Stochastic Group:

Case A. Isotropic search for a stochastic background

With S3, we begin to approach a terra incognita where a detection of a SGWB in the relatively narrow band 50 Hz–300 Hz is not ruled out by existing physics. While upper limits on $_{GW}(f)$ larger than unity (S1 and earlier) are obviously ruled out by cosmological observations, the uncertainties in the cosmological parameters estimated from direct measurements like WMAP[1] seem to allow room for additional unaccounted-for sources energy density such as gravitational waves on the order of $\Omega_{other} < \sim 10^{-2}$.

A SGWB can always be observed repeatedly. Consequently, only after repeated observation of the same phenomenon should LIGO be prepared to declare a detection.

Consistency checks to be made before declaring a detection will include

Time-Stationarity Checks

Are the point estimates for different epochs consistent? (This will also manifest itself as a cumulative signal-to-noise ratio which increases, on average, as $T_{obs}^{1/2}$.)

Does accrual of signal exhibit any diurnal variations? any sidereal variations? any seasonal variations? Note that while significant variation of the signal with time is inconsistent with a stationary isotropic background, a sidereal variation might be a signature of an anisotropic component (see case B below).

Frequency-Band Checks

While the spectrum ($_{GW}(f)$) of a stochastic background is a priori unknown, we can place some requirements on the spectrum that is seen:

- Is the $_{GW}(f)$ seen in each frequency band must be positive? The alternating sign of the overlap reduction function for non-colocated detector pairs makes this a non-trivial check, as it predicts a correlation or anti-correlation at different frequencies.
- Additionally the spectrum $_{GW}(f)$ should change believably from band to band.
- Is the same spectrum seen in different detector pairs?
- Is the same spectrum seen for different epochs of observation.

Cross-Pair Checks

Is the same effect seen in different, independent, pairs of interferometers? LIGO can rely on H1-L1 and H2-L1; as other international partners come on-line, their instruments can also be cross-correlated with LIGO interferometers. Of course, the overlap reduction function will impose a severe penalty for pairs separated by long baselines and which are not well aligned; nonetheless, these pairs of instruments should be examined.

This check will be initially problematic, as improving detector sensitivities will likely mean there is initially a period where only one pair of detector sites can detect a background, and other pairs can simply set a consistent upper limit. On the other hand, the H1-H2 pair, which is unlikely to make a conclusive detection because of the possibility of intra-site instrumental correlations, should be able to confirm an observation made with one or both LLO-LHO pairs.

Non-GW Constraint Checks

- Is the effect correlated with any of the environmental channels? Examples include accelerometers, magnetometers, acoustic sensors (for H1-H2). Over a transcontinental baseline, effort will be required to demonstrate conclusively that electromagnetic correlations in B(f) are not coupling into the test masses via their actuation systems. There have been a number of order-of-magnitude calculations based on estimating plausible B(f) and its gradients and how these would couple to unbalanced dipole or quadrupole moments of the magnet assemblies. Alternatively, electromagnetic effects could, in principle, could directly to drive electronics through EMI or RFI and thus impress correlated forces on the test masses.
- Is the null H1-H2 channel showing a measurement consistent with zero?
- Is the magnitude of the effect consistent with known cosmological or astrophysical limits or constraints?

Case B. Directed search for an astrophysical stochastic Background

Here, the plausible assumption is made that stochastic GW emission of astrophysical origin follows the matter distribution of the nearby universe. In this case, the detection technique is to track different patches of the sky during cross-correlation by implementing a directed optimal filter rather than a sky-integrated one. Now the detection statistic between two different patches of sky, one of which is expected to contain a signal.

While all the above tests still apply, here, the fact that it is possible to discern an effect by comparing two or more measurements involving different patches of the sky is important.

References

[1] D. N. Spergel et al., Ap. J. Supp. 148, 175 (2003); astro-ph/0302209

CW:

Outline: continuous-wave detection validation steps

I: CW pipelines produce candidates that pass a number of tests.

- A: The SNR is above a threshold set by a false alarm rate.
- B: The candidate is not vetoed by coincidence test(s).
  - 1: SNRs match in all IFOs within expected error.
  - 2: Frequencies match in all IFOs within expected error.
  - 3: Sky positions match in all IFOs within expected error.
  - 4: Spindowns match in all IFOs within expected error.
- C: The candidate is not vetoed by "goodness-of-fit" test(s).
  - 1: These test might be applied before or after the coincidence test(s).
  - 2: These tests have to undergo Monte Carlo simulations to set their false dismissal rates.
  - 3: A chi-squared test in frequency-domain code has been implemented.
  - 4: Other possible tests:
    - a: Test line width (instrument lines will be broadened by doppler demodulation).
    - b: Test that SNR grows as $T^{1/2}$ on average.
    - c: Test SNR vs. sky position.
    - d: Time-domain code could test chi-squared value for parameters that minimize the posterior pdf.
- D: The signal is not vetoed as a known instrument line.
- E: Many candidates will survive this step.
  - 1: Large false alarm rates will be used.
  - 2: Very small (ideally zero) false dismissal rates will be used.
  - 3: These rates are probably chosen to produce the approximate number of candidates that Step II can handle.
  - 4: The exact value of these rates will be found by Monte Carlo simulations using software and hardware injections.

II: Follow-up studies are done on candidates that survive Step I.

- A: A coherent search on a fine-grid parameter space surrounding the candidate's parameters is done on the same data.
- B: Fine tune "goodness-of-fit" test(s).
  - 1: Check that minimum chi-squared or maximum likelihood value is consistent with a signal (i.e, that a CW model for the signal is not rejected based in this value).
  - 2: Fine tune SNR vs. time tests.
    - a: Check that SNR grows as $T^{1/2}$ on average (if not done in step I).
    - b: Check that SNR varies consistently with the diurnal antenna pattern.
    - c: Estimate parameters and perform chi-squared test of SNR vs. time using JKS equations for SNR.
  - 3: Fine tune other "goodness-of-fit" test(s)?

C: Check that a joint coherent analysis using all IFOs is consistent.
D: Reproduce the results using data from a prior or subsequent run.
E: If any inconsistencies occur, check if a possible pulsar type "glitch" can account for it (i.e, does the data indicate the frequency changed discontinuously at some point, and can a better fit be found by modeling this).
F: Few candidates will survive this step.
1: This step should reduce the false alarm rate to a very small value.
2: The false dismissal rates should be kept as small as possible.
3: The exact value of these rates will be found by Monte Carlo simulations using software and hardware injections.

III: Candidates that survive Steps I and II should have very small false alarm rates and will be consistent with a real signal. Thus, it is time to find confidence intervals for h_0 or A_1, A_2, A_3, and A_4.
A: Predetermined unbiased approach(s) must be used to determine the confidence intervals.
B: Intervals for several levels of confidence could be found (e.g., 90%, 95%, 99.9%).
C: The method(s) should give σ's for the estimated parameters.
D: Frequentist approach:
1: Parameters are estimated from minimizing chi-squared or maximizing the likelihood.
2: A fake signal with the parameter estimates is injected into the noise many times (at different frequencies). The parameters are re-estimated each time.
3: The σ's of the parameters are found.
4: A boundary is drawn that contains x percent of the estimates. The boundary would be determined by one of the following criteria:
a: A boundary of constant $\Delta\chi^2$ or constant likelihood ratio is used. (For example see Numerical Recipes and Feldman and Cousins)
b: A boundary that gives the central confidence interval is used. (In 1D this gives equal probability of finding a measurement below or above the acceptance interval).
c: A boundary based on the σ's is used.
E: Bayesian approach: The method would be similar to the Frequentist approach, except the σ's and confidence interval would be drawn from the posterior pdf.
F: How to handle the nuisance parameters.
1: Don't. Give the confidence ellipsoid for A_1, A_2, A_3, and A_4; display the result by projecting the ellipsoid onto each axis of this 4D parameter space. (For example see Numerical Recipes.)
2: Marginalize.
3: Use worst-case nuisance parameters.
G: Candidates survive this Step based on whether zero amplitude is not in the confidence interval(s) and/or on how many σ's an estimated

amplitude is from zero.

IV: Candidates that survive Steps I, II, and III will be \gold-plated" potential detections. Thus, it is time to rule out all other possibilities that could produce such a signal.

- A: Review the validation of the software again.
  - 1: Have any new bugs turned up?
  - 2: Are any new validation tests or additional Monte Carlo simulations indicated?
- B: Independent code should verify the result (this may already been done as a part of Step II).
  - 1: If the frequency-domain code found the candidate use the time-domain code to verify this and vice versa.
  - 2: Incoherent methods not already applied to this candidate might be run as further validation.
- C: Check key results using independent SFTs.
- D: Check the raw frames if the candidate is found in RDS data, and vice versa.
- E: Check elogs for problems with excitations, DAQ corruption of data, etc....
- F: Understand periodicities that can occur in the DAQ system that may have not already been vetoed.
- G: Check excitation channels (make sure no accidental injection was done).
- H: Check PEM and other channels for environmental causes.
- I: Check frequencies of computer monitors and other electronics that might not already have been vetoed.
- J: Check if up/down conversion can happen in the electronics and get into GW channel?
- K: Check for other harmonics. Is there a signal at f/4, f/2, 2f, 4f or at ratios of the harmonics of the r-modes? Thus, can we determine if the signal is due to spin, precession, or a mode? (This may be very hard to do.)
- L: Check if the parameters make astrophysical sense. (If not then this could be something really new, but do we require greater confidence in that case?)
- M: Is there a known astronomical object associated with the candidate (e.g., pulsar, x-ray source, etc...). (If not this is not a problem; if so can we think of further consistency checks with astrophysical EM data for the source?)
- N: If a significant problem is found, we may need to adjust the pipeline in Steps I, II, and III and repeat Monte Carlo simulations.
  - 1: Do we know how to do this without introducing bias?
  - 2: How much of this to do we have to decide upon a priori?

V: If a candidate survives Steps I, II, III, and IV, should we seek corroboration?

- A: Ask for astronomical data to seek EM counterpart?
- B: Ask for data from other GW detectors?

Inspiral:

This is a summary list of things we've talked about and done when investigating detection candidates. It has been updated based on our discussions during a teleconference on 21 May 2004. It is our goal to follow up on the most significant candidates in any search with these methods. The words in [..] indicate whether we have followed these procedures in S1 or S2 when following up candidates:

1. Identify candidates: if they have a low (< 10-20%) false alarm probability per week (is this the right time?). [S1,S2]

1a. What does the AS_Q time series look like near to the event candidate? [S1,S2]

1b. Is there anything visible in the burst analysis of the same data? [S1,S2]

1c. What about ringdown? [no]

2. What data quality flags may have been on when these candidates were identified? [S1,S2]

3. Was there anything strange happening in the instrument according to the sci-mons or the operators in e-log? [S1,S2]

4. Is there any data corruption evident between the data used in the analysis and the raw data archived at Caltech? [no]

4a. Were there any injections being made? Are injection channels clear? [S1,S2]

5. Are the candidates stable against changes in segmentation? [S2]

6. Are the candidates stable against small changes in calibration consistent with systematic uncertainties? [S2]

7. What are the parameters of the event? Are the masses reasonable? Is the distance reasonable? What position information is available via the time-delays? Are distances as measured in both instruments consistent with position information? Can the harmonics give useful information (See papers by Sintes and Vecchio and by Brady and Evans.) [S1,S2 but not completely done]

Remember to measure the parameters more accurately after the identification. Monte-Carlos to assess our understanding of many of the listed items will be necessary.

7a. If we cut signals by regions of parameter space, does this change the false alarm probability per week? [no]

8. What does the reconstructed waveform look like? (The tools for this don't really exist at this time!) [no]

9. Where does the candidate lie in parameter space of snr, chisq, masses, etc. (Including many different tests that we have discussed -- code is not ready for these!) [S1,S2]

9a. Did the template bank ring-off all over? Is this consistent with a signal? (This is really one of our chisq tests?) [S1,S2]

9b. How does the snr v time and chisq v time plot look? [S1,S2]

10. Make a follow up with coherent multi-detector code. How does it look? [no, kinda in S2]

11. Are there are any auxiliary or PEM channels which indicate that the instrument was not behaving correctly? [S1,S2 not complete]

The procedure for PEM channel checks is being actively fleshed out and followed up. To date (05/2104), we have said that we would like to have an understanding of physical origins of any glitches if we were to use them in the pipeline after opening the box. Some explicit things that are not in the e-mail list are:

Earthquakes
Thunderstorms

We will continue to make such lists in advance of an analysis, but there has never been any evidence that there are real physical couplings.

12. Are there any EM triggers in coincidence? [S1, not yet in S2]

13. Were any other detectors operational during the period when the candidate was identified? [S1,S2]

In general, many of the items of the list could be automated, but have not been as yet. We should work toward this.

Discussion of our detection investigations list:

Patrick and Gaby circulated a note with the list of detection checks that we have discussed and used in the past. The e-mail can be found at

http://www.lsc-group.phys.uwm.edu/mailman/private/iulgroup/2004q2/002384.html

The following items were discussed during our meeting today:

The procedure for PEM channel checks is being actively fleshed out and followed up. To date, we have said that we would like to have an understanding of physical origins of any glitches if we were to use them in the pipeline after opening the box. Some explicit things that are not in the e-mail list are:

Earthquakes
Thunderstorms

Alan Weinstein asked if it is possible to make such a list in advance of an analysis? Gaby said that we always have done. But there has never been any evidence that there are real physical couplings.

Alan mentioned that we should be able to add as much information to the list we have and add it to the DCC as a statement of the group's intent. This is counter to our current stance where we feel uncertain about our understanding of all cuts until we have investigated in detail for each run. The group plans to add this to the web page for reference in our investigations.

Sathya commented that they are having problems of GEO identifying good veto channels. The experimentalists are not comfortable in the understanding to say that any channels are good vetos.

Stas reminded us that we should remember to measure the parameters more accurately after the identification.

Peter liked the non-judgmental way the list is made and thinks we should continue in this way.

Patrick indicated that Monte-Carlos to assess our understanding of many of the listed items will be necessary. In general, many of the items of the list could be automated, but have not been as yet. We should work toward this.

Burst:

Outline of steps in burst group detection validation claim

- Understand the statistical significance of result
    - Convincing statistics essential
    - Other ETGs?
    - Robustness to ETG tuning?
- Data validation
    - elog check
    - suspicious GPS times, times within second, time within ETG stride
    - Possible unintended injection
    - Check RDS data against original data frames
- Software validation
- Investigate possible external causes for events
    - PEM channels ("blind" testing on H1-H2 coincidences)
    - Compare PEM channels with measured suseptibilities
    - Clear evidence of an auxiliary channel cause on single interferometer level
- Consistency of individual events with GW expectations
    - Signals have the same sign?
    - H1-H2 amplitudes consistent?
    - Waveform reconstruction?
    - T-F spectrograms?
    - Consistency of time delays?
    - If multiple events, consistent with plausible sky distribution (e.g., not too many events coming in along L1 minimum…)?
- Seek any corroborative indicators
    - Other LIGO searches
    - Other GW detectors
    - Gamma-ray telescopes

Date: Mon, 23 Feb 2004 14:38:11 -0500 (EST)
From: Erik Katsavounidis <kats@ligo.mit.edu>
To: Laura Cadonati <cadonati@ligo.mit.edu>,
Sergei Klimenko <klimenko@phys.ufl.edu>,
Igor Yakushin <igor@ligo-la.caltech.edu>,
Ken Yoshiki Franzen <franzen@ligo-la.caltech.edu>,
Szabolcs Marka <smarka@ligo.caltech.edu>, mit-ldas@ligo.mit.edu,
Keith Thorne <thorne@gravity.psu.edu>,
Peter Saulson <saulson@physics.syr.edu>,
Alessandra Di Credico <dicredic@physics.syr.edu>,
Robert Schofield <rmss@conch.uoregon.edu>,
Alan Weinstein <ajw@hep.caltech.edu>, Lee Samuel Finn <LSFinn@PSU.Edu>,
Patrick Brady <patrick@gravity.phys.uwm.edu>,
Shourov Keith Chatterji <shourov@ligo.mit.edu>
Cc: Stan Whitcomb <stan@ligo.caltech.edu>
Subject: Re: [Bursts] Intriguing result from the waveburst analysis

Hi all,

Here is a compiled list of proposed investigations one would like to pursue in examining closer the observed excess of waveburst events that pass the r-statistic. It is mostly based on a list earlier last week Peter as well as myself attempted to make; it also tries to capture comments that most of you have made during last week's telecon and in separate e-mail exchanges.

We should be merging ideas and checks in the next day or so and try to make a plan of how to respond to this 'firedrill' as Szabi very successfully called it.

The list is indicative, kind of thematically organized and not necessarily prioritized. If you could please run through it and provide any comments as well as add yourself what you would like to see in it? Our goal is to involve the larger burst group in the making and carrying out of a prioritized plan starting as soon as possible.

Erik and Stan

Proposed investigations
======================

Zero-level sanity checks:
- Convert GPS times to Calendar times, and check for suspiciousness.
  (done, nothing obvious)
- Read elogs for the times in question. Was there anything anomalous going on?
  E.g., injections, people in LVEA, problems noted by operators,...
  also, where in lock stretch was event (start/mid/end)?

Data integrity:
- Check for undocumented/unauthorized/spontaneous hardware injections.
  Some test points are recorded in frames, some are not. How do we
  go about it?
- What checks can we make concerning data tampering?
  We can check RDS against raw frames. What ways do we have to
  check the integrity of the frames themselves?

Other ETGs/detectors:
- Do other burst ETGs find these events (not really- but how much of this
  can change with different thresholds/coincidence criteria)?
- Similar analysis on other ETGs; multi-threshold analysis;
  is there any zero-lag excesses?
- Coincidence analysis with LIGO's inspiral search for BBH/BNS;
  are there counterparts of the burst events there?
- Coincidence analysis with LIGO's BH Ringdown search.
- TAMA was on in 3/4 events (none were seen); push for a coincidence
  analysis of the high frequency WaveBurst events with TAMA events
- Rome bar detectors data.
- HETE extended GRB catalog for S2 (none were seen).

WaveBurst:
- What changed from v2 and v3?
- How's the peak time reported by Waveburst relates to the times of the
  r-statistic; actually, at this point it will be good to have the
  entire Waveburst event structure.

- Run the waveburst with the data pre-filtered according to Shourov's HP and LPEF's.
- Examine 2-IFO coincidences before folded into the triple coincidences; (distributed Waveburst events are triplets).
- Pursue an H1-L1 and H2-L1 analysis along the same lines of the present H1-H2-L1; apply to double/triple coincidence playground.
- Examine details of Waveburst events involved; are they 'typical'? something that makes them 'non-standard'?
- What was the detection efficiency for Waveburst calculated the nearest to these events (or even at the events)? Was it average/low/high?
- Use time-lags randomly chosen in the [-110,+110]s interval as opposed to fixed steps of 5s.

R-statistic:
- Check whether the colleration between H1-H2 interferometers has a positive sign (consistent with GWs) or negative sign (inconsistent with GWs)
- Perform the same (with the one of the 0.1-1KHz) r-statistic analysis for the 1-4kHz range of waveburst triggers. Is same excess observed?
- Perform a coherence test of all waveforms involved in the events; e.g., use the L1 waveform from the first event to check its coherence with the L1 of the fourth event.
- Use external trigger r-statistic code to study coherence of the full waveburst measured/background events.
  Use exttrig search parameters as well as those chosen by Laura.
  How does the result change with the 'exttrg' search parameters, are the results reproduced?
- Does r-statistic run at random times yield zero-lag excesses?
- r-statistic run on "events" produced by ETGs with time shifts: is there any zero-lag excess?
- Implement a chi^2 test on the r-statistic; to start with, no need to retune, but simply obtain the same distribution and see if the 4 outliers following the K-S test remain this way once a chi^2 test is invoked.
- Trend of r-statistic with its internal time shift
- Comparison of r-statistic between physical and unphysical time shifts
  Does ± 10 ms produce similar or different results than longer shifts.
  This could be a way of looking for environmental influences that don't necessarily come from precisely midway between ifos, or that travel

at less than light speed.
- Run the r-statistic with Sergei's HP and whitening.
- Plot rate of events vs. lag-time together with the zero-lag measurement
  as a function of r-statistic's beta threshold.

Instrument and environment anomalies:
- Examination of problem of H1 vs. H2 correlation
  Would we do better to adopt Stochastic Group's method of producing a single correlated Hanford channel to correlate with LLO? Does that have simpler statistics?
- Studies of veto channels:
  + One method is to return to previous glitchMon/WaveMon runs of various candidate veto channels, and look for something unusual that is coincident with these events. Need to choose the right list of channels, and then to pick a criterion of unusualness that doesn't give too large a false-dismissal probability for real gravity wave events. (Is 1% false dismissal over all channels a good choice?) Alessandra has started this.
  + Shourov has produced spectrograms (Q-grams) of various channels. They provide a wealth of information, but it is hard to know how to process them without defining event parameters and doing statistical study like in previous paragraph. He has started on this.
  + We should probably break into two separate categories: PEM channels and ifo diagnostic channels. In particular, for PEM channels with measured transfer functions (or influence coefficients), we have a better criterion than "unusualness": we can instead require that a signal be large enough to produce a false event. Robert Schofield has this information for some channels, or at least data from which it can be derived.
  + We should produce histograms of PEM channel event strengths anyway.
- Check violin mode excitations:
  Local mechanical events at one test mass could excite violin modes. That can be looked for as a diagnostic.
- Ask the true experts to examine ifo data for anomalies.

Common features of emerging events:
- Can one see match in time series? How to treat different SNR for each

detector?
- Can one extract the candidate waveforms? (same SNR considerations)
- Can one see events, and degree of match, from spectrograms?
  Shourov has provided these. Again, need to learn criteria, but at first look only Event #3 looks like what one would hope, that is blobs at same t-f coords in all 3 spectrograms.
  In other 3 events, one sees many equally strong blobs at other times.
- Can one see events, and degree of match, from corrgrams?
  Szabi has provided these. Similar mixed picture: It isn't clear that there is anything special about these events OTHER THAN the combined r-statistic confidence. Then again, that is what we thought we wanted to look for ...
- Use external trigger code to estimate coherent power and burst times.
- What is the statistical significance of the counting experiment's observed number of events? How many S2's playgrounds would take in order to observe this number of events from fluctuations of background?
- confront the distribution of strength of the events at zero vs non-zero lags; are these two the same (KS/chi^2/Mann-Whitney)?
- Can we estimate strength, and then link to any possible astrophysical model?
  Strength (h_rss) we need to do, but without clues to distance scale we can't say too much astrophysical. Still, we could pick a couple of fiducial distances (e.g. from 1 kpc up to 20 Mpc.) Stan sketched a graph like this at the last f2f.
- with enough events in hand, perform sidereal time analysis, point source analysis.

E. Katsavounidis
Massachusetts Institute of Technology | LIGO Lab, NW17-161
175 Albany Street | tel#: 617 258 9218
Cambridge, MA 02139 | fax#: 617 253 7014