The Bayesian toolbox in the observational era:

Parallel nested sampling and reduced order models

Rory Smith

ICERM 11/16/20

overview

- The last year in observations
  - What do we need to do the best astrophysics
- Challenges in Bayesian inference
- Parallel nested sampling
- Reduced order models
- Looking to O4 and beyond
  - Rapid sky localization

Observations in O3

The last couple of years have been interesting...

GW190521: A Binary Black Hole Merger with a Total Mass of $150\ M_\odot$

R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration)
Phys. Rev. Lett. **125**, 101102 – Published 2 September 2020

Physics See Viewpoint: A Heavyweight Merger

GW190412: Observation of a binary-black-hole coalescence with asymmetric masses

R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration)
Phys. Rev. D **102**, 043015 – Published 24 August 2020

OPEN ACCESS

GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object

R. Abbott[1], T. D. Abbott[2], S. Abraham[3], F. Ace
V. B. Adya[8], C. Affeldt[9,10], M. Agathos[11,12] + S

Published 2020 June 23 • © 2020. The American A

OPEN ACCESS

GW190425: Observation of a Compact Binary Coalescence with Total Mass ~ $3.4\ M_\odot$

ott[1], T. D. Abbott[2], S. Abraham[3], F. Acernese[4,5], K. Ackley[6], C. Adams[7],
Adya[8], C. Affeldt[9,10] + Show full author list

19 • © 2020. The Author(s). Published by the American Astronomical Society.

nal Letters, Volume 892, Number 1

Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo

B. P. Abbott[1], R. Abbott[1], T. D. Abbott[2], S. Abraham[3], F. Acernese[4,5], K. Ackley[6], C. Adams[7],
R. X. Adhikari[1], V. B. Adya[8,9], C. Affeldt[8,9] + Show full author list

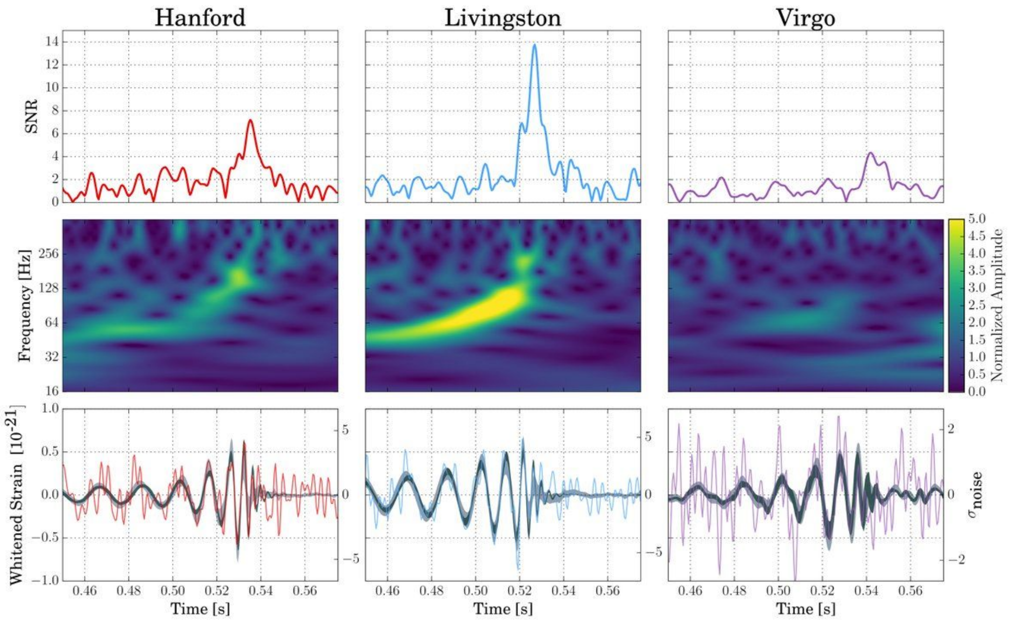
Published 2019 September 9 • © 2019. The American Astronomical Society. All rights reserved.

The Astrophysical Journal Letters, Volume 882, Number 2

Astronomy with gravitational-wave transients

Coalescing compact binaries

- Precise measurements of black hole spins
- Unambiguous measurement of asymmetric mass ratios
- Evidence for higher-order gravitational-wave modes
- Population properties and formation scenarios



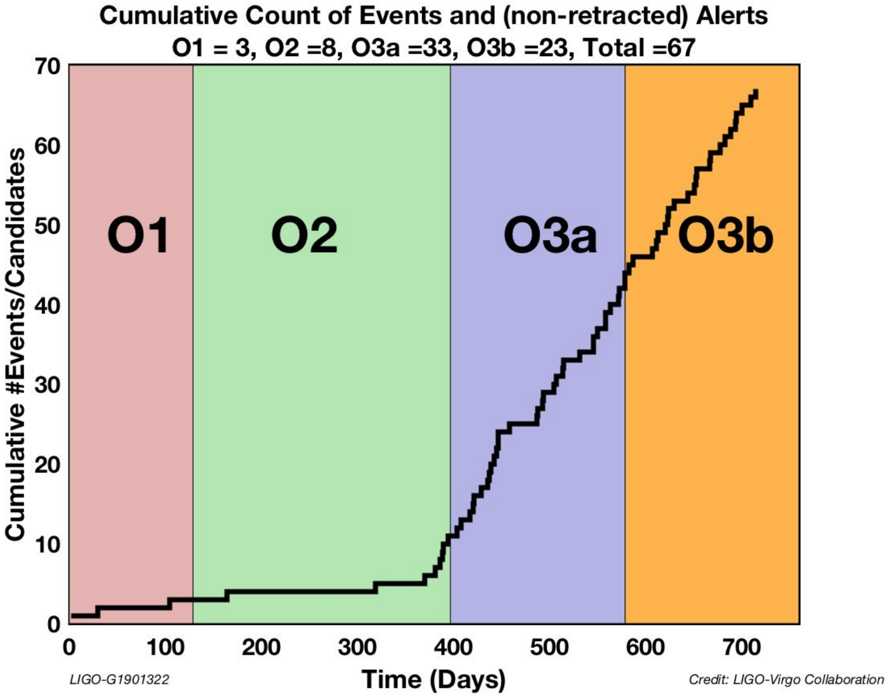
Extracting this information pushes the limits of our data analysis methods

What we need to do astronomy in O4 and beyond

- Compact binary waveform models with:
  - Higher order mode content
  - Precession
  - Calibration to NR (NR surrogates)
  - High mass ratios
  - Eccentricity (important for future BBH observations)
  - Tidal disruption (for future NSBH merger observations)
- Inference tools that can use the best, cutting edge models

What we need to do astronomy in O4 and beyond

- GW Astronomy requires scalable inference algorithms and accurate models models to keep up with event rate



Bayesian inference

Bayesian inference

Parameter estimation and hypothesis testing in a unified framework

$$p(\theta|d,\mathcal{H}) = \frac{\pi(\theta|\mathcal{H})\mathcal{L}(d|\theta,\mathcal{H})}{Z(d|\mathcal{H})}$$

- Unknown source parameters, e.g., masses & spins
- Experimental data
- Hypothesis/model of the data

Bayesian inference

Parameter estimation and hypothesis testing in a unified framework

$$p(\theta|d,\mathcal{H}) = \frac{\pi(\theta|\mathcal{H})\mathcal{L}(d|\theta,\mathcal{H})}{Z(d|\mathcal{H})}$$

- **Posterior**: Probability of parameters after analyzing data

- **Prior**: probability of the parameters before analyzing the data
- **Likelihood**: probability of the *data* given parameters and an hypothesis
- **Evidence**: Probability of the data given the hypothesis (marginalized over all parameters)

Bayesian inference: parameter estimation

$p(\theta|d,\mathcal{H})$ example: 1D & 2D projection of the full (17+)D probability distribution

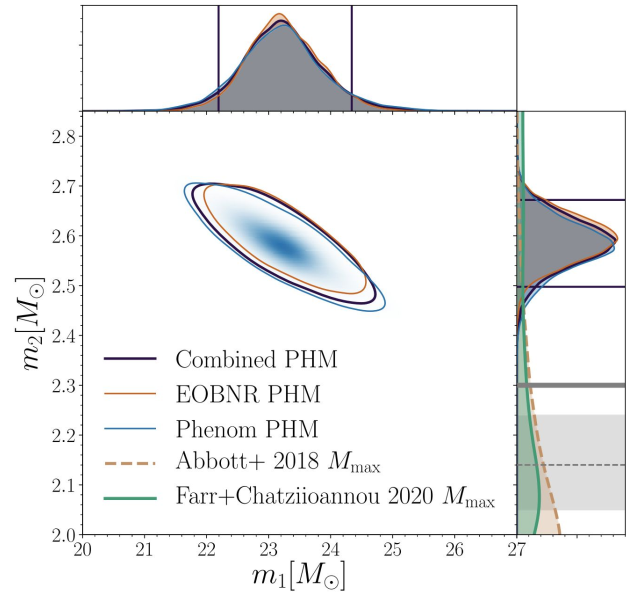


Table 1. Source Properties of GW190814: We Report the Median Values Along with the Symmetric 90% Credible Intervals for the SEOBNRv4PHM (EOBNR PHM) and IMRPhenomPv3HM (Phenom PHM) Waveform Models

	EOBNR PHM	Phenom PHM	Combined
Primary mass $m_1/M_\odot$	$23.2^{+1.0}_{-0.9}$	$23.2^{+1.3}_{-1.1}$	$23.2^{+1.1}_{-1.0}$
Secondary mass $m_2/M_\odot$	$2.59^{+0.08}_{-0.08}$	$2.58^{+0.09}_{-0.10}$	$2.59^{+0.08}_{-0.09}$
Mass ratio q	$0.112^{+0.008}_{-0.008}$	$0.111^{+0.009}_{-0.010}$	$0.112^{+0.008}_{-0.009}$
Chirp mass $\mathcal{M}/M_\odot$	$6.10^{+0.06}_{-0.05}$	$6.08^{+0.06}_{-0.05}$	$6.09^{+0.06}_{-0.06}$
Total mass $M/M_\odot$	$25.8^{+0.9}_{-0.8}$	$25.8^{+1.2}_{-1.0}$	$25.8^{+1.0}_{-0.9}$
Final mass $M_f/M_\odot$	$25.6^{+1.0}_{-0.8}$	$25.5^{+1.2}_{-1.0}$	$25.6^{+1.1}_{-0.9}$
Upper bound on primary spin magnitude χ_1	0.06	0.08	0.07

GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object, ApJL (2020)

Bayesian inference: hypothesis testing

Hypothesis testing encoded in the Bayesian "evidence" $Z(d|\mathcal{H})$

- Allows for data-driven hypothesis testing, e.g.,
  - "How much more likely is it that GW190814 was described by a signal containing higher order modes than a signal without higher order modes?"
  - This would be expressed in a Bayesian way using a **Bayes factor**:

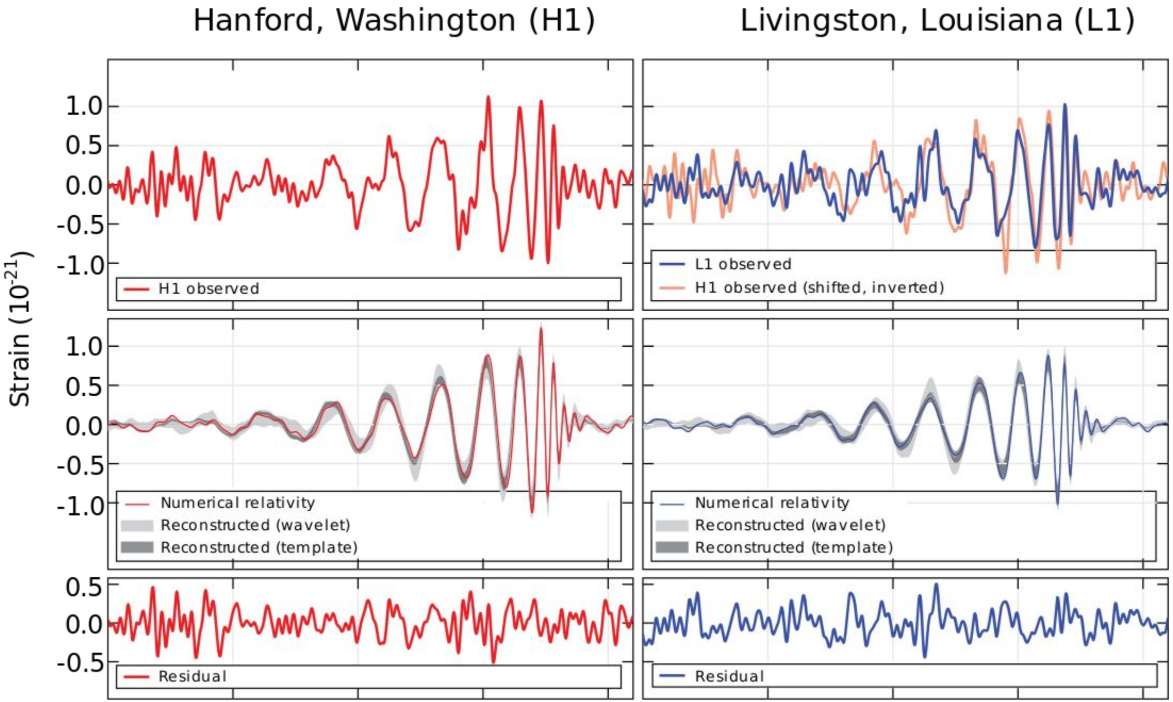
$$B = \frac{Z(d|\mathcal{H}_1)}{Z(d|\mathcal{H}_2)}$$

Challenges

Challenges in Bayesian inference

Expensive models

- Computing PDFs and evidences requires comparing signal models to data

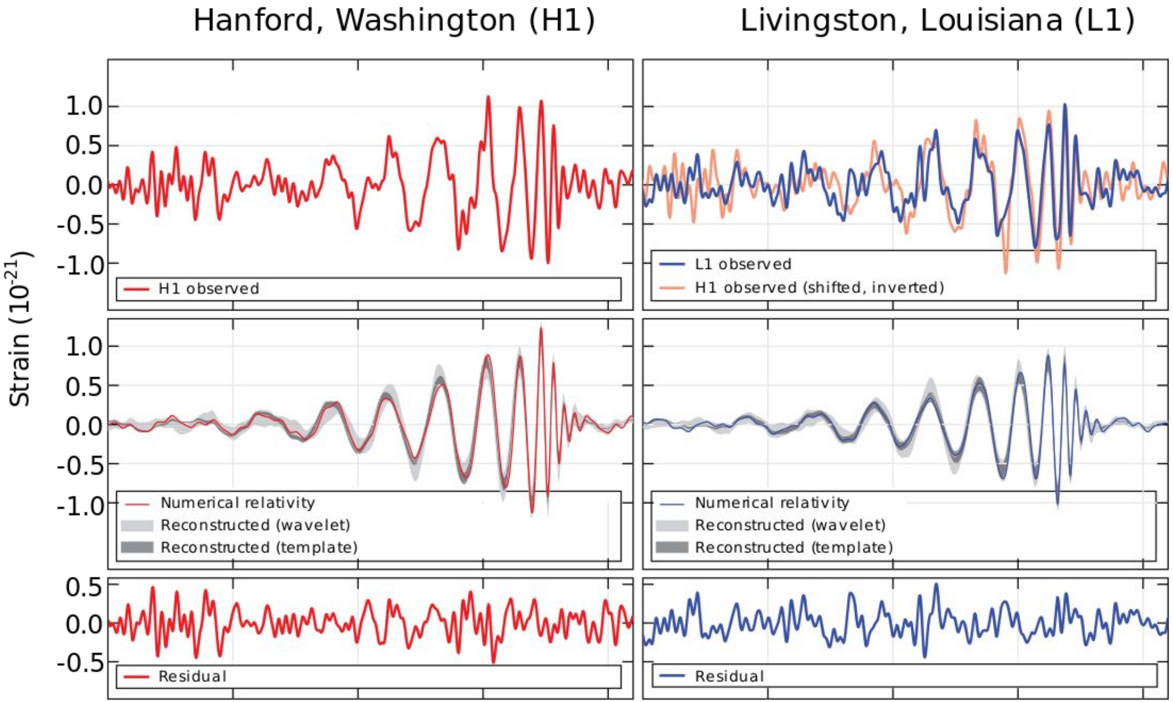


GW150914

Challenges in Bayesian inference

Expensive models

- Computing PDFs and evidences requires comparing signal models to data
  - When used “out of the box”, inference can take anywhere between **hours to years**
  - Most expensive, e.g.,
    - HoMs, precession, beyond GR effects etc...

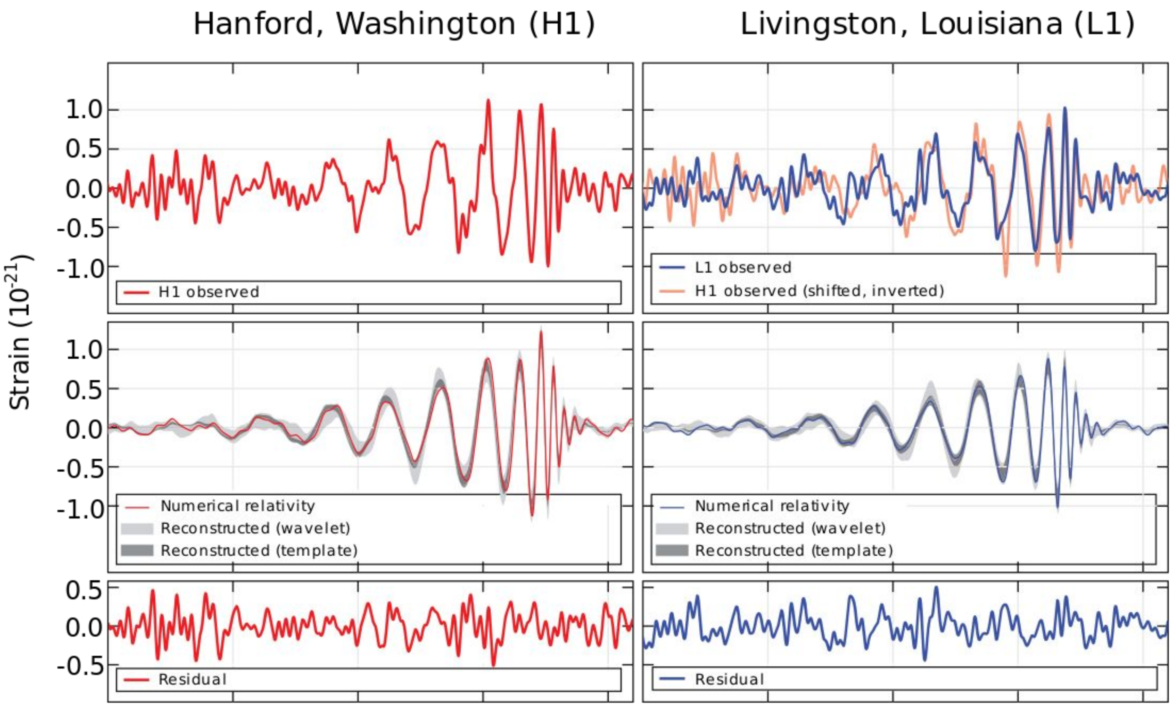


GW150914

Challenges in Bayesian inference

Expensive models

- Computing PDFs and evidences requires comparing signal models to data
  - In some cases reduced order models exist that are cheaper to evaluation
  - But these often take time to develop



GW150914

Challenges in Bayesian inference

“Curse of dimensionality”

- Astrophysical parameter spaces are 15D (binary black holes) and 17D (binary neutron stars)
- Additional 20 parameters *per GW detector* that encode uncertainty about detector calibration
  - Between 50-70 parameters that have to be inferred simultaneously

Challenges in Bayesian inference

Big data. Sort of…

In practice, often use stochastic samplers to explore parameter spaces

- ❖ Nested sampling and MCMC

- Roughly 100Tb-1Pb of data generated and analyzed *per event* to produce parameter estimates
    - Model space much much MUCH bigger than the strain data
- Population inference takes as input millions of posterior samples

Main costs

1. Template waveform generation is expensive
2. Large number of likelihood(waveform) calls
    - Around 50-100M per analysis

These problems compound

Some solutions

- Parallel sampling methods :
    - Reduce the wall time of inference by producing more samples per s, but overall CPU time is roughly conserved (and high)
- Reduced order models:
    - Reduce overall CPU time by making likelihood(waveform) evaluations cheaper
    - Can be stand ins (surrogates) for full Numerical Relativity

(I'm only going to focus on classical sampling methods, i.e., no machine learning, which is also interesting for astrophyiscal inference)

Parallel nested sampling

Parallel nested sampling

For O3, we needed a method that was

- *Accurate*
  - Don’t cut corners or make approximations (if you can avoid it)
- *Flexible*
  - Use all of the best signal models to analyze each event! Update models when new ones become available
  - Useful for wide range of problems, not just for CBCs
- *Scalable*
  - Should handle a growing amount of work by throwing more CPUs/GPUs at it

Nested sampling

- Designed for high-dimensional integration of the Bayesian evidence (Skilling 2006):

$$Z(d|\mathcal{H}) = \int d\theta \, \pi(\theta|\mathcal{H}) \mathcal{L}(d|\theta, \mathcal{H})$$

In our case, this is integral is around 50-70 dimensional

As a byproduct, nested sampling produces posterior samples

- Accomplishes both tasks of inference

Nested sampling

The “trick” of nested sampling is to replace a high-D integral with a 1D integral:

$$Z(d|\mathcal{H}) = \int_0^1 dX\, \mathcal{L}(d|X,\mathcal{H})$$

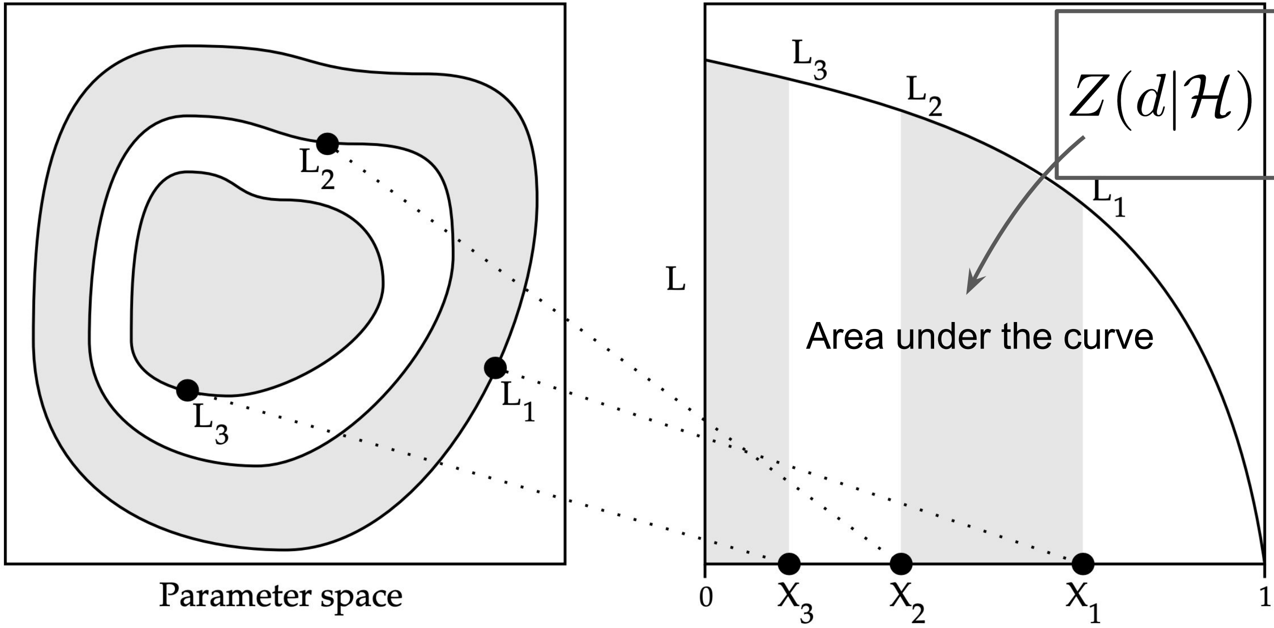


Figure 3: Nested likelihood contours are sorted to enclosed prior mass X.

Skilling 2006 (Nested sampling for general Bayesian computation)

Nested sampling

Algorithmically, we:

0. Initialize: draw M samples ("live points") from the prior and rank them from highest to lowest likelihood

1. Draw a sample from the ***prior***
   a. Accept if the likelihood is greater than the lowest live point
   b. Otherwise, repeat
2. Replace lowest-likelihood live point with new sample
3. Estimate evidence
4. Repeat until change in evidence is below some threshold

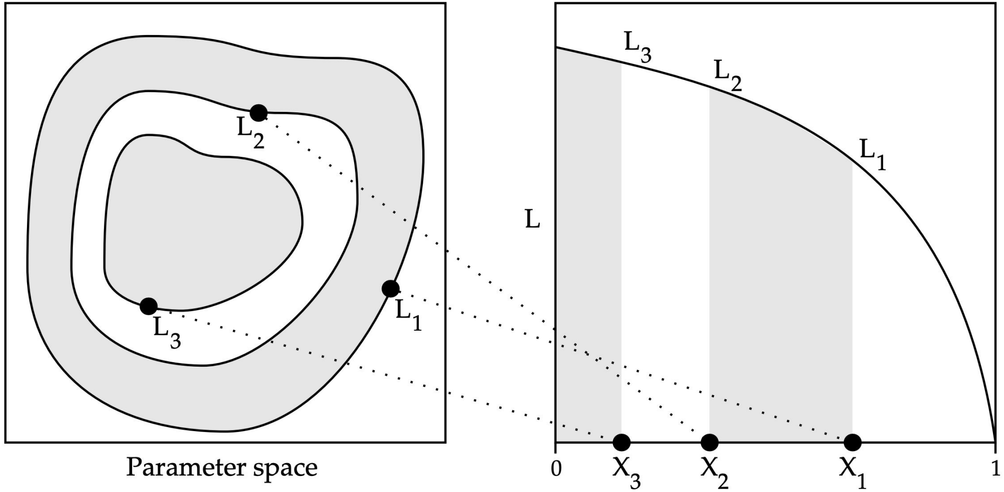


Figure 3: Nested likelihood contours are sorted to enclosed prior mass X.

Nested sampling

Algorithmically, we:

0. Initialize: draw M samples ("live points") from the prior and rank them from highest to lowest likelihood

1. **Draw a sample from the prior**
   a. **Accept if the likelihood is greater than the lowest live point**
   b. **Otherwise, repeat**
2. Replace lowest-likelihood live point with new sample
3. Estimate evidence
4. Repeat until change in evidence is below some threshold

We know the prior (by definition) *a priori* so we can draw N samples simultaneously on each iteration

Provides a theoretical speedup of

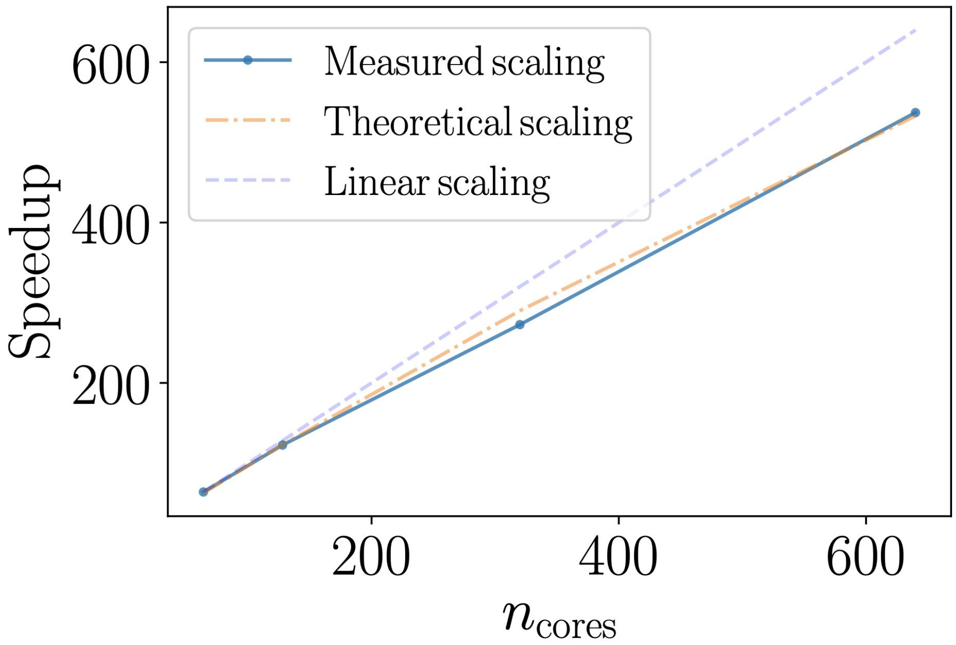
$$S = M_{\mathrm{live}} \ln\left(1 + N_{\mathrm{cores}}/M_{\mathrm{live}}\right)$$

Not perfect scaling: probability of accepting samples < 1

Smith et al 2020, Handley et al 2015

Main results

- Scales well up to around 800 cores
- Implemented within the parallel bilby ([pBilby]) library.
- Uses the dynesty nested sampler parallelized with `mpi4py`
  - Production code in the LVC since around March



Smith et al MNRAS Vol. 498 Issue 3 (2020)

Main results

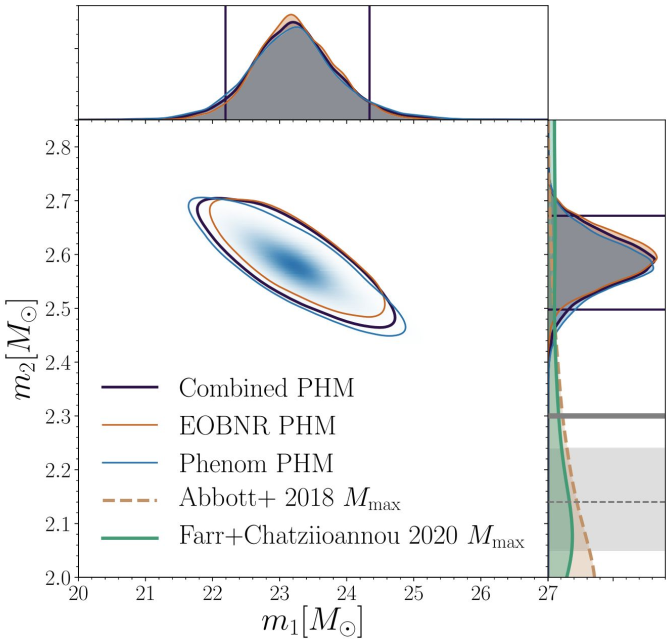
	IMRPhenomPv3HM			SEOBNRv4PHM			IMRPhenomPv2NRT		
Number of CPUs	16	64	640	16	64	640	16	64	640
GW150914	3.9 d	23.3 hr	2.8 hr	*83.7* d	*21.2* d	2.5 d	–	–	–
GW190425	–	–	–	—	—	—	*30.7* d	*7.8 d*	22 hr
GW190412	*60.3 d*	*15.3 d*	1.8 d	*2.9* yr	*276.1* d	11.53 d	—	—	—

Table 1. Wall times for selected events using n_{cores} = (16, 64, 640) CPUs. Measured wall times are non-italicized and estimated wall times are *italicized*.

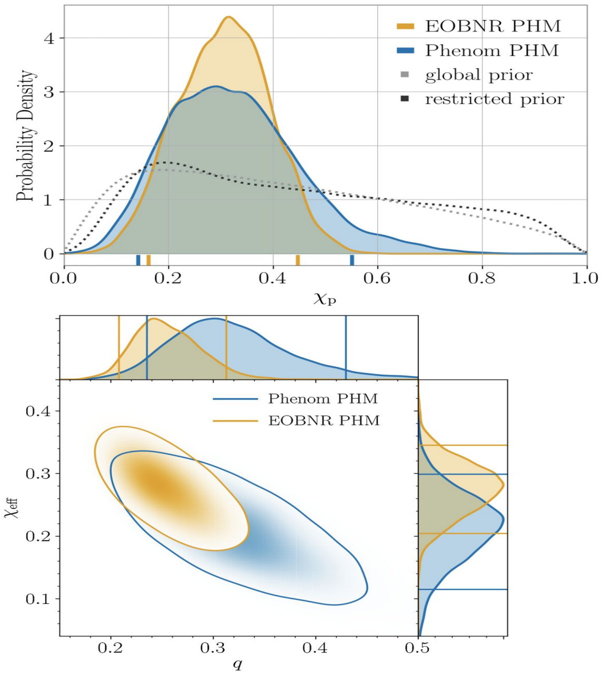
- Submission of our paper was before publication of GW190814
  - Similar scalings and run times for SEOBNRv4PHM

Smith et al MNRAS Vol. 498 Issue 3 (2020)

Use in the LVC



GW190814



GW190412

Reduced order models
(ROMs)

Reduced order models

- Directly address the overall cost of inference (reduce CPU time)
  - Can be “surrogate” models for full numerical relativity simulations
  - ...or faster-to-evaluate versions of approximate waveform models
  - Important for keeping up with event rate in O4+
  - Can enable ***fast and optimal sky localization*** for electromagnetic follow up

Reduced order models: what are they?

Represent the waveform as a weighted sum of basis elements

Usually, the basis set is *sparse*, i.e., only need a small number of elements

$$\text{Frequency domain: } h(f;\theta) = \sum_i h(F_i;\theta) B_i(f)$$

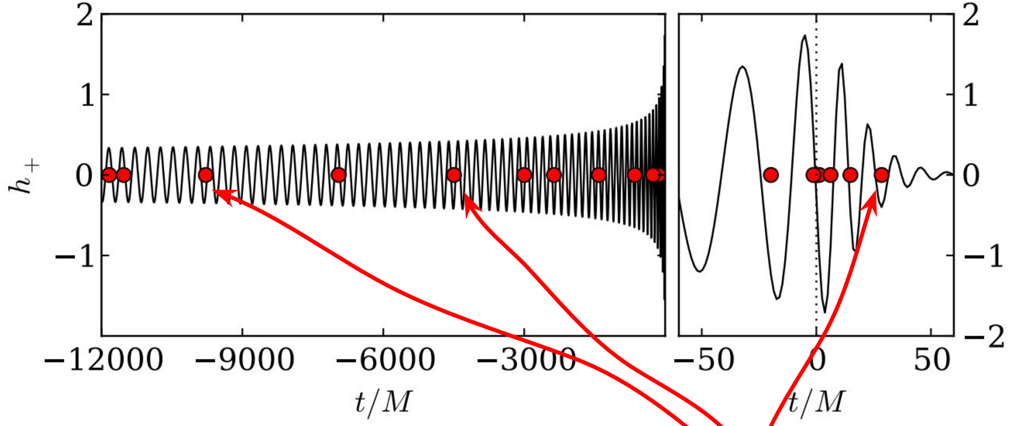
$$\text{Time domain: } h(t;\theta) = \sum_i h(T_i;\theta) B_i(t)$$

“Empirical interpolation” nodes (using EIM greedy algorithm)

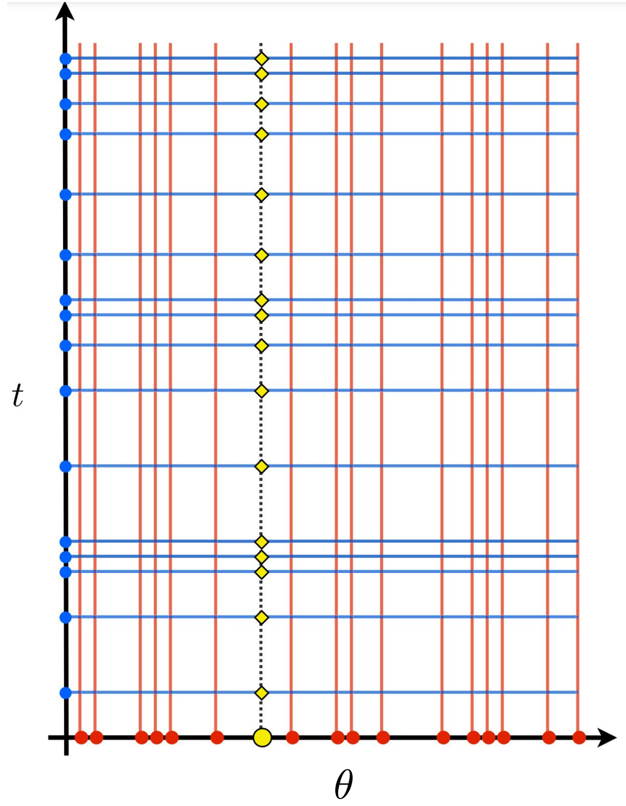
basis set via Greedy algorithm (judiciously chosen templates)

Field et al Phys. Rev. X **4**, 031006 (2014)

Reduced order models: what are they?



Time domain: $h(t;\theta) = \sum_i h(T_i;\theta) B_i(t)$



Field et al Phys. Rev. X **4**, 031006 (2014)

Reduced order models: why are they useful?

- Only need to compute waveform at nodes
  - *Reduces overall CPU time when templates are dominant cost of an analysis*
  - *Compress large inner products that appear in the likelihood function (reduced order quadrature -- **ROQ**)*

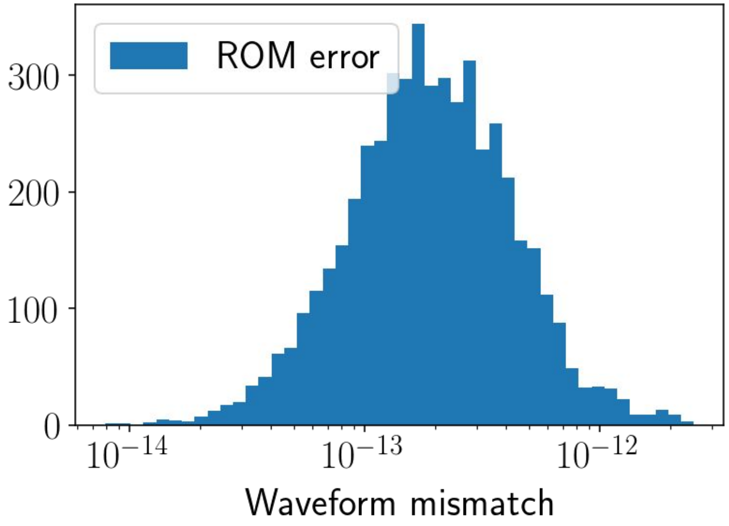
f (Hz) Min	f (Hz) Max	Waveform duration T	Δf (Hz)	$\mathcal{M}$ ($M_\odot$) Min	$\mathcal{M}$ ($M_\odot$) Max	Basis size Linear	Basis size Quadratic	Speedup
20	1024	$1.5\text{s} \leq T \leq 4\text{s}$	1/4	12.3	23	300	197	8
20	1024	$3\text{s} \leq T \leq 8\text{s}$	1/8	7.9	14.8	388	278	12
20	2048	$6\text{s} \leq T \leq 16\text{s}$	1/16	5.2	9.5	360	233	54
20	2048	$12\text{s} \leq T \leq 32\text{s}$	1/32	3.4	6.2	524	254	83
20	2048	$23.8\text{s} \leq T \leq 64\text{s}$	1/64	2.2	4.2	749	270	127
20	4096	$47.5\text{s} \leq T \leq 128\text{s}$	1/128	1.4	2.6	1253	487	300

$$\text{speedup} = \left[(f_{\text{high}} - f_{\text{low}})/\Delta f\right]/(N_{\text{bases}})$$

Smith et al Phys. Rev. D **94**, 044031 (2016)

Reduced order models: why are they useful?

- Useful representation for numerical relativity surrogates → helps inference by allowing us to use stand ins for full NR
- Extremely accurate (as measured by the mismatch)



More details in, e.g., Smith et al Phys. Rev. D **94**, 044031 (2016), Canizares et al Phys. Rev. Lett. **114**, 071104

Reduced order models: why are they useful?

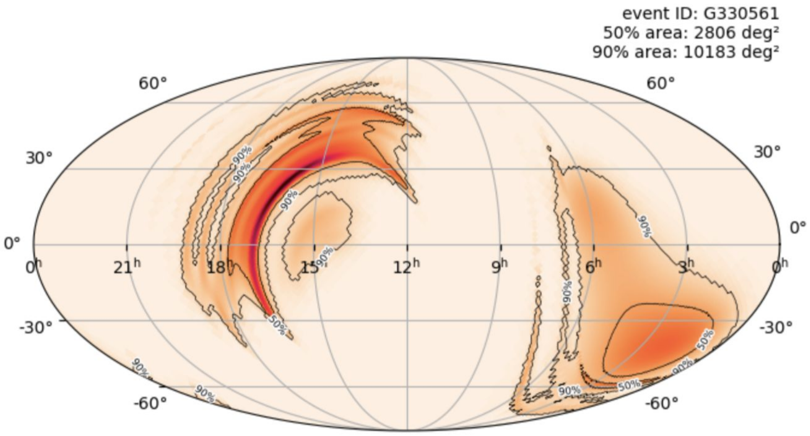
Why they will be useful in O4+

- Need ROMs/Surrogates with as much physics as possible
    - Expect to get more exceptional events as observations continue
        - Non-zero eccentricity?
        - More higher order mode content → better tests of GR
        - Asymmetric mass ratios

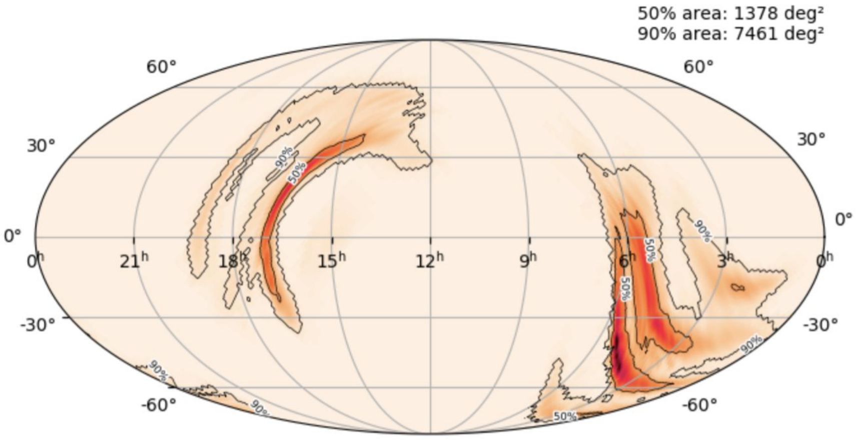
- **Fast and optimal Bayesian sky localization**

Fast sky localization

After a few seconds (BAYESTAR)



After a few hours (bilby)

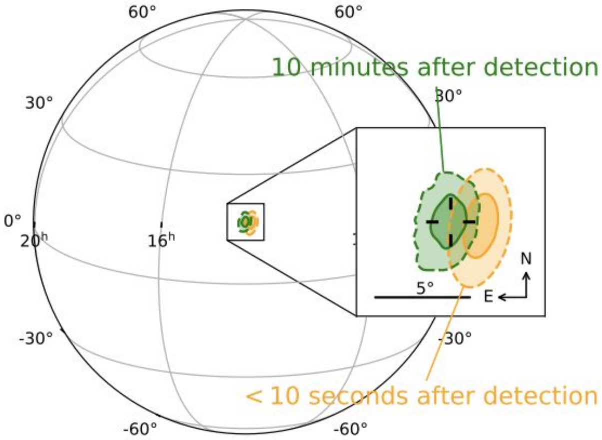


In general, full inference can reduce sky uncertainty by factors of a few, to factors of ten or more

GW190425

Fast sky localization

- Morisaki & Raymond (2019) demonstrated that extremely compact ROMs can be build for binary neutron star mergers
- They demonstrated full Bayesian localization on the order of tens of minutes (around 30-60 mins)

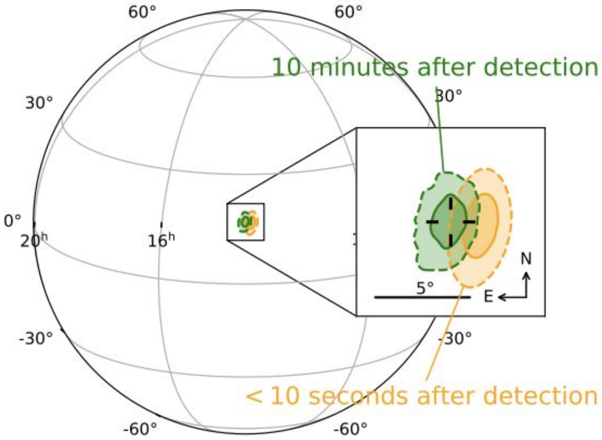


Morisaki & Raymond Phys. Rev. D **102**, 104020 (2020)

Fast sky localization

- Morisaki & Raymond (2019) demonstrated that extremely compact ROMs can be build for binary neutron star mergers
- They demonstrated full Bayesian localization on the order of tens of minutes (around 30-60 mins)

- **Combining ROMs with parallel nested sampling (pbilby) can reduce this time to only a couple of minutes**



Morisaki & Raymond Phys. Rev. D **102**, 104020 (2020)

Reduced order models + parallel sampling

cores	Sampling time (minutes)
64	2.2
16	8.6
8	16.9
2	43.4
1	83.7

Morisaki & Smith (in prep)

Summary

Parallel nested sampling and ROMs are *practical* and *readily available* methods for performing inference on GWs, incorporating detailed physics of BBHs, BNSs and mixed binaries

- Bilby and Parallel Bilby tutorial on Thurs
  - https://git.ligo.org/lscsoft/parallel_bilby
  - https://git.ligo.org/lscsoft/bilby

Should be useful to anyone interested in using bleeding edge waveform/population models for precision astrophysics

Scalable tools for inference will be crucial going forward as event rate increases

- This is an active area of research in and out of the LSC: lots of room to contribute!