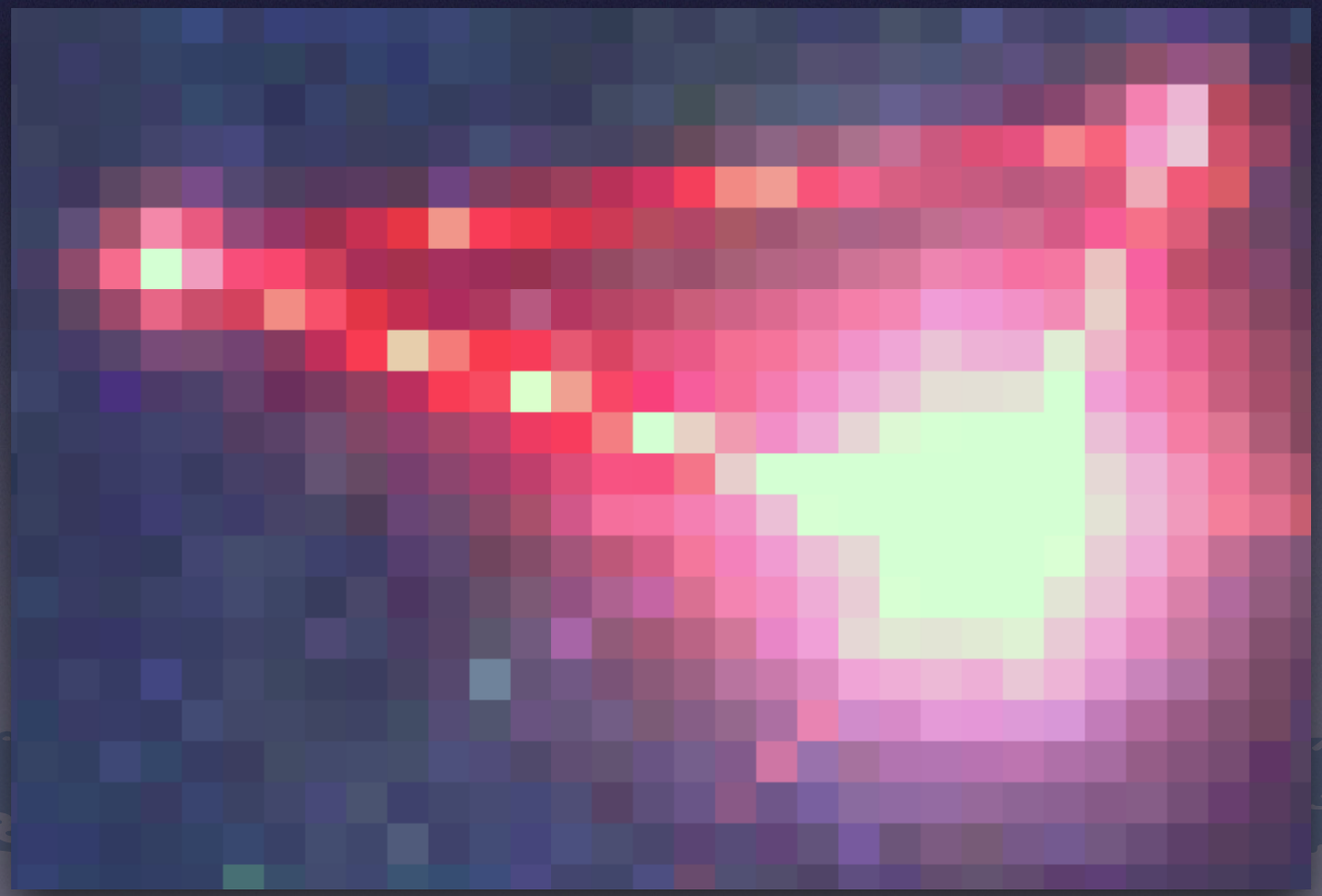
LISA Data Challenges: Status and future prospects

N. Karnesis

on behalf of the LDC team
APC Paris

Gravitational Waves, Black Holes and Fundamental Physics
13-16 January 2020
Trieste

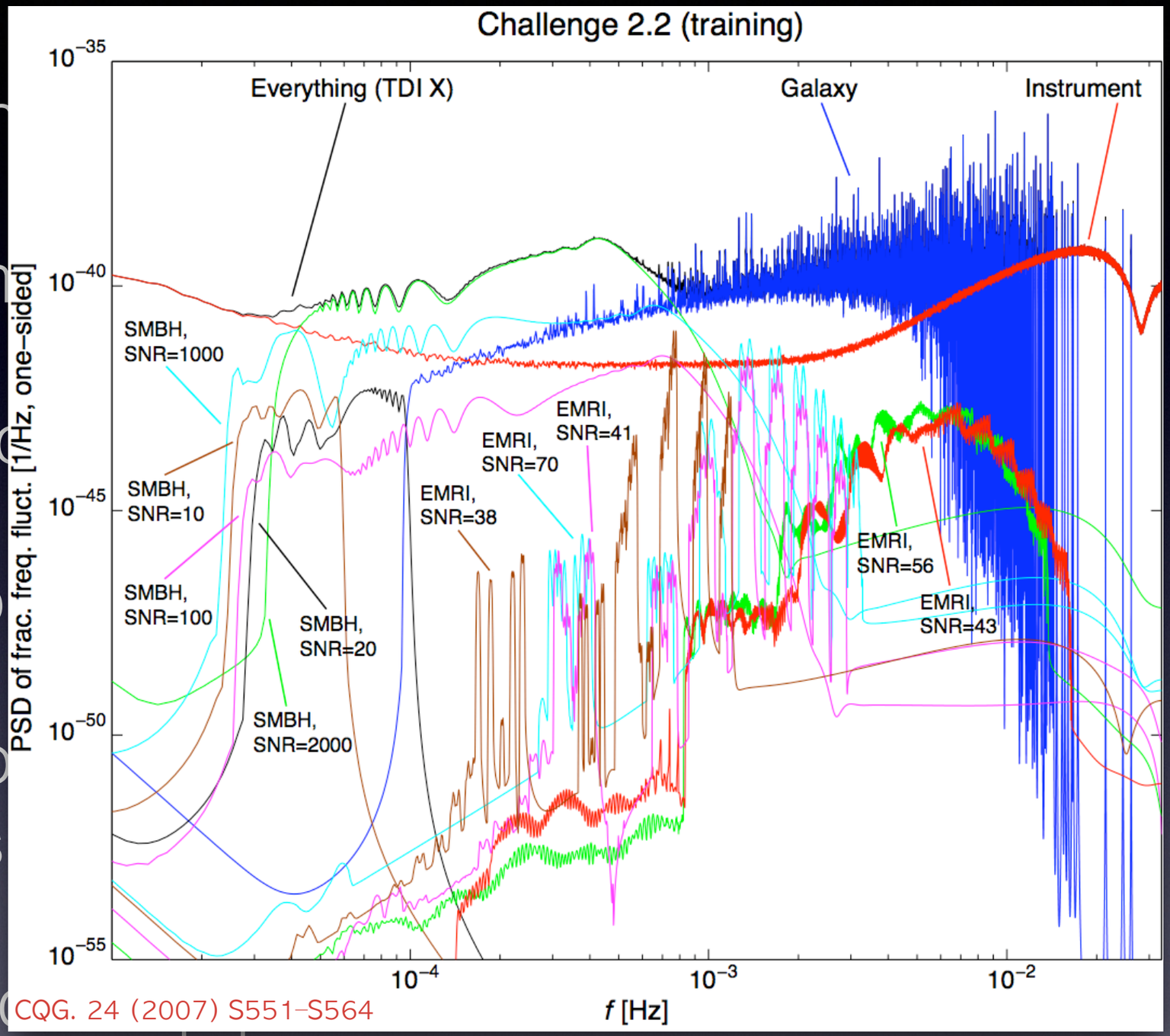


LDCs: Definition

- Involve community
- Build a common DA ground for the community
- Gather methods & techniques
- Oversee/coordinate the implementation
- Generate mock data-sets, organise challenges/ competitions, gather results, compare.
- Propose/build DA pipelines based on the successful methods.

LDCs: Definition

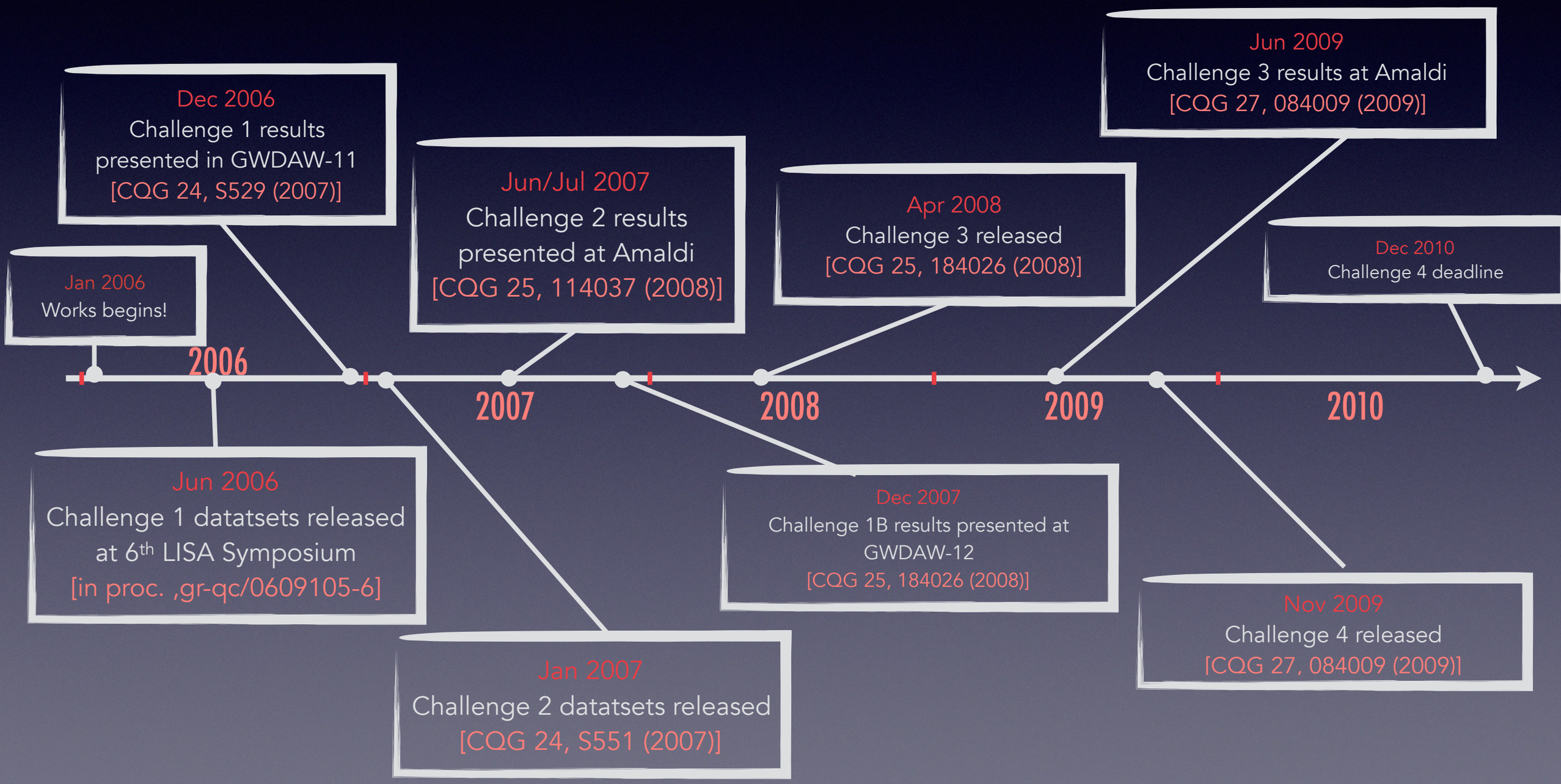
- Involve comm
- Build a comm
- Gather metho
- Oversee/coo
- Generate mo
  competitions
- Propose/buil
  successful methods.



CQG. 24 (2007) S551–S564

LDCs: History

- The first “generation” of LDCs started at 2006 with the last data-set being released in 2010.



LDCs: History

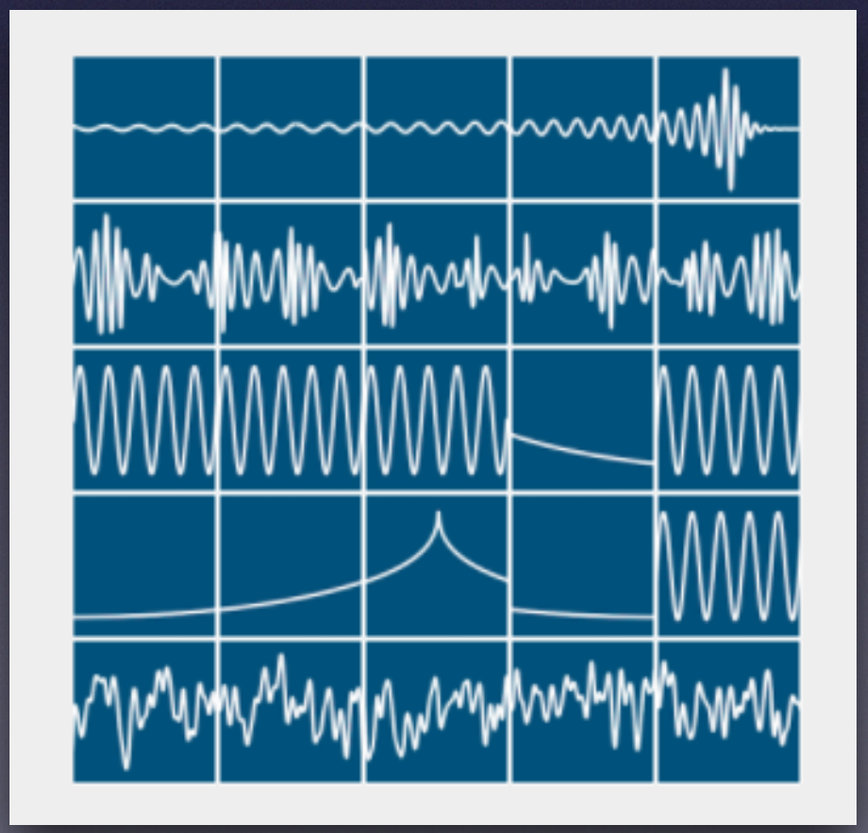
	MLDC 1	MLDC 2	MLDC 1B	MLDC 3	MLDC 4
Galactic binaries	• Verification • Unknown isolated • Unknown interfering	• Galaxy 3×10^6	• Verification • Unknown isolated • Unknown interfering	• Galaxy 6×10^7 chirping	• Galaxy 6×10^7 chirping
Massive BH binaries	• Isolated	• 4-6x, over "Galaxy" & EMRIs	• Isolated	• 4-6x spinning & precessing over "Galaxy"	• 4-6x spinning & precessing, extended to low-mass
EMRI		• Isolated • 4-6x, over "Galaxy" & MBHs	• Isolated	• 5 together, weaker	• 3 x Poisson(2)
Bursts				• Cosmic string cusp	• Poisson(20) cosmic string cusp
Stochastic background				• Isotropic	• Isotropic

- Five challenges completed
- 70 participants, 25 institutions, 30+ publications
- Chairs: Alberto Vecchio, Michele Vallisneri
- Demonstrated the detection and parameter estimation of all major LISA source classes, using a great variety of methods
- Provided methods for ground-based parameter estimation

LDCs: Current status

https://lisa-ldc.lal.in2p3.fr/

- The new “generation” of LDCs was launched a year ago with the *Radler* data-set.
- Besides the already mentioned purposes, now:
- Support mission studies: gap & glitches project; official SNR/parameter-estimation calculators
- Support LSG and LDPG development
- Work closely with Simulation WG

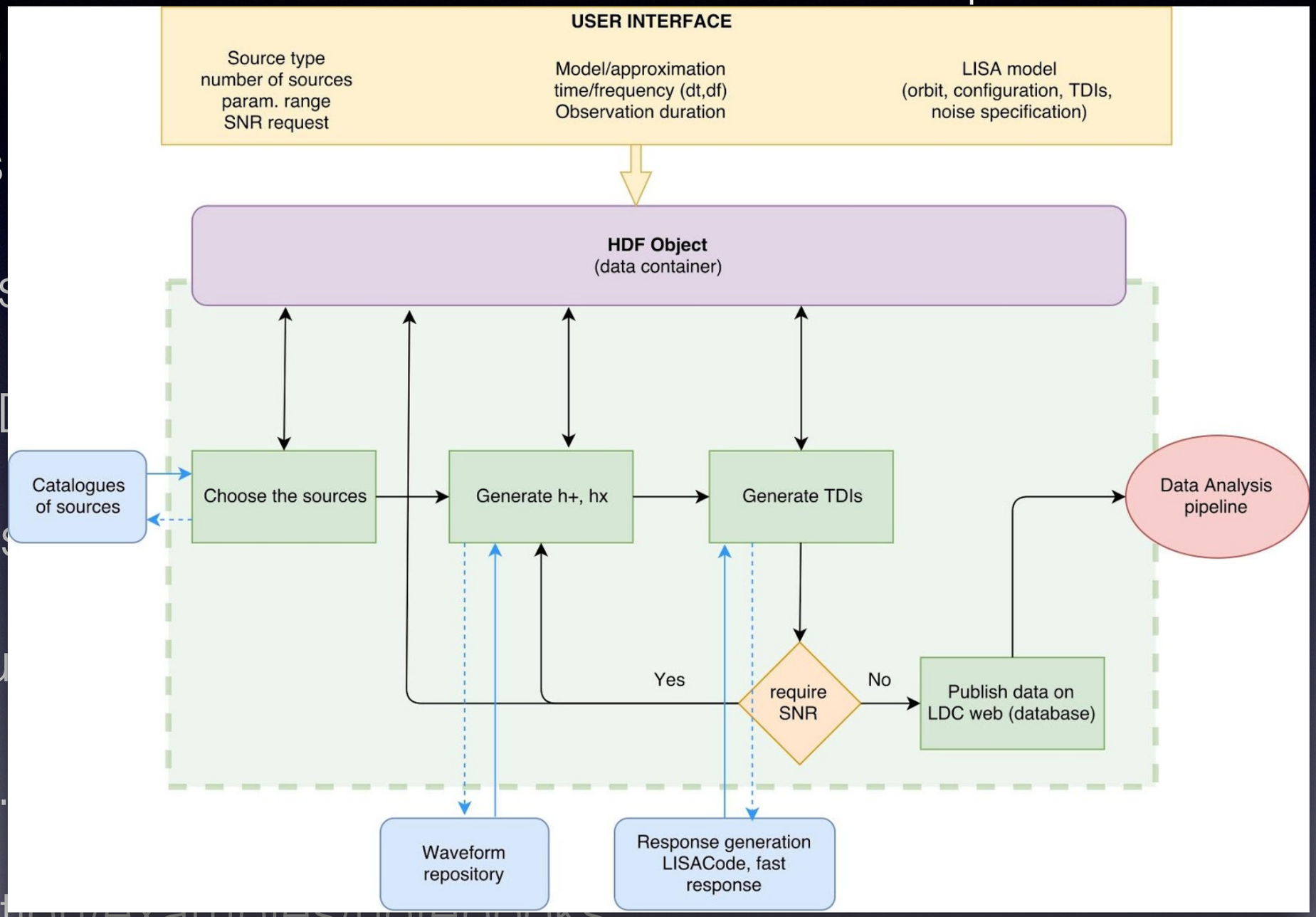


LDCs: The software

- The software contains all the basic tools in order to perform LISA simulations, and basic data analysis. In particular:
- Wave forms (MBHBs, SOBBHs, EMRIs, GBs).
- LISA Code simulator.
- Fast LISA FD TDI response.
- Analytic noise curves (TDI X, Y, XY, A, E, T).
- SNR computation tools.
- Catalogues.
- Documentation/examples/notebooks.

LDCs: The software

- The software contains all the basic tools in order to perform LISA simulations,
- Wave forms
- LISA Code s
- Fast LISA F
- Analytic nois
- SNR compu
- Catalogues.
- Documentation/examples/notebooks.



LDCs: Current status

- The software lives here: https://gitlab.in2p3.fr/stas/MLDC,

  officially moving to https://lisa-ldc.lal.in2p3.fr/code (releases, etc.)

- The first data-set was produced: the *Radler*. It is located here: https://lisa-ldc.lal.in2p3.fr/ldc

- The LDC Team also manages the results submission tools and interface: https://lisa-ldc.lal.in2p3.fr/form

LDC: more details

- **MBHBs & SOBBHs:**
  - Merger+ring-down, PhenomD WF, adapted by S. Babak, S. Marsat, & M. Pürrer (arxiv:508.07250, 1508.07253)
  - TODO: fast, higher mods WF with eccentricity.
- **EMRIs:**
  - Older analytic WFs from Barack & Cutler (PhysRevD.69.082005).
  - New WFs from A J K Chua & J R Gair are going to be integrated.
  - Fast tools welcome!
- **SGWB:**
  - LISA Code & pixel sources in the sky isotropically emitting a stochastic signal given a power law.
  - Possibility to simulate data for any given spectral shape.

LDCs: the Radler

LDCs: the Radler

- The Radler data-set is the first new data generated by the new LDCs. It aims to:
- Introduce new researchers to LISA data analysis.
- Tackle main LISA sources separately under idealized instrument noise.
- Rehabilitate existing analysis codes: Resurrect & update old ideas, propose & develop new ideas, gather the tools and codes for the analysis.
- Establish a normalized LDC process and introduce standards and basic infrastructure.

LDCs: the Radler contents

	Type
EMRIs	Single EMRI_AK in instrumental noise
	Single EMRI_AK, noiseless
Galactic Binaries	Verification Binaries in Instrumental noise
	Verification Binaries, noiseless
	Complete catalogue Binaries in instrumental noise
	Complete catalogue Binaries, noiseless
SOBBHs	Complete population*, in instrumental noise
	Complete population*, noiseless
	Bright SOBBHs (SNR>5) in instrumental noise
	Bright SOBBHs (SNR>5), noiseless
MBHBs	Single MBHB in instrumental noise
	Single MBHB, noiseless
SGWB	Power law signal in instrumental noise
	Power law signal, noiseless

- Instrumental noise is simulated w/ LISA Code, configured to SciRD noise levels.
- Now discussions on the submitted results (~15 submissions!).
- Submissions refer to Verification Binaries, MBHBs, and SGWB.

* PRL 116, 231102

LDCs: the Radler contents

- Submissions refer to Verification Binaries, MBHBs, and SGWB.
- Preference to Matched filtering.
  - If $d(t) = s(t, \vec{\theta}) + n(t)$, and s=h if template matches the signal.
  - Then, with Gaussian noise, we form the posterior as:

$$p\left(d\right) \propto \exp\left[-\frac{1}{2}\left(d - h(\vec{\theta})|d - h(\vec{\theta})\right)\right] p(\vec{\theta})$$

with

$$(a|b) \equiv 4\mathcal{R}\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}\mathrm{d}f$$

- Stochastic methods are also preferred (Markov Chain Monte Carlo, PTMCMC, Nested sampling …)
- Final aim is to train for the “Global fit”, due to the signals overlap in frequencies and time.

* PRL 116, 231102

LDCs: the Radler contents

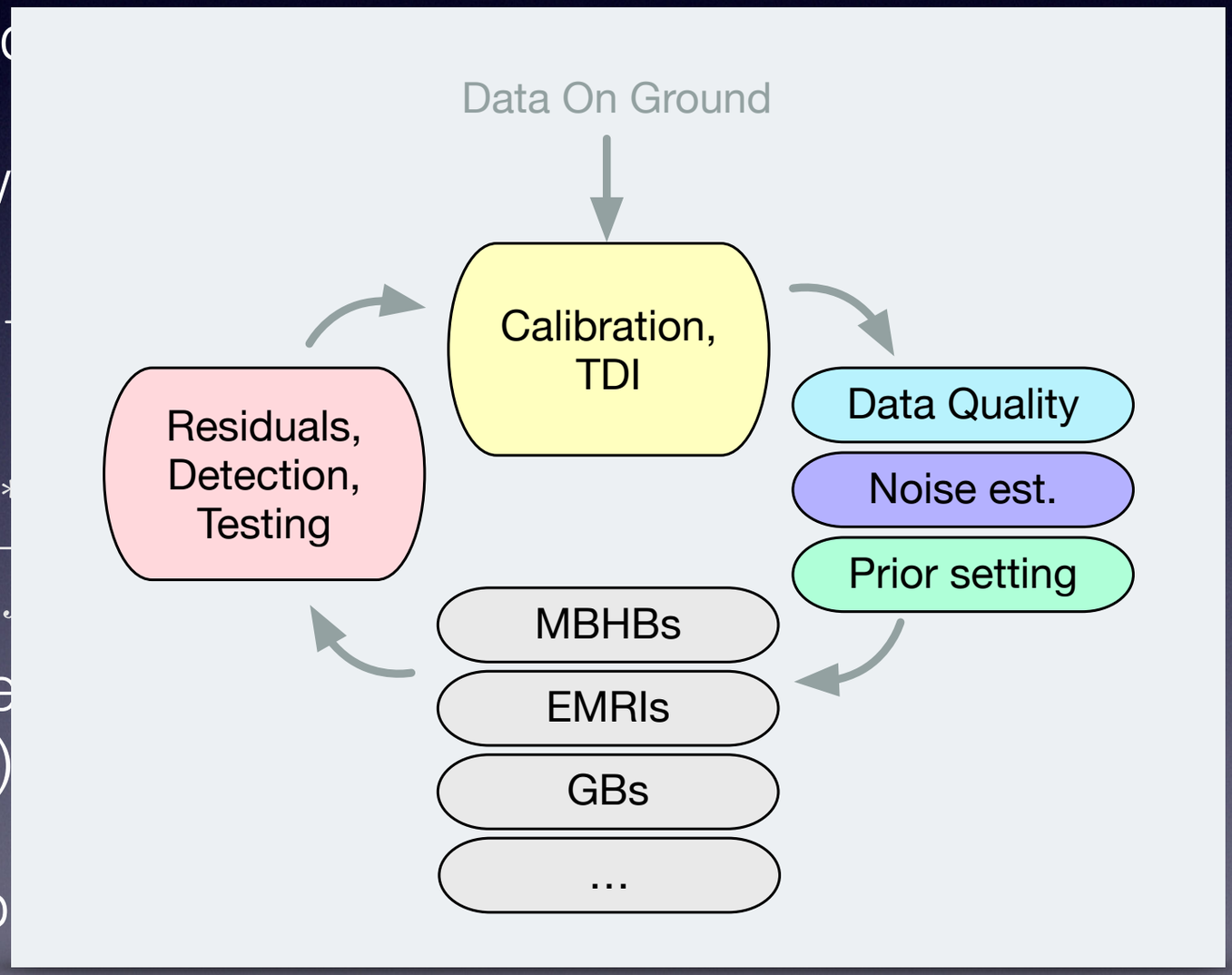
- Submissions refer to Verification Binaries, MBHBs, and SGWB.
- Preference to Matched filtering.
  - If $d(t) = s(t, \vec{\theta}) + n(t)$, and
  - Then, with Gaussian noise, w

$$p(d) \propto \exp\left[-\right.$$

with

$$(a|b) \equiv 4\mathcal{R}\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*}{S_n(}$$

- Stochastic methods are also pre
  PTMCMC, Nested sampling …)
- Final aim is to train for the "Glob
  frequencies and time.



* PRL 116, 231102

Another thing: Figures Of Merit!

- There is an ongoing effort to develop a FoM infrastructure with the aim of evaluating instrument configurations based on science output of said configuration.
- Latest example is the latest “study” for changes in the sensitivity for frequencies < 1e-4. We are expecting that this will happen more in the future.
- **Status:** assembling & definitions.
- We need the Science objectives of each WG and tools.
- Ideas and inputs are welcome. Implementation inputs as well.

Another thing: Figures Of Merit!

- The idea is to create a pipeline that would target specific scientific objectives of LISA and assess the detectability capabilities, given instrument configurations.

- Directly connected to the Performance Working Group.
  - The Performance WG builds a tree of noise performance of the all the subsystems of the LISA satellites. Then it produces a final estimation of the PSD of X TDI channel.

- This noise input is then used to calculate the figures of merit. The output should have three possibilities: Green, Orange and Red.

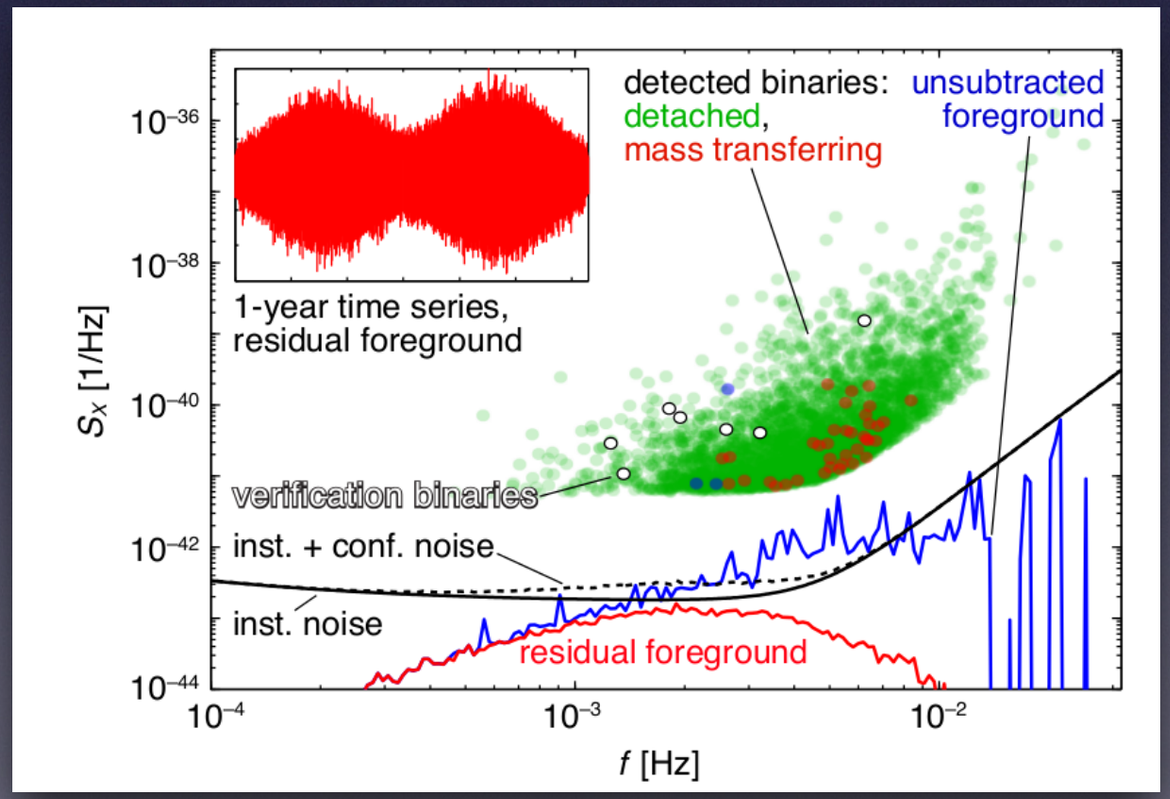
Figures Of Merit: An example with the GBs.

- A figure of merit can be constructed for the WD GBs sources that we will be able to extract in a given stage of the mission.

- Expect millions of binaries. Almost monochromatic in nature. Also Verification Binaries.

- Expect thousands of them to be resolvable.

- But how can we set up a FoM, based on these models?



S. Nissanke et al, Astrophys. J. 758(131) (2012)

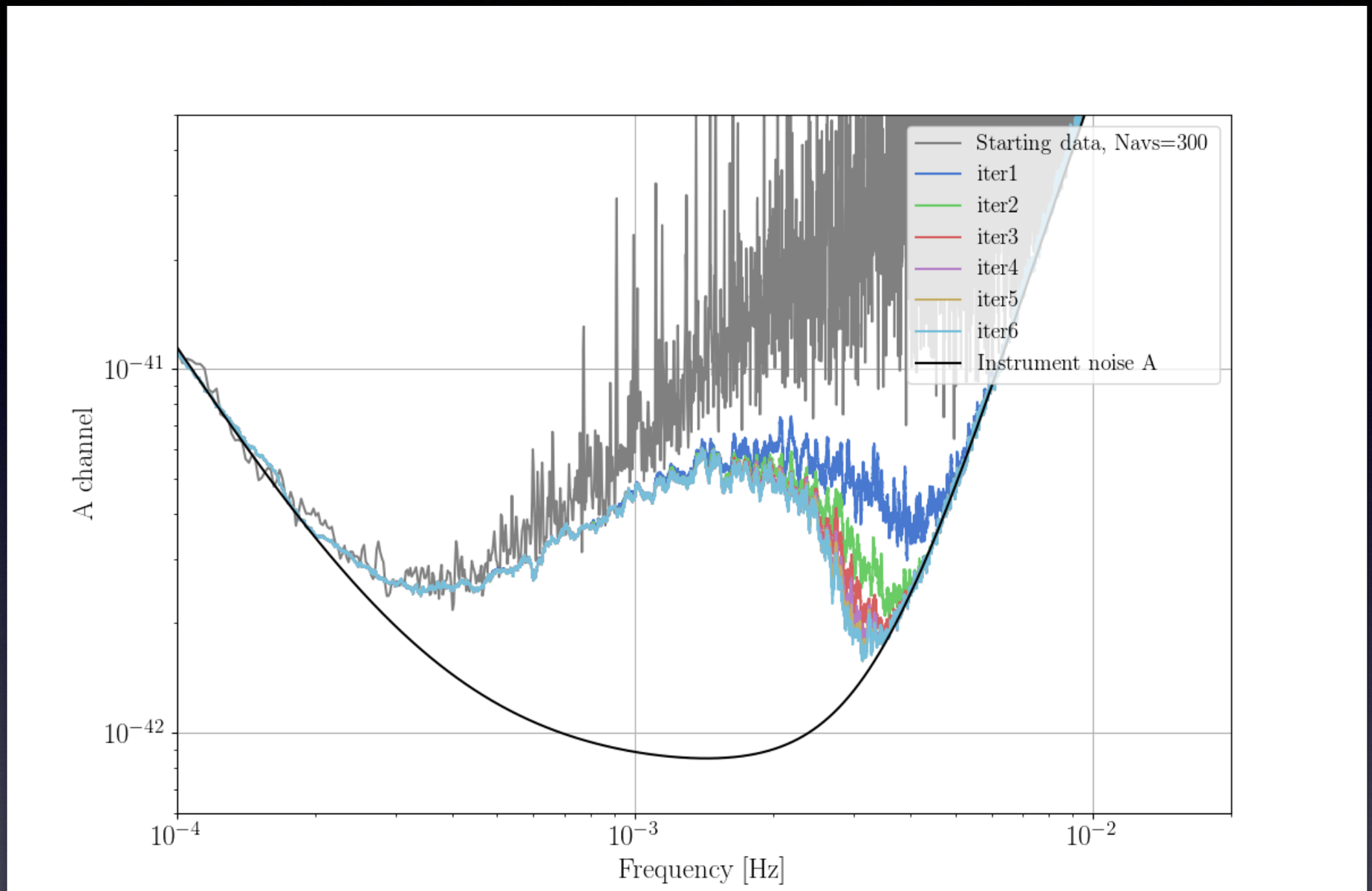
Figures Of Merit: An example with the GBs.

- A figure of merit can be constructed for the WD GBs sources that we will be able to extract in a given stage of the mission.
- For now we choose the end of the nominal mission, which is 4 yrs.
- We set the colour codes as:
  - Green: above 6e3 sources extracted,
  - Orange: between 4e3 to 6e3 sources extracted,
  - Red: below 4e3 sources :(
- We have built a multi-purpose tool within the LDC software with the aim to characterise stochastic signals, such as the residuals from the GBs. This is based on S. Nissanke et al, Astrophys. J. 758, 131, 2012.
- Multipurpose: Because it is designed to work for EMRIs and SOBBHs populations.

Figures Of Merit: An example with the GBs.

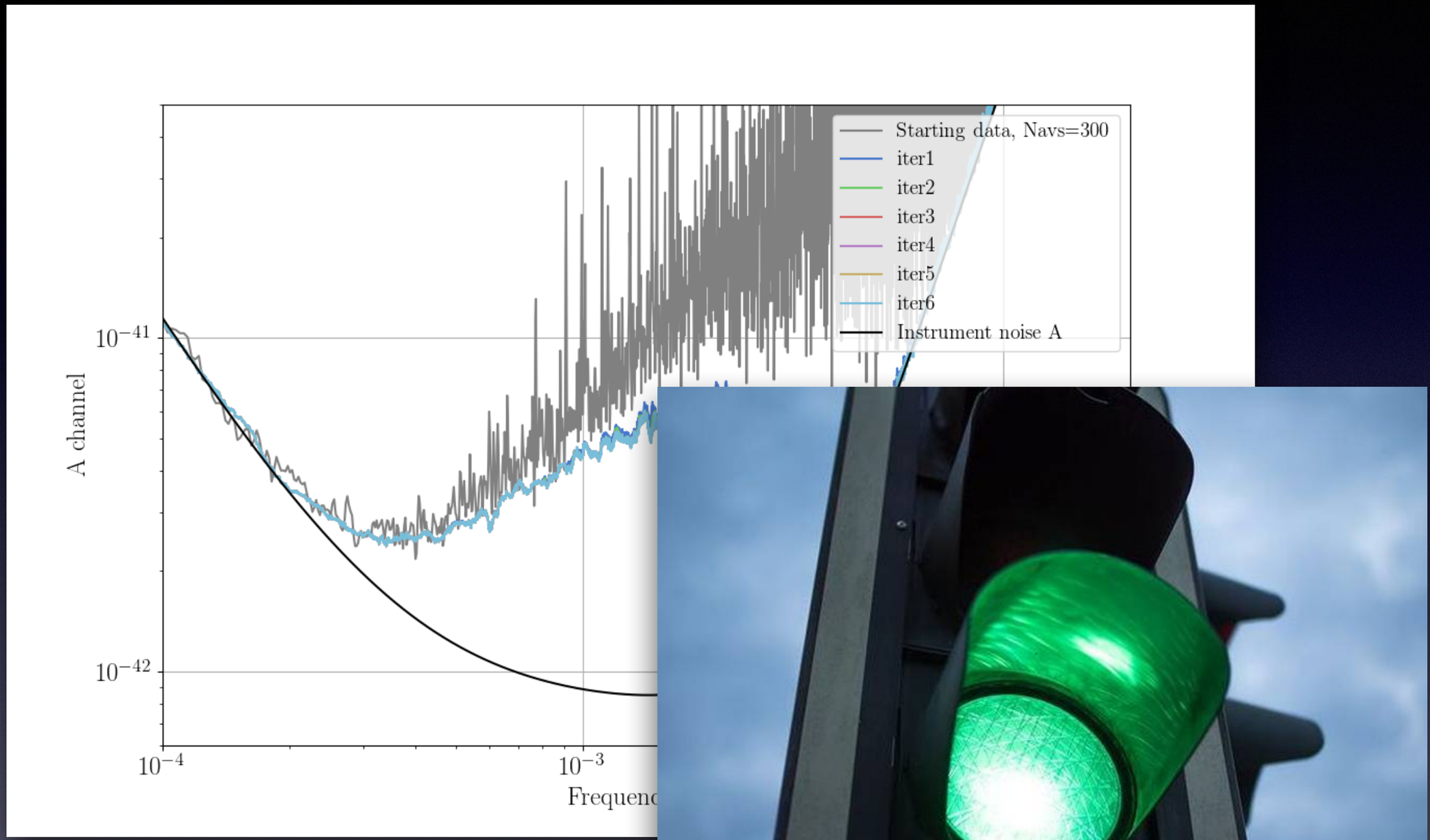
- Given $T_{\rm obs} = 4$ yrs, simulate noise with LISA Code.
- Simulate TDI AE of the complete catalogue, *and* compute & store the *optimal* $SNR_{\rm opt}$ for each source. Optimal in the sense that it is computed with respect to instrumental noise only.
- Loop:
  1. Compute a Smooth PSD of the data $S_{n,k}$, by performing a running median on the it (optionally do a polynomial or spline fit).
  2. Compute $SNR_{\rm i}$ for source i, with respect to $S_{n,k}$.
  3. If $SNR_{\rm i} > SNR_0$ subtract.
  4. If $S_{n,k} = S_{n,k-1}$ stop. Otherwise go back to 1.
- Suboptimal process (assume perfect subtraction).

Figures Of Merit: An example with the GBs.



- For the particular example, where we have used SNR0 = 7 we have recovered 14935 sources already for 2 years of observation.

Figures Of Merit: An example with the GBs.



- For the particular example, whe e have recovered 14935 sources already for 2 years of observation.

Figures Of Merit: An example with the GBs.

- Perform this analysis in different time steps of the mission.
- For example start with 0.1, 0.2, 0.5, 1, 2, 5 years data-sets.
- Evaluate an analytic model of the confusion noise that depends on Tobs.
- Integrate it to the LDC software.
- Build more FoM based on this function.

LDCs: The future Challenges

LDCs: The future Challenges

- Ongoing discussions starting now.
- Submitted results from the Radler need to be examined/reviewed, and the performance of each tool needs to be assessed.
- **Spritz:** Non-stationary instrument noise & light astrophysical content:
    - to address robustness of algorithms used in Radler for non-stationary noise. Proper definitions if non-stationarities is needed. Extrapolation from LPF data is possible.
    - to help setting some requirements on the instrument performance & Data Quality.
- **Sangria:** Mild Enchilada: Galaxy + MBHBs + EMRI+ Gaussian stationary noise:
    - Start prototyping global fit pipeline
    - Investigation: are signals aware of each other?
    - Building the catalogues
    - Assessment of required resources and hardware structure.

LDCs: Summary

- The Radler challenge is complete. Evaluation of the results is pending.
- New LDCs are underway. Now at discussions & Planning stage. Probably the next one will materialise during summer (?)
- Software is under development. Improvements are being made at all fronts (DA, WF, Catalogues, infrastructure).

LDCs: A tutorial

- Verification GBs Part I: *colab notebook link*
- Verification GBs Part II: *colab notebook link*
- MBHBs: *colab notebook link*

N. Karnesis, 7th LISA Cosmology WG Workshop, Padova, 2019