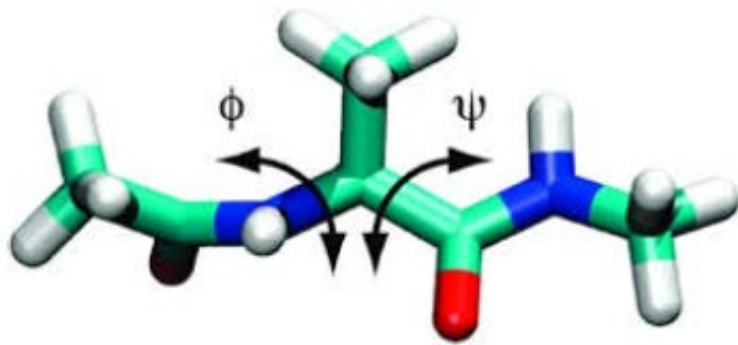
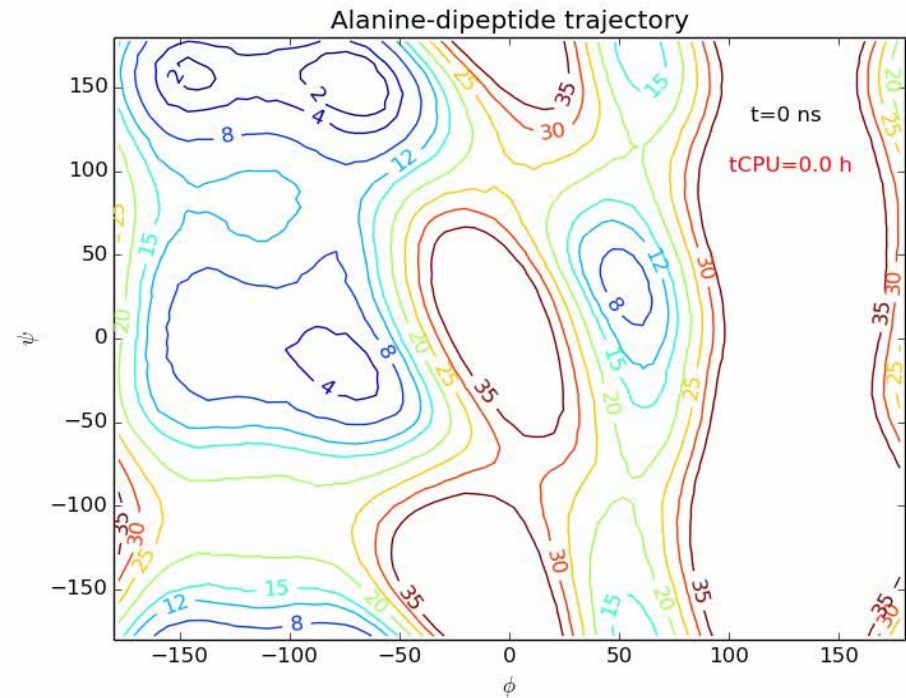


Density-based clustering and core set MSMs

Markov model from molecular dynamics



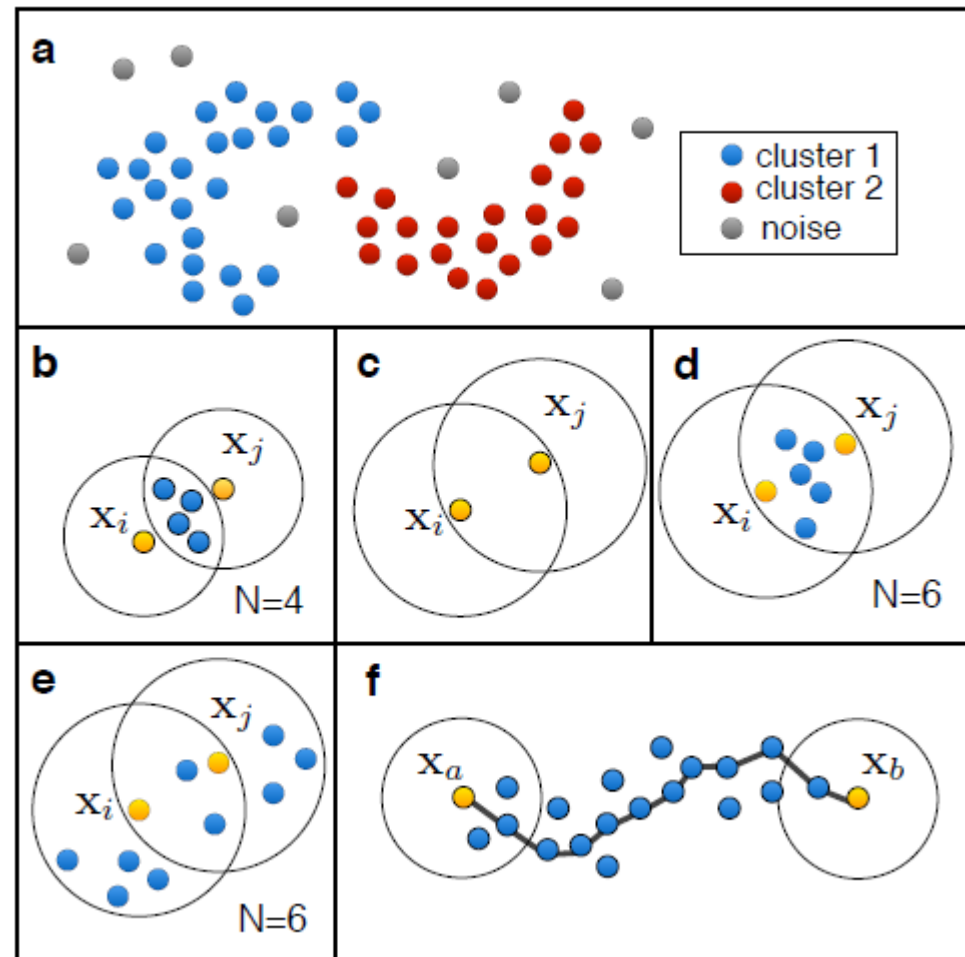
alanine dipeptide



Identify metastable states and probabilities of transitioning.

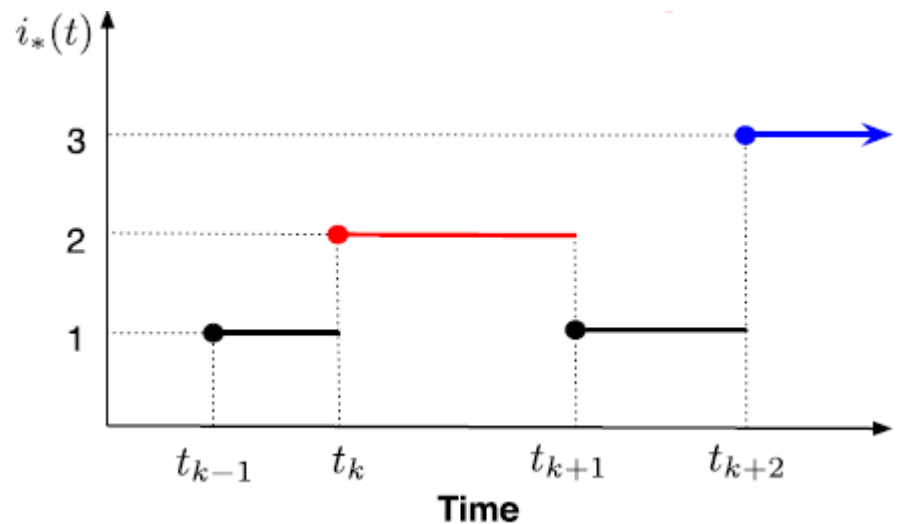
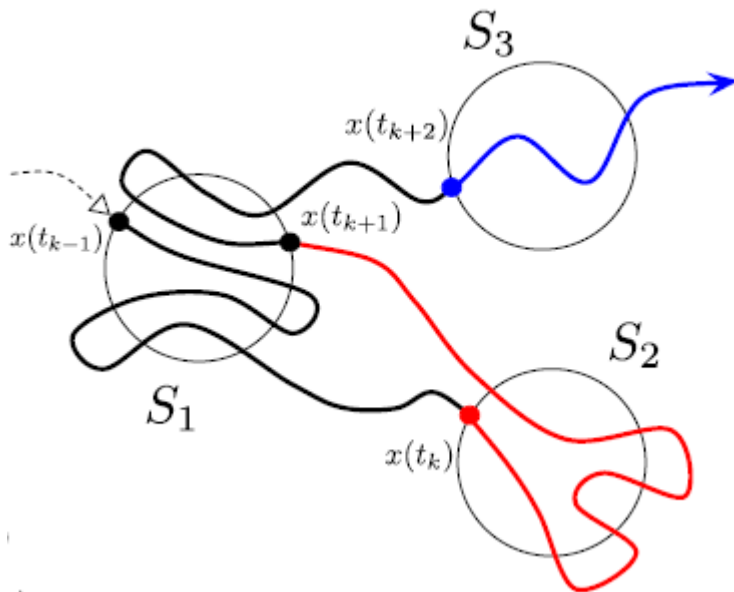
Density-based clustering

- Clusters are dense regions in the data space, separated by regions of lower object density
- A cluster is defined as a maximal set of density connected points.



Markov state models based on milestones / core set MSM

- Define some “core regions”. Assign trajectory to the last visited core + count transtions in this reassigned trajectory.



- Needed input for the algorithm: definition of cores.
- idea: take regions of relatively high density as cores.

Possible objectives

- Use insights from the project “sparse sampling” to implement a suitable density-based clustering algorithm.
- A standard trick to deal with large data sets is pre-clustering with *k-means* to sparsify the data. What is the connection to sparse sampling?
- For the core set MSM, cores should neither be too small (would always be missed) nor too large (recrossing). Experiment with different ideas to choose the core size.

Literature

- *Lemke, Keller*, “Density-based cluster algorithms for the identification of core sets” (preprint)
- *Rodriguez, Laio*, "Clustering by fast search and find of density peaks", *Science* Vol. 344, Issue 6191, 1492 (2014)
- *Schütte, Noé, Lu, Sarich, Vanden-Eijnden* "Markov state models based on milestoning" *J Chem Phys.* 134 :204105. (2011)