Hardware and Software for Gravitational Wave Data Analysis

Bruce Allen

Max Planck Institute for Gravitational Physics

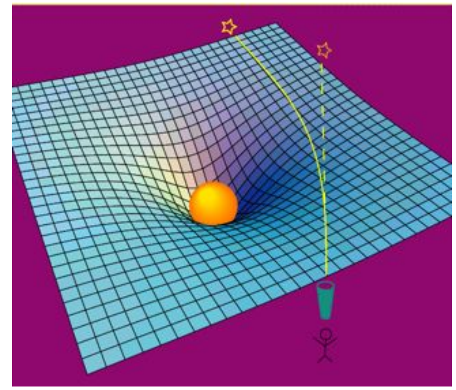
(Albert Einstein Institute, or AEI)

Hannover, Germany

Outline

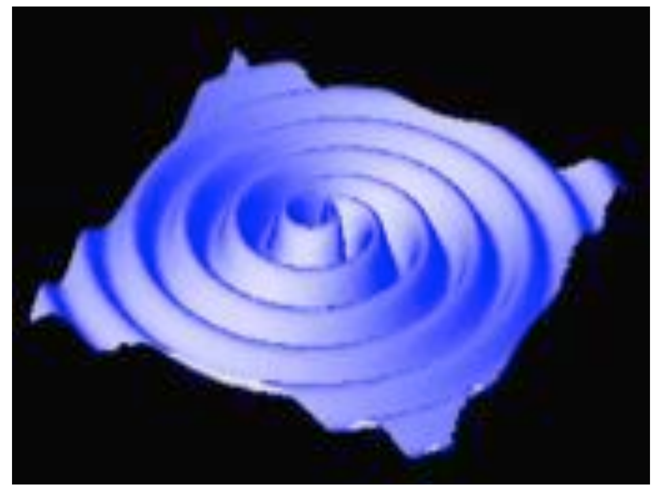
- Gravitational waves and gravitational wave detection
- How the software/computing infrastructure has been organized with LIGO (USA) and GEO (UK/German) collaborations
- The computing facilities at Albert Einstein Institute Hannover

What are gravitational waves?



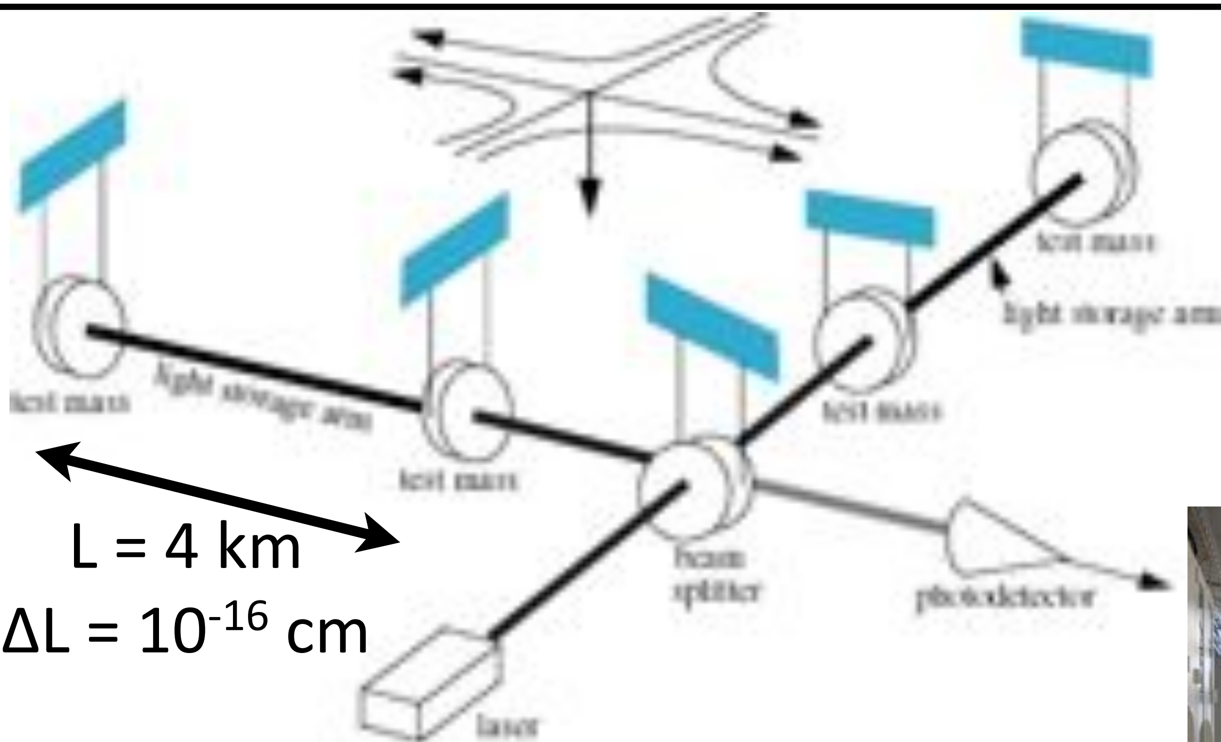
- Predicted by general relativity, Einstein, 1915-7
- Consequence of ‘nothing can go faster than light’
- Very weak. Produced by very compact massive objects undergoing large accelerations

$$L = \frac{G}{5c^5}\left(\frac{d^3}{dt^3}Q_{ab}\right)\left(\frac{d^3}{dt^3}Q_{ab}\right)$$



- Effect: differential acceleration in freely-falling test masses

Gravitational Wave Detectors

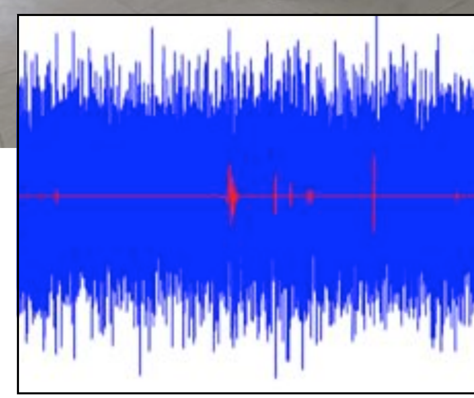


L = 4 km

ΔL = 10^{-16} cm

$$h = \frac{\Delta L}{L} \approx 10^{-21}$$

- Interferometer: gravitational wave "changes distance between mirrors"
- Laser measures mirror positions: differential motion causes fringe shifts
- Detected via error signal in locked loop
- Search for weak signals in noise

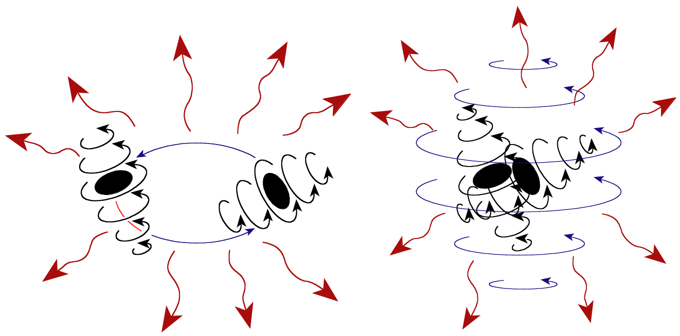


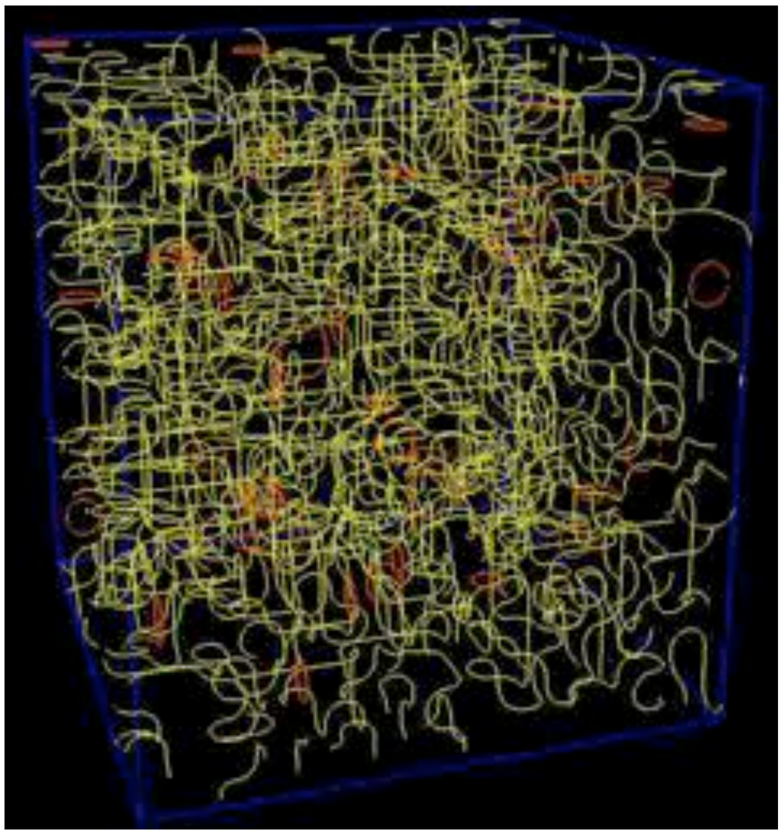
Detectors & Collaborations

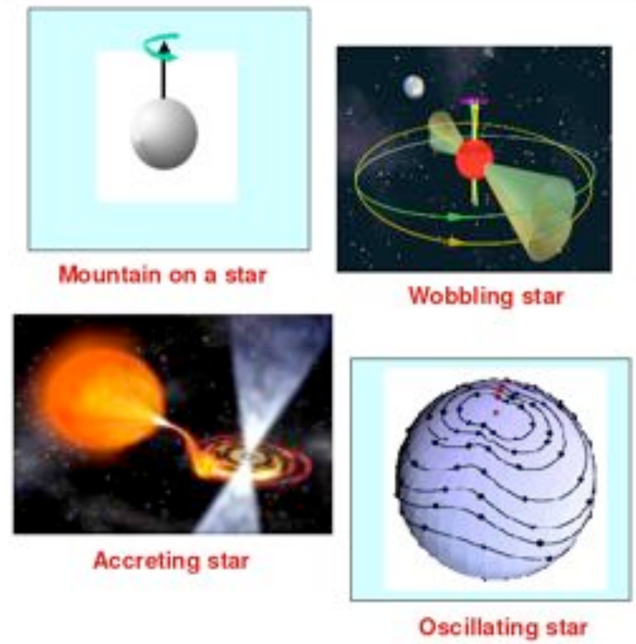
- LIGO in Hanford and Livingston, USA
- VIRGO (France/Italy) in Casinca
- GEO-600/HF (German/U.K.) in Hannover
- Large scientific collaborations (LSC, VIRGO, GEO) with O(1000) people
- Collaborations share data and publications

A “new window” on the universe

- Testing strong-field gravity and dynamic gravity near black holes and compact objects
- Neutron star structure (eccentricity, glitches)
- Population studies and stellar evolution
- Supernova mechanisms
- Origins of gamma-ray bursts
- Perhaps get information about the very early universe →







What's Coming (~ 5 years)

- Second generation detectors: advanced LIGO/VIRGO and GEO HF **FULLY FUNDED**. **First detections** on the 5-year timescale
- Japan Large Cryogenic Gravitational Telescope
- Possible "LIGO South" in Australia.
- USA: open Data policy. Will Europe follow?

Expected First Detection

Table 5. Detection rates for compact binary coalescence sources.

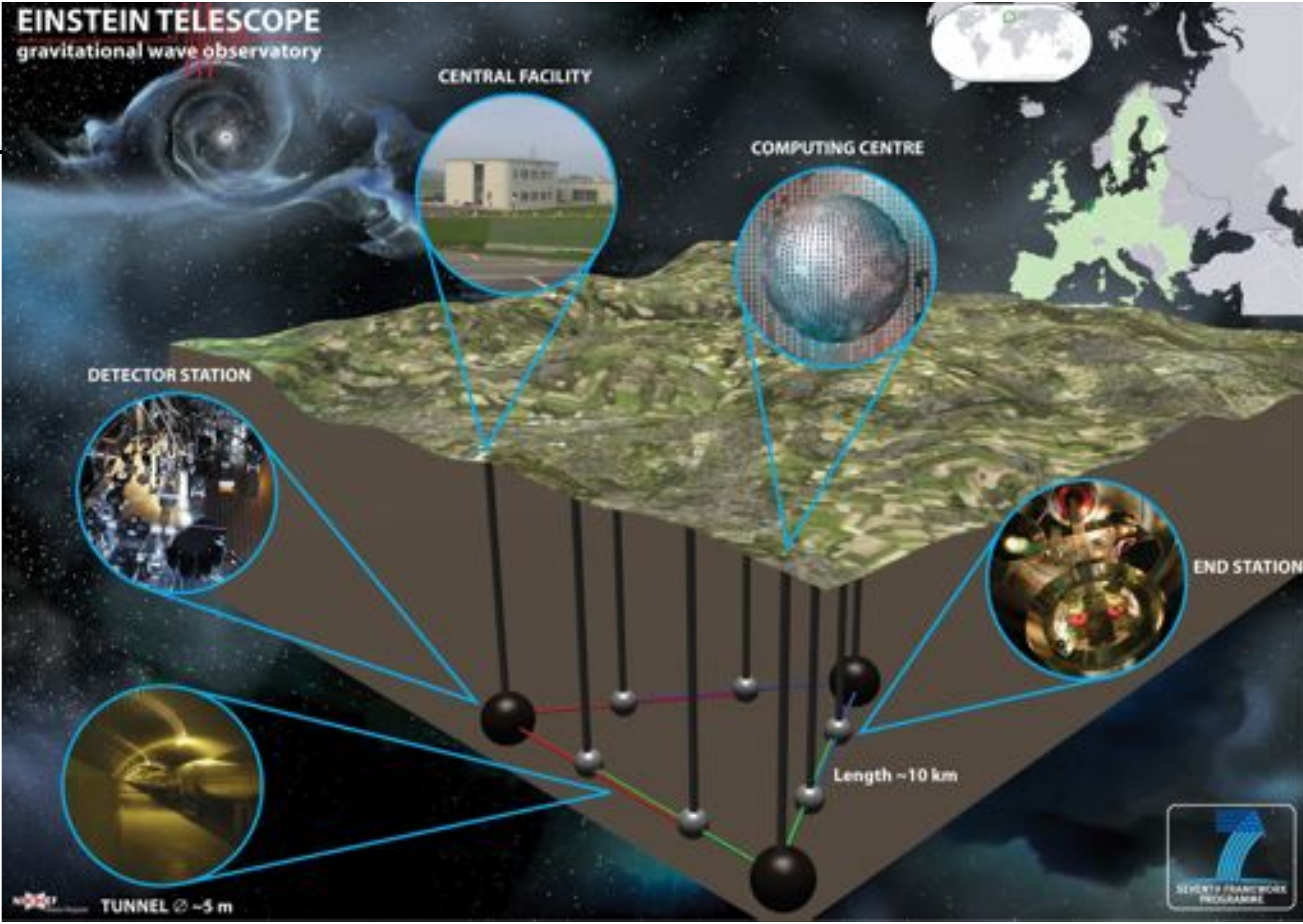
IFO	Source[a]	$\dot{N}_{\rm low}$ yr^{-1}	$\dot{N}_{\rm re}$ yr^{-1}	$\dot{N}_{\rm high}$ yr^{-1}	$\dot{N}_{\rm max}$ yr^{-1}
	NS–NS	2×10^{-4}	0.02	0.2	0.6
	NS–BH	7×10^{-5}	0.004	0.1	
Initial	BH–BH	2×10^{-4}	0.007	0.5	
	IMRI into IMBH			<0.001[b]	0.01[c]
	IMBH-IMBH			10^{-4}[d]	10^{-3}[e]
	NS–NS	0.4	40	400	1000
	NS–BH	0.2	10	300	
Advanced	BH–BH	0.4	20	1000	
	IMRI into IMBH			10[b]	300[c]
	IMBH-IMBH			0.1[d]	1[e]

J Abadie et al, Class. Quantum Grav. 27 173001 (2010)

- The "Second Generation" or "Advanced" detectors should make a first detection (compact binary inspiral systems) around 2015-17.
- Frequency band 10 Hz - few kHz
- Data analysis challenges for Advanced Detectors:
    - Sensitive at lower frequency: signals are 30 minutes long, not 30 seconds!
    - Need real-time detection pipelines and parameter estimation
    - Use GPUs efficiently

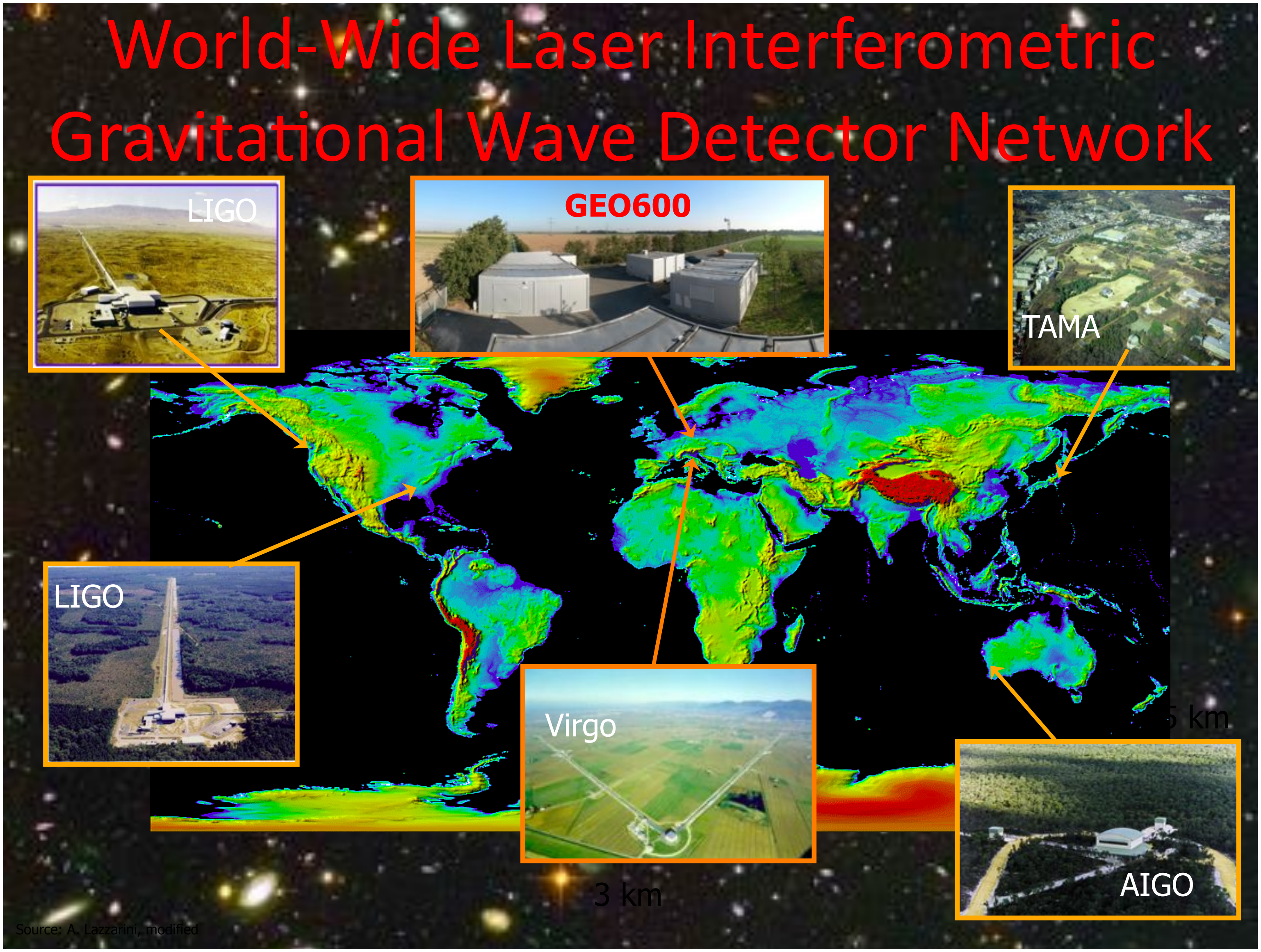
Future: Einstein Telescope (ET)

- European project, in a design study phase, 3rd generation. Operations might start around 2020 - 2025
- Additional order of magnitude sensitivity beyond Advanced LIGO/VIRGO/LGCT/GEO-HF
- Triangular configuration. Design study not yet complete, but it probably will be underground and under-mountain
- Sensitive from 1 Hz to few kHz
- Will see all NS/NS coalescences to z=2, BH/BH to z=8 (10^6 per year, one every 30 seconds).
  - High-precision cosmography
  - Mass spectrum of compact stars
  - Accurate determination of the equation of state of Neutron stars


EINSTEIN TELESCOPE
gravitational wave observatory
CENTRAL FACILITY
COMPUTING CENTRE
DETECTOR STATION
END STATION
Length ~10 km
TUNNEL ∅ ~5 m

Joint Data Sharing and Analysis

- LSC (LIGO Scientific Collaboration) and VIRGO and GEO600 all have joint data-sharing agreements.
- improved sensitivity for weak signals
- detection confidence
- position reconstruction (triangulation)
- waveform reconstruction
- Large-scale computing needs



LIGO/GEO DATA GRID

- Each detector: typically 250 TB of data per year.
- LDG has automatic data replication to sites (LDR)
- Common software installations (either CentOS or Debian based)
- About 200 users across collaboration.
- Compute farm, NOT HPC: use Condor as job manager
- Good parallel access to large data sets

LOCATION	CPU CORES
LIGO Caltech	2756
Syracuse University	380
AEI Hannover	**7024**
Cardiff University	176
LIGO Livingston	2024
U. Wisconsin - Milwaukee	1 628
LIGO Hanford	2 016
TOTAL	16 004

Software/Computing and FTEs

LSC

FTEs	
	Applications: DQ Pipelines, Low-latency Analysis, Offline Analysis, Simulations
10-12	Software Packages: I/O Libraries, LAL Suite, MatApps, DMT, …….
10-12	Services: Support, Build/Test, Monitoring, DQ Database, Gracedb, Data Replication, h(t), ..
14-16	Data Center Operations: Compute & Storage Clusters, Web Services, …..

Software/Computing

Applications: DQ Pipelines, Low-latency Analysis, Offline Analysis, Simulations

- Regime of analysis and detector characterization groups
  - follow open source model to allow improvements to get back into main software stack quickly
  - still need to define data products, especially in transient searches
  - and better document the codes for long-term viability

Software/Computing

Software Packages: I/O Libraries, LAL Suite, MatApps, DMT, ...

- Software stack is maturing:
  - Support different build systems in LIGO/Virgo and support different software environments enhancing code interoperability
  - Develop software release mechanisms that can be implemented for core libraries and online analysis while allowing incremental improvements
  - Improve documentation and ease of use
  - Provide support through help desk and training
  - Document data formats, data products and related libraries
  - Deliver robust, scalable detector characterization infrastructure

Software/Computing

Services: Support, Build/Test, Monitoring, DQ Database, Gracedb, Data Replication, h(t), ..

- Existing services span range of maturity
  - Data replication service is mature, but different ideas in LIGO and Virgo
  - DQ database worked for S5/VSR and S6/VSR2, but there is lots of administrative overhead
  - Deliver low-latency h(t)
  - Deliver reduced data sets
  - Deliver an integrated build & test
  - Deliver robust, user friendly event database
  - Deliver robust, user friendly interfaces to use distributed resources
  - Deliver robust, user friendly archive service for data products

Software/Computing

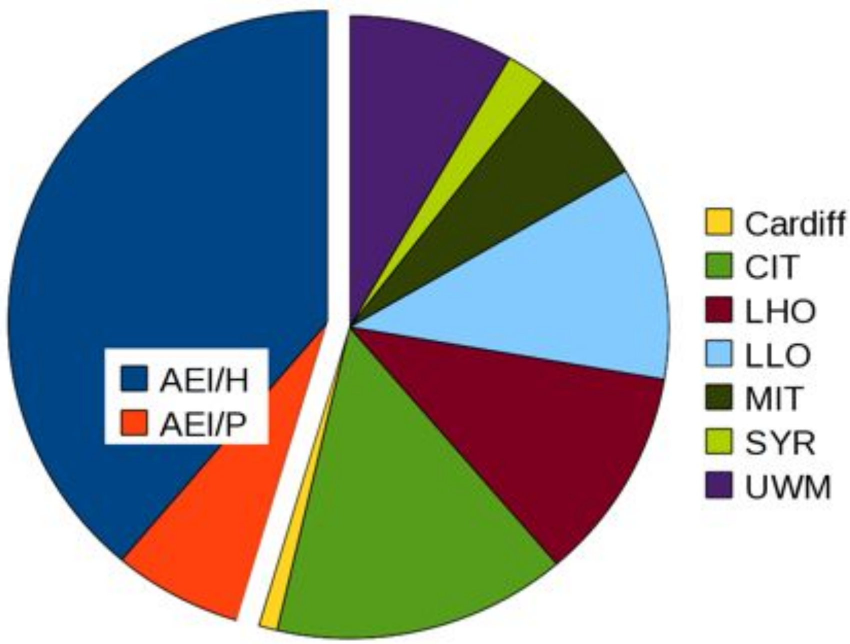
Data Center Operations: Compute & Storage Clusters, Web Services, ...

- Distributed computing: tiered data center approach with combination of US/European resources
- Technological improvements suggest we can meet computational needs, but we need to revisit/refine:
  - computational requirements
  - storage requirements
  - memory requirements
- Resource allocation and configuration for online/offline analysis should be reviewed

Managing Effort

- Scientific Applications
  - <u>Analysis groups</u> develop algorithms, software and pipelines
  - <u>Detector Characterization group</u> develops algorithms & software
- Hardware resources and software development
  - <u>Computing Committee</u> addresses policy, strategic planning
  - <u>Software Working Group</u> coordinate standards and software development, development, release & support
- Identification of responsible teams
  - <u>Data Analysis Council</u> undertakes annual planning exercise for analysis
  - <u>Computing Committee & Software Working Group</u> undertakes an annual planning exercise (in advance of MOUs being developed for LSC)
  - MOUs list the responsibilities and deliverables for LSC

Atlas at AEI



Atlas at AEI

- Dedicated facility in operation since May 2008
  - 400 square meters
  - 500 kW UPS power and cooling
- More than 7,000 CPU and 90,000 GPU cores. More than 3 Petabytes of spinning data storage, backed by a tape robot.
- Two full-time experienced PhD-physicist "sysadmins" plus ~8 undergraduate assistants.
- Designed for rapid access to large data sets, and reconfigurable.
- The biggest compute engine in the LIGO Scientific Collaboration (LSC). Widely used by all the LSC/Virgo search groups
- Atlas has done a significant fraction of the computing for ~ 25 LSC papers since 2008.
- Data-exchange hub for LIGO/VIRGO data.

Atlas at AEI

- Atlas was **#58 on the June 2008 Top-500 list** (fastest computers worldwide). Was #6 in Germany, and #11 in Europe.
- On the June 2008 Top-500 list, Atlas was the **world's fastest ethernet-based cluster**. This probably also makes it the world's most cost-effective cluster.
- Atlas won an **Infoworld-100 Award in 2008** (one of the two awards in the research category) for innovative design.
- Our networking supplier, Woven Systems (since aquired by Fortinet) gave AEI a million-dollar gift of enough additional switching equipment to double the size of the network.

AEI Hannover Computing Facilities

Atlas at AEI

Latest hardware

- Hierarchical Storage Management (HSM) system. A fast 90 TB fibre-channel disk array, backed by a1-PB expandable tape robot. Store home directories, LIGO/GEO/VIRGO instrument data, analysis results, and backups.
- 66-node GPU cluster, with 132 NVIDIA Tesla cards and 132 NVIDIA Fermi cards. Close to 200 Tflops of peak floating-point performance.

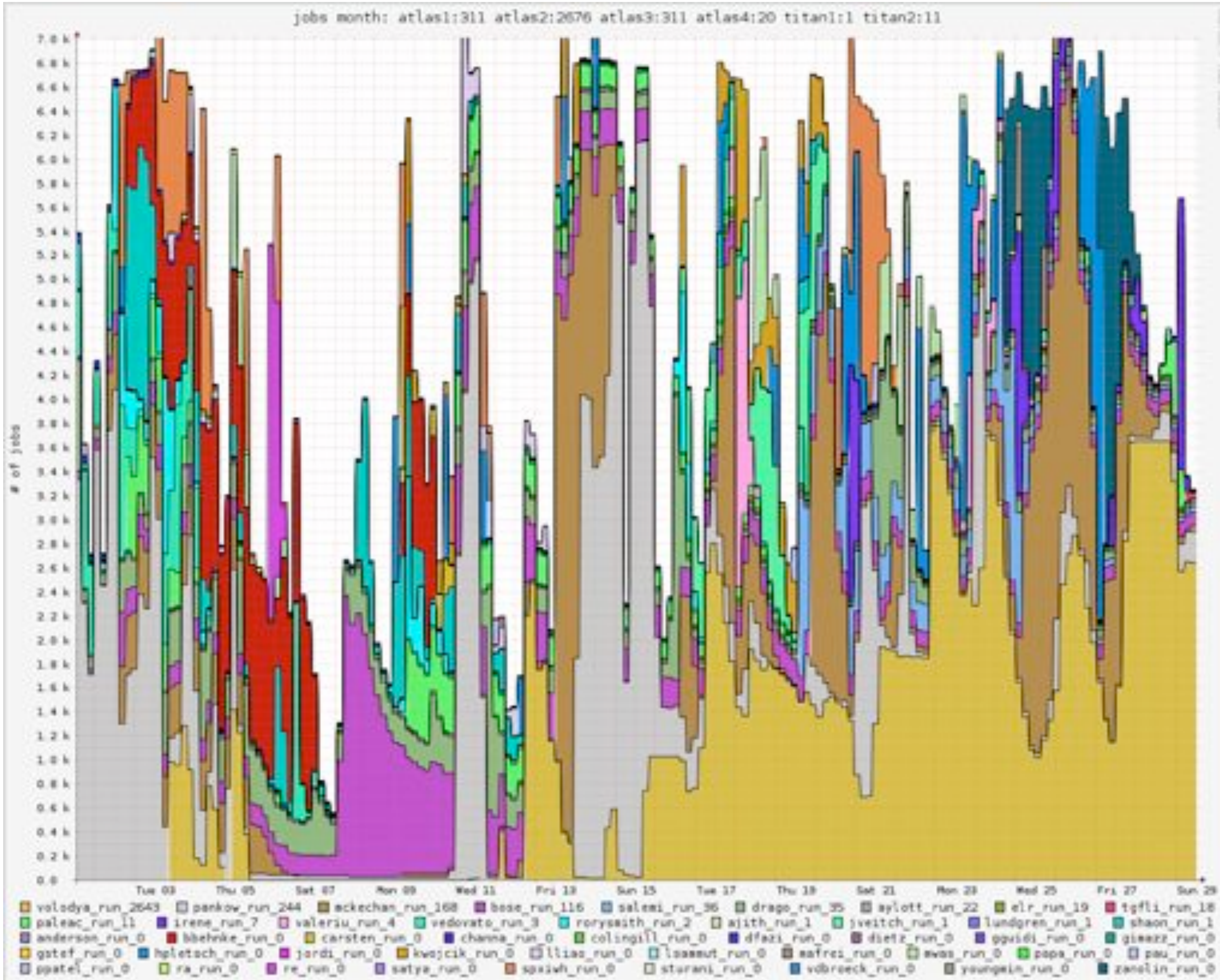
Next steps

- Early 2011: add another 400 square meters, 1800 HU of rack space, 500 kW of additional power and cooling, expand HSM cache and back-end.
- Networking already in place.
- Late 2011, Atlas 4-years old, install new compute
  cluster in new cluster room.

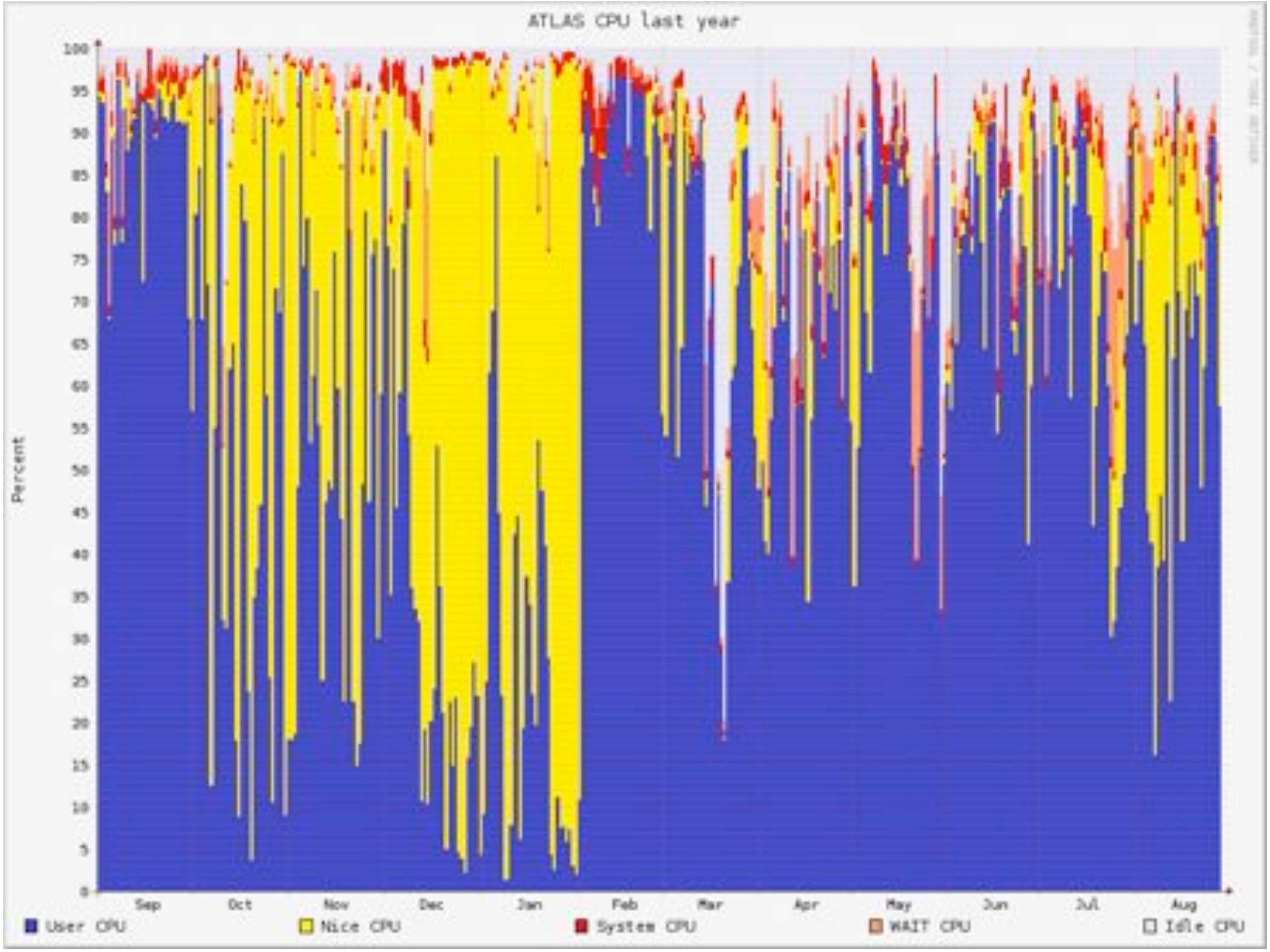
Longer term

- Flip-flop every three to four years.

Atlas Utilization



Atlas Utilization

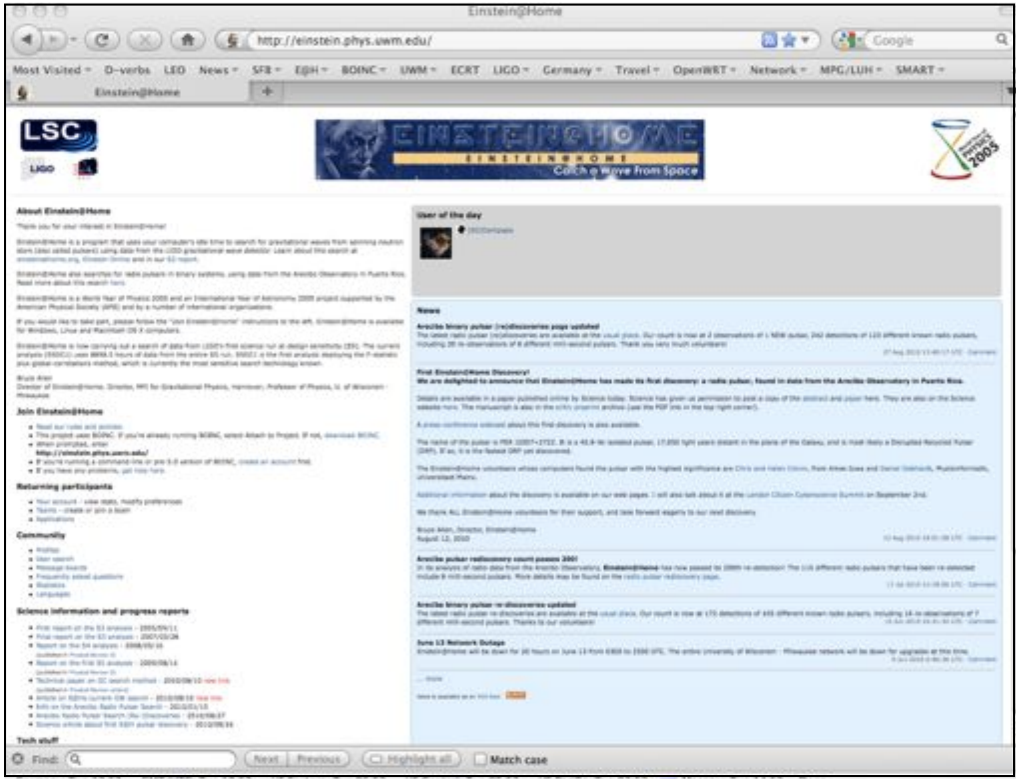


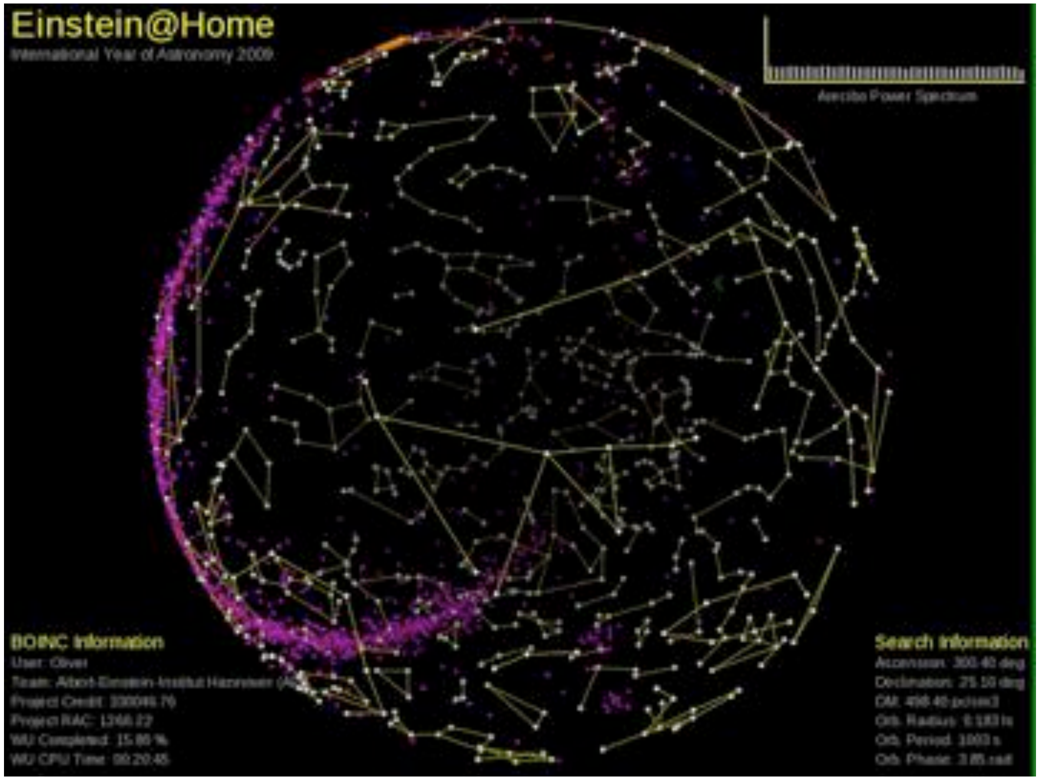
Einstein@Home: 320 Tflops

http://einstein.phys.uwm.edu/

- Science: public distributed search for gravitational waves from rapidly-spinning neutron stars, using LIGO/VIRGO/GEO data
- Started in 2005; most work done at AEI and U. of Wisconsin
- How to participate: download and install a simple “screensaver” program (2 minutes)
- Currently more than 250,000 people participate
- Computer power: at any time, day or night, about 100,000 computers (Win, Mac, Linux) are running Einstein@Home. Collectively more than 300 Tflops, 24 x 7, dwarfs the other computing resources of the LSC.

Einstein@Home





- LSC publications: all-sky blind search upper limits on continuous gravitational-wave sources in LIGO S4 and S5 data
- Significantly more sensitive upper limit paper in preparation (LIGO S5 + Hough)
- Also searching data from Arecibo Radio Observatory
- Einstein@Home recently discovered two new millisecond radio pulsars, one in a binary system

Pulsar Discovery by Global Volunteer Computing

B. Knispel,*† B. Allen, J. M. Cordes, J. S. Deneva, D. Anderson, C. Aulbert, N. D. R. Bhat, O. Bock, S. Bogdanov, A. Brazier, F. Camilo, D. J. Champion, S. Chatterjee, F. Crawford, P. B. Demorest, H. Fehrmann, P. C. C. Freire, M. E. Gonzalez, D. Hammer, J. W. T. Hessels, F. A. Jenet, L. Kasian, V. M. Kaspi, M. Kramer, P. Lazarus, J. van Leeuwen, D. R. Lorimer, A. G. Lyne, B. Machenschalk, M. A. McLaughlin, C. Messenger, D. J. Nice, M. A. Papa, H. J. Pletsch, R. Prix, S. M. Ransom, X. Siemens, I. H. Stairs, B. W. Stappers, K. Stovall, A. Venkataraman

Einstein@Home (*1*) (E@H) is a volunteer distributed computing project (*2*). Members of the public sign up their home or office computers (hosts), which automatically download work units from the servers, carry out analyses when idle, and return results. These are automatically validated by comparison with results for the same work unit produced by a different volunteer's host. More than 250,000 individuals from 192 countries have contributed; each week about 100,000 different computers download work. The aggregate computational power (0.25 Pflop/s) is on par with the largest supercomputers. E@H's primary goal is to detect gravitational waves from rapidly spinning neutron stars in data from Laser Interferometer Gravitational-Wave Observatory (LIGO) and VIRGO (*1*).

Since 2009, about 35% of E@H compute cycles have also been used to search for pulsars in radio data from the Pulsar ALFA (PALFA) project [supporting online material (SOM)] at the 305-m Arecibo Telescope (Puerto Rico). Data disks are sent to Cornell University's Center for Advanced Computing (United States), where data are archived. For E@H, data are transferred to Liebniz Universität (Hannover, Germany), dedispersed for 628 different dispersion measures (DM $\in$ [0, 1002.4] pc cm^{-3}), and resampled at 128 μs. Hosts receive work units containing time series for four DM values for one beam. Each is 2 MB in size, covering 268.435456 s. A host demodulates each time series (in the time domain) for 6661 different circular orbital templates with periods greater than 11 min (our Galaxy has even shorter period binaries). The grid of templates is spaced so that, for pulsar spin frequencies below 400 Hz, less than 20% of signal-to-noise ratio is lost. Fourier algorithms sum up to 16 harmonics. A total of 1.85% of the power spectrum is removed to eliminate well-known sources of radio frequency interference. A significance ($S = -\log_{10} p$, with p the false-alarm probability in Gaussian noise) is calculated at each grid point. After ~2 central processing unit hours, the host uploads the 100 most significant candidates to the server.

When all work units for a given beam are complete, the results are postprocessed on servers at Hannover. Candidates with $S > 15$ are identified by eye, then optimized with PRESTO (www.cv.nrao.edu/~sransom/presto/) (SOM). To date E@H has searched 27,000 of 68,000 observed beams. It has redetected 120 pulsars, most in the past 4 months, because code and algorithm optimizations sped up the search by a factor of ~7.

On 11 July, the 24-ms PSR J2007+2722 was discovered with a significance of $S = 169.7$ (Fig. 1) in survey data from February 2007. It was later re-detected in another PALFA survey observation. Follow-up observations were done by the Green Bank Telescope (GBT, United States), the Lovell Telescope at Jodrell Bank Observatory (United Kingdom), the radio telescope at Effelsberg (Germany), the Westerbork Synthesis Radio Telescope (WSRT, Netherlands), and Arecibo. The period-averaged flux density is 2.1 mJy (1 Jy = 10^{-26} W m^{-2} Hz^{-1}) at 1.5 GHz. Gridding observations using Arecibo and WSRT unambiguously associate the pulsar with a source in an archival Very Large Array (VLA) C-array observation, having 1.2 mJy flux density at 4.86 GHz, at right ascension (RA) $20^h07^m15^s.77$ and declination (Dec) 27°22′47″.7 (J2000) with uncertainty $\lesssim 1''$. The pulsar is not in a supernova remnant or globular cluster and has no counterpart in x-ray or gamma-ray point-source catalogs. The DM of 127 pc cm^{-3} implies a distance of 5.3 kpc (*3*). The full pulse width between the outer half-maxima is $W \approx 224°$. The wide pulse and initial polarization observations indicate that the pulsar likely has nearly aligned magnetic and spin axes.

The pulsar's barycentric spin frequency (*4*) is 40.820677620(6) Hz at MJD 55399.0. With the VLA position, the 2010 data give limits $|\dot{f}| < 3 \times 10^{-14}$ s^{-2}, magnetic field $B < 2.1 \times 10^{10}$ G, and spin-down age $> 21 \times 10^6$ years. These limits and lack of a companion mean that J2007+2722 is likely the fastest-spinning disrupted recycled pulsar yet found (*5*). However we cannot rule out it having been born with low B (*6*). Either way, PSR J2007+2722 is a rare, isolated low-B pulsar, which contributes to our understanding of pulsar evolution.

This result demonstrates the capability of "consumer" computational power for realizing discoveries in astronomy and other data-driven science.

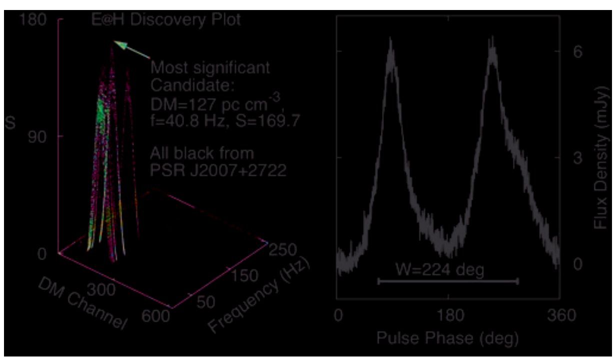


Fig. 1. (**Left**) Significance *S* as a function of DM and spin frequency (all E@H results for the discovery beam). (**Right**) The pulse profile at 1.5 GHz (GBT). The bar illustrates the extent of the pulse.

References and Notes

1. B. P. Abbott *et al.*, *Phys. Rev. D* **80**, 042003 (2009).
2. D. P. Anderson *et al.*, in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, Tampa, FL, 11 to 17 November 2006 (Association for Computing Machinery, New York, 2006), p. 126.
3. J. M. Cordes, T. J. W. Lazio, "NE2001. I. A new model for the Galactic distribution of free electrons and its fluctuations," http://arxiv.org/abs/astro-ph/0207156v3.
4. Results were obtained by using the Tempo software package (http://tempo.sourceforge.net/) and the Jet Propulsion Laboratory DE405 ephemeris.
5. K. Belczynski *et al.*, *Mon. Not. R. Astron. Soc.*, in press; available online at http://arxiv.org/abs/0907.3486v3.
6. J. P. Halpern, E. V. Gotthelf, *Astrophys. J.* **709**, 436 (2010).
7. We thank Einstein@Home volunteers, who made this discovery possible. The computers of C. and H. Colvin (Ames, Iowa, USA) and D. Gebhardt (Universität Mainz, Musikinformatik,Germany) identified J2007+2722 with the highest significance. This work was supported by Canada Foundation for Innovation, Canadian Institute for Advanced Research, fonds québécois de la recherche sur la nature et les technologies, Max Planck Gesellschaft, National Astronomy and Ionosphere Center, National Radio Astronomy Observatory, Natural Sciences and Engineering Research Council (of Canada), NSF, and Netherlands Organization for Scientific Research, Science and Technology Facilities Council; see the SOM for details.

*Affiliations are listed in the SOM.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1195253/DC1
Materials and Methods
SOM Text
References

16 July 2010; accepted 5 August 2010
Published online 12 August 2010;
10.1126/science.1195253
Include this information when citing this paper.

†To whom correspondence should be addressed. E-mail: benjamin.knispel@aei.mpg.de

Time line

41 authors from 21 Institutions

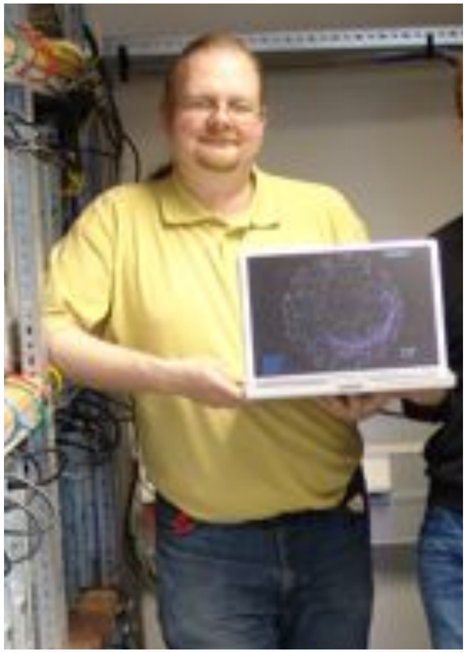
- July 11th, saw beamplot
- July 12/13th, J2007+2722 reconfirmed by Green Bank Telescope observations
- August 12th, online publication of the paper

Scientific Programming and Computing

Machenschalk and Bock

Helping “ordinary scientists” overcome the nightmare of modern large-scale scientific computing:

Maintenance of LSC software libraries, custom libraries, source-code repositories, build scripts, porting tools, E@H software infrastructure, LIGO/GEO/Virgo data access, running code in parallel on Atlas, specialized cluster data distribution, data-bases, data-base interface code, code optimization, GPU porting and optimization, specialized PHP and web-site infrastructure, ...



Computing Expertise

Since April, AEI Hannover has been proud to host Tejun Heo: the 8th-ranked Linux kernel developer world-wide.

Name	Number of Changes
David S. Miller	2,239
Ingo Molnar	2,125
Al Viro	1,981
Adrian Bunk	1,883
Takashi Iwai	1,801
Bartlomiej Zolnierkiewicz	1,651
Ralf Baechle	1,471
Tejun Heo	1,457
Stephen Hemminger	1,408
Andrew Morton	1,370

From "***Who Writes Linux***", The Linux Foundation, August 2009; Table 7. GIT commits counted over development from the 2.6.12 to the 2.6.30 kernel.

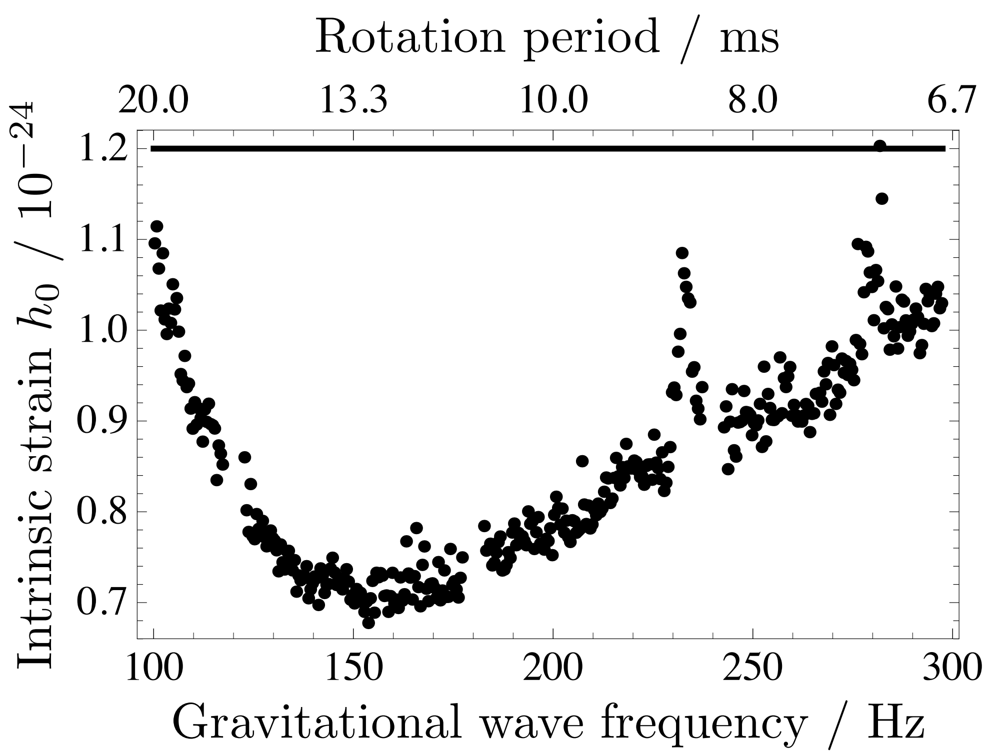
Conclusions

- Gravitational wave data analysis has a well-established "grid-like" software and hardware infrastructure, currently around 16,000 CPU cores + 30,000 Einstein@Home.
- Increases in detector sensitivity should lead to first detections around 2015-17
- Third-generation instruments (2020-2025) will leave to even larger data volumes, and harder analysis problems, but lots of fascinating science
- Because the astrophysical signals are the same, anbd computing needs are significant, there is great incentive to develop common/shared software and hardware systems for the analysis.
- Common need across our projects: high-quality distributed file systems!

Thank you for your attention!

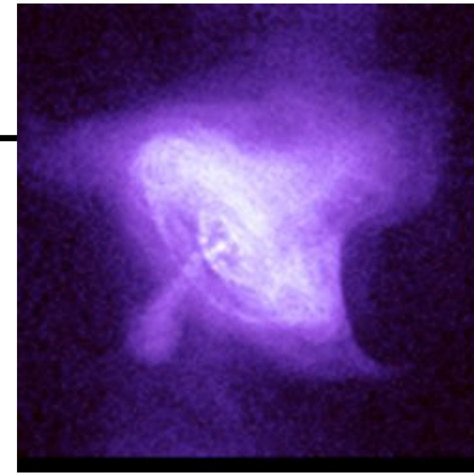
Search for GWs from the youngest neutron star

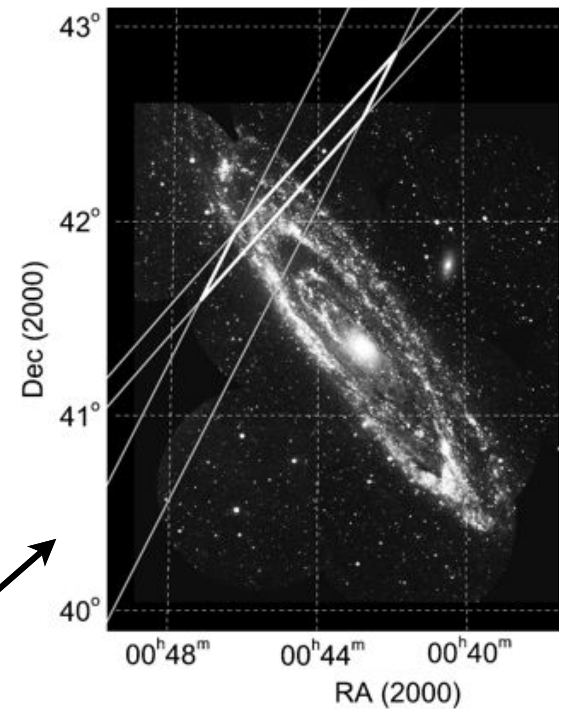
- Karl Wette (now a postdoc in the group) did the central part of his PhD work here: a directed search for continuous gravitational waves from the brightest radio source in the sky, the supernova remenant Casseopeia A (Cass A) which contains the youngest-known neutron star (born 1681 ± 19).
- Beats inferred spin-down limit
- Paper just accepted to the Astrophysical Journal
- Perfect example of how the combination of computing facilities, people, and knowledge made it possible to pull an analysis together.



Analyzing Detector Noise

- So far, the LIGO Scientific Collaboration has **published about 50 scientific journal papers**, and **detected exactly 0 signals**. Nevertheless some interesting results:
  - **Stochastic Background** (**Nature 460, 2009, p990**) correlating LIGO L1/H1 gives a limit $\Omega_{gw} < 6.9 \times 10^{-6}$ for 40-170Hz. This beats best existing limit ($\Omega_{gw} < 1.1 \times 10^{-5}$ from Big Bang Nucleosynthesis).
  - Crab Pulsar (Ap J. Lett 683, 2009, L45) less than 4-6% of the energy-loss from spin-down is going into gravitational waves.
  - **Search for gravitational waves from 116 known pulsars** (**Astrophys. J. 713, 2010, p671**) shows < 2.3 x 10^{-26} for J1603-7202 and $\varepsilon < 7 \times 10^{-8}$ for J2124-3358. **Crab pular**: GWs less than < 2% of EM.
  - **GRB 070201** (short, hard gamma-ray burst) **was NOT a binary inspiral in M31** (**ApJ 681, 2008, p1419**). Most likely an SGR giant flare in M31.



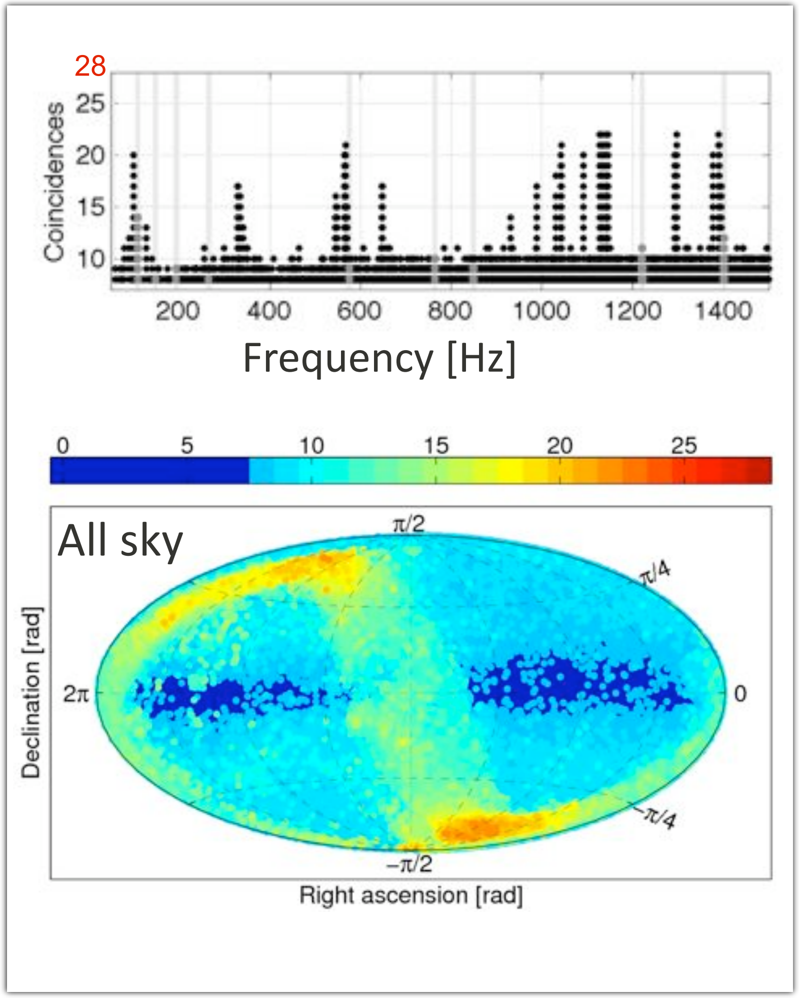


Inter-Planetary Network 3-sigma error region from Mazets et al., ApJ 680, 2008, p545

E@H all-sky blind search of early S5 data

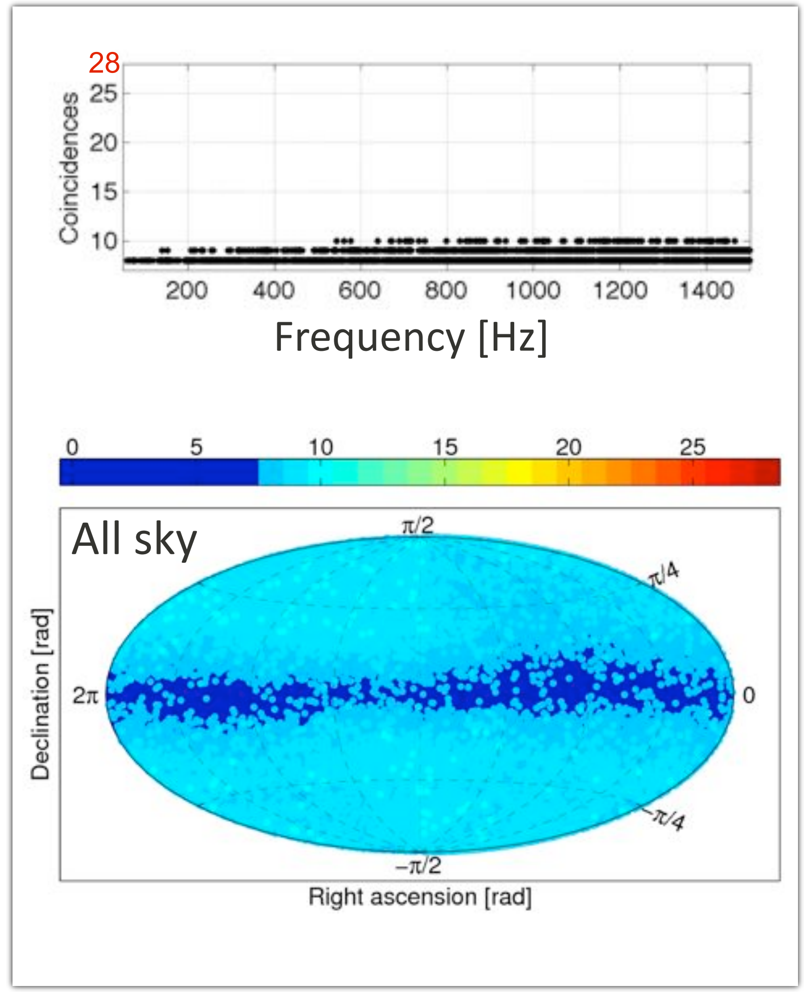
Coincidence analysis results:

The LSC, Phys. Rev. **D80**, 042003 (2009)



Further post-processing:

- Remove known instrumental noise.
- Apply parameter-space veto.



Final results:

» **No statistically significant candidate**

"It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it's wrong."

Richard P. Feynman