Inferring the properties of gravitational-wave signals using Bayesian Inference

Greg Ashton – Royal Holloway

Inference in gravitational wave astronomy

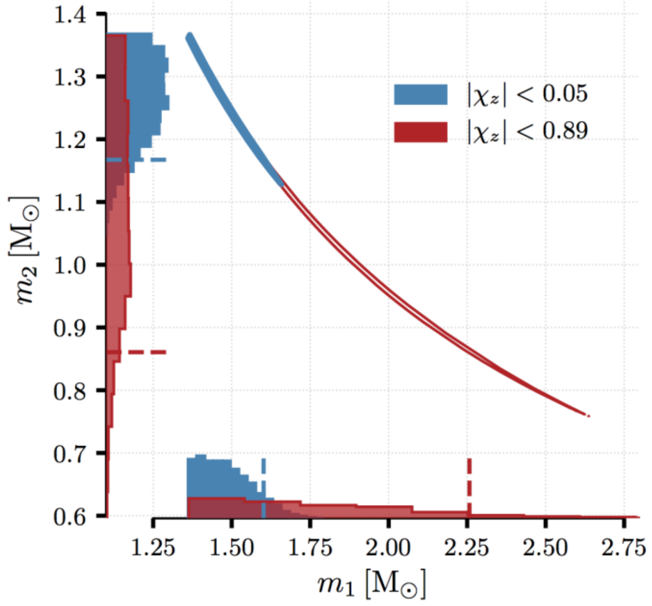
Why do you need Bayesian inference?

1. Non-linear correlations between parameters

2. Informative astrophysical priors

3. Single-event inference fits into hierarchical framework for population modelling

Mass posterior: GW170817



What is Bayesian inference?

- $P(AB)$: "Prob. of A and B" , $P(A|B)$: "Prob of A given B"
- $P(AB) = P(A|B)\,P(B)$
  $\quad\;\; = P(B|A)\,P(A)$
- "Bayes theorem"

$$\boxed{\Rightarrow P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}$$

Parameters : θ

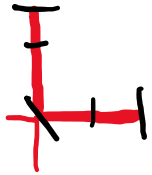
Data : d

Model : M

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

Inference: parameter estimation

The experiment give us data

 $\longrightarrow h(t)$ or $\underline{d}$

The theory $G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$

gives us: Model M

Parameters θ

Bayes theorem:

"posterior" $p(\theta | \underline{d}, M)$ $\propto$ "likelihood" $\mathcal{L}(\underline{d}; \theta, M)$ "prior" $\pi(\theta | M)$

Inference: model selection

The normalizing evidence:

$$p(\theta|\underline{d},M) = \frac{\mathcal{L}(\underline{d};\theta,M)\,\pi(\theta|M)}{Z(d|M)} \Rightarrow Z(d|M) = \int d\theta\, \mathcal{L}(d;\theta,M)\,\pi(\theta|M)$$

Can be used to compare models:

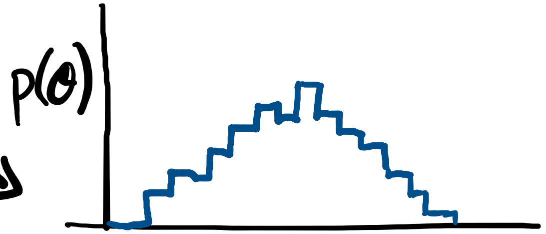
$$\underbrace{\frac{P(M_A|d)}{P(M_B|d)}}_{\text{"Odds"}} = \underbrace{\frac{Z(d|M_A)}{Z(d|M_B)}}_{\text{"Bayes Factor"}}\;\underbrace{\frac{\pi(M_A)}{\pi(M_B)}}_{\text{"Prior odds"}}$$

Stochastic sampling

- Approximate the posterior distribution with "samples"

$$p(\theta|d,M) \sim \{\theta_i\}$$

$$\bar{\theta} = \frac{1}{N}\sum \theta_i$$



- Estimate the evidence numerically

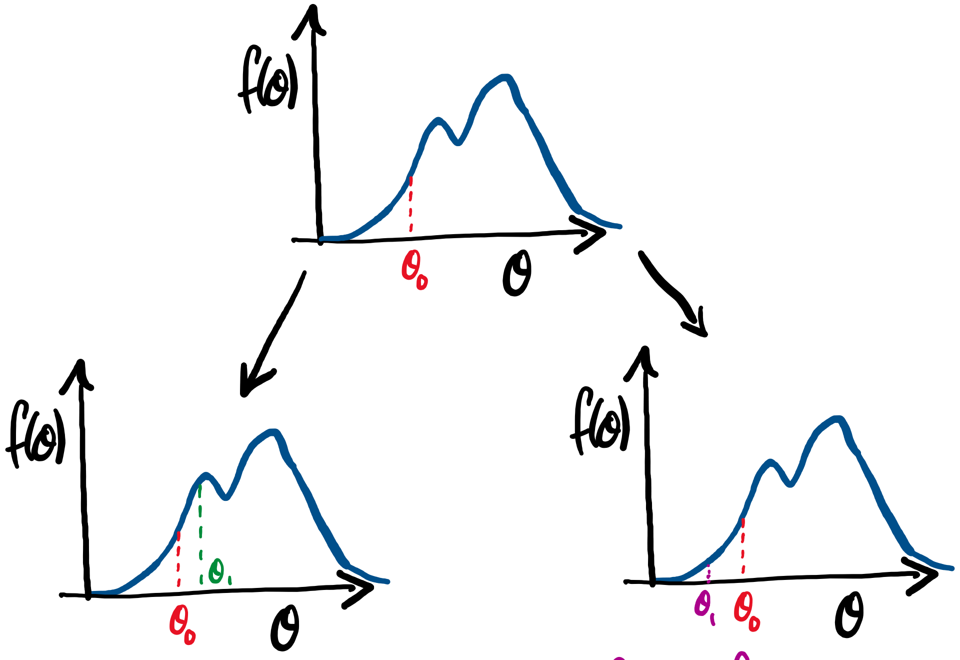
$$Z(d|M) \approx \sum_i \mathcal{L}(d;\theta_i,M)\,\pi(\theta_i|M)\,\Delta\theta$$

MCMC in detail

Algorithm 4 The Metropolis MCMC algorithm to draw samples from a target $f(\theta)$ given an initialization point θ_0.

$c \leftarrow []$	▷ Initialize an empty Markov chain
$c_0 \leftarrow [\theta_0]$	▷ Set the first element of the to an initial value θ_0
for i in range(1, N_{steps}) **do**	▷ Repeat the loop N_{steps} times
$\theta' \sim Q(\theta' \vert c_{i-1})$	▷ Draw a proposed point θ' from the proposal distribution
$u \sim U(0,1)$	▷ Draw a uniform random number u
$\alpha \leftarrow f(\theta')/f(\theta)$	▷ Calculate the acceptance ratio α
if $u \leq \alpha$ **then**	
$c_i \leftarrow \theta'$	▷ Accept the proposed point and append it to the chain
else	
$c_i \leftarrow \theta$	▷ Reject the proposed point and append the existing point to the chain
end if	
end for	

1) Target distribution: $f(\theta) = L(d;\theta)\pi(\theta)$
2) Pick θ_0 an initial value
3) Propose a new point θ_1
4) Apply Metropolis algorithm:



$f(\theta_1) > f(\theta_0) \Rightarrow$ ACCEPT

$- f(\theta_1) < f(\theta_0)$

ACCEPT if $\frac{f(\theta_1)}{f(\theta_0)} > u$

REJECT otherwise

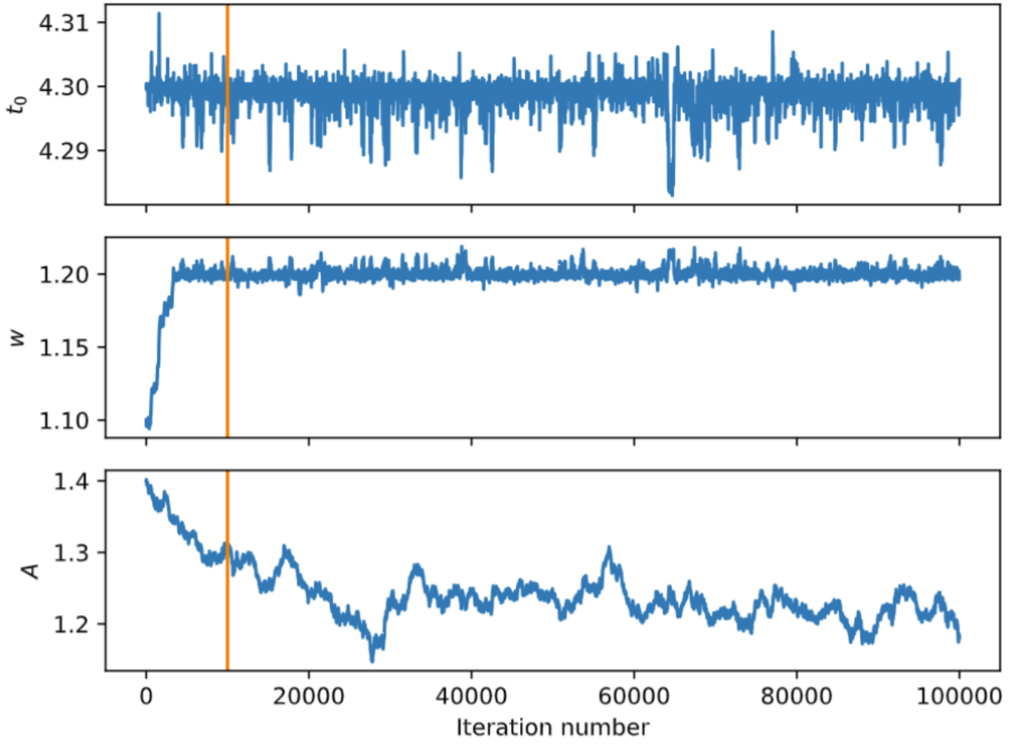
5) Repeat

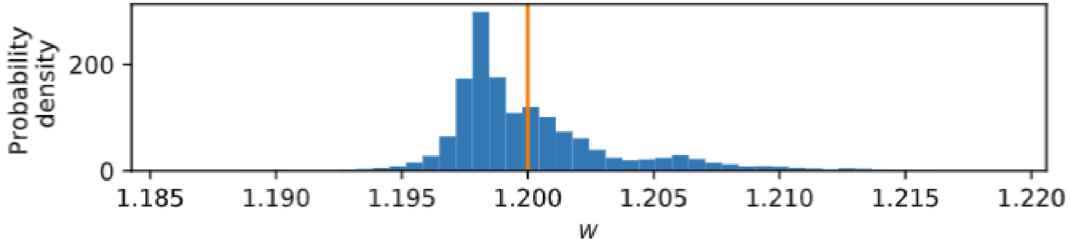
All the subtlety is in the detail

- How do you choose θ_0?
- How do you choose θ_1?
- How many steps should I take?
- Etc..

The output of an MCMC sampler

- A chain of samples
- Samples will be repeated
- Samples will be correlated
- Need to:
  - Estimate autocorrelation
  - Remove the “burn in”
  - Thin to produce uncorrelated samples
- Resulting samples can:
  - Bin in histogram to produce “posterior”
  - Calculate mean/uncertainty for summary statistics





A historical overview

- First approaches based on Markov-Chain Monte-Carlo (MCMC) [Christensen & Meyer 1998]
- Then, Nested Sampling was introduced [Veitch & Vecchio 2008]
- A grid and sampling approach has since been applied by RIFT [Pankow et al. 2015]
- LALInference [Veitch et al. 2015]:
  - Used as flagship software O1-O3a
  - Offered both an MCMC and Nested Sampling package
  - Remains in active use
  - Ports to TGR based codes likely to remain in use for some time

The new era

LALInference is great, why build anything else?

- C-based code is hard to modify/extend for new users
- Pragmatically: the developers had permanent positions and the developer base was running thin

A host of new approaches have since been developed:

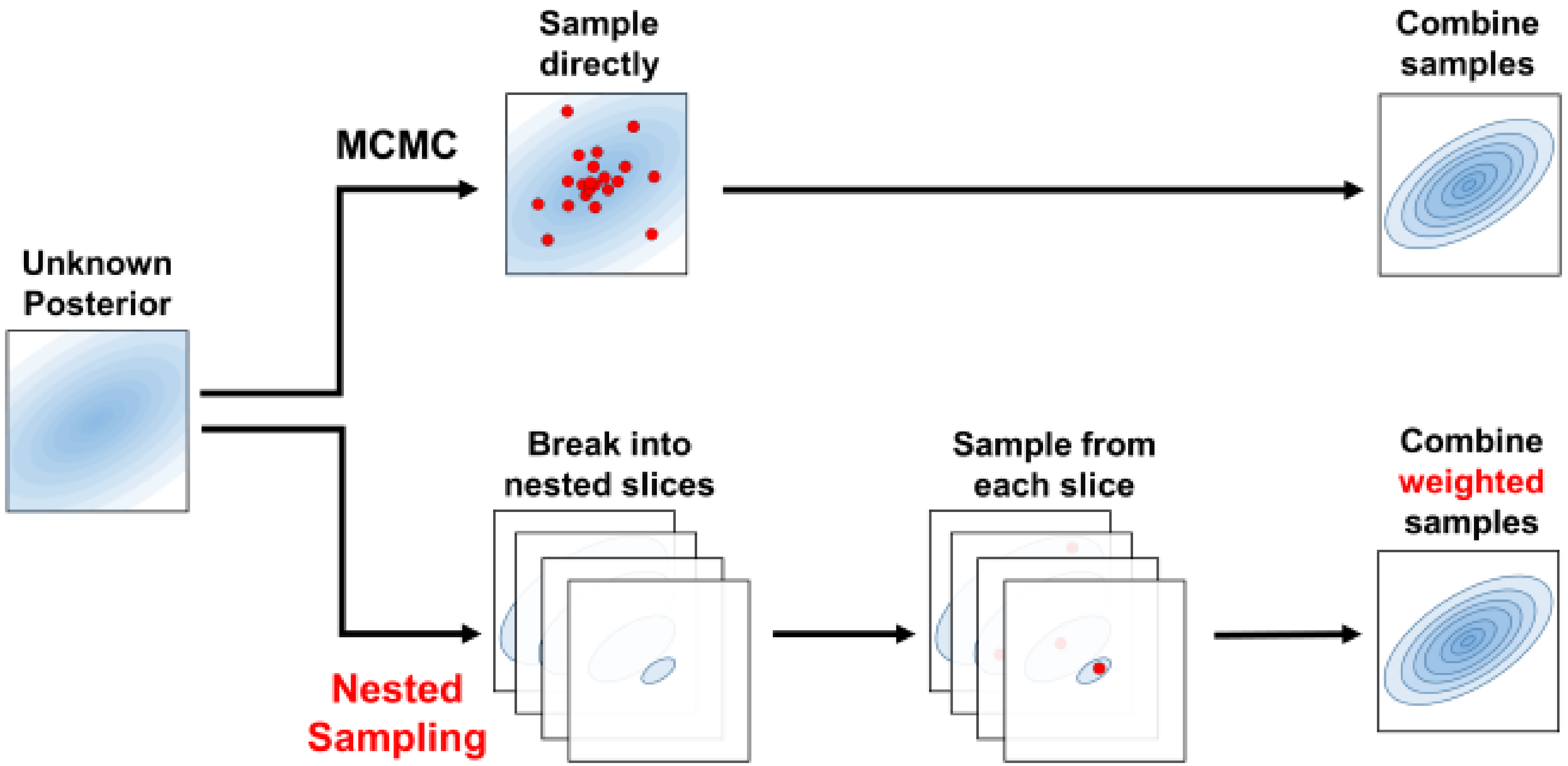
- PyCBC inference [Biwer et al. 2018]
- Bilby [Ashton et al. 2018]
- Bajes [Breschi et al. 2021]
- gwmodel [Pagano et al. in prep]
- + more which I have missed (please let me know!)

Bilby

- Package adopted by LIGO-Virgo-KAGRA for use in O3 onwards
- Interface between Bayesian inference concepts and **off-the-shelf** MCMC and NS stochastic sampling packages
- Design principles:
  - Modularity, Consistency, Generality, Usability
- These principles have:
  - Lowered the barrier to users
  - Enabled rapid development of new ideas/approaches
- Used to analyze: CBCs, CWs, FRBs, GRBs, X-ray afterglows, …

Nested Sampling vs MCMC

- Two different approaches to the same problem



Speagle (2019)

Which is better?

- **Both**: cross-checking results between samplers has been vital
- LALInference MCMC was easier to parallelize:
  - Run tens of independent jobs and combine
  - Fits the high **throughput** computing (HTC) model of the LIGO Data Grid
- NS generally produces a more robust estimate of the evidence
- MCMC generally produces a more robust estimate of the posterior

Replacing LALInference MCMC?

- Bilby enables access to several off-the-shelf MCMC packages
- None of them have been validated in Bilby

Sampler	Ensemble?	Parallel Tempering?	GW-tuned proposals?	Cross-check validated?
LALInfernce-MCMC	No	Yes	Yes	Yes
emcee	Yes	No	No	No
ptemcee	Yes	Yes	No	No*
Bilby-MCMC	**Yes****	**Yes**	**Yes**	**Yes**

- Bilby-MCMC [Ashton & Talbot 2021] was conceived to implement an MCMC sampler in Bilby with GW-tuned proposal

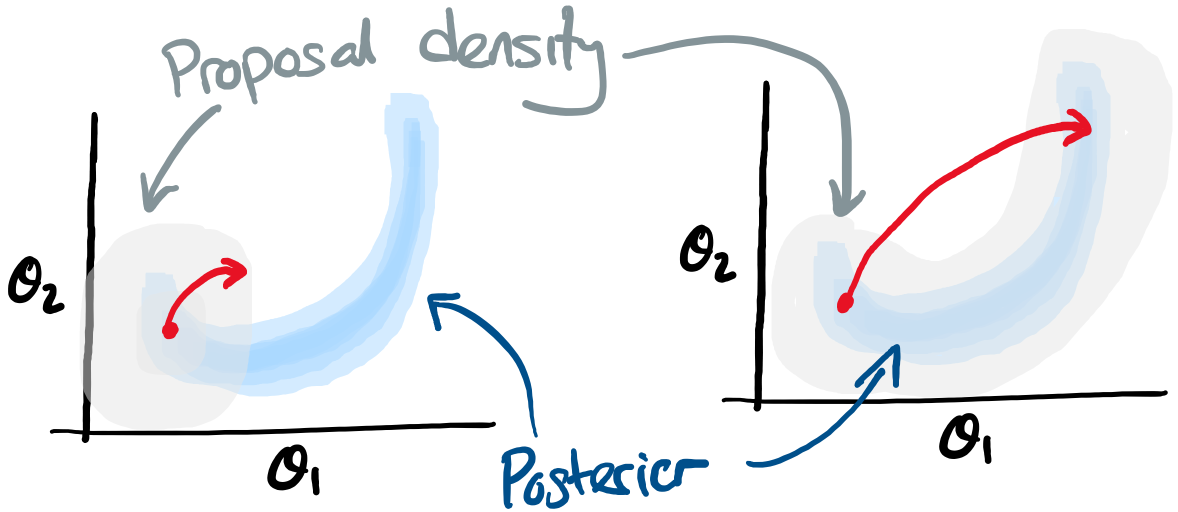
*ptemcee has been used with success by PyCBC-inference; **The ensemble approach was found to be ineffective

Bilby-MCMC

Details in [Ashton & Talbot 2021]

Improved algorithmic efficiency:

- GW-tuned proposals
- Machine-Learning proposals:
  - Use distribution to learn efficient proposal density
  - Normalizing Flows, Kernel Density Estimates, and Gaussian Mixture Models

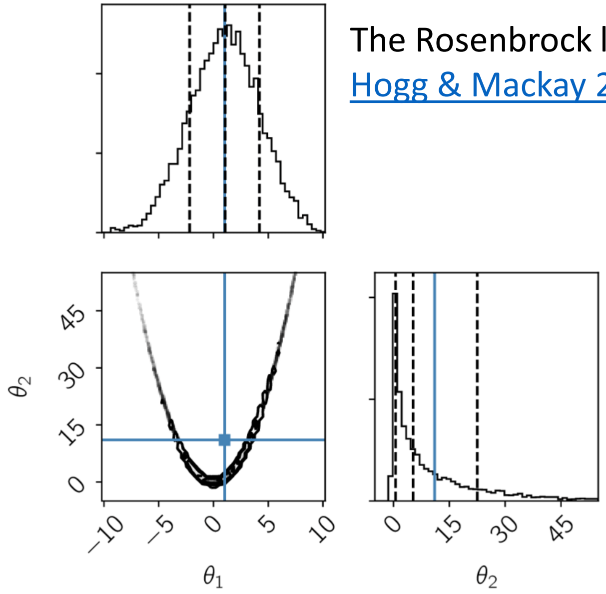


Bilby-MCMC enables the use of HTC parallelization

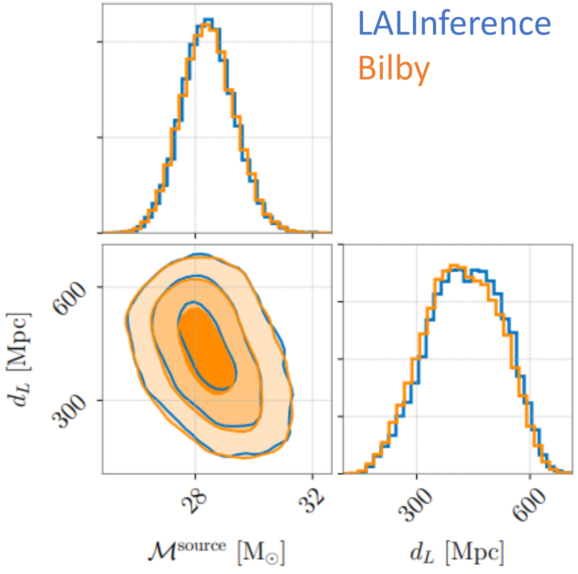
The devil is in the details

[Romero-Shaw et al. 2021](#) validated Bilby

1. Using the dynesty Nested Sampling package
2. Demonstrated robust inference using standardized tests
3. Demonstrated “statistically identical” to LALInference for CBC inference problems
4. We now utilize Bilby-MCMC and other samplers for validation checks



The Rosenbrock likelihood: [Hogg & Mackay 2018](#)



LALInference
Bilby

Toward O4: optimizing for performance

Typically, we need

$$n_L \sim 10^8 \rightarrow 10^9$$

evaluations of $\mathcal{L}(d|\theta, M)$

Meanwhile:

$$t_L \sim ms \rightarrow s$$

$\Rightarrow$ Wall time $= n_L \times t_L \sim$ days $\rightarrow$ years

For CBC systems containing a neutron star, EM observers want the best possible skymap ASAP!

Optimize to reduce the time to produce a skymap

Waveform models with "better physics" tend to be more computationally expensive

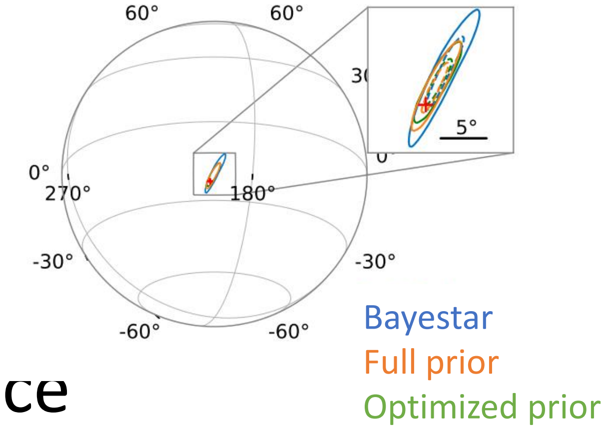
Optimize to enable the use of better waveforms

In O4, we will have hundreds of CBCs to analyze. Multiple analyses needed to investigate physics

Optimize to improve the results!

How to optimize?

- Reduce the cost of the likelihood
  - Use of ROQs and heterodyning can reduce t_ℓ by factors of 1000s
- Optimize the choice of prior
  - You, Ashton, et al. 2021
  - Can yield speed-ups by factors of 2



- Better algorithms/faster convergence
- Parallelization

Optimization: the likelihood

The waveform tends to dominate

- For BBH typically more than 90%
- For BNS can be more like 50%
- More physics in the waveform increases t_ℓ
- Some waveforms take several seconds

Waveform approximant	per-likelihood evaluation [ms]	per-waveform evaluation [ms]
IMRPhenomPv2_NRTidal	93 ± 5	53 ± 4
IMRPhenomPv2	87 ± 6	47 ± 4
TaylorF2	60 ± 8	13.3 ± 0.7

Timings for a fiducial BNS signal [You, Ashton, et al. 2021]

Two approaches available to reduce t_ℓ:

- Reduced Order Quadrature [E.g., Smith et al. 2016, Morisaki & Raymond 2020]
- Hetrodyned likelihood [Cornish 2010] / Relative binning [Zackay et al. 2018]
- Both offer reductions in t_ℓ over over 1000
- No silver bullet solution to any waveform yet..

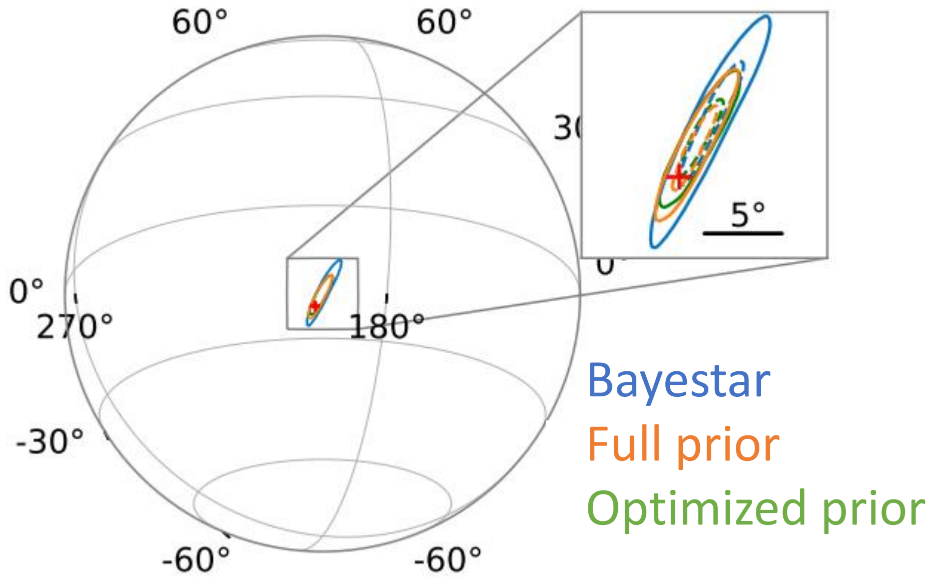
Optimization: priors

For astrophysical applications, priors should represent our prior belief

But, if we need a skymap as soon as possible:

- Optimize the prior:
  - For BNS/NSBH systems non-spinning and equal-mass can reduce wall-times by up to 50% (on top of other optimizations)

  - For BBH systems such optimization is unwise



GW170817 skymap [You, Ashton, et al. 2021]

Optimization: faster convergence

Can we reduce this?

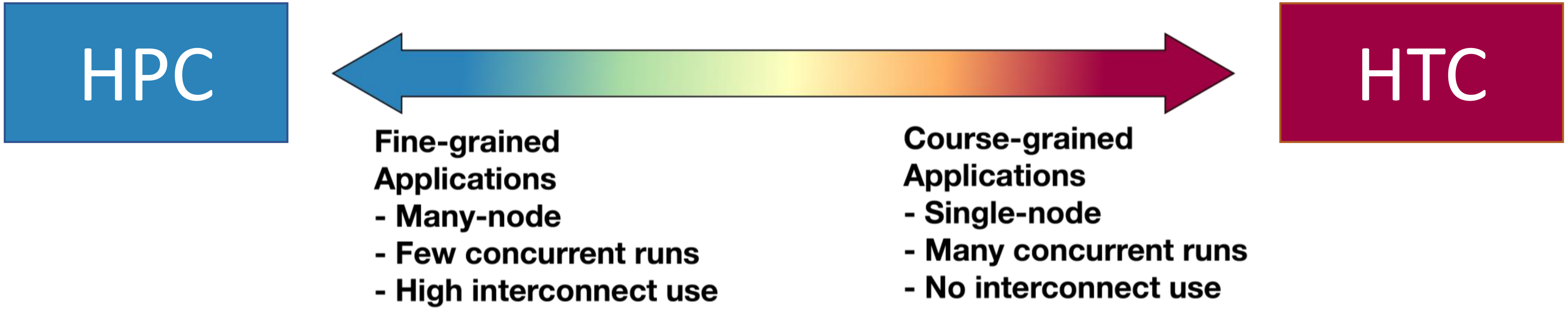
Yes: improve the convergence

1. Use analytic marginalization of parts of the likelihood [See Thrane & Talbot 2019 for a review]

2. Improve the sampling efficiency:
   - Improve the proposal density (see, e.g. Williams et al. 2021 for ML approach)
   - CBC-specific jump proposals important for the LALInference MCMC sampler

Optimization: parallelization I

- High Performance Computing (HPC)
- High Throughput Computing (HTC)

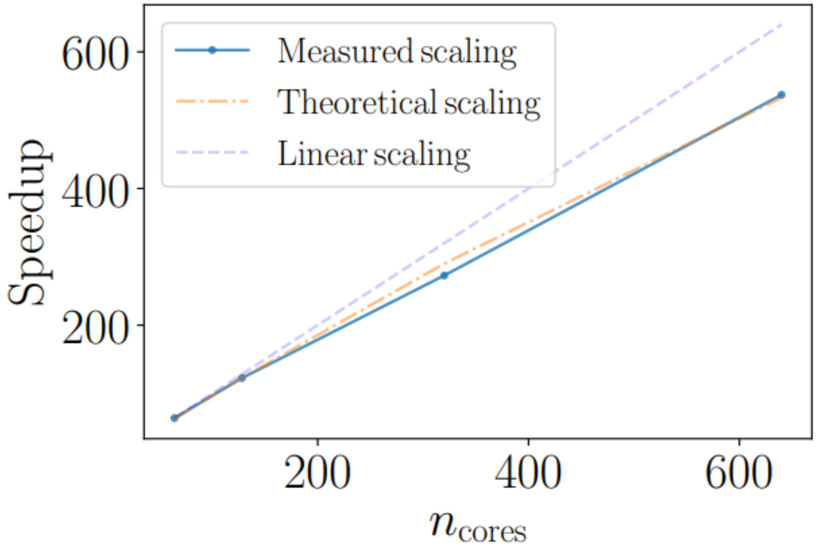


Optimization: HPC parallelization

- The nested sampling algorithm can be parallelized using HPC

- Smith, Ashton et al. 2020 introduced parallel-Bilby
  - Uses MPI to leverage HPC environments
  - Achieved near-linear scaling

- For GW190412:
  - Using 16 cores (i.e., standard analysis): ~3 years
  - Using 640 cores (entire cluster): 12 days



Optimization: HTC parallelization

- The LIGO Data Grid & Open Science Grid are predominantly HTC
- Moreover, it is difficult to organize taking over an entire cluster
- How did we do things before we had parallel-Bilby?

MCMC is trivially parallelizable

- Run N independent runs
- Each chain produces n samples
- Combine them together
- Used with great effect by LALInference MCMC

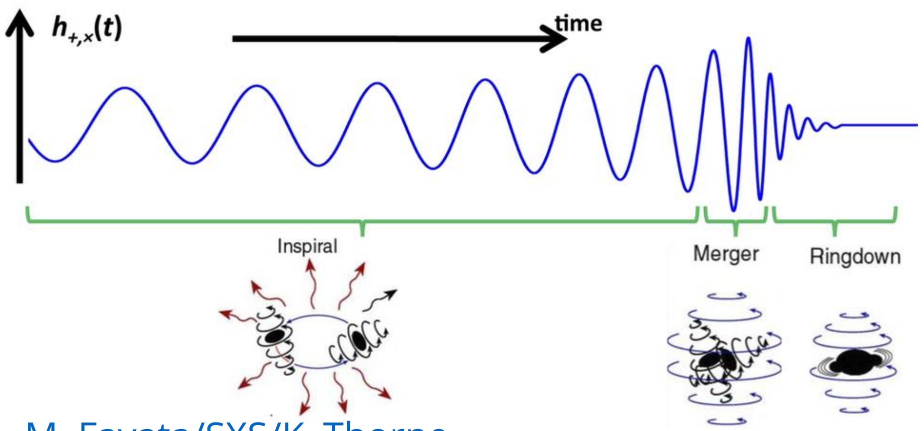
- But we don't have an effective MCMC sampler in Bilby..

Optimization: take-aways

- In O4, we could achieve posteriors for BNS/NSBH in < 1hr
  - Main speed up comes from likelihood optimization fROQ/heterodyne
  - Further optimizations from prior and model

Modelling uncertainty

- Numerical Relativity is our only means to model CBC mergers in GR
- Computationally infeasible to model the full signal
- Waveform models combine the inspiral, merger, and ringdown using different approximations
- This results in **systematic waveform uncertainty**
- Different waveforms make different predictions



M. Favata/SXS/K. Thorne

Modelling uncertainty: current approach

As of the GWTC3 catalogue, the current approach is:

1. Analyse each event with two waveform models
   - In practise, two different stochastic samplers are also applied

2. Compare between the waveform models to look for cases where they disagree

3. Combine equal numbers of samples from each waveform to produce a posterior which captures the uncertainty

Modelling uncertainty: improved approach

- The current approach **neglects the evidence**

- The evidence tells us how well each model explains the data

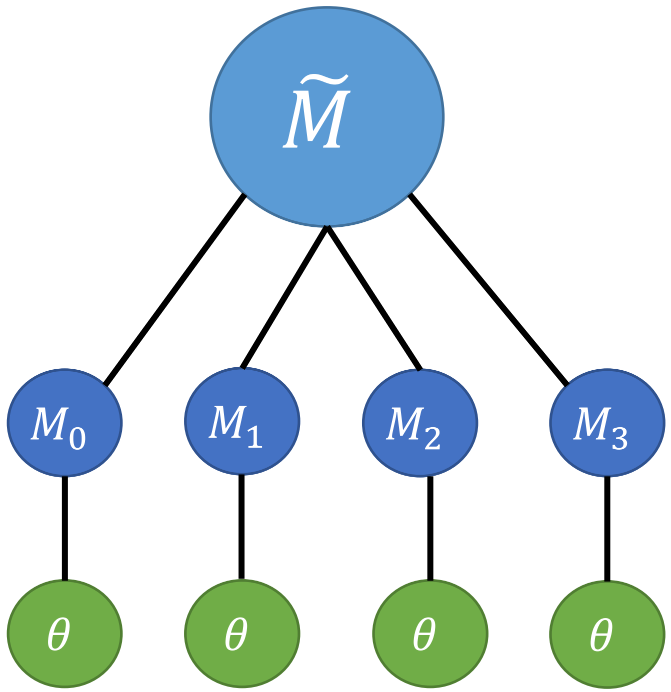
- Ashton & Khan 2020 demonstrate how to use the evidence to produce weighted posteriors

$$\mathcal{E}_i = \frac{\mathcal{Z}(d|M_i)}{\sum \mathcal{Z}_i}$$

- Pragmatic difficulties persist:
  - Relies on robust evidence estimates
  - Have both your analyses used identical data, likelihood, priors, models?

Modelling uncertainty: hypermodel approach

- Alternative to evidence-based approaches
- Define a hierarchical **hypermodel** $\widetilde{M} = \{M_0, M_1, \dots, M_N\}$
- Calculate the posterior of $p(\theta |\ \text{data}, \widetilde{M})$
  - During sampling, first propose jumps between models
  - Then analyze the likelihood under the given waveform
  - Uses a special Reversible-Jump MCMC implemented in Bilby-MCMC

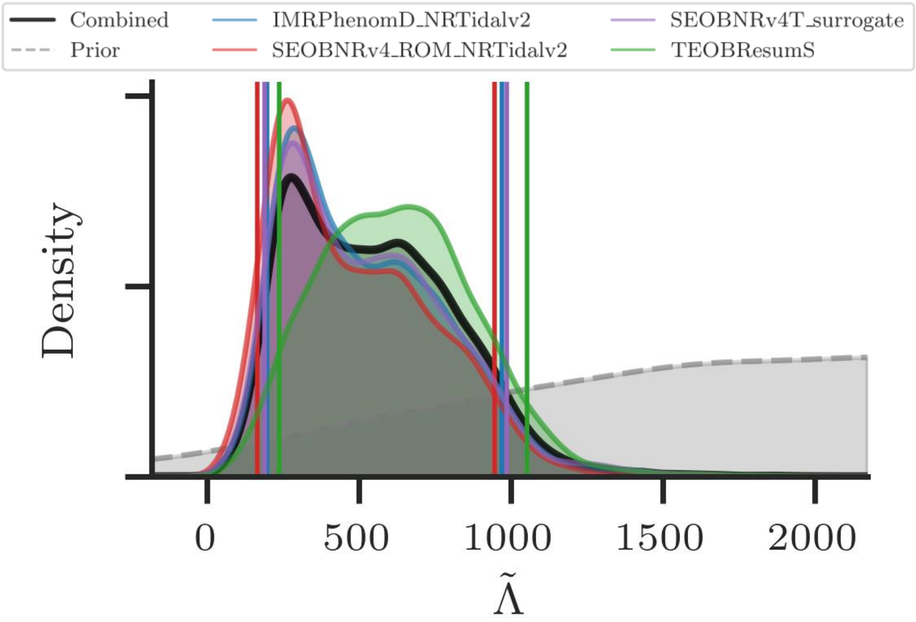


Benefits:

- "All in one" analysis approach
- Produces a posterior marginalized over $\widetilde{M}$
- Posteriors for individual waveforms can be "pulled out"
- Odds between models can be calculated

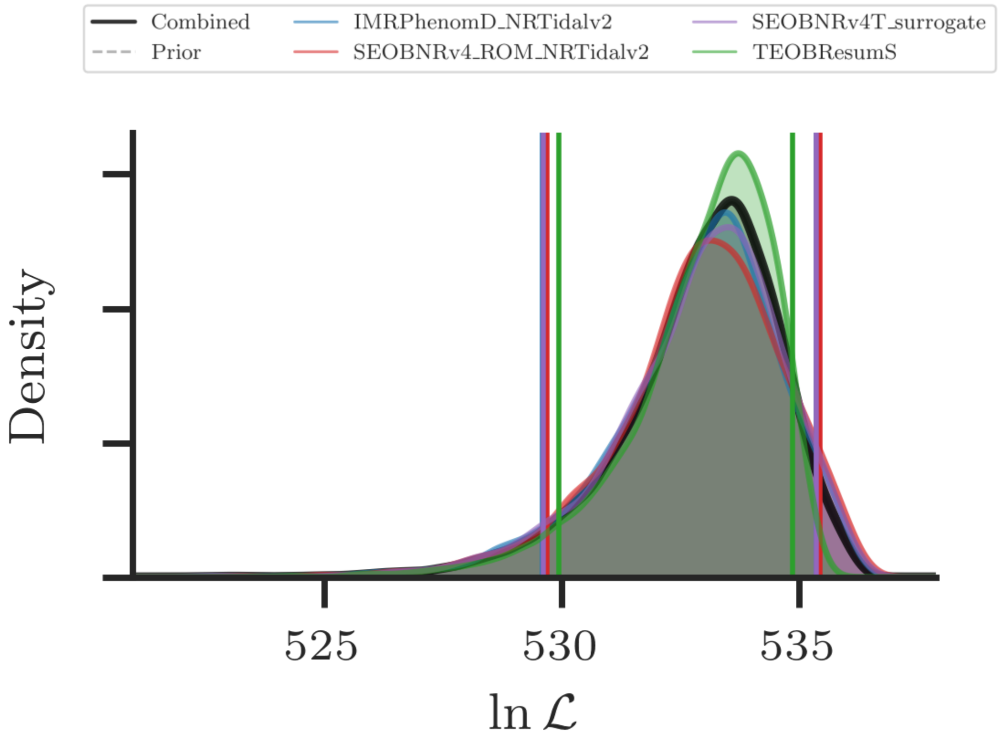
Applying hypermodels to observed BNS

- Ashton & Dietrich 2021: GW170817, GW190525,and GW200311_103121
- Four cutting-edge aligned-spin BNS models
- **Consistent preference TEOBResumS**
  - Bayesian odds range from 1.7 – 2.3
  - Not conclusive, but tantalizing
- Evidence suggests it is the tidal sector
  - Implications for BNS physics
  - Predictions of larger tidal deformability
- Combining the events in O4 will cement this result



The failure of the maximum likelihood

- TEOBResumS does not have the largest likelihood!
- It is the distribution which matters
- Important demonstration of why Bayesian approaches matter



Thank you for listening