Overview and suggested requirements for multi-channel in vivo electrophysiology
who am i?

• neurophysiologist with 15+ years experience designing, fabricating and recording with multi-electrode arrays in rodents and NHPs

• graduate work: 64 channel data acquisition system

• company: 2,048 channel data acquisition system

open source software, standardized open source data formats

= key to success of data sharing
CRCNS pvc-3 data

- anesthetized cat area 17/18
- continuous LFP (time-series)
- isolated extracellular spikes (times)
- stimuli (gratings, noise, nat. movies)
- extensive metadata describing the recording hardware, electrode config., retinotopic location of recorded cells etc.

- hierarchical data format: mixture of raw binary and text files (metadata)
CRCNS pvc-3 data

The `spike_data` directories contain five different file types:

- `.spk` files contain spike times for an individual neuron, indexed by template id (e.g. `t00.spk`, `t18.spk`, etc.). Timestamps are in microseconds with 10µs precision and stored as binary 64-bit signed integers.

- `spk_info.txt` is a text file with additional metadata describing the `.spk` files.

- `polytrode_xx.pas` is a text file with metadata describing the polytrode used in the recordings, and specifies the mapping between channel number and electrode site coordinates (in microns). The file is Object Pascal code, but it should be trivial to parse and extract the relevant metadata.

- `.tem` contain the multi-channel spike waveform template for an individual spike-sorted unit, indexed by template id (e.g. `t00`, `t18`, `t26`). The template comprises a 1ms average spike waveform from each channel in the electrode array, up-sampled to 100kHz, and stored as 54x100 32-bit real numbers (in millivolts). The data are stored in ascending order by channel, using the channel numbering scheme defined in `polytrode_x.pas`.

- `.tif` image files: `polytrode2a.tif` is an image of the probe showing the recording site configuration; `spk_templates_overlay.tif` shows the templates used for spike sorting.
The stimulus_data directories contain (up to) six different files:

din files contain the timing information for the stimulus, one entry for every vertical frame refresh of the stimulus CRT (i.e. 200Hz, so every 5ms). These are binary files comprising consecutive pairs of 64-bit timestamps/64-bit condition indices. For the drifting bar stimulus, the condition index indicates the orientation of the bar; refer to the relevant .py file for the mapping between condition index and bar orientation. For the spatiotemporal white noise and dynamic natural scene movies the condition index refers to the (zero-based) frame index of the .m file (see below) that was displayed on that refresh. Since the movie frames were updated at 50Hz, and the CRT refresh was 200Hz, there are four repeating entries per frame. As with the spike times, all timestamps are in microseconds, stored as 64-bit signed integers.

stim_info.txt is a text file with metadata describing the .din file.

.py files are text files with additional metadata describing the stimulus. They are actual Python scripts that, if executed in conjunction with Vision Egg & DimStim, will display the original stimulus presented in the recording session.

.m files contain the stimulus movie sequences. Both the spatiotemporal white noise and the dynamic natural scene movies are 64 x 64 pixels, with a resolution of 0.2°/pixel, subtending 12.8° of visual angle. Pixels are 8 bit grayscale, stored uncompressed as bytes (unsigned integers). The movies were displayed at a frame rate of 50Hz, and run for 2 minutes each, thus each movie sequence is 6000 frames long. Note that there is no movie for the drifting bar stimulus; this was generated ‘on the fly’ by Vision Egg.

edf files are native Eyelink II files and contain binocular eye-tracking data for the monkey recordings, sampled every 2ms (500Hz).

eye files are tab-delimited text files exported from the original .edf files. The data are arranged in columns, for example:
Allen Institute data

- distributed recording in awake behaving (transgenic) mice
- 384~640 channels
- binocular eye tracking
- behavior tracking…

Berkeley ‘flex’ probes

IMEC ‘neuropix’ consortium probes
top 6 requirements

• store data in native format with API to convert to SI units (losslessly)

• bind meta-data with raw data non-redundantly (HDF-5)

• retrieve (query-filtered subsets) of multi-GB files quickly (HDF-5)

• handle discontinuous recordings, different sampling rates and facilitate analysis across trials, data types (event vs. time-series)

• data source ‘paper trail’ for multiple instantiations of derivative data (e.g. spike sorting results from different algorithms)

• register data across coordinate frameworks across scale / modalities