Extraction of bouton-like structures from neuropil calcium imaging data

Kazushi Fukumasu[a], Akinao Nose[a,b] and Hiroshi Kohsaka[b*]

[a] Department of Physics, Graduate School of Science, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 133-0033 Tokyo, Japan

[b] Department of Complexity Science and Engineering, Graduate School of Frontier Science, the University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, 277-8561 Chiba, Japan

* For correspondence: (kohsaka@edu.k.u-tokyo.ac.jp, +81-4-7136-3922)

Abstract

The neuropil, the plexus of axons and dendrites, plays a critical role in operating the circuit processing of the nervous system. Revealing the spatiotemporal activity pattern within the neuropil would clarify how the information flows throughout the nervous system. However, calcium imaging to examine the circuit dynamics has mainly focused on the population of somas due to their discrete distribution. Development of a methodology to analyze the calcium imaging data of the densely packed neuropil would provide us with new insights into the circuit dynamics.

Here, we propose a new method to decompose calcium imaging data of the neuropil into populations of bouton-like synaptic structures. To this aim, we pan-neuronally expressed a membrane-bound form of the calcium probe protein, CD4::GCaMP6f in Drosophila larvae. With the probe, we succeeded in visualizing a calcium signal in the neuropil. We obtained the baseline of the signal in the spatial- and temporal-dimension by an iteration method, which enabled the alignment of the neuropil at a single bouton scale and attenuation of the sample deformation and stage drift during the imaging. To group subsets of pixels into bouton-like structures, we introduced a proximity matrix and optimized the clustering configuration by the Markov Chain Monte Carlo method with Simulated Annealing algorithm, which is established in statistical physics. We applied this procedure to the neuropil in the isolated central nervous system of Drosophila larvae and succeeded in extracting individual bouton-like structures and relating them to population neural activity for larval fictive motion. These results demonstrate the usefulness and significance of the neuropil calcium imaging and its decomposition to study the spatiotemporal patterns during the operation of neural processing.

Introduction

Calcium imaging of the cell bodies of neurons has provided rich information on the dynamic activity of neural circuits. Functional imaging of neuron populations has revealed the existence of maps in several regions of the central nervous system, including the visual cortex, auditory cortex, hippocampus, amygdala, and spinal cord (Gauthier and Tank, 2018; Gründemann et al., 2019; Hackett et al., 2011; Tinnermann et al., 2021; Zhuang et al., 2017). The configuration of cell bodies is one of the critical factors for neural circuit computation.

While the calcium imaging of somas has been widely used, the cell body is not the sole component of a neuron. For synapse transduction, axons and dendrites are major structures and they form the neuropil in the brain. Considering the physiological roles in the neuropil, calcium imaging of the neuropil would bring us several new types of information on the circuit dynamics: first, the calcium signal in the neuropil directly reflects the synaptic transmission, which is critical to understand spatiotemporal dynamics in the nervous system. Second, the activation of one neuron could generate heterogenous activity among its multiple presynaptic sites (Bilz et al., 2020). This heterogeneity can be detected by neuropil imaging but not by somatic imaging. Third, since the neural wiring diagram is embedded in the neuropil, the activity dynamics in the neuropil can represent the flow of information in the neural circuit. Whereas the neuropil imaging would be informative to study circuit dynamism, the data analysis of calcium imaging in the neuropil has been challenging. While the large proportion of cell bodies in volume cause intense fluorescence signals from them, calcium probe signals from synapses and neurites with less volume is weak. Furthermore, because synapses and neurites are densely packed in the nervous system, it is hard to identify synapses and extract their activity from pan-neuronal imaging data. Although researchers have proposed several algorithms to automatically decompose Ca2+ imaging data into cell body activities (Pnevmatikakis et al., 2016; Shibue and Komaki, 2020), automatic approaches to extract synapses remain to be developed.

In the decomposition of somatic activity, there are several requirements to be met. Activities of neuronal populations have to be sporadic so that signal traces of proximate cell bodies can be clearly differentiated, or the frame rate should be sufficient to recognize each spike in Ca2+ signal traces. However, neural activity is not always sporadic patterns but in some cases spatiotemporally coordinated ones such as global bursts or wave-like

propagations (Churchland et al., 2012; Lemon et al., 2015). In addition, the frame rate is in a trade-off relationship with the signal-to-noise (S/N) ratio or the number of focus planes scanned along the depth. Therefore, the existing algorithms might not be suitable for processing such imaging data of modest quality.

In this study, we developed a new methodology for the analysis of neuropil Ca2+ imaging and an unsupervised algorithm to extract sites and activity traces of synapses. We applied an iteration method to adaptively infer the baseline of the signals in both temporal and spatial dimensions, which enabled the alignment of the neuropil at a single bouton scale and the attenuation of the sample deformation and stage drift during the imaging. To label each bouton-like component in the aligned data, we introduced a proximity matrix and optimized the clustering configuration by the Markov Chain Monte Carlo method with Simulated Annealing algorithm, which is established in statistical physics. We applied this procedure to the neuropil of Drosophila larvae and succeeded in extracting population activity in the neuropil for motor control.

Materials and Methods

The proposed method for bouton extraction from neuropil calcium imaging data consists of three steps: baseline correction, fine and coarse structure decomposition, and pixel clustering. Figure 1 shows the method flowchart.

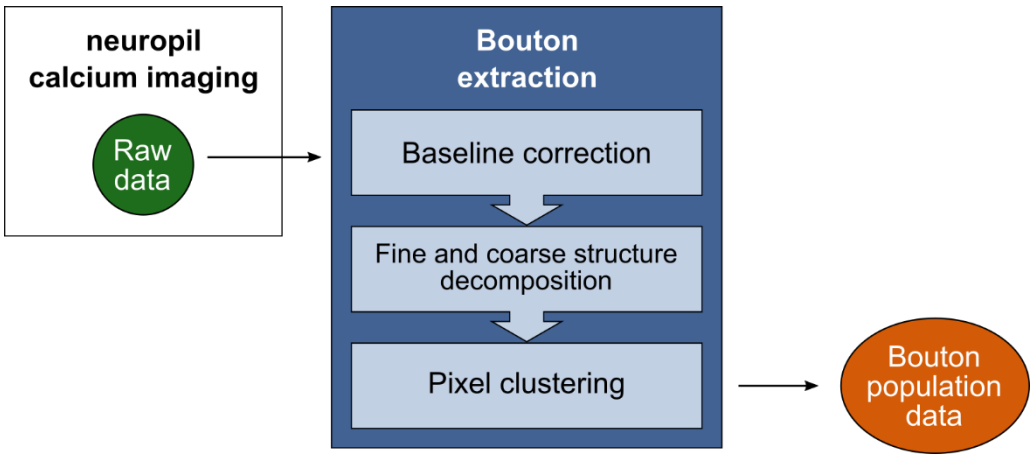


Figure 1. Flowchart of the bouton extraction method.

1. Three-dimensional calcium imaging of the neuropil

Calcium imaging data of the neuropil was obtained from the central nervous system of Drosophila melanogaster third instar larvae. The genotype of the larvae we used was UAS-CD4::GCaMP6f; nSyb-Gal4, which expresses a membrane-bound form calcium sensor (CD4::GCaMP6f) in all neurons (nSyb-Gal4: Bloomington #58763, UAS-CD4::GCaMP6f: (Kohsaka et al., 2019)). Flies were reared at 25℃ on standard cornmeal-based food.

To record neural activity in the neuropil, we used a membrane-localized calcium probe, CD4::GCaMP6f (Figure 2B). Calcium probes without any localization tags diffuse in the cytosol, which causes the accumulation of the probes in soma rather than neurites (Figure 2C). To facilitate the localization of probes to the neuropil, a membrane-bound region of CD4 protein (Han et al., 2011) was connected to the N terminal of GCaMP6f. To check whether the probe was localized to the neuropil, we performed immunostaining with an anti-GFP rabbit antibody (Af2020, Frontier science, 1:1000) and a secondary antibody (AlexaFluor 488-conjugated anti-rabbit antibody, A-11034, Invitrogen, 1:300) (Figure 2D). The staining result indicates the probe was successfully delivered to the neuropil region. This protein construct offers two advanced properties of the probe: first, the distribution is highly biased on the neurites, which are the main component of the neuropil; second, the signal efficiency is higher than the original probes diffused in the cytosol, because the Ca2+ concentration change is generally more intense in the sub-membrane region than in the cytoplasm, as Ca2+ influx flows through the cell membrane. Thanks to these properties, CD4::GCaMP6f gives high-contrast signals from neurites, especially from boutons, which are the swollen structures at pre-synaptic sites (Figure 2A).

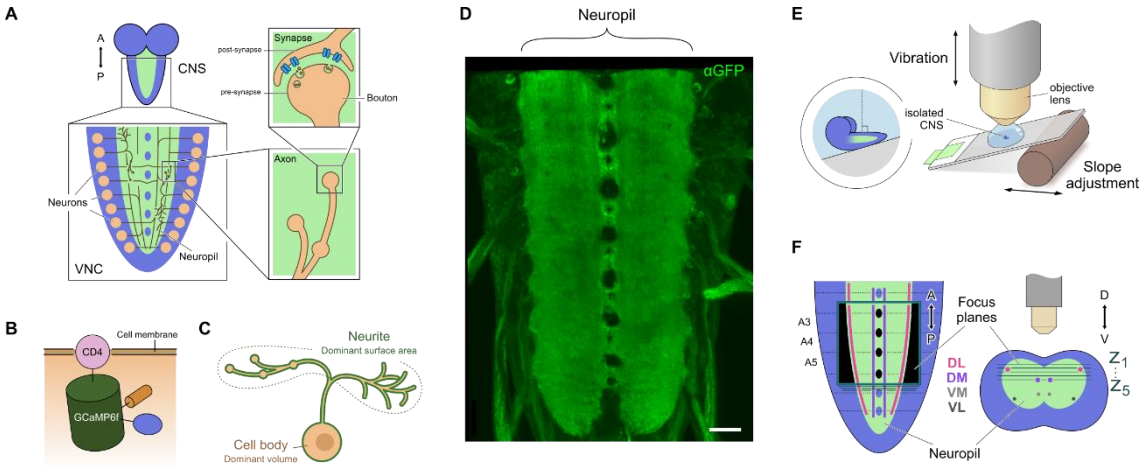


Figure 2. 3-dimensional pan-synapse Ca2+ imaging with membrane-anchored calcium probe. (A) Spatial relationship of the CNS, the VNC, the neuropil and boutons. (B) Schematic diagram showing the structure of CD4::GCaMP6f. (C) Cell bodies occupy the largest portion of the cellular volume, while neurites occupy the largest portion of the surface area of neurons. (D) Immunostaining image targeting CD4::GCaMP6f (maximum intensity projection). The scale bar: 20 μm. (E) Three-dimensional Ca2+ imaging of the isolated CNS. (F) Imaging was performed around A3-A5 segments in the dorsal-half neuropil. Four of the anti-Fasciclin 2 antibody positive tracts (Landgraf et al., 2003) are indicated: dorsolateral (DL), dorsomedial (DM), ventromedial (VM) and ventrolateral (VL).

Calcium imaging of the CNS of fly larvae was carried out as described previously (Kohsaka et al., 2014, 2019) (Figure 2E). In brief, third-instar larvae of the *UAS-CD4::GCaMP6f; nSyb-Gal4* fly were dissected with microscissors and their CNS was isolated. The isolated CNS was mounted on an adhesive slide glass (MAS coated slide glass S9215, Matsunami Glass, Japan) and soaked in an insect saline, TES buffer (TES 5 mM, NaCl 135 mM, KCl 5 mM, CaCl2 2 mM, MgCl2 4 mM, sucrose 36 mM; pH = 7.15). The three-dimensional fluorescence signal from the CNS was recorded by an EMCCD camera (iXon+ DU-897E-CS0-#BV, Andor, UK) (63x, 0.27 μm/px, 30 ms/frame, EM gain: 144, data depth: 14 bit) with attached spinning disk confocal unit (CSU21, YOKOGAWA, Japan) with a piezo unit (P-725.2CL, Physik Instrumente, Germany). The CNS preparation was illuminated by a blue laser (power: 300 μW ~ 500 μW; the wavelength: 488 nm; CSU-LS2WF, Solution Systems, Japan) through a water immersion objective lens (ACROPLAN 63X, Zeiss,

Germany). To align the larval VNC on the focal plane, the slide glass was put on a slope which is adjustable by moving a fulcrum. Each volume contains five xy planes with an interval of 5 μm (Figure 2F). The recording length was 233 seconds at three frames per second.

2. Bouton extraction

Tens of thousands of boutons are embedded in the neuropil. By taking advantage of their temporal profiles in the calcium imaging data, each of the single bouton-like structures was isolated in the neuropil. In this section, imaging data of each focal plane is described by the following:

$$l_{txy} \in \{0, \dots, l_{max}\}$$

where $t \in \{0, \dots, T-1\}$, $x, y \in \{0, \dots, L-1\}$ $(l_{max} = 2^{14} - 1,\ T = 700\ \text{frames},\ L = 512\ \text{px})$. The data l_{txy} was decomposed into a population of bouton-like structures by the following three steps. We used python3.7 combined with c++, and python libraries, numpy, SciPy and OpenCV for the following calculation.

2.1 Baseline correction

The signal intensity at each pixel is changed by two main factors: fluorescence change by probe function, and experimental artifacts. The former is what we need to extract in the calcium signal analysis, and the latter should be removed. The causes of the artifacts include sample deformation, stage drift, and photo bleaching of the calcium probe. Here we formalized the artifacts as the change of the baseline and assumed that the signal intensity at each pixel of the fluorescence movie is the product of the baseline and the ratiometric change in probe signal intensity as follows:

$$l_{txy}^{(1)} = b_{txy} \times \Delta s_{txy} \tag{2.1}$$

where $l_{txy}^{(1)}$: fluorescence data, b_{txy}: baseline, Δs_{txy}: signal ratio above the baseline

To suppress the effect of temporal change of the baseline, baseline detection processing was applied using one-dimensional Gaussian blur (1d-GB) (Figure 3A). 1d-GB is defined as the following:

$$G_{\sigma_t}^{(1)}(t) \coloneqq e^{-t^2/2\sigma_t^2}$$

$$f_{t(G1:\sigma_t)} \coloneqq \sum_{t'=-\infty}^{\infty} f_{t-t'} G_{\sigma_t}^{(1)}(t') \Big/ \sum_{t'=-\infty}^{\infty} G_{\sigma_t}^{(1)}(t') \tag{2.2}$$

where f: 1-dimensional discrete function

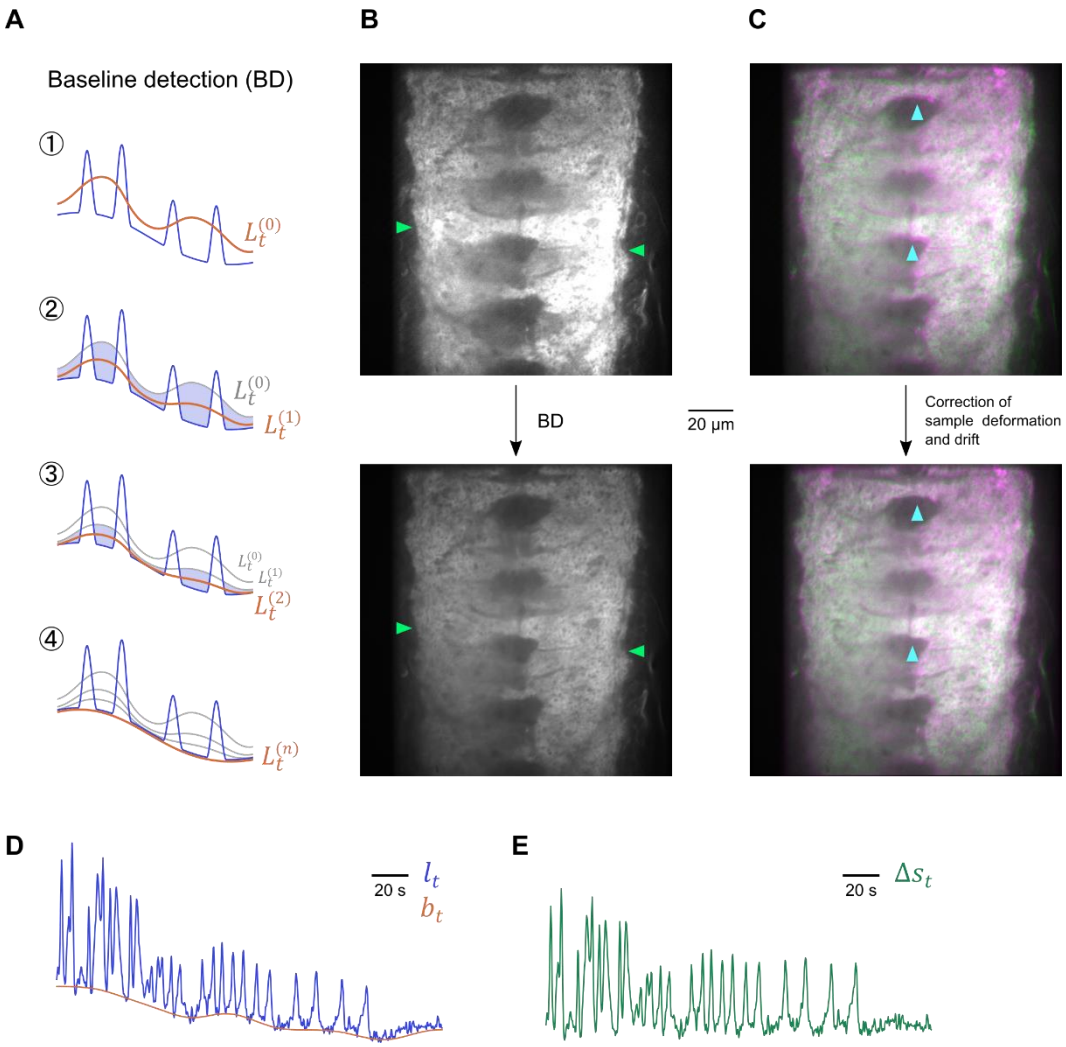


Figure 3: Baseline detection and correction. (A) Explanation of how baseline detection works. Baseline is extracted by repeated calculation of a recurrence formula. See the main text for details. (B) An imaging frame showing neural activity (top) and the baseline frame of the same frame, detected by baseline detection at each pixel. Green arrows indicate regions showing strong neural activity before BD. (C) The baseline frames of the first (magenta) and the last (green) frame numbers. Before (top) and after (bottom) correction of sample deformation and drift. Cyan arrows indicate neuropil edges whose misalignment was clearly removed after the processing. (D) An activity trace of one pixel and detected baseline. (E) Baseline-corrected trace of the same pixel in (D), based on the detected baseline.

Here, the reflection boundary condition was adopted:

When f_t is defined in $t = 0, 1, \ldots, T-1$,

$$f_t = \begin{cases} f_{-t-1} & (t < 0) \\ f_{2T-t-1} & (T \leq t) \end{cases}$$

After 1d-GB was applied to the luminosity $l_{txy}^{(1)}$ ($\sigma_t = 2.0$), five frames per six were thinned out instead of using all data (only $t = 0, 6, 12, \ldots$ out of $t = 0, 1, \ldots, T-1$ was used) to save calculation cost in the following processes:

$$l'_{\tilde{t}xy} = l_{t(G1:\sigma_t)xy}^{(1)}, \qquad t = 0, 6, 12, \ldots \tilde{T} \tag{2.3}$$

$\tilde{T}$ is the maximum multiple of six less than T. This down sampling is reasonable because the degrees of freedom are sufficient to calculate the gradually changing baseline (data not shown).

To approximate baseline curves as asymptotic lines, $L_{\tilde{t}xy}^{(n)}$ was calculated by the recurrence formula $(2.4)(2.5)(2.5)$. The deviation $\sigma_{\tilde{t}}$ was chosen to be sufficiently larger than the signal change timescale, and comparable with baseline change timescale ($\sigma_{\tilde{t}} = 8.3$).

$$L_{\tilde{t}xy}^{(0)} = l'_{\tilde{t}(G1:\sigma_{\tilde{t}})xy} \tag{2.4}$$

$$\hat{L}_{\tilde{t}xy}^{(n)} := \max\left\{L_{\tilde{t}xy}^{(n)} - l'_{\tilde{t}xy}, 0\right\}, \qquad L_{\tilde{t}xy}^{(n+1)} = L_{\tilde{t}xy}^{(n)} - \hat{L}_{\tilde{t}(G1:\sigma_{\tilde{t}})xy}^{(n)} \ (n = 0, 1, 2, \ldots) \tag{2.5}$$

The obtained $L_{\tilde{t}xy}^{(N)}$ ($N = 100$, which was large enough for convergence of the baseline curves) was interpolated to recover the full-length data. The interpolation method is explained below.

The interpolation was conducted by using Fourier transform: Suppose F_t is a discrete function which is to be interpolated. The interpolation was performed by applying Fourier transform (eq.(2.7)), and then inverse Fourier transform (eq.(2.8)) to F_t. Before applying Fourier transform, the inverse trace $F_{T-1}, \ldots, F_1, F_0$ is appended behind the original trace $F_0, F_1, \ldots, F_{T-2}, F_{T-1}$ to avoid artifacts (eq.(2.6)). If the first and the last numbers of a trace are largely different, Fourier transform produces too large components of high wave number.

$$F'_t = \begin{cases} F_t & (t \leq T-1) \\ F_{2T-t-1} & (t > T-1) \end{cases} \qquad (t = 0, 1, \ldots, 2T-1) \tag{2.6}$$

Then, Fourier and inverse Fourier transform is applied to F'_t.

$$\tilde{F}(k) = \sum_{t=0}^{2T-1} F'_t e^{i\frac{\pi k t}{T}} \tag{2.7}$$

$$F(t) = \frac{1}{2T} \sum_{k=0}^{2T-1} \tilde{F}(k) e^{-i\frac{\pi k t}{T}} \tag{2.8}$$

A continuous function $F(t)$ was produced in eq.(2.8), which can be assumed as a smooth interpolation of F_t. By this calculation, the function F'_t, which is defined at discrete time points $(t = 1,2, \dots, T-1)$, is converted to a continuous function $F(t)$ of which domain is real number t, where $F(t) = F_t$ $(t = 1,2, \dots, T-1)$. By the procedure described above, we interpolated $L^{(N)}_{\tilde{t}xy}$ into $L^{(N)}_{xy}(\tilde{t})$, then $b_{txy} \coloneqq L^{(N)}_{xy}(t/6)$.

The extracted b_{txy} is the temporally shifting baseline for each pixel. In other words, b_{txy} is a fluorescence distribution without calcium transient (Figure 3B). Therefore, b_{txy} was first used to correct the sample deformation and drift. Each frame $\{b_{txy}|x, y\}$ was mapped by homography to a new coordinate $\{b_{tx',y'}|x', y'\}$ so that the result was overlapped with an image at reference frame t_{ref} $(\{b_{t_{\mathrm{ref}}xy}|x, y\})$ as closely as possible. The reference time t_{ref} was manually selected from around the middle of the data. Thanks to the correction of deformation and drift, the sample stability within the imaging frame was improved (Figure 3C). Then, the baseline detection was applied again to fluorescence with mapped coordination, $l^{(1)}_{tx'y'}$ and new $b_{tx'y'}$ were calculated (Figure 3D). Finally, the temporal profiles of calcium signal without baseline change were extracted (Figure 3E). The pure neuronal signal $\Delta s_{tx'y'}$ can be calculated by rearranging the eq.(2.1):

$$\Delta s_{tx'y'} = \frac{l^{(1)}_{tx'y'}}{b_{tx'y'}} \tag{2.9}$$

The primes of the spatial variables are omitted from the next section: Δs_{txy}

2.2 Fine and coarse structure decomposition

The data Δs_{txy} obtained above contained blurred signals from activities out of focus, because the focal planes were placed in a three-dimensional object filled with calcium probes (Figure 4A). These blurred and coarse signals made it difficult to extract individual bouton activities. Δs_{txy} is formulated as a summation of fine (well-focused) and coarse (out-of-focus) components:

$$\Delta s_{txy} = f_{txy} + c_{txy} \tag{2.10}$$

where f: fine component, c: coarse component

The data Δs_{txy} was decomposed into f_{txy} and c_{txy} based on difference in spatial scale of intensity change. In each frame of the data, the coarse component looked like a gently changing baseline, and narrow peaks of the well-focused component were mounted on that baseline (Figure 4B). Therefore, like the decomposition of $s(t)$ and $b(t)$ in section 0, the spatial structure of f_{txy} and c_{txy} can be segregated by utilizing the scale difference of intensity change (Figure 4C). Accordingly, we applied two-dimensional baseline detection to Δs_{txy}, using two-dimensional Gaussian blur (2d-GB):

$$G_{\sigma_s}^{(2)}(x,y) \coloneqq e^{-(x^2+y^2)/2\sigma_s^2}$$

$$f_{xy(G2:\sigma_s)} \coloneqq \sum_{x',y'=-\infty}^{\infty} f_{(x-x')(y-y')} \cdot G_{\sigma_s}^{(2)}(x',y') \Big/ \sum_{x',y'=-\infty}^{\infty} G_{\sigma_s}^{(2)}(x',y') \qquad (2.11)$$

where f: 2-dimensional discrete function

Again, we adopted the reflection boundary condition:

When f_{xy} is defined in $x, y = 0, 1, \dots, L-1$,

$$f_{xy} = \begin{cases} f_{-x-1\,y} & (x < 0) \\ f_{2L-x-1\,y} & (L \le x) \end{cases}, \quad f_{xy} = \begin{cases} f_{x\,-y-1} & (y < 0) \\ f_{x\,2L-y-1} & (L \le y) \end{cases}.$$

By using 2d-GB, we applied fine and coarse decomposition processing for each frame t_0:

$$S_{txy}^{(0)} = \Delta s_{txy(G2:\sigma_s)} \qquad (2.12)$$

$$\hat{S}_{txy}^{(n)} \coloneqq \max\left\{S_{txy}^{(n)} - \Delta s_{txy}, 0\right\}, \qquad S_{txy}^{(n+1)} = S_{txy}^{(n)} - \hat{S}_{txy(G2:\sigma_s)}^{(n)} \quad (n = 0, 1, 2, \dots) \qquad (2.13)$$

Coarse component c_{txy} was assumed as $S_{txy}^{(100)}$, and the fine component f_{txy} as $\Delta s_{txy} - c_{txy}$ (cf. eq.(2.10)). Here, σ_s should be sufficiently larger than the typical scale of the fine structure, and comparable with the scale of the coarse structure ($\sigma_s = 6.5$ pixels). In the following, these data are being expressed in the following form:

$$f_{\mu t} \coloneqq f_{txy}\,, \qquad c_{\mu t} \coloneqq c_{txy} \qquad (2.14)$$

where $t = 0, 1, 2, \dots, T-1$: frame number, μ: id of pixels

The significance of this decomposition was observed in the comparison between forward and backward propagation wave in fictive motion. A bouton #1 in Figure 4D exhibits activation at both forward and backward waves. However, the activation at the forward wave had disappeared after the decomposition, which means that the signal had originated from neighboring boutons. Similarly, a bouton #2 demonstrates the opposite case: the activation during forward waves was eliminated by the decomposition. It has been shown that there are several direction specific interneurons in the CNS (Carreira-Rosario et al., 2018; Fushiki et

al., 2016; Hiramoto et al., 2021; Kohsaka et al., 2019). Accordingly, the decomposition step is critical to extract the proper activity of the neurons.

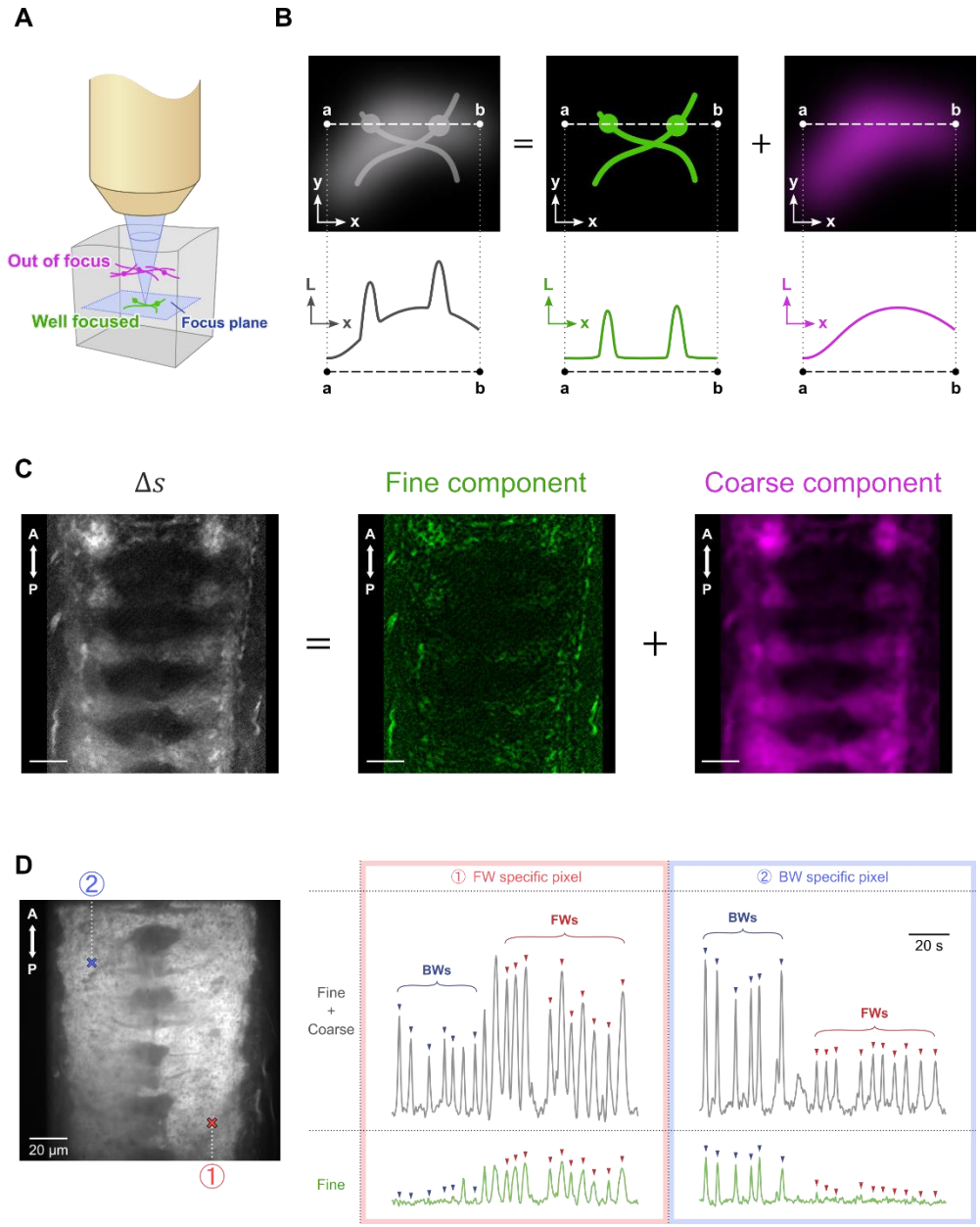


Figure 4. The data were successfully decomposed into fine and coarse structures. (A) Data is described as a summation of fine and coarse structures. (B) Fine and coarse components on frames correspond to well-focused and out-of-focus structures of the CNS, respectively. (C) An example of fine-coarse decomposition on an actual frame. Scale bar: 20 μm. (D) FW- or BW-specific pixels can be recognized only in fine components.

2.3 Pixel clustering

To detect bouton regions within imaging frames in an unsupervised manner, pixel clustering based on the temporal profile comparison was applied to the data $f_{\mu t}$. By the pixel clustering, the imaging data were spatially divided into bouton-sized regions (pixel clusters, each of them consisting of about 20 pixels) based on the temporal similarity in the intensity changes between neighboring pixels. This procedure was devised with the idea that the calcium signal from the points within the same bouton should be similar. Because boutons are densely packed in the neuropil, it is difficult to recognize them solely by their morphological features. This is why we introduced the clustering criteria utilizing the temporal information.

For the pixel clustering, it is important to choose proper definitions for the activity similarity, clustering score, and optimization algorithm. As the simplest indexes showing activity similarity, there are correlation coefficients and the reciprocal of Euclid distance. Here, the correlation coefficient ignores the difference between two norms of the activity traces, whereas Euclid distance is scaled corresponding to the average of two norms of the traces. However, we needed the similarity to reflect the norm difference without being scaled depending on the average of them. Therefore, we defined the similarity between pixels as follows (Figure 5D):

$$\text{Similarity:} \quad A_{\mu\nu} = \arctan\left(\frac{\|\vec{f}_\mu + \vec{f}_\nu\|}{\|\vec{f}_\mu - \vec{f}_\nu\|}\right) \left(\vec{f}_\mu \neq \vec{f}_\nu, \text{ otherwise } A_{\mu\nu} = \frac{\pi}{2}\right) \tag{2.15}$$

where μ, ν: pixel ids, $\vec{f}_\mu = \left(f_{\mu 0}, f_{\mu 1}, \dots, f_{\mu(T-1)}\right)$: activity data of μ

This definition was derived by the following process:

Since similarity should grow larger when data vectors $\vec{f}_\mu$ and $\vec{f}_\nu$ get closer, the reciprocal of Euclid distance is a good indicator for it:

$$\frac{1}{\|\vec{f}_\mu - \vec{f}_\nu\|}$$

To make it independent of the scale of the vectors, the above was converted into a dimensionless quantity by multiplying the norm of the sum vector:

$$\frac{\|\vec{f}_\mu + \vec{f}_\nu\|}{\|\vec{f}_\mu - \vec{f}_\nu\|}$$

To eliminate divergence around $\vec{f}_\mu = \vec{f}_\nu$, we took its arctangent: eq.(2.15).

Next, we defined the clustering score. While many of the existing clustering algorithms, including k-means or hierarchical clustering, need to set the resulting number of clusters, an unsupervised clustering algorithm developed by a previous work (Humphries, 2011) does not require a decision on the number of clusters beforehand. Because it is difficult to predict the exact number of boutons, the method seemed to meet our demands of pixel decomposition. To categorize elements into clusters, the algorithm calculates similarity between elements of all combinations, then a similarity network among the elements is constructed. Next, the algorithm defines the clustering score, which is maximized when the elements are divided into clusters so that connections of similarity are relatively dense within clusters and sparse between clusters.

However, it was hard to directly apply the previous clustering score to our pixel clustering because the number of pixel combinations was too large to manipulate. The method needs to calculate similarities between every combination of elements, which in our case involved data for more than one million pixels, a task that is impossible to carry out on a standard computer. Instead, by utilizing information about spatial position of the pixels, the similarity network can be reduced to a manageable size. The similarity calculation between pixels was localized so that only pairs of spatially close pixels (within five pixels) were calculated. Thanks to this modification, the memory cost of pixel similarity calculation was drastically reduced. Moreover, this method enabled us to control the typical size of clusters by adjusting the maximum distance between a pair of boutons for calculation without fixing the number of clusters.

According to the modified definition of the clustering score, a matrix $B_{\mu\nu}$ was calculated first:

$$B_{\mu\nu} = A_{\mu\nu} - (expected\ \langle \mu, \nu \rangle\ similarity)$$

Here, the second term on the right side was estimated from average similarity around the "neighborhood" of pixels μ and ν. Therefore, $B_{\mu\nu}$ means the relative similarity between μ and ν compared with other pixel pairs around them. If each "neighborhood" region is large enough to include all pixels, the definition of $B_{\mu\nu}$ comes back to the original algorithm (Humphries, 2011).

Then, the clustering score Q was defined as follows:

$$Q = \sum_{\langle \mu,\nu \rangle} B_{\mu\nu}$$

Here, $\langle \mu, \nu \rangle$ denotes pairs of pixels that are both within the same cluster and within the "neighborhood" of each other. The clustering score Q is assumed as a function of cluster assignment. Therefore, the cluster assignment $\vec{S}^* = \underset{\vec{S}}{\operatorname{argmax}}\, Q(\vec{S})$, is the goal of the clustering. $\vec{S}$ is an n-dimension vector (n: the number of pixels in an image) whose components denote a cluster identity that each pixel belongs to (integers, $1 \sim c$, $c = 100$). Since the calculation cost is proportional to the maximum number c, c should be set as small as possible. The number c does not have to be as large as the bouton number, because it is enough to assign different identities to spatially adjacent regions. On the other hand, a too small degree of freedom makes it difficult to find the goal assignment $\vec{S}^*$. Here, $c = 100$ was enough to extract bouton sized clusters robustly.

The detailed definition of Q is shown in the equations below (Through this article, the average is represented as ${}_{i=1}^{n}\langle x_i \rangle = \sum_{i=1}^{n} x_i / n$):

Neighborhood of pixel μ: $\mathcal{V}_1^{(\mu)} = \{\, \rho \mid \mathcal{L}(\rho, \mu) < r \,\}$

Neighborhood of mid-point $\mathrm{M}_{\mu\nu}$: $\mathcal{V}_2^{(\mu\nu)} = \{\, \rho \mid \mathcal{L}(\rho, \mathrm{M}_{\mu\nu}) < r \,\}$

where ρ: pixel id, $\mathcal{L}$: Euclidean distance between two pixels,

$\mathrm{M}_{\mu\nu}$: mid-point of line segment $\mu\nu$, r: typical scale of clusters (Figure 5A)

$$d_\mu = {}_{\rho \in \mathcal{V}_1^{(\mu)}}\langle A_{\mu\rho} \rangle, \qquad m^{(\mu\nu)} = {}_{\rho \in \mathcal{V}_2^{(\mu\nu)}}\langle d_\rho \rangle \tag{2.16}$$

$$B_{\mu\nu} = A_{\mu\nu} - \frac{d_\mu d_\nu}{m^{(\mu\nu)}} \tag{2.17}$$

Clustering score: $$Q(\vec{S}) = \sum_{\mu} \sum_{\nu \in \mathcal{V}_1^{(\mu)}} B_{\mu\nu} \delta_{S_\mu S_\nu} \quad \left(\delta_{xy} = \begin{cases} 1 \ (x = y) \\ 0 \ (x \neq y) \end{cases} \right) \tag{2.18}$$

where S_μ is the cluster identity of pixel μ (integer, $1 \leq S_\mu \leq c$), $\vec{S} = (S_0, S_1, \ldots)$

Here, only values with $\mathcal{L}(\mu, \nu) < r$ in $A_{\mu\nu}$, $B_{\mu\nu}$ and $m^{(\mu\nu)}$ need to be calculated $Q(\vec{S})$. The radius r corresponds to the typical scale of clusters in the clustering results, and we chose here $r = 5$ px as typical size of boutons.

Finally, the clustering score Q was maximized by finding a proper clustering configuration. Here, as the optimization algorithm, we adopted the Simulated Annealing (SA)

with Markov Chain Monte Carlo (MCMC) method (Figure 5B), which is computationally efficient, especially when interaction between elements is localized. In SA, a system develops randomly according to a thermal state distribution of an energy (here, $E = -Q$) and a temperature. As the temperature gradually approaches absolute zero, the system finally settles into the energy minimum state (then, Q is maximized). The thermal state distribution was produced by the MCMC method. MCMC simulates the distribution without knowing the global landscape of the energy function, and this is why SA works as an optimization method. The calculation cost of MCMC is drastically reduced if the interaction among variables is localized.

The optimization processing with SA and MCMC is formulated as follows. In MCMC, the cluster identity of each pixel was updated iteratively by Gibbs sampling:

$$\text{Probability of } S_\mu = a: \quad P(S_\mu = a) = \frac{e^{\beta Q(S_\mu = a)}}{\sum_b e^{\beta Q(S_\mu = b)}} \tag{2.19}$$

where β: inverse temperature ($\beta \geq 0$)

Because of the localized connectivity, most items in the score Q are independent of S_μ. Accordingly, the fraction of probability is reduced to the following:

$$P(S_\mu = a) = \frac{\exp\left(\beta \sum_{\nu \in \mathcal{V}_1^{(\mu)}, \nu \neq \mu} B_{\mu\nu} \delta_{a S_\nu}\right)}{\sum_b \exp\left(\beta \sum_{\nu \in \mathcal{V}_1^{(\mu)}, \nu \neq \mu} B_{\mu\nu} \delta_{b S_\nu}\right)} \tag{2.20}$$

Apparently, Q itself does not have to be calculated explicitly to perform MCMC according to eq.(2.20), which drastically reduces the calculation cost. According to the probability P, identity of every pixel was updated within one iteration in a random order, and the inverse temperature was enhanced according to $\beta_i = 0.01 \times i$ (i: iteration, $i = 0,1,2,\ldots$) and to $\beta_{max} = 200$. As seen in Figure 5C, the cluster configurations of $\beta = 100$ and $\beta = 200$ were almost the same, which means the system was sufficiently converged at $\beta = 200$.

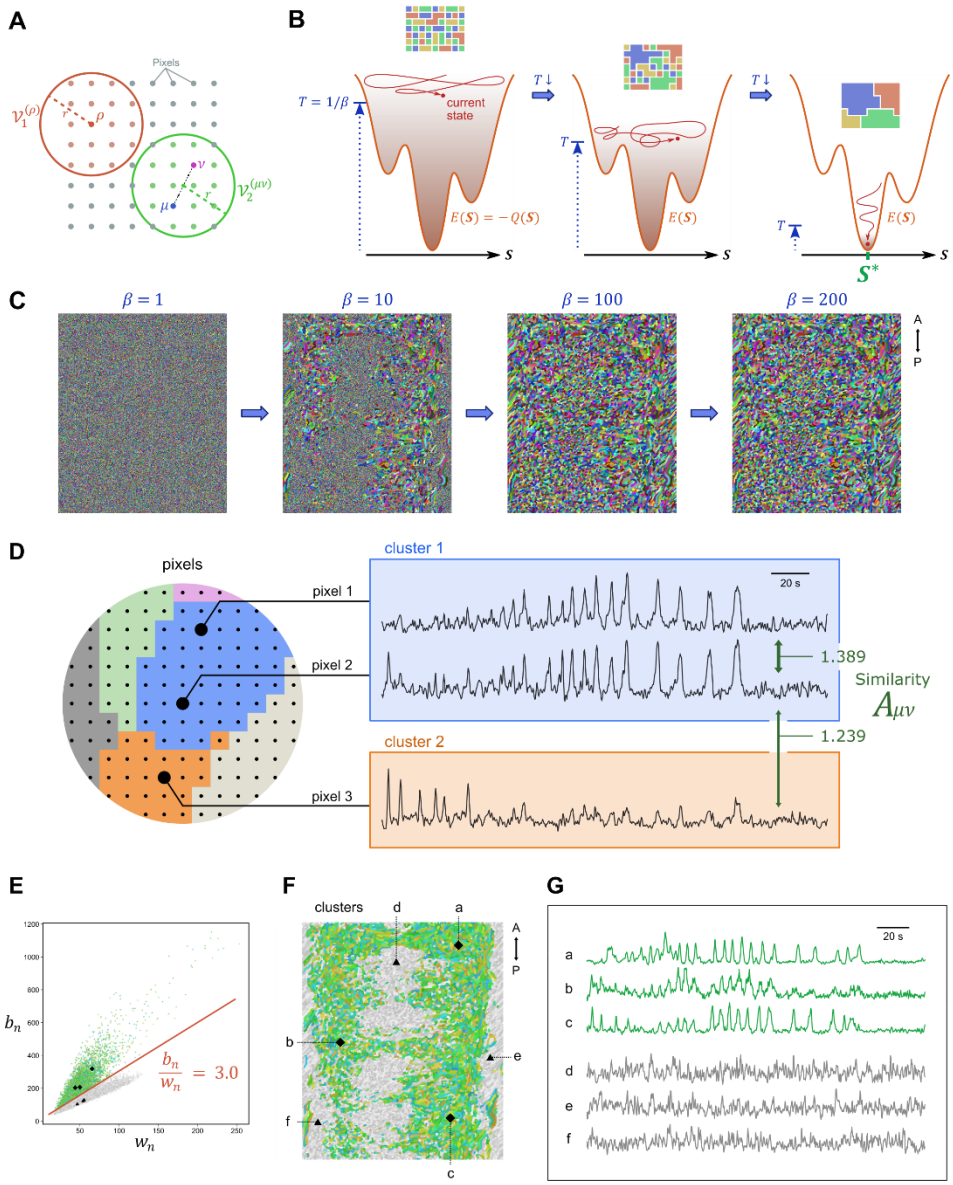


Figure 5. Pixels in imaging data were clustered into boutons. (A) Definition of $\mathcal{V}_1^{(\mu)}$ and $\mathcal{V}_2^{(\mu\nu)}$. (B) Schematic diagram explaining how SA works. (C) Transition of cluster identity configuration during elevation of inverse temperature β. (D) Examples of pixels with activity traces. Pixel 1 and 2 were categorized into cluster 1, whereas pixel 3 belongs to cluster 2. (E)(F)(G) Pixel clusters with sufficient S/N ratio (green) and discarded clusters (gray). (E) Scatter diagram of pixel clusters. Clusters above the orange line were adopted. (F) Cluster configuration on a focus plane. (G) Signal traces of clusters whose position is shown in (F).

The clustering result $\vec{S}^*$ was used to define boutons. First, was decomposed into

spatially connected components $\{R'_n | n = 1,2,\dots,N'\}$, and the signals $f_{\mu t}$ in each R'_n were averaged $\left(F'_{nt} = {}_{\mu \in R'_n}\langle f_{\mu t}\rangle\right)$. Here, these contain regions that show almost no calcium signal, such as regions outside the neuropil. Therefore, to eliminate such regions, we performed threshold processing to selectively extract regions with sufficient S/N ratio from $\{R'_n\}$. To evaluate the S/N ratio, the following two quantities were compared:

$$w_n^2 = {}_t\left\langle\left(\frac{F'_{n(t+1)} - F'_{nt}}{2}\right)^2\right\rangle, \qquad b_n^2 = {}_t\left\langle\left(\frac{F'_{n(t+1)} + F'_{nt}}{2}\right)^2\right\rangle - {}_t\left\langle\frac{F'_{n(t+1)} + F'_{nt}}{2}\right\rangle^2$$

w_n is the average of deviation of neighboring data points, and b_n is the deviation of the average of neighboring data points. Both values are proportional to signal strength and noise level, but w_n is mainly affected by noise, and b_n is mainly affected by signal intensity. Therefore, the fraction b_n/w_n can be assumed as an S/N ratio indicator (white noise shows $b_n/w_n = 1$). By choosing pixel clusters with more than $b_n/w_n = 3.0$, clusters $\{R_n | n = 1,2,\dots,N\}$ ($N = 5{,}000 \sim 15{,}000$ for each sample) were extracted (Figure 5E, F, G):

$$F_{nt} = {}_{\mu \in R_n}\langle f_{\mu t}\rangle, \qquad \frac{b_n}{w_n} > 3.0 \quad \text{for each } n = 1,2,\dots,N$$

The extracted clusters were bouton-sized regions in which pixels show relatively similar activities. Because it is natural that points within a single bouton show uniform activity, the regions of the clusters should reflect the distribution of boutons.

Results

Boutons were extracted from calcium imaging data of the neuropil

To demonstrate our methods in fly larvae, we measured neural population activity by three-dimensional calcium imaging. Previous recordings for neural population activity were largely performed on cell bodies as signal sources rather than neurites (Lemon et al., 2015; Pulver et al., 2015). However, because the actual communication between neurons occurs at synapses, there should be some important features on spatial structures in the synaptic population. Therefore, we performed populational recordings of boutons, which are swollen pre-synaptic structures, instead of somas.

In the imaging, spontaneously propagating forward activity waves (FW) and backward activity waves (BW) were recorded on the isolated CNS, with the objective lens vibrating along the z-direction to record in three dimensions. We recorded synapse population activity in the VNC neuropil, where synapses are densely packed. During the imaging, the objective lens was vibrating along the z-direction to record the three-dimensional volume of the neuropil. As a result, we succeeded in recording spontaneous FW and BW in the isolated CNS corresponding to fictive motion on multiple focus planes (six specimens, including 23 FW events and 29 BW events in total) (Figure 6A).

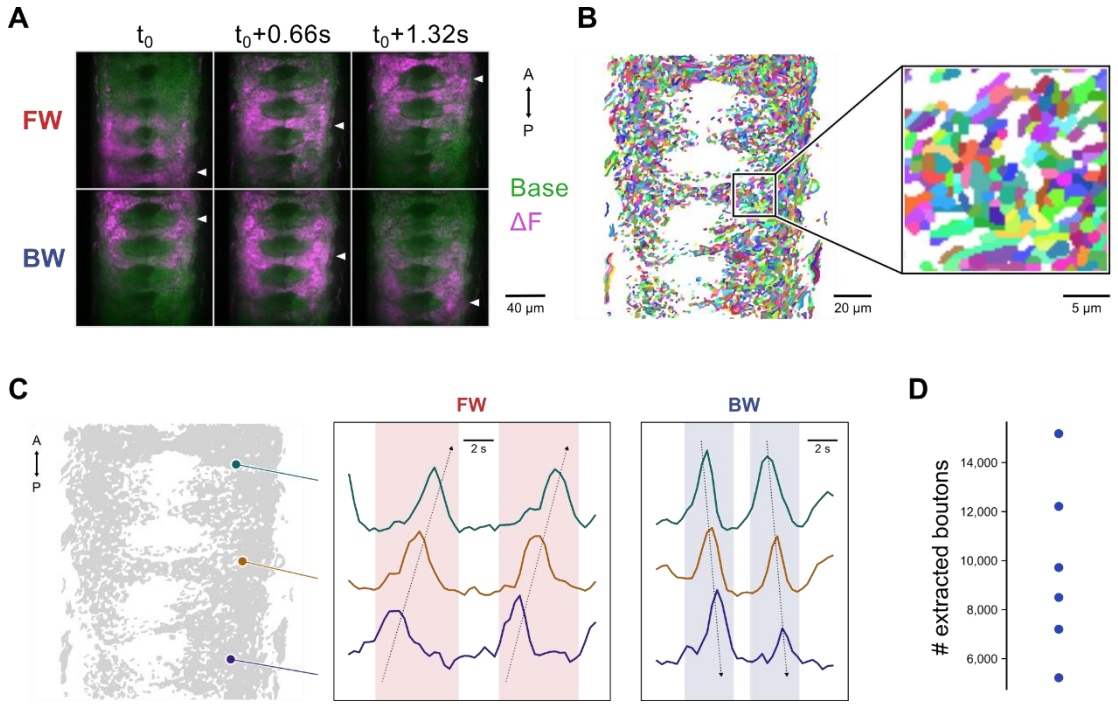


Figure 6. Boutons were extracted from three-dimensional calcium imaging of the neuropil. (A) Representative FW and BW of a sample, recorded by three-dimensional Ca2+ imaging. One focus plane out of five is displayed. Base signal of a resting frame (green) and difference signal against the base signal (magenta). (B) Extracted bouton configuration on the same plane with (A). The whole configuration (left) and an expanded image (right). (C) Signal traces of exemplified three extracted boutons in (B). Two FW events and two BW events are displayed. (D) The number of extracted boutons for each sample.

Next, our bouton extraction method was applied to the imaging data. As a result, the data were decomposed into pixel cluster regions of about 2-3 μm diameter, which is comparable to the typical size of boutons (Figure 6B). For each bouton-sized region, we obtained a calcium signal trace, including FW and BW events (Figure 6C). In the signal traces, we could observe FW and BW activity waves in which signal peaks propagate among the bouton-sized regions toward anterior and posterior, respectively. The number of boutons per sample was about 5,000-15,000 (Figure 6D). Here, the number of boutons depends on data quality, because we discarded pixel cluster regions with insufficient S/N ratio. Using our method, we obtained an extraordinarily large number of calcium transients, which are linked with the spatial locations of putative boutons.

Discussion

In this study, we established a novel method to decompose Ca2+ imaging data into synapse-sized pixel clusters. We applied baseline correction, fine structure extraction, and pixel clustering to calcium imaging data of the neuropil of Drosophila larvae. This procedure allowed us to extract the dynamics of activity of each bouton from the densely packed neuropil.

First, we developed a method to eliminate photobleaching, sample deformation and drift, and out-of-focus signals from imaging data. Because of the high density of synapses, we had to utilize temporal information of activity along with spatial information to decompose synapses. Therefore, it was important to acquire proper temporal profiles of pixels. Actually, the sample deformation and drift severely affect the synapse extraction, because the synapse size is so small that a gentle sample deformation can displace synapses more than their diameter. Accordingly, the pre-processing we developed is crucial to detecting synapses.

Next, we engineered a pixel clustering algorithm. One of the major advantages in our algorithm is the small number of parameters that need to be determined manually. Basically, only the typical size of pixel clusters needs to be determined, based on anatomical evidence, whereas the number of clusters is automatically obtained. In addition, the algorithm does not need information of any physiological features, so no parameter such as spike width or other time constant has to be adjusted. By simply fixing the typical cluster size, frames of imaging data are decomposed into regions with proximal sizes to that parameter.

Another advantage of our pixel clustering is the relatively small computational and memory cost required for the pixel clustering. Comparison of only neighboring pixels instead of all-to-all comparison to calculate the similarity matrix drastically reduced the calculation time as well as the required memory. Similarly, with the calculation of the similarity matrix, the calculation cost for each update during the MCMC iteration is dramatically reduced, because of the localized interaction between pixels. This is because the change of the clustering score based on the localized interaction is only affected by interaction between the pixel updating at each time and its neighbors.

Thanks to the small computational cost, we could extract as much as 5,000–15,000 boutons per sample using a standard desktop computer (maximum speed: 3.60 GHz, 4-core CPU, 8 GB RAM). Actually, the number is extremely large for the clustered activity units

compared with previous examples. Previous methods decomposed pan-neuronal imaging data into typically less than 1,000 cell bodies. However, analysis of the neuropil requires extraction of synapses, the number of which is 10 times or more larger than cell bodies. Therefore, it is difficult to apply previous methods to the synapse decomposition, and our achievement of improving the calculation efficiency is essential for the neuropil analysis.

It is important to note the limitation of our clustering method. Pixel clusters obtained by our methods can include not only a whole single bouton but also a fragment of a single bouton or a group of boutons. This is because our method fixes the typical size of a cluster, so a bouton larger than that scale might be split into multiple parts, or smaller boutons might be grouped into a single bouton. Our algorithm might also group spatially overlapping boutons if the focus plane has non-negligible thickness, because distinct boutons with different depths can merge onto the same pixel. While some previous algorithms can decompose a pixel into multiple overlapping components (Pnevmatikakis et al., 2016; Shibue and Komaki, 2020), our pixel clustering method distributes exactly one cluster identity to each pixel.

Nevertheless, our method provides vast accessibility to the spatiotemporal data of population activity in the neuropil. Even if some of the pixel clusters are not exactly boutons, each signal trace shows neural activity that actually exists at that location in the synapse population. Therefore, the spatiotemporal population activity data extracted by our method should contain more information on the processing conducted in the neural circuits than cell body imaging data. It is expected that analyses of the synapse population data will elucidate circuit activity dynamics with respect to its spatial structure, and enhance our comprehension of the nervous system.

Acknowledgements

This work was supported by MEXT/JSPS KAKENHI grants (17K19439, 19H04742, and 20H05048 to A.N. and 17K07042 and 20K06908 to H.K.). We would like to thank D. Miyamoto, S. Takagi, A. Hiramoto, Y. Liu and X. Zeng for informative discussions. We would like to thank Bloomington Drosophila Stock Center for providing fly lines.

References

Bilz, F., Geurten, B.R.H., Hancock, C.E., Widmann, A., and Fiala, A. (2020). Visualization of a Distributed Synaptic Memory Code in the Drosophila Brain. Neuron 963–976.

Carreira-Rosario, A., Zarin, A.A., Clark, M.Q., Manning, L., Fetter, R., Cardona, A., and Doe, C.Q. (2018). A command-like descending neuron that coordinately activates backward and inhibits forward locomotion. BioRxiv 1–28.

Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., Shenoy, K. V., and Shenoy, K. V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56.

Fushiki, A., Zwart, M.F., Kohsaka, H., Fetter, R.D., Cardona, A., and Nose, A. (2016). A circuit mechanism for the propagation of waves of muscle contraction in Drosophila. Elife *5*, 1–23.

Gauthier, J.L., and Tank, D.W. (2018). A Dedicated Population for Reward Coding in the Hippocampus. Neuron *99*, 179-193.e7.

Gründemann, J., Bitterman, Y., Lu, T., Krabbe, S., Grewe, B.F., Schnitzer, M.J., and Lüthi, A. (2019). Amygdala ensembles encode behavioral states. Science (80-.). *364*.

Hackett, T.A., Barkat, T.R., O'Brien, B.M.J., Hensch, T.K., and Polley, D.B. (2011). Linking topography to tonotopy in the mouse auditory thalamocortical circuit. J. Neurosci. *31*, 2983–2995.

Han, C., Jan, L.Y., and Jan, Y.N. (2011). Enhancer-driven membrane markers for analysis of nonautonomous mechanisms reveal neuron-glia interactions in Drosophila. Proc. Natl. Acad. Sci. U. S. A. *108*, 9673–9678.

Hiramoto, A., Jonaitis, J., Niki, S., and Kohsaka, H. (2021). Regulation of coordinated muscular relaxation by a pattern- generating intersegmental circuit. Nat. Commun. 1–14.

Humphries, M.D. (2011). Spike-train communities: Finding groups of similar spike trains. J. Neurosci. *31*, 2321–2336.

Kohsaka, H., Takasu, E., Morimoto, T., and Nose, A. (2014). A group of segmental premotor interneurons regulates the speed of axial locomotion in drosophila larvae. Curr. Biol. *24*, 2632–2642.

Kohsaka, H., Zwart, M.F., Fushiki, A., Fetter, R.D., Truman, J.W., Cardona, A., and Nose, A. (2019). Regulation of forward and backward locomotion through intersegmental feedback circuits in Drosophila larvae. Nat. Commun. *10*, 1–11.

Landgraf, M., Sánchez-Soriano, N., Technau, G.M., Urban, J., and Prokop, A. (2003). Charting the Drosophila neuropile: A strategy for the standardised characterisation of genetically amenable neurites. Dev. Biol. *260*, 207–225.

Lemon, W.C., Pulver, S.R., Höckendorf, B., McDole, K., Branson, K., Freeman, J., and

Keller, P.J. (2015). Whole-central nervous system functional imaging in larval Drosophila. Nat. Commun. *6*.

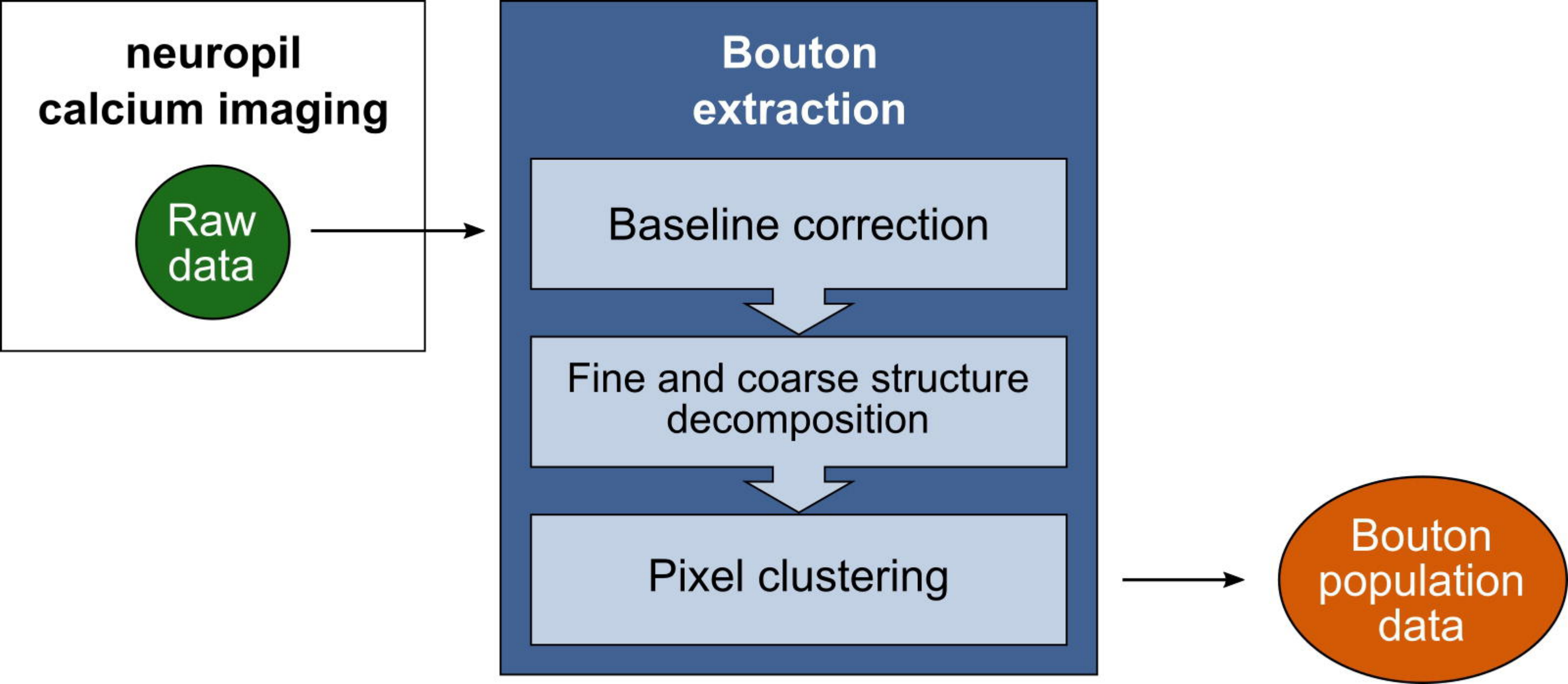
Pnevmatikakis, E.A., Soudry, D., Gao, Y., Machado, T.A., Merel, J., Pfau, D., Reardon, T., Mu, Y., Lacefield, C., Yang, W., et al. (2016). Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. Neuron *89*, 285.

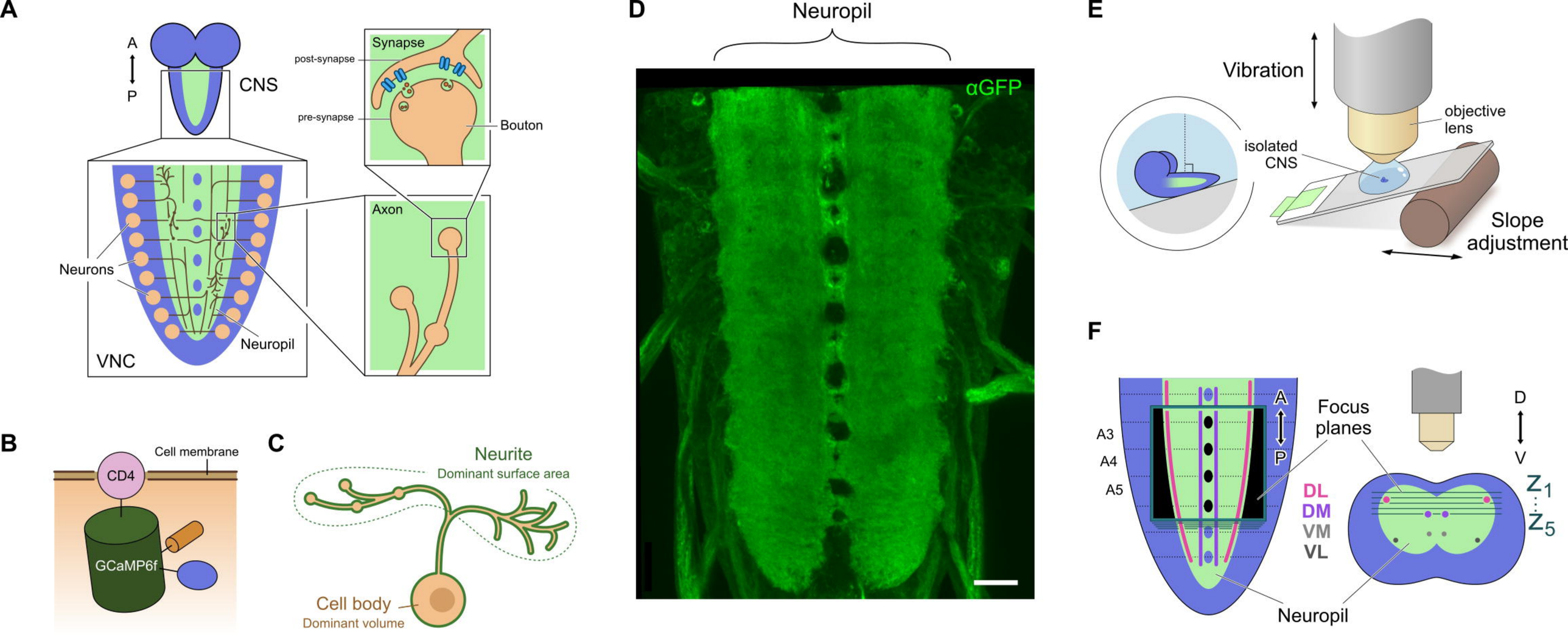
Pulver, S.R., Bayley, T.G., Taylor, A.L., Berni, J., Bate, M., and Hedwig, B. (2015). Imaging fictive locomotor patterns in larval Drosophila. J. Neurophysiol. *114*, 2564–2577.

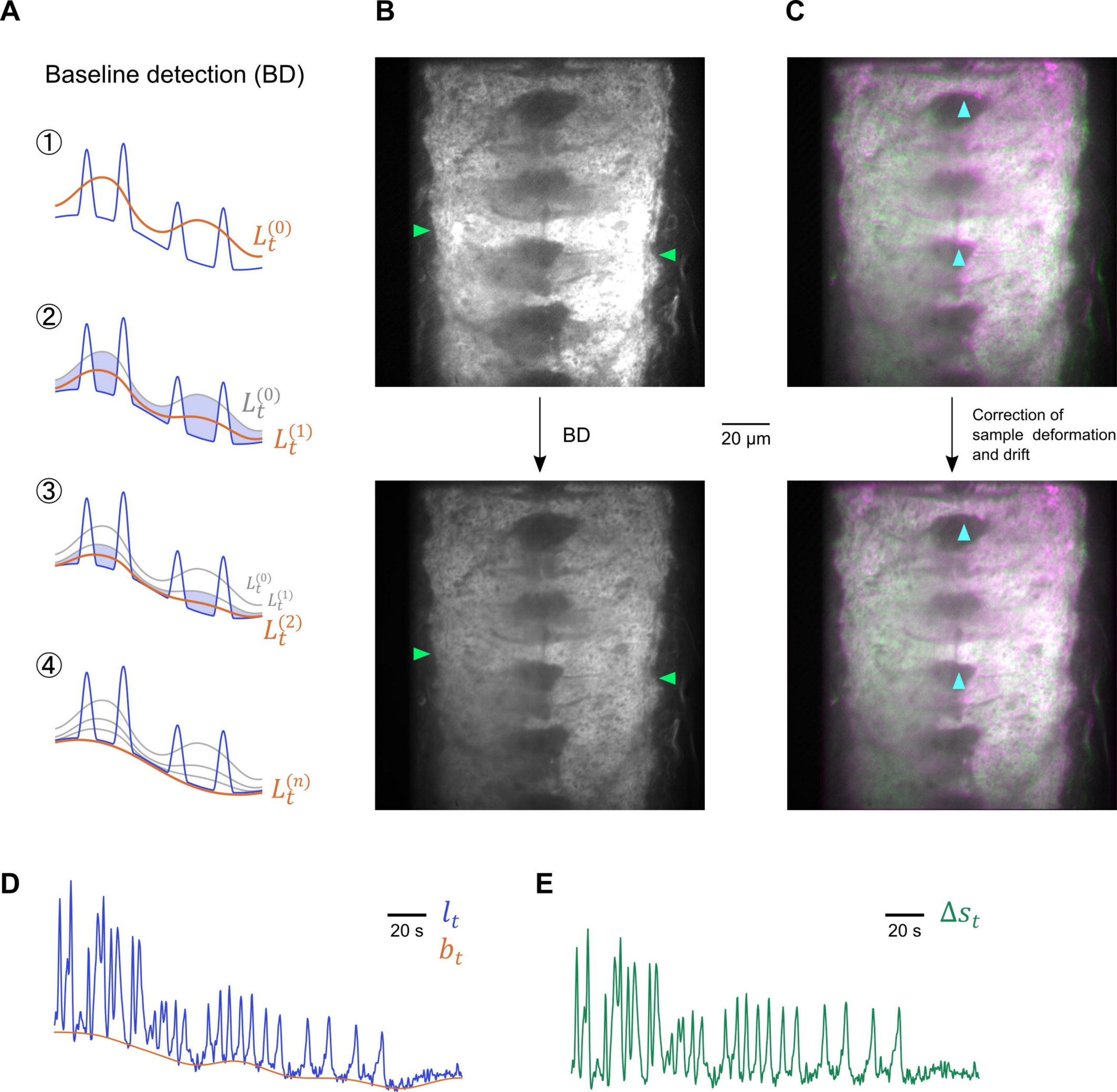
Shibue, R., and Komaki, F. (2020). Deconvolution of calcium imaging data using marked point processes. PLoS Comput. Biol. *16*, 1–25.

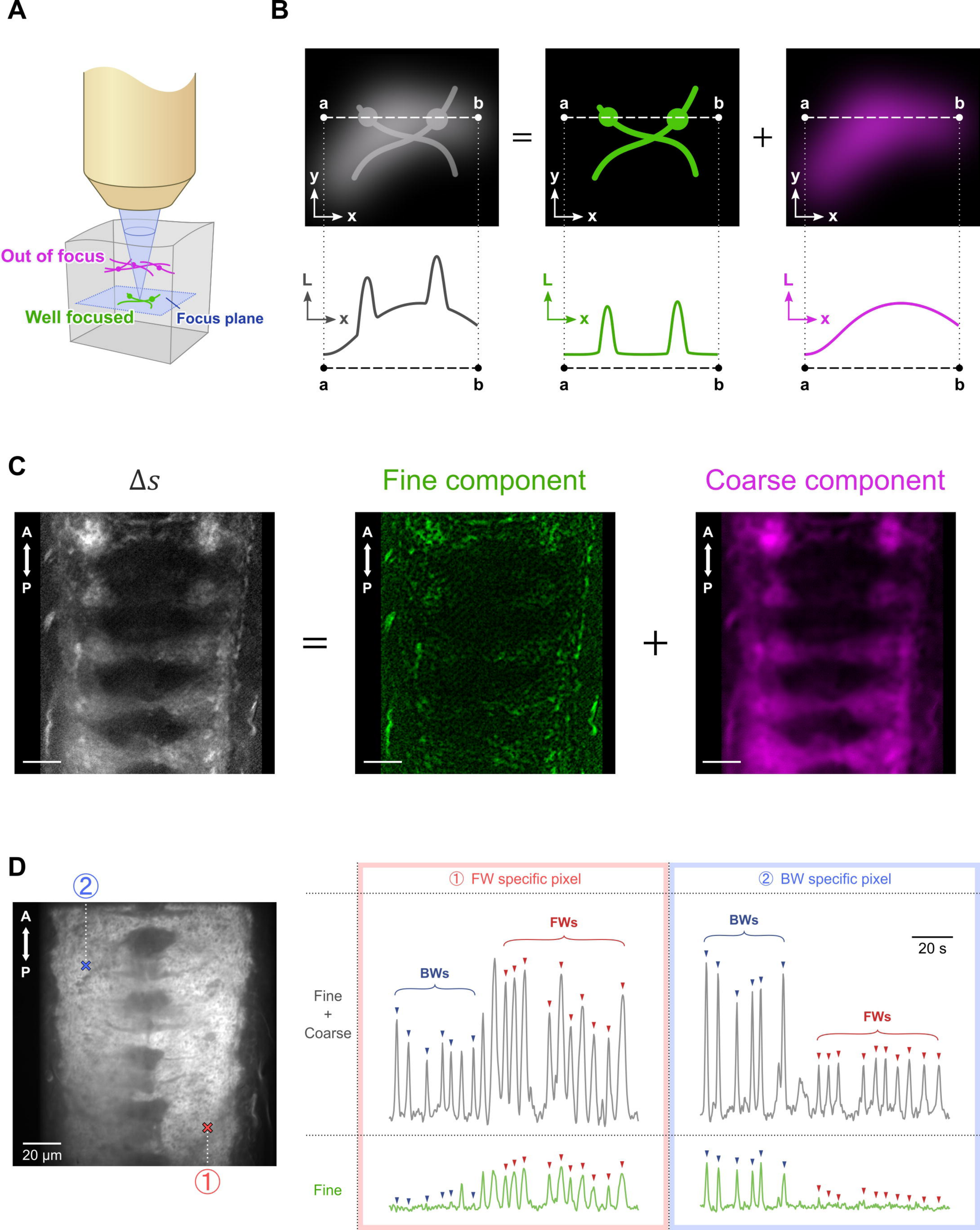
Tinnermann, A., Büchel, C., and Cohen-Adad, J. (2021). Cortico-spinal imaging to study pain. Neuroimage *224*, 117439.

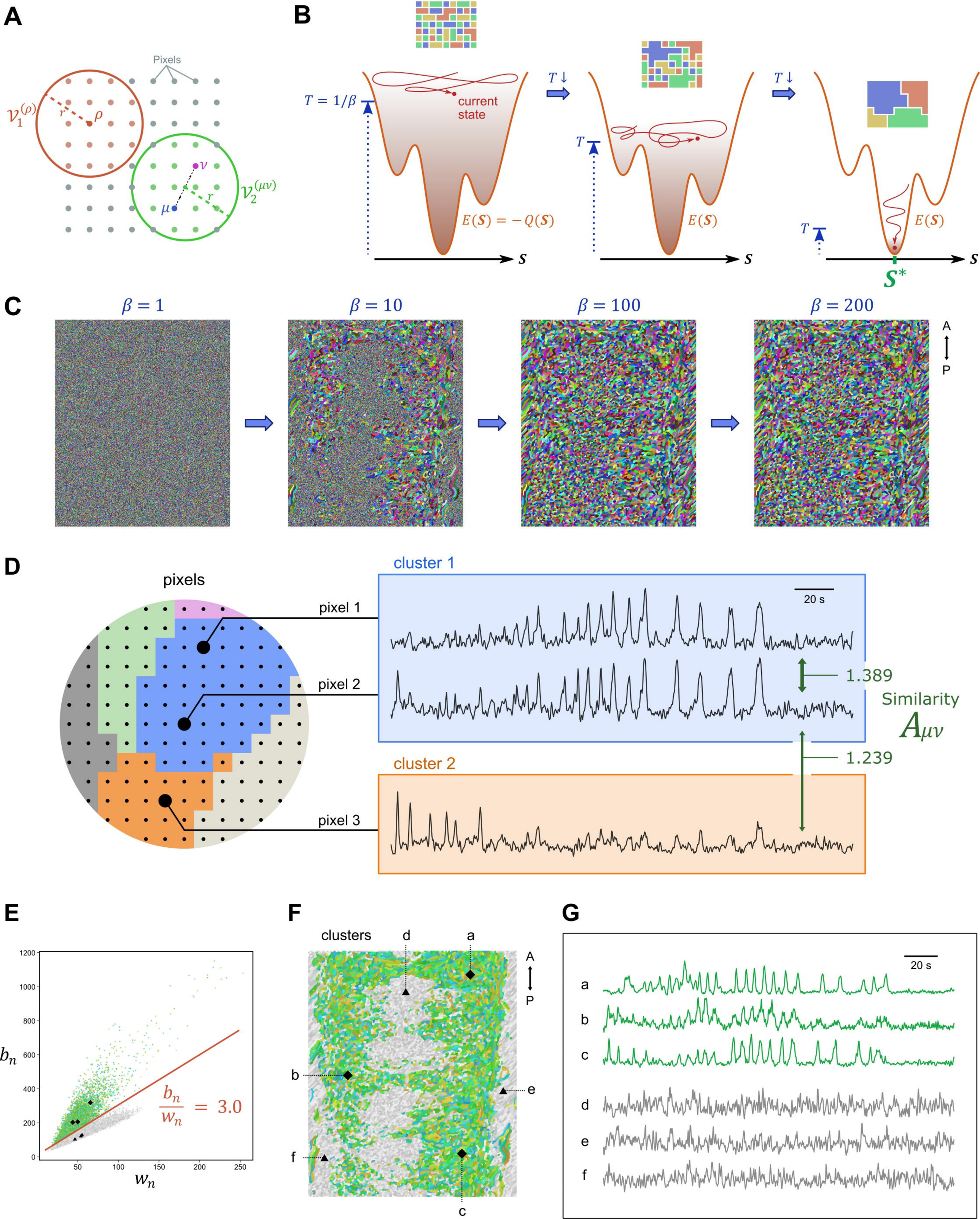
Zhuang, J., Ng, L., Williams, D., Valley, M., Li, Y., Garrett, M., and Waters, J. (2017). An extended retinotopic map of mouse cortex. Elife *6*, 1–29.

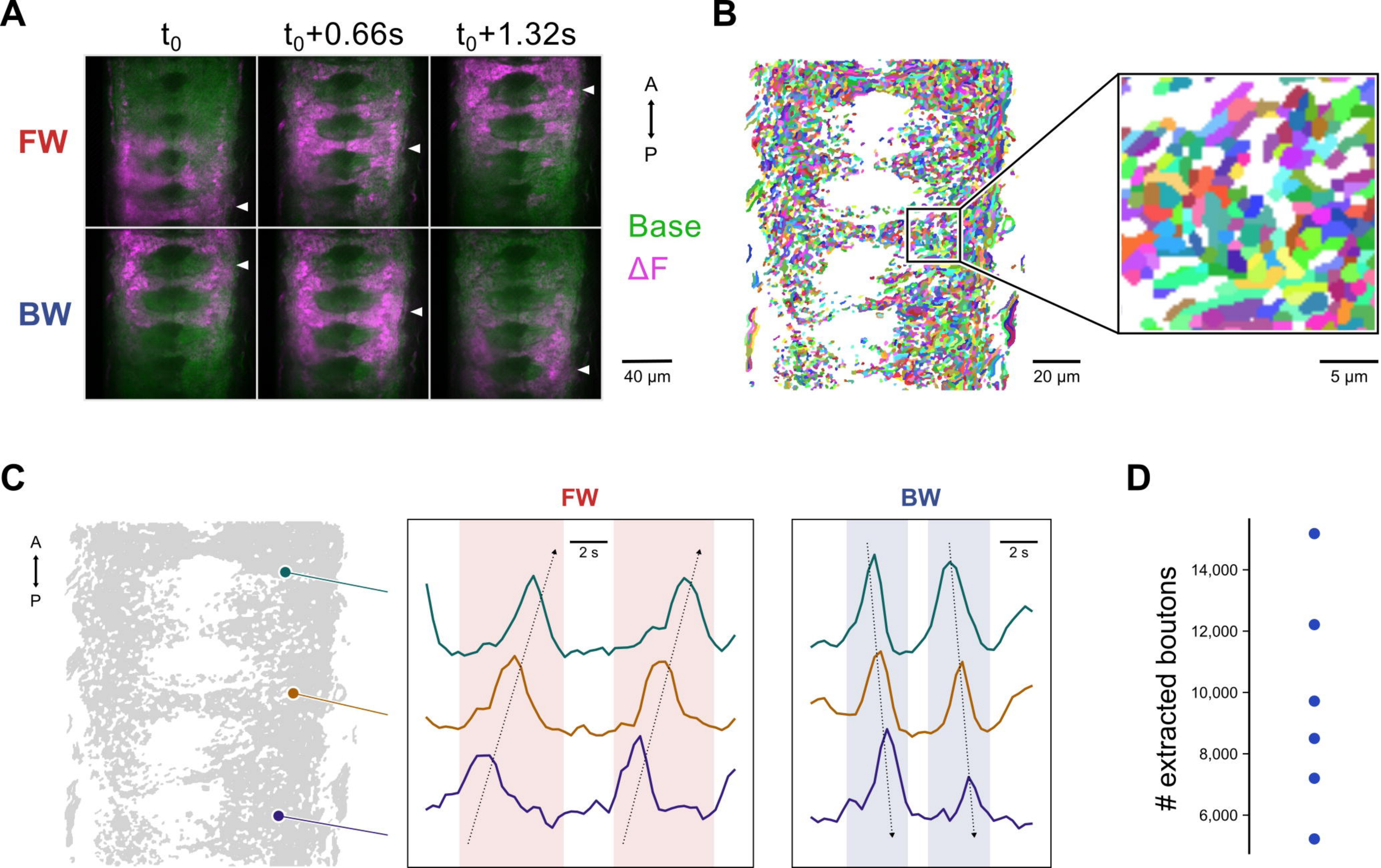

neuropil
calcium imaging
Raw
data
Bouton
extraction
Baseline correction
Fine and coarse structure
decomposition
Pixel clustering
Bouton
population
data


A
A
P
CNS
Synapse
post-synapse
pre-synapse
Bouton
Axon
Neurons
Neuropil
VNC
B
Cell membrane
CD4
GCaMP6f
C
Neurite
Dominant surface area
Cell body
Dominant volume
D
Neuropil
αGFP
E
Vibration
objective lens
isolated CNS
Slope adjustment
F
A3
A4
A5
A
P
Focus planes
DL
DM
VM
VL
D
V
Z1
Z5
Neuropil


A
Baseline detection (BD)
①
$L_t^{(0)}$
②
$L_t^{(0)}$
$L_t^{(1)}$
③
$L_t^{(0)}$
$L_t^{(1)}$
$L_t^{(2)}$
④
$L_t^{(n)}$
B
BD
20 μm
C
Correction of sample deformation and drift
D
20 s
l_t
b_t
E
20 s
Δs_t


A
Out of focus
Well focused
Focus plane
B
a
b
y
x
L
=
+
C
Δs
Fine component
Coarse component
A
P
D
②
①
20 μm
① FW specific pixel
② BW specific pixel
Fine
+
Coarse
Fine
BWs
FWs
20 s


A
Pixels
$\mathcal{V}_1^{(\rho)}$
r
ρ
ν
μ
r
$\mathcal{V}_2^{(\mu\nu)}$
B
$T = 1/\beta$
current state
$E(\boldsymbol{S}) = -Q(\boldsymbol{S})$
$\boldsymbol{S}$
$T\downarrow$
T
$E(\boldsymbol{S})$
$\boldsymbol{S}$
$T\downarrow$
T
$E(\boldsymbol{S})$
$\boldsymbol{S}$
$\boldsymbol{S}^*$
C
$\beta = 1$
$\beta = 10$
$\beta = 100$
$\beta = 200$
A
P
D
pixels
pixel 1
pixel 2
pixel 3
cluster 1
20 s
1.389
Similarity
$A_{\mu\nu}$
1.239
cluster 2
E
b_n
w_n
$\frac{b_n}{w_n} = 3.0$
F
clusters
a
b
c
d
e
f
A
P
G
20 s
a
b
c
d
e
f


A
t_0
t_0+0.66s
t_0+1.32s
FW
BW
A
P
Base
ΔF
40 µm
B
20 µm
5 µm
C
A
P
FW
BW
2 s
2 s
D
extracted boutons
14,000
12,000
10,000
8,000
6,000