Data Analysis Techniques for LIGO

Laura Cadonati, M.I.T.

Trento, March 1-2, 2007

Lesson Plan

Today:

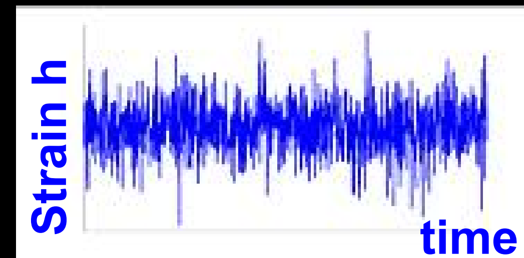
1. Introducing the problem: GW and LIGO
2. Search for Continuous Waves
3. Search for Stochastic Background

Tomorrow:

4. Search for Binary Inspirals
5. Search for Bursts
6. Network Analysis

Detecting Stochastic GWB

What does the signal look like?



Not very different from the detector noise!
Waveform is unknown, but it is always there.

How do we quantify it?

$$\Omega_{GW}(f) = \frac{1}{\rho_c} \frac{d\rho_{GW}(f)}{d\ln f}$$

$$\int_0^\infty (1/f)\,\Omega_{GW}(f)df = \frac{\rho_{GW}}{\rho_{critical}}$$

$$\rho_{GW} = \frac{c^2}{32\pi G} < \dot{h}_{ab}\,\dot{h}^{ab} >$$

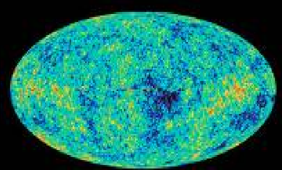
Gravitational wave energy density per unit logarithmic frequency, divided by the critical energy density to close the universe.

How do we look for it?

Template-less search based on cross-correlation of data from pairs of detectors

Isotropic Stochastic Gravitational Wave Background

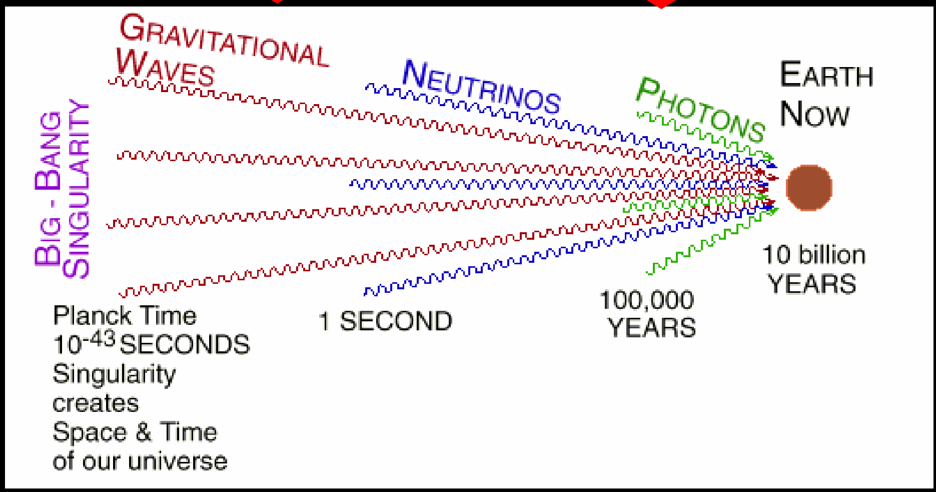
Cosmological: Big Bang



NASA, WMAP

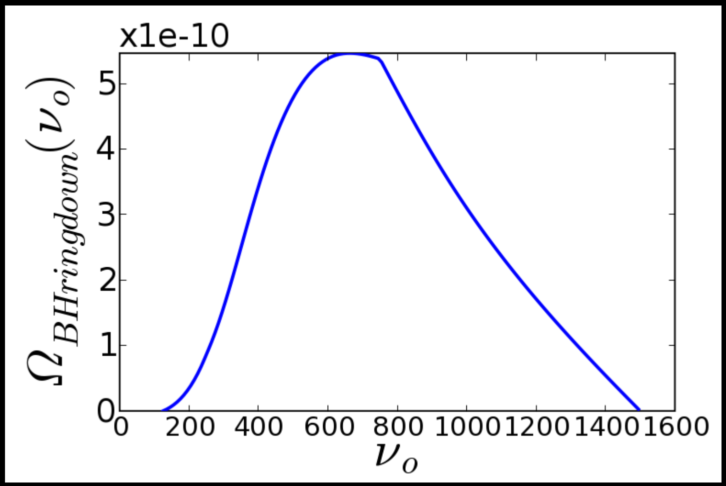
GWs in the LIGO frequency band were produced 10^{-22}s after the Big Bang

Cosmic Microwave Background (10^{+12}s)



Astrophysical: random superposition of unresolved and independent events

e.g.: black hole mergers, binary neutron star inspirals, supernovae



GW spectrum due to ringdowns of 40-80 $M_\odot$ black holes out to z=5 (Regimbau & Fotopoulos)

Stochastic GWB

The strain in an interferometer (IFO) due to a GW signal depends on the space-time metric and on the detector geometry:

$$h(t) \equiv h_{ab}(t, \vec{x}_0) \frac{1}{2} (\hat{X}^a \hat{X}^b - \hat{Y}^a \hat{Y}^b)$$

perturbations of the space-time metric

IFO vertex coordinates

Unit vectors in the direction of the IFO arms

$S_{gw}(f)$ = one-sided power spectrum of the gravitational strain

$$\lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} dt |h(t)|^2 = \int_0^\infty df S_{gw}(f)$$

$$S_{gw}(f) = \frac{3 H_0^2}{10 \pi^2} f^{-3} \Omega_{gw}(f)$$

H_0=Hubble expansion rate

Allen, Romano PRD 59, 102001 (1999)

Strain scale:

$$h(f) = 6.3 \times 10^{-22} \sqrt{\Omega_{GW}(f)} \left(\frac{100 \text{ Hz}}{f} \right)^{3/2} \text{Hz}^{-1/2}$$

Detection Statistics

Output of detector *i*: $s_i(t) \equiv h_i(t) + n_i(t)$

Cross-correlation of the output of two detectors takes advantage of the fact that sources of noise in each detector are independent.

$$Y \equiv \int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{T/2} dt_2 s_1(t_1) Q(t_1 - t_2) s_2(t_2)$$

T=measurement time

$Q(t_1-t_2)$=(real) filter function, chosen to maximize the SNR of Y

Optimal choice for Q falls off rapidly for $|t_1-t_2|$>>D/c=10ms (Livingston-Hanford)

$\Rightarrow$ we can extend the integral to infinity and express Y in the frequency domain:

$$Y \approx \int_{-\infty}^{\infty} df \int_{-\infty}^{\infty} df' \, \delta_T(f-f') \tilde{s}_1^*(f) \tilde{Q}(f') \tilde{s}_2(f')$$

$$\tilde{a}(f) \equiv \int_{-\infty}^{\infty} dt e^{-i2\pi f t} a(t)$$

Detection Statistics

$$Y = \int_{-\infty}^{+\infty} df\; \tilde{s}_1^*(f)\; \tilde{s}_2(f)\; \tilde{Q}(f)$$

Assuming the noise is:
- Stationary over the measurement time T
- Gaussian
- Uncorrelated between different detectors
- Uncorrelated with the stochastic GW signal

The expected value of the cross correlation only depends on the stochastic signal

$$\mu_Y \equiv \langle Y \rangle = \frac{T}{2} \int_{-\infty}^{\infty} df\, \gamma(|f|) S_{gw}(|f|) \tilde{Q}(f)$$

The variance is dominated by noise

$$\sigma_Y^2 \equiv \langle (Y - \langle Y \rangle)^2 \rangle \approx \frac{T}{4} \int_{-\infty}^{\infty} df P_1(|f|) |\tilde{Q}(f)|^2 P_2(|f|)$$

SNR = μ_Y / σ_Y is maximized when: $$\tilde{Q}(f) \propto \frac{\gamma(|f|) S_{gw}(|f|)}{P_1(|f|) P_2(|f|)}$$

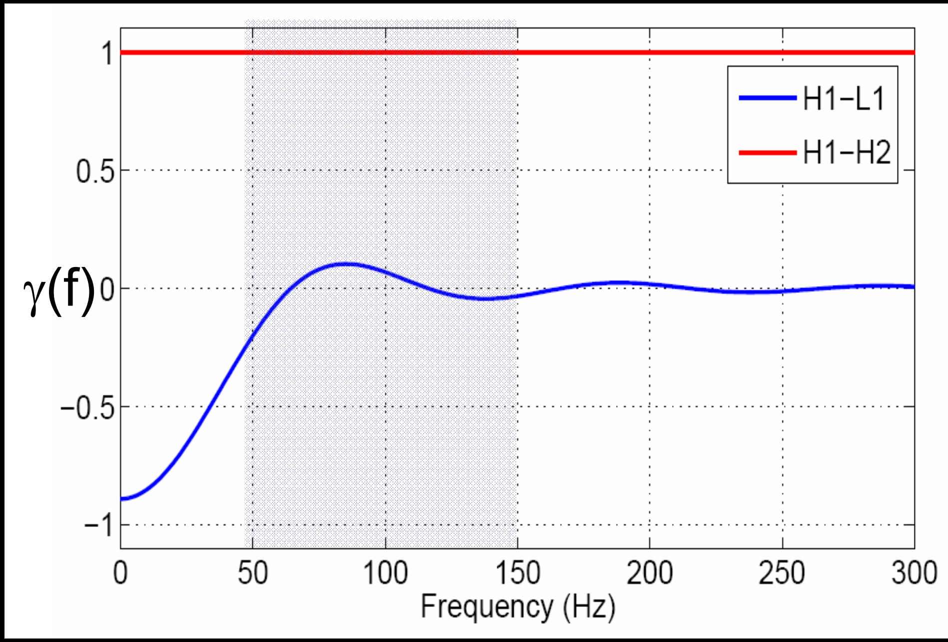
Detection Statistics

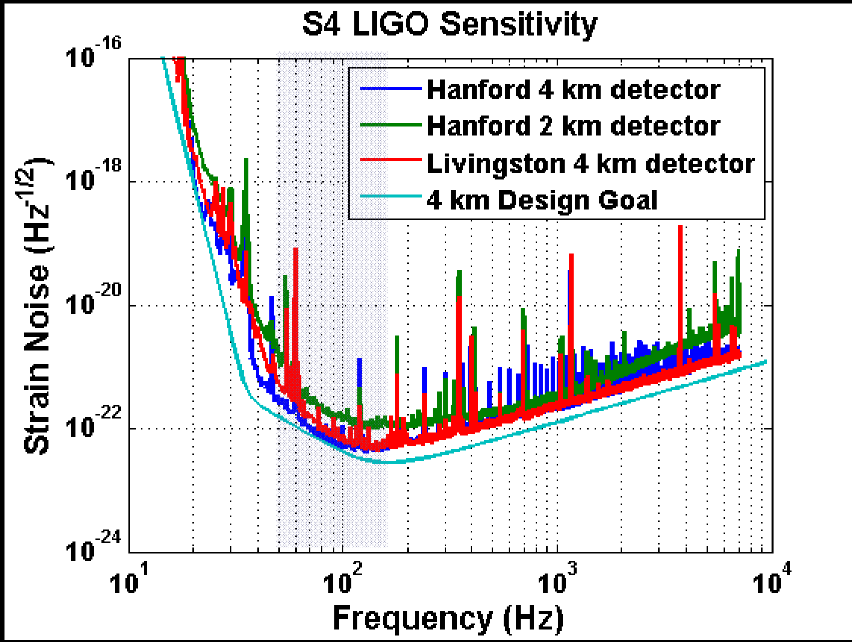
Optimal filter for all-sky search :

$$\tilde{Q}(f) = \frac{1}{N} \frac{\gamma(f)\, \Omega_t(f)}{f^3\, P_1(f)\, P_2(f)}$$

"Overlap Reduction Function"
(determined by network geometry)

One-sided detector noise spectra





Detection Statistics

Optimal filter for all-sky search : $\tilde{Q}(f) = \frac{1}{N} \frac{\gamma(f)\, \Omega_t(f)}{f^3\, P_1(f)\, P_2(f)}$

We need an hypothesis on how Ω depends on frequency: power law

$$\Omega_t(f) = \Omega_\alpha (f/100\ \mathrm{Hz})^\alpha$$

⇒ Choose N such that: $< Y >= \Omega_\alpha T$

For instance: for α=0 and $\Omega = \Omega_0$ =constant $\tilde{Q}(f) = \mathcal{N} \frac{\gamma(|f|)}{|f|^3 P_1(|f|) P_2(|f|)}$

$$\mathcal{N} = \frac{20\pi^2}{3H_0^2} \left[\int_{-\infty}^{\infty} df \frac{\gamma^2(|f|)}{f^6 P_1(|f|) P_2(|f|)} \right]^{-1},$$

$$\sigma_Y^2 \approx T \left(\frac{10\pi^2}{3H_0^2} \right)^2 \left[\int_{-\infty}^{\infty} df \frac{\gamma^2(|f|)}{f^6 P_1(|f|) P_2(|f|)} \right]^{-1}.$$

$$\langle \rho_Y \rangle = \frac{\mu_Y}{\sigma_Y}$$

SNR, prop to Ω and to the square root of the observation time

$$\approx \frac{3H_0^2}{10\pi^2} \Omega_0 \sqrt{T} \left[\int_{-\infty}^{\infty} df \frac{\gamma^2(|f|)}{f^6 P_1(|f|) P_2(|f|)} \right]^{1/2}$$

Implementation in LIGO run S4

Ref: astro-ph/0608606v2 to appear in Ap.J.

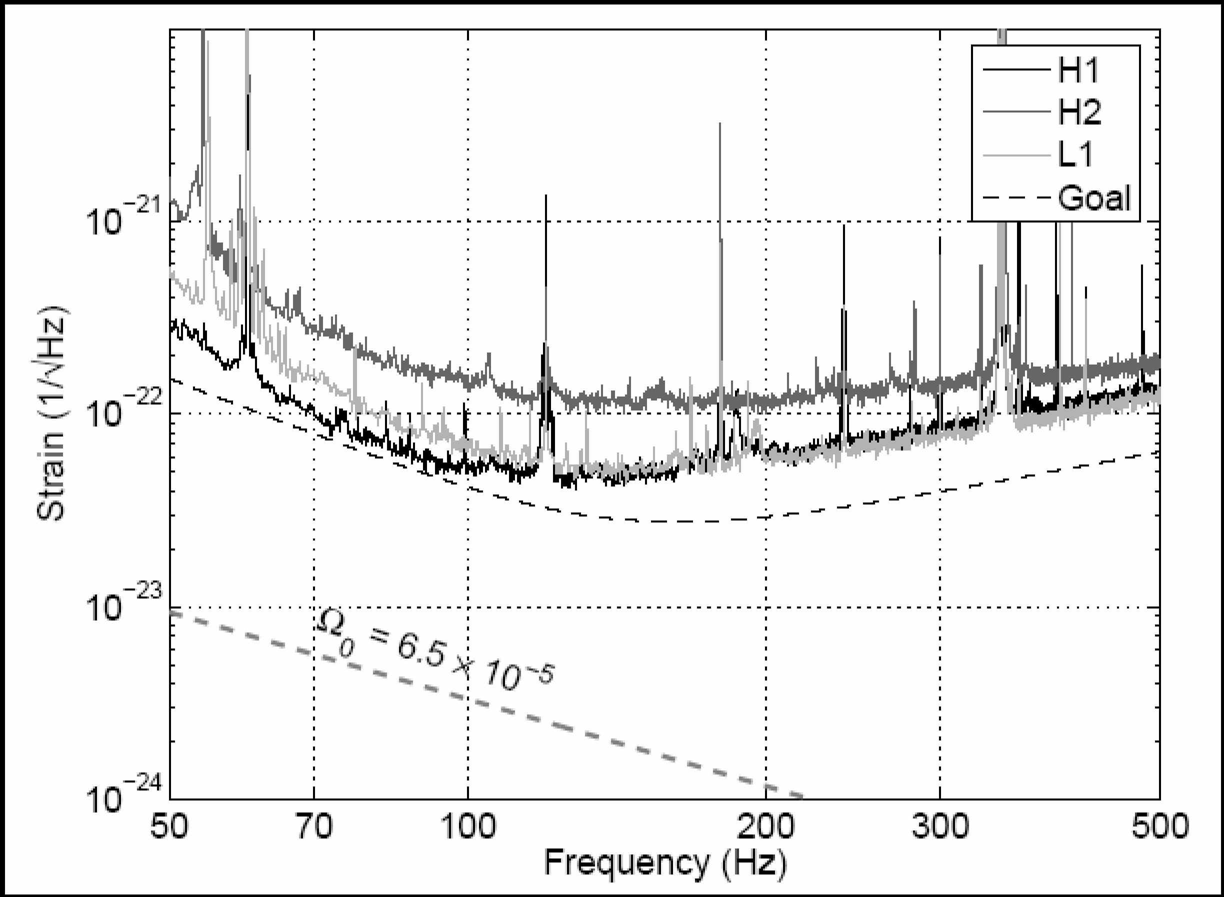


Fig. 1.— Typical strain amplitude spectra of LIGO interferometers during the science run S4 (solid curves top-to-bottom at 70 Hz: H2, L1, H1). The black dashed curve is the LIGO sensitivity goal. The gray dashed curve is the strain amplitude spectrum corresponding to the limit presented in this paper for the frequency-independent GW spectrum $\Omega_0 < 6.5 \times 10^{-5}$.

Implementation Details

$$Y = \int_{-\infty}^{+\infty} df\ \tilde{s}_1^*(f)\ \tilde{s}_2(f)\ \tilde{Q}(f)$$

$$\sigma_Y^2 \approx \frac{T}{2} \int_0^{+\infty} df\ P_1(f)\ P_2(f)\ |\ \tilde{Q}(f)\ |^2$$

$$\tilde{Q}(f) = \frac{1}{N} \frac{\gamma(f)\ \Omega_t(f)}{f^3\ P_1(f)\ P_2(f)}$$

Divide data into intervals
Compute Y_i and σ_i for each interval
Take the weighed average.

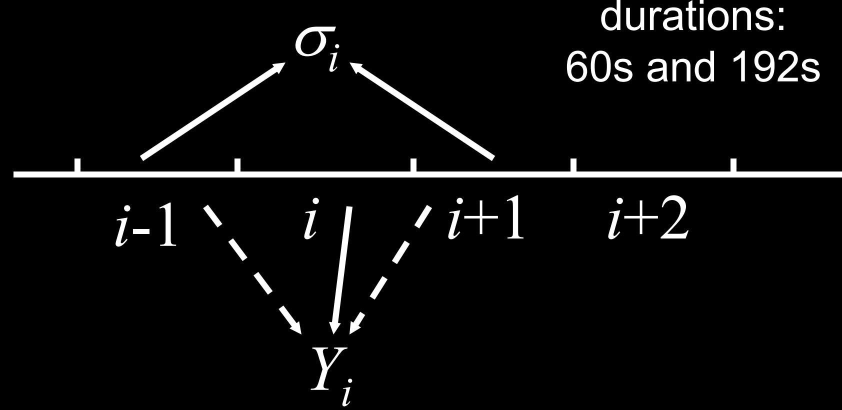
$$Y_{\text{opt}} = \frac{\sum_i \sigma_i^{-2} Y_i}{\sum_i \sigma_i^{-2}}$$

$$\sigma_{\text{opt}}^{-2} = \sum_i \sigma_i^{-2}$$

$$\Omega_\alpha = Y_{\text{opt}} / T$$
$$\hat{\sigma}_\Omega = \sigma_{\text{opt}} / T$$

Used two interval durations:
60s and 192s

The PSDs $P_1(f)$ and $P_2(f)$ for the filter and the variance are computed from near neighbor intervals, to avoid a bias in the point estimate due to non-zero covariance between the cross-power Y and power spectra estimated from the same data



Data manipulation: down-sample to 1024 Hz; High-pass filter (40 Hz cutoff), FFT, calibrate
50% overlapping Hann windows; overlap in order to recover the SNR loss due to windowing.

Technical Challenges

- We are digging deep into instrumental noise, looking for small correlations.
- We need to be mindful of possible non-GW correlations
  - common environment (two Hanford detectors)
  - common equipment (could affect any detector pair!)

- In the following slides, we show the techniques used to “clean the data” in the LIGO stochastic S4 analysis, focusing on the H1-L1 interferometer pair (Hanford-4km, Livingston-4km)

Plots and discussions are extracted from astro-ph/0608606v2, to appear in Ap.J.

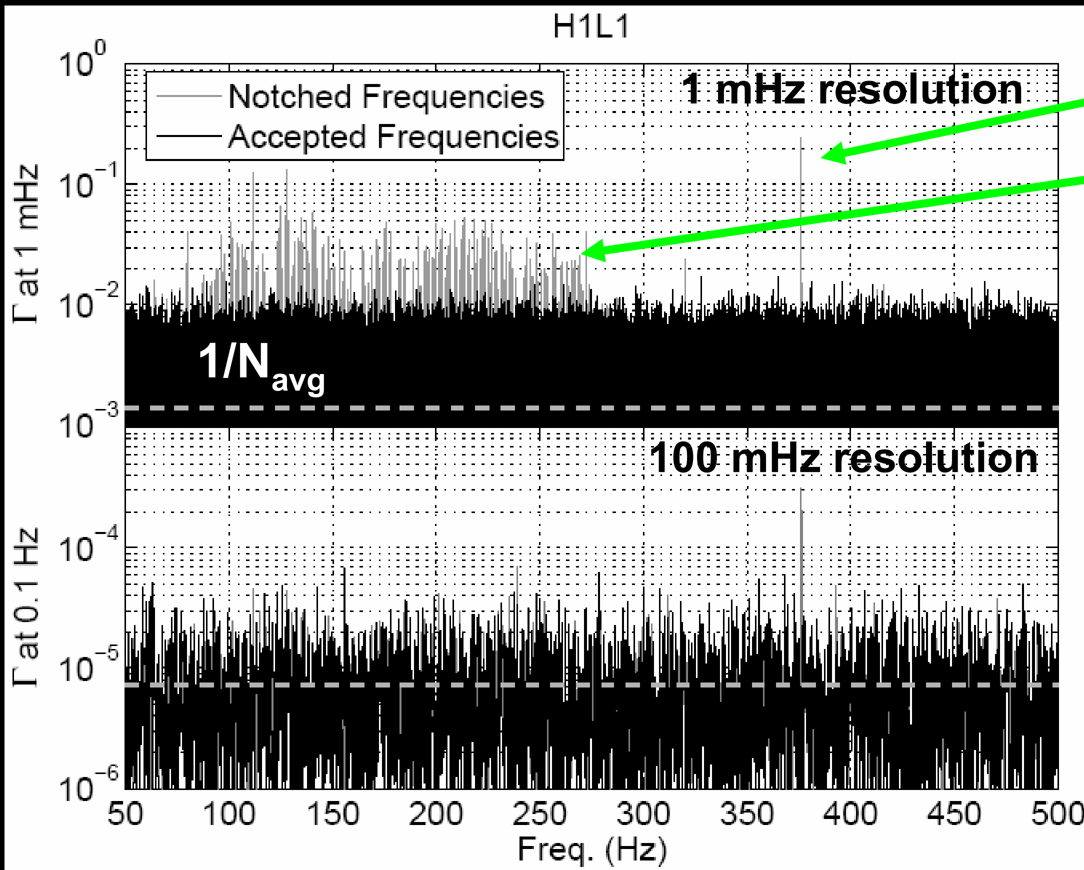
Correlated Instrumental Lines

We look for them by calculating the coherence over the entire data period

$$\Gamma(f) = \frac{|s_1^*(f)s_2(f)|^2}{P_1(f)P_2(f)}$$

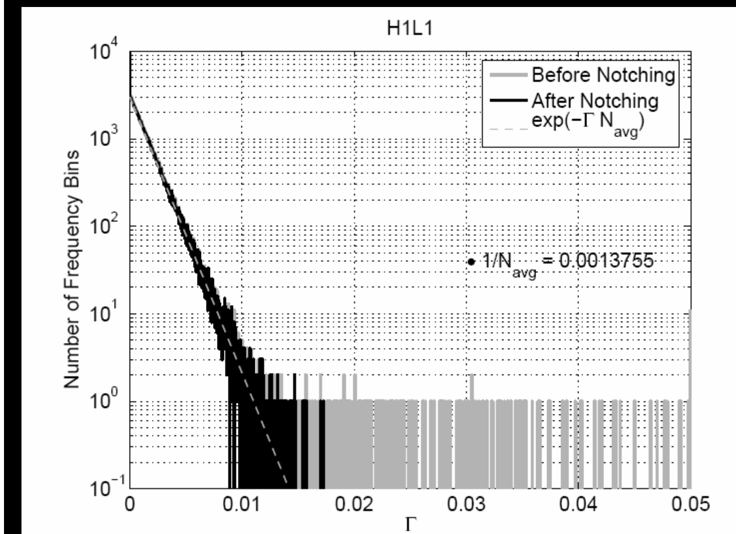
Cross-spectral density (CSD)

Power Spectral density (PSD)



Simulated pulsar line

1 Hz harmonics, from the GPS synchronization signal (improved in S5)



60 sec vs 192 sec analysis: blind or not blind?

S4 data was analyzed twice:

- T = 60 sec, 1/4Hz frequency resolution
  - Better sensitivity to noise transient
  - Better suited for definition of data stationarity cuts
- T=192 sec, 1/32 Hz frequency resolution
  - Higher frequency resolution
  - Better suited for the removal of sharp lines (1Hz and 60Hz harmonics, simulated pulsars, 3% bandwidth loss)

Blind: tuning with a 0.1s shift between data streams

Not blind: we already had the 60sec results

The 60s analysis was tuned blindly and completed before discovery of the sharp lines in the 1ms coherence.

Once these were recognized as an instrumental effect, it was decided, on scientific ground, to remove them and repeat the analysis with T=192sec.

The 192sec analysis is not blind, but more correct (and more conservative).

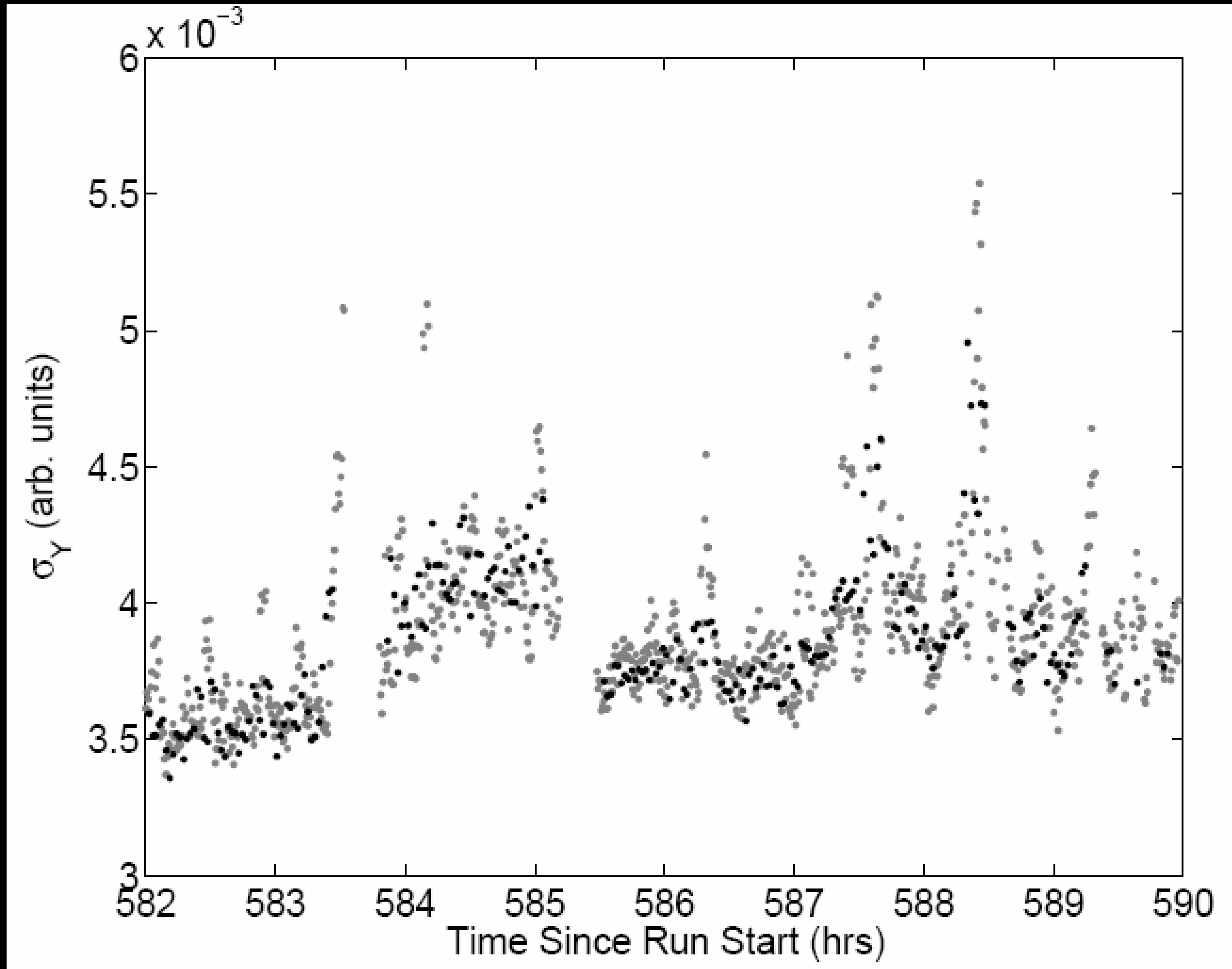


Fig. 7.— Trend of σ_Y for the 60-sec analysis (gray) and 192-sec analysis (black) over a short period of time. The two bands were scaled to overlap. Note that the gray band is wider, indicating that the 60-sec analysis is more sensitive to noise variations.

Data Quality Cuts

1. Removal of known features:

   30 sec before lock loss; saturations.

2. Glitch cut:

   reject intervals known to have large "glitches" in one interferometer, identified as discontinuities in the PSD trends over the S4 run

3. Large σ cut:

   Reject intervals with anomalously large value of σ.

   These would not contribute much to the final result due to the $1/\sigma^2$ weighting, but are removed anyways.

   Overlap with the glitch cut.

4. $\Delta\sigma$ cut:

   Remove periods where the noise changes quickly, by removing intervals with:

   $$\Delta\sigma = |\sigma_{Y_I} - \sigma'_{Y_I}|/\sigma_{Y_I} > \zeta$$

   Where σ is the theoretical variance, computed on neighbor intervals, and σ' is computed on the interval itself. The threshold ζ was chosen to compromise between data loss and data quality.

Data Quality Cuts

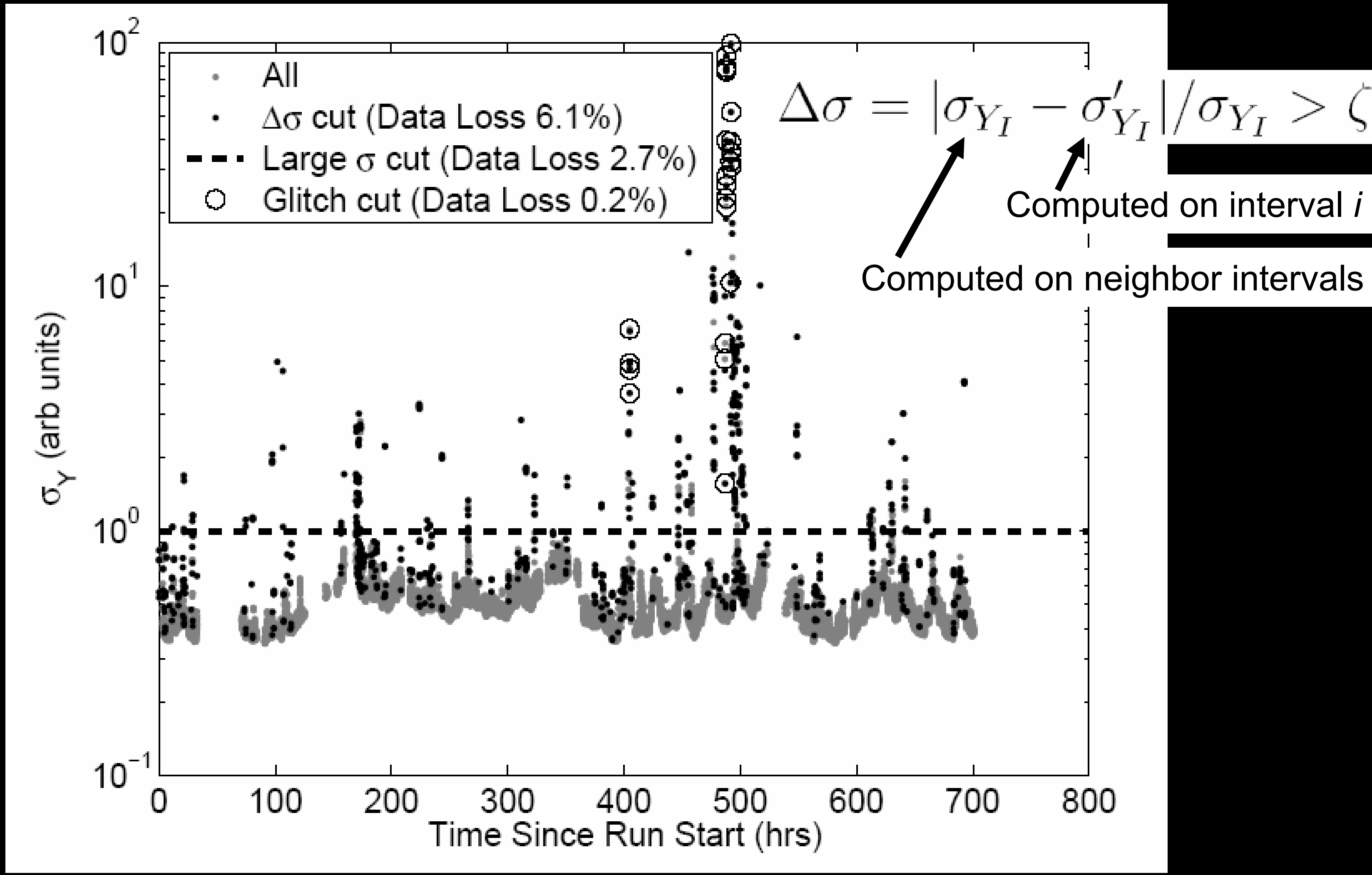


Fig. 5.— Trend of σ_{Y_I} over the whole S4 run for the 192-sec intervals of H1L1 pair. The dashed line denotes the large σ cut: segments lying above this line are removed from analysis.

Residuals

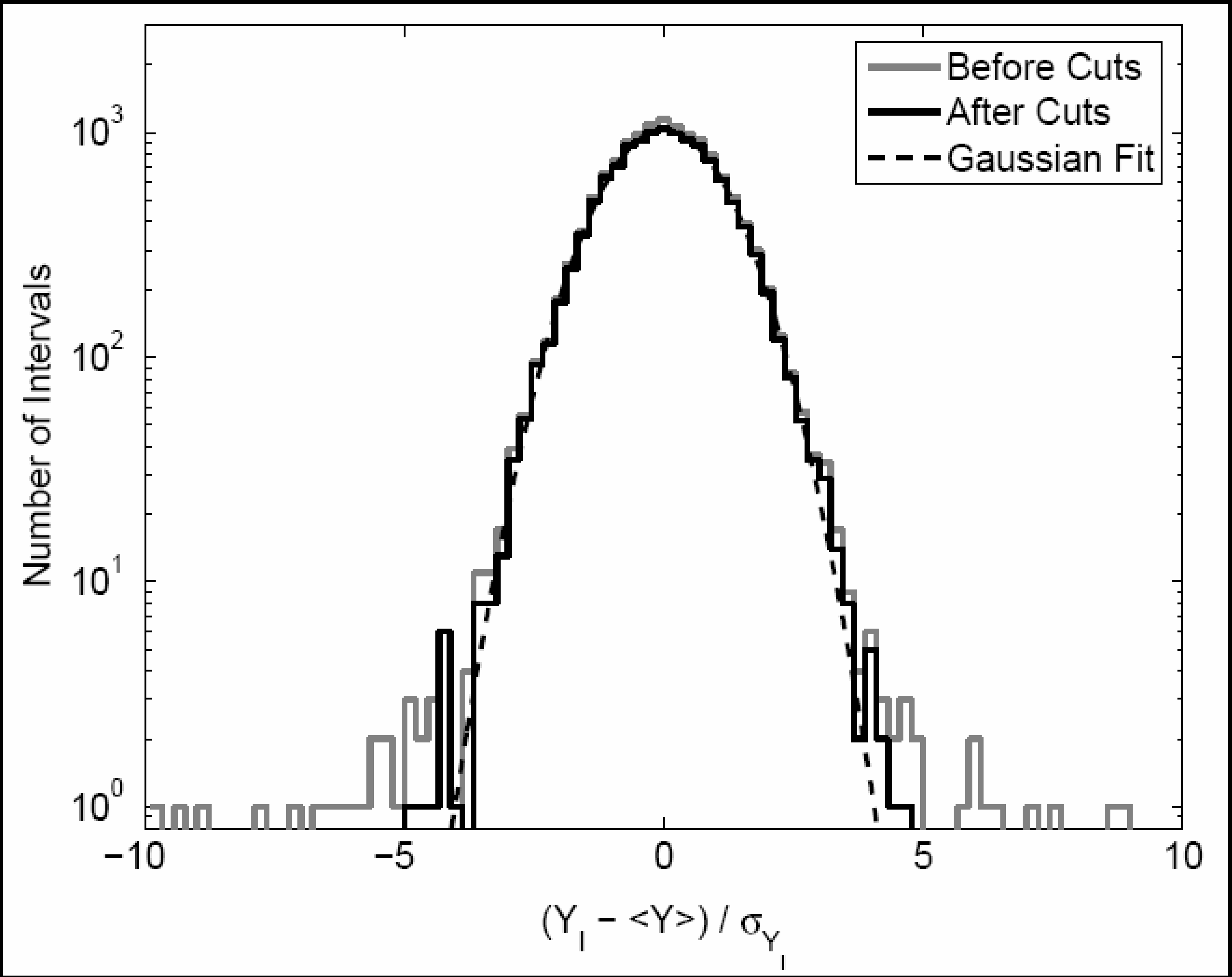
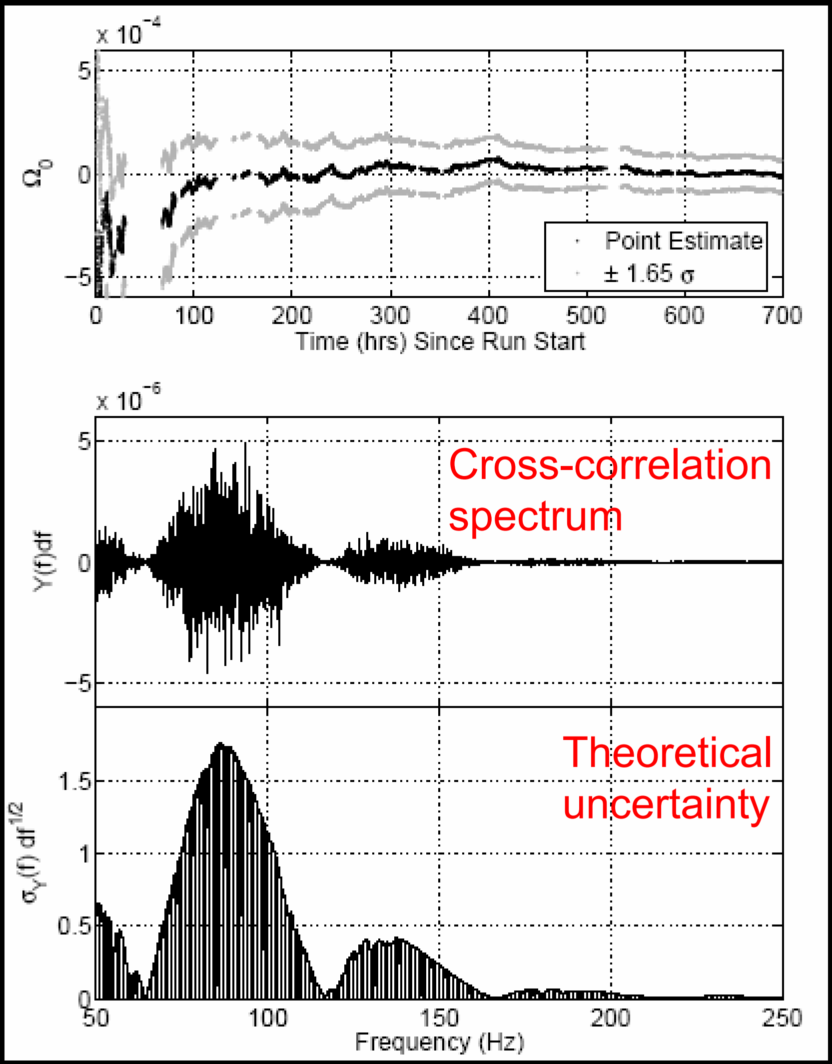


Fig. 6.— Distribution of residuals for the H1L1 pair with 192-sec segments: all data are shown in gray, data that passes data quality cuts are shown in black, and the Gaussian fit to the black histogram is shown as a dashed curve.

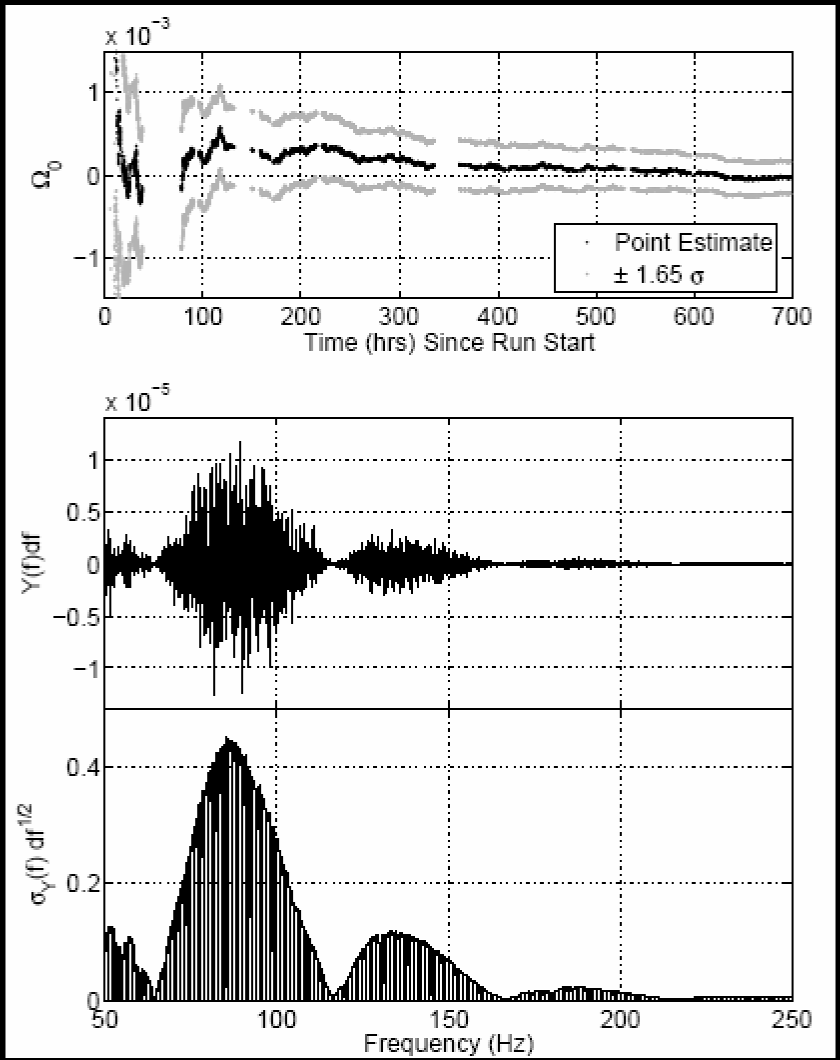
H1L1 and H2L1

Frequency-independent GW spectrum: $\Omega_{GW}=\Omega_0$

H1L1 pair



H2L1 pair



Combined Results

- Cross-correlate Hanford-Livingston
  - Hanford 4km – Livingston
  - Hanford 2km – Livingston
  - Bin-by-bin weighted average of two cross-correlations.
  - Do not cross-correlate the Hanford detectors.

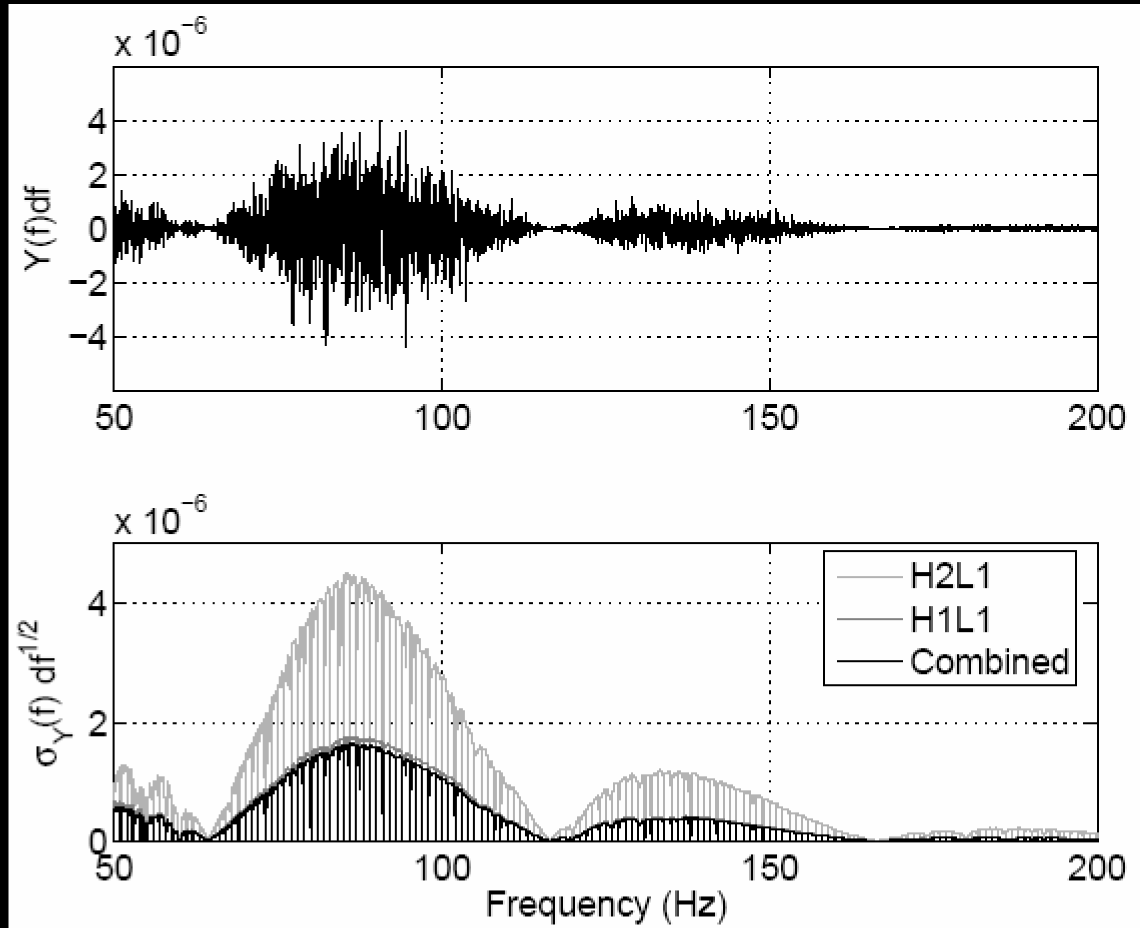


Fig. 10.— Combined H1L1 + H2L1 result, 192-sec analysis with $\zeta = 0.3$. Top: Combined cross-correlation spectrum. Bottom: Theoretical uncertainty $\sigma_Y(f)$.

Setting an Upper Limit

Uncertainties:

- Dominant systematics: the amplitude calibration uncertainty:

  5% (L1) and 8% (H1 and H2), frequency-independent.
- The uncertainty in the phase of the interferometer strain response is negligible compared to the magnitude and statistical uncertainties.
- negligible effect of timing errors=0.4 μs, 0.2% effect on point estimate.

Bayesian upper limit:

❖ Assume a Gaussian distribution for the amplitude calibration uncertainty (mean=1 and standard deviation=sqrt(0.052 + 0.082) = 0.093), and marginalize over it.

❖ Assume the prior distribution for Ω_0 to be the posterior distribution obtained in our previous analysis of the S3 data.

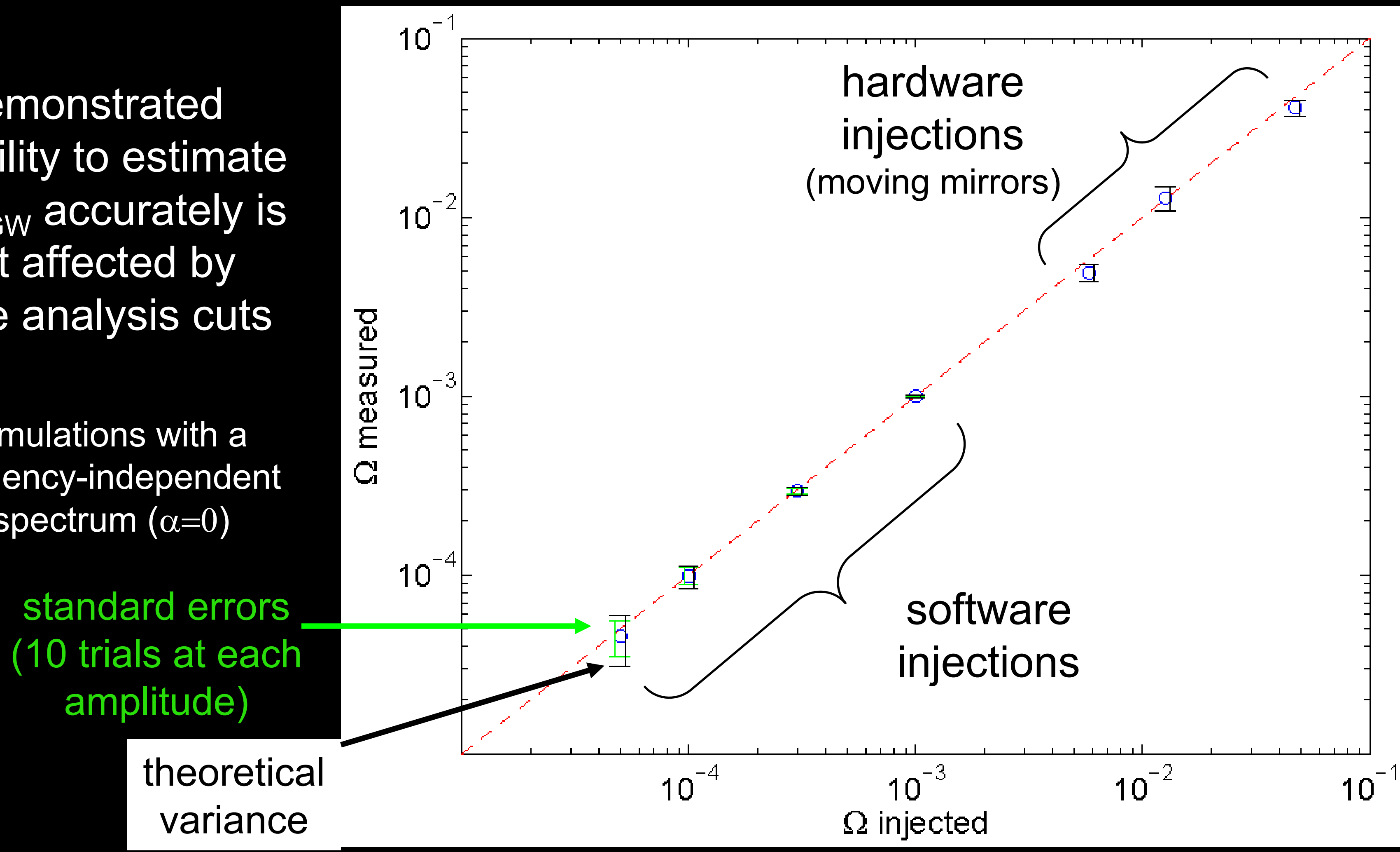
❖ The 90% upper limit is the value of Ω_0 for which 90% of the posterior distribution lies between 0 and the upper limit.

❖ This procedure yields the Bayesian 90% UL on $\Omega_0 < 6.5 \times 10^{-5}$.

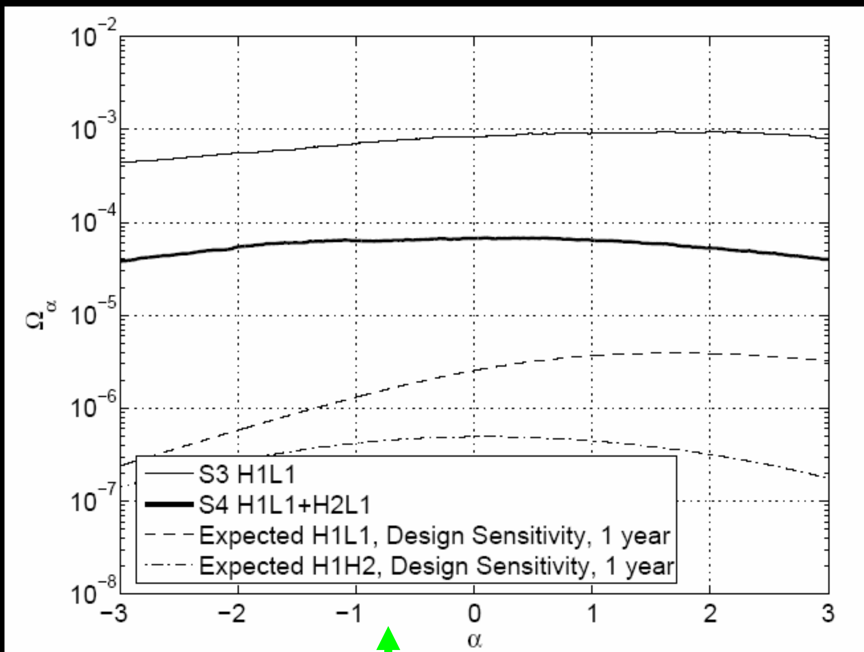
Signal Recovery

Demonstrated ability to estimate Ω_{GW} accurately is not affected by the analysis cuts

All simulations with a frequency-independent GW spectrum ($\alpha=0$)



Different Frequency Dependencies



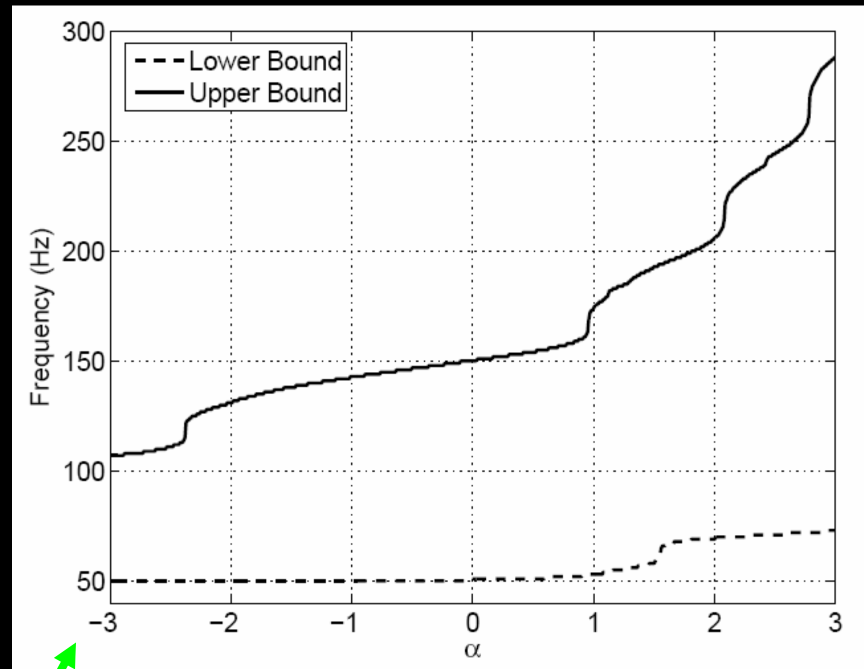


Fig. 11.— Top: 90% UL on Ω_α as a function of α for S3 H1L1 and S4 H1L1+H2L1 combined, and expected final sensitivities of LIGO H1L1 and H1H2 pairs, assuming LIGO design sensitivity and one year of exposure. Bottom: Frequency band containing 99% of the full sensitivity (as determined by the inverse variance) is plotted as a function of α for the S4 result.

Directional Search

- Stochastic GW Background due to Astrophysical Sources is not isotropic if dominated by nearby sources

  ➔Do a *Directional Stochastic Search*

- Source position information from
  - Signal time delay between different sites (sidereal time dependent)
  - Sidereal variation of the single detector acceptance

➔Time-Shift and Cross-Correlate!

Ref. 2006 Class. Quantum Grav. 23 S179-S185

Detection Strategy, Point Source

- Cross-correlation estimator

$$Y = \int_{-\infty}^{+\infty} df\ \tilde{s}_1^*(f)\ \tilde{s}_2(f)\ \tilde{Q}(f)$$

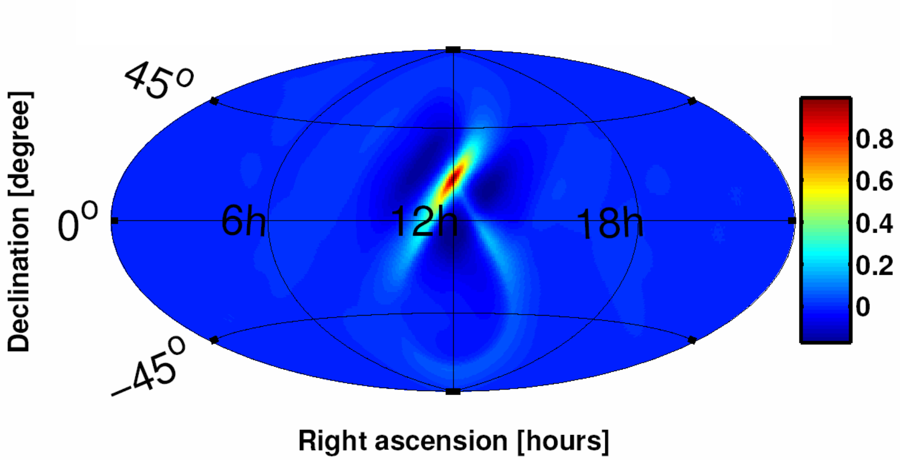
- Theoretical variance

$$\sigma_Y^2 \approx \frac{T}{2} \int_0^{+\infty} df\ P_1(f)\ P_2(f)\ |\ \tilde{Q}(f)\ |^2$$

- Optimal filter depends on sidereal time

$$\tilde{Q}(t,f) = \frac{1}{N} \frac{\gamma_{\text{point}}(t,f) H(f)}{P_1(f) P_2(f)}$$

Point Spread Function

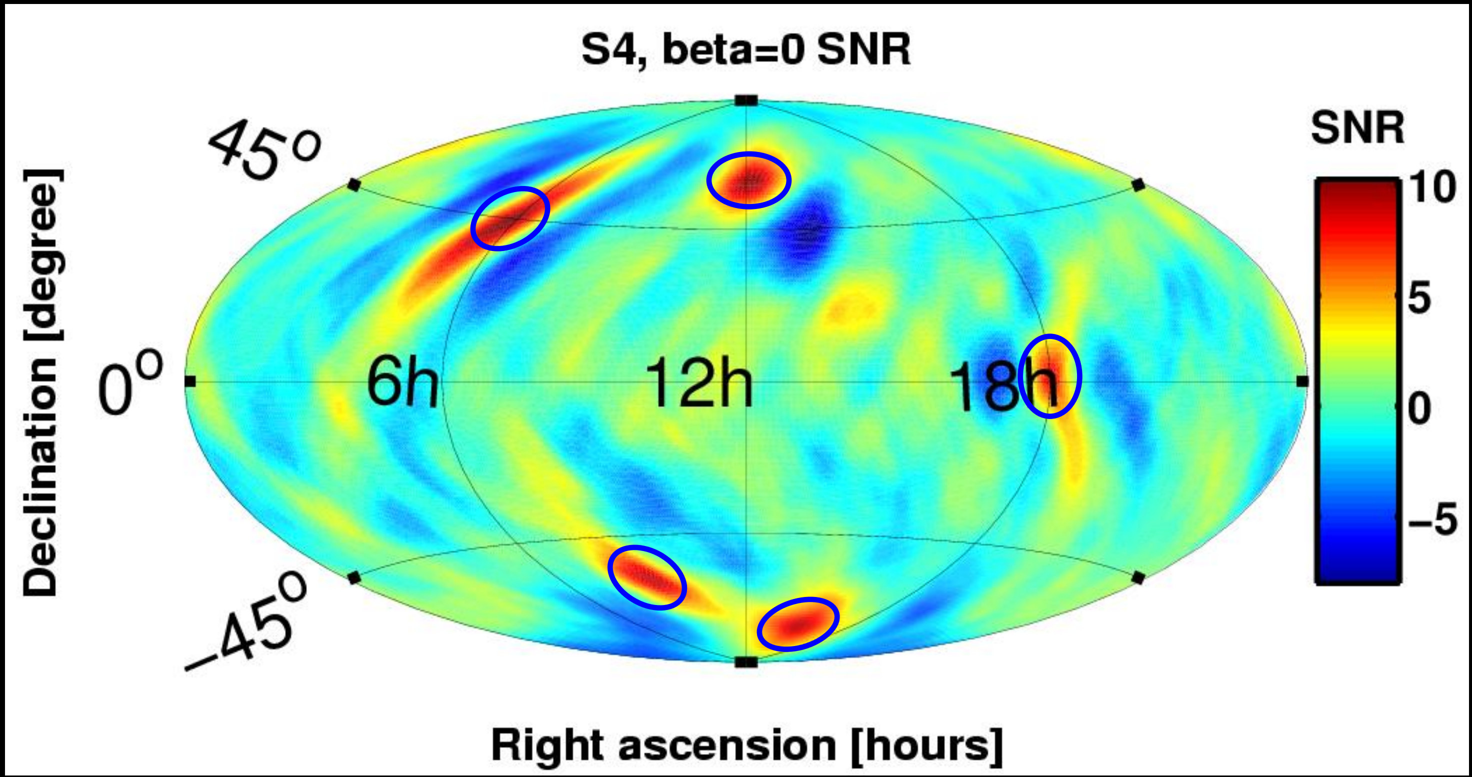


$$\gamma_{\text{point}}(t,f) = \sum_{A=+,\times} e^{i 2\pi f \hat{\Omega} \frac{\Delta \vec{x}(t)}{c}} F_{1,t}^A(\Omega) F_{2,t}^A(\Omega)$$

Strain Power: $H(f) = H_\beta (f / 100Hz)^\beta$

Choose *N* such that: $< Y >= H_\beta$

Simulated Point Sources

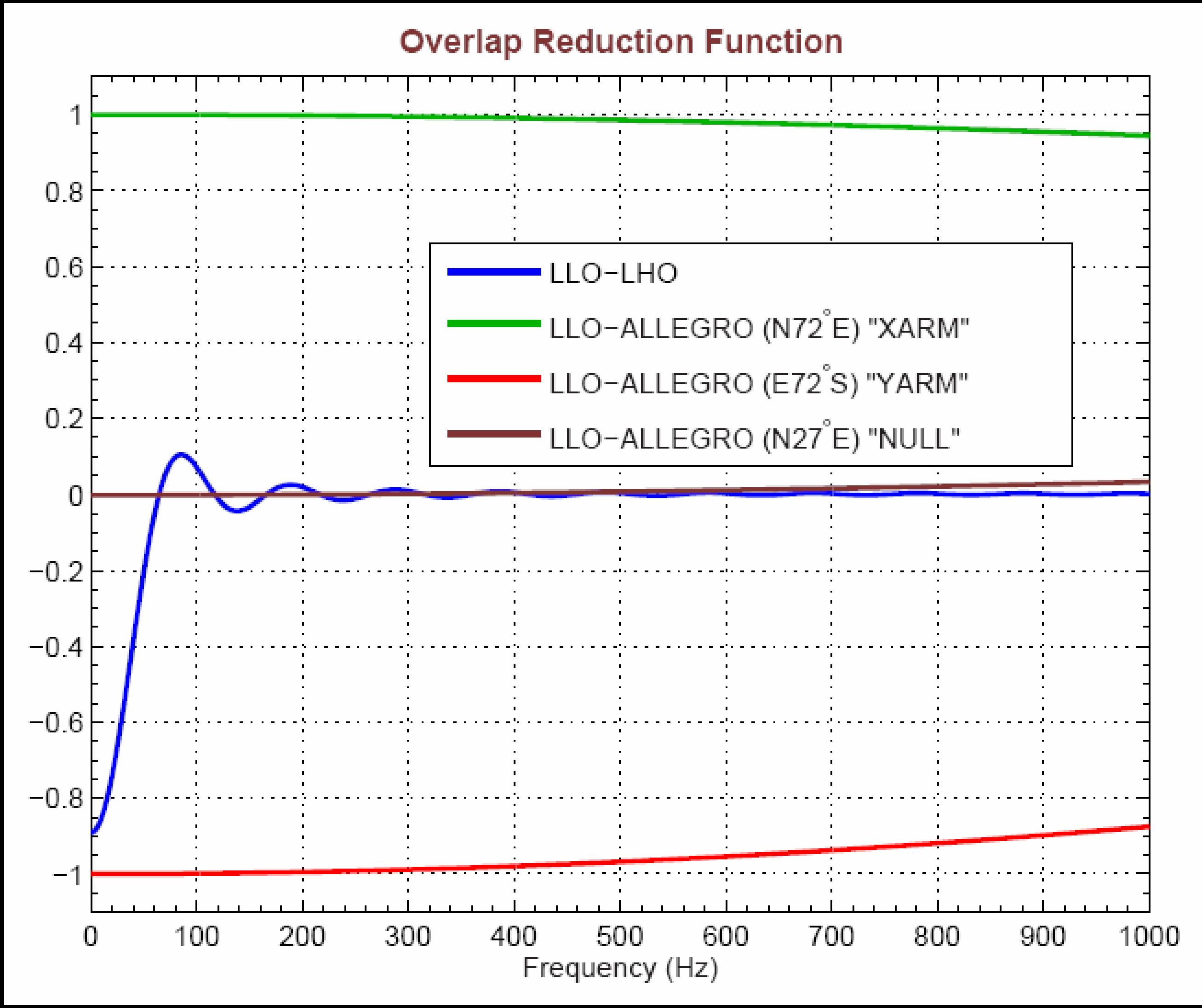


Five point sources with $H_{\beta=0}=10^{-47}\text{strain}^2\text{Hz}^{-1}$ were injected in software (blue circles). Analysing the same injection for an isotropic background only yields SNR 1.2

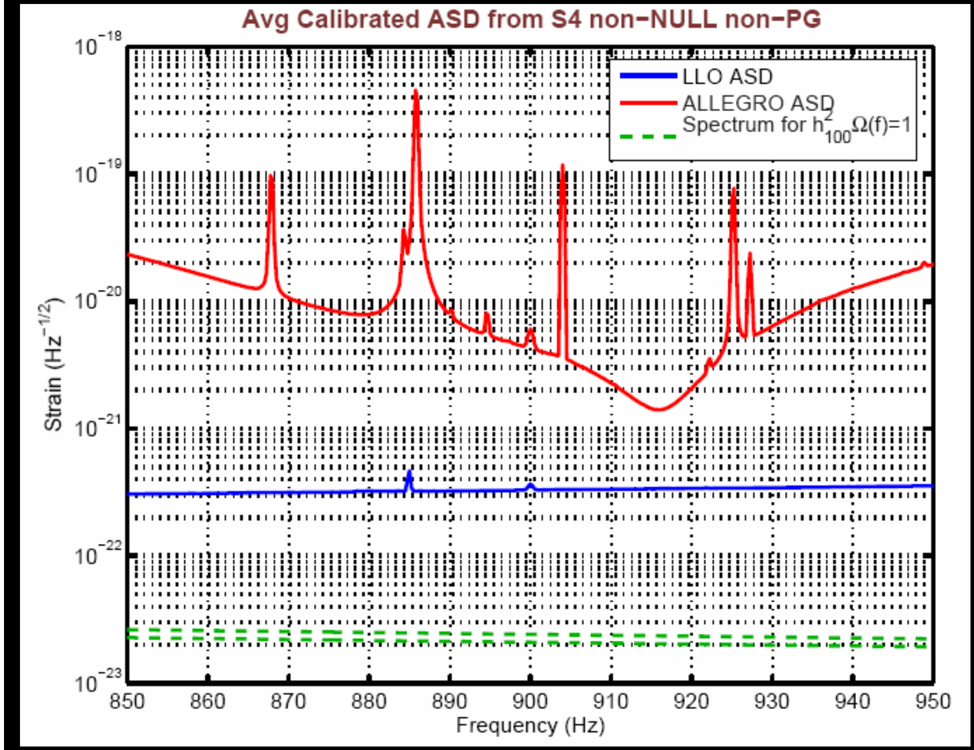
LIGO-Allegro

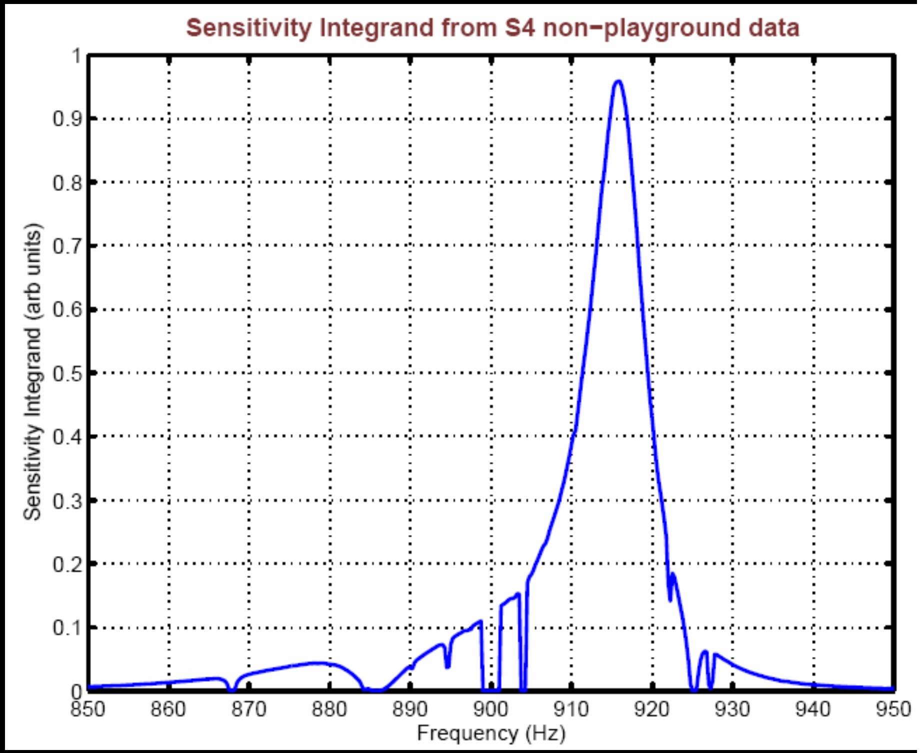
Ref. LIGO-G050633-00-Z – paper in preparation

Correlate LIGO with ALLEGRO resonant bar located within ~40 km or each other so delay time vs. point on sky not an issue



Different Spectra





Probes higher frequency band than IFO-IFO correlations:
~850Hz – 950Hz