A neural networks method to search for long transient gravitational waves

Francesca Attadio, Leonardo Ricca, Marco Serra, Cristiano Palomba

Arxiv: https://arxiv.org/abs/2407.02391

Credits: ESO/L. Calçada

State of art

Goal: develop a computationally feasible procedure to search for gravitational waves (GW) emitted by newly born magnetars

Key points of our work

→ Standard search techniques are **computationally unfeasible** (matched filtering) or **very demanding** (sub-optimal semi-coherent methods), (B. P. Abbott+, APJL 851, L16 (2017)).

→ We developed a **classifier** exploiting machine learning techniques, previous work: A. L. Miller+, PRD 100, 062005 (2019).

→ To help the classification task, we developed a **denoiser** .

We have used **simulated noise** (according to the detectors noise curve) and **simulated signals** to measure the performances of this procedure

Master plan

→ **Final goal**: get candidates for the signal (second phase detailed study to extract signal parameters).

→ Use a **fraction of the dataset to train the neural networks** and then **perform the analysis on the remaining sample.**

→ **Feasibility study of this approach on simulation**, verification of "what" is needed for training and computational load.

→ Address **training on real data**.

Physical problem

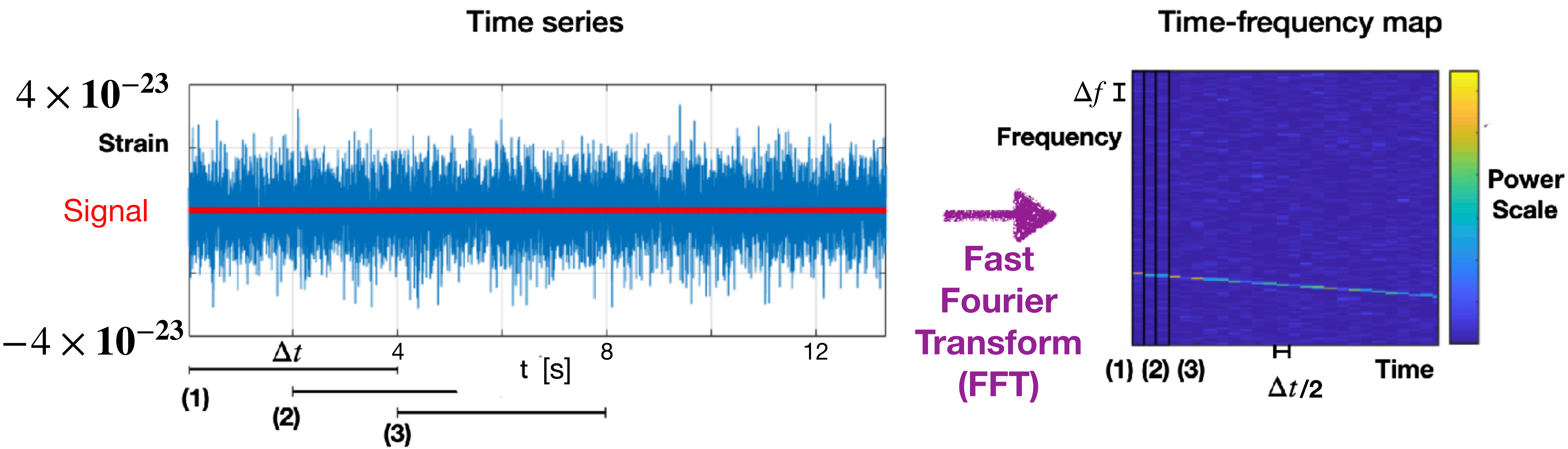
- Isolated neutron star (NS) spinning with a non axi-symmetric asymmetry
- GW frequency linked to the rotational frequency
- Ellipticity (ϵ) measure of the asymmetry

Why study GW emitted by newly born magnetars?

- Strong inner magnetic field ($B \sim 10^{15} - 10^{16}$ G) can induce significant ellipticities.
- Magnetars represent up to 20% of NS.

1)S. Dall'Osso+ (Springer International Publishing, 2021) p. 245–280
2) V. M. Kaspi+, ARAA 55, 261–301 (2017)
3)P.Beniamini+, MNRAS 487, 1426 (2019).

Time-frequency maps



- We construct time frequency maps with a **time resolution** of $\Delta t/2 = 2$ s and **frequency resolution** of $\Delta f = 0.25$ Hz.
- Squared maps with a **time interval** of 1200 s and a **frequency interval** of 150 Hz
- These choices were suggested by **physical properties of the signal** and **computational considerations.**

Signal

GW amplitude

$$h_0(t) = \frac{4\pi^2 G}{c^4} \frac{I\, f(t)^2}{d} \epsilon$$

Distance of the source

Frequency variation

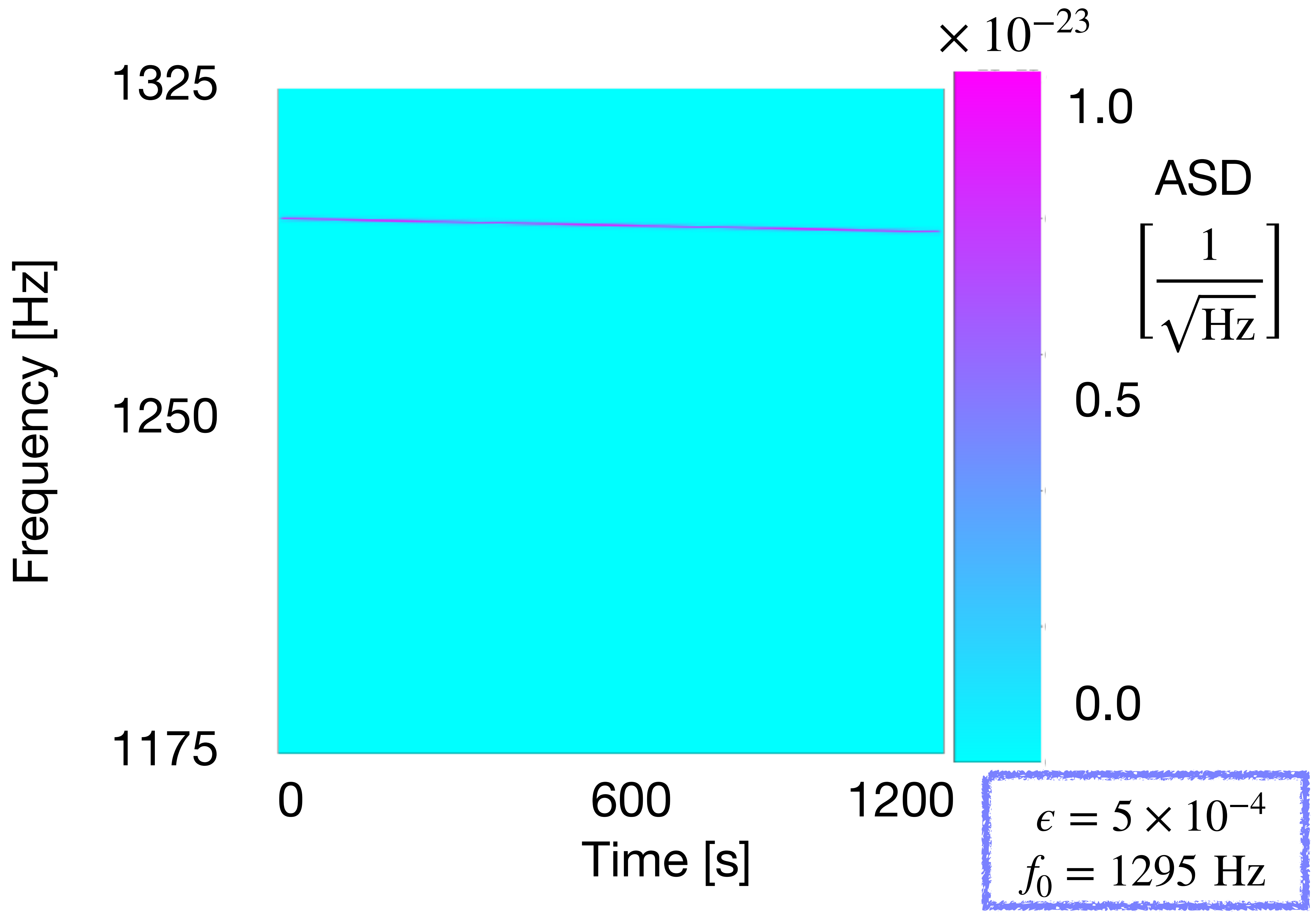
$$\dot{f}(t) \propto -\epsilon^2 f(t)^5 \longrightarrow f(t) = f_0 \left(1 + \frac{t\, \epsilon^2 f_0}{const}\right)^{-\frac{1}{4}}$$

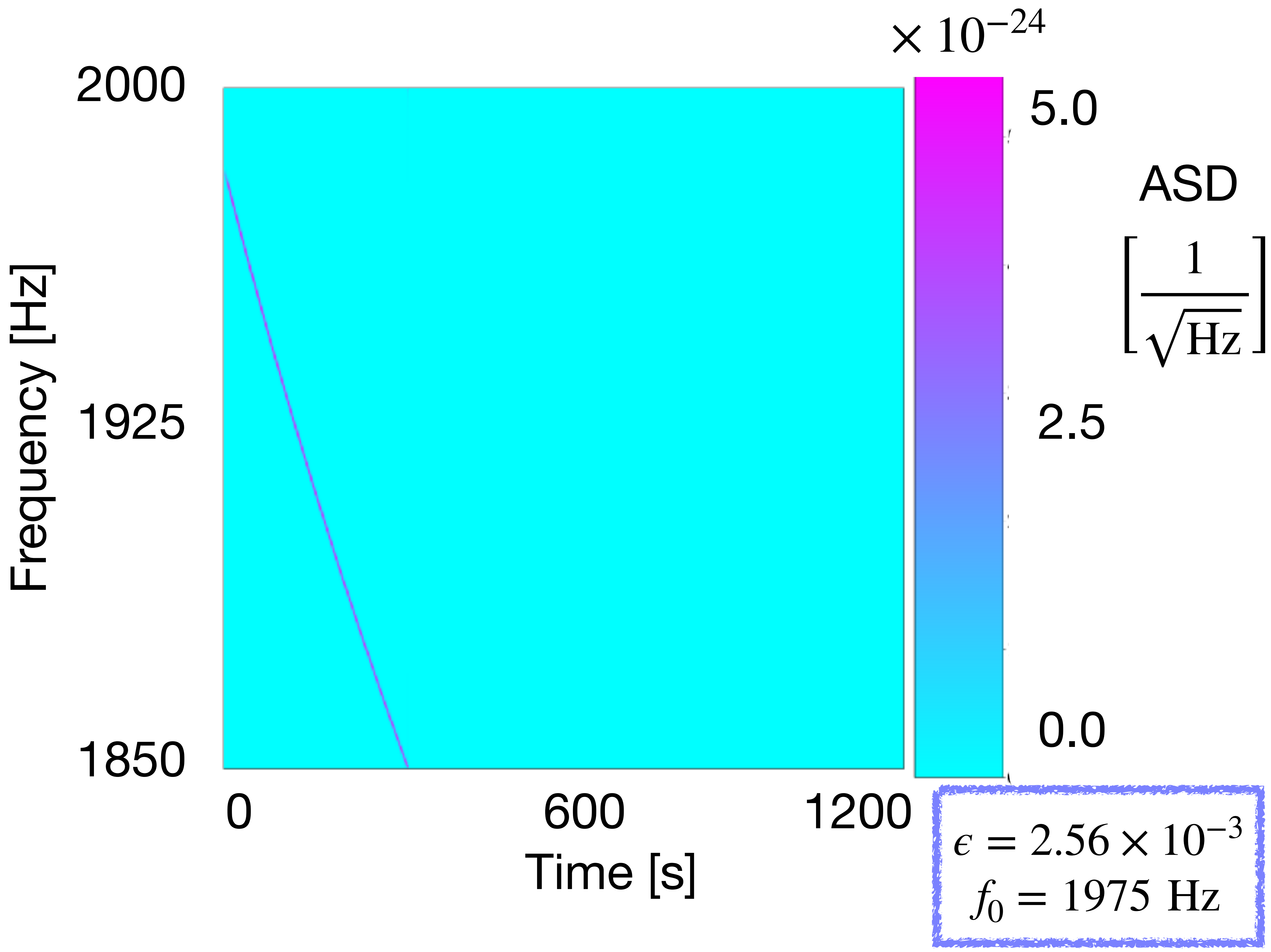
- **Reference initial amplitude :** 2×10^{-23} **(different amplitudes tested)**
- **Fixed inclination angle:** $\iota \sim 56°$
- **Fixed momentum of inertia:** $I = 1.4 \times 10^{38} \text{kg m}^2$

Parameters range: $\epsilon \in [3,30] \times 10^{-4}$ $f_0 \in [1.25,2.00]$ kHz

S Dall'Osso+, Monthly Notices of the Royal Astronomical Society, vol. 480, no. 1, pp. 1353-1362, July 2018

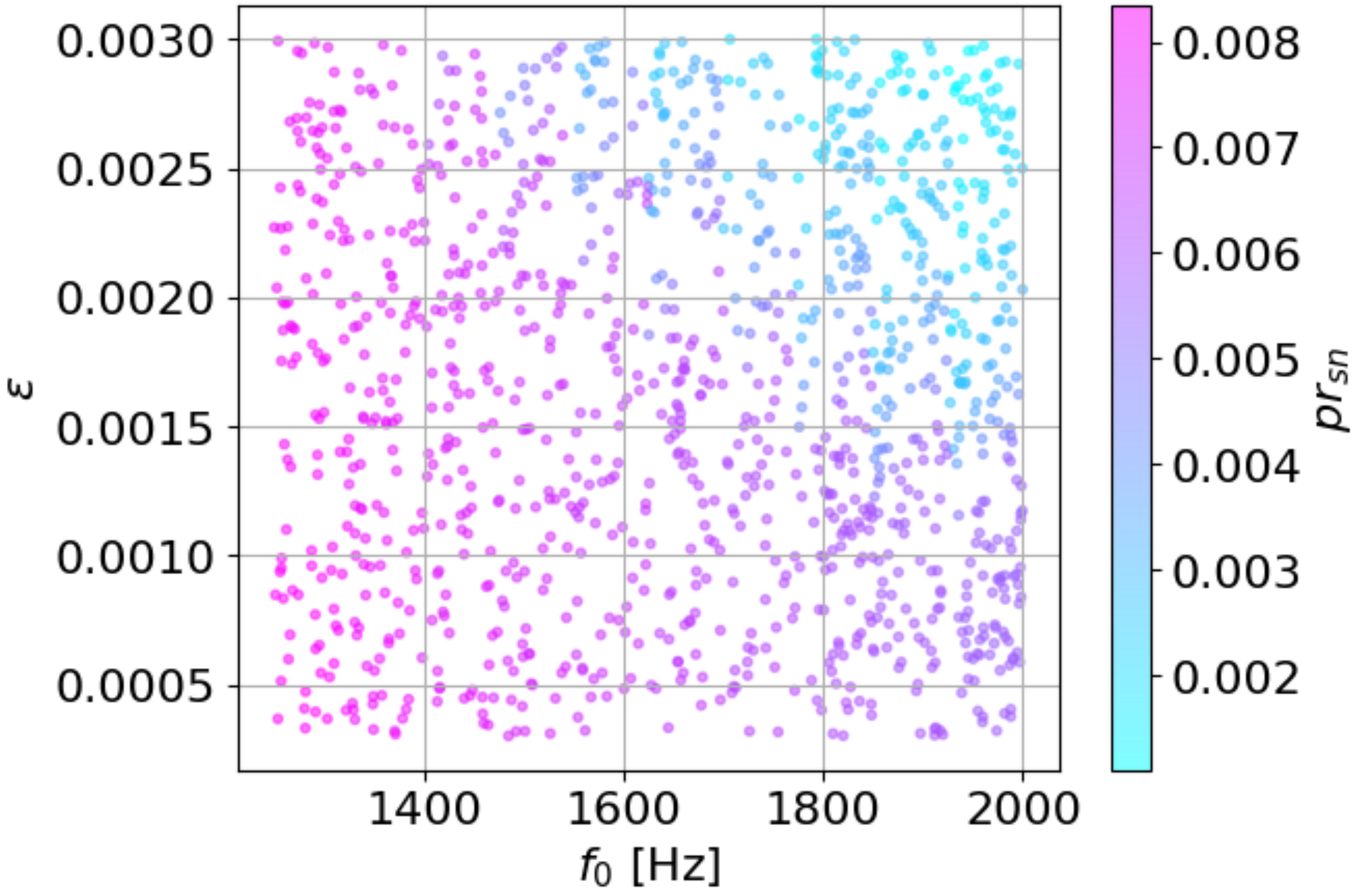
Signal





→ We are not focusing on standard continuous waves.

Pixel Signal-to-Noise ratio (pr_{sn})



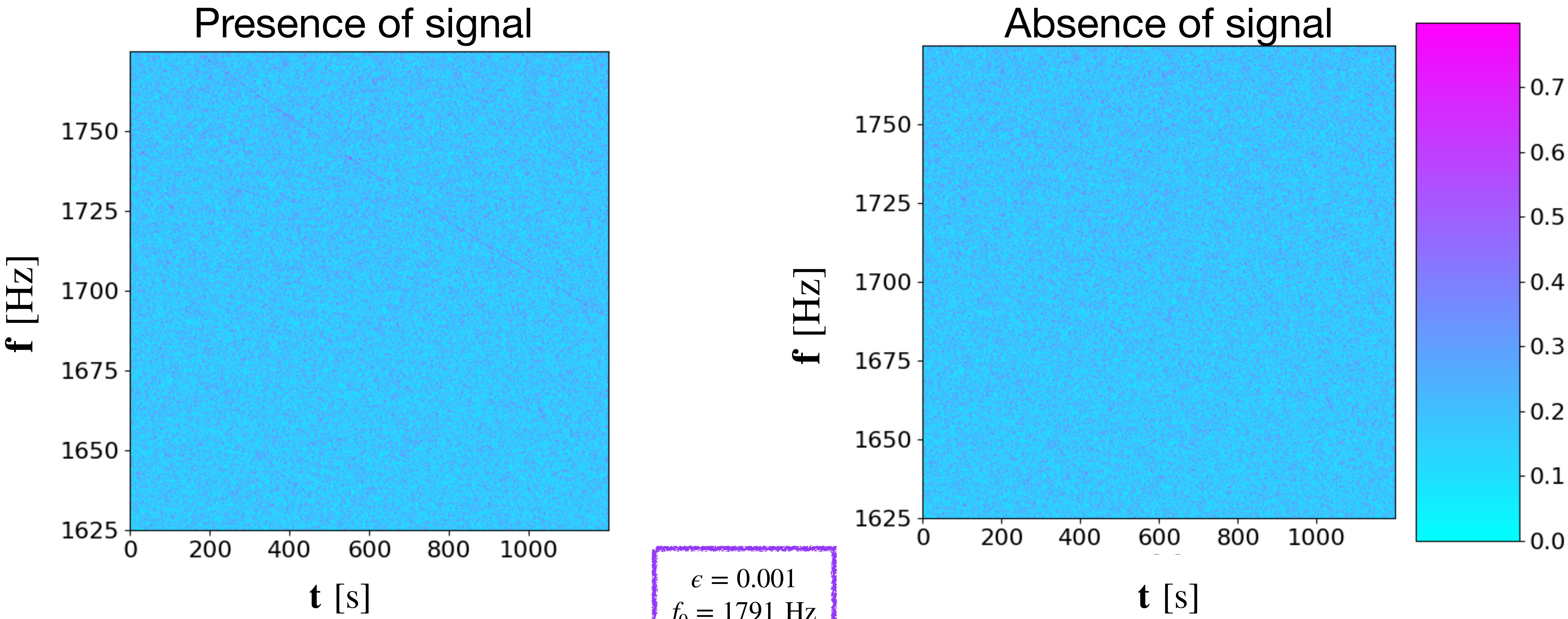
$$pr_{sn} = \frac{1}{N_{pix}} \sum_{N_{tf} \neq 0} \frac{S_{tf}}{N_{tf}}$$

→ It estimates **how much of the signal emerges from the noise level** (before the denoiser)

→ As f_0 and ϵ increase, pr_{sn} decreases.

Classifier

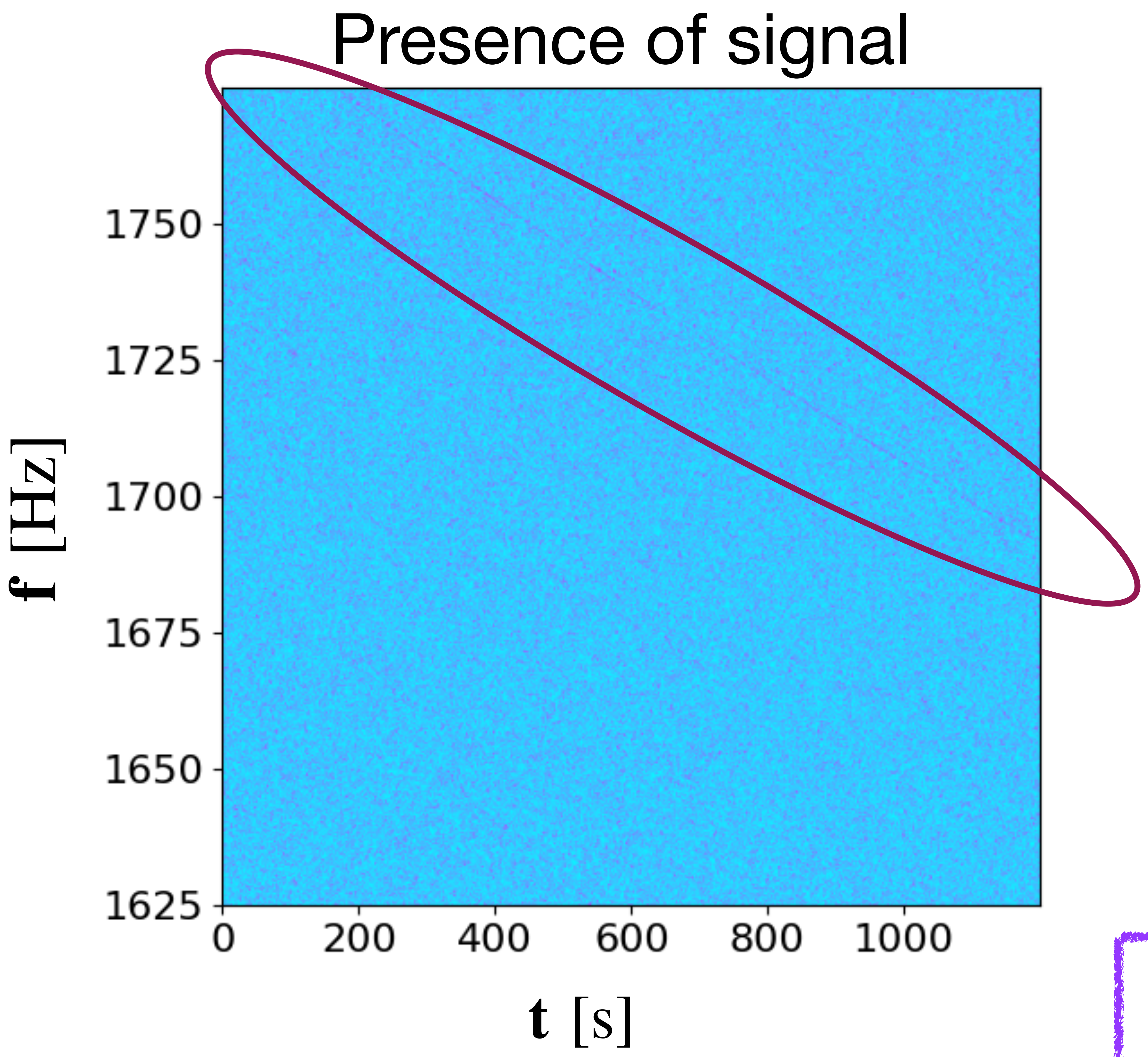
Classification of time-frequency maps



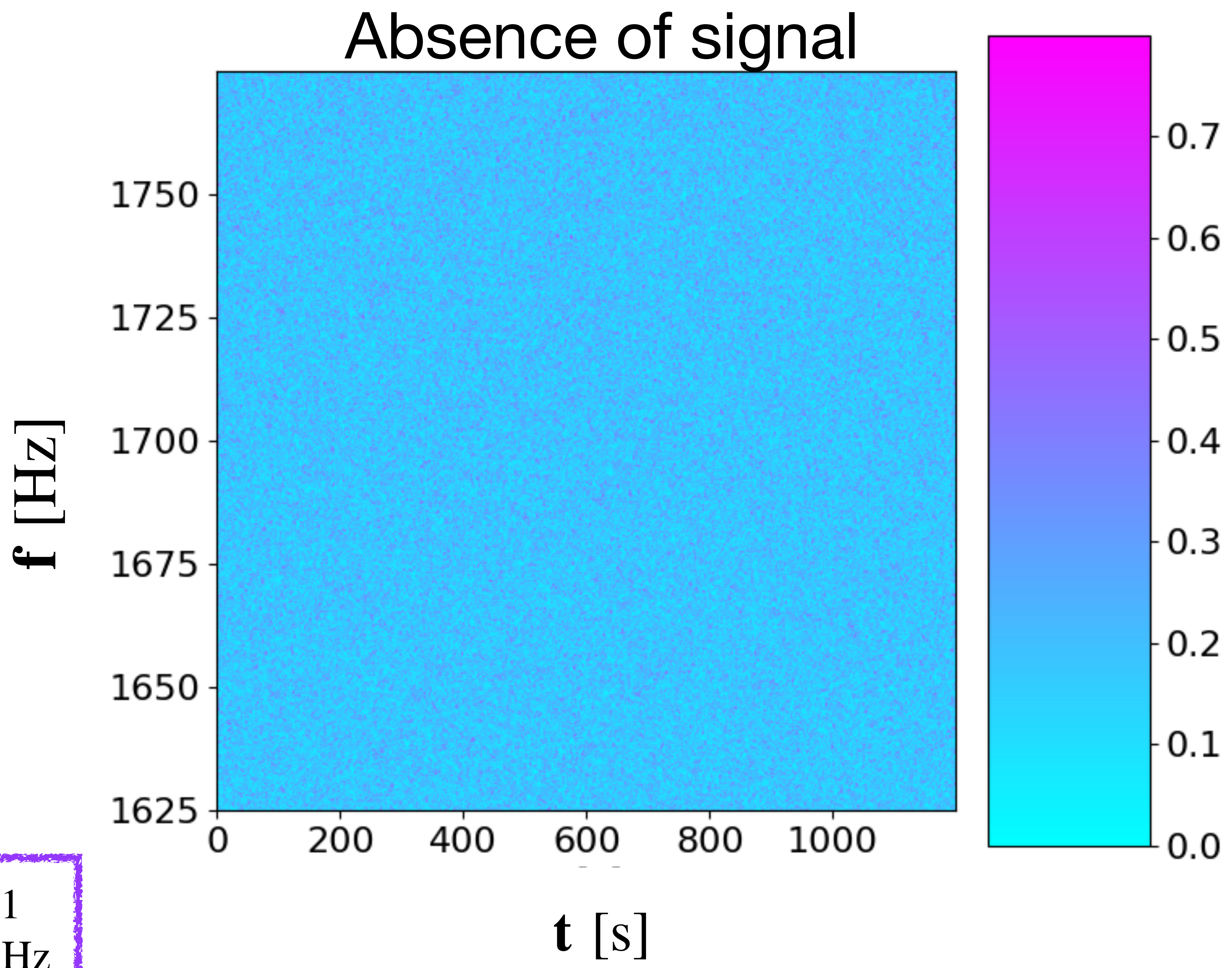
$\epsilon = 0.001$
$f_0 = 1791$ Hz

Classifier

Classification of time-frequency maps

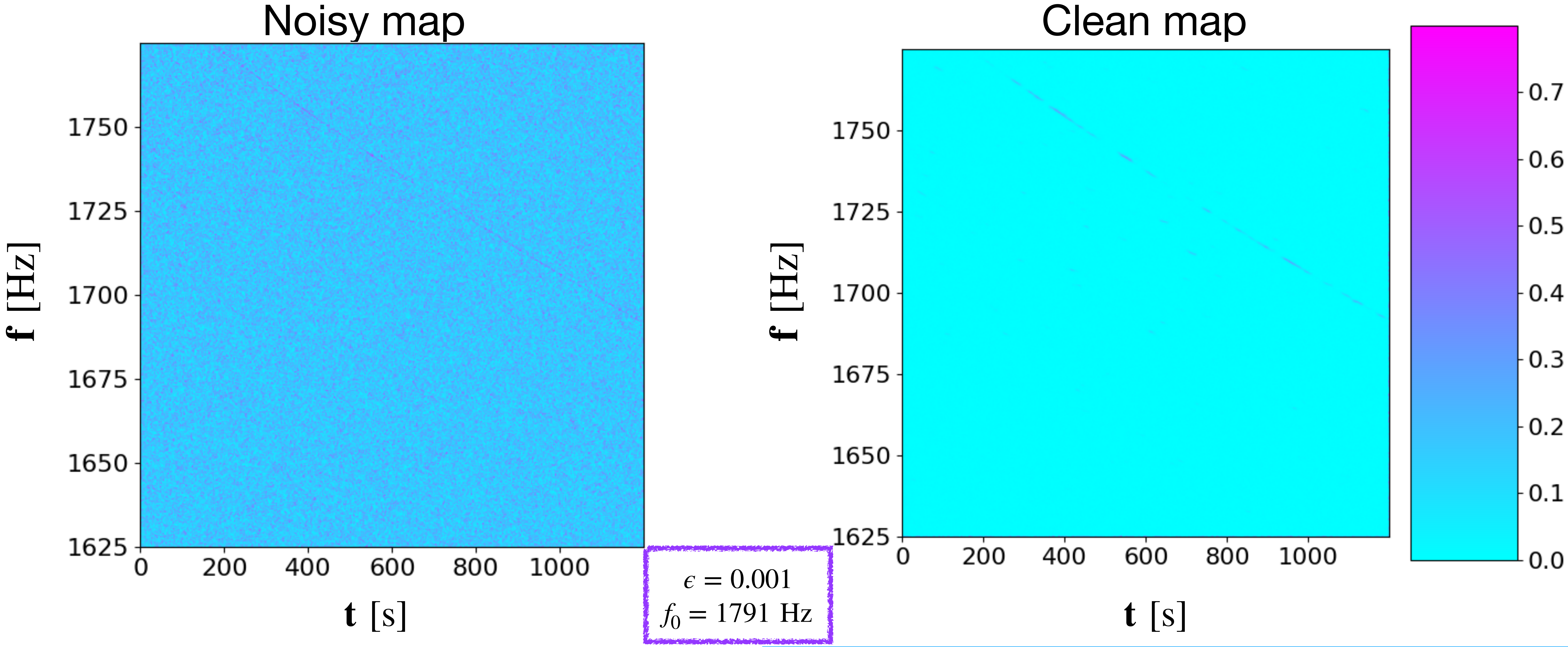


$\epsilon = 0.001$
$f_0 = 1791$ Hz



Denoiser

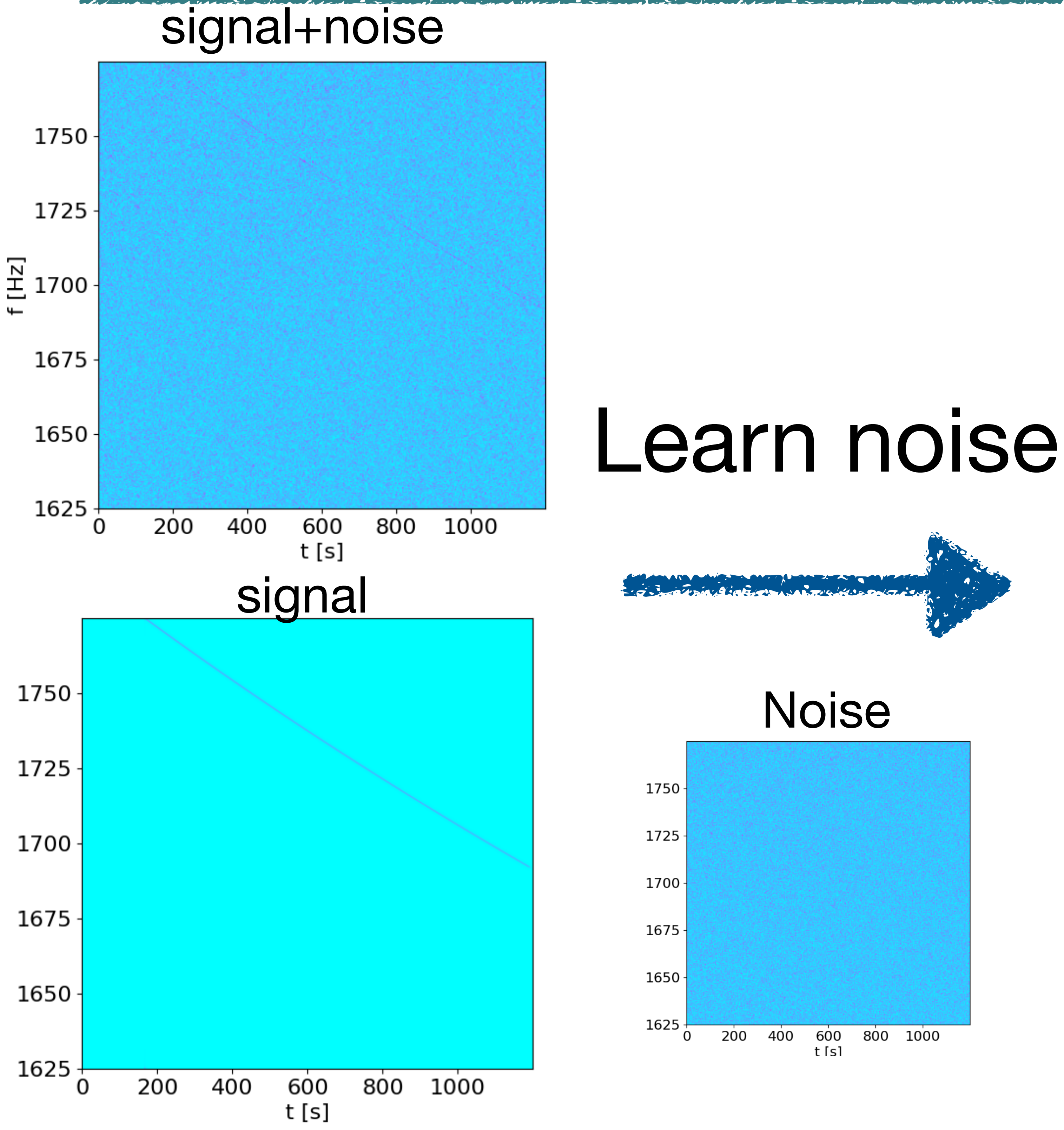
It reduces the noise level of the image while preserving a significant fraction of the signal

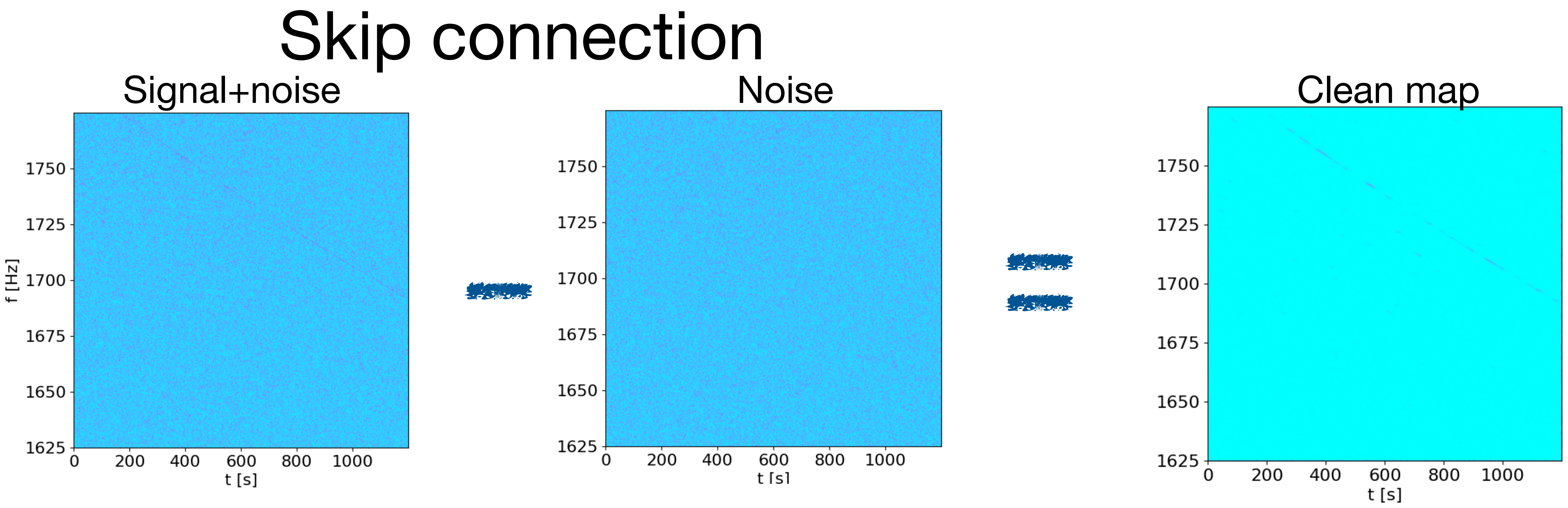


$\epsilon = 0.001$
$f_0 = 1791$ Hz

Denoising process

Residual learning approach: our neural network learn the noise behavior.





K. Zhang+, IEEE Transactions on Image Processing 26, 3142 (2017).

→ Our goal is to train the denoiser with real data.

Useful classification glossary

The frequency can vary rapidly in time and **the signal can cross more maps**

→ **Positive trigger**: at least one map is correctly tagged

→ **Efficiency per signal (Eff):** $\frac{\#positive\ trigger}{\#signals}$

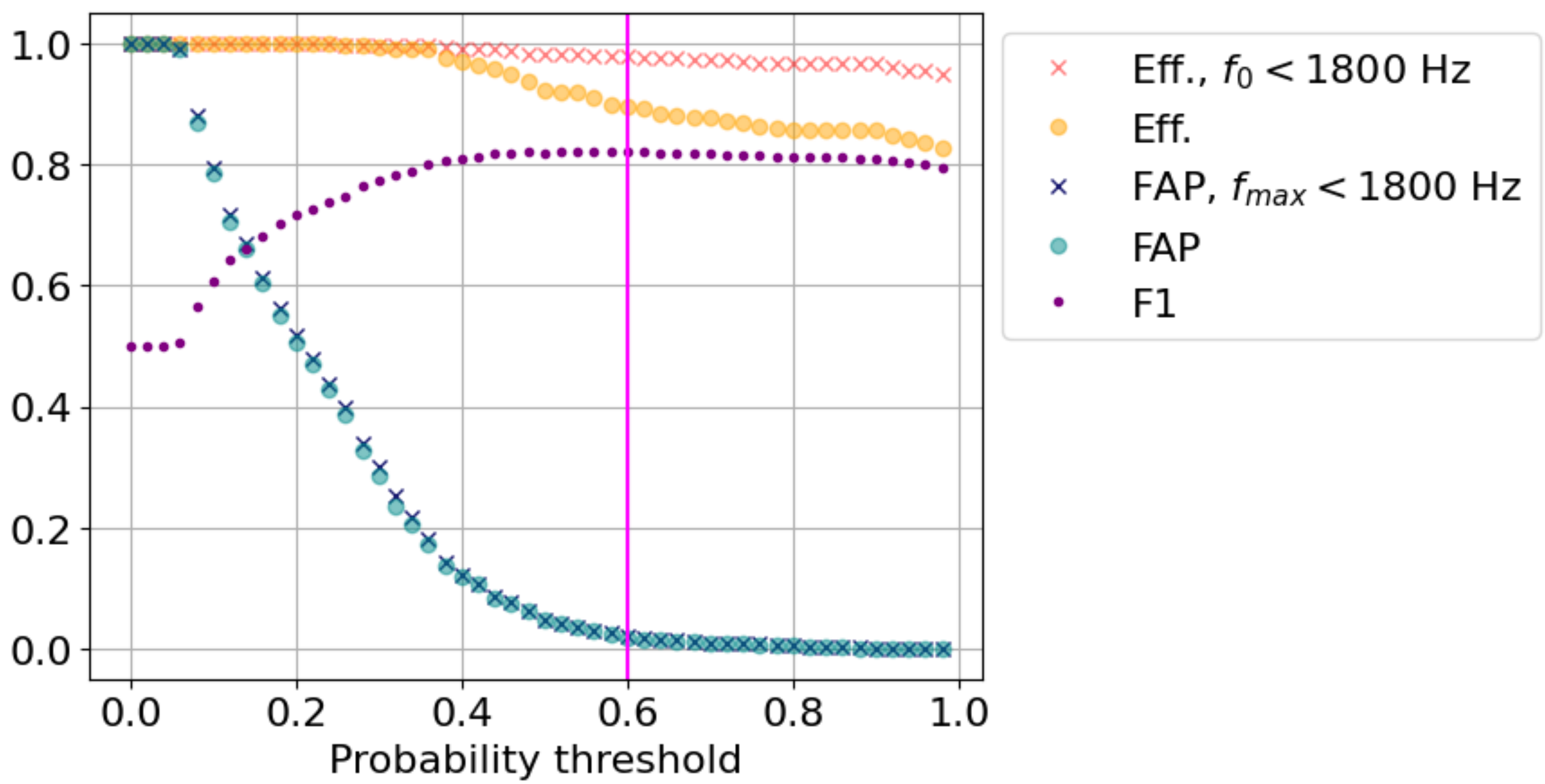
→ **False alarm probability (FAP)**: $\frac{\#false\ positives}{\#noise\ maps}$

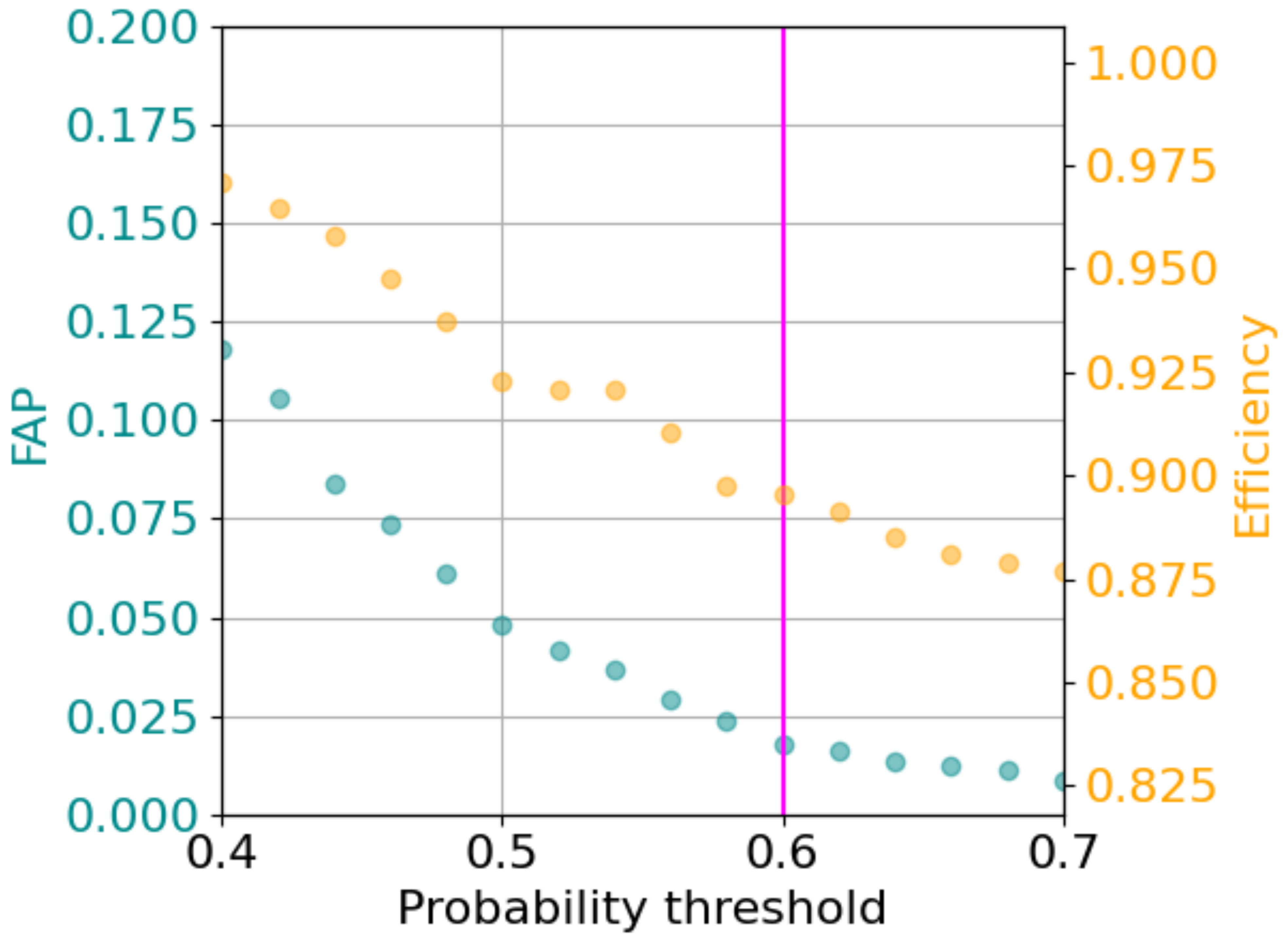
To obtain a **good balance** between Eff and FAP, we use the **F1 score**

$$\mathrm{F}_1 = 2\frac{p \cdot r}{p + r} \qquad p = \frac{true\ positive}{predicted\ positive} \qquad r = \frac{true\ positive}{true\ positive + \ false\ negative}$$

Classification process

Classifier output: probability of presence and absence of signal.



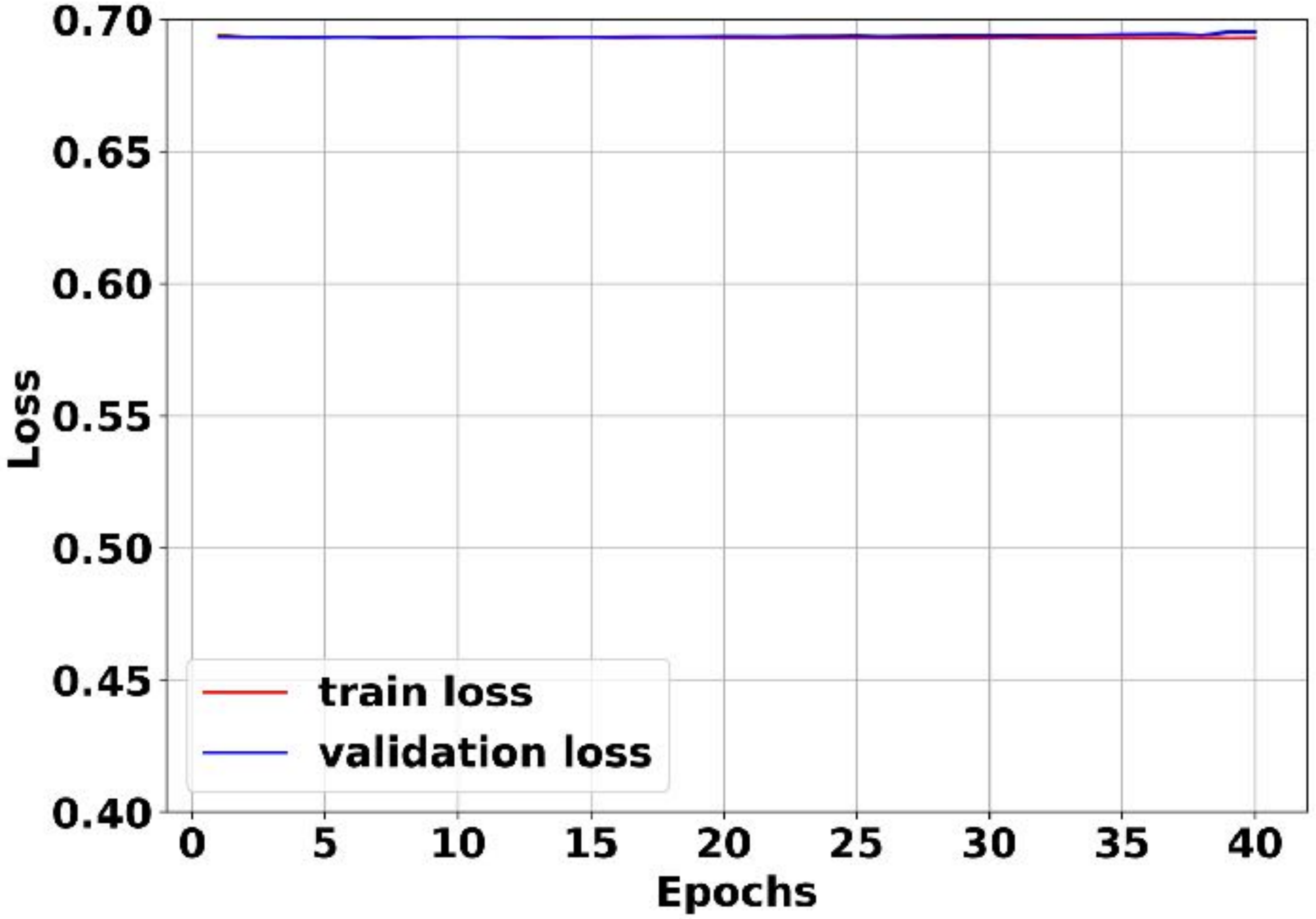


$$\mathrm{FAP} = 2\,\% \qquad \mathrm{FAP}_{1800} = 2\,\%$$
$$\mathrm{Eff} = 90\,\% \qquad \mathrm{Eff}_{1800} = 98\,\%\,.$$

We are using a single detector

Classification

Noisy maps

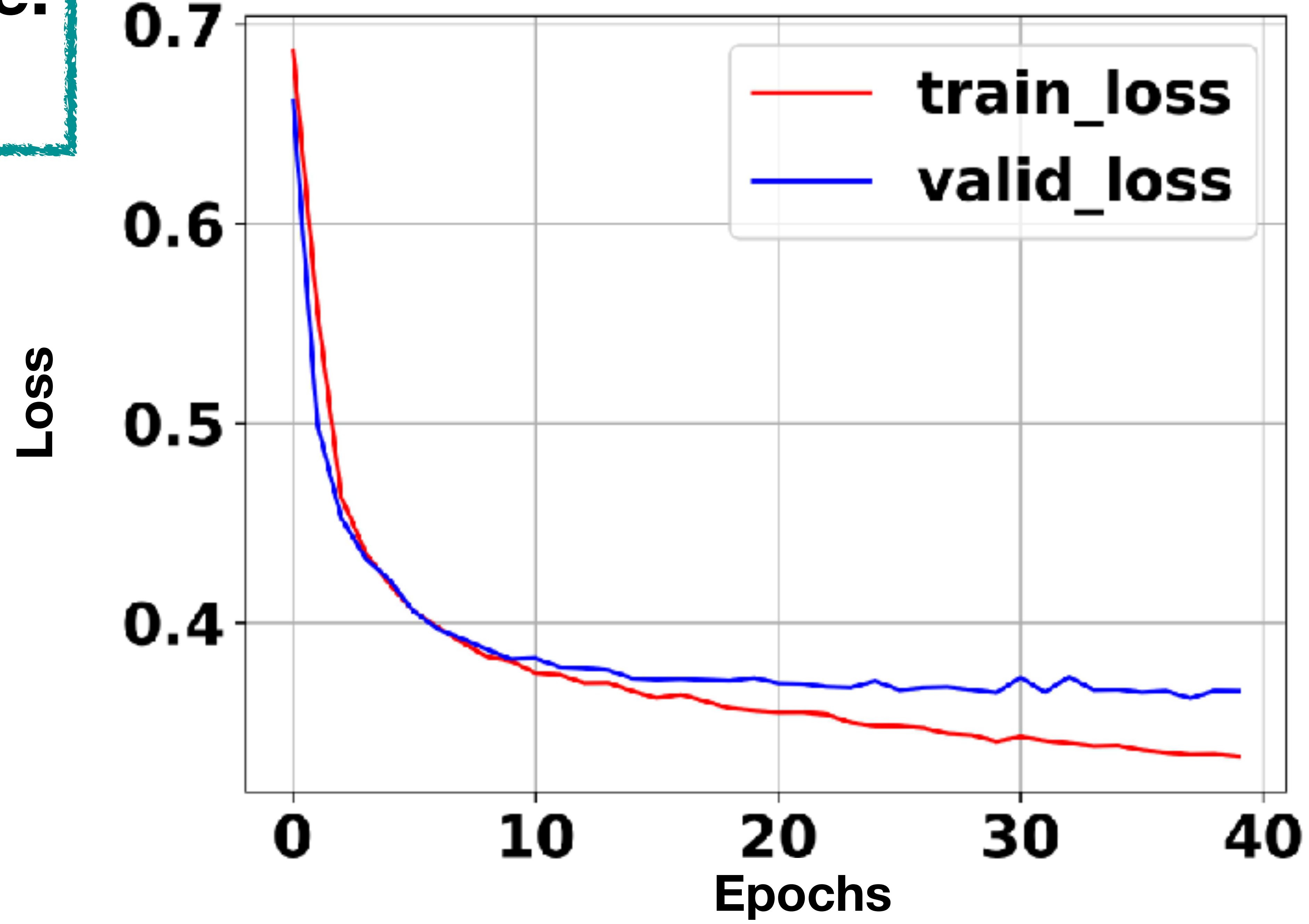


Initial signal amplitude:
2×10^{-23}

Train loss

Validation loss

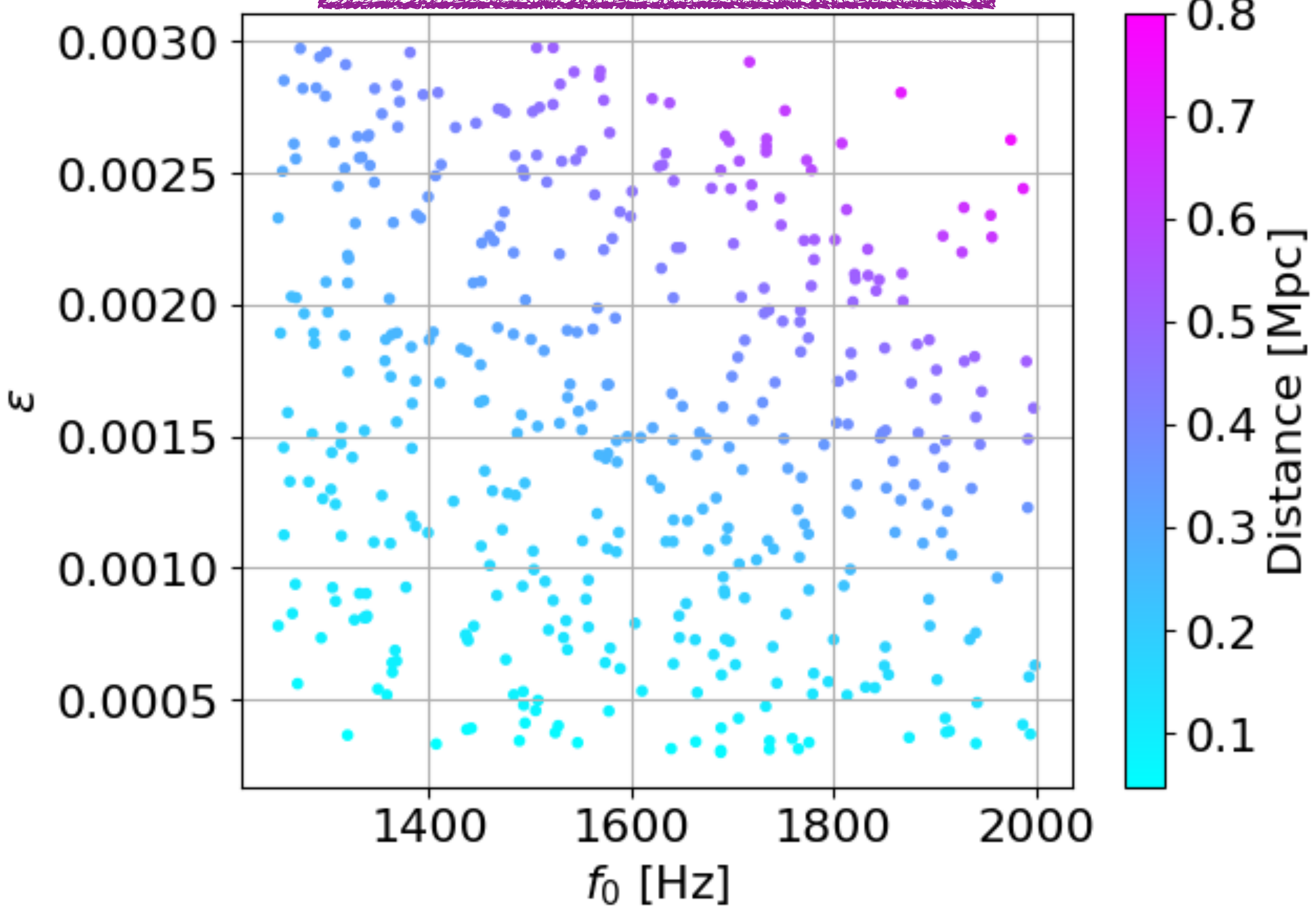
Clean maps



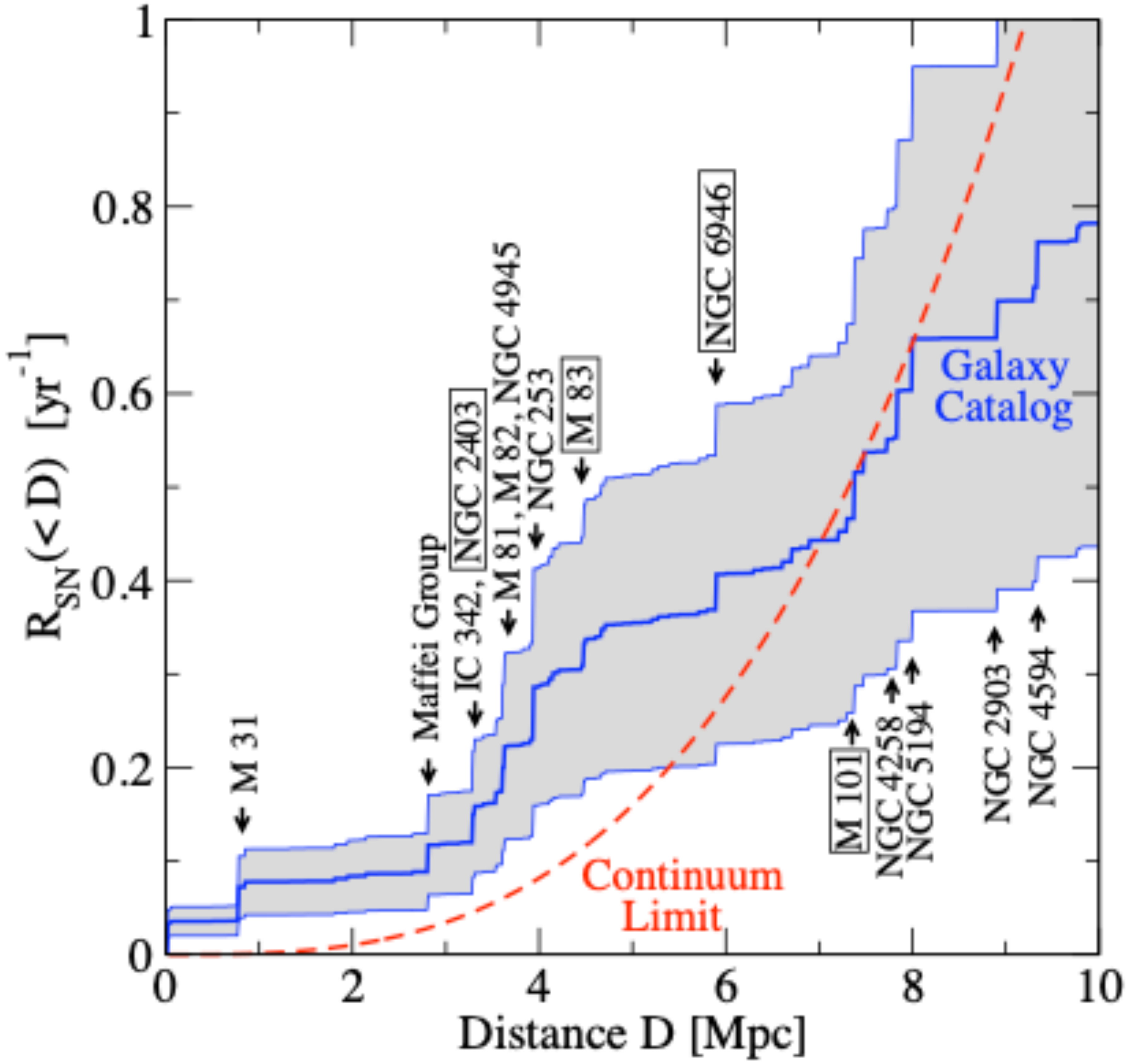
- With noisy maps the **classifier do not learn** to discriminate between presence or absence of signal.
- **The denoiser is crucial!!!**

Distance

Positive triggers



Supernova rate



S. Ando+ Phys. Rev. Lett. 95, 171101 (2005)

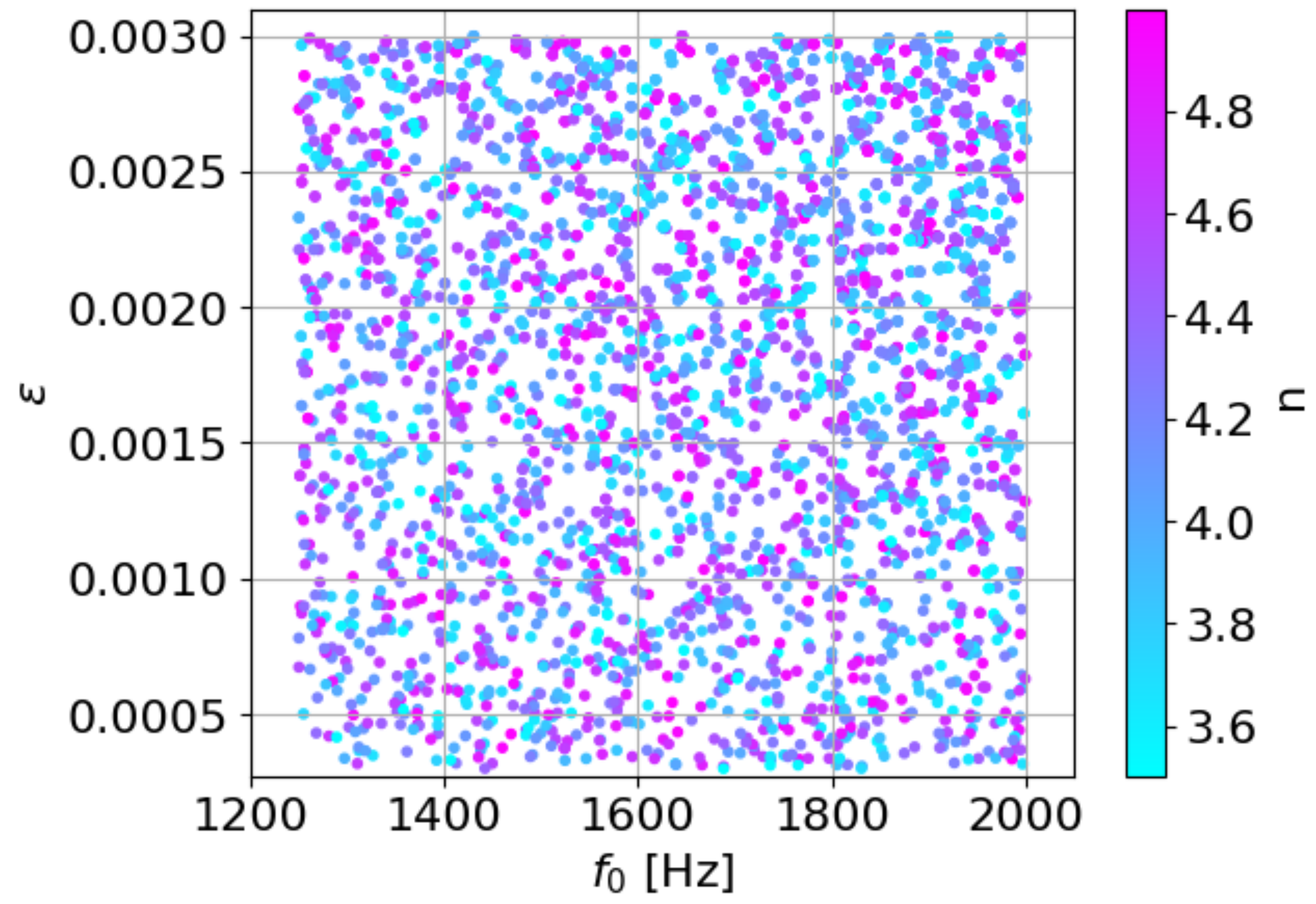
We need to reach higher distances to increase the probability to see an event

Changing braking index

Frequency variation

$\dot{f}(t) = -kf(t)^n$ → Up to now we assumed n = 5, i.e. **purely gravitational radiation** → $f(t) = f_0\left(1 + \frac{t\,\epsilon^2 f_0}{const}\right)^{-\frac{1}{4}}$

We tested our trained denoiser (n=5) on n $\in [3.5, 5]$, keeping the same k.



$$\text{FAP} = 1\,\% \qquad \text{FAP}_{1800} = 1\,\%$$

$$\text{Eff} = 71\,\% \qquad \text{Eff}_{1800} = 67\,\%\,.$$

→ Testing with simulations our method is robust to different values of n.

Computational load

Key points: the rapidity of this technique and the limited requirements in terms of computing power

Denoiser

- **Training time***: ~ 3 hours (2198 + 550 maps)
- **Denoising time***: $\sim$ 20 minutes (5296 maps)

Classifier

- **Training time***: 30 minutes (3389 + 848 maps)
- **Classifying time***: $\sim$ minutes (1059 maps)

Real search: 40 days of interferometer data should be sufficient to train the whole NN pipeline.

*Detailed info in the reference paper

Conclusions

According to our study, it should be possible, with the current architecture, to search for GW emitted by newly born magnetars and it seems to be computationally feasible.

→ We need to **improve the denoiser** to explore greater distances

→ With this procedure we can reach **distances** up to ∼0.8 Mpc (O4)

Future developments

→ Conduct our study on **real data**

→ Combine **different interferometers**

→ Use sensitivity curve **future detectors** (e.g. ET).

Final goal: GW search in “O4” (2023-2025)

Conclusions

According to our study, it should be possible, with the current architecture, to search for GW emitted by newly born magnetars and it seems to be computationally feasible.

→ We need to **improve the denoiser** to explore greater distances

→ With this procedure we can reach **distances** up to ∼0.8 Mpc (O4)

Future developments

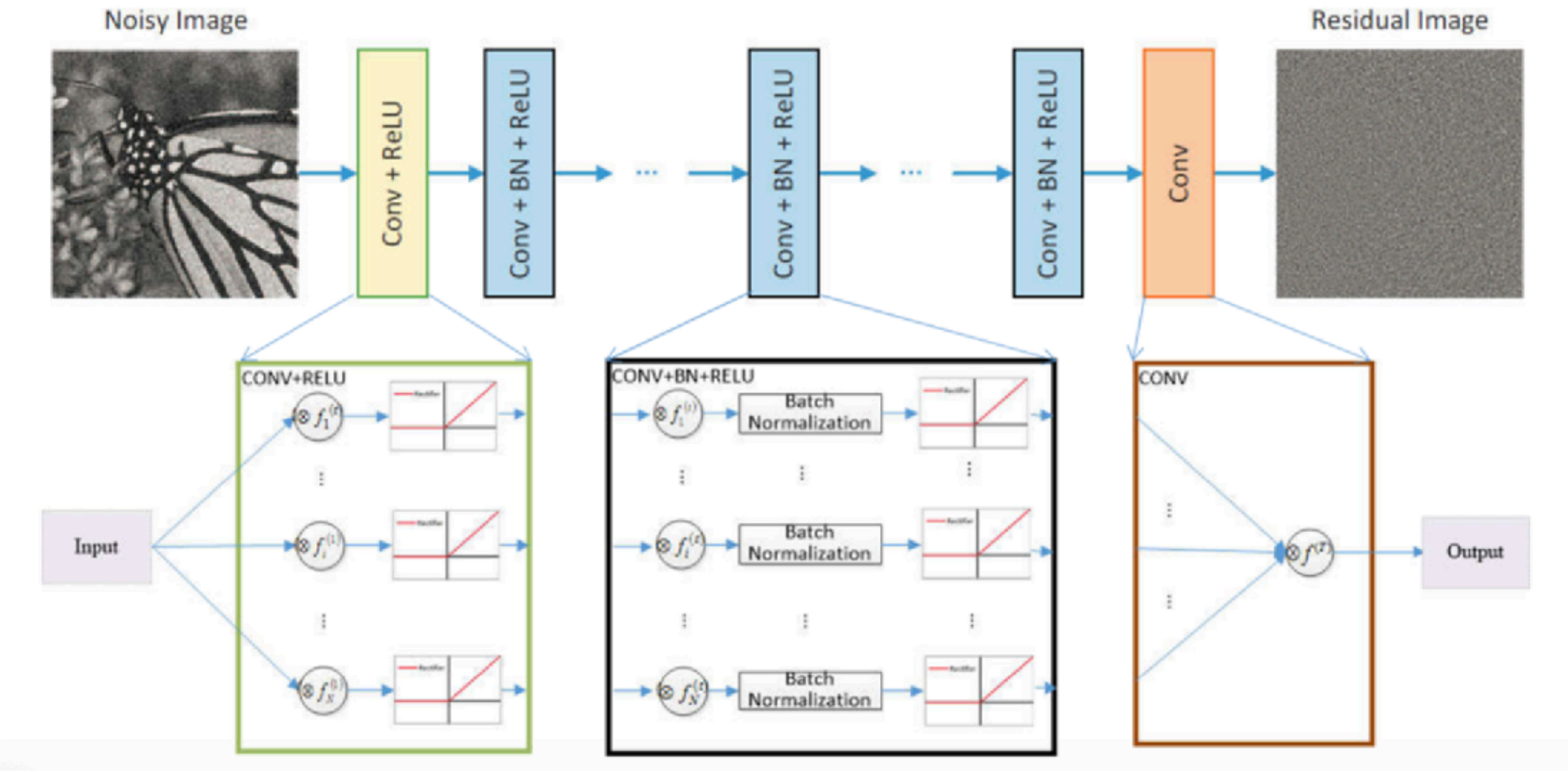
→ Conduct our study on **real data**
→ Combine **different interferometers**
→ Use sensitivity curve **future detectors** (e.g. ET).

THANK YOU FOR YOUR ATTENTION

Final goal: GW search in “O4” (2023-2025)

Backup slides

Model architecture: denoiser

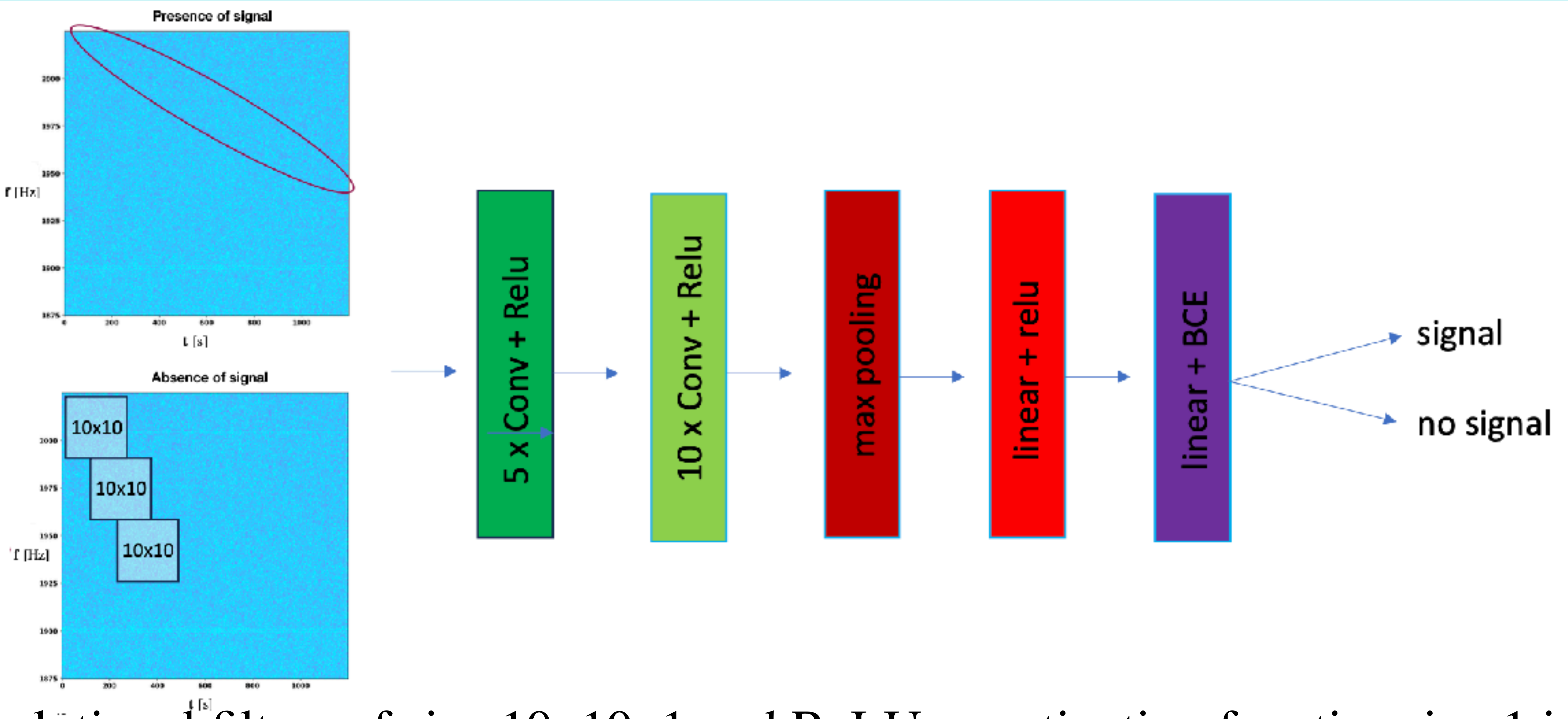


- **First layer**: 64 convolutional filters of size 3x3x1 and ReLU as activation function
- **Second group of layers**: six layers, each one with 64 convolutional filters of size 3x3x64, later, we have batch normalization and finally ReLU as an activation function.
- **Last layer**: 1 convolutional filter of size 3x3x64.
- **Skip connection**

Model architecture: classifier

Presence of signal

Absence of signal



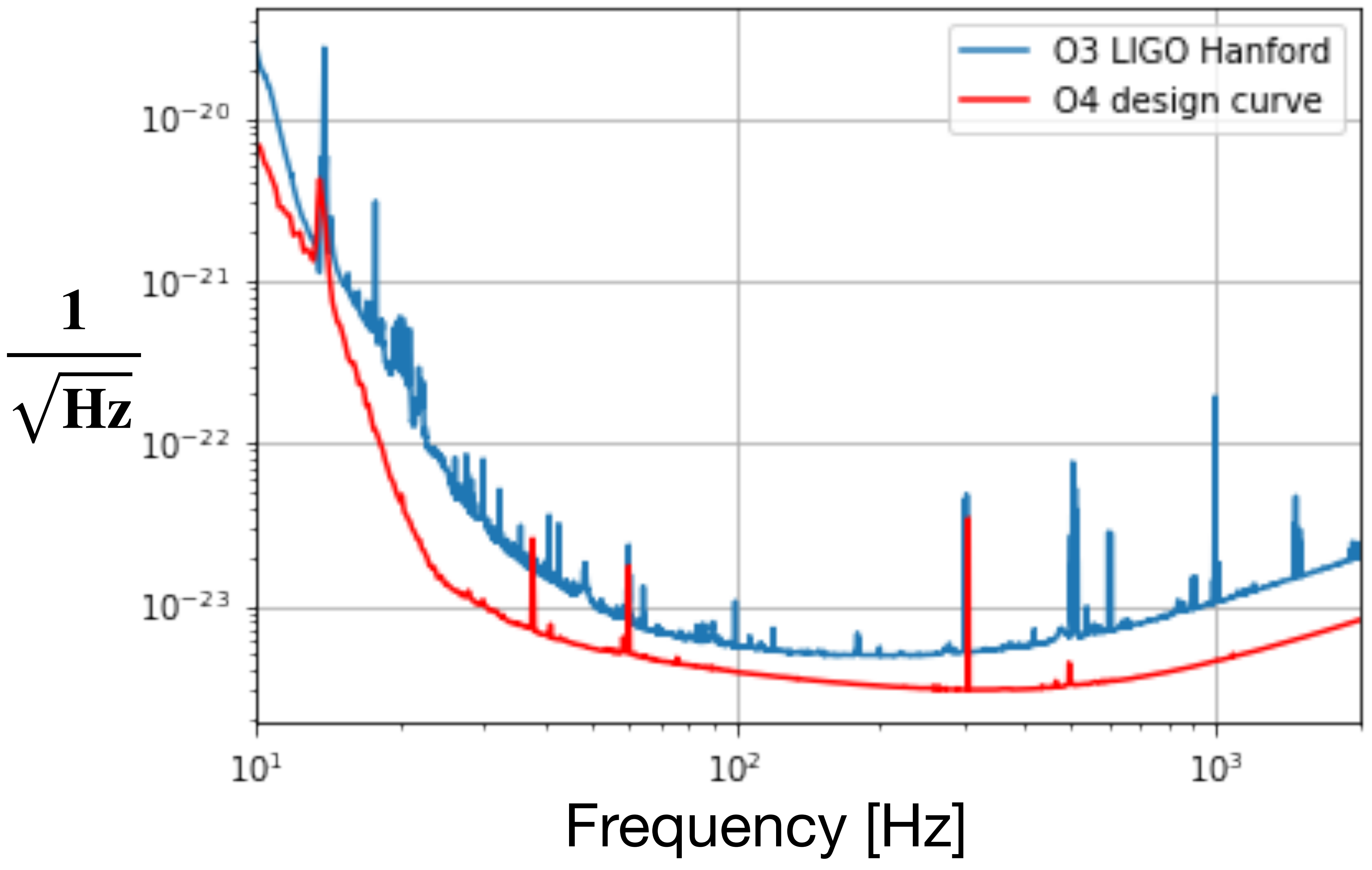
- **First layer**: 5 convolutional filters of size 10x10x1 and ReLU as activation function, i.e. 1 input channel and 5 output channels. The filters move with a stride of 5.
- **Second layer**: 10 convolutional filters of size 6x6x1 and ReLU as activation function, i.e. 5 input channels and 10 output channels. The filters move with a stride of 3.
- **Third layer**: 1 max pooling layer with a kernel of size 5x5x1.
- **Fourth layer**: one linear layer that reduces the dimension from 1690 to 84 and ReLU as an activation function.
- **Last layer**: one linear layer that passes from 84 numbers to 2.

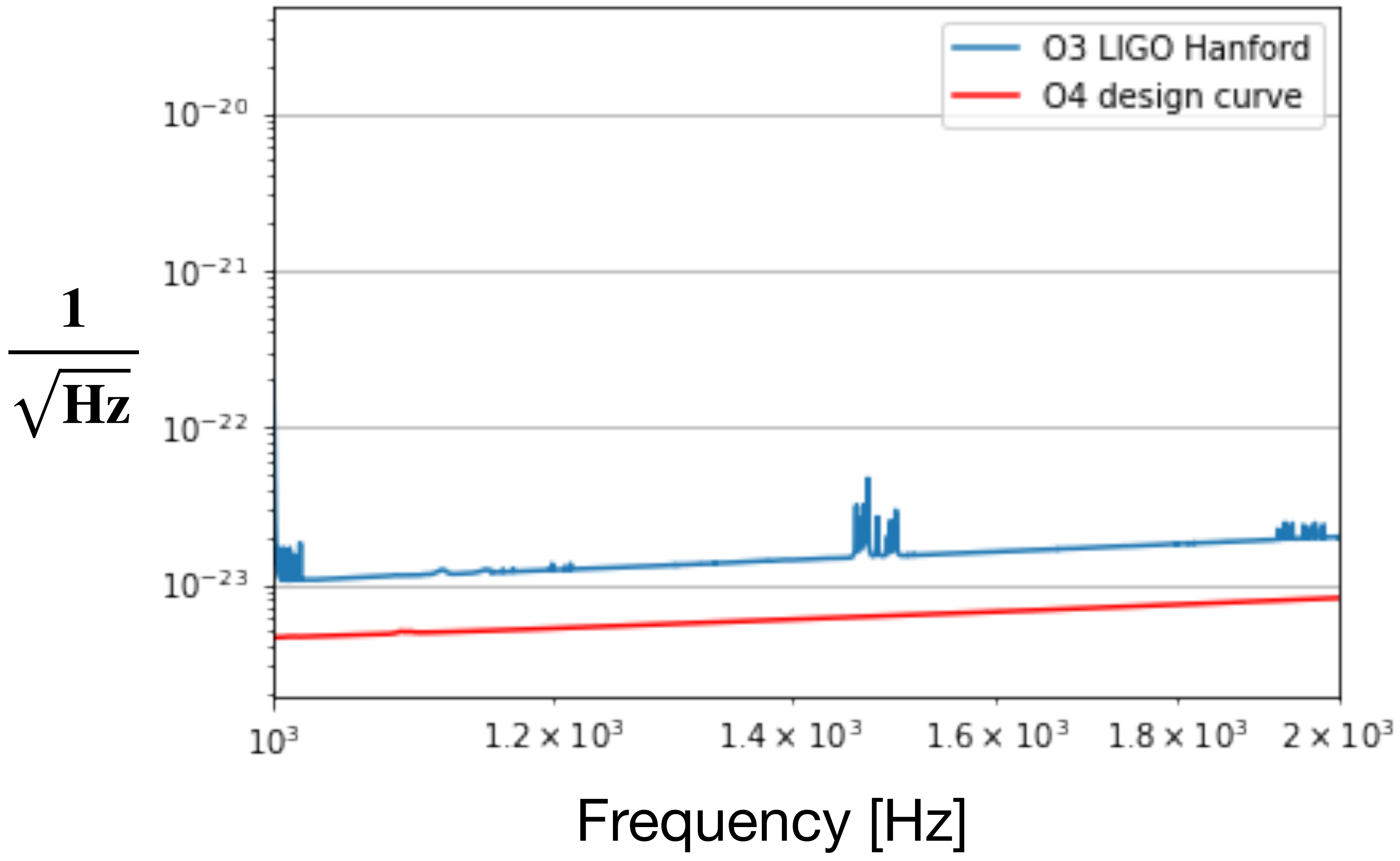
Noise

Simulated data:

→ Simulated noise according to the noise curve

Gaussian frequency dependent noise





Noise curves used for Simulations in the update of the Observing Scenarios Paper
LIGO Document T2000012-v2

Signal

GW amplitude

$$h_0(t) = \frac{4\pi^2 G}{c^4} \frac{I\, f(t)^2}{d} \epsilon$$

Distance of the source

Frequency variation

$$\dot{f}(t) \sim -\epsilon^2 f(t)^5 \rightarrow f(t) = f_0 \left(1 + \frac{t\, \epsilon^2 f_0}{const}\right)^{-\frac{1}{4}}$$

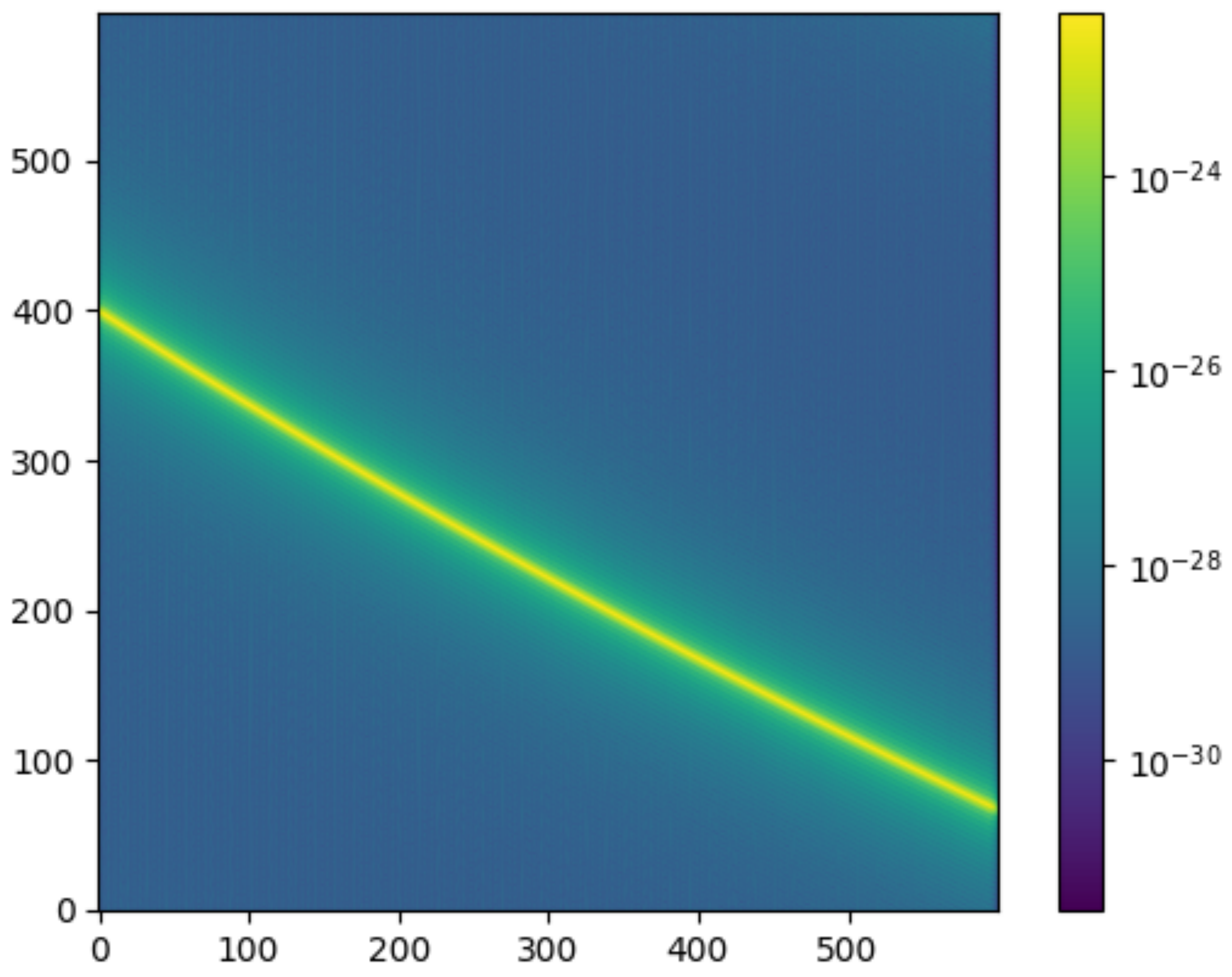
→ **Fixed initial amplitude :** 2×10^{-23}

→ **Fixed inclination angle:** $\iota \sim 56°$

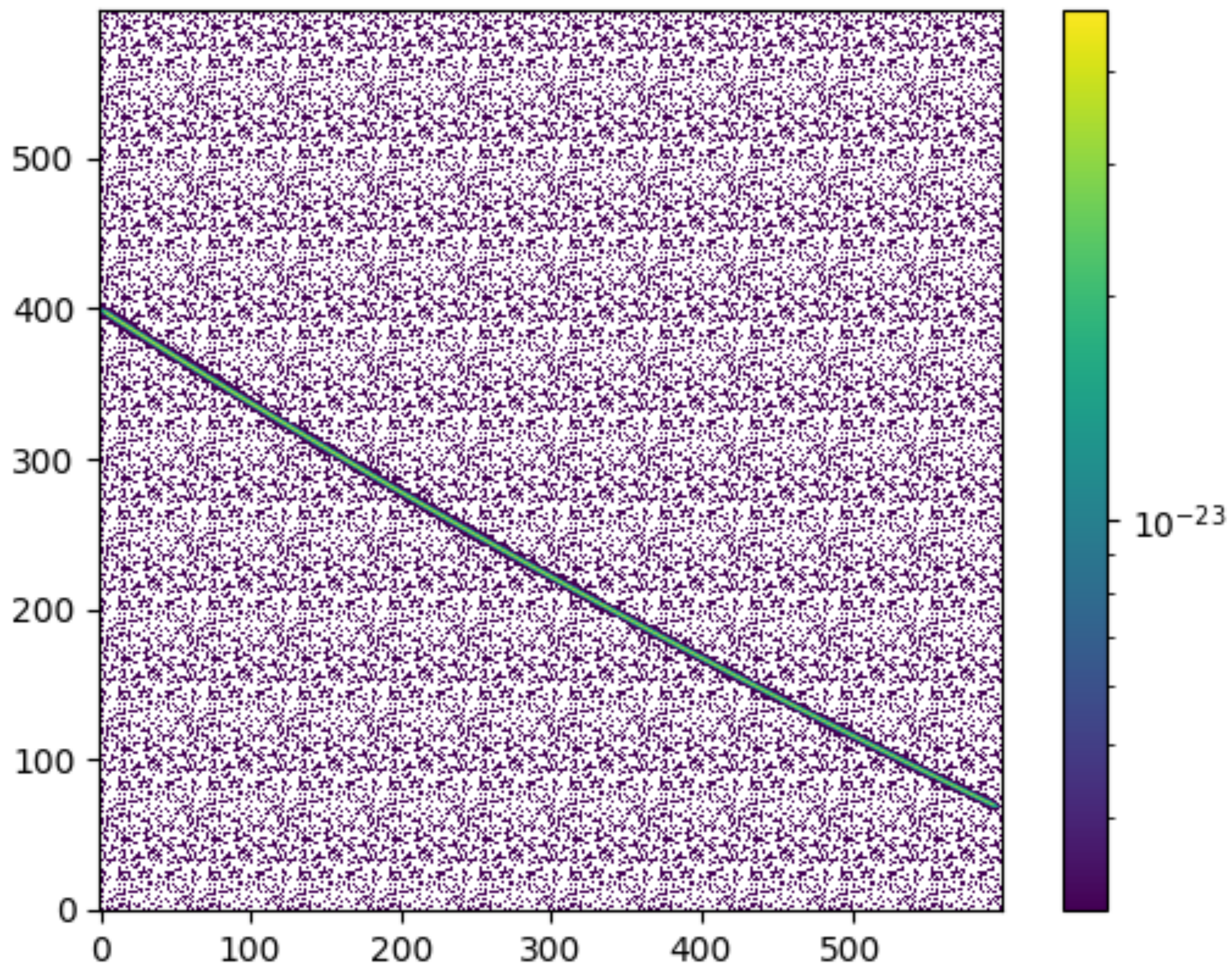
Parameters range: $\epsilon \in [3{,}30] \times 10^{-4}$ $\quad f_0 \in [1.25{,}2.00]$ kHz

We are not focusing on standard continuous waves

Artefacts



500
400
300
200
100
0
0
100
200
300
400
500
10^{-24}
10^{-26}
10^{-28}
10^{-30}



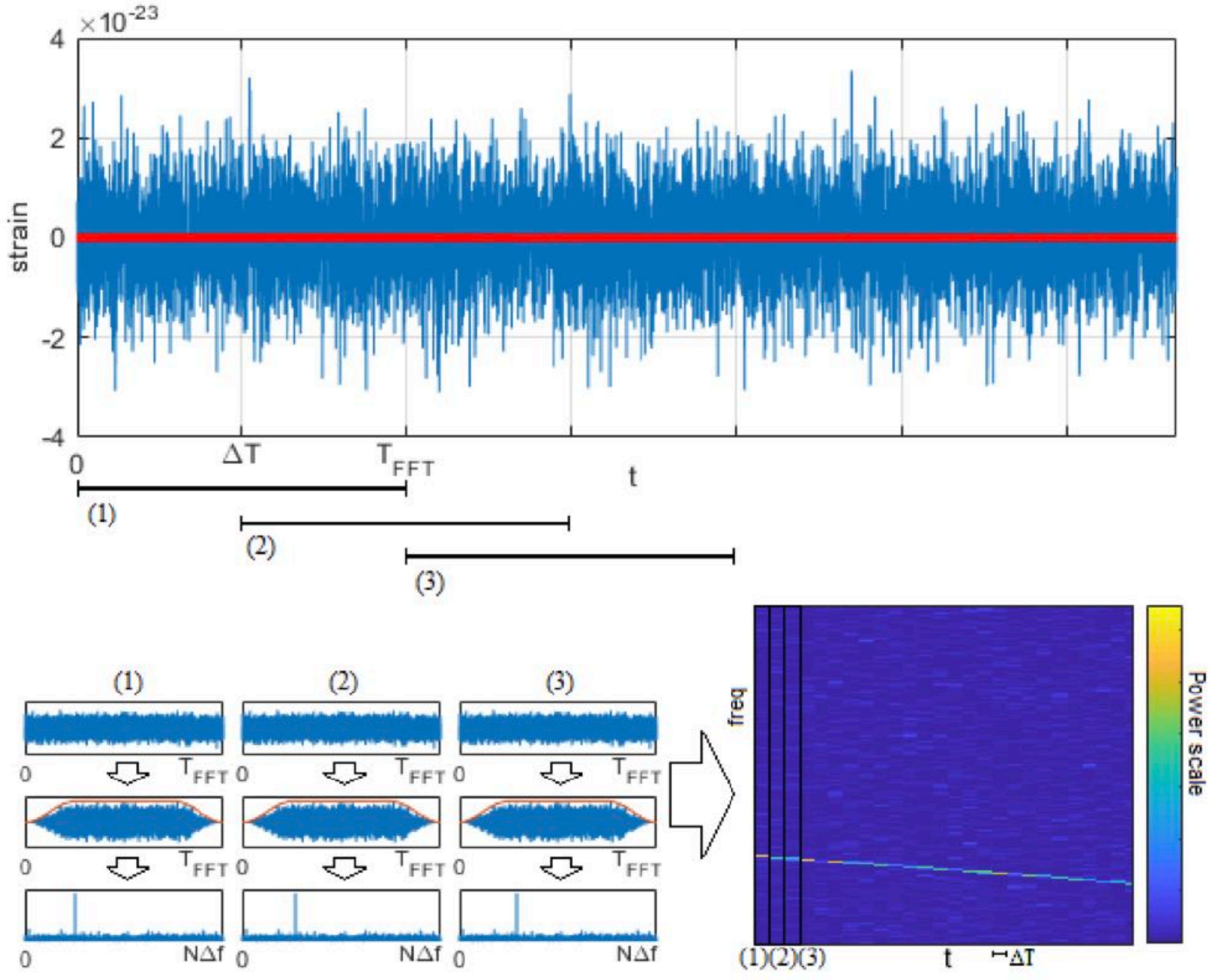
500
400
300
200
100
0
0
100
200
300
400
500
10^{-23}

Flat-top window

To avoid **spectral leakage** we used a flat-top window

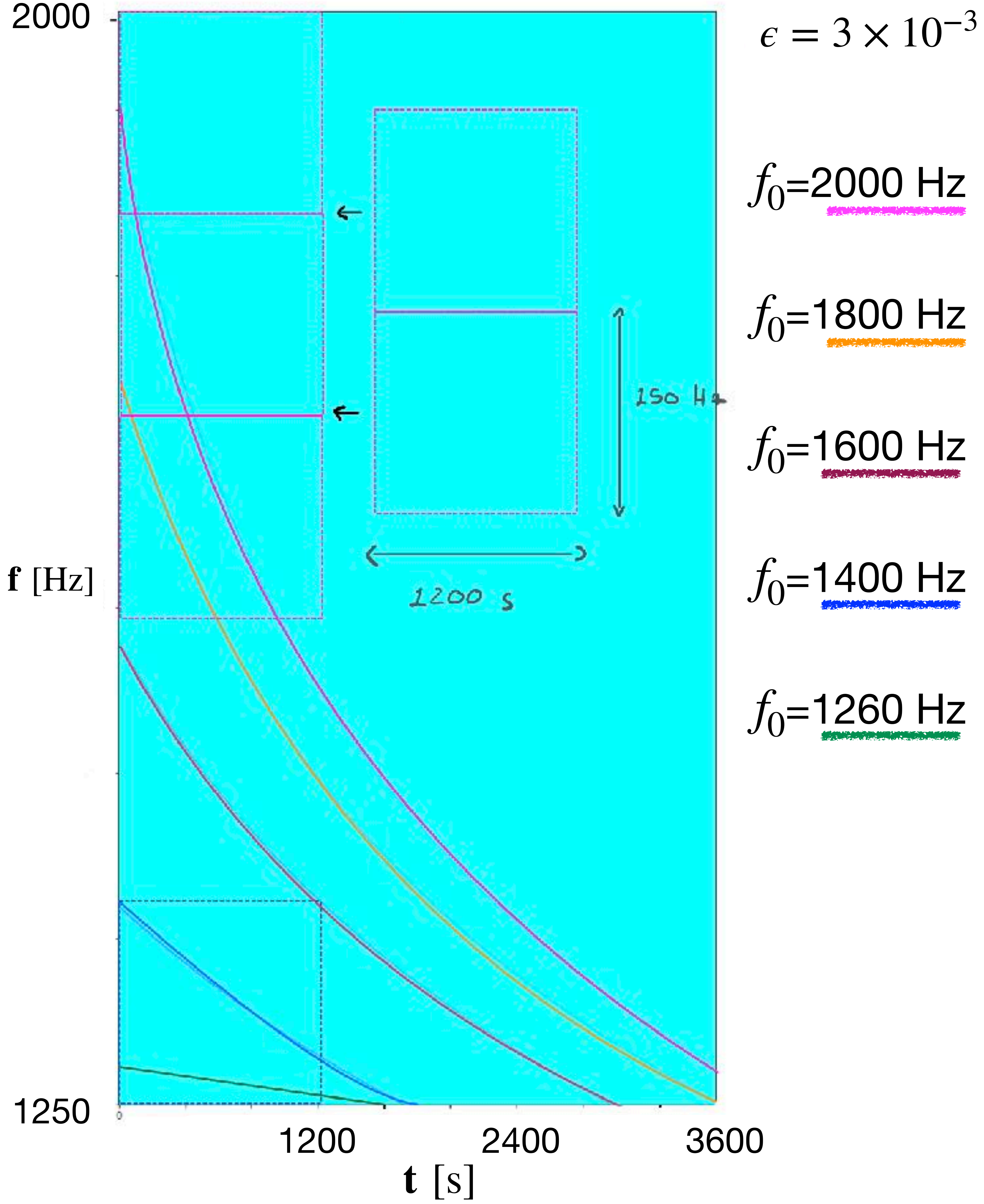
$$f_{\cos}(x_i) = \sum_{k=0}^{4} (-1)^k a_k \cos\left(\frac{2k\pi x_i}{N}\right)$$

x_i samples, N number of samples

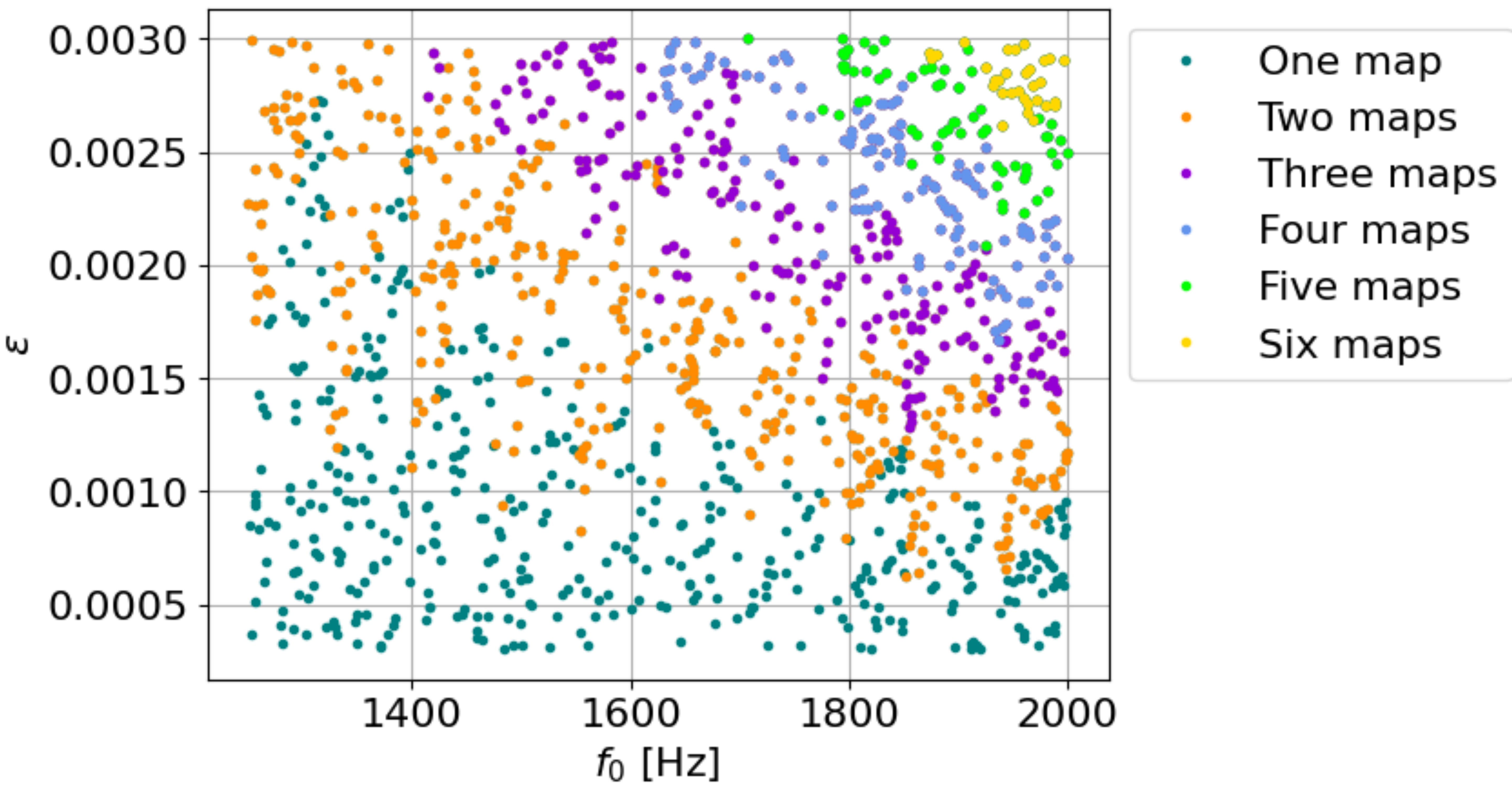


Courtesy of Lorenzo Pierini

Maps construction



Number of maps crossed by a signal

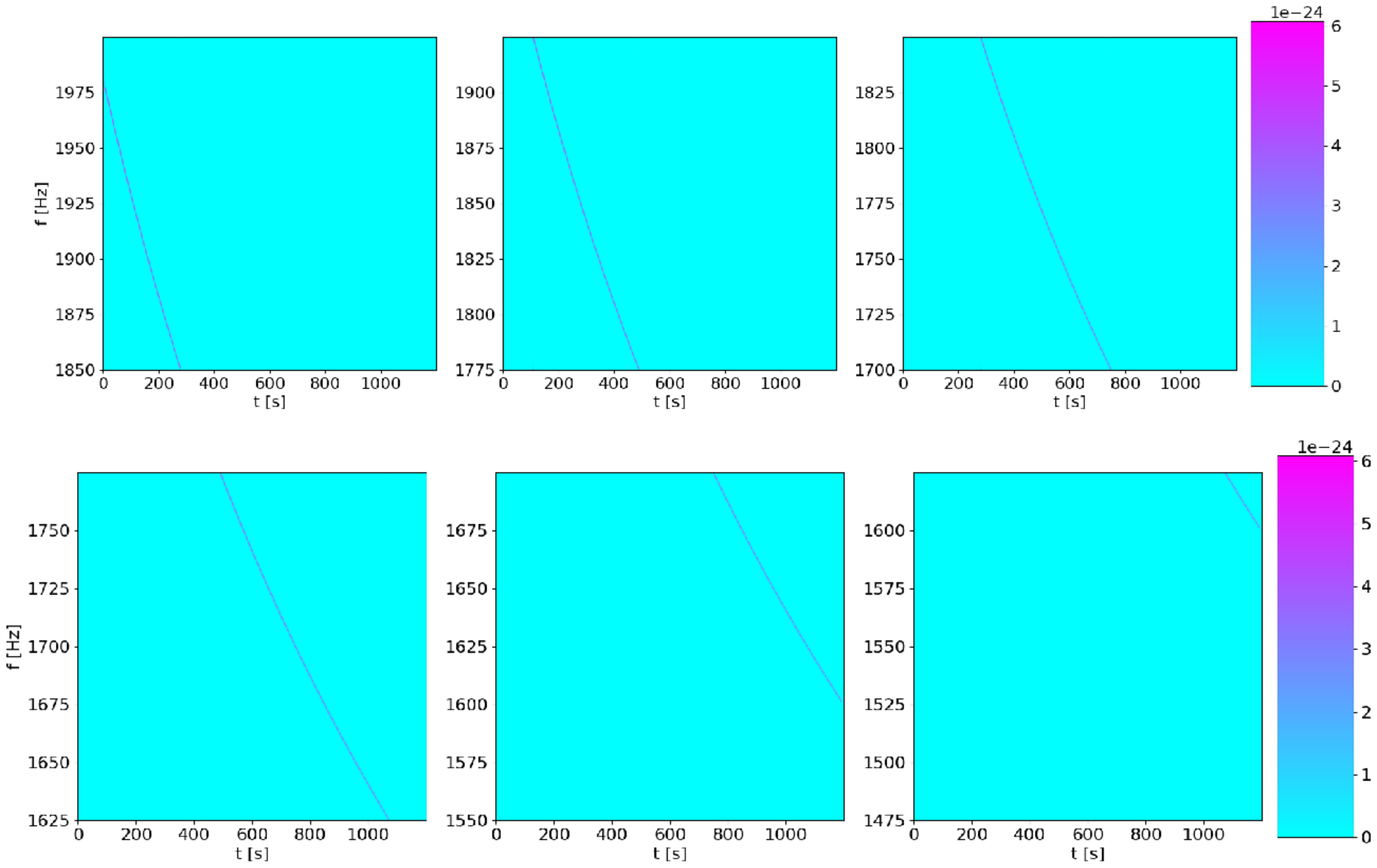


The number of maps per signal depends on the parameters

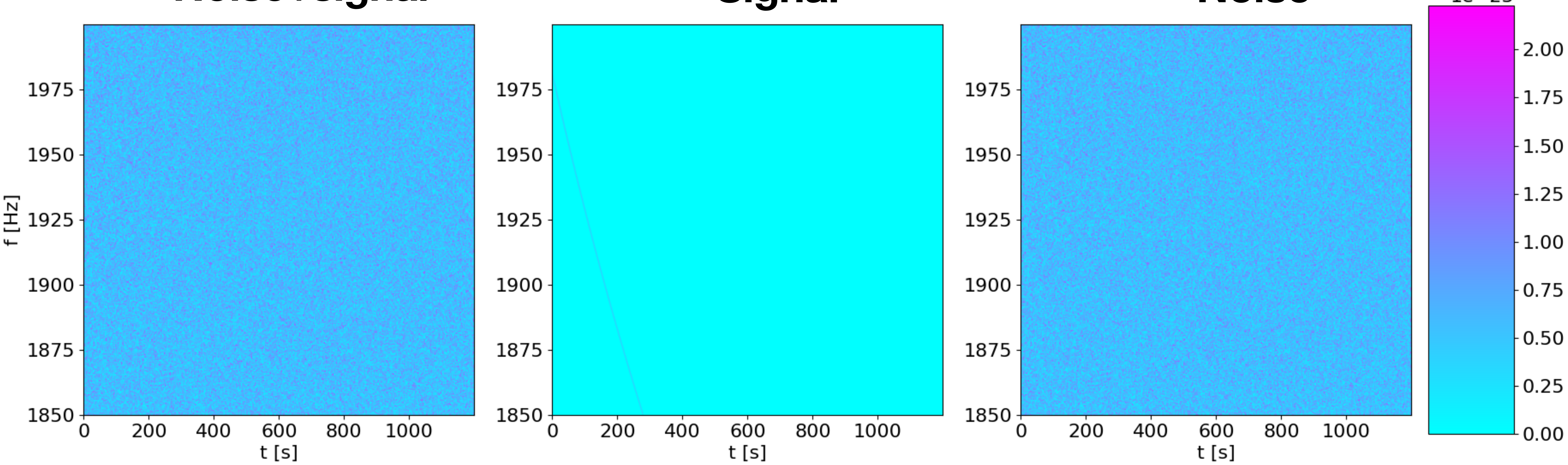
It is enough to tag right one map to have a trigger

Example of a signal that crosses multiple maps

$\epsilon = 0.003$
$f_0 = 1982$ Hz



Noise+signal	Signal	Noise



$\epsilon = 0.003$
$f_0 = 1982$ Hz

Useful tools

Overlap

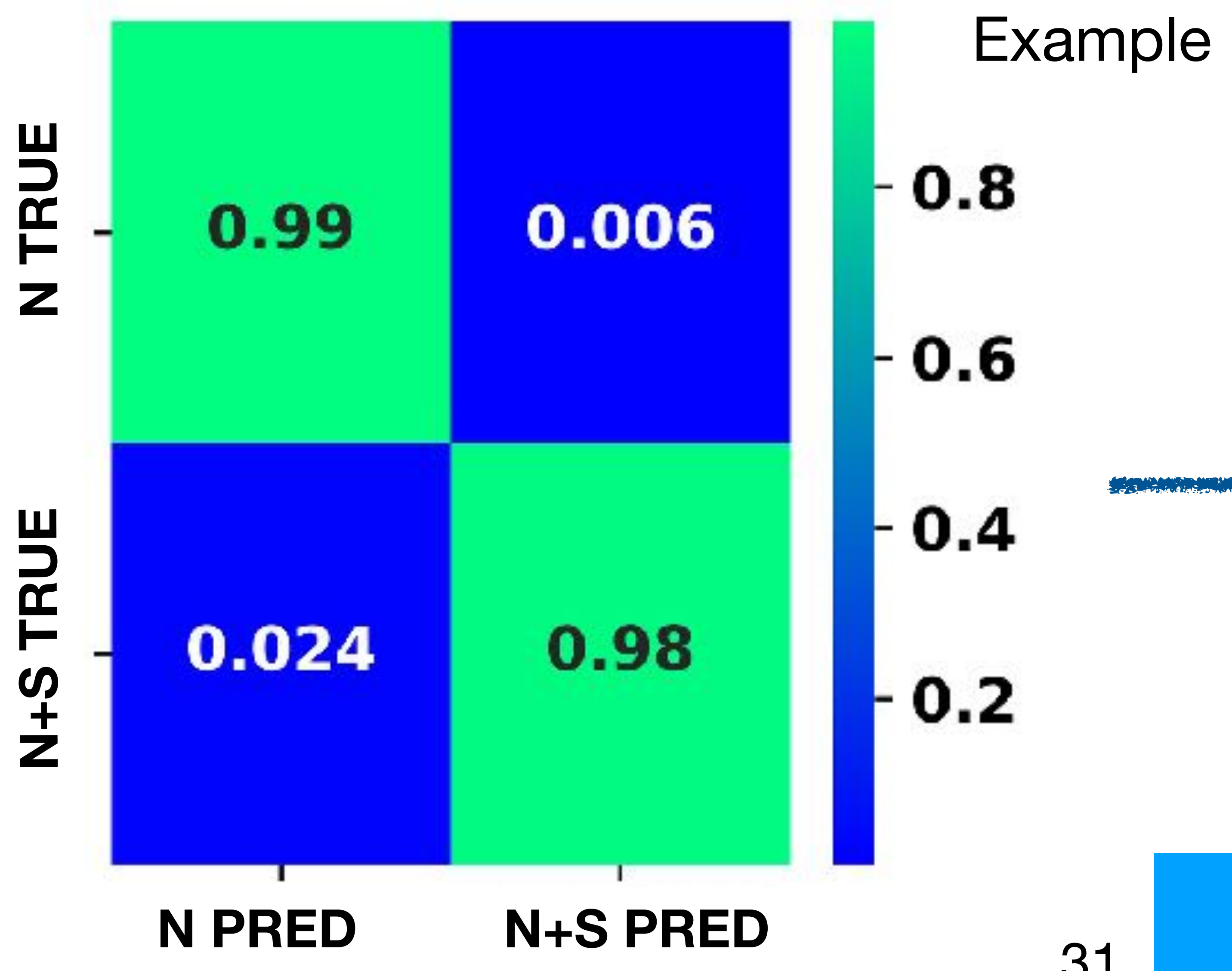
$$\mathcal{O} = \sqrt{\sum_{tf} S_{tf} h^d_{tf} \left(\sum_{tf} S_{tf} S_{tf} \right)^{-1}}$$

→ It calculates the percentage of the signal preserved by the denoiser

Confusion matrix

ROWS
True classes

COLUMNS
Predicted classes



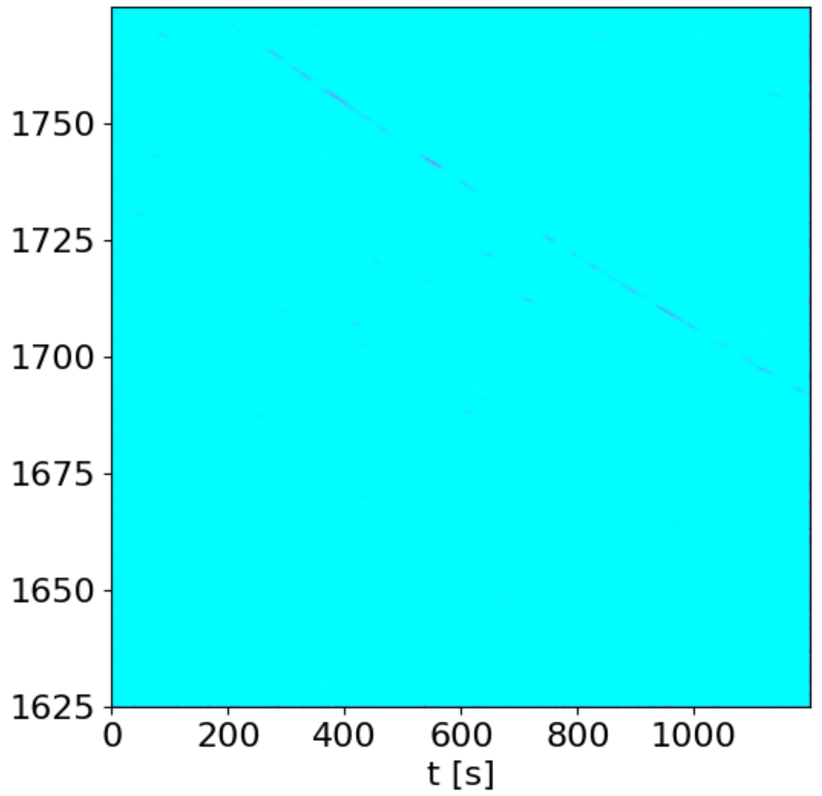
→ It gives a visual representation of the outcome of our classifier.

Masked loss

We tried to highlight the structure of the signal and help the model learn it.

1.0	0.0	0.0	0.0	0.0
0.0	1.0	0.0	0.0	0.0
0.0	0.0	1.0	0.0	0.0
0.0.	0.0	0.0	1.0	0.0
0.0	0.0.	0.0	0.0	1.0

X

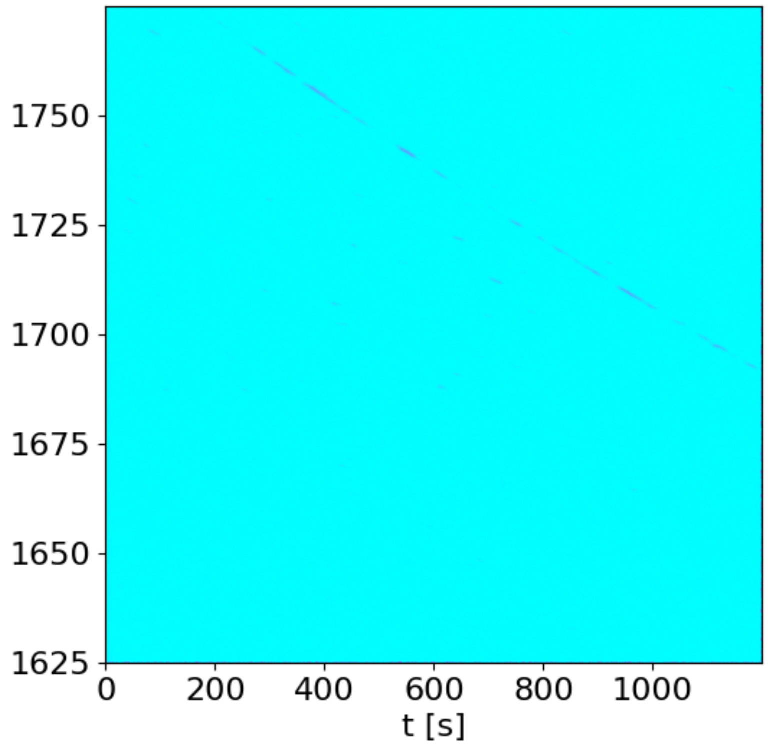


First epochs: each map was multiplied by a matrix that had 1 on the signal pixels and 0 otherwise.

We computed the loss function with that.

1.0	0.5	0.2	0.0	0.0
0.5	1.0	0.5	0.2	0.0
0.2	0.5	1.0	0.5	0.2
0.0.	0.2	0.5	1.0	0.5
0.0	0.0.	0.2	0.5	1.0

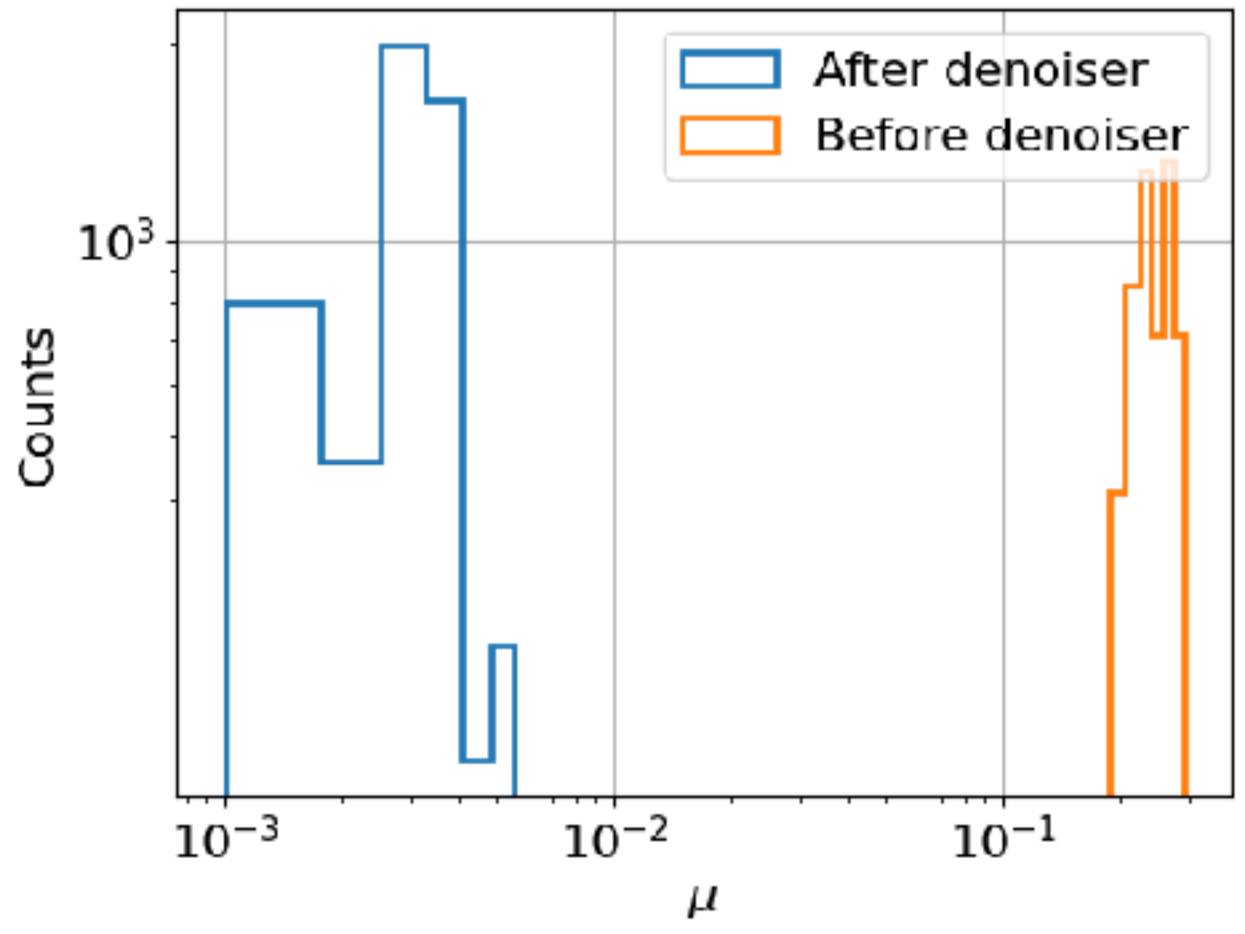
X



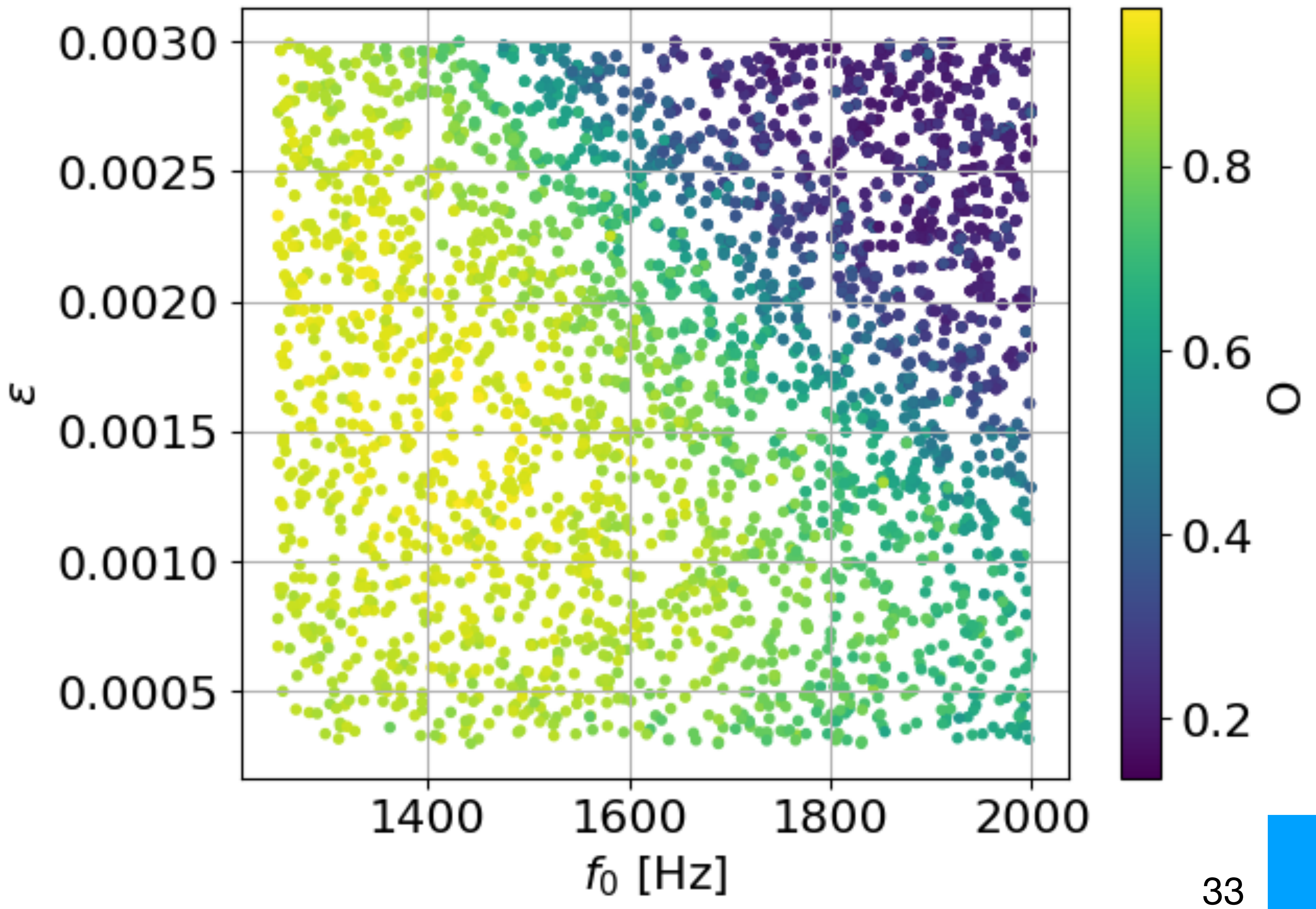
As the training proceeded, we enlarged the area of the map that was not set to zero.

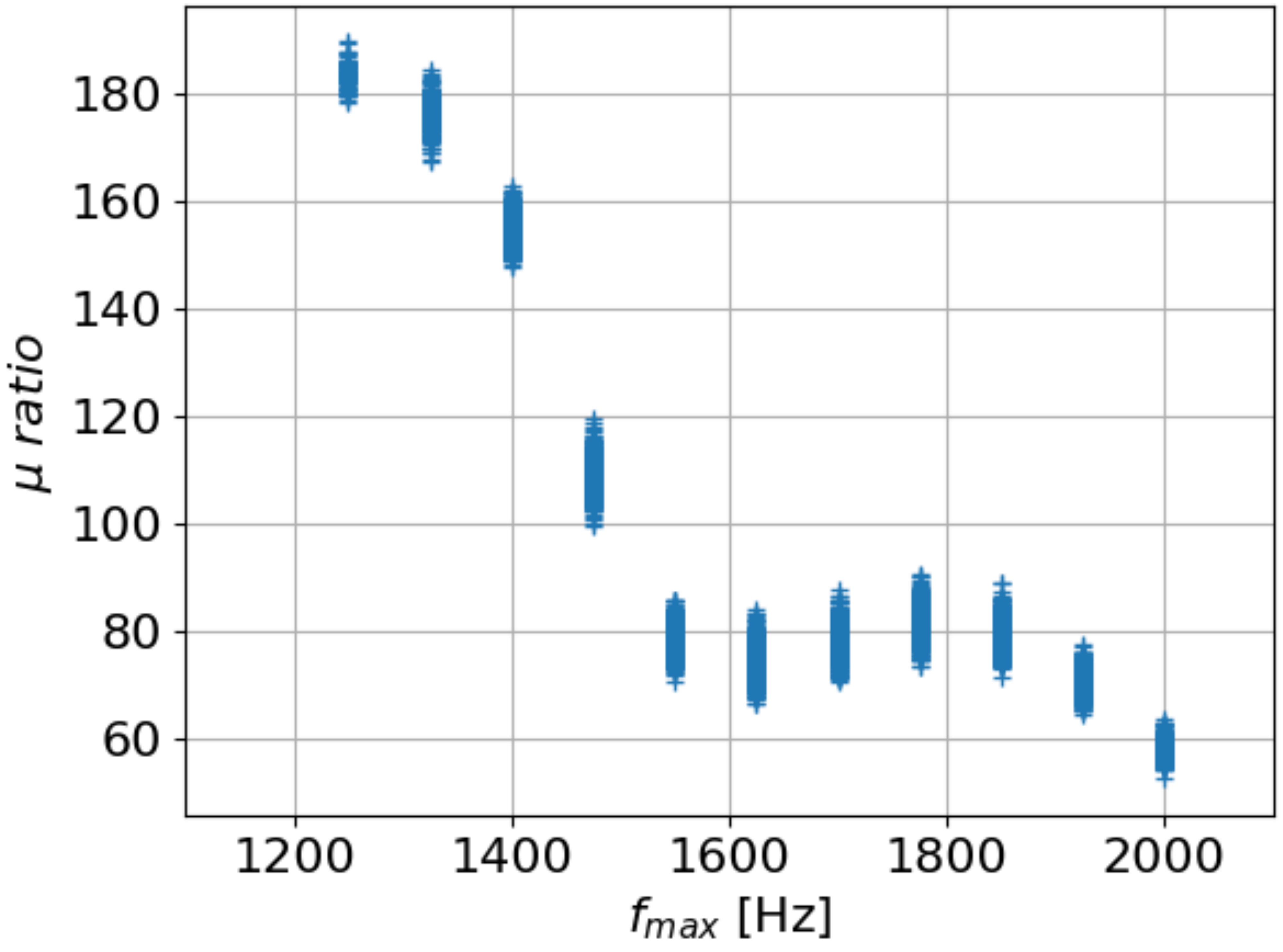
At the end, we computed the loss function with the entire map.

Denoiser



Removing noise



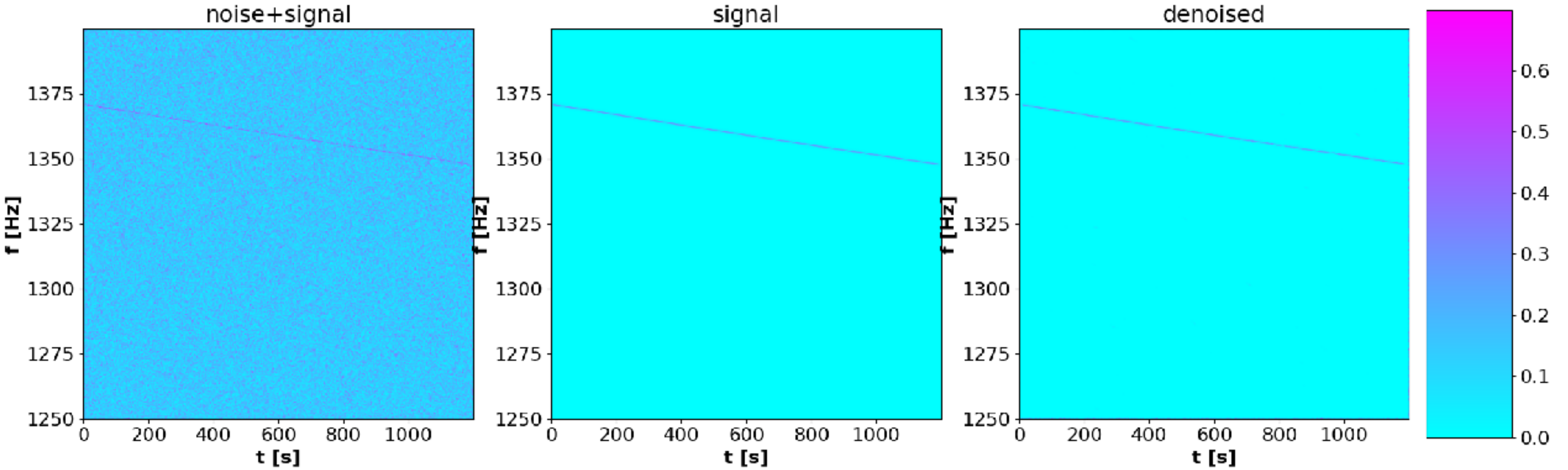


Preserving the signal

Overlap

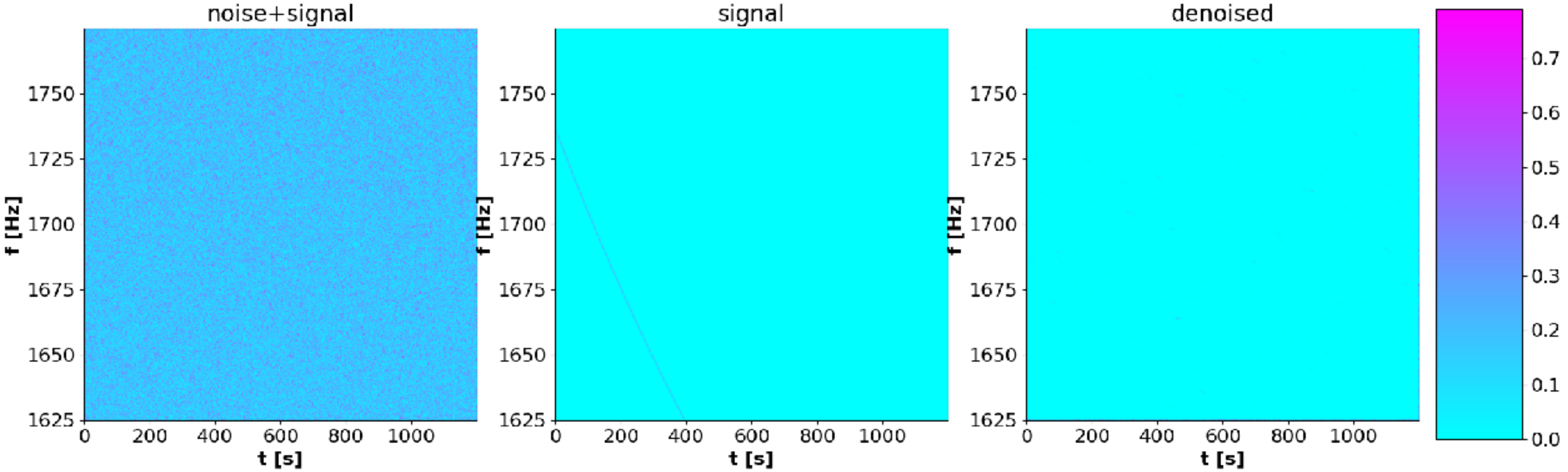
$$\epsilon = 1.3 \times 10^{-3}$$
$$f_0 = 1370 \text{ Hz}$$

$$\mathcal{O} = 0.96$$

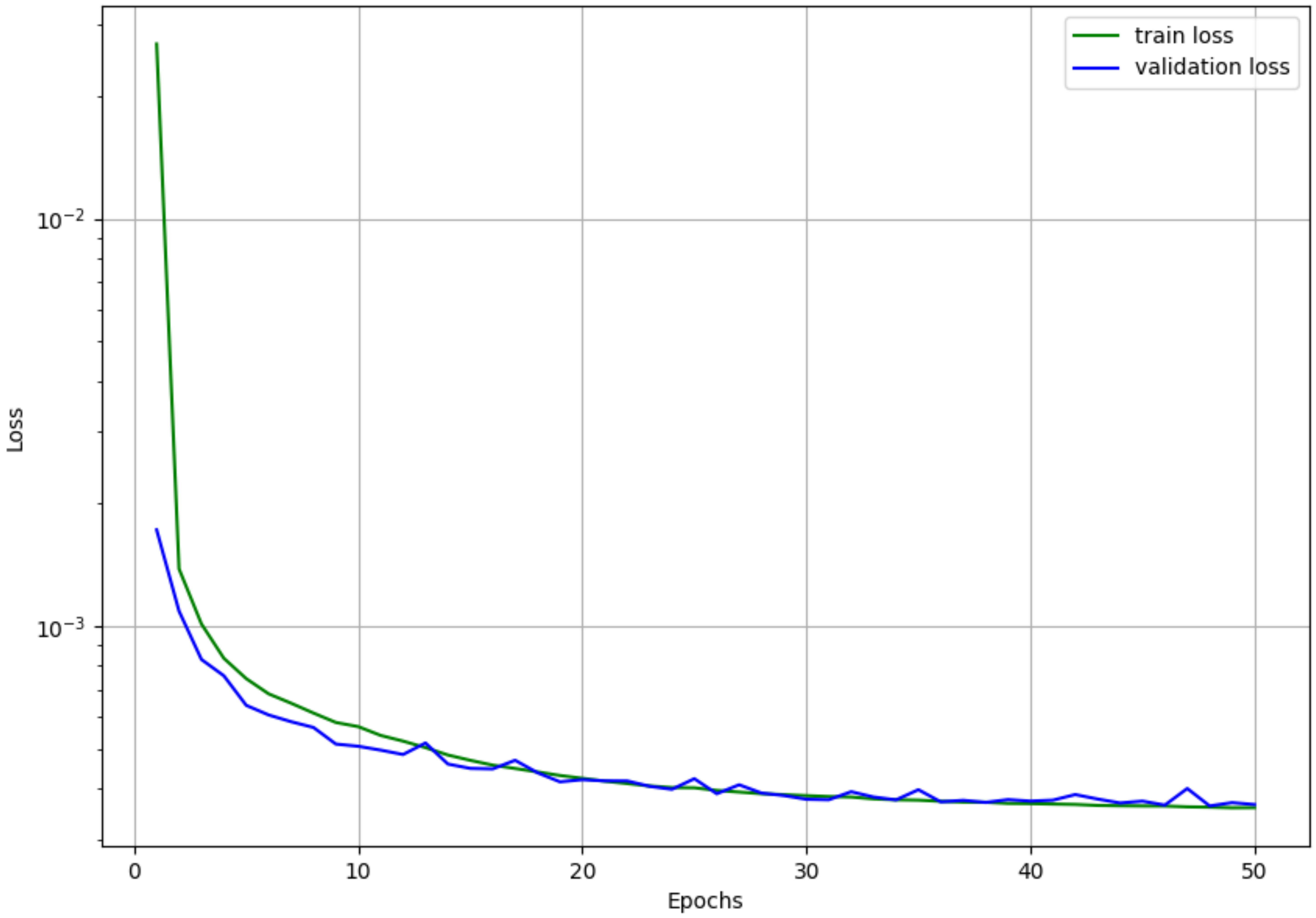


$$\epsilon = 3 \times 10^{-3}$$
$$f_0 = 1737 \text{ Hz}$$

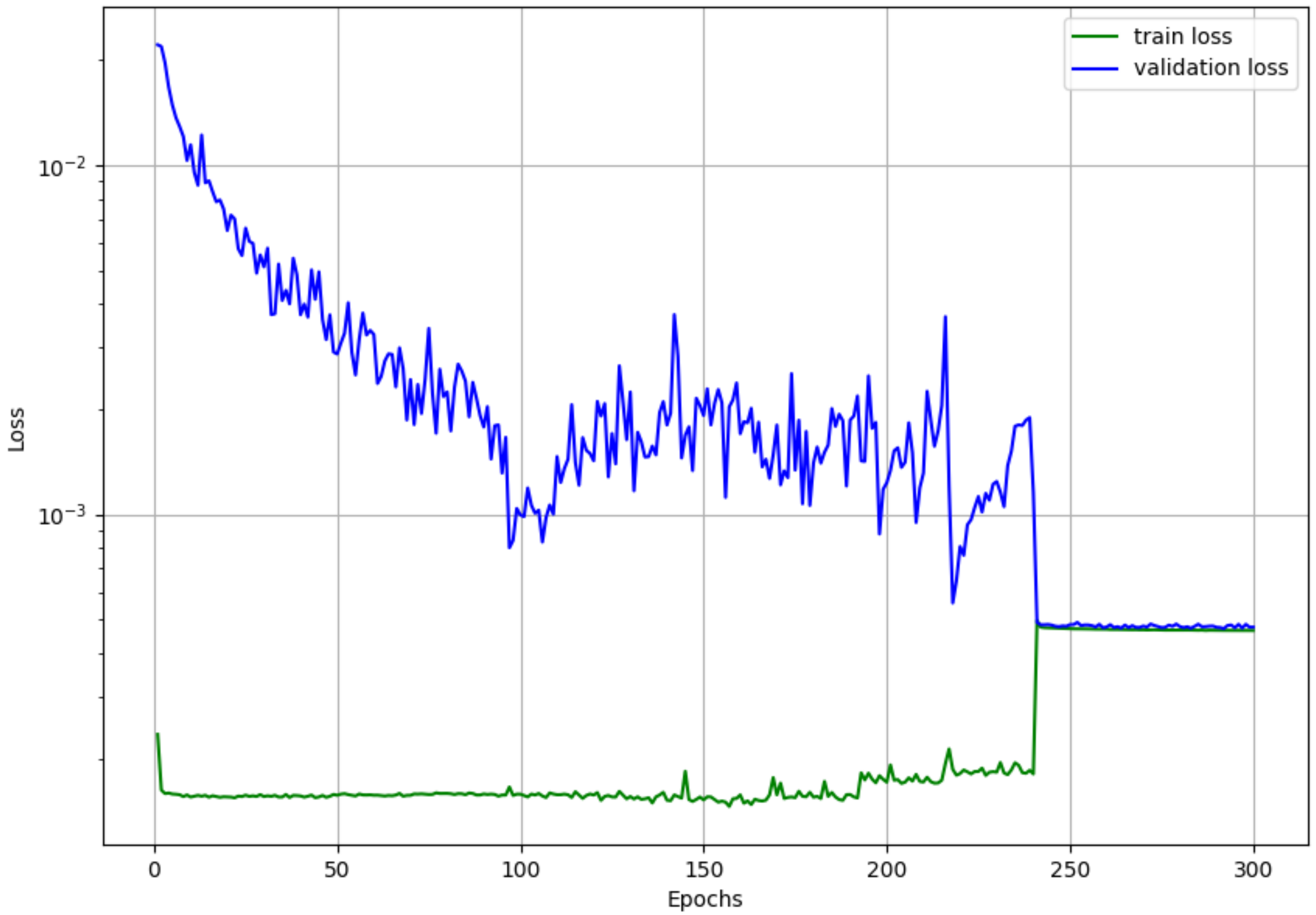
$$\mathcal{O} = 0.11$$



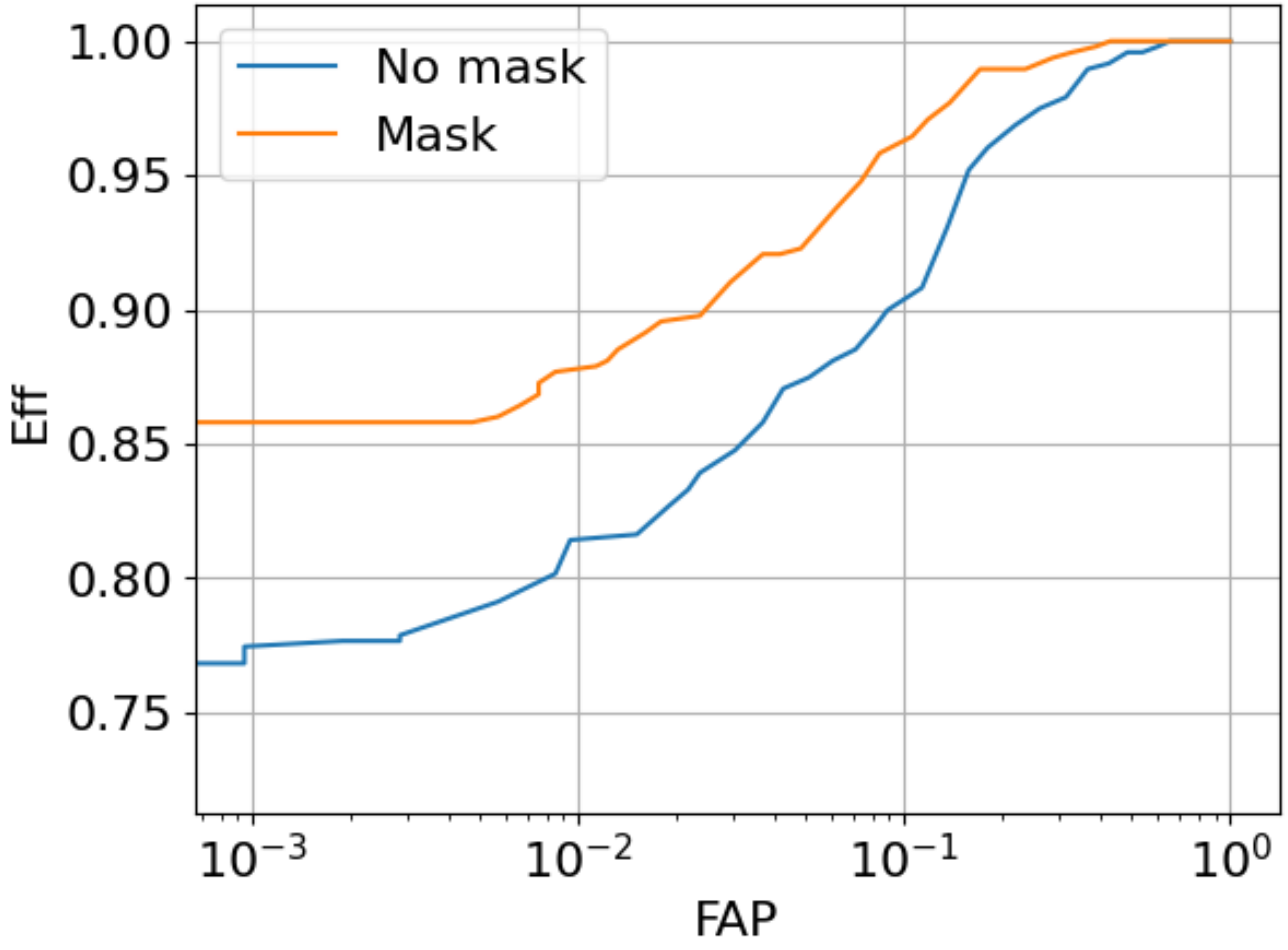
Loss function, denoiser no masked loss



Loss function, masked loss

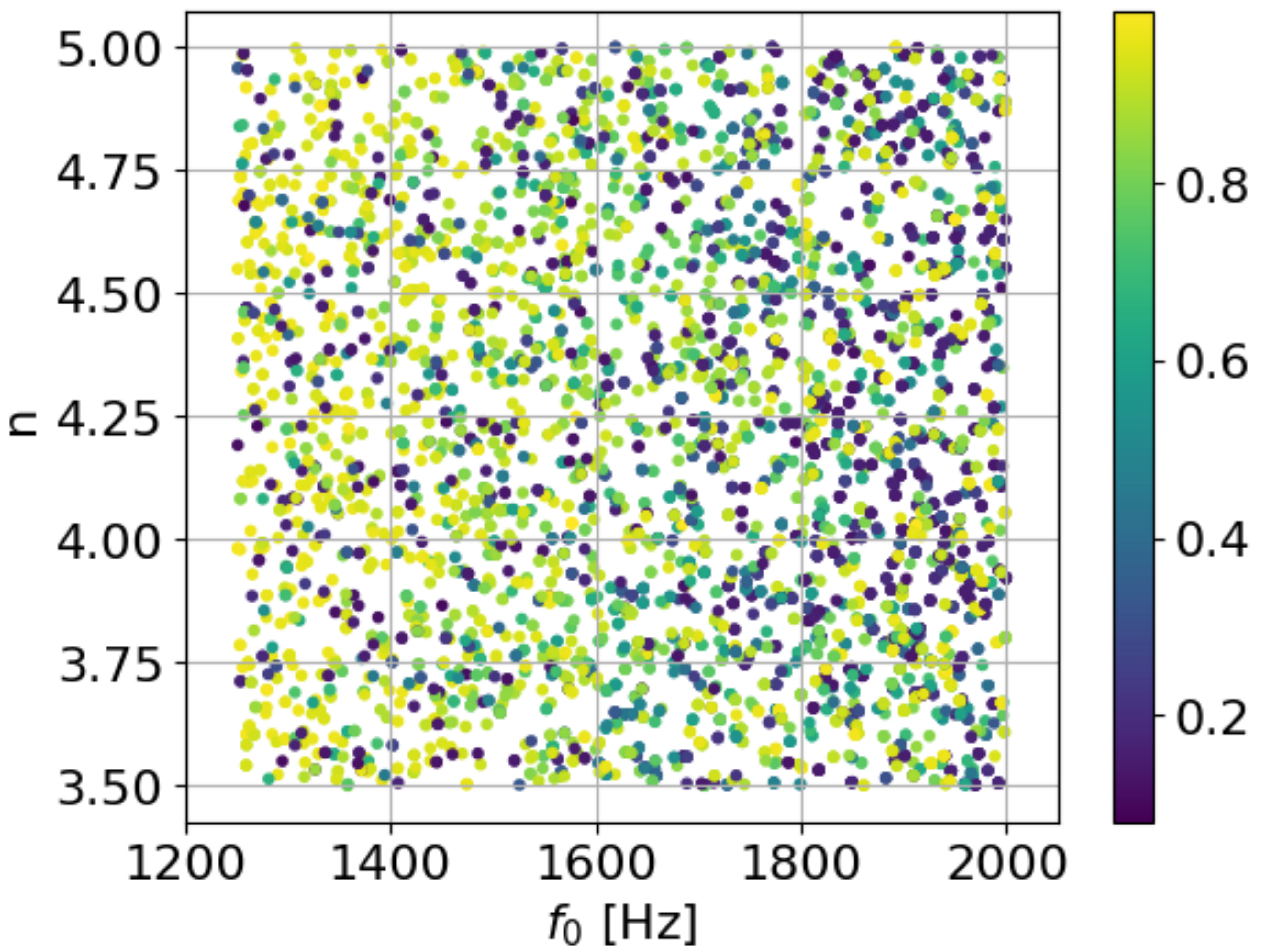


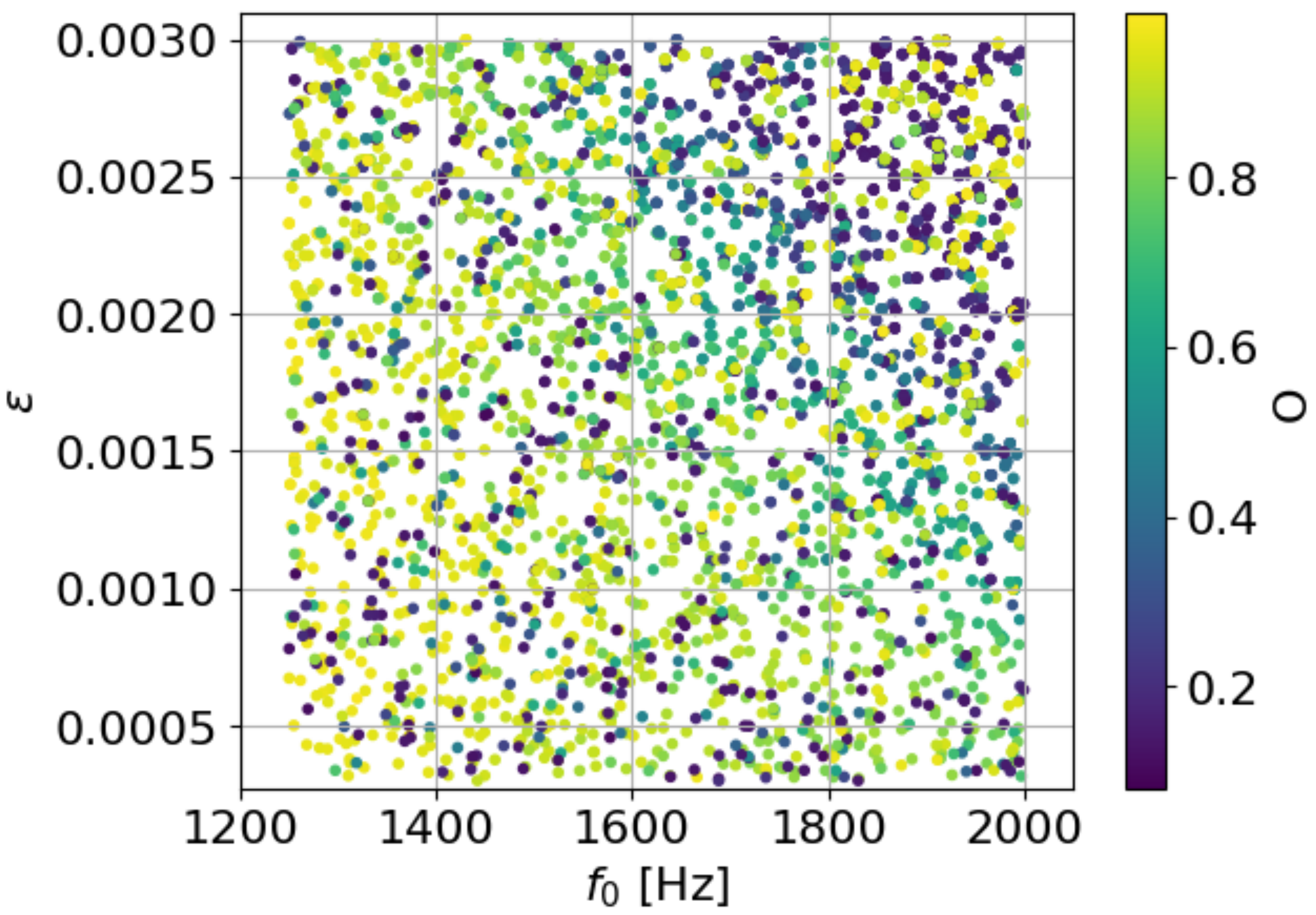
ROC curve



→ Improvement of the classification performances due to the masked loss

Changing braking index (Denoiser)





→ No trend with the braking index.

→ Improvement in the upper right corner.

Comparison with other methods for long-transient signals

Collaboration paper: *Search for Gravitational Waves from a Long-lived Remnant of the Binary Neutron Star Merger GW170817, Abbott et al. 2019*

Generalized FrequencyHough

O2

$\epsilon = 1.44 \times 10^{-3}$ $\quad f_0 = 1740\ \mathrm{kHz}$ $\quad \Delta t = 2\ \mathrm{s}$ $\quad I = 4.34 \times 10^{38} \mathrm{kg\ m}^2$ → $d_{FrH} = 0.242\ \mathrm{Mpc}$

This method

Computational cost: 1 GPU for ~4 hours, smaller than GFh

O4

$\epsilon = 1.77 \times 10^{-3}$ $\quad f_0 = 1753\ \mathrm{kHz}$ $\quad \Delta t = 2\ \mathrm{s}$ $\quad I = 1.4 \times 10^{38} \mathrm{kg\ m}^2$ → $d = 0.402\ \mathrm{Mpc}$

→ Detector sensitivity improved by a factor of 3 in the [1700,1800] frequency band

We gained a factor of ~2 in distance