Standardized assessment of extracellular single unit isolation quality

SA Neymotin[1], AV Olypher[2], HY Kao[3], E Kelemen[4], AE Jozwicka[3], WW Lytton[4,5], AA Fenton[4]
Physiology & Pharmacology(4), Neurology(5), Biomedical Engineering Program(1), Neural & Behavioral Sci. Program(3), SUNY Downstate, Brooklyn, NY
Biology Department, Emory University(2), Atlanta, GA

SUNY Downstate Medical Center

690.12

Introduction

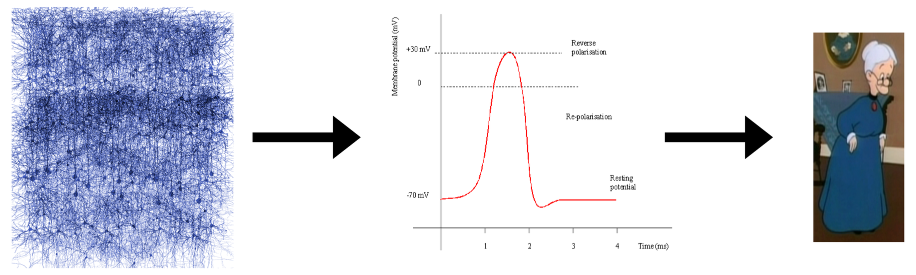
Why do we need to read spike identities?

Information processing in brain

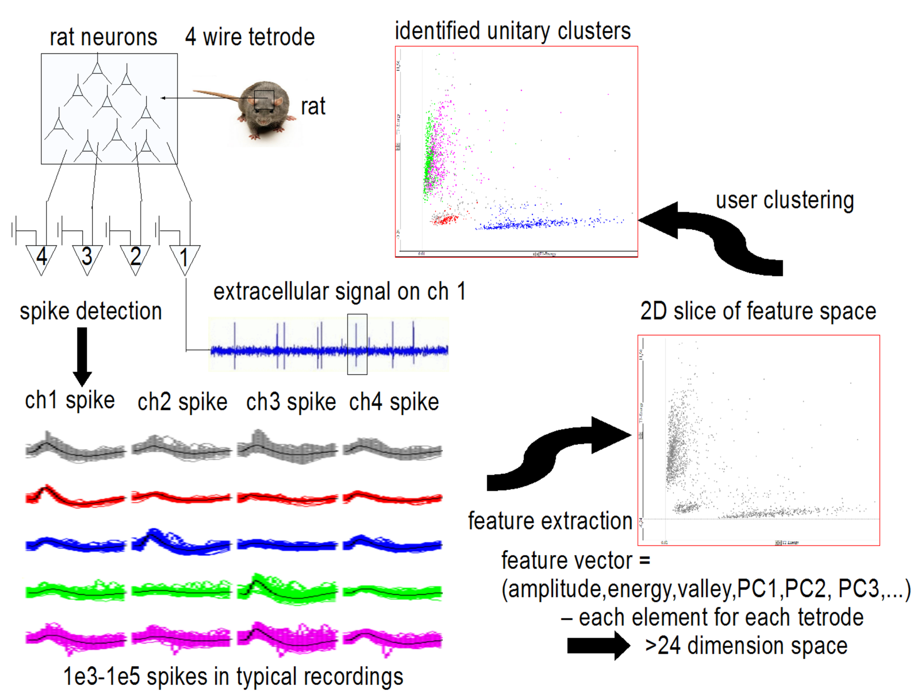
Ensemble codes

Dedicated coding

How do we "read" the codes?

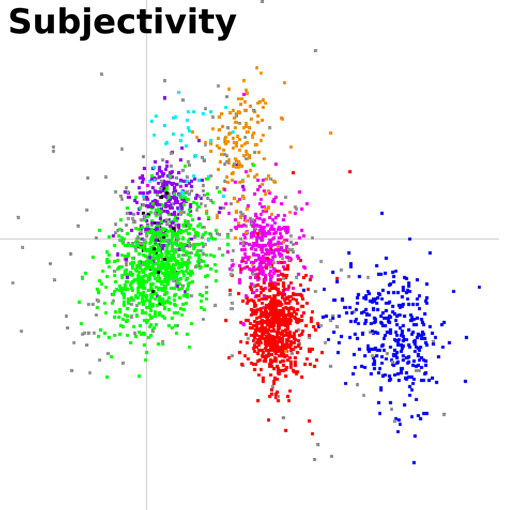


Recording & Clustering Overview

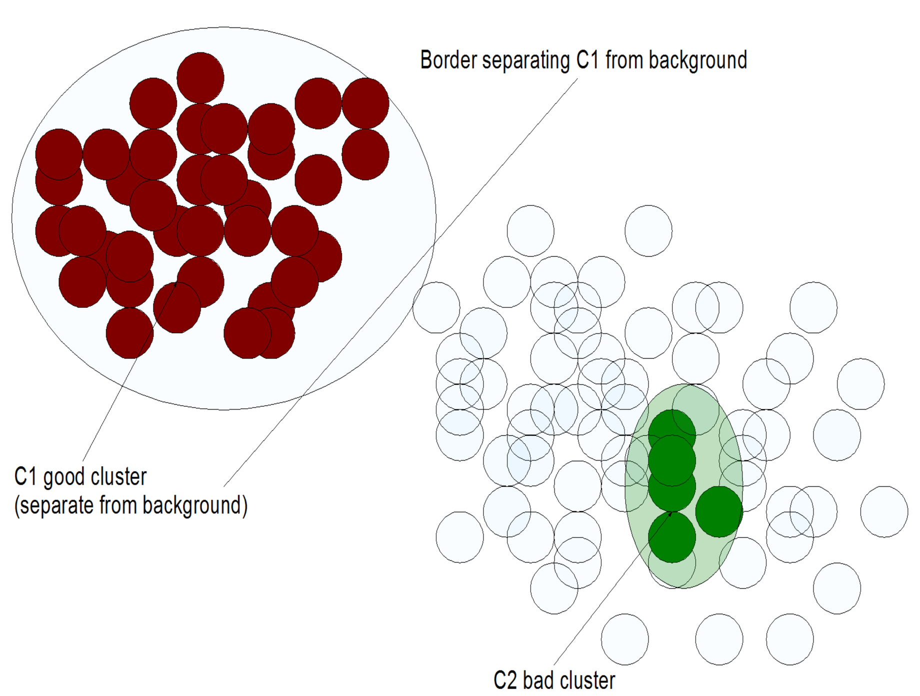


Real Clusters & Subjectivity

Monday: Alice says "Red Cluster Good"
Tuesday: Alice says "Red Cluster Bad"
Wednesday: Alice says "I'm just not sure about the red cluster."
Which Alice do we trust?



Good and Bad Clusters: Cartoon Example



Good = isolated from background, compact, unique

Bad= everything else

Methods

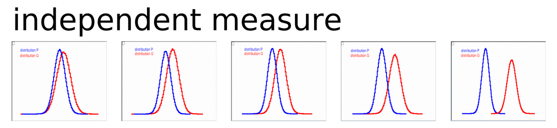
Kullback-Leibler Divergence (KLD)

Measures distinction between two distributions: cluster P from cluster Q:

$$\mathbf{D}_{KL}(P\|Q) = \sum_{i \text{ in } P} p(i)\cdot\log\left(\frac{p(i)}{q(i)}\right)$$

p(i)=probability of i in cluster P
q(i)=probability of i in cluster Q
Relative probabilities are key!!

KLD is a non-symmetric, parameter-independent measure

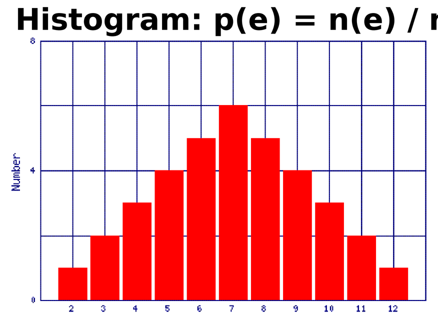


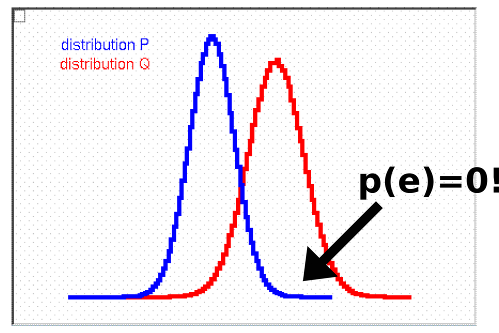
Increasing KLD

Symmetrize KLD into a distance measure with resistor average: $RA(P,Q) = \left(\frac{\mathbf{D}_{KL}(P\|Q)\cdot\mathbf{D}_{KL}(Q\|P)}{\mathbf{D}_{KL}(P\|Q)+\mathbf{D}_{KL}(Q\|P)}\right)$

How do we estimate probabilities?

Histogram: p(e) = n(e) / n

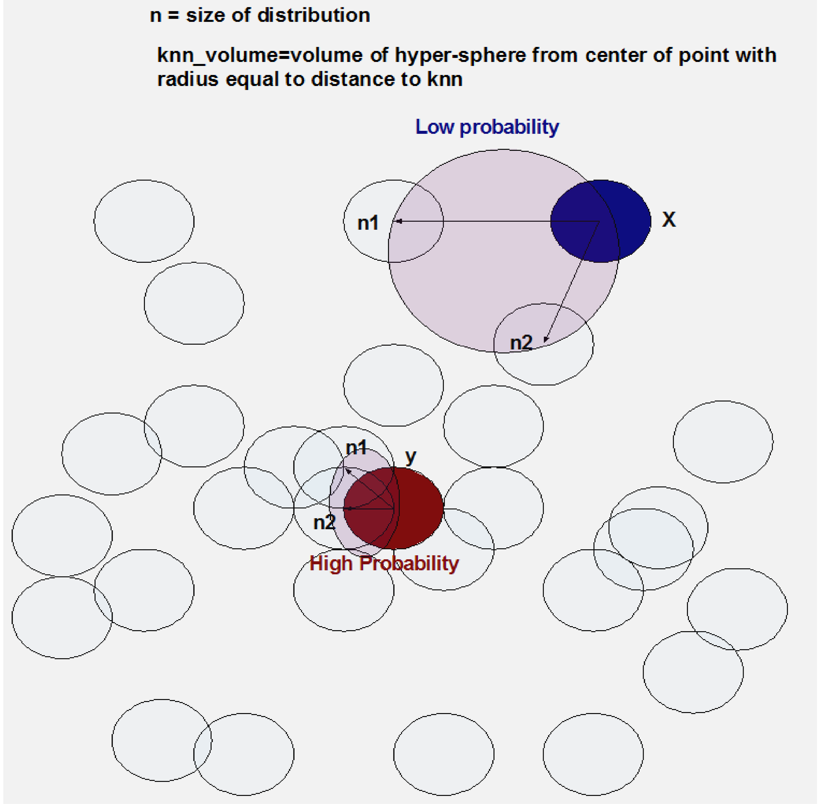




Histograms can have p(e)=0 -> KLD undefined!!

Solution: use k-nearest neighbor (KNN) estimates

KNN Probability estimates
p(x) ~ (k/n)/knn_volume



KNN-based KLD Estimator for clusters P,Q:

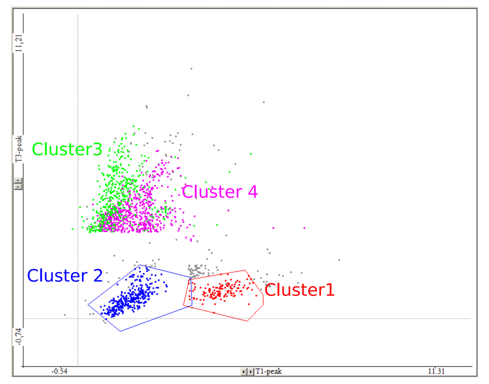
$$\frac{d}{|P|}\sum_{P_i\in P}\log\frac{\min\left(dist\left(q_j\in Q, P_i\right)\right)}{\min\left(dist\left(P_j\in P\setminus P_i, P_i\right)\right)} + \log\frac{|Q|}{|P|-1}$$

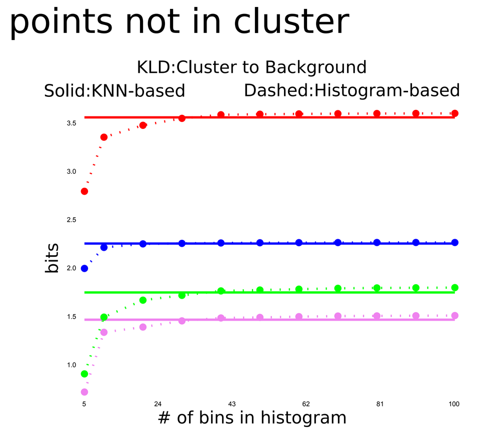
min(dist(Qj,Pi))=min. distance of an element in Q to Pi
min(dist(Pj,Pi))=min. dist. of other element in P to Pi
ratio of these distances is key!!
d=number of features, |P|=# elements in P
|Q|=# elements in Q

(Adopted from "A Nearest-Neighbor Approach to Estimating Divergence between Continuous Random Vectors",Wang. et al, 2006)

Compare: Histogram-based KLD vs KNN-based KLD

Background (BG) = all points not in cluster





Results - properties

Symmetric KLD between clusters as Isolation-Information (ISO-I)

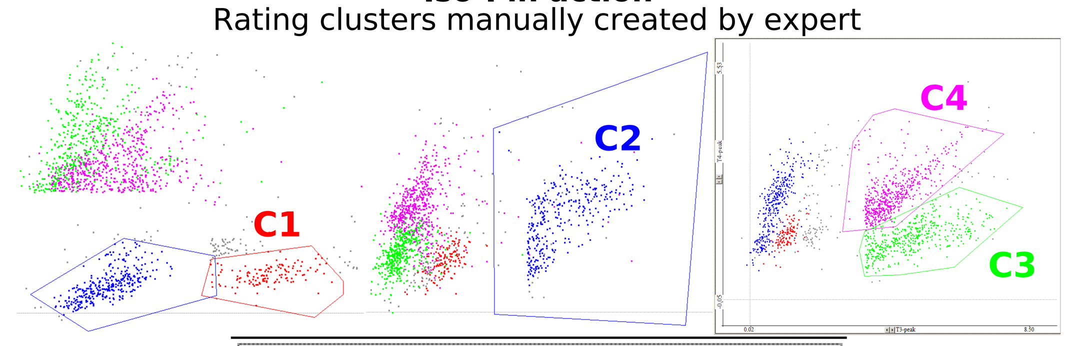
ISO-I(BG) Measures distance from cluster C to background distribution (BG) - measures isolation & compactness

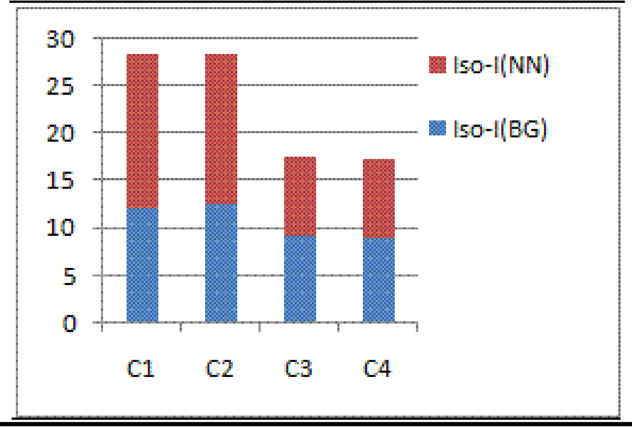
ISO-I(NN) Measures distance from cluster C to nearest neighboring cluster - measures uniqueness and compactness

Both use units of information(bits): intuitive and feature-independent

Iso-I in action

Rating clusters manually created by expert



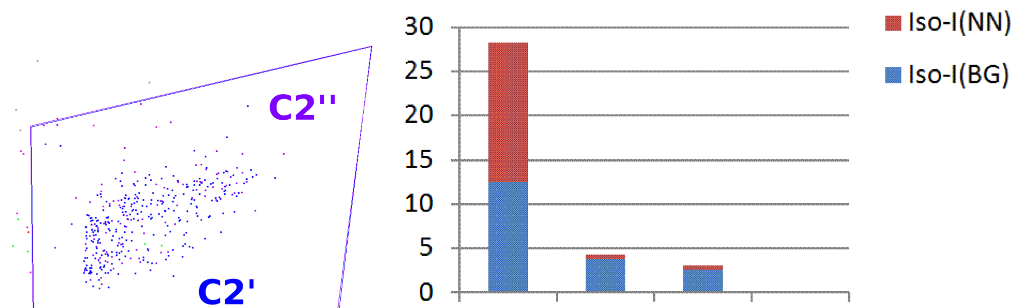


Iso-I Sensitivity to perturbations

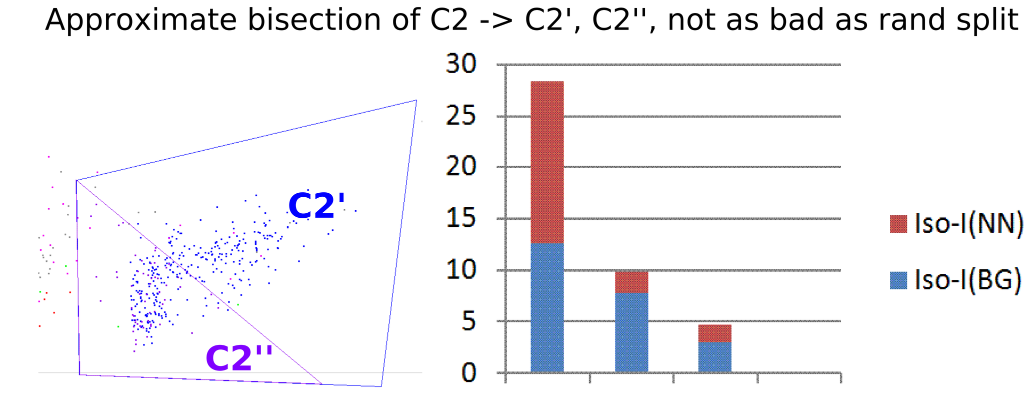
Apply degradations/enhancements to clusters ,track changes in Iso-I

Demonstrate that when we damage a cluster, the information content of the clustering decreases.

Randomly split C2 into C2' , C2''. C2',C2'' have negligible Iso-I(NN) values.



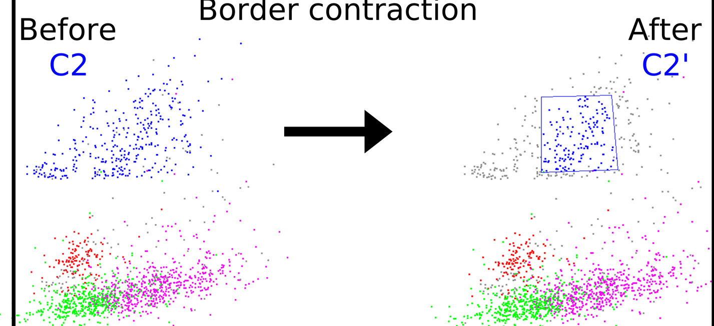
Approximate bisection of C2 -> C2', C2'', not as bad as rand split

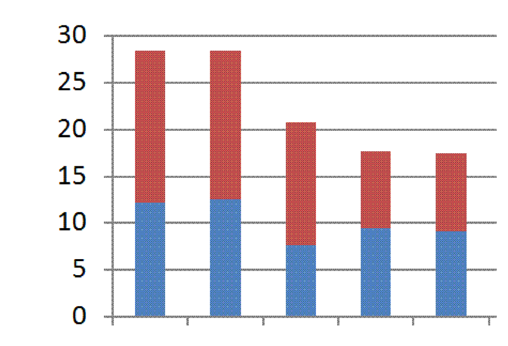


Border contraction

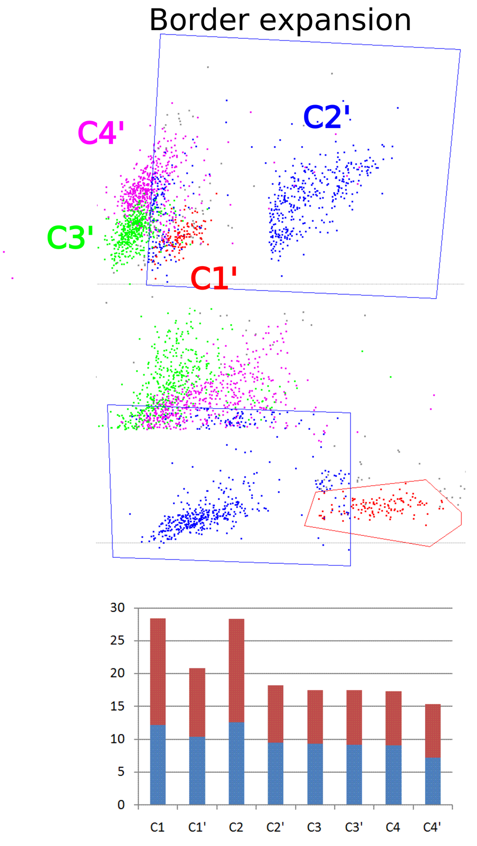
Before

After





Border expansion



Results - compare Iso-I with other cluster quality measures

Isolation distance (Iso-D) :
Calculates Nth largest squared mahalanobis distance (MD) of spikes not in cluster to cluster center, where N is size of cluster. Large value indicates good separation of cluster from background.

L-ratio: $L(C) = \sum_{i\notin C} 1 - CDF\chi^2_{df}(D^2_{i,C})$

Estimates separation of cluster from surrounding noise with the summation above using Chi-squared cumulative distribution function with 8 degrees of freedom and the squared MD: (Di)^2, for all points in the noise distribution. Then normalizes by N. Small values indicate a "moat" around the cluster.

Standard Iso-D and L-ratio operate in feature space of 1st principal component of energy-normalized spikes and energy of each spike on each microwire of a tetrode (8 features). (Quantitative Measures Of Cluster Quality For Use In Eextracellular Recordings, Schmitzer-Torbert et. al, 2005)

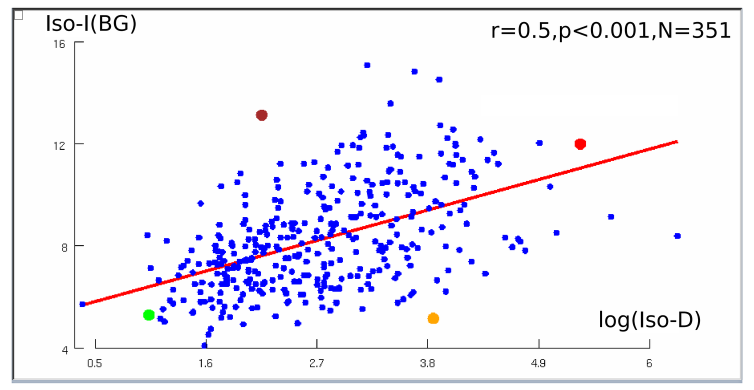
Silhouette-Width(SW): $s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}}$

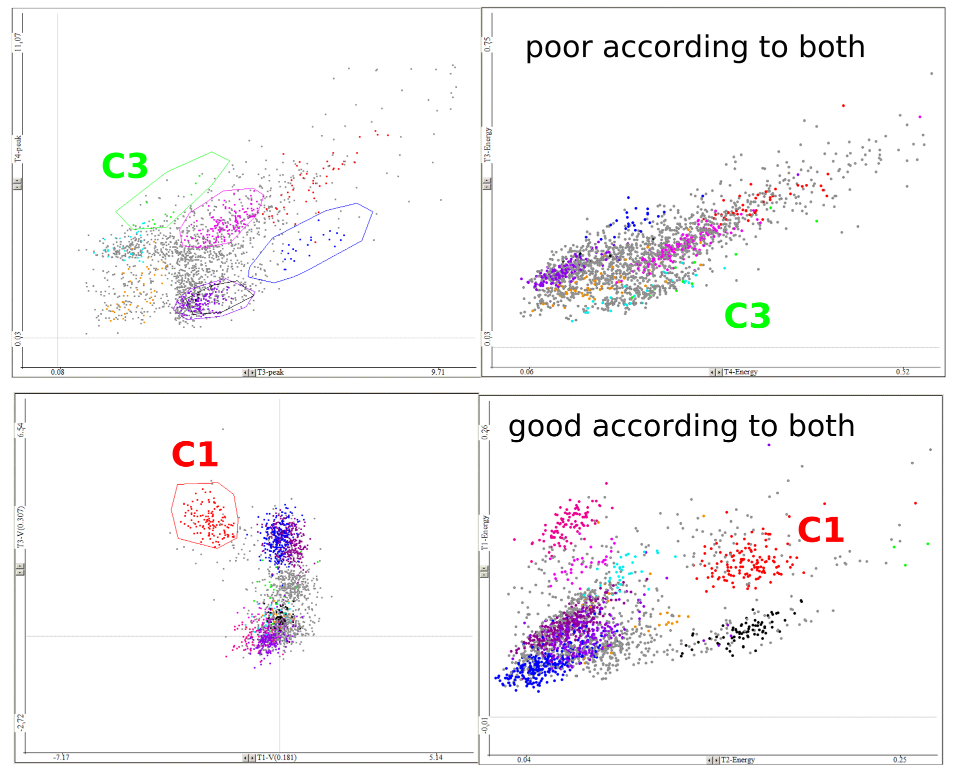
Given a(i) as distance of nearest neighbor(nn) of element i in cluster A and b(i) as distance of nn of element i in cluster B, average s(i) for elements in cluster A, measures confidence of points belonging to cluster A. Values [-1,1] are [bad,good]. Operates in arbitrary feature space.
(Graphical aid to cluster analysis, Rousseeuw PJ,1987)

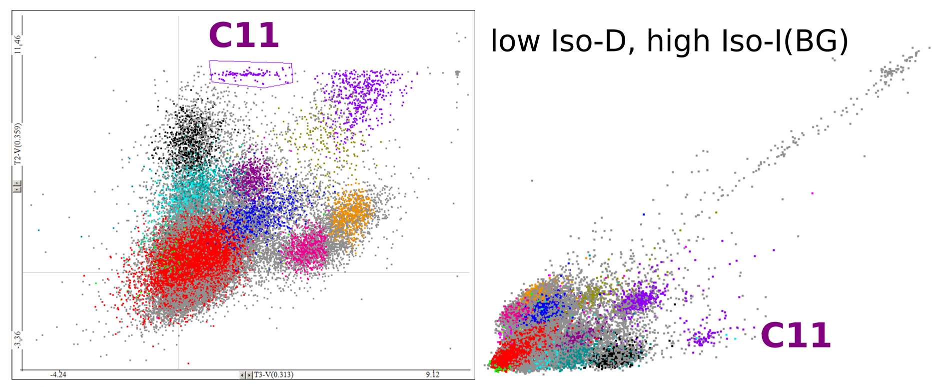
	Iso-I(BG)	Iso-I(NN)	log(Iso-D)	log(L-Ratio)	SilhouetteWidth
Iso-I(BG)	•	▪	▪	▪	▪
Iso-I(NN)	0.55	•	▪	▪	▪
log(Iso-D)	0.5	0.47	•	▪	▪
log(L-Ratio)	-0.57	-0.46	-0.91	•	▪
SilhouetteWidth	0.27	0.63	0.43	-0.44	•
Spikes	-0.02	0.21	0.37	-0.26	0.07

Table of correlations between different measures and number of spikes in a cluster.

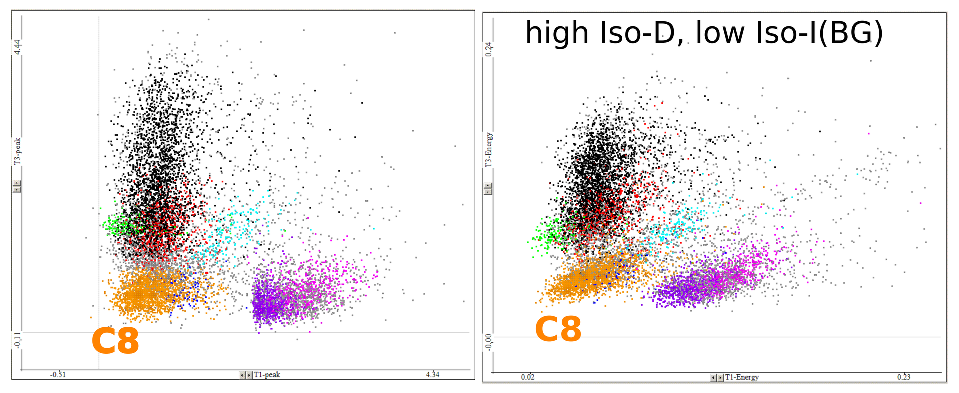
Iso-I(BG) less sensitive to cluster size than Iso-D





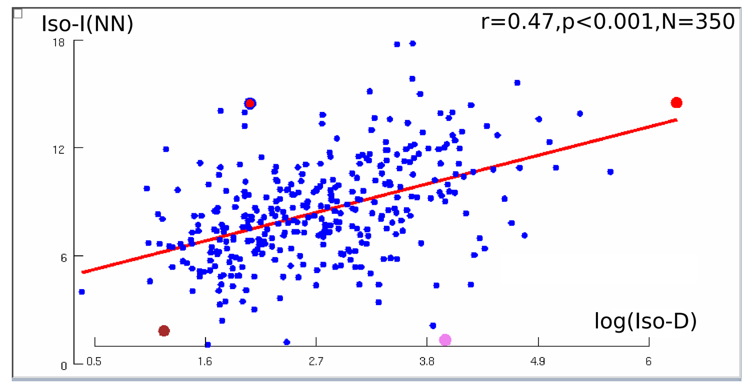


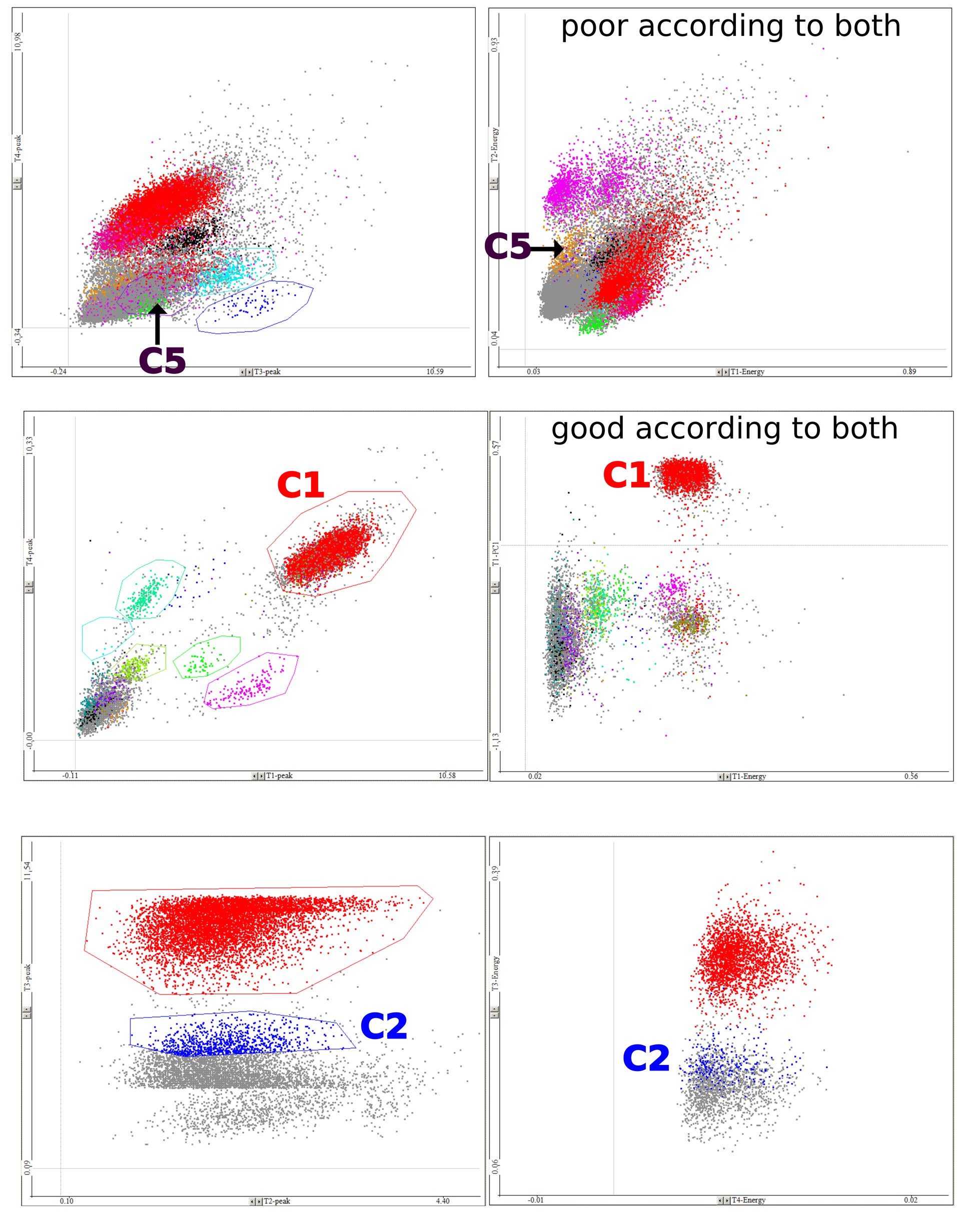
C11, reasonable separation, but few elements, -> low Iso-D



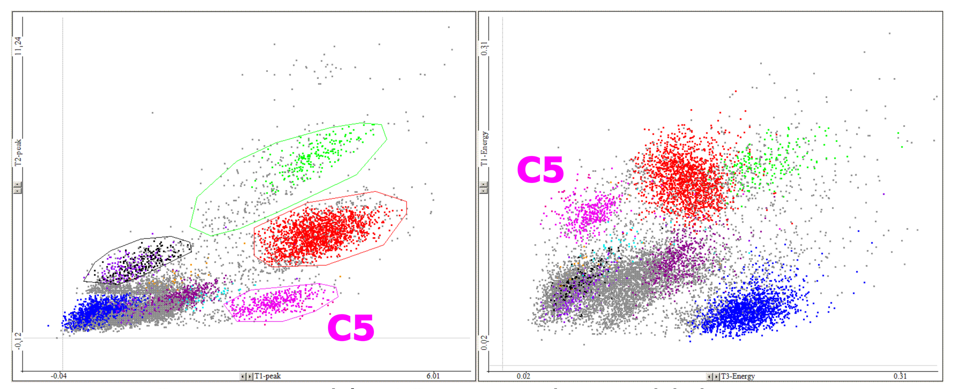
C8, poor separation, but many elements (~28% of spikes) -> inflated Iso-D

Iso-I(NN) sensitive to uniqueness of cluster compared to Iso-D





C2, poor background separation -> low Iso-D, Iso-I(BG)
C2, good inter-cluster separation -> high Iso-I(NN)
Iso-I(NN) AND Iso-I(BG) can detect different types of problems.



C5, reasonable BG separation -> high Iso-D
C5, poor inter-cluster separation (low uniqueness) -> low Iso-I(NN)

Conclusions

Cluster assessment critical for quality control of spike sorting.

Various cluster quality measures are useful for detecting different types of cluster features/errors.

Information measures are feature independent and show good concordance with intuitive sense of quality.

Standard Iso-D is oversensitive to number of elements in a cluster, limited to specific features, redundant with L-ratio, and not sensitive to uniqueness of clusters.

Iso-I(BG),Iso-I(NN) are less redundant cluster quality measures that can detect cluster background-separation, uniqueness, and compactness.

We invite submissions of spike cluster data for evaluation.

Acknowledgments

NS045612

National Science Foundation (IOS-0725001)