

Methods for Analyzing GCaMP Calcium Imaging

Houde He^{1*}, Sheng Huang²

¹*Department of communication engineering, Xi Dian University, Xian, China*

²*Living Word Shanghai, Shanghai, China*

**Corresponding Author. Email: 1745148575@qq.com*

Abstract: GCaMP is a genetically encoded calcium indicator that is most widely used to observe the activities of populations of as many as thousands of neurons simultaneously using modern fluorescence imaging techniques. GCaMP is widely used to monitor neural activity in living animals. For example, researchers can use GCaMP to observe calcium transients in neurons, which are proxies for action potentials and synaptic activity.(GCaMP - an Overview | ScienceDirect Topics, 2015). However, calcium imaging itself has several limitations. First, the calcium imaging has a poor signal to noise ratio(SNR) that makes the imaging difficult to detect dynamic signal and subtle fluctuations. Second, the fluorescence of GCaMP is slower than the action potentials from neuron activities especially in quick succession. The third issue is that there are millions of neurons that researchers investigated, but each neuron presents different activity and relation. Therefore, accurately targeting a specific group of neurons that perform similar tasks is challenging in the experiment. Aimed at these challenges, four successive methods including High pass filter, Gaussian Mixed Model, correlation matrix, and deconvolution were used to improve the analysis of neuron activity. By using these methods to analyze the data set from two GCaMP6 (a particular version of GCaMP) fluorescence recording data sets containing the time series traces of hundreds of neurons in the mouse primary visual cortex (VI) residing within a three-dimensional volume approximately $800\mu\text{m} \times 800\mu\text{m} \times 100\mu\text{m}$ in size. The results from the data analysis showed that the filtering effect of FIR high pass filter is the most significant because it significantly enhance the SNR and reduce noise. Through Gaussian Mixed Model and correlation coefficient, it clearly presents the connectivity of each neuron in a 233 times 233 matrix R and indicates the distribution of neuron activities by fitting into the Gaussian curve. The deconvolution successfully infer potential spikes. These methods efficiently enhance the noise reduction, network connectivity and temporal resolution of the analysis of imaging.

Keywords: GCaMP, filtering, High-Pass Filter, Gaussian Mixed Model, Correlation coefficient, Deconvolution

1. Introduction

The indicator consists of three key components: Green Fluorescent Protein (GFP), Calmodulin (CaM), and the M13 peptide. When calcium ions bind to the calmodulin, the protein undergoes a conformational change that increases the fluorescence of GFP, allowing researchers to visualize calcium

activity. As GCaMP is now a very popular genetic indicator that could be used to track different cell's activity through the fluorescence triggered by the calcium pump in the cell. There are large amounts of paper related to methods of analyzing the calcium imaging. For instance, some paper explore the software used to analyze the calcium data optimally and consistently across different groups. They addressed this unmet need by incorporating recent software solutions, namely NoRMCorre and CaImAn, for motion correction, segmentation, signal extraction, and deconvolution of calcium imaging data into an open-source (Cantu et al., 2020). However, researchers might need to get a more accurate and conclusive result from the imaging. There are several ways to achieve that, for instance, the development of fluorometric calcium sensors, new approaches for targeted labeling with these sensors and the implementation of powerful imaging techniques, especially two-photon microscopy. (Rocheffort et al., 2008). What we did is using succession of methods from primary processing of data set to the final spike inference to get a more accurate result.

2. Method

It's a research of the methods of optimizing data analysis of GCaMP calcium imaging. For setting up all of the methods Matlab should be installed for programming the data. And the data set should be prepared for analysis. The data we found is from two GCaMP6 (a particular version of GCaMP) fluorescence recording data sets containing the time series traces of hundreds of neurons in the mouse primary visual cortex (VI) residing within a three-dimensional volume approximately $800\mu\text{m} \times 800\mu\text{m} \times 100\mu\text{m}$ in size. For every neuron n , the GCaMP fluorescence recording sample at each time point m is given as a relative change in fluorescence with respect to a baseline fluorescence level. Let N be the total number of neurons in the aforementioned three-dimensional volume field-of-view and M be the total number of time sample points. Each data file contains three variables: FR, allLoc,data.

$$\frac{\Delta F}{F_0}[m] = \frac{F[m] - F_0}{F_0} \quad (1)$$

2.1. High pass filter

To build a high pass filter, we need to first set up a $N \times M$ matrix. Then, applying Fourier transform on every neuron and compute the average Fourier magnitude. Eventually, using the plot method to visualize the average Fourier magnitude spectrum. Now, setting up the FIR high pass filter with a specific order $L = 50$ and cutoff frequency of $\omega = 0.0001$. In the frequency domain, the ideal high-pass filter is defined, for some cutoff frequency ω_c , as follows:

$$H_{\text{ideal}}(e^{i\omega}) = \begin{cases} 0, & \text{if } \omega \leq \omega_c \\ 1, & \text{otherwise} \end{cases} \quad (2)$$

Filter each neuron by the FIR high pass filter and choose one of the neurons to plot. Now, to construct a more ideal high pass filter, the order of the filter is halved and the sin function was used to get an ideal cutoff frequency. In addition, it's optional to use Hanning Window to continue to reduce the side lobe effect and noise to improve the SNR. Then, plotting the coefficient $h[n]$ of the ideal high pass filter. Implement the approximately ideal high-pass filter in the time domain for some large L . Filter the raw ΔF data using the filter taps in the time domain using convolution. At the end, comparing two filtered data from one neuron to determine whether it's efficient to filter the noise. Trying different types of the high pass filters such as Butterworth filter, Chebyshev filter. It gives the researcher more possibilities to find out the most efficient high pass filter to reduce the noise in the calcium imaging.

2.2. Gaussian Mixed Model

Before fitting the Gaussian Mixture Model, you must first apply a high-pass filter to remove low-frequency noise and extract the relevant neural signal. Initializes an array of constructs to store the Gaussian mixture model for each neuron. The result of this preprocessing is the high-pass-filtered signal, denoted as y_n . The next step is to fit a two-component GMM to the data. This will model the data as a mixture of two Gaussian distributions, representing the baseline state $N(\mu_{n0}, \sigma^2_{n0})$ and the excited state $N(\mu_{n1}, \sigma^2_{n1})$ of neuronal activity. The probability density function of the GMM fit is plotted. The GMM defines the probability density function $f_n(x)$ of the high-pass-filtered data as a mixture of these two Gaussian distributions:

$$f_n(x) = \omega_0 f_{N(\mu_{n0}, \sigma^2_{n0})}(x) + \omega_1 f_{N(\mu_{n1}, \sigma^2_{n1})}(x) \quad (3)$$

For each neuron n , we fit a two-component GMM to the high-pass-filtered data using the Expectation-Maximization (EM) algorithm, which iteratively estimates the model parameters that best describe the data as a combination of two Gaussian distributions. Analyze the fitted parameters to differentiate between the baseline and excited neural activity states including the means (μ_{n0}, μ_{n1}), the variances ($\sigma^2_{n0}, \sigma^2_{n1}$), and the mixing coefficients (ω_0, ω_1). The image instructs to use Python or MATLAB to fit the two-component GMM to y_n . This involves using the expectation-maximization (EM) algorithm or other optimization methods built into libraries like scikit-learn (Python) or fitgmdist (MATLAB) to find the optimal values for $\mu_{n0}, \mu_{n1}, \sigma_{n0}, \sigma_{n1}, \omega_0, \omega_1$.

2.3. Correlation Coefficient Heat Map

Before calculating the correlation coefficients, the fluorescence time-series data for each neuron must be standardized. Standardization ensures that differences in the scale of fluorescence signals do not bias the correlation calculations. Each neuron's time series is transformed into a z-score representation, normalizing it so that the data has a mean of zero and a standard deviation of one.

$$z_i = \frac{x_i - \mu_i}{\sigma_i} \quad (4)$$

This step ensures that all neurons are on the same scale before calculating correlation values. Once the data is standardized, we calculate the pairwise Pearson correlation coefficients between the time-series data of every pair of neurons. The Pearson correlation coefficient measures the linear relationship between two neurons' activities. It ranges from -1 to 1, the Pearson correlation coefficient for two neurons i and j is calculated as:

$$R_{ij} = \frac{\sum_{t=1}^T (z_{i,t} - \bar{z}_i)(z_{j,t} - \bar{z}_j)}{\sqrt{\sum_{t=1}^T (z_{i,t} - \bar{z}_i)^2} \sqrt{\sum_{t=1}^T (z_{j,t} - \bar{z}_j)^2}} \quad (5)$$

The matrix R is symmetric, with diagonal elements $R_{ii} = 1$, representing the self-correlation of each neuron. To gain insight into the connectivity patterns, we visualize the correlation matrix. Plotting the matrix as a heatmap allows for a quick assessment of how neurons interact with each other. In the heatmap. The resulting correlation matrix provides insights into the functional connectivity within the neural population. Highly correlated neurons might belong to the same network or functional group, suggesting cooperative activity. To perform further network analysis, metrics such as degree centrality, betweenness centrality, and clustering coefficients can be derived from the correlation matrix.

2.4. Deconvolution

The observed GCaMP fluorescence trace y is modeled as a convolution of the underlying spike train s with the calcium kernel c , followed by the addition of noise:

$$y(t) = c(t) * s(t) + \epsilon(t) \quad (6)$$

$s(t)$: the unobserved spike train (discrete events corresponding to neuronal spikes).

$c(t)$: the calcium kernel, which describes the response of the calcium indicator to a spike, usually characterized by a rapid rise and slow decay.

$\epsilon(t)$: added observational noise.

$*$: denotes the convolution operation.

The calcium indicator GCaMP has a temporal resolution in the range of hundreds of milliseconds, which is significantly slower than the millisecond-scale duration of individual spikes. As a result, the observed fluorescence signal lacks the temporal precision to directly resolve the timing of neuronal spikes. To overcome this limitation, deconvolution methods are applied to estimate the underlying spike train $s(t)$ from the convolved fluorescence signal $y(t)$. Before calling the deconvolution function, it's important to hypothesize all the raw data is from filtered data. To recover the underlying spike train s , the OASIS (Online Active Set method to Infer Spikes) deconvolution algorithm is applied. OASIS first estimates the GCaMP kernel, which represents how calcium levels rise and fall following a spike. The kernel is essentially the shape of the calcium response to a single action potential. The deconvolution process involves reversing the convolution of the spike train and the calcium dynamics. OASIS attempts to "undo" the smoothing effect of the calcium transients by estimating when spikes occurred based on the observed calcium signal. The algorithm identifies the most likely times when spikes occurred by finding points in the observed fluorescence where the deconvolved signal suggests that a spike must have happened. The average firing rate is calculated by dividing the total number of spikes by the total recording time, which is determined by dividing the number of time points M by the frame rate FR (frames per second). Finally, it creates a figure with three subplots: The raw filtered calcium signal for that neuron, the denoised calcium signal (c) after OASIS deconvolution, the inferred spike train (s), represented as vertical lines (spikes) over time. This provides a visual comparison between the raw data, the denoised signal, and the inferred spikes.

3. Result

3.1. High-Pass Filtering

3.1.1. Determining the Cutoff Frequency

After applying the high-pass filter, significant noise reduction was observed in the calcium imaging data. The Fourier magnitude spectrum revealed that low-frequency components were effectively attenuated, preserving higher-frequency components associated with neural activity. Specifically, the cutoff frequency was determined to be 0.001 Hz based on the Fourier spectrum in figure 1.

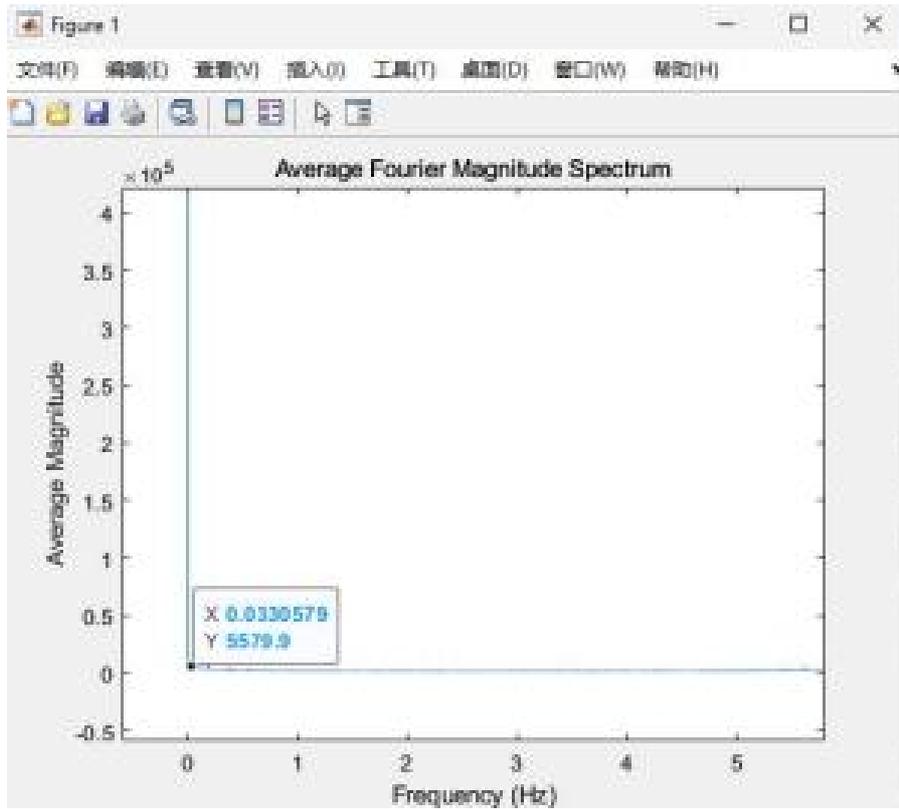


Figure 1: Average Fourier Magnitude Spectrum of the signal, displaying frequency (Hz) on the x-axis and the average magnitude on the y-axis, scaled by 10⁵. A significant peak is observed at very low frequencies near 0 Hz, with a magnitude of approximately 55,779.9, as highlighted in the plot. The remainder of the spectrum shows relatively low magnitudes across higher frequencies, indicating that the signal contains most of its power at near-zero frequency components.

3.1.2. High-Pass Filter Coefficients

The coefficients of the high-pass filter were calculated using the inverse discrete-time Fourier transform (IDTFT) of the ideal frequency response. Figure 2 presents the filter coefficients for FIR filter.

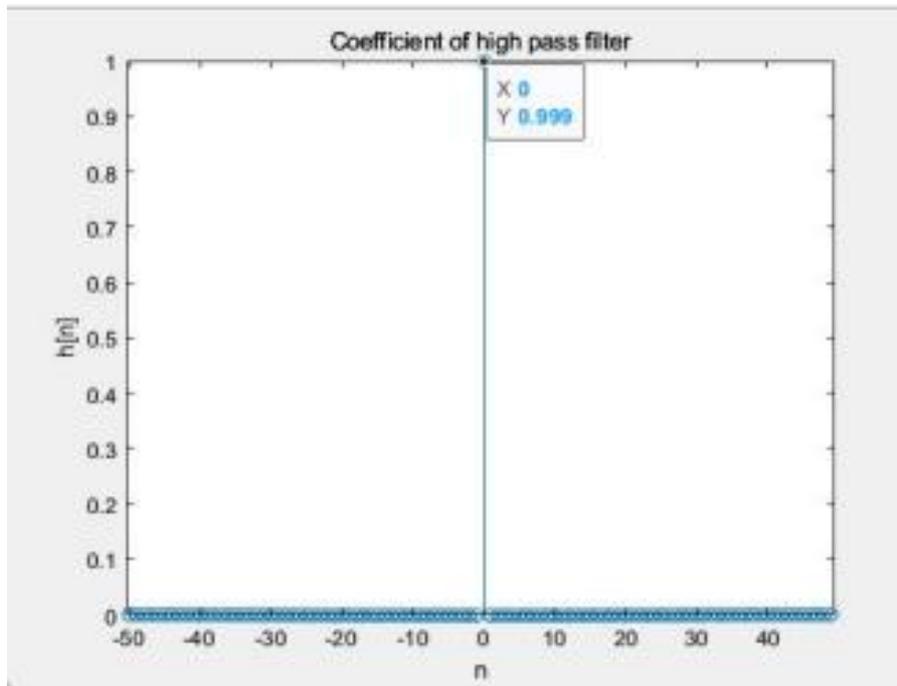


Figure 2: Coefficient distribution of a high-pass filter. The plot shows the filter coefficients $h[n]$, with the x-axis representing the index n and the y-axis representing the coefficient magnitude. A significant peak occurs at $n=0$ with a coefficient value of approximately 0.999, while all other coefficients remain close to zero. This illustrates the filter’s sharp response, primarily allowing high-frequency components to pass through while attenuating lower frequencies.

3.1.3. Filter Implementation

We implemented several types of high-pass filters to compare their performance. The signal-to-noise ratio (SNR) was improved after applying the filters, with the FIR filter achieving an SNR of 26.11, while the Butterworth and Chebyshev filters yielded SNRs of -5.96, -7.89 respectively (Table 1). Among these, the FIR filter provided the best noise attenuation with minimal signal distortion.

Table 1: Comparison of Signal-to-Noise Ratio (SNR) for Different Filter Types

Filter Type	SNR (dB)
FIR Filter	26.11
Butterworth	-5.96
Chebyshev	-7.89

Figure 3 shows a comparison of the calcium traces for a representative neuron before and after filtering (FIR filter).

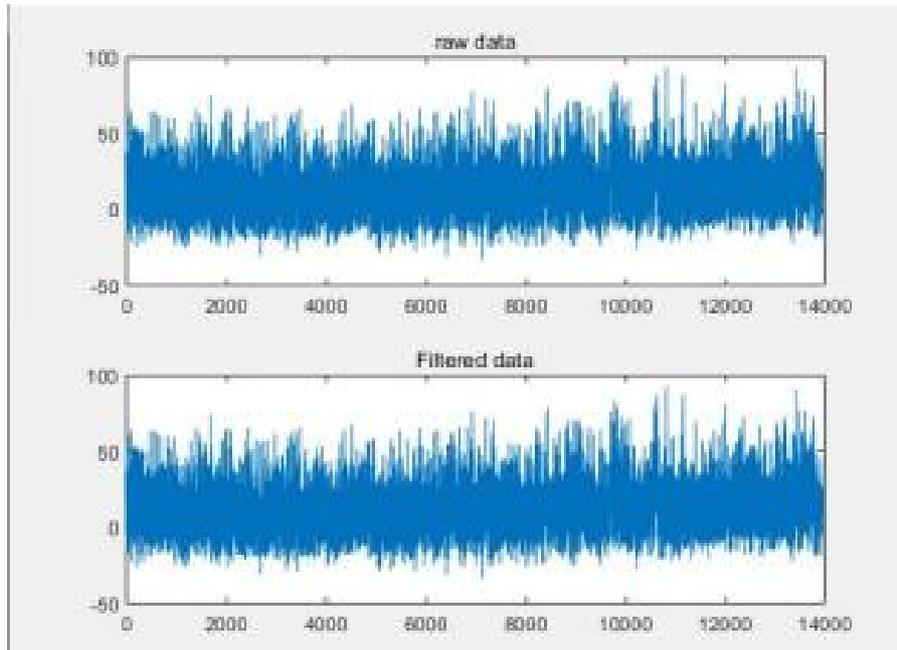


Figure 3: Comparison of raw and filtered data. The top plot illustrates the raw data signal, which exhibits considerable noise and fluctuations throughout the entire duration. The bottom plot shows the same signal after filtering, with noticeable reduction in noise and smoother overall behavior, while preserving key features of the original signal. This demonstrates the effectiveness of the filtering process in removing unwanted noise from the data.

3.2. Gaussian Mixture Model (GMM)

Given the large dataset consisting of 233 neurons, it is impractical to present the results for each neuron individually. Therefore, Neuron 29 was selected as a representative example to demonstrate the application of the Gaussian Mixture Model (GMM) on the calcium imaging data.

After applying the high-pass filter, the GMM with two components was fitted to the calcium imaging data from Neuron 29 (Figure 4) . The model aimed to differentiate between baseline and excited neural states based on the calcium signal dynamics.

For Neuron 29, the GMM fitting results are summarized in Table 2. The model indicated that 74.52% of the data belonged to the excited state and 25.48% to the baseline state.

The baseline state for Neuron 29 had a mean of 47.9129 and a standard deviation of 30.9359, whereas the excited state exhibited a mean of 9.6364 and a standard deviation of 11.2521. The mixing weights, which represent the proportion of data points associated with each state, were 0.2548 for the baseline state and 0.7452 for the excited state.

Table 2: Statistical Characteristics of Baseline and Excited States

Component	Mean (μ)	Standard Deviation (σ)	Weight (α)
Baseline State	47.9129	(σ)	0.2548
Excited State	9.6364	11.2521	0.7452

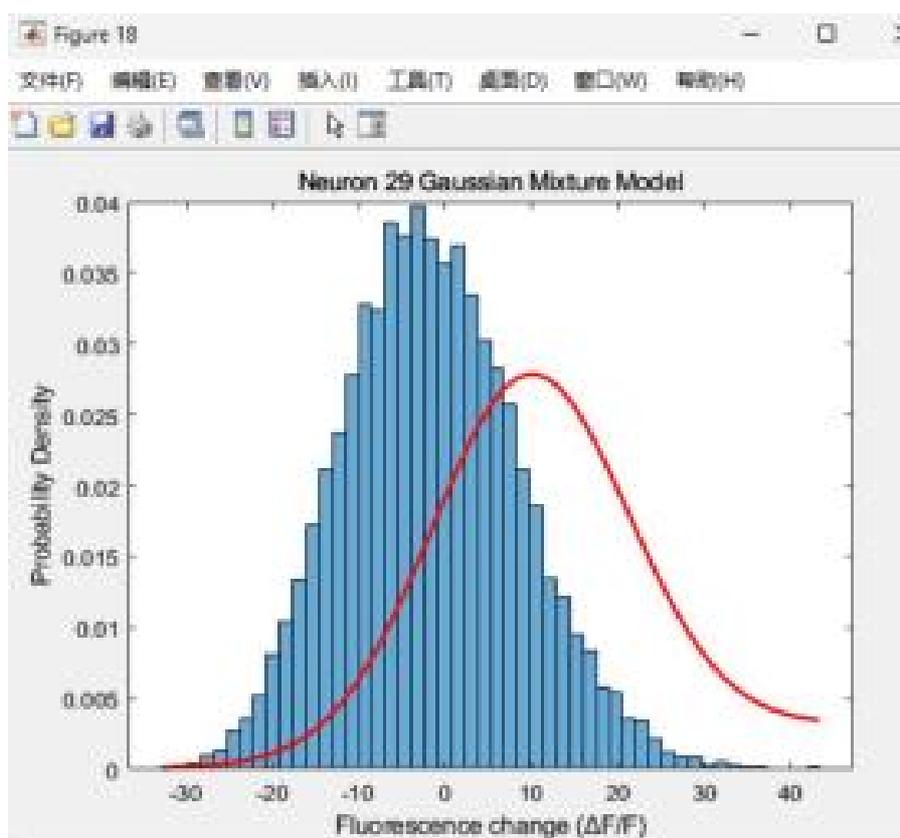


Figure 4: Probability density distribution of fluorescence change ($\Delta F/F$) for Neuron 29, modeled using a Gaussian Mixture Model. The histogram represents the distribution of fluorescence changes, while the red curve shows the fitted Gaussian mixture model. The model highlights the underlying statistical structure of the data, capturing both the central tendency and the spread of fluorescence changes around the mean, with peaks near 0.

3.3. Correlation Coefficient Analysis

The Pearson correlation coefficient heat map (Figure 5) provided insights into the functional connectivity between neurons. Neurons with high correlation coefficients (>0.7) formed clusters, indicating potential functional networks. These networks suggest synchronized activity across multiple neurons, which may reflect cooperative behavior in response to visual stimuli.

The matrix revealed several strong inter-neuron correlations, suggesting a networked response to the external environment. This clustering was particularly prominent in neurons located in close proximity within the visual cortex, indicating local synchrony.

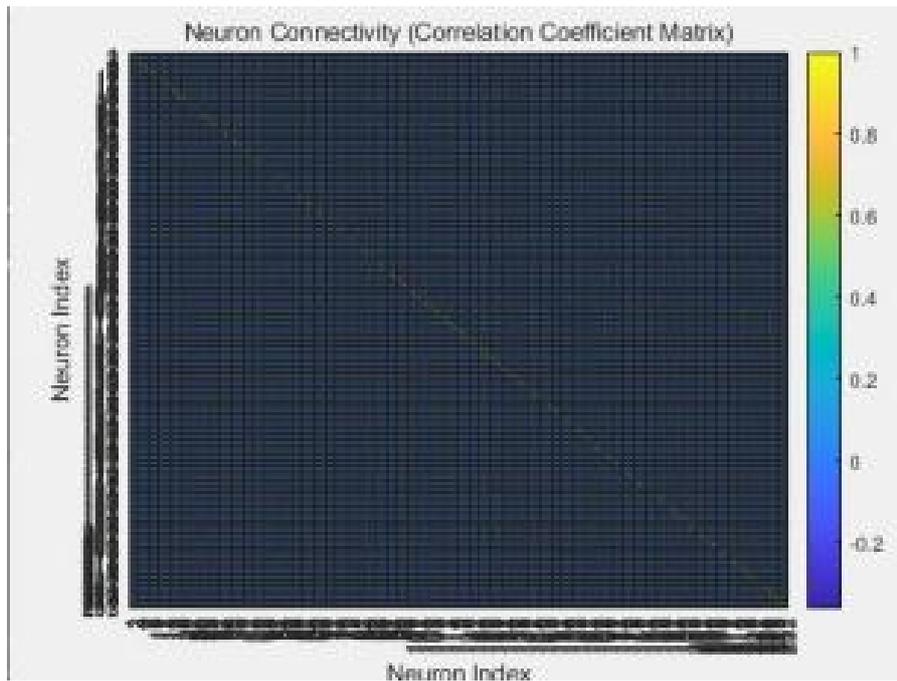


Figure 5: Neuron connectivity matrix represented by a correlation coefficient matrix. Each entry in the matrix corresponds to the correlation coefficient between the activity of two neurons, with neuron indices along both the x- and y-axes. The color bar on the right indicates the strength of the correlation, where values range from -1 (strong negative correlation) to +1 (strong positive correlation). High positive correlations are represented by warmer colors (yellow), while negative correlations are represented by cooler colors (blue).

Due to the large size of the 233x233 matrix, we have presented a 5x5 submatrix instead as shown in Figure 5(b).

1.0000	0.0486	-0.0426	0.0472	0.0226
0.0486	1.0000	-0.0301	0.0248	-0.0454
-0.0426	-0.0301	1.0000	-0.0151	0.0008
0.0472	0.0248	-0.0151	1.0000	-0.0041
0.0226	-0.0454	0.0008	-0.0041	1.0000

Figure 5(b): 5x5 submatrix

3.4. Deconvolution of Calcium Imaging Data

Deconvolution was performed to infer neuronal spike events from the calcium imaging data. The OASIS deconvolution algorithm effectively recovered the underlying spike trains from the filtered fluorescence traces. The resulting deconvolved data allowed for precise identification of spike events, with a temporal resolution significantly improved compared to the raw calcium signals.

The inferred spikes for a selected neuron (29) are shown in Figure 6, where each vertical line corresponds to an action potential. The average firing rate was calculated as 7.4172 spikes per second, demonstrating the increased temporal precision achieved through deconvolution.

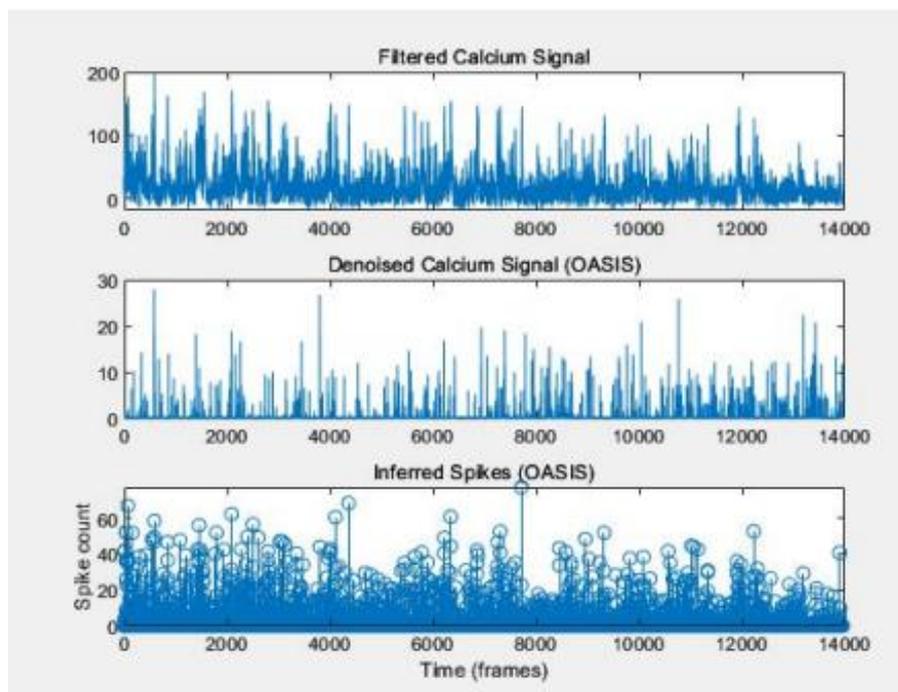


Figure 6: Analysis of calcium imaging data across three stages. The top panel shows the Filtered Calcium Signal, representing the raw calcium signal after applying a filtering process to reduce noise. The middle panel displays the Denoised Calcium Signal processed using the OASIS algorithm, highlighting significant fluctuations corresponding to neural activity while minimizing noise. The bottom panel shows Inferred Spikes derived from the denoised signal using the OASIS spike inference method, with circles representing detected spikes across time frames.

4. Discussion

The results from the experiment indicate that the four methods of analyzing the calcium imaging are effective. They further overcome the challenge from temporal resolution, low noise to signal ratio, network connectivity of neurons, and variability in signal. For the first method, several high pass filter were tested and found out the one that best filter the noise is the FIR high pass filter. In fact, the FIR filter is the only appropriate one shown in the Table 1 as all the other ones showed negative SNR, which is abnormal in this case. This is because the Butterworth high pass filter is used in audio processing and communication systems. The Chebyshev high pass filter is used in radar systems. (Chebyshev Filter - an Overview | ScienceDirect Topics, n.d.)(Ellis, 2012) Secondly, both the Gaussian mixed model and the correlation matrix showed the behavior of each neuron clearly. Moreover, the deconvolution algorithm predicted the potential spikes accurately. Our paper not only provided new methods for other researchers to use but also support their experimental conclusion and hypothesis. For instance, according to the paper of "Calcium imaging analysis - how far have we come?"(Robbins et al., 2021) researchers tried to use several methods including the denoising method to improve the SNR. They used local filters such as anisotropic filter and Gaussian smoothing filter, which are com-

putationally light but have clear limitations. Using the high pass filters might help them solve this challenge can contribute to more perfect results. Of course, there are limitations in our experiment. First, all the data we chose is from the same source, which could be not convincing to prove that the result is conclusive. It becomes difficult to validate whether the observed patterns are robust across different subjects or whether they hold under varying experimental conditions. This limits the overall scope of the experiment and makes it challenging to draw conclusions beyond the specific dataset. Besides, the single dataset might not reflect the diversity of neural responses present in a population of cells or across multiple trials. Repeated use of the same dataset increases the risk of overfitting the analysis to one specific set of conditions, thus reducing the experiment's ecological validity. As a consequence, further studies should focus on whether the four methods would indicate the similar result as the neurons are at different regions of mice and with mice that performing different tasks. The following experiment is a sample from (Precision Calcium Imaging of Dense Neural Populations via a Cell-Body-Targeted Calcium Indicator, [www.https://doi.org/10.1016/j.neuron.2020.05.029](https://doi.org/10.1016/j.neuron.2020.05.029))

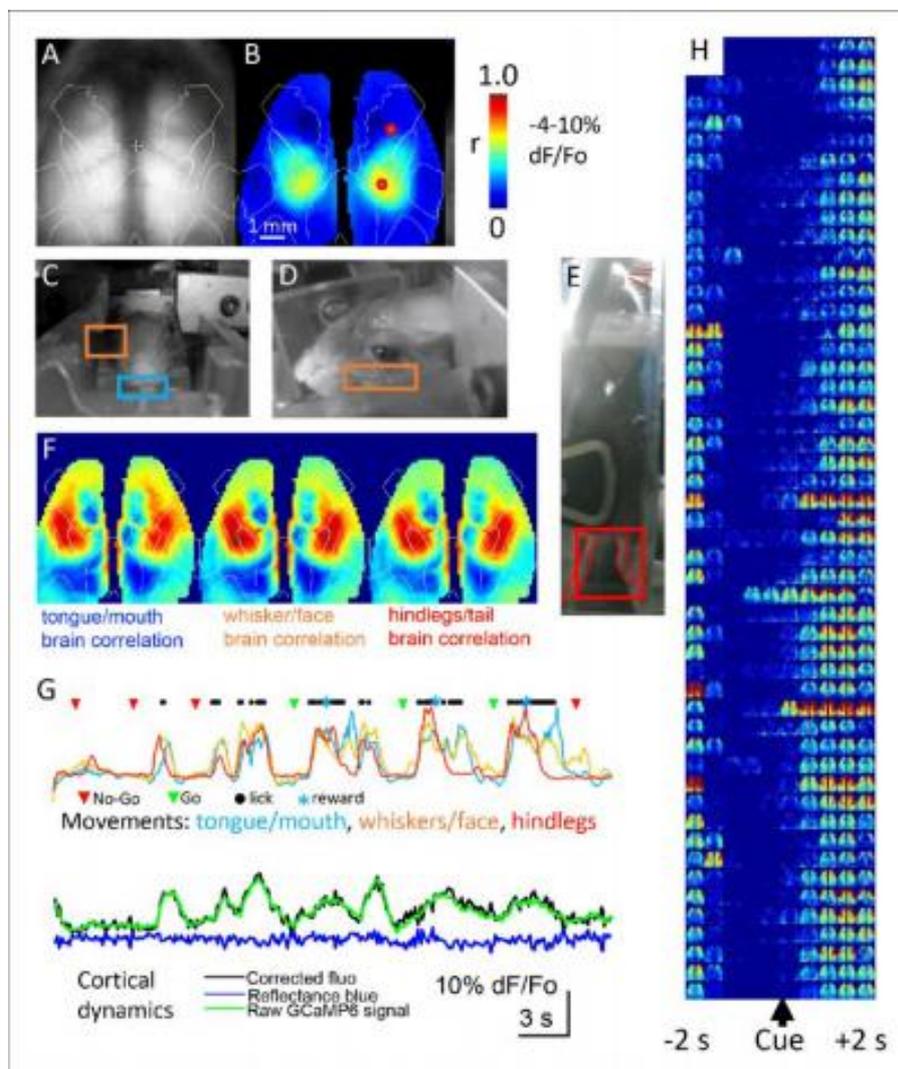


Figure 7: Imaging data showing cortical dynamics and behavioral correlations during a task.

Panels A and B: (A) Anatomical reference image and (B) correlation map showing areas of the cortex activated in relation to specific movements. Colors represent the correlation coefficient (r), with red indicating higher correlation between calcium signals and movement, measured as a percentage change in fluorescence ($\Delta F/F_0$).

Panels C-E: Example frames from the experiment depicting different behaviors, including tongue/mouth movements, whisker/face movements, and hindleg movements (highlighted in colored boxes: orange and blue). Panel E shows a zoomed-in view of the hind leg movement.

Panel F: Heatmaps showing the brain regions correlated with specific movements, with color-coded labels for tongue/mouth (blue), whisker/face (orange), and hindlegs/tail (red). The heatmaps highlight cortical activity associated with each type of movement.

Panel G: Temporal dynamics of movements (top) aligned with cortical activity (bottom). The movement traces show specific task events (Go, No-Go, Lick, and Reward) aligned with the activity of different body parts (tongue/mouth, whiskers/face, hind legs), and cortical fluorescence dynamics are represented using GCaMP6 signals.

Panel H: A summary heatmap representing population neural activity relative to task cues. Time is shown from -2 seconds before the cue to +2 seconds after the cue, with color intensity indicating activity levels in various regions of the cortex.

Supplementary materials

PPT Source:

..\..\..\..\..\Desktop\2025application\Berkeley Neuroscience\Final Presentation(2).pptx

References

- [1] Cantu, D. A., Wang, B., Gongwer, M. W., He, C. X., Goel, A., Suresh, A., Kourdougli, N., Arroyo, E. D., Zeiger, W., & Portera-Cailliau, C. (2020). EZcalcium: Open-Source Toolbox for Analysis of Calcium Imaging Data. *Frontiers in Neural Circuits*, 14. <https://doi.org/10.3389/fncir.2020.00025>
- [2] “Chebyshev Filter - an Overview | ScienceDirect Topics.” www.sciencedirect.com, www.sciencedirect.com/topics/engineering/chebyshev-filter. Robbins, M., Christensen, C. N., Kaminski, C. F., & Zlatic, M. (2021). Calcium imaging analysis – how far have we come? *F1000Research*, 10, 258. <https://doi.org/10.12688/f1000research.51755.2>
- [3] Ellis, George . “Butterworth Filter - an Overview | ScienceDirect Topics.” www.sciencedirect.com, 2012, www.sciencedirect.com/topics/computer-science/butterworth-filter.
- [4] Friedrich, Johannes, et al. “Fast Online Deconvolution of Calcium Imaging Data.” *PLOS Computational Biology*, vol. 13, no. 3, 14 Mar. 2017, pp. e1005423–e1005423, <https://doi.org/10.1371/journal.pcbi.1005423>. Accessed 30 Apr. 2023.
- [5] Hanning Window - an overview | ScienceDirect Topics. (n.d.). www.sciencedirect.com. <https://www.sciencedirect.com/topics/engineering/hanning-window>
- [6] Hann or Hanning or Raised Cosine. (n.d.). ccrma.stanford.edu. https://ccrma.stanford.edu/~jos/sasp/Hann_Hanning_Raised_Cosine.html
- [7] j-friedrich. “GitHub - J-Friedrich/OASIS: Deconvolution of Calcium Imaging Data.” GitHub, 28 Feb. 2023, github.com/j-friedrich/OASIS. <https://doi.org/10.1371/journal.pcbi.1005423.g001>. Accessed 5 Oct. 2024. `\Project3_data\MACOSX\Project3_data\OASIS_matlab-master\packages\oasis’);`

- [8] Qin, Hao. “Computational Analysis of GCaMP Fluorescence Data in Neuronal Activity.” *Theoretical and Natural Science*, vol. 46, no. 1, 31 July 2024, pp. 163–172, www.researchgate.net/publication/382759620_Computational_analysis_of_GCaMP_fluorescence_data_in_neuronal_activity, <https://doi.org/10.54254/2753-8818/46/20240628>. Accessed 7 Oct. 2024.
- [9] Rochefort, N. L., Jia, H., & Konnerth, A. (2008). Calcium imaging in the living brain: prospects for molecular medicine. *Trends in Molecular Medicine*, 14(9), 389–399. <https://doi.org/10.1016/j.molmed.2008.07.005>
- [10] Shemesh, Or A., et al. “Precision Calcium Imaging of Dense Neural Populations via a Cell-Body-Targeted Calcium Indicator.” *Neuron*, vol. 107, no. 3, Aug. 2020, pp. 470-486.e11, <https://doi.org/10.1016/j.neuron.2020.05.029>.