Science & Integration Meeting

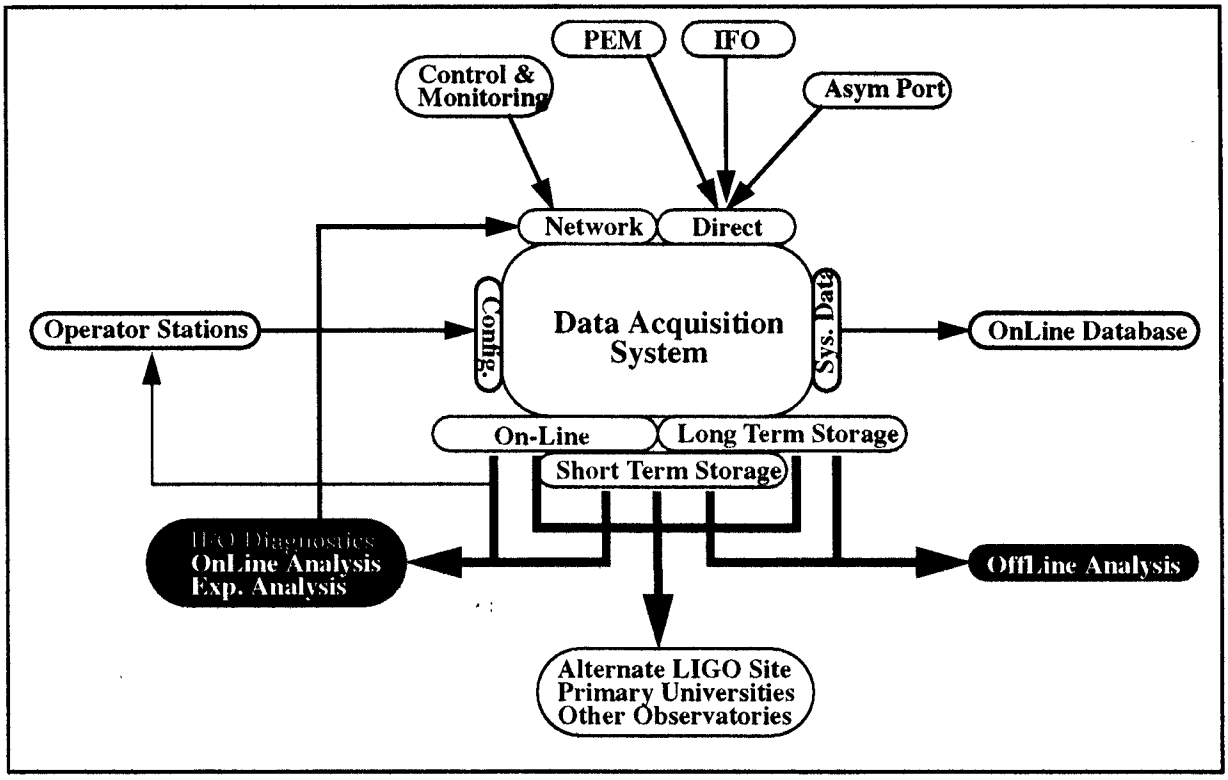
Agenda

- Information Management
  - ›› Electronic submission to DCC — Althouse
- LIGO Data Analysis
  - ›› Report from the MIT Data Analysis Meeting — Weiss
  - ›› Prototyping data analysis at CACR — Blackburn
  - ›› Forum to review baseline approach proposed to NSF — Lazzarini

G960277-00-E
LIGO-G950000-00-M

Data Analysis System for the Initial LIGO

- CDS DAQ Design - May 1996



DAQS Interfaces

- On line analysis & diagnostics are closely related

Data Analysis System for the Initial LIGO

- CDS DAQ provides for on-line short term storage (ref.: CDS DAQ CDR, DRD)
  - 63GB (commissioning) -> 400 GB (later)
  - 8 hr complete data frames (CDF); 10 days limited data frames (LDF)
  - 25MB/s I/O
- Data analysis system will use same resources and augment them as needed

Data Analysis System for the Initial LIGO

- On-line analysis
  - ›› Diagnostics - ensure instrumental sensitivity at all times
    - performance metrics -- Gaussian noise
      - $h_{rms}[t]$
      - statistics of h in limited frequency bands
    - performance metrics - non-Gaussian noise
      - templates (limited range: $m_{NS} < m_{Sun}$)
      - wire resonances -- $\delta A[t]$ - $\delta\phi\,[t]$ phasor diagram
      - PEM (vetos)
    - calibrations - lines/broadband - h[t] extraction, if possible
    - "quick-look" analyses - real time

Data Analysis System for the Initial LIGO

- On-line analysis
  - ›› Astrophysics
    - detection of transient phenomena for which coincident operation with other (highly) directional instruments is feasible
    - Supernovae
      - limited to Milky Way -- rare (.025/yr)
      - ν detectors (omnidirectional) - can be done with post processing correlation
      - γ /visible light curves - want to track curve from onset of SN explosion
        - LIGO could provide trigger;
        - 2 sites give cone -- insufficient information;
        - telescopes slewed to position;
        - requires site-to-site communication correlation in real time
      - maximum few hour latency allowed - TBD
      - Supernova signature is of short duration
      - No templates -- wavelet characterization; datastream cross-correlation

Data Analysis System for the Initial LIGO

- On-line analysis
  - ›› Astrophysics
    - — NS/NS coalescence
      - LIGO detection would not be able to affect other directional detectors
        - End point may produce fast/short burst of EM radiation, but duration does not allow slewing of directional instruments
        - Template analysis has built-in latency (collection time)
      - EM detection (γ burst) can always be correlated off-line since LIGO is omnidirectional
      - On-line templates will be needed to characterize non-Gaussian noise performance
  - ›› Delaying data reduction exacerbates problem later!

Data Analysis System for the Initial LIGO

- Off-line analysis -- must keep up with data stream
  - "Production" of data -- how to deal with 500 TB/yr problem
    - consolidation/refinement of data for permanent archive
    - data product generation -> h[t] for scientific analysis
    - "quick-look": 1 day -> 1 month
    - "Data QA Board"
  - Scientific analysis
    - multiple analyses possible simultaneously
    - substantial resources required
      - 10 - 50 GFLOPS
      - 500+ GB disk farm(s) for tape dumps
      - SCC-class systems
    - computational resources resident with archive
    - network (wide area) access
    - some work may require additional SCC resources (SDSC, NCSA, NPAC, etc.)

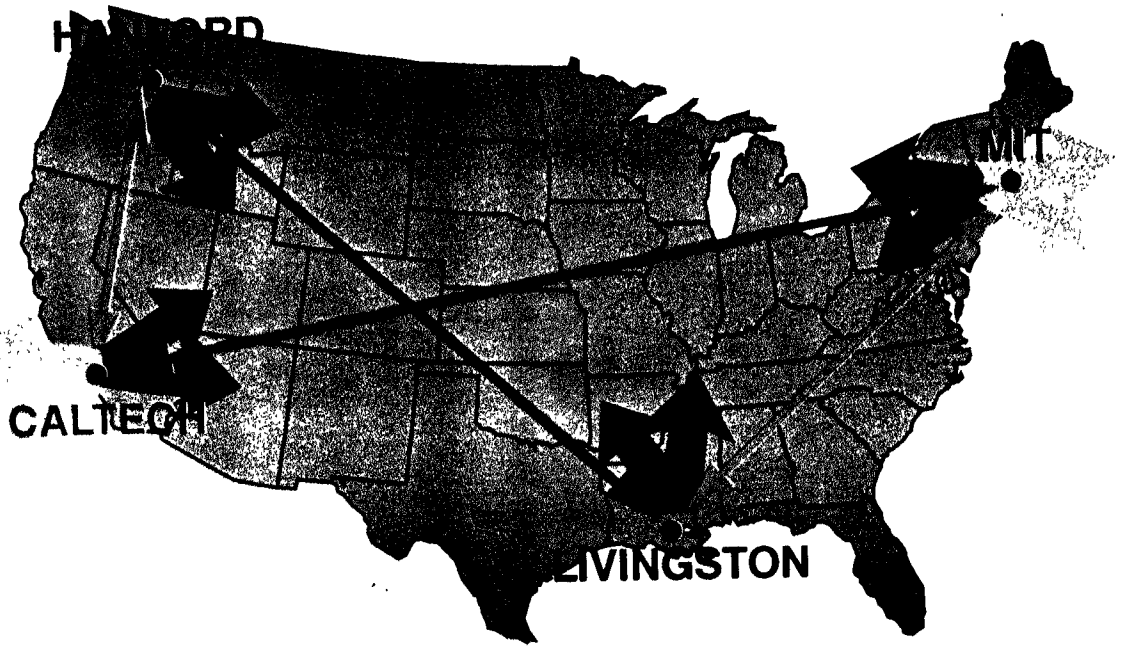
Data Analysis System for the Initial LIGO

- Data "QA" Board
  - regular (frequent) meetings to review detector performance and statistics of data from previous period
    - % locked/up time
    - $h_{rms}[t]$ statistics
    - event lists (template filters)
  - responsibility is to produce triage on archived data
    - what to keep
    - what to recycle (keep limited data frames -- GW channel) -- annouce to collaborations: if someone wants 100% data, he/she can retrieve it, store it, etc.

LIGO-G950000-00-M

Data Analysis System for the Initial LIGO

- Networks



›› Hanford-Livingston link permits real-time cross-correlations among instruments

›› Caltech-MIT link provides high speed link to data archives; data tapes to be archived at university.

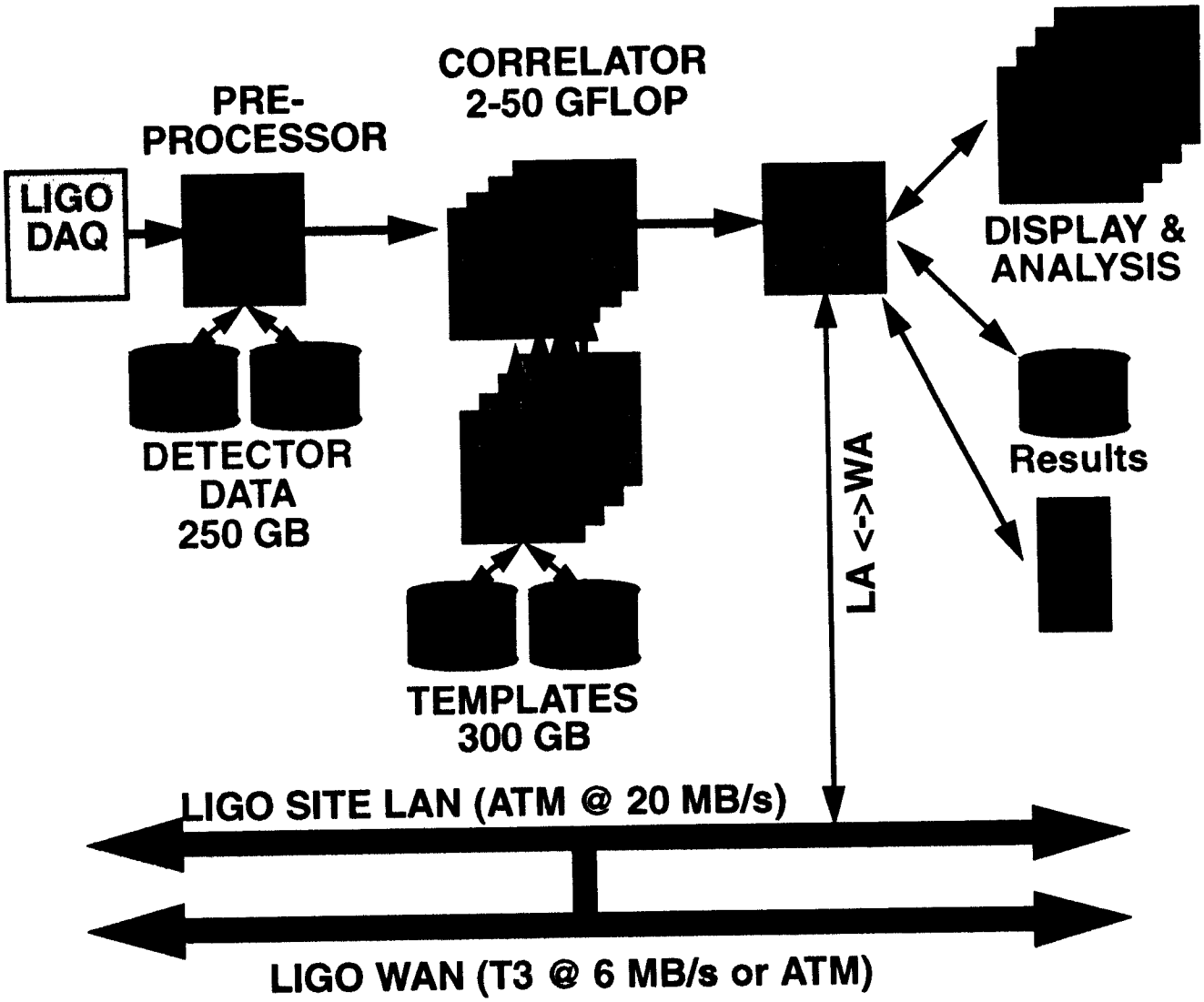
›› Site-University links provides site scientific staff access to archived data

›› University gateways provide broader access to database

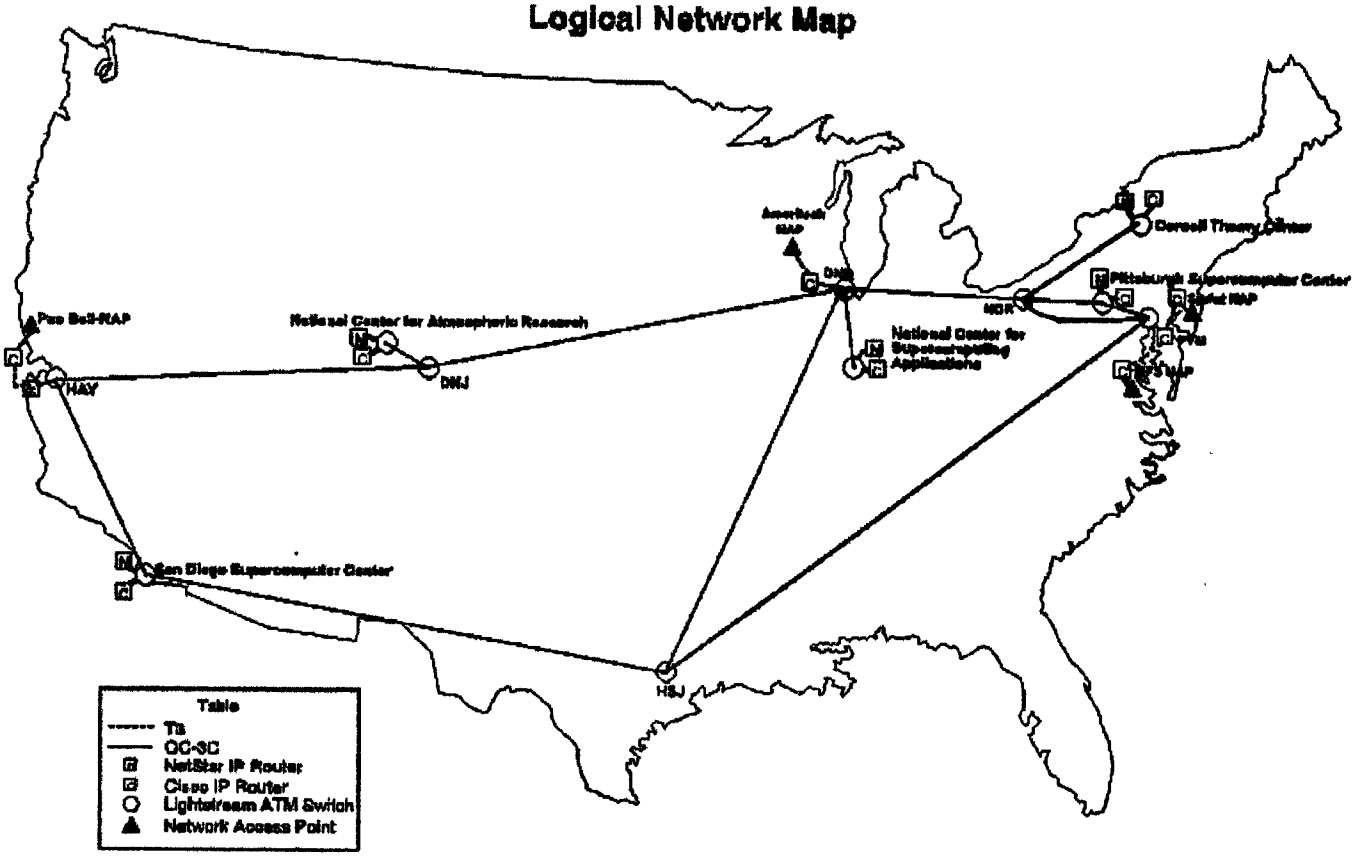
›› Data tapes transported to University repository

Data Analysis for Initial LIGO

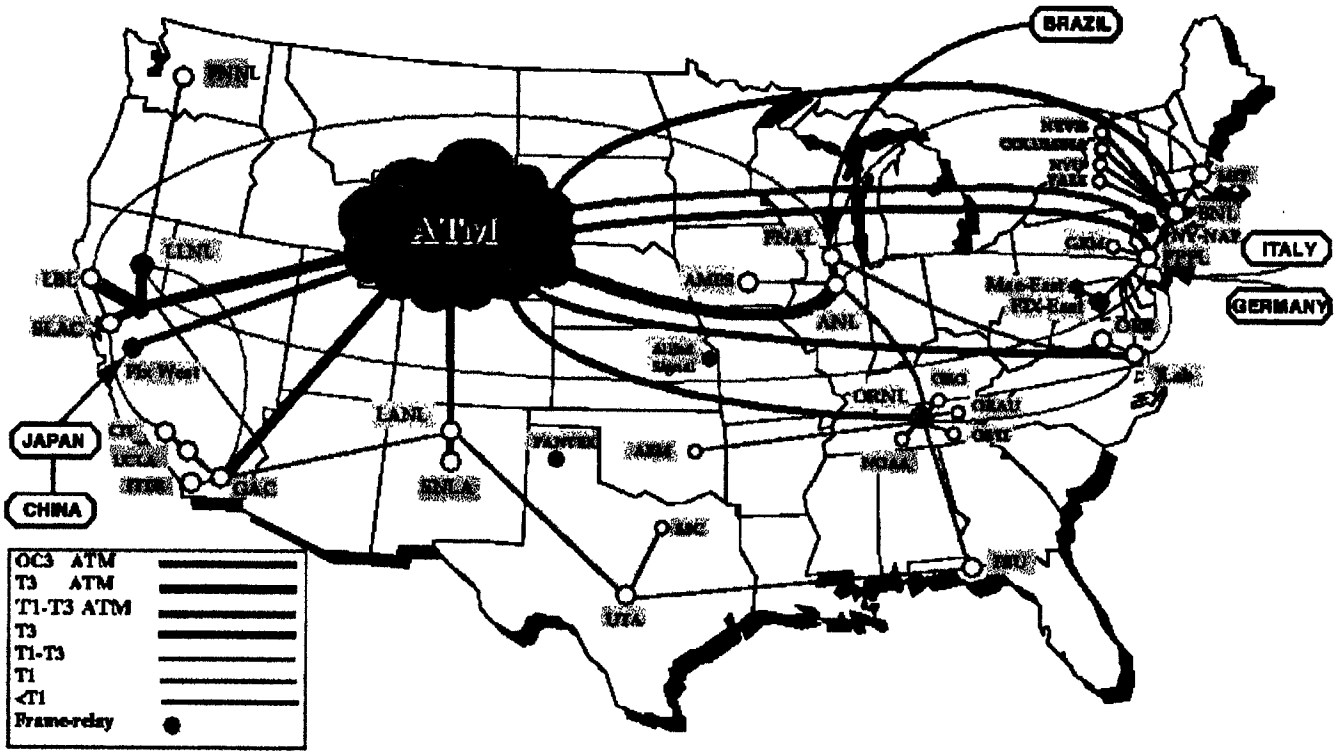
On-line Processing Computing Resources & Distribution



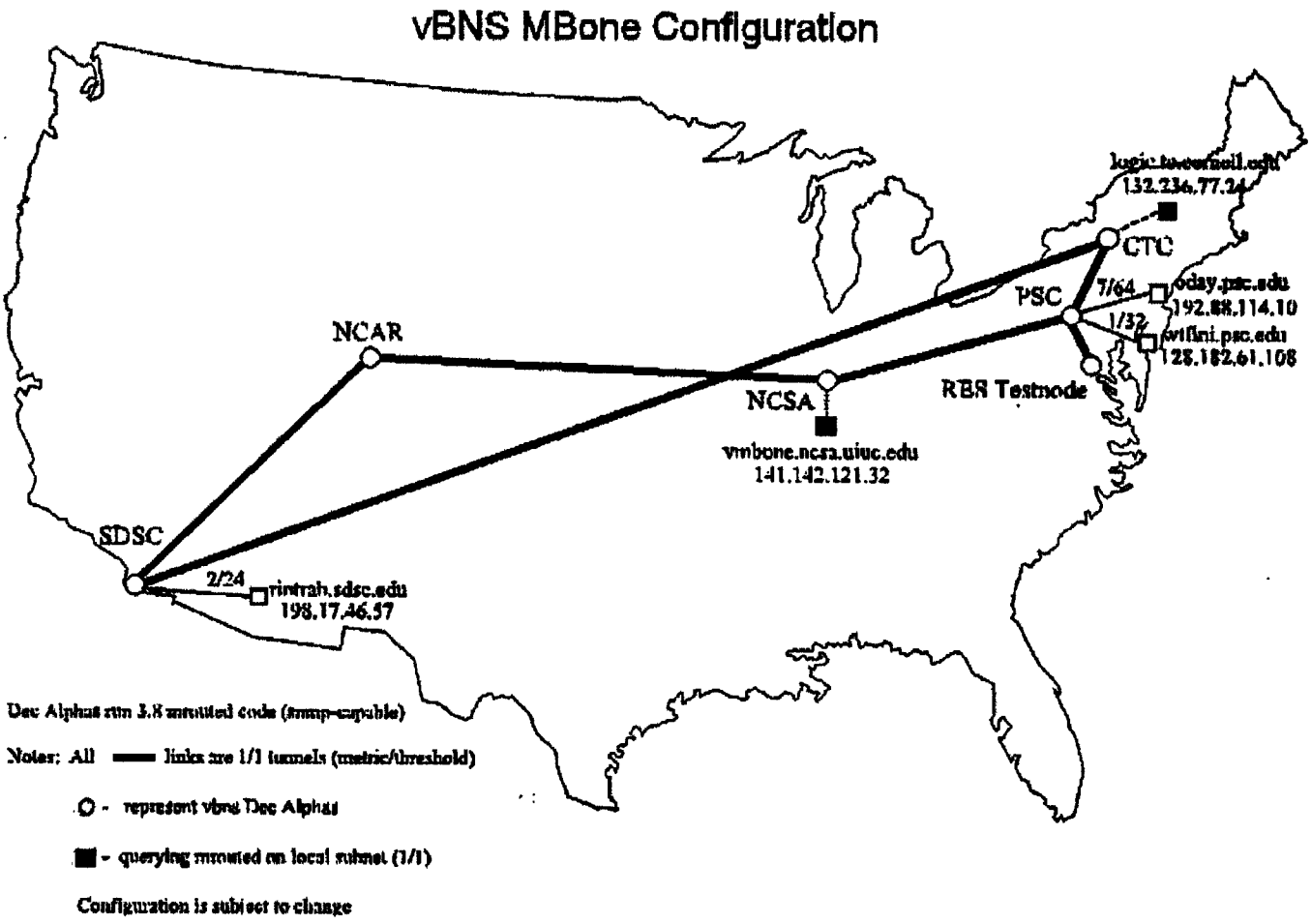
The National Solence Foundation Very-High-Speed Backbone Network Service
Logical Network Map



ESnet BACKBONE
Mid 1996



vBNS MBONE MAP



Data Analysis System for the Initial LIGO

- Software
  - Layered, standardized, stylized, documented (users' manuals,etc)
  - I/O libraries to access archived frames
  - concatenation of frames
  - extraction of specific channels
  - cross-correlation among channels
  - diagnostic software - time and frequency domain
  - analysis software - filtering algorithms
  - 2D/3D display & visualization
    - AVS (=GRID) or similar varients
      - IBM's DATA EXPLORER
      - SGI IRIX EXPLORER

LIGO-G950000-00-M

Data Analysis System for the Initial LIGO

- What needs to be done?
  - by 5/97:
    - Data Analysis System Requirements
      - Hardware - location/storage/performance
      - Communications - WAN requirements; bandwidth; up time; connectivity
      - Software - functional requirements/specifications/standards
    - Conceptual Design
      - Hardware architecture
      - WAN architecture
      - Software module specification; user environment; implementation approach

Planned Activities
Timeline for Development

Milestone or Event	Date	Communications	Hardware	Software
Begin Coincidence Operations	7/00	Common		
On-Line System Available	1/00	Common		
	3/99-12/99	Definition, Agreements, Implementation	Procurement & Integration	Development, Verification
	11/98		Specifications	Specifications
System FDR	11/98		Design & Prototyping	Design & Prototyping
System PDR	11/97			
System DRR	5/97			

Data Analysis Requirements

Science & Computational Requirements

Initial LIGO Sources and Estimated Analysis Capability Requirements

	Sources	Initial LIGO Performance Estimate	Data Analysis Requirements		
			CPU	Storage	Comments
Burst Signals $\Delta T < 1s$	Supernovae	$\Re_0 \sim 2-3/$ yr @ 15 Mpc If sufficiently asymmetric	Minimal for straightforward correlation; *if optimal filters are discovered, problem may increase in complexity.*	Minimal Need PEM/houskeeping data for veto	• *On-line analysis* desirable for correlation with other astrophysics: **Electroweak** • visible/radio/γ (HETE. GRO) • ν (Super-K/SNO) **Gravity** • VIRGO/GEO • Resonant bars • Waveforms unknown • 2x/3x IFO correlation • Off-line analysis to enhance SNR
	BH/BH Collisions	$\Re_0 \sim 1/$ yr(?) @ 500 Mpc; $M_{BH} \sim 30 - 200 M_{SUN}$			
Chirped Waveform $10s < \Delta T < 1000s$	NS/NS Inspirals	$\Re_0 \sim 3/$ yr @ 23 Mpc; $\Delta T \sim 4 \times 60$ s $M_{NS} \sim M_{SUN}$ $\Delta T \sim 4 \times 500$ s $M_{NS} \sim 0.3 M_{SUN}$	Templates/Data ~ 2 GFLOPS ~ 50 GFLOPS	Templates/Data ~20 GB /~1 GB ~500 GB /~10 GB	• *On-line analysis* for $M_{NS} > M_{SUN}$ can be done; appears feasible down to ~ 0.3 M_{SUN} • 2x/3x correlations feasible depending on SNR. • Coalescence event may generate correlated (EW) signals as above. • PEM/housekeeping needed for vetoing • Template matching (Wiener filtering) or wavelet analysis in f-t domain. • Off-line analysis to enhance SNR
	BH/BH Inspirals	$\Re_0 \sim 1/$ yr @ 150 Mpc; $\Delta T \sim 4 \times 10$ s $M_{NS} \sim 10 M_{SUN}$;	~ 2 GFLOPS	~20 GB /~1 GB	

Data Analysis Requirements

Science & Computational Requirements

Initial LIGO Sources and Estimated Analysis Capability Requirements

	Sources	Initial LIGO Performance Estimate	Data Analysis Requirements		
			CPU	Storage	Comments
Periodic Signal $\Delta T \sim 10^6 - 10^7$ s	Pulsars with mass asymmetry $h \propto \left(\frac{\varepsilon}{10^{-6}}\right)\left(\frac{10\text{kpc}}{r}\right)\left(\frac{1\text{ms}}{P}\right)^2$	$\varepsilon = 3\times10^{-5}$; r=10kpc ; P=1ms $T_{int} = 10^6$s SNR ≈ 5	Directed searches (e.g., galactic center, known pulsars) require minimal resources All-sky searches require tens of TFLOPS -- beyond anticipated capabilities	10 GB for 10^6s (GW waveform)	• *Off-line analysis* • Detection less sensitive to non-Gaussian noise; more sensitive to calibration drifts&drop-outs • Detection techniques as for pulsars -- narrow line sources with modulated frequency. • Correlations among interferometers may be performed (if needed) after detection. • All-sky search requires decomposition of 4π sr into $>10^{10}$ pixels, each region requiring a different spectral transformation of same dataset.
Broadband Signals $\Delta T \sim 10^6 - 10^7$ s	Stochastic Background $\Omega \equiv \frac{\Omega_g}{\Omega_0}$	$\Omega \geq 3\times10^{-6}$ $\Delta f, f \approx 100$Hz $T_{int} = 10^7$sec	Minimal requirements -- analysis maybe done on single workstations		• *Off-line analysis* • Requires multiple interferometers to be correlated; may use PEM to imprive SNR.

CSCC HARDWARE

FEATURE	DELTA	TREX	RAPTOR	JPL CRAY
MODEL	PROTOTYPE	XPS L38	XPS A4	T3D
GFLOPS	30.7	38.4	4.3	38.4
NODES, PE's PER NODE	513, 1 PE	512, 1 PE & 1 Comms Procsr.	57, 1 PE & 1 Comms Procsr.	128, 2PE's
CPU	i860 XR	i860 XP	i860 XP	DEC 21064
SPEED	40 MHZ	50 MHZ	50 MHZ	150 MHZ
MFLOPS/CPU	60	75	75	150
MB/NODE	16	32	32	64
TOTAL GB	8.2	16.4	1.8	16.4
DISKS IN GB's	93 (RAID0)	67.2 (RAID3)	14.4 (RAID3)	103
TOPOLOGY	2D (16X36)	2D (16X36)	2D (16X4)	3D TORUS