



---

# Deep Learning for Real-time Classification of Astronomical Radio Signals: Current Status

---

Andrei Kazantsev, Ramesh Karuppusamy, Yunpeng Men, Rishi  
Kumar, Michael Kramer

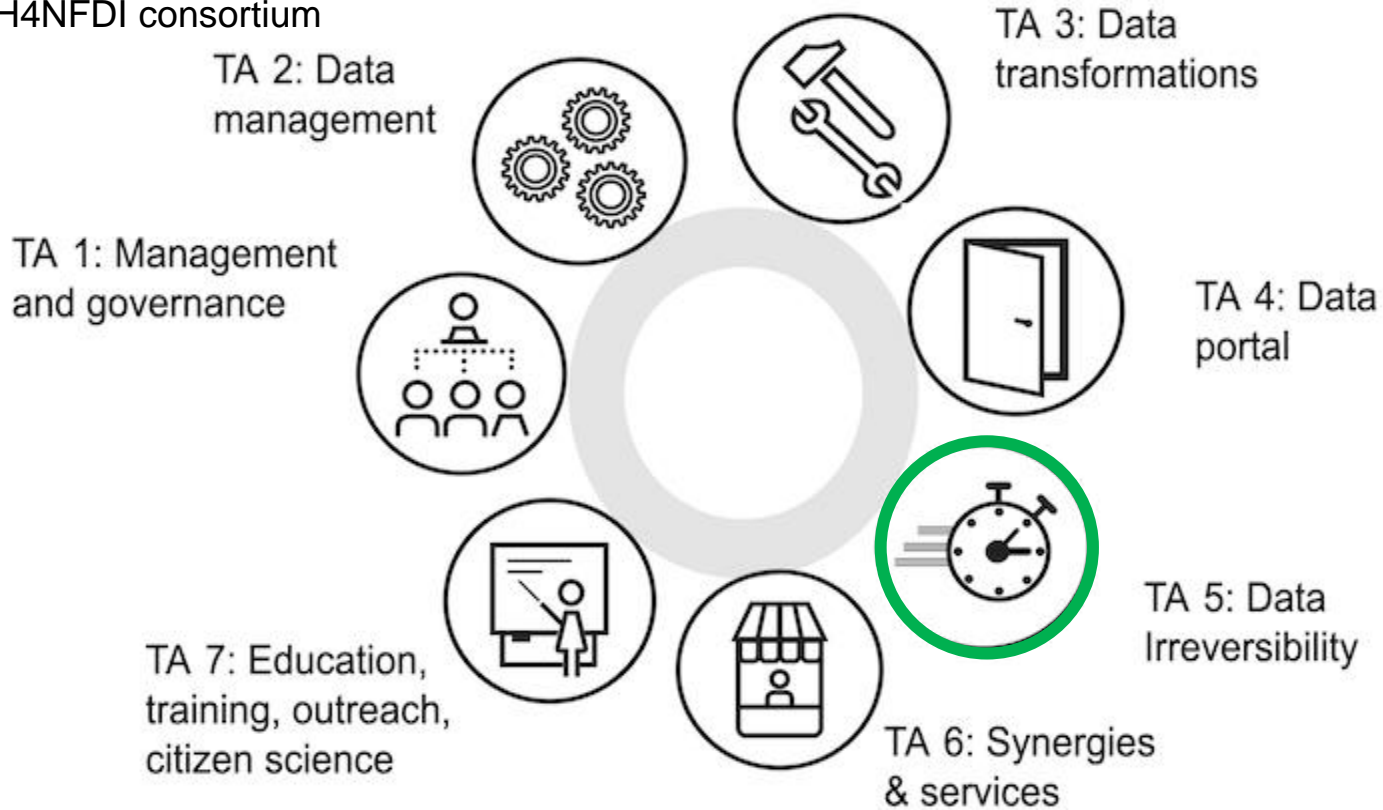
Max Planck Institute for Radio Astronomy  
Bonn, Germany

---



# PUNCH4NFDI consortium

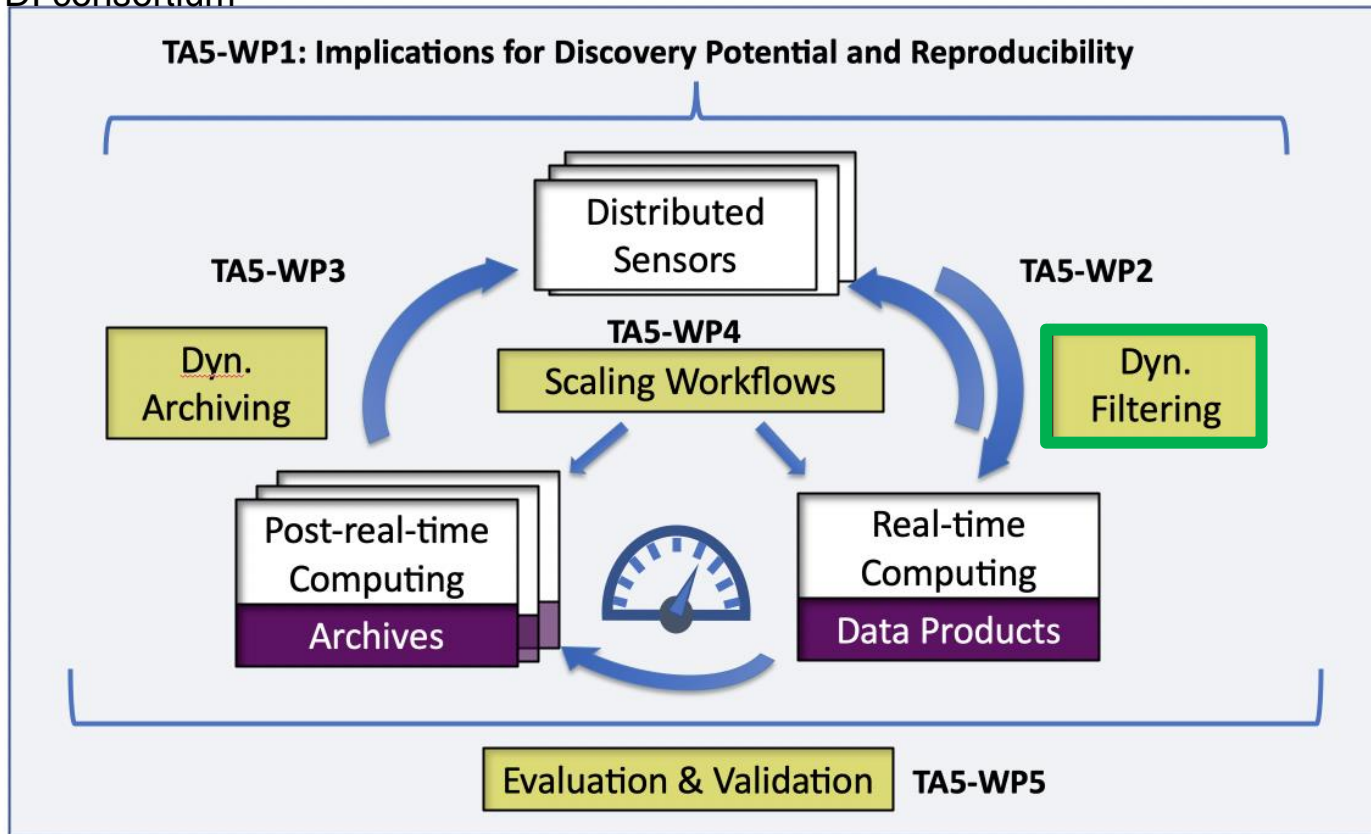
©: PUNCH4NFDI consortium



The task areas of the PUNCH4NFDI consortium

# Work Packages in Task Area 5

©: PUNCH4NFDI consortium

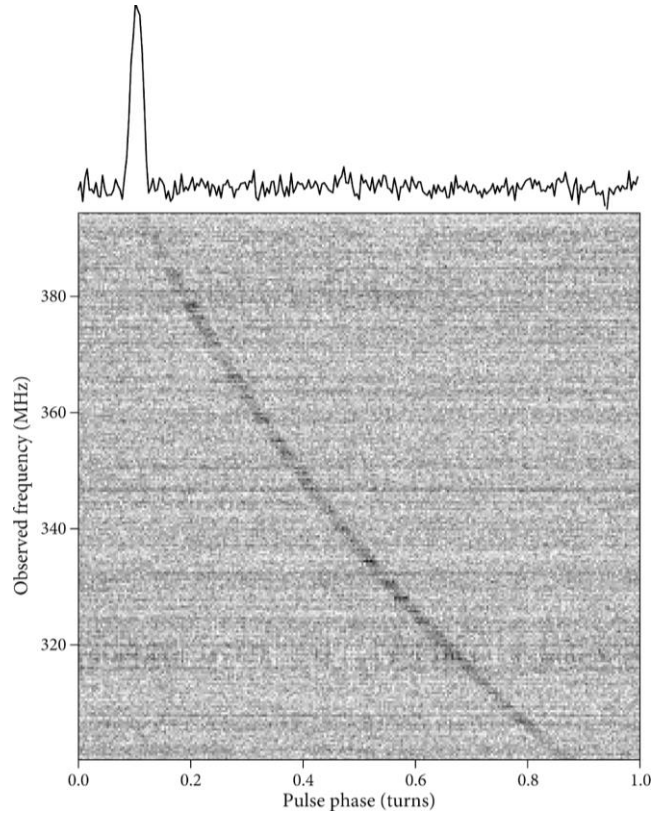
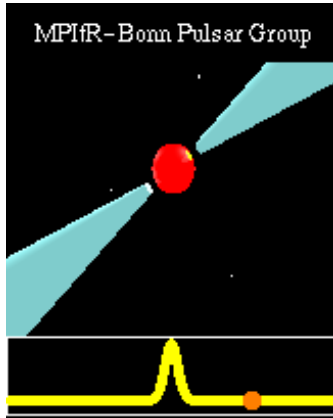


## Dynamic Life Cycle Model

The work packages of TA5 of the PUNCH4NFDI consortium

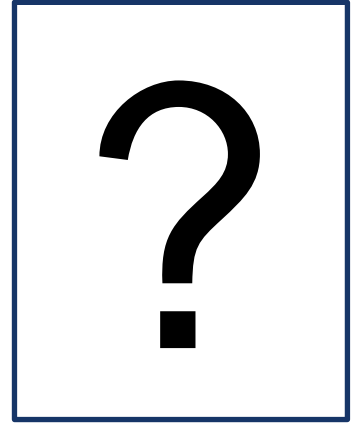
# Astronomical Radio Signals

## Pulsars

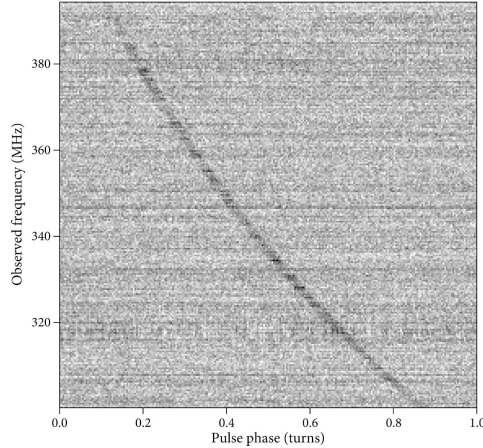


*Example of dispersion delay for an individual pulse from J1800+5034*

## Fast Radio Bursts

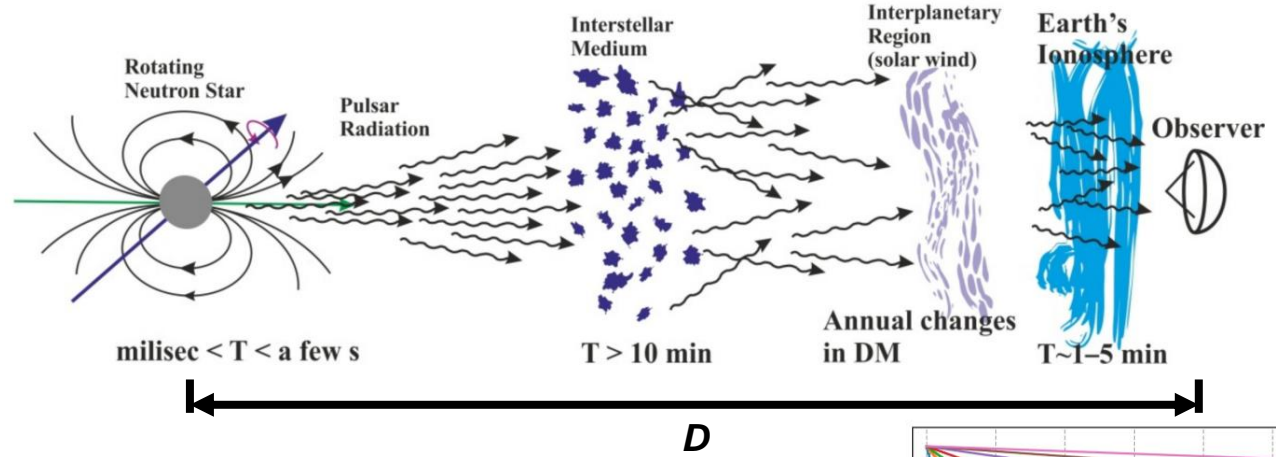


# Astronomical Radio Signals



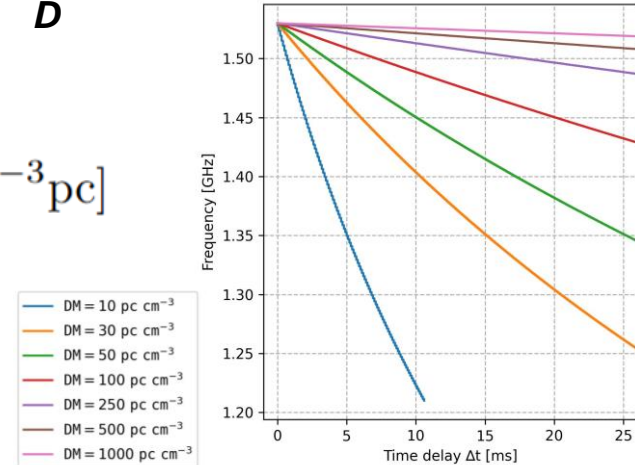
Example of dispersion delay for an individual pulse from J1800+5034

## The Flux of Pulsars' Signal Moderation Areas © Leszek Błasziewicz



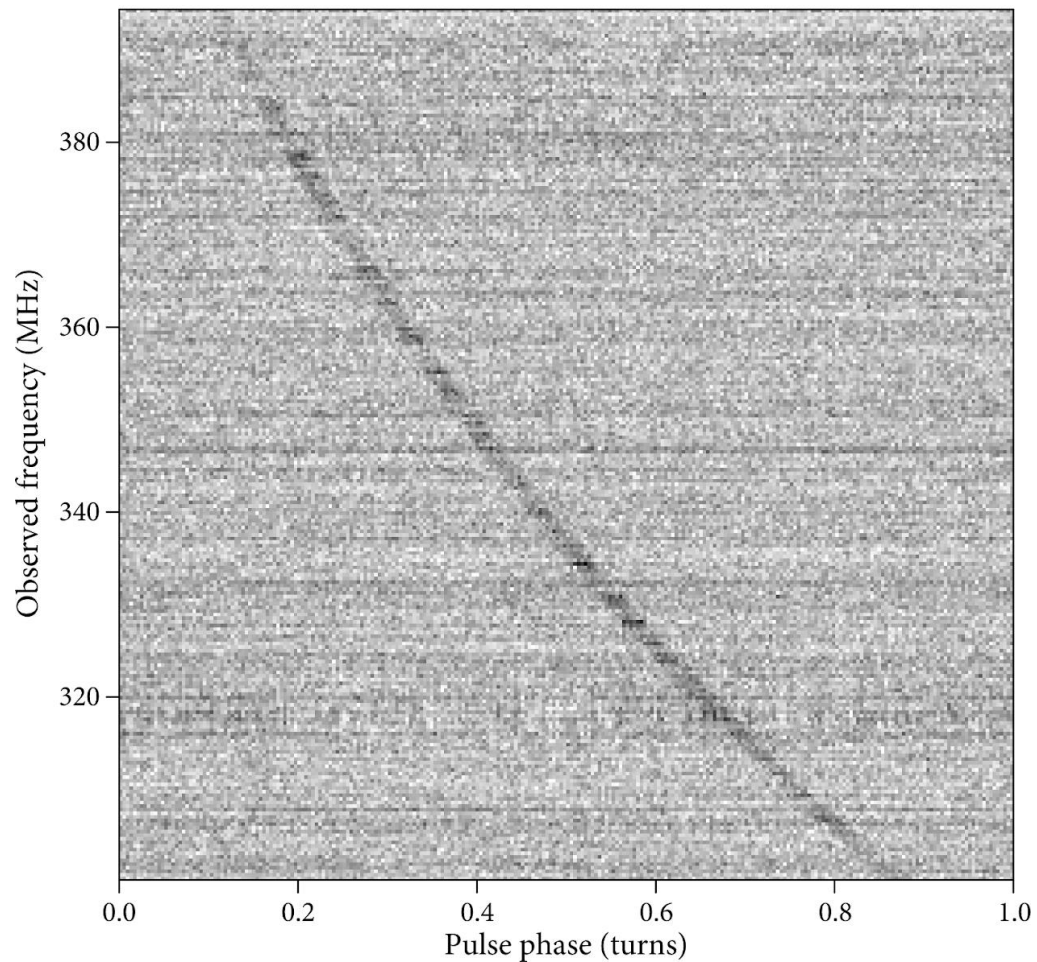
$$\Delta t = 4.12 \text{ ms} \left( (f_{\text{LO}}[\text{GHz}])^{-2} - (f_{\text{HI}}[\text{GHz}])^{-2} \right) \text{DM} [\text{cm}^{-3} \text{pc}]$$

$$\text{DM} = n_e [\text{cm}^{-3}] D [\text{pc}]$$

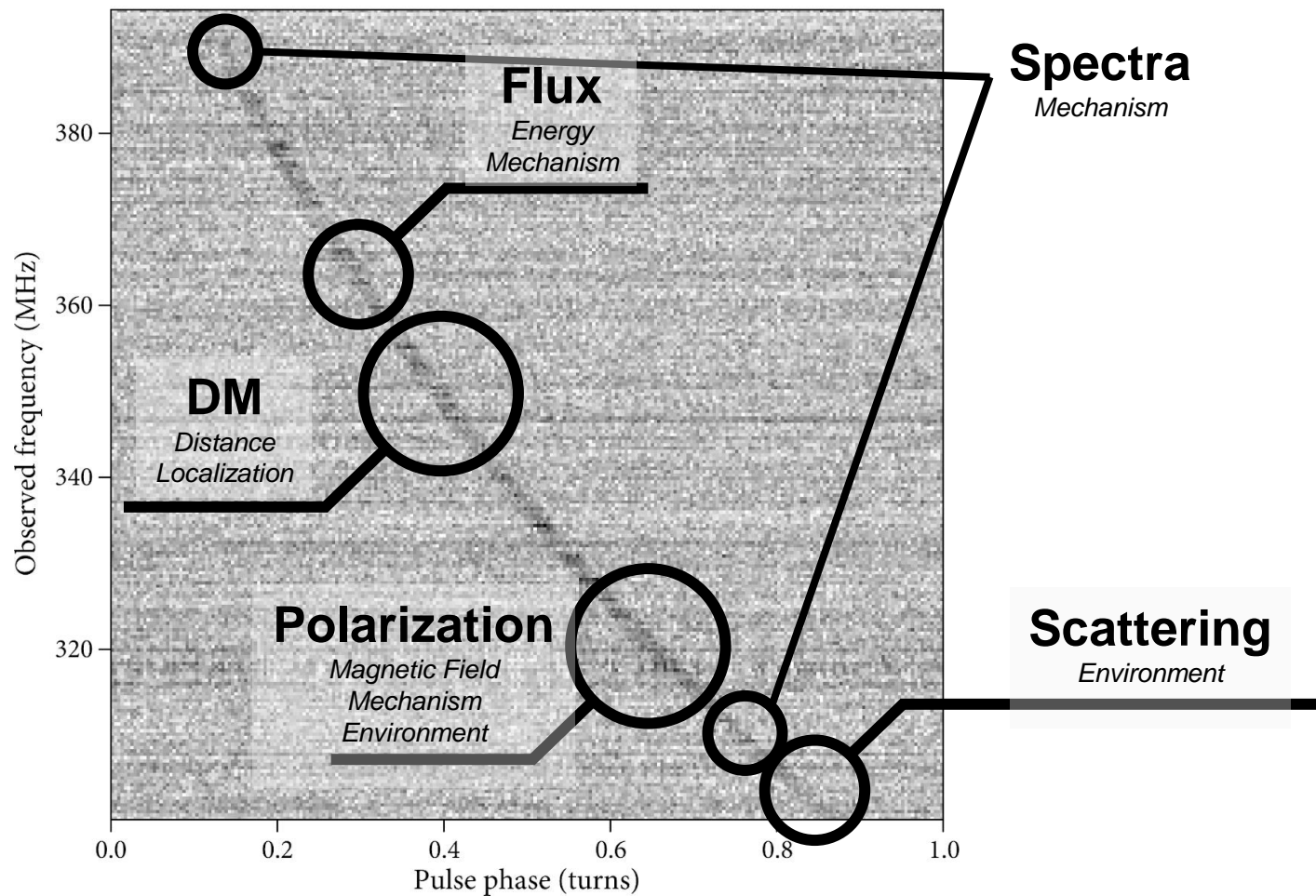


# Information from the Signal

6



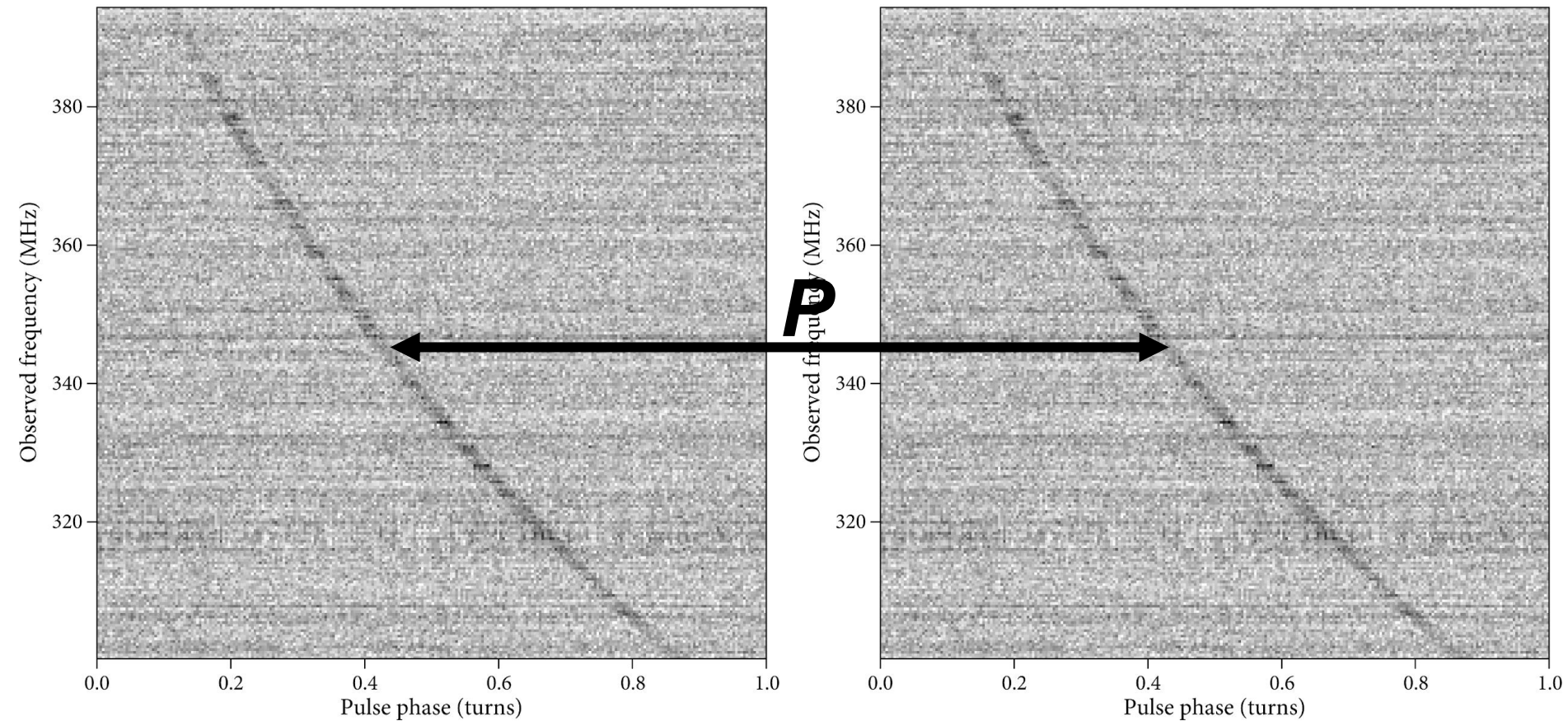
# Information from the Signal





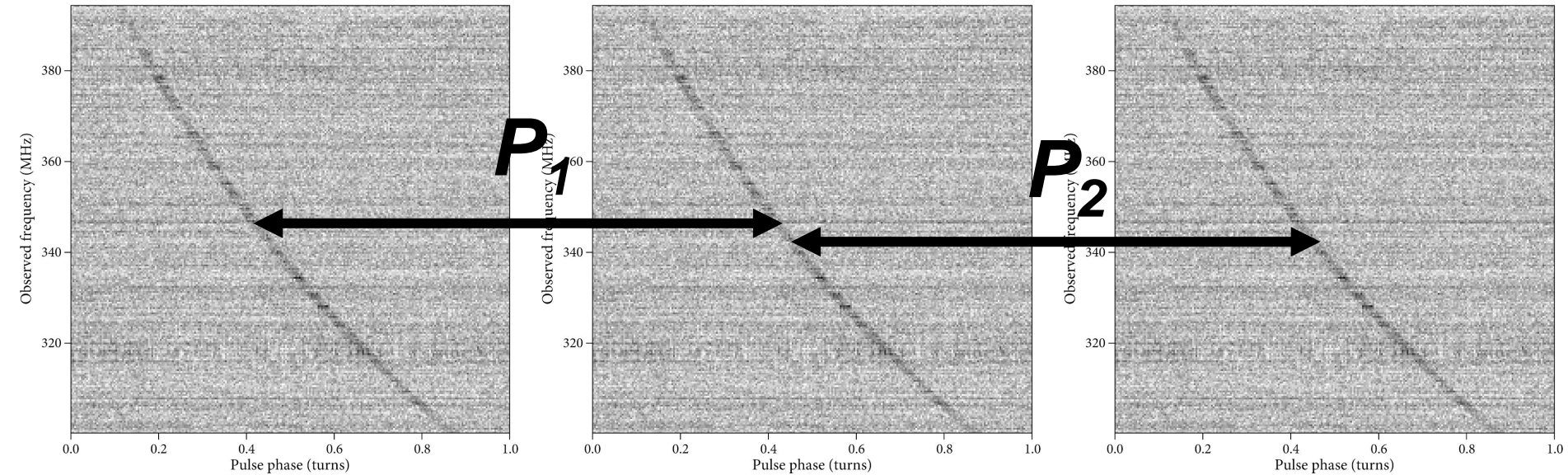
# Information from the Signal

8



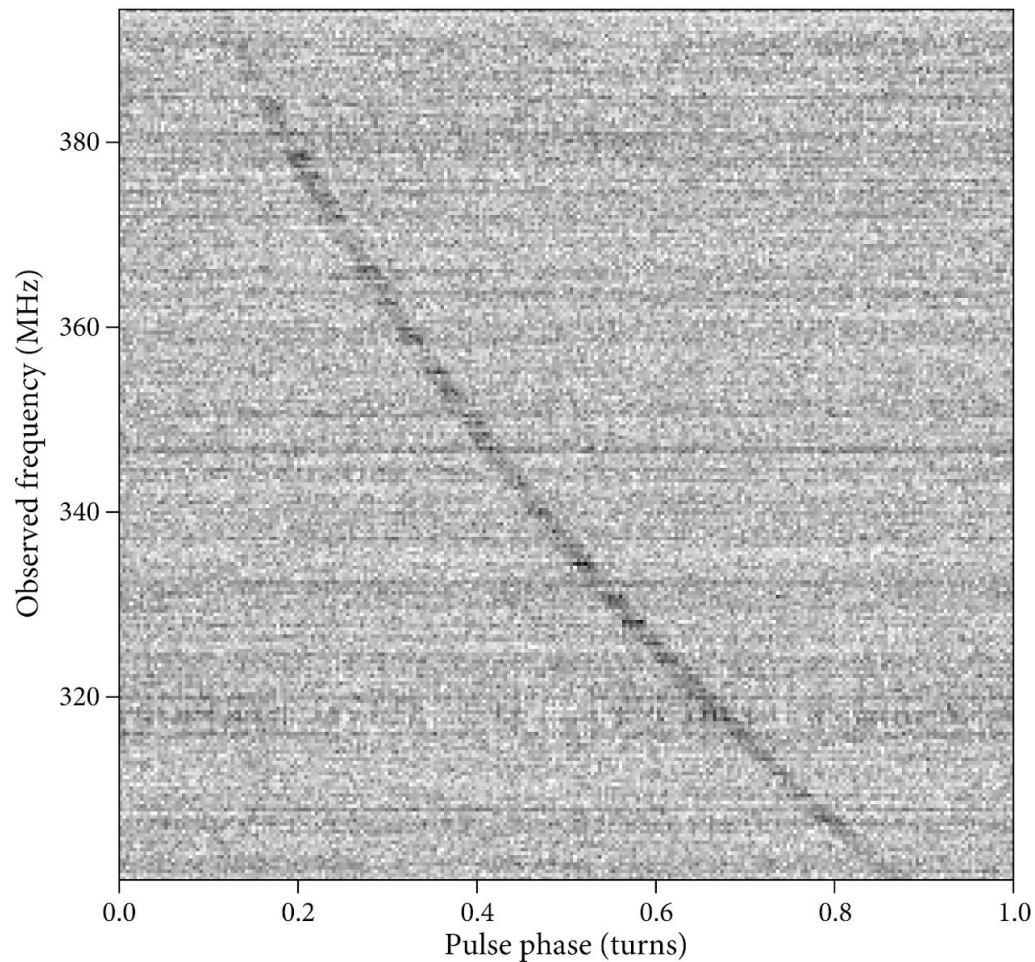


# Information from the Signal



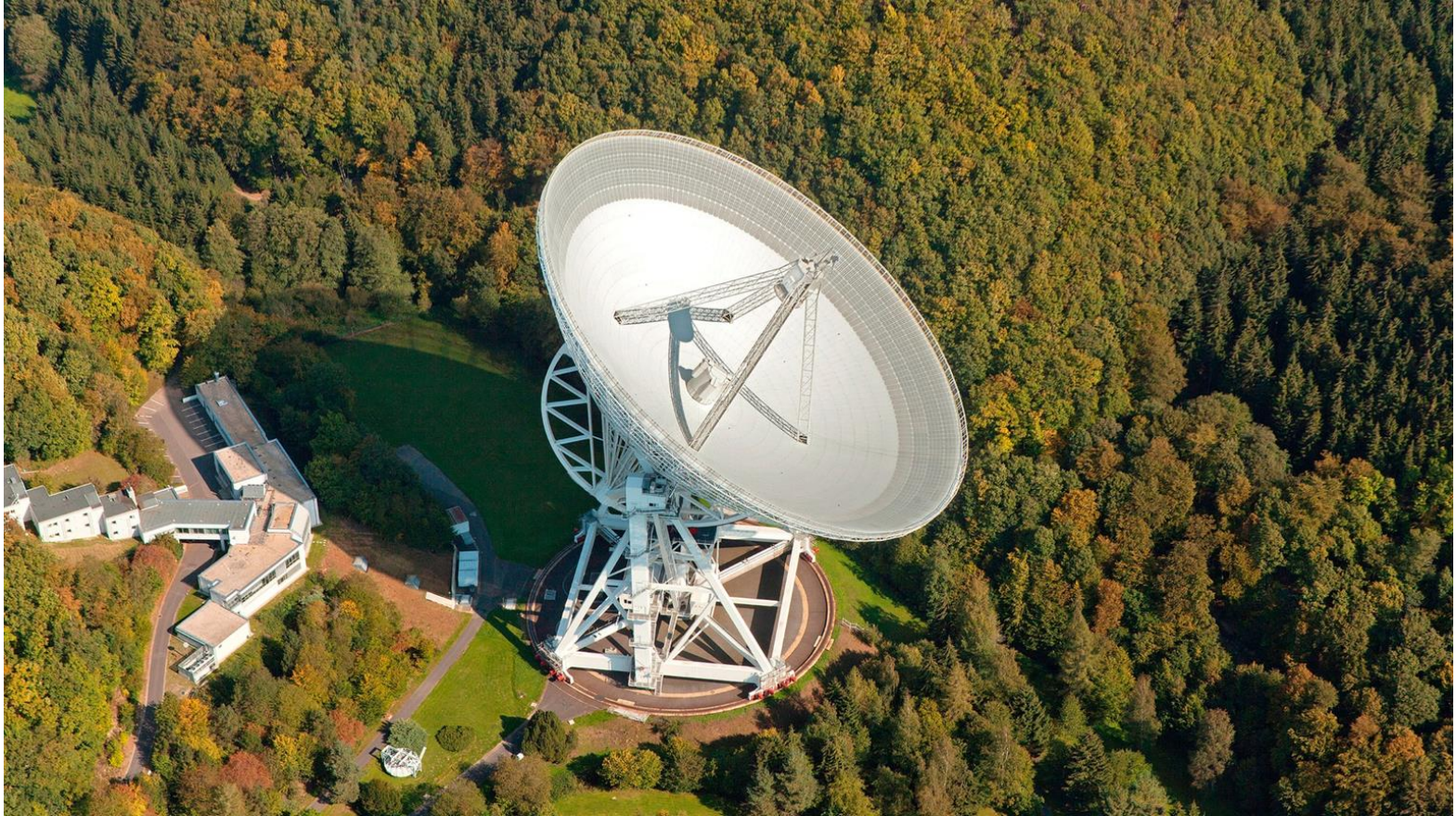
# Roseta Stone for Pulsar Astrophysics

10



# Detection of the Astronomical Radio Signals

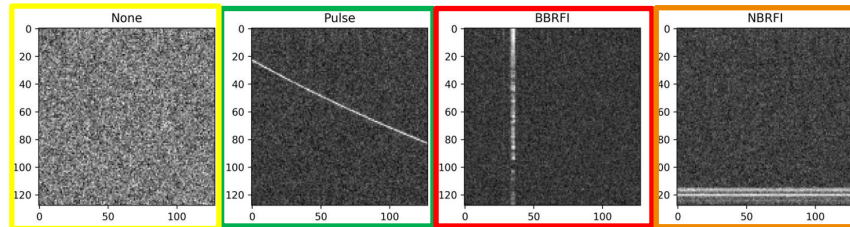
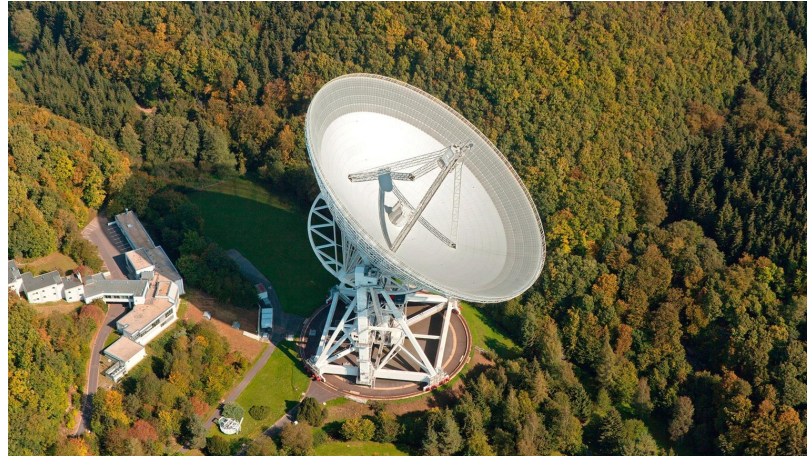
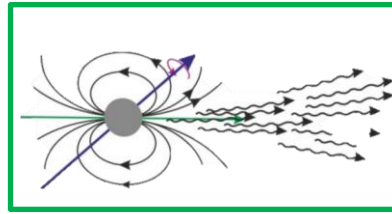
11



*Radio Telescope Effelsberg, Bad Münstereifel, Germany © Hans Blossey*



# Astronomical and non- Astronomical Radio Signals

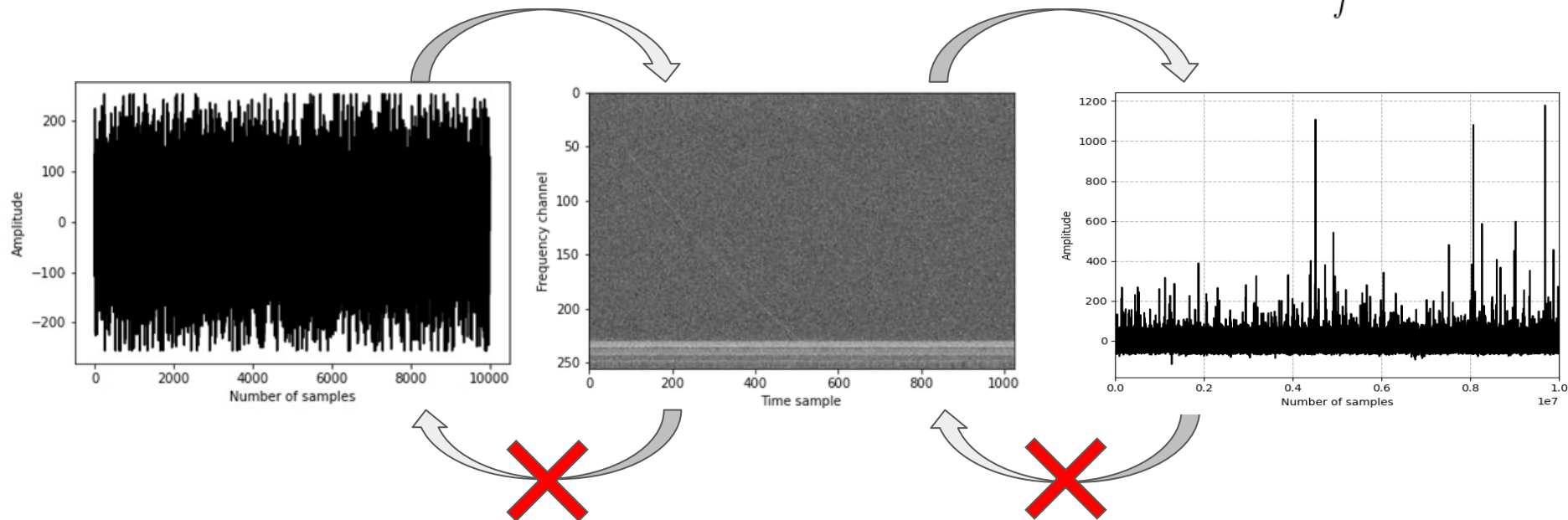


# Data Formats in Radio Astronomy

$$X(t)$$

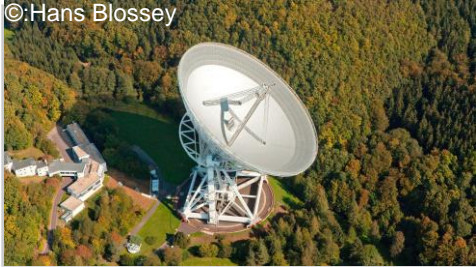


$$S(f, t) = \left| \int_{-\infty}^{\infty} X(t) e^{-2\pi i f t} dt \right|$$

$$T(t) = \sum_f S(f, t)$$



Format	Baseband data (dada files)	Spectrograms (filterbank file)	Time series
21 minutes	1.6 TB	3 GB	49 MB

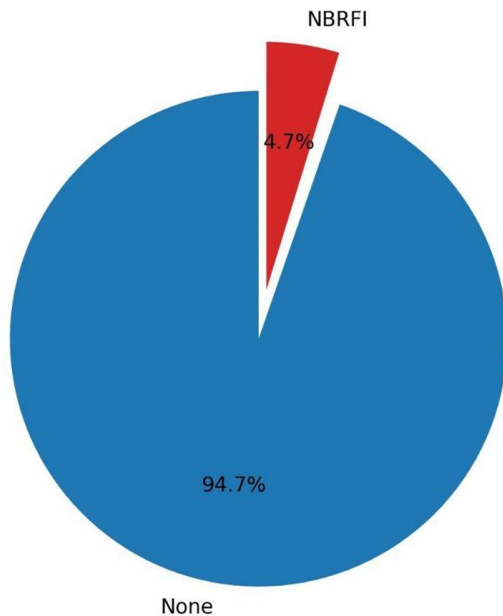
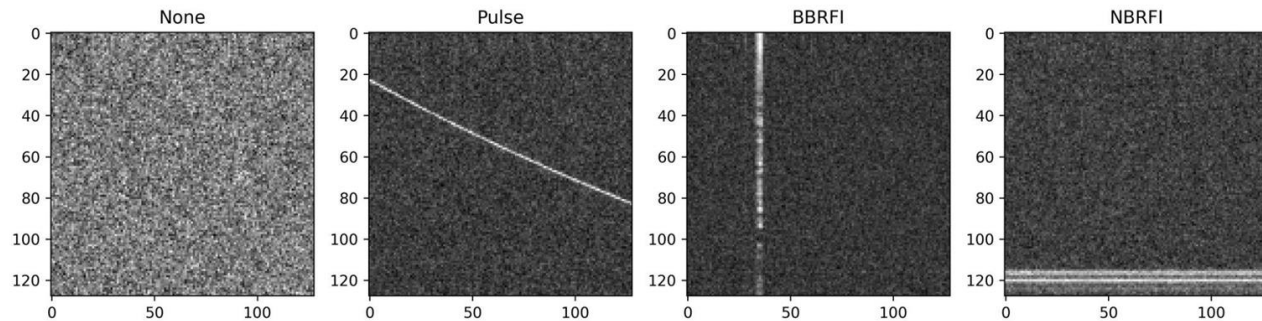
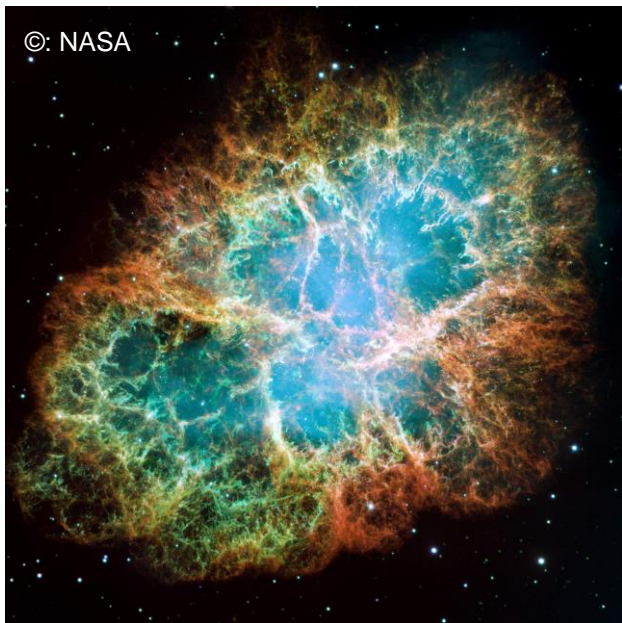
# Data Rate in Radio Astronomy

Radio telescope name	Radio telescope exterior	Bitrate per beam	Total bitrate
Effelsberg	©:Hans Blossey 	P210-7: 11.04 Gb / s	77 Gb / s (7)
		UWB: 290 Gb / s	290 Gb / s (1)
MeerKAT	©:picture alliance / Xinhua / CSIRO 	107 Mb / s	1.7 Tb / s (~1024 beams)
Square Kilometer Array	©:SKAO 	~ 1 Gb / s	20 Tb / s (>2200 beams)



# Statistics for One Observation

©: NASA



Object: **Crab pulsar (B0531+21)**

Data: **2020-05-31**

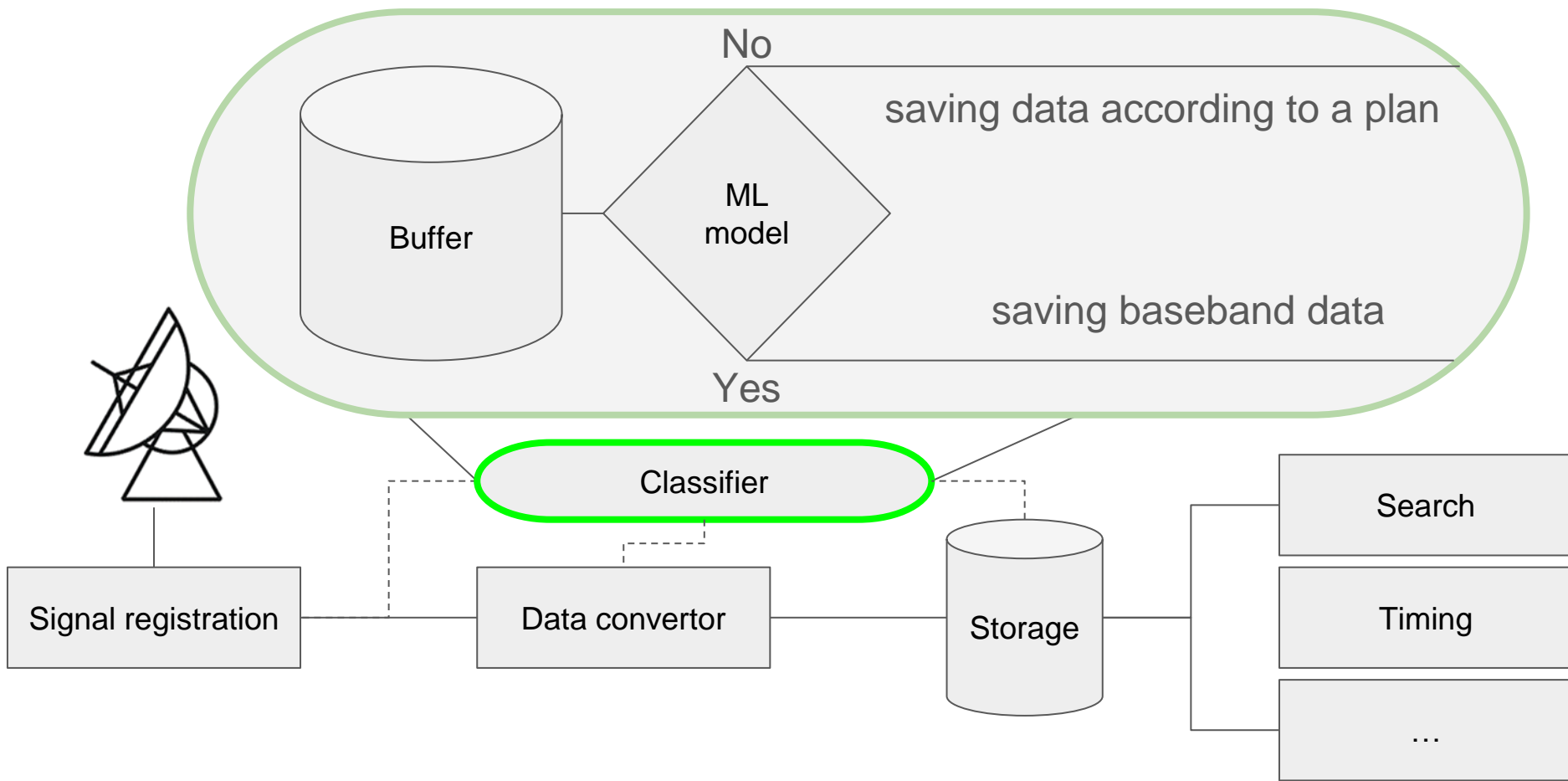
Time resolution: **102.4  $\mu$ s**

Telescope: **Effelsberg**

Bandwidth: **1210 - 1530 MHz**

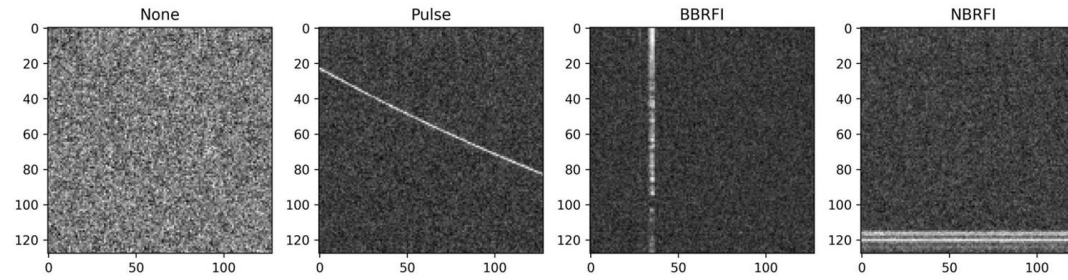
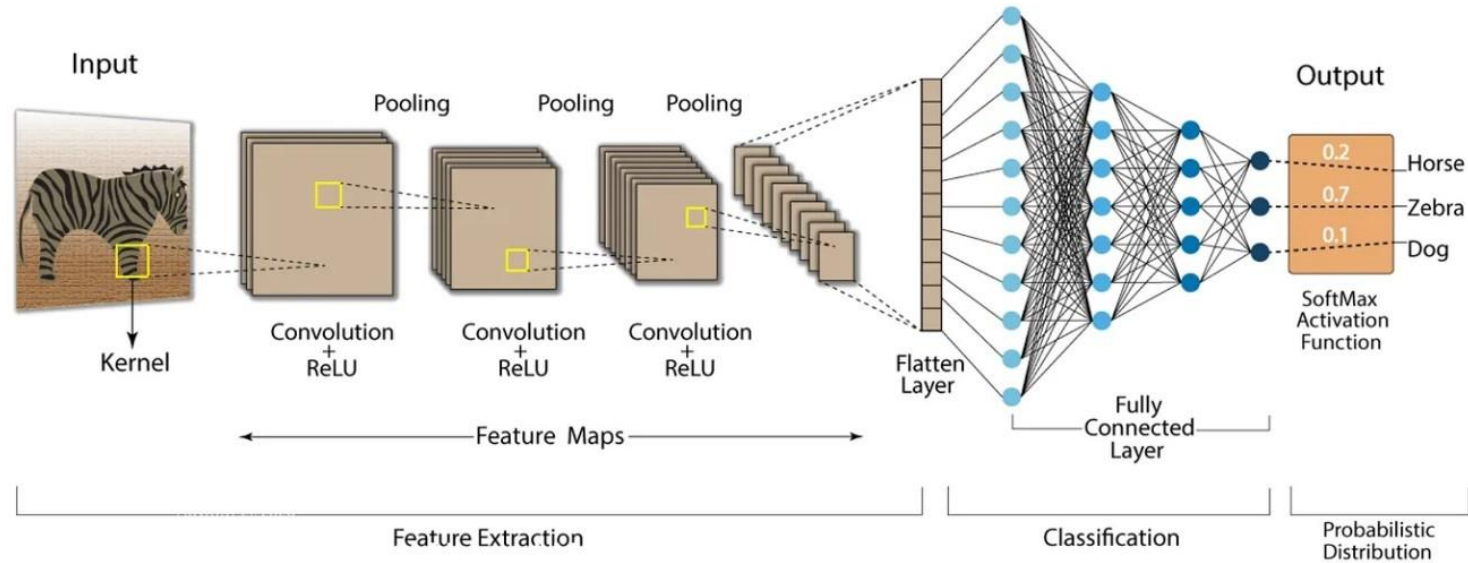
Duration: **21 minutes**

# Conception of the Task

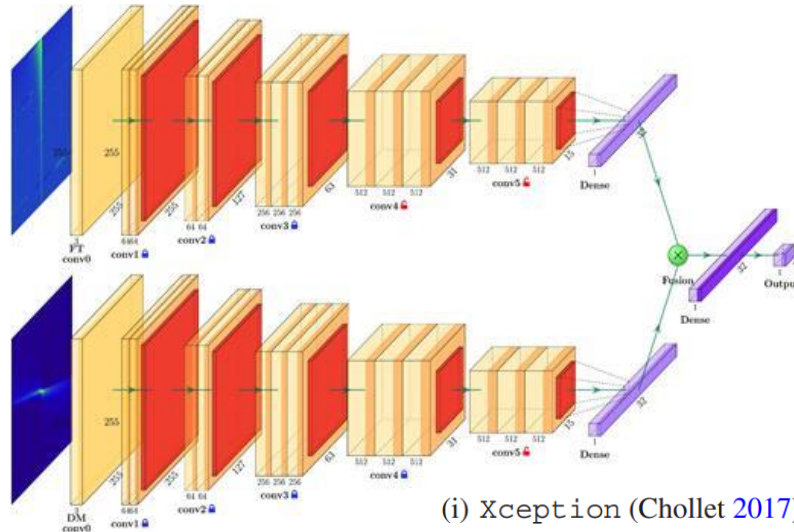
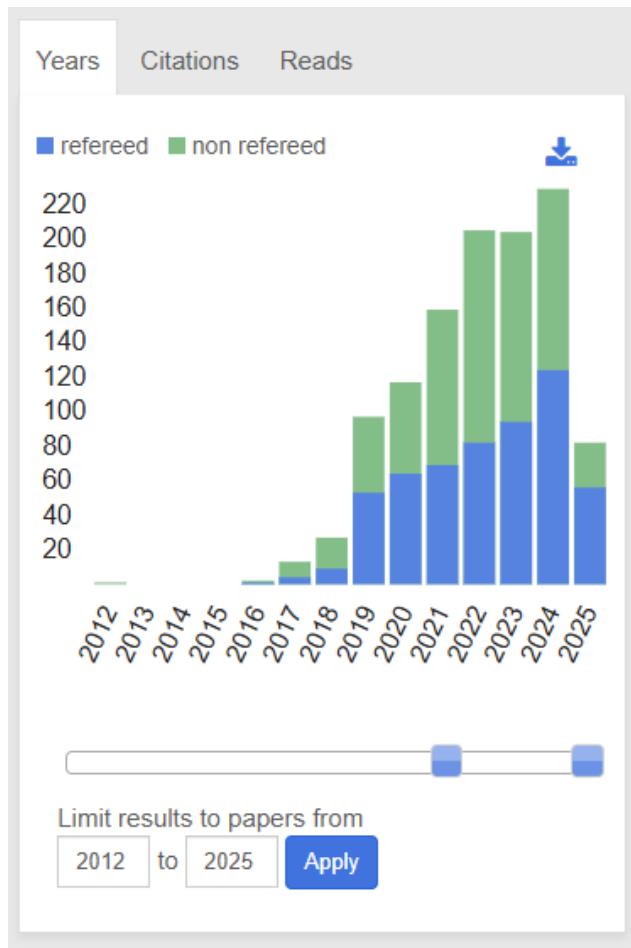


# CNN Usage for Pattern Recognition

**Convolution Neural Network (CNN)** © Kh. Nafizul Haque



# CNN Usage in Astronomy



- (i) Xception (Chollet [2017](#))
- (ii) VGG16, VGG19 (Simonyan & Zisserman [2014](#))
- (iii) ResNet50 (He et al. [2016](#))
- (iv) DenseNet121, DenseNet169, DenseNet201 (Huang et al. [2017](#))
- (v) InceptionV3 (Szegedy et al. [2016](#))
- (vi) InceptionResNetV2 (Szegedy et al. [2017](#))
- (vii) MobileNet (Howard et al. [2017](#))
- (viii) MobileNetV2 (Sandler et al. [2018](#))

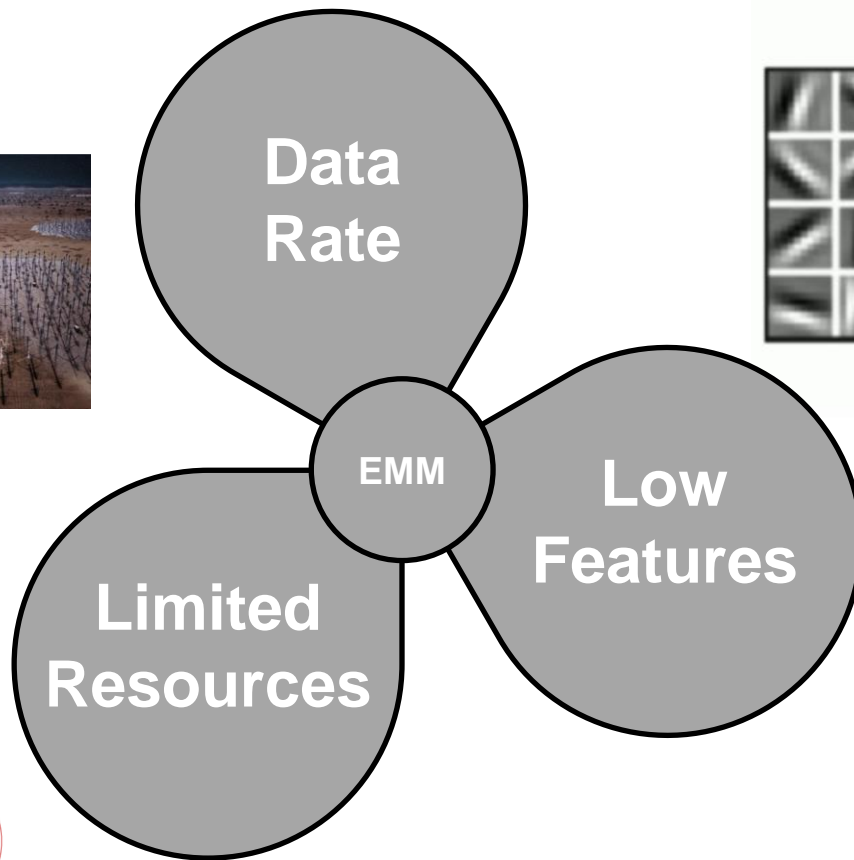
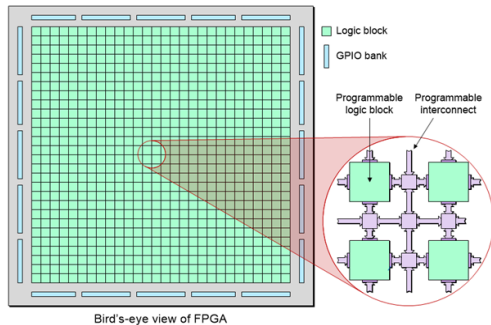
Devansh Agarwal, et al., MNRAS, 2020,  
<https://doi.org/10.1093/mnras/staa1856>

# Usage of Extremely Minimalistic Model

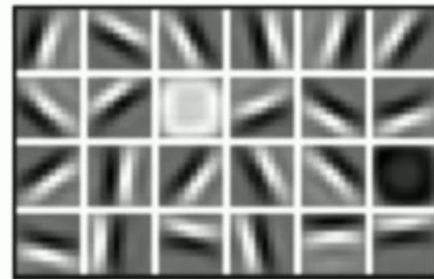
©: SKA



©: www.latticesemi.com



Low Level Features



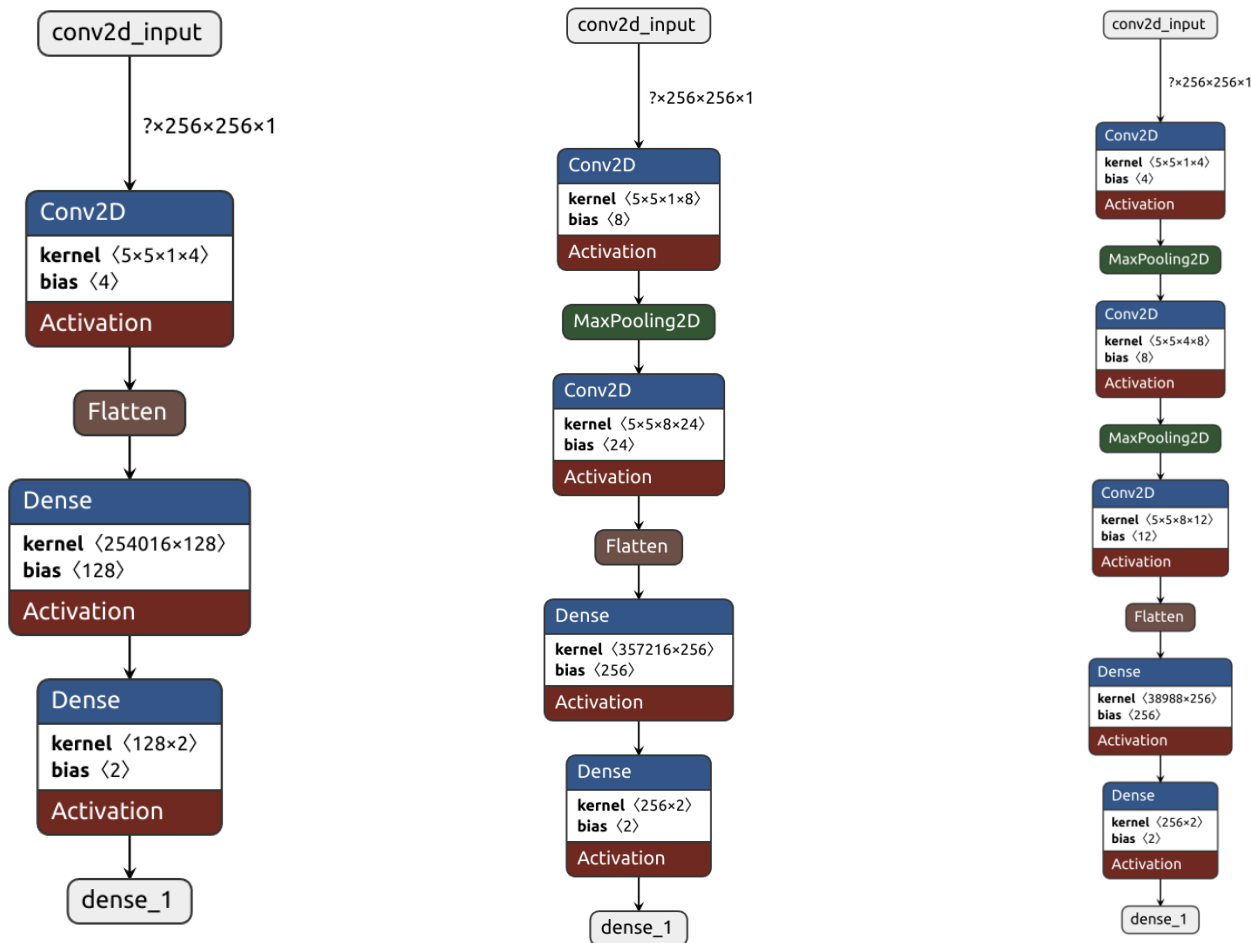
Lines & Edges

High Level Features



Facial Structure

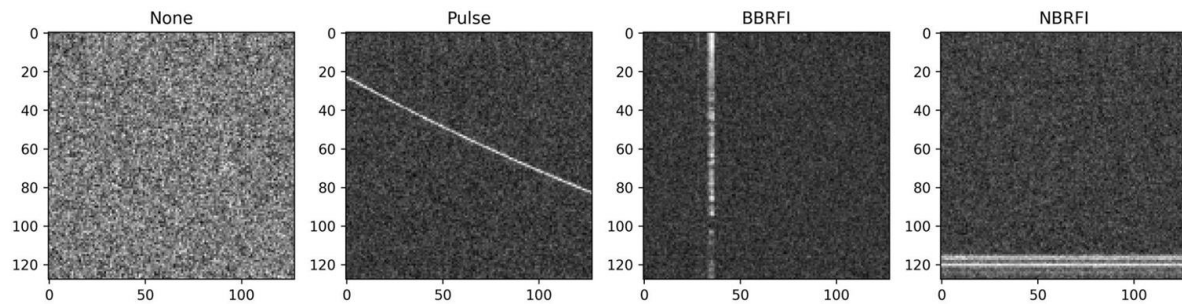
# Extremely Minimalistic Classifier Architecture



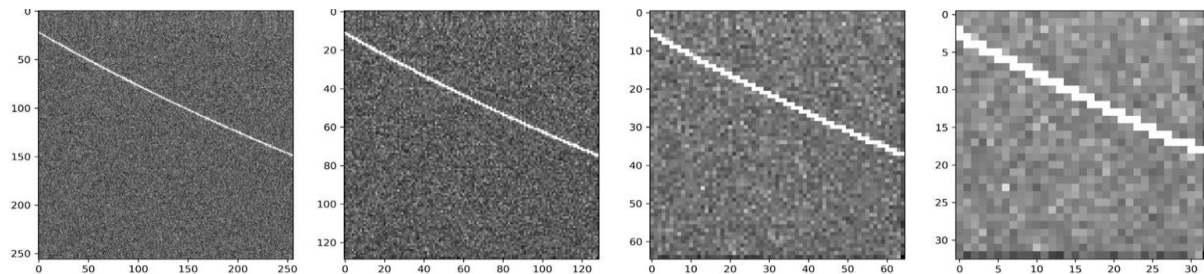


# Spectrogram Classifier

©: NASA



**Number of classes: 8**



**Test with four image resolutions**

Data Preparation

Model Selection

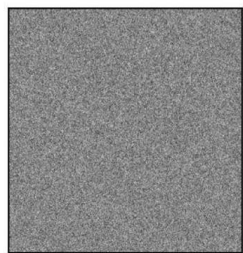
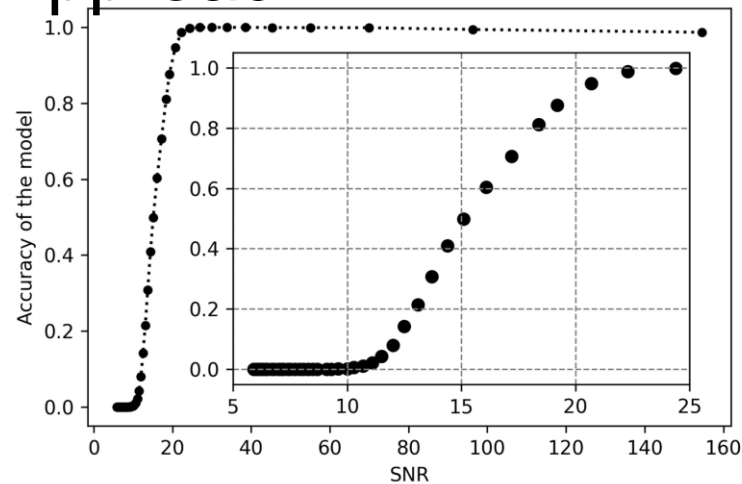
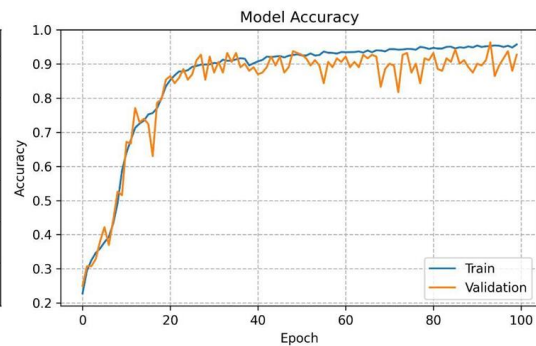
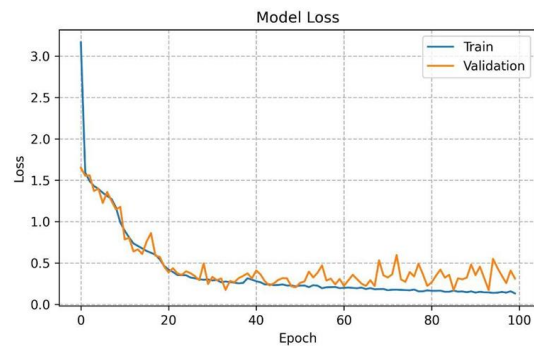
HP Tuning

Training

Validation

Validation with  
additional dataset

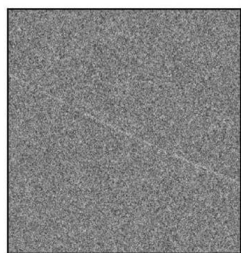
# Limitations of the “Spectrogram” Approach



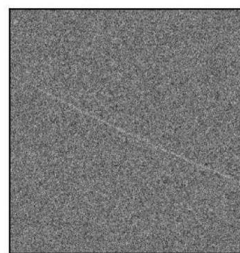
5.0



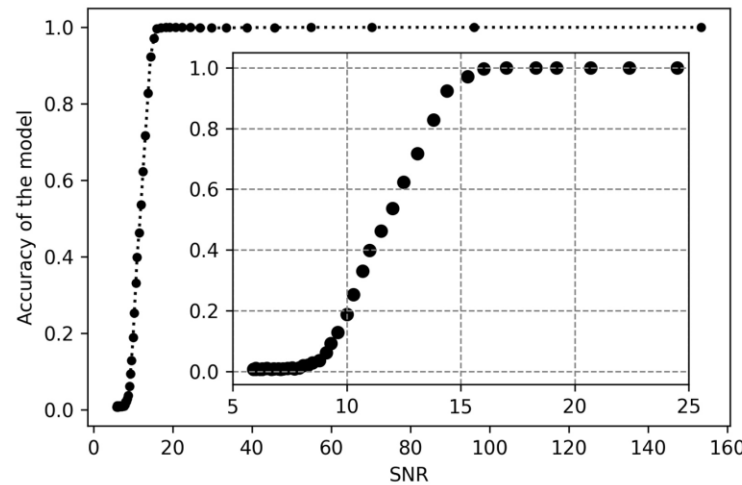
8.6



11.4



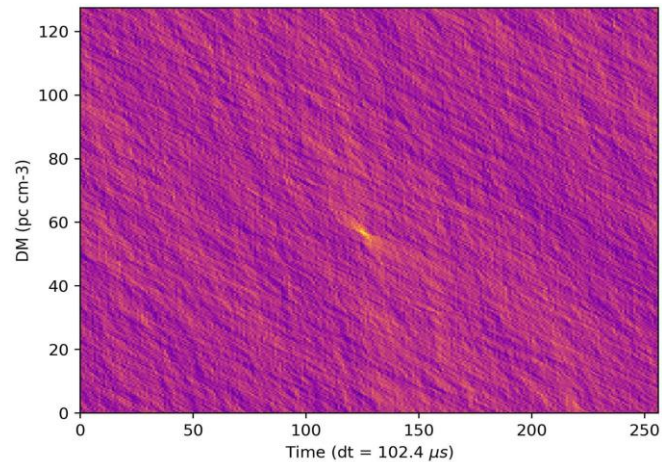
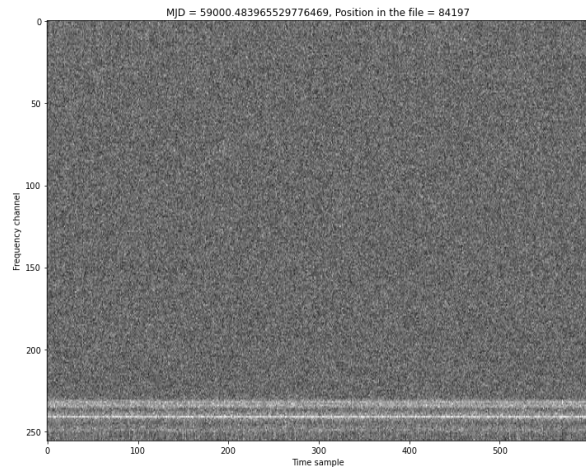
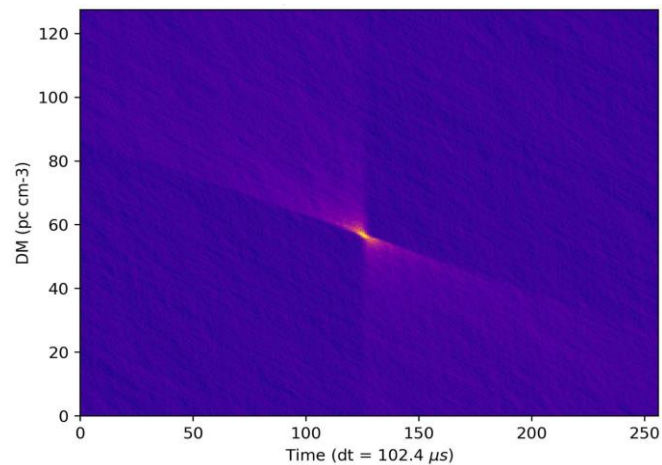
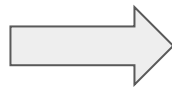
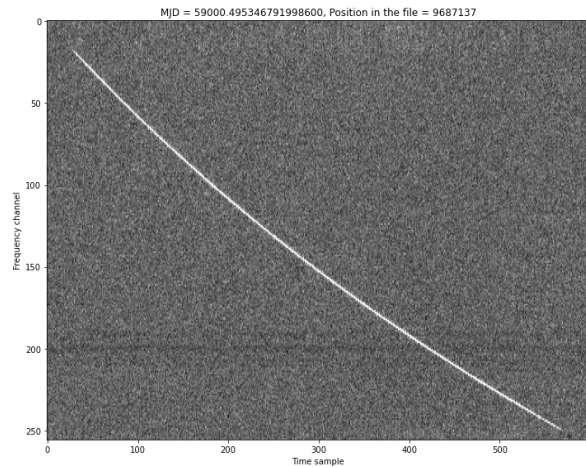
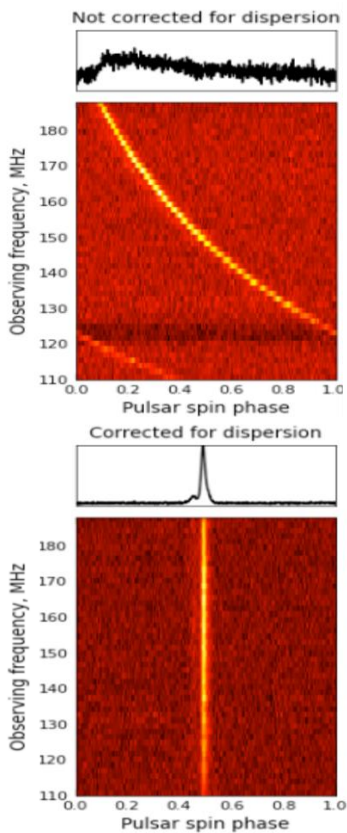
15.6





# Spectrograms vs. DM-Images

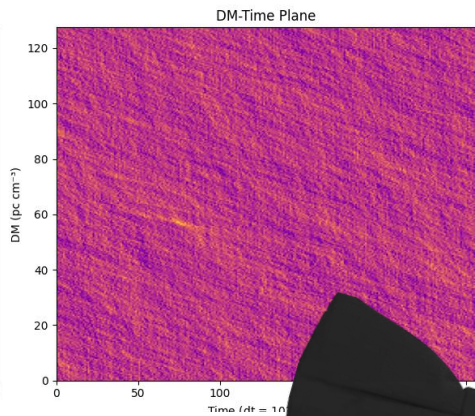
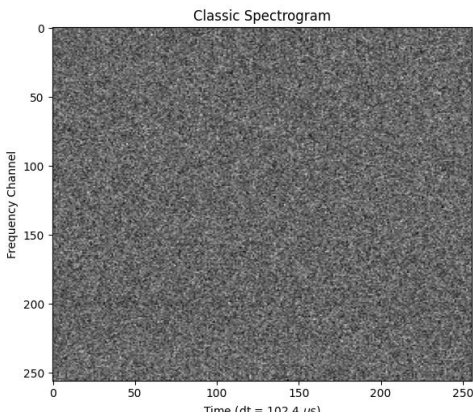
©: Anya Bilous



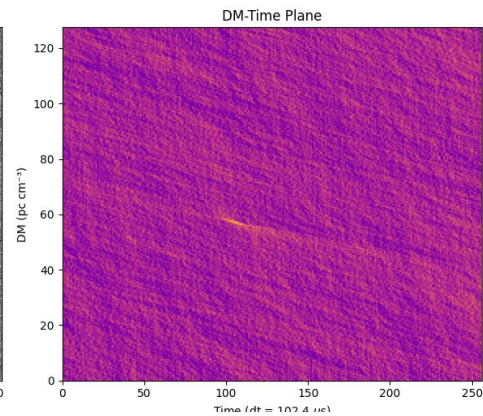
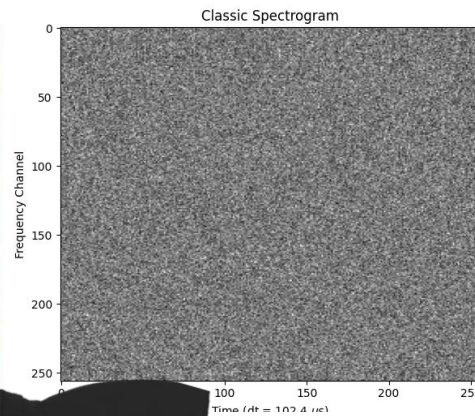


# “Bow-tie” with Different SNR

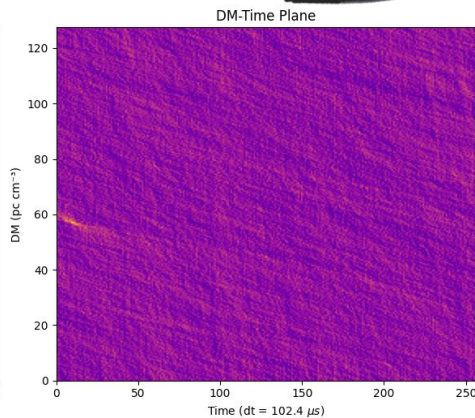
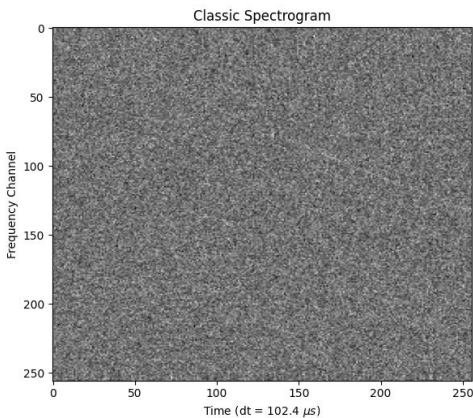
Pulse Example with 7.5 SNR



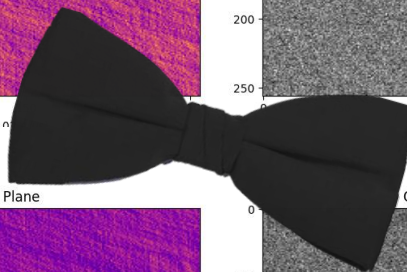
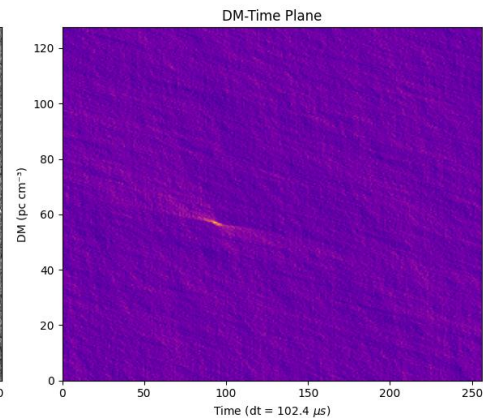
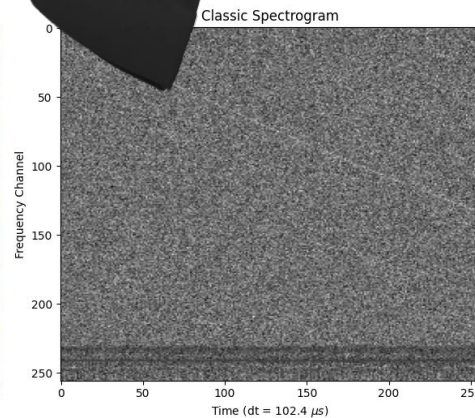
Pulse Example with 10.5 SNR



Pulse Example with 12.5 SNR

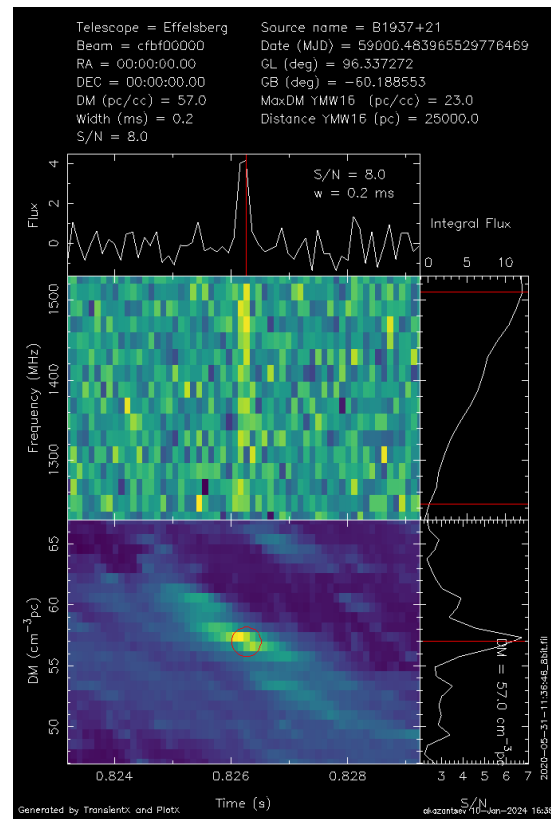
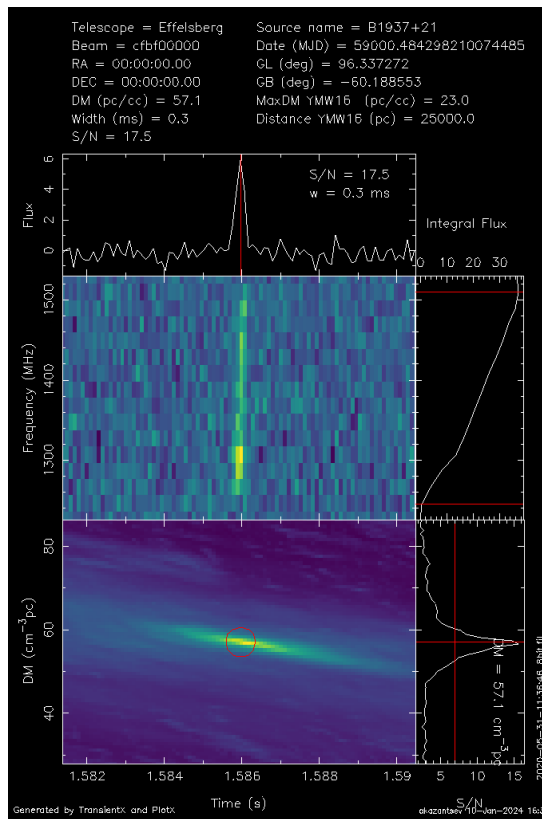


Pulse Example with 21.3 SNR



# TransientX as a Fast Solution for Dedispersion

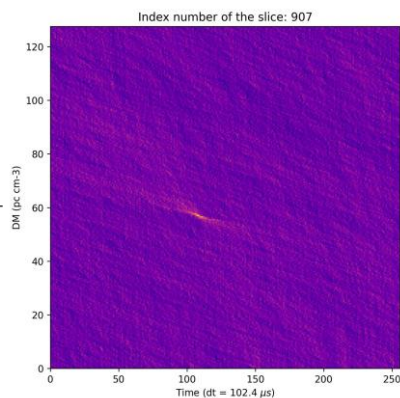
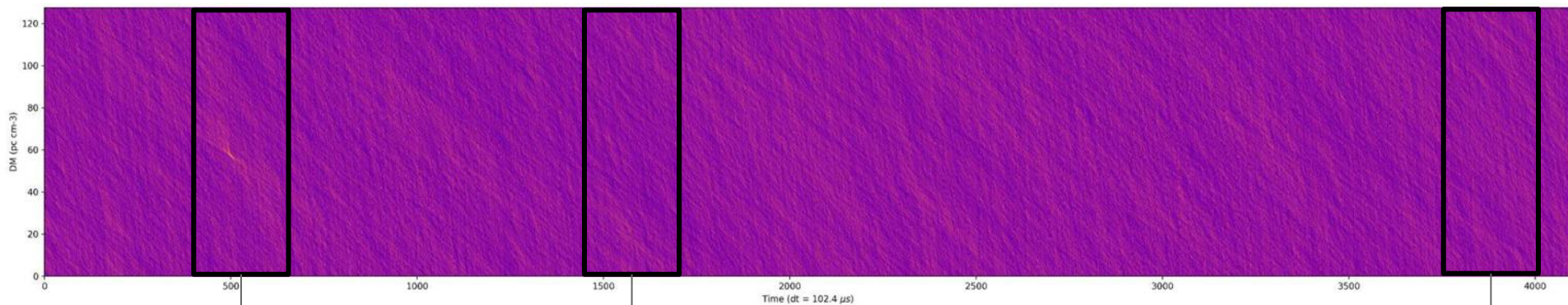
a one command line high performance transient search software.



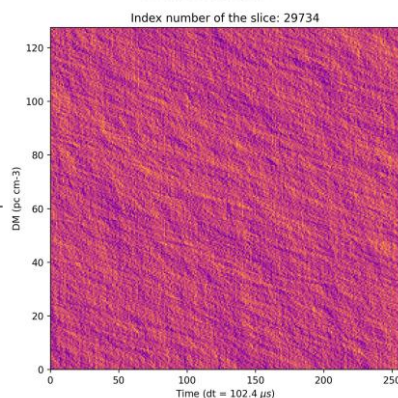
<https://github.com/ypmen/TransientX>



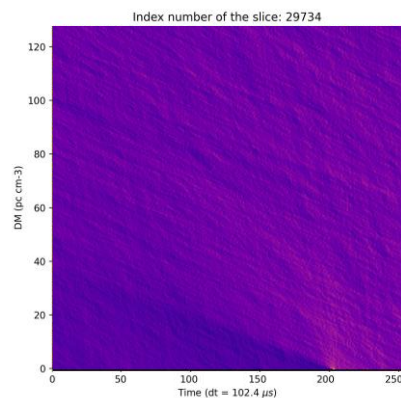
# DM-time Data from TransientX



Pulse



Noise

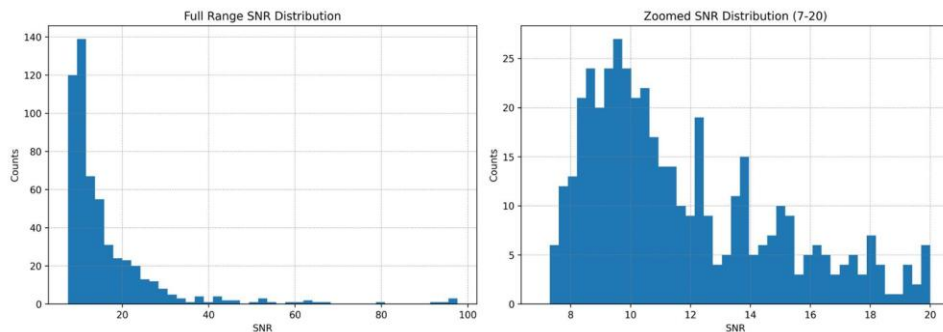


BBRFI



# DM-time dataset

**297 472** DM-time image in total. A big portion of this data (**80%, 237 977**) has been used for training sessions, The rest of the data (**20%, 59 495**) has been saved as an unseen dataset to calculate key metrics after training.

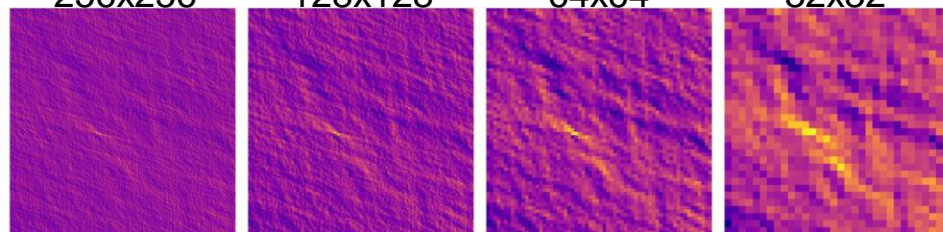


256x256

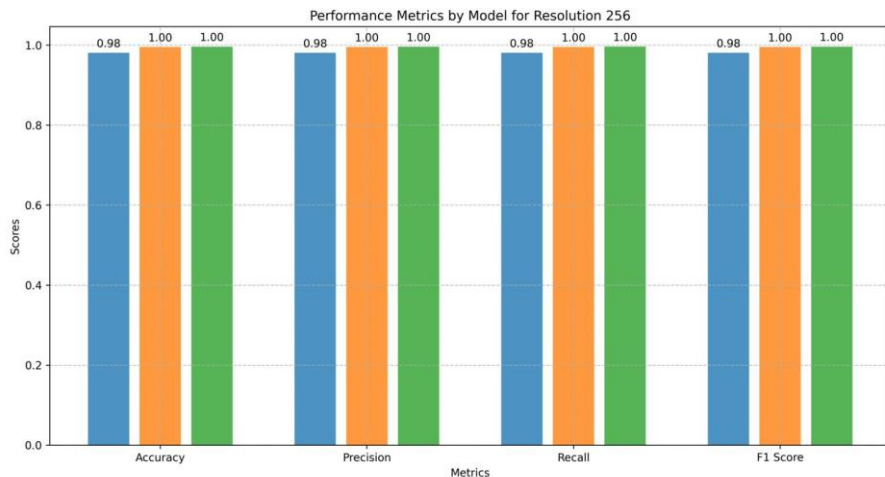
128x128

64x64

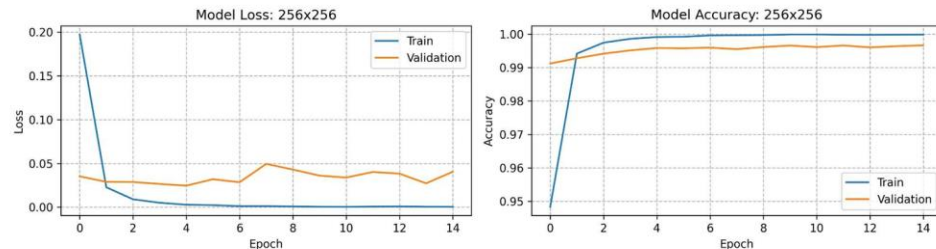
32x32



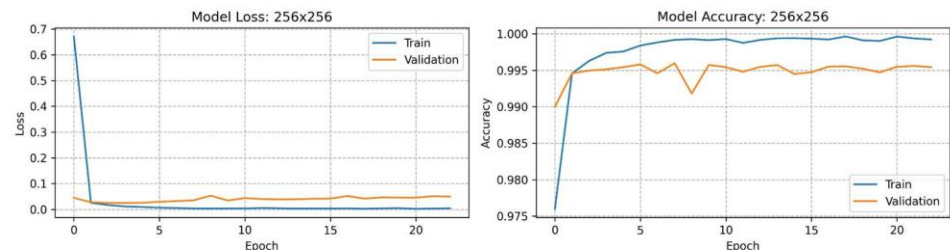
# Results of Training and Validation



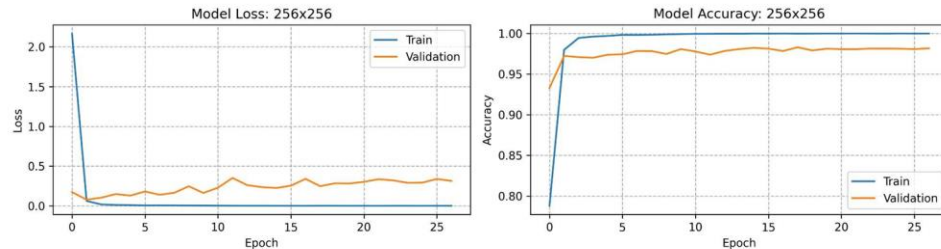
Model DM\_time\_binary\_classifier\_240315\_3 performance for 256x256 resolution



Model DM\_time\_binary\_classifier\_240315\_2 performance for 256x256 resolution



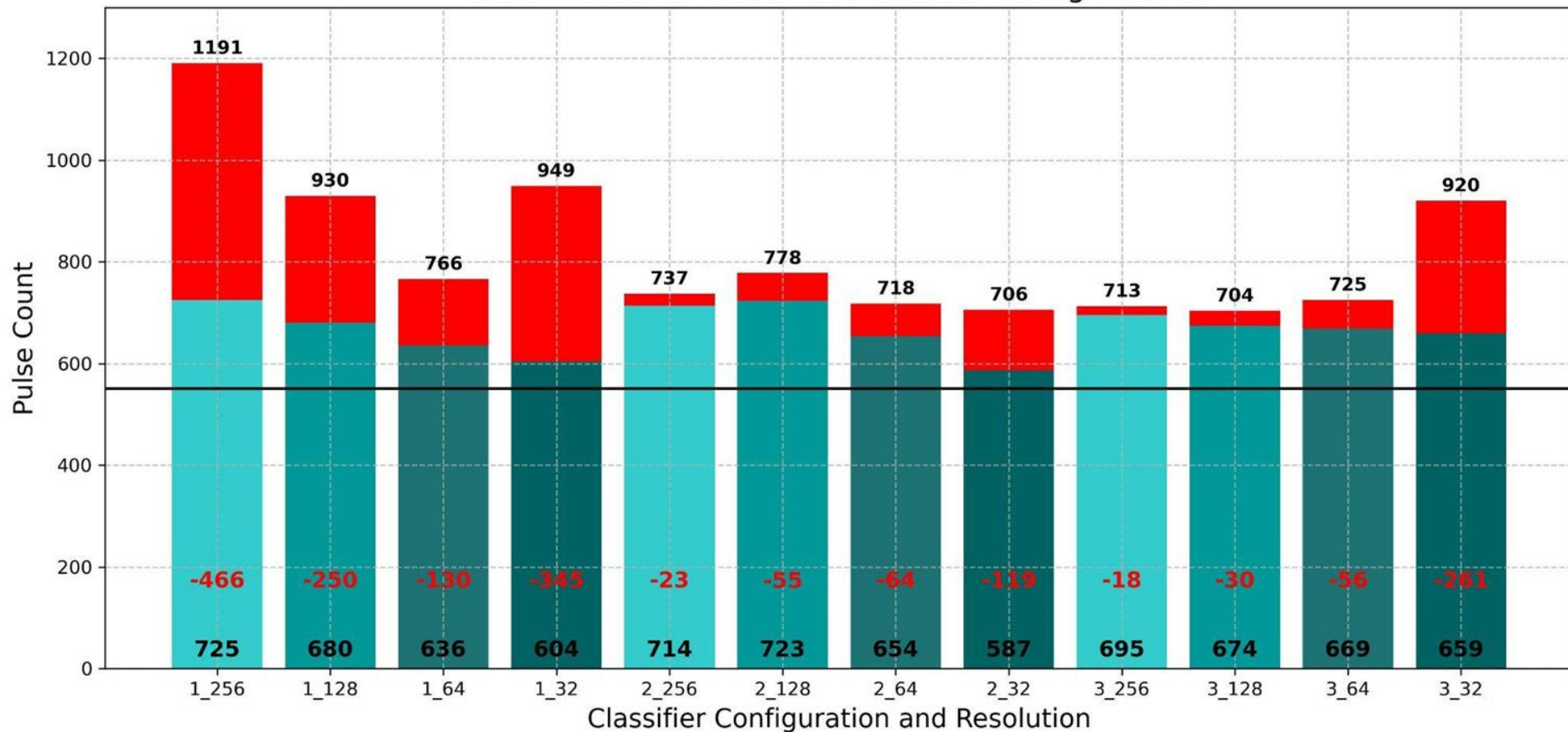
Model DM\_time\_binary\_classifier\_240315\_1 performance for 256x256 resolution



Legend: DM\_time\_binary\_classifier\_240315\_1 (Blue), DM\_time\_binary\_classifier\_240315\_2 (Orange), DM\_time\_binary\_classifier\_240315\_3 (Green)

# Comparison with TransientX Results

Pulse Count for Different Classifier Configurations



# Results of Tests with Real Data

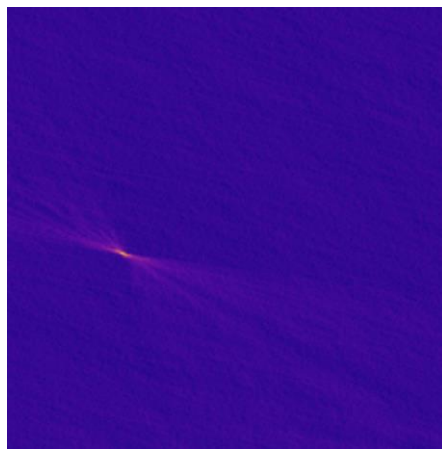
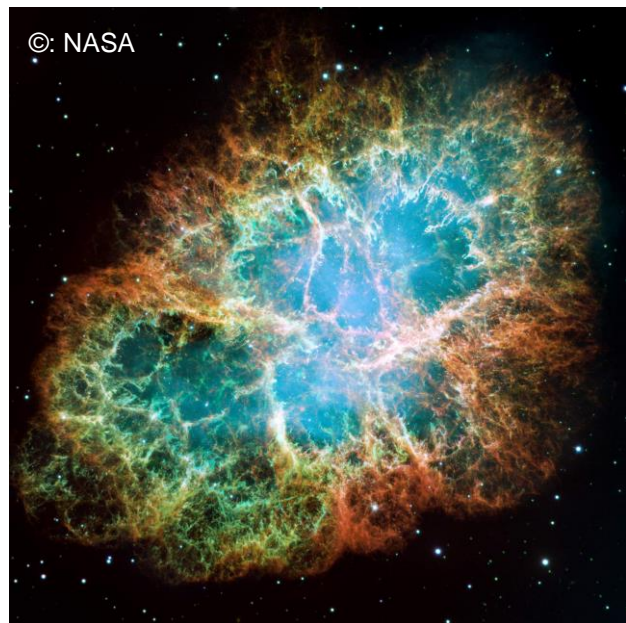
Model name	%	Model did not find	Model found additionally	Unique for the model
2_256	<b>93.5</b>	22	210 / <b>110</b>	25
2_128	<b>94.4</b>	19	212 / <b>90</b>	14
2_64	<b>94.4</b>	19	274 / <b>99</b>	11
2_32	<b>75.2</b>	87	164 / <b>33</b>	1
3_256	<b>94.6</b>	18	125 / <b>85</b>	11
3_128	<b>94.6</b>	18	97 / <b>66</b>	5
3_64	<b>94.4</b>	19	150 / <b>68</b>	4
3_32	<b>85.6</b>	50	477 / <b>58</b>	7



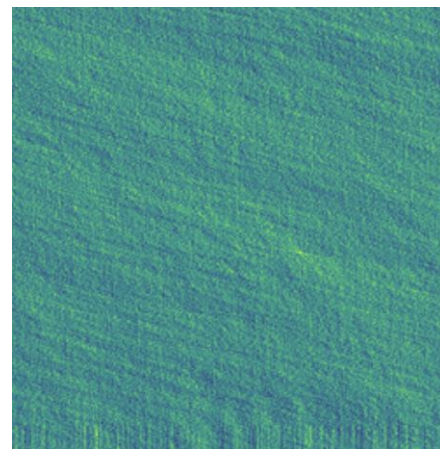


# Testing with New Data

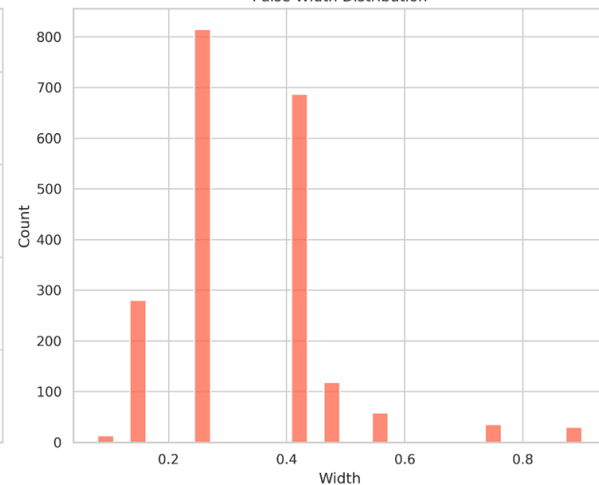
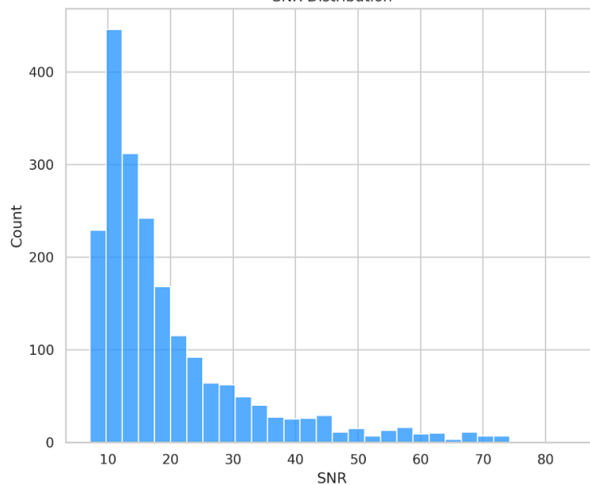
©: NASA



SNR Distribution



Pulse Width Distribution



Object: **Crab pulsar (B0531+21)**

Data: **2024-06-21**

Time resolution: **81.92  $\mu$ s**

Telescope: **Effelsberg**

Bandwidth: **1281 - 1600 MHz**

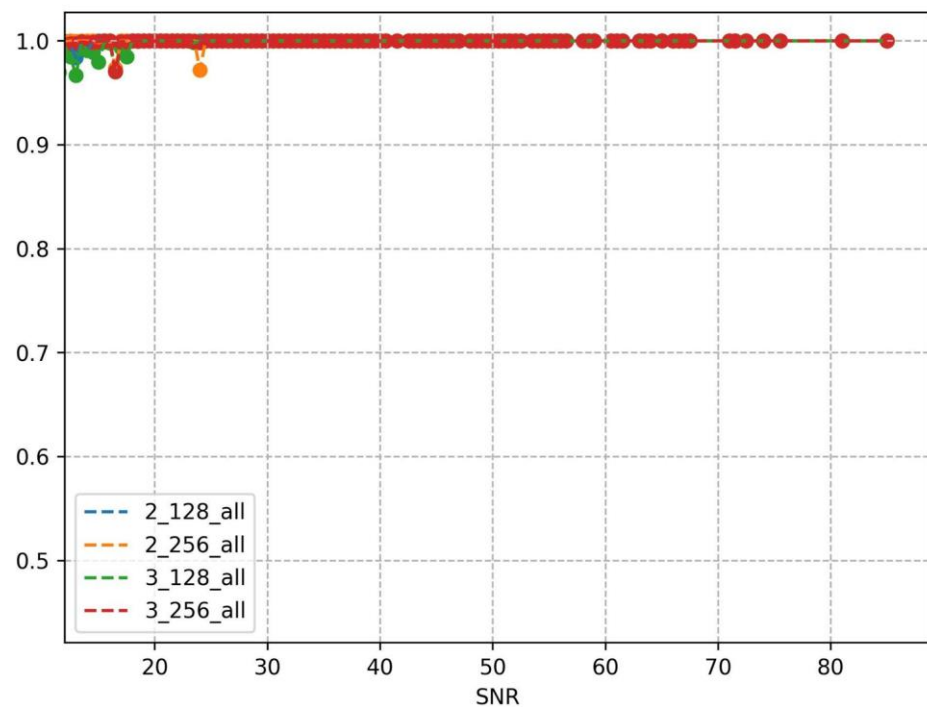
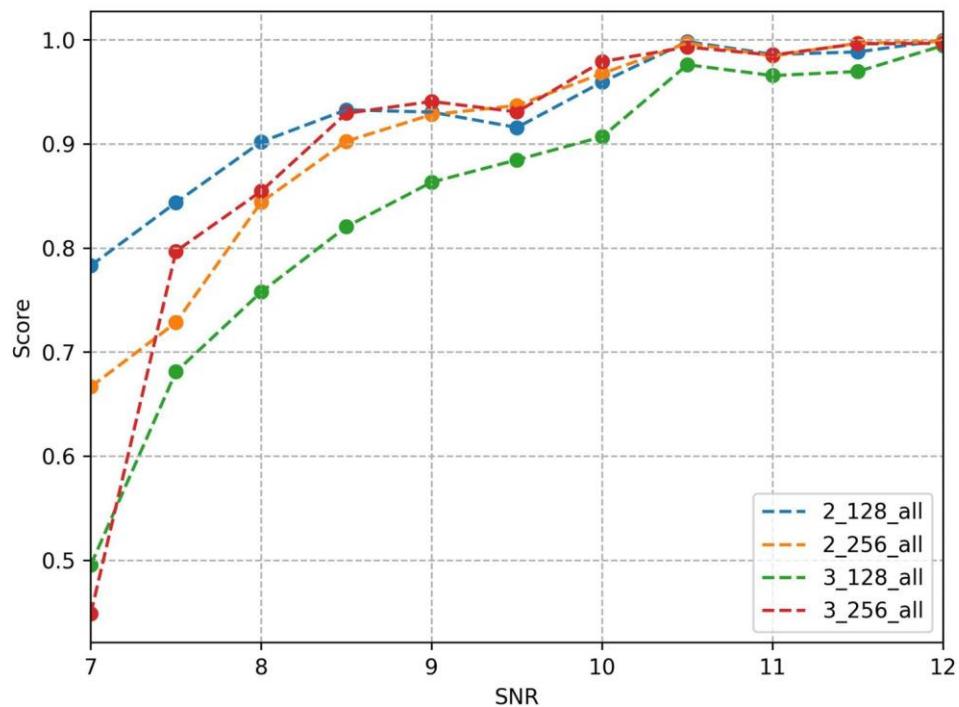
Duration: **30 minutes**



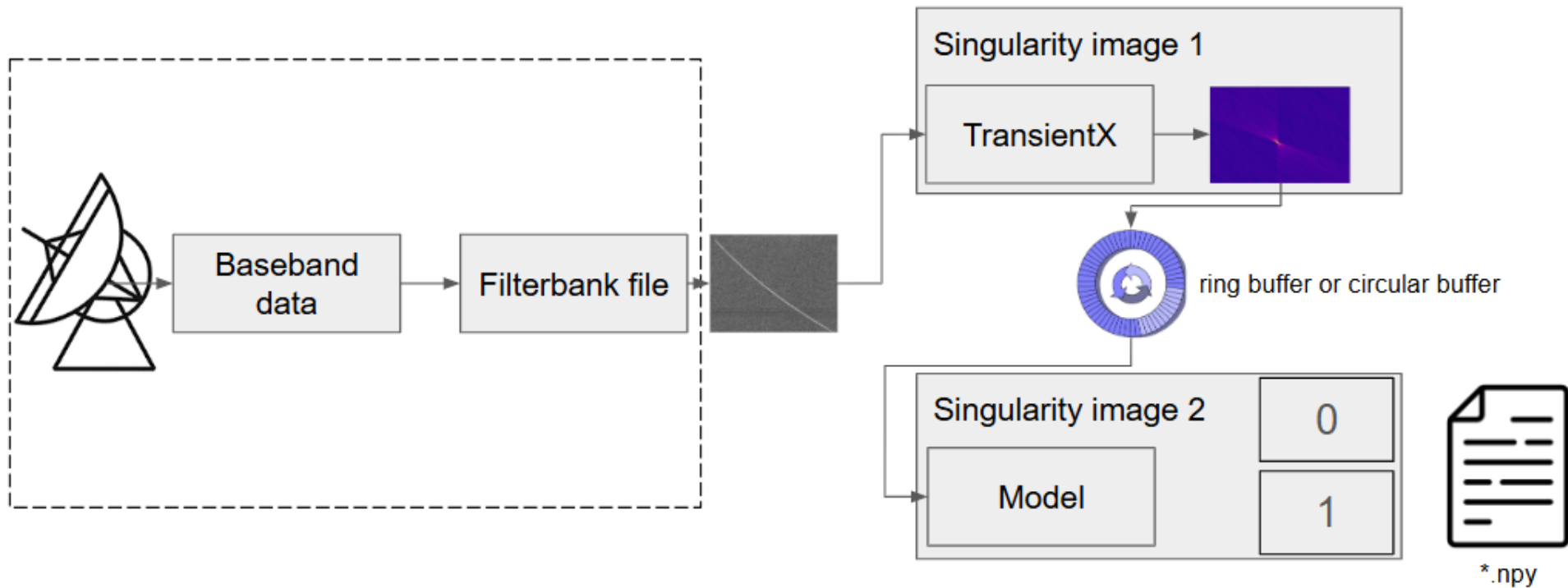
# Key Metrics for the Models on New Data

2_256	0.97	0.99	0.97	0.98
2_128	0.97	0.99	0.97	0.98
2_64	0.90	0.99	0.89	0.94
2_32	0.81	0.99	0.79	0.88
3_256	0.97	0.99	0.97	0.98
3_128	0.94	0.99	0.94	0.97
3_64	0.92	0.99	0.92	0.95
3_32	0.87	0.99	0.86	0.92
	Accuracy	Precision	Recall	F1 Score
Metrics				

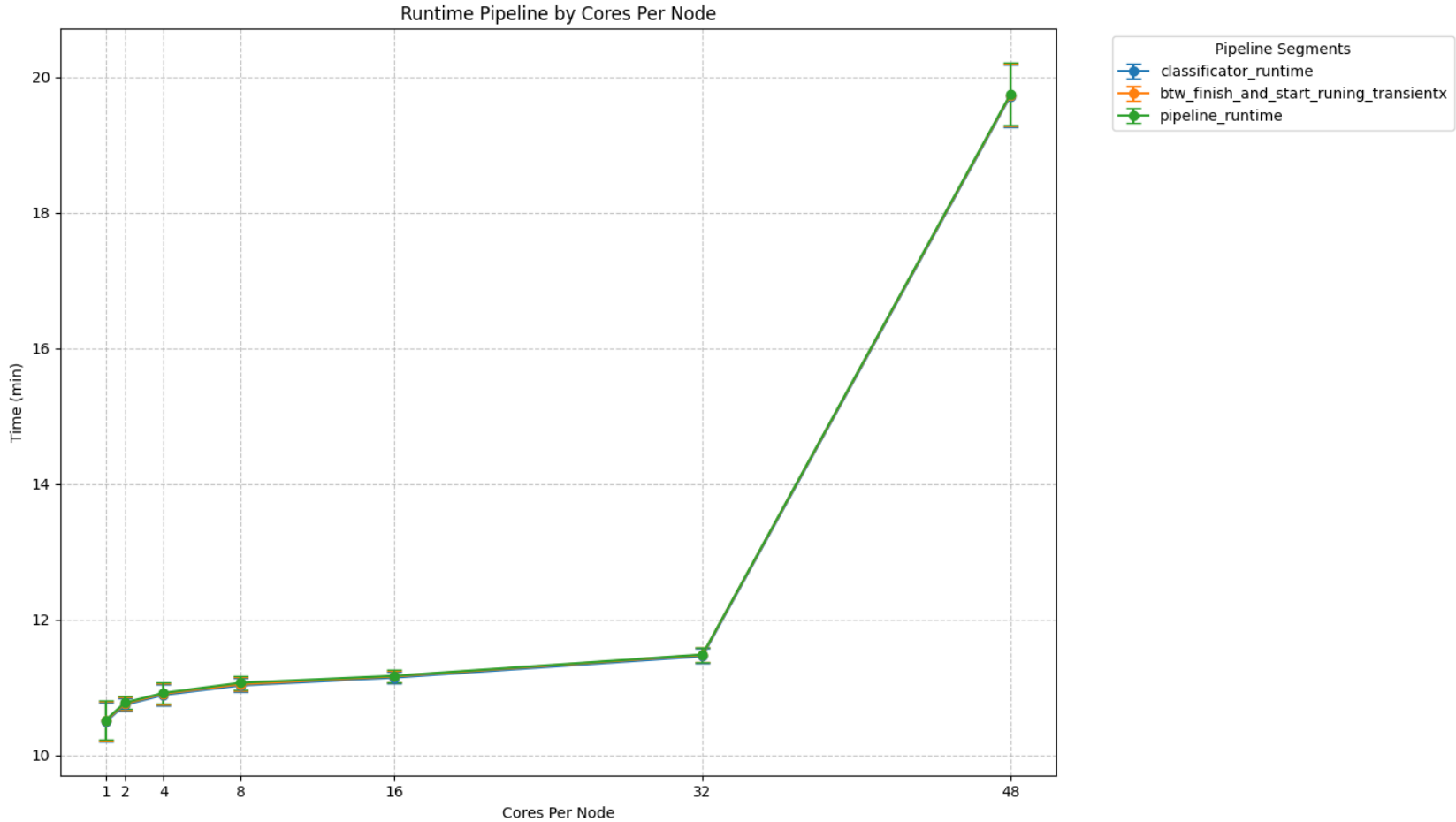
# SNR Sensitivity on the New Data



# The Pipeline for Pulse Classification in a Filterbank File<sup>39</sup>




# Performance Metric of the Pipeline






# Data Availability on the Edmond Platform

Search ▾ Help Support Log In

## Realtime identification of Dispersed Radio signals using ML - A Case Study on the Crab Pulsar

Version 1.0



Kazantsev, Andrei; Karuppusamy, Ramesh, 2025, "Realtime identification of Dispersed Radio signals using ML - A Case Study on the Crab Pulsar", <https://doi.org/10.17617/3.HQYC8O>, Edmond, V1

[Cite Dataset ▾](#) [Learn about Data Citation Standards](#)

[Access Dataset ▾](#)  
[Contact Owner](#) [Share](#)

Dataset Metrics ?

0 Downloads ?


**Description** ?

Robust realtime identification of dispersed radio astronomical signals that last much less than a second is challenging. Here we explore the utility of machine learning techniques to identify such signals and use data taken on the Crab pulsar using the Effelsberg 100m Radio Telescope. The data corresponds to the frequency range of 1240–1510 MHz, and contains 20 minutes of the pulsar signal. In addition, the DM-time data generated by the realtime pipeline, the associated Tensorflow CNN model is included and the training dataset are included. The data are intended for machine learning tasks focused on single-pulse detection and classification.

**Subject** ?

Physics

**License/Data Use Agreement**

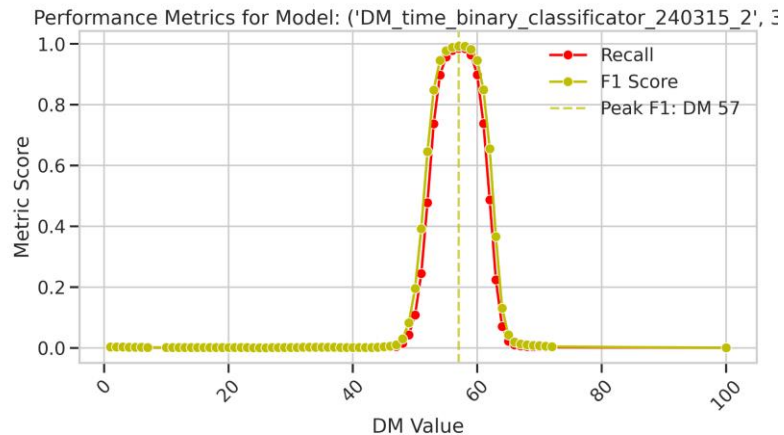
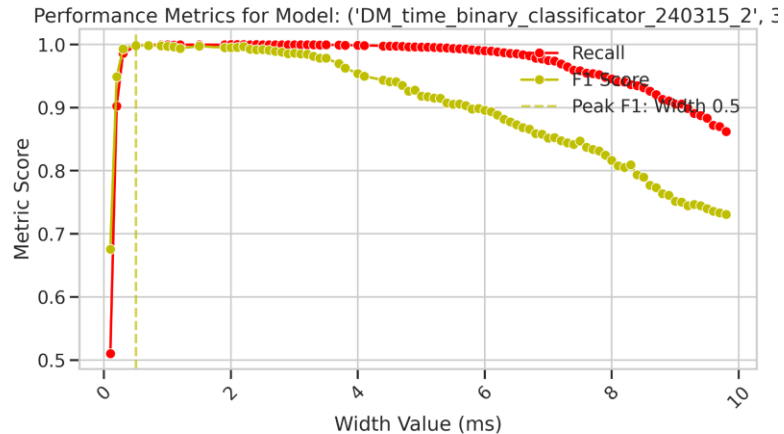
 [CCO 1.0](#)

<https://doi.org/10.17617/3.HQYC8O>

# Summary for the Current Stage

- The extremely minimalistic CNN model has been trained on DM-time images to detect individual pulses from the Crab pulsar with acceptable accuracy and sensitivity.
- The processing pipeline has been created to recognize individual pulses from Crab pulsar in filterbank files that includes TransientX and the trained model.
- The performance of the pipeline meets the set requirements.

# Next Steps



Rishi Kumar

Questions? Suggestions?