Results From An International MHD Data Mining Collaboration

D. Pretty[1] B. Blackwell[1] S. Yamamoto[2] E. Ascasíbar[3] F. Detering[4] R. Jiménez-Gómez[3] K. Nagasaki[2] and S. Sakakibara[5]

[1]Plasma Research Laboratory, Australian National University.

[2]Institute of Advanced Energy, Kyoto University.

[3]Laboratorio Nacional de Fusión, Asociación Euratom/Ciemat, Madrid.

[4]Diversity Arrays Technology P/L, Yarralumla, ACT 2600, Australia

[5]National Institute for Fusion Science, Gifu, Japan

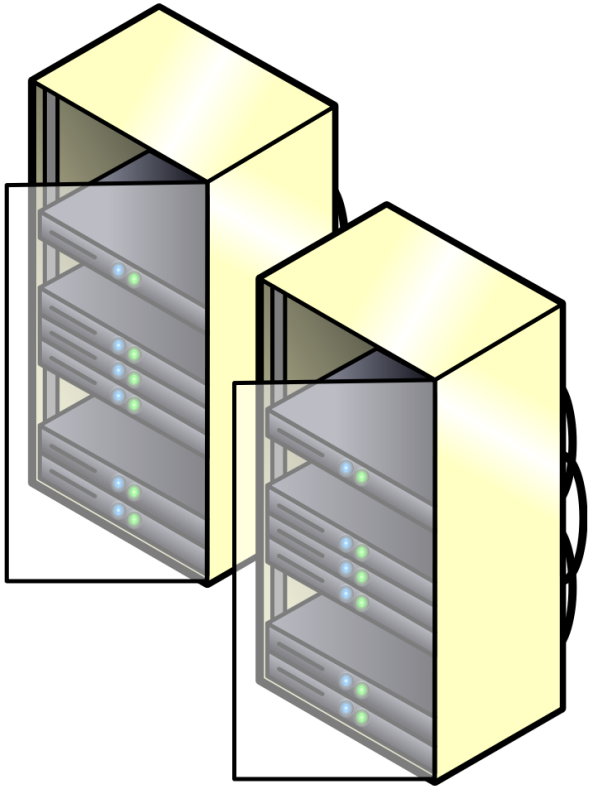
International Stellarator/Heliotron Workshop, 2009

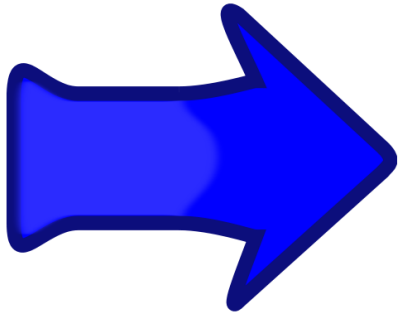
Outline

1. Introduction
2. Method
   - Overview
   - Preprocessing
   - Clustering
3. Results
   - H-1
   - TJ-II
   - Heliotron-J

What Is Data Mining?

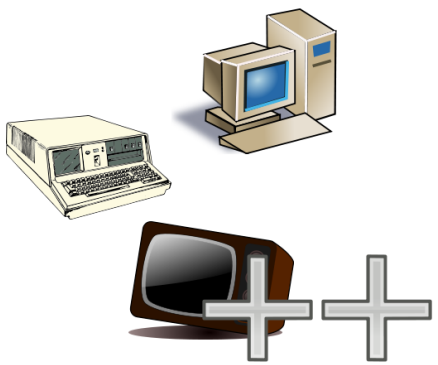
Very Large Databases





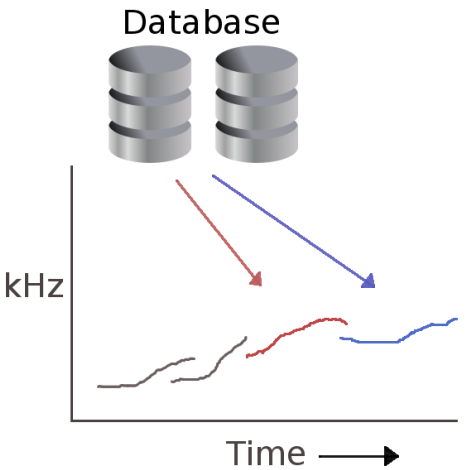
New Knowledge

Motivation



Can we obtain new results from old data?

Can we contribute to other MHD database projects?

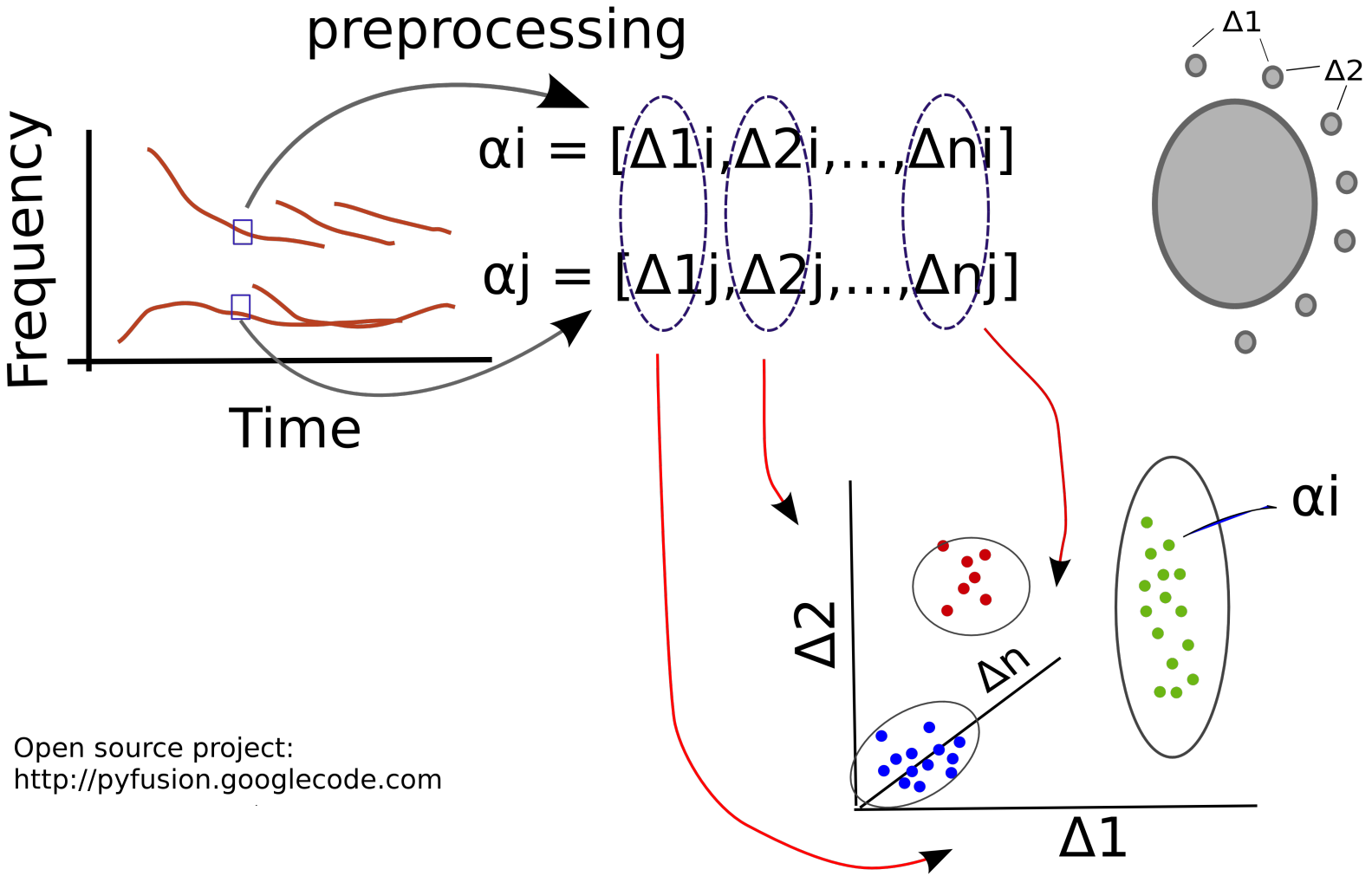


Can we use a MHD documentation database for identification of new data? in real-time?

Outline

1. Introduction
2. Method
   - Overview
   - Preprocessing
   - Clustering
3. Results
   - H-1
   - TJ-II
   - Heliotron-J

Overview



Open source project:
http://pyfusion.googlecode.com

clustering in n-dimensional phase space

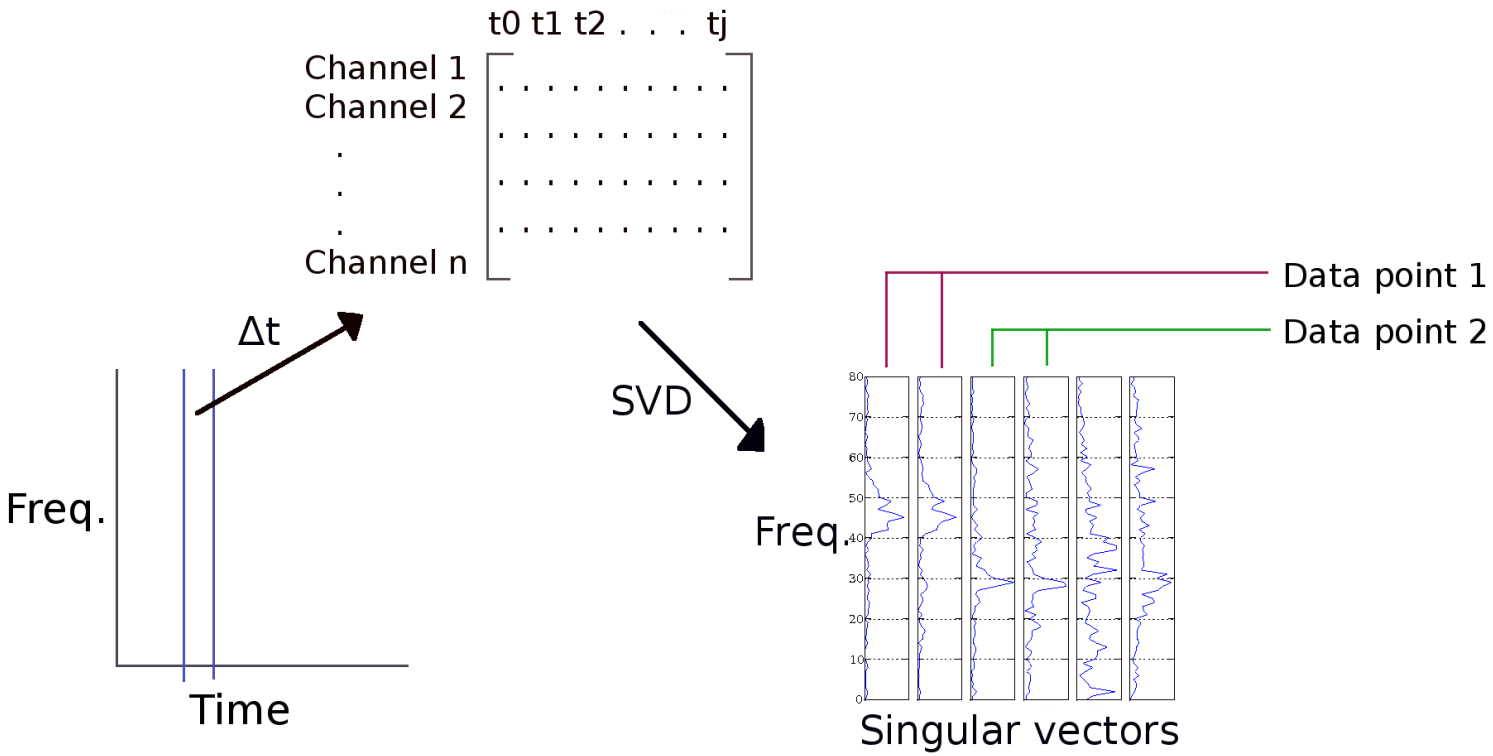
Outline

1. Introduction
2. Method
   - Overview
   - Preprocessing
   - Clustering
3. Results
   - H-1
   - TJ-II
   - Heliotron-J

Preprocessing

1) Dataset selection: choose shots with some set of Mirnovs at sufficient sampling rate.

2) Separate and filter modes, map to $\Delta\varphi$-space



Outline

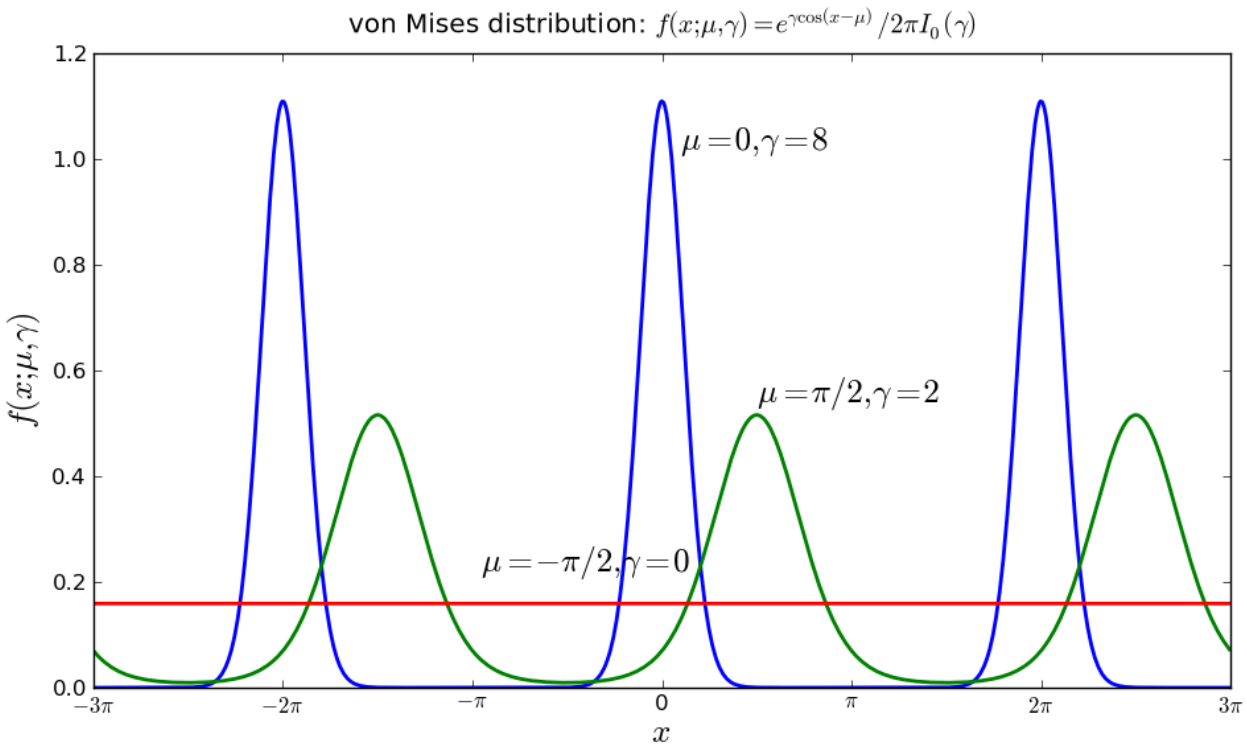
1. Introduction
2. Method
   - Overview
   - Preprocessing
   - Clustering
3. Results
   - H-1
   - TJ-II
   - Heliotron-J

Clustering 1

- Common approach, *Expectation Maximisation* algorithm, with Gaussian clusters.
- But, $\Delta\phi$ is periodic. $\Delta\phi_i = \Delta\phi_i \pm 2n\pi$
- Solution 1: use $\sin(\Delta\phi)$, $\cos(\Delta\phi)$, however Gaussian approximation is poor as $\sin, \cos \to \pm 1$.

Clustering 2

- von Mises distribution is periodic analogue of Gaussian: $f(\Delta\phi; \mu, \gamma) = \frac{e^{\gamma \cos(\Delta\phi - \mu)}}{2\pi I_0(\gamma)}$. $I_0(\gamma)$ is Bessel function of order 0.
- Bessel function complicates EM
- Now using Minimum Message Length (MML) clustering with von Mises distributions.

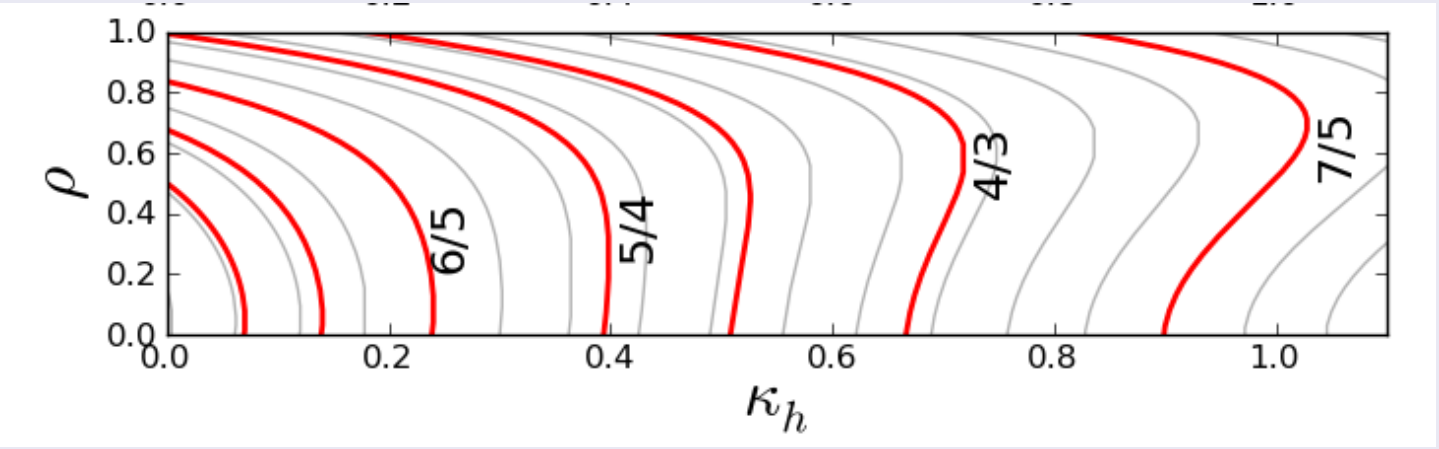


Outline

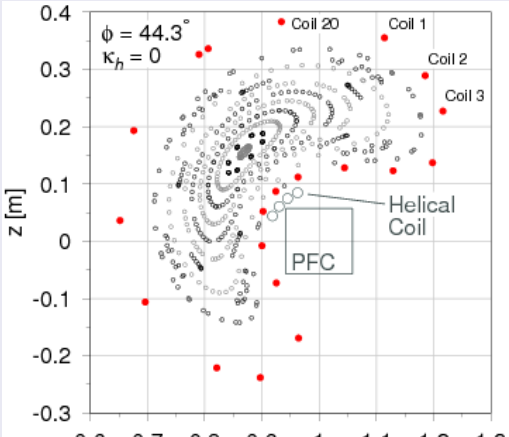
1. Introduction
2. Method
   - Overview
   - Preprocessing
   - Clustering
3. Results
   - H-1
   - TJ-II
   - Heliotron-J

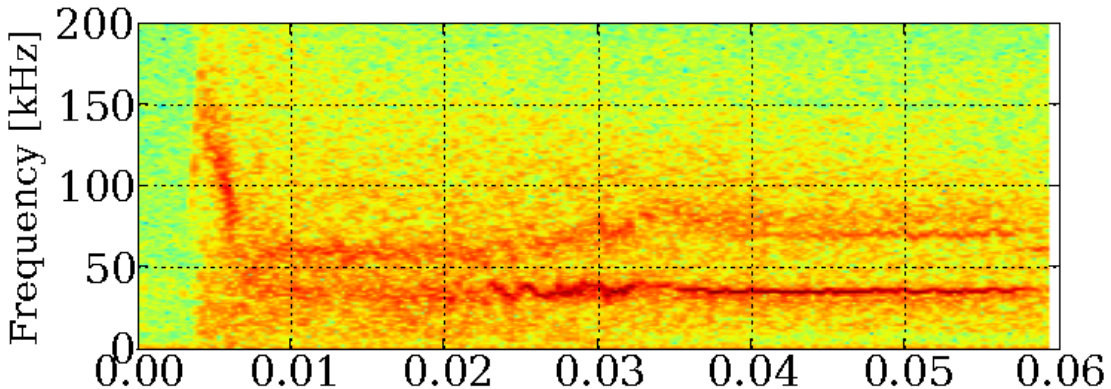
Rotational transform (ι) parameter scan

Configuration scan



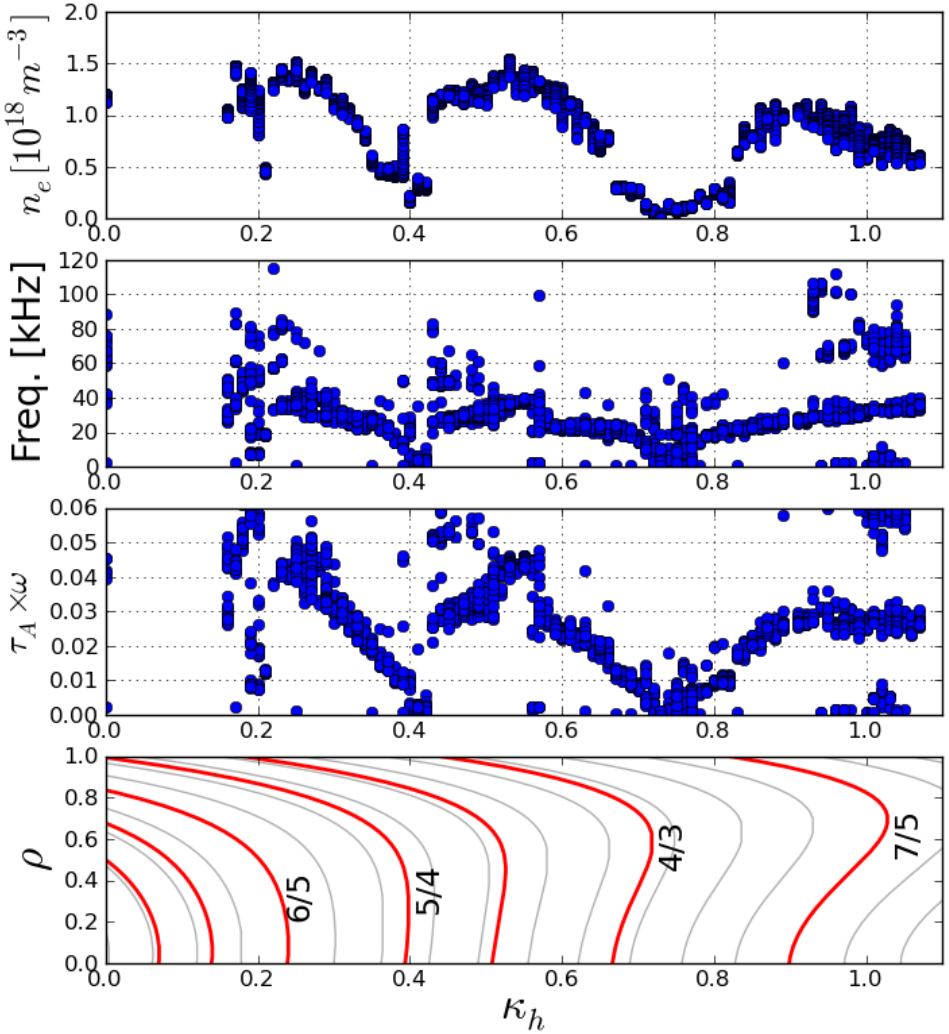
Mirnov coils



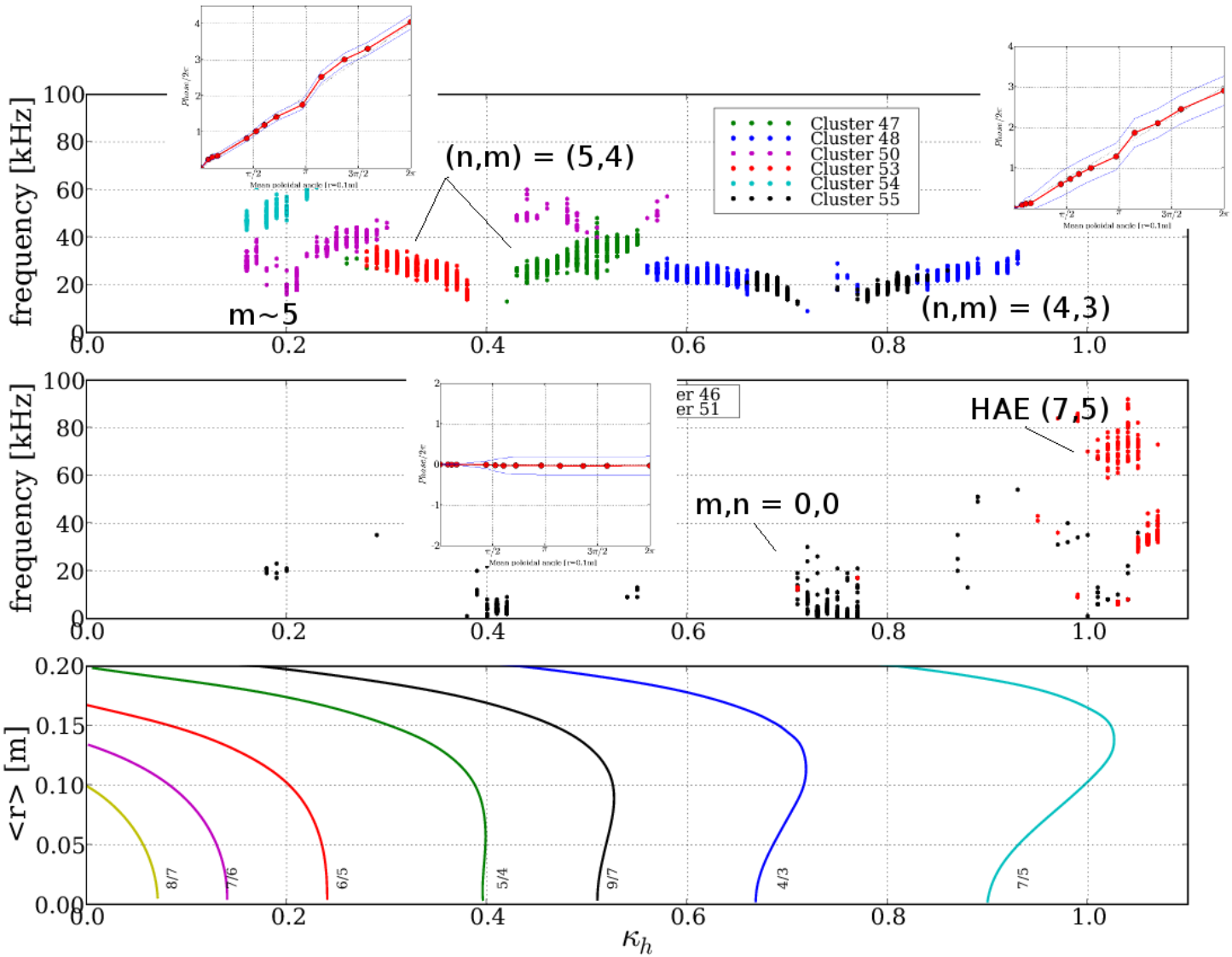


H-1 ι-scan, data mining

- Each data point from 1 ms time sample, each κ_h coordinate is a separate shot.
- Scaled by Alfvén transit time $\tau_A = R_0\sqrt{\mu_0\rho}/B_0$, the resonance modes are GAE-like ($f \propto |n - \iota m|$), but at 1/3 frequency.

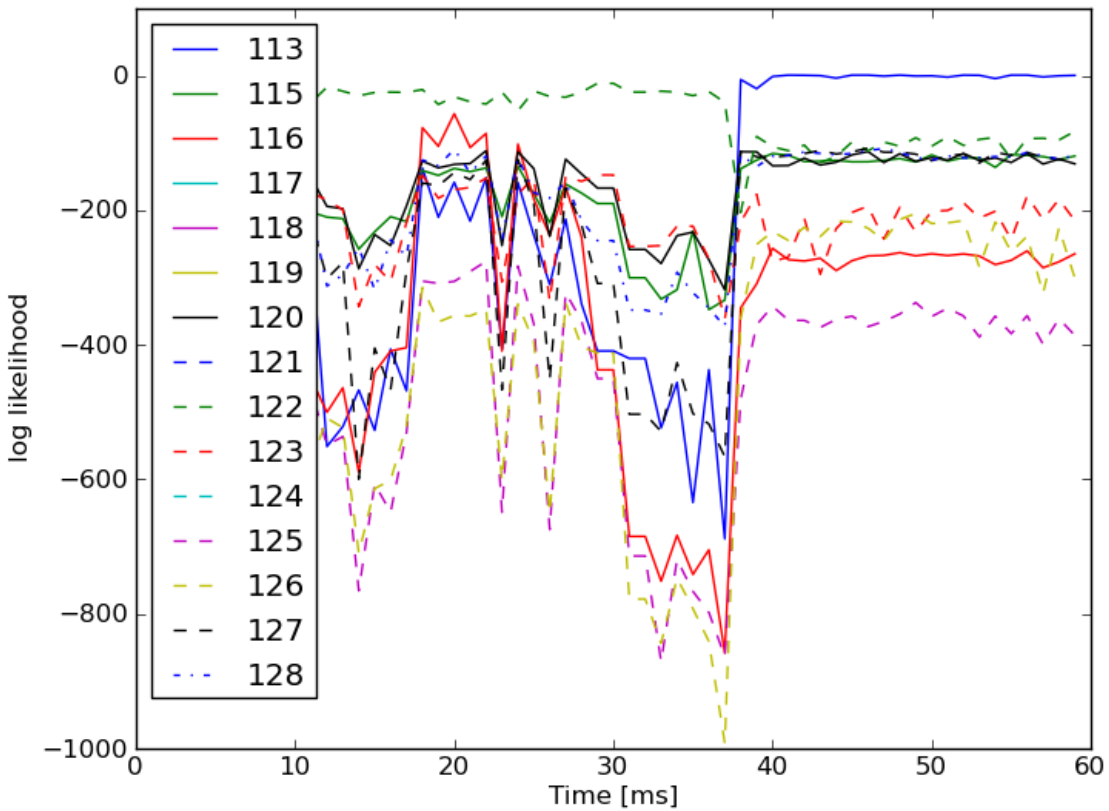


H-1 t-scan, results



H-1, probabilistic mode identification

Because modes are represented as multivariate von Mises distributions, we can trivially compute the likelihood of any new data being of a certain type of documented mode.

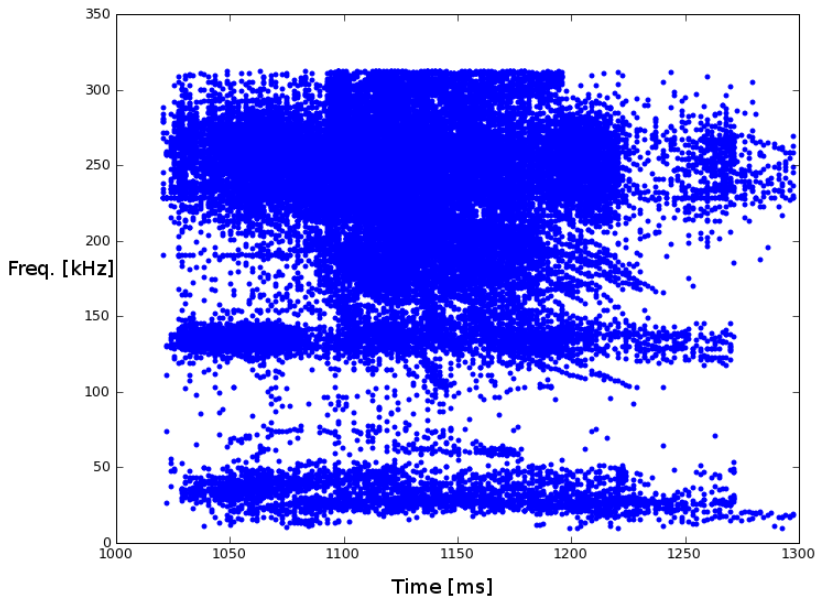


Outline

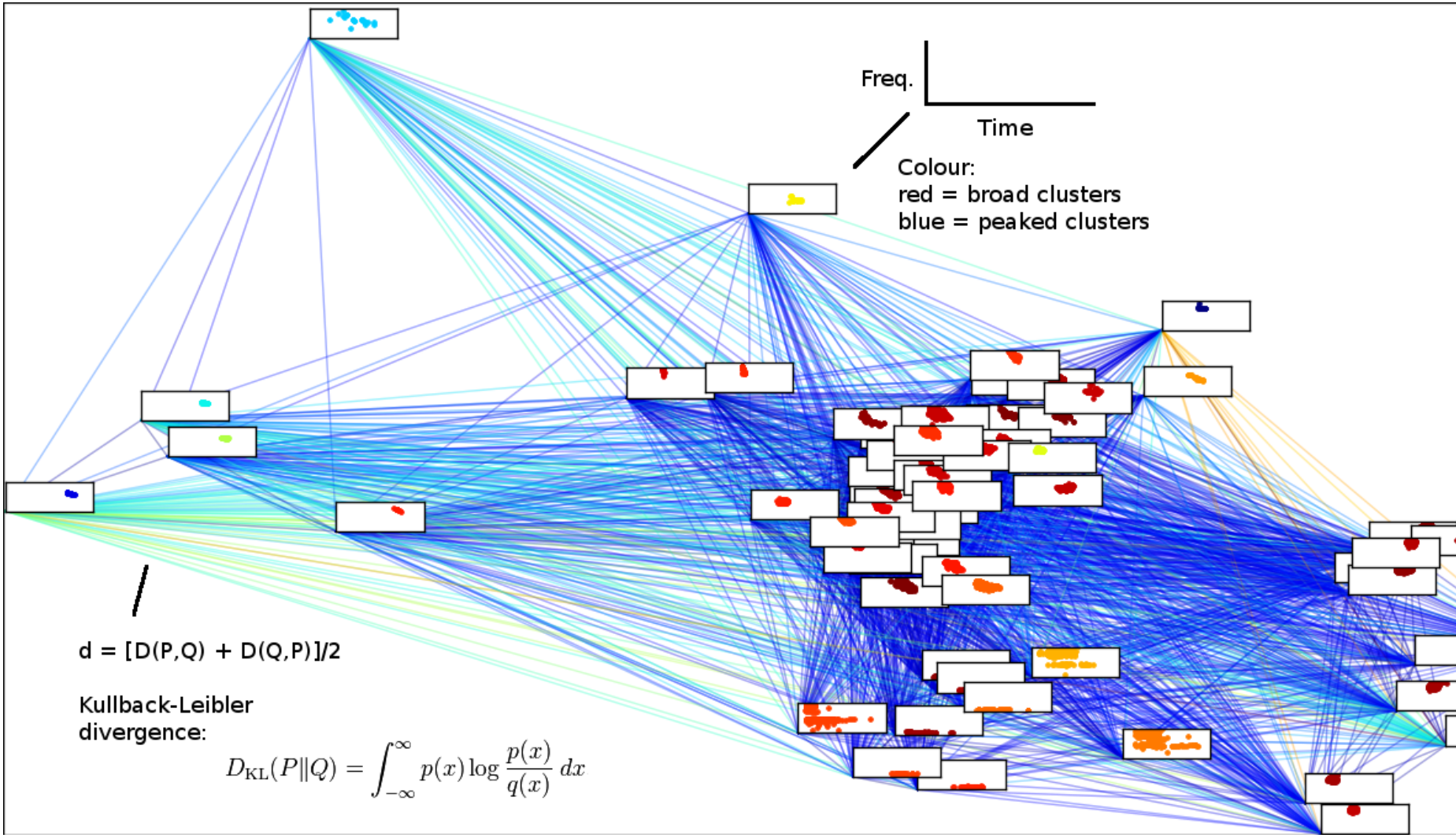
1. Introduction
2. Method
   - Overview
   - Preprocessing
   - Clustering
3. Results
   - H-1
   - TJ-II
   - Heliotron-J

TJ-II Dataset

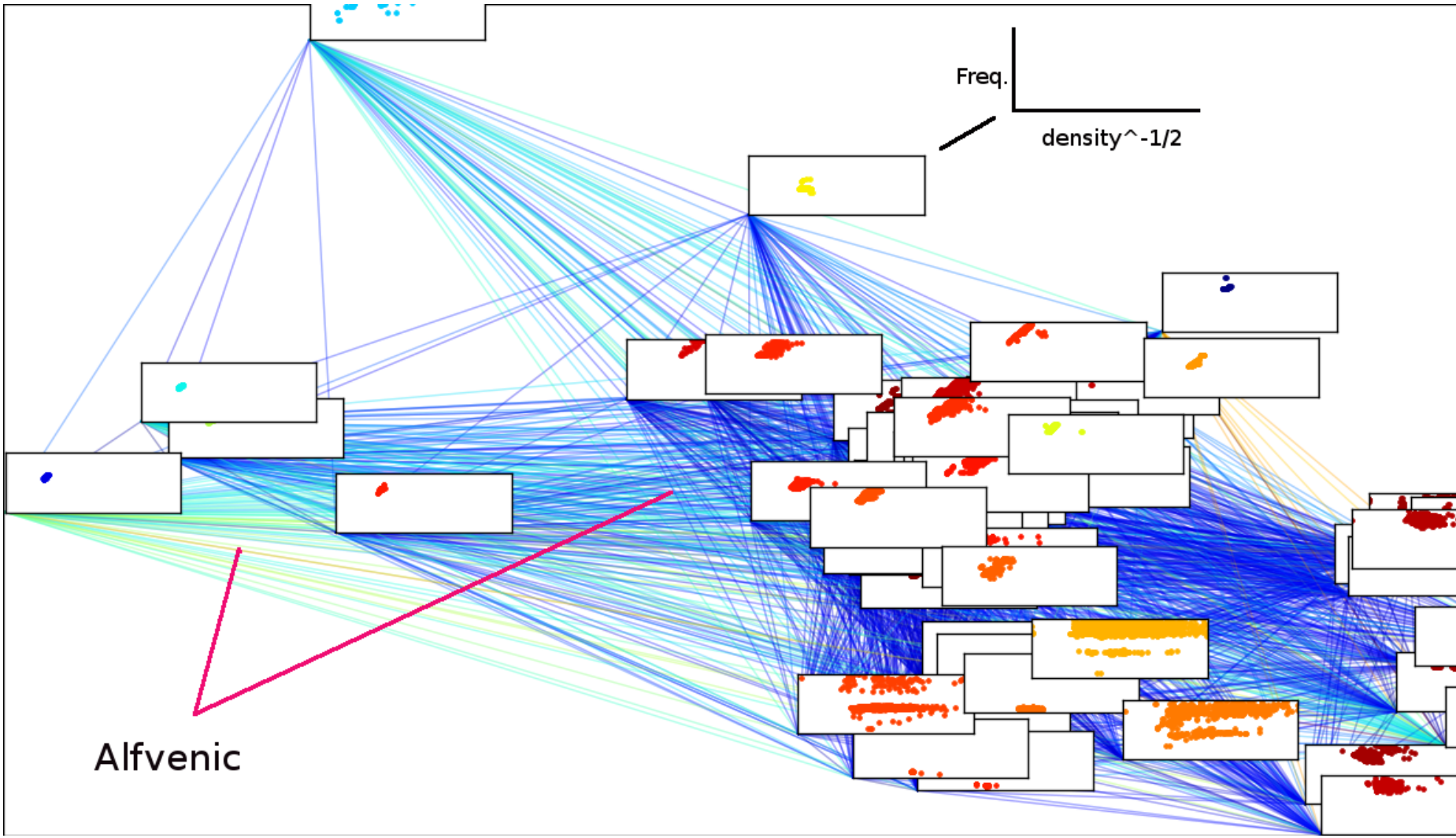
- Several datasets (different sets of Mirnovs) have been considered in TJ-II. Largest portion of database: 15 poloidal Mirnov coils.
- 3753 shots. 7.1×10^6 datapoints $\rightarrow$ 3.2×10^4 after filtering ($f > 2\,\text{kHz}$, $p > 0.3$, $S_{Mir1(RMS)} > 1.2$)



TJ-II Clusters showing time-freq.

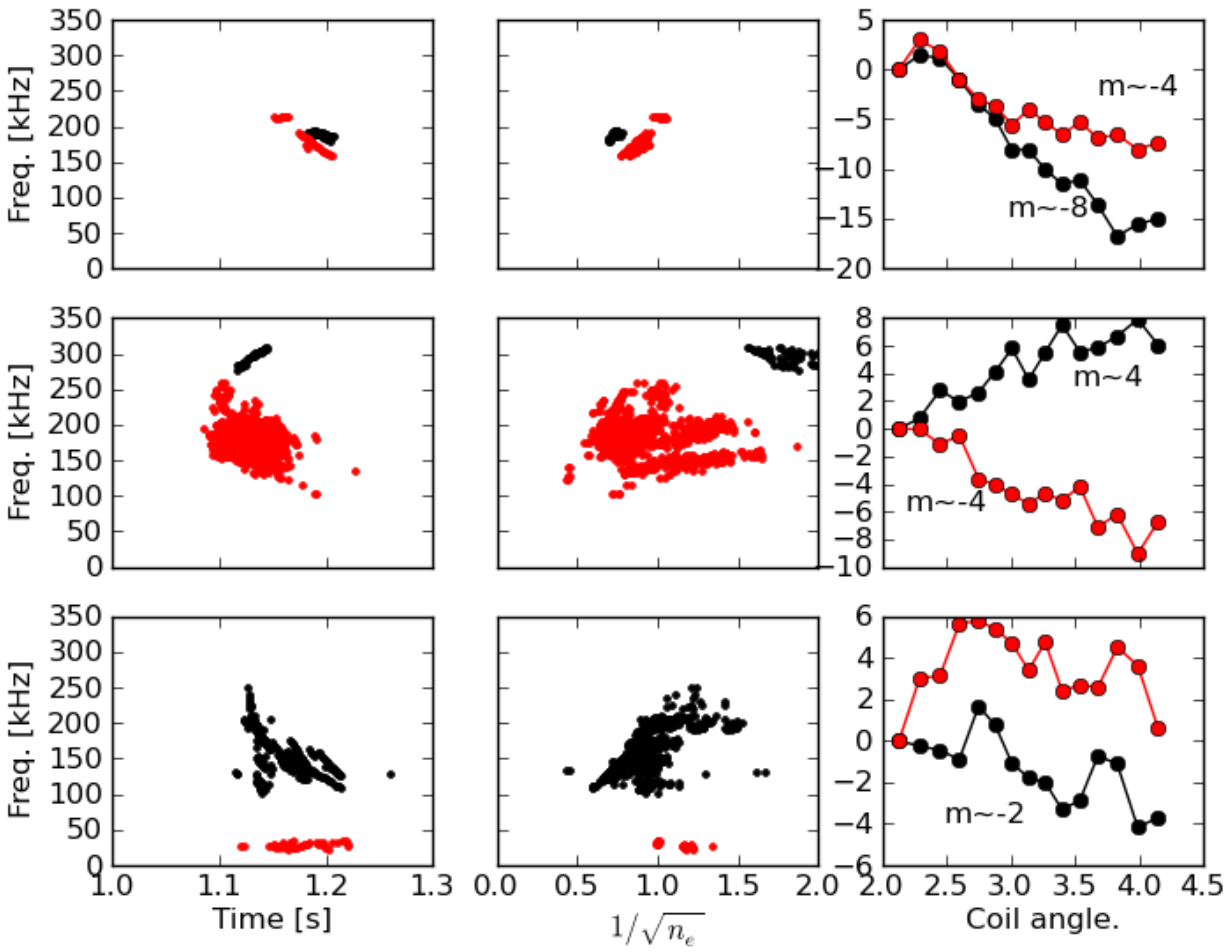


TJ-II Clusters showing $n_e^{-1/2}$-freq.



TJ-II Mode structure

Various TJ-II modes. Phase-angle defined by von Mises clusters.

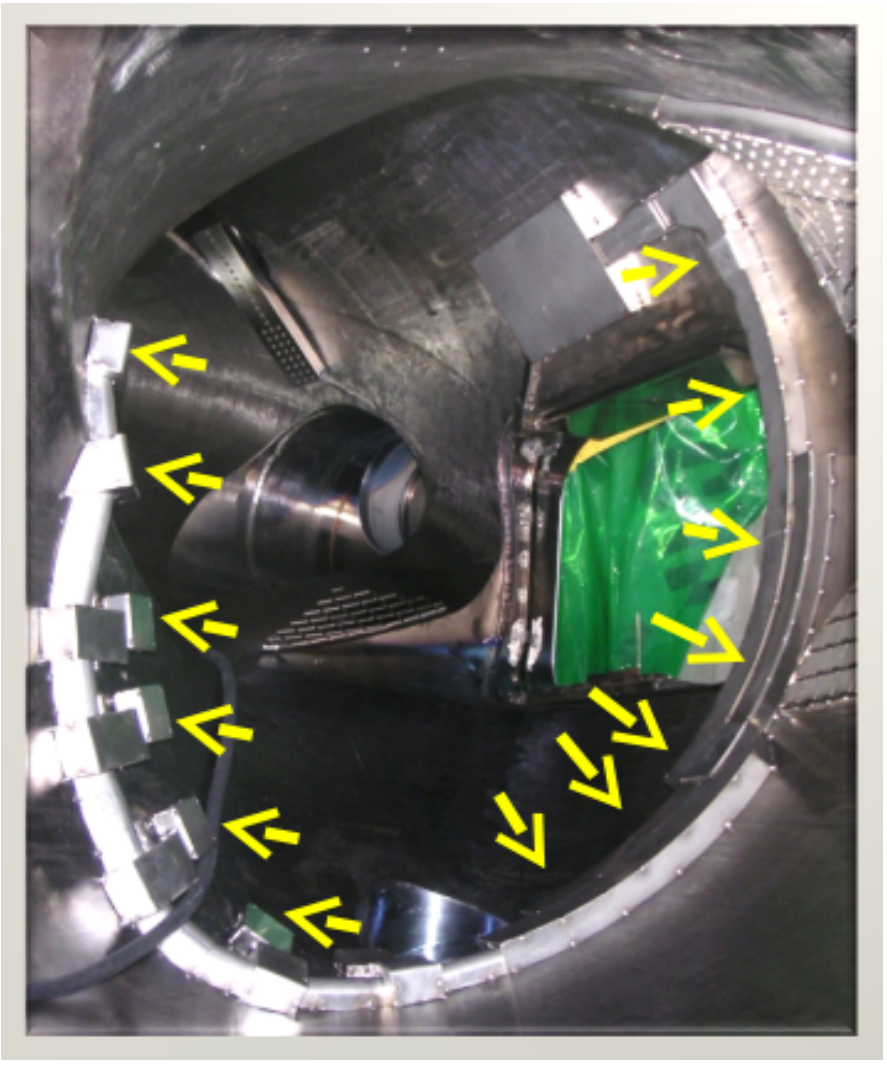


Outline

1. Introduction
2. Method
   - Overview
   - Preprocessing
   - Clustering
3. Results
   - H-1
   - TJ-II
   - Heliotron-J

Heliotron J dataset

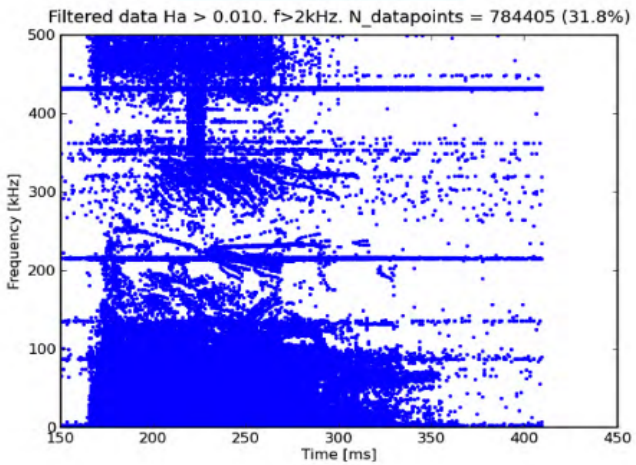
- 14 Mirnov coils with same timebase $\rightarrow$ 3786 shots
- Using 1ms samples, get 2.5 million datapoints (incl. pre-, post-shot noise etc.)



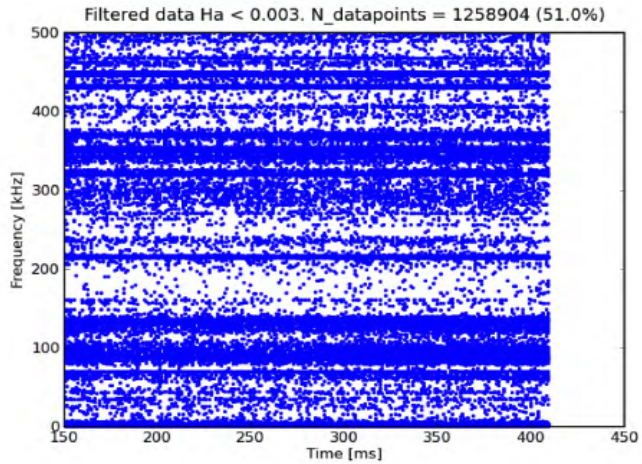
Heliotron J, filtering

Remove non-plasma signal noise by keeping only time-slices where $H_\alpha > 0.01$

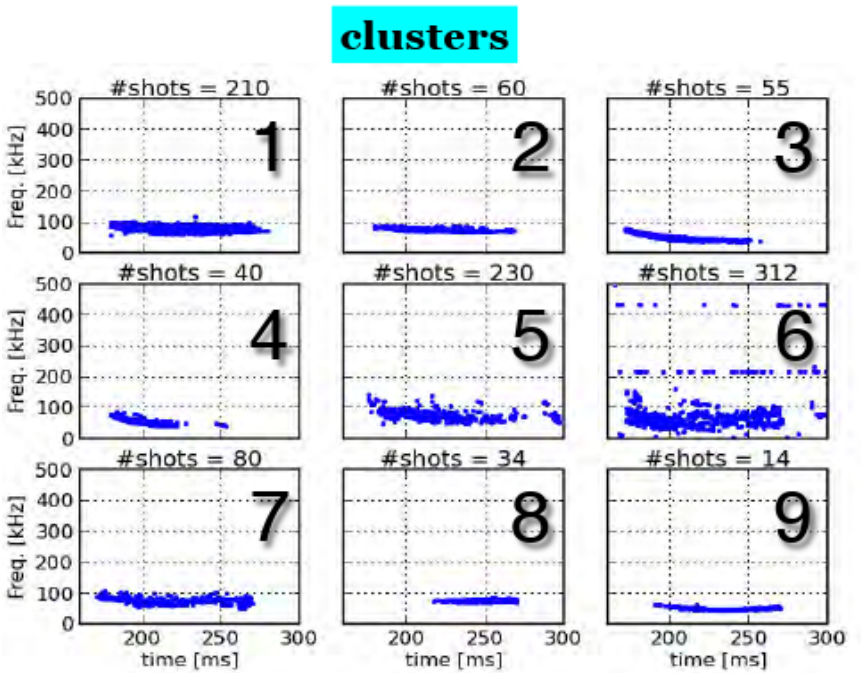
Hα filtered data

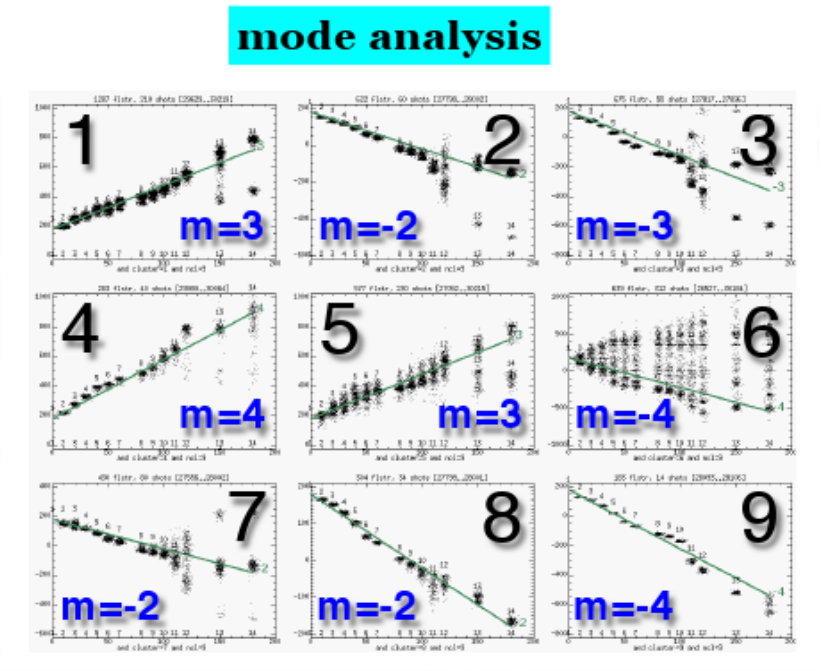


Removed "noise"



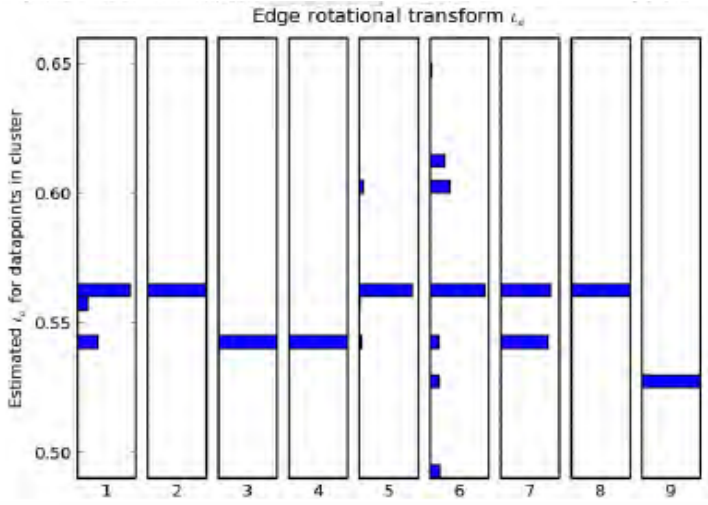
Heliotron J results

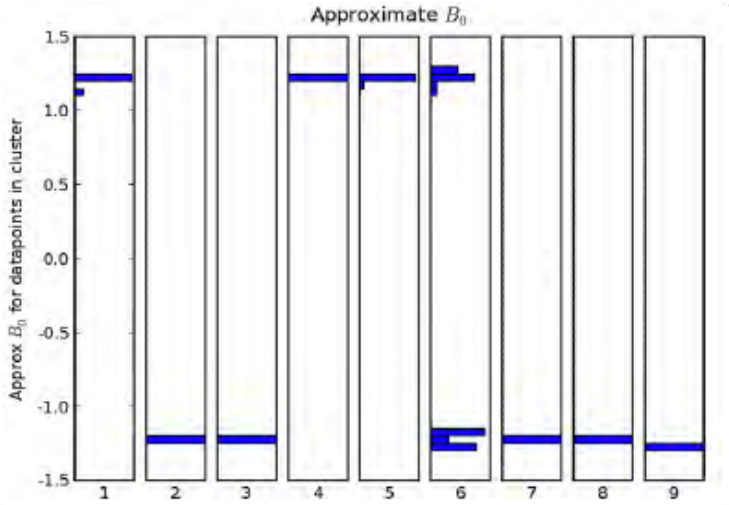




Heliotron J, mode parameters

B_0 flips phase-angle plot: modes propagate in ion diamagnetic drift direction.





Summary

- We have a working, scalable method for extraction and unsupervised classification (clustering) of coherent MHD activity in fusion databases.
- Clustering results have been used to investigate physical properties of the modes.
- We have demonstrated using clustering results for probabilistic identification of modes in new data.

Status and Plans

H-1

Use new Mirnovs. Explore mass, B dependence of modes.

Heliotron J

Expand database. Identify low, broad frequency MHD activity

TJ-II

Find dependence of modes on other parameters, diagnostics.

LHD/CHS

Status: Dedicated data mining server installed at NIFS.

W7-AS

Status: Viable range of shots selected, working on connection between pyfusion and new W7-AS webservices interface.