LIGO Data Analysis System (LDAS)

PAC 11 Meeting

California Institute of Technology
November 29, 2001

James Kent Blackburn
LIGO Laboratory, Caltech

LIGO-G010404-00-E

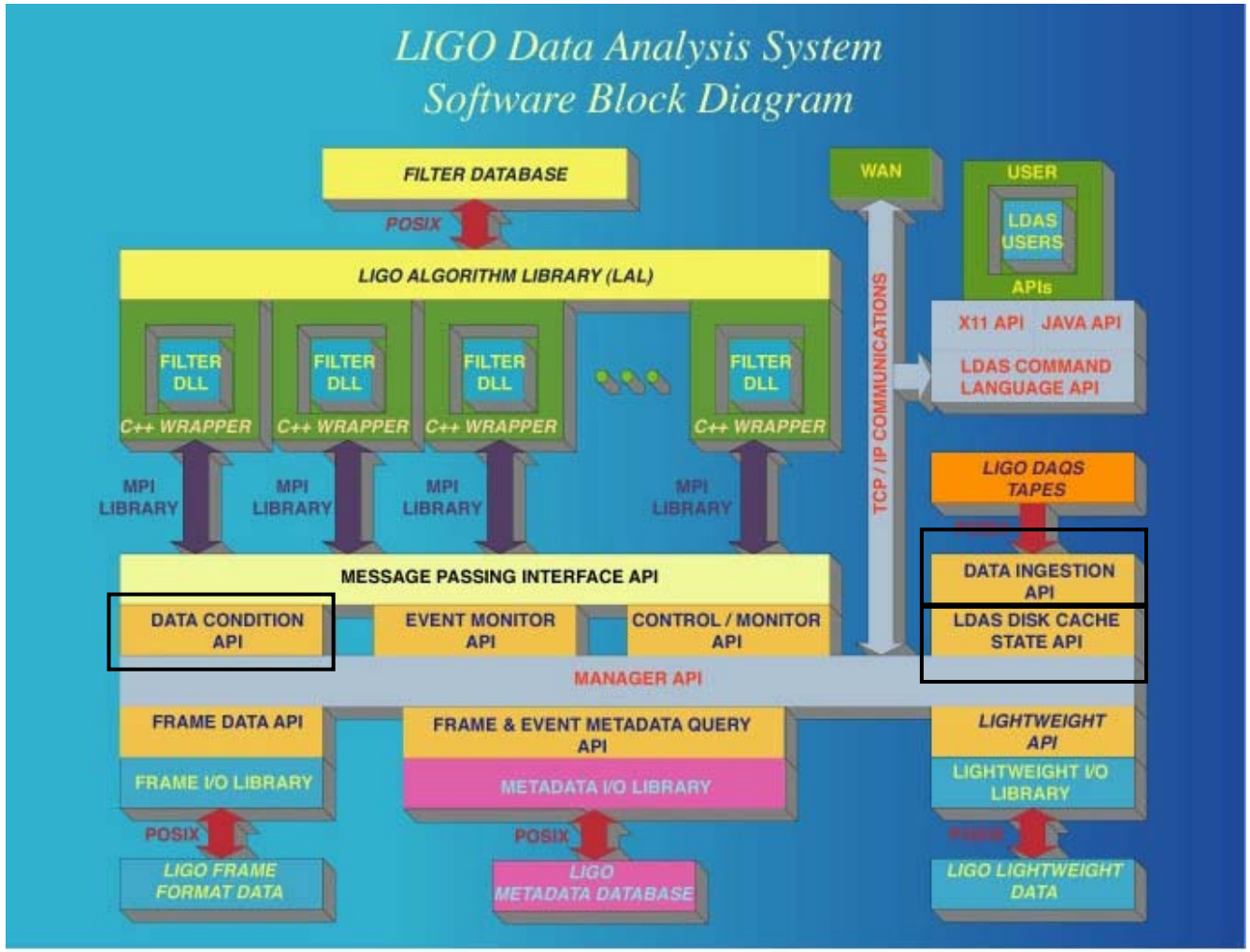
Three remaining modules under active development

Current release of LDAS: LDAS-0.0.22 (*alpha*)

dataConditionAPI:
- First module to be tested in an MDC.
- Used extensively in later MDCs and Engineering Runs.
- Infrastructure nearly complete, only support for state information and cleanup of metadata remain.
- Still plenty of signal processing functionality to be implemented.

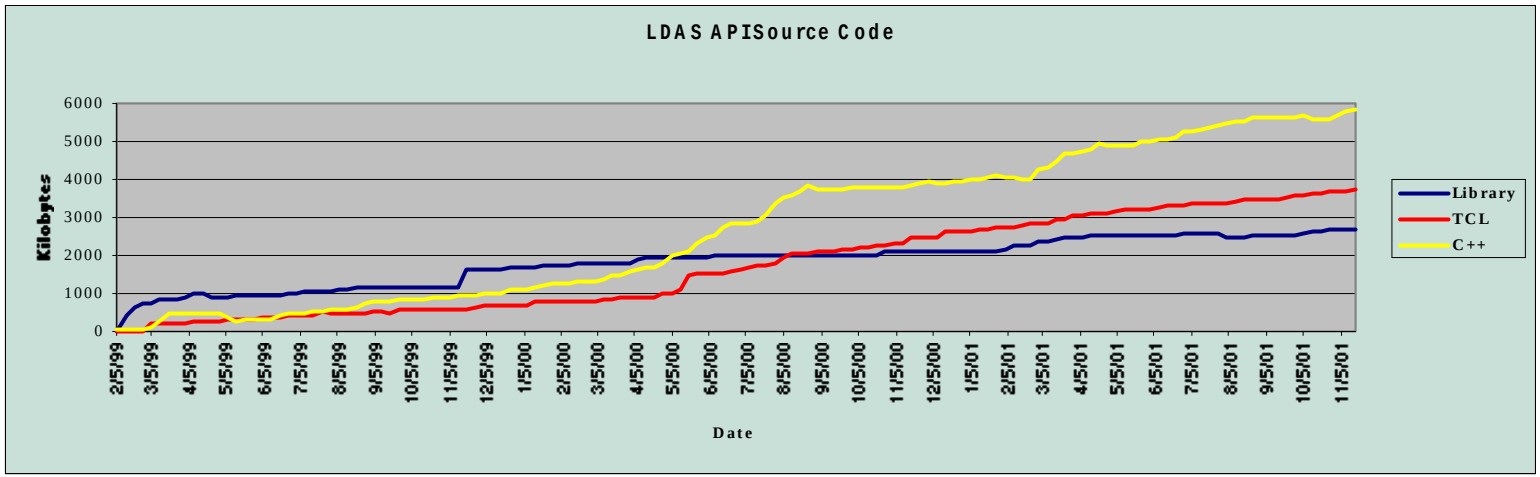
dataIngestionAPI:
- Basic functions needed with the beginning of the Engineering Runs.
- Scripts external to LDAS developed to carry out these functions.
- Script to recording data to tapes at sites can be controlled and monitored using LDAS's controlMonitorAPI.

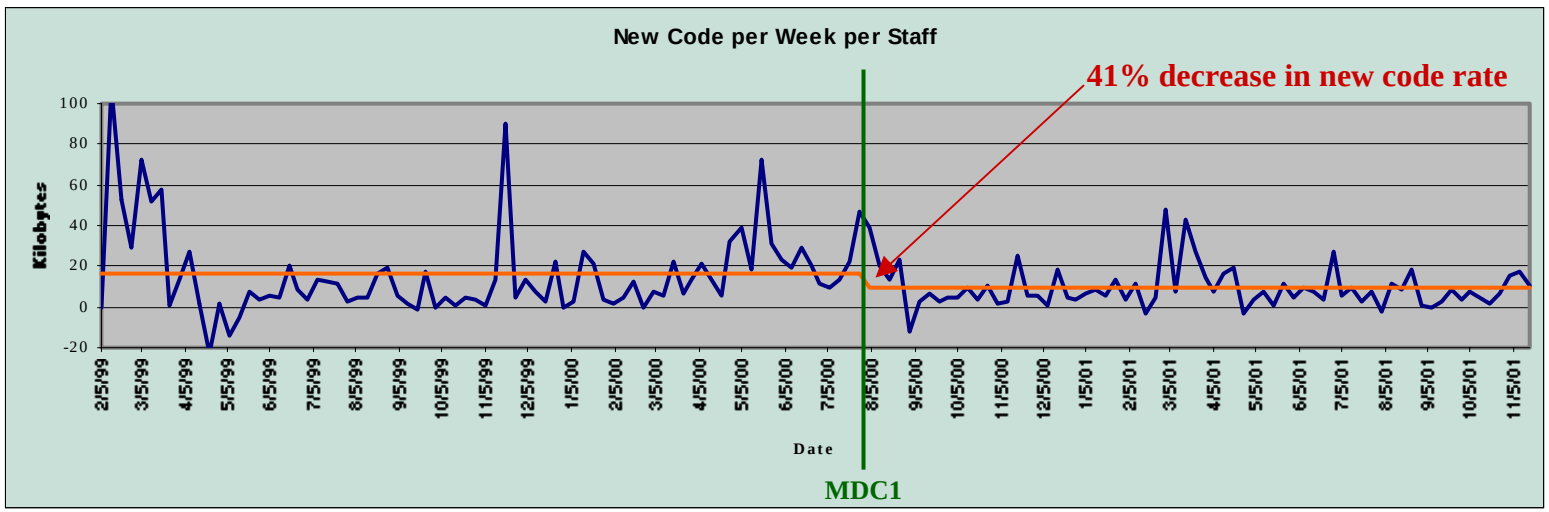
diskCacheAPI:
- Coupled to selection of archival data storage technology(ies) to be used.
- Basic functions needed in support of MDCs and Engineering; Temporarily implemented in the frameAPI!
- Anticipate removing from frameAPI later this year.
- First "beta" release of LDAS after this API is implemented.



Code Development Trends

(more LSC support & testing!)





LIGO

Problem Tracking

category	open	analyzed	suspended	feedback	closed		Total
build process	3	0	0	0	85		88
controlMonitorAPI	0	0	0	0	46		46
data archive	0	0	0	0	1		1
dataConditionAPI	16	3	8	1	176		204
dataIngestion	0	0	0	0	0		0
dbaccess	0	0	0	0	4		4
diskCacheAPI	0	0	0	0	0		0
distribution	1	0	0	0	1		2
documentation	13	0	0	0	95		108
eventMonitorAPI	1	0	0	0	16		17
frameAPI	11	0	0	0	105		116
frameCPP	1	0	0	0	12		13
general library	1	0	0	0	1		2
genericAPI	4	0	1	0	95		100
ilwd	0	0	0	0	25		25
ilwdfcs	0	0	0	0	0		0
integration tests	0	0	0	0	5		5
lightWeightAPI	1	0	0	0	47		48
managerAPI	8	3	0	0	72		83
MDC ready	0	0	0	0	3		3
MDC run	0	0	0	0	0		0
metaDataAPi	1	1	0	0	68		70
mime	0	0	0	0	0		0
mpiAPI	3	0	0	0	33		36
remoteAPI	0	0	0	0	0		0
SWIG	0	0	0	0	1		1
sys admin	4	5	0	0	206		215
userAPI	0	0	0	0	8		8
wrapperAPI	0	0	0	0	14		14
Totals:	**68**	**12**	**9**	**1**	**1119**		**1209**

93%

LDAS systems up and running around the world!

LIGO Laboratory:

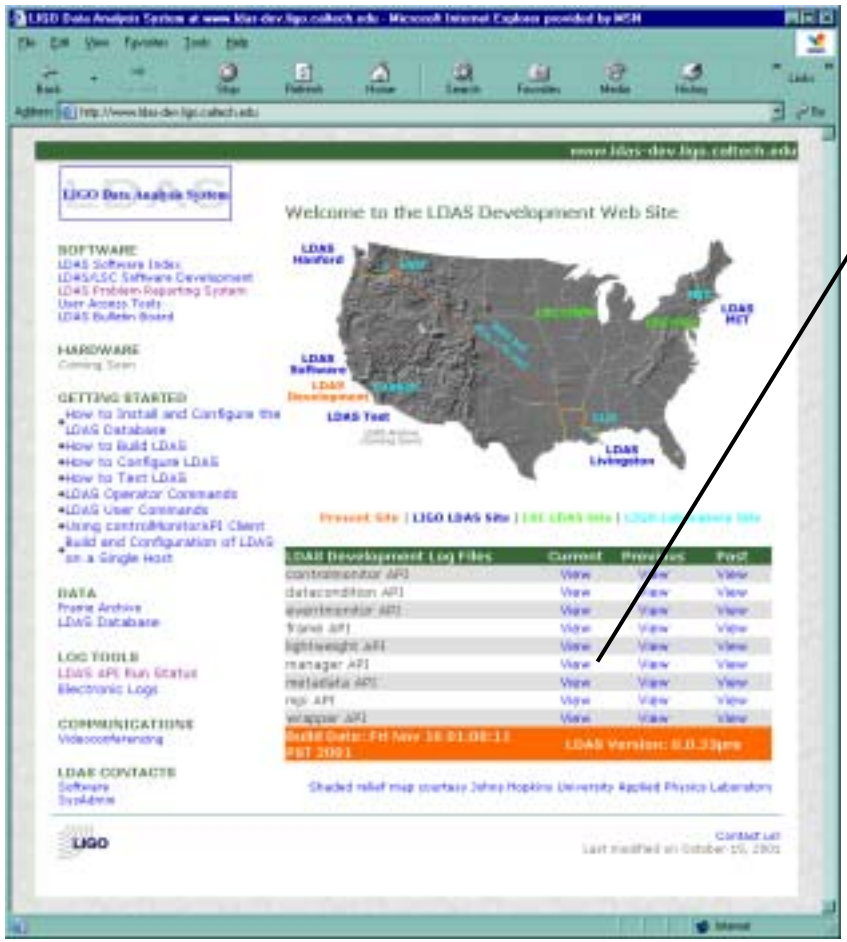
- Caltech
  - Development system
  - Test system
  - *Main system soon!*
- Hanford
- Livingston
- MIT
- *LDAS staff work closely with others setting up LDAS systems.*

LSC Institutions:

- University of Wisconsin Milwaukee†
- Penn State University†
- University of Texas Brownsville†
- Australia National University†
- University of Florida

† Sites with LDAS databases (*IBM's DB2*)

System web interface provides window onto *all* sites



LIGO

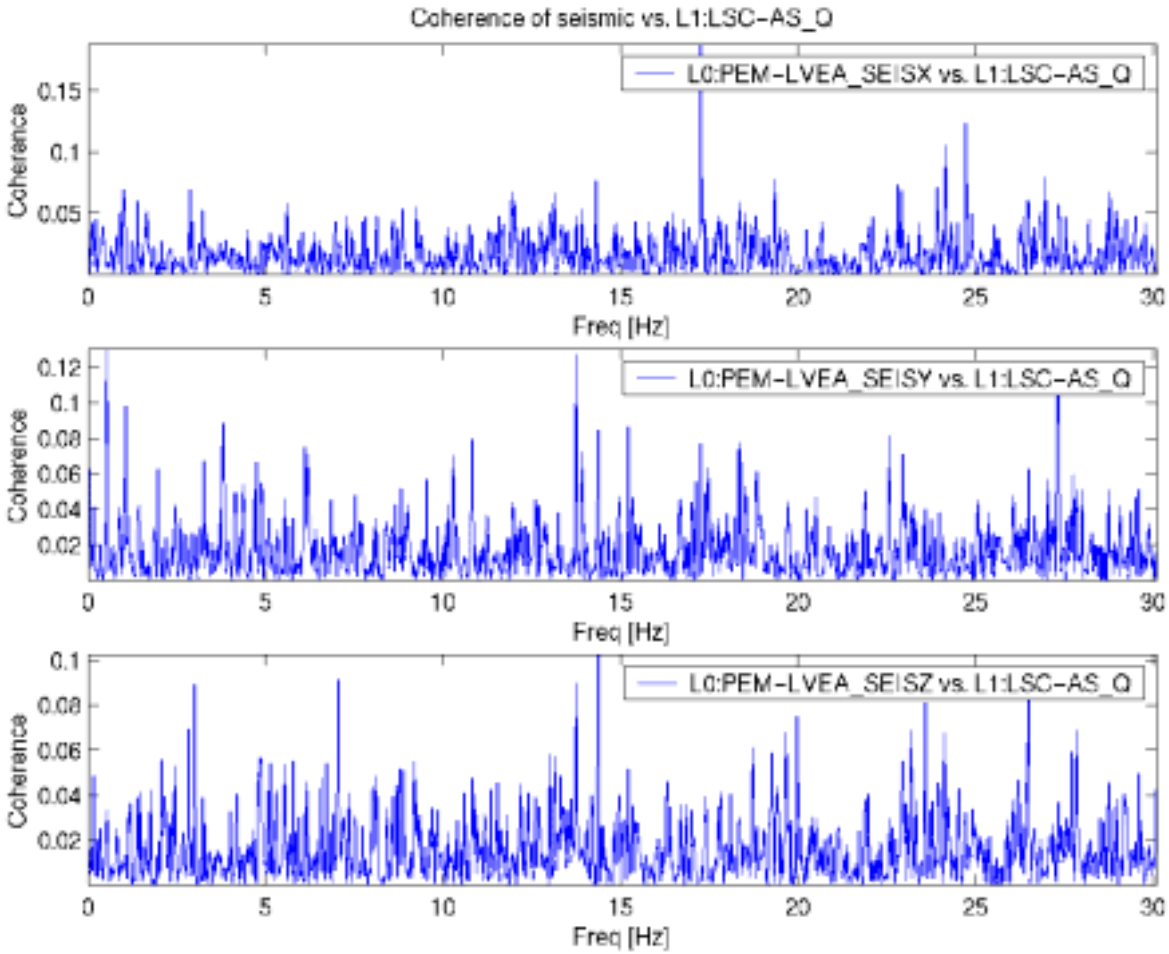
Participation with the LSC

- **Weekly dataConditioning Group**
  (CIT, ANU, PSU, UTB, UF, MIT)
- **Weekly (Message Passing Interface) MPI Group**
  (CIT, UWM, UTB, LHO)
- **Participate in LSC Software Coordination Meetings**
  (AL, KB, SA)
- **LDAS staff active in LSC Upper Limits Search Groups**
  (AL, KB, SA, PS, PC, LW, MB, PE)
- **Participate in Upper Limits Chairs Meetings**
  (AL, KB, SA, PS)
- **Participate in Detector Characterization Group**
  (AL, PS, KB)
- **Also active in Non-LSC specific areas**
  (gravitational wave network analysis, GriPhyn, and LIGO's own CDS/GDS)
- **Joint LSC/LDAS tutorial on developing analysis codes for the *Diagnostic Monitoring Tool, wrapperAPI* and *dataConditionAPI* last June at Caltech.**

Participation in Engineering Runs

- **E1**: (***ldas-0_0_10***) supported ingestion of triggers from Hanford Diagnostic Monitor Tool (DMT) using LDAS Commands.
- **E2**: (***ldas-0_0_12***) supported ingestion of triggers(~$7x10^5$) from LHO DMT and archiving of frames to tape.
- **E3**: (***ldas-0_0_15***) supported ingestion of triggers from LHO and LLO DMTs and archiving of frames to tape …
  - E3 release allowed frame channel data to be analyzed using signal processing tools to dataConditionAPI and stored in database
    - Power spectral densities & Cross spectral densities
- **E4**: (***ldas-0_0_16***) supported ingestion of triggers from DMTs and archiving of frames to tape
- **E5**: (***ldas-0_0_19***) supported ingestion of triggers from DMTs and archiving of frames to tape.
- **E6**: (***ldas-0_0_22***) supported ingestion of triggers from DMTs and archiving of frames to tape…
  - LDAS/LSC driver script polled database for locked segments, then submitted dataPipeline jobs to conduct an unmodeled burst search:
    - 95 segments of 180 second long each analyzed to produce 218933 burst events in database!

Cross Spectral Densities generated online with LDAS in E3 run



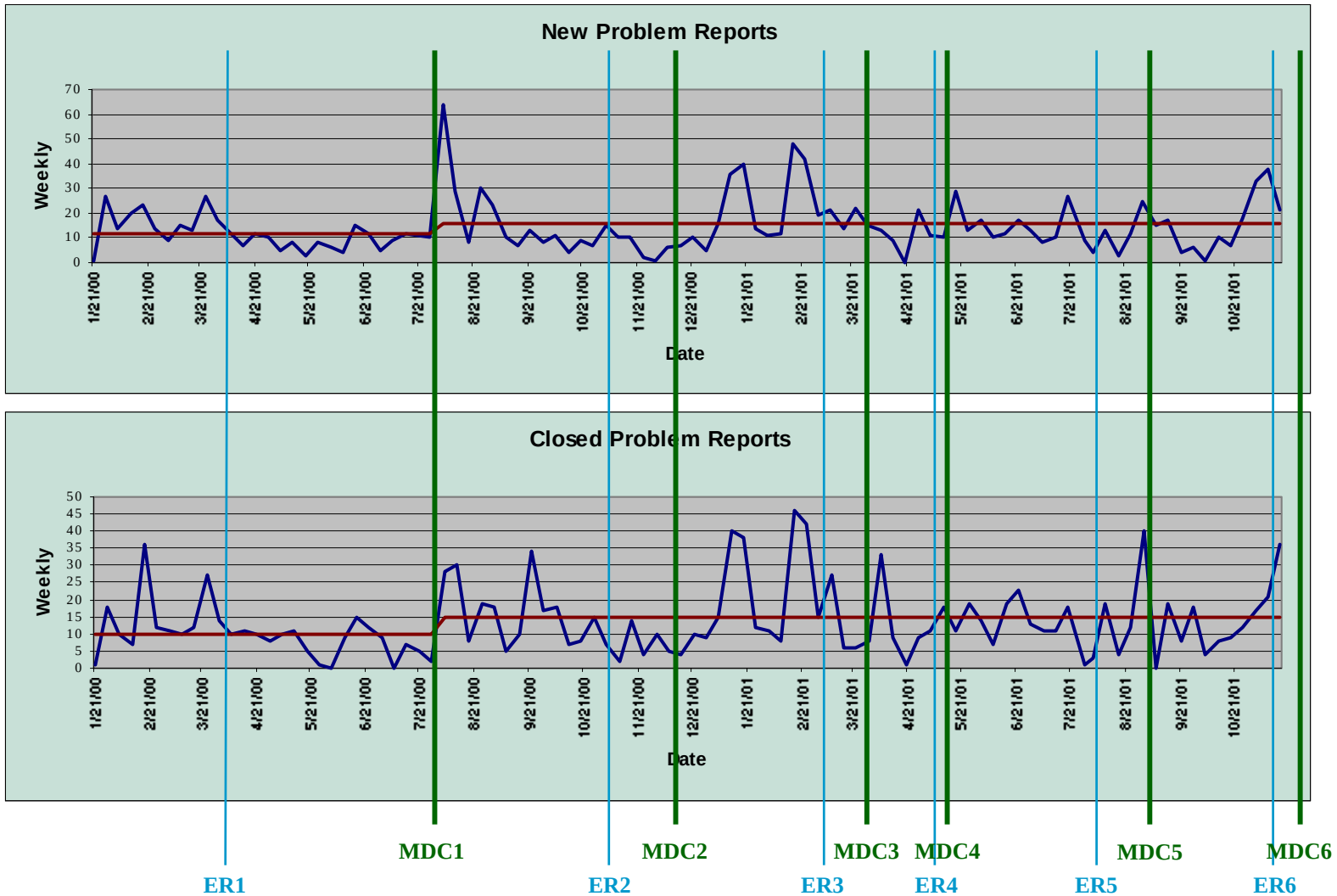
Coherence between single arm length signal & difference between end & vertex station seismic channels

Participation in MDCs

(*mock data challenges*)

- dataConditionAPI MDC1: August 2000, release (***ldas-0_0_10***)
  - Tested managerAPI & *dataConditionAPI*.
- Message Passing Interface MDC2: December 2000, release (***ldas-0_0_12***)
  - Tested managerAPI, mpiAPI, wrapperAPI, & (LSC's: LAL and LALwrapper).
- Database MDC3: January 2001, release (***ldas-0_0_12 - ldas-0_0_15***)
  - Tested managerAPI, metaDataAPI, lightWeightAPI, & database table design.
- Inspiral MDC4: May 2001, release (***ldas-0_0_17***)
  - Tested managerAPI, frameAPI, dataConditionAPI, mpiAPI, wrapperAPI, eventMonitorAPI, metaDataAPI & (LSC's: inspiral dynamically loaded shared object (dso) search codes).
  - Data products inserted into database and into ilwd data objects.
- Joint burst/stochastic MDC5: September 2001, release (***ldas-0_0_20***)
  - Tested managerAPI, frameAPI, dataConditionAPI, mpiAPI, wrapperAPI, eventMonitorAPI, metaDataAPI, lightWeightAPI & (LSC's burst and stochastic dso search codes).
  - Data products inserted into database, frame files and ligo_lw(xml) data objects.
- Periodic MDC6: November 2001, release (***ldas-0_0_22***)
  - Tested managerAPI, frameAPI, dataConditionAPI, mpiAPI, wrapperAPI, eventMonitorAPI, metaDataAPI, lightWeightAPI & (LSC's periodic source dso search codes).
  - Data products inserted into database, frame files and ligo_lw(xml) data objects.

MDC influence on problem reports



LIGO

LIGO and GriPhyN

- LDAS working closely with GriPhyN collaboration on LIGO related components.
- Current prototype combines several LDAS systems using GLOBUS components to carry forward user requests.
- First prototype demonstrated at SC 2001 Conference.
  - For each request for data, determined if it already exists and if so where, if not how to get it using LDAS.
  - Data moved between LDAS system at UWM and CIT to carry out request if computation needed.
- Next phase of prototype will integrate components from continuous wave (periodic source) searches.
  - Data products will involve the common “Short Fourier Transform” SFT frames produced by search codes running on the LDAS Beowulf clusters.
- Security to LDAS systems and data assured by Globus proxy credentials and GridFTP over secure socket connections.

LIGO

Open Issues

- GriPhyN activities starting to ramp up across collaboration(ocean).
  - Caltech, UWM, PSU, UTB, GEO (UK, Germany).
  - VIRGO, Others.
- Different potentials for how GriPhyN evolves from demos to real infrastructure.
  - Extensions of LDAS.
  - LIGO @ Home (a.k.a. SETI @ Home, a possibility for periodic searches).
  - GEO model (Java based TRIANA).
  - Grand Challenge model (CACTUS).
- No coherent body in the LSC to provide guidance on how to proceed.
  - Want to prevent redundant/conflicting efforts given that we have so few people to work on infrastructure.

Open Issues (continued)

- In LIGO the Tier 1-2-3 capacity ratio is close to [1:1:1], unlike the HEP ratio of close to [1:1/N_{Tier2}:1/N_{Tier3}] model.
  - How Tier 2 centers evolve actually impacts LIGO Lab significantly!
  - Tier 2 centers are hardware heavy and people light.
  - Unforeseen (in 1996-1997) was the propagation of LAB LDAS systems to collaboration sites, this has pluses and minuses:
    - (+) user experience/expertise defusing quickly.
    - (+) feedback; faster problem identification and resolution.
    - (-) LAB support function grown to greater than anticipated.
    - (-) becoming progressively more difficult to update LDAS/LAL releases due to inertia and lack of people.
- Tier 2 centers either will not or cannot adopt a pure LDAS hardware model.
  - heterogeneous environment complicates maintenance of LDAS/LAL releases.

Open Issues (unavoidable tension)

- 20 terabytes of archival data at Caltech already!
  - collaboration starting to implement analysis capabilities that will demand data transmission/moving.
  - pressures certain LDAS functionality (scope) be added against original timeline.
- How data are reduced as they filter down the Tier Levels still undefined by collaboration.
  - worst case => everything everywhere.
- Individuality vs. Conformality of Tier 2 design / functionality.
  - Identify subset of each Tier 2 that mirror LIGO configuration of LDAS.
    - fraction / partition can be changed according to load.
- Bandwidth limits the model of all data everywhere.
  - data reduction needs to waterfall through the Tier Hierarchy.
    - logarithmic data reduction required for optimization.