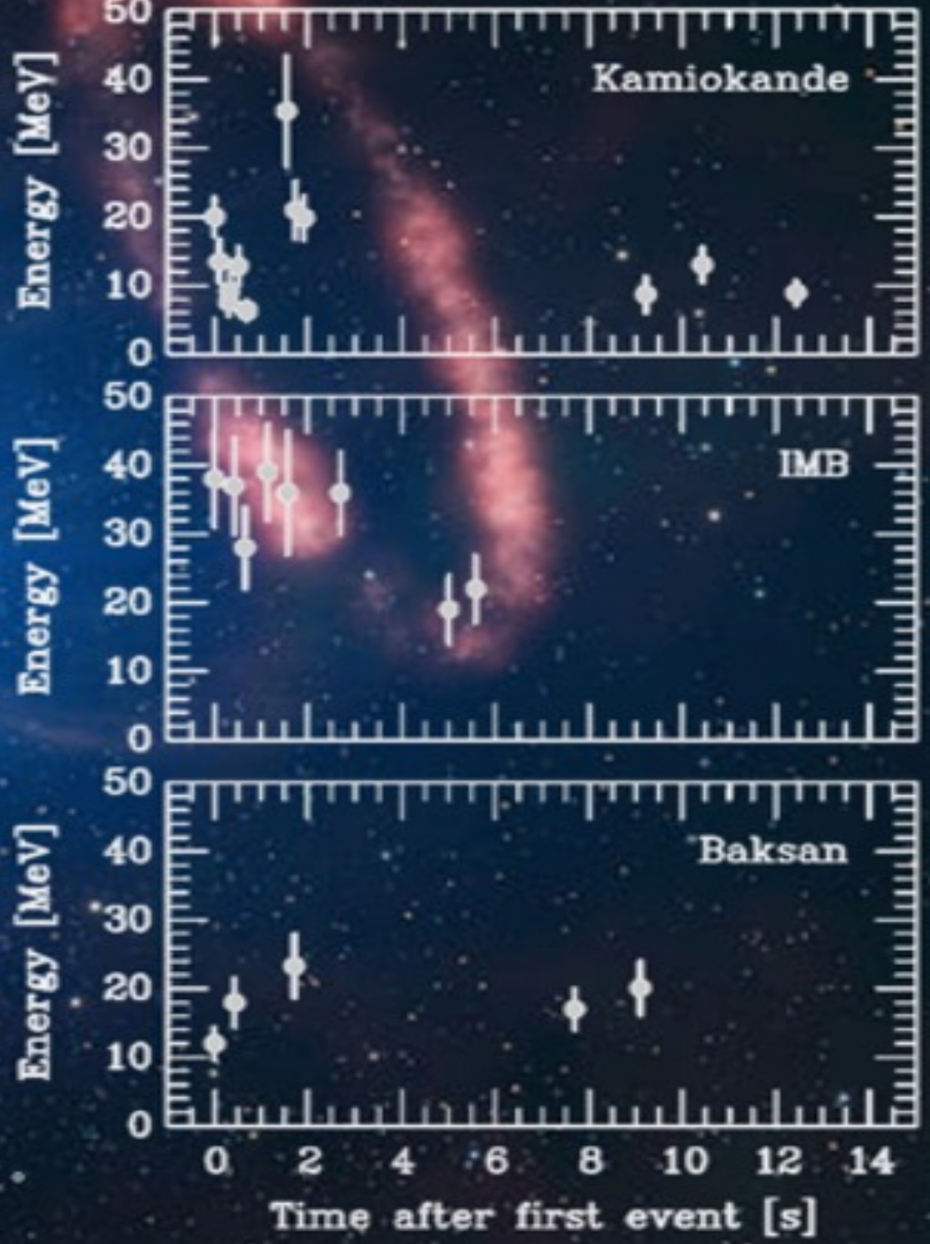
A machine learning approach to discover core collapse supernovae

Irene Di Palma
Sapienza University of Rome
Irene.DiPalma@roma1.infn.it

Motivations

- In a supernova explosion, GWs are generated in the inner core of the source, so that this messenger carries direct information of the inner mechanism.
- Although the phenomenon is among of the most energetic in the universe, the amplitude of the gravitational wave impinging on a detector on the Earth is extremely faint.
- For a CCSN in the center of the Milky way, a rare event, we could expect amplitudes of the metric tensor perturbations ranging between $10^{-21} - 10^{-23}$.
- To increase the detection probability we should increase the volume of the universe to be explored and this can be achieved both by decreasing the detector noise and using better performing statistical algorithms.

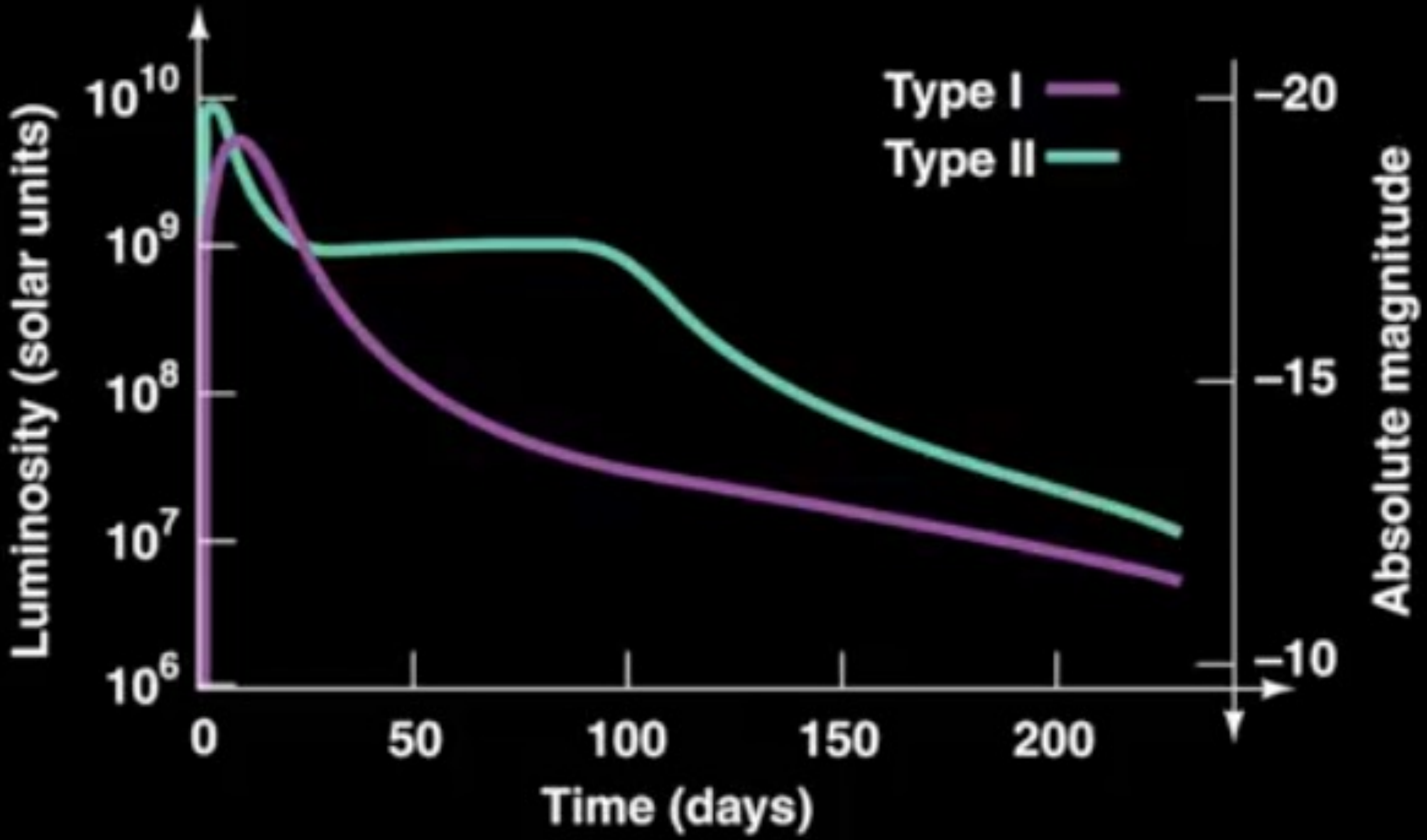
MeV Neutrinos from SN1987A



February 23, 1987.

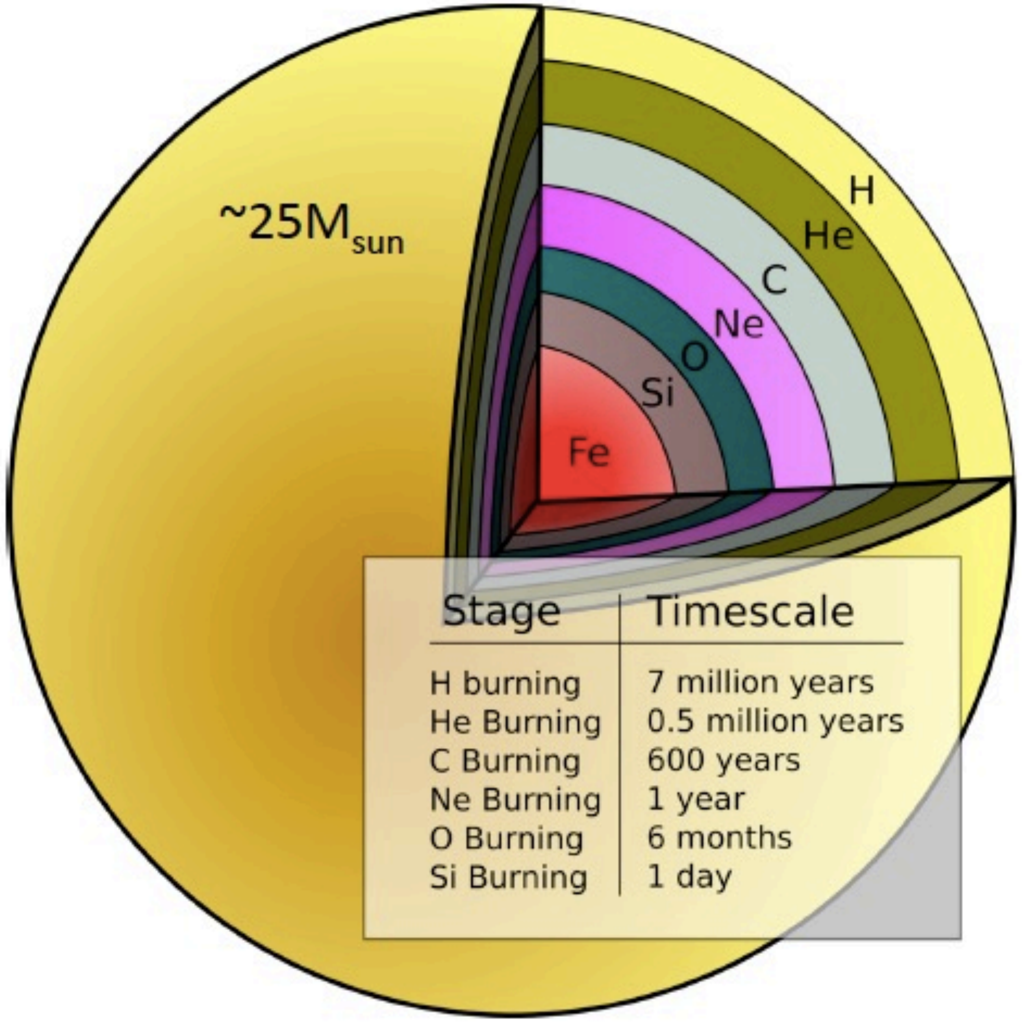
Supernovae

- **Thermonuclear Supernovae: Type Ia**
  - Caused by runaway thermonuclear burning of white dwarf fuel to Nickel
  - Roughly of 10^{51} ergs released
  - Very bright, used as standard candles
  - No remnant
- **Core Collapse Supernovae: Type II, Ib, Ic**
  - Result from the collapse of an iron core in an evolved massive star (M_{ZAMS} >8-10 M_{SUN})
  - Few x 10^{53} ergs released in gravitational collapse, most (99%) radiated in neutrinos
  - Spread stellar evolution elemental products throughout galaxy
  - Neutron star or black hole remnant



Massive Stars: Burning stages

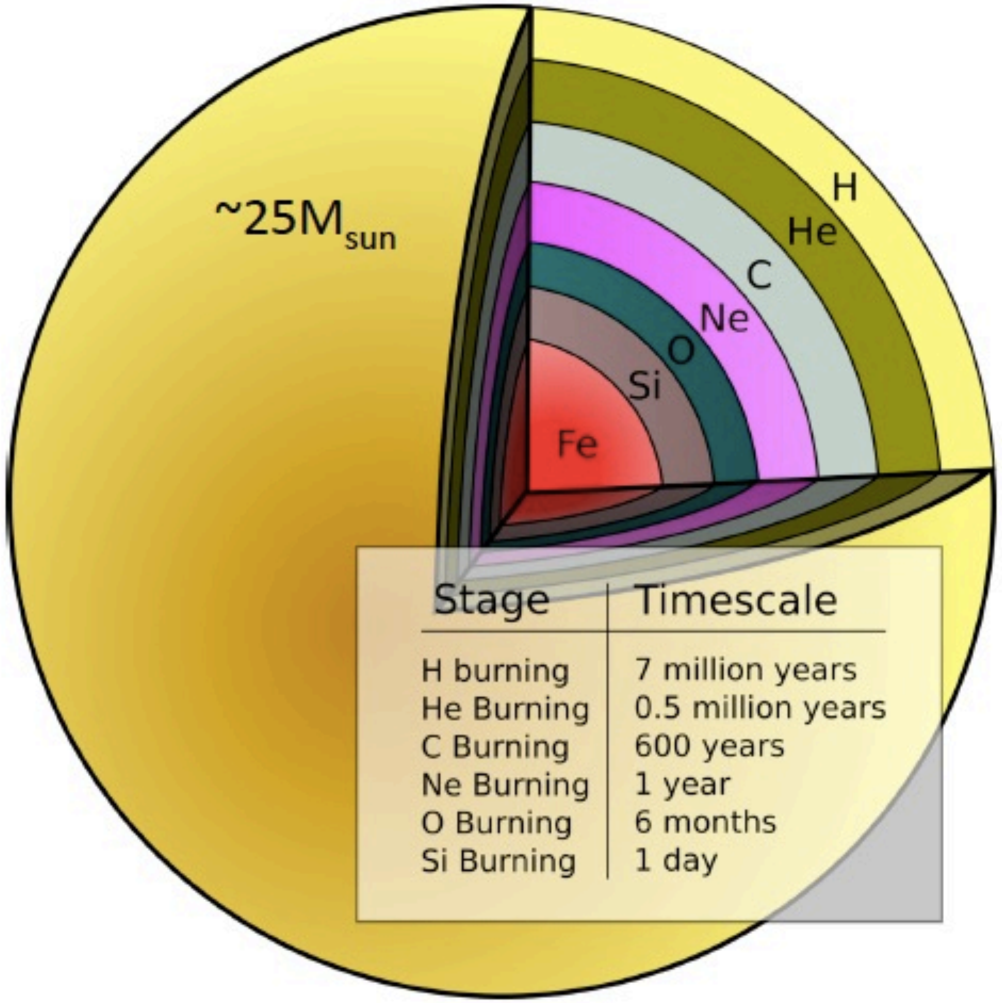
- Stars spend most of their lives burning hydrogen.
- The product – helium – settles in the core and will burn when temperatures increase sufficiently.
- For massive stars ($M > 8\text{-}10M_{sun}$), the process continues through carbon, oxygen, … , up to iron.
- This process does not continue past iron as iron is one of the most tightly bound nuclei.
- Iron core builds up in center of star.



A. C. Phillips, *The Physics of Stars*, 2nd Edition (Wiley, 1999).

Massive Stars: End Stage

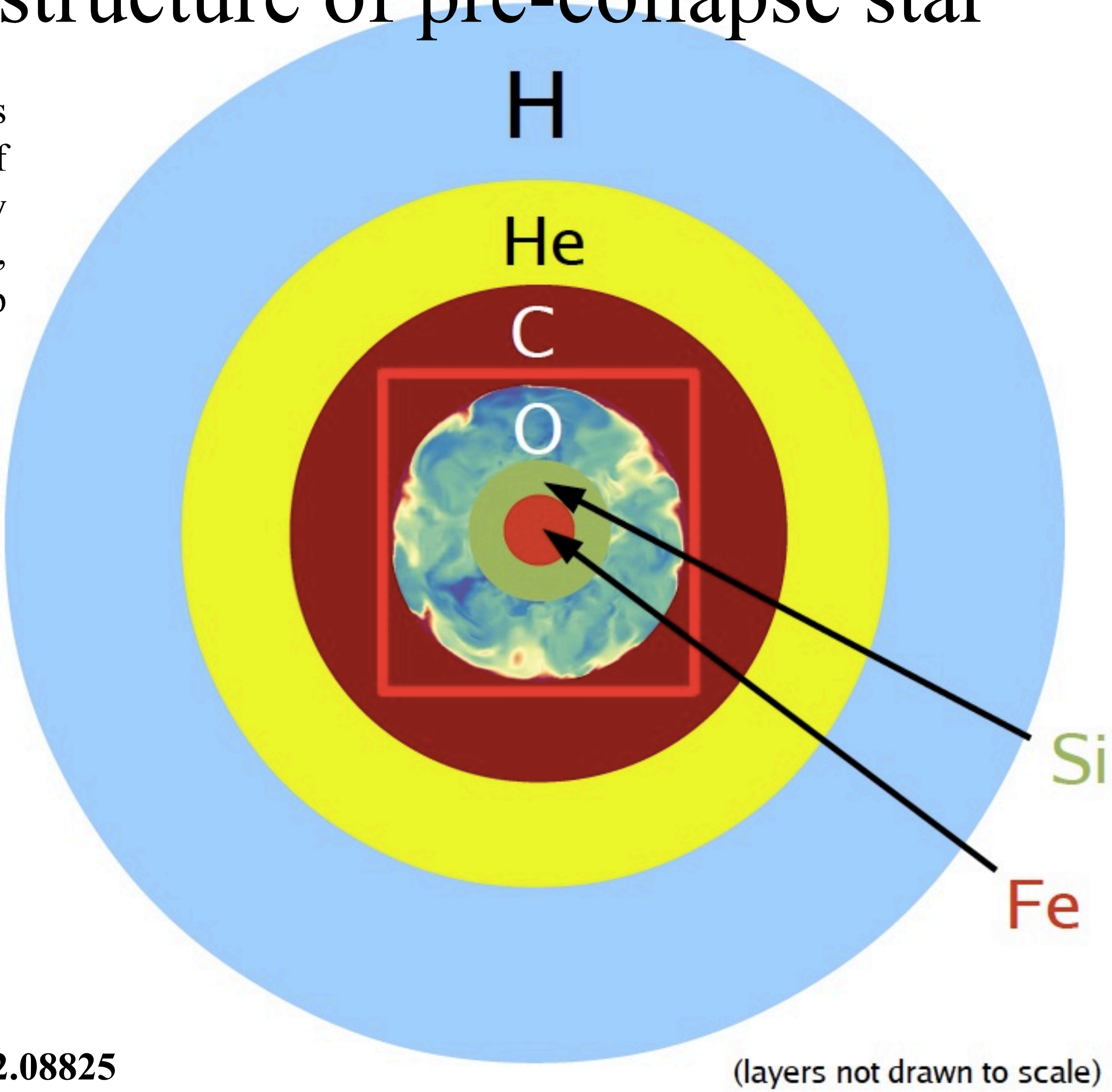
- Stars are, for the majority of the time, in hydrostatic equilibrium because the radiation pressure of the photons from nuclear reactions balance gravity.
- Iron cores however are supported by electron degeneracy pressure, much like a white dwarf, there is a maximum mass that electron degeneracy pressure can support.



A. C. Phillips, *The Physics of Stars*, 2nd Edition (Wiley, 1999).

Onion shell structure of pre-collapse star

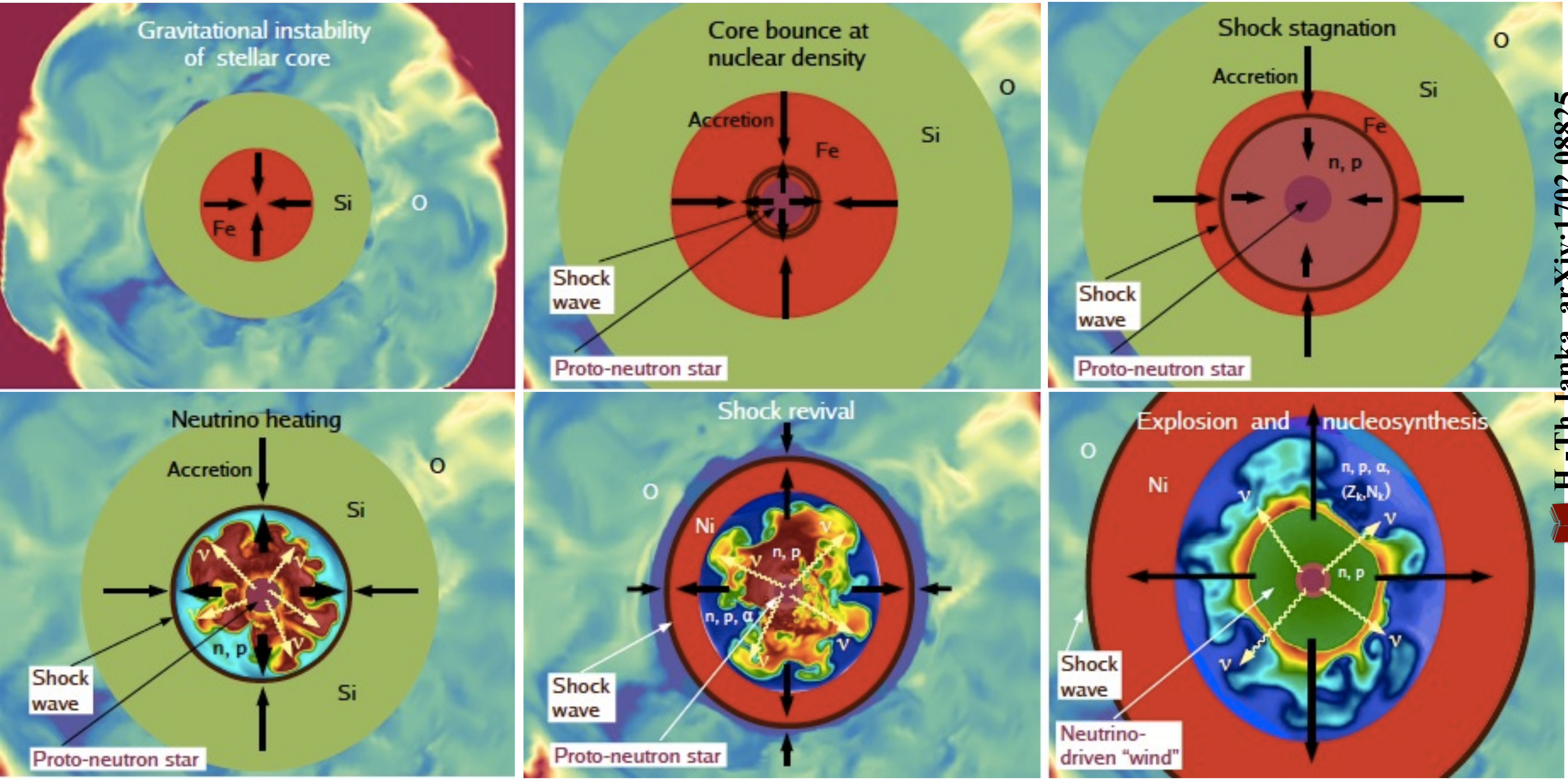
Shells of progressively heavier elements contain the ashes of a sequence of nuclear burning stages, which finally build up a degenerate core of oxygen, neon and magnesium or iron-group elements at the center.



Convective burning can lead to large scale velocity and density perturbations in the oxygen and silicon layers (as indicated for the O-shell).

H.-Th Janka, arXiv:1702.08825

Dynamical phases of stellar core collapse and explosion



H.-Th Janka, arXiv:1702.08825

Before and after the collapse

At the start of the Iron Core Collapse, the core properties are:

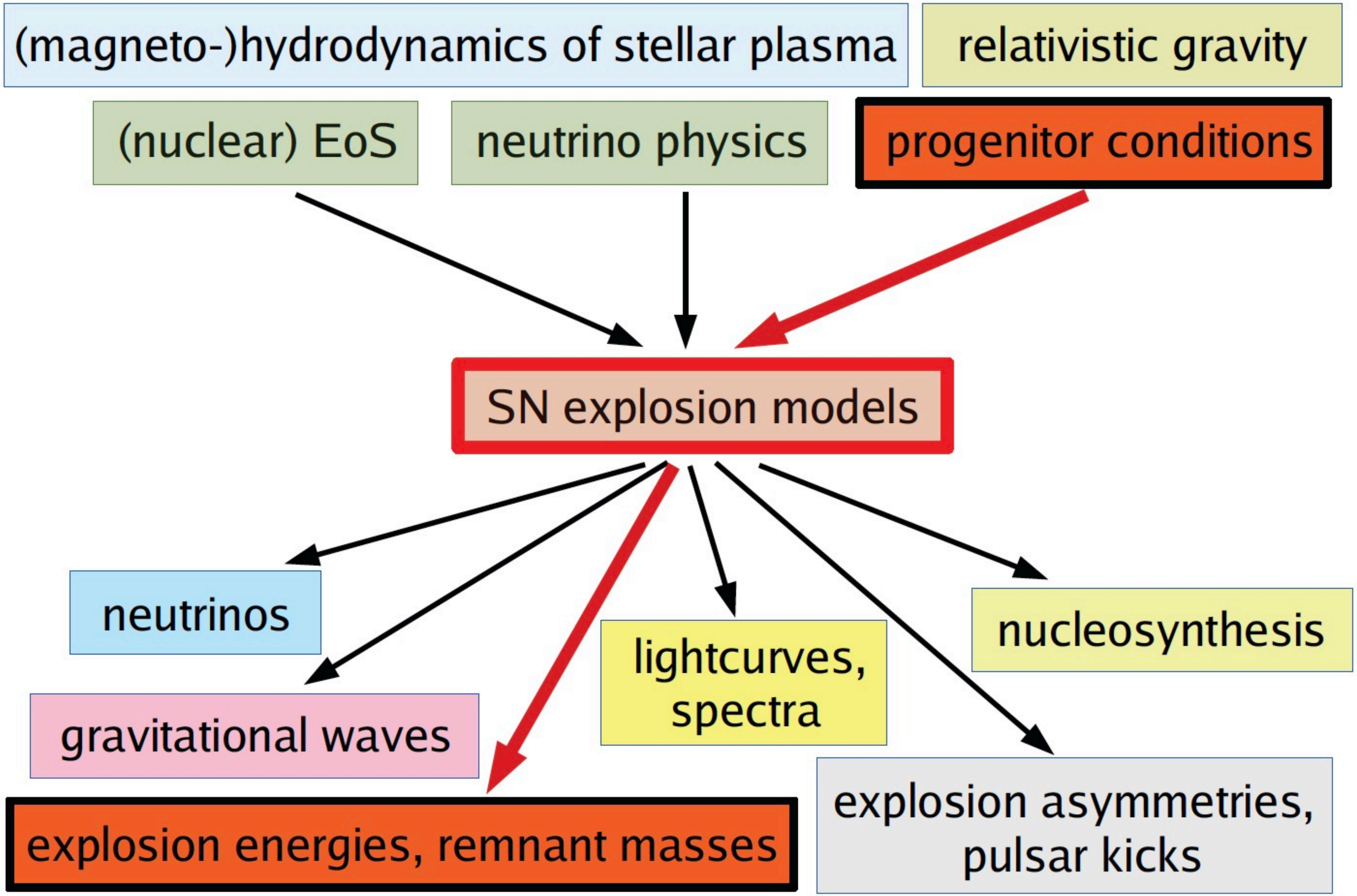
- Radius: ~6000 km (about the size of Earth)
- Density: 100,000,000 grams per cubic centimeter

ONE second later, the properties are:

- Radius: ~30-50 km
- Density: 100,000,000,000,000 grams per cubic centimeter

HTTP://CHANDRA.SI.EDU

Predictions of Signals from Supernovae



A new gravitational wave signature from standing accretion shock instabilities in supernovae

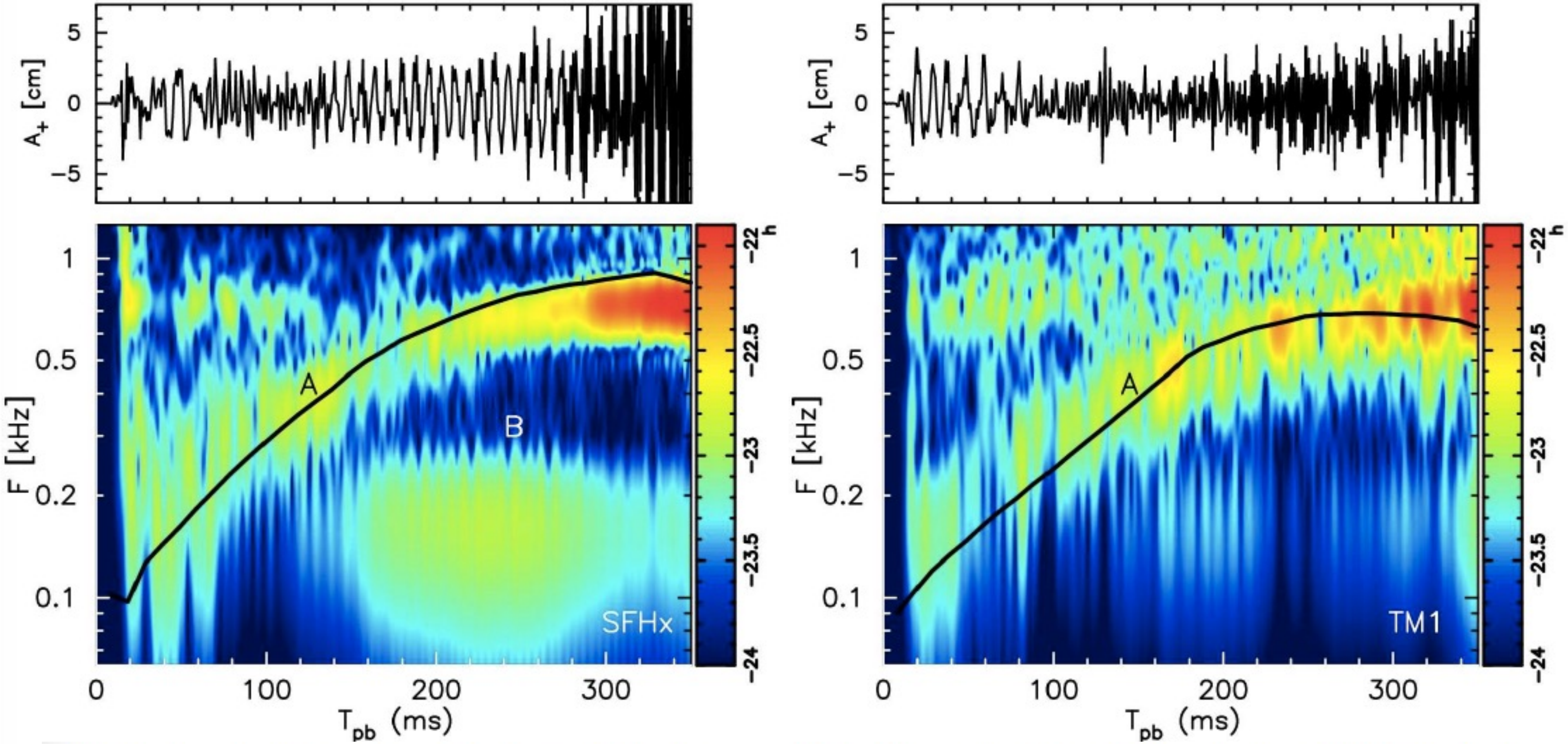
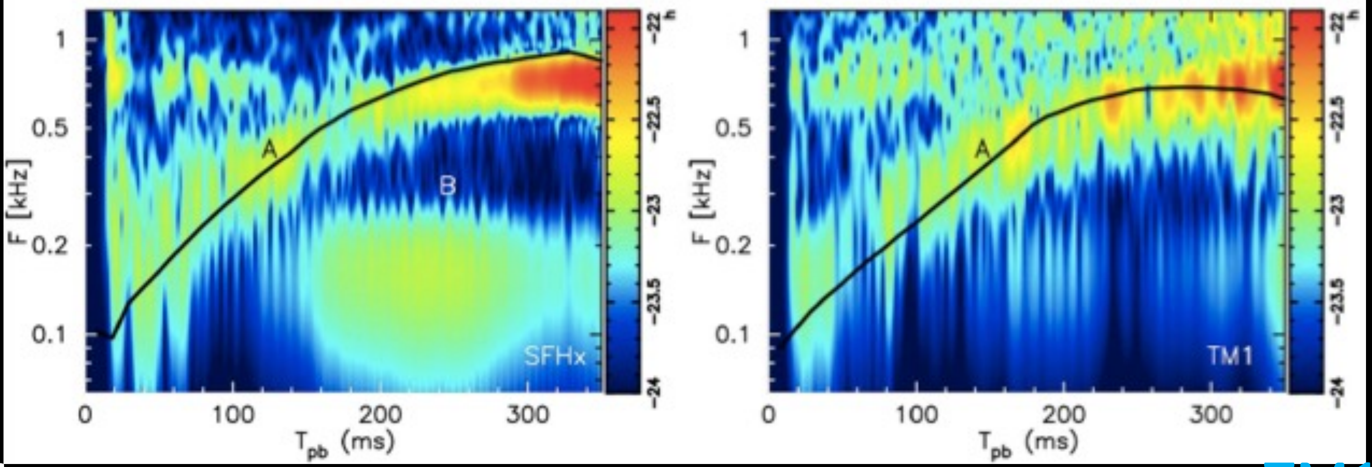


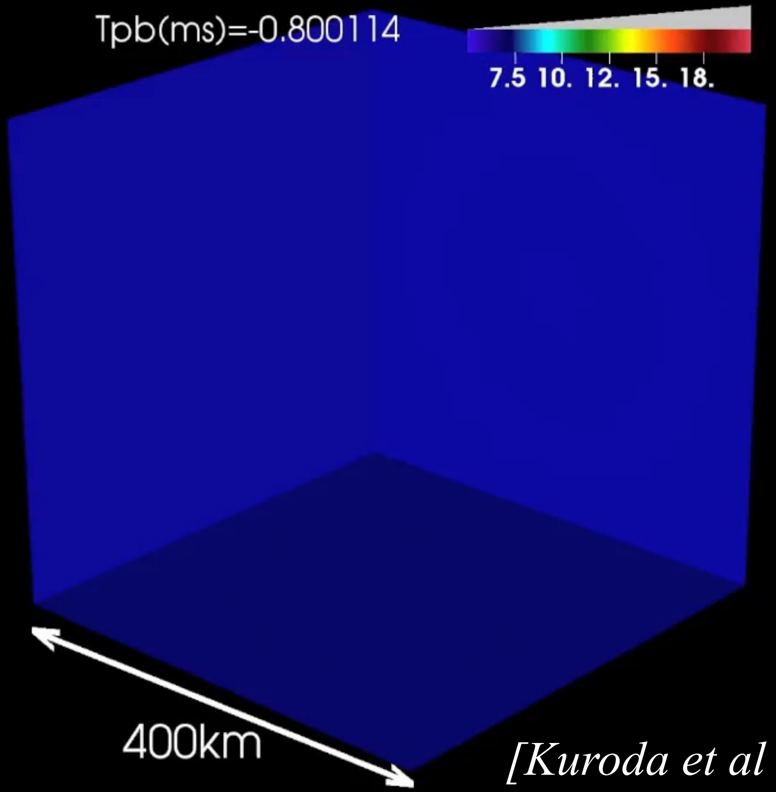
FIG. 1.— In each set of panels, we plot, top; gravitational wave amplitude of plus mode A_+ [cm], bottom; the characteristic wave strain in frequency-time domain $\tilde{h}$ in a logarithmic scale which is over plotted by the expected peak frequency $F_{\rm peak}$ (black line denoted by "A"). "B" indicates the low frequency component. The component "A" is originated from the PNS g-mode oscillation (Marek & Janka 2009; Müller et al. 2013). The component "B" is considered to be associated with the SASI activities (see Sec. 3). Left and right panels are for TM1 and SFHx, respectively. We mention that SFHx (left) and TM1 (right) are softer and stiffer EoS models, respectively.

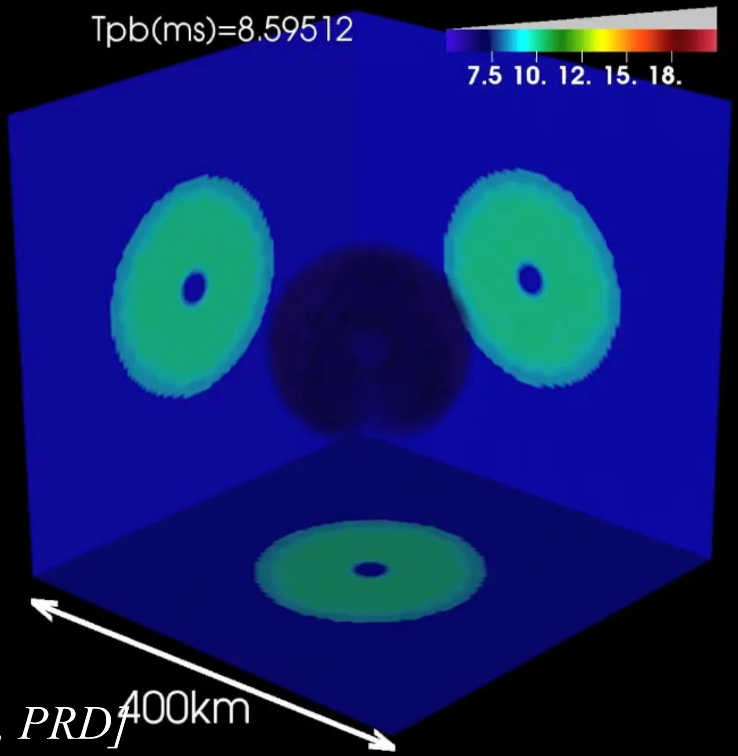
T. Kuroda et al., Astrophys.J. 829 (2016) no.1, L14



SFHx :softer

TM1 :stiffer



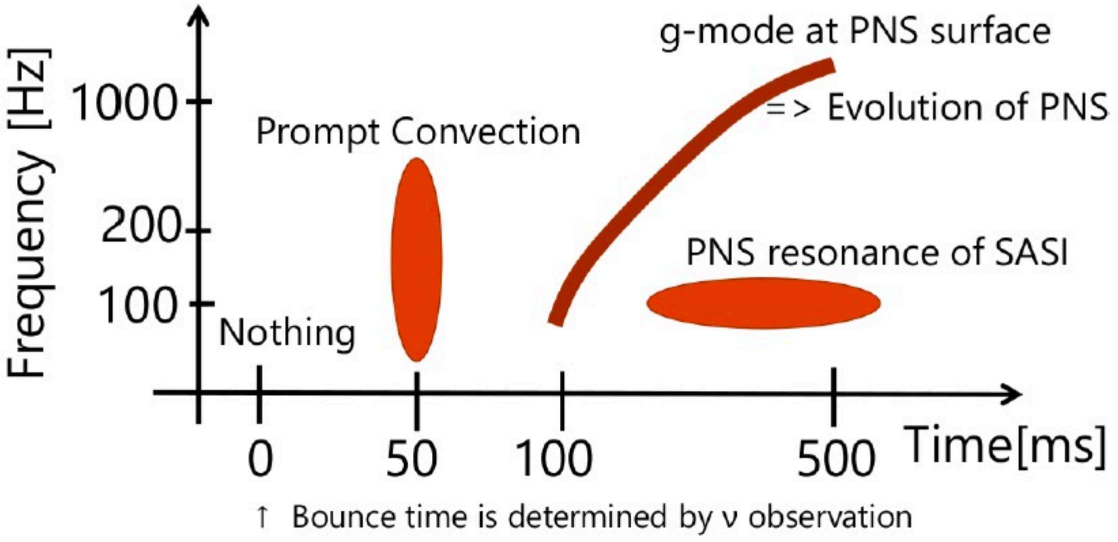


[Kuroda et al 2016, ApJL, 2014, PRD]

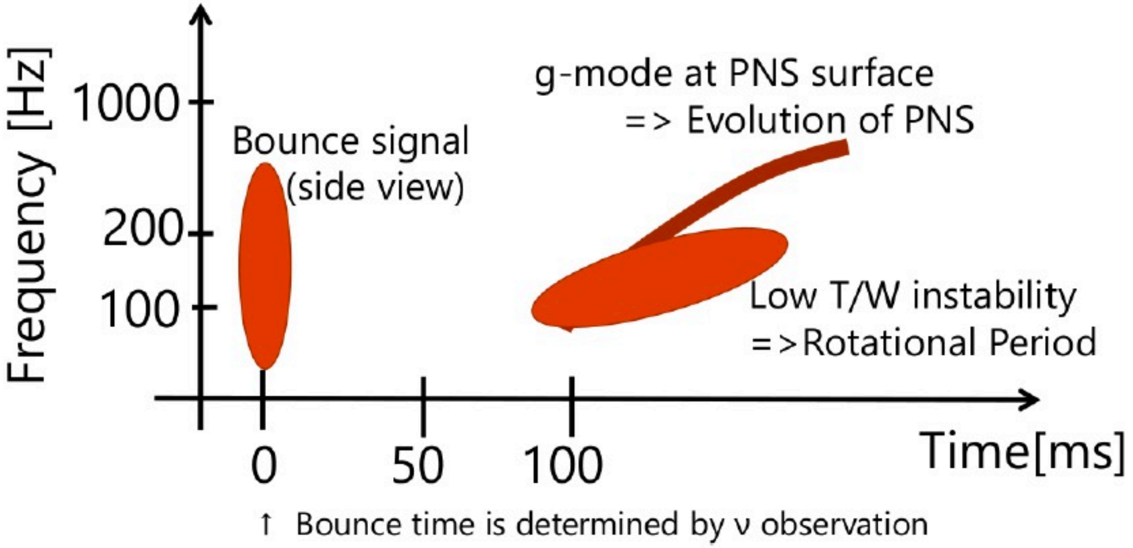
✓ **SASI activity higher for softer EOS** (due to high growth rate, e.g., Foglizzo et al. ('06)).

Different scenarios

Non rotating scenario

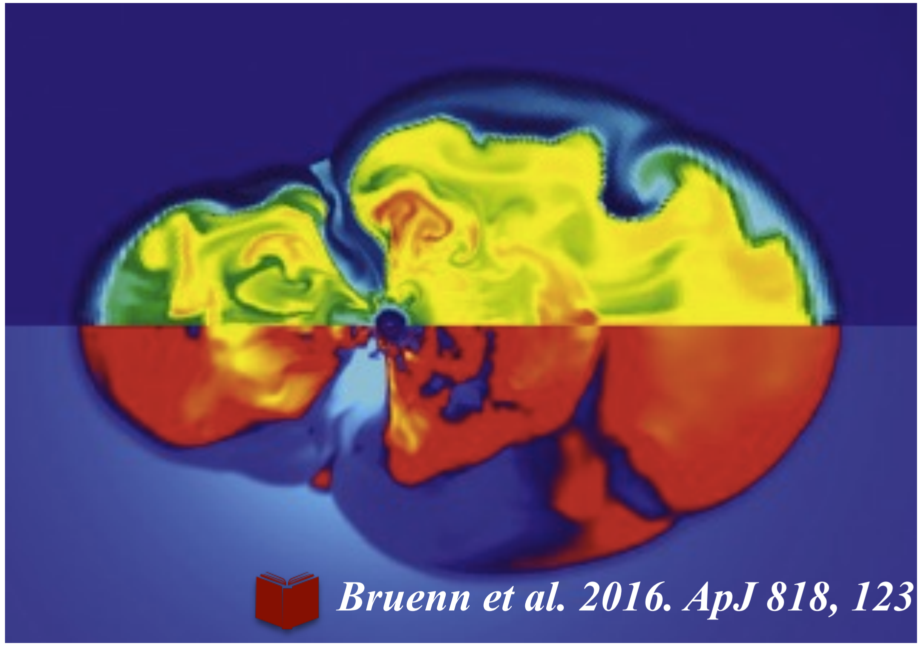


Rapidly rotating scenario

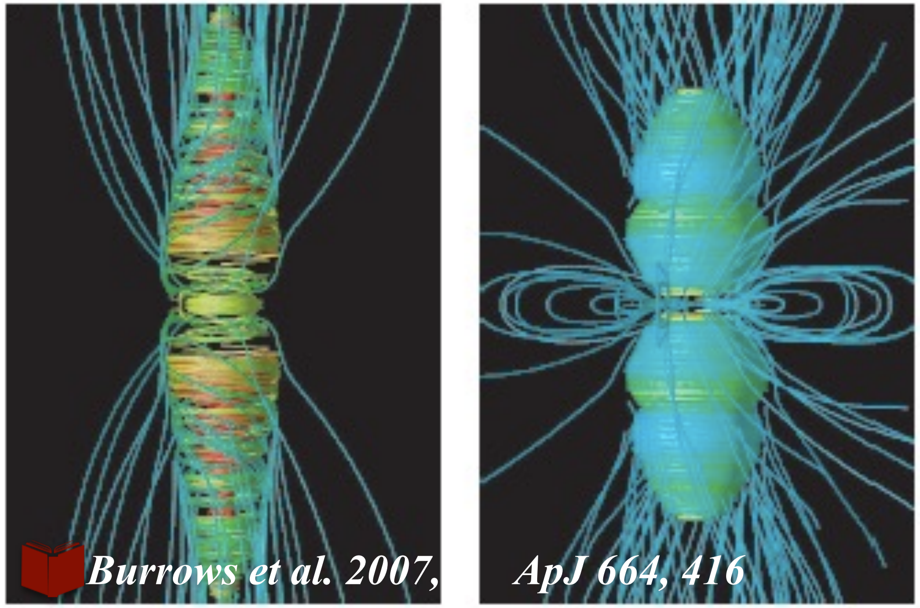


Credit: Tomoya Takiwaki

Neutrino driven CCSNe



Magneto-rotationally-driven CCSNe



Phenomenological Waveforms

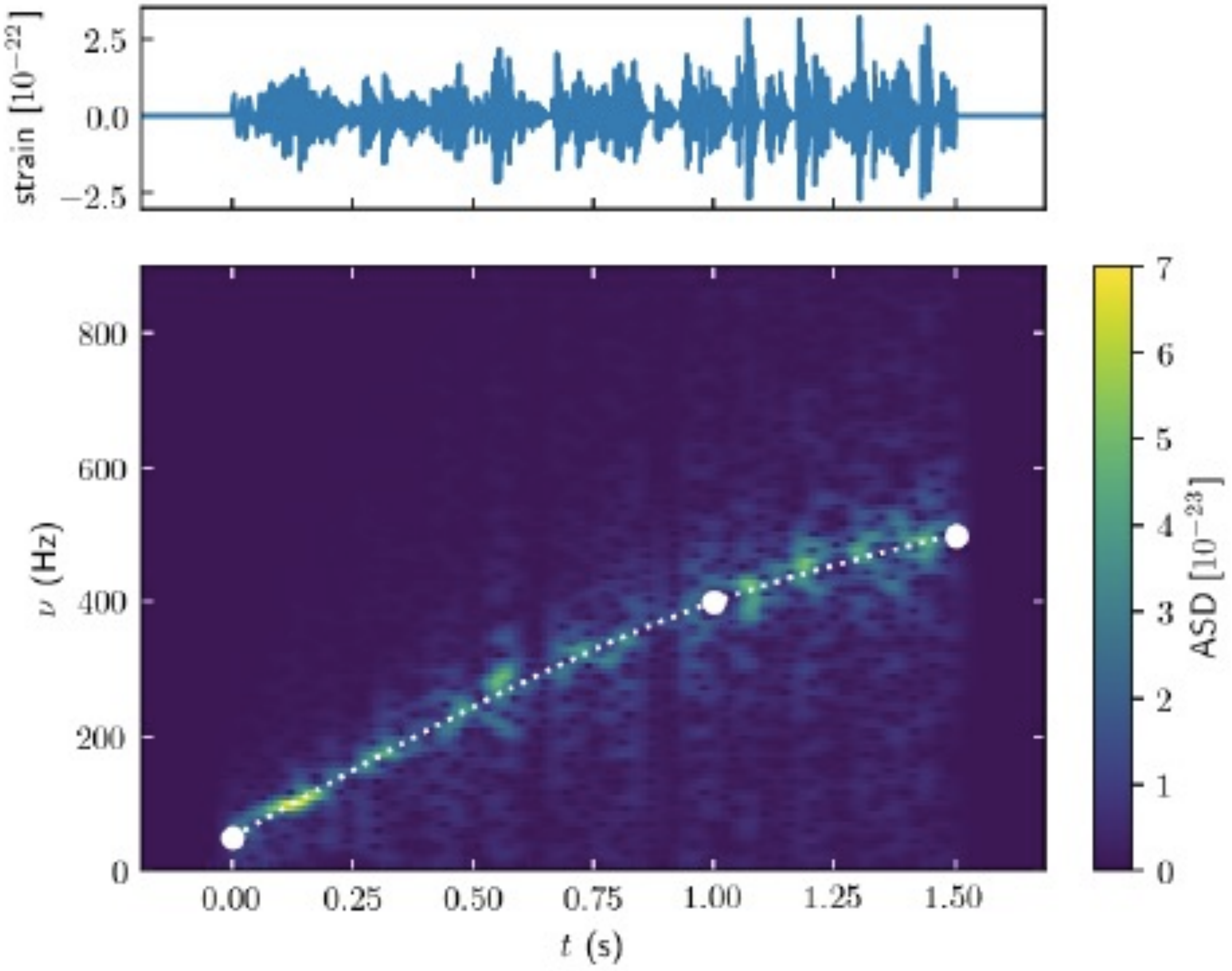
- The aim of our phenomenological template is to mimic the raising arch observed in core-collapse simulations.

- The idea is that at each time in the post-bounce evolution, the PNS is in quasi-hydrostatic equilibrium and any perturbation will excite the eigenmodes of the system, in particular g-modes.

- These modes are continually being excited by the hot bubble surrounding the PNS, in particular by convective motions and SASI. At the same time these excited modes are damped by the PNS conditions (e.g. by the existence of convective layers that do not allow for buoyantly supported waves) and by the presence of non-linearities and instabilities (e.g. turbulence).

- The GW emission can be modelled as a damped harmonic oscillator with a random forcing, in which the frequency varies with time.

Phenomenological Waveforms



parameter	min.	max.	Δ	description
t_{ini} [s]	0	0.2	0.1	beginning of the waveform
t_{end} [s]	0.2	1.5	0.1	end of the waveform
ν_0 [Hz]	50	150	50	frequency at bounce
ν_1 [Hz]	1000	2000	500	frequency at 1 s
ν_2 [Hz]	1500	4500	1000	frequency at 1.5 s
ν_{driver} [Hz]	100	200	100	driver frequency
Q	$(1, 5, 10)$			quality factor
D [kpc]	$(1, 2, 5, 10, 15)$			distance to source

- New and more flexible parametrisation for the frequency evolution.
- The distance is used as a parameter.

Phys.Rev.D 103 (2021) 6, 063011

Gravitational Wave Observatories

LIGO, Livingston, LA

LIGO, Hanford, WA

Virgo, Cascina, Italy

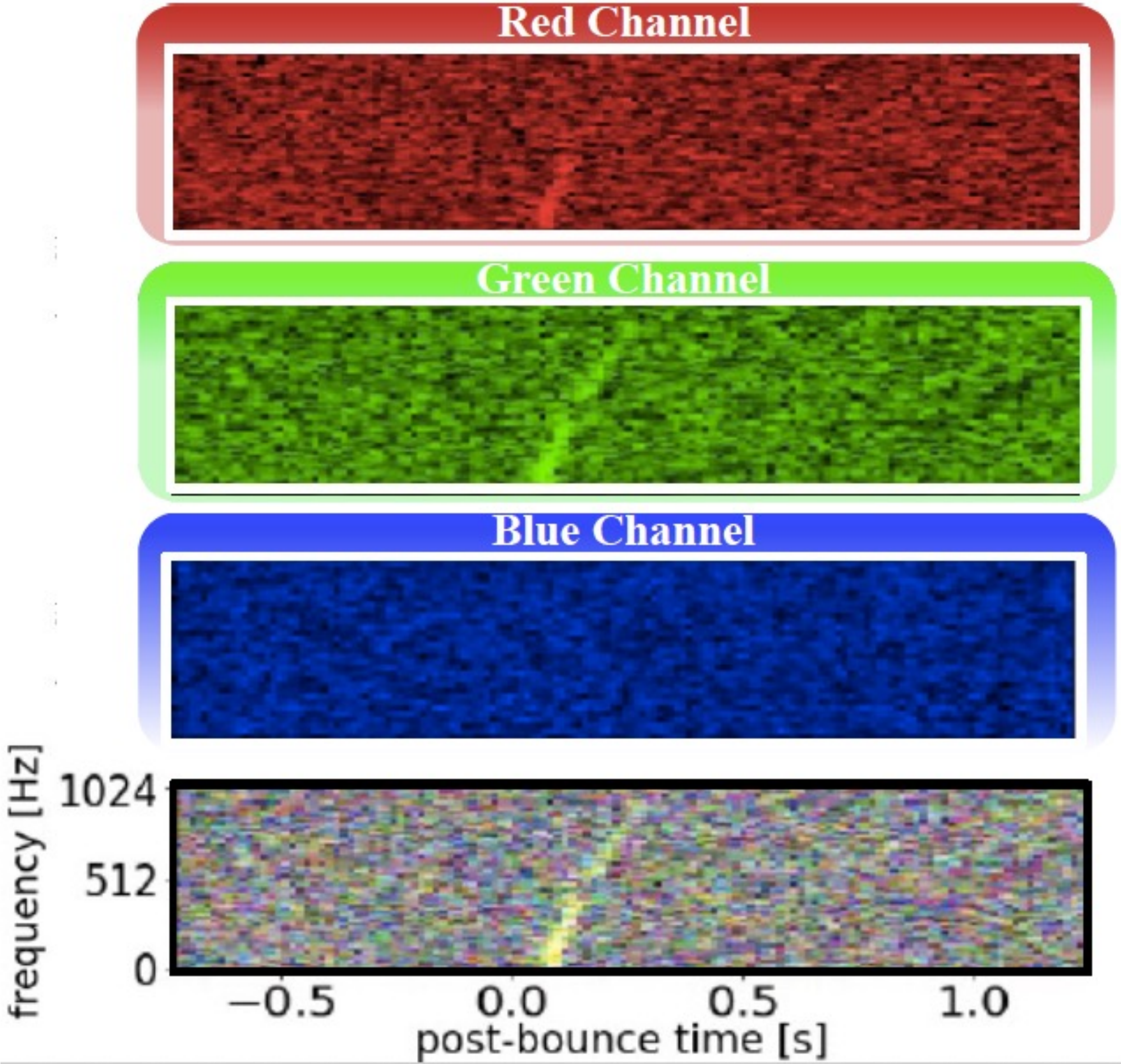
KAGRA, Gifu, Japan

Aim of our Convolutional Neural Network

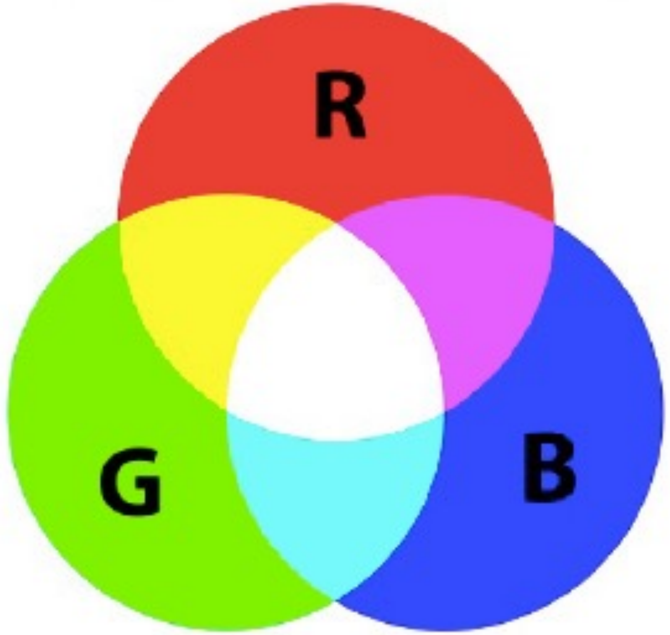
- We want to perform signal detection as an image recognition task, classifying the images in two classes: *Real detector Noise* and *Signal+Real detector Noise.*
- The input images are the RGB multi-detector scalograms.
- The aim is to build a pipeline for a data-driven weakly-modelled robust search.
- Our RGB approach allows us to straightforwardly exploit coincidences among different detectors.

RGB time-frequency plane

Coincidences among detectors



Additive colour synthesis



LIGO Hanford = red

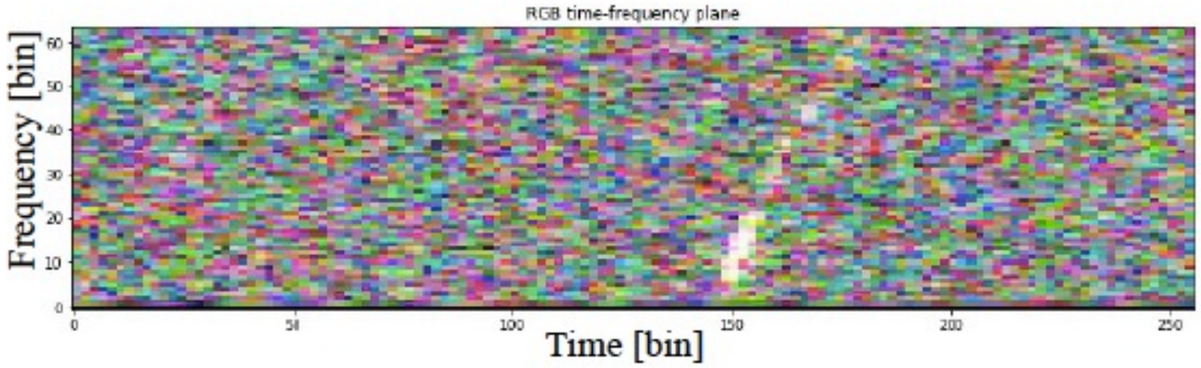
LIGO Livingston = green

Virgo = blue

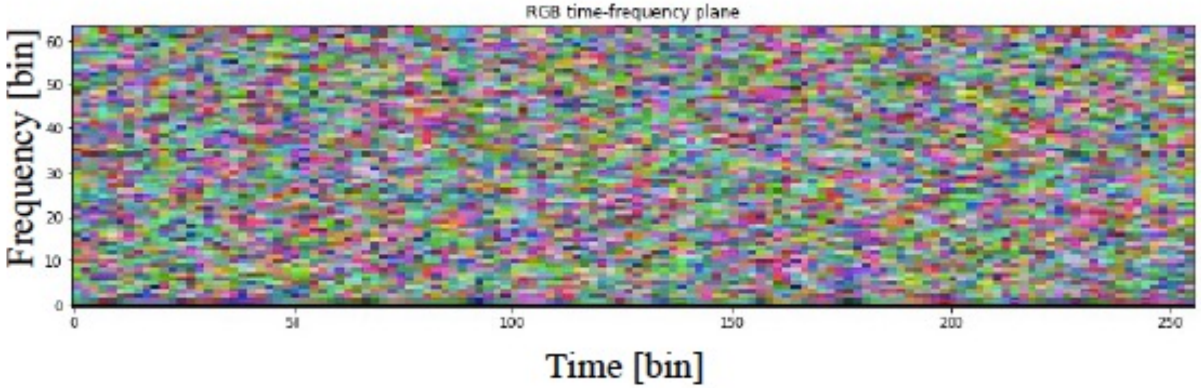
RGB time-frequency plane

Coincidences among detectors

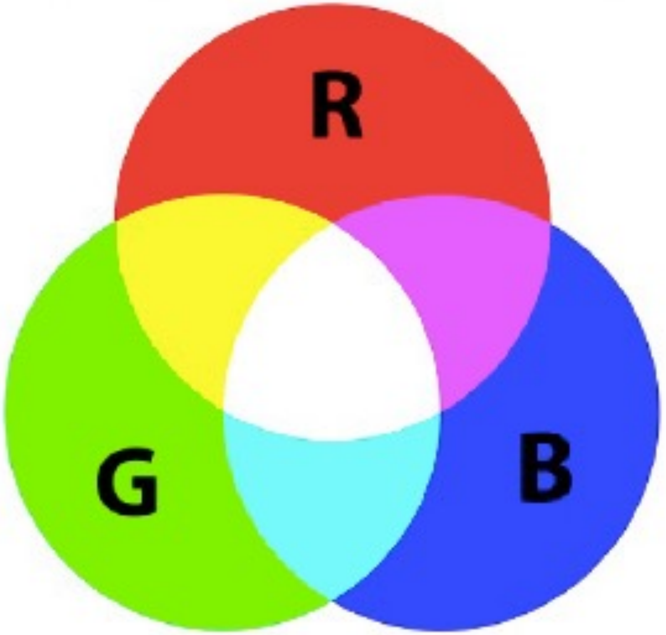
Signal+Noise



Only Noise



Additive colour synthesis



LIGO Hanford = red

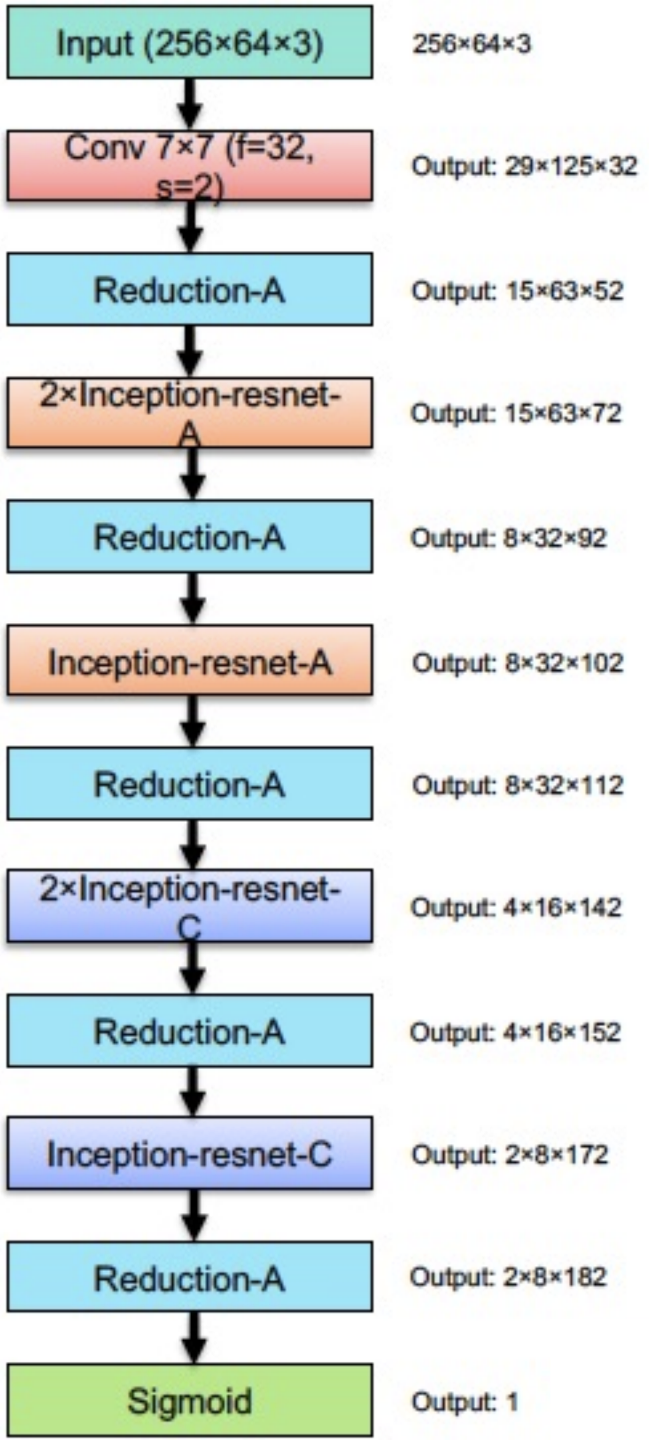
LIGO Livingston = green

Virgo = blue

Phys.Rev. D 98 (2018) 12, 122002

Architecture of the deep learning algorithm

- Mini Inception Resnet v1: reduced version of Inception-Resnet
- Keras framework, based on the TensorFlow backend
- Total number of parameters: 98997
- 30 times more complex than previous network
- The task is treated as a multi-class classification problem with two classes: the event class and the noise class, by using the binary cross entropy.
- The training and validation phase, performed in the real detector noise, is done in 2 h and 21 min using a GPU Nvidia Quadro P5000, while predicting the test set takes 3 ms for each 2 s long image.



Data: from Gaussian noise to real noise

Gaussian noise

(Previous work)

Previous set: 10^4 images for each value of Network SNR ∈ [8,40]

Real noise

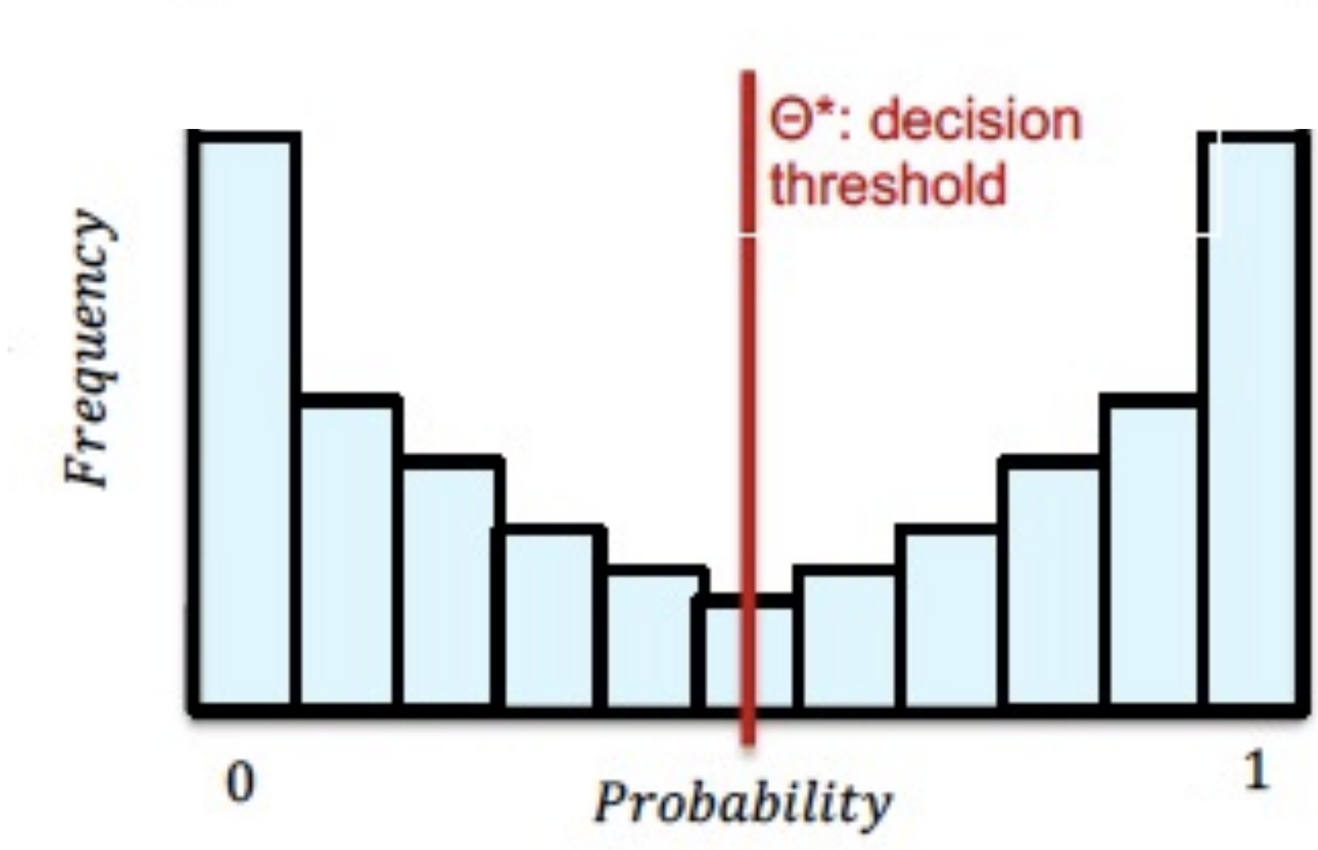
(O2 – August 2017)

- **Training set – phenomenological waveforms:** 7 x 10^4 images for each distance ∈ [0.2, 3] kpc and random sky localisation.
- **Blind set – phenomenological waveforms:** 26 x 10^4 images with distances chosen in a uniform distribution ∈ [0.2, 15] kpc. NOT involved in the training or validation procedure.
- **Test set - numerical simulations from the literature:** 6.5 x 10^4 images with distances ∈ [0.1, 15] kpc

In particular, we chose a stretch of real data even containing glitches, taken during August 2017, when Virgo joined the run. The period includes about 15 days of coincidence time among the three detectors and we used this data set to generate about 2 years of time-shifts data to train and test the neural network as noise class.

Phys.Rev.D 103 (2021) 6, 063011

Measuring and constraining the learning



- The output of the network is a probability vector ϑ, which contains the probabilities of the template belonging to one class or another.
- The classification task is performed according to a threshold ϑ^*, the template will be classified as event class only if its porbability overcomes ϑ^*.

Confusion matrix

		Actual class	
		Event	*Noise*
Predicted class	*Event*	True positive (TP)	False positive (FP)
	Noise	False negative (FN)	True negative (TN)

Efficiency:

$$\eta_{CNN} = \frac{\text{correctly classified signals}}{\text{all the signals at CNN input}} = \frac{TP}{TP+FN}$$

False Alarm Rate:

$$FAR_{CNN} = \frac{\text{misclassified noise}}{\text{all classified events}} = \frac{FP}{FP+TP}$$

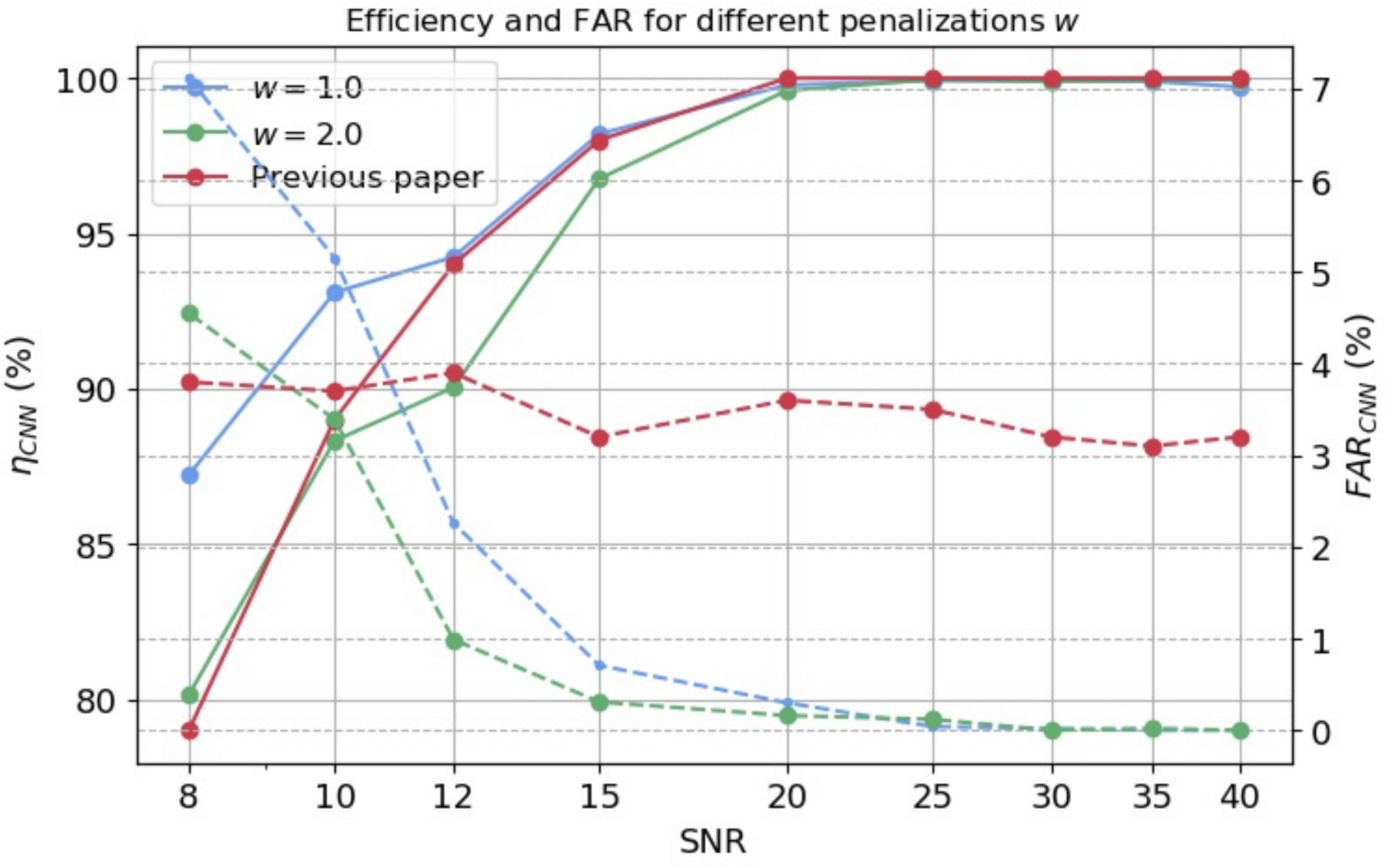
False Positive Rate: $$FPR = \frac{FP}{FP+TN}$$

Comparison with previous work in Gaussian noise

***Weighted binary cross-entropy*:**

w=1 correctly classify the noise class or the event class is the same

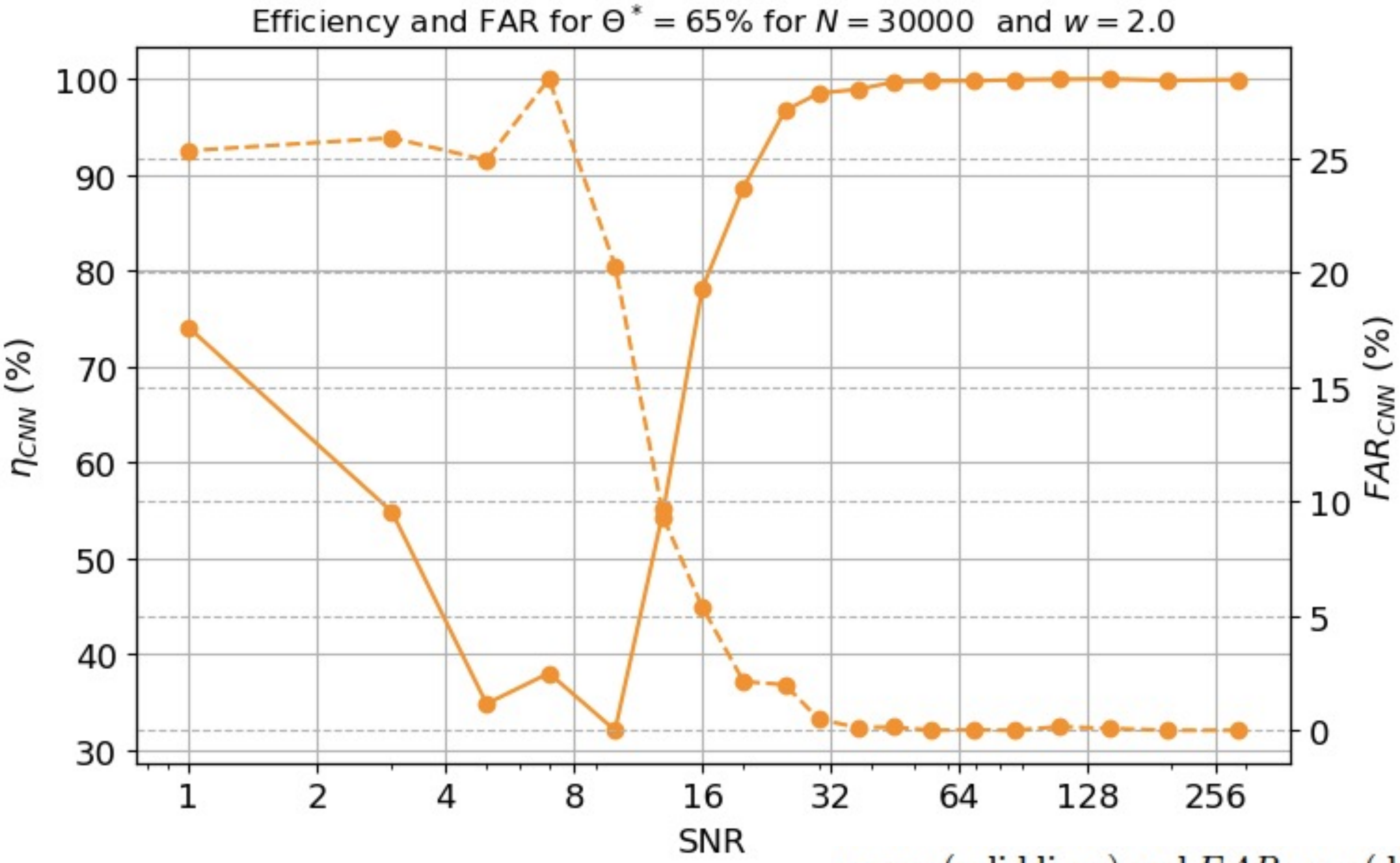
w=2 it is 2 times more important to correctly classify the noise class rather than the event class.



η_{CNN} (solid lines) and FAR_{CNN} (dashed lines)

Phys.Rev. D 103 (2021) 6, 063011

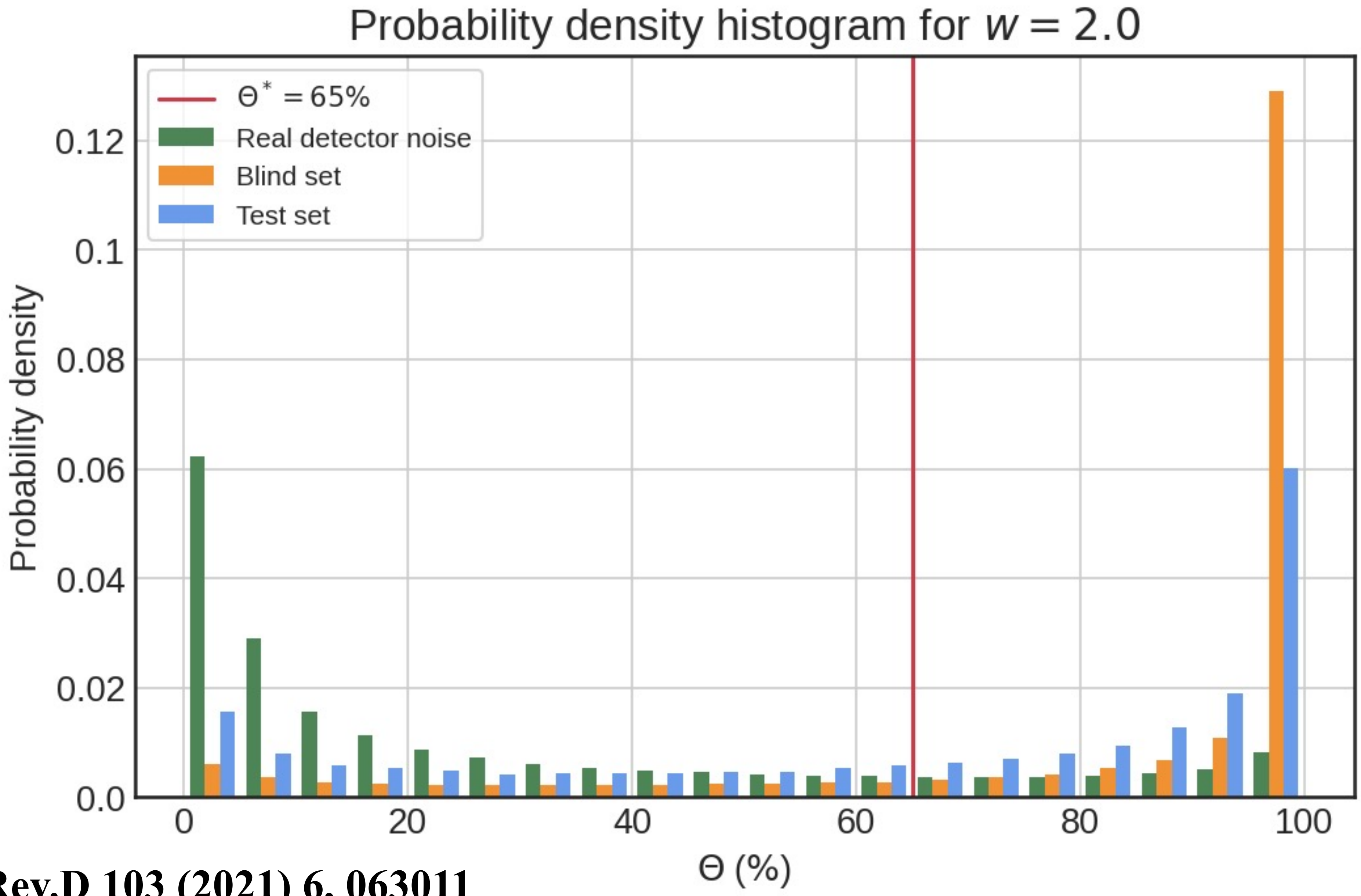
Validation process in real detector noise



η_{CNN} (solid lines) and FAR_{CNN} (dashed lines)

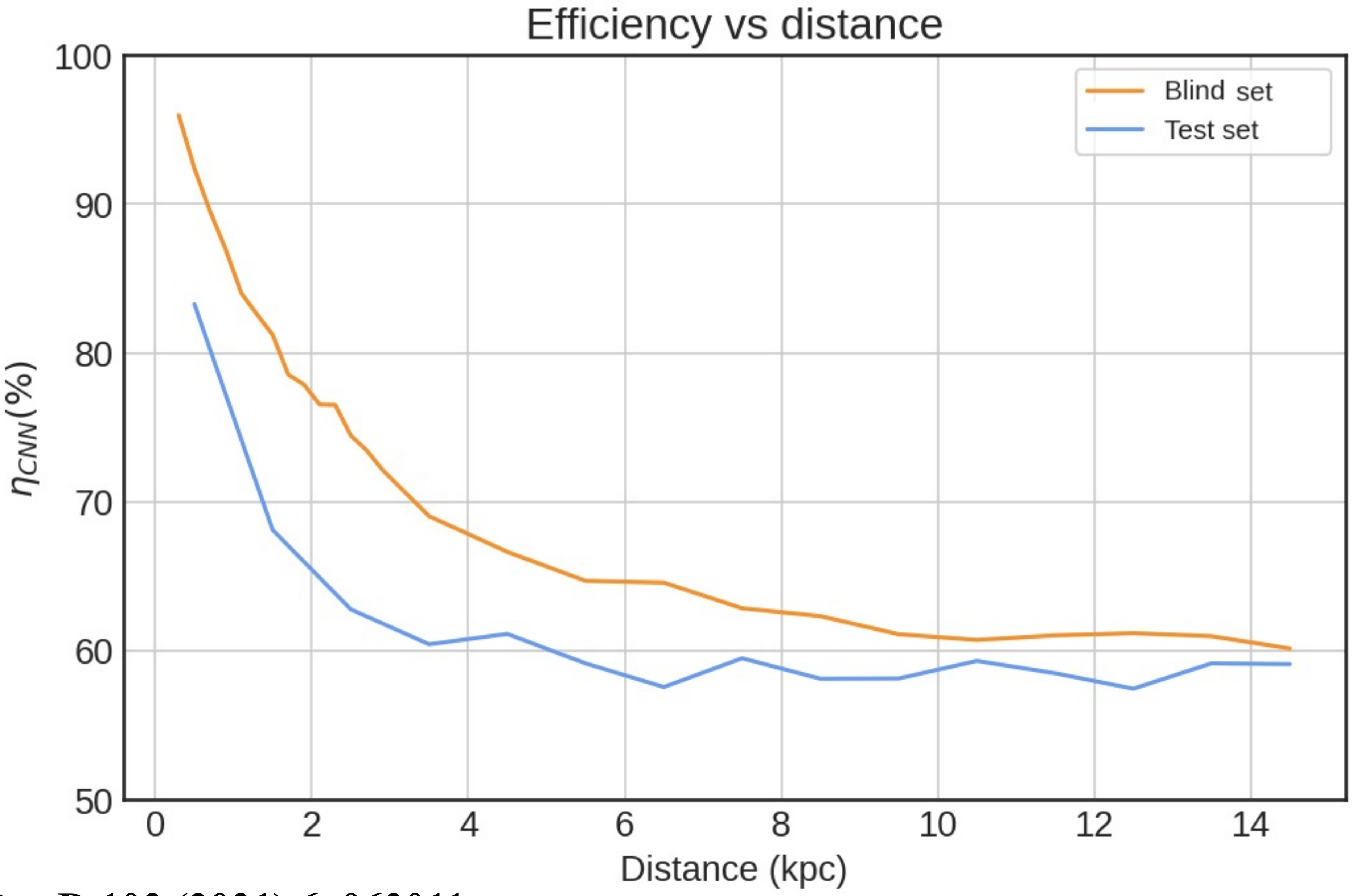
Phys.Rev.D 103 (2021) 6, 063011

Results in real detector noise



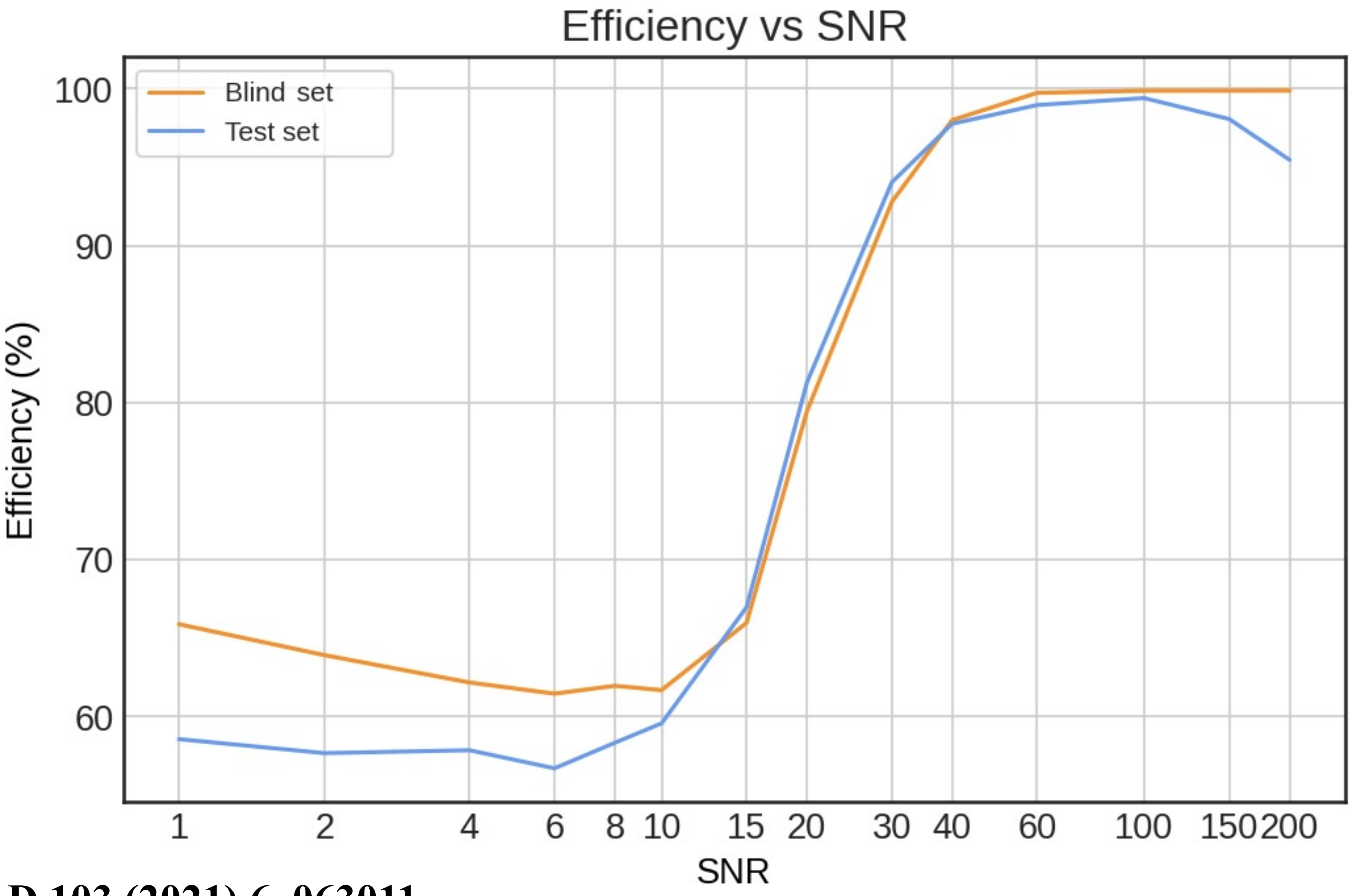
Phys.Rev.D 103 (2021) 6, 063011

Results in real detector noise



Phys.Rev.D 103 (2021) 6, 063011

Results in real detector noise



Phys.Rev.D 103 (2021) 6, 063011

Conclusions

- We trained a newly developed Mini-Inception Resnet neural network using time-frequency images corresponding to injections of simulated phenomenological signals, which mimic the waveforms obtained in 3D numerical simulations of CCSNe.
- We computed the detection efficiency versus the source distance, obtaining that, for signal to noise ratio higher than 15, the detection efficiency is 70 % at a false alarm rate lower than 5%.
- In the case of O2 run, it would have been possible to detect signals emitted at 1 kpc of distance, whilst lowering down the efficiency to 60%, the event distance reaches values up to 14 kpc.
- These results are very promising for future detections and the algorithm has multiple possible extensions.

Waveforms for the test set

TABLE II: List of models of the test set used in the injections. $M_{\rm ZAMS}$ corresponds to the progenitor mass at zero-age in the main sequence (ZAMS). Unless commented, all progenitors have solar metallicity, result in explosions and their GW signal do not show signatures of the standing-shock accretion instability (SASI).

Model name	reference	$M_{\rm ZAMS}$	comments
s9	[47]	$9M_{\odot}$	Low mass progenitor, low GW amplitude.
s25	[47]	$25M_{\odot}$	Develops SASI.
s13	[47]	$13M_{\odot}$	Non-exploding model.
s18	[48]	$18M_{\odot}$	Higher GW amplitude.
he3.5	[48]	-	Ultra-stripped progenitor ($3.5M_{\odot}$ He core).
SFHx	[49]	$15M_{\odot}$	Non-exploding model. Develops SASI.
mesa20	[50]	$20M_{\odot}$	
mesa20_pert	[50]	$20M_{\odot}$	Same as mesa20, but including perturbations.
s11.2	[31]	$11.2M_{\odot}$	
L15	[28]	$15M_{\odot}$	Simplified neutrino treatment.

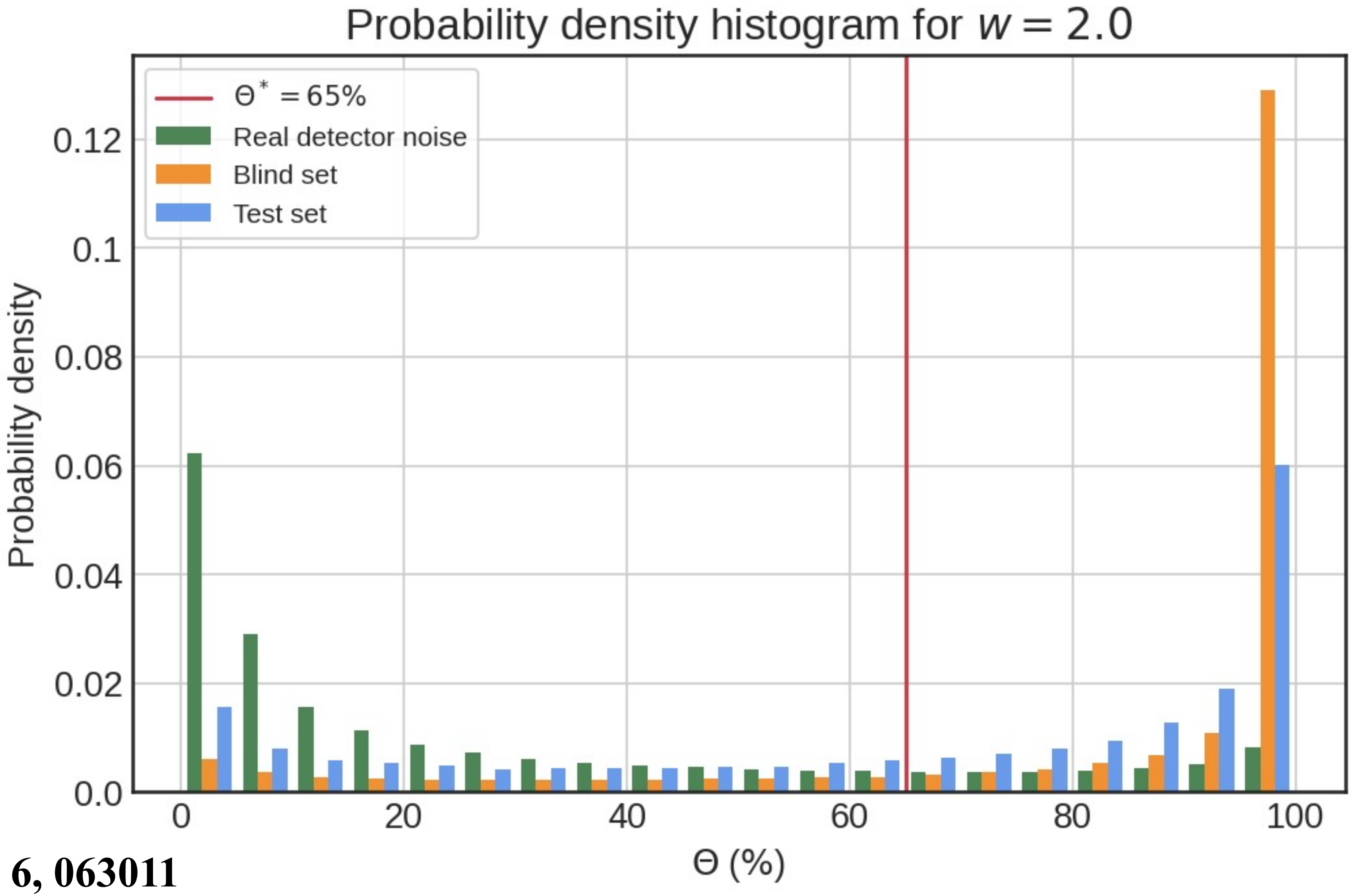
Phys.Rev.D 103 (2021) 6, 063011

Results in real detector noise

Given the counts of the ith bin c_i and its width b_i, we define the probability density as

$$c_i/(\sum_i^N c_i \times b_i),$$

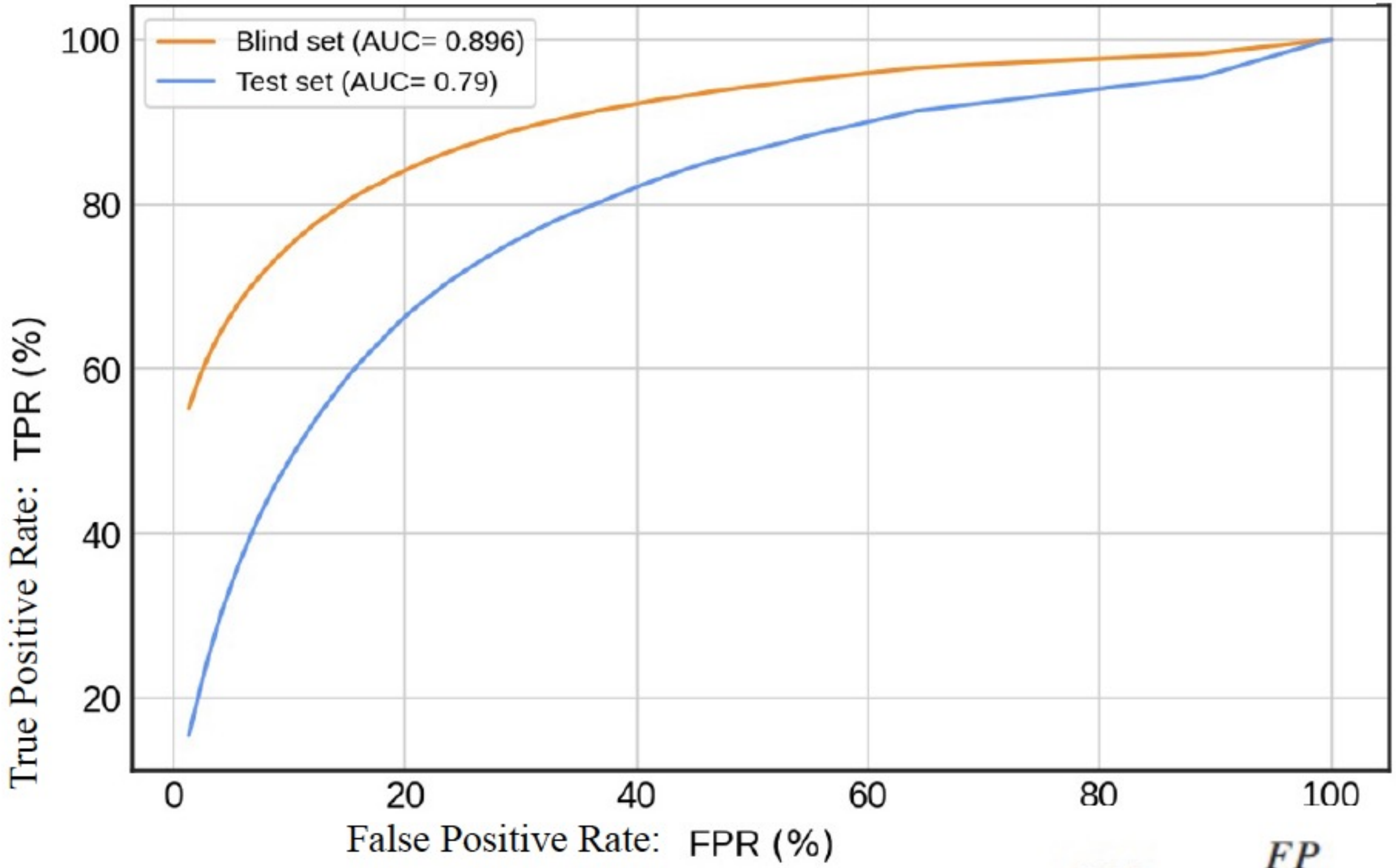
where N is the total number of bins of the histogram.



Phys.Rev.D 103 (2021) 6, 063011

ROC curves

$$\eta_{CNN} = \frac{\text{correctly classified signals}}{\text{all the signals at the input of CNN}} = \text{TPR}$$



$$FPR = \frac{FP}{FP + TN}$$

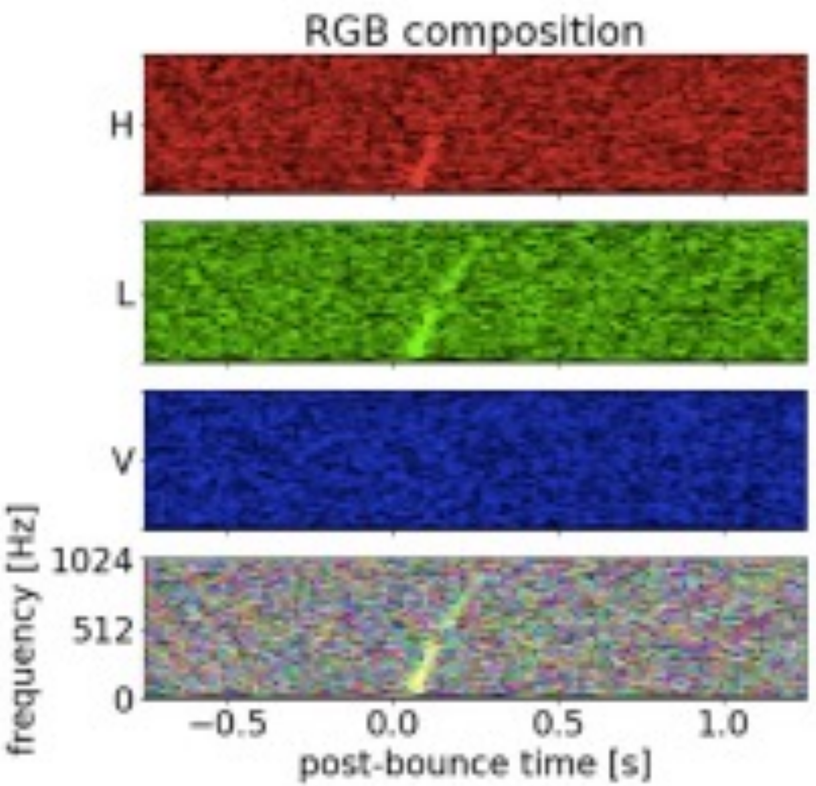
Previous work

Task: classification problem

Classes: 0 class (noise) and 1 class (event) with different level of noise (SNR)

Learning: curriculum learning

Data: Gaussian noise



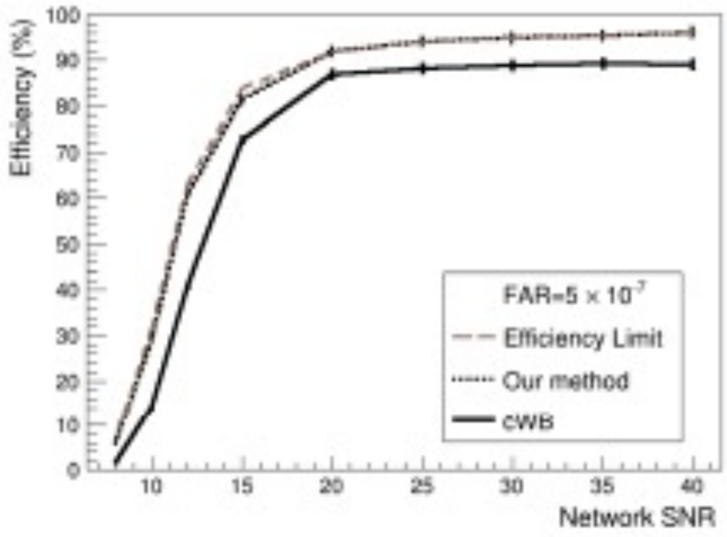


Figure 8: *Efficiency vs SNR in the case of complete cW… (continuous) and our method (dashed) for all SNRs. We re… port also the curve that shows the ratio between the inpu… events of the CNN and the total injected events in function o… SNR (brown). This curve sets the maximum efficiency th… our method can achieve*

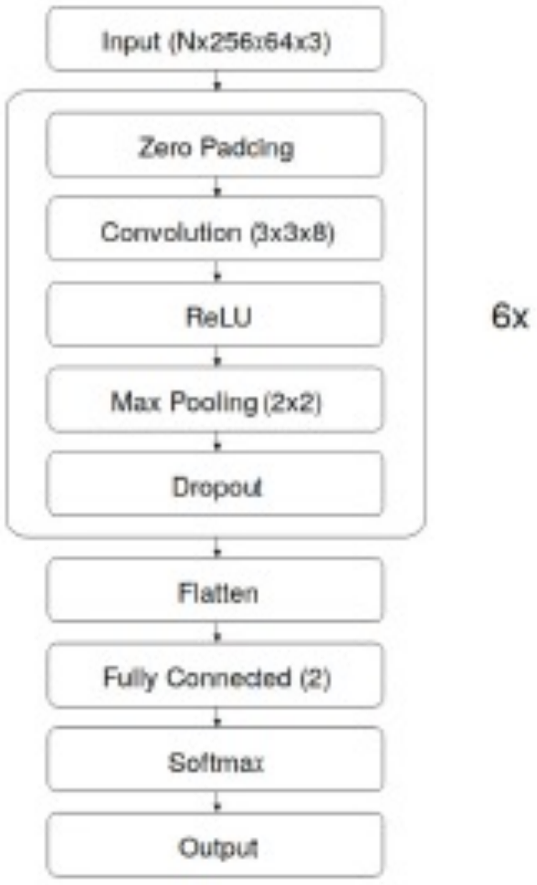


Figure 4: *Sketch of the architecture of our model.*

Phys.Rev. D 98 (2018) 12, 122002