OpenSorter User's Guide

OpenSorter User's Guide

Copyright

Licenses and Trademarks

Updated: 11/20/2012 4:51 PM

Table of Contents

Before You Begin

Requirements

TDT Drivers and the OpenEx Suite must be installed before installing OpenSorter.

The recommended operating system for all TDT systems is Windows 7®.

Organization of the Manual

This manual is organized in the following sections:

- ➢ Overview
  Briefly describes the OpenSorter application and how it works with the OpenEx Software Suite

- ➢ OpenSorter Basics
  Describes the visual interface and how to select and view data.

- ➢ Sorting
  Details the available sorting algorithms, manual sort techniques, eliminating outliers, and viewing statistics.

- ➢ Processing Multiple Data Sets
  Details techniques for batch processing and concatenating data.

- ➢ Menu and Toolbar Reference
  A quick reference for the toolbars and menus.

- ➢ Glossary and Index

Overview

OpenSorter is stand-alone neurophysiological spike sorting software for TDT's OpenEx software suite. OpenSorter is designed to sort spike event data saved in the OpenEx DataTank format. Sorting is based on principal components feature space, waveform parameter feature space, or waveform shape. Semi-automated, K-Means, Closest Centers, and manual sorting methods are supported. Sorting results can be manually edited and spikes re-assigned using editors in the feature space, timeline, or waveform space displays. Sorting analyses, such as Pseudo-F statistics, isolation distances, L-ratios and Silhouette indices, display sort quality and help to guide any manual resorting and editing that may be required.

OpenSorter features:

- ➢ Fully manual to fully automated spike sorting modes
- ➢ Sorting and editing in both the feature space and the waveform space
- ➢ Supervised or unsupervised batch mode
- ➢ Data set level and cluster level outlier elimination
- ➢ Dynamic sort editing
- ➢ Selectable sorting dimensions
- ➢ Tank Filtering
- ➢ Statistical analysis
- ➢ Stepwise undo/redo for interactive data exploration, editing, and analysis

Sorting operations are carried out on the active data set displayed in the main display and editing window. Data sets are defined at the channel level and are selected in the Tanks panel using an expandable tree structure similar to that used in many Windows applications.

Two general sorting methods are available: 1) graphical or *manual* selection and assignment of spikes using OpenSorter's mouse-based selection tools and 2) algorithm-based sorting methods that automatically assign sort codes to each spike based on user-defined algorithm parameter settings. Sort codes can then be saved into a *SortID*.

SortID

A *SortID* is a new level of the data tank format that contains the sort codes assigned to each spike in a data set. Since OpenEx offers real-time spike sorting capabilities, spikes read into OpenSorter may already be labeled with pre-existing sort codes. In the event that real-time sorting was not used during acquisition, each spike will still be labeled with a sort code (0) and will appear as unsorted in the OpenSorter display and editing window. For all datasets, sort codes assigned during acquisition (online) appear as a read-only SortID labeled TankSort. The TankSort SortID and thus the sort codes stored during acquisition cannot be overwritten. This SortID is marked with a padlock icon in the Tank Browser window. Sort codes assigned offline using OpenSorter are instead saved to new SortIDs with names assigned by the end-user. Data sets can be sorted or edited any number of times, with the resulting sort codes stored to multiple SortIDs for each dataset.

The new SortIDs can be used with OpenEx Client applications such as OpenBrowser and OpenExplorer or via user-developed programs using OpenDeveloper calls. Events with associated SortIDs appear in the Events lists of these programs as an event followed by the name of their SortID (for example, SNIP_MySort). When you select these events, the named event (in this case SNIP), is selected along with the sort codes associated with the specified SortID (*MySort*).

OpenSorter Basics

This section introduces the various displays, tools, and sub-windows that make up the workspace.

Topics include:

- ➢ Exploring The Workspace
- ➢ Selecting a DataSet with the DataTank Browser
- ➢ Using the Settings Panel
- ➢ Viewing Data in the Tabbed Window
- ➢ Tools for Viewing Data
- ➢ Navigating the TimeLine
- ➢ Viewing Large Datasets with Display Steps
- ➢ Exploring Data in the Feature Space

The Workspace

OpenSorter provides a versatile workspace where users can explore data sets and edit sort codes in a graphical environment. The customizable workspace includes collapsible panels that can be docked or floated and provides auto hide push pin tools to provide the most flexible workspace possible.
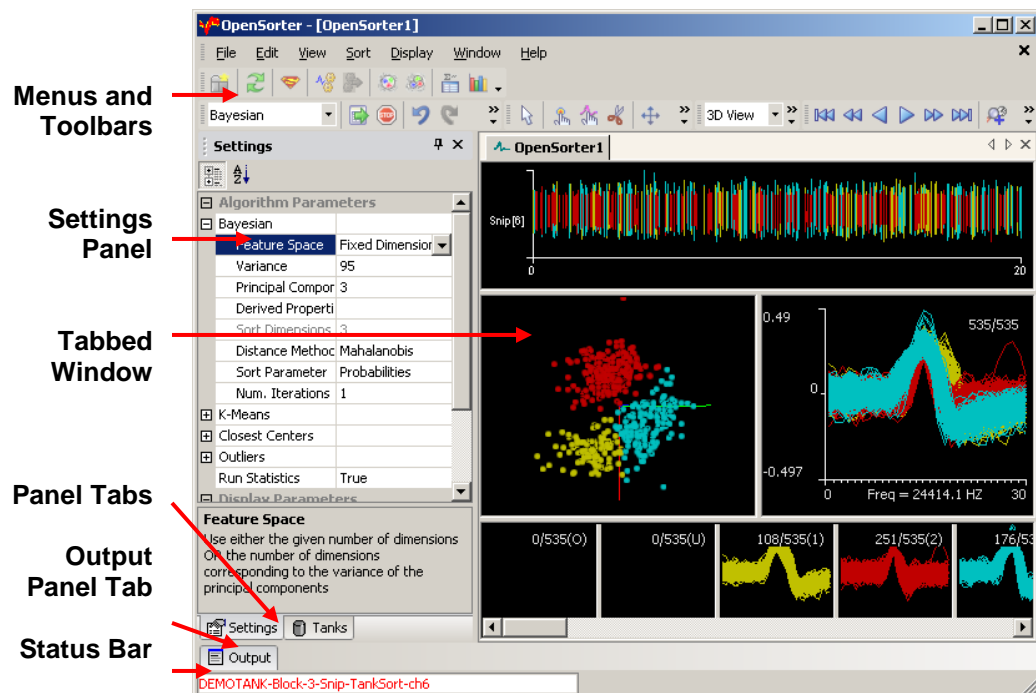
Tabbed Windows

The main workspace feature is a multi-pane tabbed window for displaying and editing data. The window includes waveform space and feature space panes as well as a time line pane and individual unit waveform display. Multiple tabbed windows can be opened for viewing multiple data sets. See *Viewing Data in the Tabbed Window,* page 11, for more information.

Menus and Toolbars

A comprehensive set of menus and toolbars provides easy access to commands and tools. Frequently-used commands are available via toolbar buttons. Move the mouse pointer over a toolbar button to display a *tool tip* that describes the function of that button. You can move the toolbars and dock them to the top or bottom of the screen or rearrange and customize the contents of the toolbars using the View|Toolbars sub-menu. See the *Workspace Reference*, page 47, for a complete list of commands and tools. Context sensitive menus are available by right-clicking the various panes of the Tabbed Window.

Tanks Panel

A collapsible panel provides a convenient method of selecting data sets. Recently used DataTanks are displayed for quick selection.

Tank Filtering Panel

A collapsible panel provides access to applying tank filter settings for event data.

Settings Panel

A collapsible panel provides access to property settings for each algorithm as well as outlier methods, statistics, and feature space display properties. This window is described in greater detail in the relevant sections.

Sort State Panel

A collapsible panel displays current algorithm, settings, and statistics.

Output Panel

Error messages are displayed in the Output panel.

Status Bar

A status bar along the bottom of the window identifies the current data set, SortID, and channel.

Working with Panels, Panes, and Windows

OpenSorter makes use of tabbed windows, panes, and panels to organize the workspace. The flexible design allows the workspace to be reorganized or reconfigured quickly to meet work habits or particular tasks.

Panels

The collapsible panels organize task related features, such as selecting data and setting algorithm parameters. When not in use, they can be easily hidden, or minimized, to provide more space for displaying data or editing sort codes in the tabbed window.

To close a panel:

➢ Click the **X** in the top right corner of the panel.

To open a closed panel:

➢ Click the panel name on the **View** menu.

To hide a panel:

> ➢ Click the ![push pin] **push pin** in the top right corner of the panel.
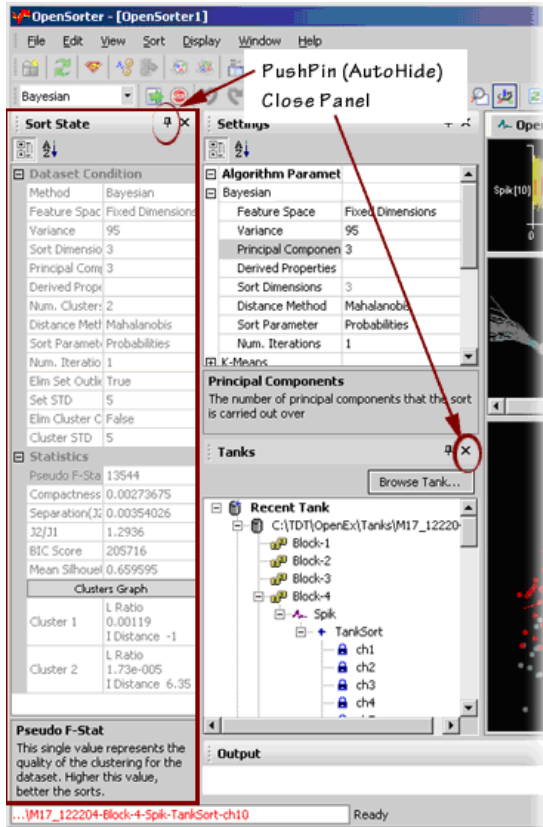
When a panel is hidden, it is represented by a tab-like button along the edge of the workspace.

To quickly show a hidden panel:

> ➢ Point to the panel's button.
>
> **Note:** the panel will be displayed, but will hide again when the pointer is moved away.

To unhide the panel:

> ➢ Point to the panel's button. When the panel is displayed, click the **push pin** button to *pin* the panel to the workspace.

You can resize an open panel by dragging an edge of the panel that is adjacent to another sizeable area.

The gray dotted line along the left edge of the panel's title bar indicates that it can be dragged to a floating position or to a new position in the workspace. Floating panels can be positioned anywhere on the desktop, including areas outside the boundaries of the workspace.

To float or move a panel:

> ➢ Click the title bar and drag it to a new location.
>
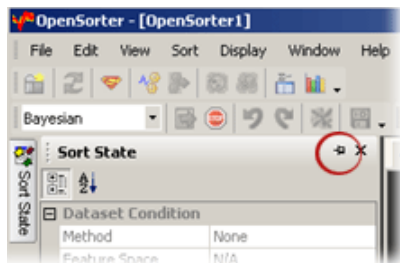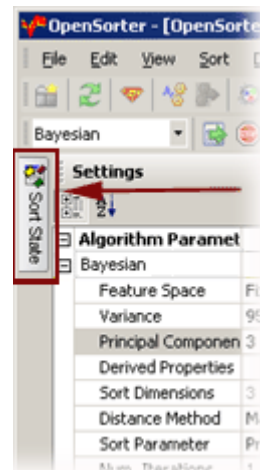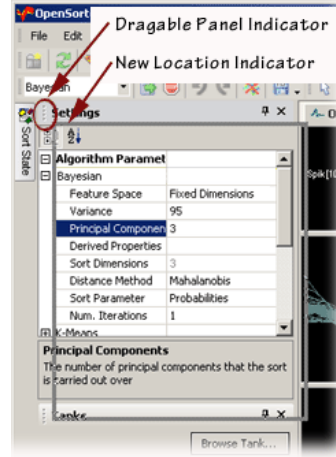> A gray box appears as the panel is moved indicating the new location.

To float the panel:

> ➢ Point to a location away from any edge of the window.

To move the panel to a new docked position:

> ➢ Point to a location along the edge of the workspace.

To create a tabbed panel:

> ➢ Drag the panel over another panel until a tab shape appears along one edge.
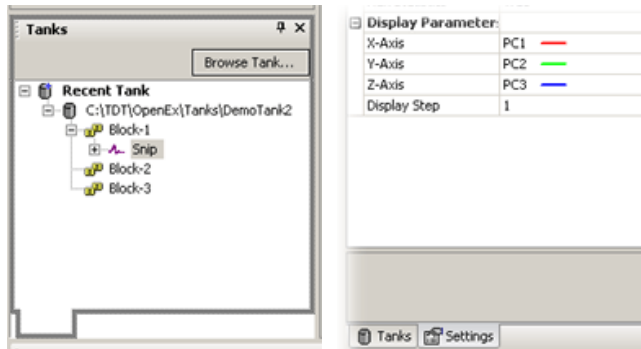
In a tabbed panel you can toggle between open panels by pointing to the tabs.

Windows and Panes

The tabbed window is the primary workspace. It is divided into several panes, to provide different views of the data for display and editing. Unlike the panels described above, these panes cannot be rearranged or hidden. All panes are in view whenever the window is open. However, the panes can be resized.

To resize a pane:

> ➢ Point to the splitter bar until a double sided arrow is displayed, then click and drag to resize the pane.

Multiple Tabbed Windows may be opened to view several data sets simultaneously. The tabbed window provides several different views of the data set. See *Viewing Data in the Tabbed Window*, page 11, for more information.

8

Selecting a Data Set in the Tanks Panel

The Tanks panel is a collapsible panel that can be docked or floated depending on the user's needs. The panel displays DataTanks in a tree structure and allows the user to select the data set for display in the tabbed window.

Browsing For a Tank

If the desired DataTank is not in view, click the **Browse Tank** button in the Tanks panel, then select the desired tank in the **Find Tank** dialog box.
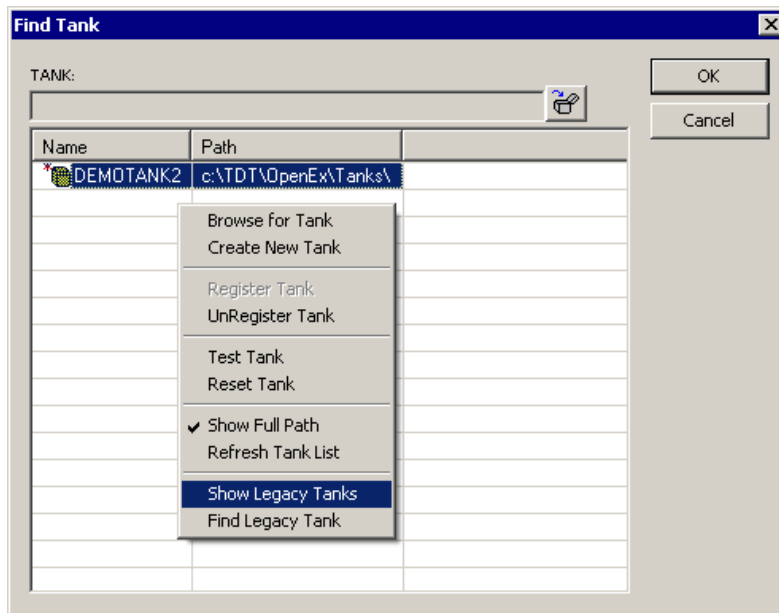
By default, the Find Tank dialog displays registered tanks in the current tank format. The shortcut menu (right-click) provides tools to locate additional tanks.

Show Legacy Tanks: view legacy tanks (tanks generated in version prior to v2.0)

Browse for Tank: locate unregistered tanks in the current tank format

Find Legacy Tanks: locate unregistered legacy tanks

Note: Tanks can be registered or unregistered in the Find Tank dialog box at any time using **Register Tank** and **UnRegister Tank** on the shortcut menu. If a previously registered tank is deleted or moved without first being unregistered, the tank will still show up in the Tanks window.

Displaying Data for Sorting

To display a dataset in the tabbed window, expand the tree structure to reveal the desired data set then double-click to select it. The data is displayed in the tabbed window.
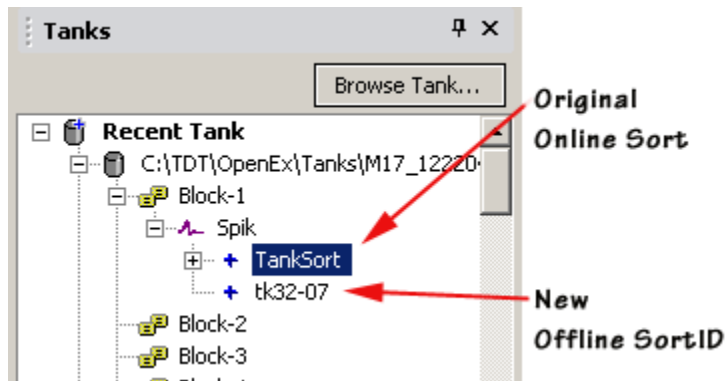
About the Tanks Panel Tree Structure on OpenSorter

The DataTank structure can be thought of as hierarchical. Each DataTank might include multiple blocks, each block might contain several spike events and each block might include several channels. When viewed in OpenSorter, only snippet events are displayed. The original data is displayed as it was acquired under a Sort ID called TankSort. Post hoc sorts are displayed with a user specified SortID, such as *tk32-07* above.

Navigating Using the Keyboard

If the Tanks window is the active window, the arrow keys can be used to move up/down in the tree. The right arrow key expands the branches.

The Save Changes Prompt

By default, each time you navigate away from the active data set in the Tanks panel you will not be prompted to save any changes you have made. You can change this default behavior from the Tree Navigation command on the File menu to prompt you to save changes or save changes automatically.

If set to prompt, a save prompt will be displayed when you modify and attempt to navigate away from the original online sort data set (TankSort). If you selected yes, a new branch is created in the tree structure to represent the new Sort ID and any sort results or applied tank filters are saved in a separate folder within the block folder. The original data set (TankSort) remains unaltered.

If you have been working with a user-defined SortID, and choose to save the changes these will be saved over your existing work in that SortID. To avoid losing your earlier work, click Cancel and save your work from the File menu or using the Save SortID button.

# *Using the Settings Panel*

The Settings panel is a collapsible panel that can be docked or floated depending on the user's needs. It includes parameters for sorting algorithms and display parameters. Tips about each parameter are displayed in the bottom section of the window when the parameter is selected.

Note: options for some settings can be selected by double-clicking the desired setting box in the list. This will cycle through the choices one at a time. You can also manually enter the desired value(s) for the setting.

Algorithm Parameters - This section contains parameters for sorting algorithms, calculating outliers, and enabling statistical information.

These parameters are applied each time a sorting algorithm is run. See the *Configuring Your Algorithm*, page 20, for more information on the settings in the Algorithm Parameters.

Important!: If set to True, the outliers will be calculated and information about outliers defined manually or using the outlier buttons on the Standard toolbar will be discarded. See *Outliers*, page 27, for more information.

Display Parameters - This section contains parameters for controlling how data is displayed in the tabbed window. See *Display Parameters*, page 12, for more information.

Viewing Data in the Tabbed Window

The selected data set is displayed in the tabbed window. Throughout all displays, each sort code (or "unit") is represented by a single color. Unsorted waveforms (U) are shown in light gray and outliers (O) are shown in dark gray. The Mouse toolbar provides tools to zoom, pan, or highlight data in any pane (see *Mouse toolbar*, page 50). Clusters can be hidden, unhidden, or reordered from a short-cut menu (right-click) available in the Unit Display Pane (see below). Multiple Tabbed Windows may be opened to view several data sets simultaneously. The tabbed window provides several different views of the data set.
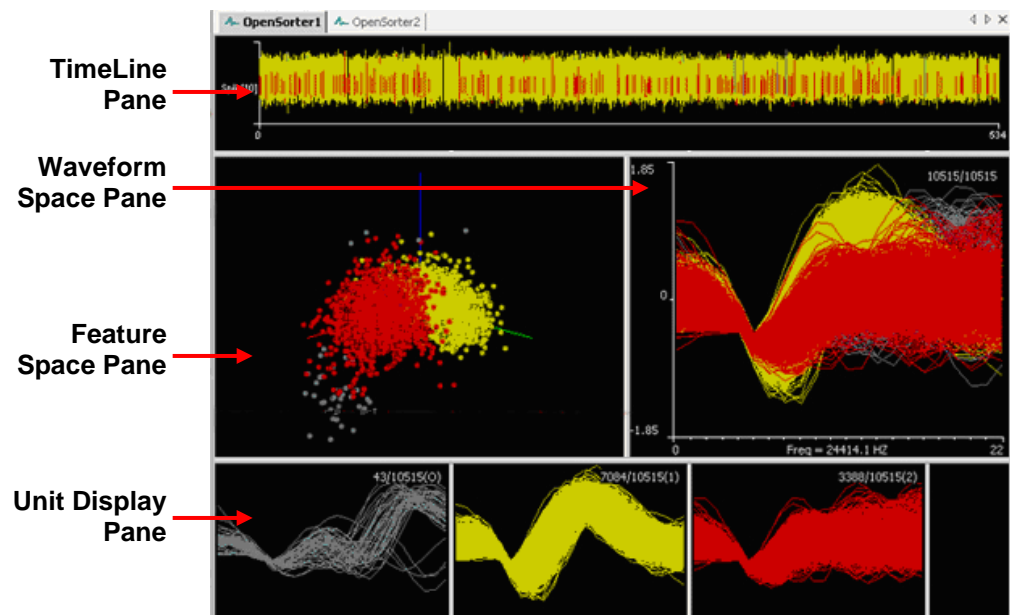
TimeLine Pane

A TimeLine pane runs along the top of the window to display spike waveforms across time. By default, the entire time span of the dataset is displayed.   Waveform amplitude is plotted the y-axis and time is plotted along the x-axis. The event label and channel number are also displayed near the y axis. If an individual trace has been selected using the mouse tools, a red pointer is displayed below the x-axis to mark its position. The Navigation toolbar provides a means of scrolling through the dataset and a Zoom TimeLine Region Tool can be used to examine traces in a visually selected time span. See *Navigating the TimeLine*, page 15, for more information.

Waveform Space Pane

By default, the waveform space displays all spike waveforms in the data set. The waveforms are aligned by sample number and colored by unit, or sort code. The number of waveforms in the data set is displayed for easy reference. In this pane, users can manually classify waveforms based on the actual waveform shape by *drawing* vertical/diagonal time-amplitude lines. All those spikes that pass through the given range of amplitudes at times as specified by the line are assigned to

one sort code. See *Manual Sorting*, page 26, for more information on using the waveform space for manual spike sorting or refining sort results.

Feature Space Pane

In the feature space, waveforms can be viewed in terms of their pertinent properties, such as amplitude or slope. This automatically reduces the dimensionality of the data while making it visually easier to comprehend. By default, the feature space represents spikes in terms of their first three principal components with PC1, PC2, and PC3 represented along the X, Y, and Z axes. (The feature space used for sorting is actually calculated with as many principal components as there are samples for each spike snippet and the display axes selected to be any of the first six principal components). Derived properties, such as the amplitude, slope and area of the spike waveforms can also be computed and displayed in the Feature Space pane. Users can specify the axis in the Settings panel under Display Parameters. This allows the user to explore and consider the best combination of features to provide maximum cluster separation.

Unit Display Pane

The unit display pane provides a convenient means of viewing and comparing waveform shapes across different units, or sort codes. A row of individual waveform plots display the sorted waveforms by unit—each individual plot displays all waveforms classified with a single sort code. Units are displayed from left to right according to their numerical sort code.

When present in the data set, unsorted waveforms are assigned a sort code of "0" and will be displayed to the left of all other sort codes except outliers. Outlier waveforms are classified with a reserved sort code of "31" and will always be displayed in the first plot when present. If there are more units than can be displayed within the width of the tabbed window, a scroll bar is displayed. A numerical display in the top right corner indicates the number of waveforms displayed for the unit followed by the total number of waveforms in the data set. A sort code identifier is displayed in parenthesis to the right of these numbers, with labels indicating the sort code number. An "O" or "U" in this field indicates outliers or unsorted waveforms, respectively.

 The unit display pane supports all of the manual sorting tools and serves as an alternative waveform space in which to edit or refine sort results. A short-cut menu (right-click) provides menu options to hide, unhide, or reorder clusters in the other three panes. When a cluster is hidden an X is displayed across it's subplot in the Units Display Pane.

Display Parameters

Parameters for controlling how data is displayed in the tabbed window can be found in the Settings Panel. These settings only affect display. They do not affect the results of automatic sorting methods.

X-Axis: Select a derived property or principal component to be displayed as the x-axis in the feature space pane.

Y-Axis: Select a derived property or principal component to be displayed as the y-axis in the feature space pane.

Z-Axis: Select a derived property or principal component to be displayed as the z-axis in the feature space pane.

Display Step: Enter a value for n where every nth trace is displayed in the tabbed window panes. 1 to display all traces, 2 for every other trace, 3 for every third trace and so forth.

Tools for Viewing Data

The OpenSorter toolbars provide a variety of tools for viewing data. The Navigation toolbar and the Rotate toolbar group tools used in a specific pane of the tabbed window. The Mouse toolbar comprises a mix of viewing and sorting tools, most of which can be used in any pane. See *Manual*

Sorting, page 26, for more information on tools for fine tuning sort results or performing all sorting manually.

Each of these tools can be accessed via their respective Menus or using the toolbar icons. A subset of tools specific to each pane can also be accessed by right-clicking in each of the pane.

To zoom or pan any pane:

> On the Mouse toolbar, click button for the desired tool, **Zoom** or **Pan,** point to the desired pane then drag the cursor across the pane.

To reset a pane to its default settings:

> On the Mouse toolbar, click the **Restore Default Display** button, point to the desired pane, then click the pane.

To manually rotate the 3D view of the feature space:

> On the Mouse toolbar, click the **Rotate 3D Display** button, click and drag in the Feature Space pane.

To reset the mouse to the default state:

> On the Mouse toolbar, click the **Reset Mouse State** button or right-click the display.

To highlight a trace in any pane:

> On the Mouse toolbar, click the **Pick Nearest Trace** tool, click the desired spike trace in any pane.

The selected trace is highlighted in each of the display panes.
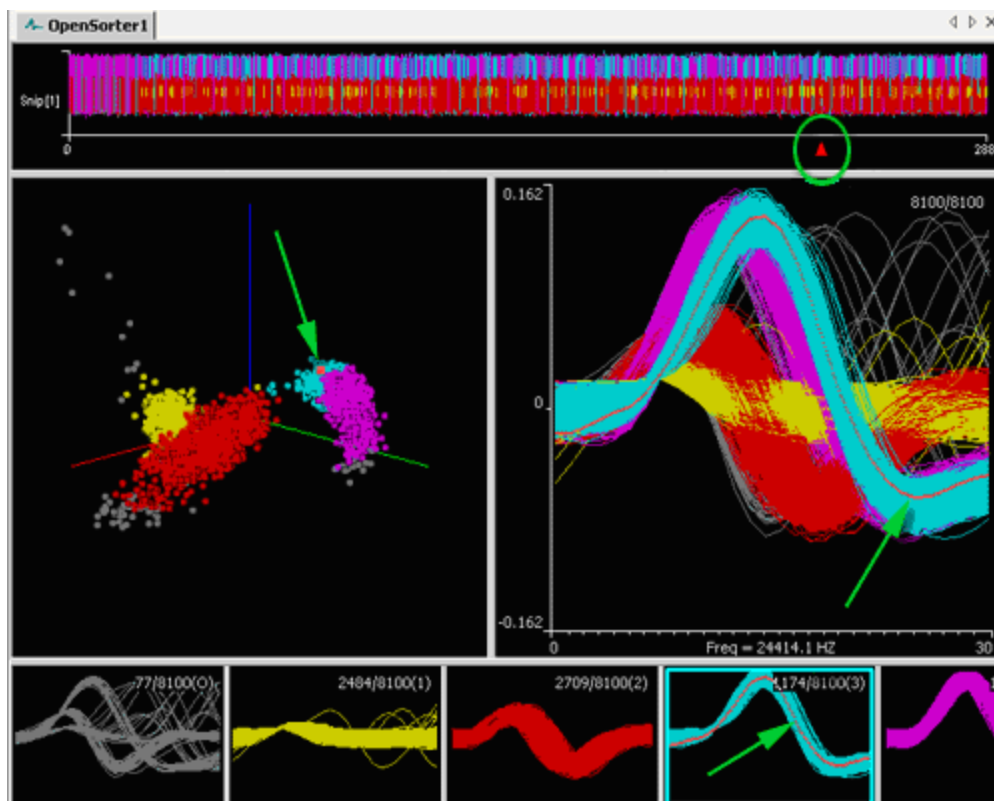
When a trace is selected a shortcut (right-click) menu is available and can be used to:

➢ Highlight the cluster (unit) to which the trace is assigned.

➢ Send the event (spike waveform) or cluster (unit) to another cluster (unit) or outlier group.

➢ De-select the event (spike waveform).

To reorder, hide, or unhide a cluster or clusters:

➢ Right-click the desired cluster in the Units Display Pane and select the desired option from the short-cut menu.
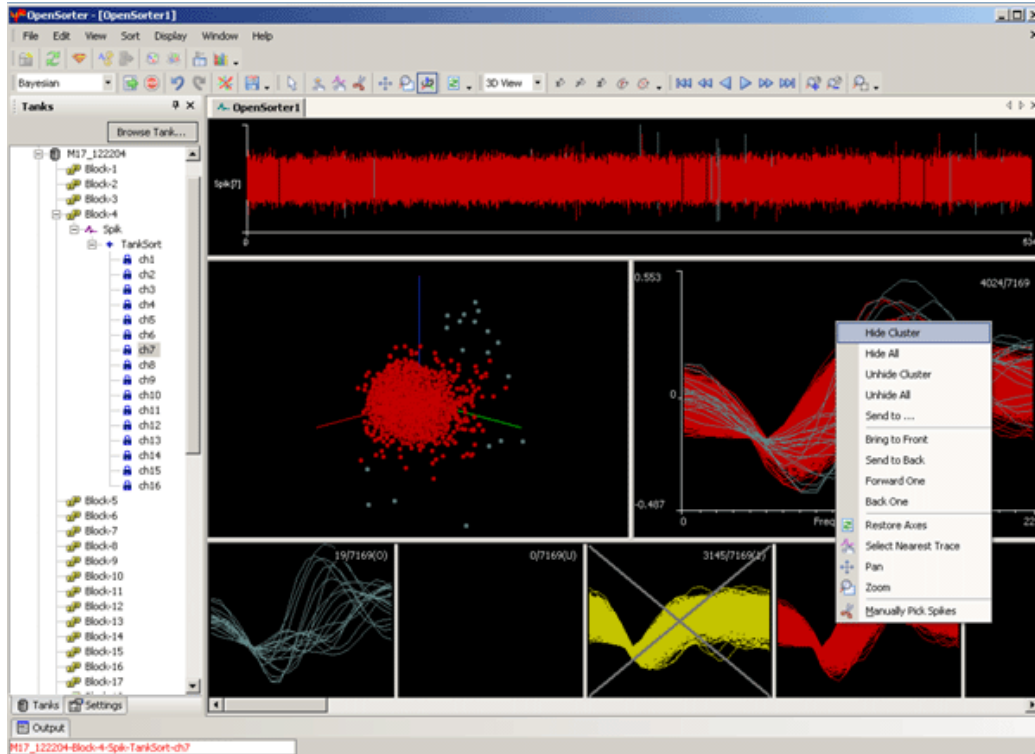
To view a selected time span in the TimeLine:

> ➢ On the Mouse toolbar, click the 🔍 **Zoom TimeLine Region Tool** and drag across the desired region of the TimeLine pane.

To navigate through the TimeLine:
> ➢ On the Navigation toolbar, click the desired button to move through the data incrementally.

Navigating the TimeLine

By default, the entire time span of the dataset is displayed in the TimeLine pane.
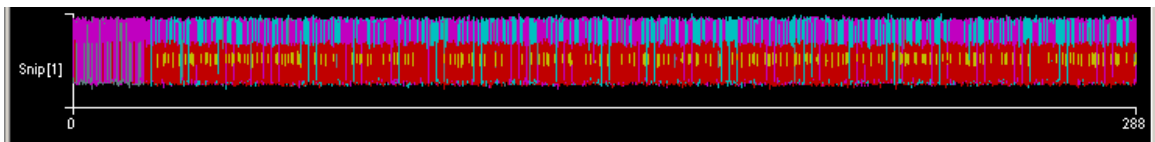
The Navigation toolbar provides tools for expanding or shrinking the displayed waveforms. Any manipulation of the timeline scale retains the earliest displayed time point in the display.

Expand Waveforms (Zooms in the timeline display by a factor of 2)

Shrink Waveforms (Zooms out the timeline display by a factor of ½)

When the timeline is expanded, you can also move the timeline left or right to browse through the data (Duration refers to the visible width of the timeline pane).

Move Timeline Left 1/10th of the Duration

▷ Move Timeline Right 1/10th of the Duration

◁◁ Move Timeline Left 1x Duration

▷▷ Move Timeline Right 1x Duration

◁◁| Move To Start

|▷▷| Move To End

To examine traces in a visually selected time span:

> Click the 🔍 **Zoom TimeLine Region Tool** on the **Mouse** toolbar and drag across the desired region of the timeline.
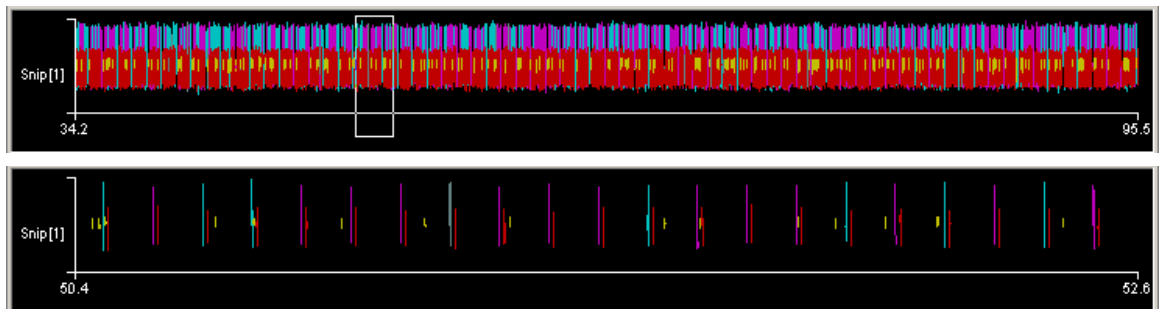
To reset the TimeLine to the default setting:

> Click the 🔄 **Restore Original Display** button on the **Mouse** toolbar then click the TimeLine pane.

Note that the timeline can be expanded or shifted such that no spikes are displayed at the current settings. The Restore Original Display, Move to Start or Move to End buttons are helpful to reconfigure the timeline to display useful data.

Viewing Large Datasets with Display Steps

By default, the entire dataset is displayed in the tabbed window. When working with large datasets it is sometimes helpful to view only a representative sample. In these cases, the user may choose to display every n^{th} trace

To view a representative sample of the data:

> Type the desired value for **Display Step** in the **Display Properties** section of the **Settings panel**. Type 1 to display all traces, 2 for every other trace, 3 for every third trace and so forth.

Exploring Data in the Feature Space Pane

In the feature space pane, waveforms can be represented in terms of their principal components and/or derived waveform properties, such as the amplitude, slope, or area. By default, the feature space display is set for 3D viewing, but the interface provides a great deal of flexibility. You can rotate the display, select from several 2D views, or specify the axes to explore the best

combination of features and viewpoints to provide maximum cluster separation. Users can use the mouse to define cluster centers (see *Closest Centers*, page 20) or to manually pick clusters (see *Manual Sorting*, page 26) in the Feature Space pane. Although the Feature Space can be displayed and man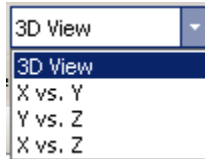ipulated in three-dimensional space, all points are mathematically transformed into the two dimensional display space for the purposes of manual sorting and editing. All mouse-based sorting tools operate in the two-dimensional display space.

To select the 3D or 2D view mode:

➢ On the Standard toolbar, click the **View Angle** box and select the desired view.

3D View ▼
3D View
X vs. Y
Y vs. Z
X vs. Z

To manually rotate the 3D view of the feature space:

➢ On the Mouse toolbar, click the **Rotate 3D Display** button, click and drag in the feature space pane.

The Rotate toolbar provides additional tools for rotating the data in the feature space, including tools to rotate along a specific axis or to start and stop automated, free rotation.

To reset a pane to its default settings:

➢ On the Mouse toolbar, click the **Restore Original Display** button, point to the desired pane, then click the pane.

To specify any axis of the feature space display:

➢ Expand the **Display Parameters** in the **Settings** panel.

➢ Select the principal component or derived property from the drop-down menu adjacent to the axis to be changed. The axes in the Display Parameters menu are color-coded and match the three axes displayed in the feature space.

Settings ⊓ ✕

⊟ **Algorithm Parameters**
⊞ Bayesian	
⊞ K-Means	
⊞ Closest Centers	
⊞ Outliers	
Run Statistics	True

⊟ **Display Parameters**
X-Axis	PC1 ▼
Y-Axis	PC2 —
Z-Axis	PC3 —
Display Step	1

X-Axis
What to display along the X axis
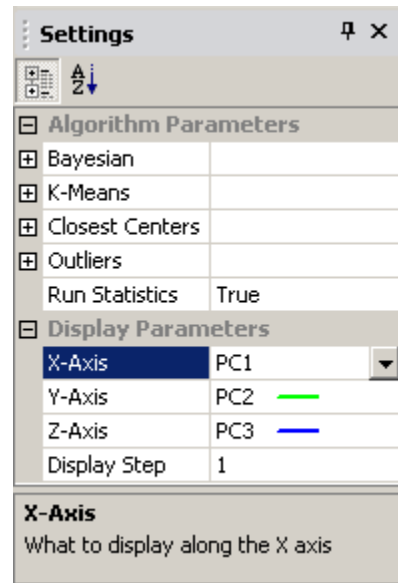
Sorting

OpenSorter offers manual, semi-automatic, and fully automatic sorting methods. This section provides detailed information about using each method.

Topics include:

- ➢ Selecting a Sort Method
- ➢ Using Automated Sort Methods
    - o Fully Automated Sorting – The Bayesian Algorithm
    - o Semi-Automated Sorting – The K-Means Algorithm
    - o Semi-Automated Sorting – The Closest Centers Algorithm
- ➢ Configuring the Algorithm
- ➢ Using Manual Sort Methods
- ➢ Eliminating Outliers
- ➢ Examining Sorting Statistics
- ➢ Exporting Statistics
- ➢ Tank Filtering

Selecting a Sort Method

OpenSorter offers two general sorting methods: 1) automated, algorithm-based sorting methods that automatically assign sort codes to each spike based on user-defined algorithm parameter settings and 2) graphical or "manual" selection and assignment of spikes using OpenSorter's mouse-based selection tools. You can use any method or a combination of methods to yield the best possible sort results. Automated sorting results can later be edited and spikes re-assigned using Manual sorting methods.

Fully Automated Sorting

No user input is required. The number of clusters, or units, is decided algorithmically and waveforms are assigned a sort code automatically.

> **Bayesian** - Sorting based on expectation-maximization analysis of Bayesian probabilities.

Semi-Automated Sorting

User selects a sorting method and specifies key sorting parameters.

> **K-Means** – User inputs the number of clusters, or units, based on visual observation of the Feature Space and then runs the K-Means algorithm that assigns sort codes automatically.

> **Closest Centers** – User selects the location of any number of cluster centers in the Feature Space window and then runs the Closest Centers algorithm that assigns sort codes automatically.

Manual Sorting

Manually specify the unit assignment for each waveform.

> **Waveform Time/Amplitude Method** – Waveforms are manually selected as belonging to a unit by drawing lines across waveforms to create a time amplitude window in a waveform space.

> **Cluster Boundary Method** – Waveforms are manually selected in the feature space by drawing an arbitrary shape around a visible cluster.
>
> **Time Boundary Method** - Waveforms are manually selected in the timeline by drawing a rectangle around a segment of the timeline.

Automated Sort Methods

To select an automated sorting method, select the sorting method on the Select Sort Algorithm drop-down list on the Sort toolbar.
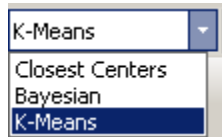
The selected algorithm will appear highlighted in the Settings panel.

The Bayesian Algorithm

The Bayesian algorithm provides fully automated sorting. With this algorithm, OpenSorter evaluates the specified sorting Feature Space of the data set and automatically computes the number of units present in the data and the waveforms that comprise those units. Initially, the entire data set is treated as one parent cluster. This is split to form two child clusters in an iterative process that continues as long as the Bayesian Information Criterion (or BIC score) of the split data set is better as a result of the split and the distortion statistics calculated for the children are not scaled chi-squared distributions. [1, 2]

The K-Means Algorithm

The K-means algorithm is a semi-automated sorting method. The only input required from the user is the desired number of clusters that the data set is to be divided into. A binary split algorithm uses this number as an input and attempts to find the optimum locations of the cluster centers using an iterative process. Data points are then assigned to those clusters based on either their distances away from the cluster center (smallest value) or their probabilities of being allocated to each of the clusters (largest value). [3]

The Closest Centers Algorithm

The Closest Centers algorithm allows you to specify the location of cluster centers in the feature space before sorting. Data points are assigned to clusters based on the nearest defined center. This distance measure could either be in Mahalanobis or Euclidean distance depending on the user's preference. Probabilities are not valid sort parameters when this algorithm is run.

Configuring Your Algorithm

Configuration settings for all automated sort algorithms are located in the Settings Panel. Here users may select from several parameters that are used when the algorithm is run.

Bayesian Setting Parameters

While this algorithm is fully automated, the user may select some parameters in the Settings panel. Default values suitable for most cases are provided.

To modify settings, enter a value or select from available options. If a default (non-number) option in displayed, double-clicking the desired setting box in the list cycles through the setting options one at a time.
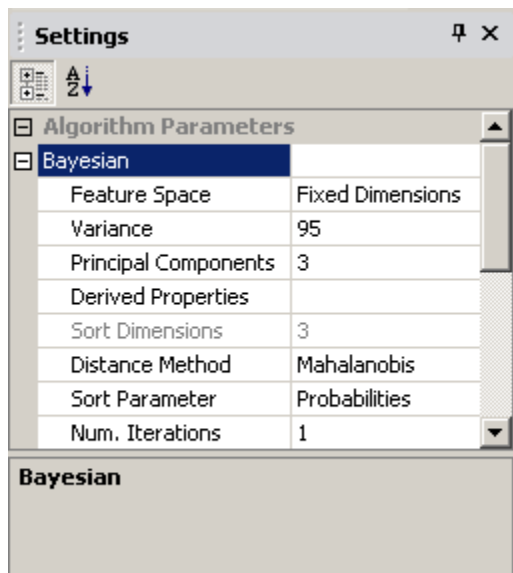
Feature Space: Fixed Dimensions or Fixed Variance

Select whether the number of dimensions for sorting are based on a fixed number of specified dimensions or on the number of dimensions explaining a specified percentage of the variance in the data.

> **FIXED DIMENSIONS:** The number of desired principal components *plus* any specified derived properties (VMax, VMin, MaxSlope, Area) are used to sort the selected data. To use only derived properties, set the principal components field to zero and select the derived properties you wish to use from the derived properties drop-down menu.

> **FIXED VARIANCE:** Uses a variable number of Principal components corresponding to the number of Principal component dimensions required to explain the specified percentage variance in the data set being sorted. The same percentage variance could represent a different number of principal components for different data sets. Note that derived properties are not used when Fixed Variance is selected.

> **Note:** parameters that are not used by the Feature Space setting selected are ignored. For example, when Fixed Dimensions is selected the value set for the Variance parameter is ignored.

Variance: Specify the percentage of variance to be accounted for. Used only when the Feature Space parameter is set to Fixed Variance.

Principal Components: Specify the number of principal components to be used. Used only when the Feature Space parameter is set to Fixed Dimensions.

Derived properties: Select any derived properties to be used as a dimension for sorting. Used only when the Feature Space parameter is set to Fixed Dimensions.

Options for Derived Properties are: VMax, VMin, MaxSlope, and Area. To use multiple options separate each with a comma.

Distance Method: Specify Mahalanobis or Euclidean distance to be used when sorting using distances.

Sort Parameter: Probabilities or Distance

SORTING USING DISTANCES: Data points are assigned to clusters based on distances from the centers of the clusters. Also select the Distance Method (Mahalanobis or Euclidian).

SORTING USING PROBABILITIES: Data points are assigned to a given cluster based on the probability of belonging to that cluster. Bayesian probabilities are computed for the data points and are used to make the assignment decision.

Num. Iterations: The Bayesian algorithm is repeated for the specified number of iterations, and the sort with the best score is automatically selected. After the first iteration, the center of the initial parent cluster is assigned at random, rather than along the axis of maximum variance of the data set. Repeated iterations may produce better results than a single iteration in some cases.

K-Means Setting Parameters

In addition to the number of clusters to partition the data set into, the user may select the following parameters in the Settings panel.

To modify settings, enter a value or select from available options. If a default (non-number) option in displayed, double-clicking the desired setting box in the list cycles through the setting options one at a time.
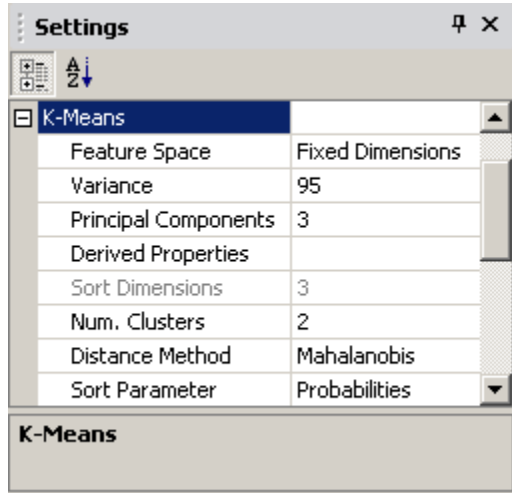
Feature Space: Fixed Dimensions or Fixed Variance

Select whether the number of dimensions for sorting are based on a fixed number of specified dimensions or on the number of dimensions explaining a specified percentage of the variance in the data.

FIXED DIMENSIONS: The specified number of principal components *plus* any specified derived properties (VMax, VMin, MaxSlope, Area) are used to sort the selected data. To use only derived properties, set the principal components field to zero and select the derived properties you wish to use from the derived properties drop-down menu.

FIXED VARIANCE: Uses a variable number of principal components corresponding to the number of principal component dimensions required to explain the specified percentage variance in the data set being sorted. The same percentage variance could represent a different number of principal components for different data sets. Note that derived properties are not used when Fixed Variance is selected.

Note: after selecting Fixed Dimensions or Fixed Variance, the user must ensure that the corresponding parameters, such as variance or principal components, are set appropriately.

Variance: Specify the percentage of variance to be accounted for. Used only when the Feature Space parameter is set to Fixed Variance.

Principal Components: Specify the number of principal components to be used. Used only when the Feature Space parameter is set to Fixed Dimensions.

Derived properties: Select any derived properties to be used as a dimension for sorting. Used only when the Feature Space parameter is set to Fixed Dimensions.

Options for Derived Properties are: VMax, VMin, MaxSlope, and Area. To use multiple options separate each with a comma.

Num Clusters: Specify the desired number of clusters (based on visual inspection of the data).

Distance Method: Specify Mahalanobis or Euclidean distance to be used when sorting using distances.

Sort Parameter: Probabilities or Distance

> **SORTING USING DISTANCES:** Data points are assigned to clusters based on distances from the centers of the clusters. Also specify the Distance Method (Mahalanobis or Euclidean) when using this option.

> **SORTING USING PROBABILITIES:** Data points are assigned to a given cluster based on the probability of belonging to that cluster. Bayesian probabilities are computed for the data points and are used to make the assignment decision.

Closest Centers Setting Parameters

Before setting parameters may be configured for the Closest Centers algorithm, the cluster centers must be defined.

Defining Centers

Before marking centers use the ⟨icon⟩ **Rotate 3D Display** tool to position the view of the feature space display at the best possible angle for marking centers. The display parameter settings can also be used to configure the 3D display. See *Exploring Data in the Feature Space*, page 16, for more information.

Centers are specified by clicking the ⟨icon⟩ **Mark Center** button of the Mouse toolbar then clicking the desired center in the feature space pane. The illustration below shows the feature space pane with centers marked.
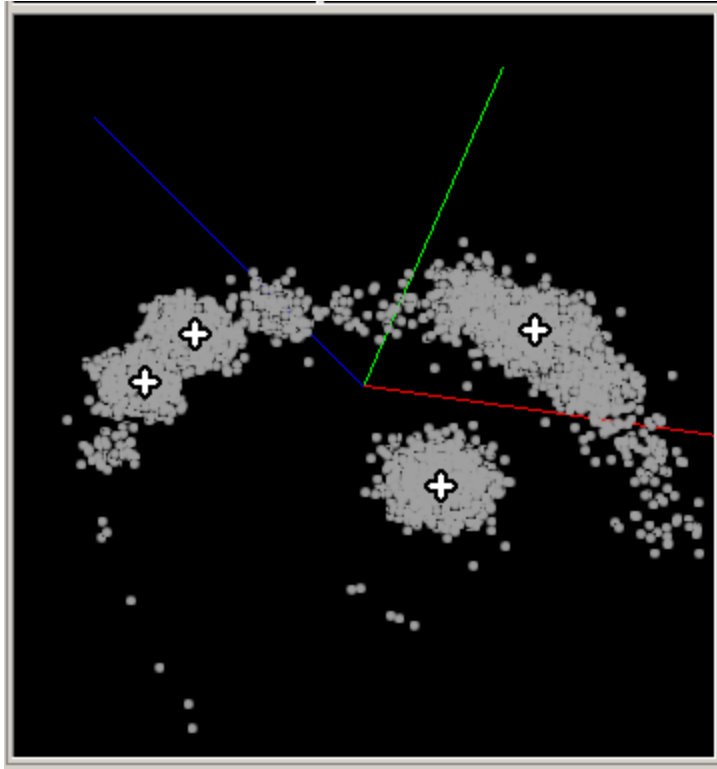
For the purposes of determining the location of marked centers, the feature space is flattened to the two dimensions in view.

If a center is placed incorrectly, all centers must be cleared and placed again. You can clear the centers by repainting the display.

After the centers are defined, the user may select the following parameters in the Settings panel.

To modify settings, enter a value or select from available options. If a default (non-number) option in displayed, double-clicking the desired setting box in the list cycles through the setting options one at a time.
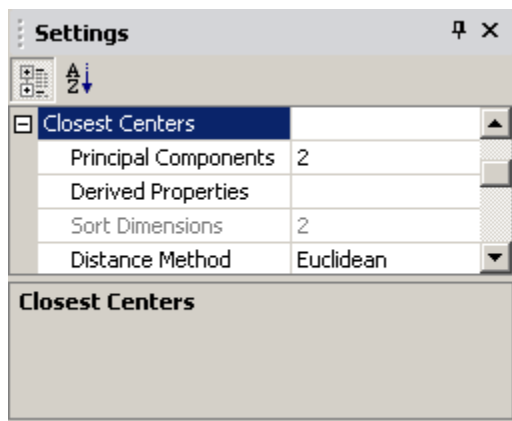
Principal Components: Specify the number of principal components to be used. **Note:** the combined total of principal components and derived properties must equal 3.

Derived Properties: Select derived properties to be used as a dimension for sorting.

Options for Derived Properties are: VMax, VMin, MaxSlope, and Area. To use multiple options separate each with a comma.

Distance Method: Mahalanobis or Euclidean, data points are assigned to clusters based on distances from the centers of the clusters.

Outliers Setting Parameters

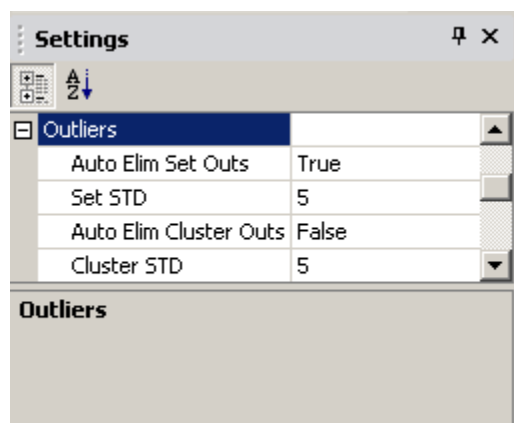OpenSorter supports eliminating outliers at the data set level and/or at the cluster level and allows users to specify the number of standard deviations (max = 10) beyond the mean to set as outlier threshold. See *Eliminating Outliers,* page 27, for more information.

Settings	⊞ ✕
⊟ Outliers	▲
Auto Elim Set Outs	True
Set STD	5
Auto Elim Cluster Outs	False
Cluster STD	5 ▼

Outliers

Auto Elim Set Outs: Set to **True** to eliminate data set level outliers. If set to True, the set level outliers will be calculated and eliminated from the data set before the automated sorting algorithm is run. Set level outliers can also be calculated after the sort is complete using buttons on the Standard toolbar and edited or refined using manual sorting tools.

Set STD Enter the number of standard deviations to be used in identifying data set level outliers.

Auto Elim Cluster Outs: Set to **True** to eliminate cluster level outliers. If set to True, the cluster level outliers will be calculated after the automated sorting algorithm has been run. Cluster-level outliers can be calculated after the sort is complete using buttons on the Standard toolbar and edited or refined using manual sorting tools.

Cluster STD: Enter the number of standard deviations to be used in identifying cluster level outliers.

Run Statistics Parameter

Set to **True** to automatically calculate and display sorting statistics with the selected algorithm. See *Examining Sorting Statistics*, page 29 for more information.

Display Parameters

For a detailed description of the display parameters settings see *Display Parameters*, page 12.

Running the Sorting Algorithm

When the desired parameters are set and the sorting method is selected, click the [icon] **Run Selected Algorithm** button on the toolbar or press F5 on the keyboard.

Important! The sorting algorithm is run based on values set in the Settings panel. Modifications to the feature space display have no effect on sorting. Further, any sorting (manual or automatic) completed prior to running an automated sorting algorithm is discarded when the sorting algorithm is run.

Saving the sort results

After running the desired sort method you will need to save the results to a SortID.

To save the sort results to a SortID

> ➢ Click the 🔢 **SaveSortID** button on the **Standard** toolbar.

Using Manual Sort Methods

Manual sorting provides the greatest degree of flexibility and control at the expense of more time and effort by the user. Manual sorting can be used as the primary sorting method or to edit and re-assign units after automated sorting.

All manual methods use the same tool, but in different sub-windows. They can be used independently or in combination.

Before beginning manual sorting, various settings and tools can be used to display the data set in a favorable way. In the feature space the display can be manipulated to better display cluster separation. See *Exploring Data in the Feature Space*, page 16, for more information. In the timeline pane the display can be scaled to better view individual waveforms. See *Navigating the TimeLine*, page 15, for more information.

To assign units to clusters using any manual method:

1. Click the ✂ **Manual Pick Cluster** button on the Mouse Toolbar (or hold down Ctrl + Shift and click the left mouse button).

    Waveform Time/Amplitude Method – use the mouse to draw a time-voltage line across a bundle of waveforms in the waveform space pane or units display.
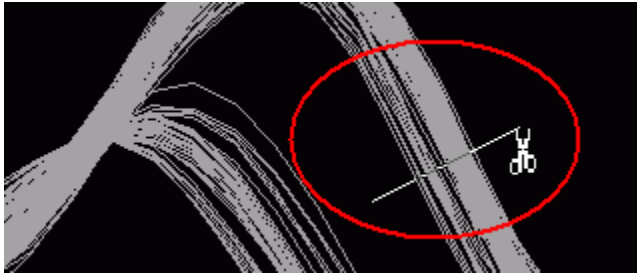
    Cluster Boundary Method –use the mouse to draw an arbitrary shape around a visible cluster in the feature space pane.
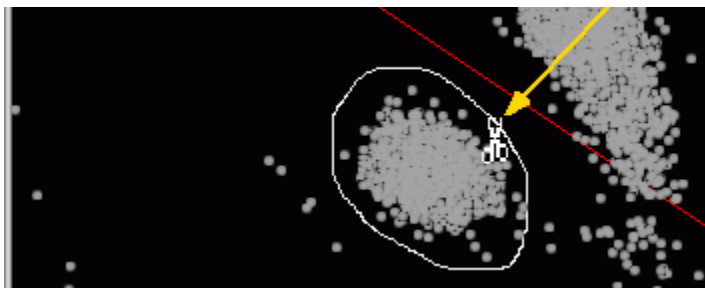
    Time Boundary Method - use the mouse to draw a rectangle around a segment of the

timeline in the timeline pane. Be sure the rectangle is large enough to include waveforms in their entirety. Only waveforms that fall completely within the boundary will be selected.

2. When the mouse button is released, the **Send Events to** dialog will open.

3. Select an existing or empty cluster from the list to send the selected waveforms to the corresponding cluster, or sort code.

   Note: you may use this method to eliminate Outliers by setting the desired waveforms to the Outliers sort code.

4. Repeat this process as needed.

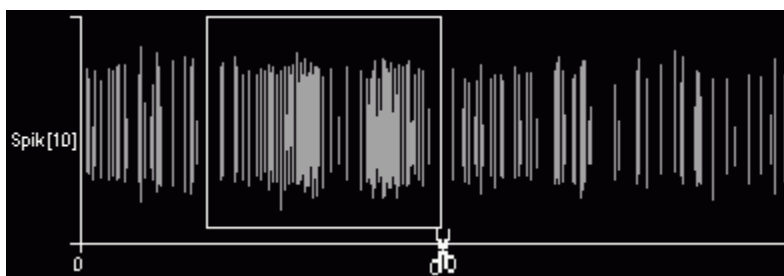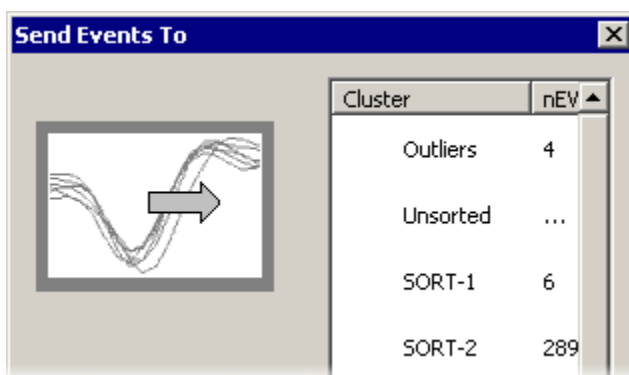5. To save results, click the **SaveSortID** button on the **Standard** toolbar.

To clear all sorts and begin again:

➤ Click the **Clear All Sorts** button on the **Standard** toolbar.

Eliminating Outliers

OpenSorter supports eliminating outliers, either manually or automatically. Manual elimination is performed using mouse-based sorting and editing tools. Automated algorithms are also available that automatically calculate outliers at the data set and/or cluster levels. The function of the automated outlier methods is controlled by parameter settings in the Settings panel. Here, outliers are defined as points that are further than a specified number of standard deviations (max = 10) beyond the mean of the data under consideration.

Data-Set-Level Outliers

Spike shapes that are drastically different from the majority of the spike shapes in the data set tend to lie away from the center of the corresponding feature space representation.

These spikes can be excluded from being sorted by assigning them as outliers. To compute set-level outliers, the Mahalanobis distances of all data points are computed from the center of the entire data set. The mean and the standard deviation of these distances are then used to set a boundary around the data set, n standard deviations away from the mean. Any point lying beyond this boundary is classified as an outlier.

Data set level outliers can be computed either before or after automated sorting algorithms are run. The order in which these operations are run will influence the results of automated sorting algorithms. This is because set-level outliers are not considered as valid input by the automated algorithms. Select True for Elim Set Outliers in the settings panel to eliminate outliers in the data set before automated algorithms are run. To consider all data points in the automated sorting

algorithms, set this field to False. Set-level outliers can later be discarded by clicking the **Calculate Set Level Outliers** button in the toolbar.

Cluster-Level Outliers

After the data has been sorted into different clusters, outliers between clusters can also be eliminated. The logic used is similar to that at the set-level, but here, the Mahalanobis distances of all of the points in each cluster from their respective cluster center are computed. A boundary is then set around each cluster as n standard deviations of those cluster point distances from the cluster mean. Cluster points that lie beyond their cluster boundary are eliminated at this stage as cluster-level outliers. Since, by definition, cluster-level outliers are not calculated until sorting is complete, cluster-level outliers do not influence output of the automated sorting algorithms.

To eliminate outliers automatically when the sort is implemented:

1. If you wish to eliminate data set level outliers, set **Auto Elim Set Outs** to **True** under Outliers in the Settings panel and enter the number of standard deviations to be used in identifying outliers in the **Set STD** value box.

2. If you wish to eliminate cluster level outliers, set **Auto Elim Cluster Outs** to **True** under Outliers in the Settings panel and enter the number of standard deviations to be used in identifying cluster level outliers in the **Cluster STD** value box.

3. After all algorithm settings are also complete, click the **Run Selected Algorithm** button on the **Standard** toolbar.

To eliminate data-set-level outliers in a separate step:

1. In the **Set STD** value box under Outliers in the Settings panel, enter the number of standard deviations to be used in identifying outliers.

2. Click the **Calculate Set Level Outliers** button on the **Standard** toolbar.

   Important Note: if the sort is performed again using the Run Selected Algorithm button, outliers will be recalculated based on settings (True or False) in the Settings panel.

To eliminate cluster level outliers in a separate step:

1. In the **Cluster STD** value box under Outliers in the Settings panel, enter the number of standard deviations to be used in identifying cluster level outliers.

2. Click the **Calculate Cluster Level Outliers** button on the **Standard** toolbar.

To eliminate outliers manually:

1. Click the **Manual Pick Clusters** button on the **Mouse** toolbar. This transforms the mouse into a cutting tool, represented by scissors. This tool can be used in pane of the tabbed window.

2. In the waveform space pane or units display, use the manual picking tool to drag/draw a time/magnitude line across the desired waveforms.

   or

   In the feature space, click and drag to draw a curve around the desired data points.

   or

   In the timeline pane, click and drag a boundary around the desired waveforms.

3. When the mouse button is released, the **Send Events To** dialog opens.

4. Select the **Outliers** cluster.

   The selected waveforms are added to the outliers cluster, that is, they are assigned a reserved sort code of "31" to identify them as outliers.
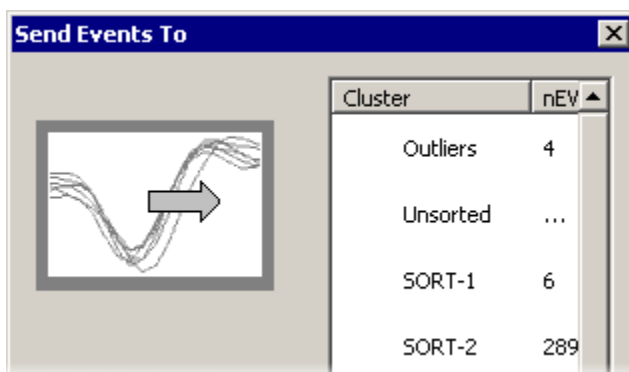
   Important Note: if the sort is performed again using the Run Selected Algorithm button, outliers will be recalculated based on settings (True or False) in the Settings panel.

Examining Sorting Statistics

Statistical measures of your sorted data set can help make determinations concerning the optimum number of clusters, the degree of separation between clusters, and how well individual data points have been classified to their respective clusters. The statistics available in OpenSorter are described below.

Pseudo-F Statistic (PFS)

The PFS is a ratio of variances. It represents a scaled version of the sum of the variance values between clusters over the sum of the variance values calculated within each cluster. This measure is used to help determine the optimum number of clusters. The higher the value, the greater the separation between clusters. [4] The Pseudo-F statistic cannot be computed for single cluster data sets.

Bayesian Information Criterion (BIC) Score

The BIC score is a measure of the log-likelihood of all of the points in the data set being correctly classified. Each data point is assigned a Gaussian point probability based on the parameters of the cluster it is assigned to, and a score for each cluster is then computed. The BIC score of the entire data set is the sum of the BIC scores for all clusters in that data set. Given two clustering models for the same data set, the model with the higher BIC score is preferred. [1] Since the BIC score is a

sum dependent on the number of included points, it should not be used to compare sets with different numbers of points.

J1

The J1 is a measure of the variance of the data points within each cluster. It indicates the compactness of the clusters. The J1 value for the data set is averaged over the J1 values for all the clusters in the data set. A smaller J1 value indicates better compactness. [7]

J2

This measure indicates the separation between clusters and is averaged over all the clusters in the data set. A higher J2 value indicates better separation. Since J2 compares separation between clusters, it cannot be computed for single cluster data sets. [7]

J3 (J2/J1)

As indicated by the equation, the J3 value is a combined measure of the compactness and the separation of the clusters in a data set. A higher value for the J3 index indicates a well-separated data set with compact clusters. Since J3 includes a comparison between clusters, it cannot be computed for single cluster data sets. [7]

Silhouette Index

The Silhouette Index indicates how well each data point has been classified to its assigned cluster as compared to the other possible clusters in the data set. The measurement is made in terms of the average Euclidean distance of a given point to its own cluster, compared to the average Euclidean distance to the points in each of the other clusters in the data set. The Silhouette Index for the data set is the mean of this measure for all points in the data set and ranges from –1 to 1, where 1 indicates a good classification, -1 indicates a bad classification and 0 indicates that the classification could go either way. Since the Silhouette Index compares classification between clusters, it cannot be computed for single cluster data sets. The Silhouette Index is an excellent tool for comparing sort quality between data sets.

L-Ratio

The L-Ratio is an indication of how separated a given cluster is from other clusters. It is the normalized sum of the probabilities with which non-cluster points belong to that cluster. Ideally it should have a value of zero. Since the L-Ratio compares separation between clusters, it cannot be computed for single cluster data sets. [5]

Isolation Distance

This is a Mahalanobis distance measure of the nearness of a cluster to the non-cluster points. The greater the value of the isolation distance, the better the separation. The Isolation distance for the largest cluster is not defined if the number of points in the largest cluster is greater than the sum of the non-cluster points. [5]

Silhouette/Cluster

The Silhouette/Cluster is an extended version of the Silhouette Index values, where a normalized value is computed for each cluster based on the silhouette indices for all the points that comprise that cluster. As with the Silhouette Index, the values can range from –1 to 1, whereby 1 indicates a good classification, etc.

The Silhouette/Cluster is an excellent tool for comparing sort quality between clusters in individual data sets. [6]

Calculating Statistics

Statistics can be calculated for the current sort or a previously sorted SortID and can be calculated automatically when the algorithm is run or post hoc. Statistics can also be calculated across numerous data sets using the Stats only batch method. See *Processing Multiple Data Sets*, page 41 for more information.

Calculating Statistics Automatically

To calculate statistics when the algorithm is run, set the Run Statistics property to True in the Settings panel. The sorting results may take slightly longer to be displayed as the additional task of calculating the statistics is performed before the sorts are updated in the feature space pane, but statistics will be immediately available in the Sort State panel. You can view saved statistics for the current event data by clicking the grey **Clusters Graph** button on the Sort State panel.

Statistics	
Pseudo F-Stat	1222.66
Compactness (J1	1.1921e-007
Separation (J2)	1.9696e-006
J2/J1 (J3)	16.522
BIC Score	4112.24
Mean Silhouette	0.8476
Clusters Graph	
Cluster 1	L Ratio 9.078e-006 I Distance
Cluster 2	L Ratio 0.0003164 I Distance
Cluster 3	L Ratio 0 I Distance 13.6108

Statistics are displayed for the entire sort and for each cluster.

To display saved statistics in a group of graphs, click the gray **Cluster Graph** button.

Calculating Statistics Post Hoc

When the Run Statistic property is set to false, statistics are not calculated when a sorting algorithm is run. To calculate statistics for the current data set after running a sorting algorithm, click the [icon] **Calculate Statistics for Current Set** button on the Standard toolbar. The Statistics window (shown below) will be displayed automatically and the statistics will be available in the SortState panel. In addition, clicking the **Compute Cluster Statistics** button located in the SortState panel can be used to calculate statistics for the current data set and display the Statistics window (shown below).

Statistics	
Pseudo F-Stat	N/A
Compactness (J1	N/A
Separation (J2)	N/A
J2/J1 (J3)	N/A
BIC Score	N/A
Mean Silhouette	N/A
Compute Cluster Statistics	

Statistics are displayed for the entire sort and for each cluster.

To compute cluster statistics for the current data set, click the Compute Cluster Statistics button.

Note: this can also be used to compute the cluster statistics prior to running a sorting algorithm.

The Statistics dialog is automatically displayed whenever statistics are computed post hoc.

Below is an example of the statistics as they are displayed in the Clusters Graph. The **Cluster Pairs** tab displays cluster to cluster comparison statistics.
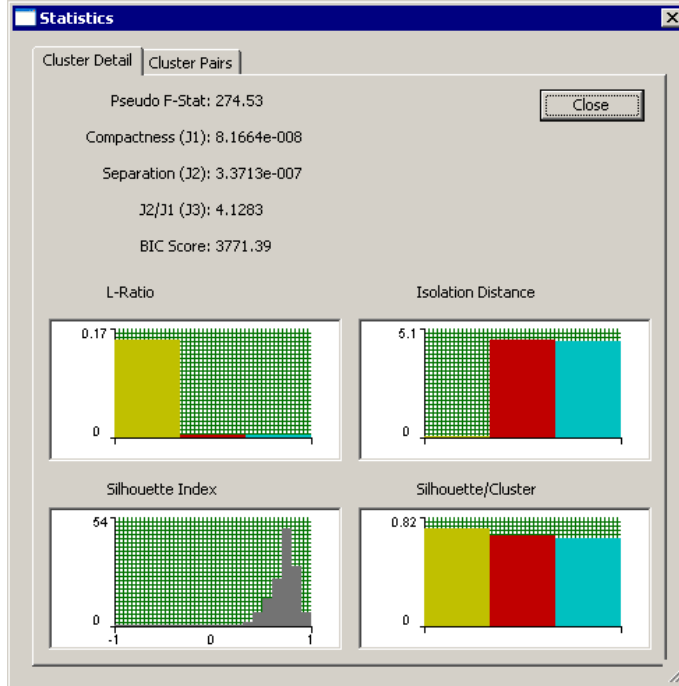
A cluster to cluster comparison allows statistics such as compactness and separation to be computed between each cluster pair. See *Cluster to Cluster Comparison* page 34, for more information.

Viewing Statistics in More Detail

After statistics have been calculated and the Sort has been saved to a SortID (see page 3), you can view statistics in a table in the Statistics Report window. The table provides a convenient way to view or export sort result statistics. It also provides a cluster comparison and allows statistics to be viewed for multiple data sets within a SortID.

To view previously computed statistics in a table format:

1. Select the **Display Saved Statistics** option from the **Sort** menu or click the
    Statistics Table button.

    The Statistics Data Set dialog is displayed.
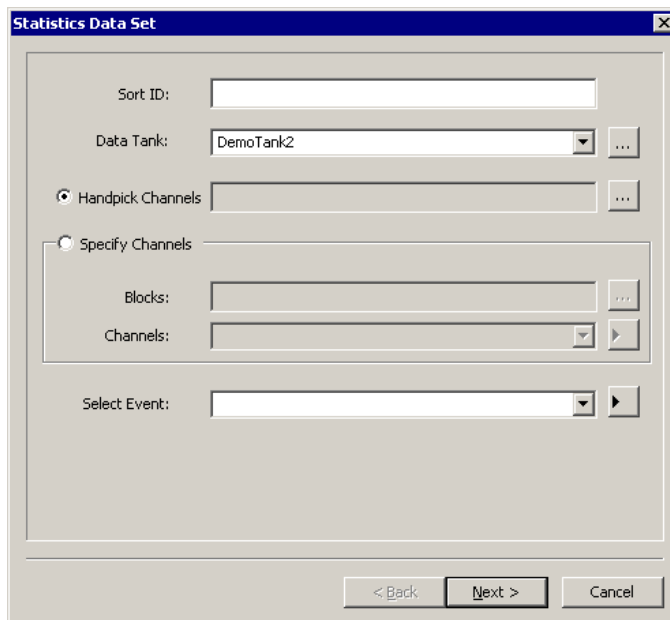
2. Type the desired SortID in the **SortID** box.

3. To select the data tank, click the drop-down box or use the ellipsis (...) next to the **Data Tank** field to browse to the desired data tank.
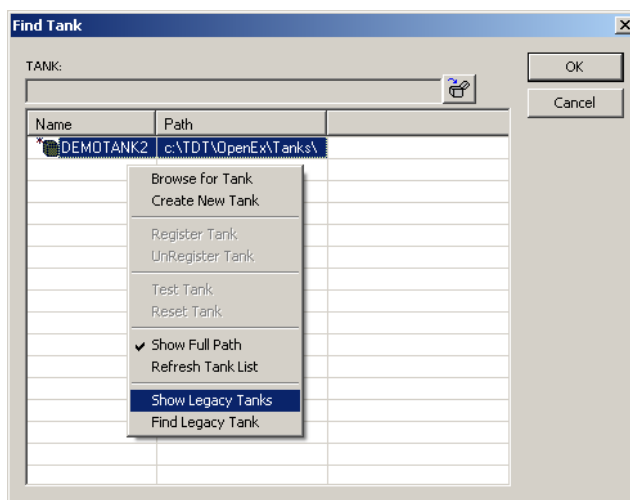

    a. Select a tank from the list or Browse for a legacy format or unregistered tank.

    b. Click **OK**.

4. Next you must select the sorted data sets of interest.

    a. To specify differing channels within each block to be processed (for example, channel 2 of block 3 and channel 4 of block 5), choose the **Handpick Channels** option.

        i. Click the ellipsis (...) next to the **Handpick Channels** option.

        ii. In the **Pick Data Set** dialog, expand the desired blocks by double-clicking the desired blocks and select desired channels using standard

> windows keyboard methods (click, Shift + click to select contiguous blocks or Ctrl + click to select non-contiguous blocks).
>
> iii. Click **Done**.

    b.   To specify a fixed group of channels to be processed all within a selected block (for example, channel 2 of blocks 3, 4, and 5), choose the **Specify Channels** option.

> i. Click the **Specify Channels** option.
>
> ii. Click the ellipsis (...) next to the **Blocks** field. In the Pick Data Set dialog, select the desired blocks using standard windows keyboard methods (click, Shift + click to select contiguous blocks or Ctrl + click to select non-contiguous blocks).
>
> iii. Click **Done**.
>
> iv. In the Channels box, type the desired channel numbers separated by a comma or semicolon or by a dash to select a range. Leaving this field blank will assume all channels are intended to be used.

    c.   To process all channels in all blocks selected, choose either option and select only at the block level. Do not pick or specify any channels.

5.   In the **Events** box, leave the field blank to include all spike events or type the four character event code to select a specific event.

6.   Click **Next**.

Statistics Report

Tank: DEMOTANK2
Event: Snip
SortID: SortID_1

Export | Close

Data Set | Cluster Detail | Cluster Pairs

Block	Chan	Size	Clusters	Pseud...	Comp....	Sep. (J2)	J2/J1 (...	BIC Sc...	μ Me...	σ St...
Block-20	1	152	3	1222.66	1.1921...	1.9696...	16.522	4112.24	0.8476	0.1005
Block-20	2	152	3	1222.66	1.1921...	1.9696...	16.522	4112.24	0.8476	0.1005
Block-20	3	152	3	1222.66	1.1921...	1.9696...	16.522	4112.24	0.8476	0.1005
Block-20	4	152	3	1222.66	1.1921...	1.9696...	16.522	4112.24	0.8476	0.1005
Block-21	1	795	5	5474.85	2.2644...	6.3090...	27.862	21982.74	0.7613	0.1785
Block-21	2	795	5	5474.85	2.2644...	6.3090...	27.862	21982.74	0.7613	0.1785
Block-21	3	795	5	5474.85	2.2644...	6.3090...	27.862	21982.74	0.7613	0.1785
Block-21	4	795	5	5474.85	2.2644...	6.3090...	27.862	21982.74	0.7613	0.1785

Click the Cluster Details or Cluster Pairs tabs for further statistics or export the statistics to a .csv file using the Export button. See *Exporting Statistics*, page 35 for more information.

Cluster to Cluster Comparison

While the general statistics provided by the Clusters Graph are useful, some researchers may require more in depth statistics. The general statistics average the compactness and separation over all clusters which may dilute the accuracy of the statistic. A cluster to cluster comparison allows statistics such as separation (J2) and combined measurement (J3) to be computed between each cluster pair.

Cluster to cluster comparison statistics are displayed in the **Cluster Pairs** tab of the **Statistics Report** window.

To access the Cluster Pairs tab:

➢ Open the Statistics Report window as described above in Viewing Statistics in More Detail and click the **Cluster Pairs** tab.

or

➢ Click the [icon] **Sort Result Statistics** button on the Standard toolbar and click the **Cluster Pairs** tab.

or

➢ Click the **Clusters Graph** button in the **Sort State** panel and click the **Cluster Pairs** tab.

Block	Chan	Cluster A	Cluster B	Sep. (J2)	J1/J2 (J3)
Block-1	1	0	1	4.7917e-006	2.7267e-001
Block-1	1	0	2	9.3934e-008	5.3531e-003
Block-1	1	0	3	6.1728e-006	3.5134e-001
Block-1	1	1	2	3.8100e-007	1.4572e+001
Block-1	1	1	3	1.5823e-005	3.2881e+002
Block-1	1	2	3	2.1164e-008	9.5519e-001

Cluster statistics are displayed in a tabular arrangement where each permutation of cluster comparison is displayed. Statistics for separation (J2) and combined measurement (J3) are provided and color coded for each cluster pair. The cluster pair being compared is identified under the Cluster A and Cluster B columns.

Exporting Statistics

Statistics can be exported directly from OpenSorter. This allows portability of statistics across multiple platforms. Statistics are exported from OpenSorter in *.csv format and can be easily viewed through spreadsheet applications such as Microsoft Excel.

To export saved statistics:

1. Access the saved statistics for the desired SortID(s) as described on page 32.
2. Click the **Export** button on the **Statistics Report** dialog box.
3. Enter the desired filename for the *.csv file and click **Save**.

Tank Filtering

A variety of filter options are available in OpenSorter to include or exclude spike events based on time, epocs, or artifact rejection. These filters may be customized to work together (such as selecting all data outside a particular time interval and within a specified voltage range). This flexibility allows the end user to further manipulate the data set before or after sorting the waveforms.

The Tank Filtering dialog is accessed either by the Tank Filtering tab located at the bottom of the Tanks panel or by selecting the **Tank Filter** option in the **View** menu.

View	Sort	Display
Toolbars	▶	
✓ Status Bar		
Tanks		
Tank Filter		
Settings		
Output		
Sort State		
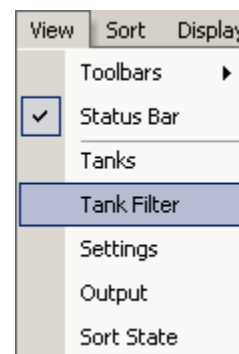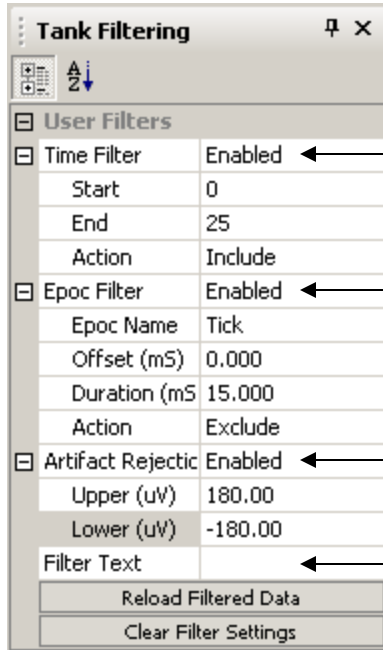
Time filter specification – When enabled, specify start and end times to include or exclude events based on the timestamp of the events.

Epoc filter specification – When enabled, designate an Epoc and filter events based on the time offset of that epoc.

For example, the filter to the left excludes events that are within 15ms after the onset of the Tick epoc.

Artifact Rejection specification – When enabled, designate a voltage range (in µV) for which values outside of the range are rejected as artifacts. A value of 0 turns a particular limit off.

Filter Text specification – Enables users to quickly apply filters through text entry. The Filter text box supports standard logical operators, such as: =, <, >, and, or and is always enabled.

Note: the Filter Text specification allows multiple filters to be combined for more versatility. For example, a filter for (Time > 5 AND Time < 20) OR (Tick < 15) OR (SORT < 3) would include events for the timestamp range 5-20 as well as the epoc Tick for values less than 15 and events with a sort code less than 3.

Applying and Clearing the Filters

When the necessary filters have been specified and **enabled**, click the **Reload Filter Data** button to apply the filtering conditions to the dataset. Data is reloaded from the tank or saved sort and filters applied.

Note: any unsaved sorting information is discarded when data is reloaded. After the filters have been applied, the **(filtered)** label is displayed in the top right corner of the waveform space to indicate that the dataset is currently filtered.

When the Reload Filter Data button is pressed, the filter specifications are applied in the following order: Time Filter, Epoc Filter, and Artifact Rejection. However, filters defined using the Filter Text specification take precedence over any other filters of the same type. The duplicated filter settings are ignored. For example, if a time filter is specified in the Filter Text specification then start and end times defined under the Time Filter specification are ignored.

Clicking the **Clear Filter Settings** button will disable all tank filters for the currently selected channel.
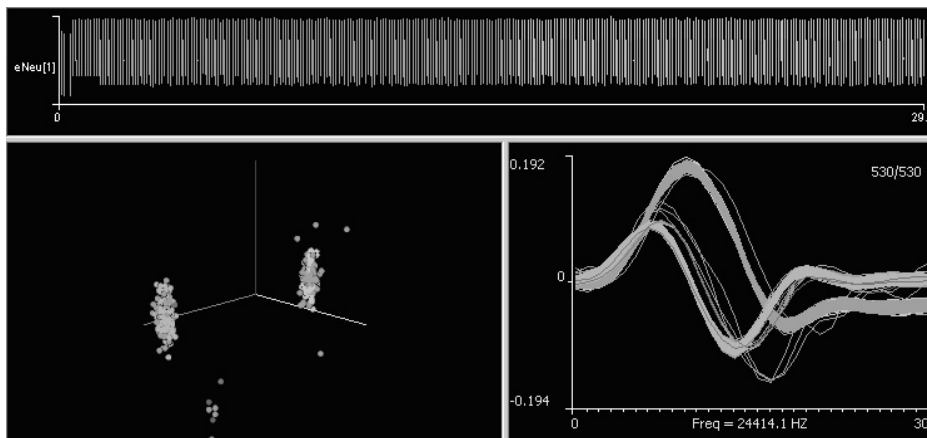
The image above shows the unfiltered data set. The image below shows the data after the filter settings above are applied. The excluded events are removed from the tabbed window display. Note that all filter characteristics have been applied including, a time range of 5-22 seconds, a voltage range -180 µV to 180 µV, and 500 ms of events that occurred 2 ms after the onset of the Tick epoc. Notice that the text (filtered) now appears in the top right corner of the waveform space.

Indicates that Tank Filtering is active for the current channel.
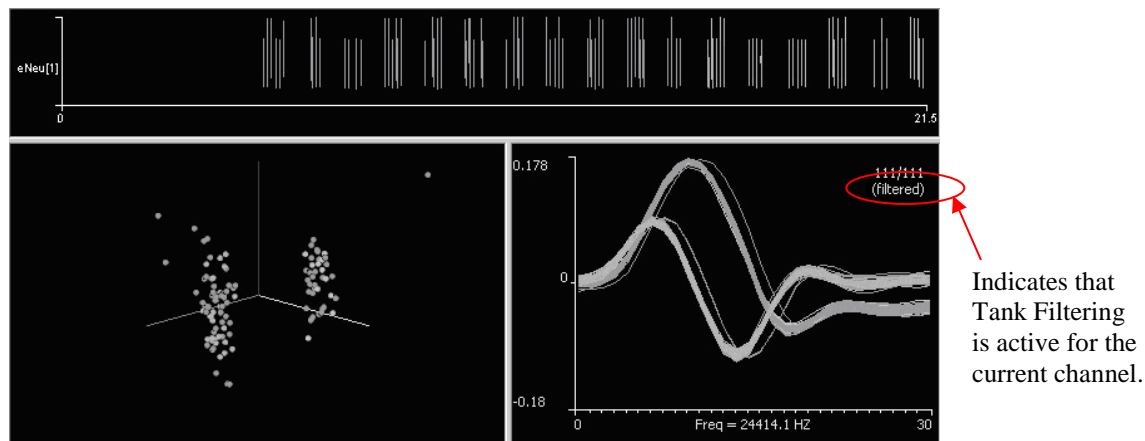
The main advantage Tank Filtering offers is a means to remove unwanted events or anomalous data from the dataset prior to calculation of the statistics. This allows end users to optimize the statistics for events that best represent the characteristics of the spikes. Once saved as a user sortID, the filtered and sorted dataset can then be exported in OpenBrowser for further processing.

Saving Filtering Characteristics

After filtering a dataset, Tank Filtering characteristics are saved when the sort results are saved to a SortID. Each channel may contain its own unique filter characteristics which are loaded whenever the channel is selected. Loaded filter characteristics can then be modified further before being saved again. If another channel is selected, the filter characteristics of that channel are loaded (if available). If there are no saved filter characteristics to load the currently configured characteristics will remain unchanged.

Applying Tank Filters to an Automatic or Semi-Automatic Batch Process

Tank filtering is an interactive tool which is configured for an individual data set. Applying tank filtering characteristics across numerous data sets simultaneously can be done through the OpenSorter Batch Processing option.

To apply a single Tank Filter to multiple data sets:

1. Configure the desired filter characteristics in the **Tank Filtering** Panel and click the **Reload Filtered Data** button.

2. Click **Set-up Batch Process** from the **Batch Processing** option in the **Sort** menu or click the **Batch Sort** button to open the Batch Options dialog.
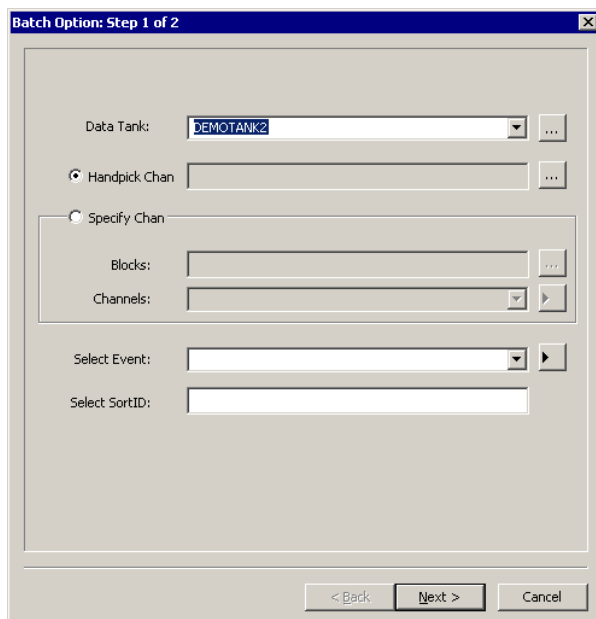
3. In the Batch Options dialog, select the data tank in the **Data Tank** field.

4. Next you must select the data sets of interest.

    a. To specify differing channels within each block to be processed (for example, channel 2 of block 3 and channel 4 of block 5), choose the **Handpick Channels** option.

        i.   Click the ellipsis (...) next to the **Handpick Channels** option.

        ii. In the Pick Data Set dialog, expand the desired blocks by double-clicking the desired blocks and select desired channels using standard windows keyboard methods (click, Shift + click to select contiguous blocks or Ctrl + click to select non-contiguous blocks).

        iii. Click **Done**.

    b. To specify a fixed group of channels to be processed all within a selected block (for example, channel 2 of blocks 3, 4, and 5), choose the **Specify Channels** option.

        i.   Click the **Specify Channels** option.

        ii. Click the ellipsis (...) next to the **Blocks** field. In the Pick Data Set dialog, select the desired blocks using standard windows keyboard methods (click, Shift + click to select contiguous blocks or Ctrl + click to select non-contiguous blocks).

        iii. Click **Done**.

        iv. In the Channels box, type the desired channel numbers separated by a comma or semicolon or by a dash to select a range. Leaving this field blank will assume all channels are intended to be used.

5. To process all channels in all blocks selected, choose either option and select only at the block level. Do not pick or specify any channels.

6. In the **Select Event** box, leave the field blank to include all spike events or type the four character event code to select a specific event.

    Note: the Select SortID box is currently not supported.

7. Click **Next.**

8. Select the desired **Batch Method**. For more information on automated batch methods see the *Processing Multiple Data Sets* section below.

9. Click the **Use currently configured filters** checkbox to apply the currently configured Tank Filter characteristics to the selected data sets.

10. Enter a name for the SortID to be created after the batch process completes.

11. Click **Finish**.

Processing Multiple Data Sets

This section details techniques for batch processing and concatenating data.

Topics include:

➢ Batch Process Multiple Blocks

➢ SuperBlock: Concatenating Data from Multiple Sets

Batch Process Multiple Blocks

OpenSorter's Batch processing mode minimizes the time and effort required to sort multiple sets of data. Batch processing can be performed in fully or semi-automatic modes. Fully automatic batch processing is fastest and provides a means to calculate statistics for multiple data sets. Here, the algorithm properties specified in the Settings panel are used to automatically sort all selected data sets without displaying the results or requiring any further user input. Semi-automatic batch processing sequentially steps through each of the selected data sets and allows the user to sort, view, and edit results for each data set before proceeding to the next.

Note: automatic batch processing supports tank filtering. Filtering may also be used during semi-automatic batch processing after each data set has been sorted and before iterating the batch process to the next data set. See *Tank Filtering*, page 35 for more information.

To batch sort data, perform the following steps:

1. Click **Set-up Batch Process** from the **Batch Processing** option in the **Sort** menu or click the ![Batch Sort icon] **Batch Sort** button to open the Batch Options dialog.
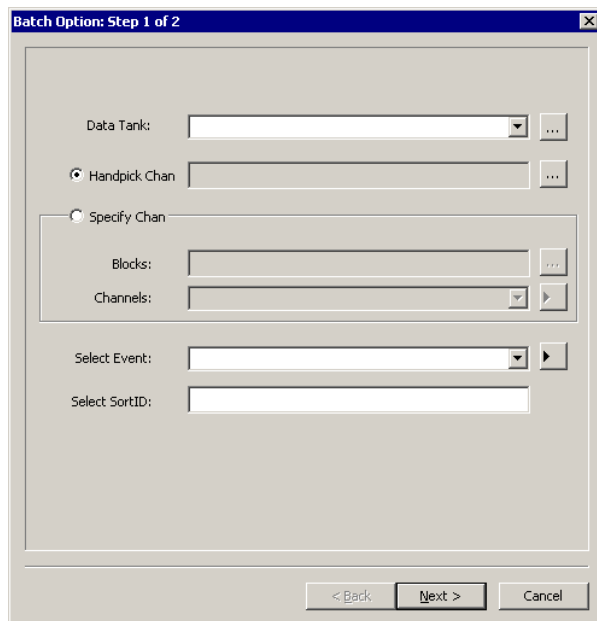
2. To select the data tank, click the drop-down box and select the desired tank or click the ellipsis (...) next to the **Data Tank** field, then select the desired tank in the **Find Tank** dialog.
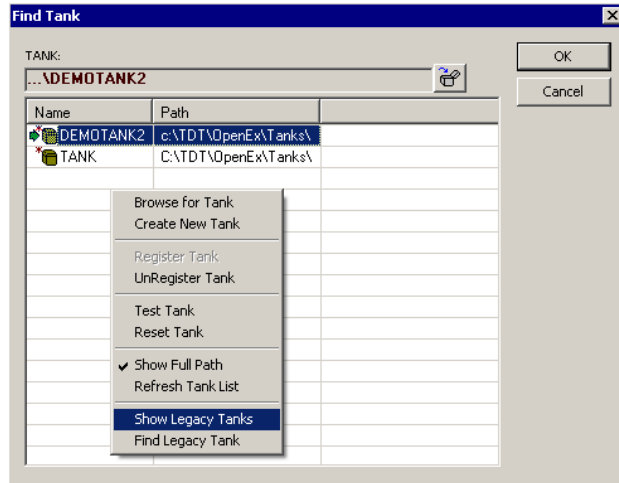
By default the Find Tank dialog displays registered tanks in the current tank format. The shortcut menu (right-click) provides tools to locate additional tanks.

Show Legacy Tanks: view legacy tanks (tanks generated in version prior to v2.0).

Browse for Tank: locate unregistered tanks in the current tank format.

Find Legacy Tanks: locate unregistered legacy tanks.

After the tank is selected, the tank name appears in the Data Tank field.

3. Next you must select the sorted data sets of interest.

    a. To specify a fixed group of channels to be processed all within a selected block (for example, channel 2 of blocks 3, 4, and 5), choose the **Specify Channels** option.

        i. Click the **Specify Channels** option.

        ii. Click the ellipsis (...) next to the **Blocks** field. In the Pick Data Set dialog, select the desired blocks using standard windows keyboard methods (click, Shift + click to select contiguous blocks or Ctrl + click to select non-contiguous blocks).

        iii. Click **Done**.

        iv. In the Channels box, type the desired channel numbers separated by a comma or semicolon or by a dash to select a range. Leaving this field blank will assume all channels are intended to be used.

    b. To specify differing channels within each block to be processed (for example, channel 2 of block 3 and channel 4 of block 5), choose the **Handpick Channels** option.

        i. Click the ellipsis (...) next to the **Handpick Channels** option.

        ii. In the Pick Data Set dialog, expand the desired blocks by double-clicking the desired blocks and select desired channels using standard windows keyboard methods (click, Shift + click to select contiguous blocks or Ctrl + click to select non-contiguous blocks).

        iii. Click **Done**.

    c. To process all channels in all blocks selected, choose either option and select only at the block level. Do not pick or specify any channels.

4. In the **Events** box, leave the field blank to include all spike events or type the four character event code to select a specific event.

    Note: the Select SortID box is currently not supported.

5. Click **Next**.
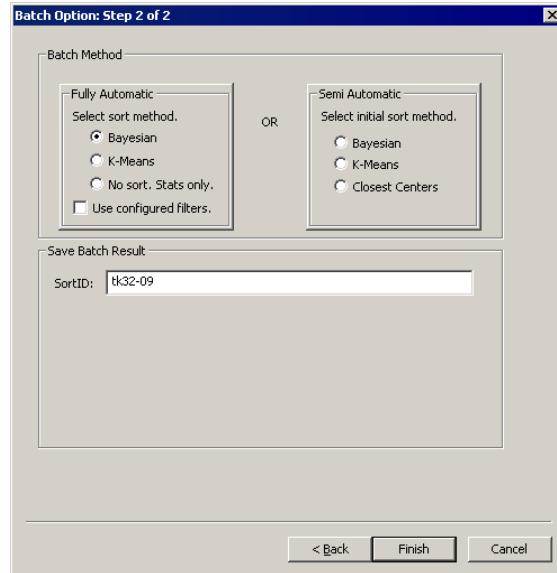
6. Select your desired automatic sort method and whether or not to include the currently configured tank filters (see page 35 for more information) for the specified data sets or select a semi-automatic batch processing mode.

    a. If you select a fully automatic method all specified data sets will be processed using the selected algorithm and parameter values set in the Settings panel. Progress during batch processing will be shown in the status bar at the lower left of the workspace.

    b. If you select **No sort. Stats only**, the specified data sets will have calculated statistics saved to the specified sortID. This is useful if you are using manual or a custom sorting algorithm.

    c. If you select a semi-automatic sorting method, each data set will be processed sequentially.

        ➢ The first data set is displayed.

        ➢ When ready run the selected algorithm, click the **Run Selected Algorithm** button or press F5.

        ➢ The data set will be sorted using the selected algorithm and parameter values specified in the Settings panel. After the sort is completed, the results are displayed and you have an opportunity to make changes prior to the next sort.

        ➢ Once you are satisfied with the sort results, click the **Continue Supervised Batch** button, or press F4, to save the results and advance to the next successive set.

7. Enter a name for the SortID to be created after the batch process completes.

8. Click **Finish**.

# *SuperBlock: Concatenating Data from Multiple Sets*

OpenSorter has the ability to concatenate data from different data sets within a tank and then sort based on the concatenated data set (SuperBlock). This feature is useful when multiple blocks were recorded without changing electrode position, allowing the user to combine electrode channels across blocks.

To create a SuperBlock, perform the following steps:

1. To open the Build Super Block dialog, click **Build SuperBlock** option in the **Sort** menu or click the [icon] **Create superblock** button.
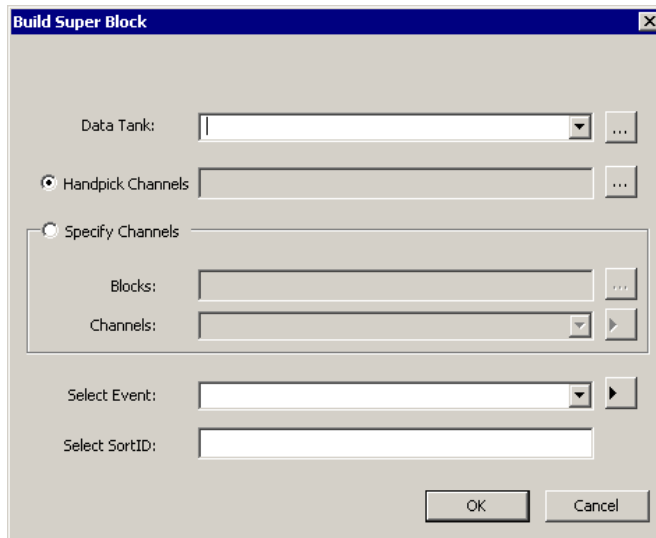
2. To select the data tank, click the drop-down box or use the ellipsis (...) next to the **Data Tank** field to select the desired tank using the **Find Tank** dialog.
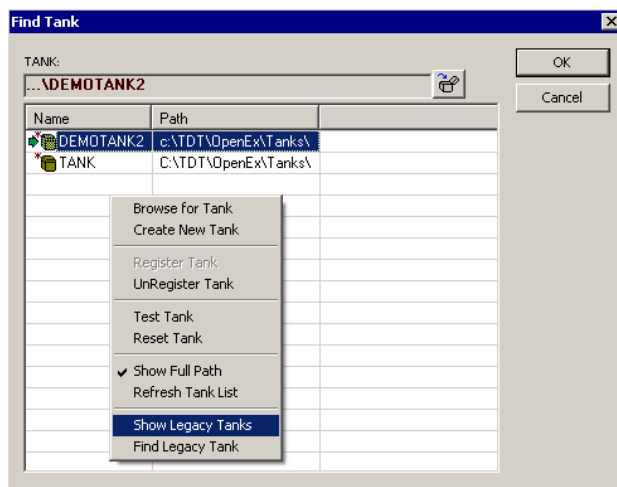
By default the Find Tank dialog displays registered tanks in the current tank format. The shortcut menu (right-click) provides tools to locate additional tanks.

Show Legacy Tanks: view legacy tanks (tanks generated in version prior to v2.0).

Browse for Tank: locate unregistered tanks in the current tank format.

Find Legacy Tanks: locate unregistered legacy tanks.

Once the tank is selected, the tank name appears in the Data Tank field.

3. Next you must select the data sets to concatenate.

   a. To specify differing channels within each block to be processed (for example, channel 2 of block 3 and channel 4 of block 5), choose the **Handpick Channels** option.

      i. Click the ellipsis (...) next to the **Handpick Channels** option.

      ii. In the Pick Data Set dialog, expand the desired blocks by double-clicking the desired blocks and select desired channels using standard windows keyboard methods (click, Shift + click to select contiguous blocks or Ctrl + click to select non-contiguous blocks).

      iii. Click **Done**.

   b. To specify a fixed group of channels to be processed all within a selected block (for example, channel 2 of blocks 3, 4, and 5), choose the **Specify Channels** option.

      v. Click the **Specify Channels** option.

      vi. Click the ellipsis (...) next to the **Blocks** field. In the Pick Data Set dialog, select the desired blocks using standard windows keyboard methods (click, Shift + click to select contiguous blocks or Ctrl + click to select non-contiguous blocks).

      vii. Click **Done**.

      viii. In the Channels box, type the desired channel numbers separated by a comma or semicolon or by a dash to select a range. Leaving this field blank will assume all channels are intended to be used.

   c. To process all channels in all blocks selected, choose either option and select only at the block level. Do not pick or specify any channels.

4. In the **Select Event** box, leave the field blank to sort all events or type the four character event code to select a specific event. You can also enter a SortID for a previously created SuperBlock to view that Superblock again.

   Note: the Select SortID box is currently not supported.

5. Click **OK**.

Progress toward creating the SuperBlock is displayed in the progress bar on the right side of the status bar. In the OpenSorter tabbed window, all spikes contained in the SuperBlock are displayed and ready for sorting. In the timeline pane, vertical white lines are displayed between each data set.

During this process, data is concatenated for sorting purposes, but no new data blocks are formed. Instead, after you sort the SuperBlock and wish to save the results, the resulting SortID is saved back into the original blocks from which the data was gathered.

Menu and Toolbar Reference

This section includes a detailed reference for the OpenSorter menus and toolbars.

Menus

The OpenSorter menus provide access to a variety of tasks. However, most common tasks will be preformed using the OpenSorter toolbars.

Use standard Windows methods to reposition or customize the OpenSorter menu bar.

File Menu

New Tabbed Window	Open a new tabbed window
Open	not yet implemented
Close	Close the active tabbed window
Save	Launch the Save Sort Results dialog
Tree Navigation	Launch the Preference dialog box, allowing the user to specify whether or not to save the active sort when the user navigates away
Save Current Config	Save the current algorithm and settings configuration
Load Config	Load a previously saved algorithm and settings configuration
Exit	Close OpenSorter

Edit Menu

Undo	Undo the last sorting action; supports multiple undo (up to 16 steps)
Redo	Redo the previously undone action; supports multiple redo (up to 16 steps)
Reset Sort Codes	Sets all sort codes in the currently displayed dataset to unsorted
Remove Set-Level Outliers	Calculates dataset-level outliers using the standard deviations value set in the Settings panel and applies the outlier sort code to the outlier waveforms
Remove Cluster Outliers	Calculates cluster-level outliers using the standard deviations value set in the Settings panel and applies the outlier sort code to the outlier events
Manually Pick Spikes	Toggles the pointer to the manual pick spikes tool allowing users to cut clusters visually in any of the tabbed window panes
Select Nearest Trace	Toggles the pointer to the select nearest trace tool

View Menu

Toolbars	Toggle toolbars in/out of view or customize toolbars
Status bar	Toggle status bar in/out of view
Tanks	Toggle Tanks panel in/out of view

Tank Filter	Toggle Tank Filtering panel in/out of view
Settings	Toggle Settings panel in/out of view
Output	Toggle Output panel in/out of view
Sort State	Toggle Sort State panel in/out of view

Sort Menu

Select Active Algorithm	Sets the active algorithm—the algorithm that will be applied the next time the *Run Selected Algorithm* command is run. Also updates the option button display on the Sort toolbar.
Run Active Algorithm	Runs the algorithm selected using the Select Active Algorithm command or using the option buttons on the Sort toolbar.
Abort Sorting Process	Aborts the sorting process. This often results in all sort codes being cleared. To return to the previous sort results use the Undo Last Action button.
Manually Pick Spikes	Toggles the pointer to the manual pick spikes tool allowing users to cut clusters visually in any of the tabbed window panes
Choose Cluster Centers	Toggles the pointer to the choose cluster centers/mark centers tool
Select Nearest Trace	Toggles the pointer to the select nearest trace tool
Save to SortID	Opens the Save Sort Results dialog, allowing you to save the current sort results to a new SortID or any existing SortID
Build SuperBlock	Opens the Build Super Block dialog, allowing you to concatenate data from different sets within a tank
Batch Processing	From the submenu you can:
	Set-up Batch Processing: launches the Batch Options dialog, allowing you to select datasets and configure batch processing
	Edit Next Set in Batch: advances to the next dataset during semi-automatic processing
Display Saved Statistics	Opens the Statistic Data Set dialog where saved statistics for multiple data sets can be selected for comparison or exported

Display Menu

Update Display	Refresh the tabbed window display
Select Nearest Trace	Toggles the pointer to the select nearest trace tool
Mouse Mode Operations	From the submenu, toggle the pointer to one of various mouse tools
Timeline	From the submenu, access commands to brows, expand, or contract the time span in the timeline pane of the tabbed window
3D View	from the submenu, access tools for viewing the data in the feature space pane.

Window Menu

New Tabbed Window	Open a new tabbed window
Cascade	Cascade open tabbed windows, allowing the title bar for all windows to be in view
Tile	Tiles open tabbed windows, allowing all windows to be in view

Help Menu

About	Display version number information
Help	Launches OpenSorter User Guide

Toolbars

Most common tasks will be preformed using the OpenSorter toolbars. Use standard Windows methods to reposition or customize OpenSorter Toolbars.

To toggle the toolbar on/or off, use the **Toolbars** submenu on the **View** menu.

Standard Toolbar

The Standard Toolbar provides access to the most common OpenSorter commands.

Open New Tab

Repaint Display (Refresh Data Set)

Create SuperBlock (Concatenate Data Set from Different Blocks)

Batch Sort (Batch Sort Multiple Data Sets)

Continue Supervised Batch (Process Next Batching Data Set)

Calculate Set Level Outliers

Calculate Cluster Level Outliers

Display Saved Statistics

Calculate Statistics for Current Data Set

Sort Toolbar

The Sort Toolbar is used to determine which sort algorithm is applied when a sort is performed using the Run Selected Algorithm button on the Standard toolbar.

Select Sort Algorithm
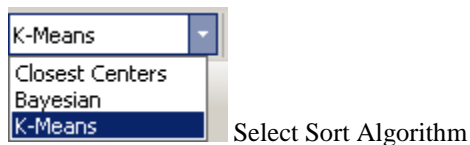
Run Selected Algorithm

Abort Current Operation

Undo the Last Action

Redo the Previous Action

Clear Sort Codes

Save to SortID

Mouse Toolbar

The Mouse toolbar controls the function of the pointer when working in the tabbed window. Each time you click a button on the Mouse toolbar, the function of the pointer changes. You can right-click any area of the tabbed window to return the pointer to its default state—Rotate 3D Display.

Reset Mouse State

Mark Center for Closet Center Sort

Highlight Nearest Trace

Manual Pick Cluster

Pan

Zoom

Rotate 3D Display

Restore Default Display

Rotate Toolbar

The Rotate toolbar provides tools for viewing the data in the feature space pane.
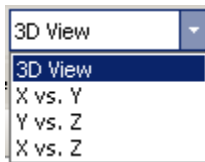
View Angle – for feature space

Rotate X Axis by 5 Degrees

Rotate Y Axis by 5 Degrees

Rotate Z Axis by 5 Degrees

Auto Rotate

Stop Auto Rotate

Navigation Toolbar

The Navigation toolbar provides a means of browsing, expanding, or contracting the time span of the timeline pane. Duration refers to the visible width of the timeline.

Move to Beginning of the Timeline

Shift Timeline Left 1X

Move Timeline Left 1/10th Of X

Move Timeline Right 1/10th Of X

Move Timeline Right 1X

Move to End of the Timeline

Expand Spike Waveforms (Zooms in the timeline display by a factor of 2)

Shrink Spike Waveforms (Zooms out the timeline display by a factor of ½)

Zoom Selected Timeline Region

Keyboard Shortcuts

The keyboard shortcuts below allow quick keyboard access to the mouse tools for viewing and working with data.

Ctrl + Shift + Click and Drag	Manually Sort Waveforms
Alt + Click to Drag	Pan
Shift + Click and Drag	Zoom
Ctrl + Click	Pick Trace
Right Click	Restore Default Mouse Function or Display Context Sensitive Shortcut Menus
F5	Run Active Algorithm
F4	Proceed To Next Data Set during Batch Processing

References

1. Pelleg D., Moore A., "*X-means: Extending K-means with efficient estimation of the number of clusters,* " in ICML 2000.

2. Hamerly G., Elkan C., *Learning the k in k-means.* In proceedings of the seventeenth annual conference on neural information processing systems (NIPS), 281-288, December 2003. (Older UCSD technical report CS2002-0716).

3. Lewicki M.S., *A review of methods for spike sorting: the detection and classification of neural action potentials.* Network: Computation in Neural Systems, 9 (4): 53-78, 1998.

4. Devore J., Peck R., Statistics: The exploration and analysis of data.

5. Schmitzer-Torbert N., Jackson J., Henze D., Harris K.D., Redish A.D. (2005) *Quantitative measures of cluster quality for use in extracellular recordings.* Neuroscience, 131:1-11

6. Bolshakova N., Azuaje F.: *Cluster validation techniques for genome expression data.* Signal Processing (2003) 825-833.

7. Wheeler B.C., *Automatic Discrimination of Single Units in Methods for Neural Ensemble Recordings*, ed. By Nicolelis, M., CRC Press, Boca Raton, 1999.

Glossary

B

Bayesian Information Criterion (BIC) Score: The BIC score is a measure of the log-likelihood of all of the points in the data set being correctly classified. Each data point is assigned a Gaussian point probability based on the parameters of the cluster it is assigned to, and a score for each cluster is then computed. The BIC score of the entire data set is the sum of the BIC scores for all clusters in that data set. Given two clustering models for the same data set, the model with the higher BIC score is preferred. [1]

C

Cluster-level outliers: After the data has been sorted into different clusters, outliers between clusters can be eliminated. The logic used is the same as that at the set-level where for each cluster the Mahalanobis distances of all the cluster points from the cluster center are computed. The boundary then set around the cluster is a function of n standard deviations of those cluster distances away from the mean. Cluster points that lie beyond their cluster boundary get eliminated at this stage as cluster-level outliers.

D

Data-set-level outlier: Spike shapes that are drastically different from the majority of the spike shapes present in the data tend to lie away from the center of the corresponding feature space representation. These spike shapes can be excluded from being allocated to a cluster by assigning them as outliers prior to the sort process. The Mahalanobis distances of the data points are computed from the center of the data set. The mean and the standard deviation of these distances are then used to set a boundary around the dataset, such that any point lying beyond that boundary gets classified as an outlier. This boundary is set as a function of n standard deviations away from the mean. Data points classified as set-level outliers are not considered when an automated or semi-automated sorting algorithm is run.

E

Euclidian distance: The straight-line distance between two points.

F

Feature space: An abstract space where each event is represented as a point in n-dimensional space. Each measurement ("feature") about the event gives the coordinate of the point along one axis of the space. The dimensionality of the feature space is equal to the number of features used to describe the event.

I

Isolation Distance: A Mahalanobis distance measure of the nearness of a cluster to the non-cluster points. The greater the value of the isolation distance, the better the separation. The Isolation distance for the largest cluster is not defined if the number of points in the largest cluster is greater than the sum of the non-cluster points. [5]

L

L-ratio: A statistical measure of how separated a given cluster is from other clusters. It is the normalized sum of the probabilities with which non-cluster points belong to that cluster. Ideally it should have a value of zero. Since the L-Ratio compares separation between clusters, it cannot be computed for single cluster data sets. [5]

M

Mahalanobis distance: The Mahalanobis distance is the straight-line distance between two points weighted by the inverse of the variance in the data set/cluster. Real-world noise that is responsible for the variance in a cluster is typically Gaussian, and Gaussian distributions form elongated/elliptical clusters. In the feature space if the data has a greater variance along one axis as compared to the others, weighting the data by the inverse of its variance along each axis ensures that data points along the direction of the maximum variance axis are not at a disadvantage just because they are further away from the cluster center as compared to those data points that are closer to the center but are along a direction with lesser variance. Hence, using Mahalanobis distances ensures an elliptical boundary around the cluster, and data points further away from the cluster center but along the direction of its maximum variance get included within that elliptical boundary.

O

Offline sort: Sort codes calculated an associated with events after data acquisition is complete.

Online sort: Sort codes associated with an event during data collection.

P

Principal component: Principal components are a multi-dimensional representation of the data set where the first principal component is that parameter that represents the maximum variance in the spike shapes; the second principal component represents the next highest variance in the spike data and so on. Transforming spike data in terms of its principal components allows representing the data in parameters that best describe the differences within the data set while also reducing its dimensionality. The number of principal components that can be obtained for a data set equals the number of sample points in each waveform. However, for data sets with well-defined spike shapes, the first two-three principal components represent about 90-95% of the variance in the data set. All the principal components taken together represent 100% of the variance. Increasing the number of principal components to represent the data may lead to a higher computation cost without a commensurate increase in the percentage variance.

Pseudo F-stat (PFS): A statistical measure that represents a scaled version of the sum of the variance values between clusters over the sum of the variance values calculated within each cluster. This measure is used to help determine the optimum number of clusters. The higher the value, the greater the separation between clusters. [4]

S

Silhouette Index: A statistical measure that indicates how well a data point has been classified to a cluster. The measurement is made in terms of its average Euclidean distance from the cluster it is assigned to, with respect to the minimum of the average distances from each of the other clusters in the data set. The Silhouette Indices can range from –1 to 1, where 1 indicates a good classification, -1 indicates a bad classification and 0 indicates that the classification could go either way.