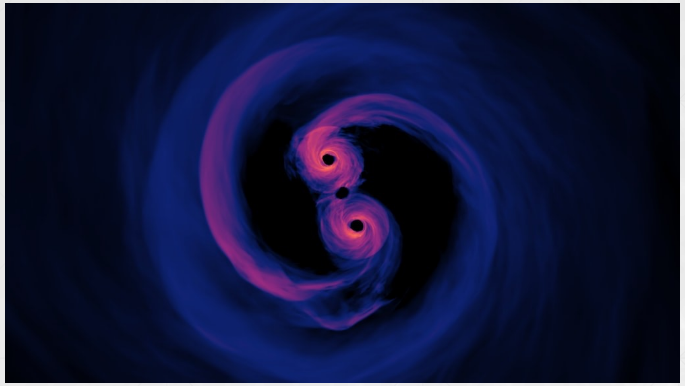
IPAM Workshop IV: Big Data in Multi-Messenger Astrophysics
Nov 29 - Dec 3, 2021

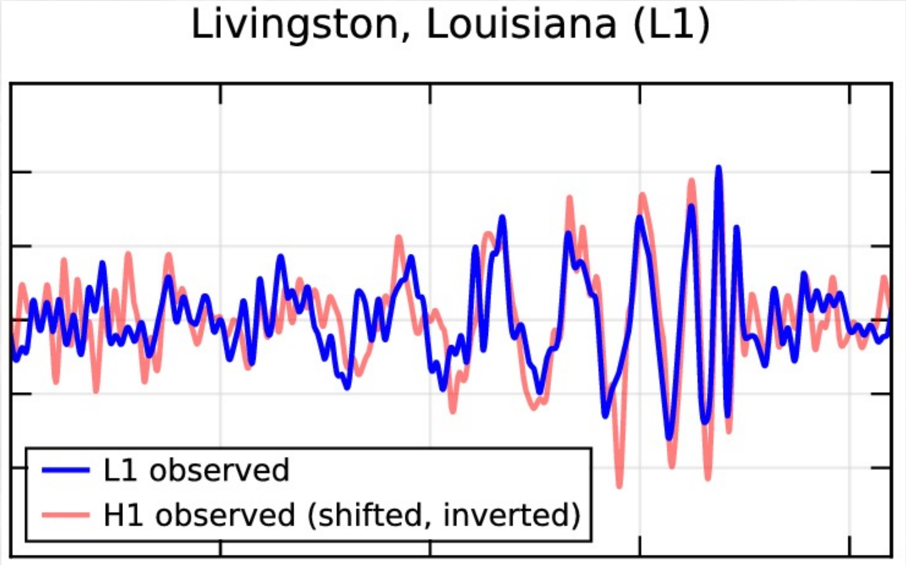
Variational models and algorithms for GW denoising and reconstruction: applications

Alejandro Torres Forné

Miquel Llorens, Ainara Sanz, Pablo Barneo, J.A.Font, Antonio Marquina, José María Ibañez and Elena Cuoco

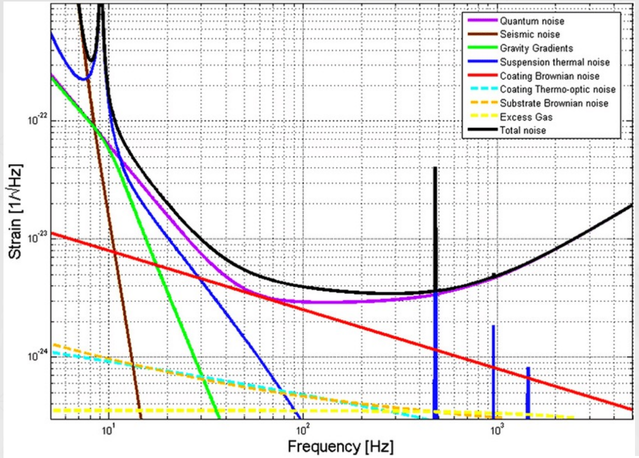
GW signal detection

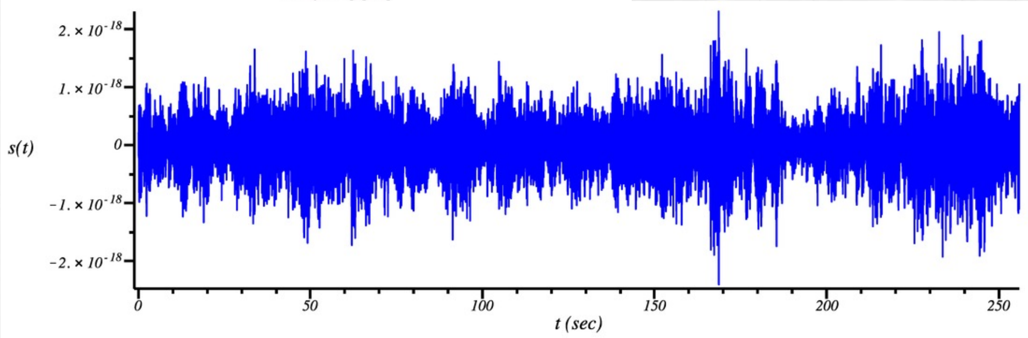


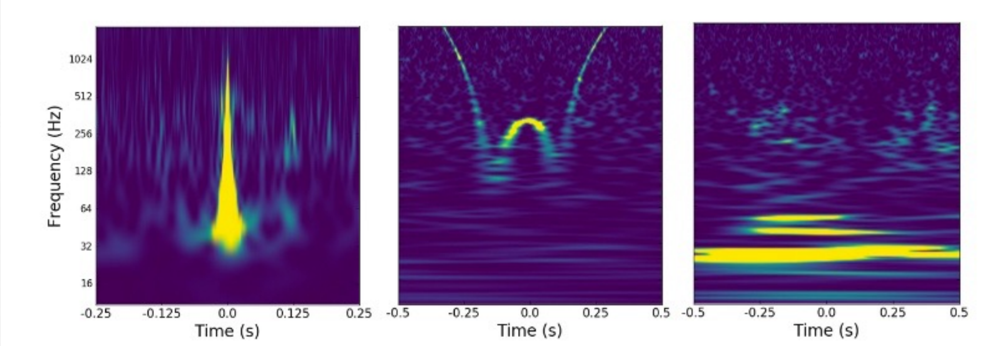


We did it!
Let’s do a press conference!

GW signal detection







Abbott et al. Classical and Quantum Gravity, Volume 37, Issue 5, id.055002 (2020)

- Online process (seconds to minutes)
  - Produce triggers.
  - Characterize detector.
  - Apply vetoes.
  - Online detection pipelines:
    - CWB
    - LAL
    - and many more….
  - Send public alerts
- Offline process (days to months)
  - Detector calibration
  - Detect and mitigate glitches
  - Detailed parameter estimation
  - Astrophysical implications

GW data analysis steps

Proof of concept

- Test with non-white Gaussian noise.
- Proof that the algorithm works with GW signals.

Application to real data

- Test with non-white non-Gaussian noise.
- Proof that the algorithm works with GW real data.

Test with standard pipelines

- Use in combination with standard pipelines.
- Proof that the algorithm improves the results.
- Proof that include new features

Mathematical Framework

Signal denoising approach

$$y = x + n$$

x: signal
n: noise

Total Variation

$$||y - u||^2_{\mathrm{L}_2} = \sigma^2$$

Smooth solution
Preserve edges

Sparse representation

$$x = D\alpha$$

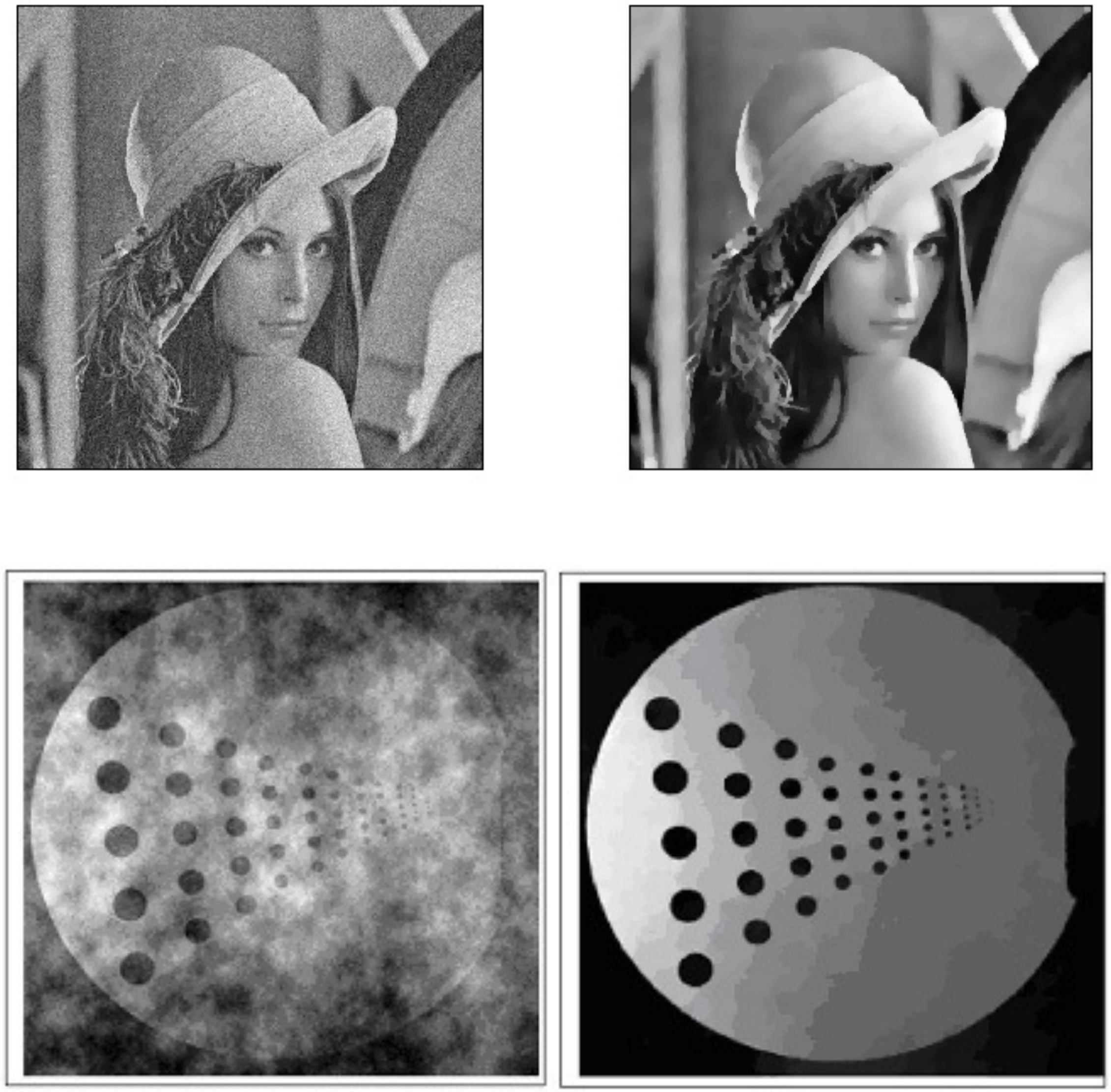
Sparse solution
Redundant dictionary

$$u = \underset{u}{\operatorname{argmin}} \left\{ \mathcal{R}(u) + \frac{\lambda}{2} \mathcal{F}(u) \right\}$$

$\mathcal{F}$: Fidelity term. Measures the similarity of the solution to the data.

$\mathcal{R}$: Regularization term. The constrain we want to impose to the data.

λ : Regularization parameter. Controls the relative weight of both sides.



Credit: IPAM/UCLA

Introduction to Total Variation Methods

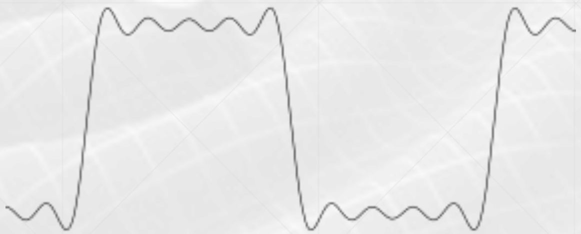
Introduction to TV methods

$$||y-u||_{\mathrm{L}_2}^2 = \sigma^2$$

Solution: Find a function *u* whose L_2-norm distance to *y* is noise standard deviation.

Least Squares
Fourier Transform

Gibbs Phenomena
Indeterminate solution



Issues overcome using an auxiliary energy prior to regularize the least-squares problem, solving a constrained variational problem:

$$\min_u \mathcal{R}(u) \quad \text{subject to} \quad ||y-u||_{\mathrm{L}_2}^2 = \sigma^2$$

Wiener filter

- Variational problem can be formulated as an unconstrained problem (Tikhonov regularization):

$$u = \operatorname*{argmin}_u \left\{ \mathcal{R}(u) + \frac{\lambda}{2}\mathcal{F}(u) \right\}$$

- Wiener filter

$$R(u) = \int |\nabla u|_2^2$$

- Good characteristics:
  - Elliptic PDE.
  - Easy to solve due to differentiability and strict convexity.
- Issues in the presence of noise:
  - Amplification of high frequencies
  - The recovered smooth solution shows spurious oscillations near steep gradients or edges.

Rudin-Osher-Fatemi model

$$u = \underset{u}{\operatorname{argmin}} \left\{ \mathrm{TV}(u) + \tfrac{\mu}{2} ||u - f||^2_{\mathrm{L}_2} \right\} \qquad \mathrm{TV}(u) = \int |\nabla u|^2_1$$

- Produces Singular distributions.
- Preserves edges and avoids spurious oscillations (ringing).
- Promotes zeros for small gradients.
- Convex problem.
- Fine scales are destroyed by the effect of TV norm.
- A good estimation of λ results in an ill-conditioned Euler – Lagrange equation.

$$\boxed{\nabla \cdot \frac{\nabla u}{|\nabla u|} + \lambda(f - u) = 0}$$

$$\mathrm{TV}_\beta(u) := \int \sqrt{|\nabla u|^2 + \beta}$$

- Associated Euler-Lagrange equation is elliptic and non-degenerate.
- Approximate solution can be obtained by e.g. a nonlinear Gauss-Seidel iterative procedure.

Split-Bregman method

$$\min_{u,d} |d| + H(u) + \frac{\lambda}{2}||d - (u)||_2^2$$

Goldstein T. & Osher S.
SIAM J. Imaging Sci., 2, 209

- Efficient implementation of L_1 regularized problems.
- Decouples into L_1 and L_2 portions.

$$u^{k+1} = \min_u \lambda||d_x^k - \nabla_x u - b_x^k||$$

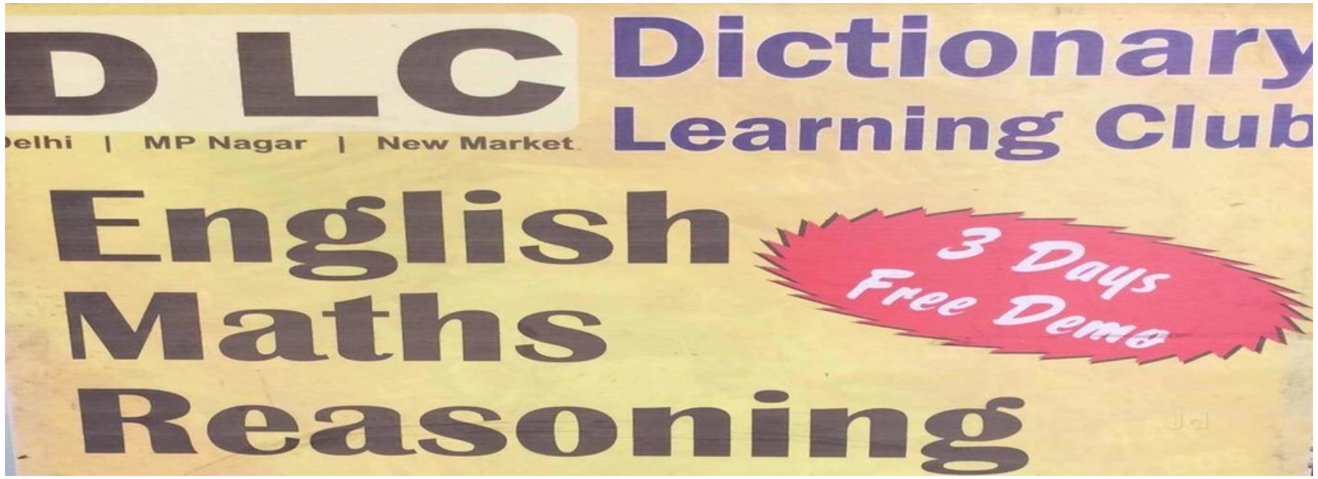
Direct Computation (Gauss-Seidel)

$$d^{k+1} = \min_d |d| + \lambda||d - u^{k+1} - b^k||_2^2$$

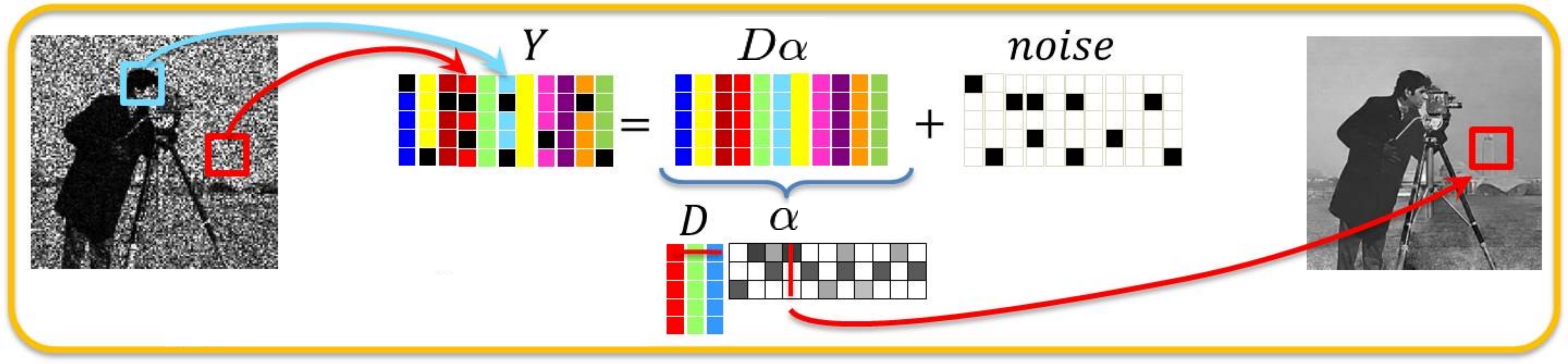
$$d_j^{k+1} = shrink(u_j + b_j^k, 1/\lambda)$$

$$shrink(x, \gamma) = \frac{x}{|x|} * max(|x| - \gamma, 0) \ ,$$

Introduction to Dictionary Learning



Sparse representation of signals



CSIRO Research

$$x = D\alpha$$

Add the constraint to assure sparsity:

$$\alpha = \underset{\alpha}{\operatorname{argmin}} ||D\alpha - y||_2^2 + \lambda||\alpha||_0$$

Not convex

NP Hard

The LASSO

- The problem can be transformed into a convex variational formulation.

$$\alpha = \underset{\alpha}{\operatorname{argmin}} ||D\alpha - y||_2^2 + \lambda ||\alpha||_1$$

Tibsibirani R. JSOR, B, Vol 58. 1996

- We have used an efficient implementation called SB-Lasso.

$$\alpha = \underset{\alpha, d}{\operatorname{argmin}} ||D\alpha - y||_2^2 + \lambda ||d||_1 + \mu ||d - u - b||_2^2$$

- Based in Split-Bregman algorithm.
- Decouples the problem into L1 and L2 terms.
- Solve them alternatively until convergence is reached.

Goldstein T. & Osher S.
SIAM J. Imaging Sci., 2, 209

Dictionary Learning problem

- We want to design the dictionary to fit a given set of signals GW templates

$S \in \mathbb{R}^N$
$x \in \mathbb{R}^n$
$n << N$

- Given a set of signal patches: $X = \{x_j\}_{j=1}^{M}$

$$\alpha = \underset{\alpha,\mathbf{D}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{m} ||\mathbf{D}\alpha_i - \mathbf{x}_i||_2^2 + \lambda ||\alpha_i||_1,$$

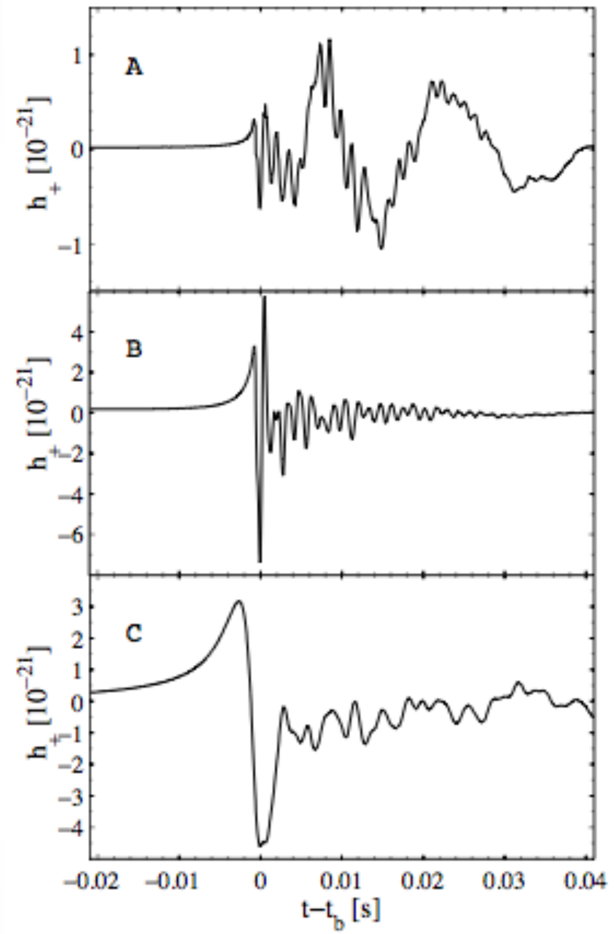
- We use block-coordinate descend method.
  - Solve D and each α_j iteratively.

Aplication to GW signals

Gravitational-wave signal catalogs

Burst from core-collapse supernova

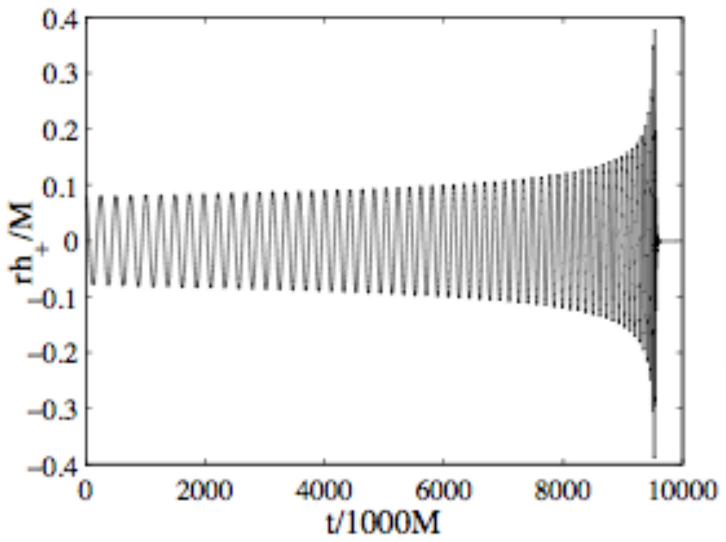
Credit: H. Dimmelmeier at al.



- 128 Waveforms.
- Short duration ~10 ms.

Binary Black Hole merger

Credit: A.H. Mroué at al.



- 174 Waveforms.
- Long duration ~ 1s.

Search Optimal Regularization Parameter

- Detector noise is not stationary, and its standard deviation is unknown.
- We must determine the **optimal** value of the regularization parameter that produces the best results. Heuristic search when the standard deviation of the noise is not known.
- Find an optimal value for λ is required.

$$\boxed{||u - f||_{L_2}^2 = \sigma^2}$$

- We use the Mean Square Error (MSE) and Structural similarity index (SSIM) to validate the results.

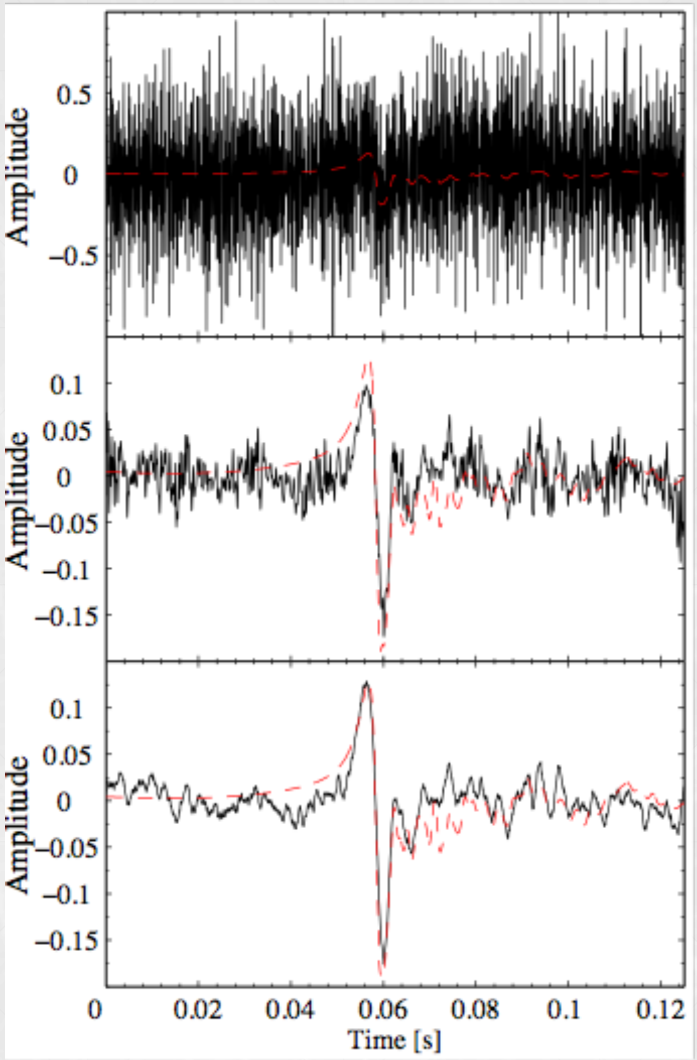
$$\boxed{\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2}$$

$$\boxed{SSIM(x, y) = \frac{2\mu_x\mu_y + e_1}{\mu_x^2\mu_y^2 + e_1}\frac{2\sigma_{xy} + e_2}{\sigma_x\sigma_y + e_2}}$$

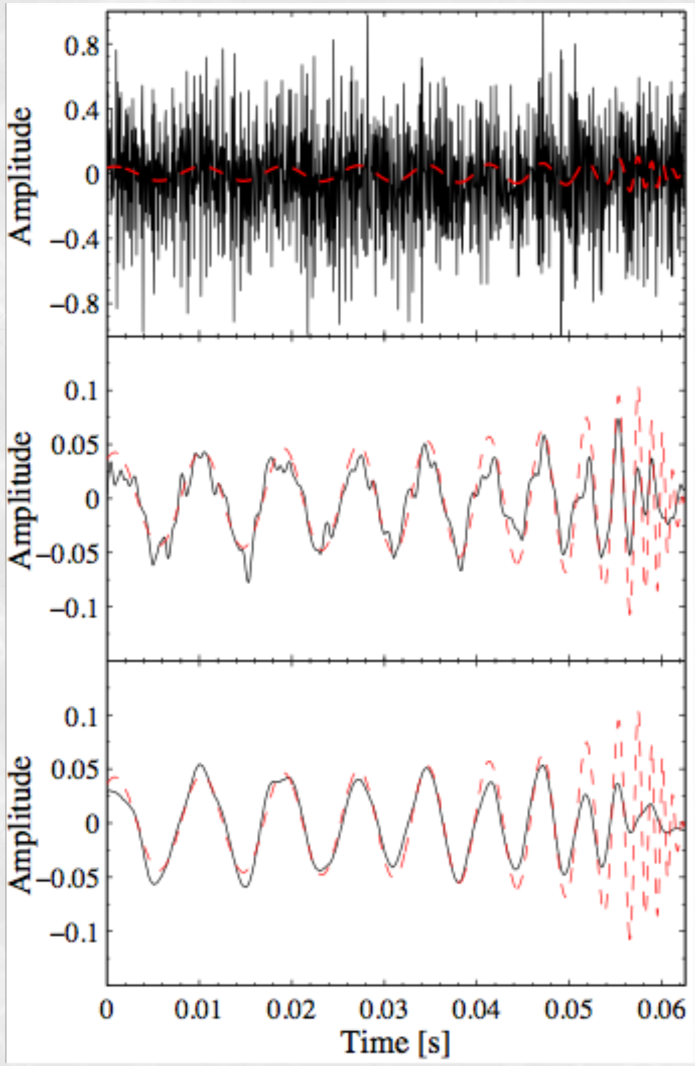
TV methods

Gaussian noise

CCSN



BBH



SNR = 20

Original

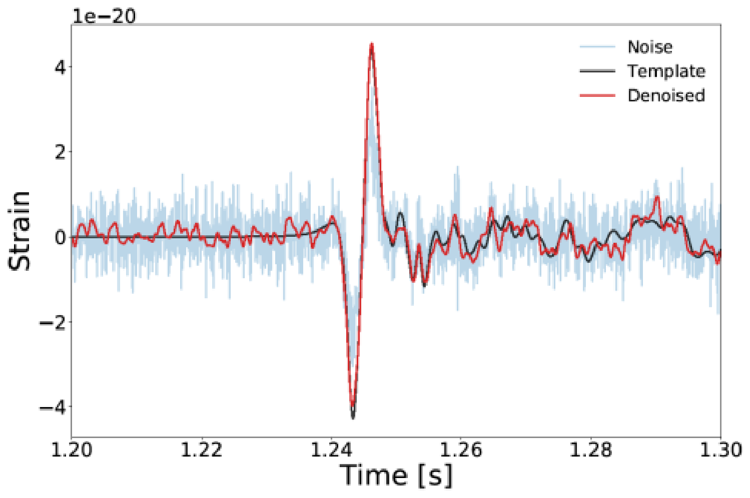
SB

rROF

TF, A. Marquina, J.A. Font and J.M Ibañez, 2014

TV methods

CCSN - AdLIGO data



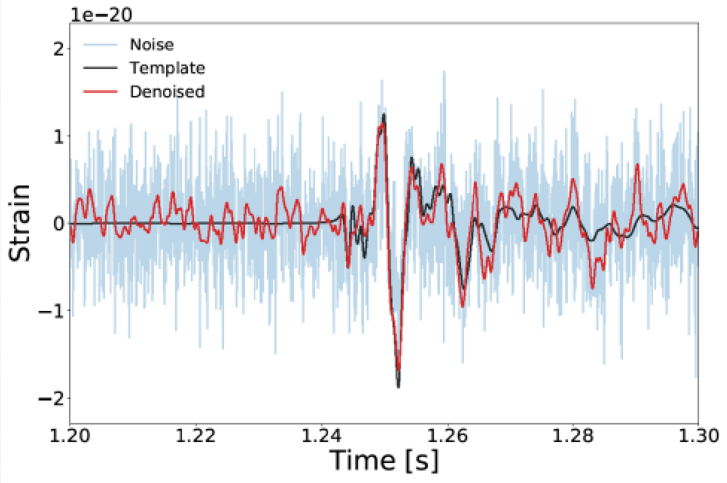


TABLE I. Values of the SSIM index for CCSN waveforms when using the optimal value of the regularization parameter for each signal, $\mu_{\rm opt}$; the mean value for all signals, $\bar{\mu}$; and multiple values, $\mu_{\rm m}$. The final column, Ref, indicates the SSIM index computed for the signal obtained after the whitening and the corresponding template.

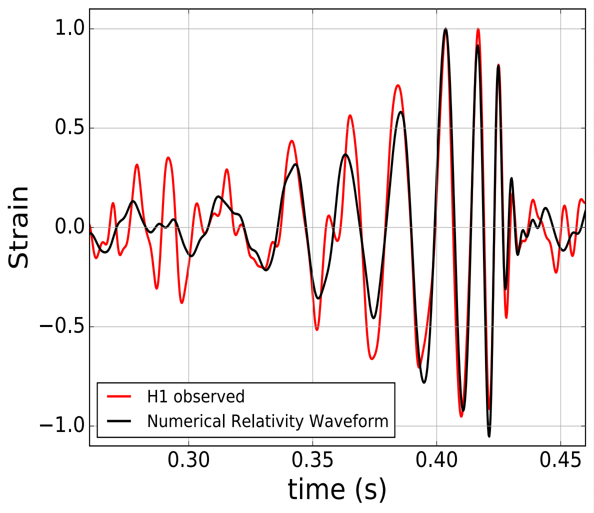
		SSIM index			
Signal	Distance (kpc)	$[\mu_{\rm opt}]$	$[\bar{\mu}]$	$[\mu_{\rm m}]$	[Ref]
A	5	0.89	0.83	0.84	0.39
	10	0.74	0.68	0.69	0.14
	20	0.54	0.44	0.43	0.03
B	5	0.71	0.61	0.65	0.21
	10	0.51	0.33	0.38	0.06
	20	0.31	0.08	0.11	0.003
C	5	0.64	0.60	0.69	0.06
	10	0.40	0.46	0.51	0.012
	20	0.23	0.24	0.29	0.002

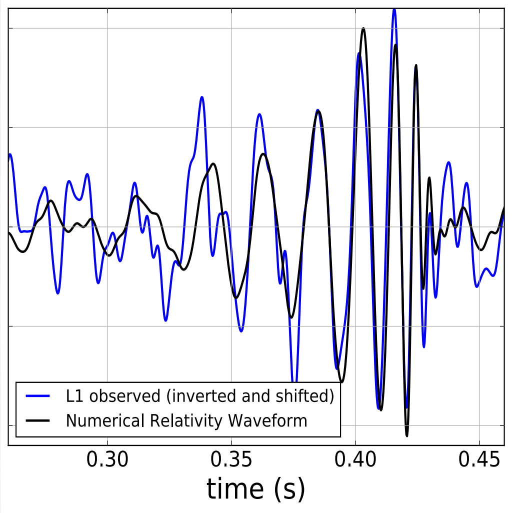
Not very strong dependence on the regularization parameter.

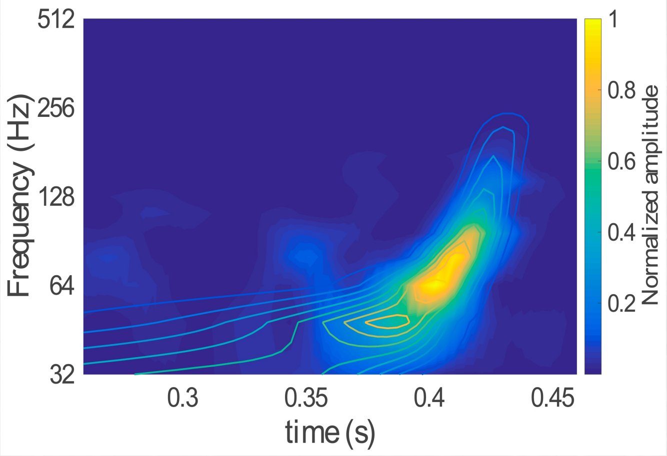
TF, E. Cuoco, A. Marquina, J.A. Font and J.M. Ibañez (2018)

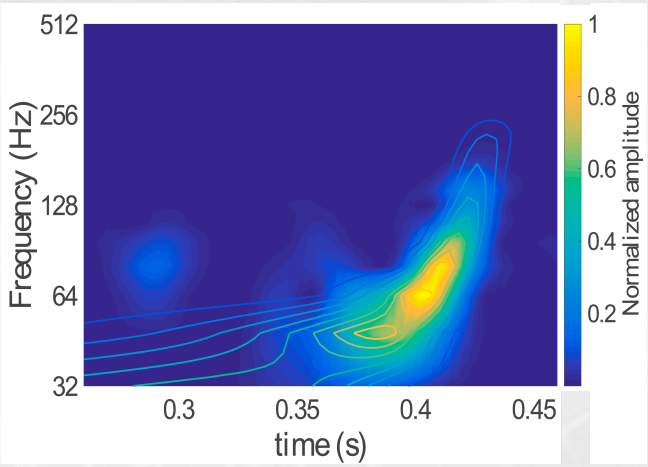
TV methods

GW150914









Integration with CWB

Barneo P.J., TF, Font J.A., Drago M., Portell J., Andrade M., and Marquina A. in preparation

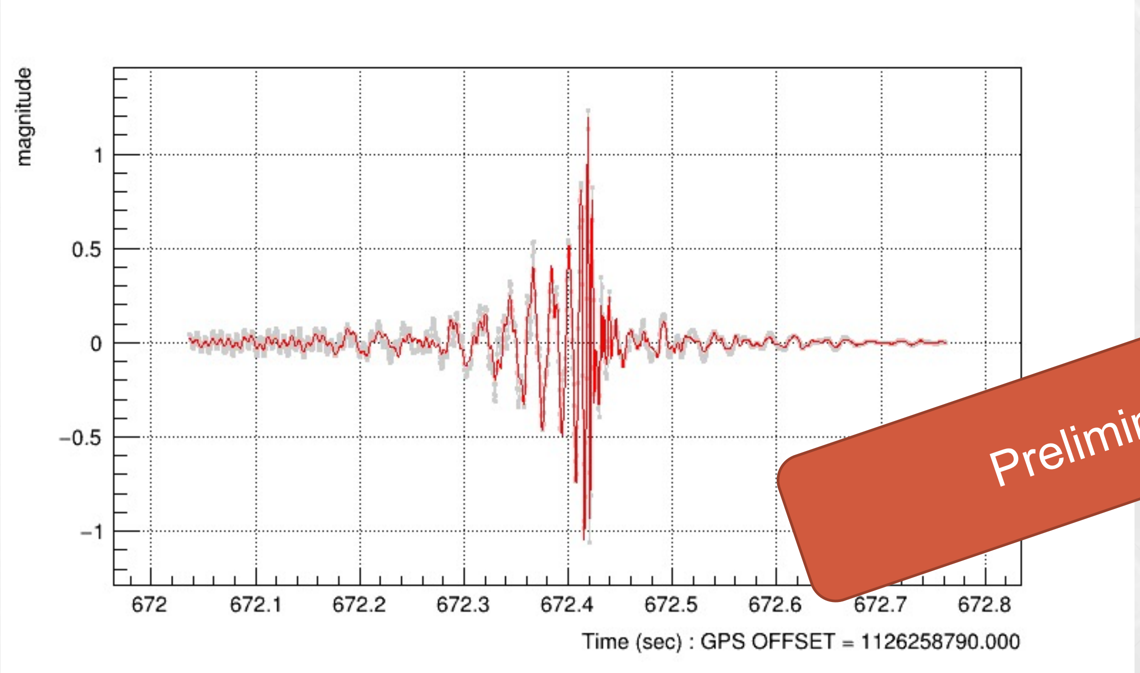
GW150914

CWB

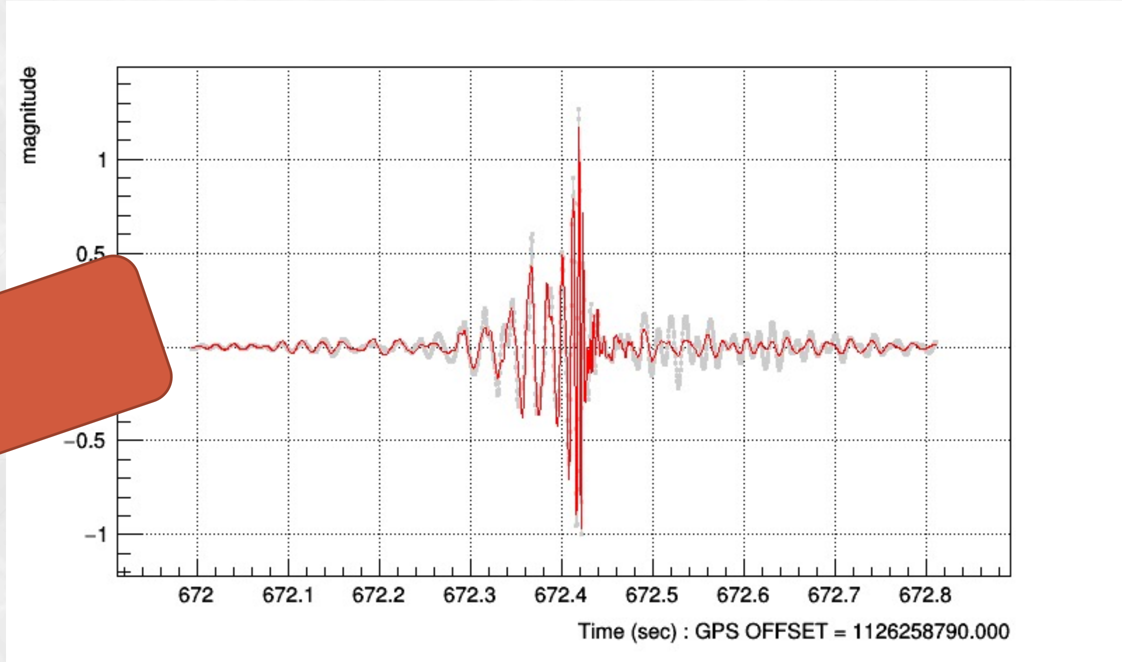
SNR
25.2

CWB + rROF

SNR
27.1

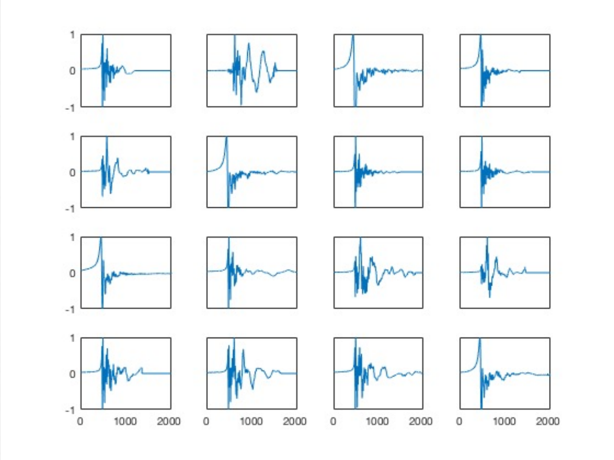


Preliminary



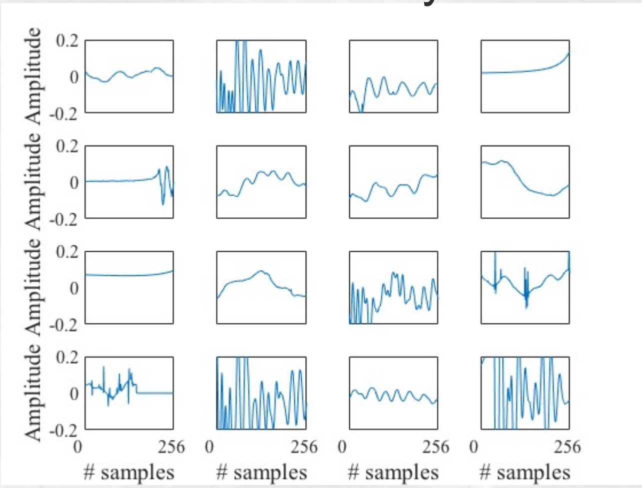
Learning process

Catalog of GW signals

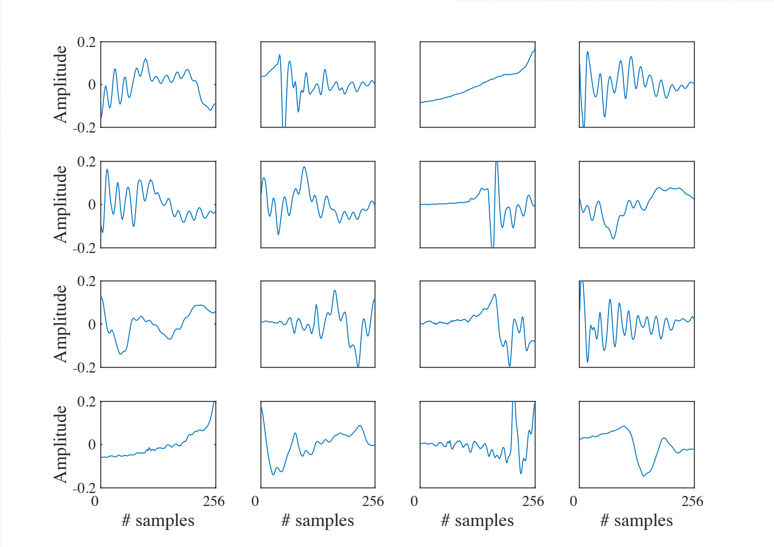


Select random patches

Initial dictionary



Perform dictionary learning



Learned dictionary

Dictionary denoising process

- We start by considering a finite number of training signals: m patches of length n.
- To obtain the trained dictionary, we add the dictionary matrix D as a variable in the minimization problem:

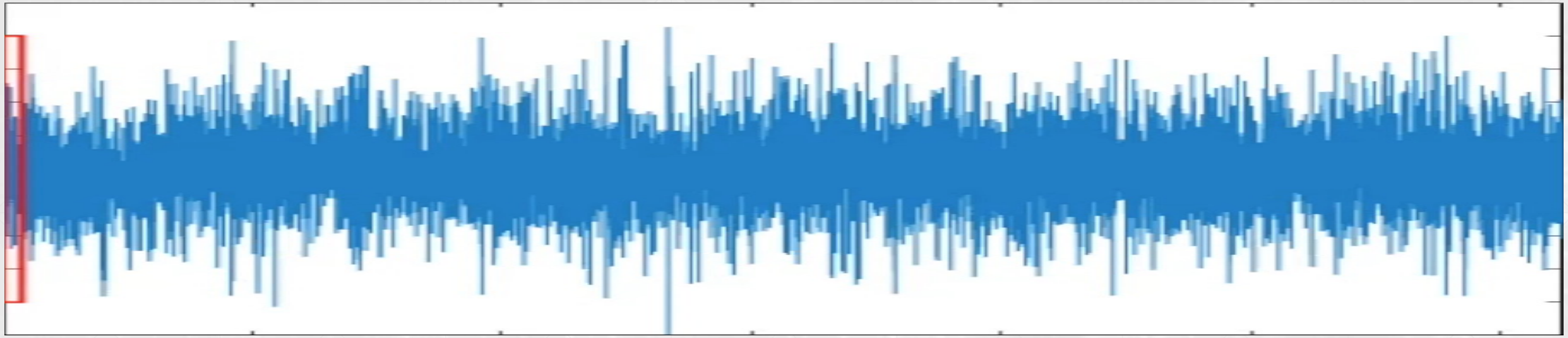
$$\alpha = \underset{\alpha,\mathbf{D}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{m} ||\mathbf{D}\alpha_i - \mathbf{x}_i||_2^2 + \lambda ||\alpha_i||_1,$$

The i-th coefficient contains the coefficients of the sparse representation of each atom in the dictionary.

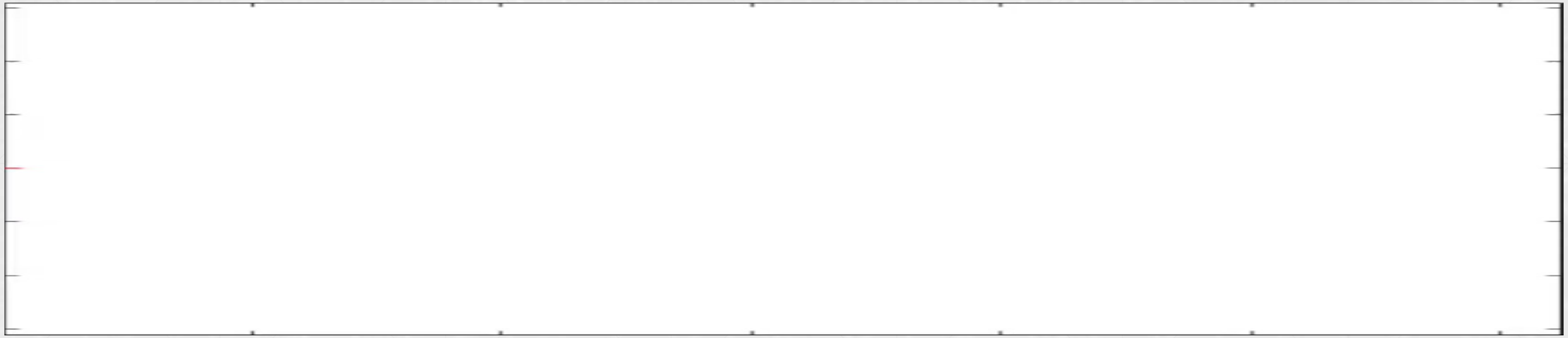
- Dictionary updated using a block-coordinate descent method (Mairal et al 2009, Tseng 2001).
- CCSN and BBH gravitational wave catalogs. 80% of waveforms to train the dictionary, 15% for method validation, and 5% to test algorithm.
- Signals shifted to be aligned with minimum peak (CCSN) or maximum peak in the merger (BBH). 2048 samples to train the dictionary

Dictionary denoising process

Noisy signal



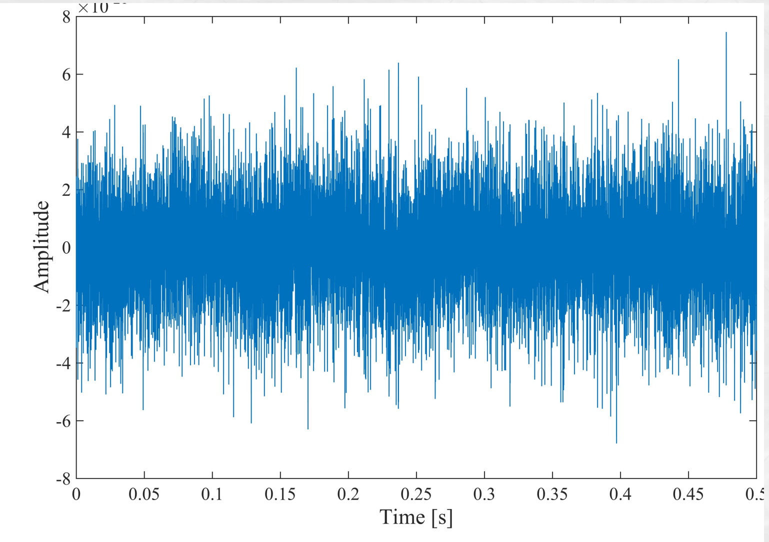
Denoised signal

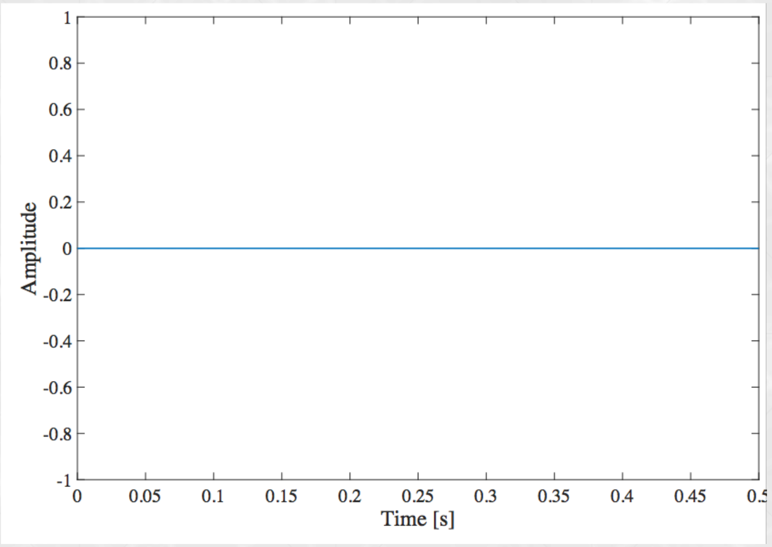


Dictionary denoising process

No signal

- Test with no signal.
- Lasso set coefficients to 0 when the signal is very different from catalog atoms.
- The selection of λ is key to avoid spurious signal reconstruction.
- In a more realistic scenario the presence of instrumental glitches could produce a false reconstruction.

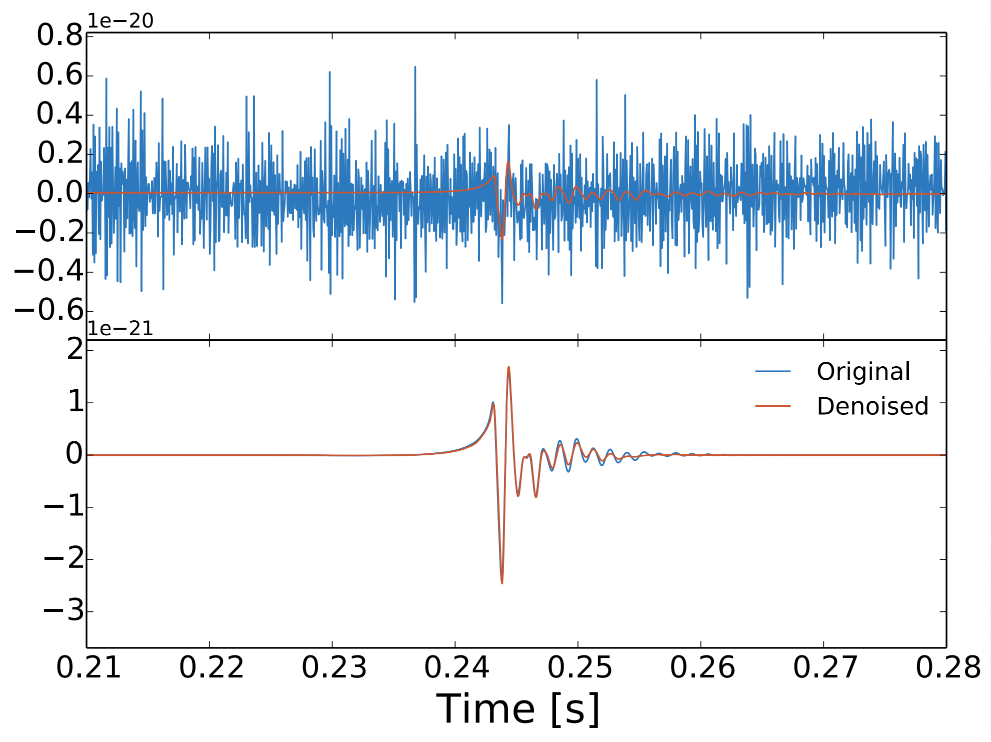


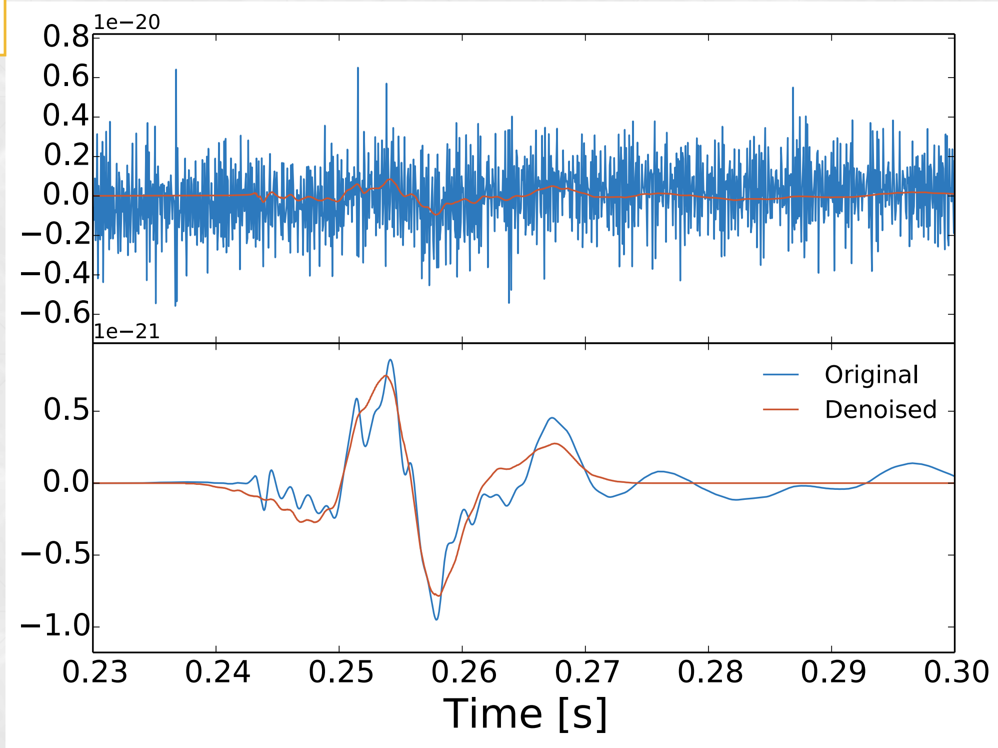


Dictionary denoising process

CCSN

SNR = 20





Peaks well recovered.
Ring down weak peaks set to zero

$$\mathrm{SSIM} = 0.98$$

$$\mathrm{MSE} = 0.018 \times 10^{-3}$$

Small oscillations are lost.
Broad morphology still captured.

$$\mathrm{SSIM} = 0.67$$

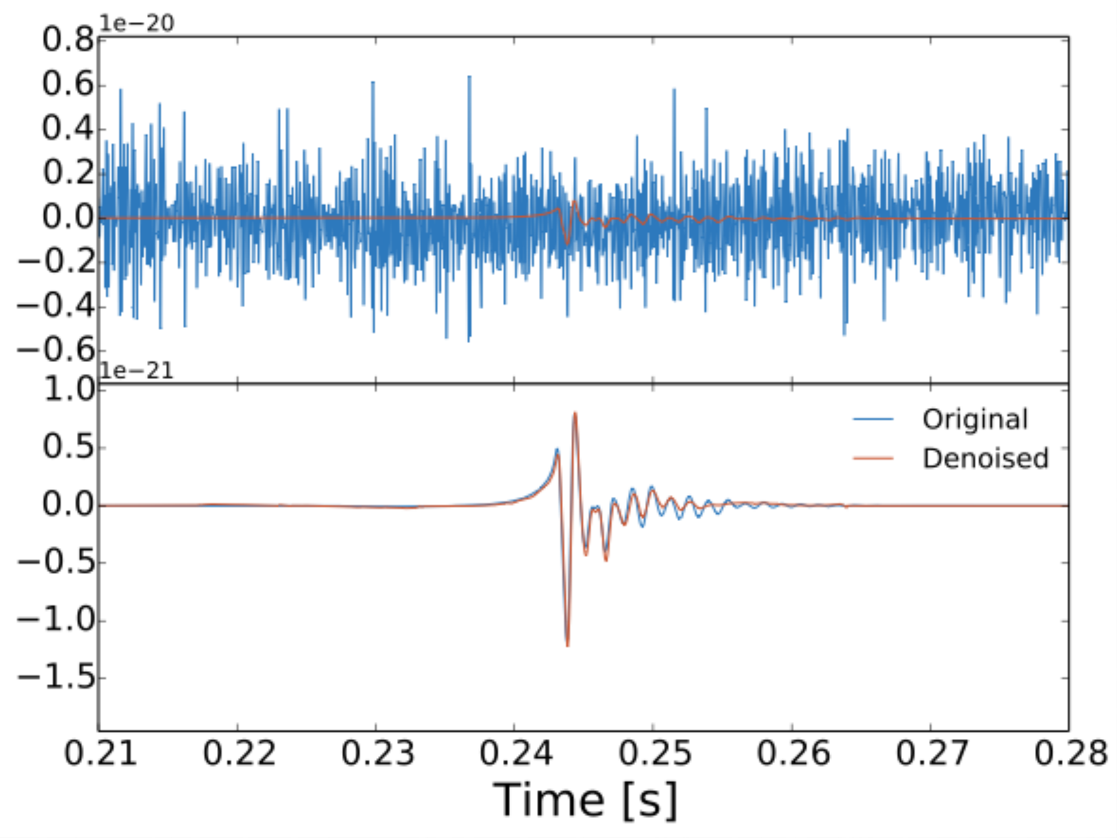
$$\mathrm{MSE} = 0.271 \times 10^{-3}$$

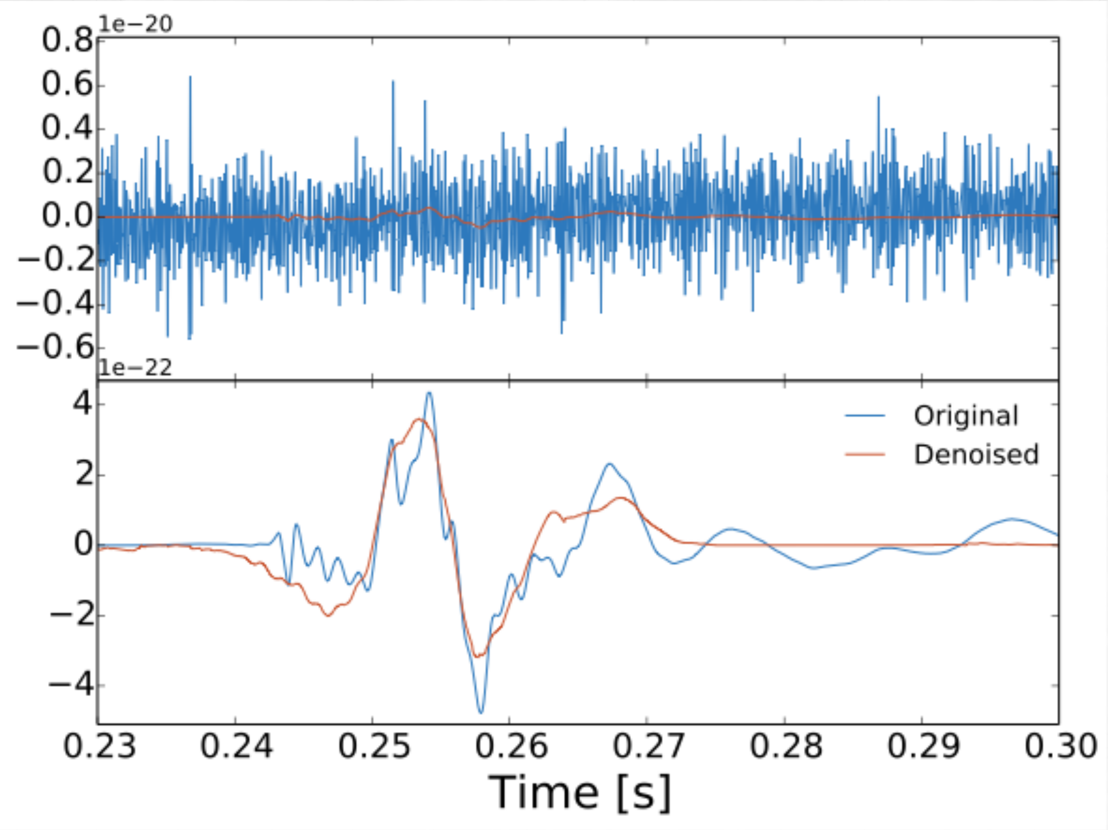
TF, A. Marquina, J.A. Font and J.M. Ibañez (2017)

Dictionary denoising process

CCSN

More challenging scenario. SNR= 10





$$\mathrm{MSE} = 0.081 \times 10^{-3}$$

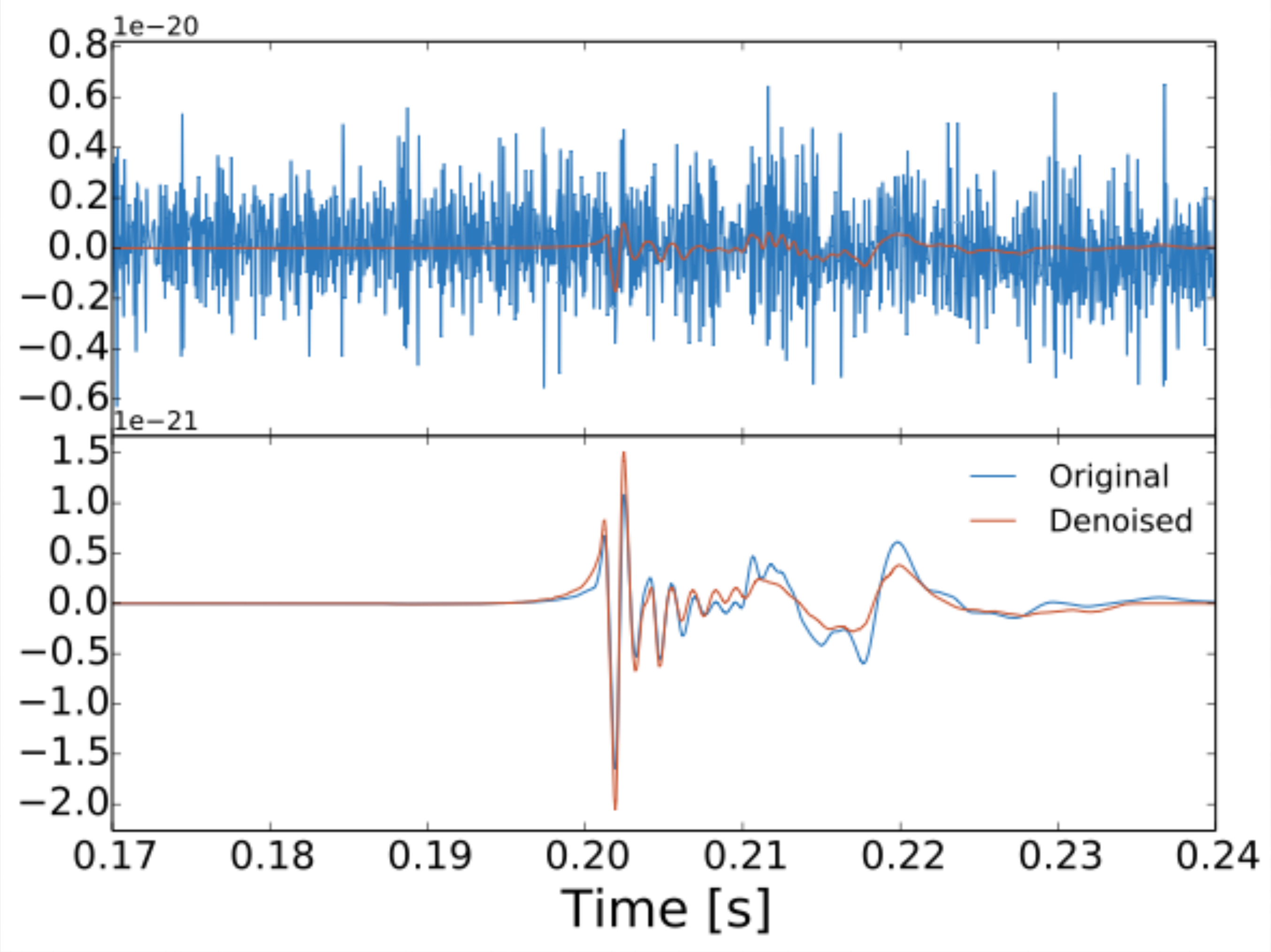
$$\mathrm{MSE} = 0.2 \times 10^{-3}$$

$$\mathrm{SSIM} = 0.91$$

$$\mathrm{SSIM} = 0.71$$

Dictionary learning results

Signal from different catalog



- We test the method with a signal from other catalog.
- Similar structure (burst).
- Accurate results.
- Some peaks missing.

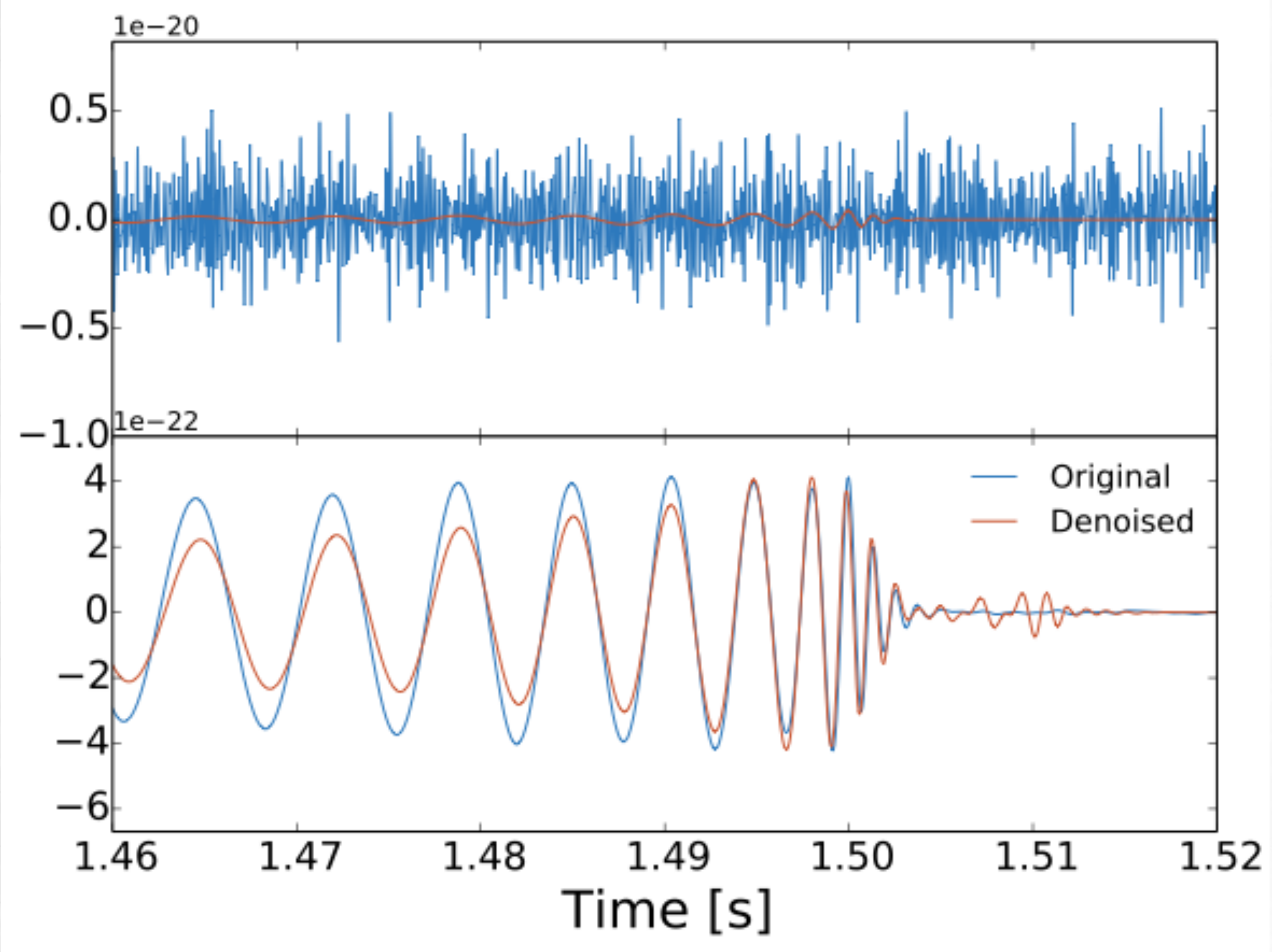
E. Abdikamalov, S. Gossan, A. M. DeMaio, and C. D. Ott, Phys. Rev. D, **90**, 044001 (2014).

Dictionary learning results

BBH

- Random signal from test set.
- Random time of arrival.
- SNR = 20.

$$\text{SSIM} = 0.86$$
$$\text{MSE} = 0.031 \times 10^{-3}$$

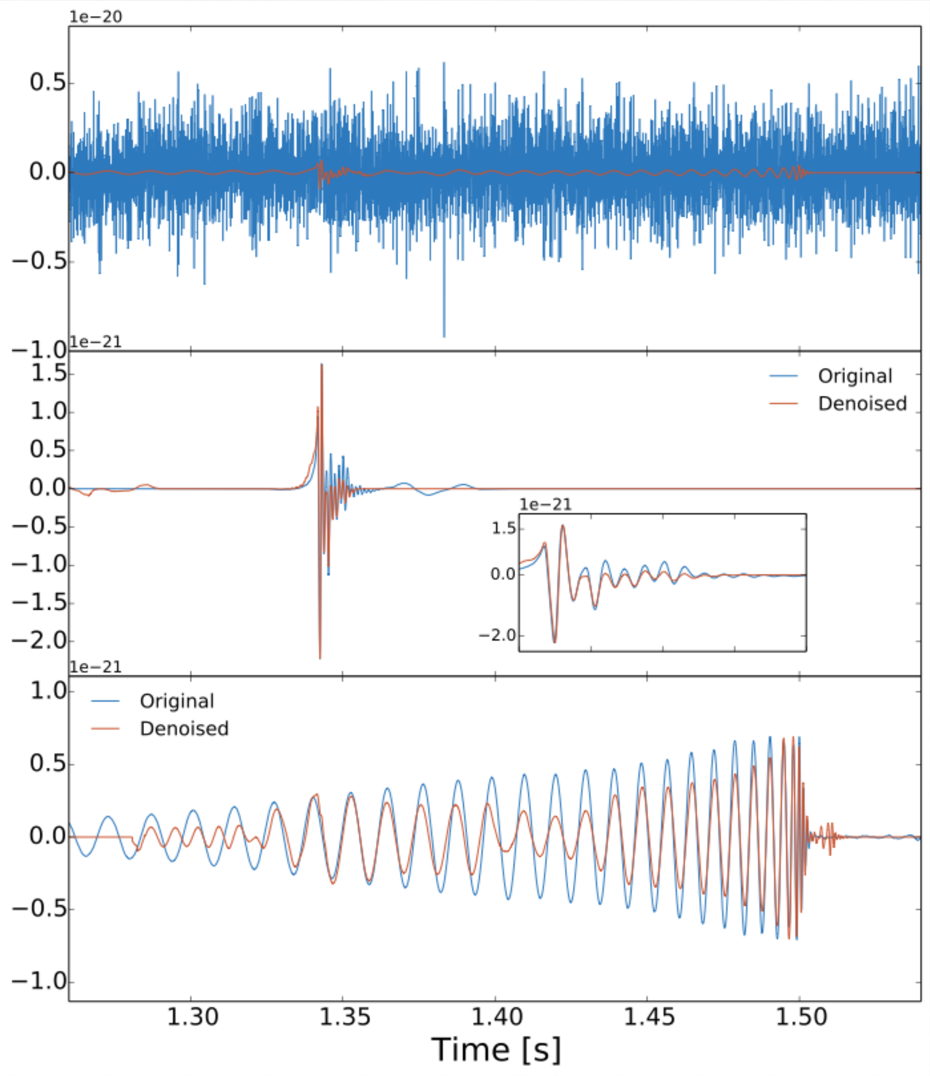


We are only interested in merger part. Use other D to recover the chirp area.

Dictionary learning results

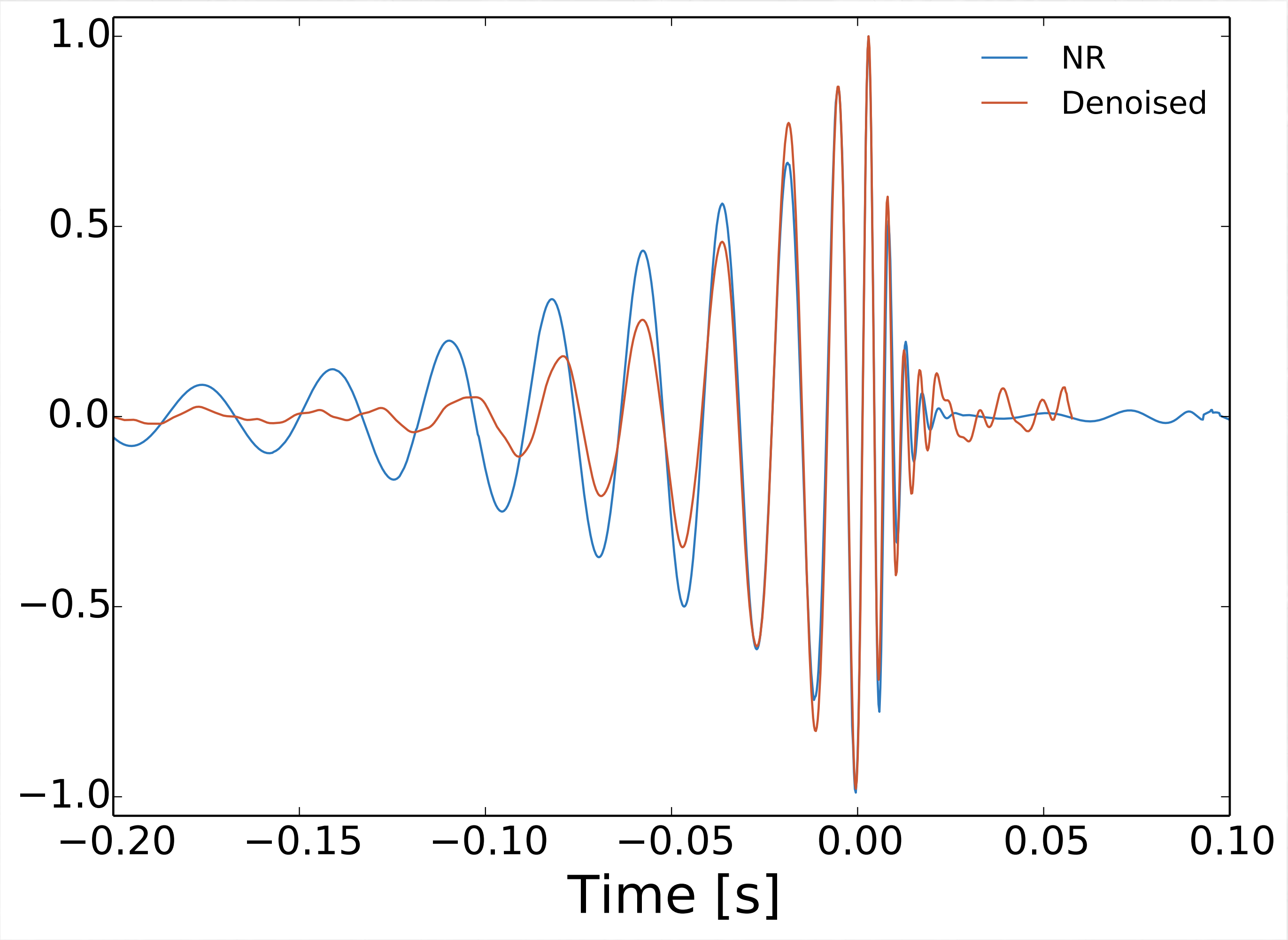
Combination of signals

- Test the dictionary when dealing with signals different from the type they are designed for.
- Use both dictionaries independently.
- Each dictionary discriminates well between the type of signal.



Dictionary denoising process

GW150914



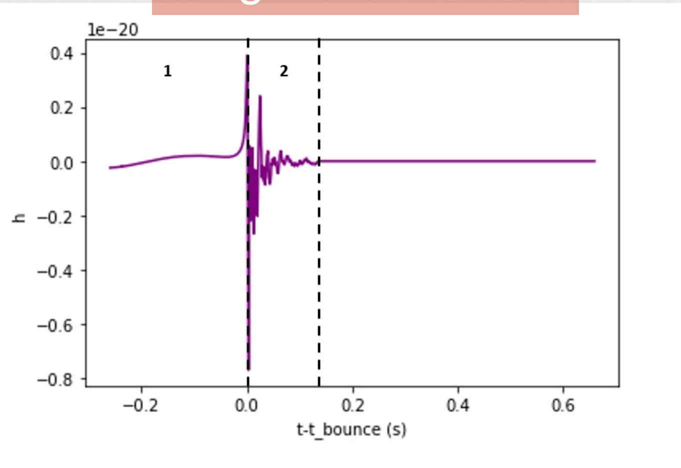
Aditional results

CCSN mechanism extraction with LASSO

Saiz-Pérez, A ; TF; Font, José A: arXiv:2110.12941

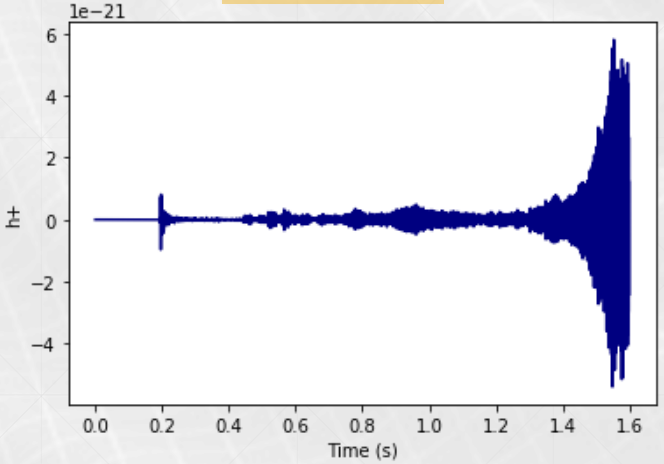
- We try to reproduce the analysis done in Logue* + 2012
- The idea is to determine the main mechanism of emission of GW.
  - Magnetorotational mechanism.
  - Neutrino mechanism.
  - Glitch model (non physic).
- We use signal from three CCSN catalogs embedded in Gaussian noise.
- Use non-trained dictionaries to determine to which dictionary belongs a random signal.

Magnetorotational



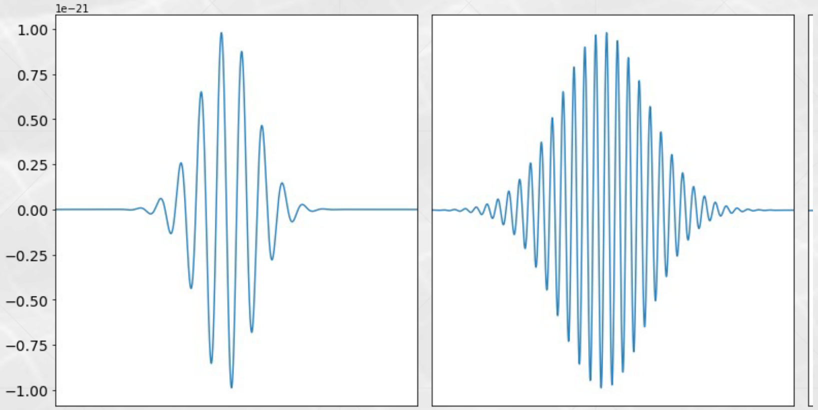
Dimmelmaier + 2008

Neutrino



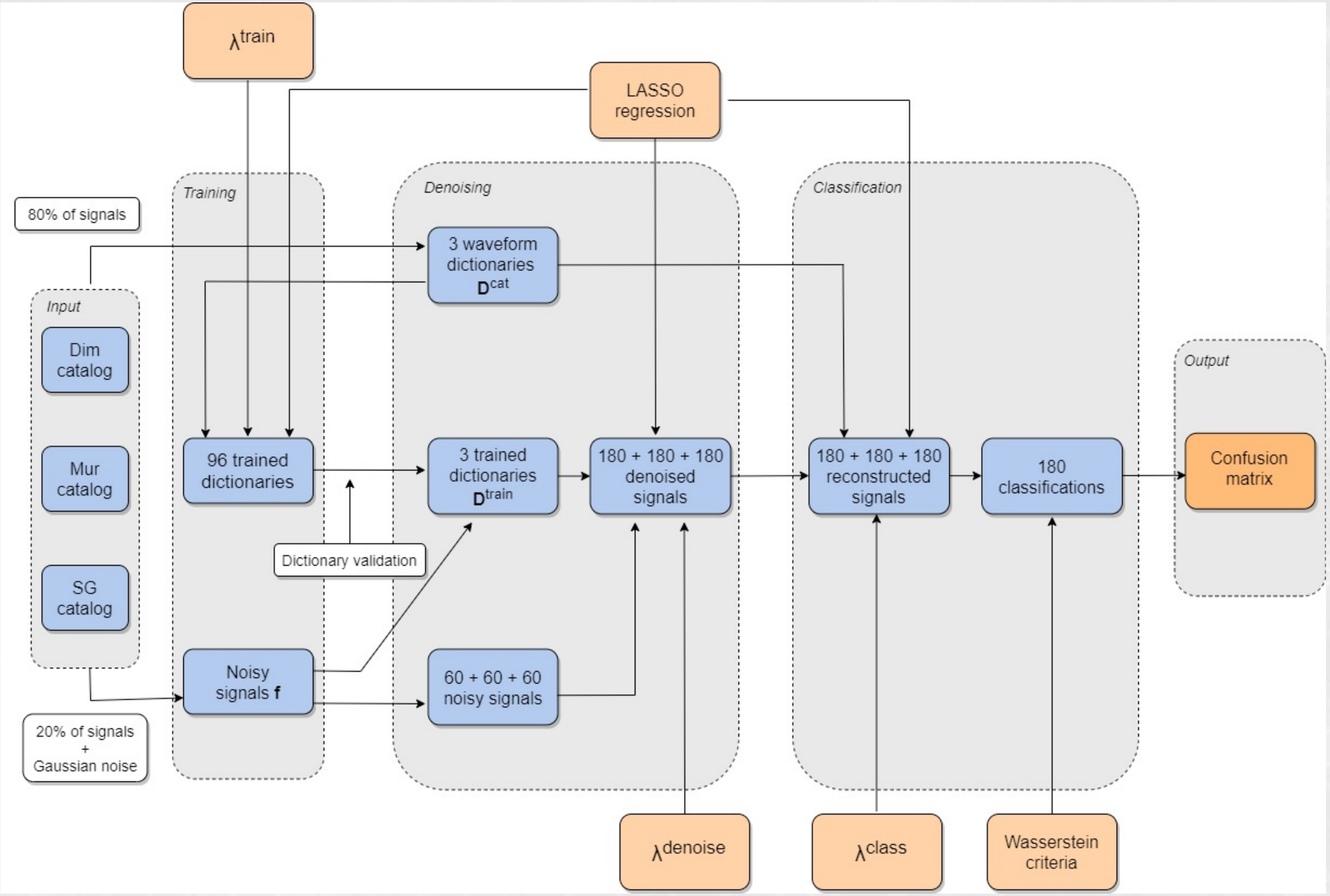
Murphy + 2009

Glitch

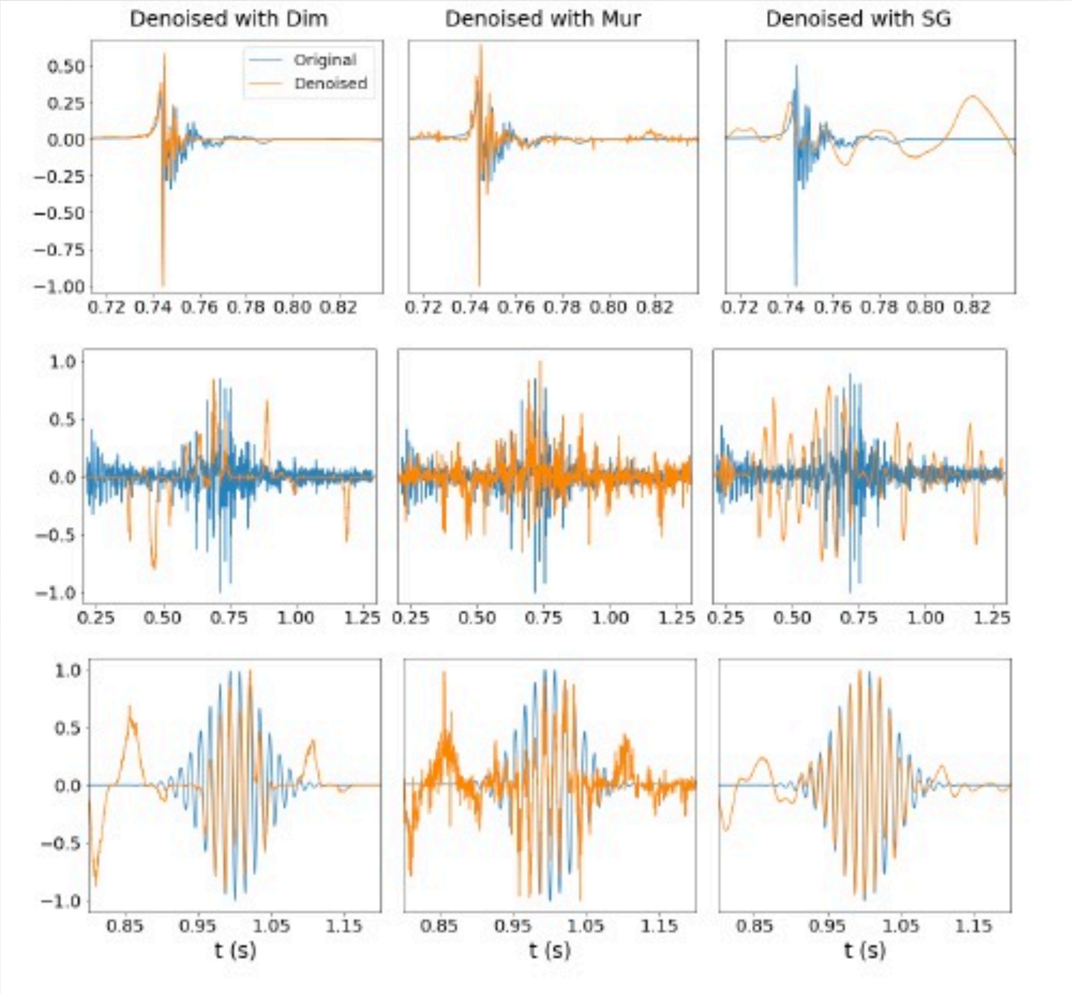


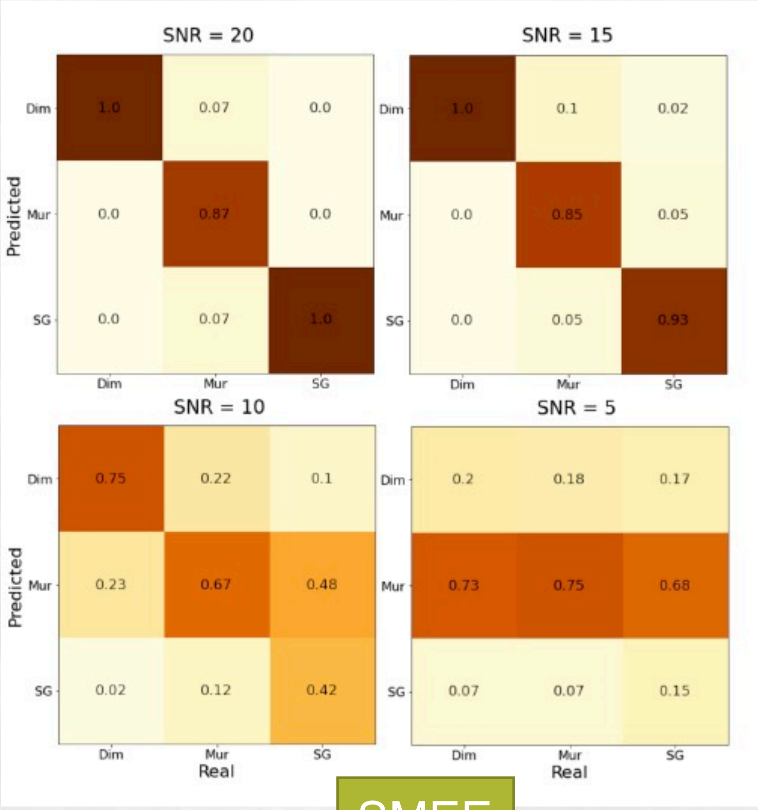
* Logue J., Ott C. D., Heng I. S., Kalmus P., Scargill J. H. C., 2012, Phys. Rev. D, 86, 044023

Classification of the core-collapse supernova explosion mechanism with learned dictionaries



CCSN mechanism extraction with DL





SMEE

100 % of magneto-rotational
95 % neutrino

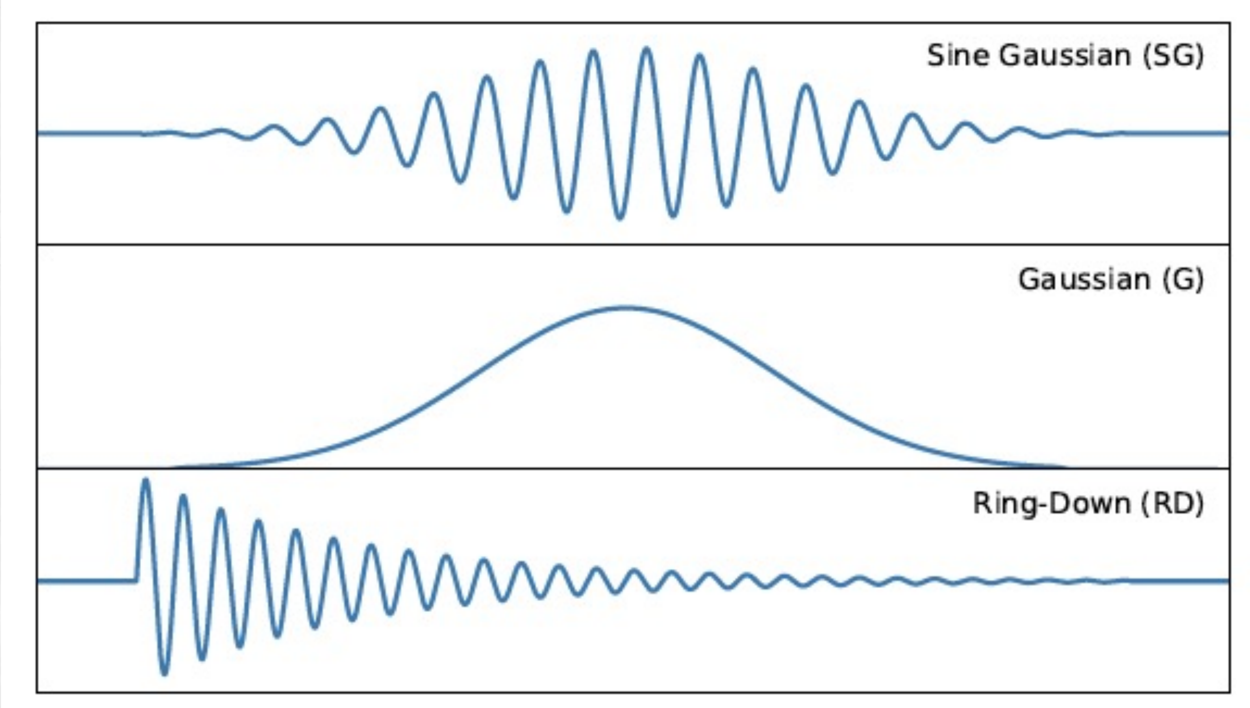
95 % of magneto-rotational
80 % neutrino

Glitch denoising and classification with dictionary learning

- In the LVC there exist diverse strategies to classify glitches in the detectors:
- Powell et al (2015, 2017):
  - PCAT: Principal Component Analysis for Transients. Uses PC coefficients to classify glitches using a Gaussian Mixture Model.
  - PC-LIB: Based on LAL-Inference. Computes Bayes factor for glitch selection. Supervised classification.
  - WDF-ML: Wavelet Detection Filter + (unsupervised) ML algorithm (GMM).
- Zevin et al (2017): Gravity Spy, Zooniverse Platform. Citizen science + ML.
- Mukund et al (2017): Difference Boosting Neural Network (supervised Bayesian classifier).
- George et al (2018): Deep Learning + Transfer Learning.
- Razzano & Cuoco (2018): Convolutional Neural Networks to classify glitches from their spectrograms (time-frequency evolution)
- Llorens-Monteagudo et al (2018): Dictionary learning.
- Farr et al: Transient glitch mitigation in Advanced LIGO data PRD 104, Issue 10 (2021)
- Ben Farr Talk http://www.ipam.ucla.edu/abstract/?tid=17039&pcode=GWAWS3
- Others

Glitch denoising and classification with dictionary learning

- Simulated glitches embedded in Gaussian noise to simulate the background noise of advanced LIGO in its broadband configuration.
- Data set of 3000 simulated glitches of three different waveform morphologies, comprising 1000 glitches per morphology.
- 3 simple types of glitch morphologies (following Powell et al, 2015)

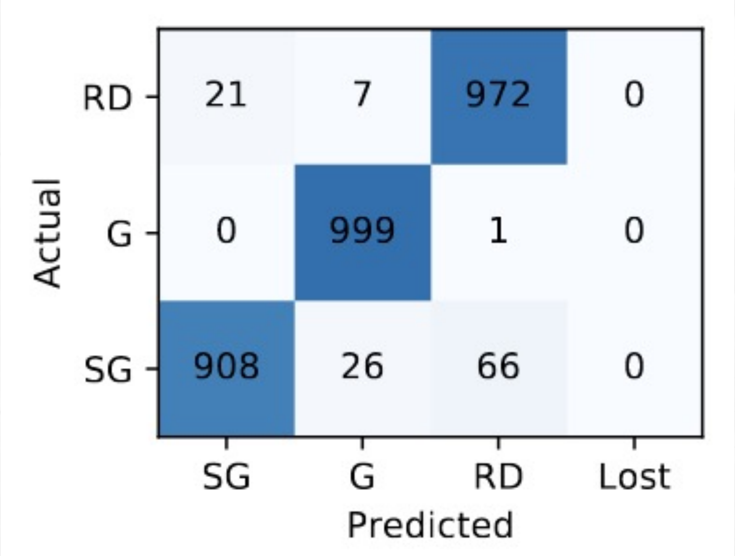


Glitch denoising and classification with dictionary learning

Llorens-Monteagudo+ (2018)

Out of 3000 glitches, 2879 (96%) are correctly classified.

Confusion matrix



Performance barely decreases with SNR (down to SNR~10).

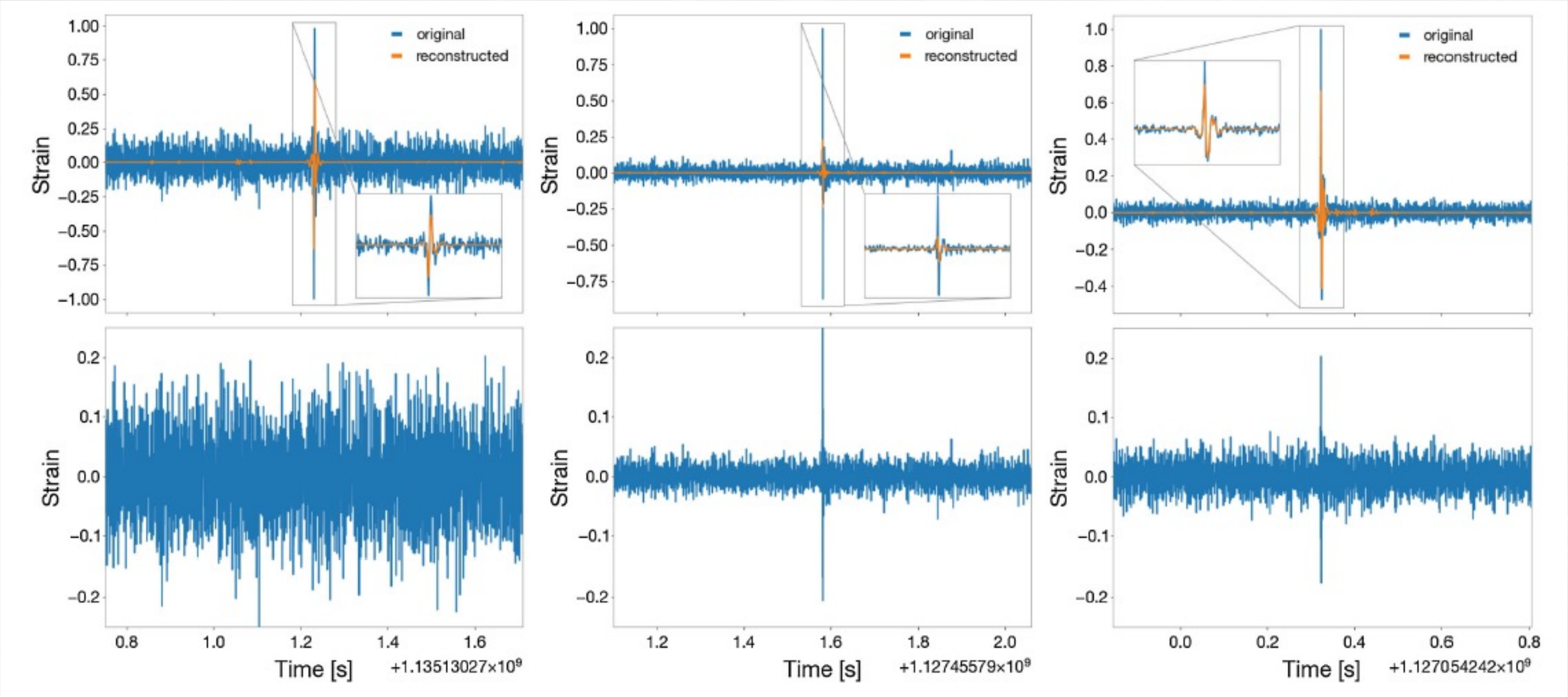
Most misclassified glitches have highest and lowest frequencies. More affected by noise than intermediate frequencies.

Blip denoising via dictionary learning

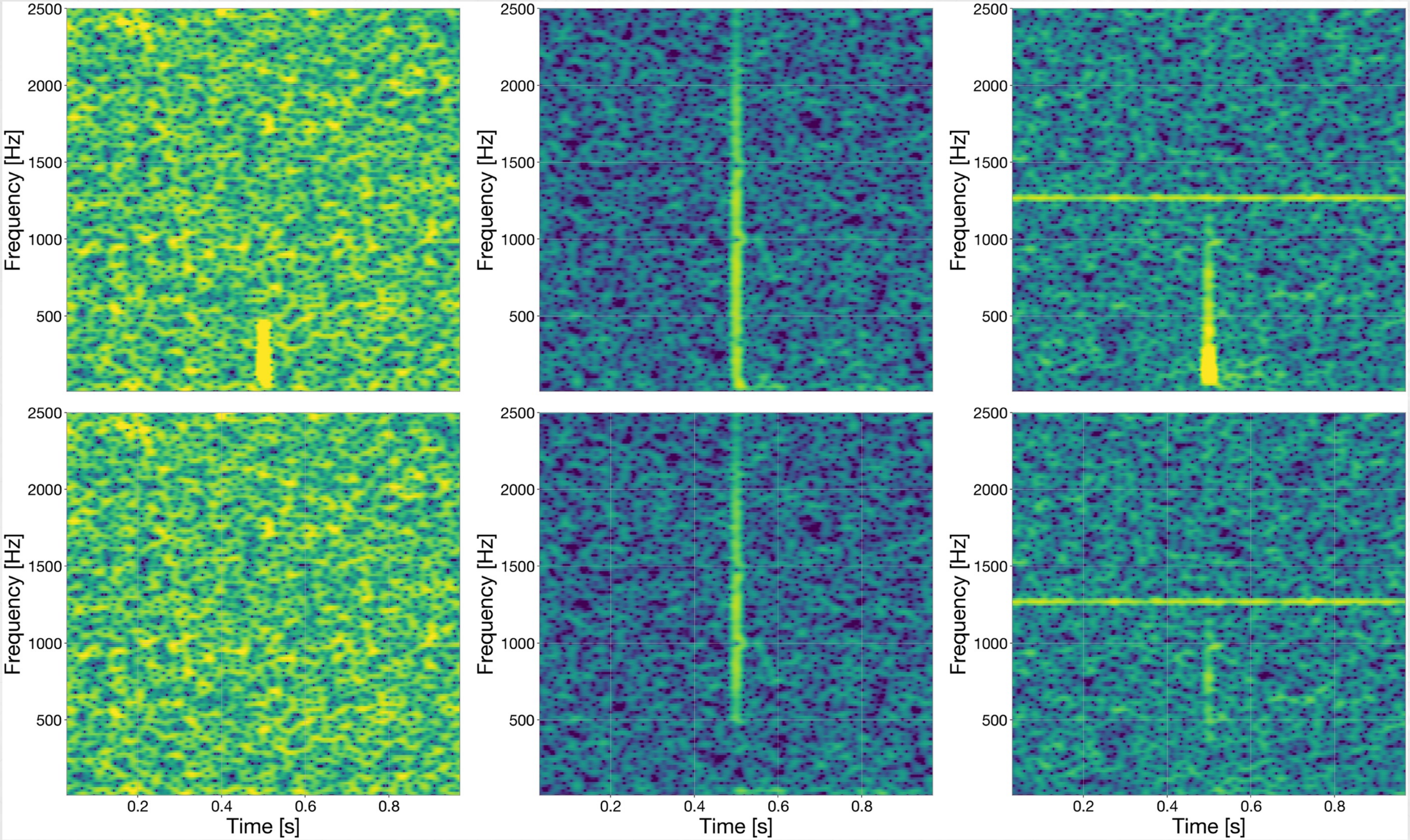
Goal: Try to create a model to remove glitches from noise.

Blip denoising via dictionary learning

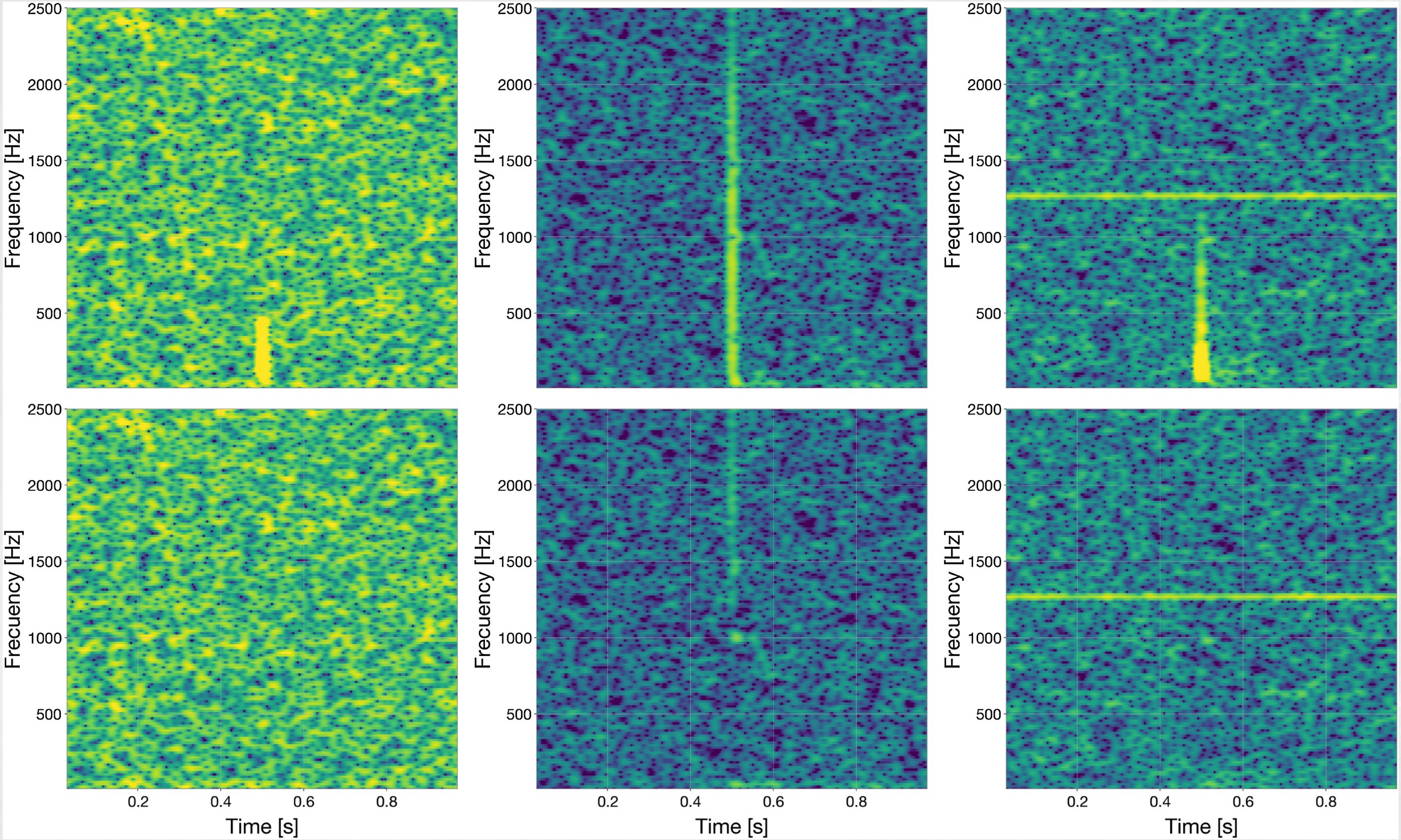
TF and al. Physical Review D, Volume 102, Issue 2, article id.023011, 2020



Blip denoising via dictionary learning

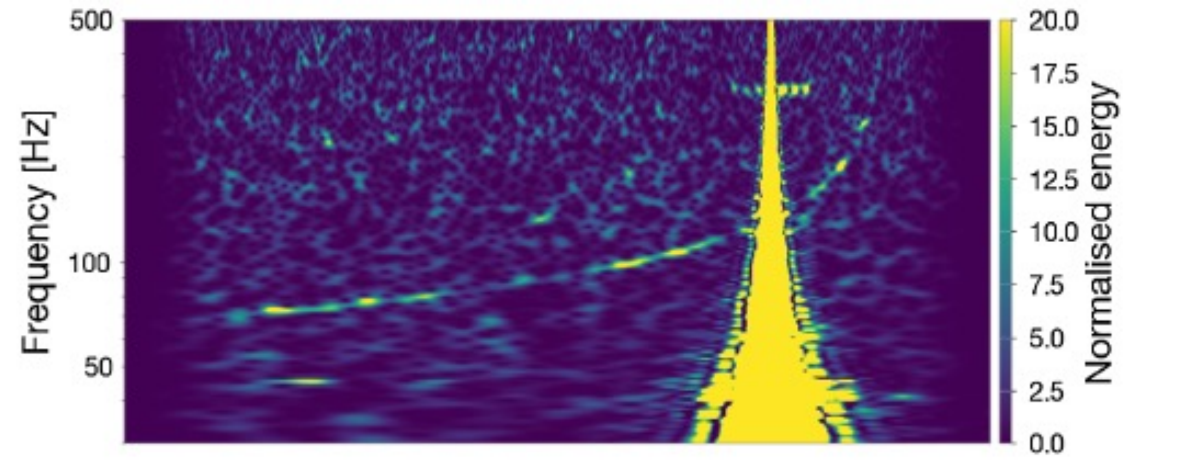


Blip denoising via dictionary learning



Blip denoising via dictionary learning

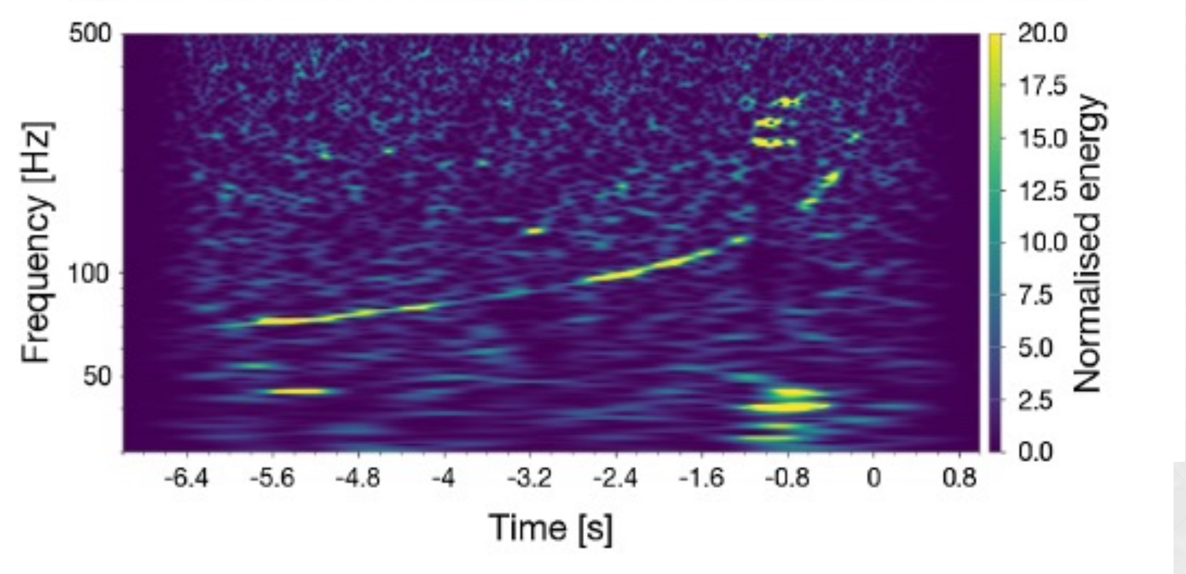
Test No.	SNR_o	SNR_{single}	SNR_{multi}	W_o	W_{single}	W_{multi}
1	5.3	1.3	1.3	0.99	0.99	0.99
2	5.2	2.4	1.3	0.94	0.94	0.93
3	9.2	1.3	1.3	0.99	0.99	0.99
4	37.1	10.2	1.5	0.84	0.92	0.97
5	18.1	2.3	1.7	0.99	0.98	0.98
6	13.5	3.8	1.7	0.98	0.98	0.98
7	4.1	1.3	1.3	0.98	0.98	0.98
8	6.3	1.3	1.3	0.96	0.97	0.97
9	13.4	9.8	2.9	0.98	0.98	0.98
10	8.7	3.0	1.7	1.00	0.99	0.99
11	7.3	1.3	1.3	0.99	0.99	0.99
12	5.8	1.3	1.2	0.99	1.00	1.00
13	4.5	1.8	1.3	0.99	0.99	0.99
14	14.2	1.2	1.2	0.99	0.99	0.99
15	15.2	2.2	1.3	0.88	0.88	0.89
16	6.2	1.3	1.3	0.99	0.99	0.99



ipam

Blip denoising via dictionary learning

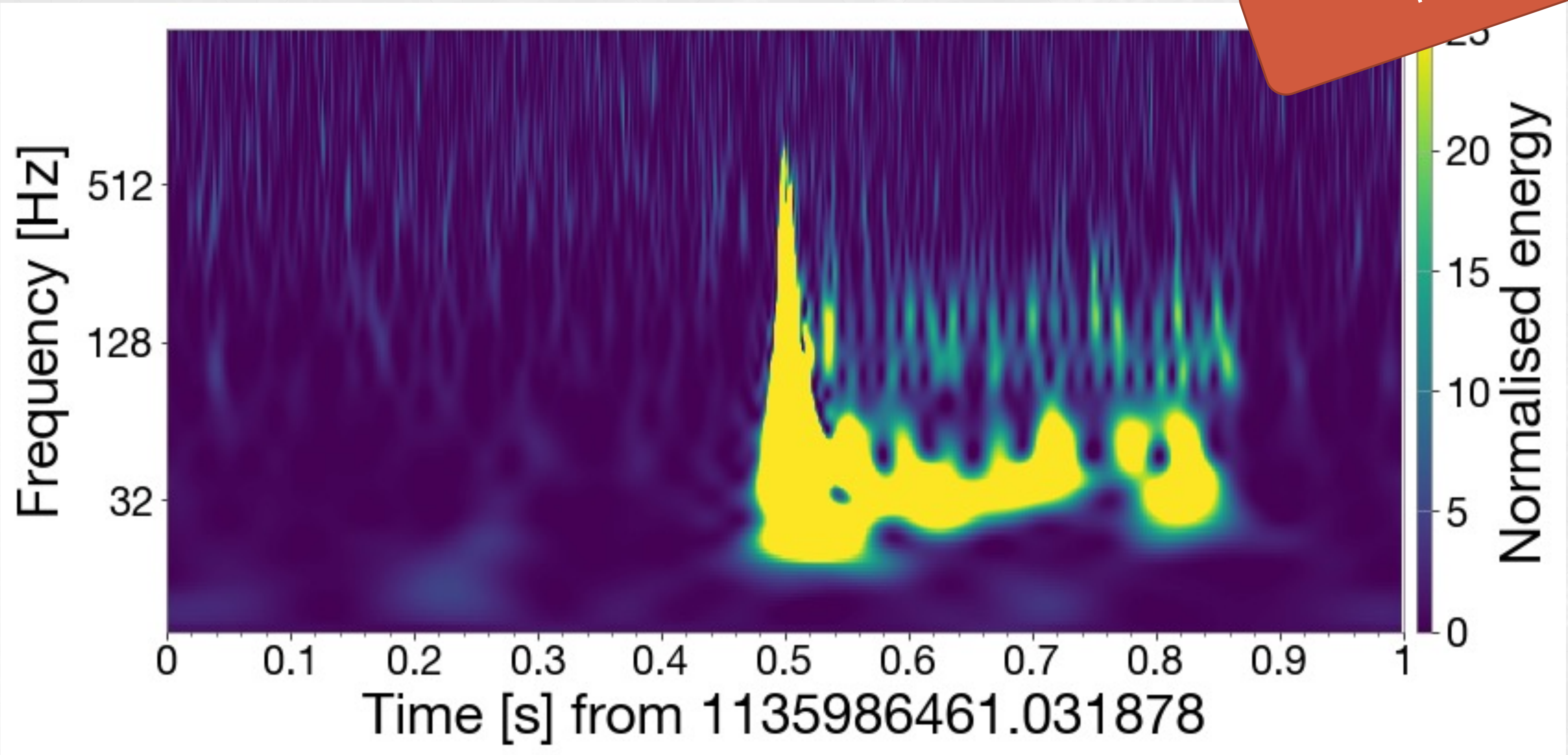
Test No.	SNR_o	SNR_{single}	SNR_{multi}	W_o	W_{single}	W_{multi}
1	5.3	1.3	1.3	0.99	0.99	0.99
2	5.2	2.4	1.3	0.94	0.94	0.93
3	9.2	1.3	1.3	0.99	0.99	0.99
4	37.1	10.2	1.5	0.84	0.92	0.97
5	18.1	2.3	1.7	0.99	0.98	0.98
6	13.5	3.8	1.7	0.98	0.98	0.98
7	4.1	1.3	1.3	0.98	0.98	0.98
8	6.3	1.3	1.3	0.96	0.97	0.97
9	13.4	9.8	2.9	0.98	0.98	0.98
10	8.7	3.0	1.7	1.00	0.99	0.99
11	7.3	1.3	1.3	0.99	0.99	0.99
12	5.8	1.3	1.2	0.99	1.00	1.00
13	4.5	1.8	1.3	0.99	0.99	0.99
14	14.2	1.2	1.2	0.99	0.99	0.99
15	15.2	2.2	1.3	0.88	0.88	0.89
16	6.2	1.3	1.3	0.99	0.99	0.99



Glitch denoising via dictionary learning

Koi Fish

Preliminary

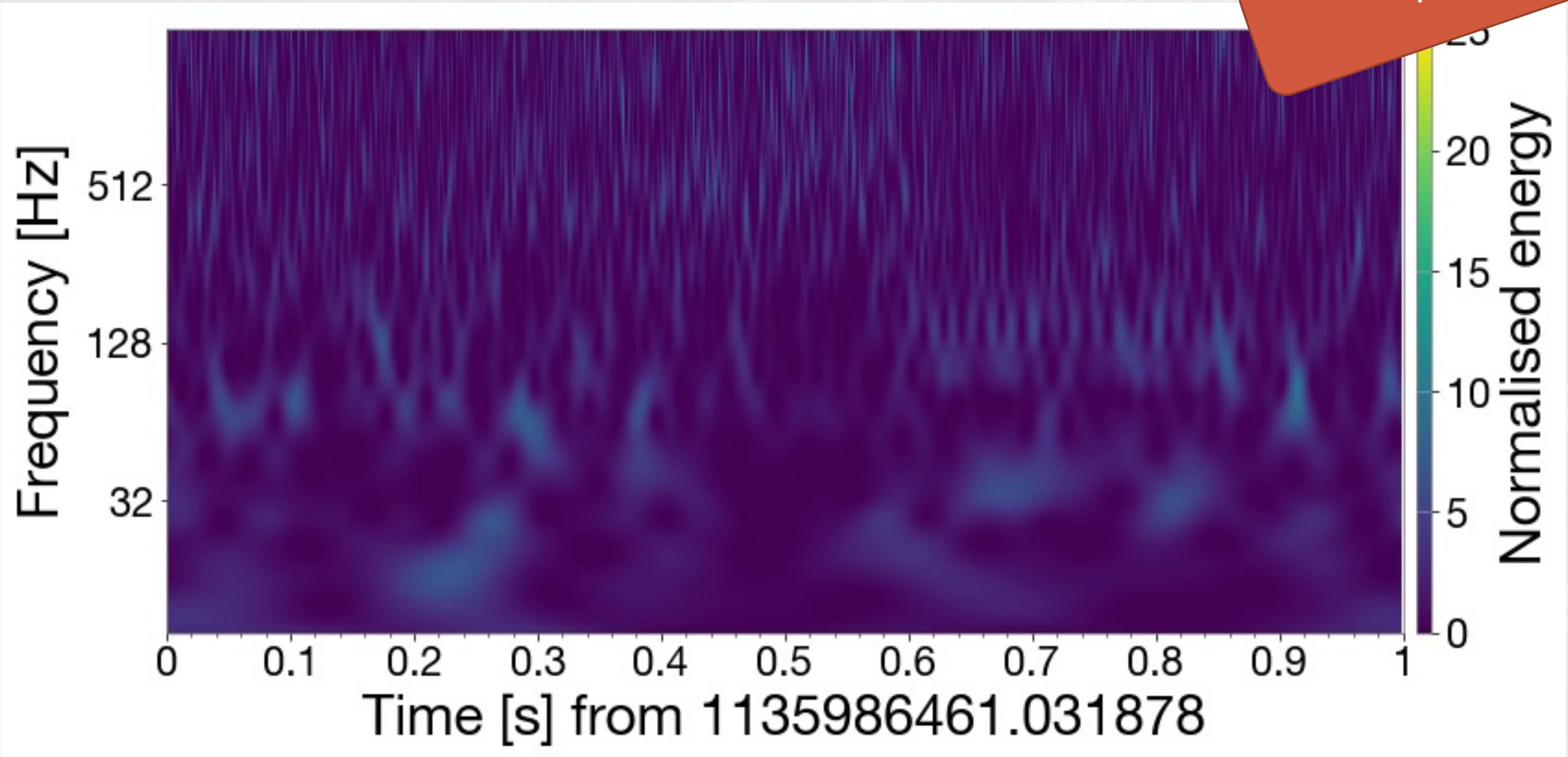


TF, E. Cuoco, A. Marquina and J.A. Font + (in preparation)

Glitch denoising via dictionary learning

Koi Fish

Preliminary

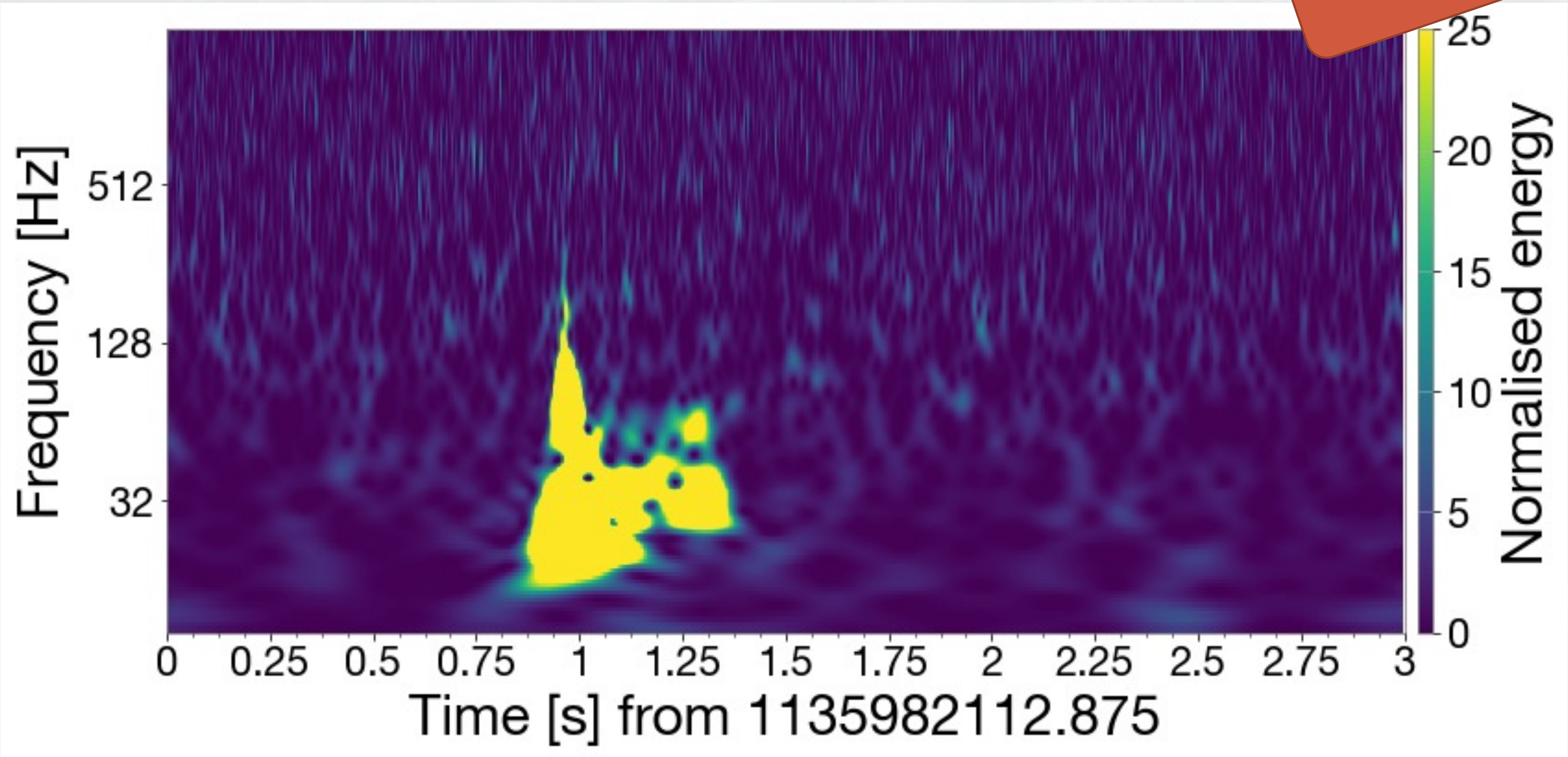


TF, E. Cuoco, A. Marquina and J.A. Font + (in preparation)

Glitch denoising via dictionary learning

Scattered Light

Preliminary

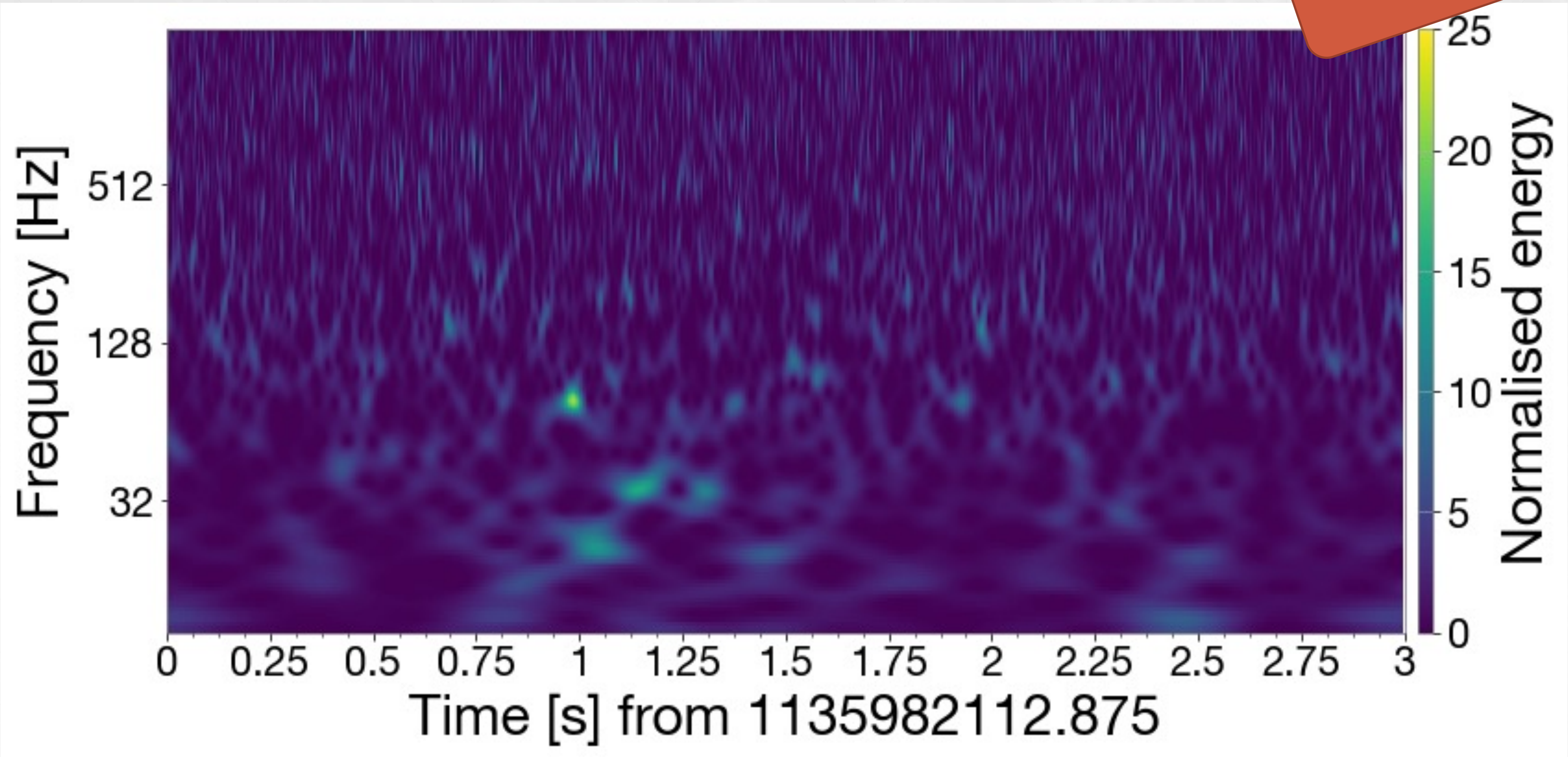


TF, E. Cuoco, A. Marquina and J.A. Font + (in preparation)

Glitch denoising via dictionary learning

Scattered Light

Preliminary

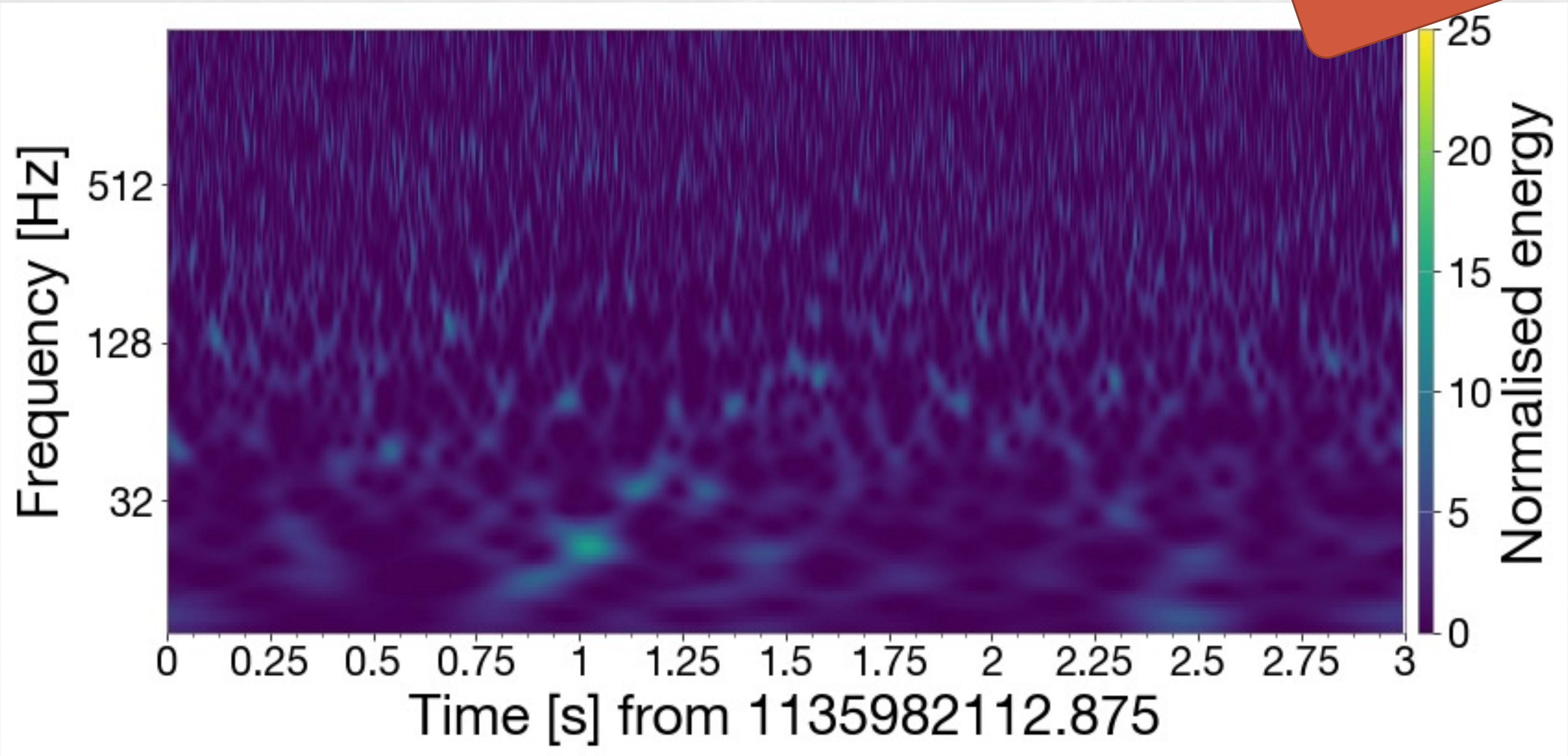


TF, E. Cuoco, A. Marquina and J.A. Font + (in preparation)

Glitch denoising via dictionary learning

Preliminary

Scattered Light



TF, E. Cuoco, A. Marquina and J.A. Font + (in preparation)

Summary and Conclusions

- We have discussed variational methods for minimization problems based on the TV-norm in the context of gravitational-wave signals.
- Novel strategy in the field.
- We have shown that TV algorithms can be useful in the field of Gravitational-Wave Astronomy as a tool to remove noise.
- Integration of rROF with CWB. Next step integrate with Bilby.

- We have shown that denoising based in Dictionary Learning produces very good results in the case of Gaussian noise.
- We have obtained promising results combining the denoising with a classification based in LASSO.
- Next step with real data for both signals and glitches is in progress.

Thank you for your atention