Discussion of Statistical Methods, Tools, and Simulations

- Review of tools and methods
- (I am only familiar with) tools and methods in use in the burst working group
  - » What are the special needs of other groups?
  - » What tools and methods used by other groups would be of general use?
- the simulation effort exercises and requires the entire analysis chain

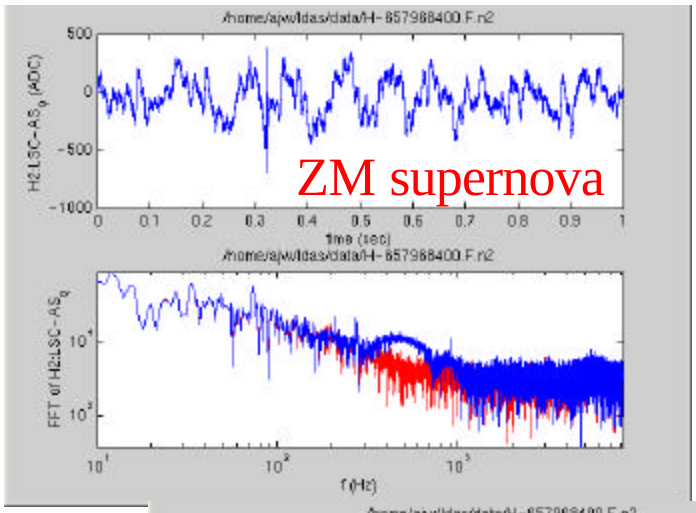
Data analysis model (burst group)

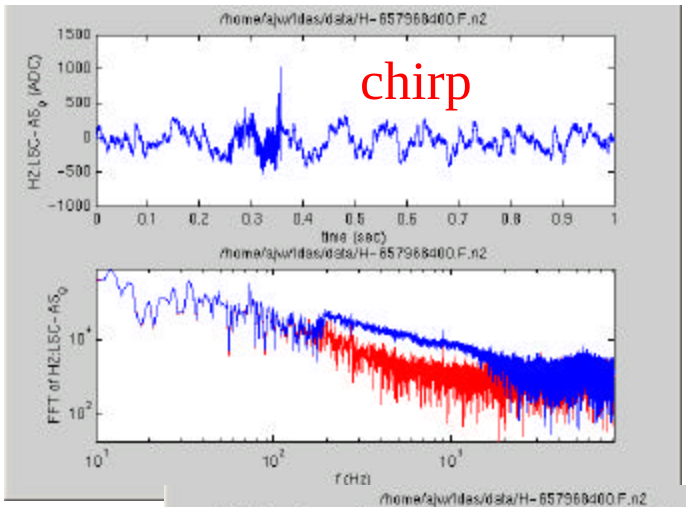
- Does the model used by burst WG fit the other WGs?
  - » generate fake signals, scaled to absolute h_strain
  - » filter fake signals through IFO response functions
  - » add to real data, or to white or colored gaussian noise
  - » run search DSOs on the resulting frames
  - » perform statistical analysis on results of DSO runs
  - » present results in a form suitable for publication(s).

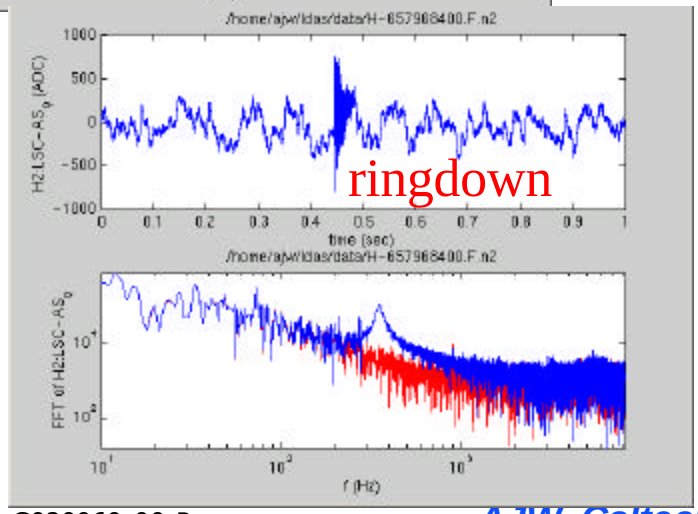
Generating simulated signals

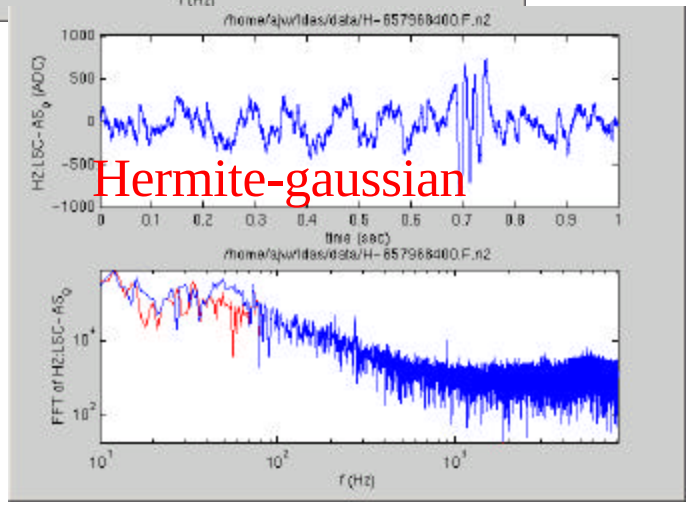
- Generating simulated signals in units of strain is the easy and fun part!
- I think that all WGs have mechanisms for doing this.
- For burst WG:
  - » Chirps
  - » Ringdowns
  - » Hermite-Gaussians and sine-Gaussians
  - » Zwerger-Muller supernova GW simulations

Waveforms buried in E2 noise, including calibration/TF









Adding simulations to the data

- How are fake signals added to the data, minimizing uncertainty associated with special handling of simulated data relative to real data?
  - » Make fake frame files
    - – c code, matlab, ROOT tools
    - – can add to data or simulated noise before making frame,
    - – or can add to noise in LDAS job, "on the fly"
  - » add signals to real data or noise frames in datacondAPI or frameAPI (burst group)
    - – need more sequence manipulation tools in datacondAPI, so that one could add a signal from a file, at a run-time-determined time, amplitude, duration
    - – could actually generate the signal in datacondAPI
    - – maybe even convert from strain -> volts there as well
    - – This would facilitate the accumulation of high statistics simulation runs, without having to manage huge frame files padded with zeros.
  - » add signals in wrapperAPI / LAL code (inspiral group)
- How to ensure reliable IFO response functions?
  - » Shall we standardize the filtering algorithm / mechanism?
  - » I use matlab *zpk* to represent response function and *lsim* to filter the fake signal from strain -> volts. Better would be to use datacondAPI or LALWrapper.

Example DatacondAPI algorithm

```
-framequery { { R $site {} $times Adc($channel) }
              { F $site /ldas_outgoing/jobs/ldasmdc_data/burst-stochastic/burstscan_e7h2.F {} Adc(0) } }
 -aliases { x = _ch0; s = _ch1; }
 -algorithms { zx = slice(x,0,5914624,1);
               zy = slice(s,0,5914624,1);
               zm = mul(zy,$mulsim);
               zs = add(zx,zm);
               zz = tseries(zs, 16384.0, $stime, 0);
               stat = all(zz);
               intermediate(,stat.ilwd,stat,psd of ch0);
               pz = psd(zz,16384);
               intermediate(,pzs.ilwd,pz,psd of ch0);
               z = resample(zz,1,8);
               m = mean(z);
               y = sub(z,m);
               q = linfilt(b,y);
               r = slice(q,2047,737280,1); }
```

Grab slice of data and simulated signal

Multiply signal by a run-time-determined multiplier,
Add to real data, make a time series

Accumulate statistics and power spectrum

Subtract mean; additional whitening

statistical analysis

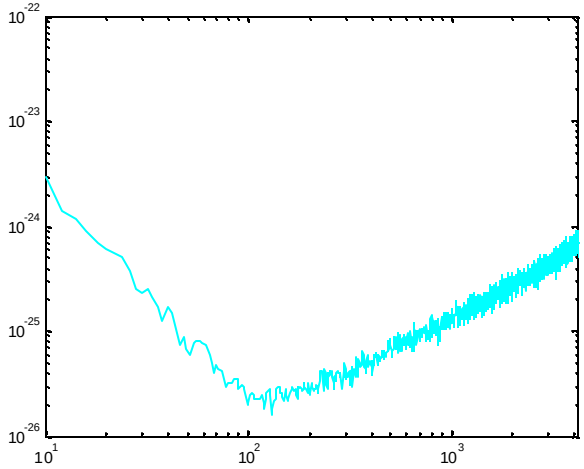
For burst and inspiral:

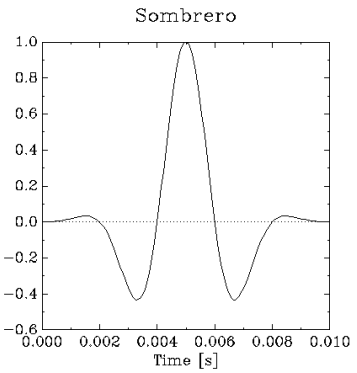
- output of DSOs are event triggers, deposited into database
- further analysis is done at the event trigger level, although in the future, will want to look back at raw data
  - » manipulate database entries
  - » merge multiple entries
  - » apply vetos from triggers on auxiliary channel (DMT?)
  - » find coincidences
  - » evaluate fake rates vs efficiencies at the same time
- Here, the Event Tool is being used
  - » Sigg and Ito. OO classes in the ROOT environment for manipulating database "event" entries.
- Can also use matlab tools, including LIGOtools (Shawhan) for ingesting xml files obtained from DB queries.
- Can also use the power of DB queries.

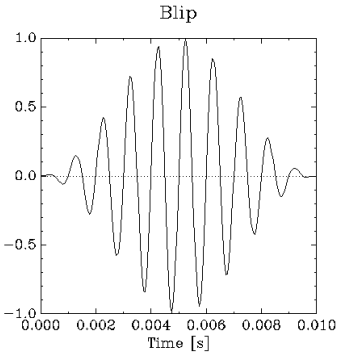
How about the stochastic and pulsar analyses?

Other sources of fake signals

- Use of *simdata* matlab package (Finn) to generate colored noise
  - » steep seismic noise spectrum makes it difficult but not impossible to generate desired spectrum
- Use of signals injected into interferometer using GDS system:
  - » bursts and inspiral signals (chirps, wiggles, blips, sombrero; Gonzalez, Marka, Shawhan, during E7)
  - » stochastic and line signals (Shawhan, Bose, Brown, during E7)
- How best to handle "model dependence"?
  - » The burst group must struggle with model dependence because we have no good models
  - » The other groups DO have good models, but will have errors and variations in those models).







LDAS and database

- What demands will be made on which LDAS installations?
  - » Shall we use the site installations? Or only LSC installations devoted to this task?
  - » How shall we coordinate these resources? (The LIGO Software Users Group has been discussing this)
  - » How/where should data streams from LHO and LLO be combined, for analyses that require cross-correlations, like stochastic?
  - » Standardized job control/bookkeeping scripts? (LSUG is developing job bookkeeping and monitoring tools, based on Shawhan's *LDASJob*.)
  - » How to facilitate multiple passes through the data? (Large RDS's at LDAS installations).
- What demands will be made on which LDAS databases?
  - » How best to avoid trashing the databases,
  - » and how/whether to get entries from one installation to another?
  - » (cf Peter Shawhan's proposals).

LDAS job bookkeeping

Run date	filter	channel	times	mul	ldas job	Tall	Tdatac	Tmpi	Ntrig	hrms	h(200-300)	h(300-400)	h(200-1000)
20020314145958	power	H2:LSC-AS_Q	693913420-693913679	0e0	mit NORMAL691	122.66	58.10	20.40	510	3.73e+04	1.27e+00	1.28e+00	Trig691.xml
20020314150231	power	H2:LSC-AS_Q	693913679-693913938	0e0	mit NORMAL693	124.84	57.86	20.43	509	3.38e+04	1.41e+00	1.42e+00	Trig693.xml
20020314150508	power	H2:LSC-AS_Q	693913938-693914197	0e0	mit NORMAL695	125.50	57.28	20.44	510	3.64e+04	1.40e+00	1.41e+00	Trig695.xml
20020314150739	power	H2:LSC-AS_Q	693914197-693914456	0e0	mit NORMAL697	120.11	57.78	19.43	510	4.09e+04	1.43e+00	1.44e+00	Trig697.xml
20020314151010	power	H2:LSC-AS_Q	693914456-693914715	0e0	mit NORMAL699	121.36	59.06	19.33	510	3.50e+04	1.52e+00	1.54e+00	Trig699.xml
20020318155516	power	H2:LSC-AS_Q	693913420-693913679	1e1	mit NORMAL766	115.09	51.75	20.44	498	3.73e+04	1.31e+00	1.40e+00	Trig766.xml
20020318155743	power	H2:LSC-AS_Q	693913679-693913938	1e1	mit NORMAL768	117.45	50.05	20.40	493	3.38e+04	1.45e+00	1.54e+00	Trig768.xml
20020318160010	power	H2:LSC-AS_Q	693913938-693914197	1e1	mit NORMAL770	118.14	52.13	20.42	497	3.64e+04	1.44e+00	1.53e+00	Trig770.xml
20020318160236	power	H2:LSC-AS_Q	693914197-693914456	1e1	mit NORMAL772	115.99	49.76	19.37	494	4.09e+04	1.47e+00	1.56e+00	Trig772.xml
20020318160506	power	H2:LSC-AS_Q	693914456-693914715	1e1	mit NORMAL774	120.77	50.49	20.43	497	3.50e+04	1.57e+00	1.66e+00	Trig774.xml

Run date	filter	channel	times	mul	ldas job	Tall	Tdatac	Tmpi	Ntrig	hrms	h(200-300)	h(300-400)	h(200-1000)
20020218205555	tfcluster	L1:LSC-AS_Q	693960121-693960481	0e0	mit NORMAL121	172.62	84.36	33.27	104	1.20e+07	1.80e+05	1.30e+05	6.59e+05 Trig121.xml
20020218205817	tfcluster	L1:LSC-AS_Q	693960481-693960841	0e0	mit NORMAL125	113.53	34.20	32.22	524	4.83e+03	7.89e-01	6.06e-02	8.64e-01 Trig125.xml
20020218210034	tfcluster	L1:LSC-AS_Q	693960841-693961201	0e0	mit NORMAL127	110.86	32.41	33.35	391	5.01e+03	7.66e-01	1.73e-02	8.01e-01 Trig127.xml
20020218210348	tfcluster	L1:LSC-AS_Q	693961201-693961561	0e0	mit NORMAL129	169.09	97.74	31.53	8	6.62e+06	9.08e+04	5.12e+04	2.71e+05 Trig129.xml
20020218210738	tfcluster	L1:LSC-AS_Q	693961561-693961921	0e0	mit NORMAL131	202.59	88.10	33.25	12	6.34e+06	8.43e+04	5.76e+04	3.09e+05 Trig131.xml
20020218211006	tfcluster	L1:LSC-AS_Q	693961921-693962281	0e0	mit NORMAL134	120.87	35.83	31.21	514	7.03e+03	9.45e-01	1.08e-01	1.08e+00 Trig134.xml
20020218211221	tfcluster	L1:LSC-AS_Q	693962281-693962641	0e0	mit NORMAL136	109.86	35.02	32.26	505	6.26e+03	9.30e-01	2.14e-02	9.71e-01 Trig136.xml
20020218211437	tfcluster	L1:LSC-AS_Q	693962641-693963001	0e0	mit NORMAL138	107.89	33.26	32.31	452	4.74e+03	9.28e-01	2.42e-02	9.75e-01 Trig138.xml
20020218211654	tfcluster	L1:LSC-AS_Q	693963001-693963361	0e0	mit NORMAL140	111.00	33.42	33.28	477	4.52e+03	9.06e-01	2.83e-02	9.64e-01 Trig140.xml
20020218211914	tfcluster	L1:LSC-AS_Q	693963361-693963721	0e0	mit NORMAL142	112.97	35.30	32.34	575	4.07e+03	9.13e-01	3.51e-02	9.90e-01 Trig142.xml
20020218212135	tfcluster	L1:LSC-AS_Q	693963721-693964081	0e0	mit NORMAL144	111.41	34.16	33.27	465	3.55e+03	8.76e-01	3.30e-02	9.49e-01 Trig144.xml
20020218212454	tfcluster	L1:LSC-AS_Q	693964081-693964441	0e0	mit NORMAL146	171.81	101.1	31.29	181	1.23e+07	2.06e+05	1.20e+05	6.13e+05 Trig146.xml
20020218212816	tfcluster	L1:LSC-AS_Q	693964441-693964801	0e0	mit NORMAL148	174.27	99.61	33.24	500	1.87e+07	2.96e+05	1.95e+05	9.20e+05 Trig148.xml
20020218213032	tfcluster	L1:LSC-AS_Q	693964801-693965161	0e0	mit NORMAL150	110.73	35.90	32.56	605	4.18e+03	1.60e+00	9.58e-02	1.76e+00 Trig150.xml
20020218213238	tfcluster	L1:LSC-AS_Q	693965161-693965521	0e0	mit NORMAL152	100.62	30.72	32.25	2	7.82e+05	2.16e+01	7.81e+00	9.14e+01 Trig152.xml
20020218213539	tfcluster	L1:LSC-AS_Q	693965521-693965881	0e0	mit NORMAL154	154.24	84.77	32.27	154	1.21e+07	1.94e+05	1.32e+05	6.82e+05 Trig154.xml

Modest proposal

- Collect useful tools and pubs in a web page
- Establish mailing list for people interested in developing or using these tools
- Coordinate with LSUG (and LDAS, datacond, etc)