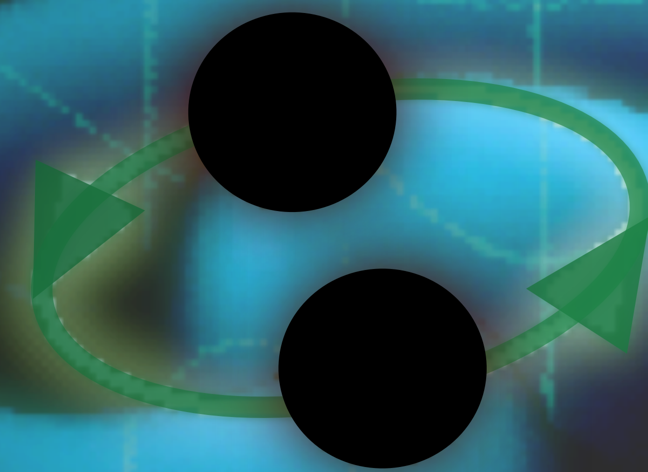
What will advanced detectors teach us about compact binaries, and how?

John Veitch

Cardiff University

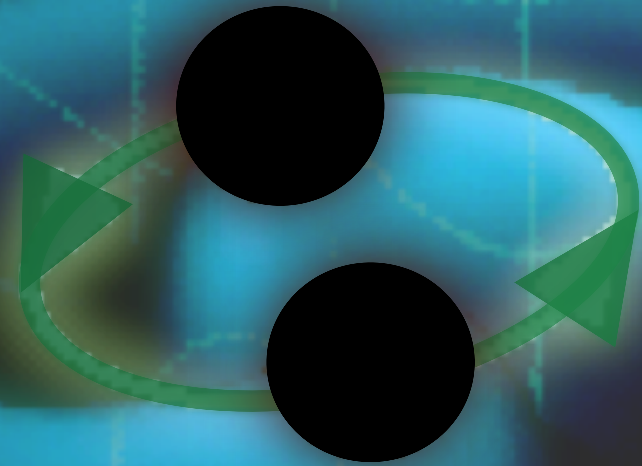
2013-4-12

Contents



- Gravitational waves from compact binaries
- Parameter Estimation
  - Status & recent developments
  - Road to advanced detectors
- CBCs as probes of astrophysics
  - Combining multiple sources

Compact Binaries

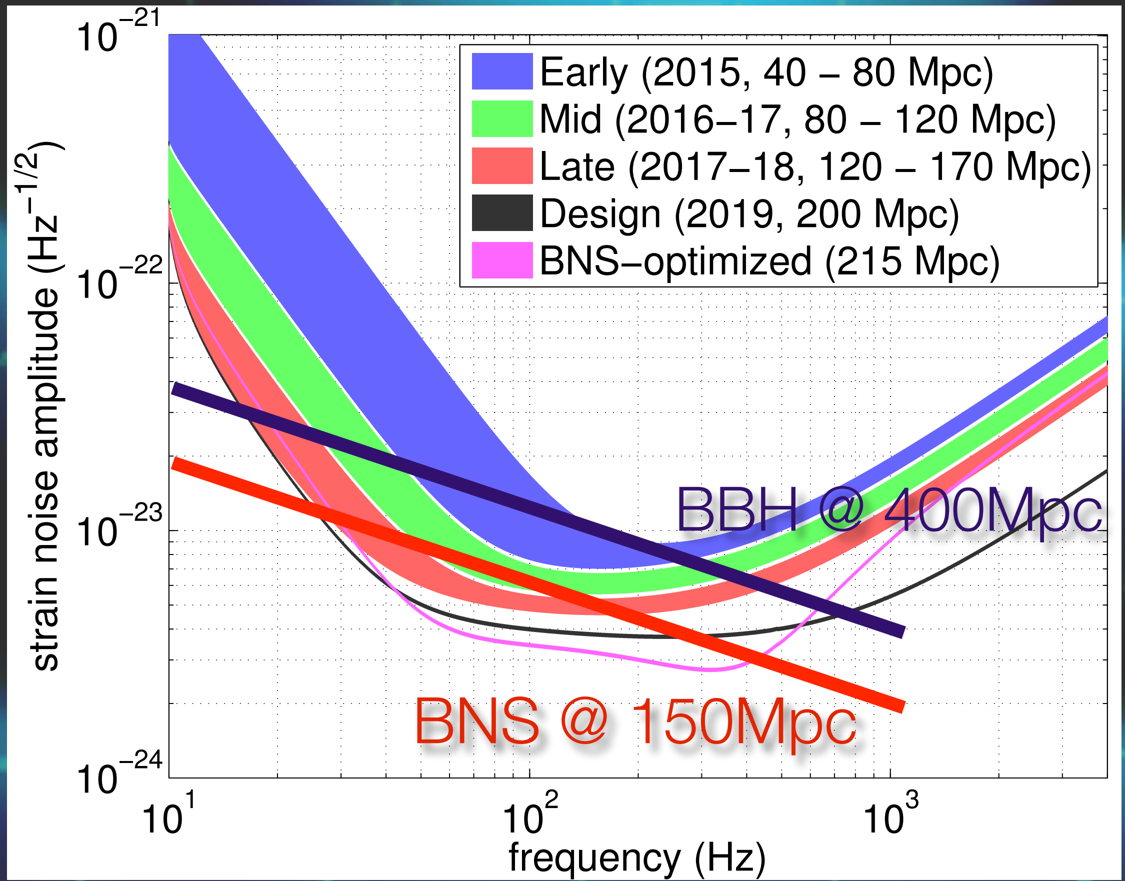


Coalescence of black holes and neutron stars in binaries are the primary candidate for ground-based GW detectors.

Astrophysical event rates uncertain:

- NS-NS: 0.01 - 10 / (Mpc^3Myr)
- NS-BH: $4x10^{-4}$ - 1 / (Mpc^3Myr) (theoretical)
- BH-BH: $1x10^{-4}$ - 0.3 /(Mpc^3Myr) (theoretical)

[Abadie *et al* Class. Quant. Grav 27 172001]



Translated to BNS detections:

Early (2015): 0.0004 - 3

Mid (2016-2017): 0.006 - 20

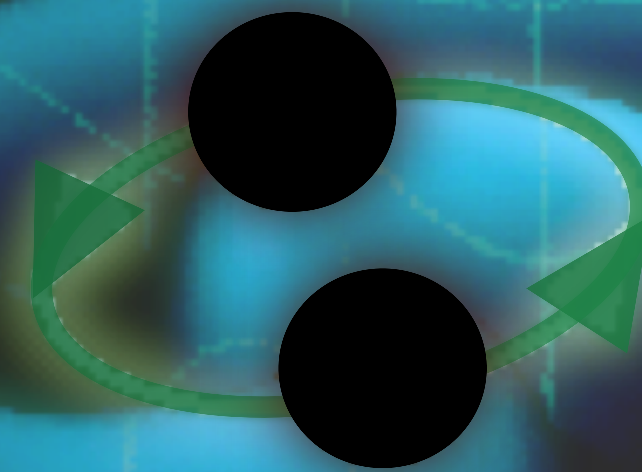
Late (2017-2018): 0.04 - 100

Final: 0.2 - 200

[LIGO-DCC P1200087]

Beyond confirming the existence of gravitational waves, what can they tell us about astrophysics?

What parameters?



What parameters are there are to measure for an individual source?

- The simplest model of a compact binary in a circular orbit has 9 parameters

2 masses, time, sky position, distance, 3 orientation angles

- Add spins for another six

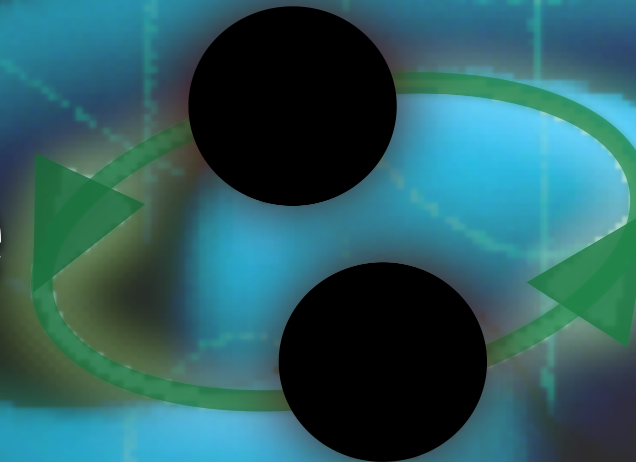
2 magnitudes, 4 orientation angles

- and neutron stars have equation of state parameters

Tidal deformability of each star λ_1, λ_2

- Then there are possible deviations from the GR model...

Model gravitational wave



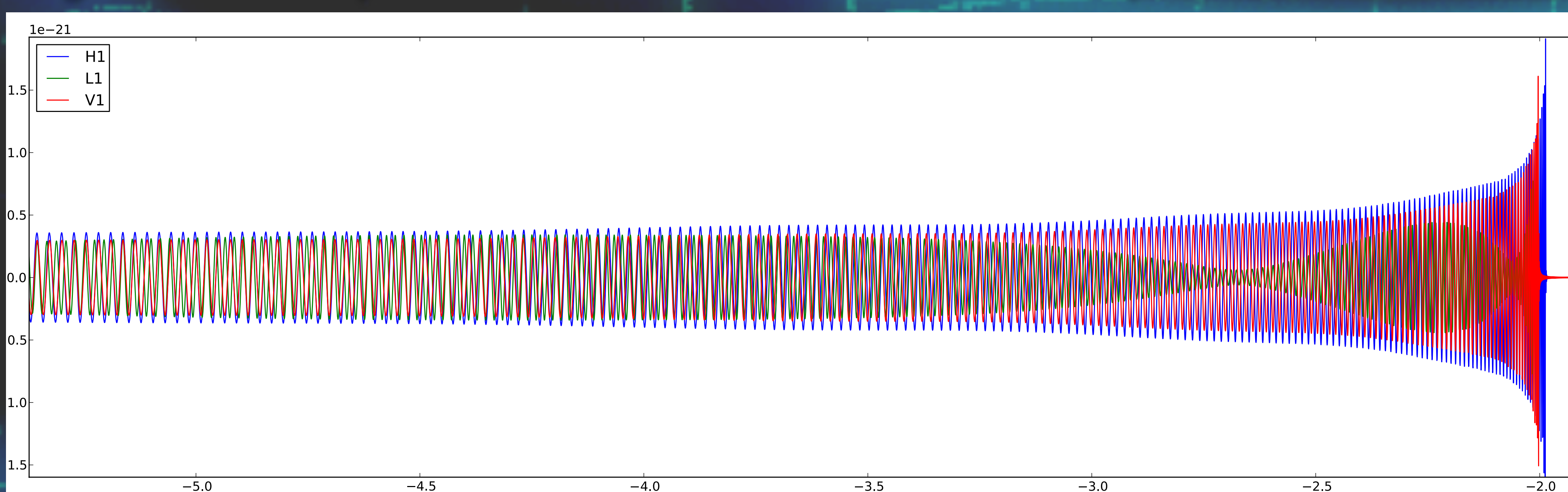
The response of a given detector to an incoming CBC signal is a complicated function of the physical parameters:

$$h_i(t) \quad = A_+(t-\Delta t_i)F_+^i(t-\Delta t_i)\exp\Phi_+(t-\Delta t_i) + A_\times(t-\Delta t_i)F_\times^i(t-\Delta t_i)\exp\Phi_\times(t-\Delta t_i)$$

Can categorise them as instrinsic: $\Phi_+(t), \Phi_\times(t) \leftarrow m_1, m_2, \vec{S}_1, \vec{S}_2, t, \phi_0$

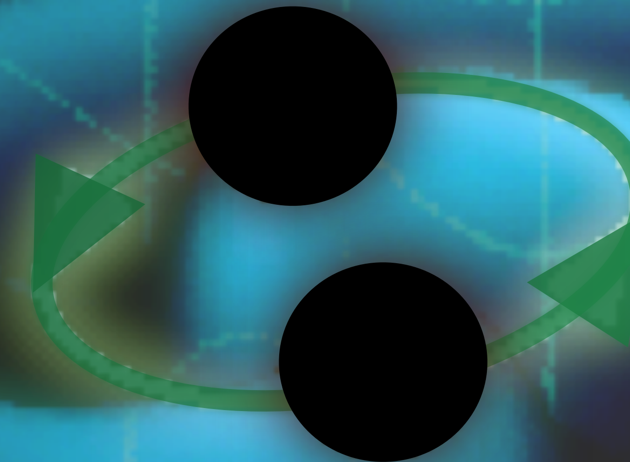
and extrinsic $A_+, A_\times, F_+^i, F_\times^i, \Delta t_i \leftarrow d_L, \alpha, \delta, \psi, \iota$

(assuming we know the position and orientation of the detectors)



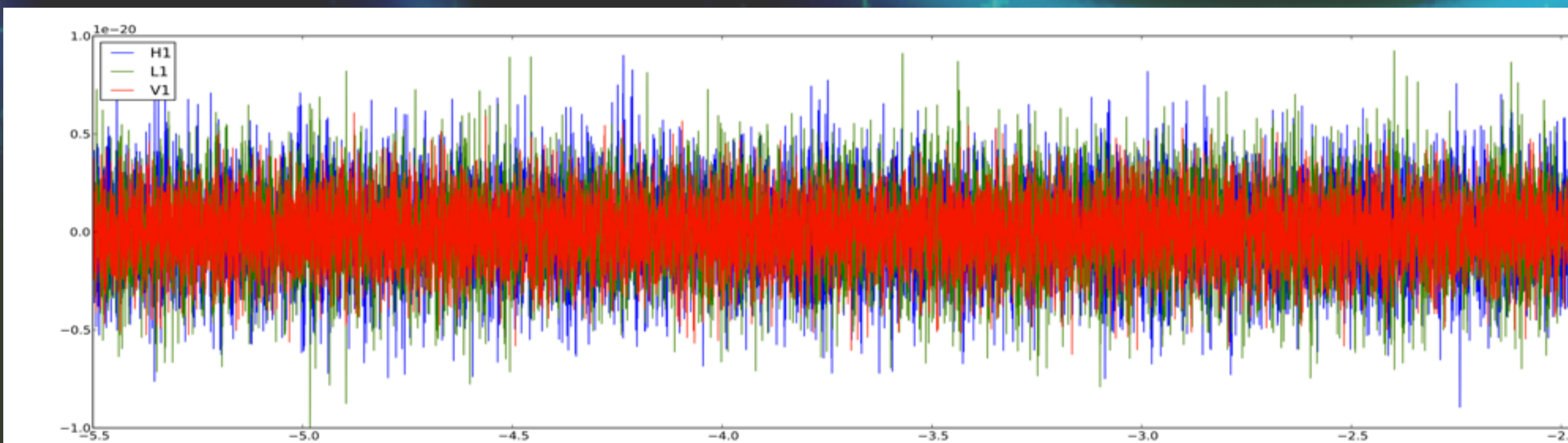
5

Noisy detectors



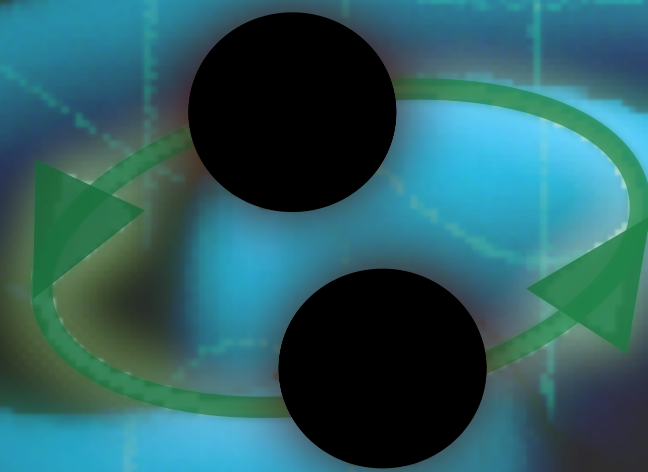
Unfortunately signal is buried in additive noise from the detector:

$$d_i(t) = h_i(t) + n_i(t)$$



Must also model the noise!

Noise model



Simplest noise model makes few assumptions:

- zero mean: $\langle n_i \rangle = 0$
- known variance: $\langle n_i^2 \rangle = S_h(f_i)\, \Delta f = \sigma_i^2$

Maximum entropy distribution is Gaussian, i.e. $p(n_i|\sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{n_i^2}{2\sigma_i^2}\right]$

- Assuming independence in each frequency bin:

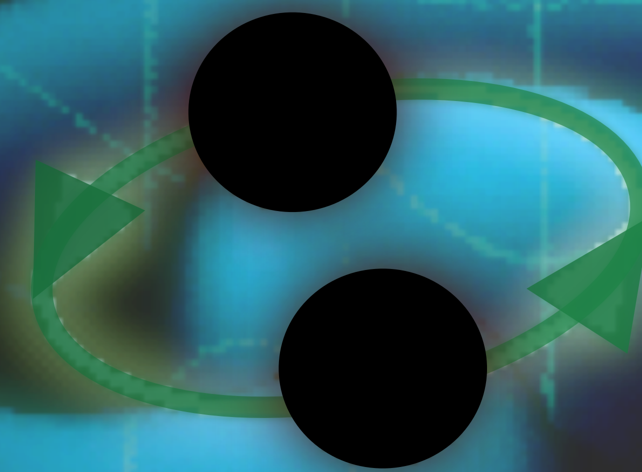
$$p(\{n_i\}|\{\sigma_i\}) = \prod_i p(n_i|\sigma_i) = \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{n_i^2}{2\sigma_i^2}\right]$$

- And independence in each detector:

$$p(\vec{n}_H, \vec{n}_L, \vec{n}_V | \vec{\sigma}_H, \vec{\sigma}_L, \vec{\sigma}_V) = p(\vec{n}_H|\vec{\sigma}_H)\, p(\vec{n}_L|\vec{\sigma}_L)\, p(\vec{n}_V|\vec{\sigma}_V)$$

- Terminology: the “likelihood” of the noise

Likelihood function



For additive noise, $d_i=h_i+n_i$, the mean of the data distribution becomes the prediction of the signal model for given parameters θ: $\langle d_i \rangle = h(f_i,\theta)$, whereas variance remains the same.

Putting everything together, we have a model of the data containing an unknown signal and noise, and can write the joint distribution of all quantities as

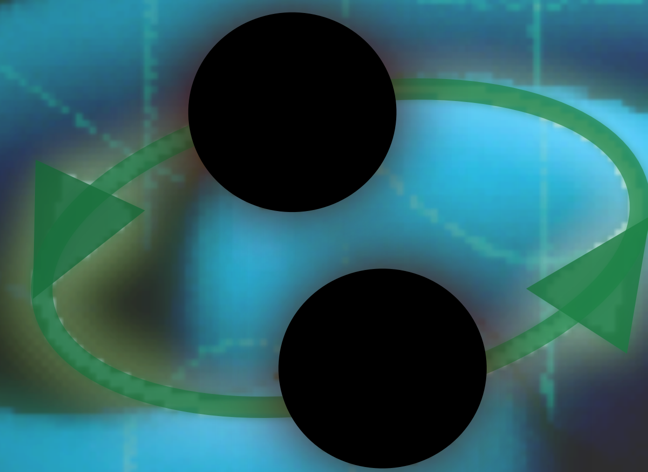
$$\begin{aligned} p(\vec{d},\vec{\theta}|\vec{\sigma},H_S) &= p(\vec{d}|\vec{\theta},\vec{\sigma},H_S)p(\vec{\theta}|\vec{\sigma},H_S) \\ &= p(\vec{\theta}|\vec{d},\vec{\sigma},H_S)p(\vec{d}|\vec{\sigma},H_S) \end{aligned}$$

If we then observe a specific set of data, we can infer the parameters of the signal by calculating the posterior probability distribution (PDF):

$$p(\vec{\theta}|\vec{d},\vec{\sigma},H_S) = \frac{p(\vec{\theta}|\vec{d},\vec{\sigma},H_S)p(\vec{d}|\vec{\sigma},H_S)}{p(\vec{d}|\vec{\sigma},H_S)}$$

where $p(\vec{d}|\vec{\sigma},H_S) = \int_{\vec{\theta}\in\Theta} d\vec{\theta}\, p(\vec{d}|\vec{\theta},\vec{\sigma},H_S)p(\vec{\theta}|\vec{\sigma},H_S)$ is the *evidence* of the model.

Model Selection



In detector data we do now know if a signal is present or not, so we can compute the relative probability of the signal and null (noise) hypotheses - “odds ratio”

$$\frac{p(H_S|\vec{d})}{p(H_N|\vec{d})} = \frac{p(H_S)\; p(\vec{d}|H_S)}{p(H_N)\; p(\vec{d}|H_N)}$$

Note: search pipelines *exclude* background, calculating $1/p(d|H_N)$ ~ “inverse false alarm probability”.

Incoherent search pipelines matched filter each detector independently: valid given the independence of noise in each detector.

Information about the signal (priors, coincidence in time, parameter consistency between detectors, multi-detector coherence...) belongs on the numerator.

Can also compute odds between different models to compare theories of gravity, neutron star equations of state, etc.

Calculating posteriors

Nothing so far is specific to searching or parameter estimation.

But, the waveform model of a CBC signal in a given detector depends on 9 or more physical parameters,

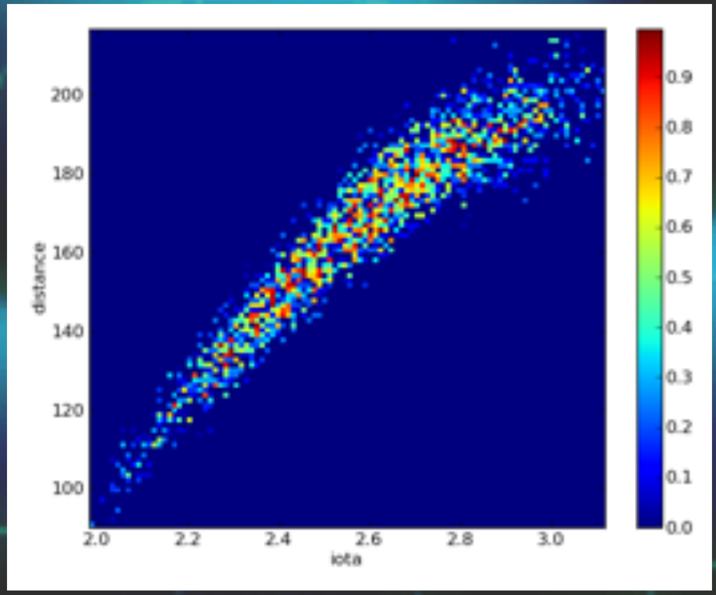
$$\vec{\theta} = \{\mathcal{M}, \eta, t_c, \phi_c, d_L, \alpha, \delta, \iota, \psi, \ldots\}$$

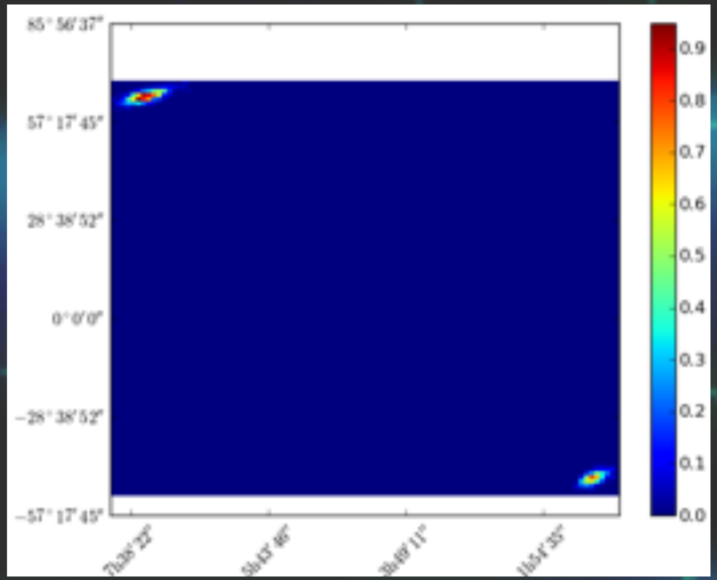
and the likelihood function is:

- non-linear: difficult to find maximum
- non-invertible: multiple maxima

To fully search this space need $>2^{30}$ templates. This is impractical, so what to do?

- Search codes: Maximise! Reduce dimensionality by eliminating extrinsic parameters
- Parameter estimation codes: Sample posterior PDF



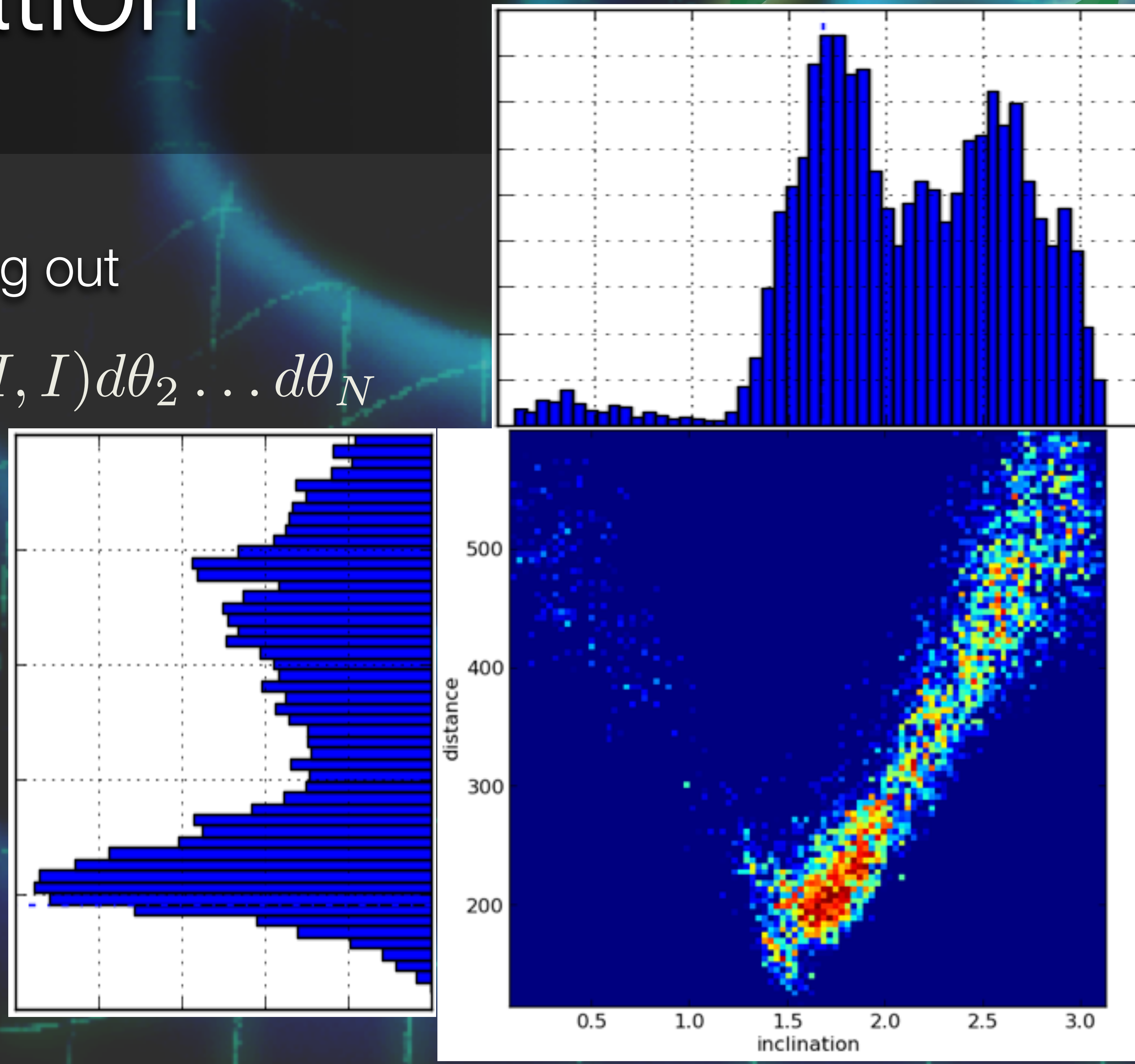


Marginalisation

Eliminate nuisance parameters by integrating out

$$p(\theta_1|H,I) = \int p(\vec{\theta}|H,I)d\theta_2 \ldots d\theta_N$$

Can be done easily with collection of posterior samples by producing histograms for the parameters of interest. Can use 2D histograms to see correlations, etc



11

Nested Sampling

Nested sampling [Skilling 2004] is designed to calculate the evidence integral probabilistically using random samples

$$Z = p(\vec{d}|\vec{\sigma}, H_S) = \int_{\vec{\theta}\in\Theta} d\vec{\theta}\, p(\vec{d}|\vec{\theta}, \vec{\sigma}, H_S)\, p(\vec{\theta}|\vec{\sigma}, H_S)$$

$$p(\vec{d}|\vec{\theta}, H_S)\, p(\vec{\theta}|H_S) = Z p(\vec{\theta}|d, H_S)$$

We know from product rule that

Define $X(\lambda) = \int_{L(\vec{\theta})>\lambda} p(\vec{\theta}|H_S)\, d\vec{\theta}$

as the prior mass enclosed by a contour of equal likelihood λ.

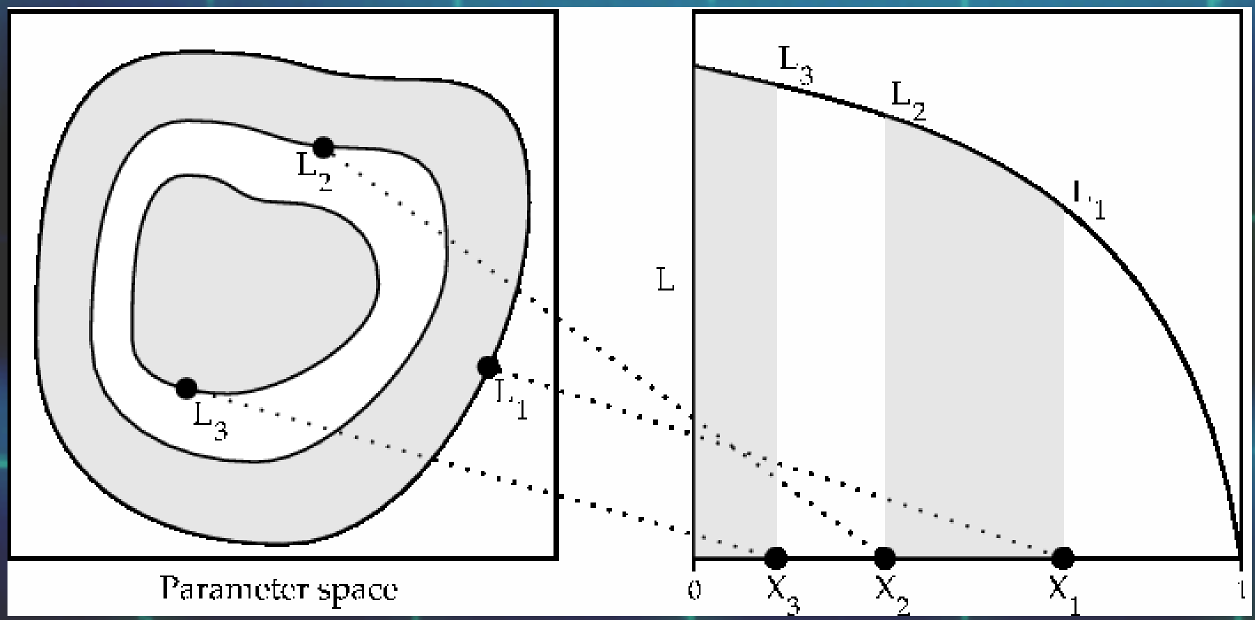
X is a monotonic function of λ between λ_{min} and λ_{max}, the likelihood range.

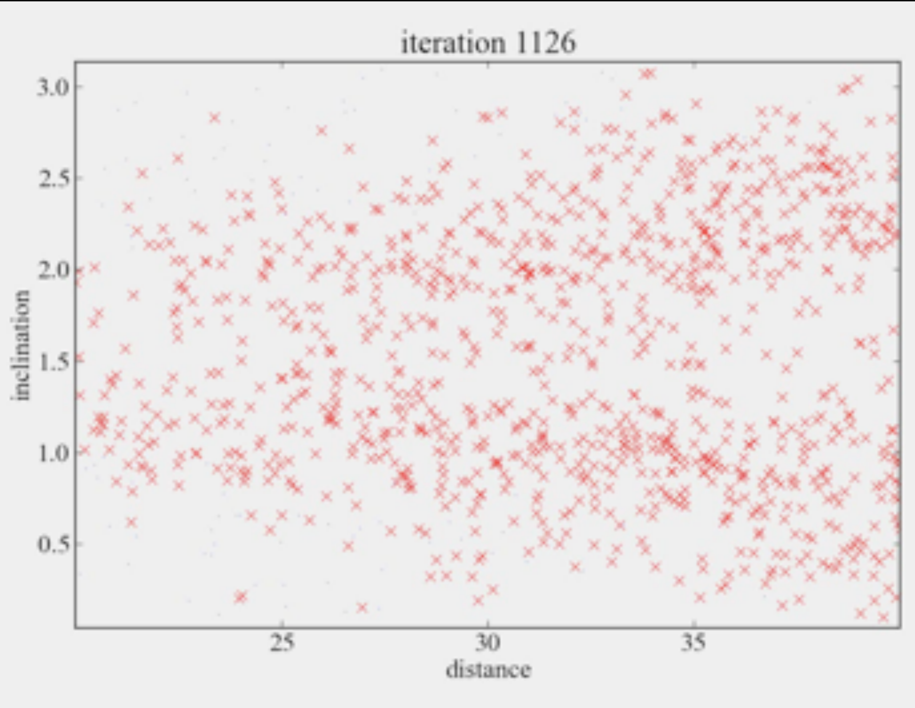
So we can also write $Z = \int_0^1 L(X)\, dX$

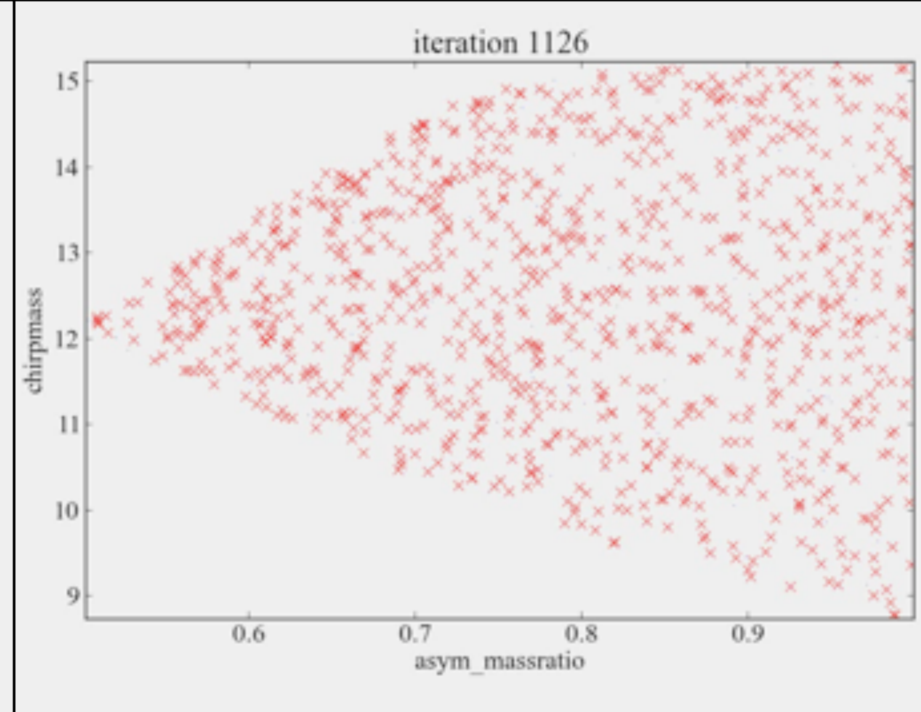
Key idea: remap the N-dimensional integral over θ to a 1-D integral over X.

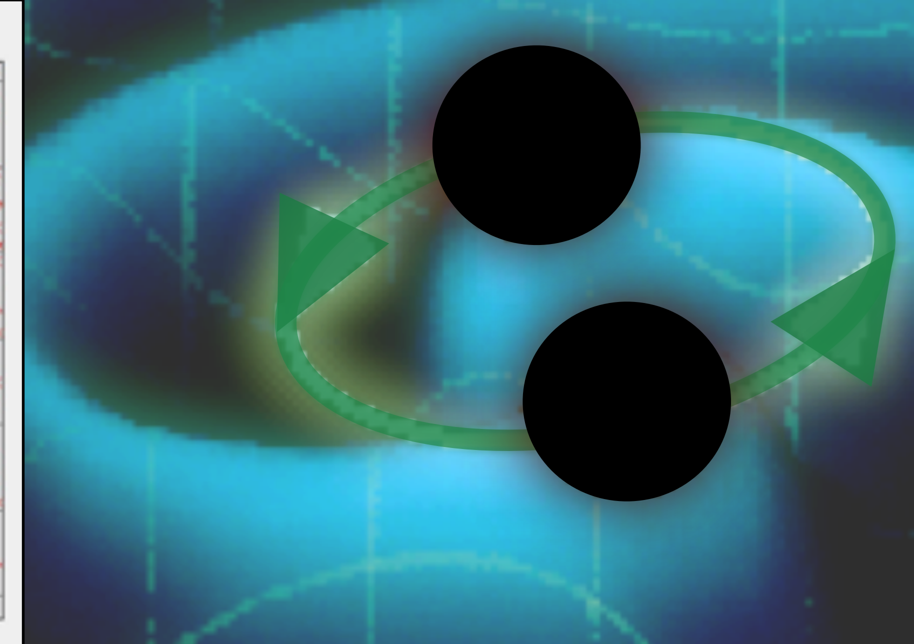
Each nested contour shrinks by (N-1) / N times. Choose X(λ) by randomly sampling N “live” points.

$$\int_{\vec{\theta}\in\Theta} p(\vec{d}|\vec{\theta}, \vec{\sigma}, H_S)\, p(\vec{\theta}|\vec{\sigma}, H_S) = Z = \int_0^1 L(X)\, dX$$





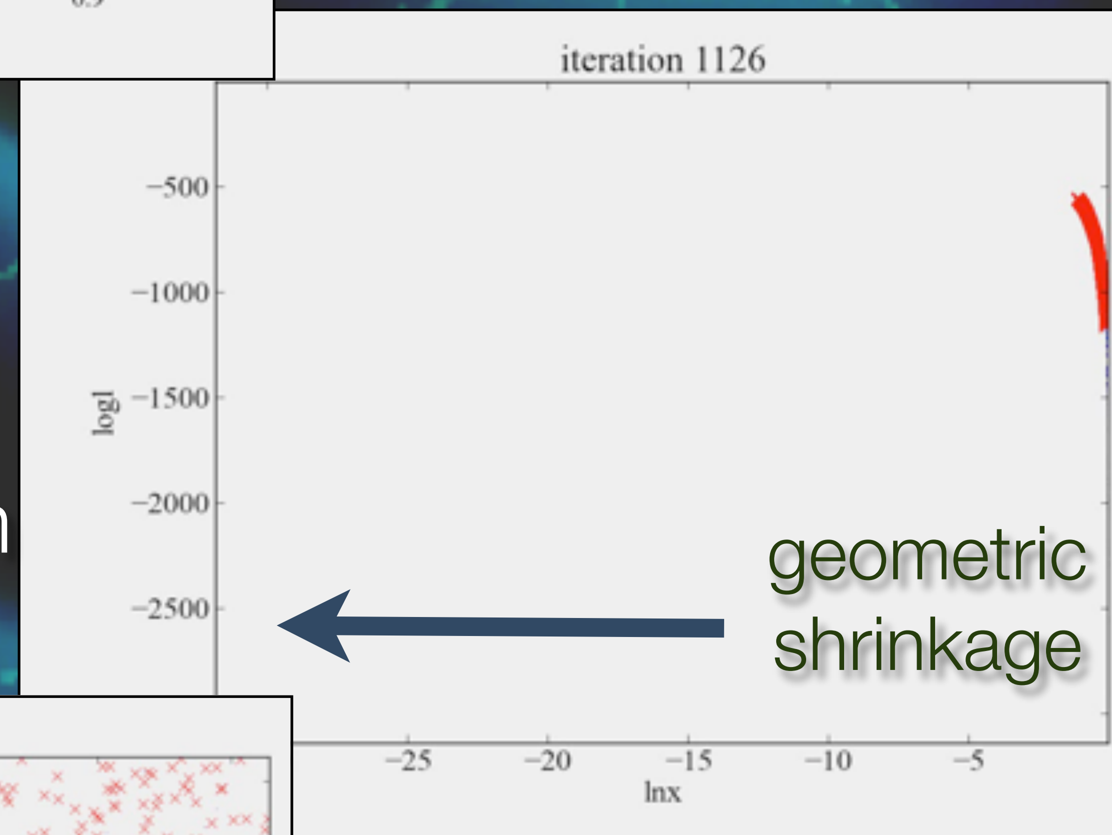


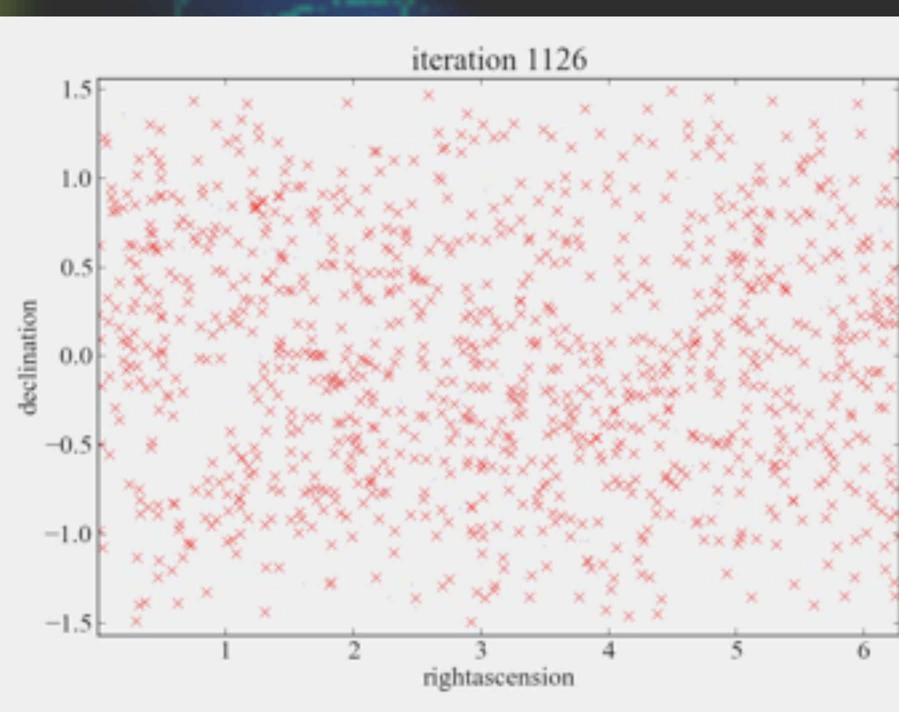


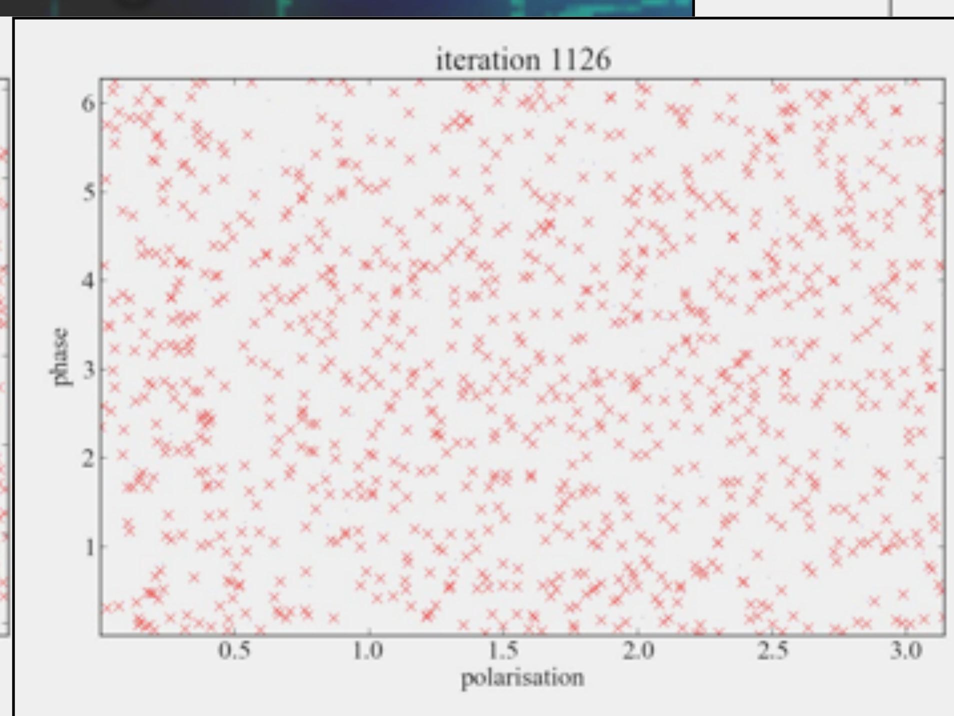
Live points (1024 samples from inside contour)

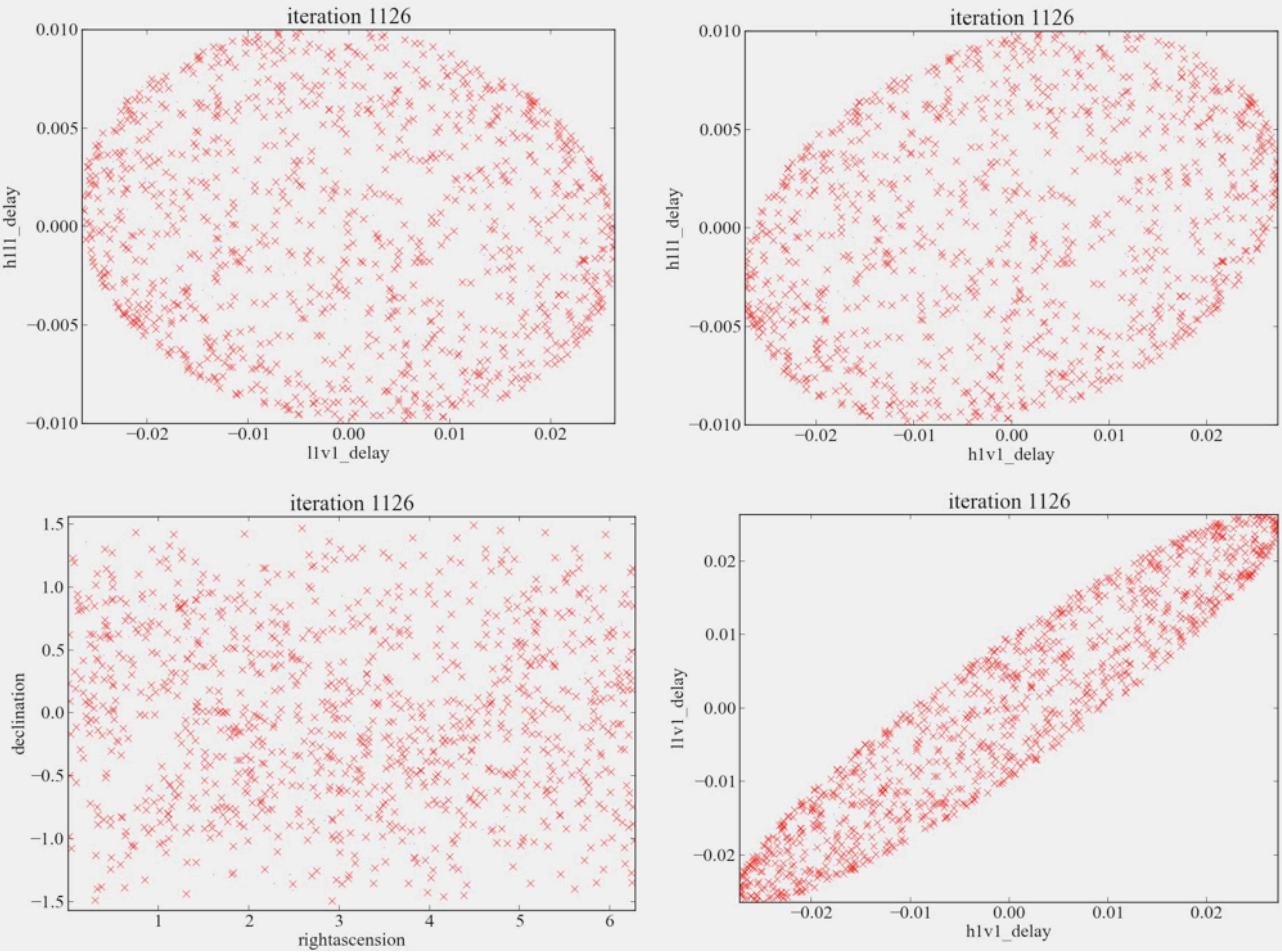
Used points

Replacement samples are drawn from inside the contour using MCMC

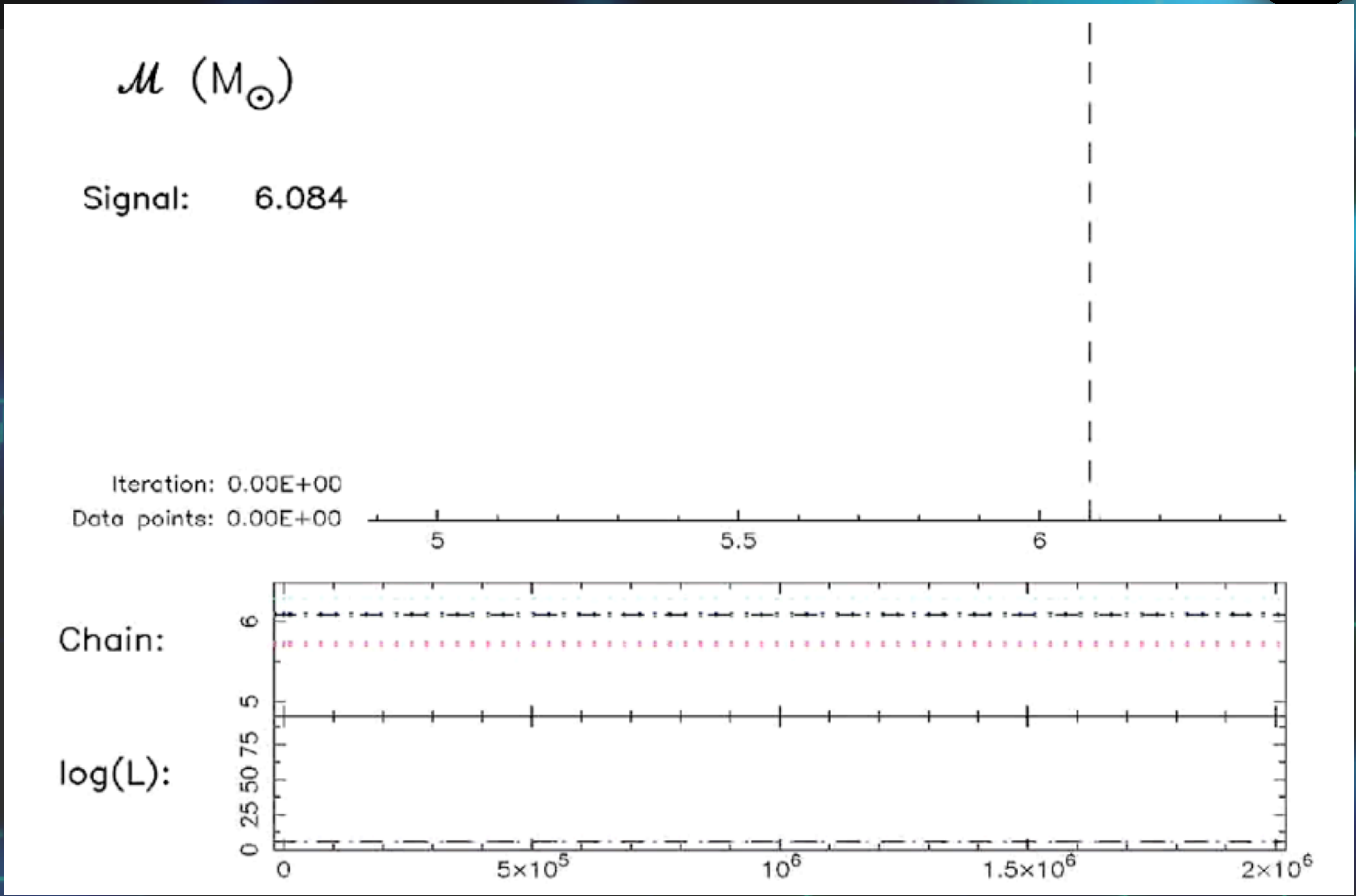






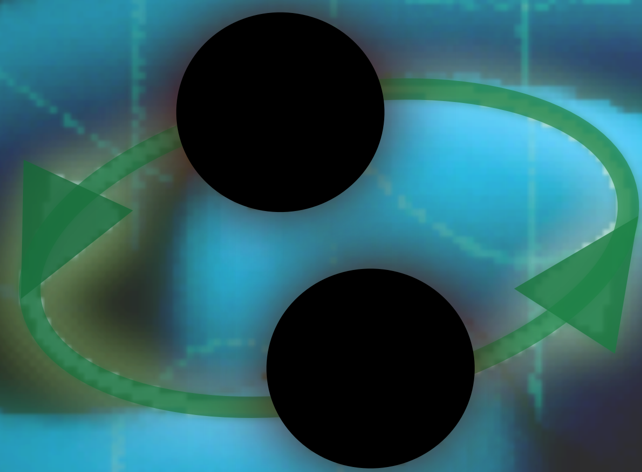

iteration 1126
h1l1_delay
l1v1_delay
iteration 1126
h1l1_delay
h1v1_delay
iteration 1126
declination
rightascension
iteration 1126
l1v1_delay
h1v1_delay

MCMC



$\mathcal{M}$ ($M_\odot$)
Signal: 6.084
Iteration: 0.00E+00
Data points: 0.00E+00
5
5.5
6
Chain:
6
5
log(L):
0 25 50 75
0
5×10^5
10^6
1.5×10^6
2×10^6

LALInference



Generating samples directly from posterior PDF is hard.

- Need to know the shape of the distribution, which is the point of doing the analysis.

Instead, use probabilistic algorithms.

MCMC & Nested Sampling algorithms have been investigated for CBC & general GW analyses:

- MCMC: Röver, *et al* CQG 23 (2006) [4 params]; PRD 75 (2007) [9 params] ... van der Sluys *et al* ApJL 688 L61 (2008) [15 params] .... Raymond *et al* CQG 26 (2009) ... Littenberg & Cornish PRD 80 (2009) ....
- Nested Sampling: Veitch & Vecchio PRD 78 (2008) [4 params] ... Feroz *et al* CQG 26 (2009) ... Veitch & Vecchio PRD 81 (2010) [9 params] ....

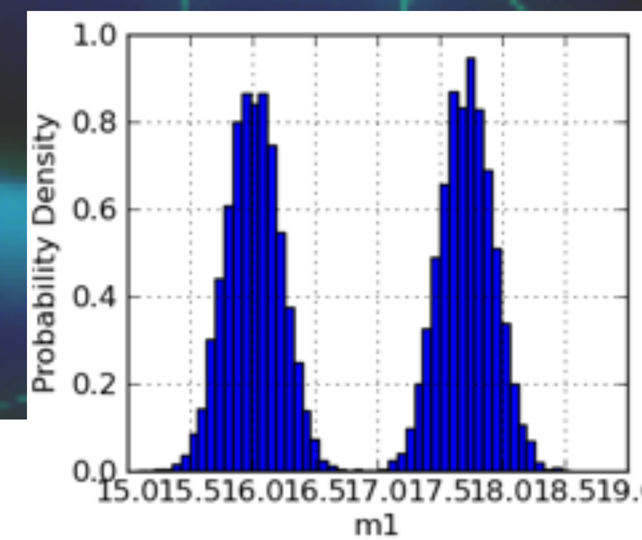
Several groups combined efforts to implement common library for analyses of real data: LALInference.

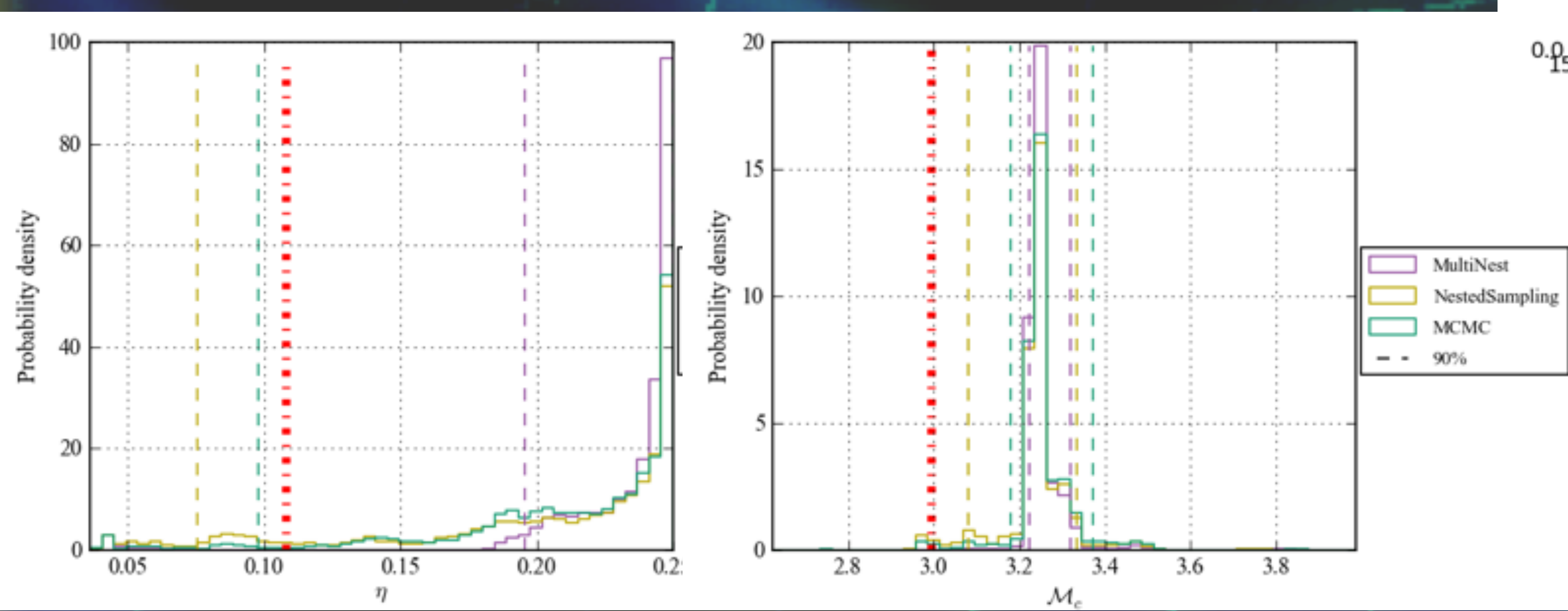
- Free software, download from https://www.lsc-group.phys.uwm.edu/daswg/projects/lalsuite.html
- Reviewed and robust waveforms, likelihoods, priors, samplers.
- Uses MCMC, Nested Sampling or MultiNest for P.E. & model selection

Consistency checks

How can we be confident we have solved the problems?

- Cross-checks with different samplers
- Check against known likelihood functions

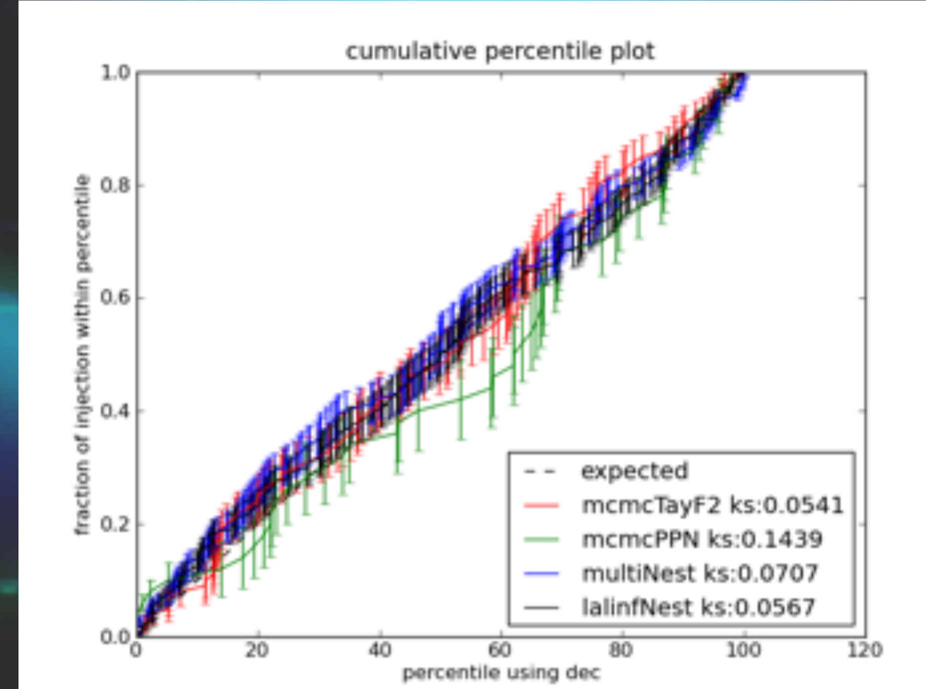


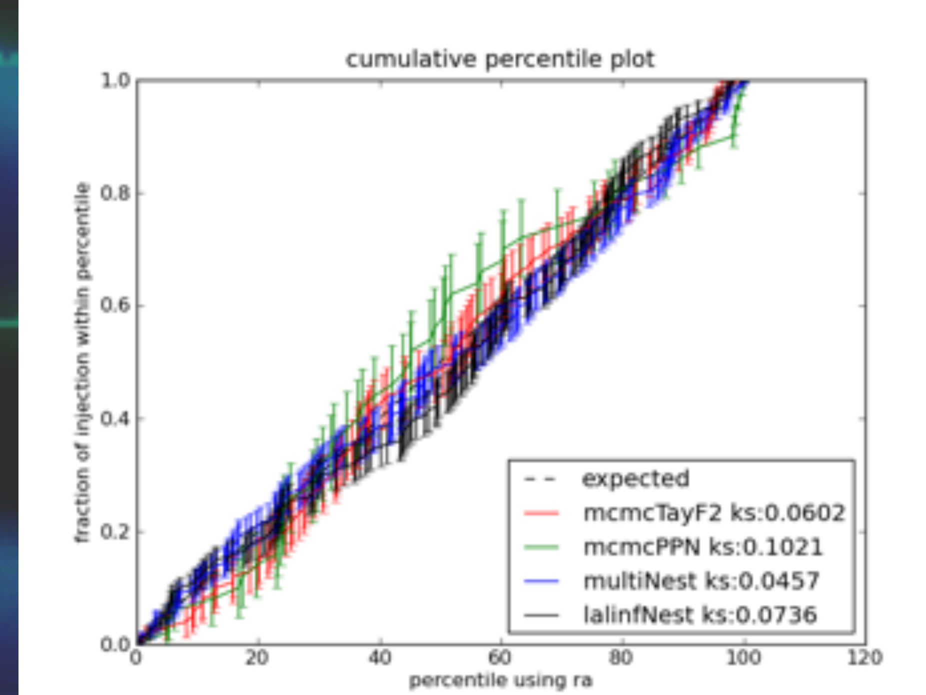


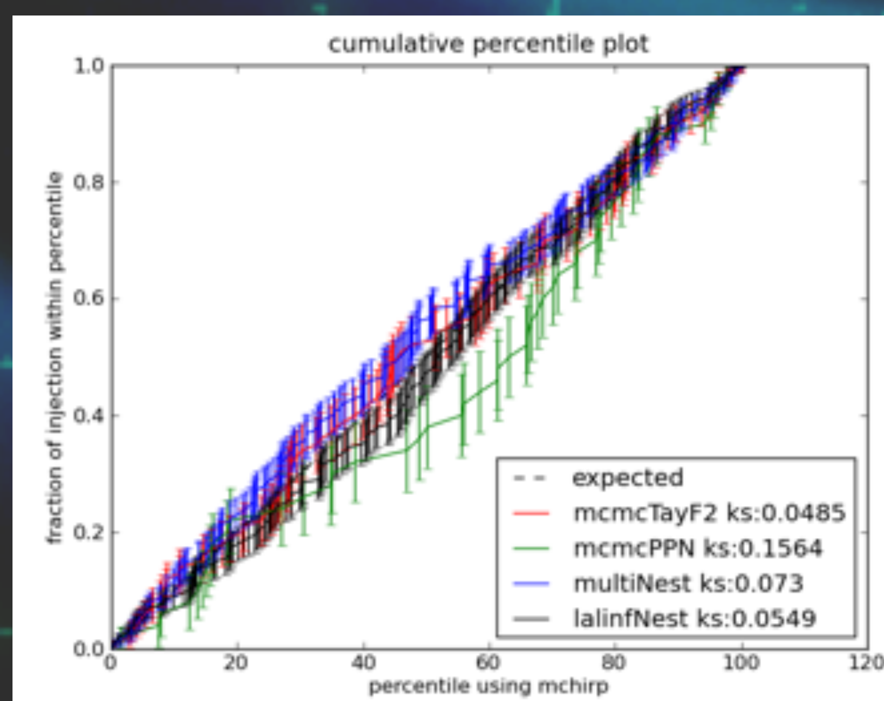
Consistency checks

Also use probability - probability plots to compare Bayesian probability intervals with frequency of result in a Monte Carlo over many injections

- Do 5% of injections fall inside the 5% probability interval?

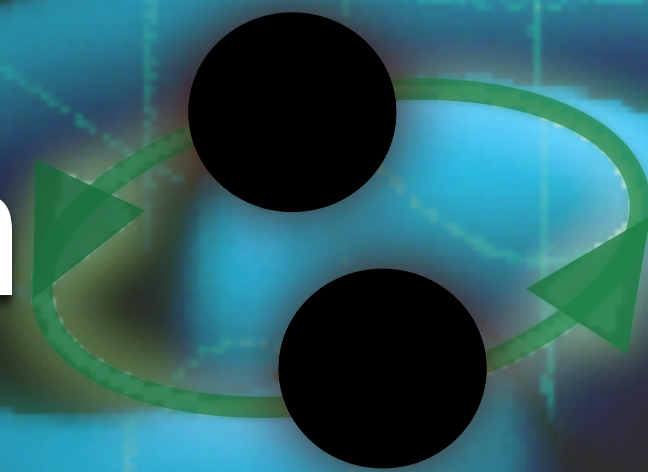






18

S6 Parameter Estimation



S6 parameter estimation paper hit the arXiv this week! arXiv:1304.1775

- Put LALInference into practice on hardware and software injections
  - BNS, NSBH and BBH
  - spinning and non-spinning
  - Comparison of results with different waveform approximants

Parameter Estimation for Compact Binary Coalescence events with first-generation ground-based Gravitational-Wave Detector Network.

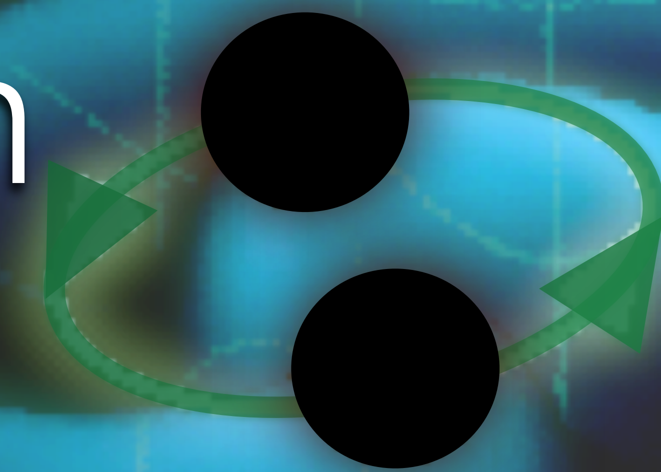
The LIGO Scientific Collaboration[1] and The Virgo Collaboration[2]

[1] *The LSC*
[2] *Virgo*

(RCS Id: s6pe.tex,v 1.115 2013/02/03 13:10:32 john.veitch Exp ; compiled 4 February 2013)

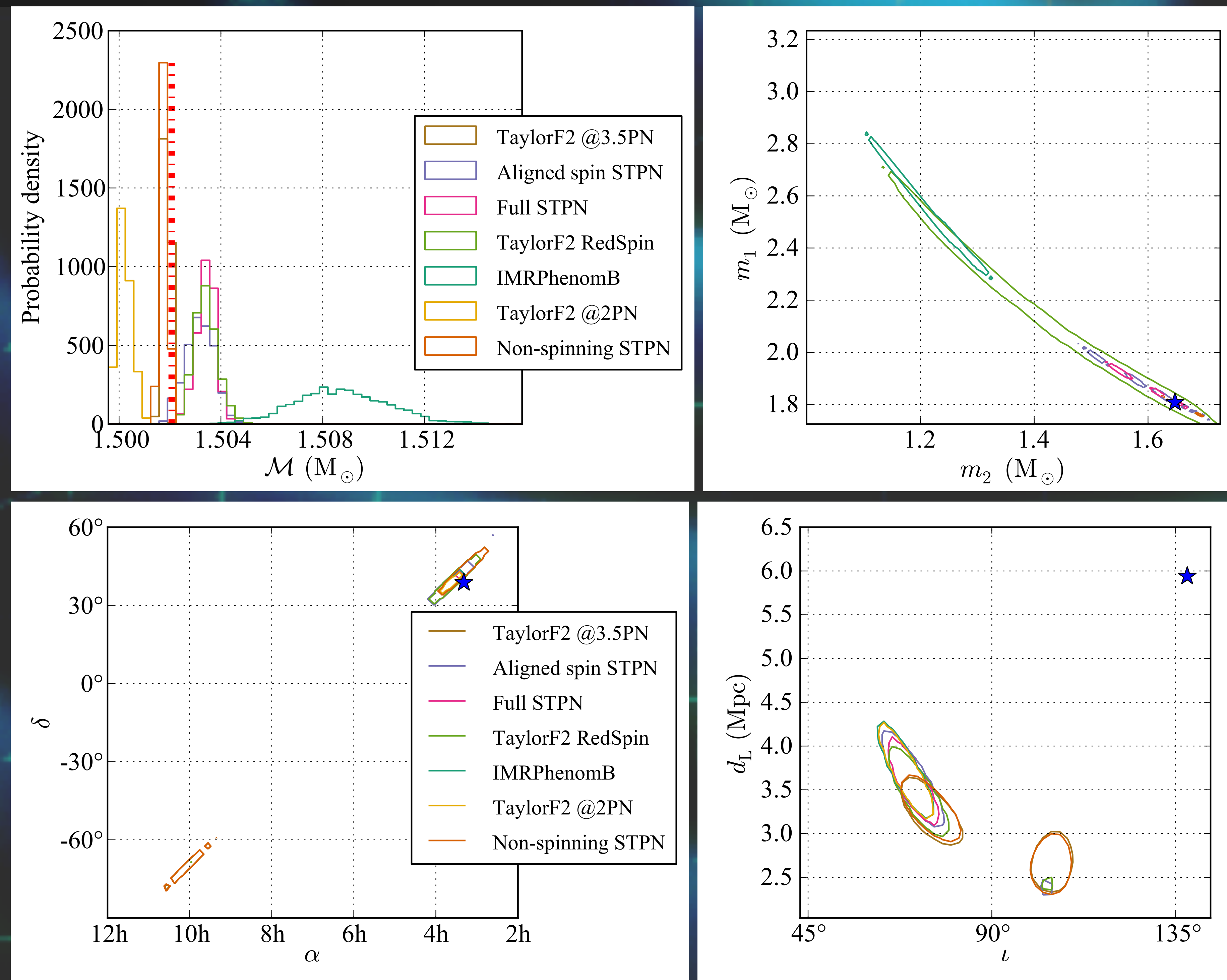
Gravitational waves from coalescing compact binaries, a primary source for ground-based gravitational-wave detectors, encode information about source physics. Detailed models of the anticipated waveforms from these coalescences enable inference on parameters including component masses, spins, sky location, and distance. Parameter estimation and model selection in multi-dimensional parameter spaces are complex but necessary in order to pursue gravitational-wave astronomy and astrophysics. In this paper, we describe the application of tools that we developed for LIGO-Virgo parameter estimation to several simulated signals in the most recent detector data and we present the results in a qualitatively similar way to what we expect will be delivered in the advanced detector era. We demonstrate that we can successfully recover the known parameters of these signals, subject to statistical measurement uncertainties and systematic biases due to differences between the waveform families used for simulation and analysis.

S6 Parameter Estimation BNS (HW)

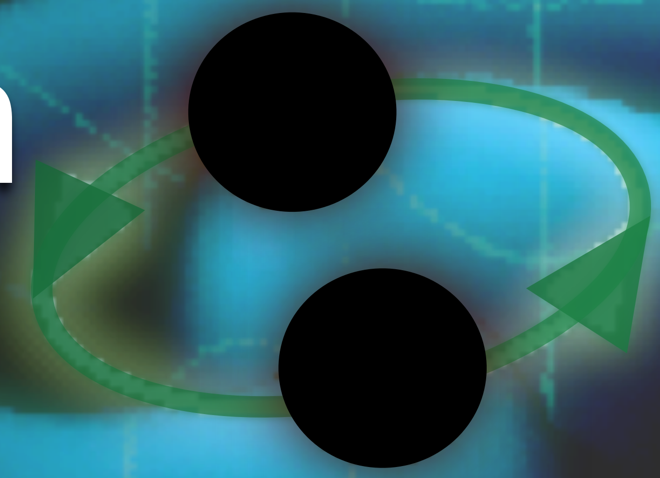


BNS HW injection

- SNR ~36
- EOBNR waveform
- 1.8 - 1.65 Msun
- Distance poorly recovered
- 32s template took MONTHS to analyse using SpinTaylorT4

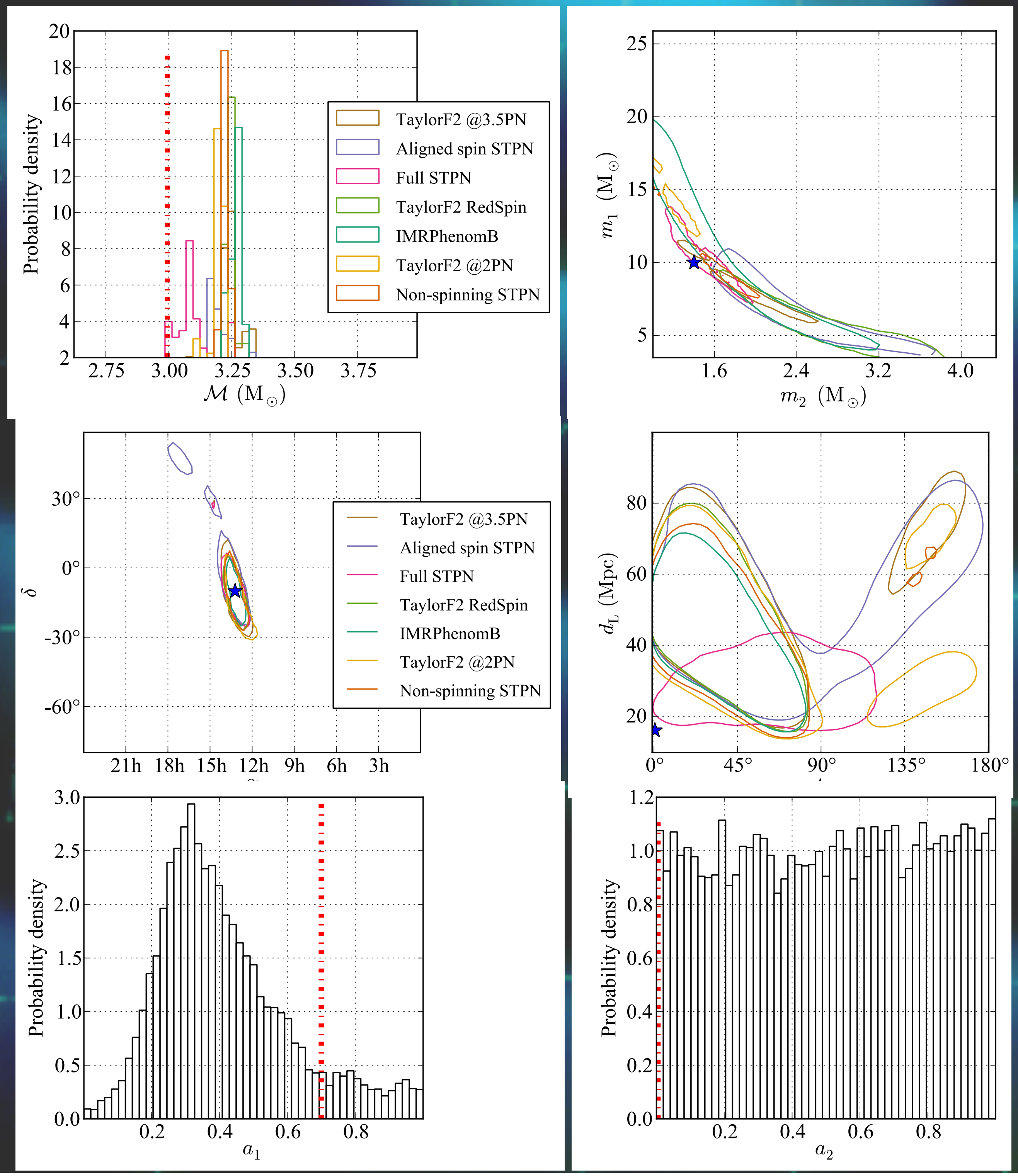


S6 Parameter Estimation NSBH (SW, spinning)

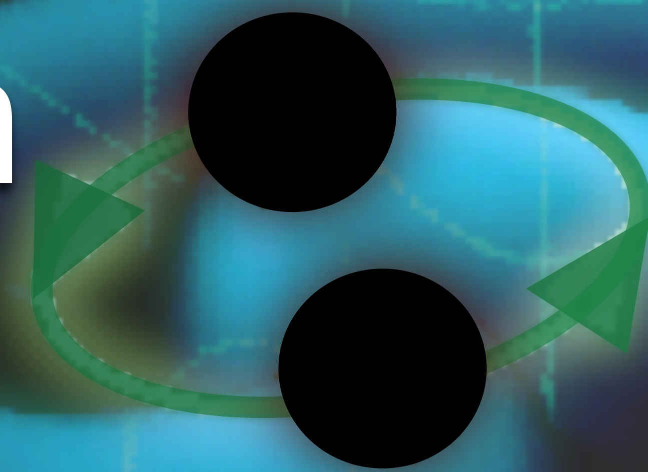


Spinning NSBH software injection

- SpinTaylorT4 waveform
- 10 - 1.4 Msun
- SNR ~13
- Only fully spinning waveform accurately estimates masses & distances
- Good evidence for spin (logB ~ 12)
- Discovered NSBH is surprisingly difficult to analyse. Had to re-run with higher accuracy.

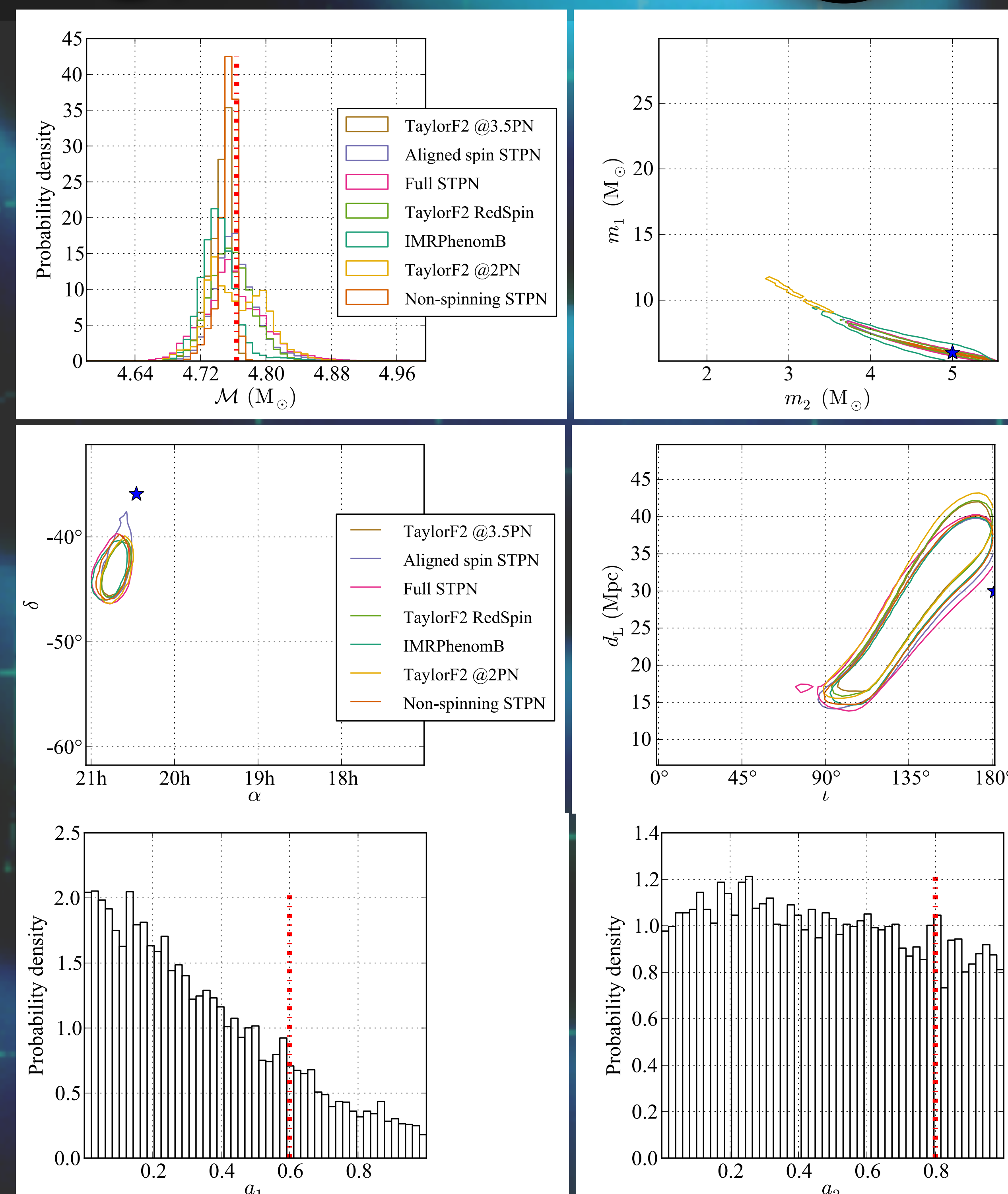


S6 Parameter Estimation BBH (SW, spinning)

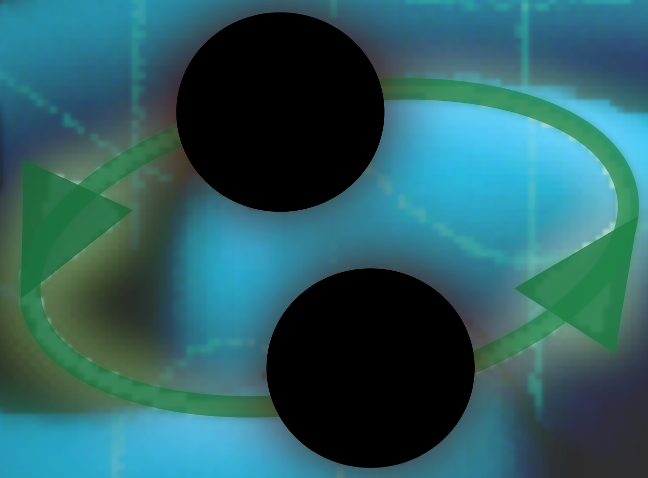


Spinning BBH SW injection

- 6 - 5 Msun
- SNR ~19
- Relatively easy to analyse
- Systematic bias surprisingly small for all but TF2_2PN
- Spins poorly constrained (c.f. non-spinning PDFs) with no significant evidence in favour of fully spinning waveform

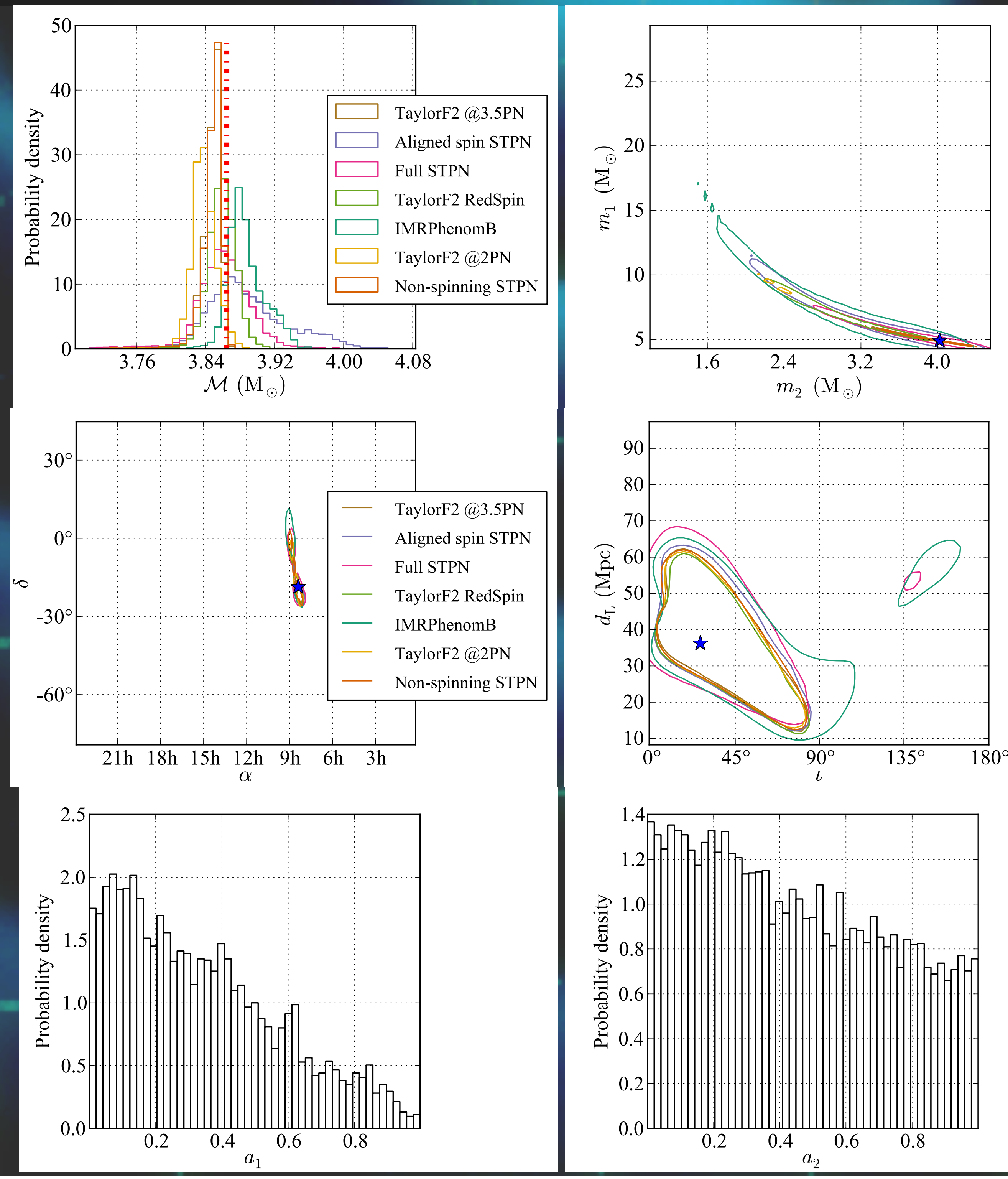


S6 Parameter Estimation BBH (HW, non-spinning)



Non-spinning BBH hardware injection

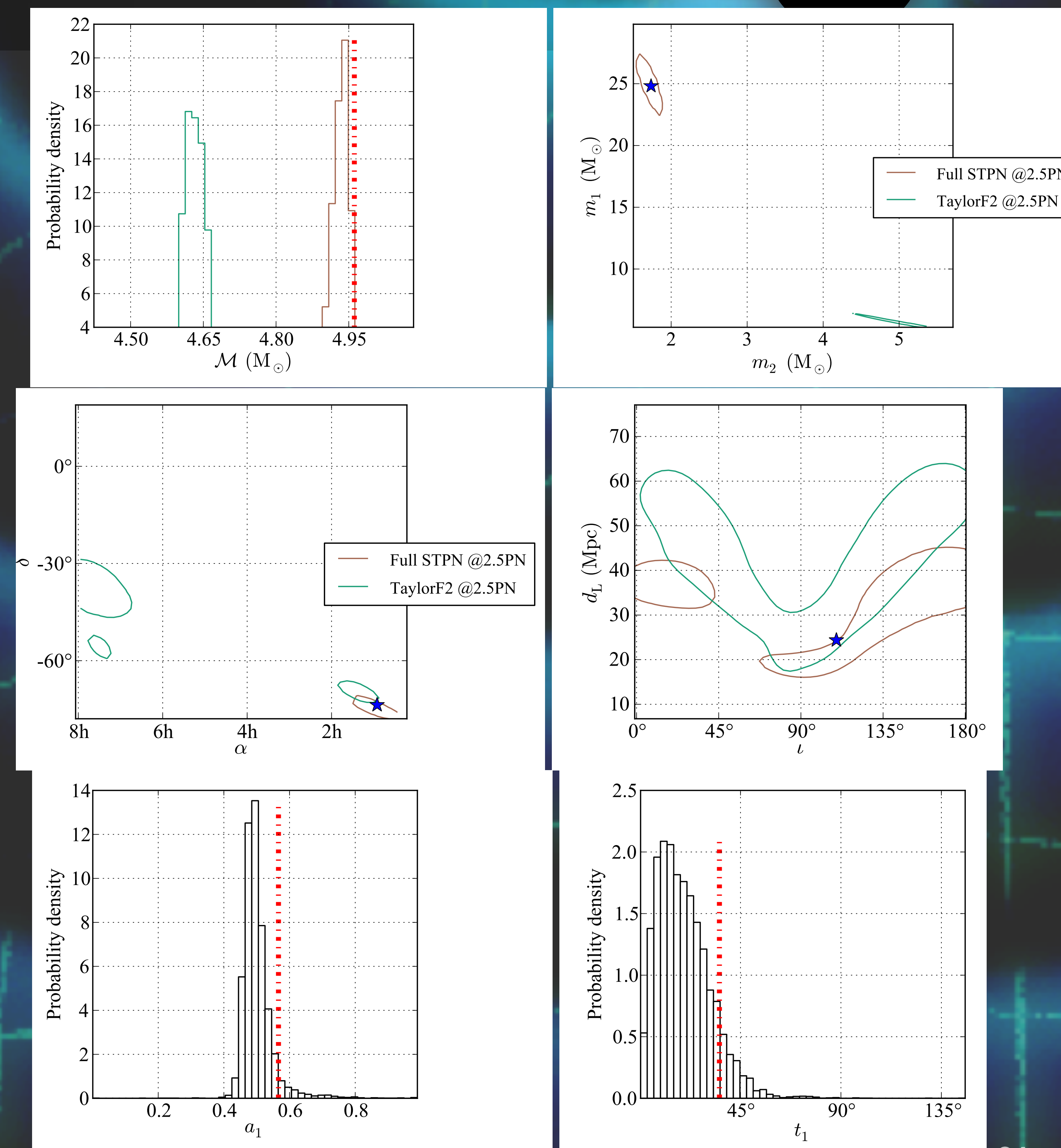
- 4.9 - 4.0 Msun
- SNR 13
- Easy to analyse
- Most templates suffice to recover masses
- Very similar PDFs on spins as in the spinning BBH case!
  - need longer waveforms?



S6 Parameter Estimation Big Dog

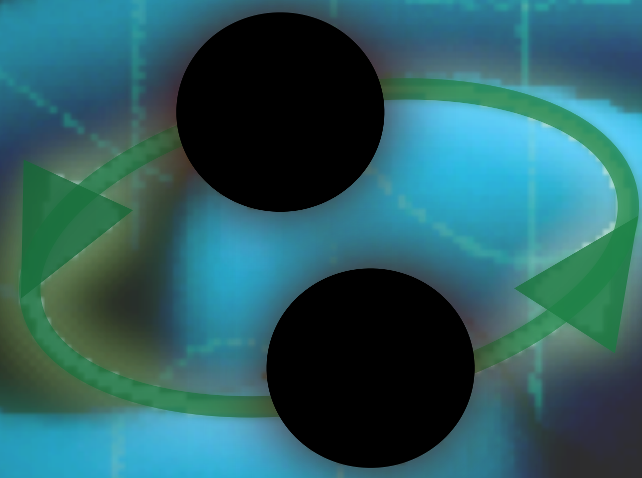
Blind injection!

- 24.8 - 1.7 Msun
- SNR ~16
- Had a number of issues with waveform
- High asymmetry makes longer template, better spin resolution



24

Towards advanced detector era

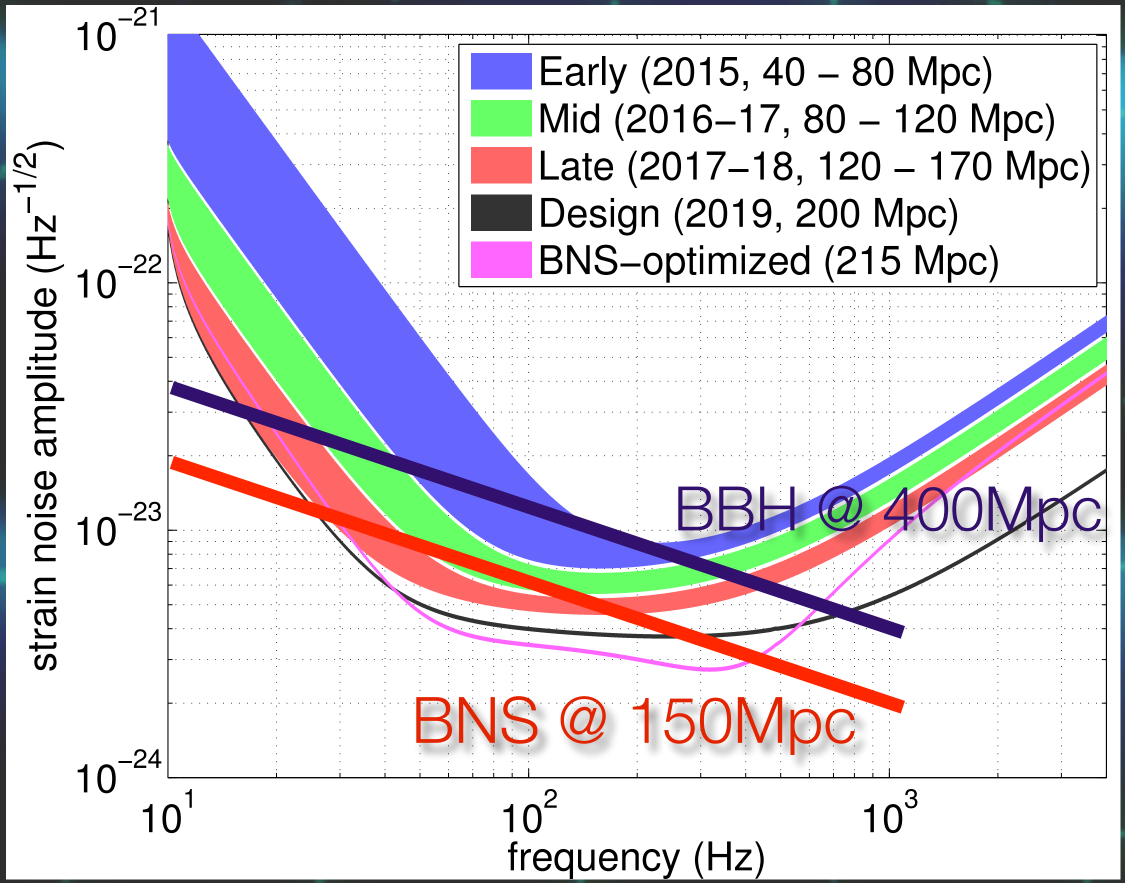


Advanced detectors will allow lower frequency cutoffs

- Longer templates

f_{low}	chirp length (1.4-1.4)	chirp length (5-5)
40 Hz	25 s	3 s
30 Hz	55 s	6.5 s
20 Hz	160 s	19 s
15 Hz	345 s	42 s

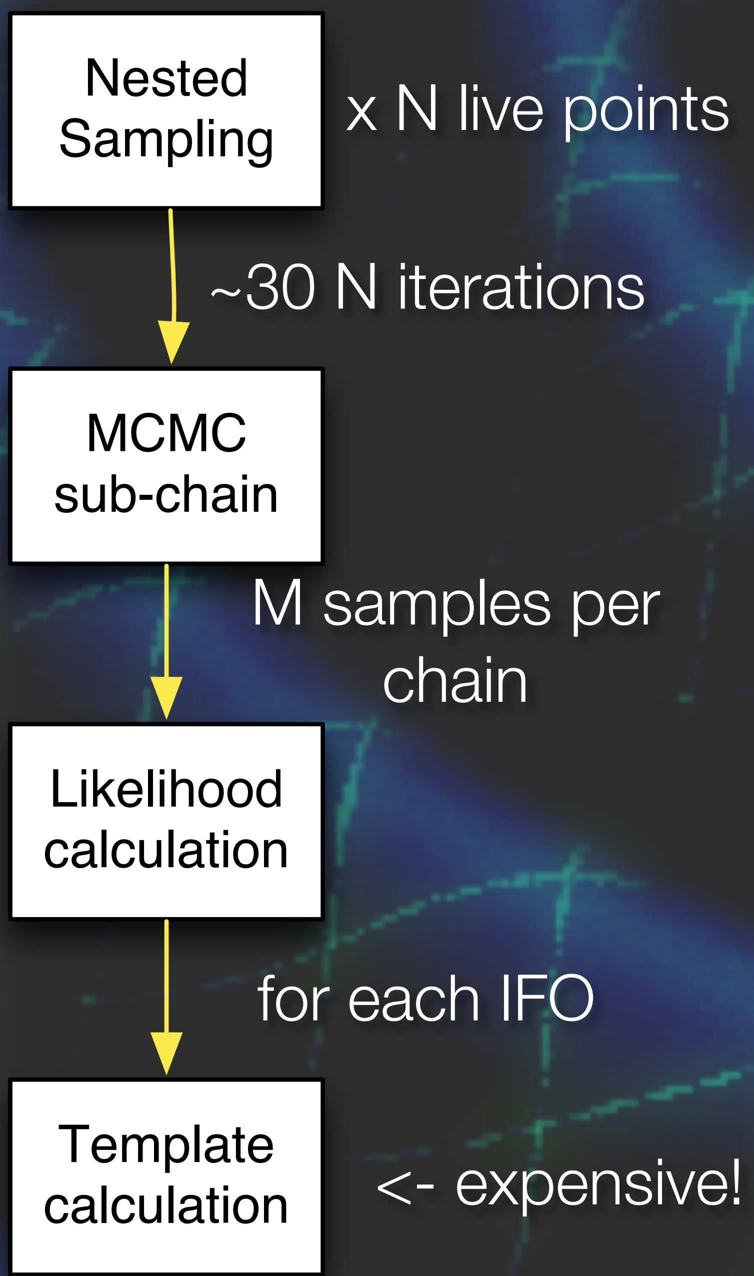
Want to push low frequencies for measuring masses, spins. Not so relevant for sky localisation (bandwidth)



- But we have 3 month detection -> publication schedule!

- Especially difficult for BNS!

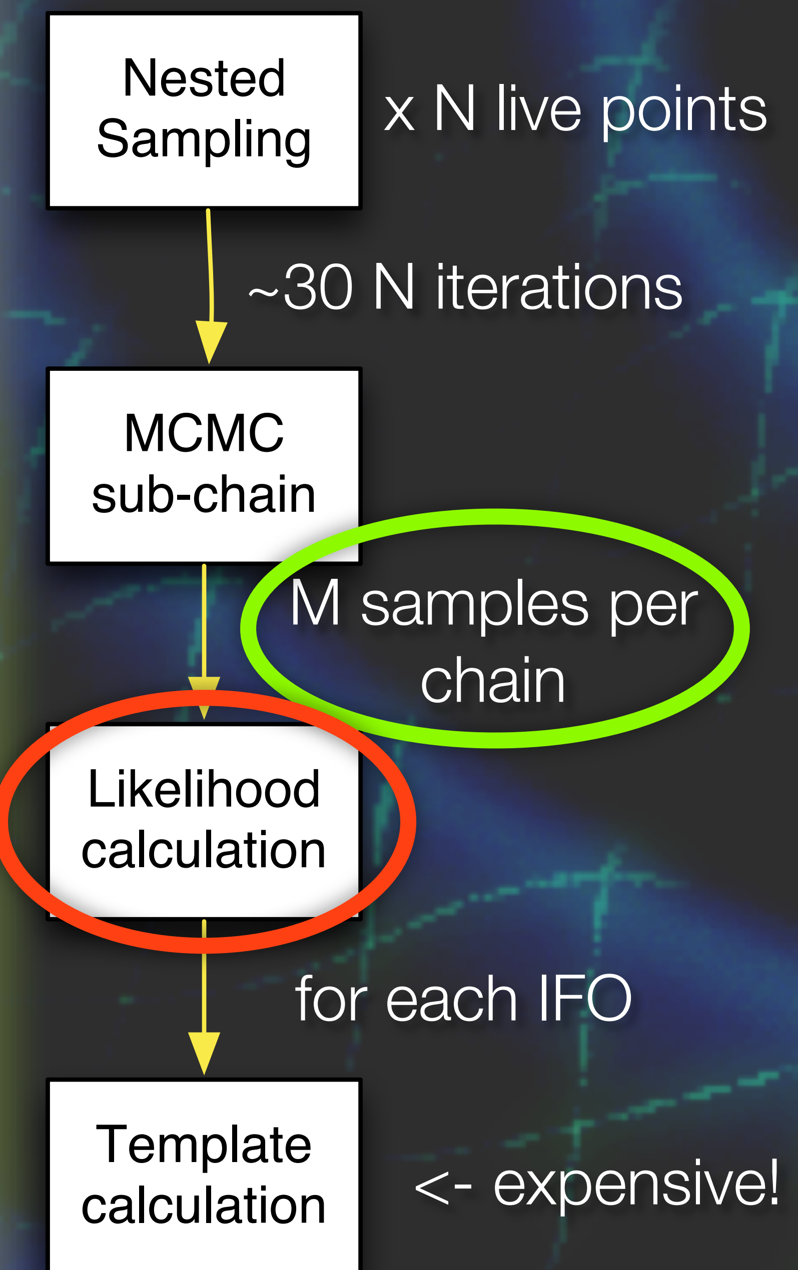
Computational Cost



Run-time is dominated by

- Template calculation (60-99% run-time)
  - including FFT for time-domain signals
  - scales with length of template and sampling frequency
  - Already vastly improved by LALSimulation (~x2)
- Overlap calculation (remainder of run-time)
  - Once per detector
  - Already caches template to avoid recalculation for each detector (x3 speed-up)
  - Uses accelerated trigonometry approximation

Optimisations: Adaptive MCMC chains

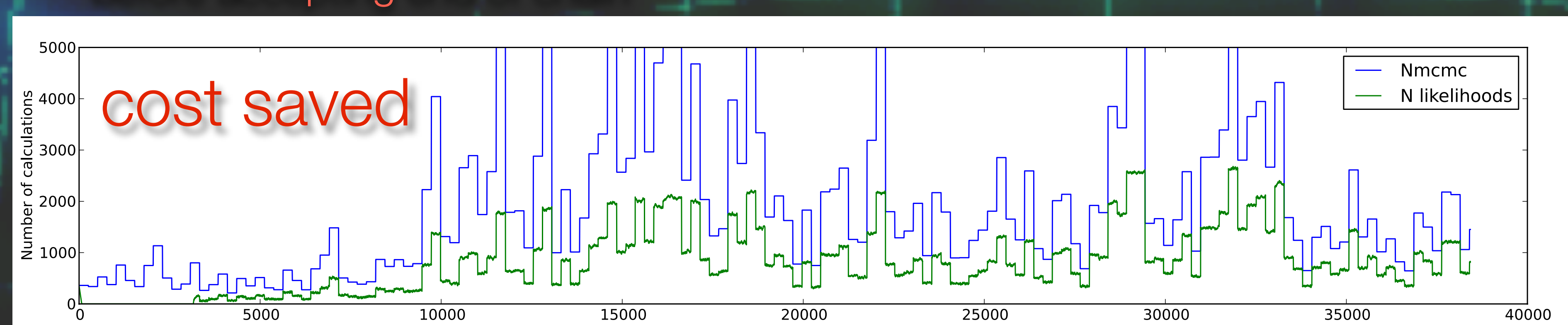


MCMC stage draws uncorrelated samples from the prior bounded by a likelihood contour.

- Requires likelihood to be checked for each sample
- Difficulty of sampling this distribution varies at different stages of the run
- Required number M must be large enough to accommodate the most difficult part of the run

Optimisations:

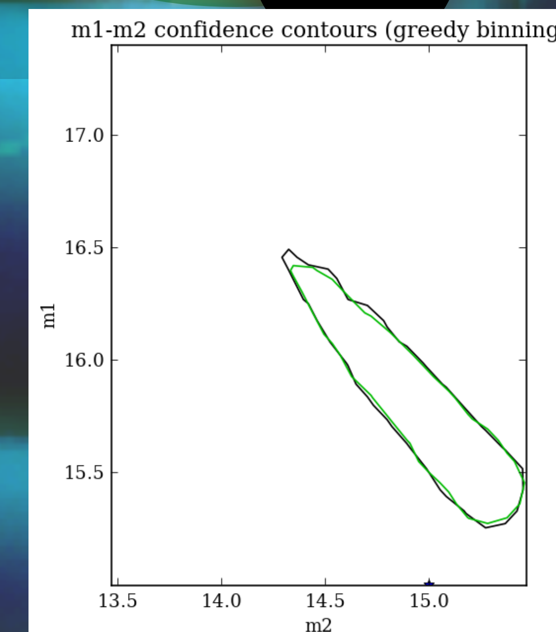
- Automatically determine the necessary M as the run progresses
- Don't need to compute Likelihood for every sample in the chain, just check before accepting end of chain

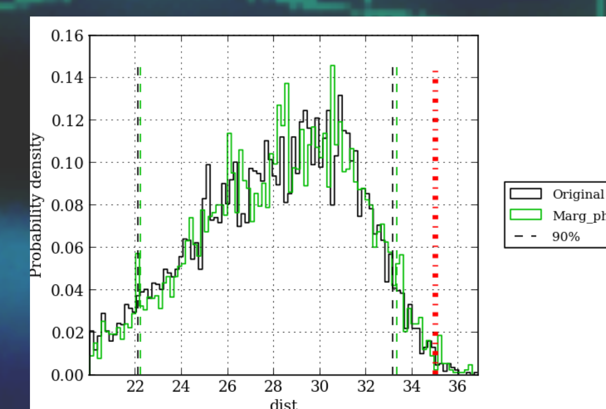


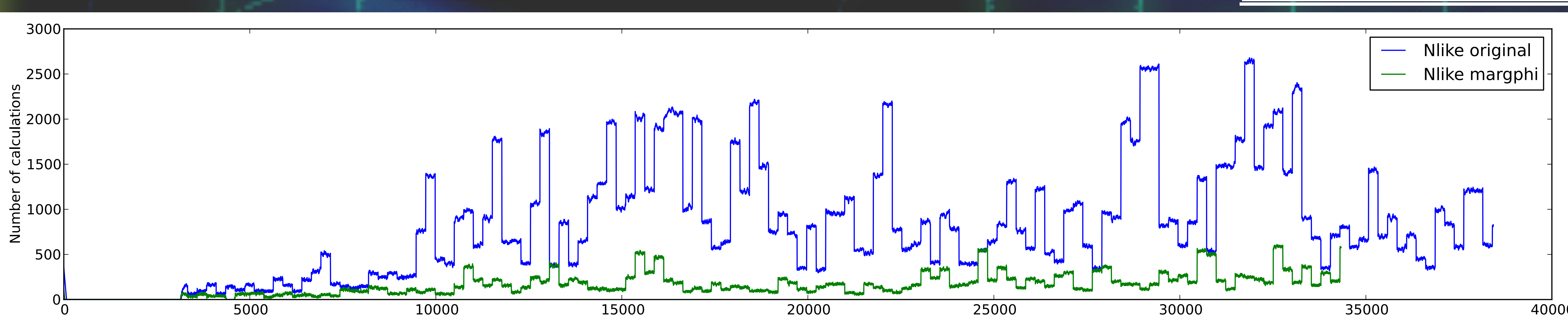
Optimisations: Analytic Marginalisation

Don't care about phase parameter! - analytically marginalise it away

- Modified likelihood: $p(n_i|\sigma_i) \quad \propto \exp\left[-\sum_i \frac{|d_i|^2-|h_i|^2}{2\sigma_i^2}\right] \mathrm{I}_0\left[\sum_i \frac{|d_i\bar{h}_i|}{\sigma_i^2}\right]$
- Eliminates 1 dimension (reduces information and therefore run time)
- logZ before: 368.60, after: 368.62
- Run-time reduced by factor 4 !
  - Smoothes likelihood for easier sampling, fewer MCMC steps needed!
- Details in LIGO-T1300326







28

Optimisations: Parallelisation

Nested Sampling — x N live points

~30 N iterations

MCMC sub-chain

M samples per chain

Likelihood calculation

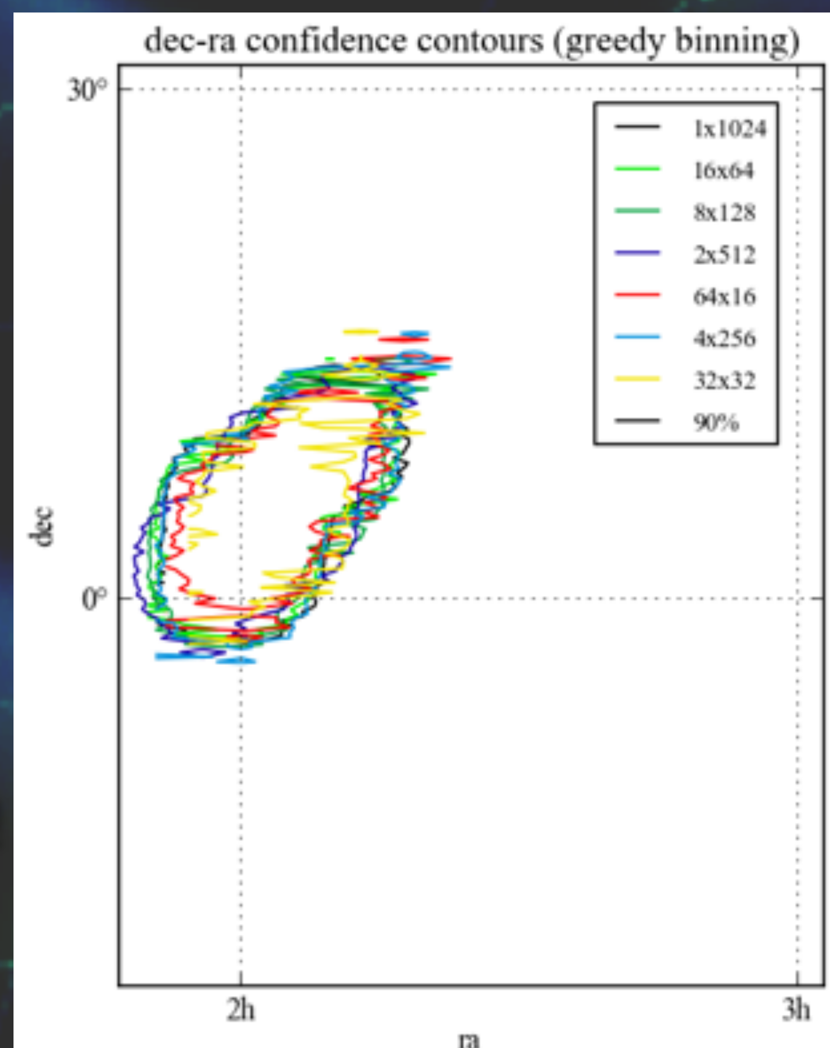
for each IFO

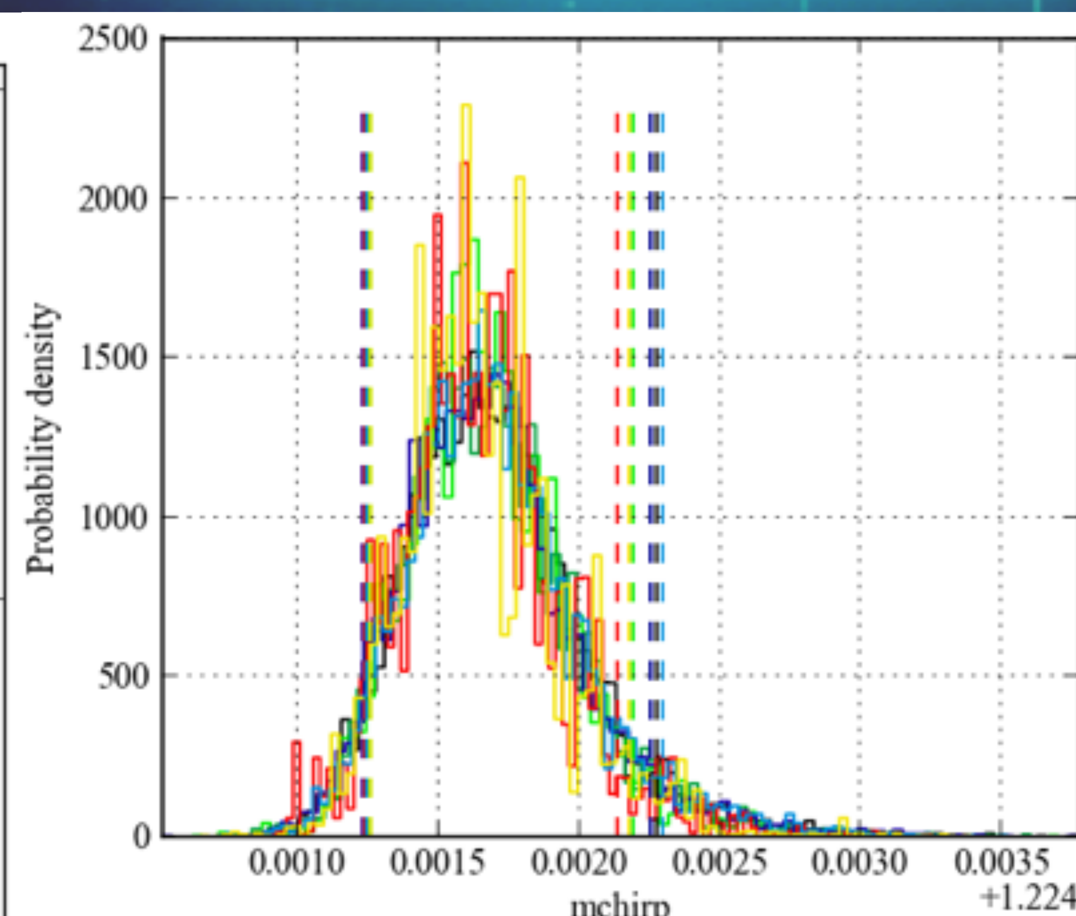
Template calculation

<- expensive!

Parallelise work by running more chains with fewer live points each.

- Same amount of work but done in less clock time





Prospects for further speed

Further optimisations that might be possible:

Algorithm:

- Adaptive number of live points
- Parallelisation
- different algorithm (dNest?)
- More efficient proposals

Template generation / overlap calculation:

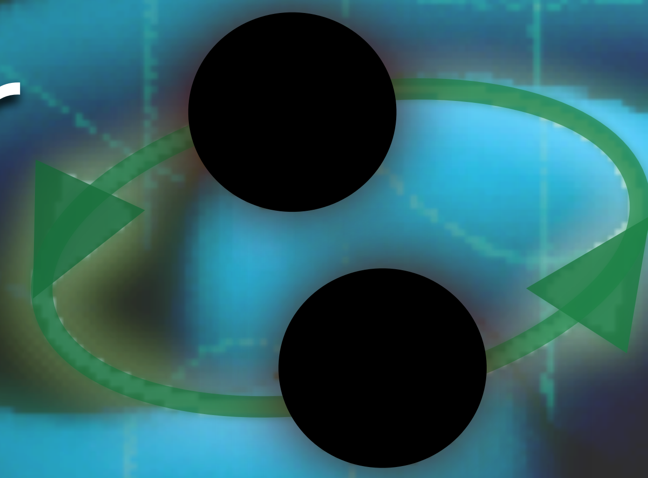
- lookup tables for precalculation
- GPUs?
- multi-bandwidth (MBTA-style)?
- Interpolated templates (R. Smith)

Further analytic marginalisation:

- Closed forms probably impossible
- Rapidly convergent Taylor expansion... maybe

Probably another order of magnitude to be gained at least!

CBC Advanced Detector Science



★ Find astrophysical rate of BNS, NSBH, BBH

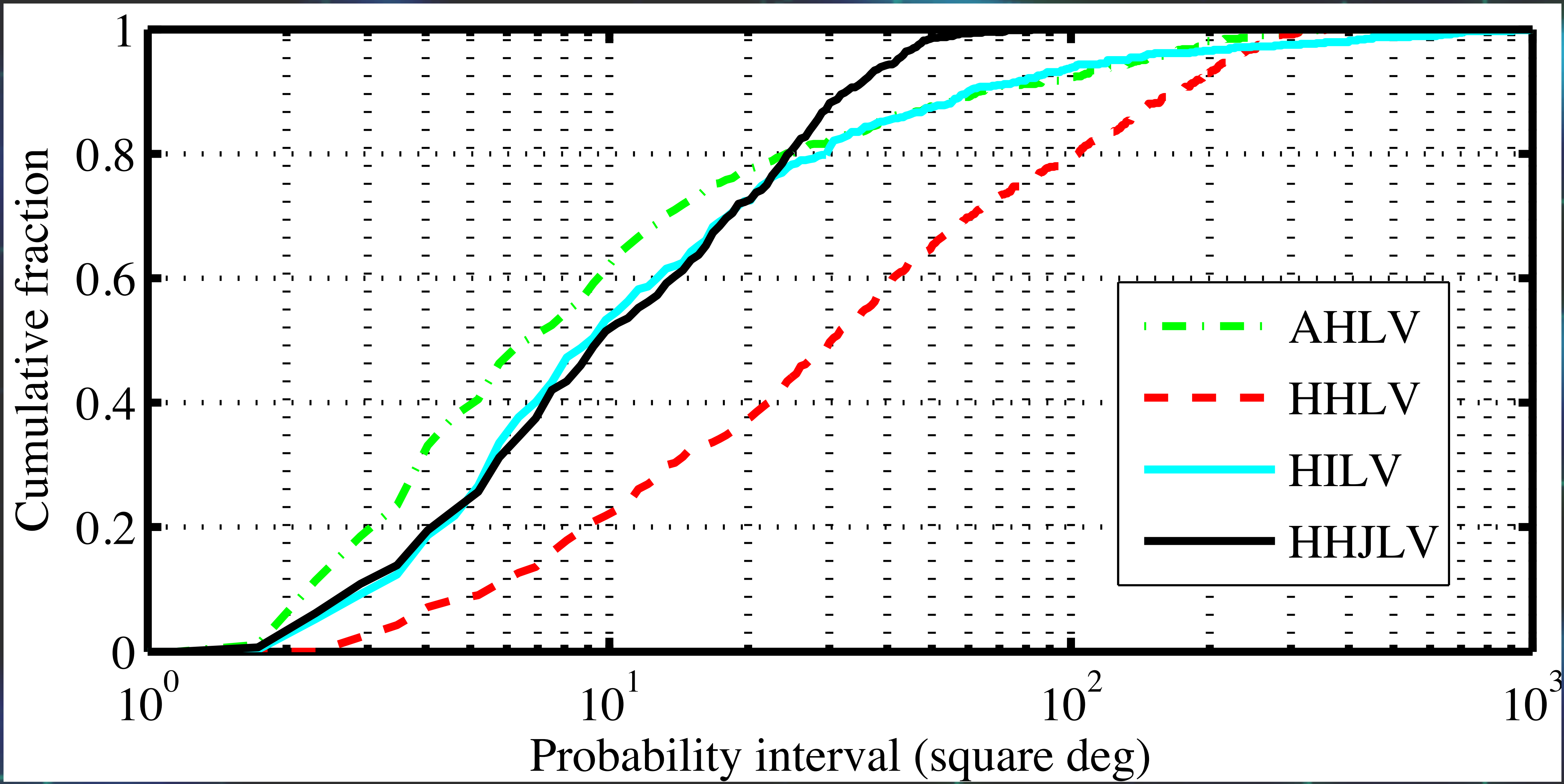
★ Measure mass distributions of neutron stars and black holes

- Stellar evolution
- Maximum mass of NS?
- Equation of state from tidal interactions
  - Physics of quark matter at core
- sGRB central engine
- Test GR orbital prediction

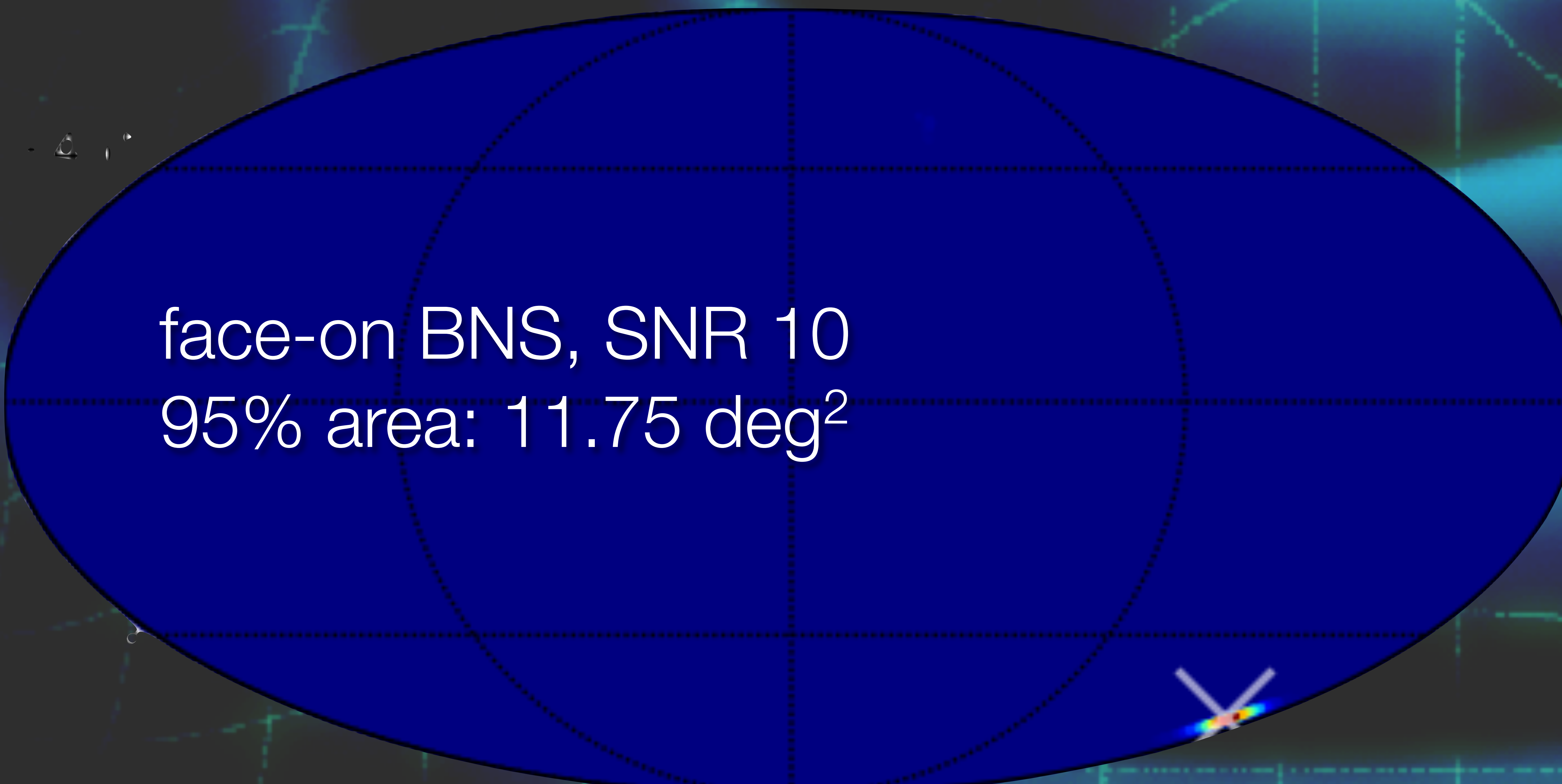
- Confirm predicted existence of NSBH/BBH binaries
- Measure spins
  - “kick” from supernova
- Lower mass limit of BH?
- Test no hair theorem
- Tidal disruption in NSBH?

Sky localisation

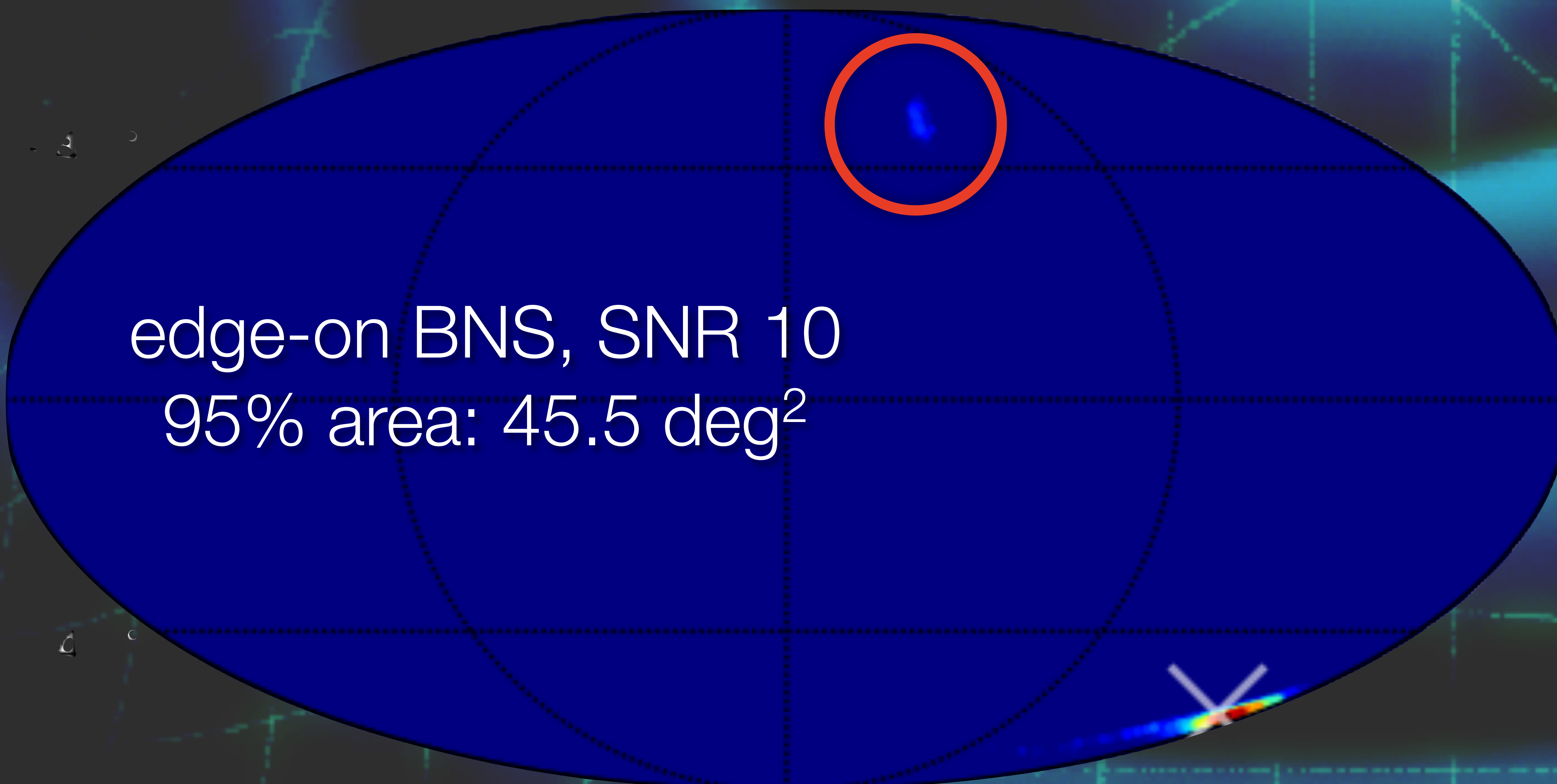
Simulation with 750 BNS signals in different networks [Veitch *et al* PRD 85 2012]



Breaking sky position degeneracy

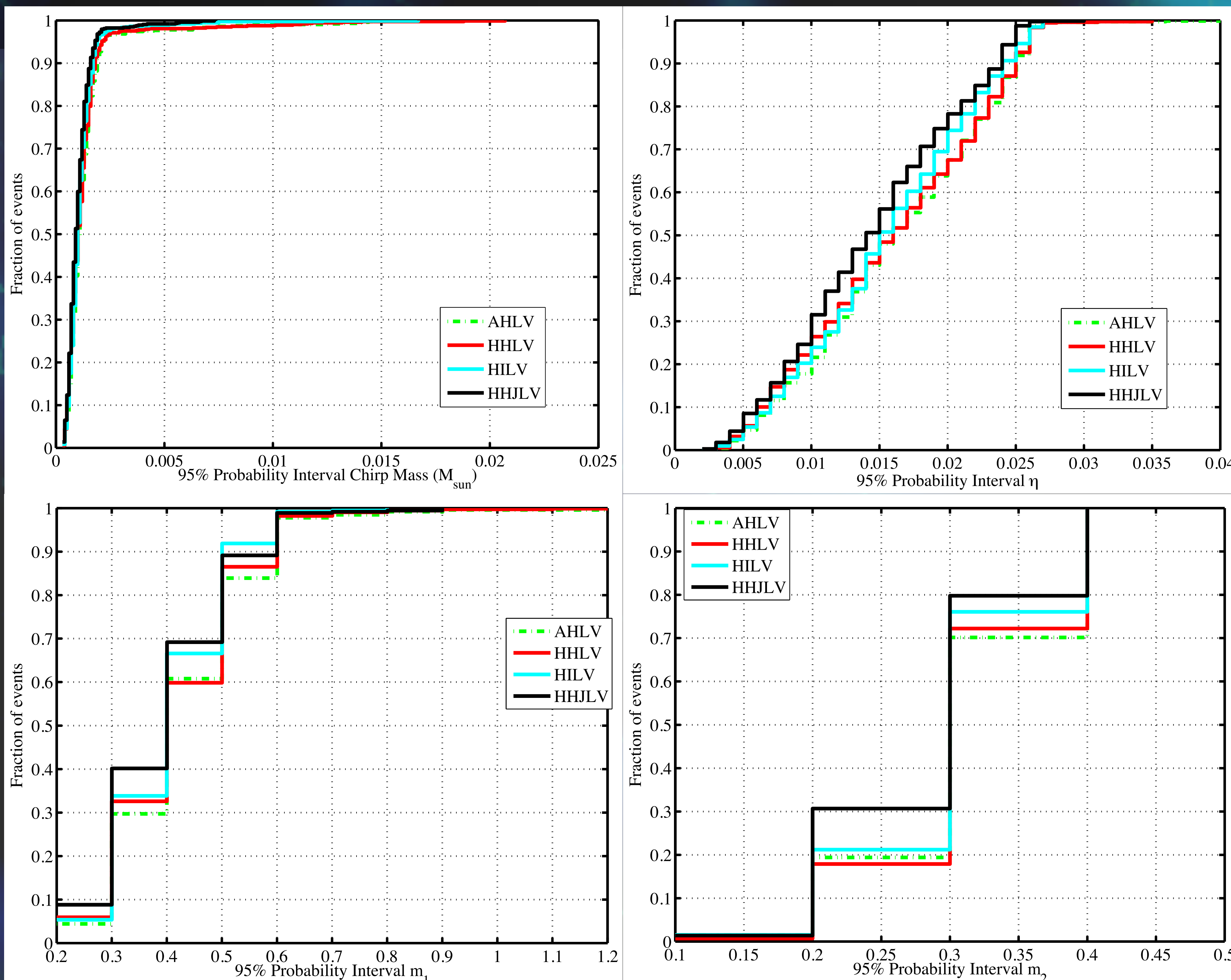


Breaking sky position degeneracy

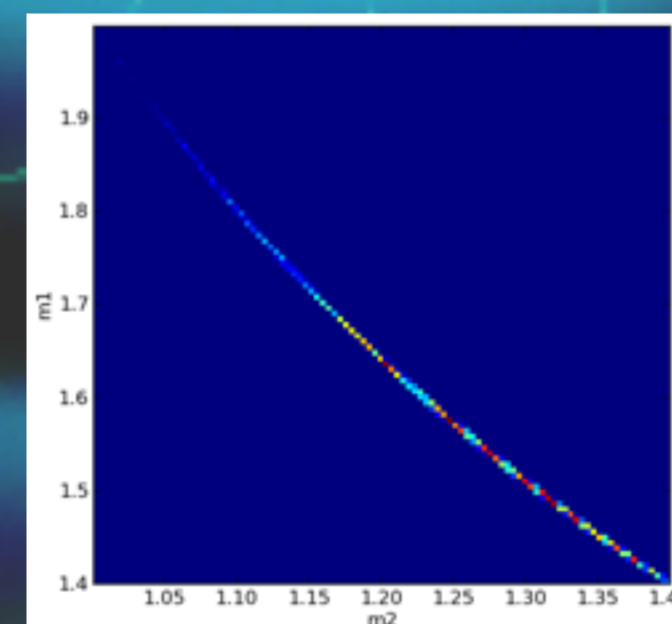


34

Masses



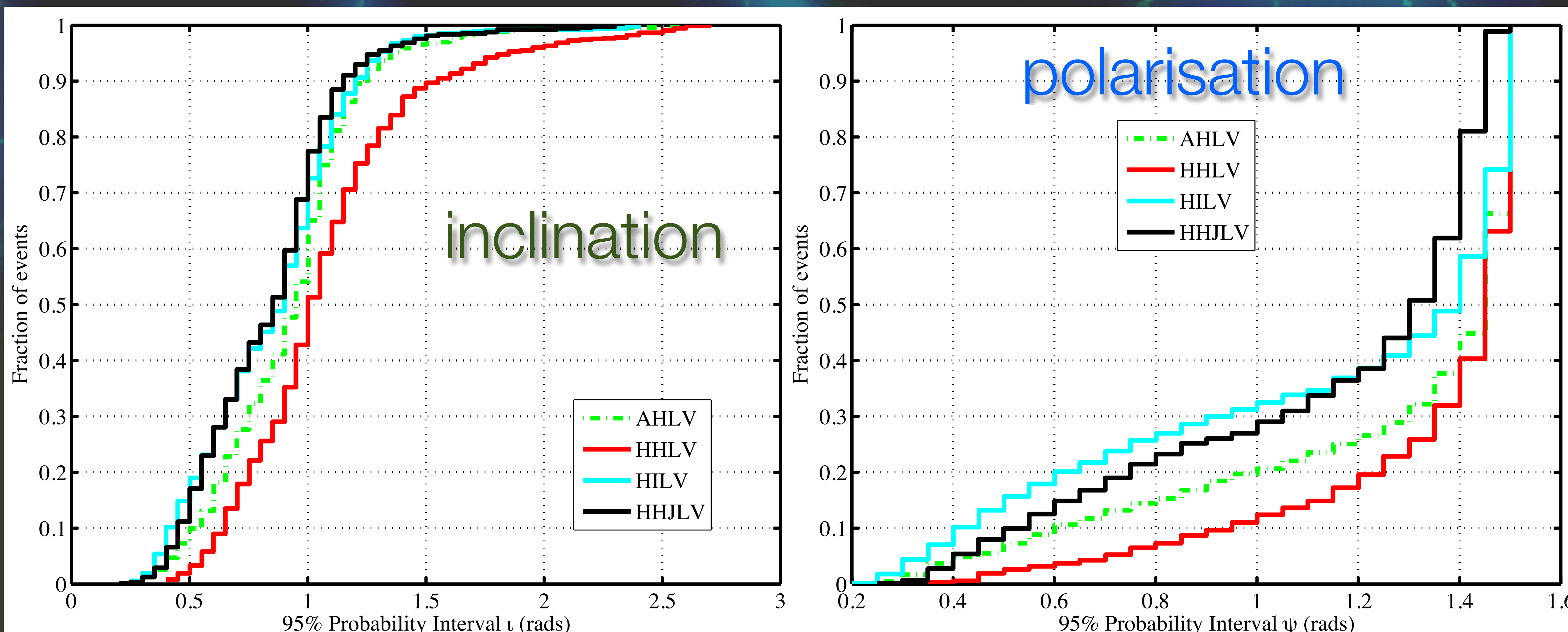
Fraction of events
95% Probability Interval Chirp Mass (M_{sun})
95% Probability Interval η
95% Probability Interval m_1
95% Probability Interval m_2
AHLV
HHLV
HILV
HHJLV



m1
m2

35

Extrinsic parameters



Measuring populations

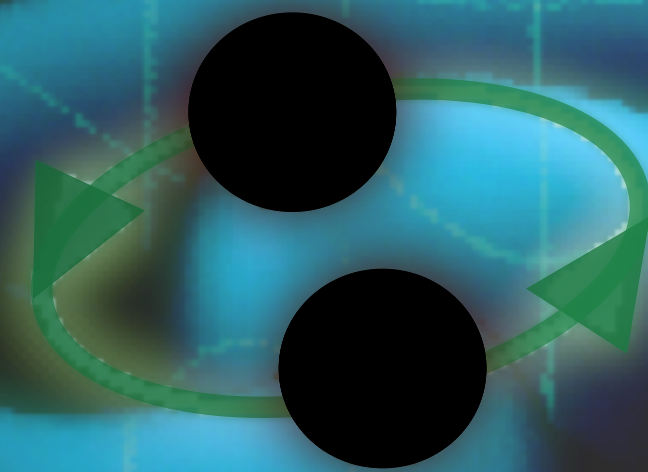
Rate and mass distributions are not properties of any single source. By definition, require consideration of populations.

3 key things to bear in mind:

- Population of detected signals ≠ population of present signals
- Trigger ≠ real signal
- non-trigger ≠ no signal

Considering these facts, what does an analysis look like?

Thresholding



Assume the search applies a threshold on detection statistics $x > x_{th}$. If we receive a GW trigger (D^+), we can look at the data *x* and write the posterior on global parameters γ.

$$p(\boldsymbol{\gamma}|D^+, x, I) = \sum_{\{\mathcal{H}\}} p(\boldsymbol{\gamma}, \mathcal{H}|D^+, x, I)$$

$$= \frac{p(\boldsymbol{\gamma}|I)}{P(x|I)} \overbrace{\sum_{\{\mathcal{H}\}} p(x|\boldsymbol{\gamma}, \mathcal{H}, I) P(\mathcal{H}|\boldsymbol{\gamma}, I)}^{\text{trigger likelihood } p(D^+, x|\boldsymbol{\gamma}, I)}$$

which is summed over the evidence for each model

$$p(x|\boldsymbol{\gamma}, \mathcal{H}, I) = \int_{\Theta_{\mathcal{H}}} d\boldsymbol{\theta}_{\mathcal{H}} \, p(x|\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\gamma}, \mathcal{H}, I) p(\boldsymbol{\theta}_{\mathcal{H}}|\boldsymbol{\gamma}, \mathcal{H}, I),$$

If we do NOT receive a trigger (D^-), we still have information about the global parameters γ! But this time we don't look at the data *x.*

$$p(\boldsymbol{\gamma}|D^-, I) = \frac{p(\boldsymbol{\gamma}|I)}{P(D^-|I)} \overbrace{\sum_{\{\mathcal{H}\}} P(D^-|\boldsymbol{\gamma}, \mathcal{H}, I) P(\mathcal{H}|\boldsymbol{\gamma}, I)}^{\text{non-trigger likelihood } p(D^-|\boldsymbol{\gamma}, I)}$$

Where the unknown value of *x* is marginalised out.

$$P(D^-|\boldsymbol{\gamma}, \mathcal{H}, I) = \int_{\Theta_{\mathcal{H}}} d\boldsymbol{\theta}_{\mathcal{H}} \int_{x \le x_{\mathrm{th}}} dx \; p(x|\boldsymbol{\theta}_{\mathcal{H}}, \boldsymbol{\gamma}, \mathcal{H}, I) p(\boldsymbol{\theta}_{\mathcal{H}}|\boldsymbol{\gamma}, \mathcal{H}, I)$$

Multiple Sources

Divide the run into multiple independent time chunks. Can write

$$p(\boldsymbol{\gamma}|\{D\}, I) = \frac{p(\boldsymbol{\gamma}|I)}{P(\{D\}|I)} \prod_{i=1}^{N} p(D_i|\boldsymbol{\gamma}, I)$$

decomposed into the triggers and non-triggers

$$p(\boldsymbol{\gamma}|\{D\}, I) = \frac{p(\boldsymbol{\gamma}|I)}{P(\{D\}|I)} \prod_{j=1}^{n} p(D^+, x_j|\boldsymbol{\gamma}, I_j) \prod_{k=1}^{N-n} P(D^-|\boldsymbol{\gamma}, I_k),$$

Make the time chunks small enough that mean number of signals per chunk <<1.

Summing over signal and non-signal models, have

$$\begin{aligned} p(\boldsymbol{\gamma}|\{D\}, I) &= \frac{p(\boldsymbol{\gamma}|I)}{P(\{D\}|I)} \\ &\times \prod_{j=1}^{n} \left[p(D^+, x_j|\boldsymbol{\gamma}, \mathcal{H}^+, I_j)P(\mathcal{H}^+|\boldsymbol{\gamma}, I_j) + p(D^+, x_j|\boldsymbol{\gamma}, \mathcal{H}^-, I_j)P(\mathcal{H}^-|\boldsymbol{\gamma}, I_j)\right] \\ &\times \prod_{k=1}^{N-n} \left[P(D^-|\boldsymbol{\gamma}, \mathcal{H}^+, I_k)P(\mathcal{H}^+|\boldsymbol{\gamma}, I_k) + P(D^-|\boldsymbol{\gamma}, \mathcal{H}^-, I_k)P(\mathcal{H}^-|\boldsymbol{\gamma}, I_k)\right]. \end{aligned}$$

where no individual chunk is ever conclusively labelled “signal” or “background”.

likelihood of trigger given signal

likelihood of trigger given noise (FAP)

false dismissal probability (1-eff.)

True dismissal probability (1-FAP)

Example analysis
Mass distribution

Estimate the mean and std. dev. of the distribution of BNSes.

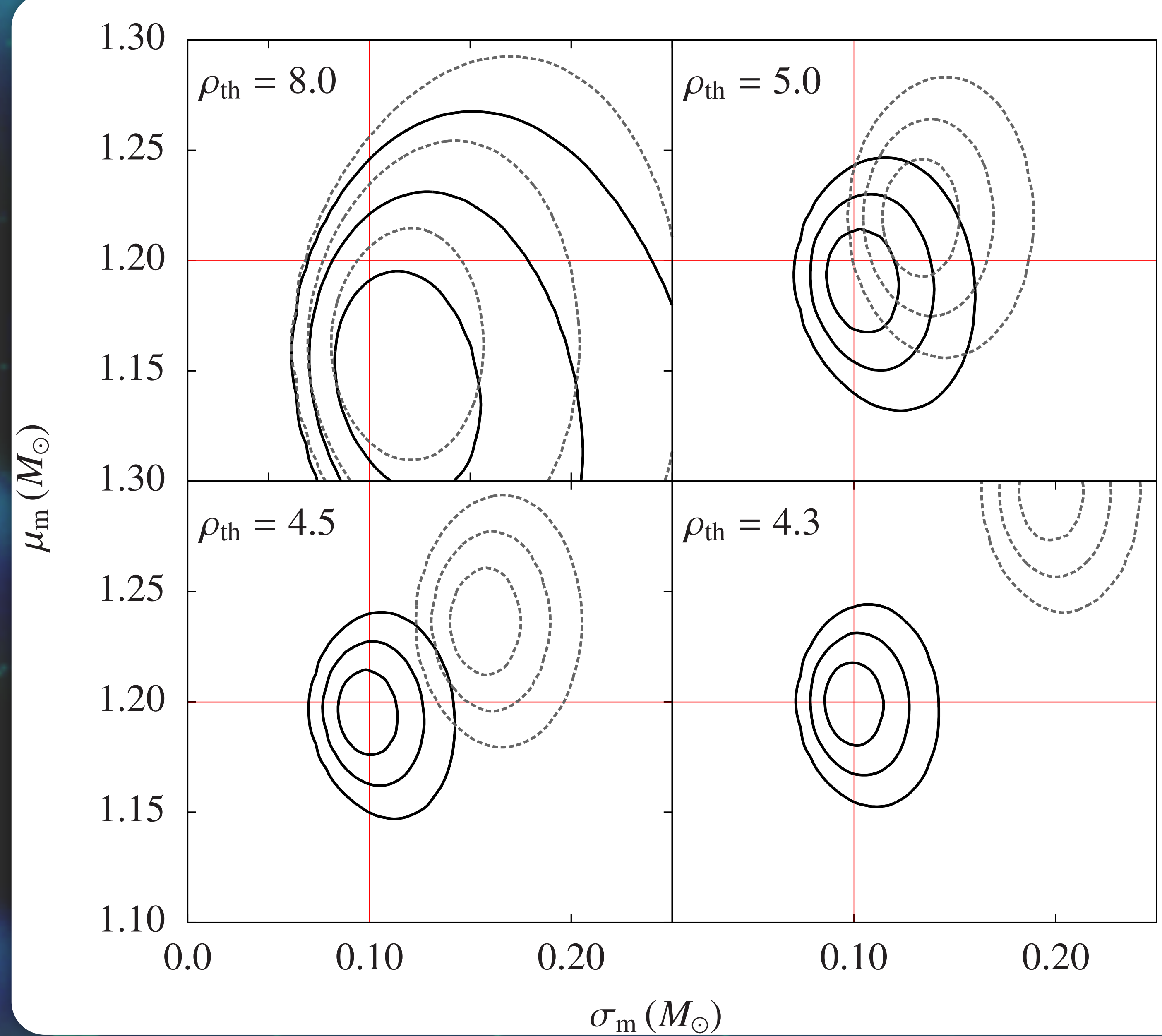
Toy example - assumes masses are measured exactly.

- solid lines - our method
- dashed lines - assuming all triggers are signals

Lowering threshold leads to narrower distributions.

Using only triggers, get increasingly biased result.

Also true for testingGR, cosmology, anything that uses multiple signals



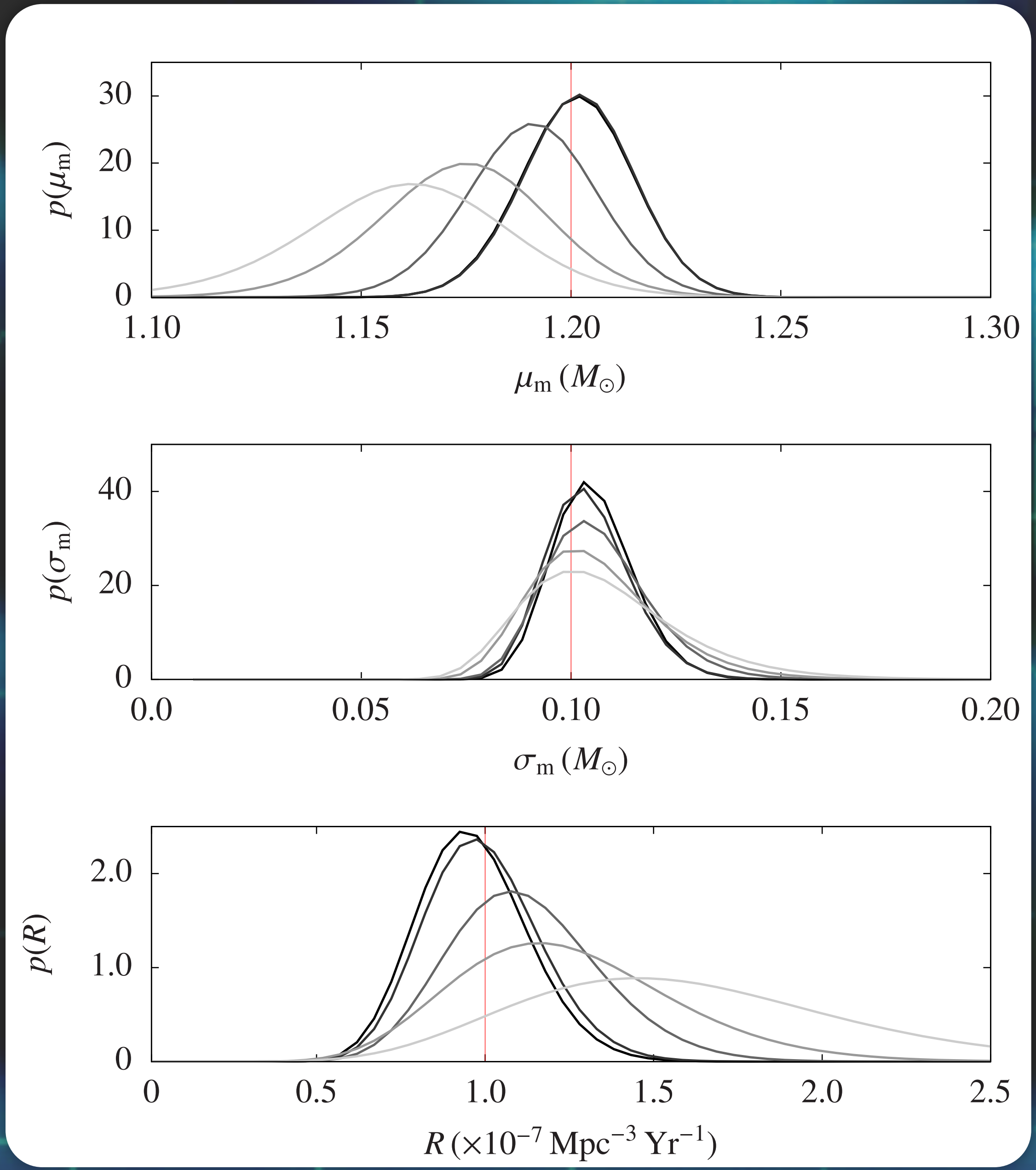
[Messenger & Veitch 2013]

Rate estimation

Lowering threshold again leads to more precise results without bias.

Rate is **special** - determines the prior odds of signal and noise.

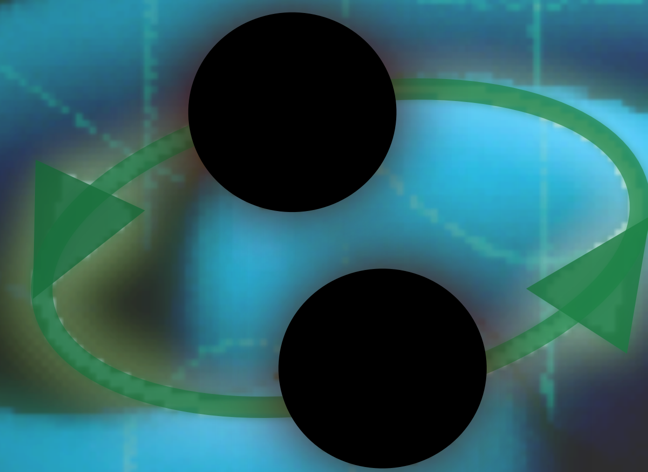
In general, inseparable from other population parameters.



Further work

- Currently extending our population analysis to process posterior samples from parameter estimation codes instead of point estimates.
- Comparing parametric and non-parametric methods of recovering the mass function of NS and BH in binaries [Veitch, Messenger, Del Pozzo (in prep.)]
- Use our method to probe the actual impact of systematic errors from waveforms in realistic scenarios.
- Can we constrain population synthesis models?

Summary



We are now developing the tools that will be used for GW astronomy.

- Astronomy generally involves statistics of multiple sources.
- Searching is the first step. Detected signals are the raw ingredients
  - But it must be well-understood or we will have problems later
  - Selection effects are likely to be important
- Parameter estimation for single sources is mature, and improving in speed and capability
  - Will be used to feed information to more complex models