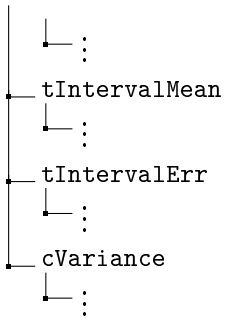
Documentation for the HDF5 data files

Aapo Tervonen

May 6, 2025

1 DATA STRUCTURE

```
file root
├── id ..............  ids for each simulation (form date_time_#)
├── stim ..............  stimulus information
│   ├── stimulusMaginitude
│   ├── stimulusDuration
│   └── stimulusTime
├── tissue ..............  tissue properties
│   ├── nTissueCols
│   ├── nTissueRows
│   ├── stimCell
│   └── distanceToTarget
├── params ..............  model parameters and their values
│   ├── kCRU
│   ├── ⋮
│   └── nSamples
├── timeTraces ..............  calcium time traces for each simulation
│   ├── ⋮
└── features ..............  calcium trace features for each simulation
    ├── nPeaks
    │   └── ⋮
    ├── cMax
    │   └── ⋮
    ├── tMax
    │   └── ⋮
    ├── cPeaks
    │   └── ⋮
    ├── tPeaks
    │   └── ⋮
    ├── fwhmPeaks
    │   └── ⋮
    └── tInterval
```



2 ID

The simulation ids are of the form YYYYMMDD_HHMMSS_i, where the time and date are based on the simulation time, and the index i is for the simulation sample index (running from 1 to nSamples.)

3 STIM

The stimulus used in all of the simulation samples. stimulusMagnitude and stimulusDuration are the amplitude and duration of the stimulus wave. stimulusTime is the time point at which the stimulus begins.

4 TISSUE

Informatino about the tissue. nTissueCols and nTissueRows describes the number of cell columns and rows in the cell array. The linear indices run top to bottom and right to left (Fig. 1). stimCell is the linear index of the stimulated cell. distanceToTarget is the neighbor distance from each cell to the target cell (vector with the length of nTissueColls*nTissueRows).

5 PARAMS

The model parameters and their values.

6 TIMETRACES

Calcium time traces for each cell in each simulation sample. There is one dataset for each simulation within the timeTraces data group, name based on the sample index (running from 1 to nSamples, e.g. root/timeTraces/1). Each sample dataset is a matrix with the size of [nCells,nTimePoints], i.e., each row contains the calcium time trace data for the cell that has the same linear index as the row.

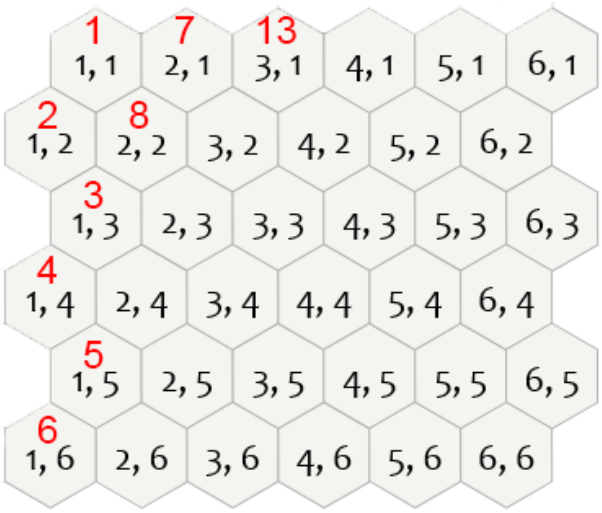


Figure 1: Indexing of the cell array. The red shows the linear indices of the cells, running top to bottom and right to left.

7 FEATURES

The features calculated from the time traces for each cell in each simulation sample. For each feature, there is a dataset under the main feature data group with the name based on the sample index (running from 1 to nSamples, e.g. root/features/nPeaks/1.

7.1 nPeaks

Number of calcium peak for each cell in the simulation sample. This is a vector with the length of total number of cells.

7.2 cMax

Maximum calcium value for each cell in the simulation sample. This is a vector with the length of total number of cells.

7.3 tMax

Time of the maximum calcium value for each cell in the simulation sample. This is a vector with the length of total number of cells.

7.4 cPeaks

The calcium value for each peak for each cell in the simulation sample. This is a matrix with each row corresponding to one cell and each column corresponding the nth calcium peak for that cell. The matrix width is based on the maximum number of peaks in any cell in the simulation. If a cell has fewer peaks than this, the rest of the values are NaN.

7.5 tPeaks

The time of each peak for each cell in the simulation sample. This is a matrix with each row corresponding to one cell and each column corresponding the nth calcium peak for that cell. The matrix width is based on the maximum number of peaks in any cell in the simulation. If a cell has fewer peaks than this, the rest of the times are NaN.

7.6 fwhmPeaks

The full width half maximum of each peak for each cell in the simulation sample. This is a matrix with each row corresponding to one cell and each column corresponding the nth calcium peak for that cell. The matrix width is based on the maximum number of peaks in any cell in the simulation. If a cell has fewer peaks than this, the rest of the times are NaN.

7.7 tInterval

The interval times between the peaks for each cell in the simulation sample. This is a matrix with each row corresponding to one cell and each column corresponding the nth calcium peak interval for that cell. The matrix width is based on the maximum number of interval in any cell in the simulation. If a cell has fewer intervals than this, the rest of the times are NaN.

7.8 meanInterval

The mean interval time between the peaks for each cell in the simulation sample. This is a vector with each row corresponding to one cell. If there are less than two peaks for a cell, this is NaN.

7.9 errInterval

The error interval time between the peaks for each cell in the simulation sample. This is a vector with each row corresponding to one cell. If there are less than two peaks for a cell, this is NaN.

7.10 cVariance

The variance in the calcium time trace data for each cell in the simulation sample. This is a vector with each row corresponding to one cell.