Open Science Grid

US ATLAS, US CMS, LIGO, (STAR) Usage and Milestones

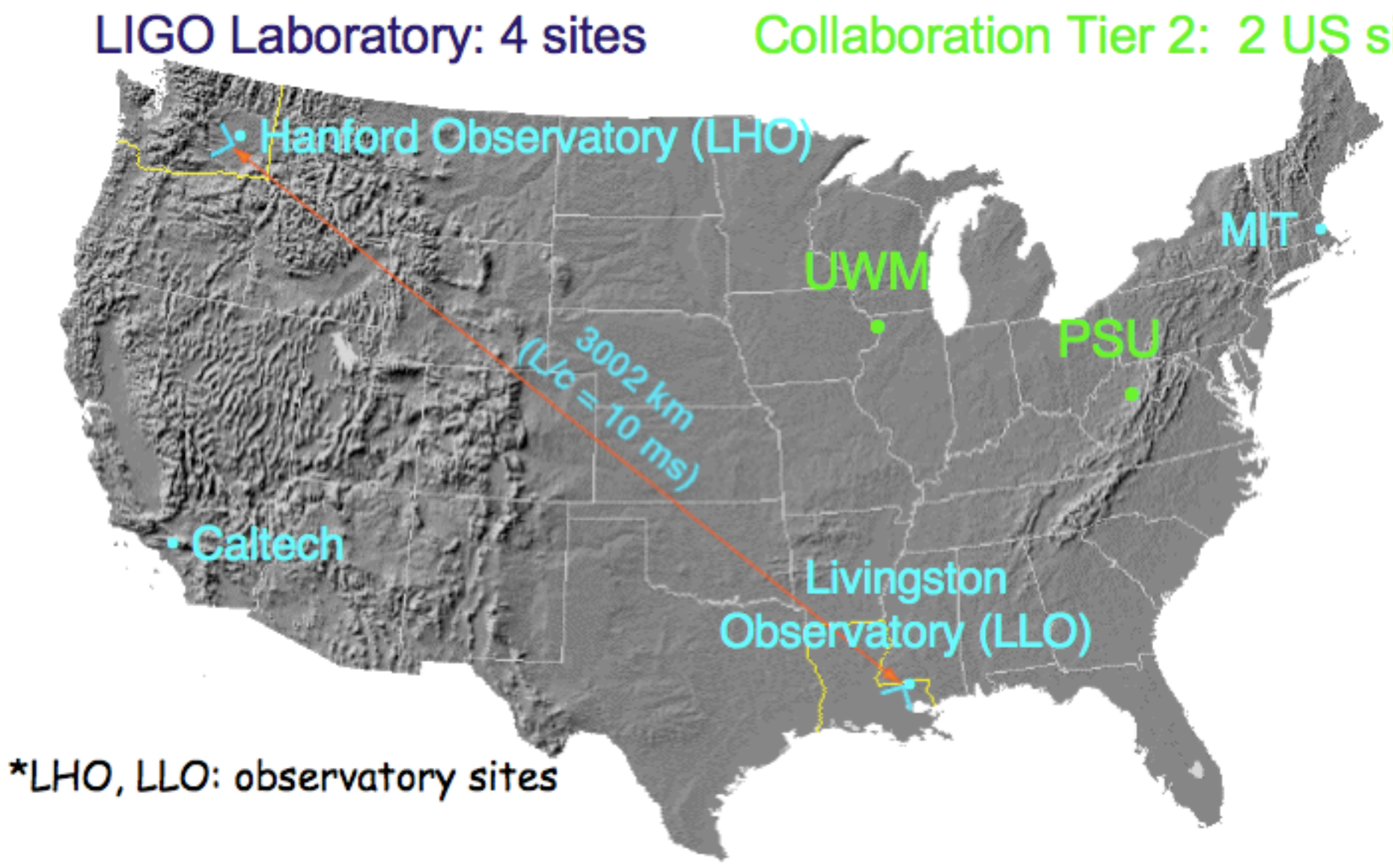
Torre Wenaus (BNL)

Open Science Grid Joint Oversight Team Meeting
February 20 2007

This Talk

- Brief survey of usage and feedback of principal OSG VOs
- Trying to be complementary (eg avoid repeating milestone content of other talks)
- Given the assigned title, I'm up here more with a 'user' hat than 'OSG management' (applications co-leader)
- Some slides and comments will reflect that
  - Intended to be useful to this group

LIGO Data Grid



- LSC - LIGO Scientific Collaboration
  - Not under organizational control of LIGO Laboratory
  - Funding provided through separate grants - NSF/EU
  - Cybersecurity policy allows them to join trust relationship with laboratory via MOUs

LIGO Data Products

LHO	Data Rate (MByte/s)	1 Yr (TByte)	1 Yr (tapes)	Comp. Ratio	Look-back
Level 0 raw	9.063	272	1800	1.7 on tape	~1 Month*
Level 1 RDS	1.391	42	460	1.2 in files	~most S5*
Level 3 RDS	0.117	3.5	39	1.2 in files	all S5
Level 4 RDS	0.029	0.87	10	1.2 in files	all S5
h(t)	0.266	8.0	88	--	all S5
SFTs	0.032	0.96	11	--	all S5
LLO					
Level 0 raw	4.406	133	970	1.5 on tape	~2 Month*
Level 1 RDS	0.750	23	249	1.2 in files	~most S5
Level 3 RDS	0.059	1.8	20	1.2 in files	all S5
Level 4 RDS	0.015	0.45	5	1.2 in files	all S5
h(t)	0.133	4.0	44	--	all S5
SFTs	0.016	0.48	5	--	all S5
Totals:	16.3	490	3701	as above	All Science Runs at CIT

** Level 0 Raw Data Transferred to Caltech Archive via Tapes. Remaining Level 1,3,4 RDS Data Transferred to Caltech and Tier II via Network.*

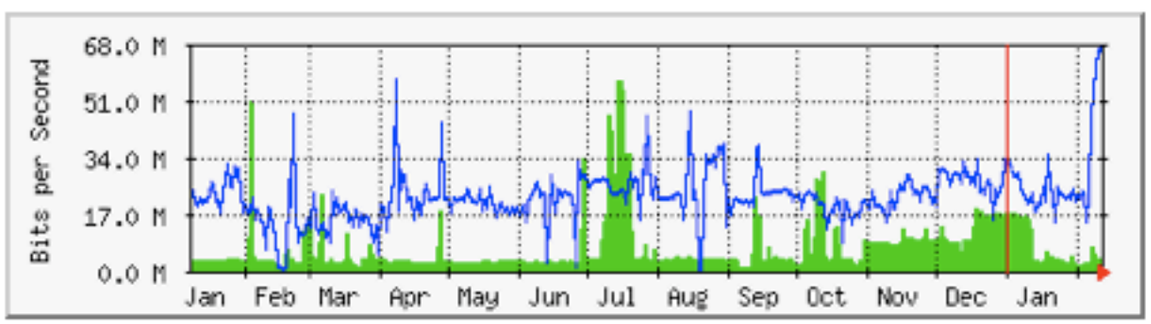
LIGO Data Distribution

- Reduced data products are distributed over the WAN using the LIGO Data Replicator (LDR), a grid tool built on top of the Globus Toolkit to manage both data and metadata transfers:
  - Metadata uses the Replication Location Server to hold mappings of filenames to paths/URL's at a given cluster location.
  - Parallel data transfers are accomplished using GridFTP.
- Data are re-distributed from CIT to Tier-2 and Tier-3 centers.
- GEO calibrated strain data are also replicated via LDR from Hannover, Germany to Caltech.
- Discussions have started with VIRGO regarding joint data exchange and possible interoperability problems between US and EU grid middleware.

LDG Networking Performance

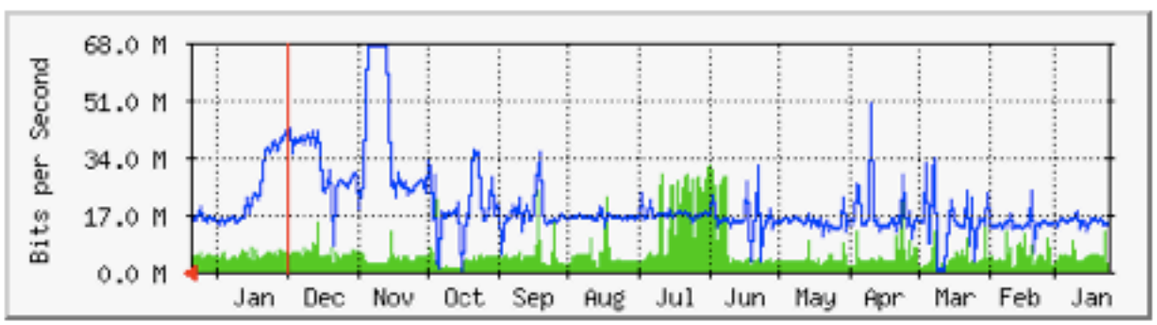
- LIGO Observatories to Caltech
  - LHO upgraded Esnet connection to GigE
    - Esnet upgraded backbone (SEA/SNV) to OC-192
    - Transferring frame data at 20-30 MByte/s
  - LLO upgraded to GigE connection to LSU
    - LSU connection to I2 is OC-3
    - 5-10 MByte/s
- CIT to Tier-2 centers
  - MIT 7 MByte/s
  - PSU 8 MByte/s
  - UWM 20 MByte/s

Yearly: LIGO Hanford Observatory



	MAX	AVG	CURRENT
IN	56.9 Mb/s (5.7%)	6.59 Mb/s (0.7%)	3.66 Mb/s (0.4%)
OUT	66.4 Mb/s (6.6%)	22.3 Mb/s (2.2%)	2.10 Mb/s (0.2%)

Yearly: LIGO Livingston Observatory



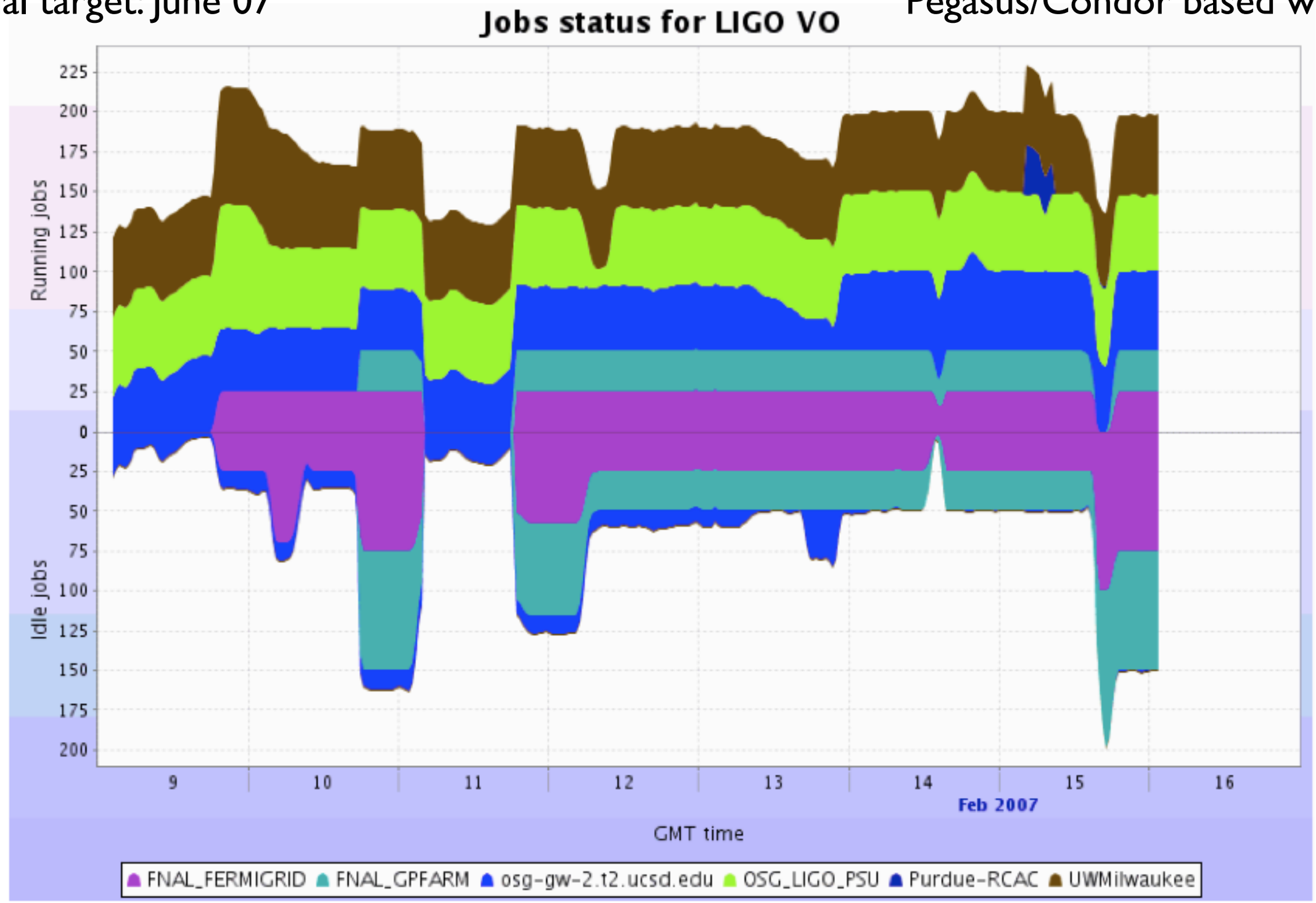
	MAX	AVG	CURRENT
IN	31.6 Mb/s (3.2%)	5.97 Mb/s (0.6%)	4.49 Mb/s (0.4%)
OUT	67.0 Mb/s (6.7%)	19.4 Mb/s (1.9%)	15.6 Mb/s (1.6%)

Second LIGO OSG Milestone

Sustaining 100+ jobs across multiple sites for 1 week

Original target: June 07

Pegasus/Condor based workflow



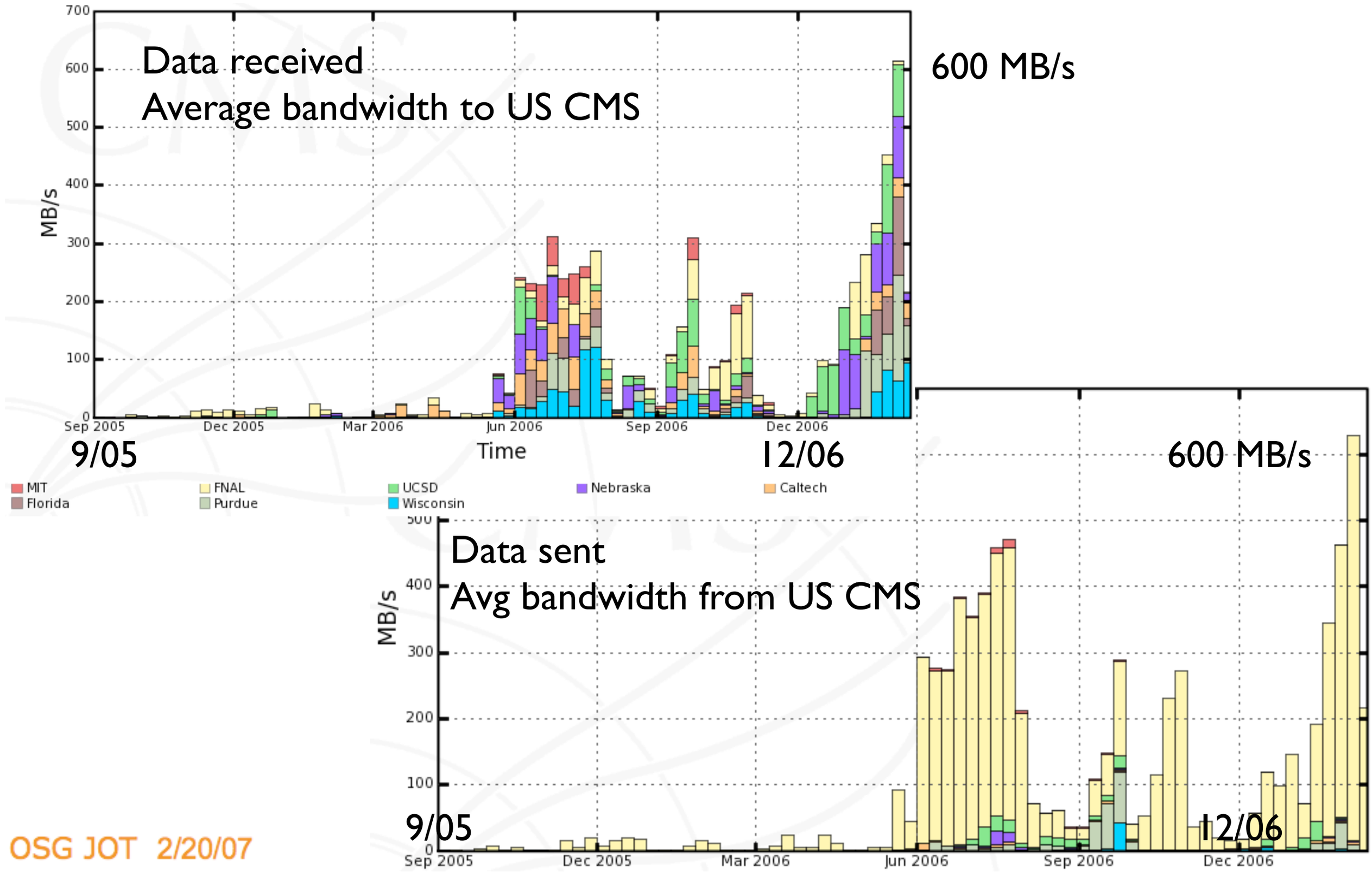
LIGO Lessons Learned

- From writeup of experiences from the first milestone (25 slots utilized for a week (in Jan) running full scale binary inspiral (HIPE) workflow)
  - With multiple running workflows, targeted slot usage was sustained well
  - Saw effects of CMS, CDF competition at UCSD, but sustained target
  - Able to run all HIPE workflows from LIGO workflow planner
  - Pegasus job clustering helps reduce gatekeeper scheduling overhead

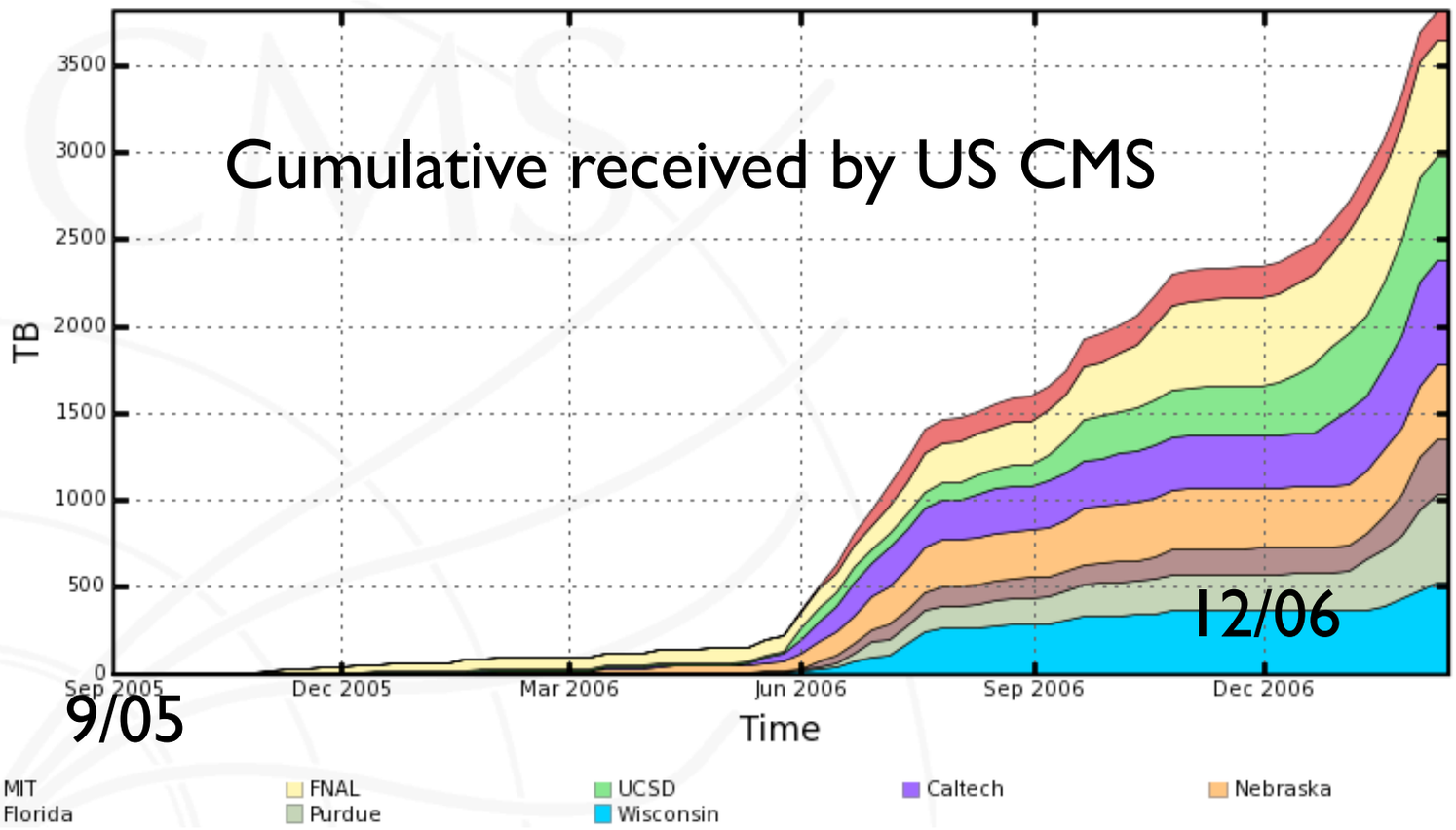
CMS Production & Scaling

- CSA06 in '06: 50M events exercising 25% of '08 scale. All goals met
  - T0 prompt reco with conditions DB, T1 skim and re-reco, T2 transfer
  - 207M events, 50k jobs/day
  - Excellent data throughput results (next slides); T0->FNAL demonstrated double the target of 150 MB/s
  - Data transfer challenges continuing this year
  - "Everything worked down to the final analysis plots"
- Currently ~1700 concurrent jobs on OSG (1.5M wall clock hours/mo), steady state, will rise to ~3000 by June
- June: test (over few days) of ~100k jobs/day across CMS; ~50k/day in OSG
- Next service challenge CSA07 in 'July'
  - 1 month, averaging 4000 jobs on OSG

US CMS Data Throughput

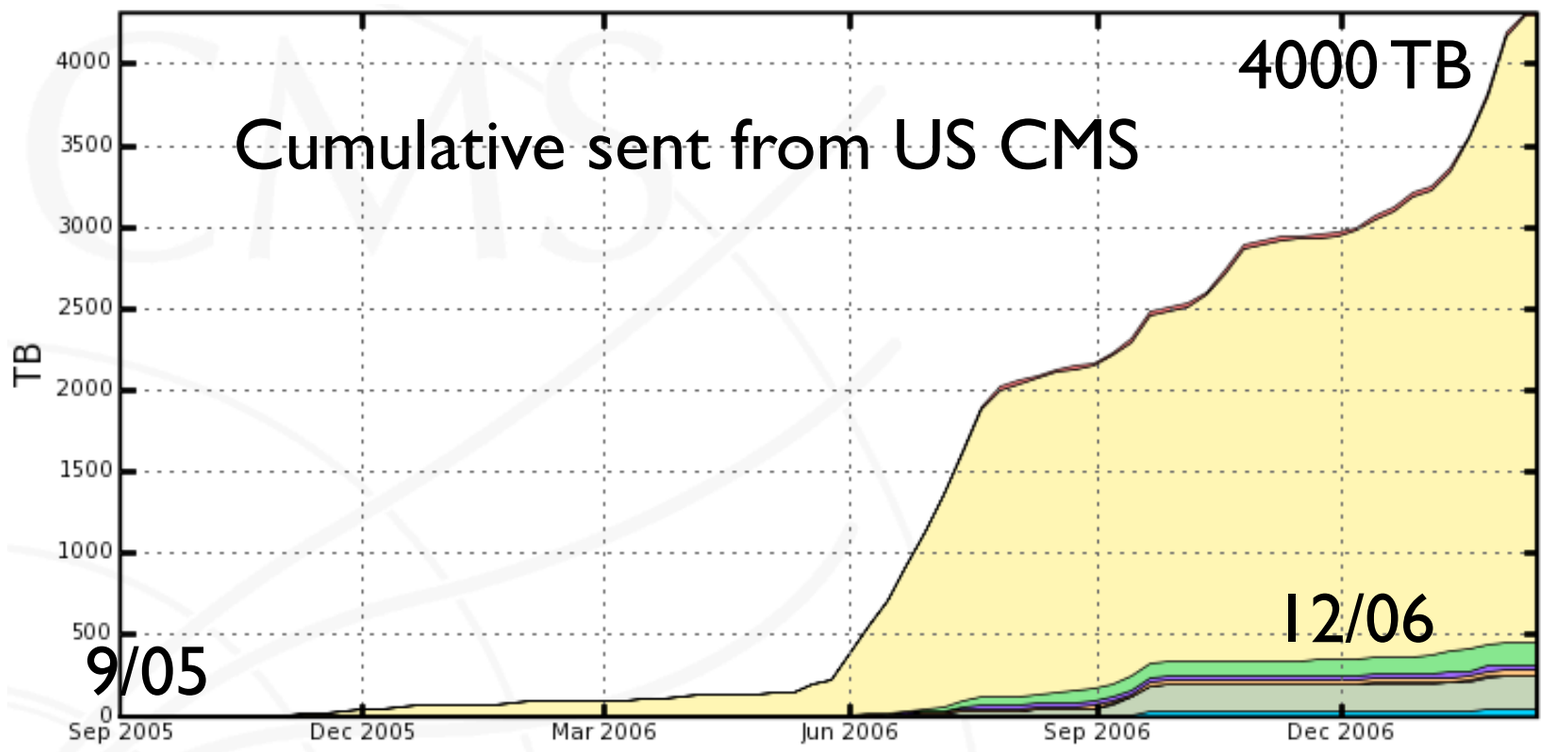


US CMS Data Throughput

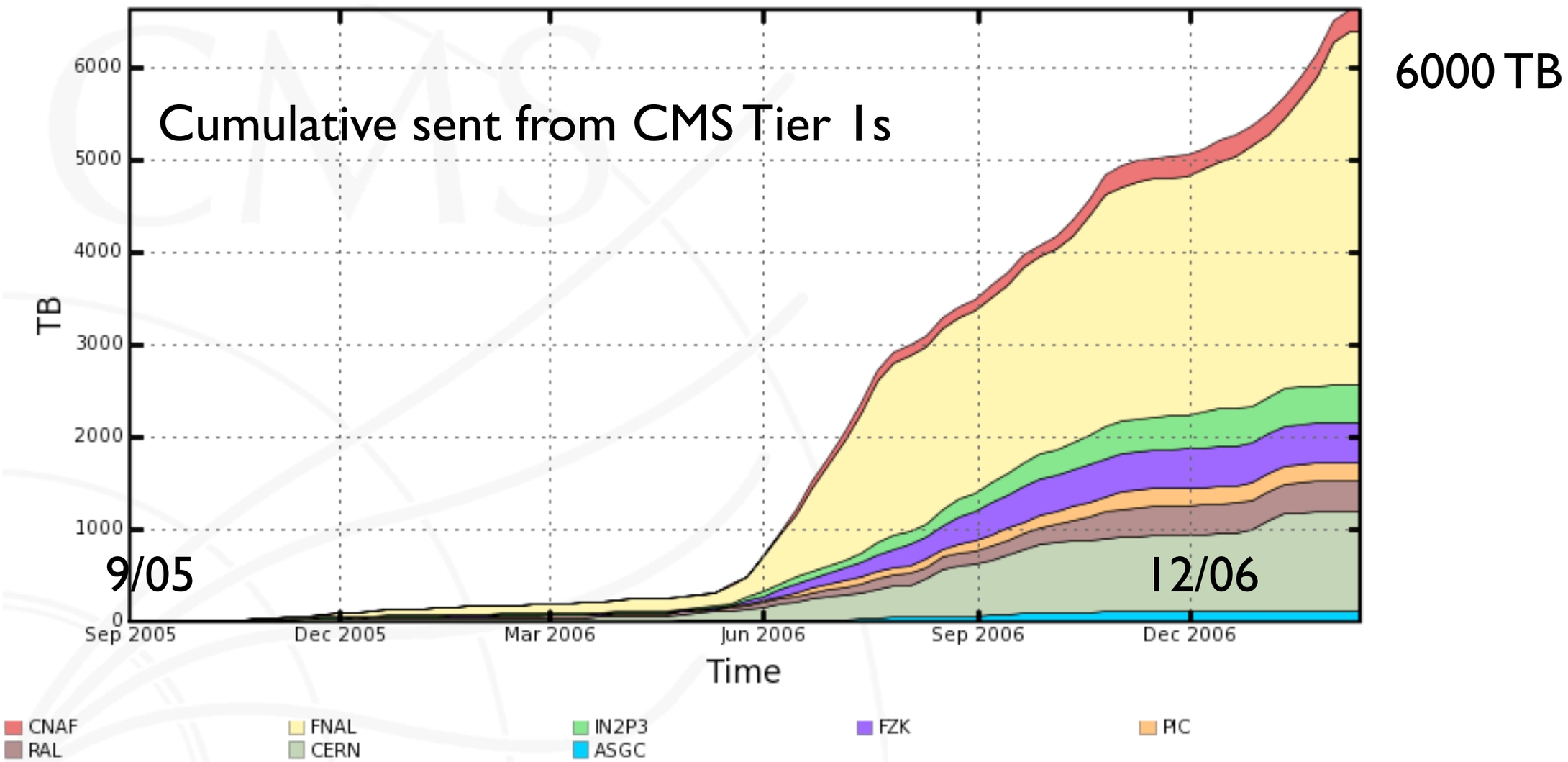


3500 TB

Note the ~.5TB 'trade surplus'



CMS Data Throughput



... done with dCache, so a big success story for dCache

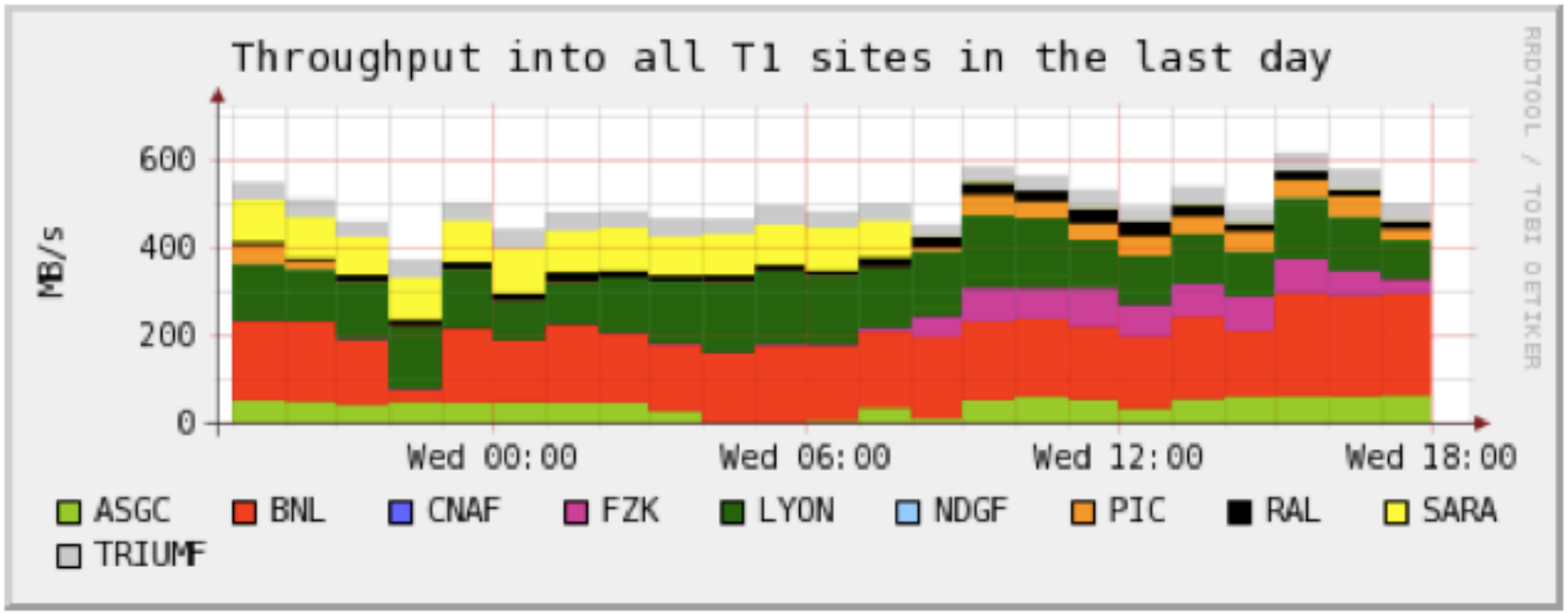
ATLAS Production & Scaling

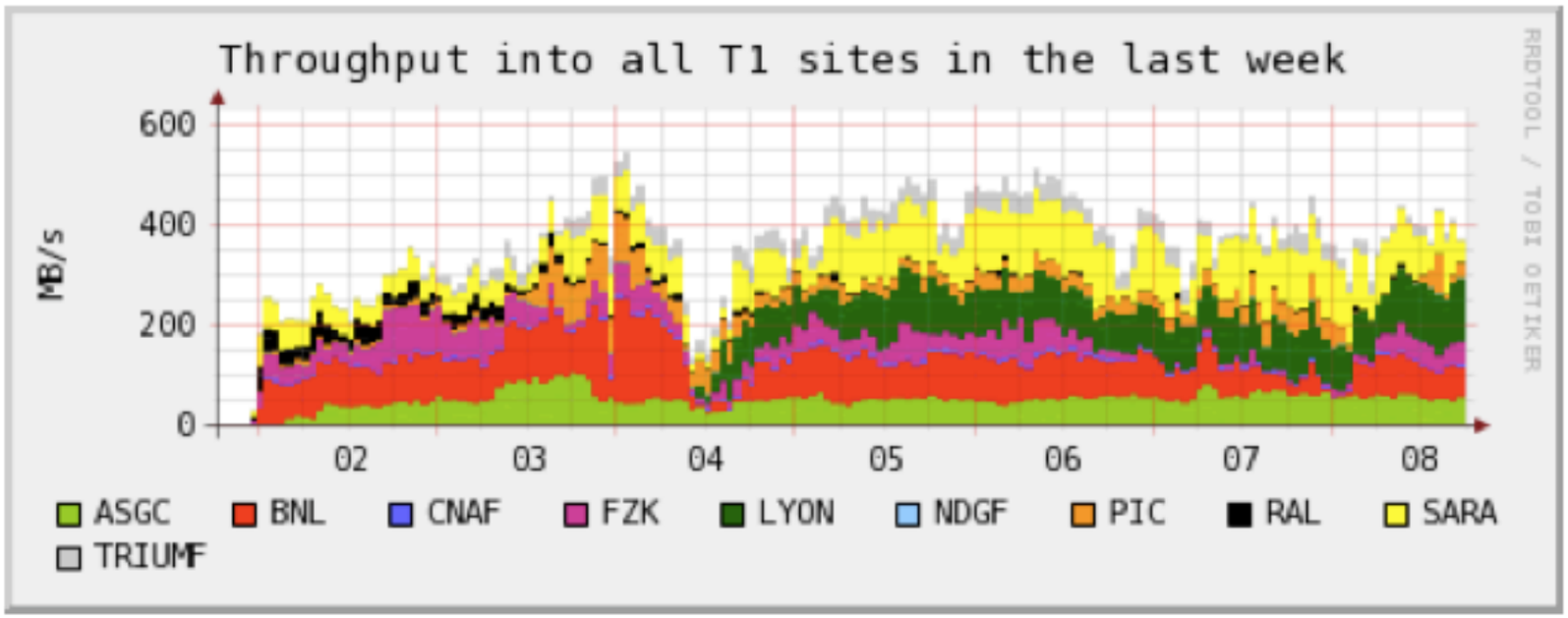
- ~50M MC events ATLAS-wide in '06, ~10k CPUs (2-3k/day continuous)
  - Processed down the full chain to AODs (analysis)
- >120TB data, with BNL and CERN the most comprehensive data pools
- OSG (Panda) production fraction: ~28% ('official share' ~19%)
  - Using US ATLAS resources (efficiently); opportunistic OSG usage in 1-2mo
  - Operations workload (exclusive of ATLAS sw issues) << 1 FTE
- OSG production currently averaging ~8000 jobs/day, ~1800 concurrent
  - Will double by the summer, double again in the fall
  - Analysis usage growing but production still heavily dominant
- OSG throughput requirements: ~20k jobs/day spring, 2-3 times that by fall
  - Not expected to tax scalability of the infrastructure
- Also pursuing expansion of the US system (Panda) into LCG (as well as OSG!)

US ATLAS Data Throughput: T0-T1

LCG Service Challenge 4

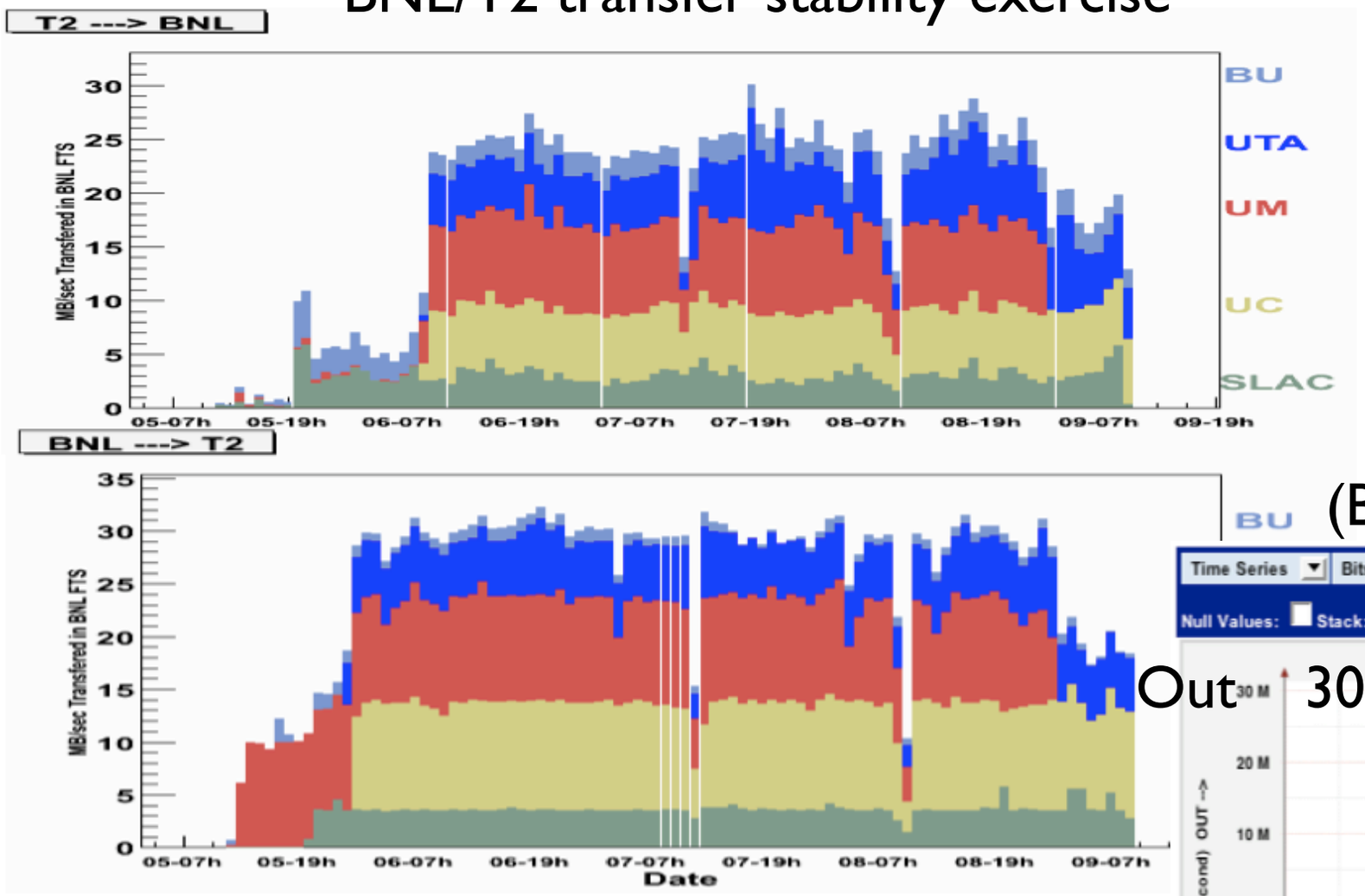
BNL intake from CERN 200MB/s nominal requirement met



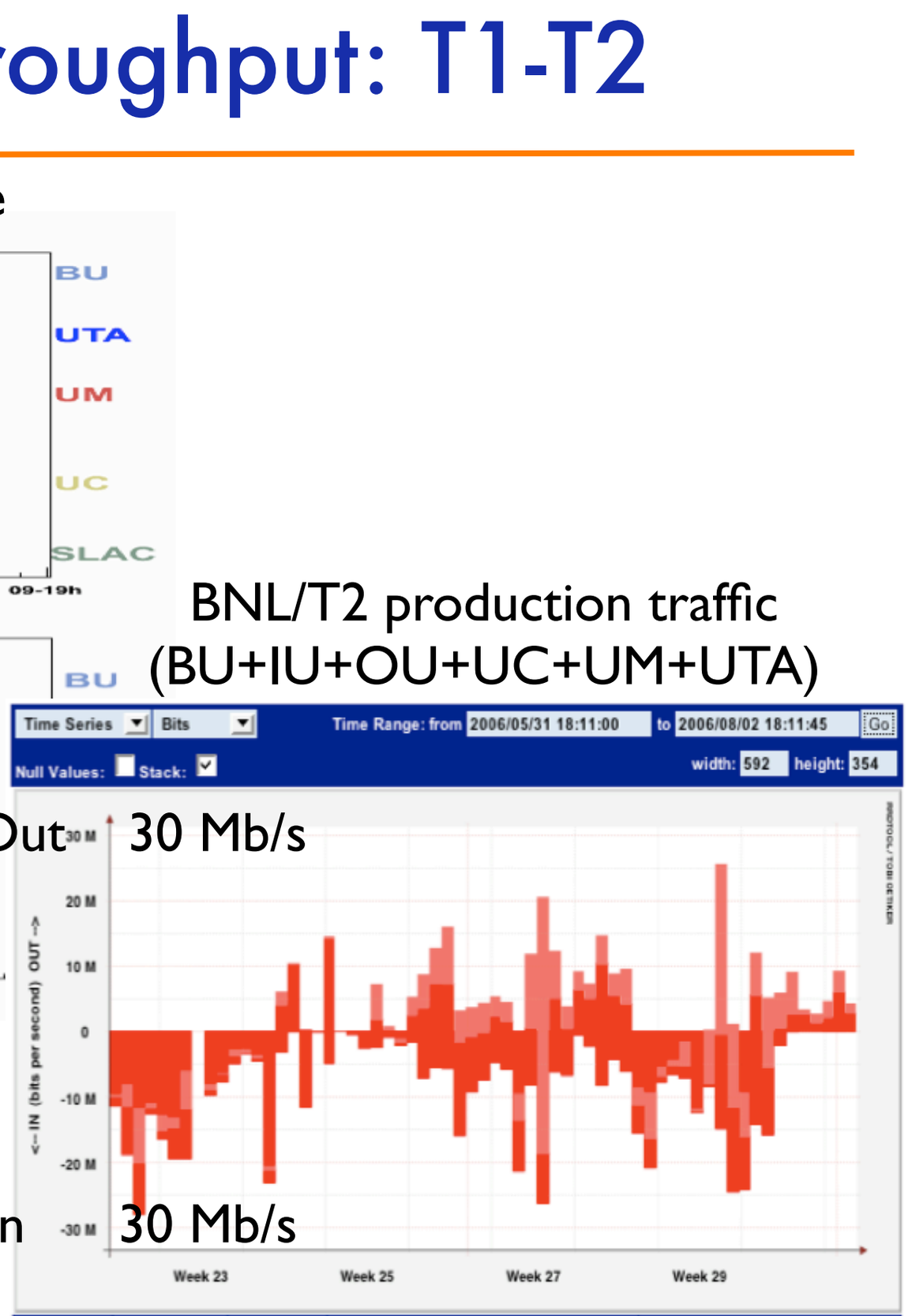


US ATLAS Data Throughput: T1-T2

BNL/T2 transfer stability exercise



BNL/T2 production traffic (BU+IU+OU+UC+UM+UTA)



Color	Site(s)	Label	Average Rate (bps)		Volume (Bytes)	
			In	Out	In	Out
	TIER-2	dst	8.878 M	5.929 M	5.494 T	3.669 T
	TIER-2	src	46.658 k	4.315 M	29.566 G	2.670 T

Using dCache, FTS

LHC Expectations from OSG

- SRM V2.2 (with dCache) deployment/interoperability 6/07
  - Follows and depends on (unrealistic?) WLCG milestone of 4/07
  - Speaking for ATLAS: we have learned the importance of minimizing SRM dependencies (via fallbacks); we can live without this until it comes as a deployed, robust, uniform service
- Submission rate scale-up for experiment production challenges 6/07
  - CMS: ~50,000 OSG jobs/day for June. O(1 Hz) submit rate
  - ATLAS: 'Dress rehearsal' from 'June'. ~20-30k prod jobs/day spring/summer
- Use of opportunistic sites for simulation production
- gLite data mover tools at OSG sites for inter-grid transfers 6/07
  - Until provided by OSG, self-installed by experiments as needed

LHC Expectations (2)

- Information services: the situation confuses me so I won't attempt to describe it to you
  - Practical (and feasible) approach today: gather info yourself, scraping OSG info sources, sites/WNs
- While it is a common (universal?) requirement among the large VOs to have (proxied! ie secure!) outbound connections from worker nodes, supporting this is not universal
  - OSG and the experiments have been working this site by site where it's an issue, usually with success (proxies are established)

ATLAS Comments wrt Control Metrics

Metrics proposed by ASCR

- User satisfaction
- System availability
- Problem response time
- Support for "capability limited problems at leadership facilities"

- US ATLAS experience to date limited to deployment of OSG-wide Panda infrastructure, with test jobs, no ATLAS prod yet
  - But now in use by CHARMM for production(!)
  - Opportunistic US ATLAS production OSG-wide in 1-2mo
- Qualitative experience so far is good
  - Deployment/debugging easier than expected; easy to get cycles
  - OSG information services imperfect but helpful
  - Ditto for 'uniformity measures' like standardized environment
  - Problem reporting and response very helpful
    - Response delays were with sites, not OSG Ops, who were effective mediator

Expectations wrt Applications Area

- Devoting major attention and resources to building upwards from the middleware foundation into higher level common services, and hardening the foundation on the way up, is a great initiative on the part of OSG
  - A comment not without bias
- Too late for a major alignment of higher level infrastructure (data management, production, analysis systems) but still much potential
  - Drawing on the commonality we do have

Principal AA/UG Common Activities

For project year 1

- SRM/dCache (and later other?) storage systems evaluation, integration, hardening, monitoring
- Workload management following the in-vogue 'just in time' approach and leveraging Condor
- Middleware studies/improvements for reliability & scalability (eg OSG CE)
- Real, non-virtual FTE support to VOs for targeted help, support, migration to OSG infrastructure, liaison, extracting feedback (NB also the related Troubleshooting activity)

Observations

- OSG well focused on requirements of primary stakeholders
  - No significant 'pull' away to costly new requirements from other domains (eg fine-grained, data security)
    - With some exceptions, our requirements are often a superset of those of others
- At the same time, OSG is proving itself serious about drawing in and supporting a broader user community
- Both these are greatly aided by OSG being operated as a cohesive project, controlling its resources and asserting its program priorities with participants

My Comments

- Do we (HENP) still have a 'grid problem'?
  - Over-promised, under-delivered
  - But very well resourced
    - And therefore, poor return for the investment
- Yes
- Is OSG part of the problem?

My Comments (2)

- Is OSG part of the problem?
  - No, in my opinion
  - OSG is trying to do its job right, and so far with promising results, in the ways that LCG tried initially to do it right, and failed
  - Where 'its job' is to turn 'the grid' into an effective, efficient science producer

My Comments (3)

- LCG's mantra in early years was 'LCG is not a middleware development project'
  - Rather, LCG would focus on middleware evaluation/feedback/ acceptance, and integration/deployment
  - And, LCG would seek to establish collaboration among the experiments on the higher levels of the infrastructure -- distribution production/data management/analysis
- Worthy, well-directed ambitions which didn't happen
- OSG *is* following this path, though, and so long as it does, stands an excellent chance of succeeding at keeping a 'science customer' focus, mediating middleware project interactions with this focus, undertaking collaborative science-focused higher-level infrastructure projects, and delivering -- at a relatively modest cost -- an effective infrastructure for producing science.

Conclusions

- OSG as a funded project is young
- But not so young that we can't see a direction
  - One that's absorbed lessons from other projects
- For the large customers, and I think evidence suggests for the small as well, the direction looks good so far
- US ATLAS reported in its last agency review that we're on a trajectory to meeting all distributed computing needs of the physics program when datataking arrives
  - This anticipates OSG continuing on its present trajectory of strong alignment with real experiment needs