

Contents

Short Keys	1
Basic Event Extraction Inputs.....	1
Acquisition and Extraction	1
Essential Levels	2
Data Boundaries.....	5
My Custom Function.....	6
Test Files.....	8

Short Keys

Left Arrow: Previous Event

Right Arrow: Next Event

Enter: View Event

f: Add to false log

r: Remove to false log

c: Clear False Events

s: Save Events

Basic Event Extraction Inputs

Acquisition and Extraction

Analysis Type: Multi Level (good for polymer chains) or Single Level (good for particles and proteins)

Frequency (kHz): Frequency at which data was acquired.

Start time: Analysis start time in seconds.

End time: Analysis end time in seconds.

Window Size (s): Data will be analyzed in segments of length specified by this input.

PDC Type: Determines whether event flagging would be based on $\sigma_{baseline}$ where $\sigma_{baseline}$ is the standard deviation of the baseline in the analysis window or ΔI_{PDC} where ΔI_{PDC} is the minimum current drop threshold for event flagging.

PDC: Events are deemed as perturbations that are at least $PDC \times \sigma_{baseline}$ or ΔI_{PDC} (i.e., let's call the minimum current deviation defined by either of these ways as ΔI_{min}).

Voltage: operational voltage (in mV. Only used to plot data analytics).

Lowpass (kHz): Lowpass filter setting.

BNB Selection: Sometimes in the preliminary event extractions, sudden/subtle changes in the baseline can also get detected as events. When such structures are fitted using the multi-level fitting option, the fit levels would be perilously close to the baseline since these are not true event structures. Thus, after the fitting, if the maximum change in current of the fit levels is less than that defined by the *PDC* option, the events are rejected.

Analysis Depth: Depth 1 is discussed in the manuscript whereas, in Depth 2, the poorly fitted events are further attempted to be fitted by increasing the cluster number.

Essential Levels

Max Levels: Maximum permitted levels in an event.

Inter Level Threshold: Used to qualify whether two adjacent levels should remain separate or not (i.e., merged). This is done either based on current ($\Delta I_{i,j}$) or standard deviation of the baseline ($n_{ILT} \cdot \sigma_{baseline}$ where n_{ILT} is the inter-level threshold) as defined by the two-way “*Threshold*” switch. If the two adjacent levels are within the threshold value defined by this option, they will be merged.

Edge PDC: A current threshold is defined as $I_{EPDC} = I_0 - n_{EPDC} \cdot \sigma_{baseline}$ where I_0 and n_{EPDC} are baseline current and Edge PDC respectively. This is especially useful with noisy baselines where false levels may appear near the baseline. Any level that is above I_{EPDC} will be rejected. This is set to 3 by default and if the baseline is very smooth, one can set this to as low as 1. If unsure about how to use this, you can set n_{EPDC} to 0.

Level PDC Fraction: Used to determine the goodness of the fit. A current drop threshold is defined as $I_{LPDC} = I_0 - n_{LPDC} \cdot \Delta I_{min}$ where n_{LPDC} is Level PDC Fraction (see PDC in Acquisition and Extraction for the definition of ΔI_{min}). The peaks and valleys of the event are found using the *findpeaks* function and if there are peaks above I_{LPDC} , the event would be deemed poor. If unsure about how to use this, you can set n_{LPDC} to 0.

Smoothing Factor: Sometimes slight fluctuations in falling/rising edges of events could be identified as levels. This step essentially smoothens levels (using Gaussian smoothing) to minimize such false identifications.

Level Goodness Threshold: Level goodness is calculated in the following manner (different from *EventPro* 1.0 which is described in 2021 ACS Analytical Chemistry paper). This is a cascading structure

Poor Event Classification

- i) Checking baseline positions: The events is deemed poor if
 - a) $|I_{base,L} - I_{base,R}| > 3 \times \sigma_{baseline}$ where $I_{base,L}$ and $I_{base,R}$ the mean value of baseline on left and right of the event respectively

- b) $|I_{fit,max} - I_{fit,min}| < \Delta I_{min}$ where $I_{fit,max}$ and $I_{fit,min}$ are the maximum and minimum currents of the fit made to the padded event (that includes left and right baselines as well!).
- c) Data between 10%-90% of each level is first selected. Then turning points (i.e., peaks and valleys) are found using the *findpeaks* function ($\{\{i_{tp}\}_{10\%}^{90\%}\}$). If $\{\{i_{tp}\}_{10\%}^{90\%}\} > I_{LPDC}$ is satisfied, event is deemed poor.

Afterward, if the event is *not poor*, the following metrics are computed.

- i) If there is only one level, the standard deviation of $(\{\{i_{tp}\}_{10\%}^{90\%}\} - \Delta I_{level})$.and $(\{\{i_{tp}\}_{10\%}^{90\%}\} - (\Delta I_{level} + 3 \times \sigma_{baseline}))$ are evaluated where ΔI_{level} is the value of the current of corresponding level from the fit. The minimum value of the quantities is assigned as $\Delta I_{\sigma,i}$
- ii) If there's more than one level, $\Delta I_{\sigma,i}$ of the i^{th} level is computed as the standard deviation of $(\{\{i_{tp}\}_{10\%}^{90\%}\} - \Delta I_{level})$.

The level goodness is then computed as,

$$L_{goodness} = 1 - \overline{(\Delta I_{\sigma}/3 \times \sigma_{baseline})}$$

Where $\overline{(\Delta I_{\sigma}/3 \times \sigma_{baseline})}$ is $\sum_{i=1}^n \Delta I_{\sigma,i}/3 \times \sigma_{baseline}$

Marginally Poor Event Classification: Only applicable to the single level analysis. When $L_{goodness} < \text{Level Goodness Threshold}$, it is classified as a marginally poor event.

Over Clusters: Typically defined as 2. However, if one needs to increase the level of segmentation sensitivity, one can do it by increasing this number.

Data Boundaries

Translocation Time UB (sec): Upper translocation time-bound in seconds. Any event with a translocation time higher than this value would be rejected.

Translocation Time LB (micro sec): Lower translocation time-bound in microseconds. Any event with a translocation time lower than this value would be rejected.

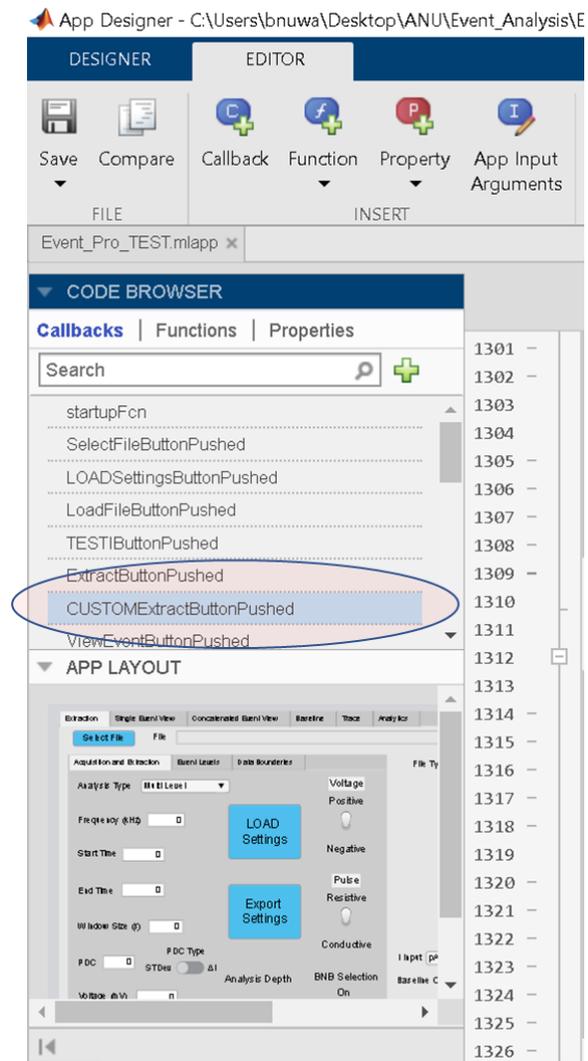
Event Pad: The number of baseline points that are added to either side of the event.

Population Factor: Only applicable to “SMD and Msbackadj” baseline option. After fitting the baseline using the Gaussian Smoothed I_0 method, the upper boundary of the baseline is determined by incrementally shifting the baseline fit n -times baseline until it harbors at least the fraction of total data points defined by the Population Factor (typically set to 0.995).

Back Step Threshold: After detecting an event, the two sides of the blockade are determined by working back until the current is $> n_{BST} \cdot \mu_{I_0}$ where n_{BST} is the back step threshold (typically set to 0.999) and μ_{I_0} is the mean baseline value.

Maximum Iterations: Maximum number of iterations permitted for level refining based on ΔI_{user} , or $n \cdot \sigma_{baseline}$ n

My Custom Function



Following Input Fields are not used in the as-supplied version (you can modify the section to make use of these if needed)

- i) Analysis Type (set to single level)
- ii) Analysis Depth (since this is only applicable to multi-level event fitting)
- iii) Threshold Switch
- iv) Inter Level Threshold
- v) Over Clusters
- vi) BNB Selection
- vii) Edge PDC
- viii) Smoothing Factor
- ix) Maximum Iterations

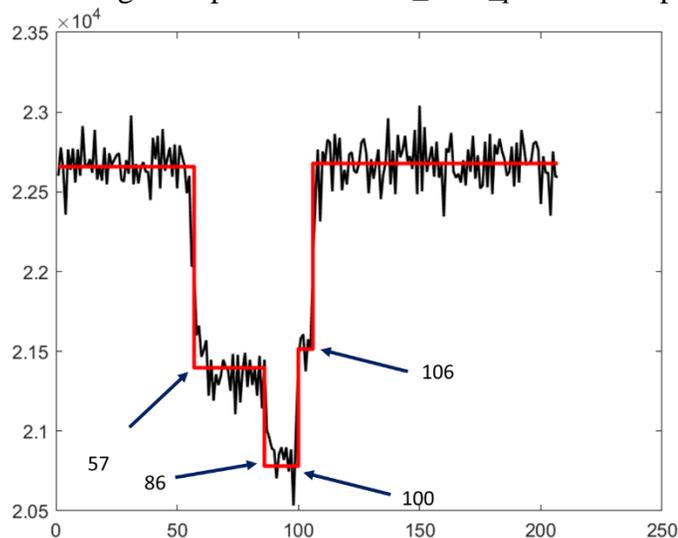
To edit the output behavior, locate the "My_CUSTOM_FUNCTION" in the function folder:

Name	Date modified	Type	Size
abfload	8/1/2018 10:14 AM	MATLAB Code	54 KB
Baseline_Fitting_Light	3/31/2021 2:52 PM	MATLAB Code	4 KB
Baseline_Fitting_UltraLight	3/31/2021 2:53 PM	MATLAB Code	4 KB
End_point_Adjustment	7/28/2021 11:08 AM	MATLAB Code	2 KB
Event_Pro_TEST	7/31/2021 1:29 PM	MATLAB App	526 KB
Event_STD_Construction	3/18/2021 5:57 PM	MATLAB Code	1 KB
Final_Level_Construction_Light	3/30/2021 4:13 AM	MATLAB Code	3 KB
FWHM	7/10/2021 12:47 AM	MATLAB Code	3 KB
level_crossing_count_Edges_MOD	7/10/2021 12:16 PM	MATLAB Code	1 KB
level_fit_goodness_MOD	7/29/2021 2:04 PM	MATLAB Code	4 KB
Level_Mean_Universal	3/15/2021 7:53 PM	MATLAB Code	1 KB
Level_Reconstruction_Light	7/28/2021 11:09 AM	MATLAB Code	9 KB
Level_Refining_Light	7/9/2021 9:27 PM	MATLAB Code	2 KB
Level_Splitting_Light	7/25/2021 2:45 PM	MATLAB Code	3 KB
Level_STD_Comp	3/16/2021 2:55 AM	MATLAB Code	1 KB
Logo	7/29/2021 2:16 PM	PNG File	1,366 KB
Lower_Level_Tuning_Light	7/10/2021 12:07 PM	MATLAB Code	2 KB
My_CUSTOM_FUNCTION	7/12/2021 6:15 PM	MATLAB Code	4 KB
Nanopore	2/24/2021 1:33 AM	JPG File	53 KB
NewIest.abf	3/3/2021 8:46 PM	ABF File	10,053 KB
Single_Level_Truncated_Event_Construction	7/28/2021 10:34 AM	MATLAB Code	3 KB

How to Modify the *My_CUSTOM_FUNCTION*

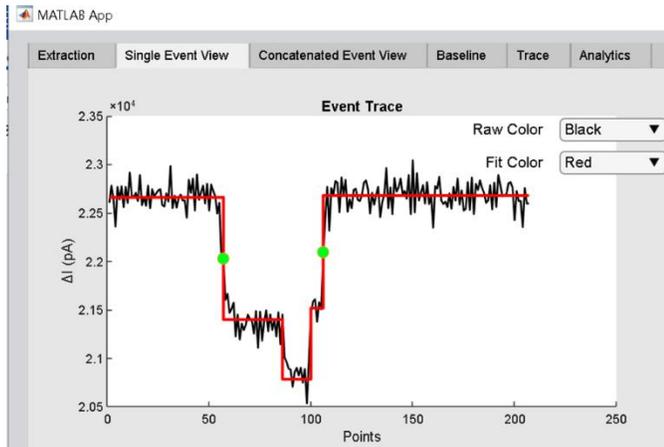
Following outputs are needed from the function for the app to function properly

- i) save_status: Whether you want to save the event (1) or not (0)
- ii) Final_IT_array: $N \times 2$ matrix containing the fit values. Column 1 (time) and Column 2 (current)
- iii) Final_dt_array: Translocation Time of each fit level in points (not in microseconds or any other units. This will be converted to time units later in the app). Save as a column vector
- iv) Final_mean_I_array: Current depth (i.e., ΔI_{level}) of each level in pico amps. Save as a column vector
- v) Final_level_pts: Boundary points of each level ($N \times 1$ column vector). Let's take the following example where *Final_level_pts* were output as [57;86;100;106]

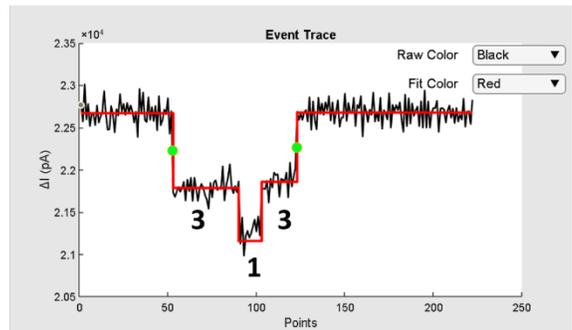
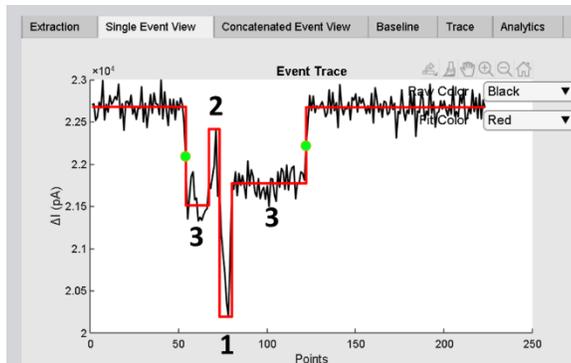


- vi) event_area: Area of the event in fC
- vii) end_to_end_trans_time: Two side of the blockade through the fit
- viii) trans_time_FWHM: Full Width Half Max Translocation time in micro seconds.

- ix) FWHM_bounds: Bounds of the FWHM time for viewing (as data points but not in time units). These are green dots that appear in the “Single Event View” tab as shown below.



- x) levels: Number of event levels
- xi) level_type: Type of each level: downward (1), upward (2) or flat (3). See examples below



- xii) I0_new: Baseline Current Value

Test Files

We have supplied two current traces in .abf (NewTest) and .mat (Test_Data) form that you can use to help develop your code.