Data Analysts Without Borders: A Case Study

LIGO-G1000994-v3

GW2010 UMN
Oct. 16, 2010

Data Analysts Without Borders: A Case Study

Warren G. Anderson
University of Wisconsin - Milwaukee

UWM
UNIVERSITY of WISCONSIN
MILWAUKEE

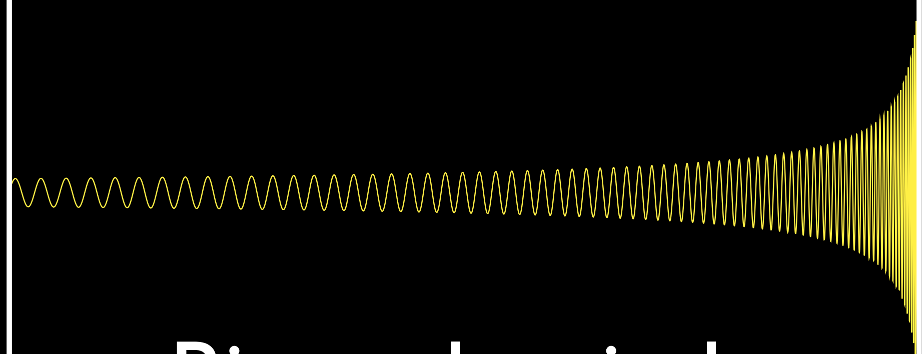
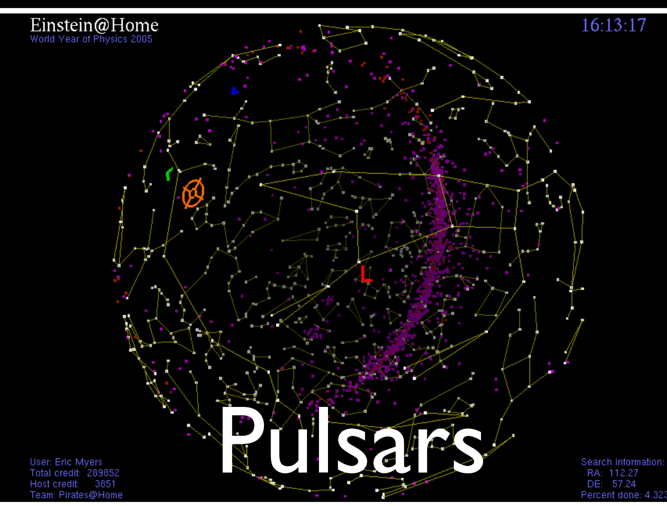
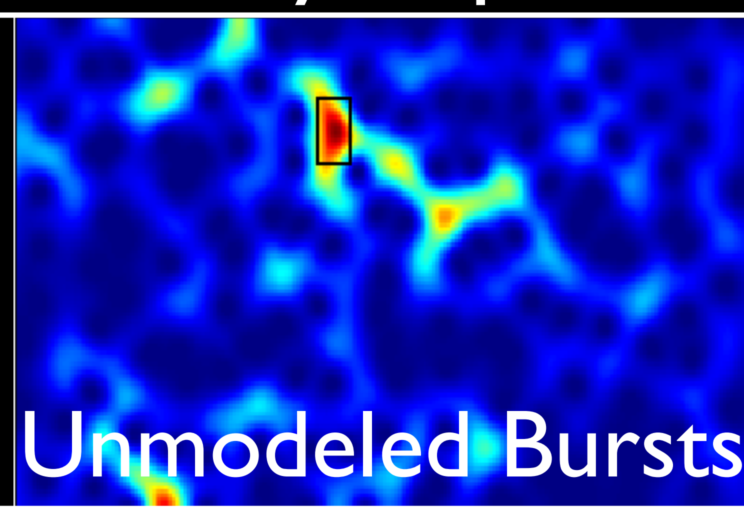
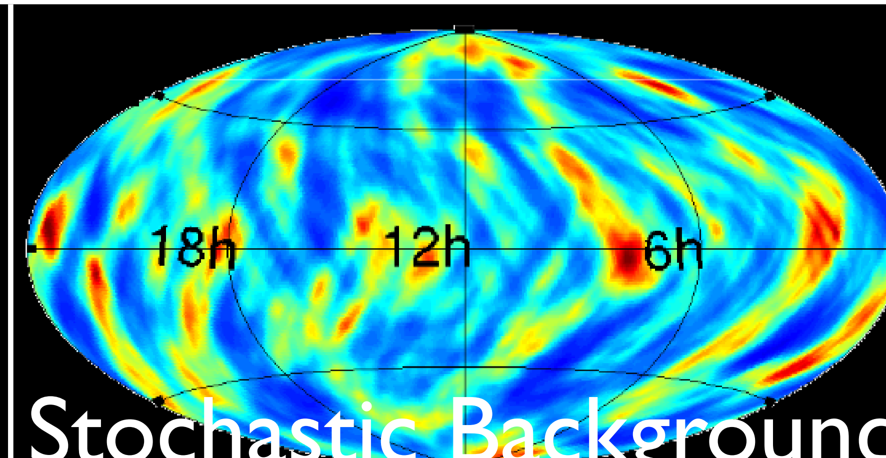
Image courtesy of NASA Goddard NR group

Outline

- Conventions
  - Four types of DA
- Foundations
  - Data, Likelihood, Power, Power^2
- STAMP
  - Motivation, Idea, Progress

Convention

- Ground-based gravitational wave DA is often divided into four categories:

Category	Short Duration	Long Duration
Theoretical Waveform	 Binary Inspirals	 Pulsars
No Theoretical Waveform	 Unmodeled Bursts	 Stochastic Background

Data

- From our instruments we get strain data $h(t)$

$$h(t) = n(t) + s(t)$$

Noise *Signal*

Data

- From our instruments we get strain data $h(t)$

$$h_i = n_i + s_i$$

Noise *Signal*

Data

- From our instruments we get strain data $h(t)$

$$h_i = n_i + s_i$$

Noise *Signal*

- Signal is deterministic (possibly $s_i = 0$)
- Noise is stochastic - eg Gaussian (ideally)

$$p(n) \propto \exp\left(-\frac{n \cdot n}{2\,\sigma}\right)$$

Likelihood

- Neyman-Pearson: Optimal statistic is ***Likelihood***

$$\Lambda[h] = \int \frac{p(h|n+\boldsymbol{s})}{p(h|n)}\,\mathcal{D}[s]$$

- The measure $\mathcal{D}[s]$ specifies the class of signals to search for by projecting, eg to search for a single specific signal $\hat{s}(t)$, let $\mathcal{D}[s] = \delta(s - \hat{s})\,ds$.

Likelihood

- Neyman-Pearson: Optimal statistic is ***Likelihood***

$$\Lambda[h] = \int \frac{p(h|n+\boldsymbol{s})}{p(h|n)} \boxed{\mathcal{D}[s]}$$

Signal

- The measure $\mathcal{D}[s]$ specifies the class of signals to search for by projecting, eg to search for a single specific signal $\hat{s}(t)$, let $\mathcal{D}[s] = \delta(s - \hat{s})\, ds$.

Likelihood

- Neyman-Pearson: Optimal statistic is ***Likelihood***

$$\Lambda[h] = \int \frac{p(h|n+s)}{p(h|n)} \underbrace{\boxed{\mathcal{D}[s]}}_{\textit{Signal}}$$

- Probabilities of h derive from distribution of n, eg

$$p(h|n+s) \propto e^{-\frac{(h-s)\cdot(h-s)}{2\,\sigma}}$$

Likelihood

- Neyman-Pearson: Optimal statistic is ***Likelihood***

$$\Lambda[h] = \int \boxed{\frac{p(h|n+s)}{p(h|n)}} \, \boxed{\mathcal{D}[s]}$$

Noise Signal

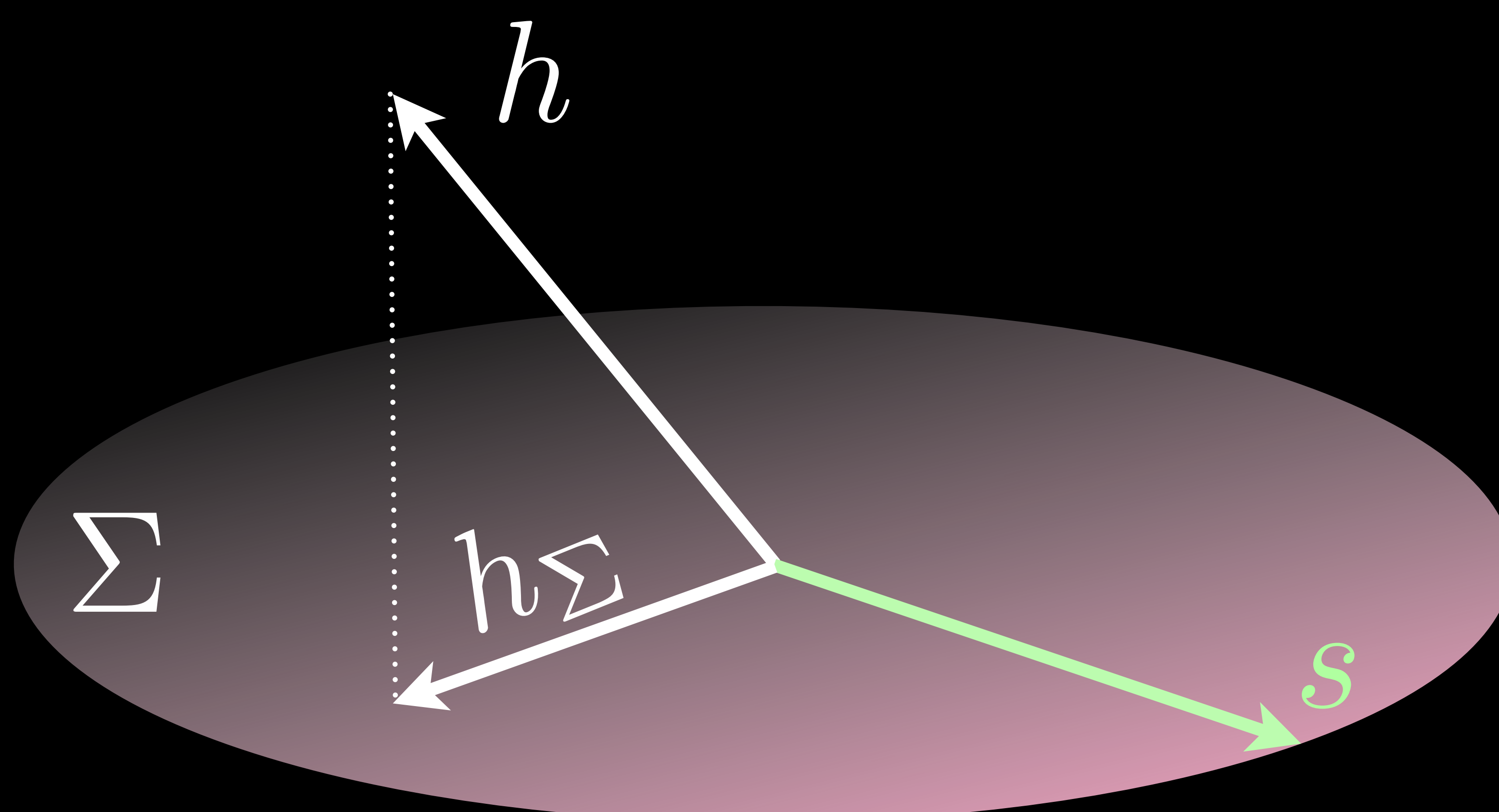
- Probabilities of h derive from distribution of n, eg

$$p(h|n+s) \propto e^{-\frac{(h-s)\cdot(h-s)}{2\,\sigma}}$$

Power

- Putting this together for Gaussian noise:

$$\Lambda \propto \int_{\Sigma} \exp(h_{\Sigma} \cdot s - s \cdot s/2)\, ds$$



Power

- Putting this together for Gaussian noise:

$$\Lambda \propto \int_{\Sigma} \exp(h_{\Sigma} \cdot s - s \cdot s/2)\, ds$$

- Note that Λ is monotonic in length of projection, $|h_{\Sigma}|$.
- ***Power*** of projected data, $|h_{\Sigma}|^2$, is eqivalently optimal statistic.

Power^2

- For more than one detector, simply substitute:

$$h = A^i(\Omega) h_i(t_i)$$

- *Eg*, for two detectors with Gaussian noise, optimal statistic is

$$\begin{aligned} |h_\Sigma|^2 &= |\ [A_1 h_1 + A_2 h_2]_\Sigma\ |^2 \\ &= A_1^2 |h_1^2|_\Sigma + A_2^2 |h_2^2|_\Sigma \\ &\quad + 2 A_1 A_2 [h_1 \cdot h_2]_\Sigma \end{aligned}$$

Power^2

- For more than one detector, simply substitute:

$$h = A^i(\Omega)h_i(t_i)$$

- *Eg*, for two detectors with Gaussian noise, optimal statistic is

$$\begin{aligned} |h_\Sigma|^2 &= |\ [A_1 h_1 + A_2 h_2]_\Sigma\ |^2 \\ &= \boxed{A_1^2|h_1^2|_\Sigma + A_2^2|h_2^2|_\Sigma \quad \text{auto power}} \\ &\quad + 2A_1A_2[h_1 \cdot h_2]_\Sigma \end{aligned}$$

Power^2

- For more than one detector, simply substitute:

$$h = A^i(\Omega)h_i(t_i)$$

- *Eg*, for two detectors with Gaussian noise, optimal statistic is

$$
\begin{aligned}
|h_\Sigma|^2 &= |\ [A_1 h_1 + A_2 h_2]_\Sigma\ |^2 \\
&= \boxed{A_1^2|h_1^2|_\Sigma + A_2^2|h_2^2|_\Sigma \quad \text{auto power}} \\
&\quad + \boxed{2A_1A_2[h_1 \cdot h_2]_\Sigma \quad \text{cross power}}
\end{aligned}
$$

Glitches

- Real interferometer noise is not Gaussian, but has "glitches".
- Two problems:
  - No analytic model - need time-slides or other methods to assess false alarm/detection stats.
  - Power statistic no longer optimal because loud glitches in one instrument cause large auto-power but real signals should have cross-power.

Idea

- Use cross-power to look for various types of signals.
- Advantages:
  - Cross-power statistics very close to Gaussian - close enough?
  - Cross-power statistic less susceptible to glitches.

STAMP

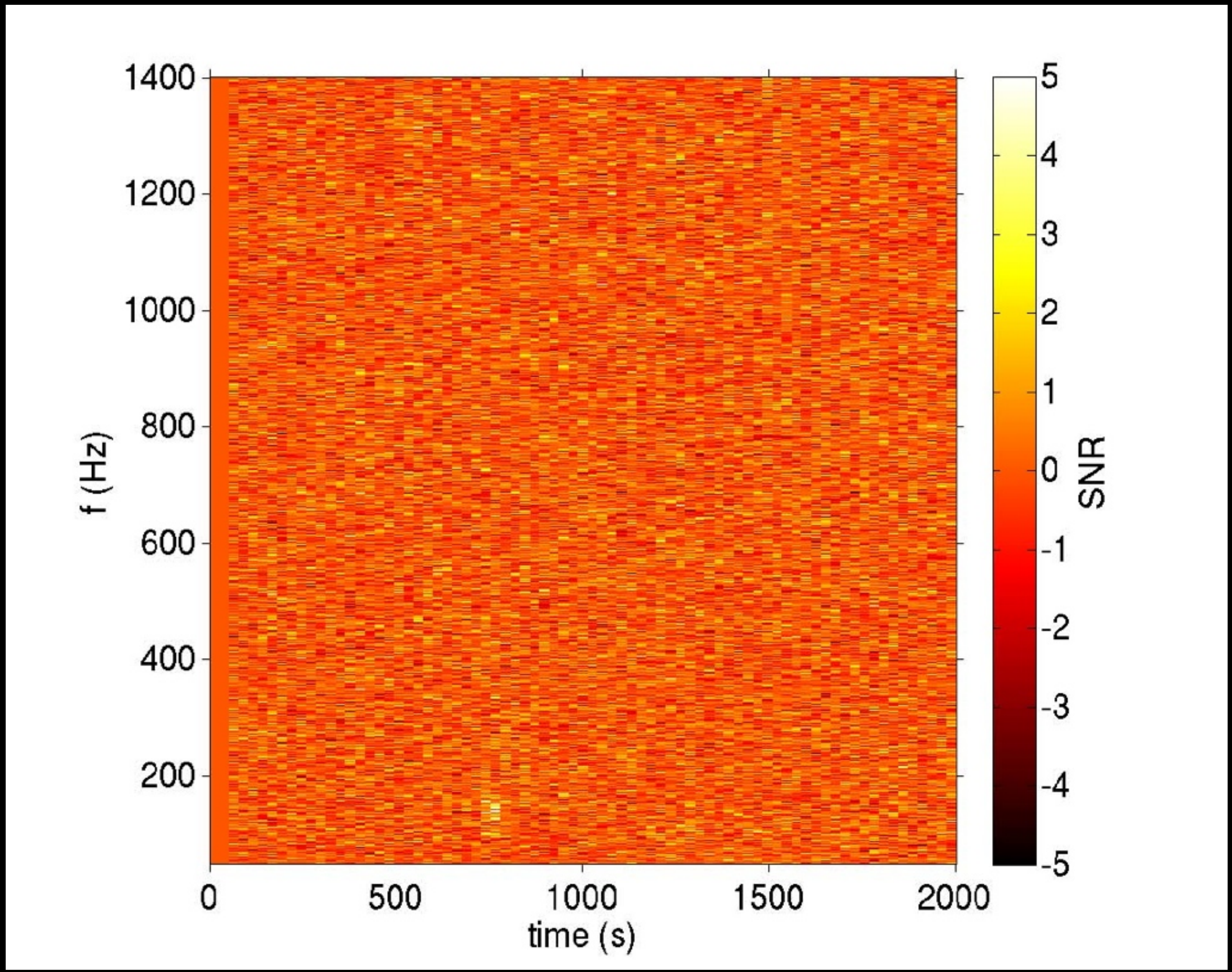
- Stochastic Transient Analysis Multi-detector Pipeline
- Participants: N. Christensen, M. Coughlin, S. Dorsher, S. Giampanis, S. Kandhasamy, V. Mandic, A. Mytidis, C. Ott, T. Prestegard, P. Raffai, E. Thrane, B. Whiting, WGA.
- Astrophysical searches for longer-lived transients (seconds - weeks) in LIGO-Virgo data.
- Searches for transients with various degrees of prior knowledge - serendipitous to targeted.
- Also useful for analyzing noise monitoring channels.

Tools

- Cross-power data
  - used in stochastic analysis - data and expertise available.
- Box search, radon transform, locust search:
  - used in unmodeled burst searches - codes and expertise available.
- Hough search:
  - used for pulsar and burst searches - codes and expertise available.

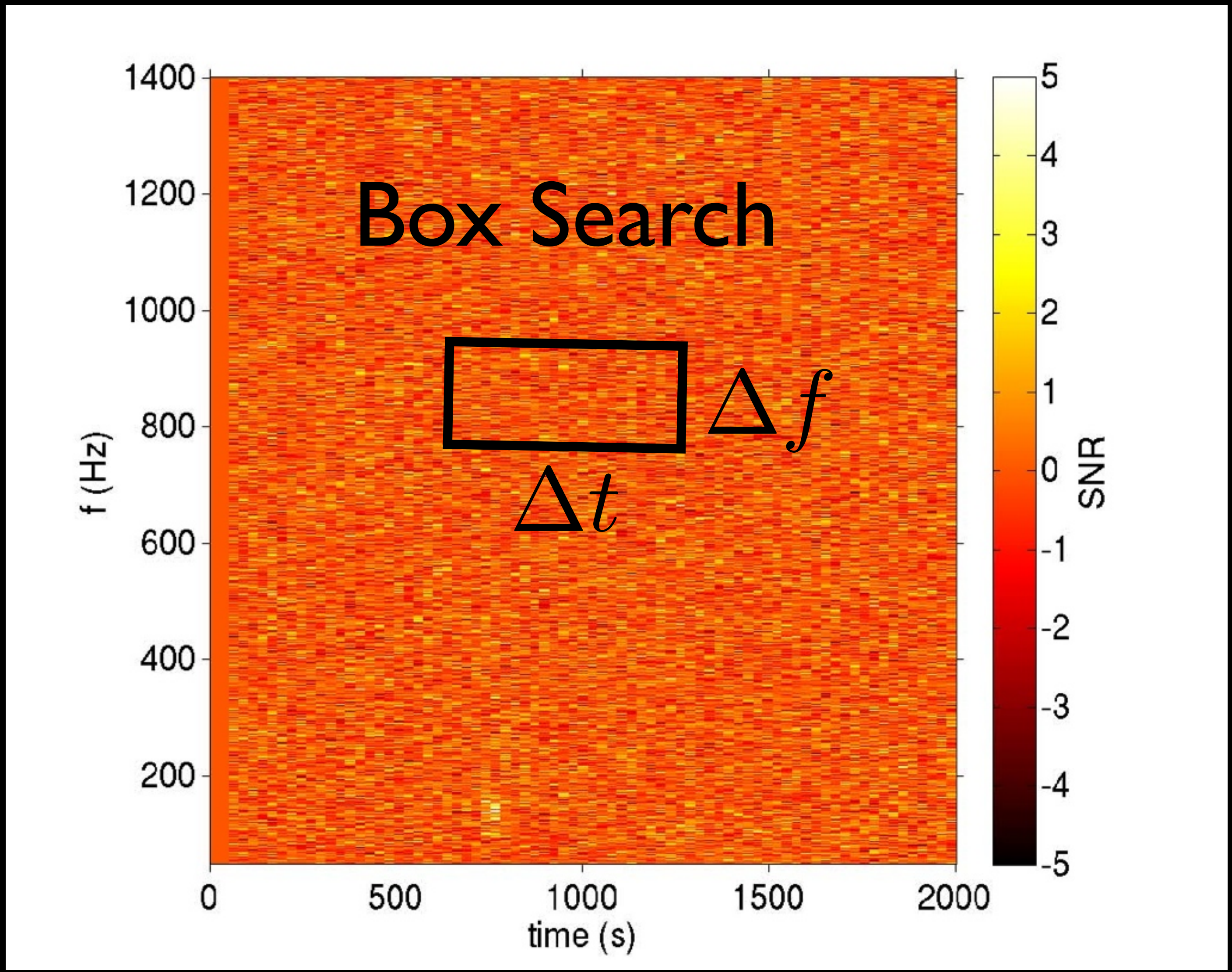
Maps

- We use a TF representation of cross-power to make projection onto signal space easier.



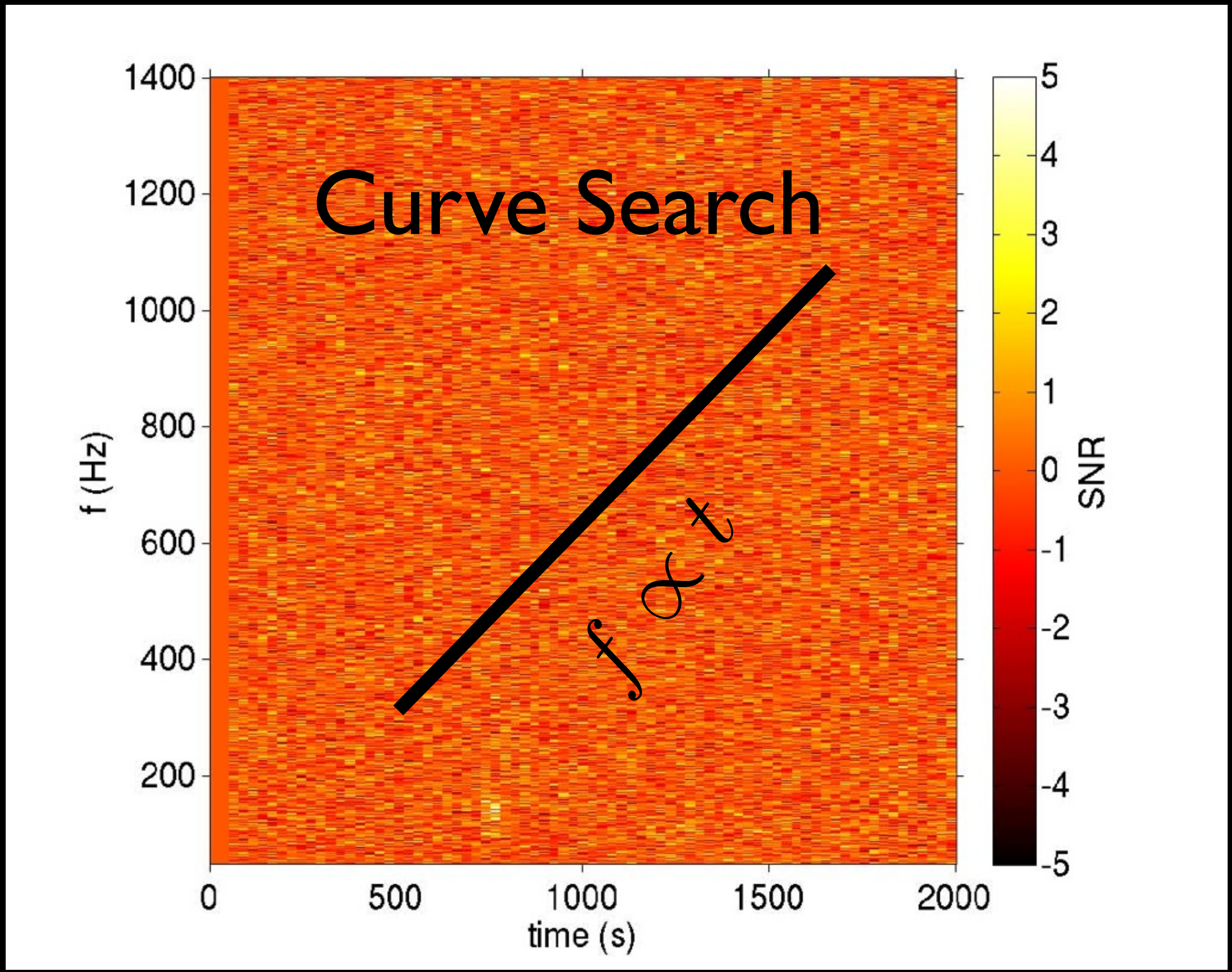
Maps

- We use a TF representation of cross-power to make projection onto signal space easier.



Maps

- We use a TF representation of cross-power to make projection onto signal space easier.



Activity

- Astrophysical sources: Ott, Giampanis, Thrane
- Statistics of cross-power: Mytidis, Whiting
- Efficiency of cross-power: Giampanis, Thrane, WGA
- Searches for astrophysical signals: Dorsher, Kandhasamy, Mandic, Thrane, Christensen, Coughlin, Dorsher, Giampanis, Prestegard, Raffai
- Noise studies: Christensen, Coughlin, Thrane
- Writing methods paper: all

Conclusions

- All "four types" of LIGO-Virgo data analysis have a lot in common:
  - use the same data
  - based on likelihood
  - the only difference is signal space we project on.
- Expertise and tools are portable across many analyses.
- STAMP is proving to be an example of fruitful interactions between analysts from different camps.

Sources

- core-collpase SN / long GRBs: protoneutron star turbulence, rotational instabilities (bar modes, r-modes), accretion disk instabilities (excitations, disintegration)
- binary inspirals: remnant protoneutron star turbulence, rotational instabilities in remnant, highly-eccentric binary black holes
- isolated neutron stars: pulsar glitches, quasi-periodic oscillations after SGR flares
- other?