Bayesian burst detection

Antony Searle (ANU)

with Patrick Sutton (Cardiff / LIGO Caltech), Massimo Tinto (JPL / LIGO Caltech), Shourov Chatterji (INFN Sezione di Roma / LIGO Caltech)

LIGO-G070425-00-Z

Bayesian inference

- Bayesian inference tells us the unique way to *change* the plausibility we assign hypotheses when we get new evidence
  - Therefore, we need to assign plausibilities to the hypotheses *prior* to receiving the evidence
- Priors are criticised as subjective from the perspective of the popular *Frequentist* paradigm
  - Bayesians note that Frequentist statistics are *not* free of priors; their priors are merely *implicit*, *unexamined* and sometimes *contradict intent*

The detection problem

- Things about the observatories we assert
  - Number
  - Locations
  - Antenna patterns
  - Noise spectra
  - Sampling rates
  - Observation time
  - ...
- Things about the gravitational wave we want to learn
  - Existence
  - Time of arrival
  - Direction of origin
  - Waveform
- We need *priors distributions* for these

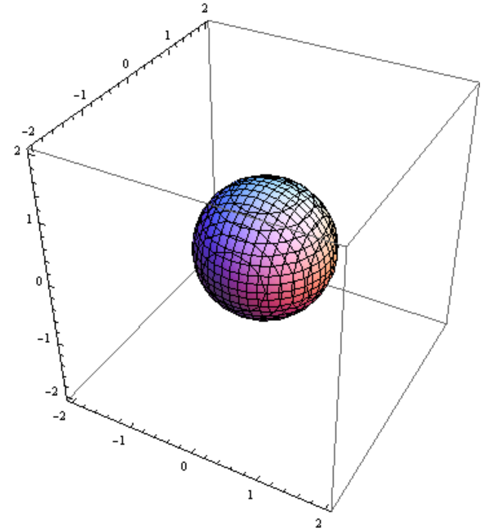
Toy model

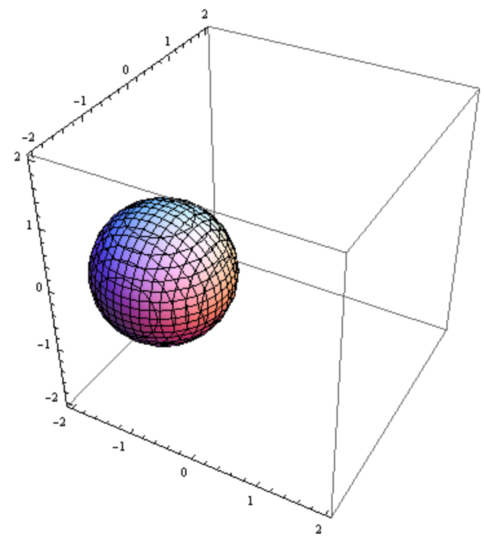
- N white detectors each make a single measurement as a postulated strain $\mathbf{h}$ from direction (θ,φ) sweeps over them

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} F_1^+(\theta,\phi) & F_1^\times(\theta,\phi) \\ F_2^+(\theta,\phi) & F_2^\times(\theta,\phi) \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_+ \\ h_\times \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \end{bmatrix}$$

$$p(\mathbf{x} \mid H_{\text{noise}}) = (2\pi)^{-N/2} \exp -\frac{1}{2}\mathbf{x}^T\mathbf{x}$$

$$p(\mathbf{x} \mid \theta,\phi,\mathbf{h}) = (2\pi)^{-N/2} \exp -\frac{1}{2}(\mathbf{x}-\mathbf{Fh})^T(\mathbf{x}-\mathbf{Fh})$$





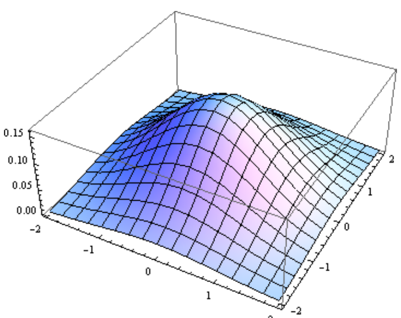
Uncontroversial priors

- How plausible is it that a gravitational wave is present?
  - This follows from the predicted event rate and is comparable to a Frequentist threshold $\frac{p(H_{\text{signal}})}{p(H_{\text{noise}})} << 1$
- When and from where?
  - Uniform over observation time and sky direction $p(\theta, \phi \mid H_{\text{signal}}) = \frac{1}{4\pi} \sin\theta$

Waveform prior

- A plausibility distribution on the space of all possible strain waveforms
- Example: a population of white noise bursts with power-law distributed energies

$$p\left(\mathbf{h} \mid H_{signal}\right) = \int_{-\infty}^{\infty} d\sigma \; p\left(\sigma \mid H_{signal}\right) p\left(\mathbf{h} \mid \sigma\right)$$



$$= \int_{1}^{\infty} d\sigma \, \frac{1}{3\sigma^4} \frac{1}{2\pi\sigma^2} \exp \frac{1}{2\sigma^2} \mathbf{h}^T \mathbf{h}$$

Marginalising away strain

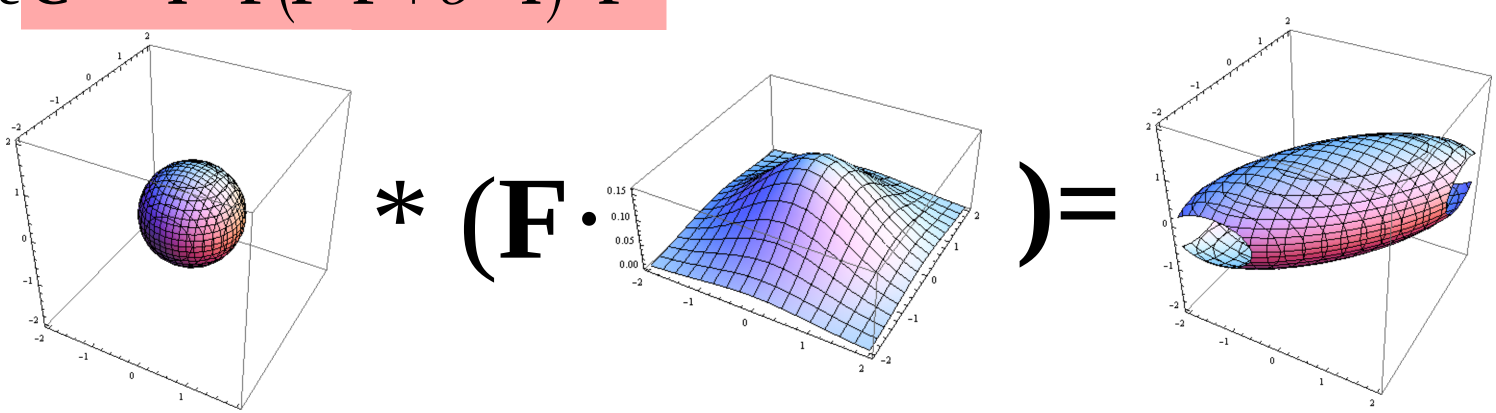
- We can analytically marginalize away strain

$$p(\mathbf{x}\mid\theta,\phi,\sigma) = \iint_{R^2} \mathrm{d}\mathbf{h}\; p(\mathbf{x}\mid\mathbf{h})\,p(\mathbf{h}\mid\sigma)$$

$$= \iint_{R^2} \mathrm{d}\mathbf{h}\,(2\pi)^{-(N+2)/2}\sigma^{-2}\exp-\frac{1}{2}\left[(\mathbf{x}-\mathbf{F}\mathbf{h})^T(\mathbf{x}-\mathbf{F}\mathbf{h})-\sigma^{-2}\mathbf{h}^T\mathbf{h}\right]$$

$$= (2\pi)^{-N/2}(\det\mathbf{C})^{-1/2}\exp-\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}$$

where $\mathbf{C}^{-1} = \mathbf{I}-\mathbf{F}\left(\mathbf{F}^T\mathbf{F}+\sigma^{-2}\mathbf{I}\right)^{-1}\mathbf{F}^T$



Result

- We have to numerically marginalize over other parameters

$$p(\mathbf{x} \mid H_{\text{signal}}) = \int_0^{\pi} \int_{-\pi}^{\pi} \int_1^{\infty} p(\theta, \phi) p(\sigma) p(\mathbf{x} \mid \theta, \phi, \sigma) \, \mathrm{d}\sigma \, \mathrm{d}\phi \, \mathrm{d}\theta$$

  – (Not very expensive)

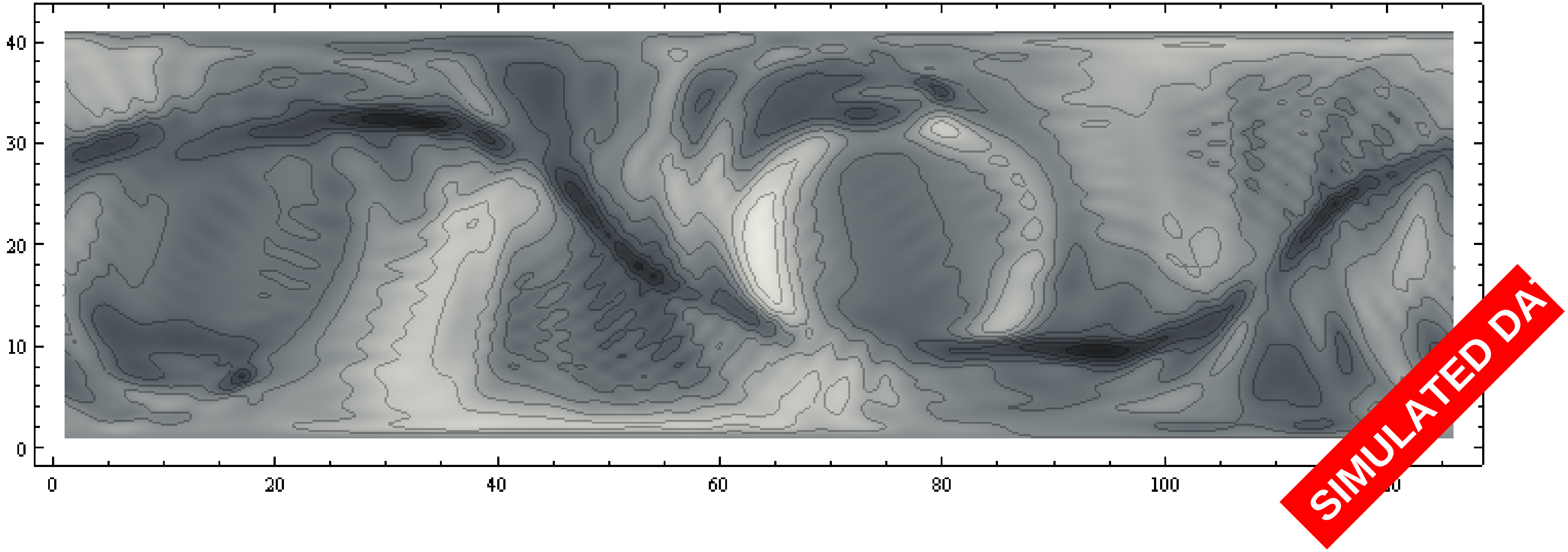
- ...to get the Bayesian odds ratio

$$\frac{p(H_{\text{signal}} \mid \mathbf{x})}{p(H_{\text{noise}} \mid \mathbf{x})} = \frac{p(H_{\text{signal}})}{p(H_{\text{noise}})} \frac{p(\mathbf{x} \mid H_{\text{signal}})}{p(\mathbf{x} \mid H_{\text{noise}})}$$

Bayesian "sky map"

$$\frac{p(\theta,\phi\,|\,\mathbf{x})}{p(H_{\mathrm{noise}}\,|\,\mathbf{x})} = \frac{p(H_{signal})\,p(\theta,\phi)\int_0^\infty p(\sigma)\,p(\mathbf{x}\,|\,\theta,\phi,\sigma)\,\mathrm{d}\sigma}{p(H_{\mathrm{noise}})\,p(\mathbf{x}\,|\,H_{\mathrm{noise}})}$$

$$= \frac{p(H_{signal})}{p(H_{\mathrm{noise}})}\,p(\theta,\phi)\int_0^\infty p(\sigma)(\det\mathbf{C})^{-1/2}\exp-\frac{1}{2}\mathbf{x}^T\left(\mathbf{C}^{-1}-\mathbf{I}\right)\mathbf{x}\,\mathrm{d}\sigma$$



Comparing with Gursel-Tinto

- “Optimal statistic is...”

$$\exp - \frac{1}{2} \mathbf{x}^T \mathbf{F} \left(\mathbf{F}^T \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{x}$$

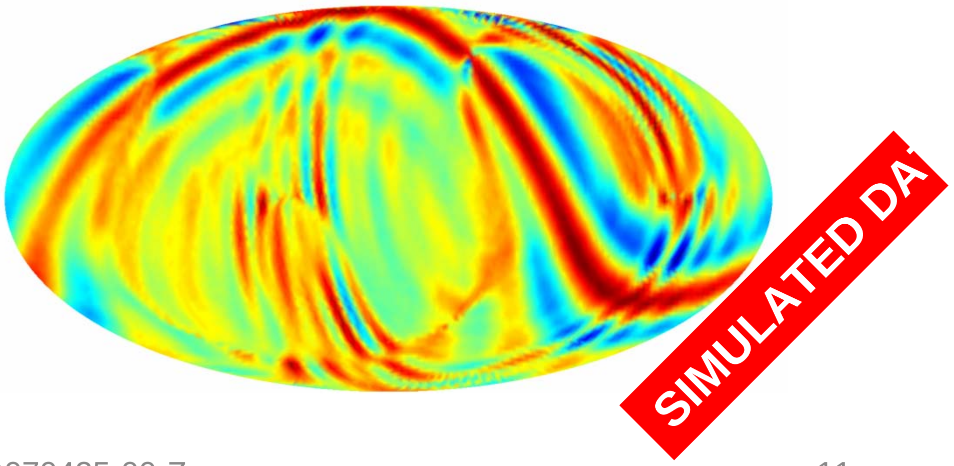
  – Y. Gürsel and M. Tinto, PRD **40**, 3884 (1989)
  – Implemented as xpipeline (ANU/Caltech/JPL)
- Bayesian sky map limits to this for

$$p(\sigma) = \delta(\sigma - a), \quad p(\theta, \phi) \propto (\det \mathbf{C})^{1/2}$$

$$\lim_{a \to \infty} \frac{p(\theta, \phi \mid \mathbf{x})}{p(H_{\text{noise}} \mid \mathbf{x})} \propto \lim_{a \to \infty} \exp - \frac{1}{2} \mathbf{x}^T \left(\mathbf{C}^{-1} - \mathbf{I} \right) \mathbf{x} = \exp - \frac{1}{2} \mathbf{x}^T \mathbf{F} \left(\mathbf{F}^T \mathbf{F} \right)^{-1} \mathbf{F}^T \mathbf{x}$$

Odd priors

- Gursel-Tinto is related to a Bayesian analysis with odd priors
  - Very (very) large signal energy!
  - Source directions distributed according to network sensitivity!
- These are consistent with GT's observed failure modes
- The priors aren't "incorrect", but...
- They certainly don't reflect Gursel and Tinto's beliefs about the universe

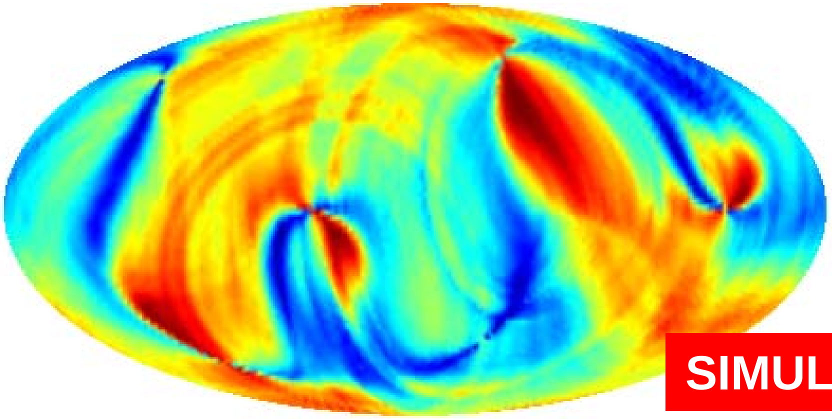


Other comparisons

- Soft constraint
  - S Klimekno *et al*, PRD **72**, 122002 (2005); Coherent WaveBurst (UFL)
  - Limiting case of infinitely small signals, non-uniform direction prior
- Tikhonov regularization
  - M Rakhmanov, CQG **23** 19 (2006) S673-S685; RIDGE (UTB/PennState)
  - Looks for signals of a particular energy, non-uniform direction prior
- All these techniques work to varying degrees...
  - Enough evidence can always overwhelm a prior
- The most effective analysis is the one whose priors best reflect reality
  - (and can be computed; the Bayesian analysis cost is comparable to Gursel-Tinto, depending on choice of strain prior)

Robust noise model

- In practice, coherent methods are easily fooled by incoherent glitches



Gursel-Tinto often mistakes glitches (shown) for gravitational waves from directions of poor sensitivity

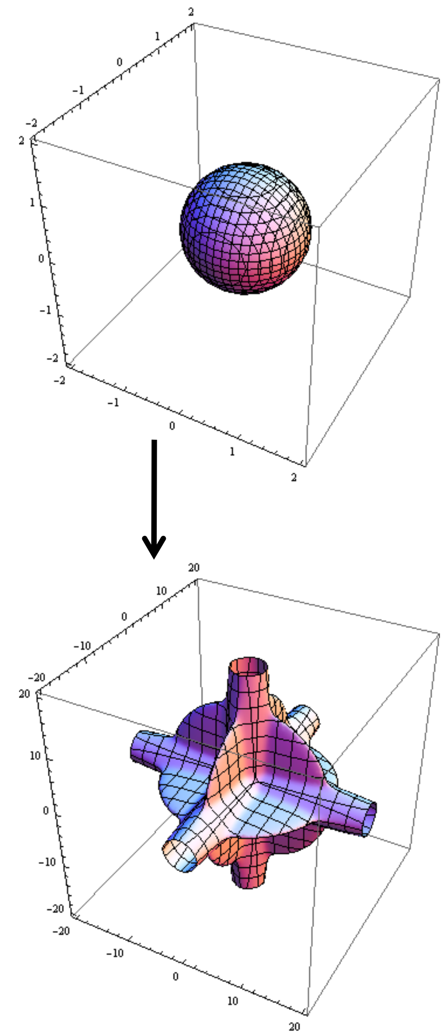
  – The analysis can only explain any excess power as a gravitational wave
  – S Chatterji *et al*, Phys. Rev. D **74**, 082005 (2006) demonstrates and proposes a more robust statistic

Robust noise model

- We can make a more realistic noise model where the detectors occasionally glitch
  - This requires a glitch model similar to the signal model, with physically motivated priors on glitch waveforms and occurrence

$$p(\mathbf{x} \mid H_{\text{noise}}) = \prod_{i=1}^{N} \left(\frac{p(H_{\text{quiet}})}{\sqrt{2\pi}} \exp \frac{-x_i^2}{2} + \frac{p(H_{\text{glitch}})}{\sigma_g \sqrt{2\pi}} \exp \frac{-x_i^2}{2\sigma_g} \right)$$

  - Easily integrated into Bayesian analysis at little extra cost



Summary

- The Bayesian approach to bursts
  - Supersedes several previously proposed methods
  - Necessarily outperforms those methods
    - Priors target more reasonable signals
  - Is an optimal test uniquely defined by making explicit assertions about the instruments and bursts
  - Is computationally tractable
    - Cost is comparable to existing methods for comparable signal models
  - Can readily incorporate glitch models for

Supplementary material

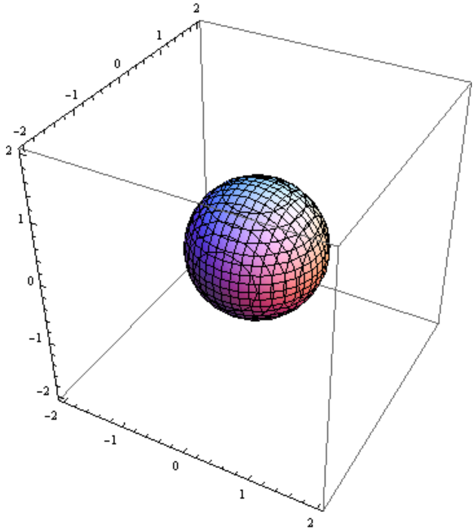
Toy model

- N white detectors each make a measurement as a postulated strain $\mathbf{h}$ from direction (θ,φ) sweeps over them

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} F_1^{+}(\theta,\phi) & F_1^{\times}(\theta,\phi) \\ F_2^{+}(\theta,\phi) & F_2^{\times}(\theta,\phi) \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_+ \\ h_\times \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \end{bmatrix}$$

- If no wave is present, the measurements are normally distributed around $\mathbf{0}$

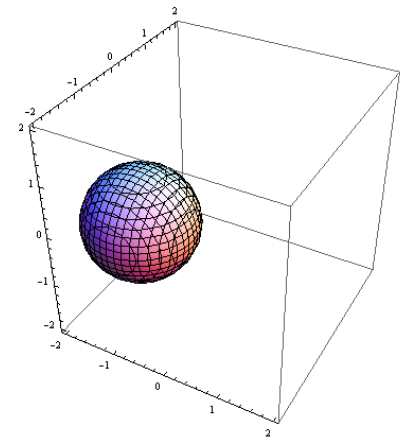
$$p(\mathbf{x} \mid H_{\text{noise}}) = (2\pi)^{-N/2} \exp -\frac{1}{2}\mathbf{x}^T\mathbf{x}$$



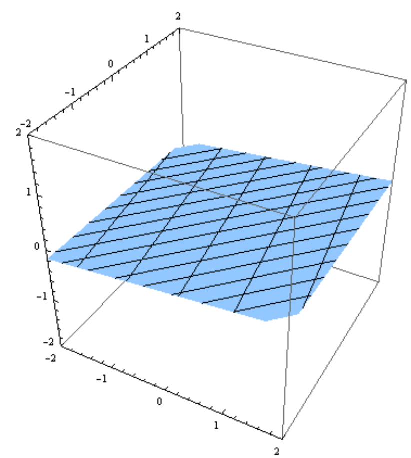
Towards a signal hypothesis

- If the wave is present, the measurements are normally distributed around the response $\mathbf{Fh}$

$$p(\mathbf{x} \mid \mathbf{h}) = p(\mathbf{x} - \mathbf{Fh} \mid H_{\text{noise}})$$
$$= (2\pi)^{-N/2} \exp - \frac{1}{2}(\mathbf{x} - \mathbf{Fh})^T (\mathbf{x} - \mathbf{Fh})$$



- Unfortunately we don't know the incoming strain, so this is not directly useful
  - Gursel & Tinto's original insight was that, as the response is constrained to $\operatorname{span} \mathbf{F}$, any function of $(\operatorname{null} \mathbf{F})^T \mathbf{x}$ was independent of the unknown $\mathbf{h}$
  - The Bayesian analysis instead uses a *prior* on $\mathbf{h}$



Prior expectations of strain

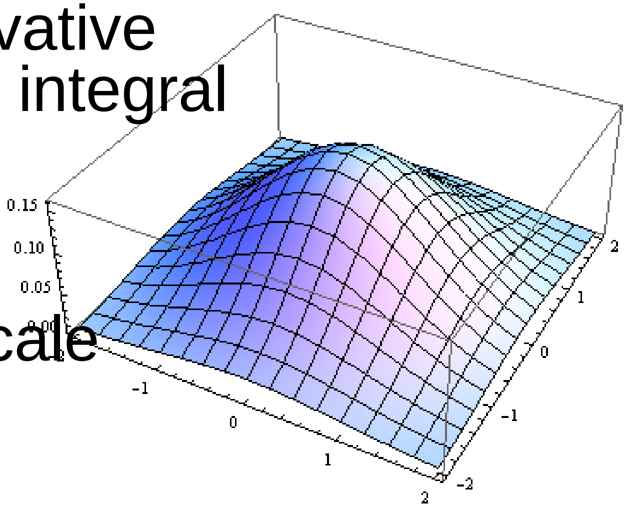
- For detection (not characterization) we want to marginalize away the nuisance parameter $\mathbf{h}$

$$p(\mathbf{x} \mid H_{\text{signal}}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(\mathbf{x} - \mathbf{F}\mathbf{h} \mid H_{\text{noise}})\, p(\mathbf{h} \mid H_{\text{signal}})\, \mathrm{d}h_{+}\, \mathrm{d}h_{\times}$$

- To do so, we need to specify how likely we think particular strains are to occur
  - A normal distribution is a conservative choice that also lets us solve the integral

$$p(\mathbf{h} \mid H_{\text{signal}}) = (2\pi)^{-1} \exp -\frac{1}{2}\sigma_h^{-2}\mathbf{h}^T\mathbf{h}$$

  - We must specify the expected scale of the strain

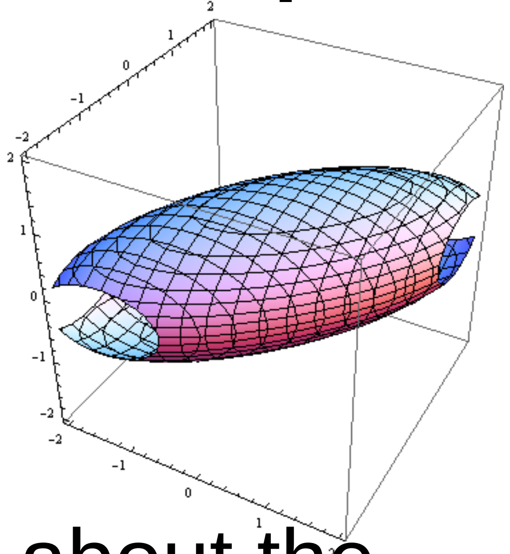


Explicit signal hypothesis

$$p\left(\mathbf{x} \mid H_{\text{signal}}\right) = (2\pi)^{-(N+2)/2} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp -\frac{1}{2}\left[(\mathbf{x}-\mathbf{F}\mathbf{h})^T(\mathbf{x}-\mathbf{F}\mathbf{h}) - \sigma_h^{-2}\mathbf{h}^T\mathbf{h}\right] \mathrm{d}h_+ \mathrm{d}h_\times$$

$$= (2\pi)^{-N/2}(\det \mathbf{C})^{-1/2} \exp -\frac{1}{2}\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}$$

$$\text{where } \mathbf{C}^{-1} = \mathbf{I} - \mathbf{F}\left(\mathbf{F}^T\mathbf{F} + \sigma_h^{-2}\mathbf{I}\right)^{-1}\mathbf{F}^T$$

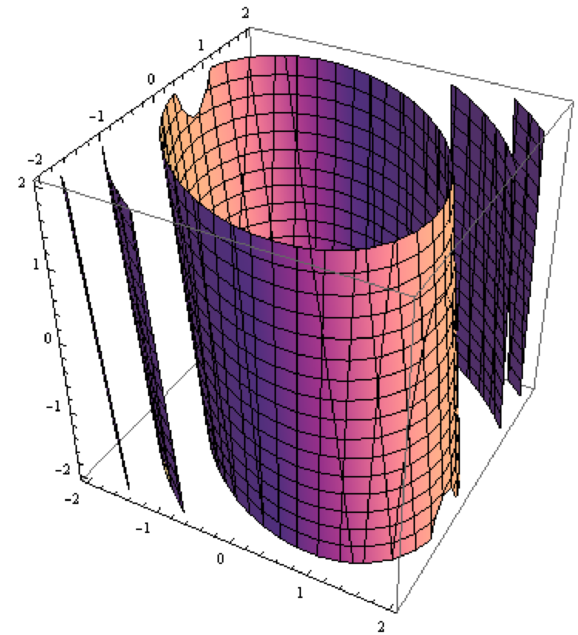


- By making a weak assumption about the strain we obtain an explicit signal hypothesis

Bayesian odds ratio

- We have a prior expectation that signals are infrequent

$$p\left(H_{\text{signal}}\right) << p\left(H_{\text{noise}}\right)$$



- We can directly compute the relative plausibilities of the competing hypotheses

$$\frac{p\left(H_{\text{signal}} \mid \mathbf{x}\right)}{p\left(H_{\text{noise}} \mid \mathbf{x}\right)} = \frac{p\left(H_{\text{signal}}\right)}{p\left(H_{\text{noise}}\right)} \frac{p\left(\mathbf{x} \mid H_{\text{signal}}\right)}{p\left(\mathbf{x} \mid H_{\text{noise}}\right)}$$

$$= \frac{p\left(H_{\text{signal}}\right)}{p\left(H_{\text{noise}}\right)} \left(\det \mathbf{C}\right)^{-1/2} \exp - \frac{1}{2} \mathbf{x}^T \left(\mathbf{C}^{-1} - \mathbf{I}\right) \mathbf{x}$$

Relationship to other methods

- Very large signal prior is like *Gursel-Tinto*

$$\sigma_h >> F^{-1},\ \mathbf{C}^{-1} - \mathbf{I} \approx \mathbf{F}\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T$$

- Very small signal prior is like *soft constraint*

$$\sigma_h << F^{-1},\ \mathbf{C}^{-1} - \mathbf{I} \approx \sigma_h^2 \mathbf{F}\,\mathbf{F}^T$$

- Physical meaning for *Tikhonov regularizer*

$$\mathbf{C}^{-1} - \mathbf{I} = \mathbf{F}\left(\mathbf{F}^T\mathbf{F} + \sigma_h^{-2}\mathbf{I}\right)^{-1}\mathbf{F}^T$$

- Previous methods are like Bayesian searches with *poorly chosen priors*
  – Their prior expectations are unexamined, not absent!

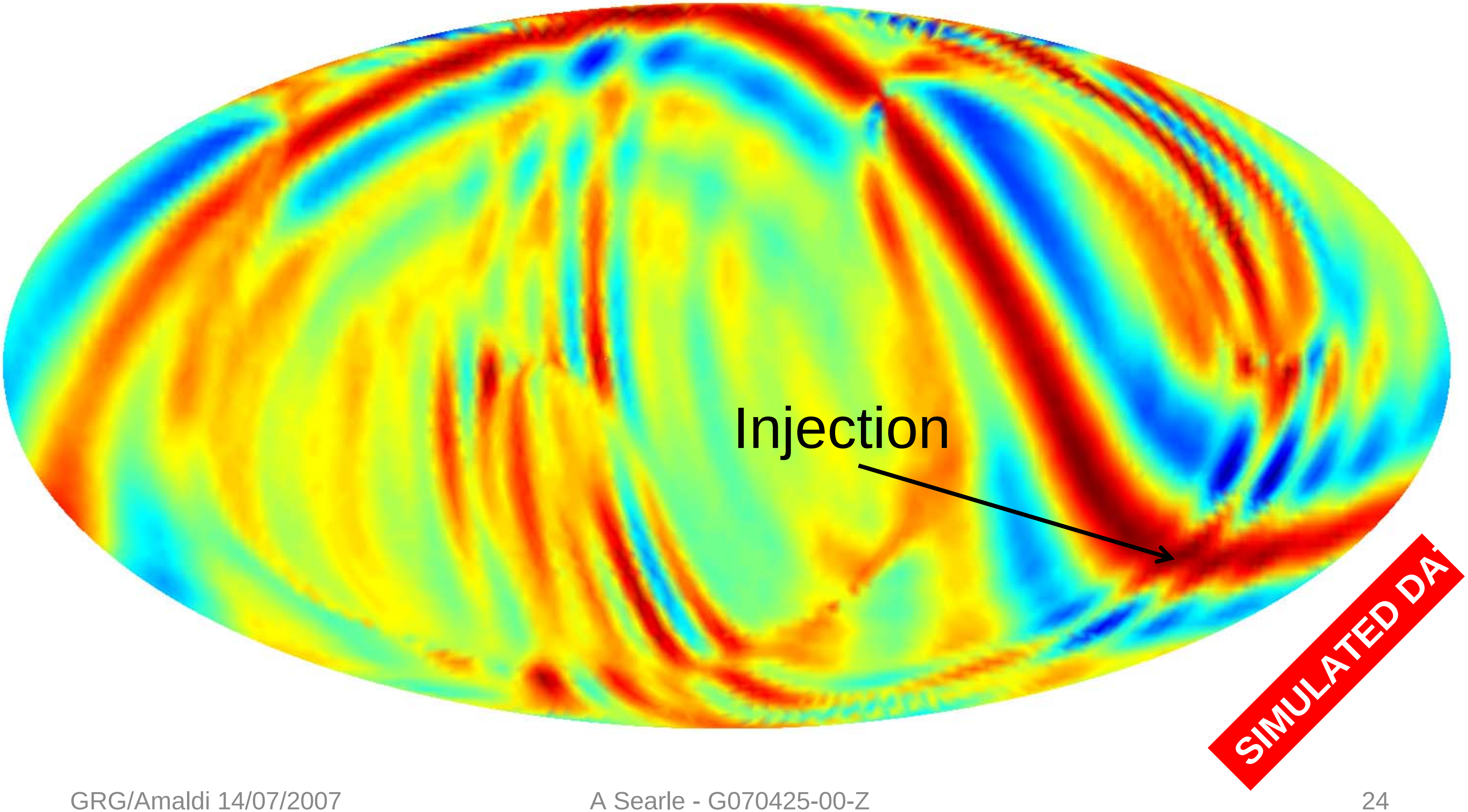
Reverse-engineering priors

- Gursel-Tinto method:
  - Y. Gürsel and M. Tinto, PRD **40**, 3884 (1989); xpipeline (ANU/Caltech/JPL)
  - "Optimal statistic is null energy $\mathbf{x}^T(\mathbf{I}-\mathbf{F}(\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T)\mathbf{x}$"
- Bayesian large signal limit

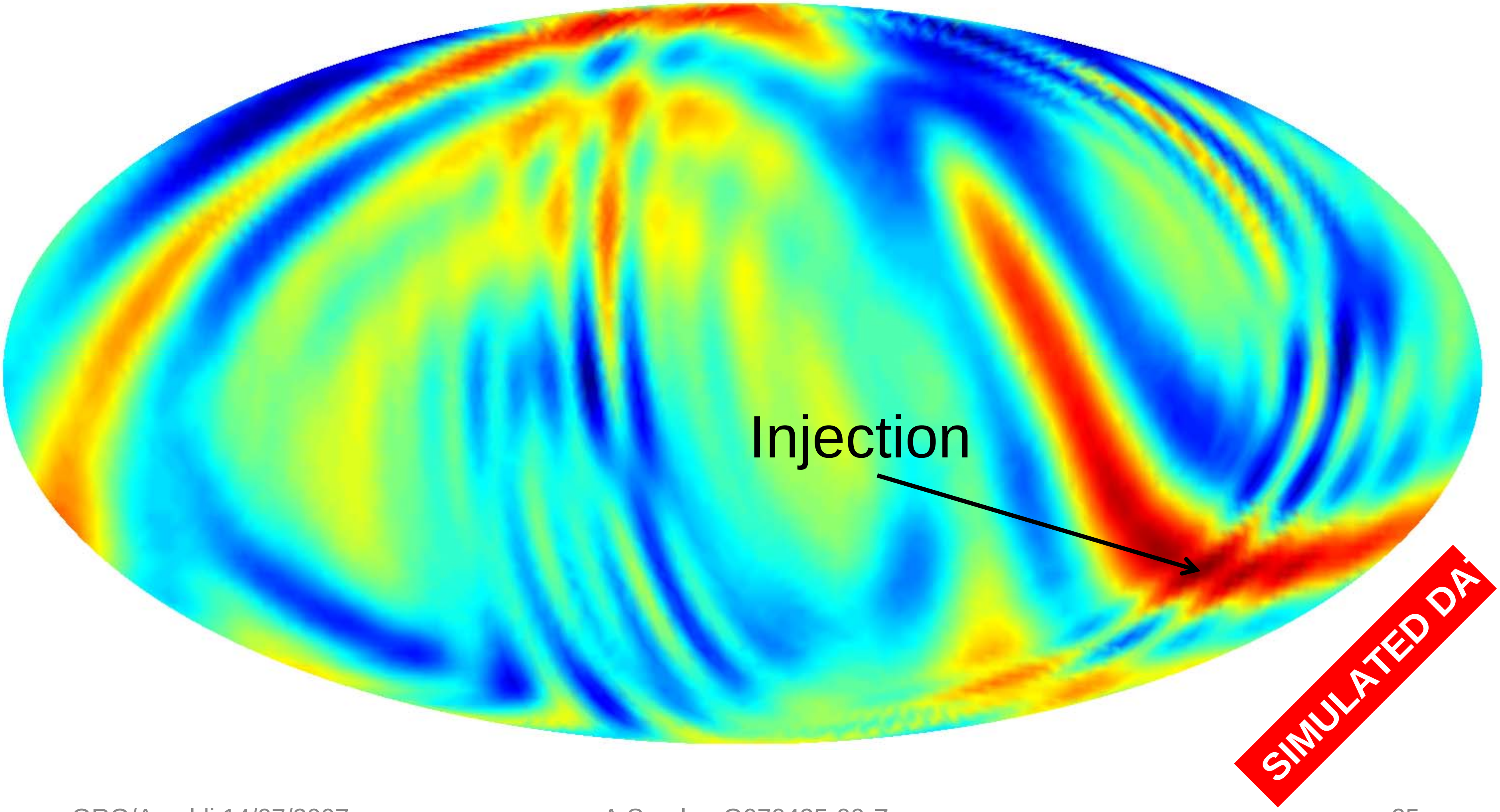
$$\sigma >> P^{-1},\ \mathbf{C}^{-1} \approx \mathbf{I}-\mathbf{F}\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T$$

  - GT is limiting case of expecting infinitely large signals (from the network's least sensitive directions!)
    - These are consistent with GT's observed failure modes
- These priors are not *wrong*, but they certainly weren't Gursel and Tinto's *intent*

Gursel-Tinto for HLV

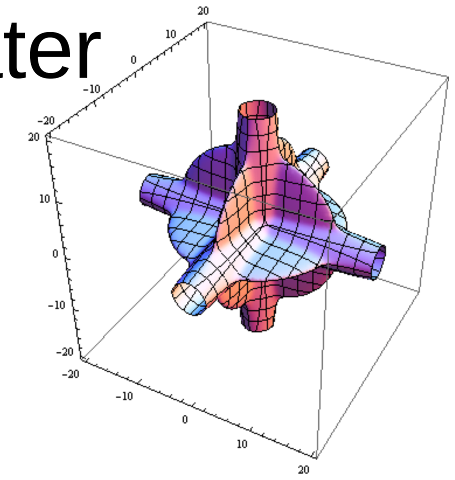


$\ln p(\theta, \boldsymbol{\varphi} | \mathbf{x}, \sigma, H_{\mathrm{signal}})$ for HLV



Advanced noise models

- None of the above methods are good at rejecting ‘glitches’
  - The models only have one way excess energy appears in a detector: a gravitational wave
- Generalize noise model for greater robustness
  - Consider a different kind of signal: infrequent uncorrelated bursts of noise
  - The Bayesian analysis can now prefer this hypothesis when appropriate



Outcome

- An expression to enable us to compute the plausibility that a gravitational wave is present
  - “Sky maps” produced by previous methods are comparable to