a place of mind

Periodicity Searches with Arrays
(a segue to machine learning)

Aaron Berndsen
(UBC)

January 20th, 2013

Aspen, Colorado

with direction from: Ingrid Stairs
lots of help from: Weiwei Zhu (UBC Postdoc)
Erik Madsen (UBC Master's student)

Outline:

- **Survey modes with arrays**
  - New possibility: bispectrum
- **Automated searches for pulsars**
  - Image-based AI search for pulsars using deep-learning algorithms

    (big props: Weiwei Zhu, Ingrid Stairs)

    (smaller props: Erik Madsen)

An array of observing modes

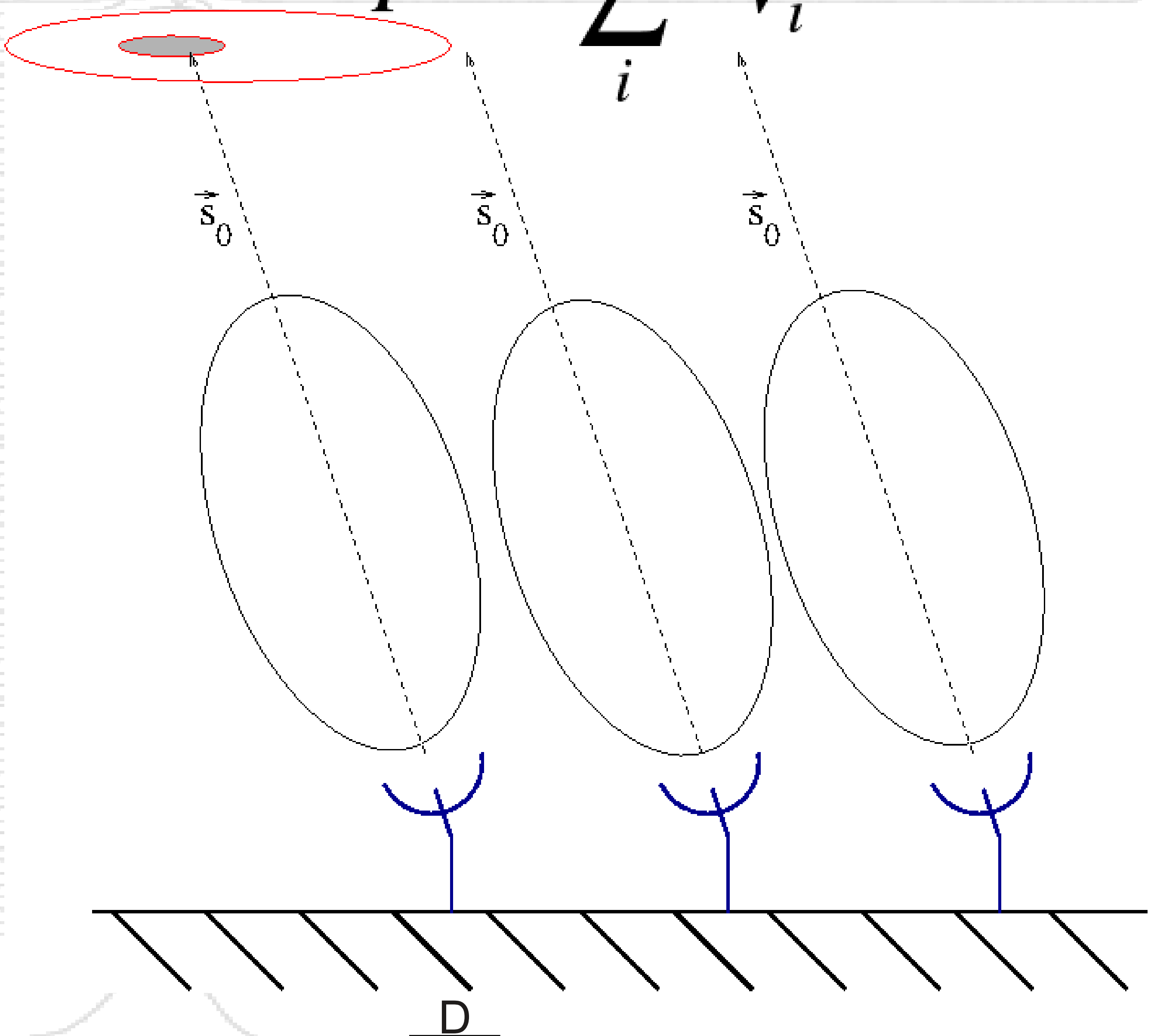
- **Incoherent Sum**

Each element provides an independent observation

Intensity ~ voltage^2

$$\left(I_\nu \propto V_\nu^2\right)$$

$$I \sim \sum_{i}^{N_{\text{antenna}}} V_i^2$$



- Resolution determined by **dish size** D. typically *O*(1deg)

$$\theta_D = 1.22\frac{\lambda}{D}$$

- FoV ~ θ_D

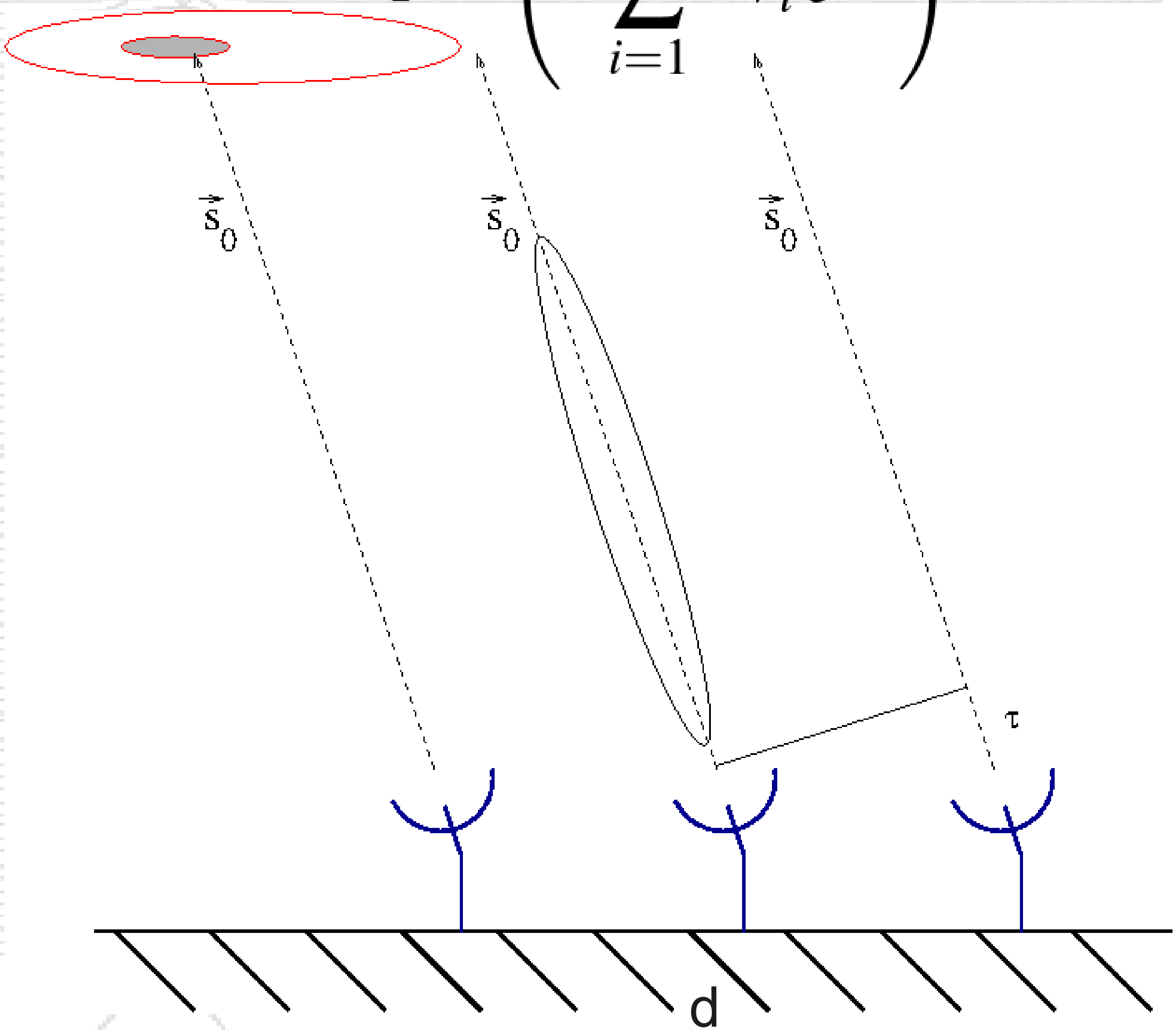
- Sensitivity ~ $\sqrt{N_a}$

- Easy to implement

An array of observing modes

- **Coherent Sum** **(phased array)** Similar to double slit experiment

$$I \sim \left(\sum_{i=1}^{N_{\text{antenna}}} V_i e^{i\phi_i} \right)^2$$



- Resolution determined by **array size** d. typically O(1")

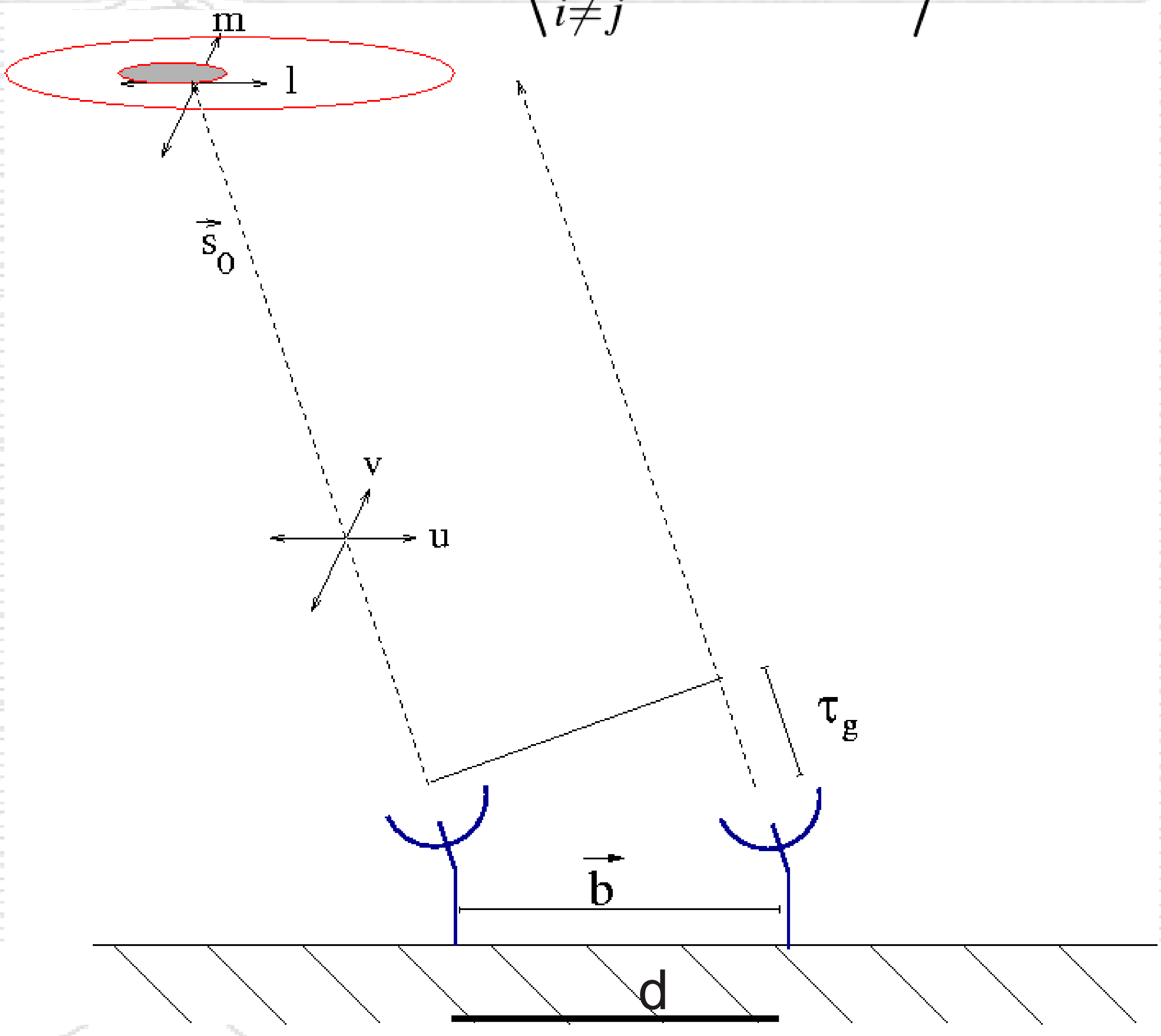
$$\theta_d \simeq \frac{\lambda}{d}$$

- FoV ~ $N_{\text{beams}}\,\theta_d$
- Sensitivity ~ N_{a}
- Survey possible if correlator can form many array beams
- demanding on correlator **and** storage-intensive

An array of observing modes

- **Cross-Correlate** Gives the spatial coherence function

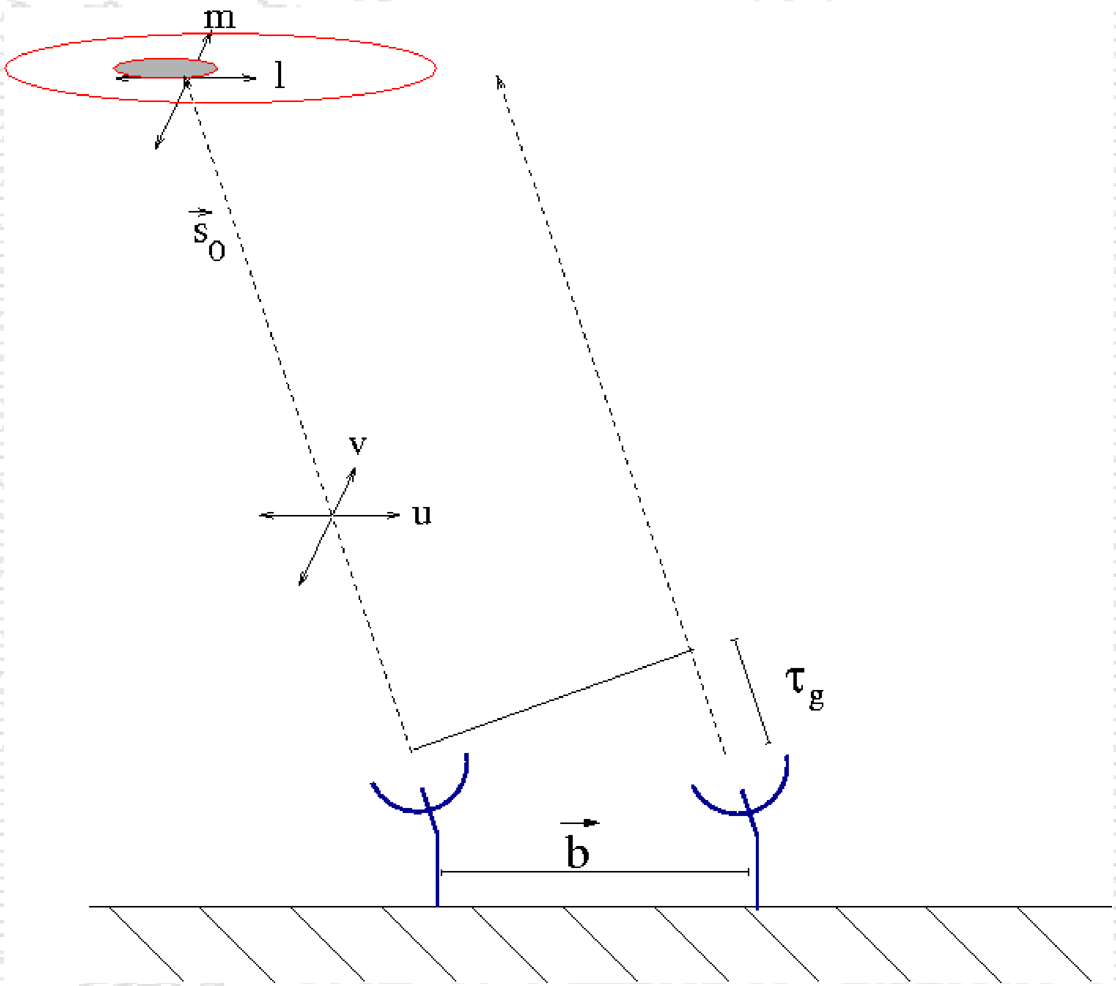
$$\left\langle \sum_{i \neq j} V_i V_j(u,v)_{ij} \right\rangle$$



- Resolution determined by **array size** d
- FoV determined by **dish size** D
- Sensitivity ~ N_a
- Easier on correlator but data intensive
- Pushes complexity to post-processing

An array of observing modes

- **Cross-Correlate**



$$N_{\text{pixels}} \sim \left(\frac{\Theta_{\text{dish}}}{\Theta_{\text{array}}}\right)^2$$

$$\sim \left(\frac{d_{\text{array}}}{D_{\text{dish}}}\right)^2$$

$$\sim \left(\frac{10^4\,\text{m}}{10^1\,\text{m}}\right)^2$$

Ultimately we want to form timeseries.
The images give *O*($10^{5\text{--}6}$) cells (beams) on the sky

Typical survey gets ~ **50 candidates/beam**

Bi-curious?

Periodicity search the bispectrum

$$b_{ijk} = \underbrace{V_{ij} \cdot V_{jk} \cdot V_{ki}}_{\text{measured}} = \underbrace{\tilde{V}_{ij} \cdot \tilde{V}_{jk} \cdot \tilde{V}_{ki}}_{\text{true}}$$

Benefits:

- Removes antenna-based errors

$$\underbrace{V_{ij}}_{\text{measured}} = \underbrace{\tilde{V}_{ij}}_{\text{true}} \underbrace{e^{i(\phi_i + \phi_j)}}_{\text{error}}$$

- $S/N \propto N_a$ Across the primary beam
- It's a single timeseries

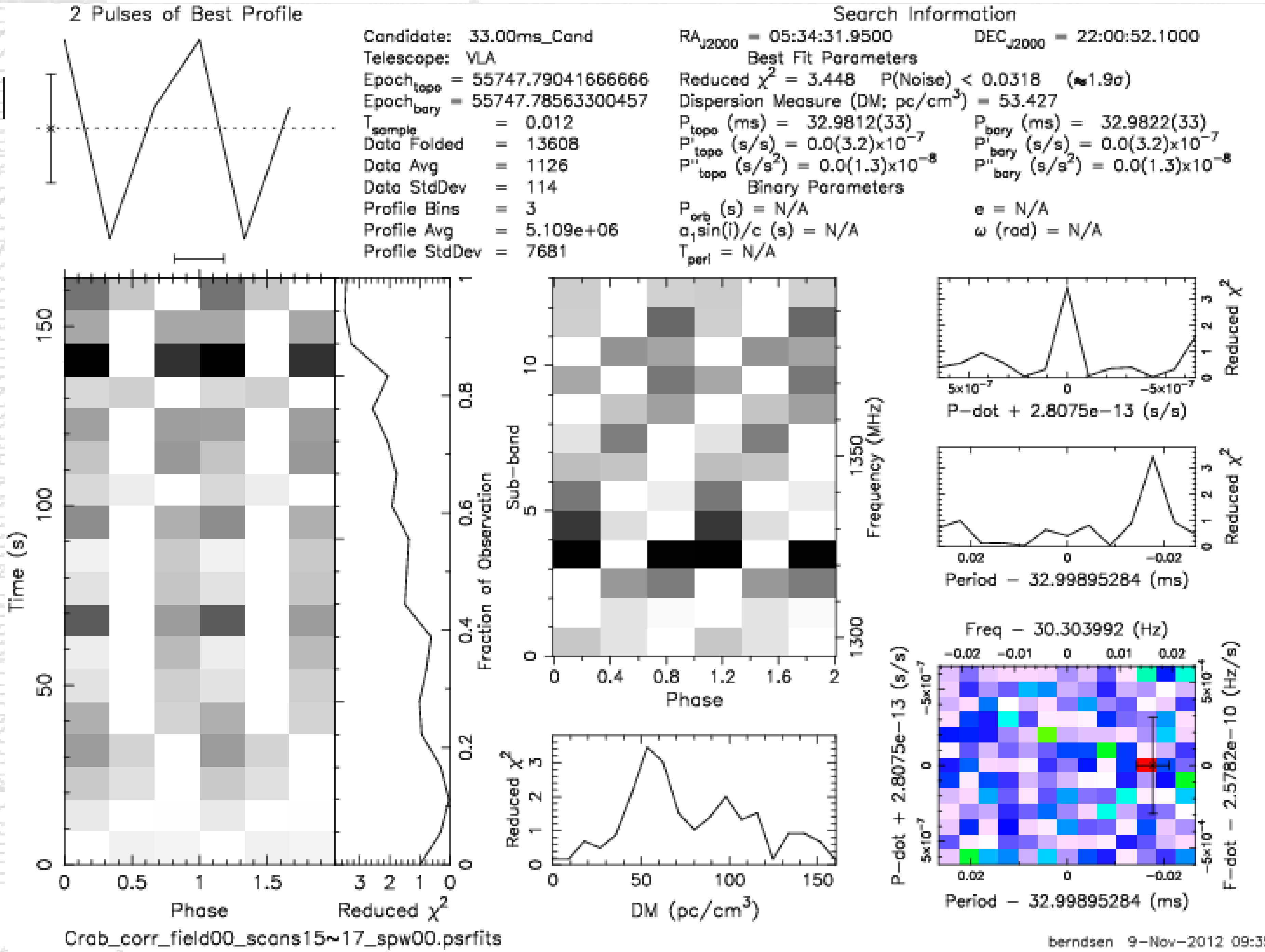
A test set: JVLA commissioning data

- The observations:
  - 7 antenna
  - 64 2 MHz channels (L-band)
  - 12 millisecond samples
  - Tracking the Crab Pulsar

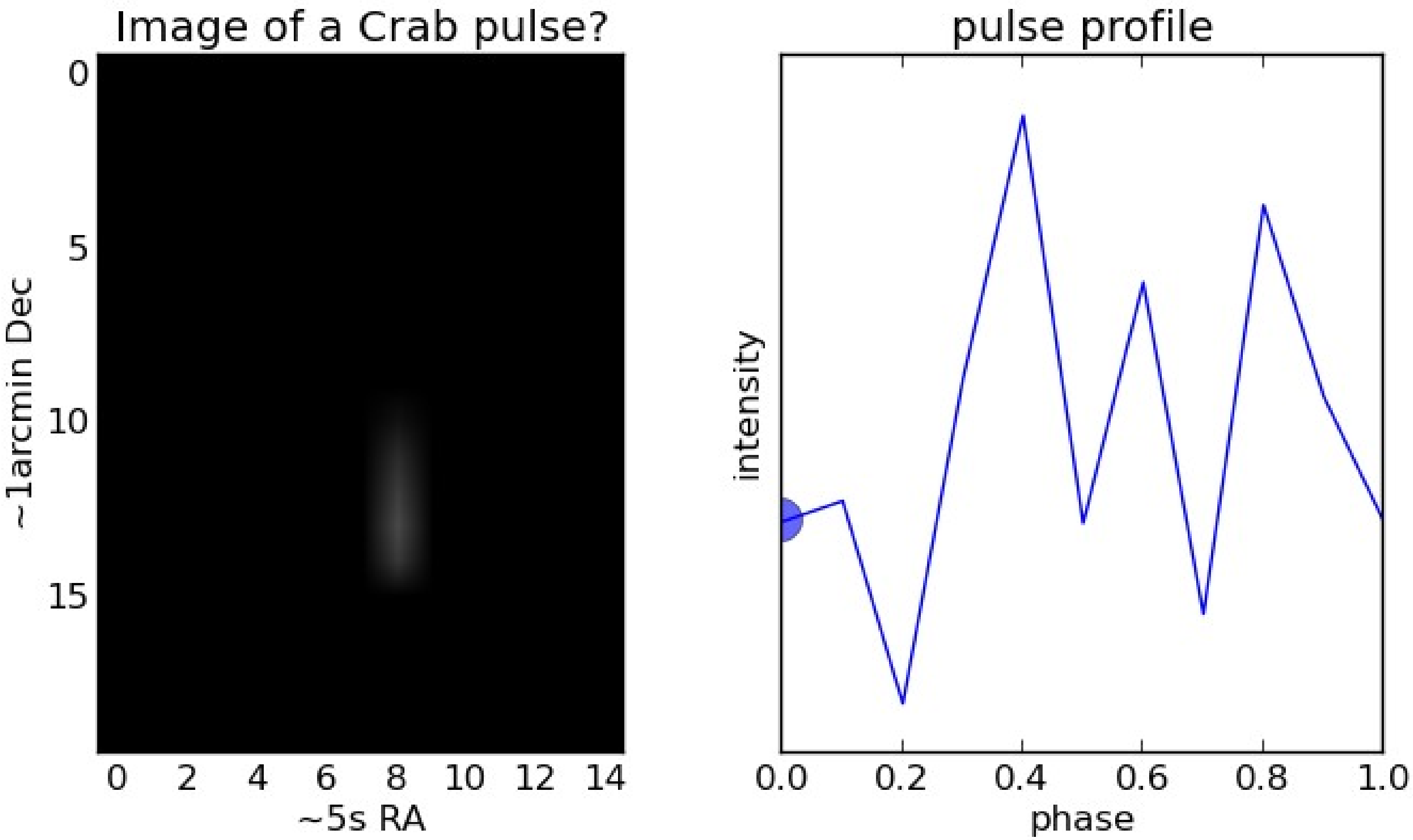
    33 ms period

A test set: JVLA commissioning data

• T

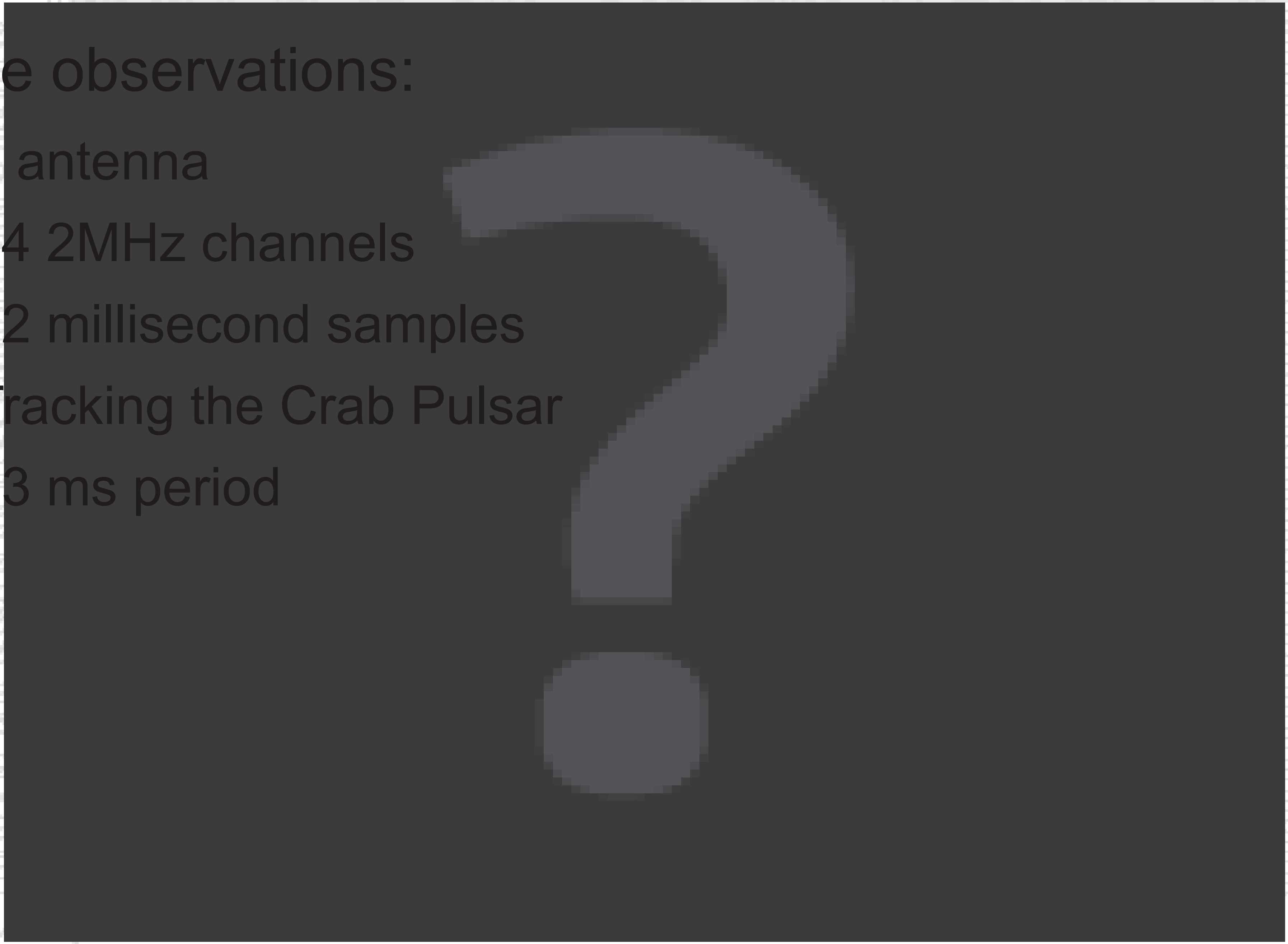


A test set: JVLA commissioning data



- The observations:
  - 7 antenna
  - 64 2MHz channels
  - 12 millisecond samples
  - Tracking the Crab Pulsar

    33 ms period



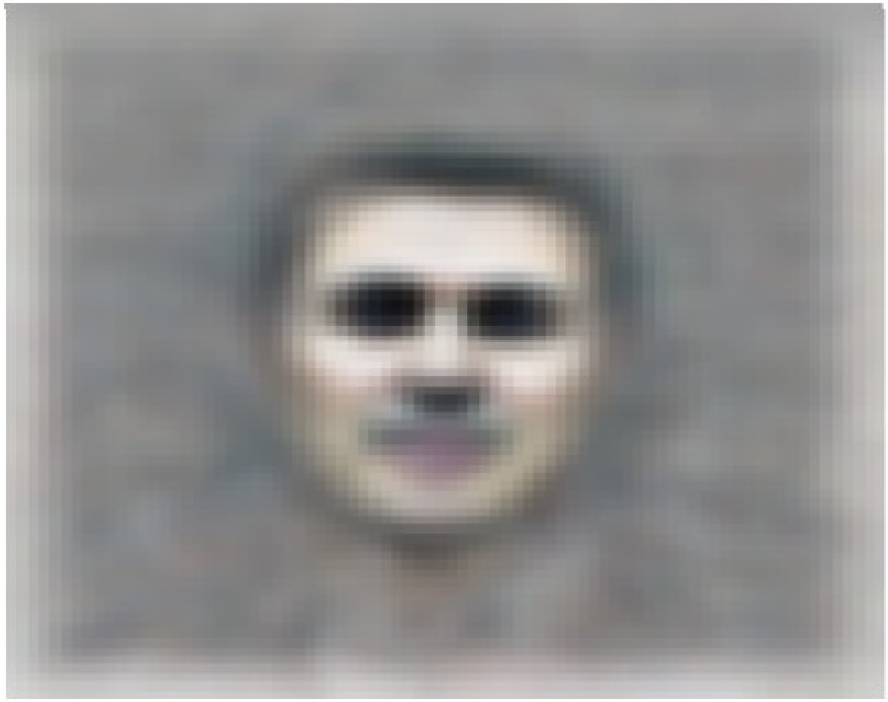
Big data is motivating the use of **automated data analysis**.

Image based Machine Learning (Pattern Recognition)

Using deep learning: The essence of youtube

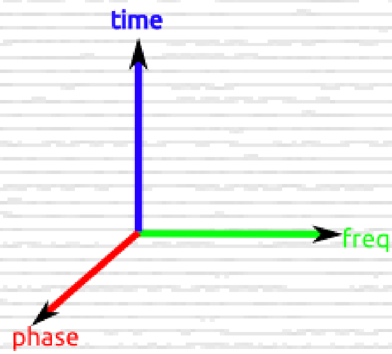
- *Building high-level features using large scale unsupervised learning* arxiv.org/abs/1112.6209v5.pdf (Ng et al.)

Feed 10 million 200x200 pixel images into a Neural Network

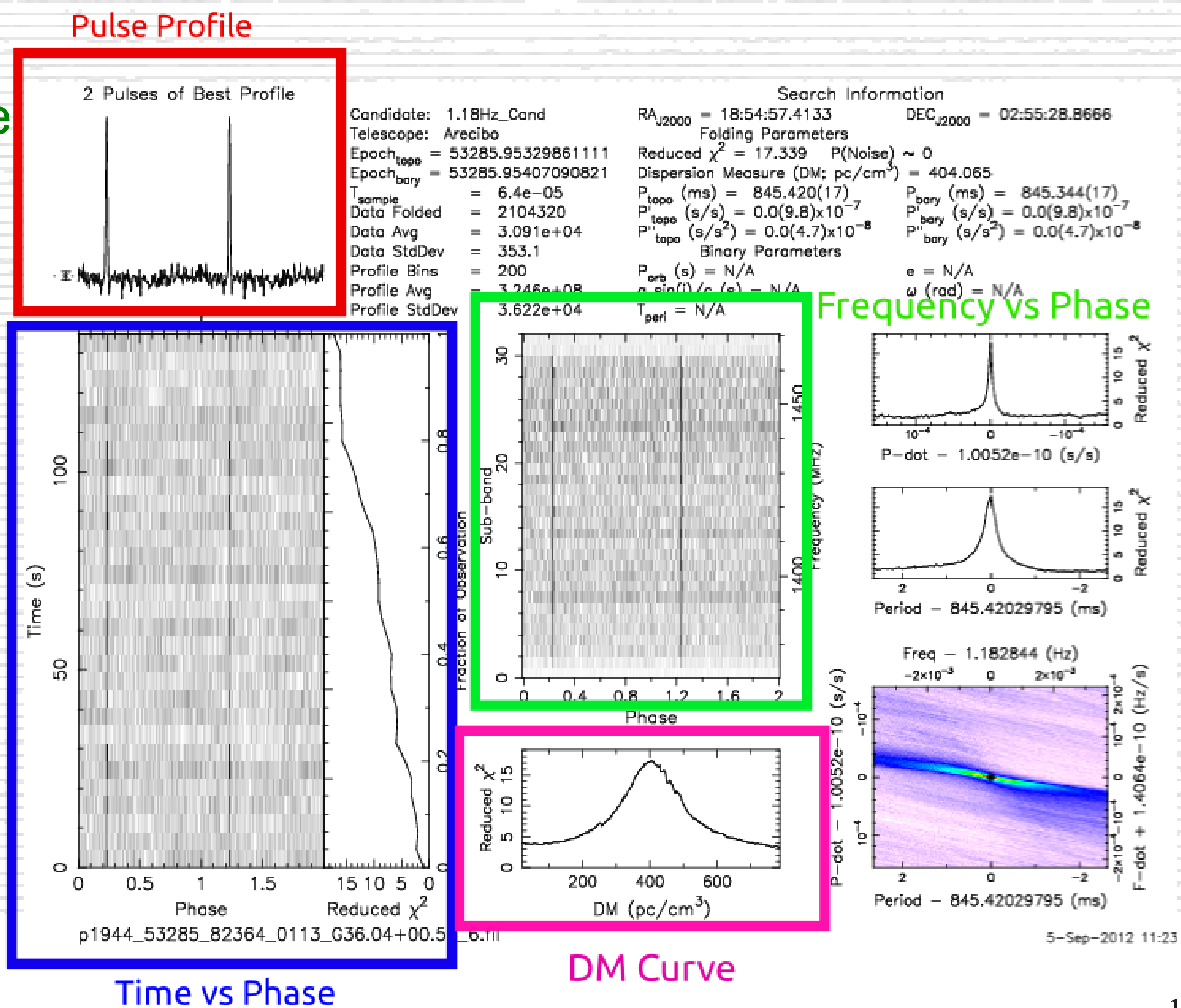


Automated Pulsar Classifier

- A combination of **4 image plots** and **3 machine learning algorithms**:
  - Logistic Regression
  - Support Vector Machine
  - (deep) Neural Network

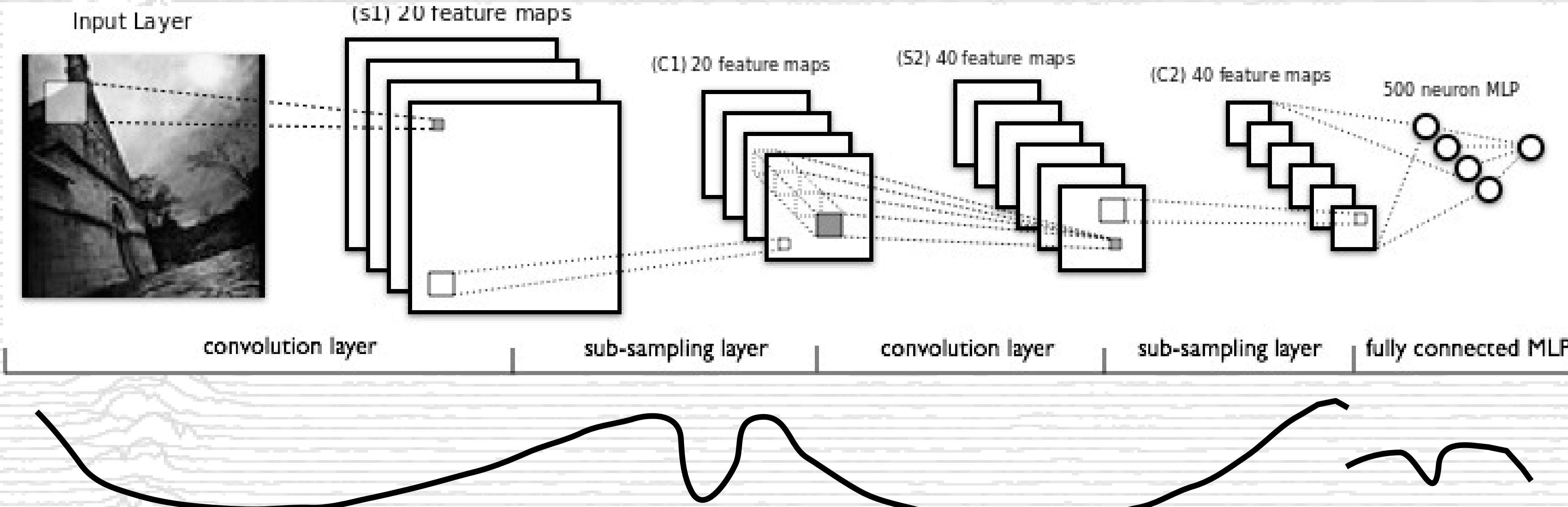


- **9 resultant predictions** are then run through another AI algorithm



Our best sub-classifier:

ConvNet, working on intervals subplot

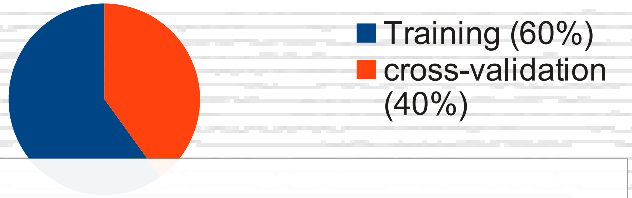


classifier

(Trainable) Feature extraction

- Convolutional layer: apply 4 filters to each **local** patch on the original image
- Subsample and apply another convolutional layer
- hierarchical learning: low-level features to mid-level invariant representations

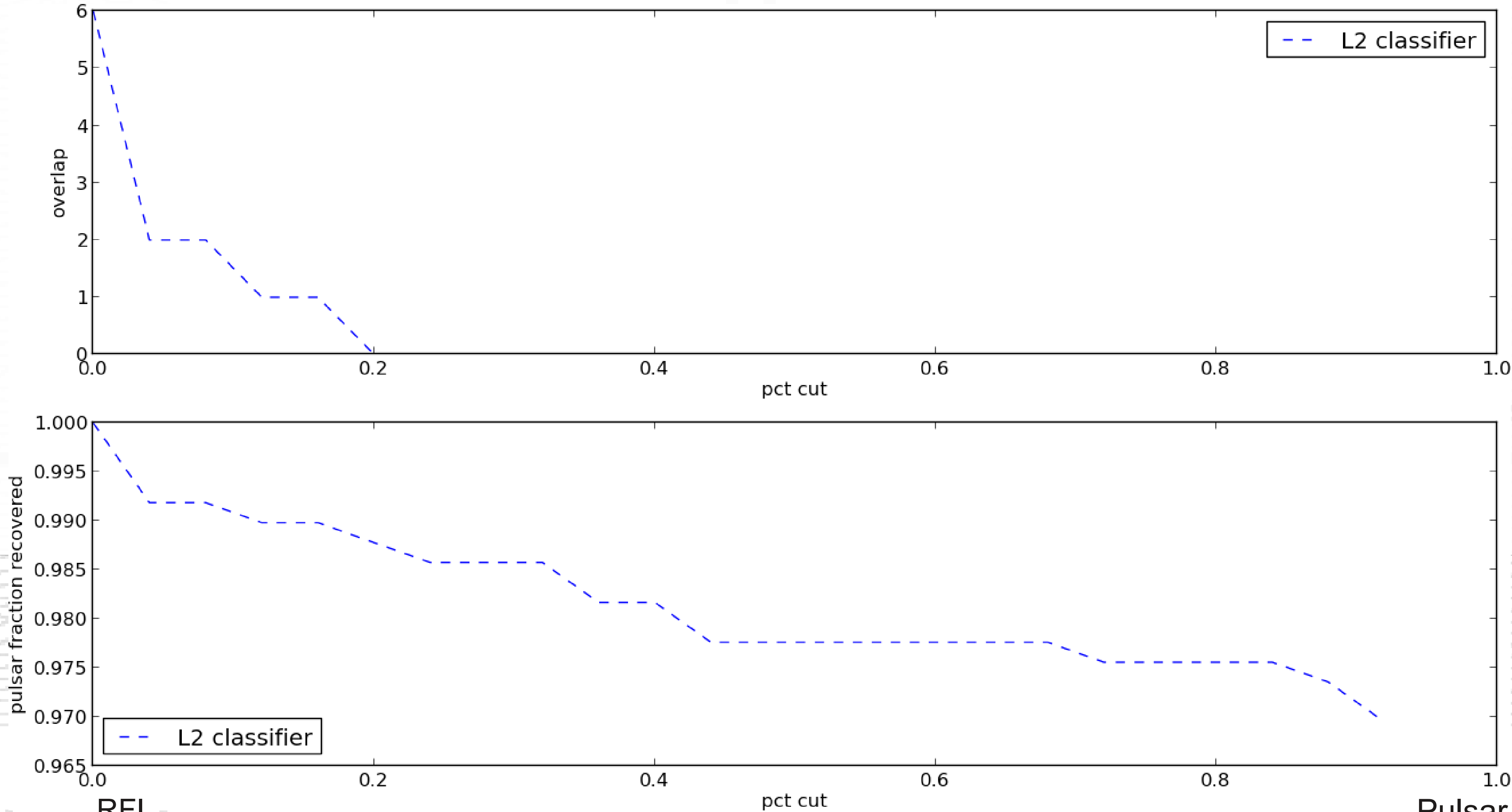
AI Performance



Current Performance:

trained on 355 pulsars, 860 harmonics, 1950 RFIs PALFA candidates

F1 = 0.995



As applied to 600,000 GBNCC candidates

- N(P>.25) = 4934 (recall: we trained on PALFA candidates)

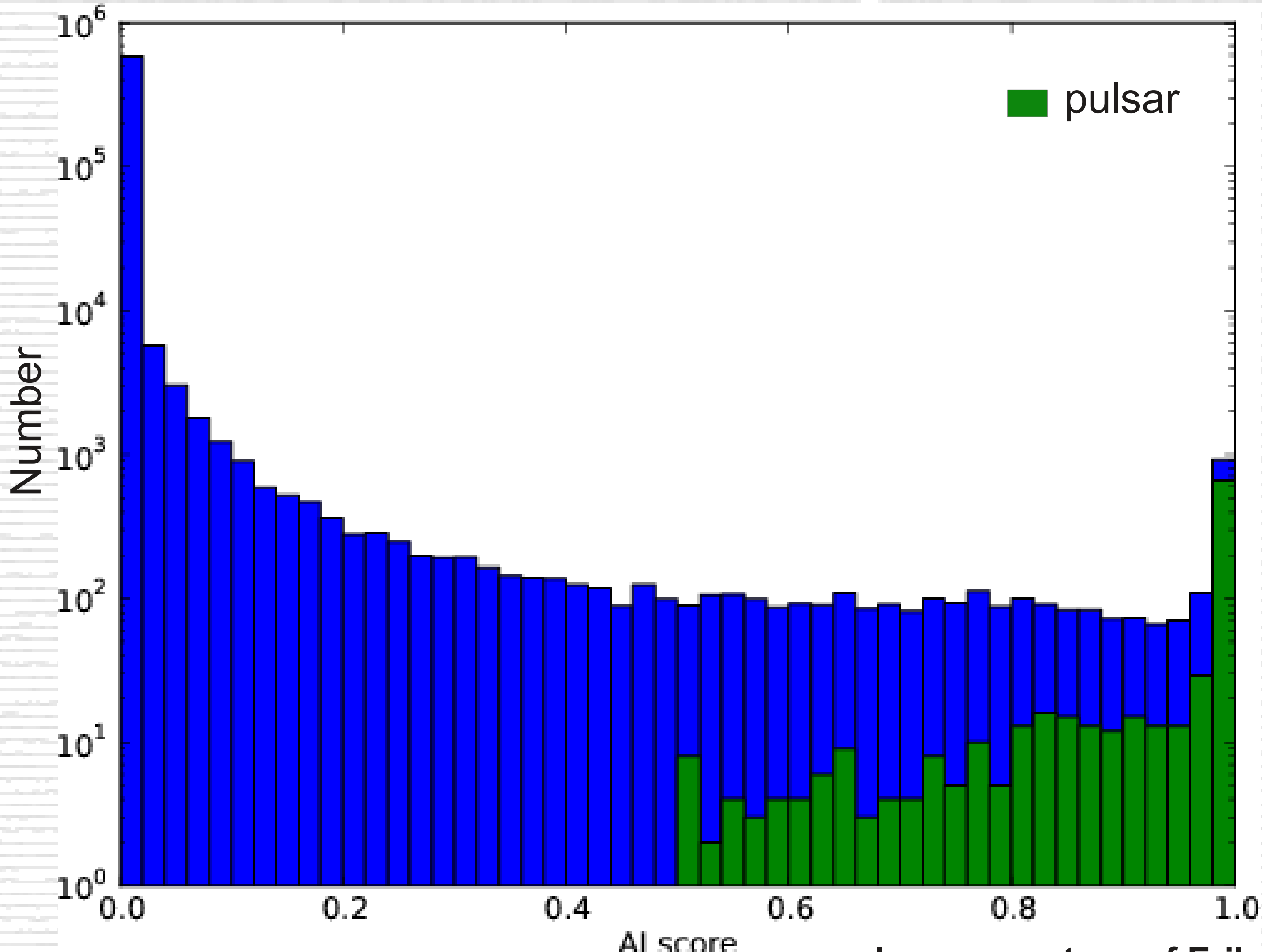


Image courtesy of Erik Madsen

AI has found 3 PALFA pulsars, waiting for follow-up observations on several more

Summary

- The **bispectrum** is a simple timeseries one can search for pulsed candidates

- **AI** and **deep learning** can play a role in pulsar data mining

  (Zhu et al., in prep)

- See the UBC team about incorporating the above into your survey

big thanks to **Weiwei Zhu, Ingrid Stairs and Erik Madsen**