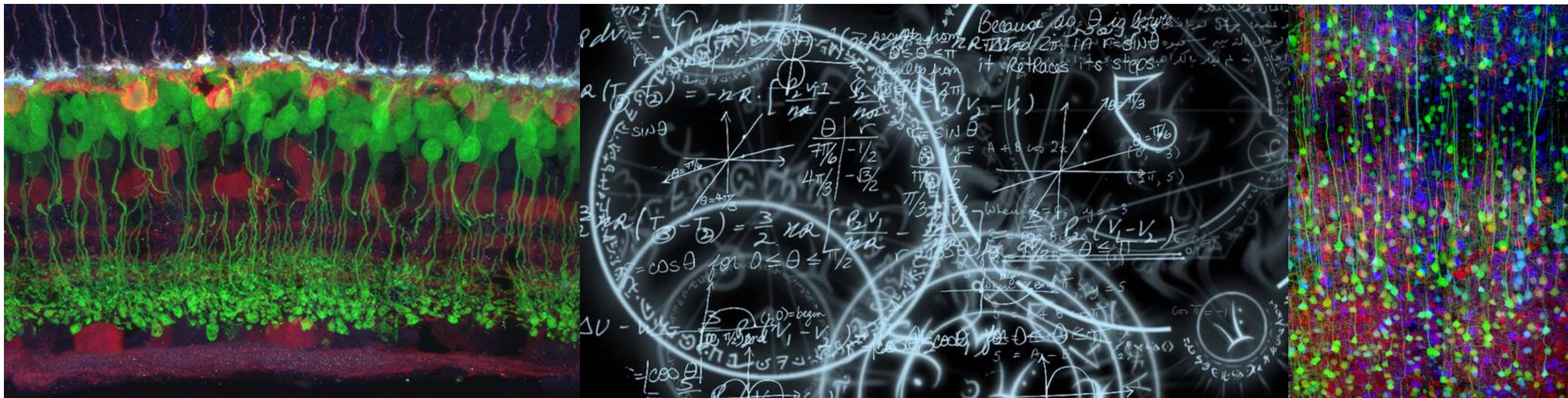
EBERHARD KARLS UNIVERSITÄT TÜBINGEN

MEDIZINISCHE FAKULTÄT

Universitätsklinikum Tübingen



Neural Data Science

Spike sorting & evaluation

Prof. Dr. Philipp Berens

Institute for Ophthalmic Research, Tübingen

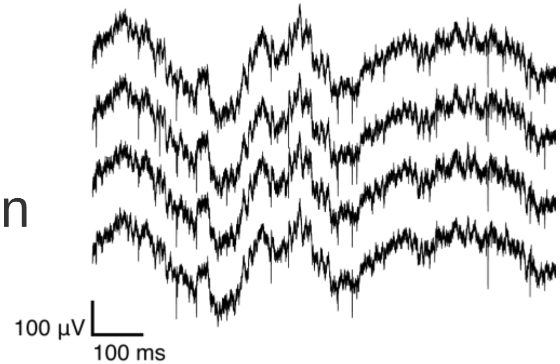
Data briefs

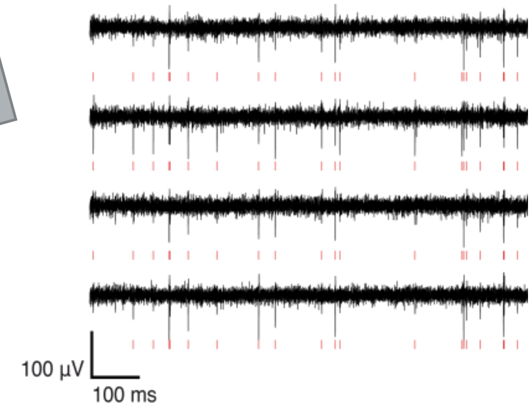
What was the average life expectancy in Europe and Africa

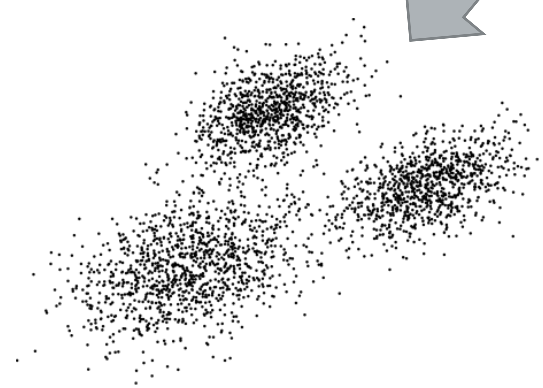
- 1800?
- 1950?
- 2015?

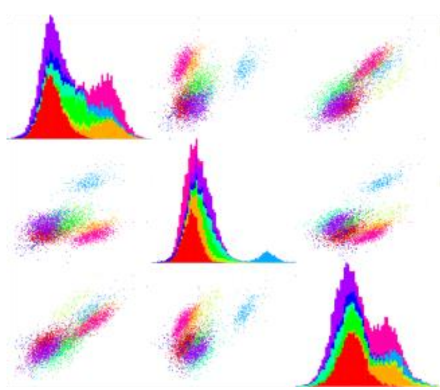
Spike sorting

- Raw data
- Spike detection
- Feature extraction
- Clustering
- Verification



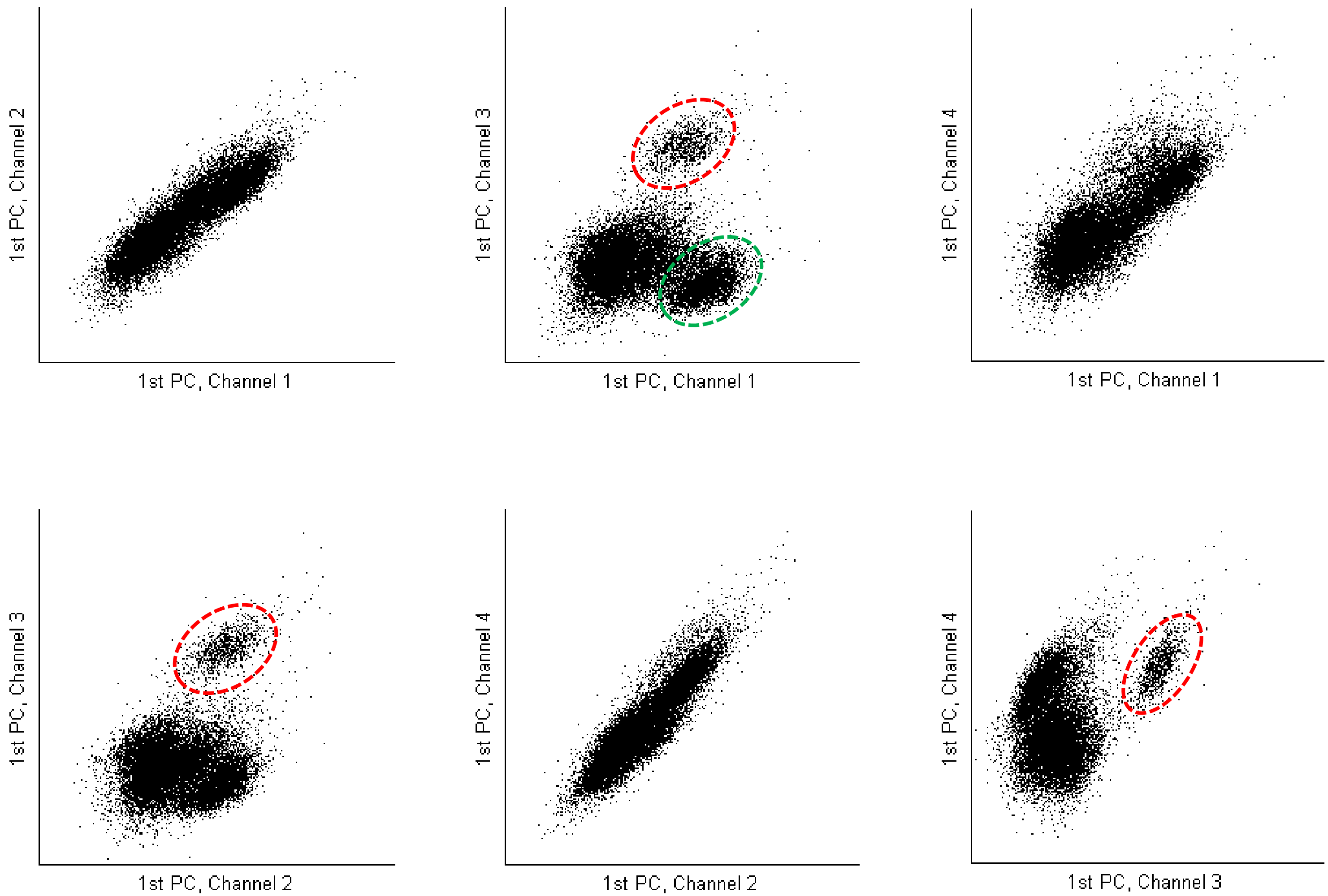




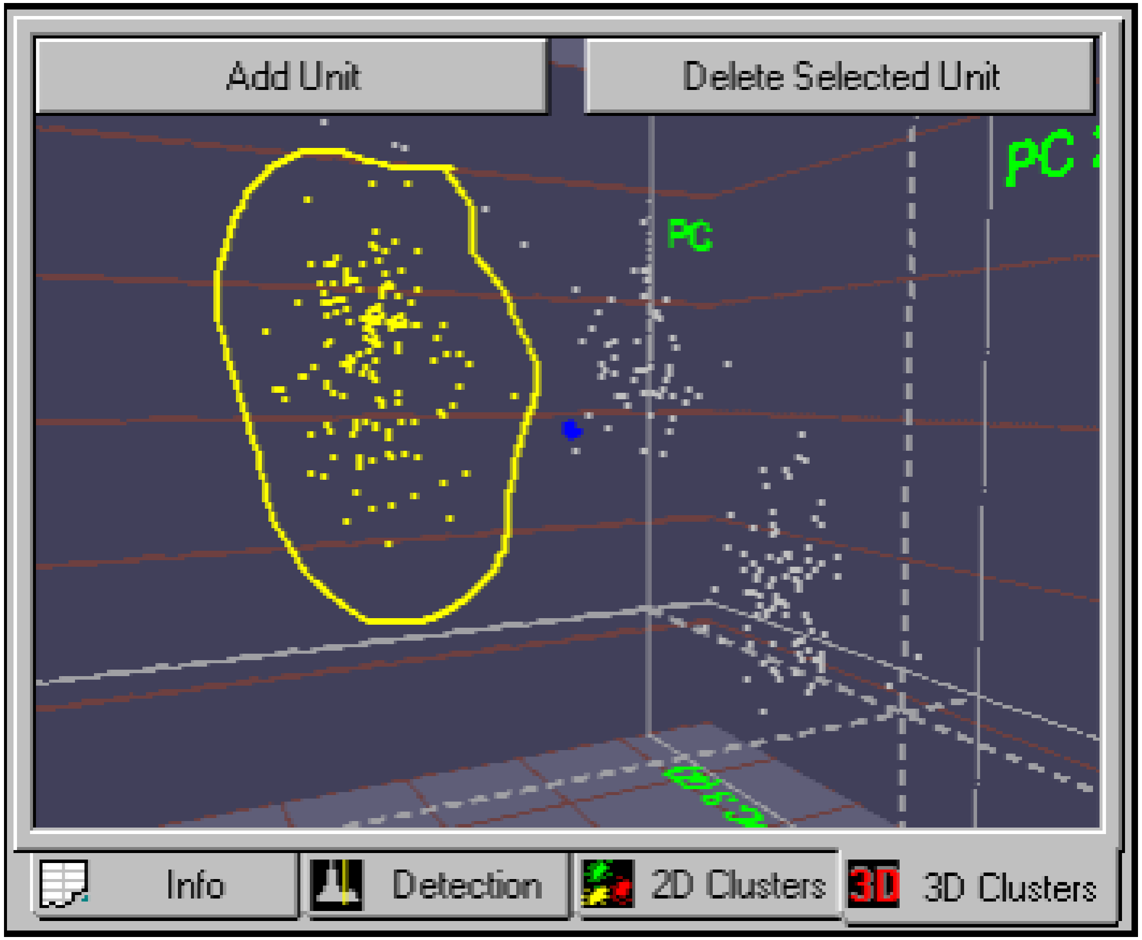


Spike sorting

Scatter plot

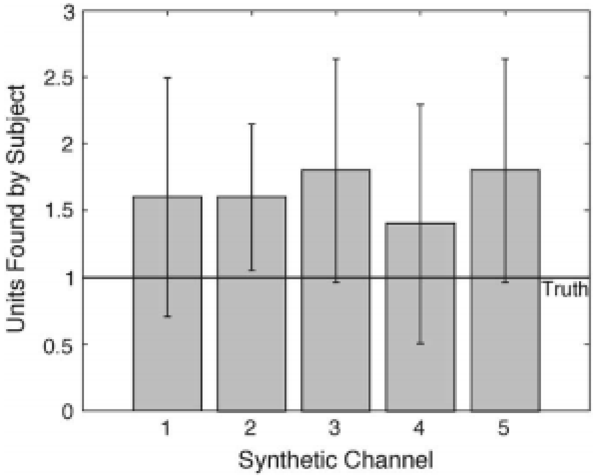


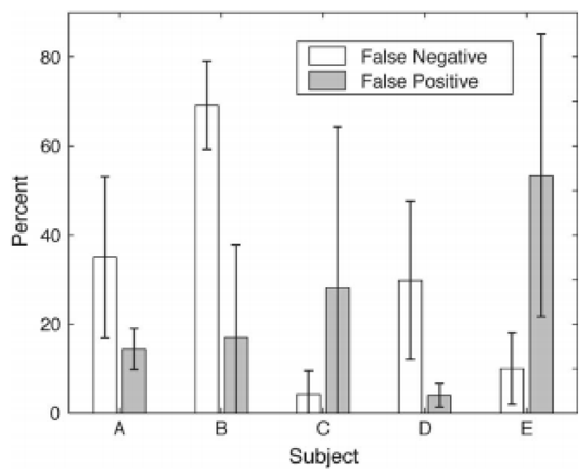
Manual cluster cutting



Why is that not a good idea?

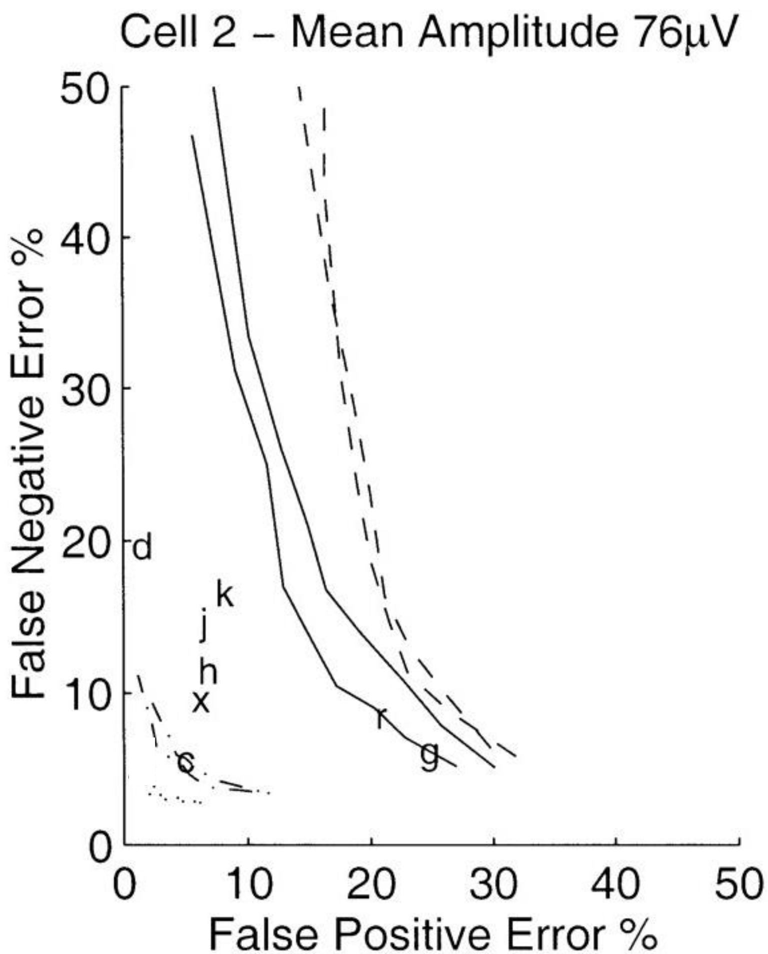
- Subjective
- High error rates
- Suboptimal boundaries
- Time consuming
- Not reproducible
- Not model based





Wood et al. 2004

Problems of manual sorting



1 PC

2 PCs

3 PCs

Peak-to-peak

Letters: human operators

Harris et al. (2000), J Neurophysiology

Automatic clustering

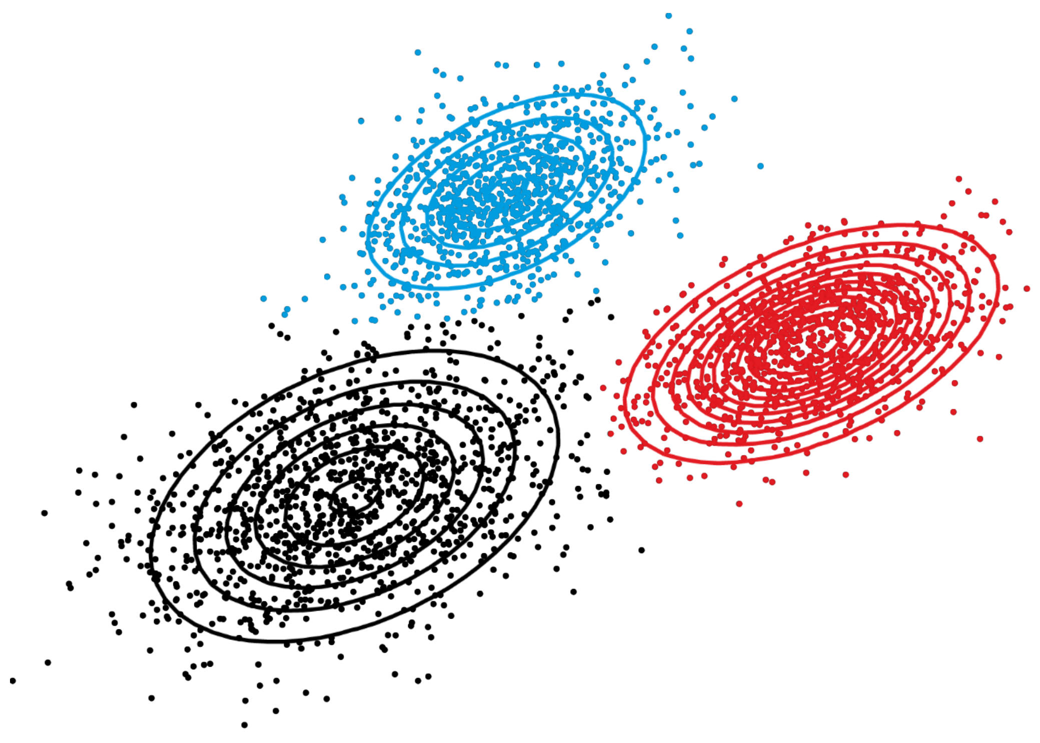
- Objective
- Model-based
- Reproducible
- Quantifiable

- K-Means
- Mixture of Gaussian
- Others?

Which algorithm?

Gaussian mixture model

$$p(x) = \sum_k \pi_k \mathcal{N}(x \mid \mu_k, \Sigma_k)$$



Log Likelihood

$$\ln p(x|\mu, \Sigma, \pi) = \sum_n \ln \sum_k \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)$$

π_k	Mixing coefficient of cluster k
μ_k	Mean of cluster k
Σ_k	Covariance of cluster k

Expectation maximization (EM)

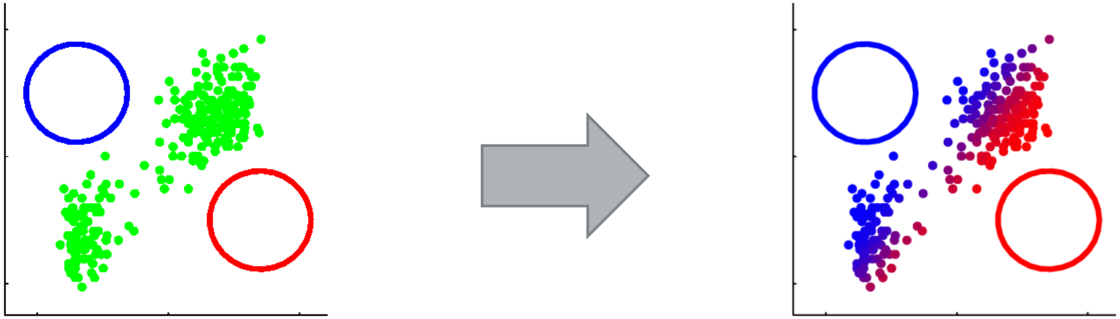
- Finds maximum likelihood estimate of latent variable model
- Parameters: θ
  Latent variables: Z
  Data: X
- Alternate until convergence:

1. **E step**
   Estimate latent variables Z given the current set of parameters θ
2. **M step**
   Update parameters θ to maximize the likelihood given current estimate of the latent variables Z

E-Step

Evaluate “responsibilities” with current parameter values (posterior probability of a data point to belong to cluster k)

$$\gamma_{k,n}^{new} = \frac{\pi_k \mathcal{N}(x_n | \mu_k, \Sigma_k)}{\sum_j \pi_j \mathcal{N}(x_n | \mu_j, \Sigma_j)}$$



M-step

Update means:

$$\mu_k^{new} = \frac{1}{N_k} \sum_n \gamma_{k,n} x_n$$

Update covariances:

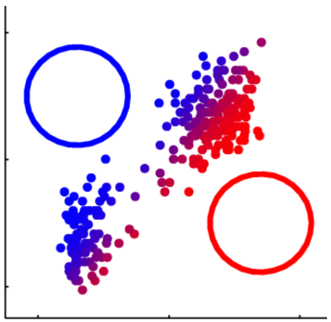
$$\Sigma_k^{new} = \frac{1}{N_k} \sum_n \gamma_{k,n} (x_n - \mu_k^{new})(x_n - \mu_k^{new})'$$

Update mixing coefficients:

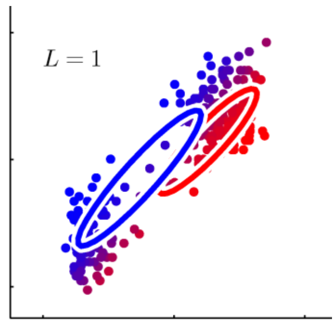
$$\pi_k^{new} = \frac{N_k}{N}$$

$$N_k = \sum_n \gamma_{k,n}$$

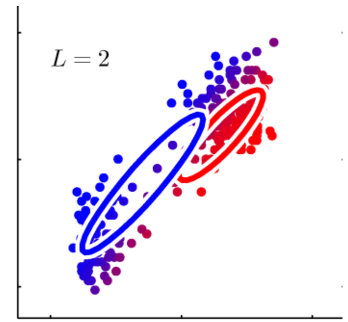
M-step



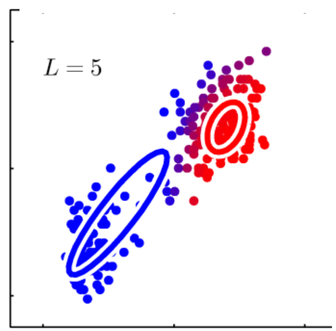
Initialization



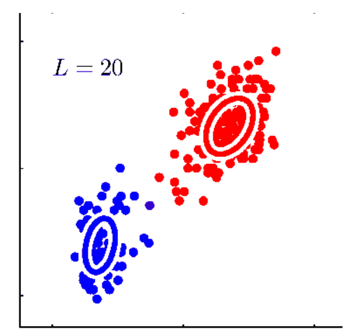
Step 1



Step 2



Step 5

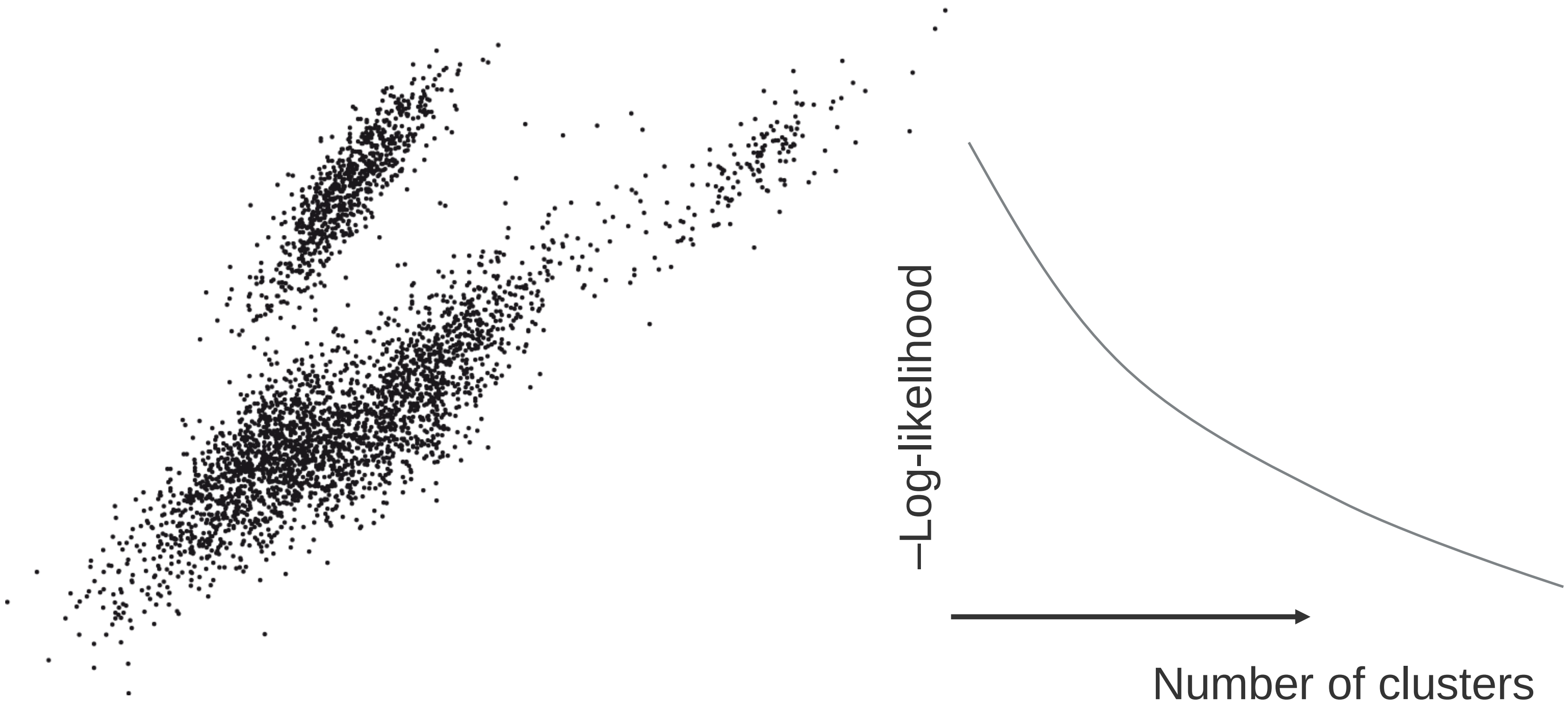


Step 20

Open issues

- How do you determine the number of clusters?
- How do you deal with local maxima?
- How do we work with outliers?
- What about clusters whose mean can drift over time?
- What about spikes occurring simultaneously?

How to determine the number of clusters?

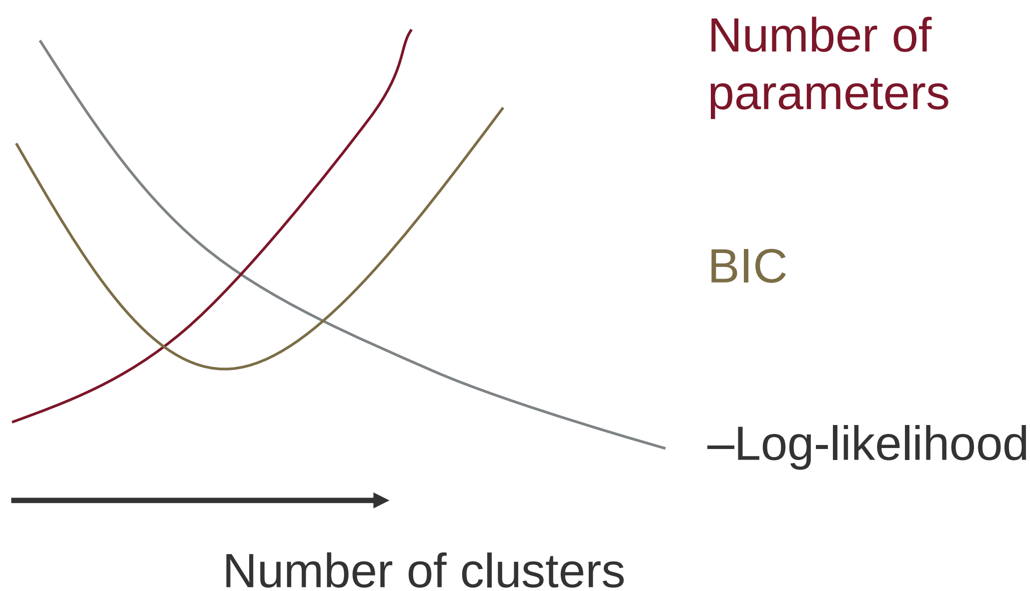


Dealing with model complexity

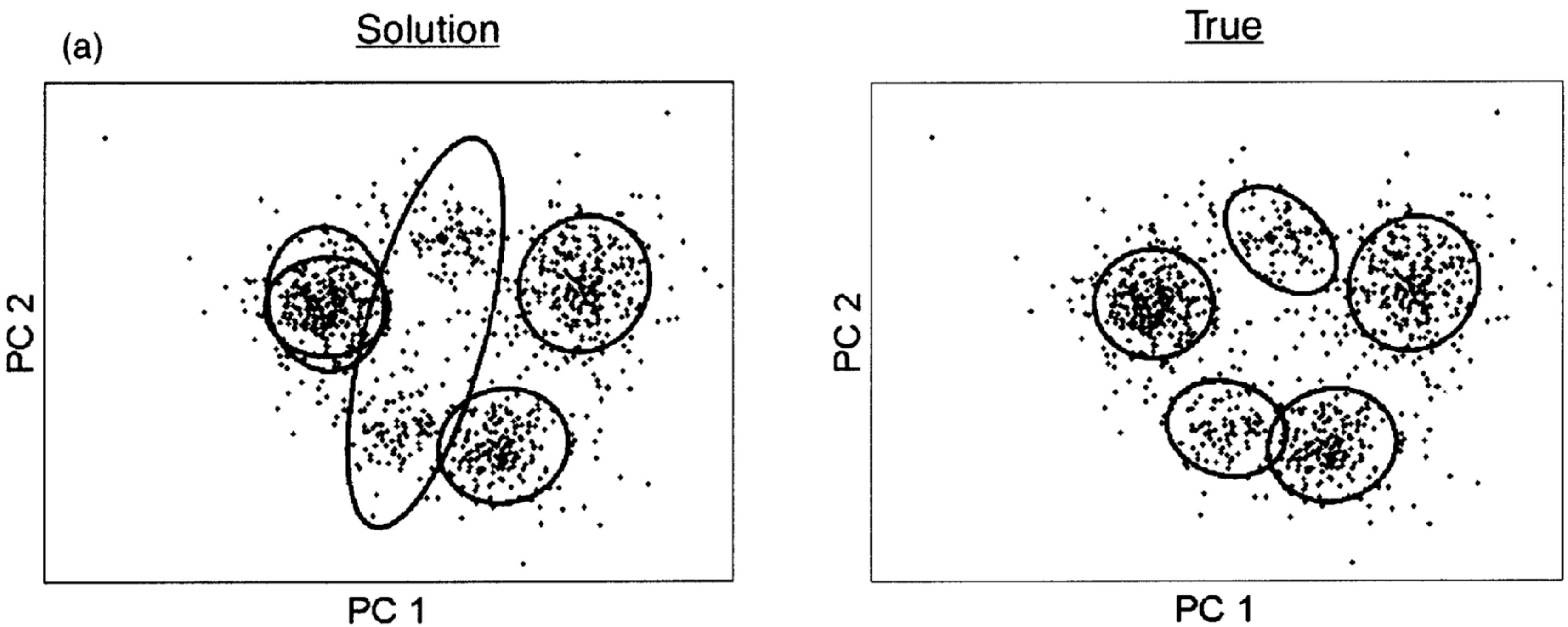
Penalize number of parameters!

Use Bayesian Information Criterion (or similar)

$$BIC = -2\ln p(x|\mu_{opt}, \Sigma_{\mathrm{opt}}, \pi_{opt}) + P\ln N$$



Dealing with local maxima

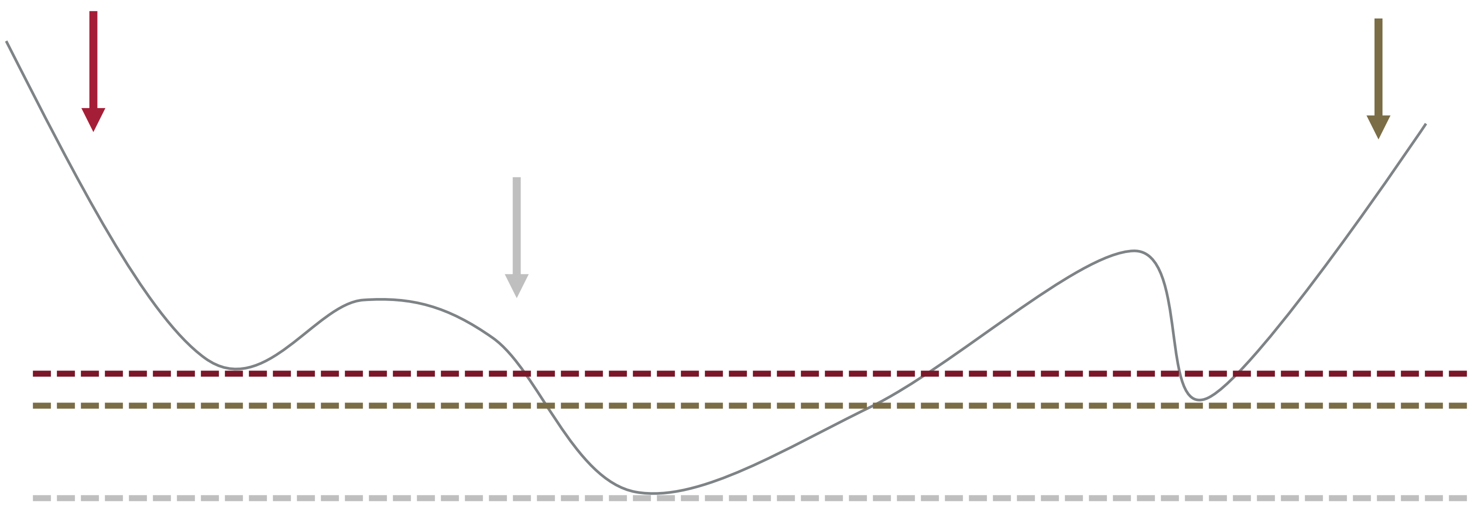


Possible solutions:

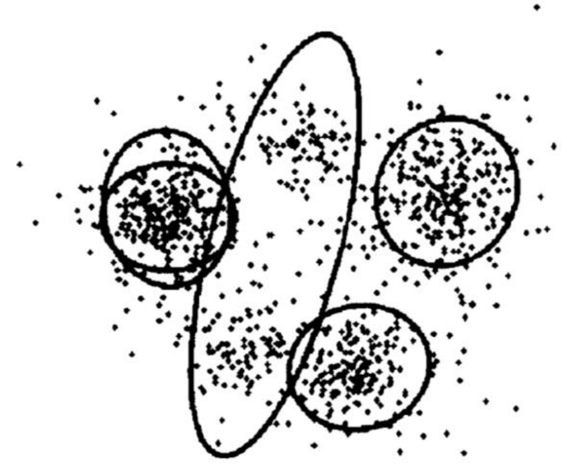
a) Pick the best result among multiple independent runs
b) Split-and-merge

Shoham et al. 2003

Restarting to Avoid local Minima

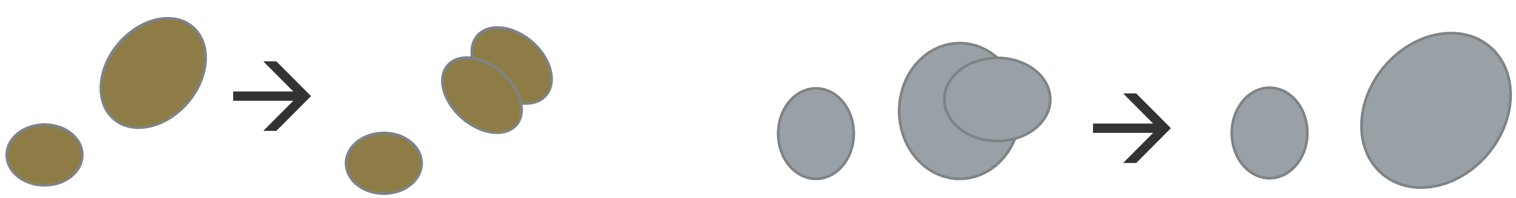


Split and Merge



From Shoham et al. 2003

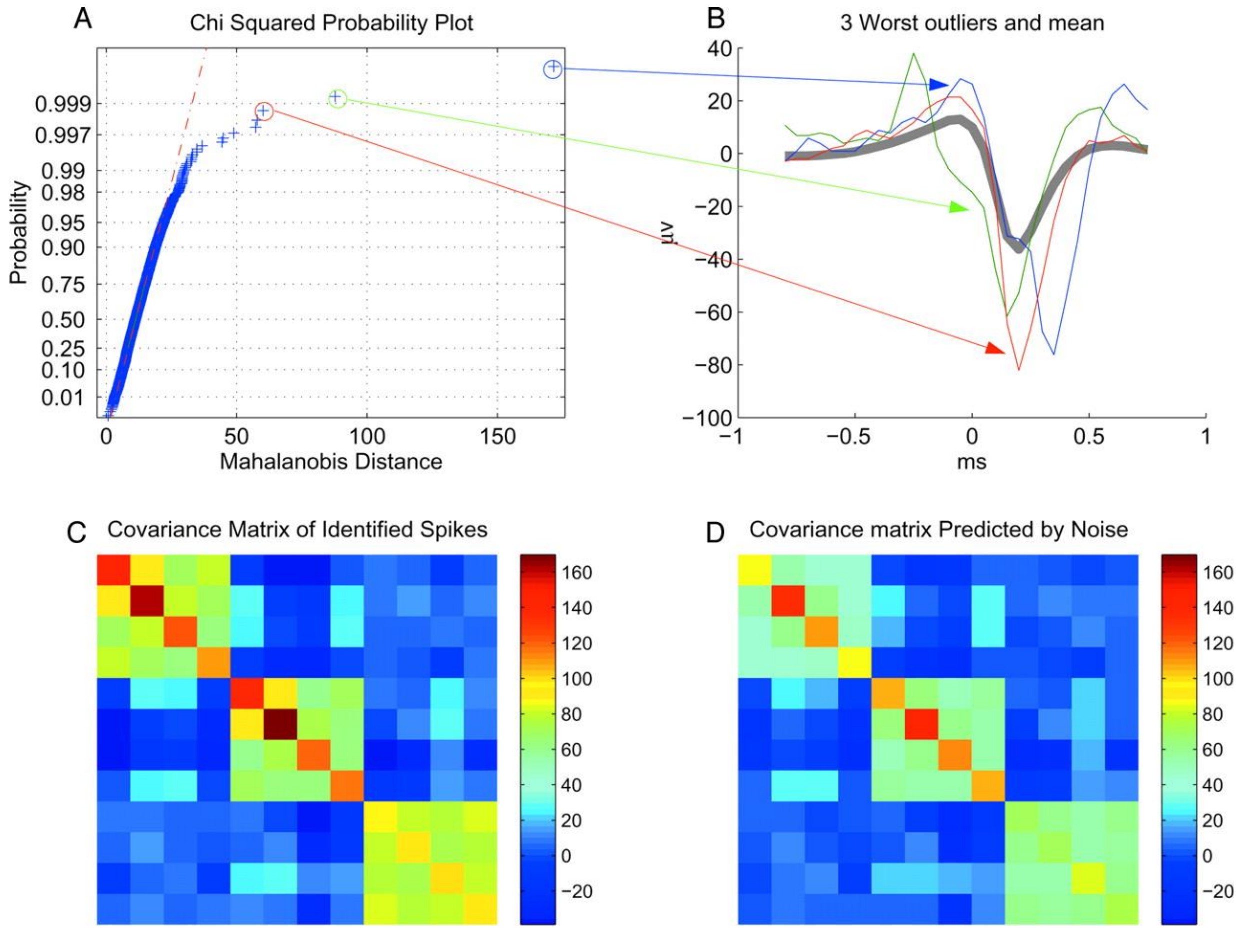
- After EM has converged, test whether BIC is decreased by:
- Splitting a cluster
  For a candidate, run ‚partial EM' with two components instead of one, keeping all other components fixed
- Merging two clusters
  Compute new model with two clusters combined into one



Modified from Ueda et al. 2002, Neural Computation

Problem: Outliers

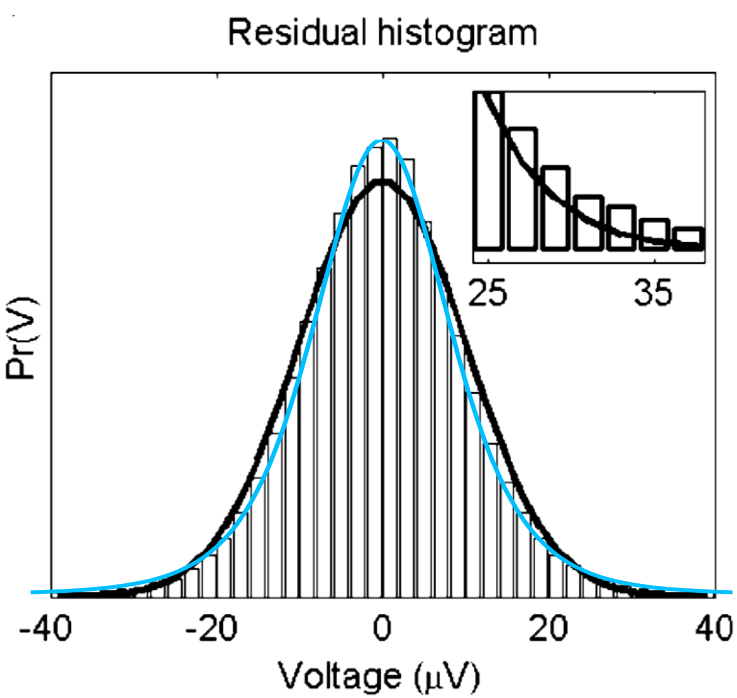
$$D = \sqrt{(x-\mu)^T C^{-1} (x-\mu)}$$

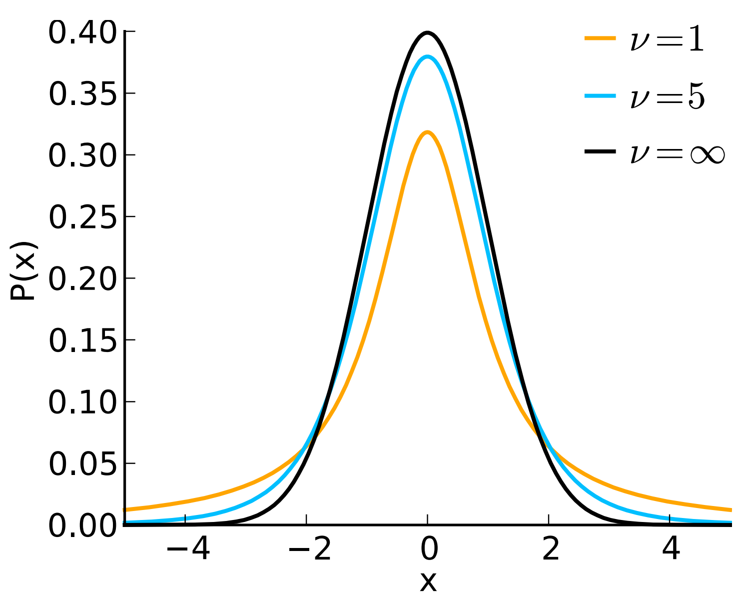


Kenneth D. Harris et al. J Neurophysiol 2000;84:401-414

Problem: Outliers

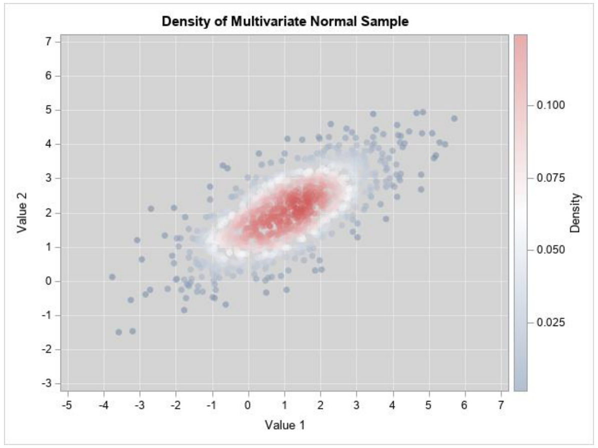
Replace Gaussian mixtures by multivariate t-distributions

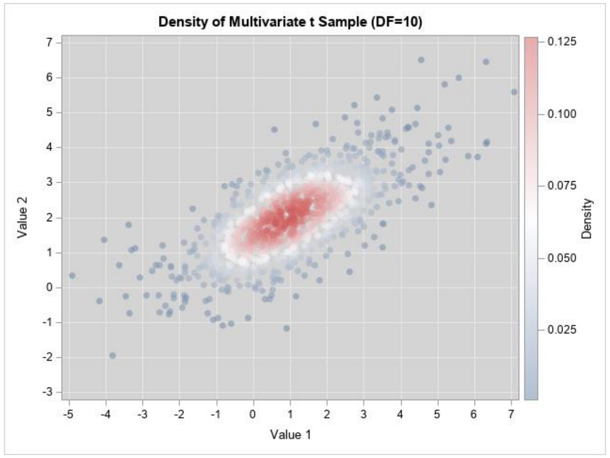




Shoham et al. 2003, Journal of Neuroscience Methods

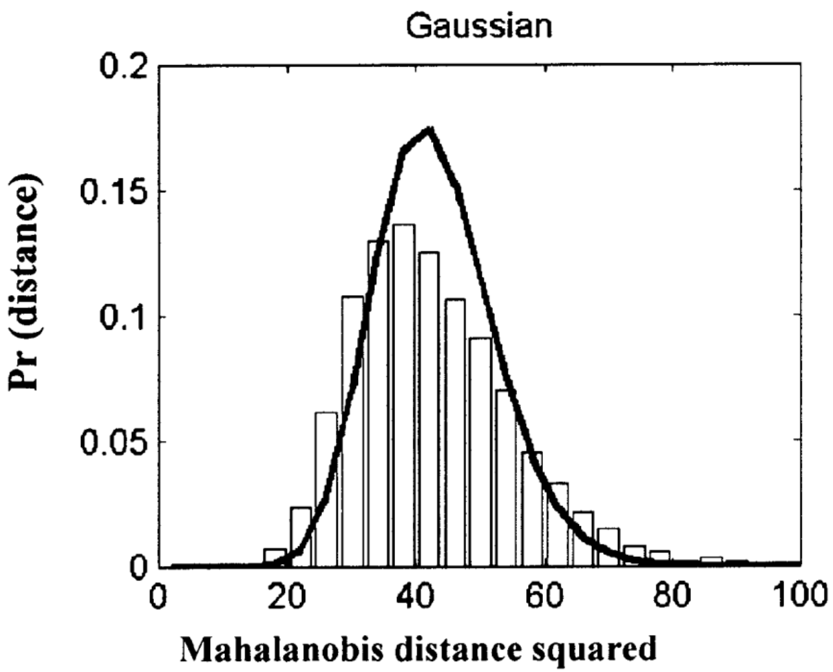
Problem: Outliers

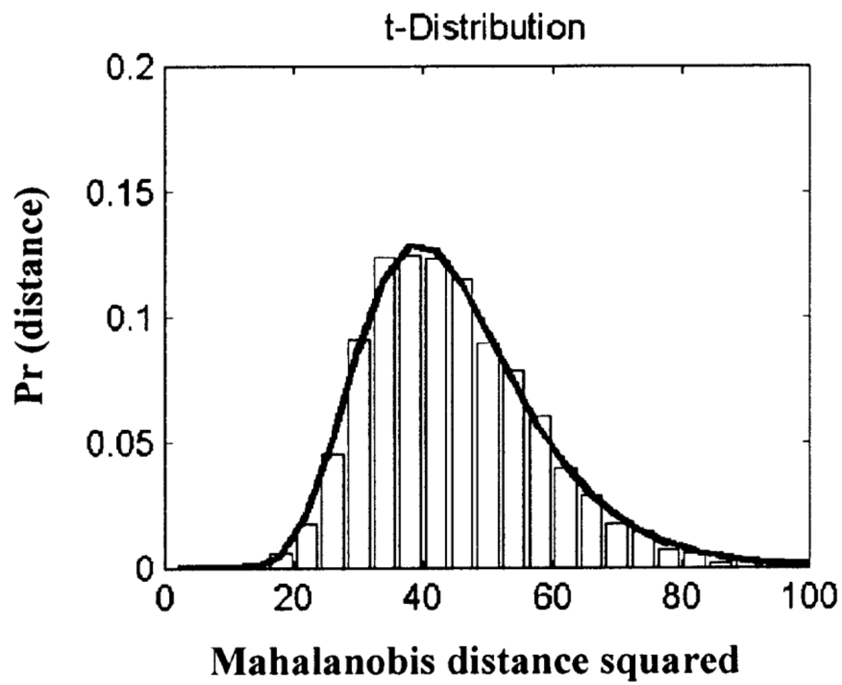




$$f(x) = \frac{\Gamma((\nu+p)/2)}{\Gamma(\nu/2)(\nu\pi)^{p/2}|\Sigma|^{1/2}} \left(1 + \frac{1}{\nu}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)^{-(\nu+p)/2}$$

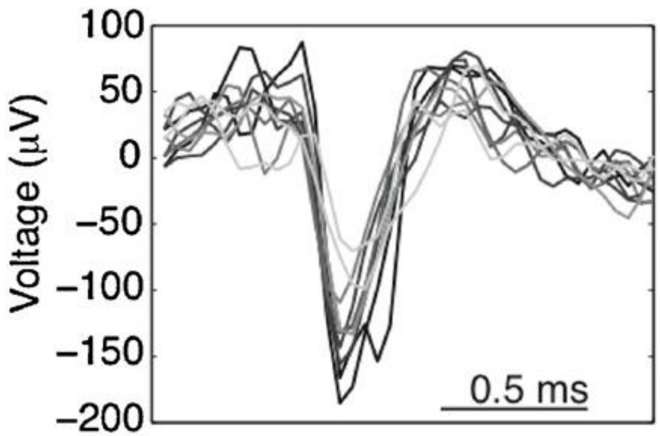
t-distributions fit data better



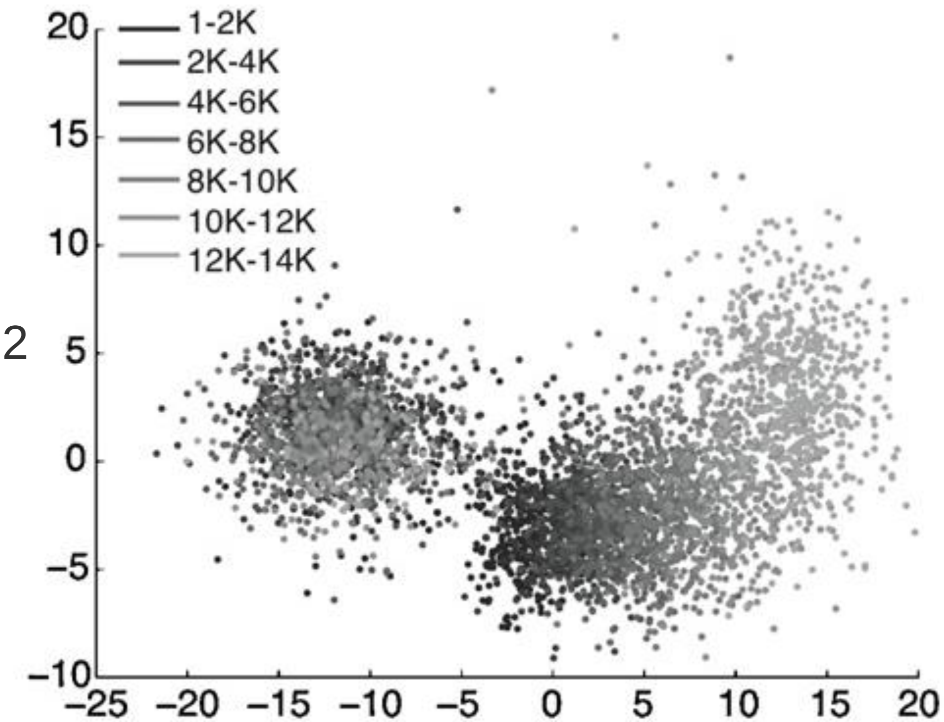


Shoham et al. 2003, Journal of Neuroscience Methods

Problem: Waveform drift



PC 2



PC 1

Calabrese & Paninski 2011
Journal of Neuroscience Methods

Kalman filter mixture model

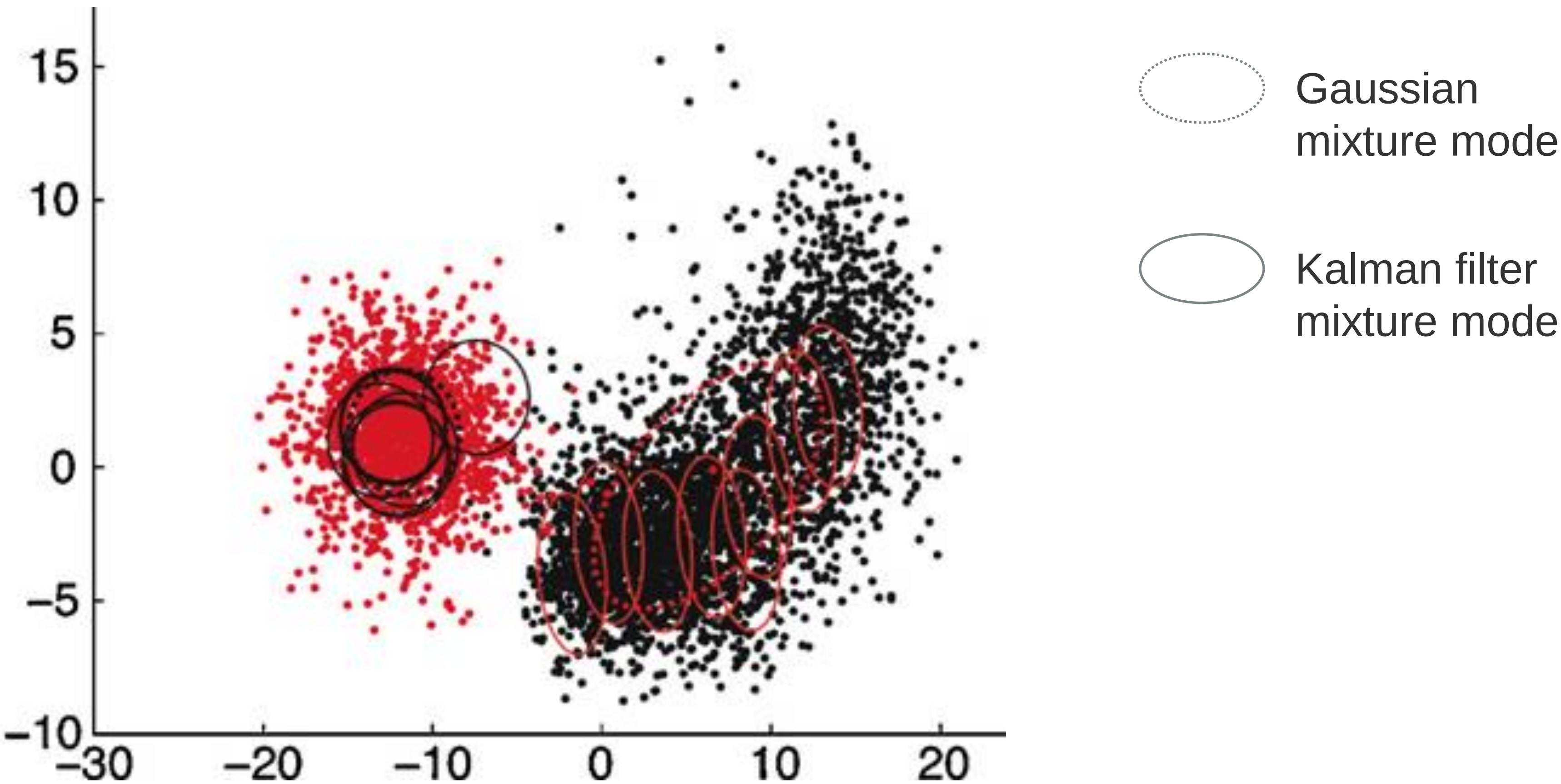
- Take a standard Gaussian mixture model
- Let cluster means μ_j drift over time

$$\mu_j^{t+1} = \mu_j^t + \epsilon_j^t, \qquad \epsilon_j^t \sim \mathcal{N}\left(0, C_j\right)$$

- Update of μ in M step via standard forward-backward Kalman recursion

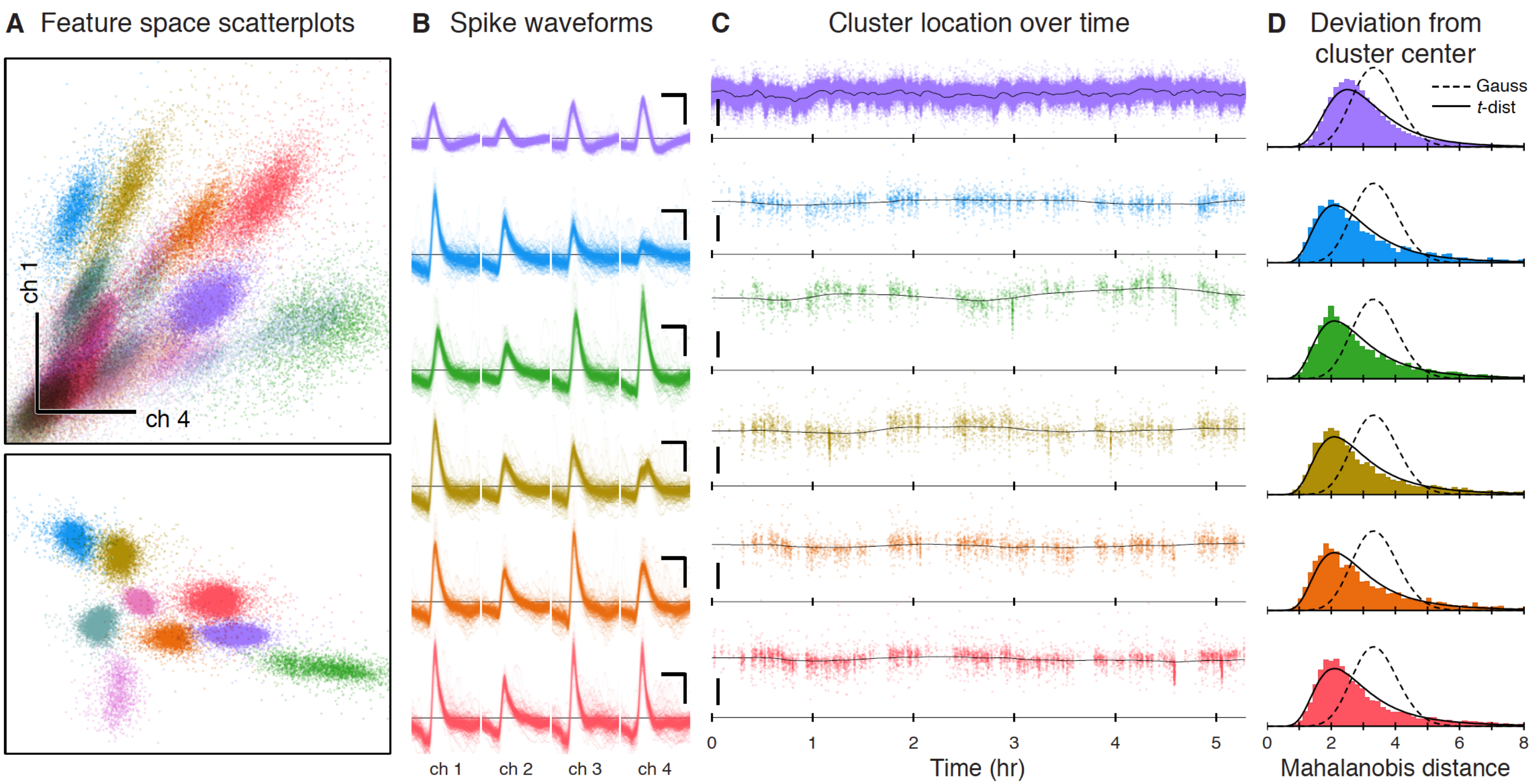
Calabrese & Paninski 2011, Journal of Neuroscience Methods

Kalman filter mixture model



Calabrese & Paninski 2011
Journal of Neuroscience Methods

Drifting mixture of t-distributions

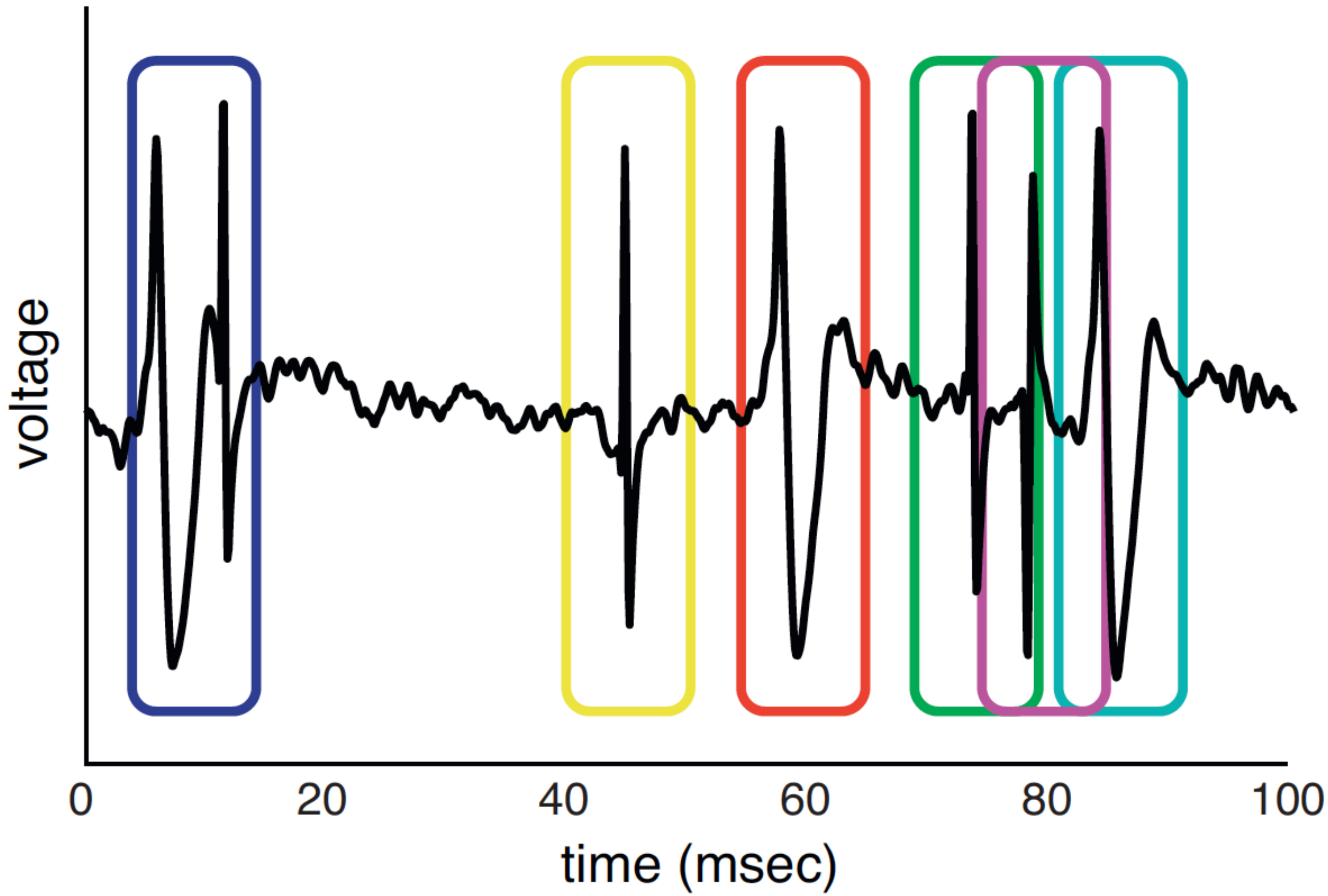


Shan et al. 2017, bioRxiv

Drifting mixture of t distributions

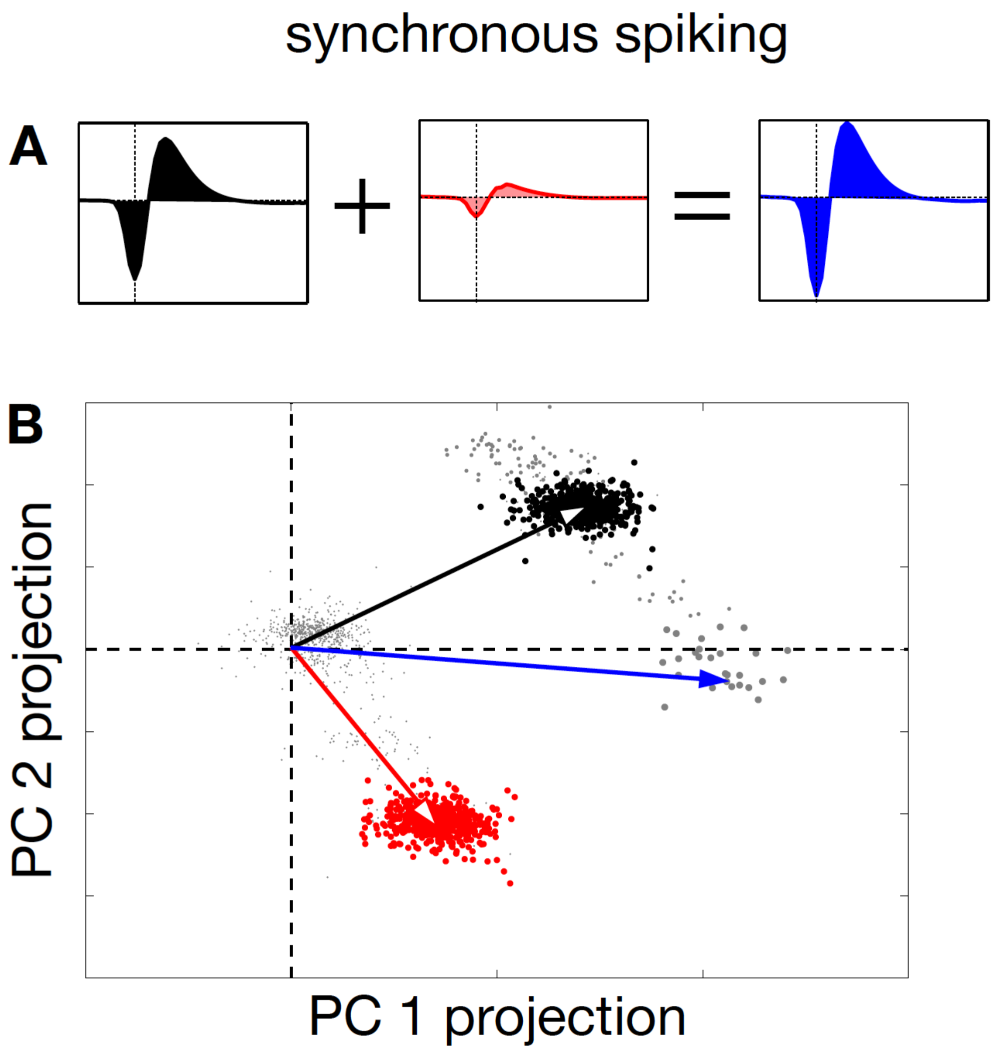
- Ecker et al. (2014): State dependence of noise correlations in macaque primary visual cortex. *Neuron*.
- Shan, Lubenov, Siapas (2017): Model-based spike sorting with a mixture of drifting t-distributions. *bioRxiv*.
- Code: https://github.com/aecker/moksm

Problem: Synchronous spikes



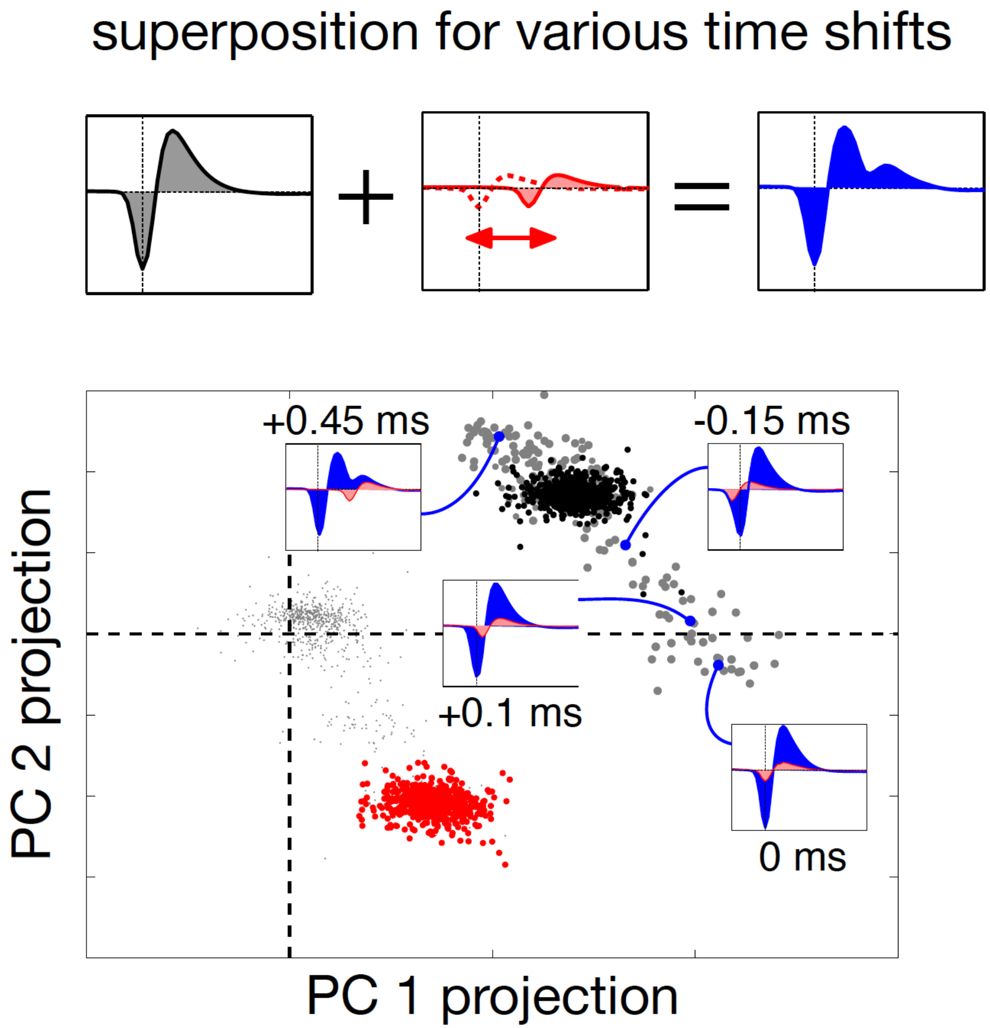
Ekanadham et al. 2014

Problem: Synchronous spikes

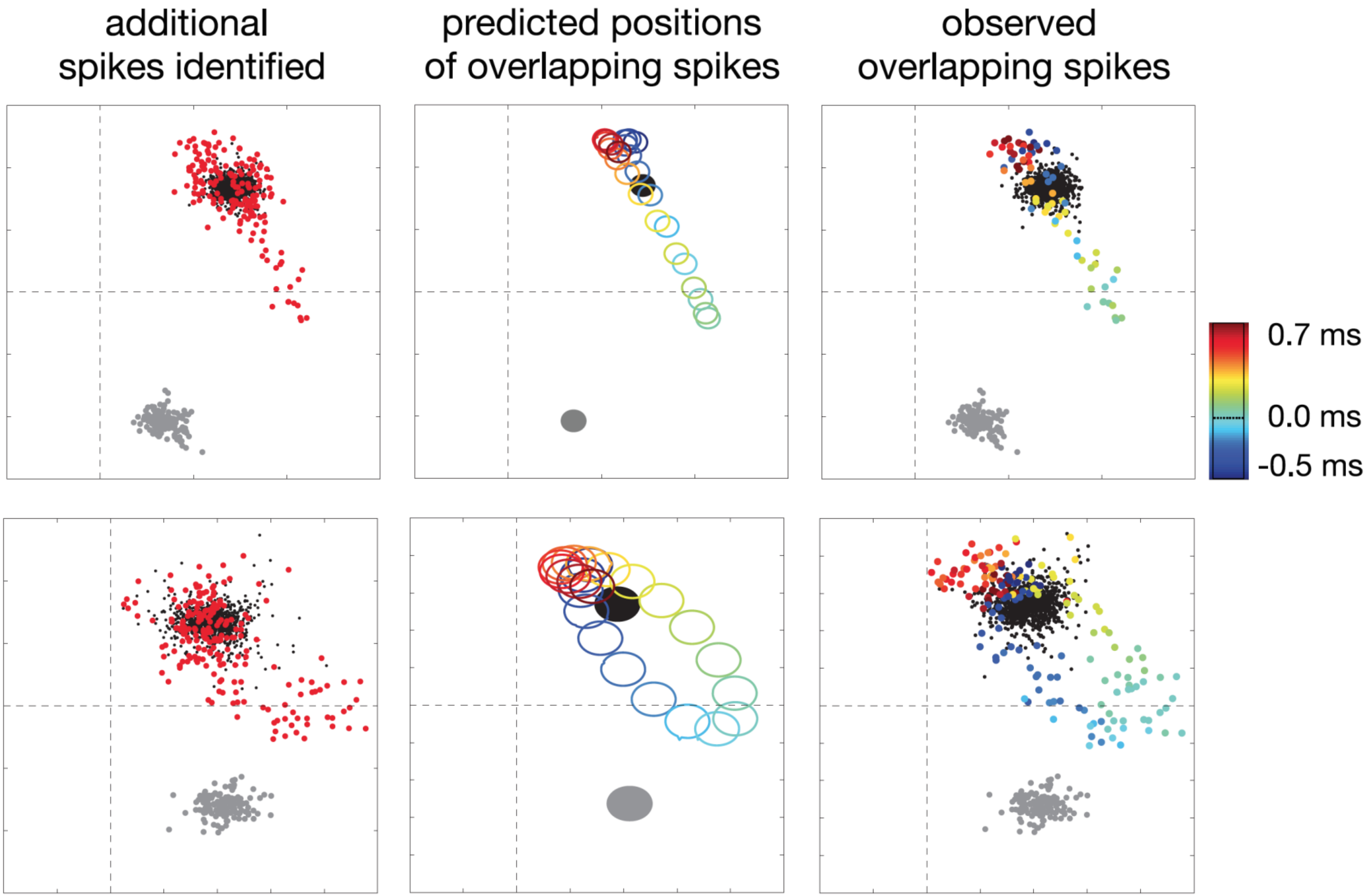


Pillow et al. 2013

Problem: Synchronous spikes

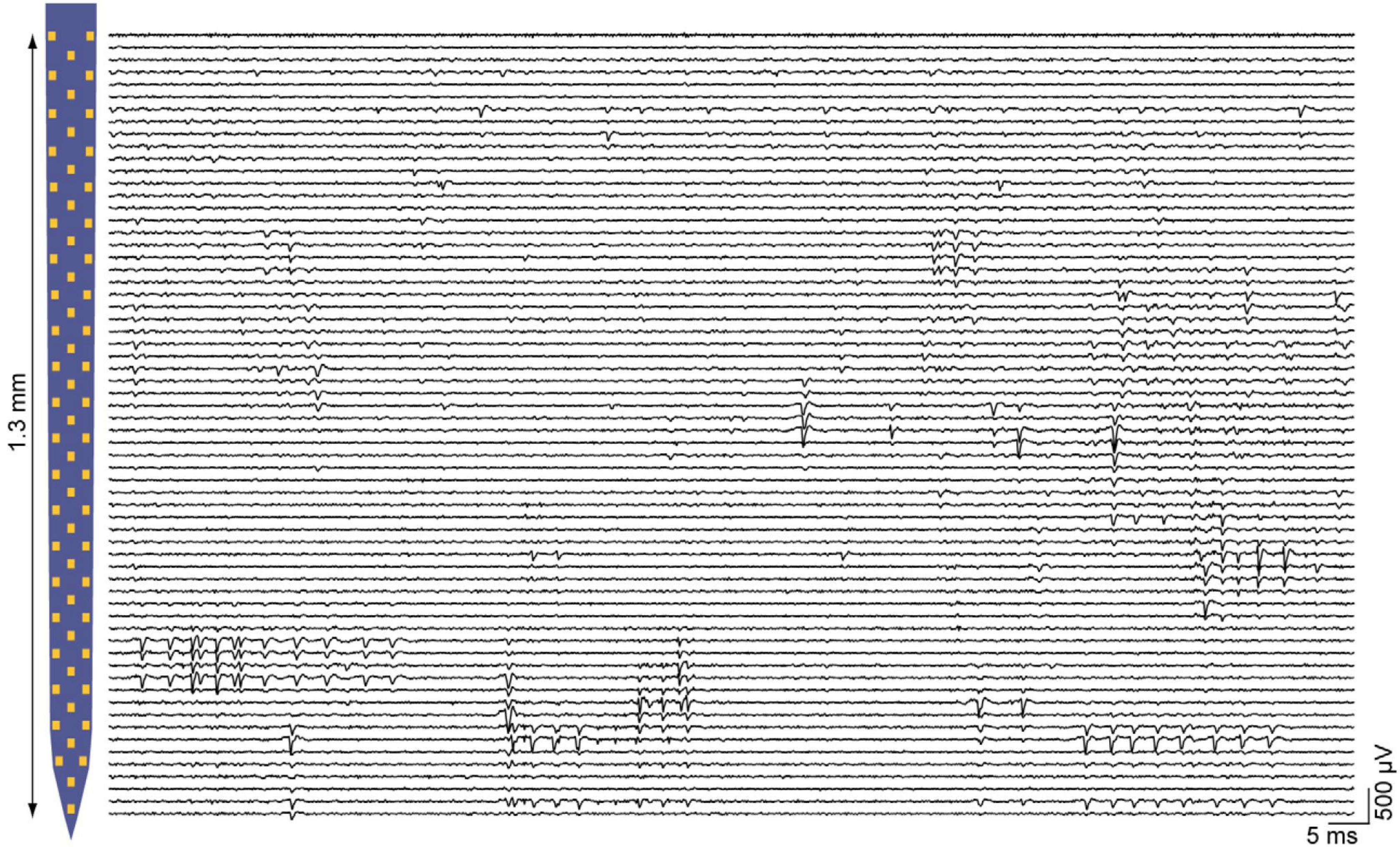


Pillow et al. 2013



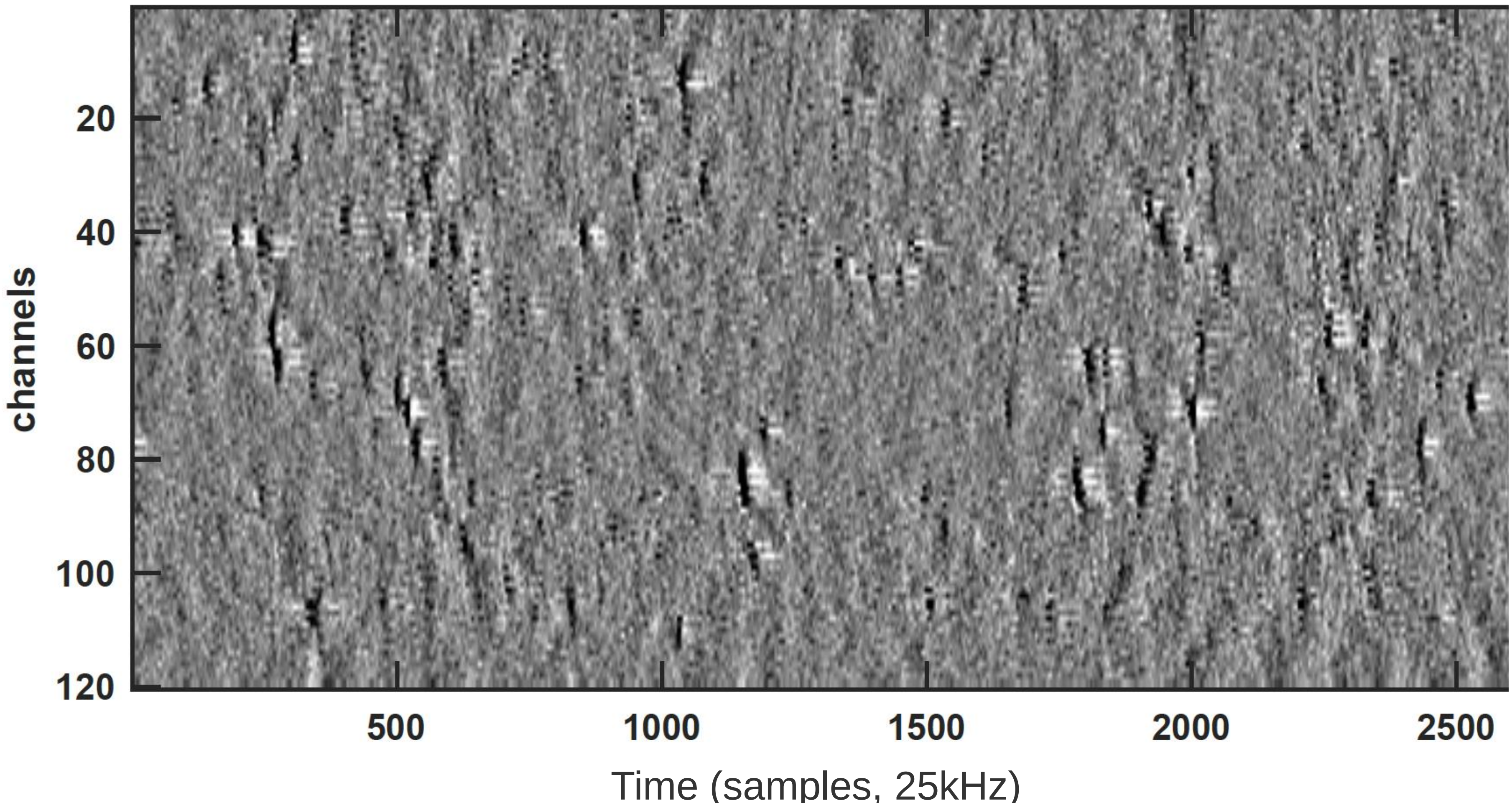
Pillow et al. 2013

Multi-channel silicon probes



Du et al. 2011

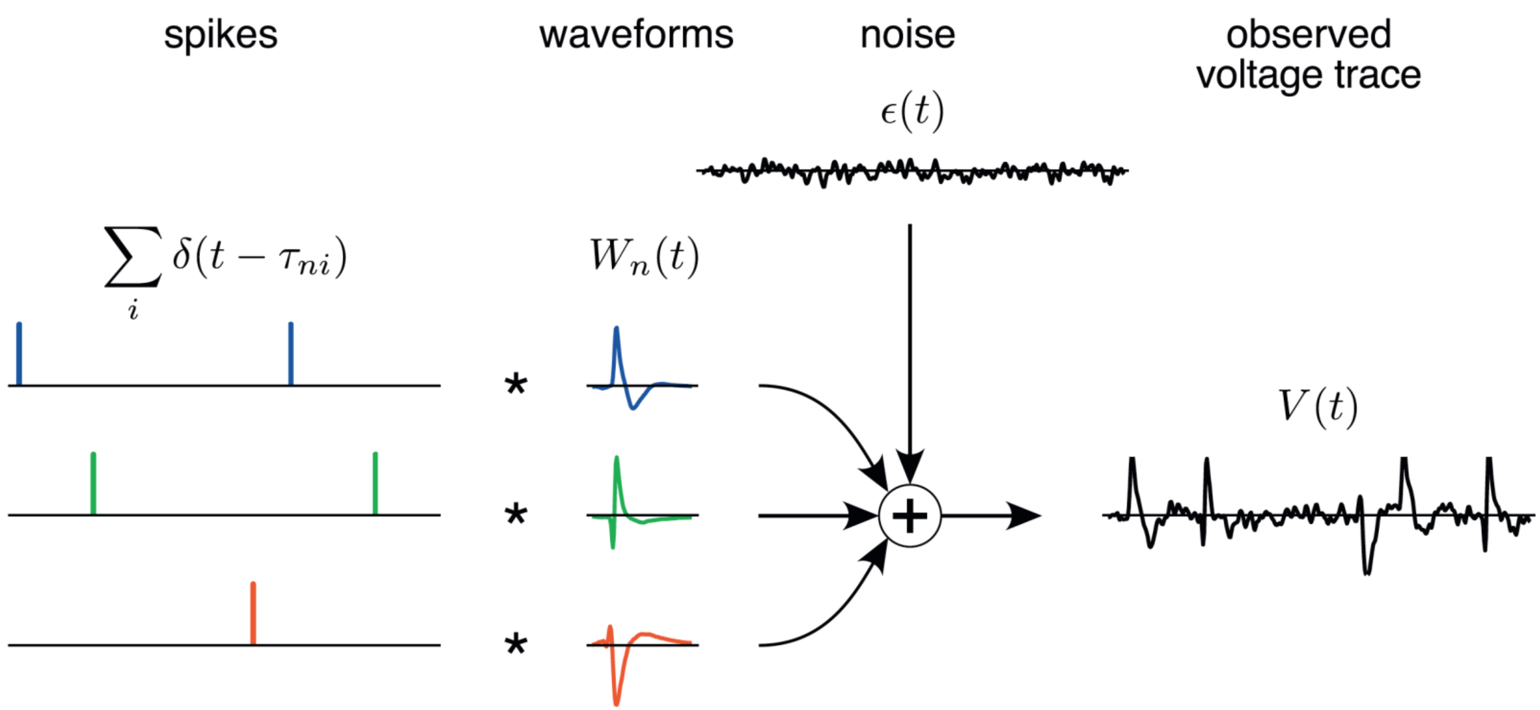
Multi-channel silicon probes



Pachitariu et al. 2017

Generative model

$$V(t) = \sum_n X_n(t) * W_n + \epsilon(t) \qquad \epsilon \sim \mathcal{N}(0, \Lambda) \qquad X_n \sim \text{Bernoulli}(p_n)$$



Pillow et al. 2013, Ekanadham et al. 2014

Generative model

$$V(t) = \sum_n X_n(t) * W_n + \epsilon(t)$$

Solve by alternating:

- Given spike times X(t): estimate waveforms by linear regression
- Given waveforms W: determine spike times by template matching

  → Binary pursuit

Binary Pursuit

Generative model:

$$V(t) = \sum_n X_n(t) * W_n + \epsilon(t) \qquad \begin{array}{l} \epsilon \sim \mathcal{N}(0, \Lambda) \\ X_n \sim \text{Bernoulli}(p_n) \end{array}$$

Loss function:

$$L(\mathbf{X},\mathbf{W}) = \left[\frac{1}{2}(\mathbf{V} - \mathbf{W} * \mathbf{X})^T \Lambda^{-1} (\mathbf{V} - \mathbf{W} * \mathbf{X})\right] + \gamma^T \mathbf{X},$$

Change in loss function after adding (deleting) spike *i* to (from) *X*

$$\Delta \tilde{L}_i = \mathbf{V}^T \mathbf{w}_i - \mathbf{w}_i^T M_w^{\setminus i} \mathbf{X}^{\setminus i} - \gamma_i - \frac{1}{2} \mathbf{w}_i^T \mathbf{w}_i,$$

Greedy algorithm

Repeat until $\Delta L_i < 0$ everywhere:

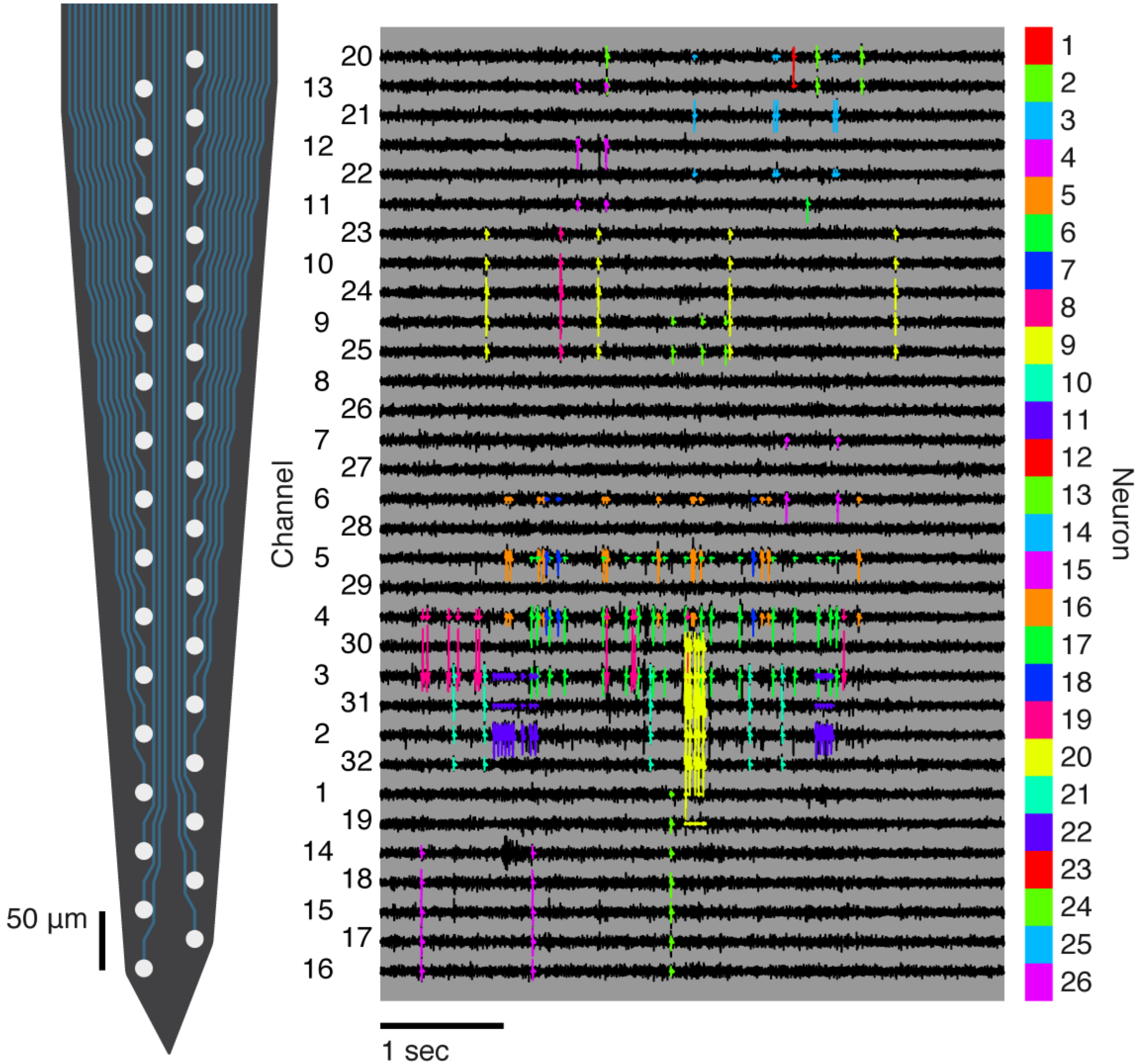
1. Flip X_i for the bin *i* for which $(1 - 2X)\Delta L_i$ is largest
2. Update ΔL_i

Binary Pursuit

Initialize by traditional clustering. Then repeat until convergence:

1. Estimate waveforms *W* by linear regression given voltage data *V* and current spike train estimate *X*
2. Compute the residuals $R = V - W * X$ and estimate noise covariance $\Lambda = \mathrm{cov}(R)$
3. Whiten by the square root of the inverse covariance:
$$\tilde{V} = \Lambda^{-0.5} V \qquad \widetilde{W} = \Lambda^{-0.5} W$$
4. Estimate prior spike probabilities for each neuron:
$$\hat{p}_j = \frac{1}{N} \sum_t X_{jt}$$
5. Estimate spike trains *X* via binary pursuit given $\tilde{V}, \widetilde{W}$ and $\hat{p}_j$

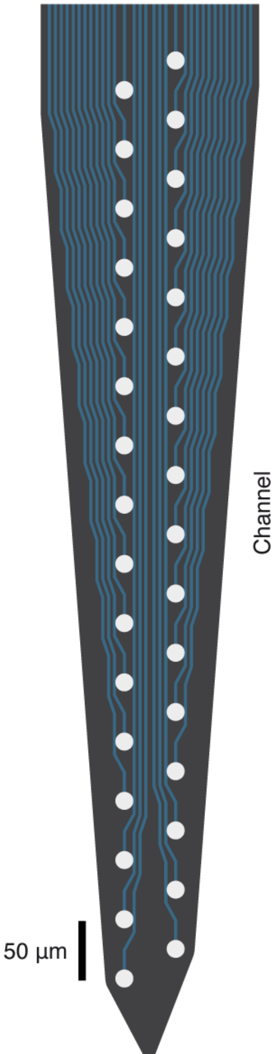
Results: binary Pursuit



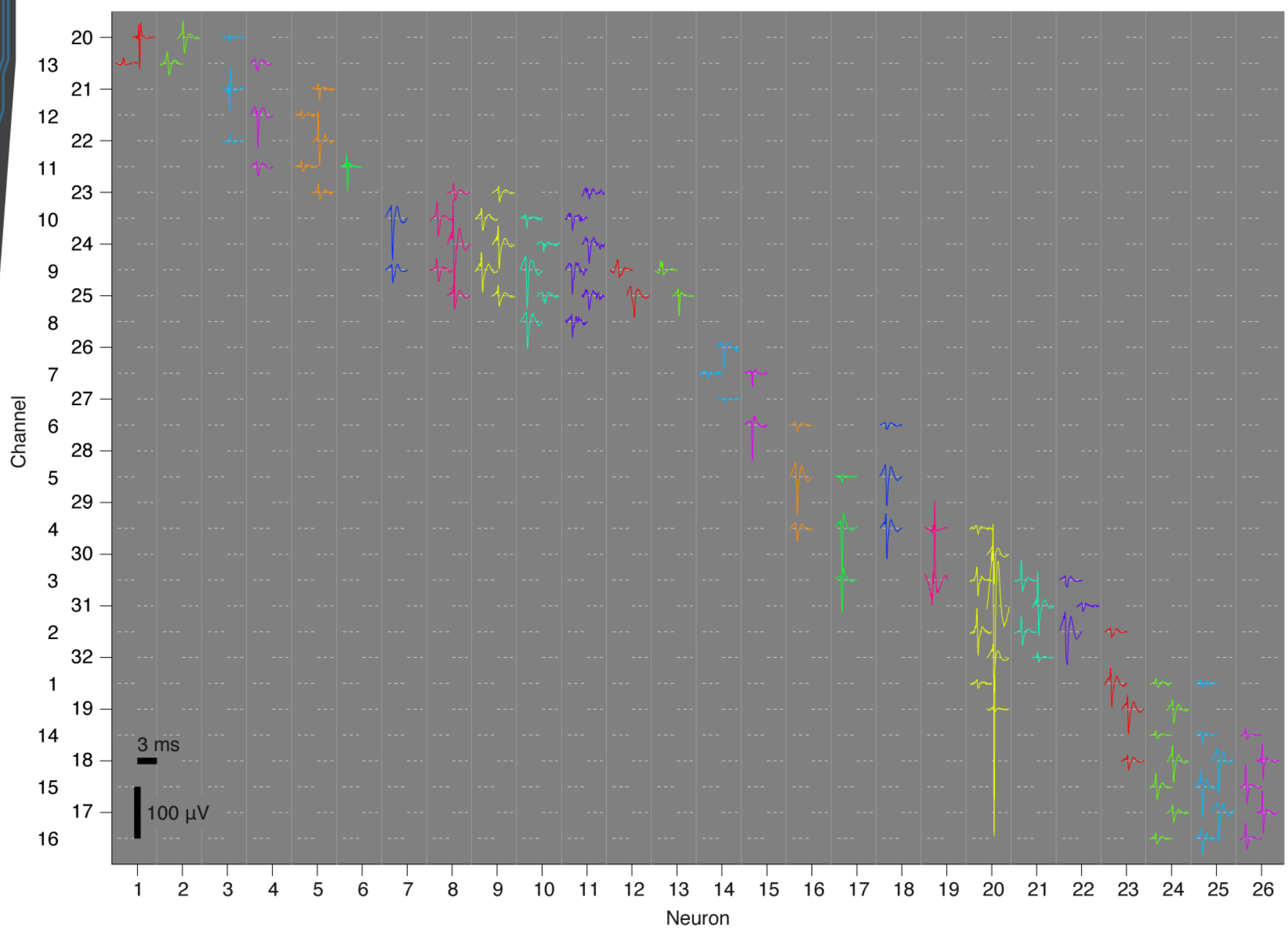
Denfield, Ecker, Tolias, unpublished

Results: Binary Pursuit

V1x32-Poly2



Single neuron waveforms (channel layout as on probe)



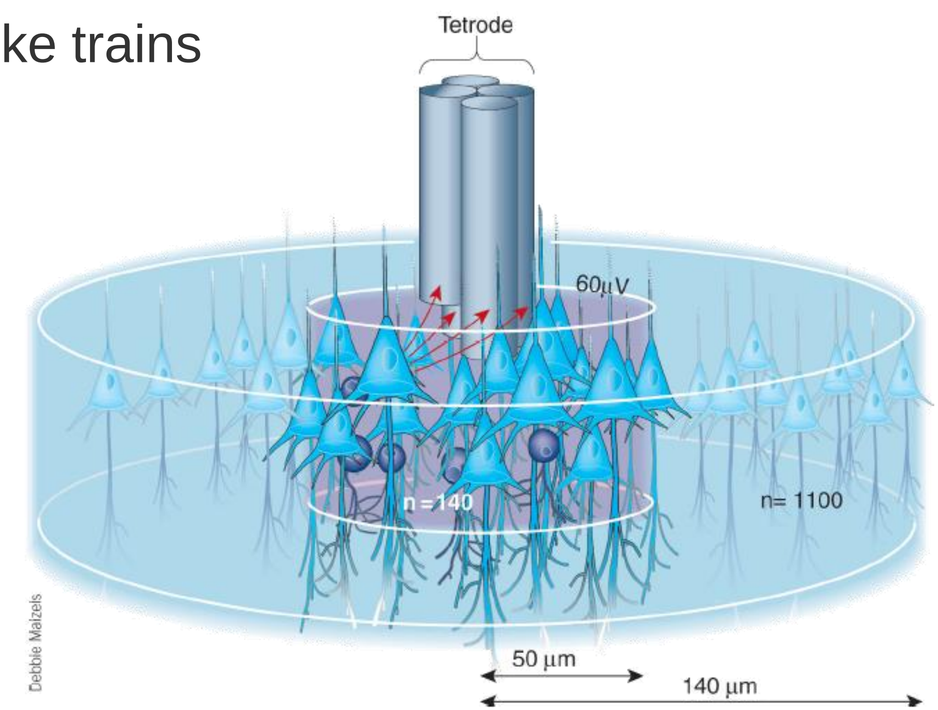
Interpreting clustering results

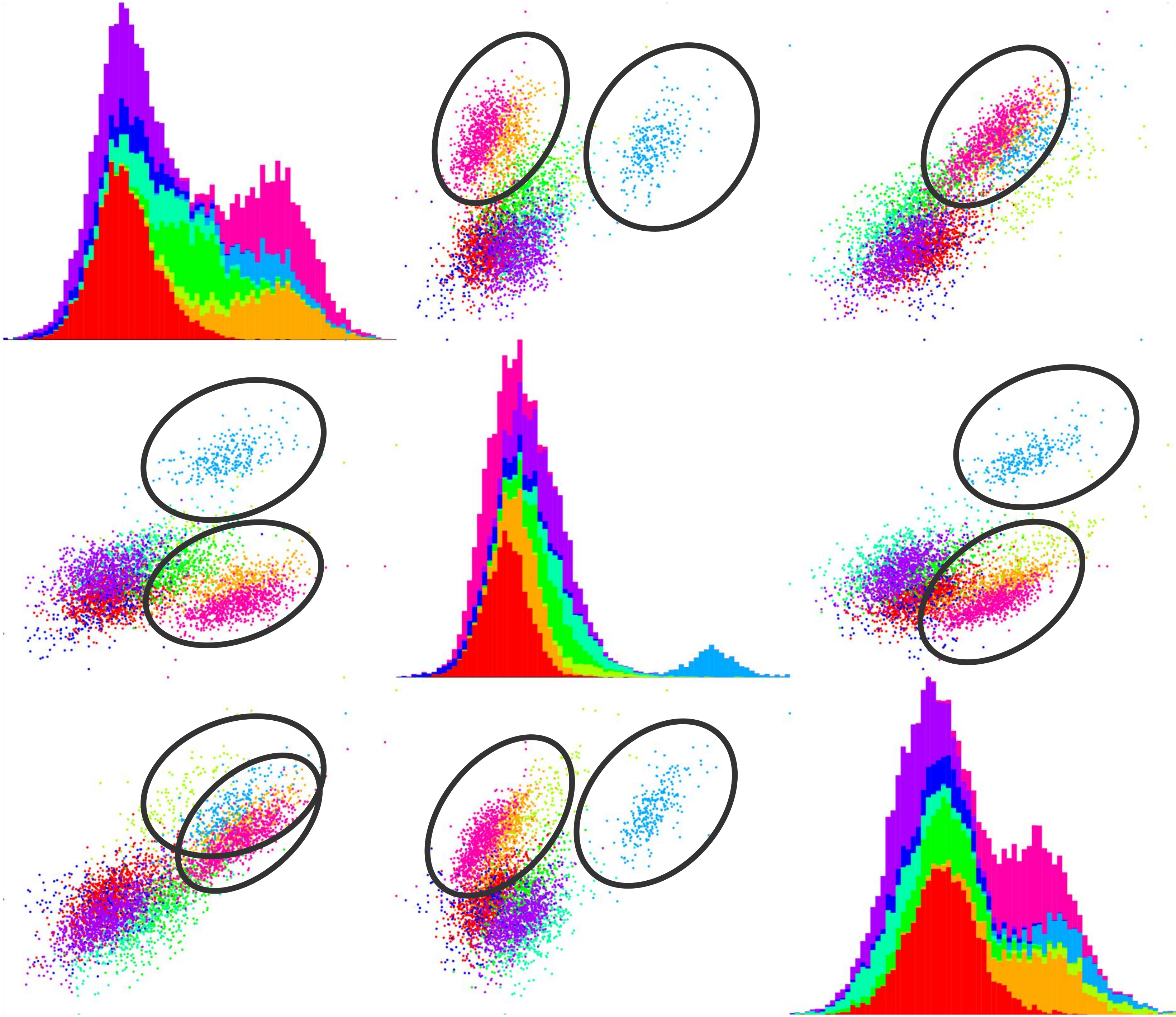
Not all clusters represent the spikes of a single neuron

- ***Single unit:*** a cluster with one neuron's spikes
- ***Multi unit:*** a cluster with multiple neurons' spikes

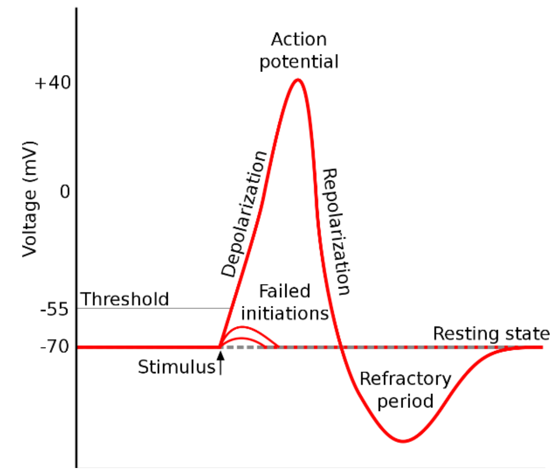
Tools for evaluating clustering:

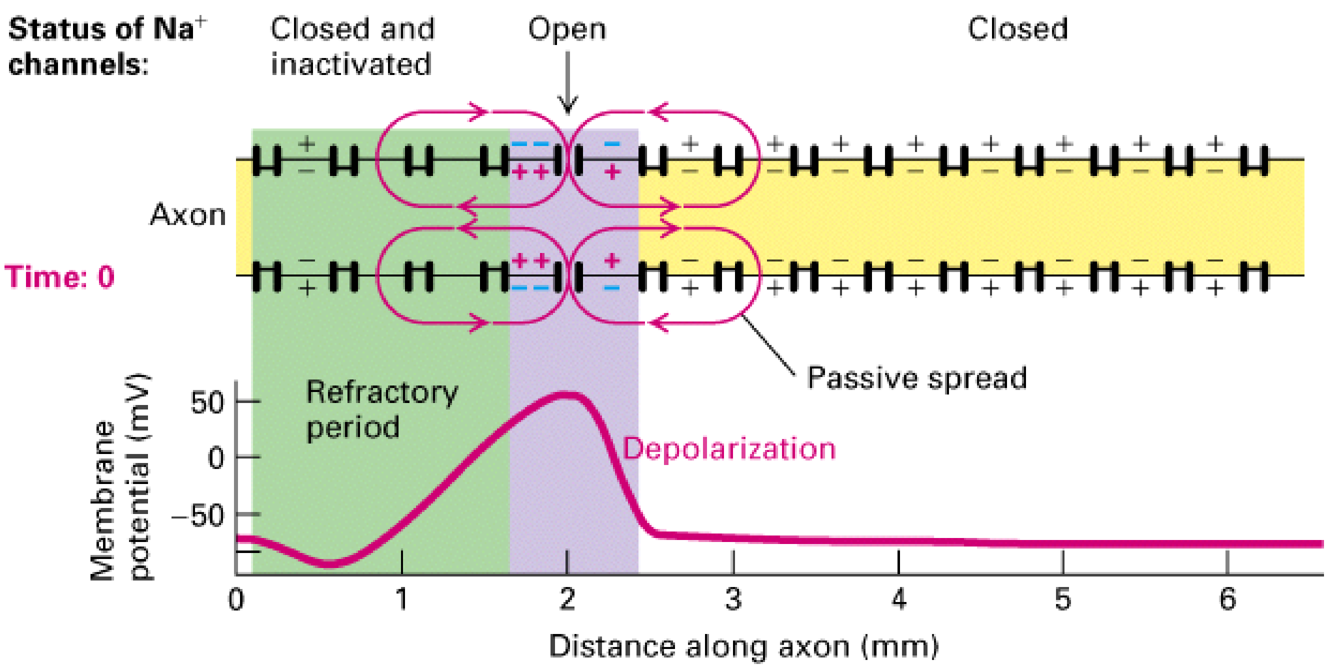
- Temporal structure of the spike trains
  → refractory period
- Quantify cluster separation





Refractory Period



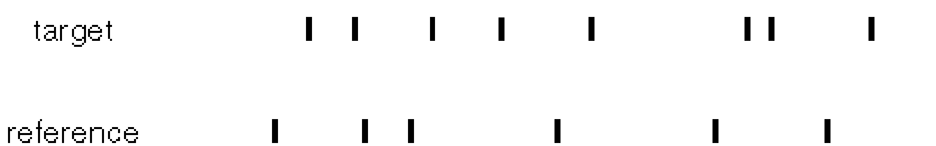


Correlogram

Cross correlation function

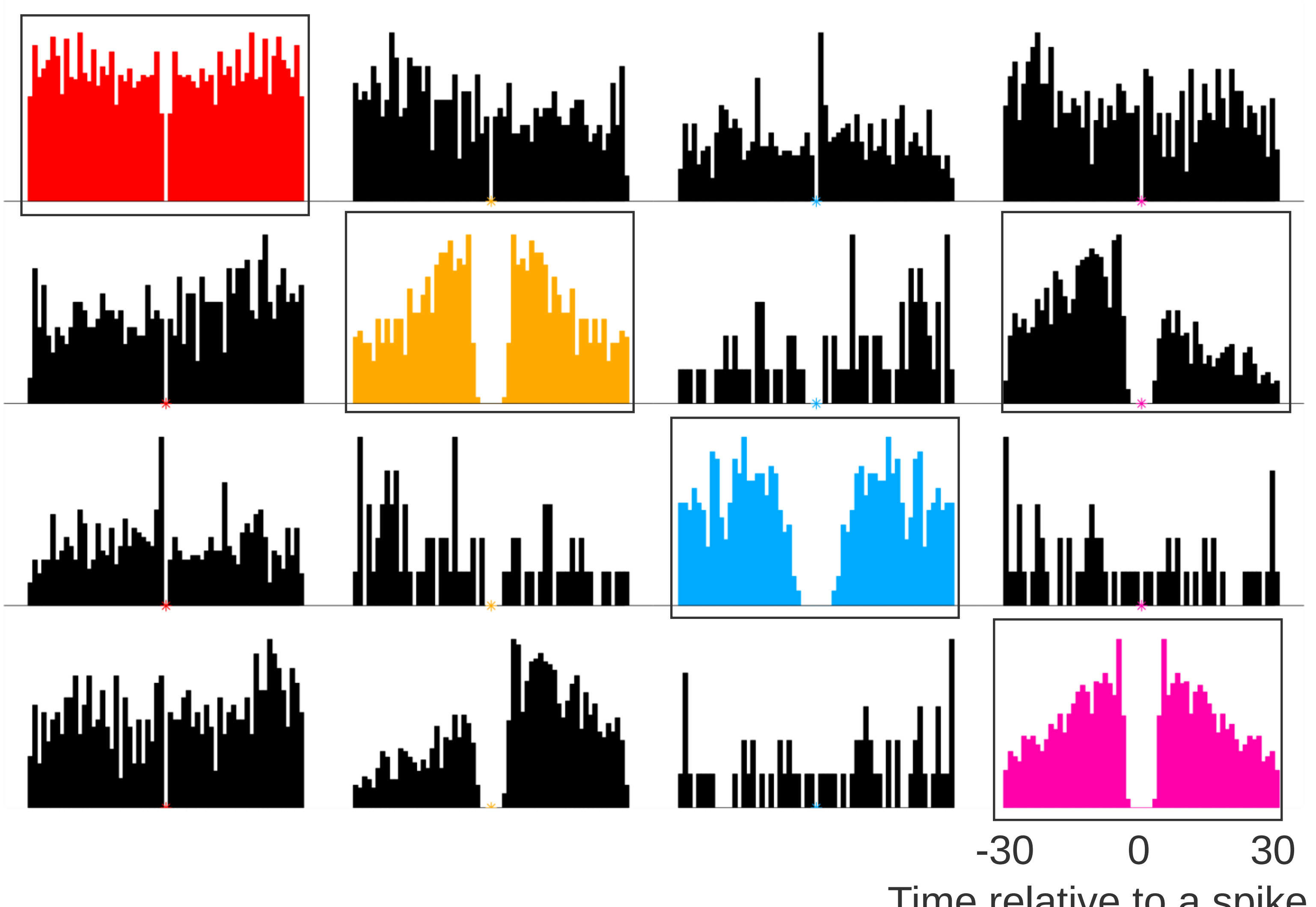
$$R_{uv}(\tau) = \int u(t)v(t-\tau)dt$$

Histogram of $\tau = t_i - t_j$ for all spike times t_i and t_j

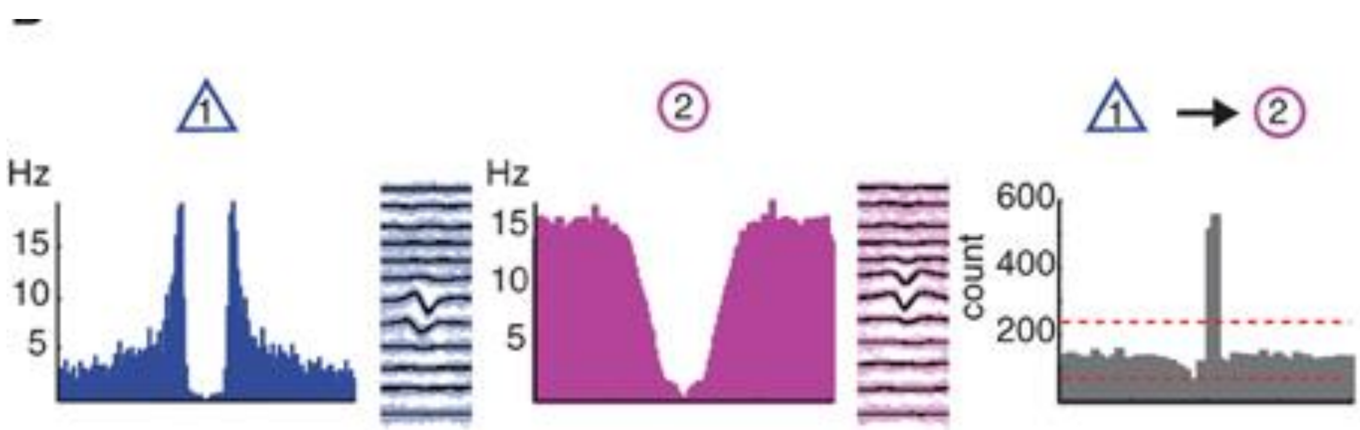


Stern et al. (1998), Nature

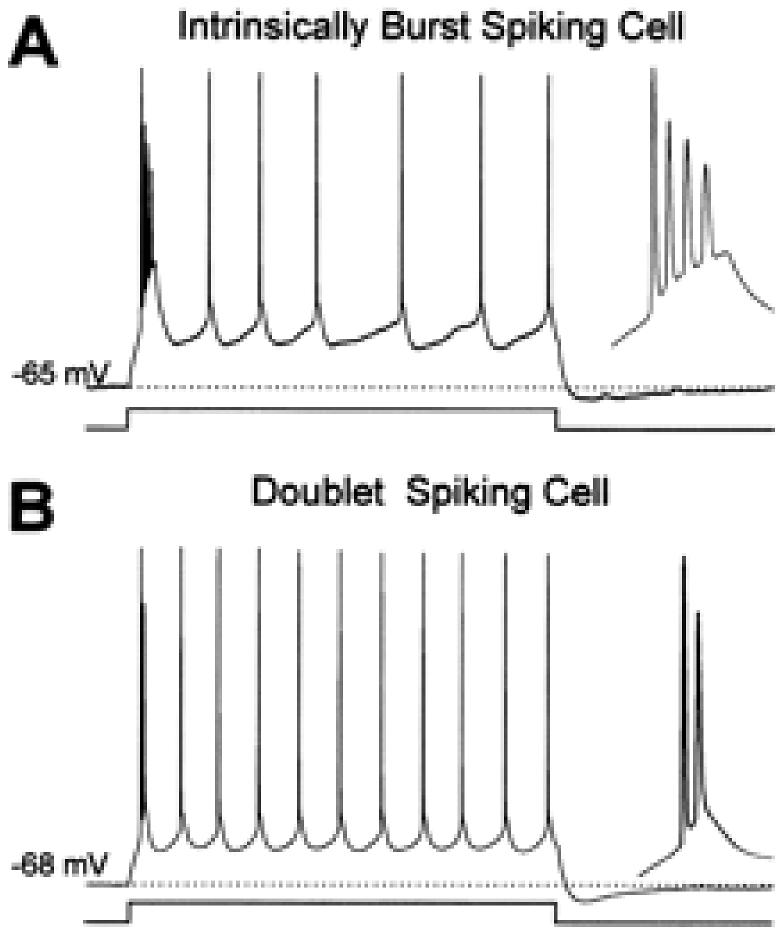
Cross-/Auto-Correlograms



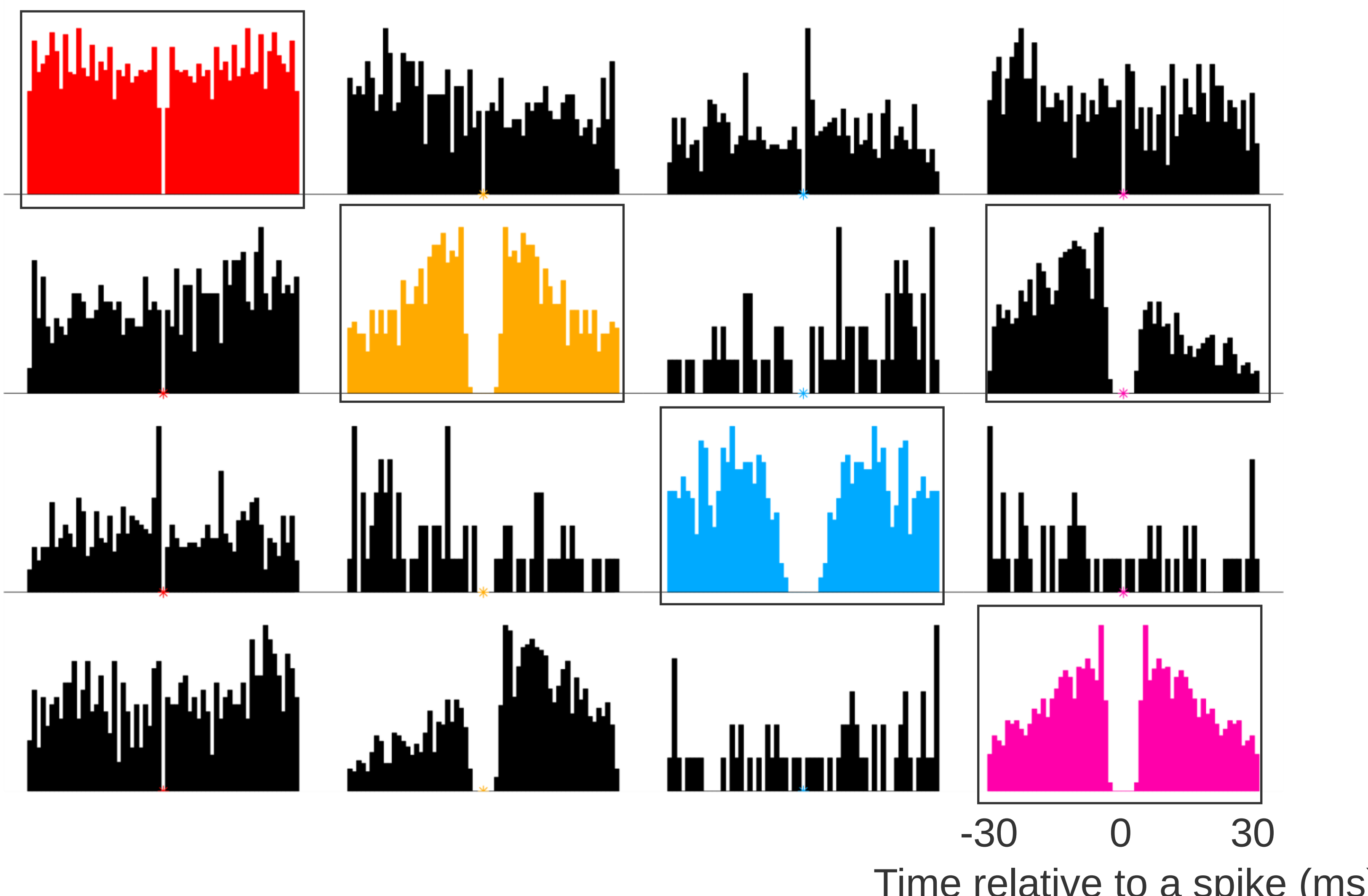
Coupled neurons

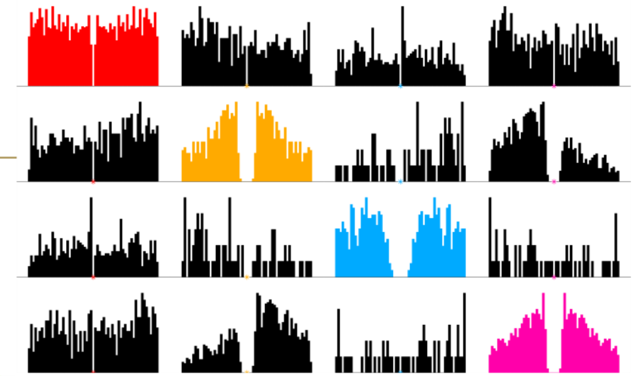


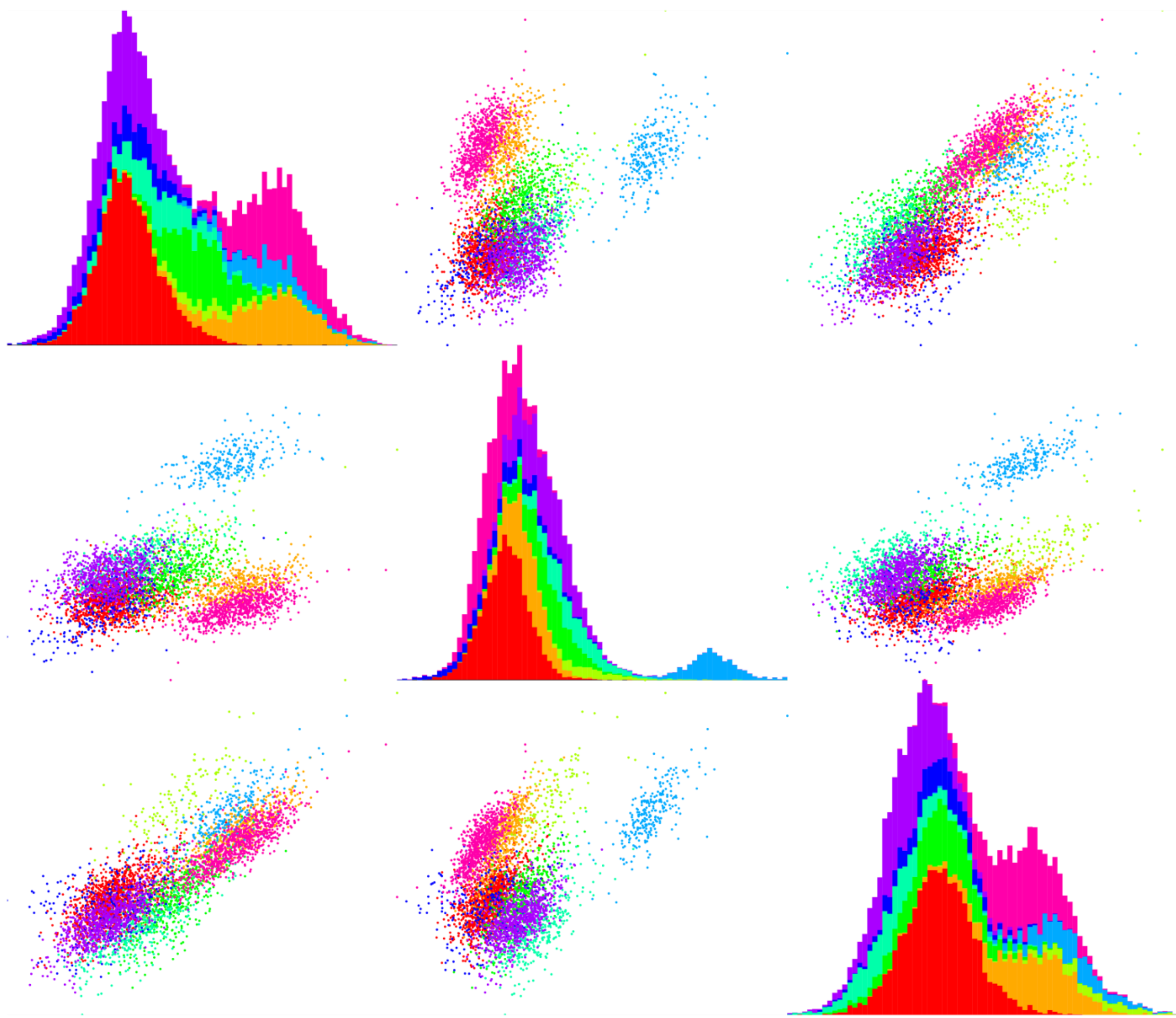
Bursting neuron with spike amplitude adaptation



Cross-/Auto-Correlograms



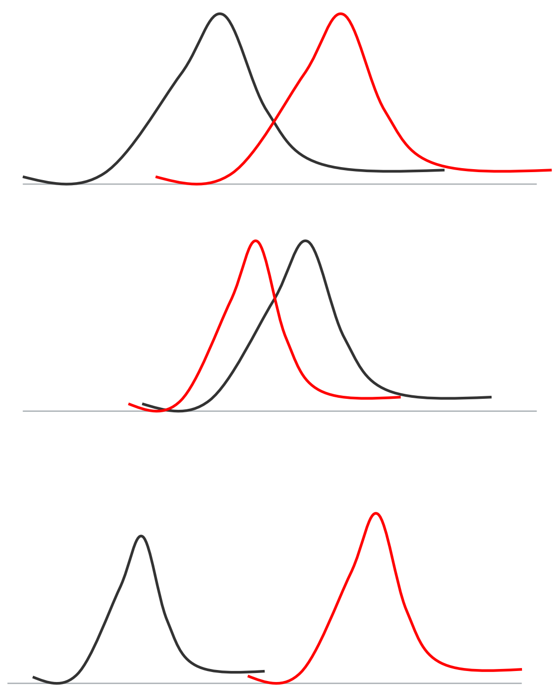


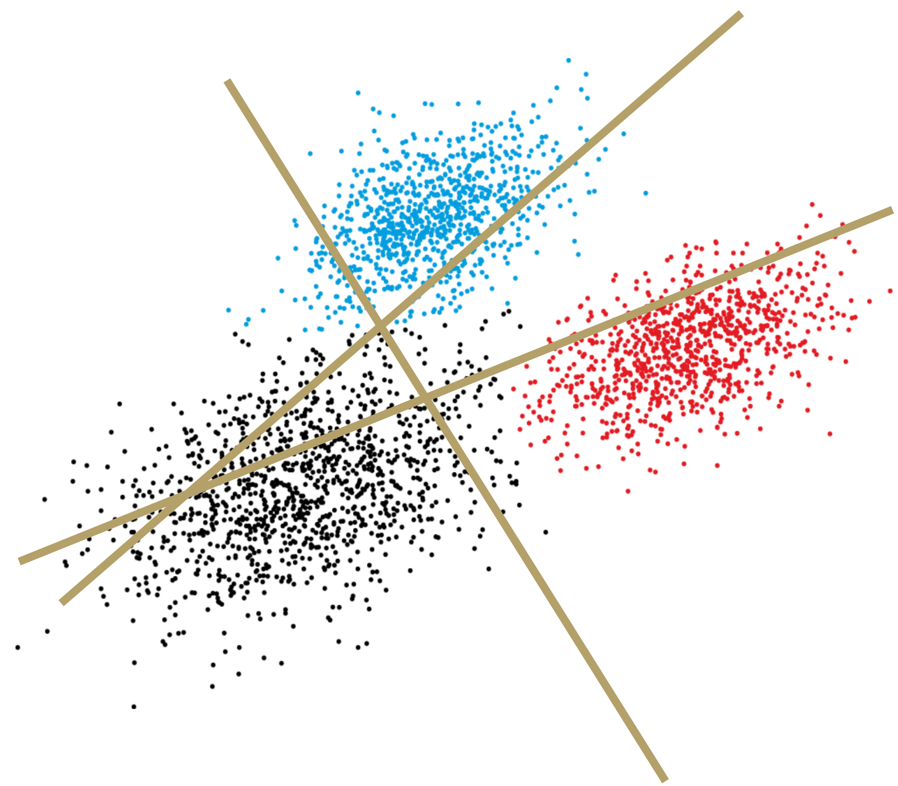


Pairwise cluster separation

Tool for visual inspection

Project clusters onto one dimension





Linear discriminant analysis

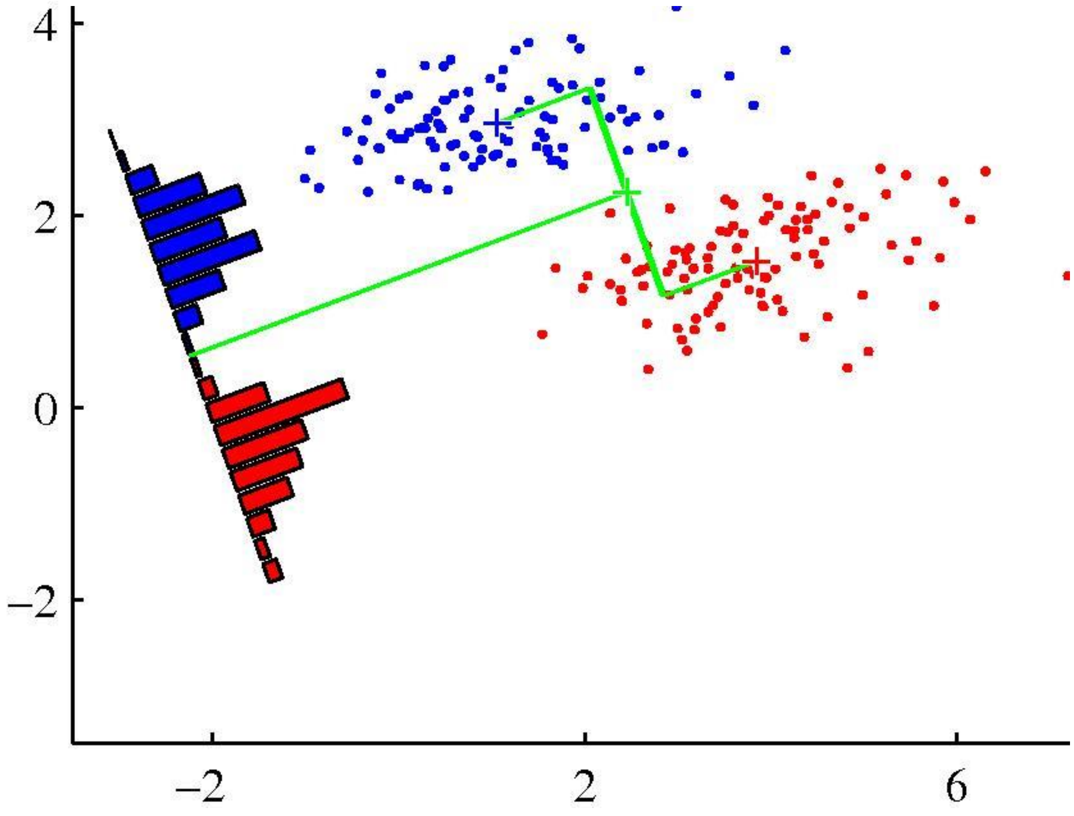
$$y = w^T x, \qquad C_1\colon\ y \geq -w_0$$

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

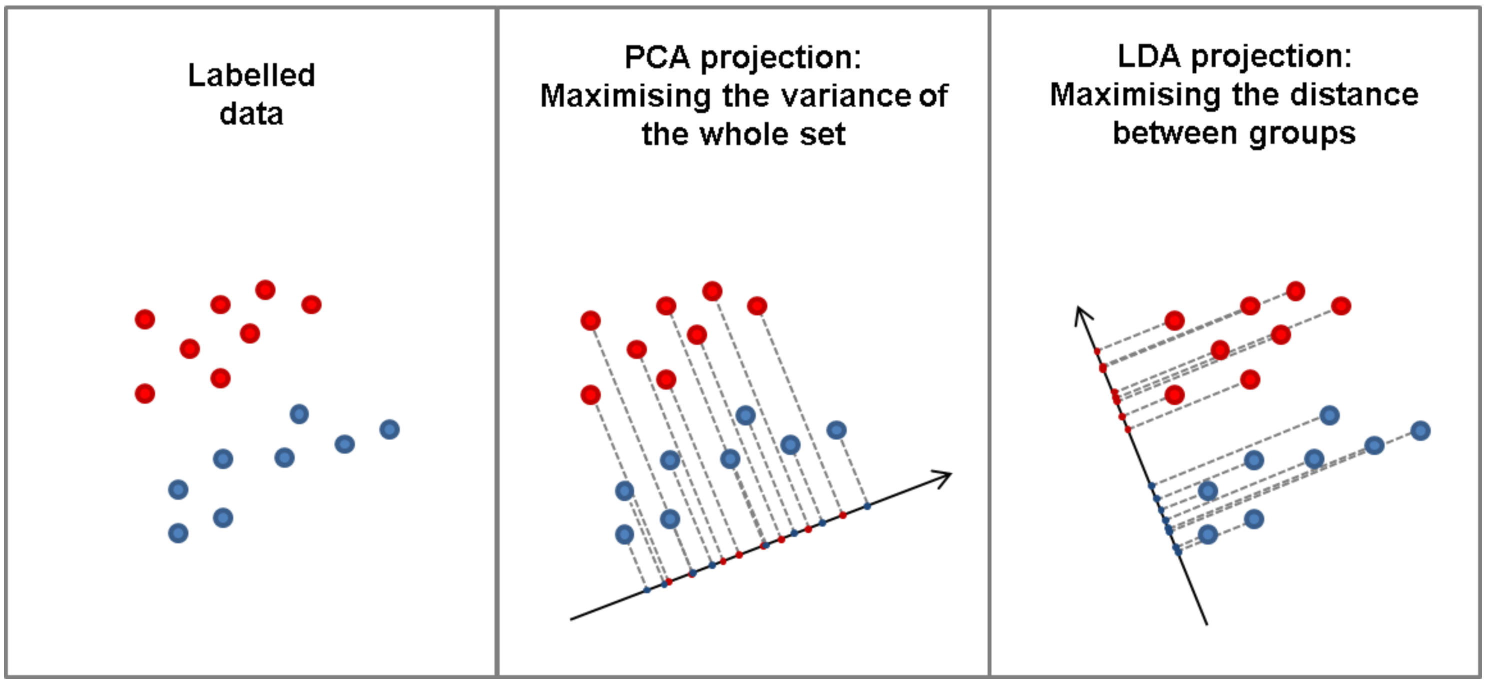
$$(w^T S_B w) S_w w = (w^T S_W w) S_B w$$

$$w = \Sigma_w^{-1}(m_2 - m_1)$$

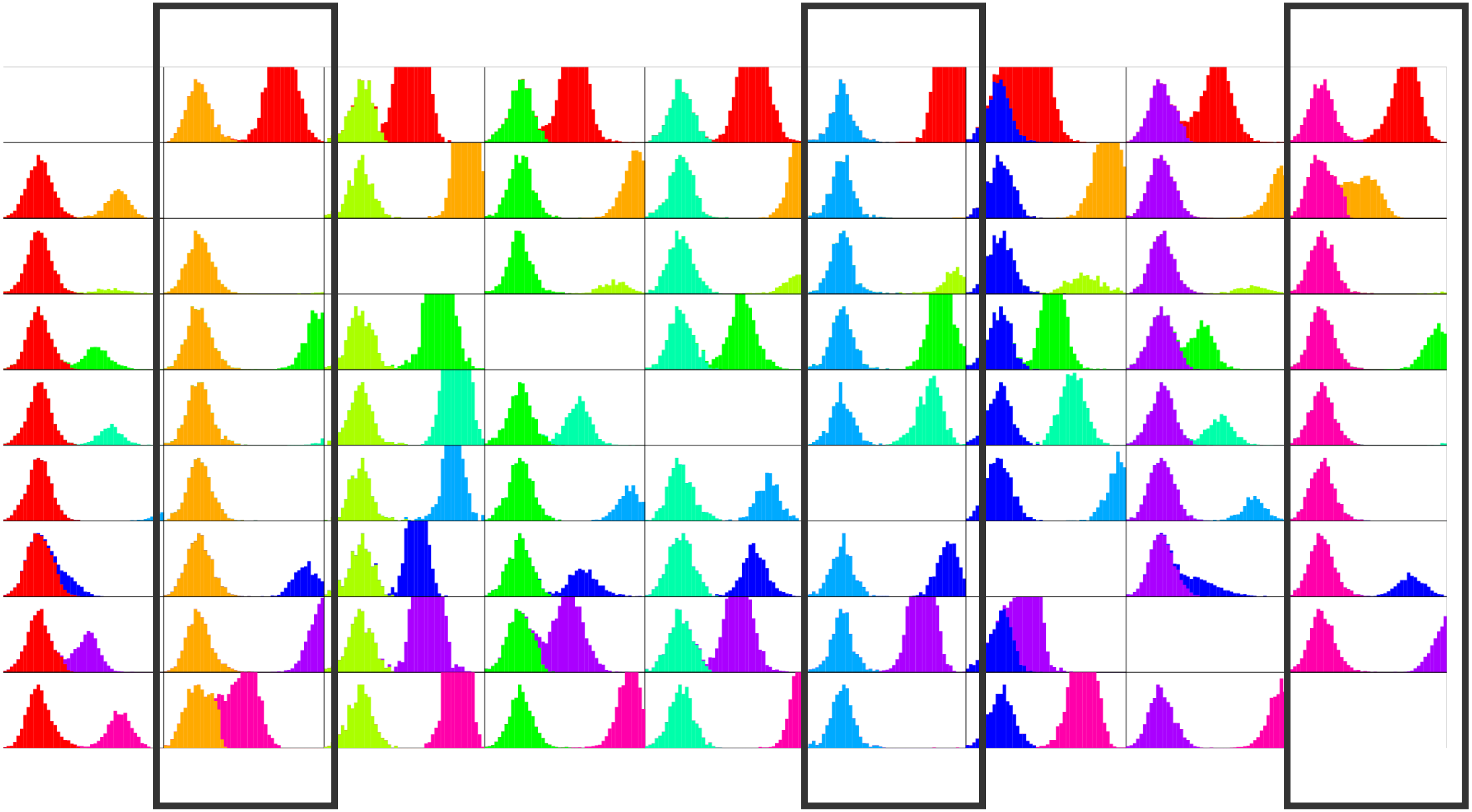
Linear Discriminant Analysis



Difference between LDA and PCA

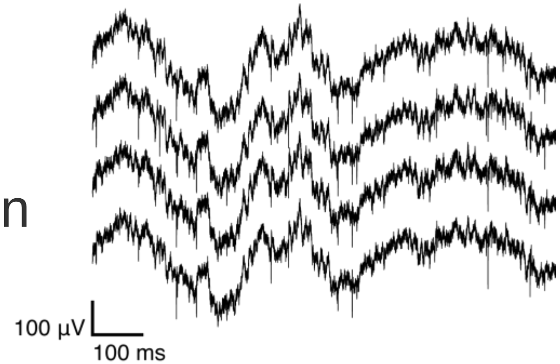


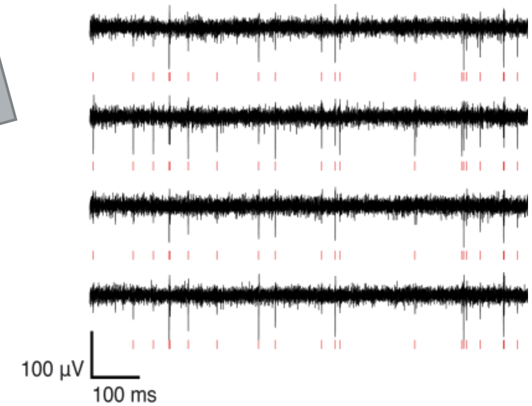
https://nirpyresearch.com/classification-nir-spectra-linear-discriminant-analysis-python/

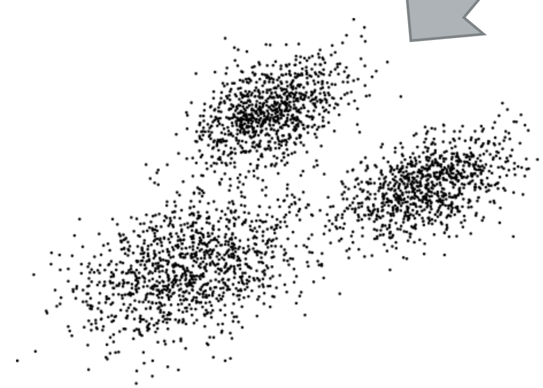


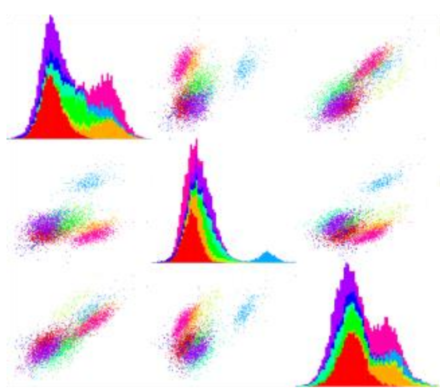
Spike sorting

- Raw data
- Spike detection
- Feature extraction
- Clustering
- Verification

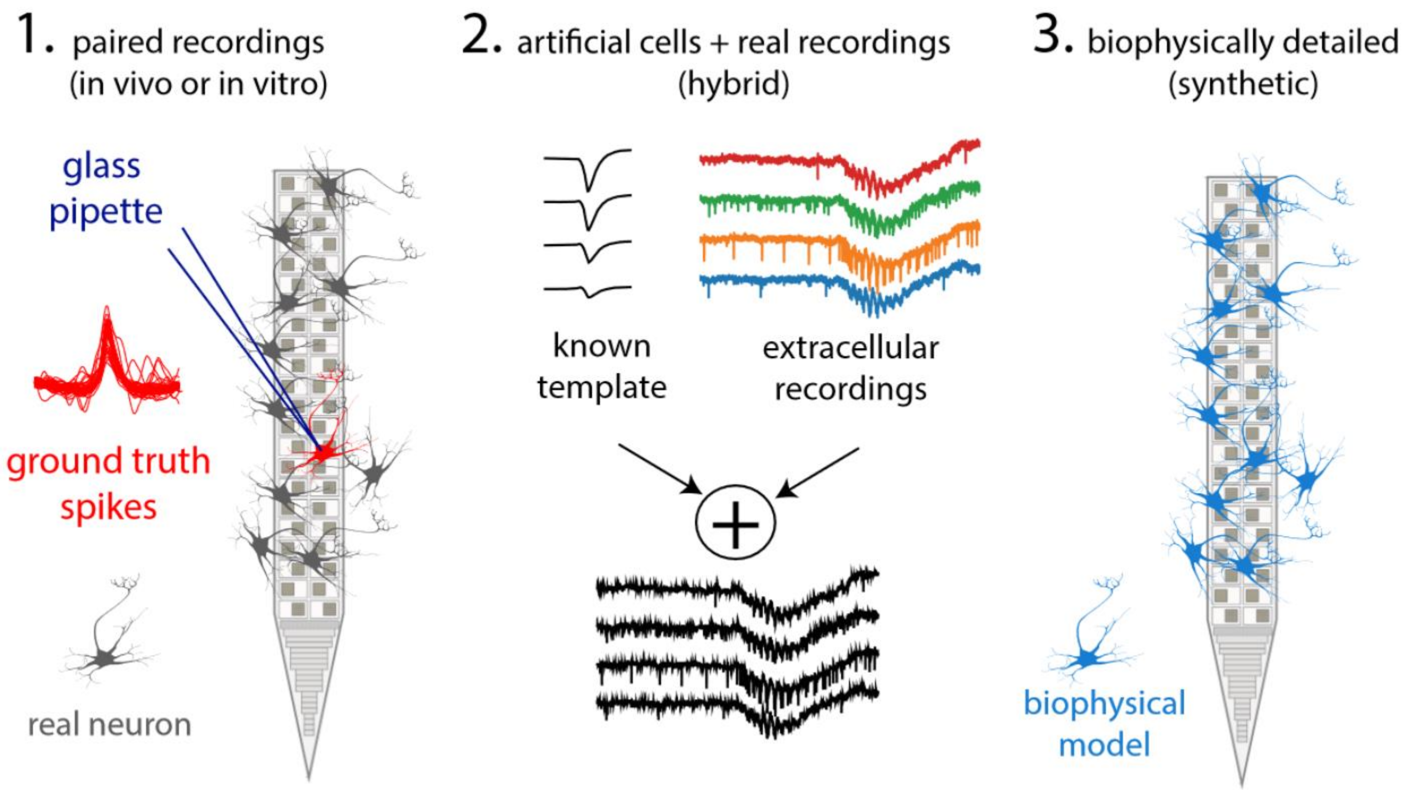




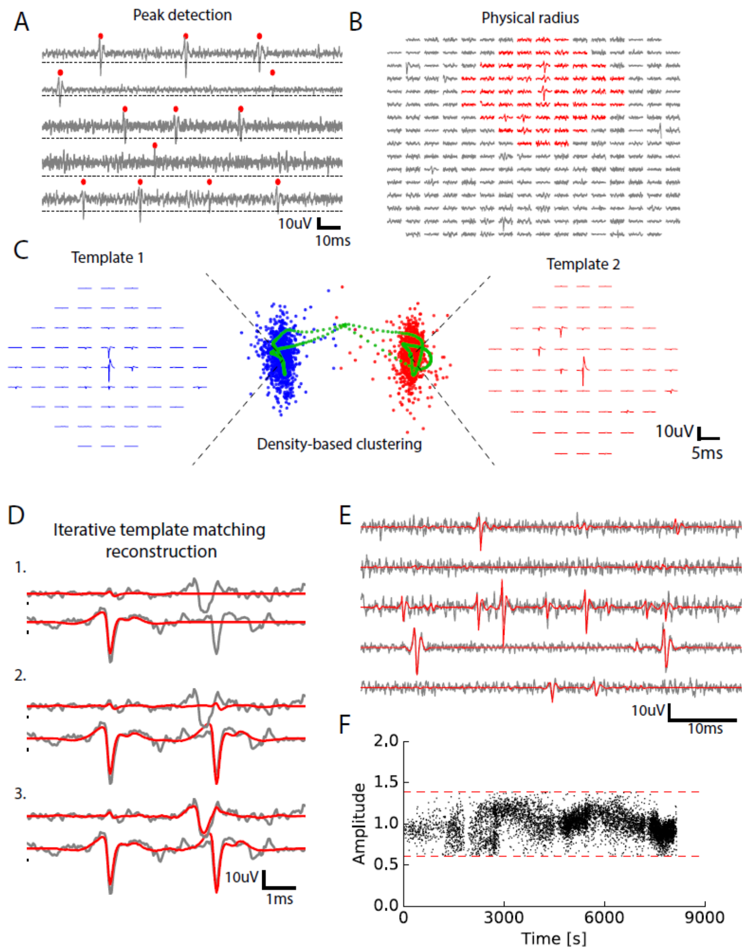




Benchmarking spike sorting algorithms



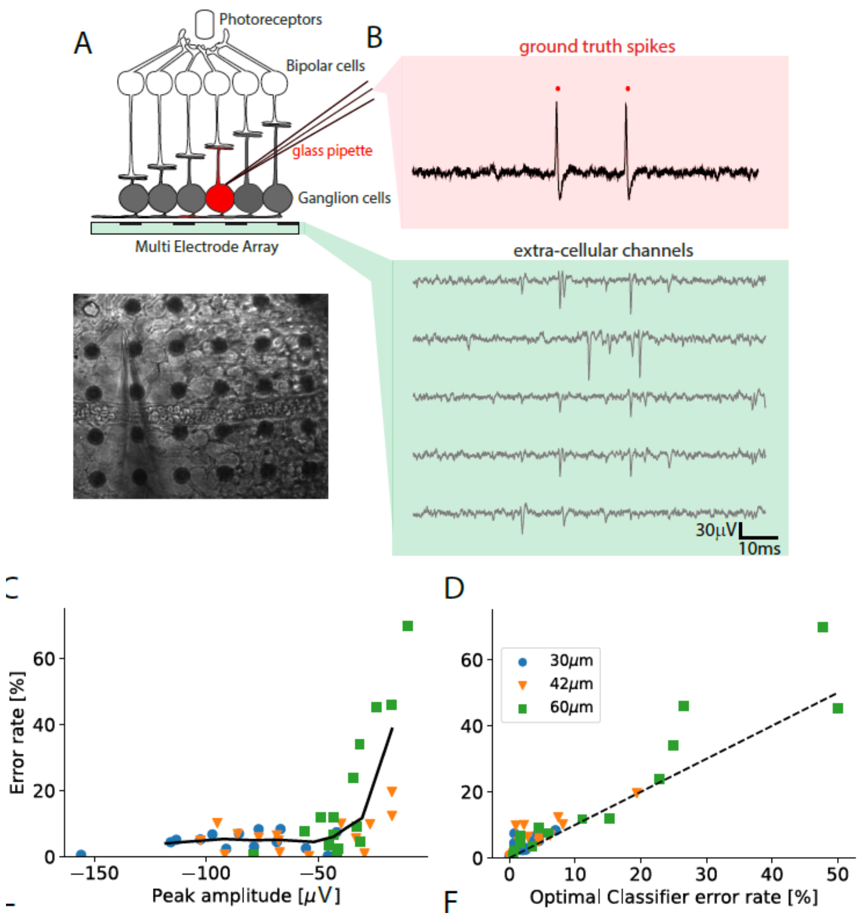
Ground truth verification



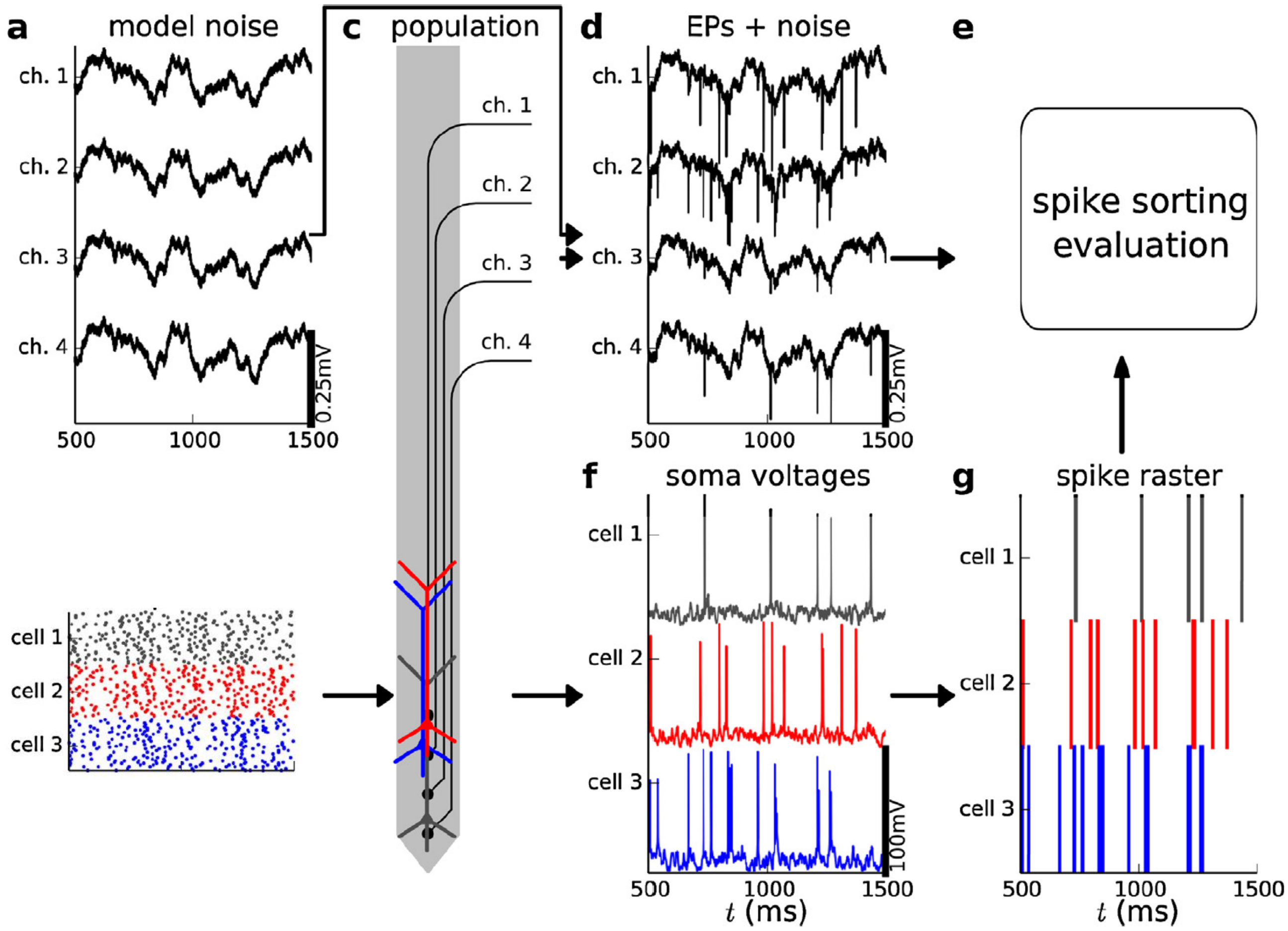
Yger et al. 2018

$$\mathrm{s}(t) = \sum_{ij} a_{ij}\mathbf{w}_j(t - t_i) + b_{ij}\mathbf{v}_j(t - t_i) + \mathbf{e}(t)$$

Verification: Patch-clamp data



Verification: synthetic data

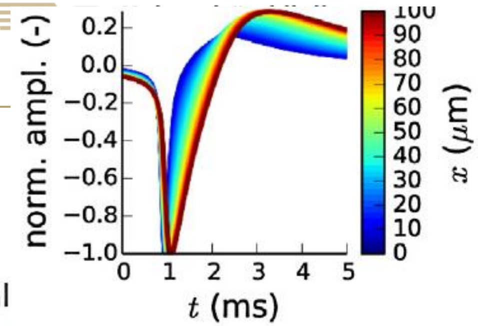


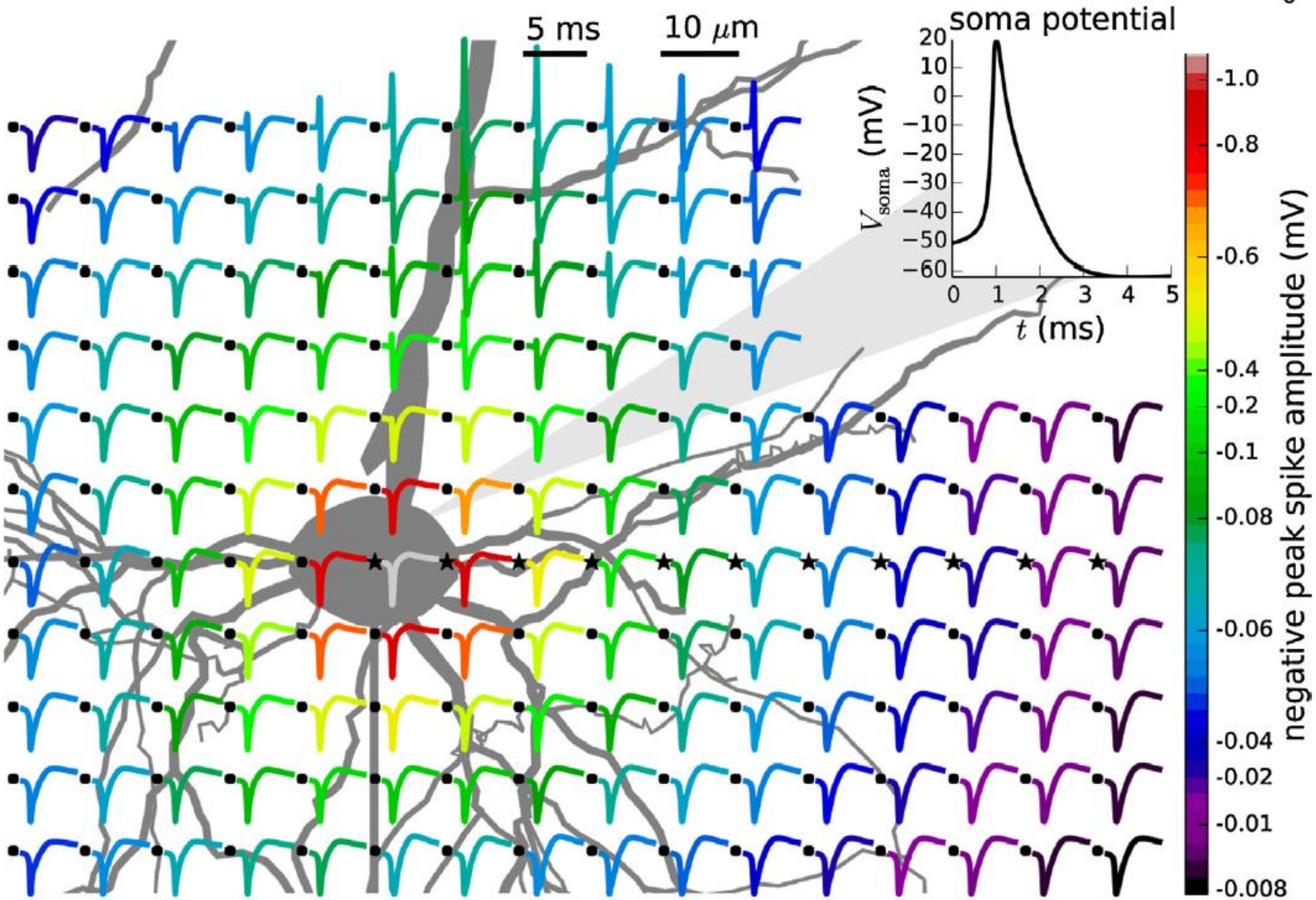
Modified from Hagen et al. (2015)

Verification: synthetic data

Hagen et al. (2015)

Verification: synthetic data

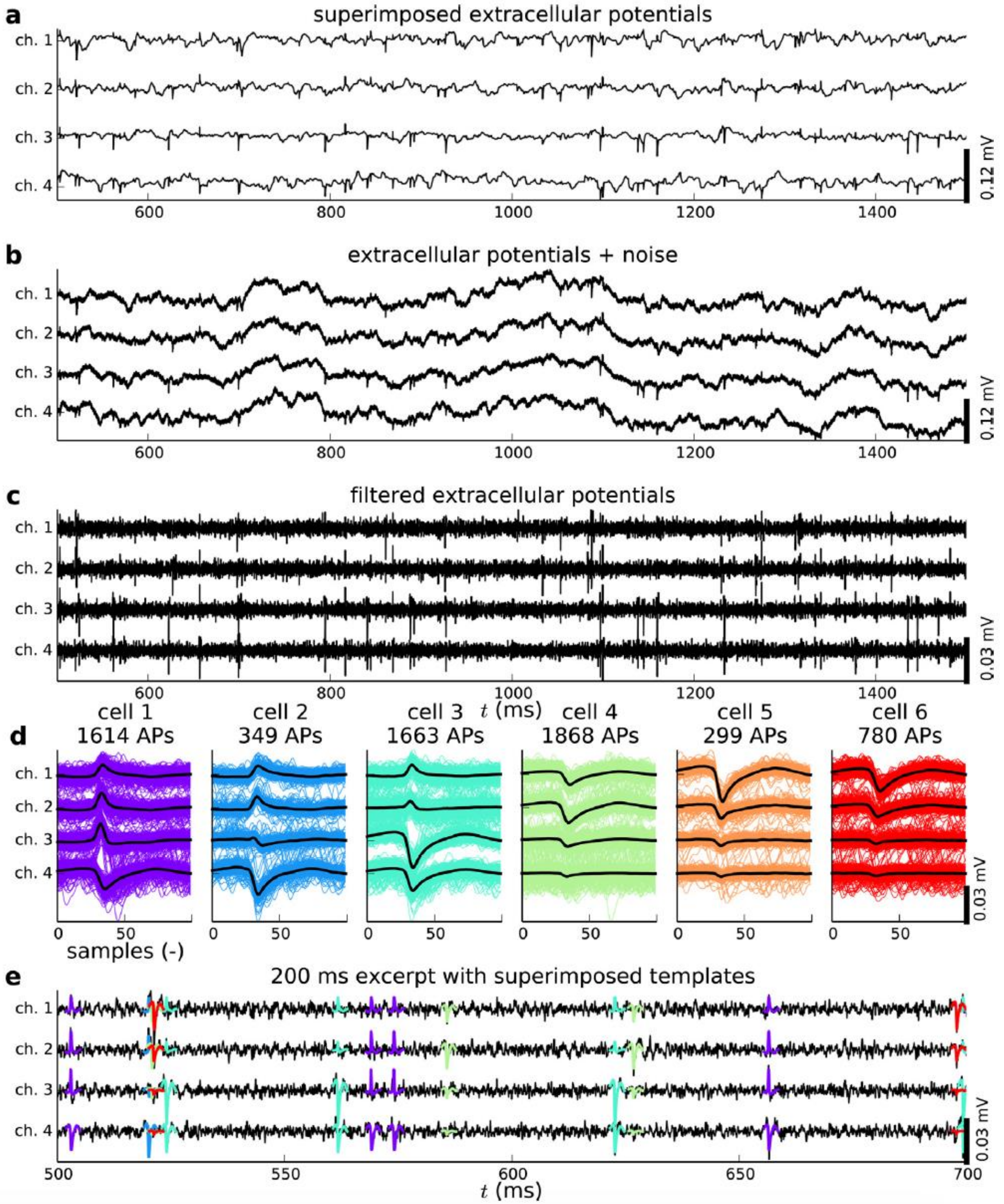




Hagen et al. (2015)

Verification: synthetic data

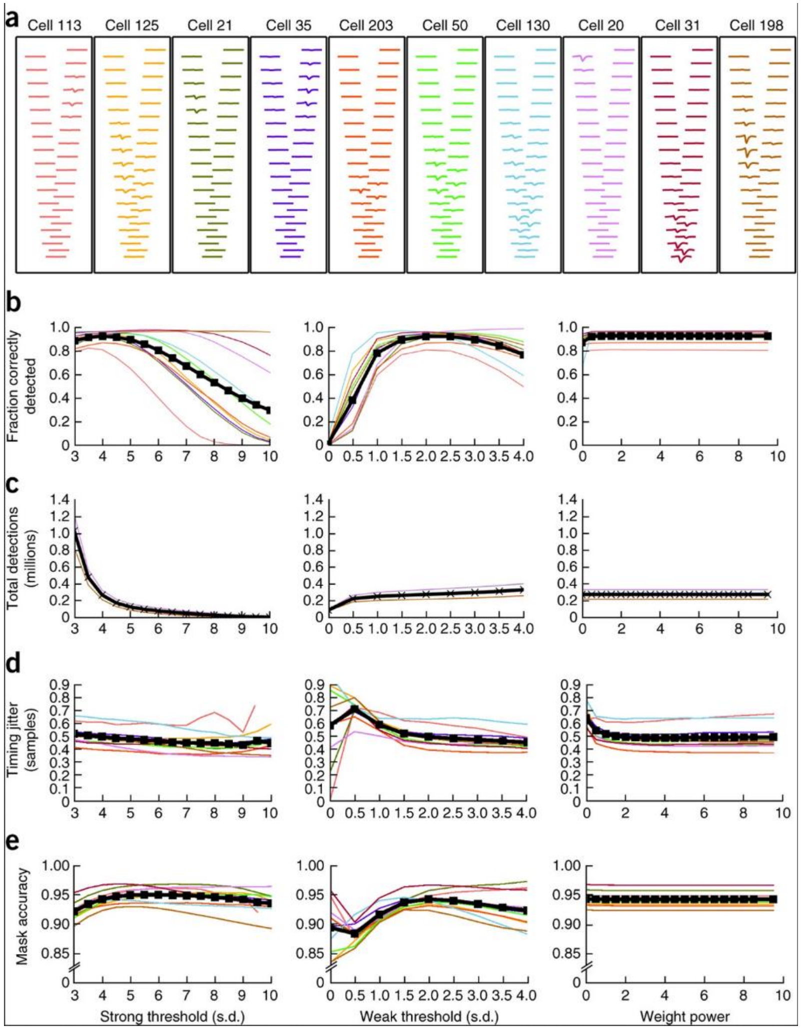
http://spike.g-node.org/spike_eval/benchmark/detail/10/



Hagen et al. (2015), Journal of Neuroscience Methods

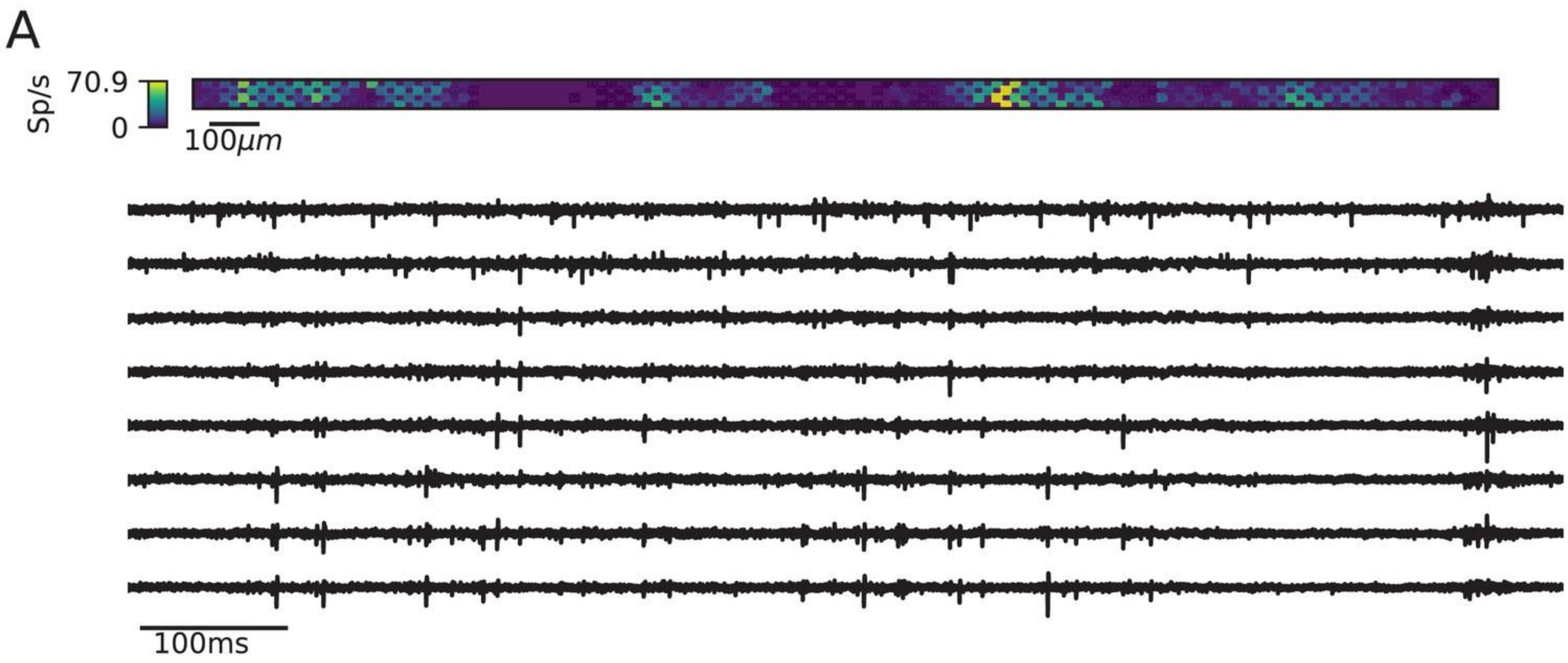
Hybrid data

- Problem: synthetic data not realistic enough! (especially noise properties)
- Use spike templates of “donor cells” and add to “acceptor” recording at known times!
- Simulate amplitude variability & drift



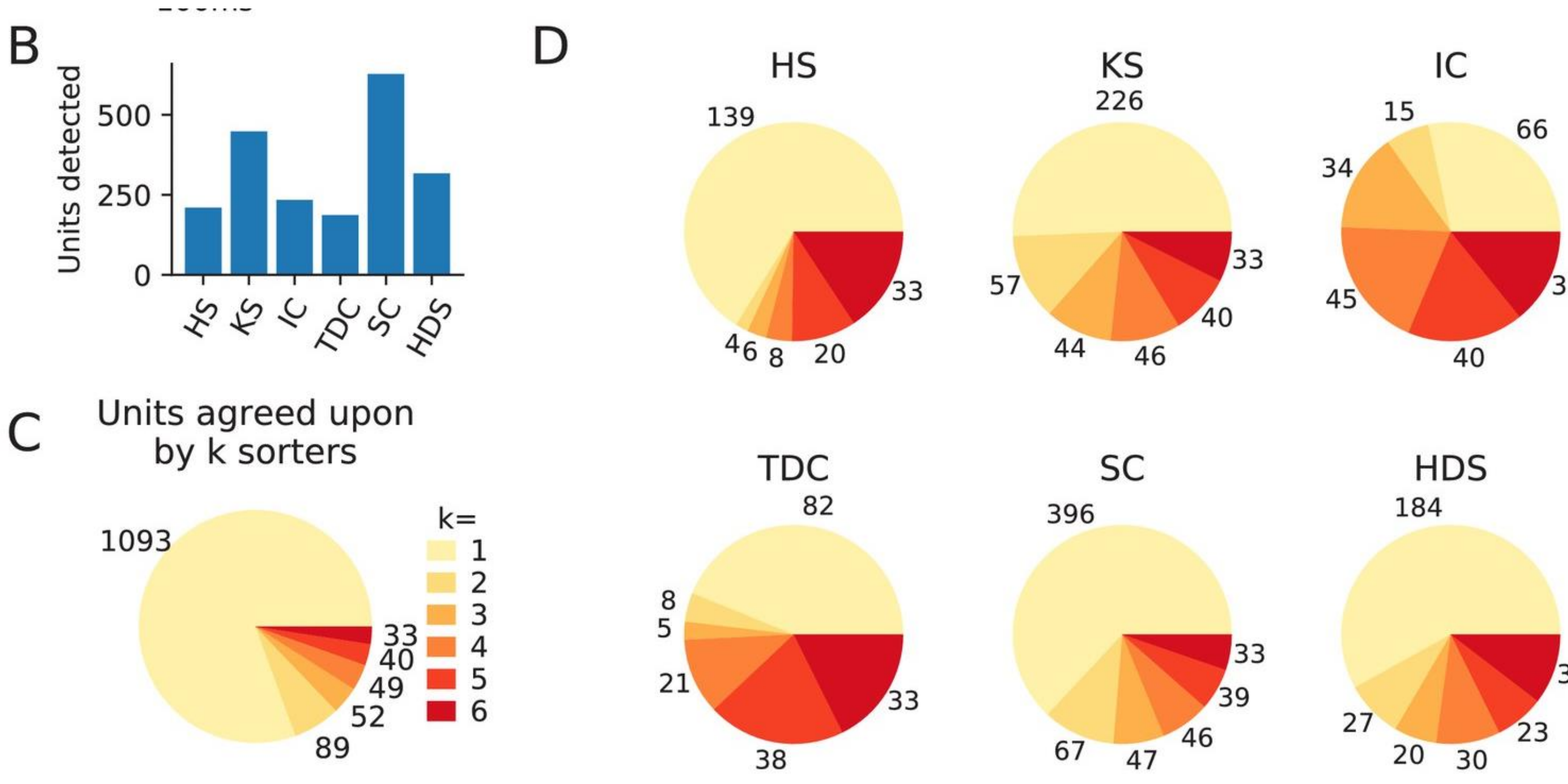
Rossant et al. (2015)

Comparison of different spike sorters



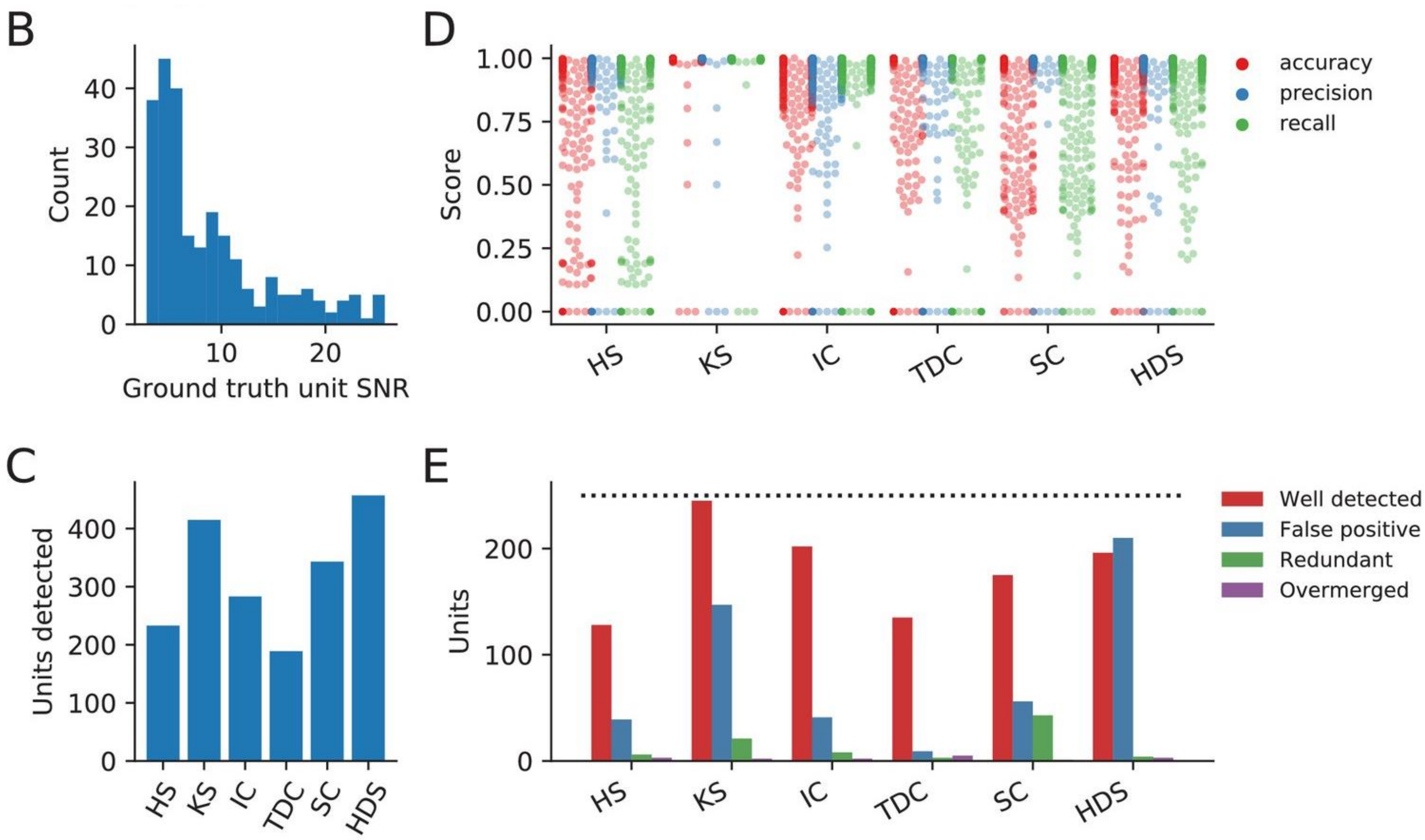
Buccino et al 2020 eLife

Comparison of different spike sorters



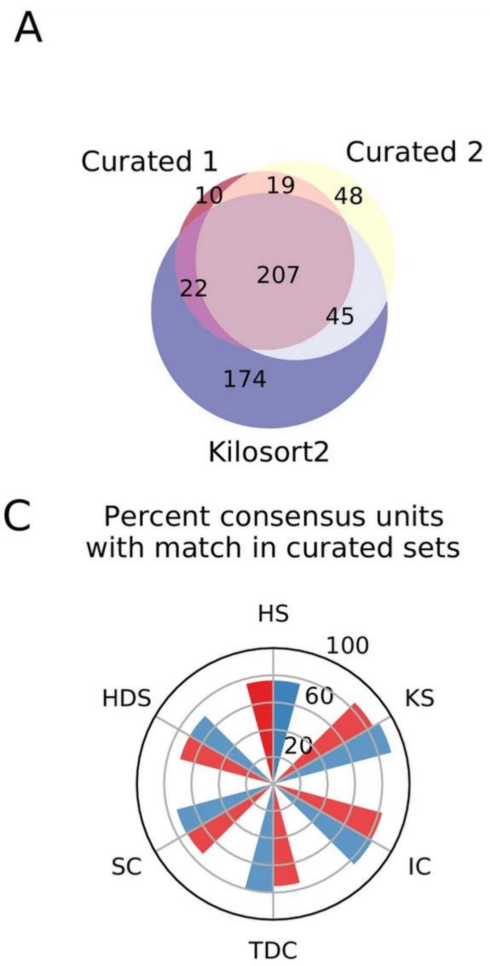
Buccino et al 2020 eLife

Comparison of different spike sorters



Buccino et al 2020 eLife

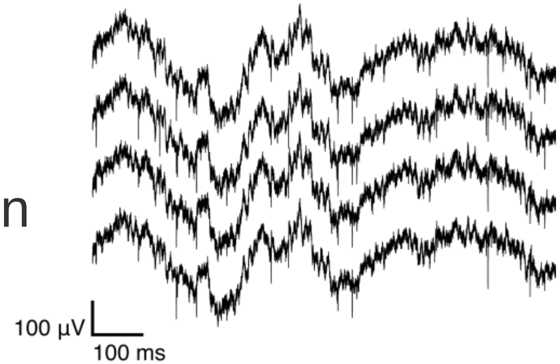
Comparison of different spike sorters

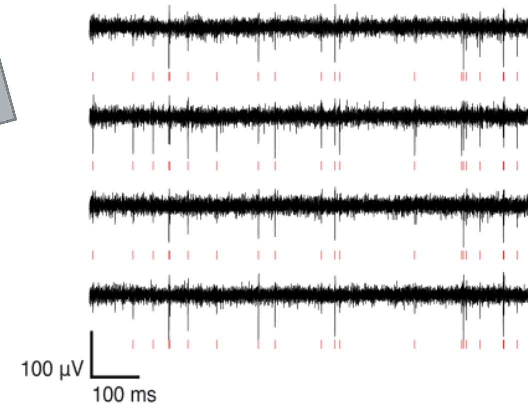


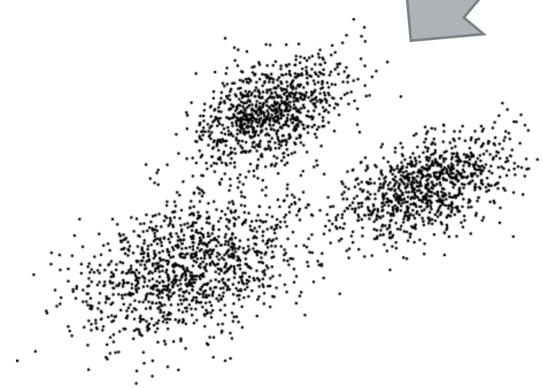
Buccino et al 2020 eLife

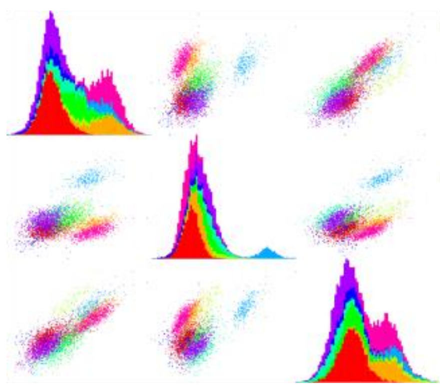
Spike sorting

- Raw data
- Spike detection
- Feature extraction
- Clustering
- Verification









See Buccino et al. Review on Illias