similar CF input kernel, with amplitudes A1c and A2c. The sum of both these convolutions was taken to be the putative calcium signal reconstruction, down-sampled to match the sampling rate of calcium imaging. The mean squared error (MSE) between the reconstruction and the true calcium imaging trace was calculated. The values for A1s, A2s, A1c and A2c were obtained by minimizing the MSE between the true calcium imaging trace and the reconstructed trace using gradient descent. To ensure that local minima were avoided, the gradient descent algorithm was repeated with randomized starting points. The kernels with these coefficients were taken to be the optimal simple spike and CF input kernels, respectively, for that cell.

The 'optimal' kernel to be used across all cells would be one that generalized best. This was tested by using the kernels from one cell to reconstruct calcium imaging for other cells for which ground truth data were available. As before, the MSE and the Pearson correlation coefficient between the reconstructed trace and the ground truth trace were calculated. The final 'optimal' kernel was obtained by taking the kernel which generalized the best (i.e. had minimum MSE and maximum correlation across all cells). To test whether the correlation was real or not, the Pearson correlation coefficient between a scrambled reconstructed trace and the ground truth trace was also calculated, for reference.

This process of convolving simple spike and CF timings with their optimal GCaMP5G kernels and summing the resulting traces was then used to generate a calcium imaging trace from both extracellular and intracellular PN recordings.

Building a state-labelled calcium imaging dataset

Intracellular and extracellular recordings from larval zebrafish PNs acquired during the period 2011–2020 by various members of the lab were compiled. Good quality recordings were those where the resting membrane potential of the cell was not below -80 mV or not above -20 mV and the recordings comprised uninterrupted gap-free recordings with no current injections and had two clearly distinguishable event types. Besides poor-quality recordings, those recordings with an ambiguous, indeterminate state were also excluded. Only

Table 1. Number of recordings compiled for the state-labelled calcium imaging dataset

	Extracellular	Intracellular	Total
Bursting	39	39	78
Tonic	11	42	53
Switch	0	5	5
Ambiguous	0	25	25
Poor Quality	0	12	12
Total	50	123	173

those recordings which met the aforementioned criteria were retained. Then, the recordings were renumbered randomly, to remove grouping of data collected by the same individual. The recordings were then classified as tonic or bursting, first by visual inspection. These classifications were then further verified by inspecting the inter-event intervals of simple spikes. Although inter-event intervals from tonic cells form a single cluster, bursting cells have inter-event intervals that trail to large values, corresponding to the interburst intervals. Lastly, the coefficient of variation of the interspike intervals was used as an extra parameter of interest. These three pieces of information were used in conjunction (Fig. 1D–G), but the final classification relied on visual inspection. If there was a switch in state within the recording, the recordings were labelled as 'switch', with the appropriate state switch timing and type (tonic-to-bursting or bursting-to-tonic) marked. A summary of the recording information is provided in Table 1.

Here, data was split into two groups: one that contained traces only with a single state throughout the recording and another that contained traces with switches in state. Once split, the spike detection protocol described above was applied to each group of recordings to detect the simple spikes and CF inputs. These event timings were then used to generate simulated calcium imaging traces as described above. Combining the reconstruction with the assigned state resulted in the state-labelled calcium imaging dataset.

Principal components analysis

Principal components analysis (PCA) was performed using the PCA module that is part of the 'sklearn' package in Python (Pedregosa et al., 2011). The input to the model was a vector with the trace features for each 10 s long trace in the reconstructions dataset (number of peaks, peak amplitude, area under the curve, mean, SD and coefficient of variation). A two-component PCA was performed and the resulting fit (Fig. 4E) explained 88% of the variance in the data.

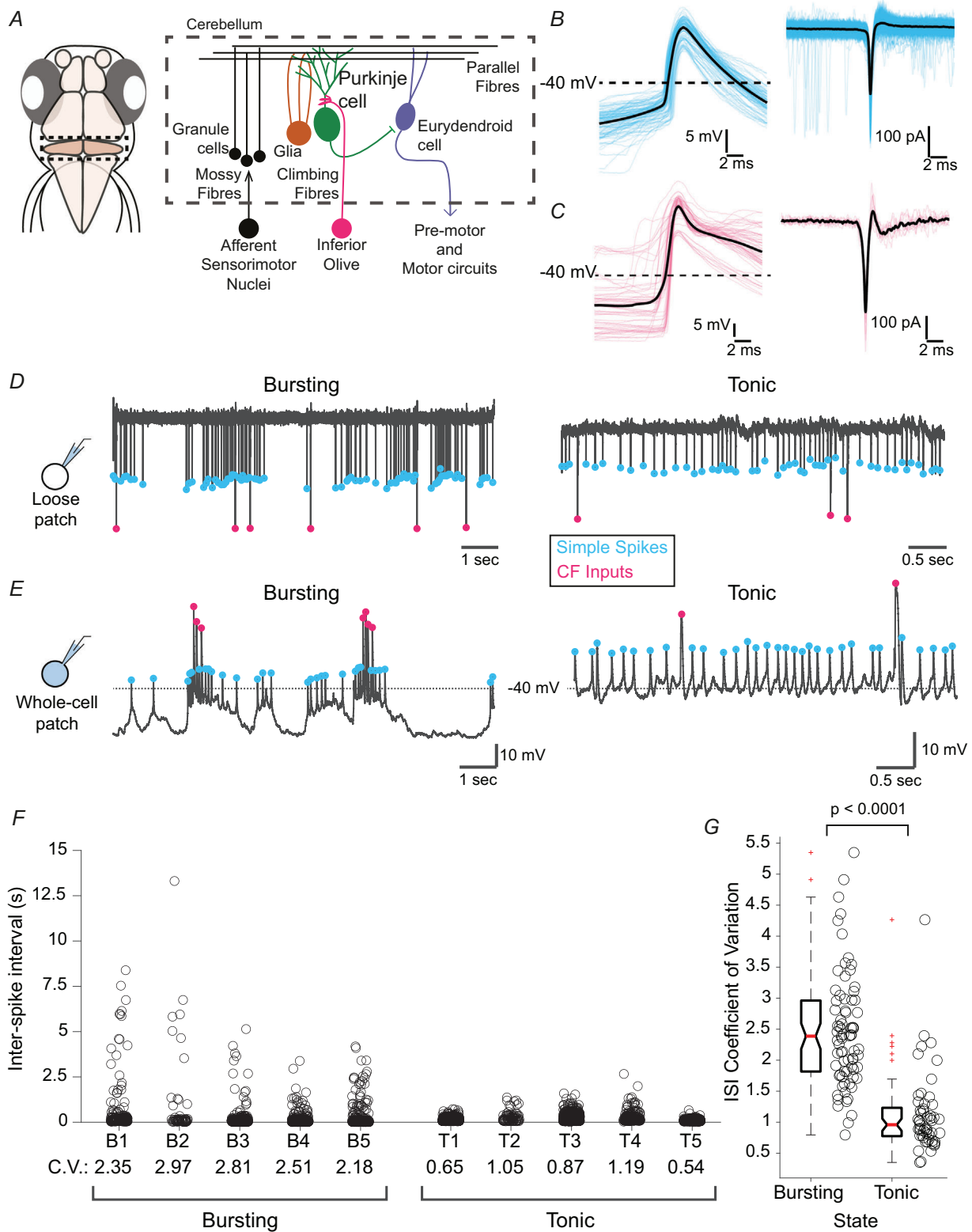


Figure 1. Cerebellar PNs in larval zebrafish exhibit membrane potential bistability

A, schematic of the cerebellar circuitry in larval zebrafish (*Danio rerio*), showing the location of the cerebellum in the brain (left, dashed line box) and the cell types and circuit architecture (right). PNs (green) are the principal neurons in this circuit and receive multiple inputs from afferent sensorimotor nuclei outside the cerebellum via parallel fibres (black) and climbing fibres (magenta). B, simple spikes, as observed intracellularly (left) and extracellularly (right). Individual events ($n = 100$, intracellular and $n = 100$, extracellular) are shown in cyan and have been

aligned and superimposed on one another. The average of these events is shown in black. *C*, climbing fibre (CF) inputs, as observed intracellularly (left) and extracellularly (right). Individual events ($n = 55$ intracellular and $n = 10$ extracellular) are shown in magenta, and their respective averages in black. *D*, representative loose patch recordings from larval zebrafish PNs showing the two modes of firing. The bursting mode recording is shown on the left, and the tonic mode is shown on the right. In each mode, electrical events can be distinguished by their amplitude, with small-amplitude simple spikes marked in cyan and the large-amplitude CF inputs marked in magenta. *E*, representative whole-cell current clamp recordings showing the two modes of firing in PNs, with the same colour scheme for events as before. Cells shown in (*D*) and (*E*) are different. *F*, example plots of simple spike inter-event intervals from individual recordings. Five example cells are shown from each class (Bursting and Tonic). The coefficient of variation of this distribution is marked under each cell as 'C.V.'. *G*, the distribution of interspike interval coefficients of variation ($n = 131$ cells), sorted by state ($n = 53$ for tonic, and $n = 78$ for bursting). The distributions are significantly different, tested using a Mann–Whitney U test ($P = 6.0733e-16$).

Training the convolutional neural network with long short-term memory (CNN-LSTM) model for classification

The dataset of reconstructed calcium imaging time series was first split into five folds to perform cross-validation (Fig. 5A). The experiments are repeated by considering each fold as a test set, with one of the folds being the validation set and the remaining three folds used as training sets such that the data split for each experiment was 60% for training, 20% for validation and 20% for testing. This splitting was performed at the level of entire traces, such that the training, testing and cross-validation sets all had a balanced representation of traces from both the tonic and bursting classes. We chose to do a five-fold cross-validation (i.e. the entire training dataset is split into five parts), each containing 20% of the data. Four of the folds are used for training and the remaining one is used for cross-validation.

The architecture of the network is shown in Fig. 5B. The input to the network is the entire trace sequence at 30 Hz, normalized using min-max normalization for every non-overlapping 10 s period, thus setting the minimum to 0 and maximum to 1. Note that, if the sampling frequency of a calcium trace was not 30 Hz, it was first linearly interpolated to a 30 Hz sampling frequency before being seen by the neural network. The first layer of processing is a CNN with kernel size 30 and stride size 30. The CNN down samples the input trace from 30 Hz to 1 Hz, and produces four vector outputs, which act as the input to the LSTM. The LSTM is bidirectional, such that it looks at the sequence in both the forward and reverse order to produce an output. The LSTM has 24 hidden layers and produces 48 numbers per time step (Fig. 5B). The feature dimensions are collapsed by a linear layer by taking their weighted average. This provides the classification logits at each position, which is passed through a softmax layer. The network is optimized using a cross-entropy loss and the Adam optimizer (Kingma & Ba, 2017). The output of a trained neural network, then, is the probability of each timestep (at 1 Hz) belonging to the bursting class, from which the probability of it belonging to the tonic class can be derived. The

class with the higher probability is assigned to each time step.

Because the model acts on the entire trace sequence, the amount of independent data for training the model is significantly reduced. To overcome this issue, we used a data augmentation strategy, which allows the network to learn features on arbitrary length sequences. For a sequence of length N , we determine a random segment size n , with n between $[k, N]$, where k is the minimum sequence length. We randomly determine a starting position x , with a constraint that $x < N - n$. In our case, $k = 300$ samples, and n varies in multiples of k (i.e. $n = 300, 600, 900 \dots$). In each iteration of the network pass during the training phase, this random sequential segment was taken as the input trace to the model and gradient descent was performed to update weights. This was carried out for all the traces in the training set, such that the network saw entire traces at least once and in a non-overlapping fashion during the 500 epochs of training. The training is stopped before the full 500 epochs, based on early stopping criteria, if the validation loss does not improve after few epochs. All models were trained using the PyTorch framework in Python 3 (Paszke et al., 2019).

For all evaluations, namely cross-validation, test and other independent validations, the network had to make predictions for the entire trace duration and not just the n -sample long chunks. Networks were evaluated using two metrics: classification accuracies and F1 scores. Classification accuracies are simply the fraction of predictions that were correct. F1 scores, on the other hand, are adjusted for class imbalances and account for the type of errors made. By definition, an F1 score is the harmonic mean of the precision and recall metrics, where precision is the fraction of all predicted positives that were true positives, and recall is the fraction of all ground truth positives that were true positives.

To improve/smoothen predictions made by CaMLSort (Calcium imaging and Machine Learning based tool to **sort** intracellular state), we used a majority voting strategy. To do this, we re-sampled the time series using rolling window sampling with a window length N . For each time step of prediction, t_n , we found all the m windows (where $m \leq N$) that contained t_n (i.e. $[t_{n-N+1}, t_{n-N+2}, \dots$

t_n, t_{n+1}], $[t_{n-N}, t_{n-N+1}, \dots, t_{n-1}, t_n]$ and so on until $[t_n, t_{n1}, \dots, t_{n+N-1}, t_{n+N}]$. For each of these windows, an average posterior score was calculated for each state, and the 'average class' taken to be the state which had the higher score. The 'majority class,' then, was the one which was the most frequent average class across all m windows. A 'confident call' was made if and only if the average posterior scores for the majority class across all m windows was strictly higher than the average posterior scores for the minority class.

Generating artificial PN spike trains

Synthetic spike trains of the tonic class were generated using a Poisson statistical model. The only parameter needed for this is an average event rate, λ , which we took to be a random number between 4 and 10 for simple spikes. Spike trains for bursting neurons do not follow Poisson statistics. Hence, we used an inverse sampling method to generate spike trains of the bursting class. For this, we used the compiled electrophysiology data to obtain the real cumulative distribution of interevent intervals for simple spikes across all bursting cells. A random number, y , is generated from the uniform distribution and it is used to find a value x from the reference cumulative interspike interval (ISI) distribution such that $P(\text{ISI} < = x) = y$. Thus, by generating events sequentially with a given next interevent interval, we obtained artificial simple spike trains for the bursting class.

CF input event trains were generated using a Poisson statistics model with an average event rate between 0.2 and 1. The distributions of interevent intervals in the synthetic datasets were compared against the real distributions using histograms and QQ plots. Independent reviewers were asked to classify unlabelled spike trains as real or artificial and managed to do so with an accuracy of $\sim 42\%$ on average.

Statistical analysis

We used a one-sided Mann–Whitney U test to compare the distributions of interspike interval coefficient of variation between tonic and bursting cells, for both PNs (Fig. 1G) and dopaminergic ventral tegmental area (VTA) neurons (Fig. 9C).

To compare the quality of reconstructions obtained from the generative forward model, we used a Kruskal–Wallis test to detect any overall group differences and followed it up with a *post hoc* Dunn's test for individual comparisons (Figs 3C and 9E).

For comparisons involving a nested data structure, such as those involving trace features between tonic and bursting cells (Fig. 4D), where there are multiple cells and multiple measurements per cell, we used a linear mixed-effects model with a maximum likelihood estimator to calculate P values.

Results

Zebrafish PNs exhibit membrane potential bistability

PNs are the principal neurons in the cerebellar circuit (Fig. 1A). In larval zebrafish, they exhibit two types of electrical events that can be distinguished solely based on their amplitude in both extracellular and intracellular recordings (Sengupta & Thirumalai, 2015) (Fig. 1B–E). The smaller amplitude events are action potentials and are called simple spikes (Fig. 1B, D and E, cyan), whereas the larger amplitude events are all-or-none synaptic input from climbing fibres (Fig. 1C–E, magenta). Individual simple spike and CF input waveforms are shown in Fig. 1B and C, respectively, for both whole-cell and juxtacellular recordings.

These neurons exhibit membrane potential bistability, such that they are found to be at one of two stable membrane potentials, depending on which they fire either tonically or in bursts. Cells can spontaneously switch between the two states, but can also be taken from one state to the other by current injection (Sengupta & Thirumalai, 2015). We observed very few spontaneous state switches in our recordings and these are dealt with later on in the present study.

The two states can be distinguished visually from extracellular loose-patch recordings based on (1) the firing frequency of simple spikes and (2) the presence of long intervals of inactivity in the burst mode (Fig. 1D, left), which are not present in the tonic mode (Fig. 1D, right). The same features can be used to distinguish between the two modes in whole-cell patch-clamp recordings, along with an extra feature: membrane potentials are more hyperpolarized in the bursting mode (Fig. 1E, left) as opposed to the tonic mode (Fig. 1E, right). Additionally, in tonic state, interspike intervals form a single cluster, whereas bursting cells have inter-event intervals that trail to large values, corresponding to the interburst intervals (Fig. 1F). This is also reflected in the values of coefficient of variation of the interspike intervals, which are larger in bursting cells than the more regularly spiking tonic cells (Fig. 1G). Given that there is a lot of variability across cells and recordings, no single cut-off value could be determined for these parameters (Fig. 1G). Instead, all the aforementioned features were taken together to determine the final state classification.

Calcium imaging reports both simple spike bursts and individual CF inputs, independent of state

Given that PNs manifest two distinct activity patterns (tonic and bursting) and two types of electrical events (simple spikes and CF events), we considered whether calcium imaging signals can reliably report these distinct patterns of activity. To do this, we performed simultaneous

wide-field calcium imaging and targeted patch-clamp electrophysiology from zebrafish PNs. We used the carbonic anhydrase (Ca8) enhancer element and a minimal promoter (*cfos*) to drive expression of the genetically encoded calcium sensor GCaMP5G in PNs (Matsui et al., 2014).

Embryos were microinjected at the one to two cell stage and those larvae that showed sparse labelling in PNs were selected for targeted patch-clamp electrophysiology and imaging (Fig. 2A). We chose to perform extracellular, loose-patch recordings aiming to be minimally invasive and to avoid the dialysis of cellular contents including GCaMP. We obtained such recordings from 15 cells in 12 larvae, out of which nine cells were naturally in the tonic state and six cells were in the bursting state.

Representative traces from this experiment for cells in the bursting and the tonic states are shown in Fig. 2B and C. The first observation we made was that many of the calcium transients (marked with a hash symbol in Fig. 2B and C) had an almost one-to-one correspondence with CF inputs (magenta lines in the raster plot) in cells of both states. In tonic cells, most simple spikes were not reported as transients, as would be expected from cells that fire at high frequencies. However, in some cases, calcium transient peaks occurred in the absence of CF inputs, during times of intense simple spike firing (Fig. 2C, asterisk and cyan lines in the raster plot). Similarly, in bursting cells, we found some calcium transients associated only with a burst of simple spikes (Fig. 2B, asterisk). Naturally, one would expect to see calcium transients when CF inputs co-occurred with a burst of simple spikes, which is what we observed (marked with open circles in Fig. 2B and C). Thus, regardless of the state of a cell, we found calcium transients that corresponded to individual CF inputs, bursts of simple spikes or a combination of both.

Although this method confirms that a calcium transient occurs after a CF input, suggesting a causal link between the CF input's arrival and the measured calcium influx into the cell, it is not useful in the case of simple spikes. First, there are multiple simple spikes per frame, ~ 2 or 3 spikes per frame when imaged at 30 Hz. This is further worsened by the timescales of GCaMP5G's rise and decay, which occurs on the order of 100 s of ms (Akerboom et al., 2012; Chen et al., 2013). Moreover, the individual effects of simple spikes and CF inputs cannot be separated out because the two events are interspersed and thus, when calculating the triggered average response for one event, there is bleed-through from the other.

To ascertain whether CFs and simple spikes indeed trigger calcium signals, we decided to compare the number of events in the time leading up to the calcium peak (Fig. 2D, region I) and during the decay period (Fig. 2D, region II). The expectation was that if an event triggered calcium activity, there would be more of them in region I, which includes the time of initiation of the

calcium transient (Fig. 2D), than in region II, when the signal is decaying. GCaMP5G has a rise time constant of ~ 100 ms, which means that the time it takes to reach a peak value from the baseline is ~ 201 ms (modelled using a difference of single exponentials, see Methods). Hence, to ensure that the time of initiation was included in region I, we considered a 300 ms interval before the peak of the calcium signal. An interval of the same 300 ms duration after the peak was taken to be region II.

We first detected calcium transient peaks in the $\Delta F/F$ signal (see Methods), of which there were a total of $n = 171$ detected across the $N = 15$ cells. The average number of simple spikes (Fig. 2E) and CF inputs (Fig. 2F) are significantly different in region I and II (linear mixed-effects model; $P < 0.0001$ for simple spikes and $P = 0.0025$ for CF inputs). There are more CF inputs in region I compared to region II indicating that these events can trigger calcium signals. However, for simple spikes, there is a small but significant increase in region II compared to region I. Although this is opposite to what might be expected, it can be explained by the high rate of simple spikes lasting over 700 s of ms in both bursting and tonic cells. In both regions I and II, there are ~ 2.5 simple spikes within the 300 ms window. These results imply that both simple spikes and CF inputs are associated with $\Delta F/F$ transients.

In summary, it is difficult to infer the underlying event train or even the cellular state of a PN directly from its calcium signal because there is no one-to-one correspondence between calcium transients and their underlying electrophysiological events, as has also been reported previously (Knogler et al., 2019; Sengupta, 2015).

The calcium signal for a PN can be reconstructed from its electrophysiological recording

Though the calcium signal in PNs could arise from more than one electrical event, we hypothesized that there might be sufficient information in the calcium signal time series to allow us to determine whether it is tonically firing or bursting. One possible approach to this problem could be the application of machine learning algorithms to build a classifier that labels cells as tonic or bursting given a time series of calcium transients. For this, it was first necessary to generate a large ground truth dataset (hundreds of cells) comprising the electrical activity patterns and their corresponding calcium transients for training the classifier. Ideally, this dataset would take the form of what is shown in Fig. 2B and C, where both electrophysiology and imaging data are simultaneously acquired. Our ground truth dataset has only 15 cells; however, the simultaneous imaging and electrophysiology experiments are low yield and technically challenging. Hence, we needed a strategy to augment our dataset.

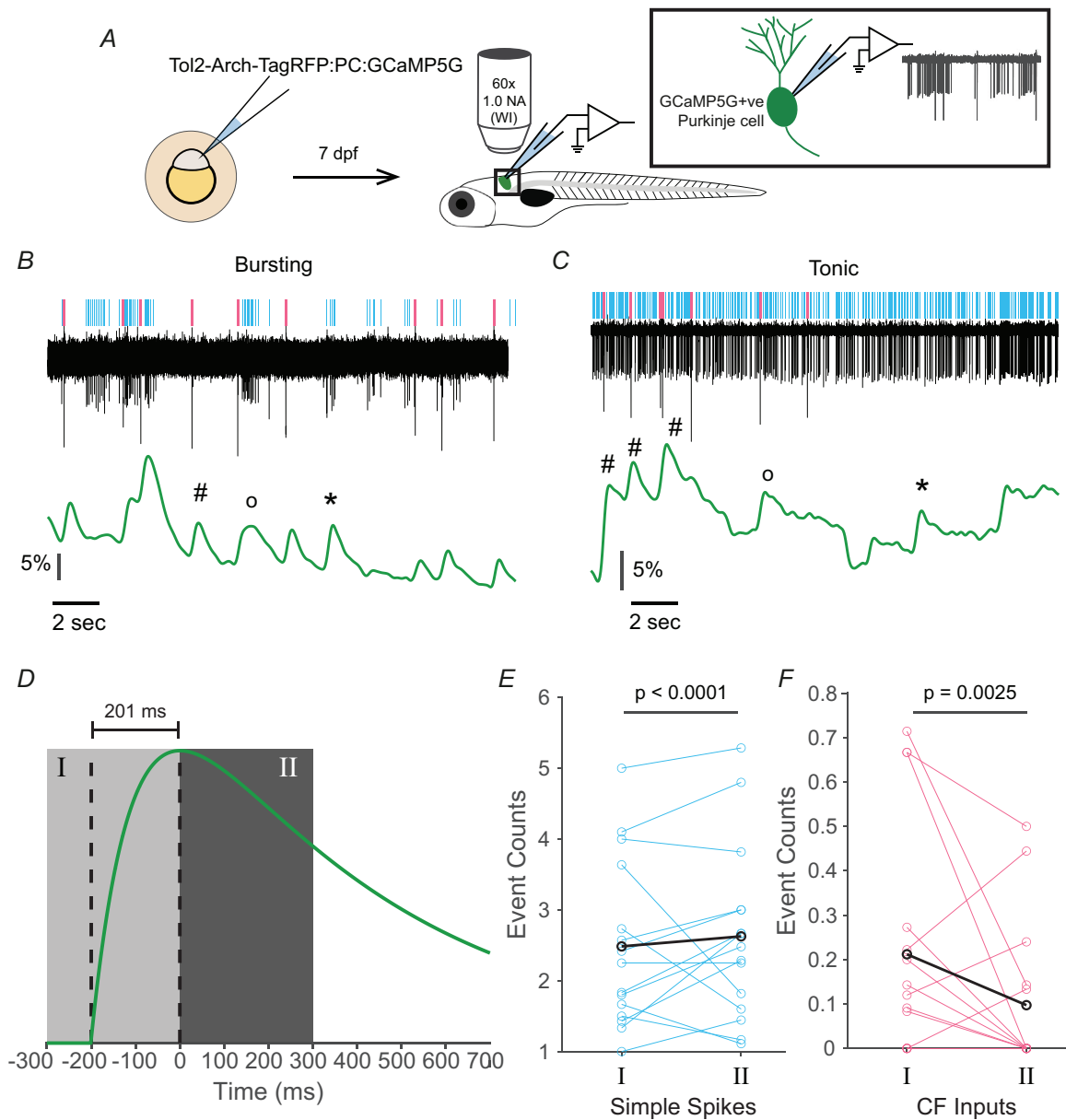


Figure 2. Both simple spikes and CF EPSPs contribute to the calcium signal in PN, irrespective of cellular state

A, schematic of the experimental design and setup of simultaneous calcium imaging and electrophysiology in larval zebrafish. **B** and **C**, representative traces obtained during simultaneous imaging and electrophysiology for cells in the bursting (**B**) and tonic (**C**) mode. Simple spikes and CF inputs in the trace are identified and marked in the raster above the trace in cyan and magenta, respectively. The simultaneously-recorded change in GCaMP fluorescence ($\Delta F/F$) is shown below in green. Calcium transients corresponding to simple spikes only (*), CF-events only (#), and both events (o) are marked. **D**, the typical GCaMP5G response to a single action potential, assuming it follows the profile of a difference of single exponentials with a half-rise time of 100 ms and half-decay time of 500 ms. The vertical dashed lines indicate the time of initiation of the calcium transient and the time when the transient peaks, which corresponds to an interval of 201 ms. Region I is the period up to 300 ms before the peak of the calcium signal, and region II is the period up to 300 ms after the peak of the transient. There were a total of $n = 171$ transient peaks in the $\Delta F/F$ signal detected across the $N = 15$ cells. **E**, average simple spike counts in the periods corresponding to regions I and II as defined in (**D**). Each cyan line corresponds to an individual cell ($N = 15$) and the average of all these lines is shown in black. **F**, average CF input event counts in the periods corresponding to regions I and II as defined in (**D**). Each magenta line corresponds to an individual cell ($N = 15$) and the average of all these lines is shown in black.

Mathematically, the calcium signal obtained by imaging can be thought of as the result of the convolution of spike trains with a calcium sensor kernel having a fast rise and a slow decay and described as a difference of exponentials. This transformation has been shown to work well in cells that are mostly quiescent and where every single action potential produces a large increase in intracellular calcium concentration (Rupperecht et al., 2021; Smetters et al., 1999; Vogelstein et al., 2010; Yaksi & Friedrich, 2006). We hypothesized that a similar transformation could be used to reconstruct the calcium signal profiles of PNs from their electrophysiological recordings alone. Such reconstructed calcium imaging traces with their corresponding electrical recording traces may then be applied to build the classifier.

However, given that these neurons have not one, but at least two electrophysiological events, each contributing differently to the calcium signal obtained from imaging, we must assign each event a separate kernel (Fig. 3A), which we assume to be independent of each other. This is a valid assumption because simple spikes and CF events arise from distinct intrinsic mechanisms (Sengupta & Thirumalai, 2015). We used simultaneous imaging and electrophysiology recordings to find optimal kernels for simple spikes and CF events. For each PN recording, we obtained simple spike and CF event rasters. Each event raster was then independently convolved with its own exponential response kernel, and the resulting 'event-specific signals' were linearly summed to produce the final reconstruction (Fig. 3A). We then calculated the correlation coefficient and the MSE between the reconstructed calcium signal time series and the experimentally obtained calcium imaging time series.

With this description in place, the problem then becomes one of finding optimal coefficients for each response kernel. This optimization was achieved by gradient descent to minimize the MSE between the ground truth and the reconstruction (Fig. 3A). The process was performed independently for each cell in the dataset. Then, these optimal kernel combinations were used to make reconstructions for other cells in the dataset, and the kernel that generalized best across all cells, as quantified by least total MSE and highest total cross-correlation between the ground truth and reconstruction, was taken to be the best kernel for the next steps.

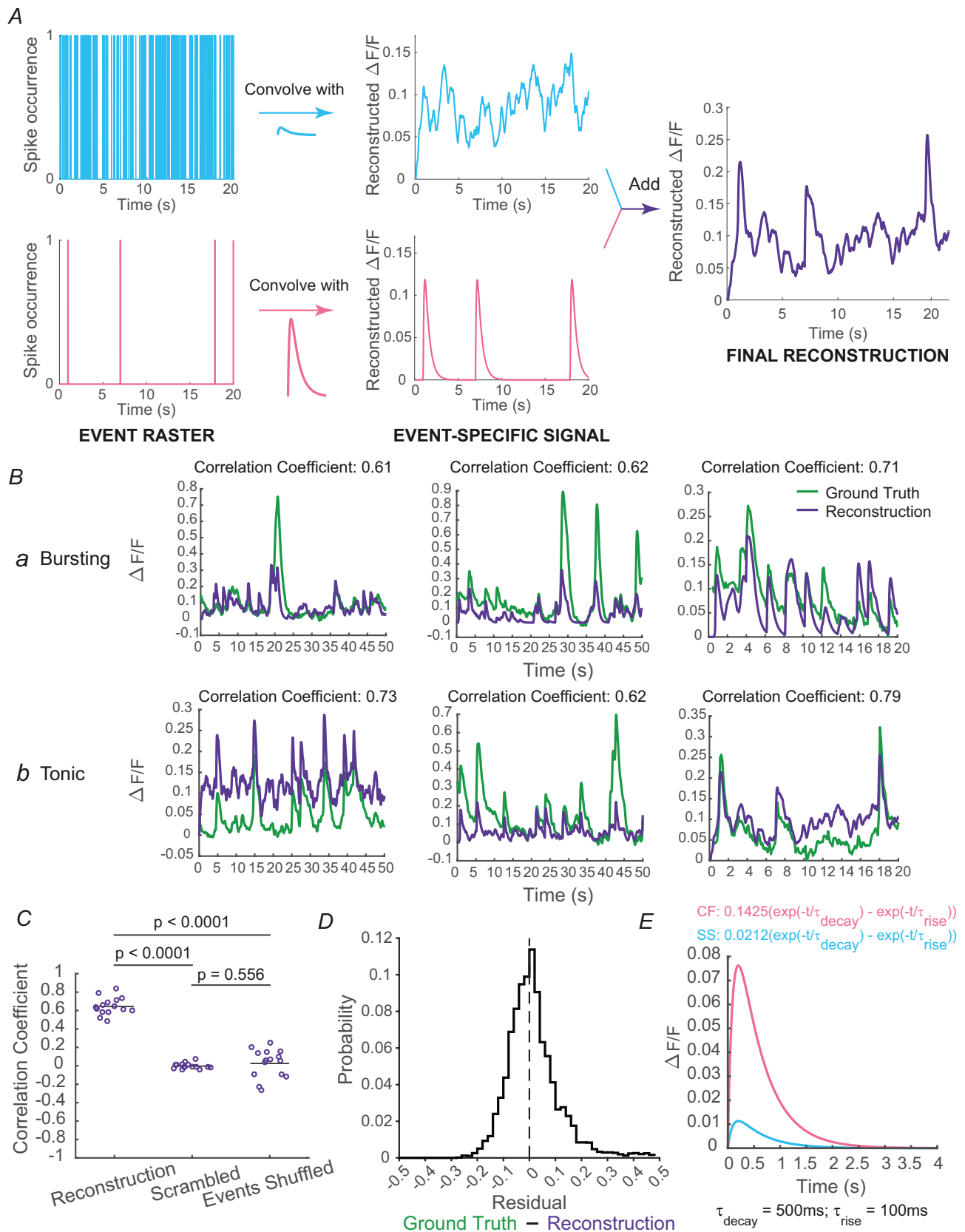
Although this is a simple forward model with calcium dynamics assumed to be linear and time-invariant, we found that this method to approximate the calcium signal works well, no matter the state of the neuron (Fig. 3B). The quality of the reconstructions was quantified by measuring the correlation coefficient between the ground truth and the reconstruction (Fig. 3C). To benchmark the correlation coefficient values obtained against those obtained simply by chance, we also calculated the

correlation coefficient between the ground truth and a scrambled version of the reconstruction, as well as that between the ground truth and a reconstruction generated after shuffling all the event times in the recording (Fig. 3C). In both the latter cases, the correlation disappears, suggesting that the reconstructions generated by this forward model are indeed faithful reconstructions. We next calculated the residuals between ground truth and reconstructions, the distribution for which is shown in Fig. 3D. It is not normally distributed and has a positive skew, which means that our reconstructions tend to underestimate the true signal. Nevertheless, the small values of the residuals further lend support to the idea that the forward model of calcium signal reconstruction works well.

Given that only groups/bursts of simple spikes produce a calcium transient amplitude comparable to that generated by an individual CF input (Knogler et al., 2019) (Fig. 2B), one would expect a much smaller response kernel for simple spikes than for the CF input. Indeed, the optimal kernels found by this method show that the CF input response kernel is ~ 7.5 times larger than the simple spike response kernel (Fig. 3E).

A state-labelled calcium signal dataset was generated by reconstructing the calcium signal for a compilation of electrophysiological traces

Having a method of converting any given PN electrophysiology trace to its corresponding calcium imaging trace allowed us to generate simulated calcium imaging traces for a compilation of all electrophysiological recordings acquired in our lab (Fig. 4A). To do so, we first pooled all the electrophysiological recordings, both extracellular and whole-cell, made during the period 2011–2020, for a total of $N = 173$ recordings. Recordings of poor quality ($N = 37$) (i.e. those with interruptions, or noisy baselines, unusually broad/filtered events, or indistinguishable simple spikes and CF inputs) were first excluded from the dataset. Retained recordings then had a state assigned to them (using criteria described above; see also Fig. 1D–G), including state switches, if there were any. This collection was split into two sets: one with those recordings that had only one state throughout ($N = 131$), and another with those that had a state switch ($N = 5$). The entire complement of recordings and their corresponding states is summarized in Table 1. Events detected in all the recordings were used to reconstruct a putative calcium signal using the kernels described in Fig. 3E. This finally produced two labelled datasets: a state-labelled reconstructed calcium signal dataset and an independent dataset with switches (Fig. 4A). The recordings used to generate this dataset were of different durations with a median duration of 60 s. To obtain multiple samples



for simple spikes, magenta for CF inputs) to generate an event-specific calcium signal time series (middle). These are then simply added together to get the final reconstruction (right). *B*, comparison of ground truth (green) and reconstructed (purple) calcium signals for three randomly chosen bursting cells (*Ba*) and three randomly chosen tonic cells (*Bb*). The Pearson's correlation coefficient between the two traces is marked above the plot for each pair. *C*, correlation coefficients for all cells calculated for either the true reconstruction, a scrambled version of the reconstruction, or the reconstruction for when simple spike and CF input event times were shuffled. A Kruskal–Wallis test yielded a *P* value of 2.108×10^{-7} , which was followed by a *post hoc* Dunn's test for individual comparisons, the latter of which is marked on the figure ($n = 15$ cells; six bursting, nine tonic). *D*, distribution of residuals between the ground truth and reconstructed calcium signal, pooled across all cells ($n = 15$ cells). *E*, the optimal calcium sensor kernels for simple spikes (cyan) and CF inputs (magenta). The formula for each kernel is mentioned above the plots, and the values for rise and decay time constants are mentioned below the plot.

per recording, we decided to split the traces into 10 s long non-overlapping chunks, representative samples from which are shown in Fig. 4*B*. This subsampling yielded a dataset of state-labelled calcium imaging traces ($n = 928$ traces of which $n = 580$ bursting, $n = 348$ tonic) that is much larger than could have been obtained experimentally. The entire complement of non-overlapping samples is shown as a heatmap in Fig. 4*C*.

To test whether there were properties of these traces that could be used to unambiguously classify their source state, we compared the number of peaks, average peak amplitude, area under the curve, mean, SD and coefficient of variation of the reconstructed calcium traces (Fig. 4*D*). Except peak amplitude, all other properties were significantly different from each other for the tonic and bursting groups. Nevertheless, given any of these properties, it was not possible to unambiguously assign state because the distributions are overlapping (Fig. 4*D*). Furthermore, a principal component analysis of these trace features also failed to separate tonic and bursting cells based on the above-mentioned features of their calcium traces (Fig. 4*E*).

Machine learning methods have successfully been used to solve such problems when the data points are not linearly separable (He et al., 2016). Thus, we turned to developing a machine learning model to identify the state of a cell from a calcium signal time series.

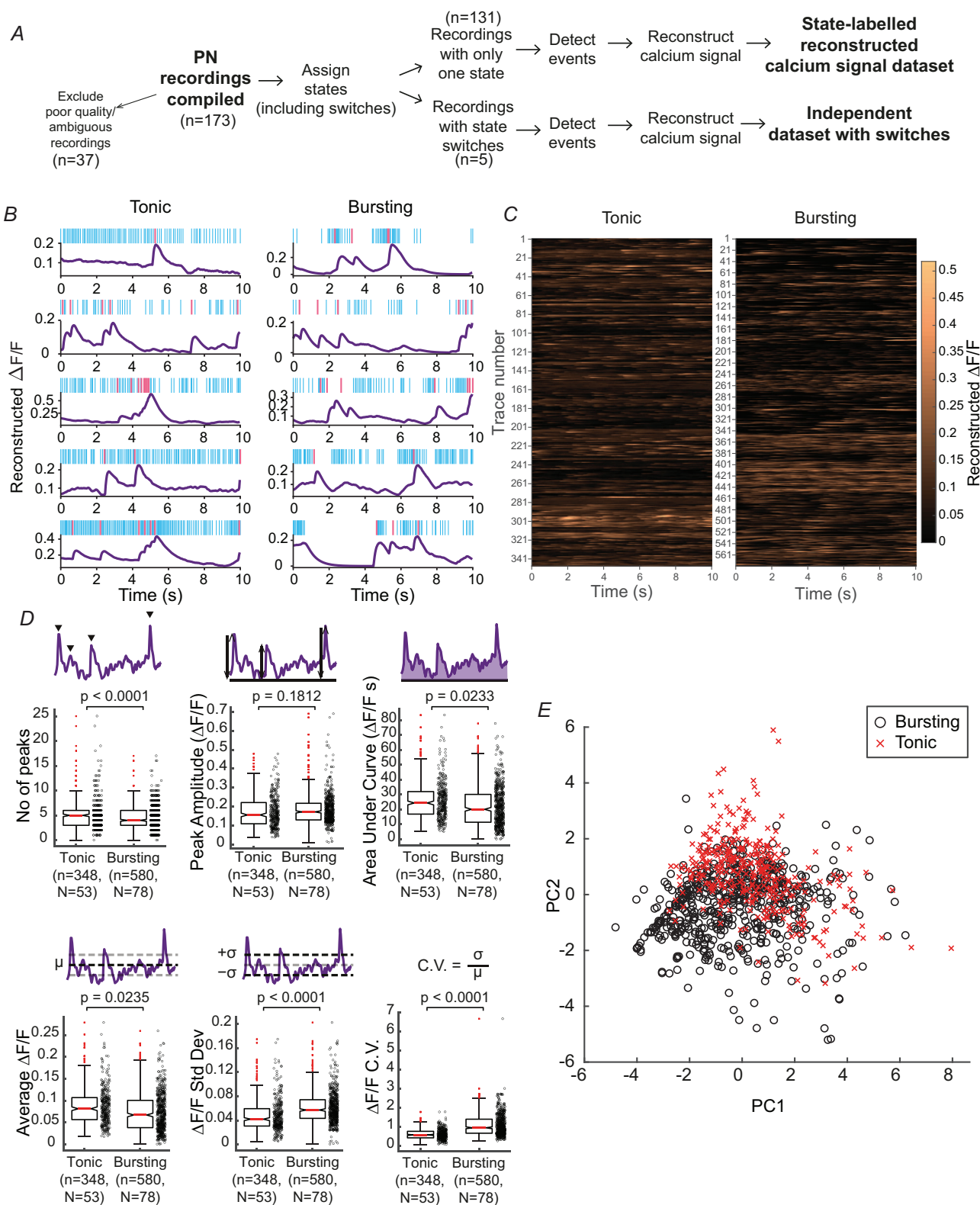
Training a convolutional recurrent neural network to distinguish state from calcium imaging data alone

To find good machine learning models to solve this classification problem, we performed five-fold cross-validation, which is used to ensure that models generalize well and do not overfit the training data (Fig. 5*A*). On each fold, only 60% of the training traces were shown to the network in the training phase (Fig. 5*A*, white boxes), the network was optimized on 20% of data used as a validation set (Fig. 5*A*, yellow boxes) and the generalizability of the trained network was assessed using the remaining 20% (Fig. 5*A*, blue boxes). Each of the five folds had a different 60:20:20 split, such that the training

and cross-validation datasets had a balance of traces from both classes.

We tested several different machine learning models on this classification problem: support vector machines (SVMs), deep neural networks (DNN), convolutional neural networks (CNNs) and convolutional recurrent neural networks (CNN-LSTM; i.e. a convolutional neural network with long short-term memory). The prediction accuracies and F1 scores (a metric that adjusts for class imbalances, as well as whether a prediction was a true/false positive/negative, as defined in the methods section) for the various models tested are shown in Table 2. Of these models, the 1D-CNN and CNN-LSTM models performed best across all folds. However, the former only predicts from snapshots of 10 s long snippets of the calcium trace, and the networks had more erroneous performance when predicting on the entire time series (data not shown). The CNN-LSTM model, however, also incorporates a recurrent neural network module (Hochreiter & Schmidhuber, 1997), which has the ability to learn features across an entire sequence and not just across neighbouring positions in a trace sequence, as CNNs do.

The architecture of the network is shown in Fig. 5*B*. The input to the network is a trace sampled at 30 Hz, which is some multiple of 10 s in duration. It is first remapped to values between 0 and 1 using min-max normalization and then processed by a CNN, which acts as a local feature extractor that also down samples the input data, such that it produces four vector outputs at a sampling frequency of 1 Hz. These outputs are passed to a bidirectional LSTM, one time-step at a time. A forward LSTM cell looks at trends in the forward direction (Fig. 5*B*, blue dotted box and arrow), whereas a reverse LSTM cell goes through the data in the reverse direction (Fig. 5*B*, red dotted box and arrow). Each of the two cells produces a 'hidden state' as its output. Given that the LSTM is a recurrent neural network, this output is fed back into the network along with the next appropriate time step from the CNN's output, a process which is repeated nine more times (for every 10 s trace). In essence, the LSTM is detecting trends across the entire sequence. The final hidden states from the forward and reverse LSTMs, produced after passing through the entire trace, are retained and concatenated.



dataset, for each state, after splitting all traces into 10-s-long non-overlapping chunks. ($n = 348$ tonic traces; $n = 580$ bursting traces). *D*, comparing the distribution of trace properties for both states. The top three plots show (from left to right) the number of peaks, peak amplitude and area under the curve of the reconstructions in the dataset. The bottom three plots show (from left to right) the mean, standard deviation and the coefficient of variation of $\Delta F/F$ values. n values are the number of 10 s long samples. N values are the number of cells used to generate the samples. P values were computed using linear mixed-effects models. *E*, principal components analysis of trace properties, with data points marked by state: tonic (red crosses) and bursting (black open circles).

Table 2. Prediction accuracies and F1 scores for various models trained to solve the classification problem

Model	Fold	Validation		Test	
		Accuracy (%)	F1 score	Accuracy (%)	F1 score
Support vector machine (SVM)	1	81.41	0.7206	70.19	0.6077
	2	72.68	0.593	79.66	0.6842
	3	75.84	0.703	72.55	0.6055
	4	73.10	0.6422	73.65	0.671
	5	70.67	0.629	74.89	0.6755
	Mean \pm SD	74.74 \pm 4.16	0.6576 \pm 0.0531	74.19 \pm 3.51	0.6488 \pm 0.0388
Deep neural network	1	83.91	0.7893	72.99	0.6649
	2	77.71	0.7205	82.89	0.7849
	3	74.85	0.7099	76.10	0.6788
	4	73.31	0.6535	75.12	0.7143
	5	77.82	0.7060	75.24	0.6629
	Mean \pm SD	77.52 \pm 4.06	0.7158 \pm 0.0486	76.47 \pm 3.77	0.7012 \pm 0.0511
1D-CNN	1	87.20	0.8163	84.00	0.7933
	2	80.67	0.7541	88.33	0.8336
	3	82.66	0.7880	80.09	0.7263
	4	78.28	0.7185	81.07	0.7719
	5	82.60	0.7604	78.34	0.7004
	Mean \pm SD	82.28 \pm 3.28	0.7675 \pm 0.0369	82.37 \pm 3.91	0.7651 \pm 0.0530
1D-CNN (3 \times)	1	87.54	0.8226	84.00	0.7893
	2	80.61	0.7535	86.97	0.8142
	3	80.61	0.7650	78.41	0.7089
	4	78.62	0.73	73.66	0.6592
	5	82.49	0.7592	76.83	0.6775
	Mean \pm SD	81.97 \pm 3.40	0.7661 \pm 0.0343	79.97 \pm 5.42	0.7298 \pm 0.0686
CNN-LSTM	1	80.15	0.7227	79.7	0.7570
	2	91.55	0.8788	92.75	0.8937
	3	79.06	0.7233	92.12	0.887
	4	83.25	0.7878	75.9	0.7219
	5	79.9	0.7302	82.02	0.8071
	Mean \pm SD	82.78 \pm 5.15	0.7686 \pm 0.0674	84.50 \pm 7.57	0.8133 \pm 0.0766

These outputs are combined by passing them through a linear layer, which performs a matrix multiplication to produce an output that is 10 samples long. At this step, each sample represents the likelihood that the trace in question comes from a cell in the bursting mode, predicted at 1 s intervals.

The likelihood of each time step having come from the tonic state can then be inferred (Fig. 5B, greyscale boxes on the right). The state with the larger likelihood score is taken to be the final classification at each time step.

Five networks were produced, one for each fold of cross-validation. The median accuracy was 80.15% for

the cross-validation phase and 82.02% for the test phase (Fig. 5C). The median F1 score was 0.7302 for the cross-validation phase and 0.8071 for the test phase (Fig. 5C). Next, we looked at the receiver operating characteristic (ROC) curve for each network fold. This curve plots the false positive rate against the true positive rate of the classifier for various classification threshold values, and the area under this curve is an indicator of the quality of the classifier. The area under the ROC curve (AUC) for each of the five network folds is plotted in Fig. 5E. The AUC values for all folds on both cross-validation and test data are much higher than 0.5,

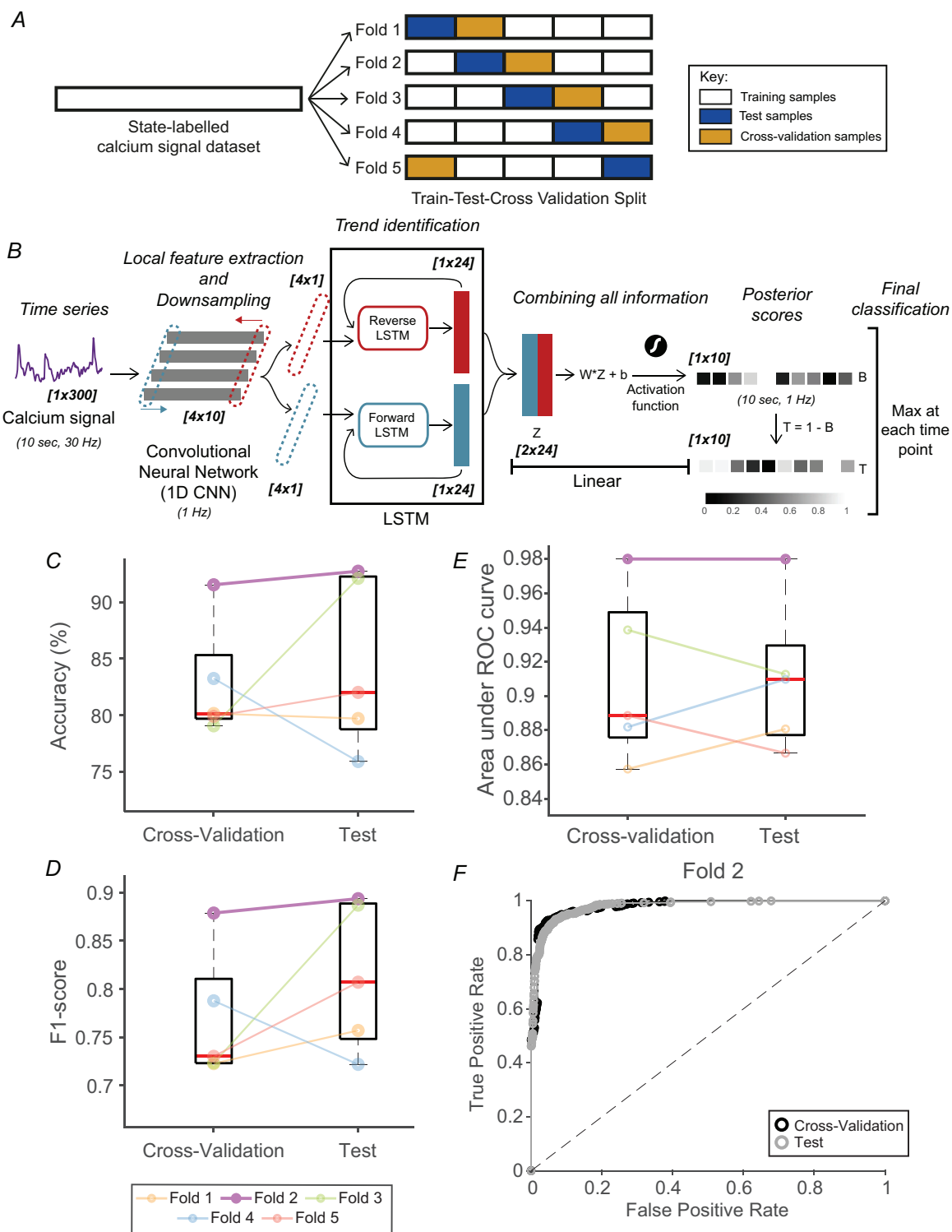


Figure 5. Design and training of CaMLsort

A, flowchart showing how the state-labelled calcium signal dataset was split into training and test datasets using a five-fold 60:20:20 split into training (white boxes), test (blue boxes) and cross-validation (yellow boxes) samples, respectively. The networks that performed well in the training phase, as assessed by performance on the cross-validation samples, were then challenged with the unseen test data (blue boxes) in each fold. An independent dataset with state switches was also used to further validate the networks obtained. **B**, the architecture of the convolutional recurrent neural networks used to solve the classification problem. The network takes an input time series trace sampled at 30 Hz (purple) and that is a multiple of 10 s long in duration (here just 10 s long). This trace is first normalized using min-max normalization, following which it is passed through a 1-D convolutional neural

network (1D-CNN), which has four kernels with a step size and stride length of 30 each. The resulting four traces (grey boxes) are down-sampled, local feature-extracted versions of the original trace. All four outputs are passed to the LSTM module (central black box) for trend identification. The LSTM is a bidirectional one, which processes traces in both the forward (blue arrow) and reverse (red arrow) directions. A single time step is taken from all the CNN outputs (red and blue dotted boxes) and passed to the respective LSTM cell, which then processes the input to produce a resultant 'hidden state' output (red and blue filled boxes). This output is passed back into the LSTM cell along with the next sample from the CNN's output traces, until every time step from it has been processed (here 10 rounds). Once this is done, the final 'hidden state' from each LSTM cell is retained and concatenated before their weighted average is taken by a linear layer. The resulting average is converted to a 'posterior score' using an activation function. These posterior scores represent the likelihood of the cell being in the bursting state at each time step. The likelihood of it being in the tonic state can be calculated from this. The final call is taken to be the state, which has a higher posterior score. The text in italics represents the net effect of each phase of the neural network. The numbers in parentheses indicate the sampling frequency and/or the duration of the resultant vectors at various stages of processing by the neural network. Similarly, the numbers in bold and italics within square brackets indicate the size of the vectors/matrices at various stages of passing through the network. *C*, average classification accuracy of trained networks at the cross-validation and the test phases. Each fold is represented in a different colour, marked in the legend at the bottom of (*D*). *D*, average F1 scores for each of the trained networks at the cross-validation and the test phases. *E*, area under the ROC (receiver operating characteristic) curve for each of the five trained CNN-LSTM networks for both cross-validation (black) and test (grey) data. *F*, ROC curve for Fold 2 of the CNN-LSTM for cross-validation (black) and test (grey) data.

indicating that these networks have been trained well and make accurate predictions.

Of all the network folds, Fold 2 looked particularly promising because it (i) consistently outperformed the other folds with respect to classification accuracy and F1 scores (Fig. 5C and D, purple) and (ii) had higher AUC values than the other network folds (Fig. 5E and F). Thus, we decided to retain this network for future use, and have dubbed it CaMLsort (Calcium imaging and Machine Learning based tool to **sort** intracellular state).

CaMLsort performs well on previously unseen calcium signal reconstructions

To ensure that networks weren't performing well just because of peculiarities in the dataset used for training and testing, we presented an independent set of calcium signal reconstructions to the network that were not used in any phase of the training. These reconstructions ($n = 19$) also had state switches in them. The ground truth class labels for each of the 19 cells (Fig. 6A, top) were compared with the raw predictions from the network (Fig. 6A, middle).

At first glance, it appears that the network is not making accurate calls because each cell's predicted state alternates between both states. However, these fluctuations are transient and the overall F1 scores for the raw predictions are >0.5 for 13 out of the 19 cells (median score 0.777). The distribution of F1 scores is shown in Fig. 6C. On closer observation, it is clear that the six cells for which F1 scores were low (< 0.5) had a switch in state during the recording. The poor prediction accuracy for these can be explained by the fact that the training dataset did not have any cases with state switches.

Given our prior knowledge that the state of a neuron doesn't flicker rapidly between tonic and bursting, we

decided to smoothen out the raw predictions using a majority voting strategy. For this, we took a rolling window of length N around each time step of the prediction and asked what the majority of the network calls were across all the windows that included the time step. If the average posterior scores for the majority class were higher than the average posterior scores for the minority class, we would confidently assign the majority class to that time step. If not, no confident call would be made for that time step. For example, to make a confident call at time t_0 with a window length of 5, we first extract all the five-step windows that include time t_0 (there will be five such windows, except at the start and end, where there will be fewer). Next, we average the posterior scores within each of the windows and assign the 'average state' as the state with the higher posterior score. In this case, let's say these were 2:3 in favour of bursting. If the average bursting posterior score for the three 'bursting' windows was greater than the average tonic posterior score for the two 'tonic' windows, the state for time t_0 would be confidently assigned as 'bursting.' If not, the prediction at time t_0 would be left blank.

Using this majority voting strategy with a window length of 7, we were able to make confident calls for $>90\%$ of the duration for all 19 cells. As shown in Fig. 6B, the jitter in the raw CaMLsort predictions was considerably smoothened out by this process, and the F1 scores showed a marginal improvement as well, with a median F1 score of 0.8194 (Fig. 6C). As before, the cases where F1 scores < 0.5 corresponded to recordings with state switches. As an extra control, we obtained CaMLsort predictions from signals that were scrambled. These predictions had much lower F1 scores (Fig. 6C). Indeed, when presented with scrambled traces, CaMLsort almost always (99.76% of the time) predicted the state to be tonic, with a very large, consistent differential between the posterior scores for

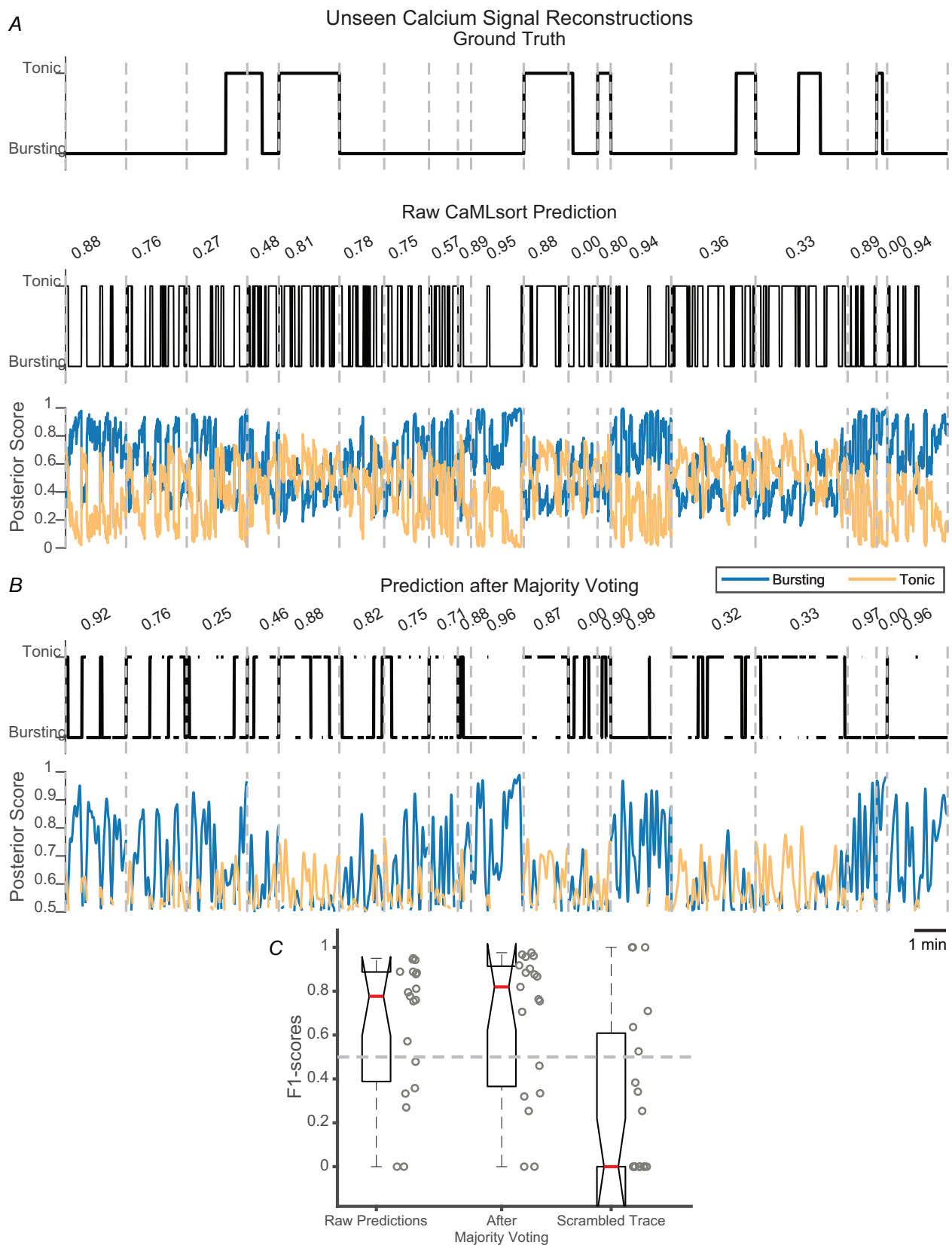


Figure 6. CaMLsort predictions from previously unseen calcium signal reconstructions

A, raw CaMLsort class label predictions from previously unseen calcium signal reconstructions ($n = 19$ recordings). Predictions from CaMLsort (middle) are compared against the ground truth state (top) at each time step. Predictions

for each recording were made independently, but have been stitched here for the purposes of representation. Each recording is separated by a vertical dashed line. The number above each prediction is the F1 score for that recording. The bottommost plot has posterior scores for each time step, with those for the 'bursting' class in blue, and those for the 'tonic' class in light orange. *B*, CaMLSort predictions (top) that remain after majority voting with a window length of seven steps has been used to retain only confident classifications (for details, see Methods and Results). The posterior score (bottom) for each class at each time step, taken as the average class posterior score across all seven-step windows that included said time step. As in (*A*), the posterior scores for the 'bursting' class are in blue, whereas those for the 'tonic' class are in light orange, and the numbers above each prediction represent the F1 scores for that recording. *C*, the distribution of F1 scores between the ground truth state and various predictions from CaMLSort, either the raw predictions (left), predictions obtained after majority voting (middle) or the raw predictions obtained by passing a scrambled version of the calcium signal as an input. The notches indicate 95% confidence intervals around the median (red line).

the tonic (mean \pm SD: 0.7810 ± 0.0565) and bursting (mean \pm SD: 0.2190 ± 0.0565) classes. This implies that CaMLSort has, indeed, captured a latent trend present only in the true signal and not in a collection of random numbers of comparable magnitude.

CaMLSort performs well on calcium signal reconstructions from simulated spike trains

As a further challenge to the trained networks, we decided to generate artificial spike trains that represent canonical cases of the tonic or bursting class, reconstruct the calcium signal for these spike trains and ask the network to classify these completely synthetic data (Fig. 7). We simulated PN simple spiking using either a Poisson model (tonic state) or by randomly sampling from the distribution of inter-event intervals from experiments (bursting state; also see Methods). The CF input spike train was considered to be independent and was generated using a Poisson model. The ground truth states for these simulated spike trains is shown in Fig. 7A (top trace). The raw CaMLSort predictions obtained for these data had a median F1 score of 0.8571 (Fig. 7A, middle trace and 7C), although these too had several reversals in the state predictions. On applying the majority voting strategy described above, however, these reversals were smoothed out (Fig. 7B), yielding a median F1 score of 0.9189 (Fig. 7C) and a confident prediction for at least 80% of the recording duration in every cell. As was observed in the previous case, scrambling the signal prior to prediction by the neural network were all of one class (tonic), yielding F1 scores that were either 1 (all right) or 0 (all wrong) (Fig. 7C).

Neural networks trained only on simulated calcium signals correctly identify PN state from real data

Thus far, all the inputs and challenges to the network were made using calcium signal reconstructions. To test the applicability of these networks on data from experiments, we went back to our original simultaneous electrophysiology-imaging data and asked if the network

could call the state of those cells correctly based on the $\Delta F/F$ traces alone.

We found that CaMLSort predicted the state of a cell from imaging data well (Fig. 8A), with a median F1 score of 0.7654 across all cells (Fig. 8C). This outcome is interesting, because the networks have only been exposed to simulated calcium signals generated using a first-order generative model in the training and test phases, and yet are able to reasonably assess the state from real data, which have several challenging features, including variable sensor expression levels, shot noise and signal bleaching, to name a few.

Majority voting did not substantially improve the overall F1 score across all cells (median 0.75). However, by retaining only confident calls, the flip-flopping of network predictions was markedly reduced (Fig. 8B), as was observed before in the case of classifications using reconstructed calcium signals.

To benchmark the quality of these predictions, we obtained CaMLSort predictions from calcium signals reconstructed from the spike trains of these cells (since CaMLSort was trained only on simulated calcium signals). The prediction quality from reconstructions was comparable to that obtained from raw imaging data (median F1 score of 0.9011) (Fig. 8C). Therefore, although CaMLSort is clearly able to classify cellular state from experimentally-obtained imaging data, its performance has room for improvement. Also consistent with our previous results were the predictions from scrambled $\Delta F/F$ traces, which were almost all (96.83%) tonic, thereby yielding F1 scores that were mostly either 0 or 1 (Fig. 8C).

In sum, we showed that CaMLSort has learnt to sort the cellular state of a PN based only on its calcium signal and is capable of producing confident calls of high quality.

CaMLSort generalizes well to calcium imaging data from other systems

Having developed a tool that can reasonably predict the bistable state of a PN given its calcium signal, we asked whether it could be broadly applicable to other neurons, too. Dopaminergic neurons of the

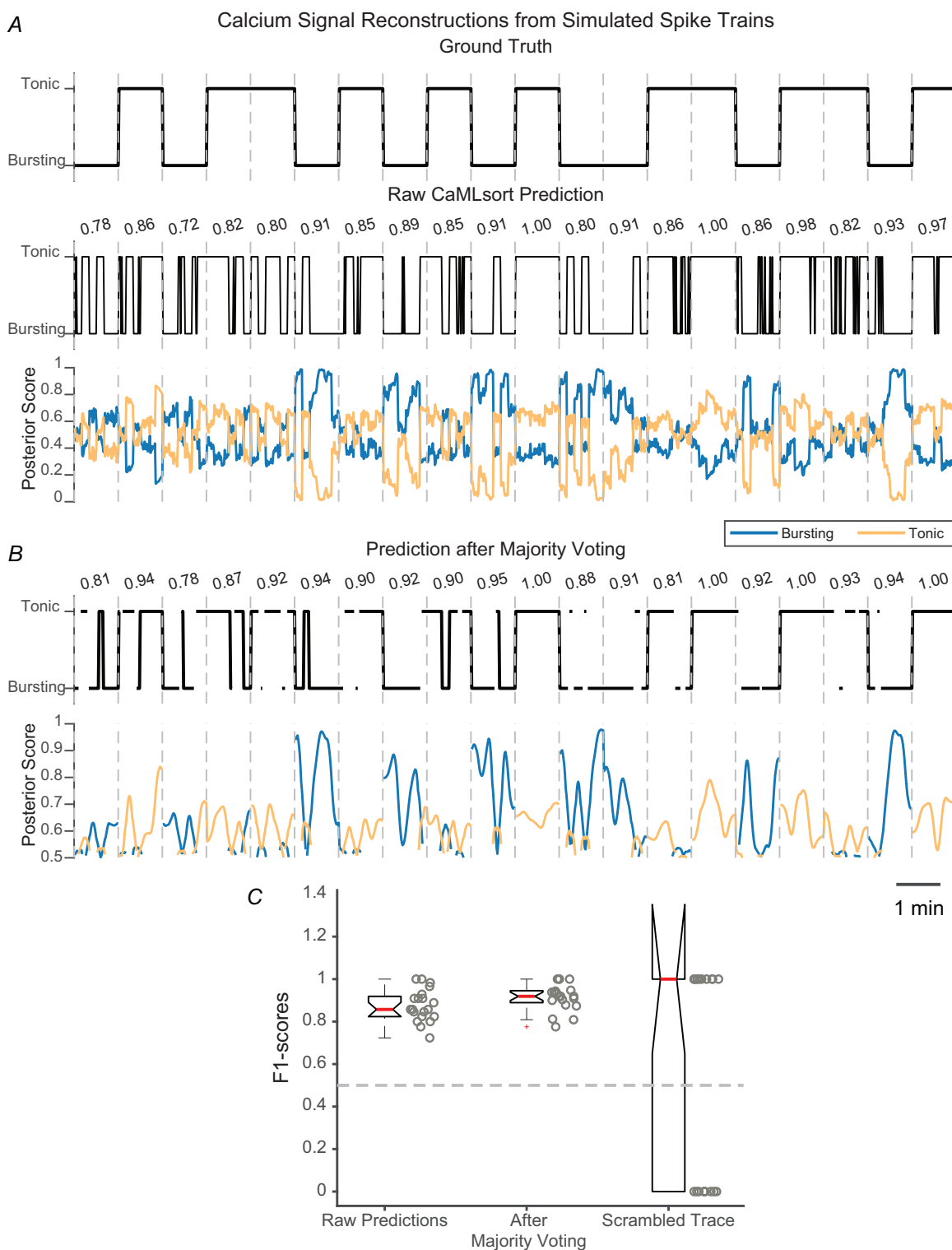


Figure 7. CaMLsort predictions from calcium signal reconstructions of simulated spike trains

A, raw CaMLsort class label predictions from previously unseen calcium signal reconstructions that were generated from simulated spike trains ($n = 20$ recordings). Predictions from CaMLsort (middle) are compared against the ground truth state (top) at each time step. Predictions for each recording were made independently, but have been stitched here for the purposes of representation. Each recording is separated by a vertical dashed line. The number above each prediction is the F1 score for that recording. The bottommost plot has posterior scores for each time step, with those for the 'bursting' class in blue, and those for the 'tonic' class in light orange. *B*, CaMLsort predictions (top) that remain after majority voting with a window length of seven steps has been used to retain

only confident classifications (for details, see Methods and Results). The posterior score (bottom) for each class at each time step, taken as the average class posterior score across all seven-step windows that included said time step. As in (A), the posterior scores for the 'bursting' class are in blue, whereas those for the 'tonic' class are in light orange, and the numbers above each prediction represent the F1 scores for that recording. C, the distribution of F1 scores between the ground truth state and various predictions from CaMLsort, either the raw predictions (left), predictions obtained after majority voting (middle) or the raw predictions obtained by passing a scrambled version of the calcium signal as an input. The notches indicate 95% confidence intervals around the median (red line).

VTA in mice also fire either tonically or phasically (Grace & Bunney, 1984a, 1984b). Fleming et al., 2021 performed simultaneous electrophysiology and calcium imaging from these neurons, which expressed either GCaMP6f or GCaMP6m. This dataset is available publicly (https://github.com/jewellsean/spike_tools) and we tested CaMLsort on this dataset.

We first compiled the dataset and sorted cells as tonic or bursting, using a combination of visual inspection of spike rasters (Fig. 9A and B) and the coefficient of variation of interspike intervals (Fig. 9C), as we did for PNs. Three independent reviewers used this method to classify states and a consensus was taken across reviewers, at the same time as taking into consideration the state labels provided in their study metadata. Data were included only if at least two out of the four labels (three reviewers plus original metadata) agreed. Using this method, 78/84 cells (92.86%) of the original dataset were included for further analysis. Of these, 16 cells were in the tonic state, whereas the remaining 62 cells were in the bursting state.

Before attempting to use CaMLsort on this data, we considered whether the convolution-based forward model of reconstructing calcium signals from electrophysiology would work on this dataset. Because there is only a single type of spike in these neurons, we used only one kernel, the amplitude, half-rise time constant and half-decay time constant, which were taken to be the mean values for GCaMP6f or GCaMP5G (Chen et al., 2013). No optimization of kernels was performed.

Visual inspection shows that reconstructions obtained from both GCaMP6f and GCaMP5G kernels were comparable to the signals imaged from VTA DA neurons, with the only difference between them being their amplitudes (Fig. 9D). To quantify the similarity between the ground truth and the two simulated calcium signals, we used the Pearson correlation coefficient. The correlation coefficients between the ground truth and reconstructions from both GCaMP6f and GCaMP5G kernels were almost identical, with a value of 0.6400 ± 0.3189 , on average, for both (Fig. 9E). Reconstructions that were either scrambled or generated after shuffling the DA neuron's spike times had no correlation with the ground truth signal, as expected (Fig. 9E).

Having confirmed that the dataset of Fleming et al. (2021) had both tonic and bursting data, and that the

mapping from electrophysiology to imaging was also possible with this data, we were ready to test CaMLsort on it. As before, we first interpolated the calcium signals to get a sampling frequency of 30 Hz before inputting them to CaMLsort. Raw predictions from CaMLsort and their corresponding ground truth labels for a subset of 12 of the 78 cells are shown in Fig. 10A. The distribution of F1 scores obtained from all 78 cells, the median for which is 0.6841, is shown in Fig. 10B.

We also calculated F1 scores for predictions made on calcium signals reconstructed from true electrophysiology data using the GCaMP5G kernel (Fig. 10B). We decided to use the 5G kernel because the reconstructions that CaMLsort was trained on followed kinetics of GCaMP5G. The predictions from signal reconstructions were comparable to those from raw imaging data, with a median of 0.8889. Lastly, if the imaging signal was scrambled before prediction, CaMLsort predicted a default 'tonic' state 89.03% of the time, yielding F1 scores that were mostly either 0 or 1 (Fig. 10B).

In all, our results from VTA DA neurons in mice were similar to those from zebrafish PNs, which means that CaMLsort generalizes well and will probably find broad use in classifying the cellular state across systems. To facilitate this use, we have made CaMLsort publicly available as a GitHub package (<https://github.com/vatsalat/CaMLsort23>).

Discussion

In the present study, we describe the development of CaMLsort, a machine learning-based tool to identify the bistable state of PNs from calcium imaging data. We develop this network and train it only on a simulated calcium signal dataset, but find that it can be used on experimental data, too. Furthermore, we demonstrate that CaMLsort is able to classify bistable states in a completely different cell type (dopaminergic neurons) from a different brain region (ventral tegmental area) in a different model system (mice), and using a different version of the calcium sensor (GCaMP6) than it was originally trained on. Although this demonstrates that CaMLsort could have wider application, more broader testing on other datasets will be essential to prove generalizability.

Methods and Results). The posterior score (bottom) for each class at each time step, taken as the average class posterior score across all seven-step windows that included said time step. As in (A), the posterior scores for the 'bursting' class are in blue, whereas those for the 'tonic' class are in light orange. C, the distribution of F1 scores between the ground truth state and various predictions from CaMLSort: raw predictions from the imaging data (far left), predictions obtained after majority voting (second from left), raw predictions obtained from calcium signals reconstructed from spike times (third from left) and those predictions from imaging $\Delta F/F$ traces that were scrambled prior to prediction. The notches indicate 95% confidence intervals around the median (red line).

Bistability in PN populations

The ability of neurons to exist in two stable states and exhibit distinct firing patterns/excitabilities in those states has been documented in invertebrate (Dickinson & Nagy, 1983; Lechner et al., 1996; Marder et al., 1996; Russell & Hartline, 1982) and vertebrate (Cowan & Wilson, 1994; Hounsgaard et al., 1984; Loewenstein et al., 2005; Misgeld et al., 1986; Schwindt & Crill, 1980) nervous systems.

We previously showed that larval zebrafish PNs are bistable and that they switch between firing either tonically or in bursts (Sengupta & Thirumalai, 2015). Such behaviour has also been reported in many other cell types such as thalamic (Jahnsen & Llinás, 1984) and cortical neurons (McCormick et al., 1985).

Although it is appreciated that bistability increases the information processing capacity of neurons, such as by allowing gating of synaptic inputs (Thirumalai & Jha, 2022), its implication for network computation or behavioural output is less well understood. This is true of PNs as well. Although it is established that these neurons are bistable and are important for motor behaviours (Engbers et al., 2013; Popa et al., 2018; Sengupta & Thirumalai, 2015), the role of bistability in PN computation or motor behaviour output has thus far not been clear.

We propose that having more than one stable state allows a cell to toggle in and out of participating in circuit computations. In the cerebellum, specifically, bistability may have significant consequences by modulating PN neuronal synchrony with other PNs. PNs make weak inhibitory synapses onto deep cerebellar nuclear (DCN) cells in mammals and eurydendroid cells in fish, with a high degree of convergence (Bae et al., 2009; De Zeeuw et al., 2011; Harmon et al., 2020; Ikenaga et al., 2006; Person & Raman, 2012). DCN cell spike times could be effectively controlled via inhibition from several PNs firing synchronously (Person & Raman, 2012). Since PNs fire tonically when synaptically-isolated and CF inputs can trigger bursts (Sengupta & Thirumalai, 2015), the bursting state is a network-embedded state whereas the tonic state is independent of network drive. By toggling between these two states PNs can increase or decrease the levels of population synchrony and thereby affect DCN firing differentially. Mapping state distribution across the PN population during various tasks will lead us to an understanding of how PNs synchronize or desynchronize during

these tasks and how that in turn shapes cerebellar output and motor behaviours.

Inferring long-term dynamics from calcium imaging data

Calcium imaging has been one of the core experimental techniques in neuroscience for more than three decades and has seen many major developments in sensor quality, microscopy and analytical algorithms and software (Chen et al., 2013; Göbel & Helmchen, 2007; Peron et al., 2015; Renninger & Orger, 2013; Tian et al., 2009; Yang & Yuste, 2017). However, the technique is largely used to identify regions of the brain that respond to presented stimuli or correlate with behavioural output under well-defined experimental conditions. This results from the fact that the predominant mode of transient calcium elevation occurs because of the opening of voltage-gated calcium channels during action potential firing (Borst & Helmchen, 1998; Göbel & Helmchen, 2007). Thus, these calcium transients are quite brief, typically lasting only tens of milliseconds and report spiking activity of neuronal populations during sensory stimulation and/or motor output. However, it is clear that neurons also exhibit longer time scale dynamics such as bursting, which carries information for downstream neurons (Zeldenrust et al., 2018). Because the currently available calcium sensors themselves are much slower compared to the calcium transients, it is obvious that the recorded $\Delta F/F$ signals must manifest signatures of such long-term dynamics. CaMLSort addresses this issue by looking at longer term signatures in the calcium imaging time series data to identify cellular state.

Broader applications of CaMLSort

The development of CaMLSort has opened up the possibility for much future research with respect to understanding PN bistability and its role in the function of the cerebellum as a whole. For example, we can now experimentally address what causes a switch in a PN's state by carrying out a high-throughput screen for chemical agents that affect bistability. We know that bistability is a cell-intrinsic property, whereas state switches could be a result of modulation of network activity. Neuro-modulators are key candidates for this because they can

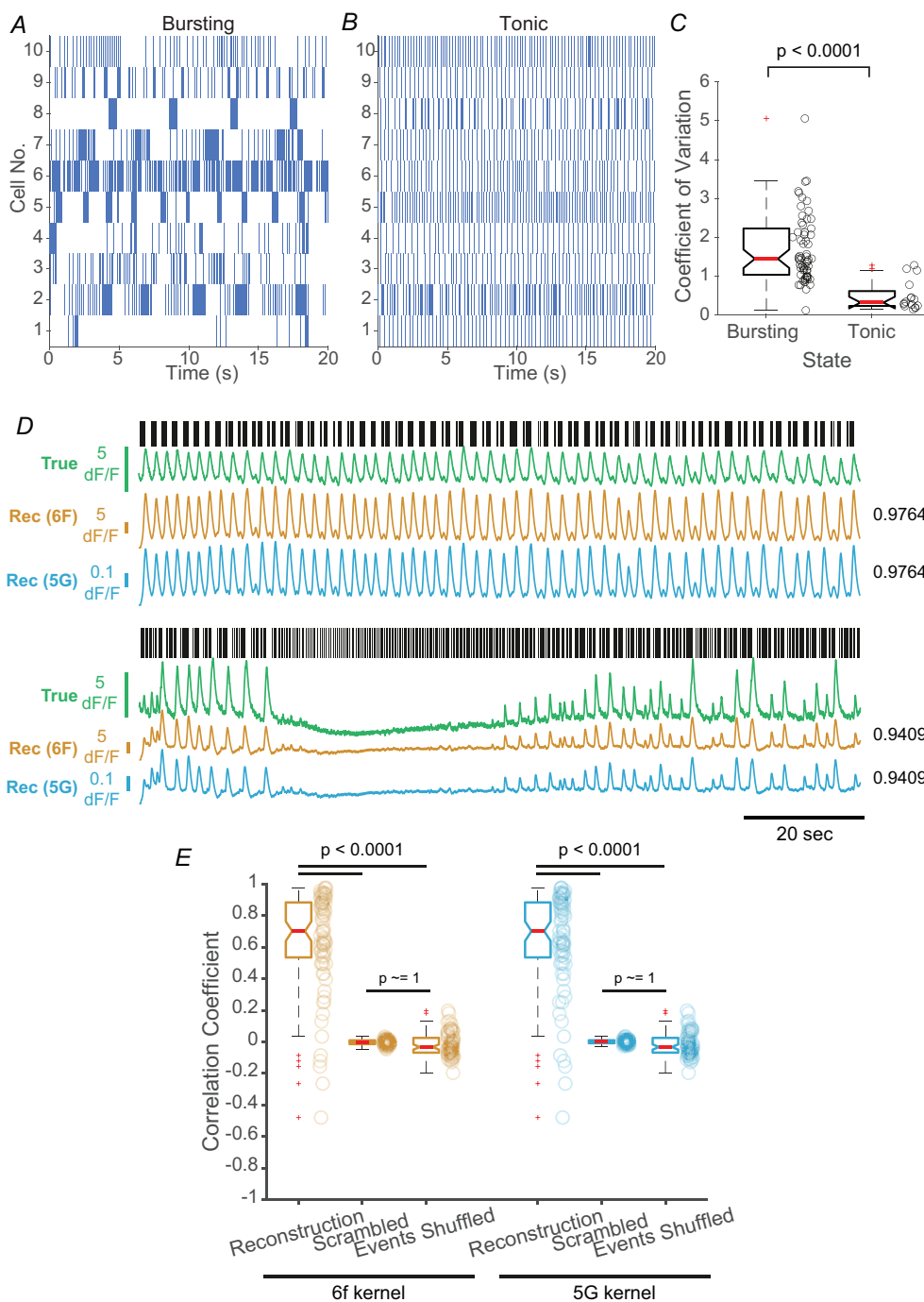


Figure 9. Reconstructing calcium signals for VTA DA neurons using a convolution-based forward model A and B, spike rasters of two representative VTA DA neurons. C, the distributions of C.V.s of interspike intervals for cells classified as either bursting ($n = 62$) or tonic ($n = 16$ cells). The P value was calculated using a one-sided Mann–Whitney U test. D, electrophysiological activity and associated calcium signals from two representative VTA DA neurons. For each cell, spike timings are shown as rasters (black) and the raw simultaneously-acquired $\Delta F/F$ signals aligned below (green). Two calcium signal reconstructions were generated, using either a GCaMP6f kernel (gold) or a GCaMP5G kernel (blue). Numbers alongside each reconstruction indicate the Pearson correlation coefficient between the ground truth and the corresponding calcium signal reconstruction. E, quantification of the reconstruction quality, as measured using the Pearson correlation coefficient between the ground truth and reconstructed signals ('Reconstruction'). For each cell ($n = 78$), the correlation between ground truth and a scrambled version of the reconstructed signal ('Scrambled'), as well as the correlation between ground truth and a reconstruction generated after shuffling spike timings ('Events Shuffled'), were measured. Individual data points are plotted alongside each boxplot. P values were calculated using a Kruskal–Wallis test and yielded a P value of 4.6130×10^{-49} , and then using the *post hoc* Dunn's test for individual comparisons, the latter of which are indicated.

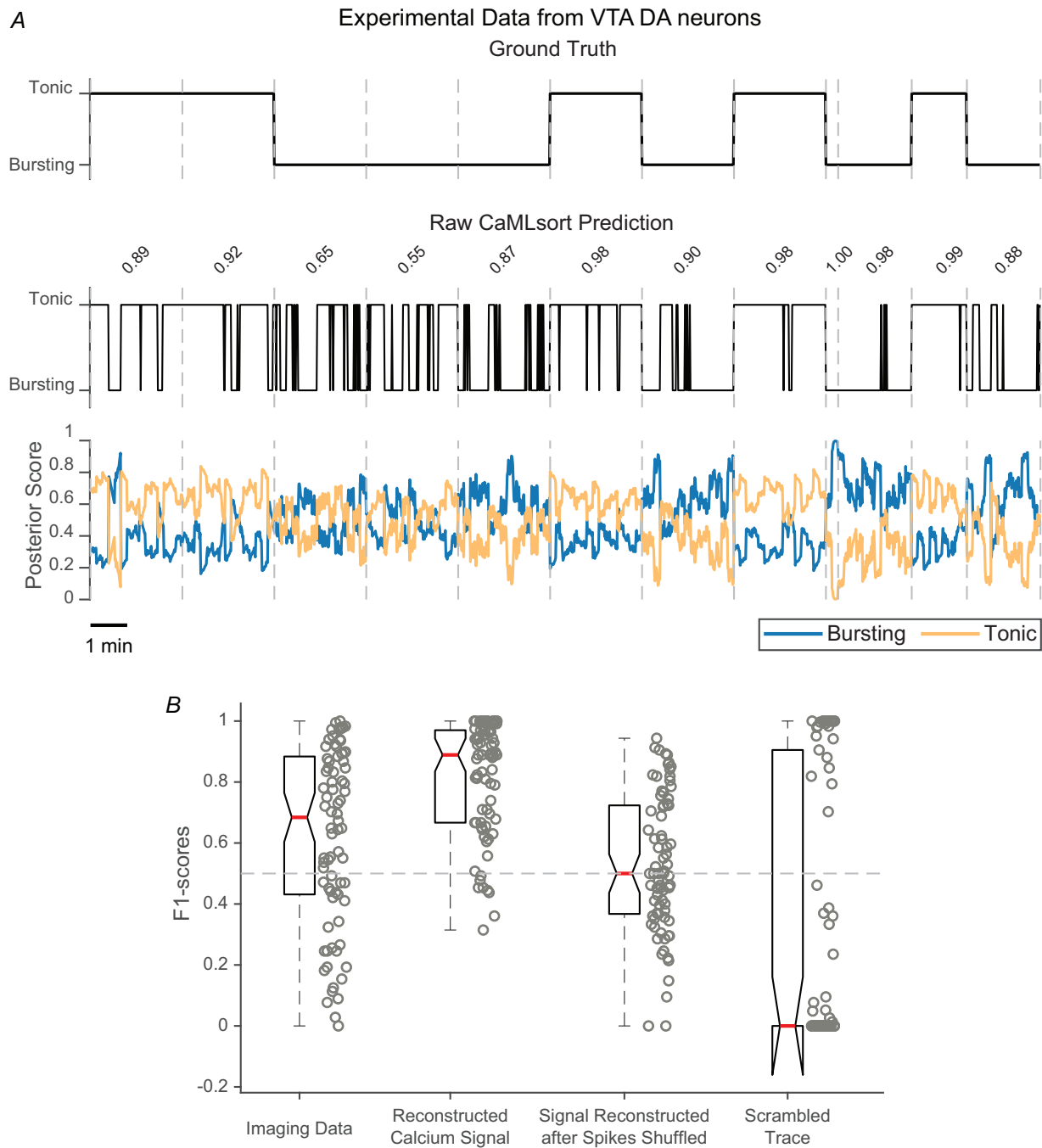


Figure 10. CaMLSort predictions of cellular state from calcium signals of VTA DA neurons
 A, a subset of 12 (of 78) raw CaMLSort predictions from imaging data acquired from DA neurons of the VTA in mice. Each prediction was independently obtained but has been stitched together to produce a single time series with cells separated by vertical dashed lines. The ground truth state (top) is used as reference to assess the quality of CaMLSort's predictions (middle) using the F1 score, which is printed above the prediction for each cell. The posterior scores for each class, bursting (blue) and tonic (light orange), are shown in the lowest plot. B, the entire distribution of F1 scores ($n = 78$) between the ground truth state and the binary classification from CaMLSort obtained from various inputs: (from left to right) experimentally-acquired $\Delta F/F$ traces, calcium signals reconstructed from true spike trains using the GCaMP5G kernel, calcium signals reconstructed from shuffled spike trains using the GCaMP5G kernel or a scrambled version of the raw $\Delta F/F$ data. Individual datapoints are shown alongside each boxplot. The horizontal dashed line corresponds to an F1 score of 0.5. The notches indicate 95% confidence intervals around the median (red line).

alter cell-intrinsic properties as well as network activity, and could thereby affect the bistability of a neuron, as has been demonstrated in multiple experimental systems (Abbinanti et al., 2012; Conway et al., 1988; Hounsgaard et al., 1988; Lechner et al., 1996; Williams et al., 2002). Another question that we can now directly address is how the state of a cell changes over development. We hypothesize that the tonic state could be a developmentally early state, with the burst-like phenotypes developing as they get integrated into the developing cerebellar circuit. By birth-dating PNs using a photoconvertible protein such as Kaede and imaging their activity with GCaMP, we can use CaMLsort to determine whether younger cells tend to be more tonic or vice versa. Although this question could potentially be addressed using electrophysiology, CaMLsort allows us to perform this experiment in a more high-throughput and minimally invasive fashion.

A number of cell types exhibit bistability including cortical neurons (Cowan & Wilson, 1994; Egorov et al., 2002), striatal neurons (Misgeld et al., 1986), thalamic neurons (Fuentealba et al., 2005) and spinal motoneurons (Hounsgaard et al., 1984). In all of these cases, cellular state dictates how synaptic inputs are integrated. Therefore, predicting the cellular state is essential to understanding circuit computation and it is here that CaMLsort will be immensely useful. It is natural to question whether CaMLsort will be equally effective in predicting cellular state in different cell types and in different model organisms. We have found that CaMLsort generalizes well to a completely different system than it was trained on: neurons from a different brain region (ventral tegmental area) in a different model organism (mouse) imaged using a later generation of GCaMP. We would caution against using CaMLsort without validating its efficacy in the new system of interest using a small ground truth dataset.

Taken together, CaMLsort is a powerful tool that opens up new avenues of research into the functioning of any circuit with bistable neurons.

Limitations of CaMLsort

Having stated all the above, CaMLsort is not free of limitations. For example, the generative model that we use to train the neural network is a linear, time-invariant model and the results may improve with the use of more accurate models, such as those that estimate the true increase in intracellular calcium concentration resulting from simple spikes and CF inputs. Furthermore, it could be that CaMLsort may not work independently of the imaging method used. In both cases for which we tested CaMLsort, widefield epifluorescence microscopy was used. The signal and noise characteristics of other, more modern and popular, imaging methods such as multiphoton and light sheet microscopy are different from

traditional widefield imaging. However, we posit that this should also be easy to resolve, either by the acquisition of a small ground truth dataset, or by comparing the traces obtained from imaging with those in the reconstructions dataset, or by denoising the data prior to prediction using CaMLsort.

References

- Abbinanti, M. D., Zhong, G., & Harris-Warrick, R. M. (2012). Postnatal emergence of serotonin-induced plateau potentials in commissural interneurons of the mouse spinal cord. *Journal of Neurophysiology*, **108**(8), 2191–2202.
- Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M., & Keller, P. J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature Methods*, **10**(5), 413–420.
- Akerboom, J., Chen, T.-W., Wardill, T. J., Tian, L., Marvin, J. S., Mutlu, S., Calderón, N. C., Esposti, F., Borghuis, B. G., Sun, X. R., Gordus, A., Orger, M. B., Portugues, R., Engert, F., Macklin, J. J., Filosa, A., Aggarwal, A., Kerr, R. A., Takagi, R., ... Looger, L. L. (2012). Optimization of a GCaMP calcium indicator for neural activity imaging. *The Journal of Neuroscience*, **32**(40), 13819–13840.
- Ali, F., & Kwan, A. C. (2020). Interpreting in vivo calcium signals from neuronal cell bodies, axons, and dendrites: A review. *Neurophotonics*, **7**(1), 011402.
- Armstrong, D. M., & Edgley, S. A. (1984). Discharges of Purkinje cells in the paravermal part of the cerebellar anterior lobe during locomotion in the cat. *The Journal of Physiology (London)*, **352**, 403–424.
- Bae, Y.-K., Kani, S., Shimizu, T., Tanabe, K., Nojima, H., Kimura, Y., Higashijima, S., & Hibi, M. (2009). Anatomy of zebrafish cerebellum and screen for mutations affecting its development. *Developmental Biology*, **330**(2), 406–426.
- Borst, J. G., & Helmchen, F. (1998). Calcium influx during an action potential. *Methods in Enzymology*, **293**, 352–371.
- Chen, T.-W., Wardill, T. J., Sun, Y., Pulver, S. R., Renninger, S. L., Baohan, A., Schreiter, E. R., Kerr, R. A., Orger, M. B., Jayaraman, V., Looger, L. L., Svoboda, K., & Kim, D. S. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*, **499**(7458), 295–300.
- Conway, B. A., Hultborn, H., Kiehn, O., & Mintz, I. (1988). Plateau potentials in alpha- motoneurons induced by intravenous injection of L-dopa and clonidine in the spinal cat. *The Journal of Physiology (London)*, **405**, 369–384.
- Cowan, R. L., & Wilson, C. J. (1994). Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *Journal of Neurophysiology*, **71**(1), 17–32.
- de Zeeuw, C. I., Hoebeek, F. E., Bosman, L. W. J., Schonewille, M., Witter, L., & Koekkoek, S. K. (2011). Spatiotemporal firing patterns in the cerebellum. *Nature Reviews Neuroscience*, **12**(6), 327–344.
- Dickinson, P. S., & Nagy, F. (1983). Control of a central pattern generator by an identified modulatory interneurone in crustacea. II. Induction and modification of plateau properties in pyloric neurones. *Journal of Experimental Biology*, **105**, 59–82.

- Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., & Alonso, A. A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature*, **420**(6912), 173–178.
- Engbers, J. D. T., Fernandez, F. R., & Turner, R. W. (2013). Bistability in Purkinje neurons: Ups and downs in cerebellar research. *Neural Networks: The Official Journal of the International Neural Network Society*, **47**, 18–31.
- Fleming, W., Jewell, S., Engelhard, B., Witten, D. M., & Witten, I. B. (2021). Inferring spikes from calcium imaging in dopamine neurons. *PLoS ONE*, **16**(6), e0252345.
- Fuentealba, P., Timofeev, I., Bazhenov, M., Sejnowski, T. J., & Steriade, M. (2005). Membrane bistability in thalamic reticular neurons during spindle oscillations. *Journal of Neurophysiology*, **93**(1), 294–304.
- Göbel, W., & Helmchen, F. (2007). In vivo calcium imaging of neural network function. *Physiology*, **22**, 358–365.
- Grace, A., & Bunney, B. (1984a). The control of firing pattern in nigral dopamine neurons: burst firing. *The Journal of Neuroscience*, **4**(11), 2877–2890.
- Grace, A., & Bunney, B. (1984b). The control of firing pattern in nigral dopamine neurons: single spike firing. *The Journal of Neuroscience*, **4**(11), 2866–2876.
- Grewe, B. F., & Helmchen, F. (2009). Optical probing of neuronal ensemble activity. *Current Opinion in Neurobiology*, **19**(5), 520–529.
- Grienberger, C., & Konnerth, A. (2012). Imaging calcium in neurons. *Neuron*, **73**(5), 862–885.
- Harmon, T. C., McLean, D. L., & Raman, I. M. (2020). Integration of swimming-related synaptic excitation and inhibition by olig²⁺ eurydendroid neurons in larval zebrafish cerebellum. *Journal of Neuroscience*, **40**(15), 3063–3074.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). <https://doi.org/10.1109/CVPR.2016.90>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, **9**(8), 1735–1780.
- Hounsgaard, J., Hultborn, H., Jespersen, B., & Kiehn, O. (1984). Intrinsic membrane properties causing a bistable behaviour of alpha-motoneurons. *Experimental Brain Research*, **55**(2), 391–394.
- Hounsgaard, J., Hultborn, H., Jespersen, B., & Kiehn, O. (1988). Bistability of alpha-motoneurons in the decerebrate cat and in the acute spinal cat after intravenous 5-hydroxytryptophan. *The Journal of Physiology*, **405**, 345–367.
- Ikenaga, T., Yoshida, M., & Uematsu, K. (2006). Cerebellar efferent neurons in teleost fish. *Cerebellum (London, England)*, **5**(4), 268–274.
- Jahnsen, H., & Llinás, R. (1984). Electrophysiological properties of guinea-pig thalamic neurones: An in vitro study. *The Journal of Physiology*, **349**, 205–226.
- Kerr, J. N. D., Greenberg, D., & Helmchen, F. (2005). Imaging input and output of neocortical networks in vivo. *Proceedings of the National Academy of Sciences*, **102**(39), 14063–14068.
- Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. <https://doi.org/10.48550/arXiv.1412.6980>
- Kitamura, K., & Häusser, M. (2011). Dendritic calcium signaling triggered by spontaneous and sensory-evoked climbing fiber input to cerebellar Purkinje cells in vivo. *Journal of Neuroscience*, **31**(30), 10847–10858.
- Knogler, L. D., Kist, A. M., & Portugues, R. (2019). Motor context dominates output from purkinje cell functional regions during reflexive visuomotor behaviours. *Elife*, **8**, e42138.
- Kwan, A. C., & Dan, Y. (2012). Dissection of cortical microcircuits by single-neuron stimulation in vivo. *Current Biology*, **22**(16), 1459–1467.
- Lechner, H. A., Baxter, D. A., Clark, J. W., & Byrne, J. H. (1996). Bistability and its regulation by serotonin in the endogenously bursting neuron R15 in *Aplysia*. *Journal of Neurophysiology*, **75**(2), 957–962.
- Lev-Ram, V., Miyakawa, H., Lasser-Ross, N., & Ross, W. N. (1992). Calcium transients in cerebellar Purkinje neurons evoked by intracellular stimulation. *Journal of Neurophysiology*, **68**(4), 1167–1177.
- Lin, B.-J., Chen, T.-W., & Schild, D. (2007). Cell type-specific relationships between spiking and [Ca²⁺]_i in neurons of the *Xenopus* tadpole olfactory bulb. *The Journal of Physiology*, **582**(Pt 1), 163–175.
- Loewenstein, Y., Mahon, S., Chadderton, P., Kitamura, K., Sompolinsky, H., Yarom, Y., & Häusser, M. (2005). Bistability of cerebellar Purkinje cells modulated by sensory stimulation. *Nature Neuroscience*, **8**(2), 202–211.
- Marder, E., Abbott, L. F., Turrigiano, G. G., Liu, Z., & Golowasch, J. (1996). Memory from the dynamics of intrinsic membrane currents. *Proceedings of the National Academy of Sciences*, **93**(24), 13481–13486.
- Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, **5**(10), 793–807.
- Matsui, H., Namikawa, K., Babaryka, A., & Köster, R. W. (2014). Functional regionalization of the teleost cerebellum analyzed in vivo. *Proceedings National Academy of Science USA*, **111**(32), 11846–11851.
- McCormick, D. A., Connors, B. W., Lighthall, J. W., & Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of Neurophysiology*, **54**(4), 782–806.
- Misgeld, U., Calabresi, P., & Dodt, H. U. (1986). Muscarinic modulation of calcium dependent plateau potentials in rat neostriatal neurons. *Pflugers Archiv: European Journal of Physiology*, **407**(5), 482–487.
- Miyakawa, H., Lev-Ram, V., Lasser-Ross, N., & Ross, W. N. (1992). Calcium transients evoked by climbing fiber and parallel fiber synaptic inputs in guinea pig cerebellar Purkinje neurons. *Journal of Neurophysiology*, **68**(4), 1178–1189.
- Moreaux, L., & Laurent, G. (2007). Estimating firing rates from calcium signals in locust projection neurons in vivo. *Frontiers in Neural Circuits*, **1**, 2.

- Narayanan, S., & Thirumalai, V. (2019). Contributions of the cerebellum for predictive and instructional control of movement. *Current Opinion in Physiology*, **8**, 146–151.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., & Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*. Curran Associates, Inc.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, **12**(85), 2825–2830.
- Peron, S., Chen, T.-W., & Svoboda, K. (2015). Comprehensive imaging of cortical networks. *Current Opinion in Neurobiology*, **32**, 115–123.
- Person, A. L., & Raman, I. M. (2012). Purkinje neuron synchrony elicits time-locked spiking in the cerebellar nuclei. *Nature*, **481**(7382), 502–505.
- Popa, L. S., Streng, M. L., & Ebner, T. J. (2018). Purkinje cell representations of behavior: Diary of a busy neuron. *The Neuroscientist*, **25**(3), 241–257.
- Ramirez, J. E., & Stell, B. M. (2016). Calcium imaging reveals coordinated simple spike pauses in populations of cerebellar purkinje cells. *Cell Reports*, **17**(12), 3125–3132.
- Renninger, S. L., & Orger, M. B. (2013). Two-photon imaging of neural population activity in zebrafish. *Methods (San Diego, California)*, **62**(3), 255–267.
- Rupprecht, P., Carta, S., Hoffmann, A., Echizen, M., Blot, A., Kwan, A. C., Dan, Y., Hofer, S. B., Kitamura, K., Helmchen, F., & Friedrich, R. W. (2021). A database and deep learning toolbox for noise-optimized, generalized spike inference from calcium imaging. *Nature Neuroscience*, **24**(9), 1324–1337.
- Russell, D. F., & Hartline, D. K. (1982). Slow active potentials and bursting motor patterns in pyloric network of the lobster, *Panulirus interruptus*. *Journal of Neurophysiology*, **48**(4), 914–937.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, **9**(7), 676–682.
- Schwindt, P., & Crill, W. (1980). Role of a persistent inward current in motoneuron bursting during spinal seizures. *Journal of Neurophysiology*, **43**(5), 1296–1318.
- Sengupta, M. (2015). Purkinje neurons in action: From single cells to ensembles. Mumbai: Tata Institute of Fundamental Research.
- Sengupta, M., & Thirumalai, V. (2015). AMPA receptor mediated synaptic excitation drives state-dependent bursting in Purkinje neurons of zebrafish larvae. *Elife*, **4**, e09158.
- Smetters, D., Majewska, A., & Yuste, R. (1999). Detecting action potentials in neuronal populations with calcium imaging. *Methods (San Diego, California)*, **18**(2), 215–221.
- Thirumalai, V., & Jha, U. (2022). Recruitment of motoneurons vertebrate motoneurons. In Press. Springer-Verlag.
- Tian, L., Hires, S. A., Mao, T., Huber, D., Chiappe, M. E., Chalasani, S. H., Petreanu, L., Akerboom, J., McKinney, S. A., Schreiter, E. R., Bargmann, C. I., Jayaraman, V., Svoboda, K., & Looger, L. L. (2009). Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods*, **6**(12), 875–881.
- Urasaki, A., Morvan, G., & Kawakami, K. (2006). Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics*, **174**(2), 639–649.
- Vogelstein, J. T., Packer, A. M., Machado, T. A., Sippy, T., Babadi, B., Yuste, R., & Paninski, L. (2010). Fast non-negative deconvolution for spike train inference from population calcium imaging. *Journal of Neurophysiology*, **104**(6), 3691–3704.
- Williams, S. R., Christensen, S. R., Stuart, G. J., & Häusser, M. (2002). Membrane potential bistability is controlled by the hyperpolarization-activated current I(H) in rat cerebellar Purkinje neurons in vitro. *The Journal of Physiology (London)*, **539**(Pt 2), 469–483.
- Yaksi, E., & Friedrich, R. W. (2006). Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca²⁺ imaging. *Nature Methods*, **3**(5), 377–383.
- Yang, W., & Yuste, R. (2017). In vivo imaging of neural activity. *Nature Methods*, **14**(4), 349–359.
- Zeldenrust, F., Wadman, W. J., & Englitz, B. (2018). Neural coding with bursts—current state and future perspectives. *Frontiers in Computational Neuroscience*, **12**, 48.

Additional information

Data availability statement

All data are available within the published article. The CaMLsort package codes are available via Github: <https://github.com/vatsalat/CaMLsort23>.

Competing interests

The authors declare that they have no competing interests.

Author contributions

A.V. and V.T. were responsible for conceptualization. A.V., S.U., M.S., P.K.G. and V.T. were responsible for investigations. A.V., S.U., M.S., P.K.G. and V.T. were responsible for analysis. A.V. and V.T. were responsible for writing the manuscript.

Funding

The following sources provided funding: Wellcome Trust-DBT India Alliance Intermediate (VT; 500 040/Z/09/Z) and Senior fellowships (VT; IA/S/17/2/503 297), Department of Biotechnology (VT; BT/PR4983/MED/30/790/2012), Science and Engineering Research Board, Department of Science and Technology (VT; EMR/2015/000595), Department of Atomic Energy (VT), and Department of Science and Technology, Govt. of India (PKG).

Acknowledgements

We thank P. T. Jagadeesh for maintaining the fish facility and members of the Fishfolk lab at NCBS for their inputs at various stages of this project. In particular, we thank Ms Meha Jadhav and Ms Shivangi Verma for their help with independently verifying parts of the manuscript. We also thank Professor Rohini Balakrishnan for initiating this collaboration. This manuscript was first published as a preprint: Varma A, Udupa S, Sengupta M, Ghosh P, Thirumalai V (2022) A machine-learning tool to identify bistable state from calcium imaging data. *bioRxiv* (doi: <https://doi.org/10.1101/2022.11.10.515941>).

Author's present address

Mohini Sengupta: Department of Biology, St Louis University, St Louis, MO, USA

Keywords

bistability, cerebellum, convolutional neural network, machine learning, Purkinje neuron, recurrent neural networks, zebrafish

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the HTML view of the article. Supporting information files available:

Statistical Summary Document

Peer Review History