

GridPP

UK Computing for Particle Physics

Tier-1 in GRIDPP5

Andrew Sansum

14th March 2011



Science & Technology Promotion Scheme
Rutherford Appleton Laboratory

- Archive (tape) - In our case an SL8500 tape robot
- Responsiveness - An on call team and daytime production
- Hardware (lots particularly disk). Eg 10PB disk > 500 servers
- Resource (commitments) - Annual procurement cycle
- Availability - Fix faults, fight multiple problems at once.
- Reliability - we have a culture of managing risk (change)
- Communications - Keep people informed in a timely way
- Core - run national and international core services
- Adaptability (agility) - meet changing VO requirements
- Trustworthiness - Do what we said, we set a good example
- Deployment (early) - Contribute to early deployment testing
- Accountability - Held to account when things go wrong

- Cost of Robotics, drives and media
- Simplistic first attempt, extend current plan, assume no new drives beyond D. Probably can reduce cost more
- Total tape costs in GRIDPP5 = £1899K (£1401K media)

	2013	2014	2015	2016	2017	2018
CAPACITY (PB)	14	15	26	38	51	57
Used Slots	9455	2511	4099	5880	7934	8909
Spent on Media	£40K	£54K	£336K	£386K	£456K	£222K
Drives and disk pools	£46K	£185K	£197K	£0	£0	£0
M&O	£52K	£53K	£70K	£75K	£77K	£79K
Total Cost	£138K	£292K	£603K	£461K	£534	£301K

- HW costs of tape 3 times less than disk (last 4 years)
- Many other costs make disk unattractive to replace tape
- Amazon Glacier list price uncompetitive

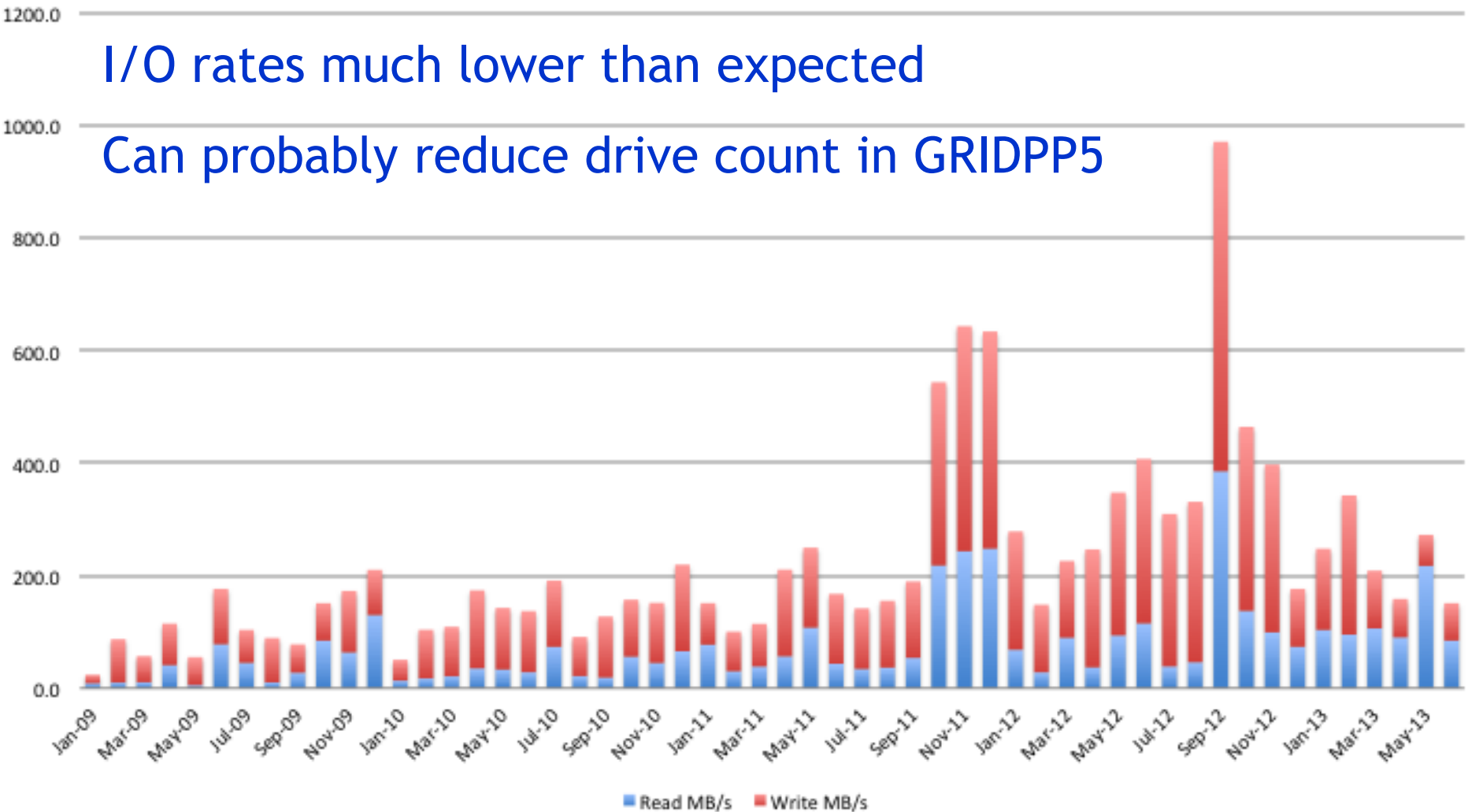
Hardware cost of Disk 2009-2013 cost/TB year	£96
Hardware cost of Tape 2009-2013 Cost/TB year	£33
Hardware cost of Tape 2014-2018 Cost/TB year (max cost)	£19
Cost of whole CASTOR team per TB/year (rough 2013 cost)	£22
Cost of electricity for disk per TB/year (rough 2013 cost)	£24
Machine room/cooling footprint, fabric/operations team disk operating costs	£???
Amazon Glacier cost per TB/year 2013 list price	£120



Tape Drive I/O Rate

I/O rates much lower than expected

Can probably reduce drive count in GRIDPP5



Role	Effort
Machine room robotics support	0.4 FTE
Robotics system admin expert intervention	0.1 FTE
CASTOR tape back end management	0.3 FTE

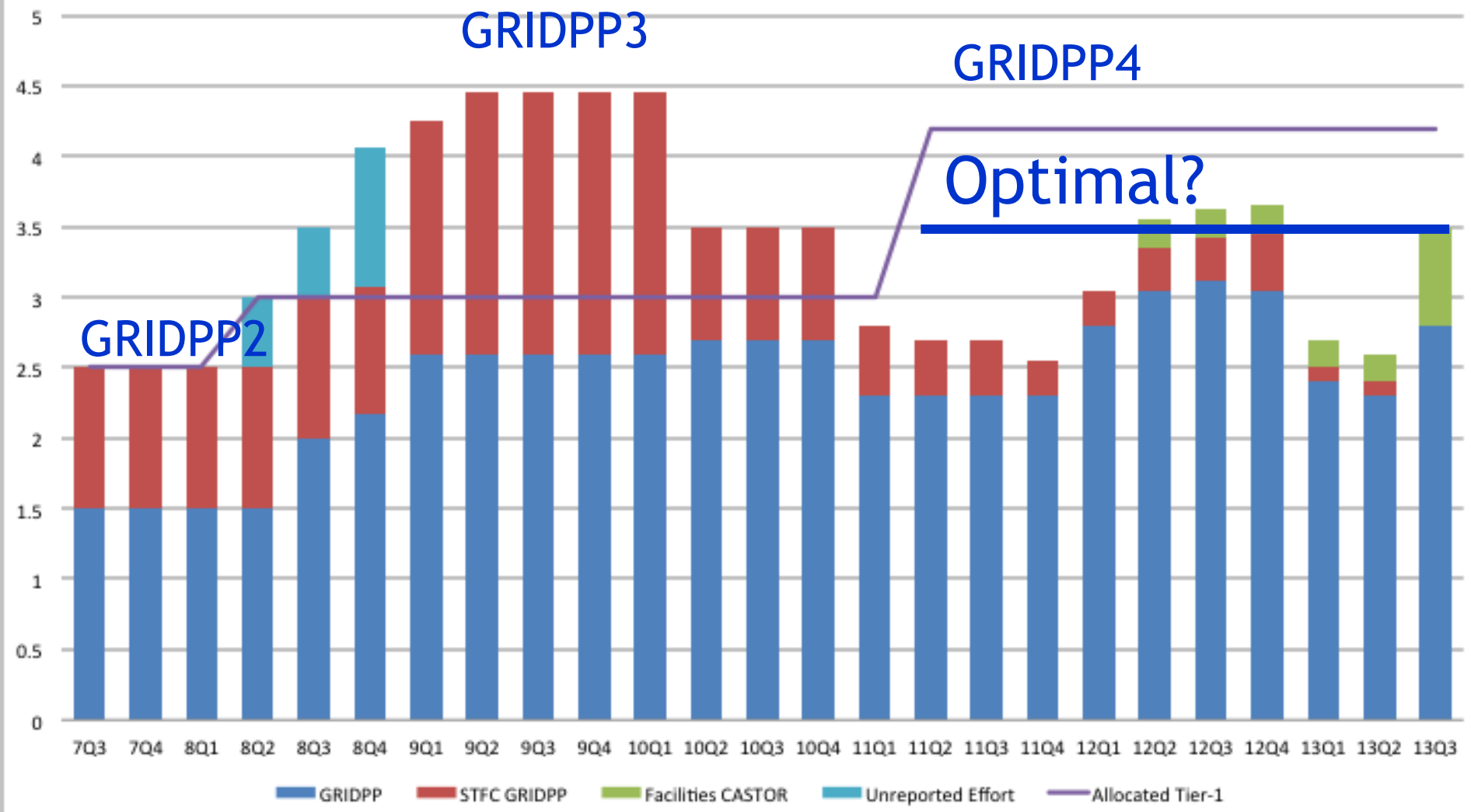
- Scientific Computing (SCD) fund most machine room staff effort, but GRIDPP4 funded 0.4 FTE to run robot
- Robotics operation costs actually actually found to be minimal (0.1 FTE)
- SCD will pick up residual cost in GRIDPP5 - reduction to GRIDPP5 of 0.4 FTE

Role	Effort
CASTOR Team Lead	0.5 FTE
CASTOR Team	1.9 FTE
CASTOR Tape back end and Robotic sysadmin	0.4 FTE

- GRIDPP4 plan projected 4.2 FTE (3.5 FTE operations 0.7 SRM development).
- Now, SRM development effort (funded by SCD) not needed
- CASTOR use (and funding) much widened. Shared cost and additional funding stream (0.7 FTE)
- GRIDPP effort can safely reduce to 2.8 FTE (from 3.5)



Castor Effort



Role	Effort
Database Team Lead	0.4 FTE
Two database admins	1.1 FTE

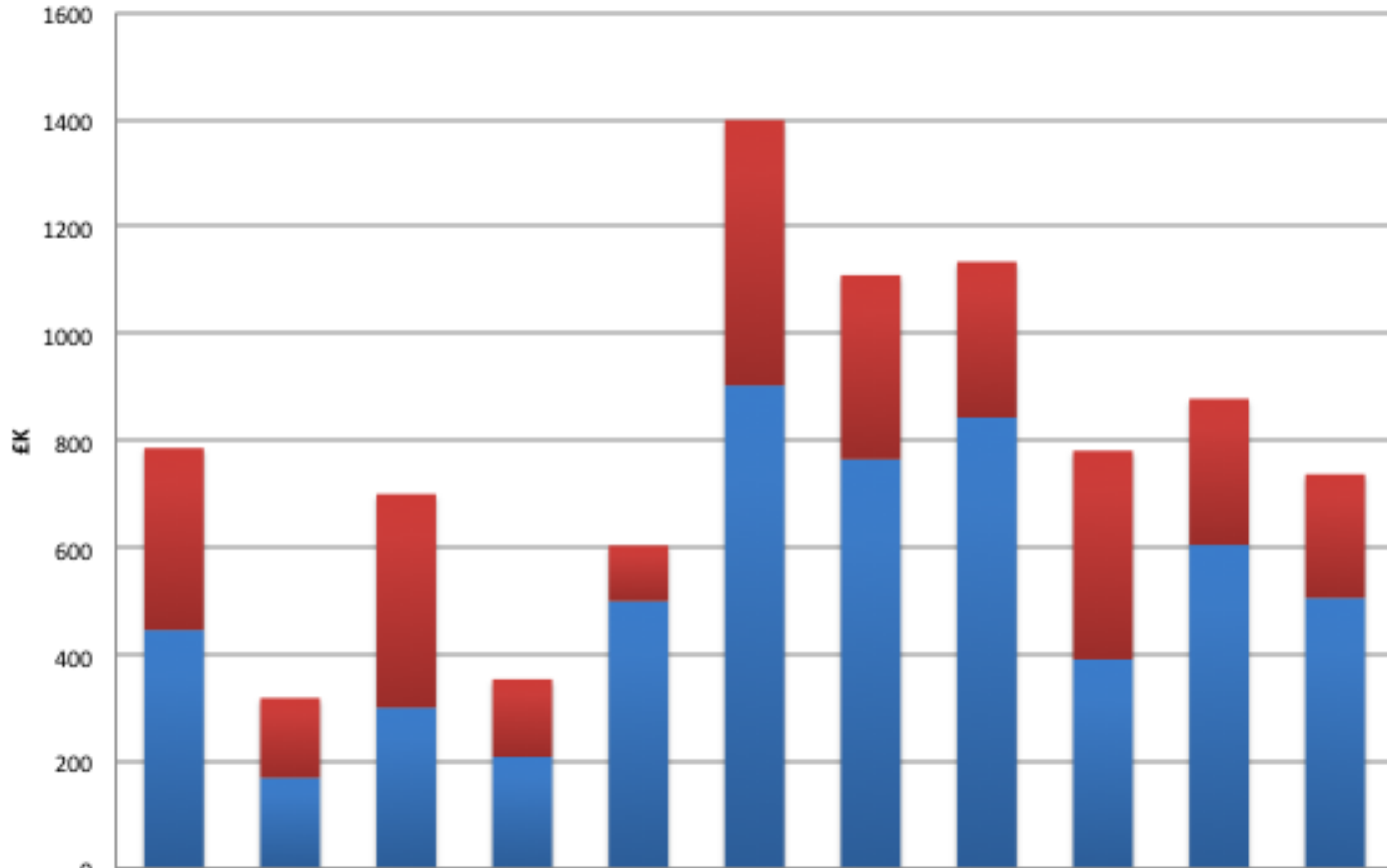
- Run the various databases. Oracle databases in particular
- CASTOR metadata catalogues, LFC, FTS and 3D
- GRIDPP4 plan projected 2.2 FTE effort. DB admin effort was at peak in 2010, for example GRIDPP4 assumed full time CASTOR D/B admin as CASTOR required daily intervention.
- CASTOR easier to run and LFC/FTS/3D databases reducing in priority.
- Team believe 1.5 FTE sufficient and have enough funding from other sources to maintain head count.

Role	Effort
Fabric Manager	0.9 FTE
System Admins	3.2 FTE
Hardware Technician	1 FTE

- Over 500 disk servers, 10000 drives, 500 worker nodes, specialised nodes (eg databases) and arrays
- 100 network switches, storage area network
- Virtualisation system and hypervisors
- New filesystems, vendor interactions and hw evaluation
- Major annual procurement and installation cycle
- Handle major m/c room incidents etc
- Quattor management but development moved to GRID team
- Slight reduction below GRIDPP4 level

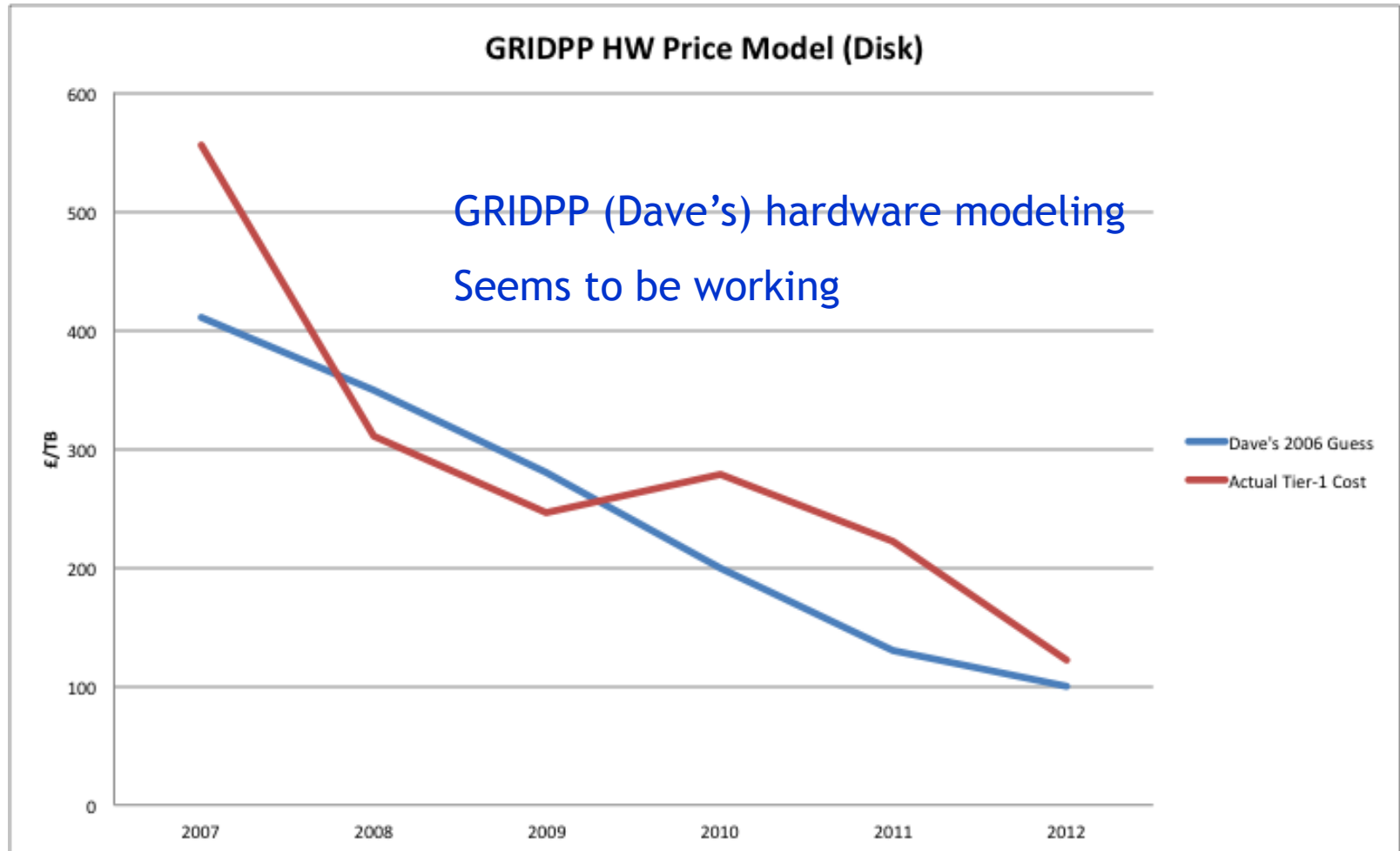


Capacity Spend - CPU and Disk - (£K)



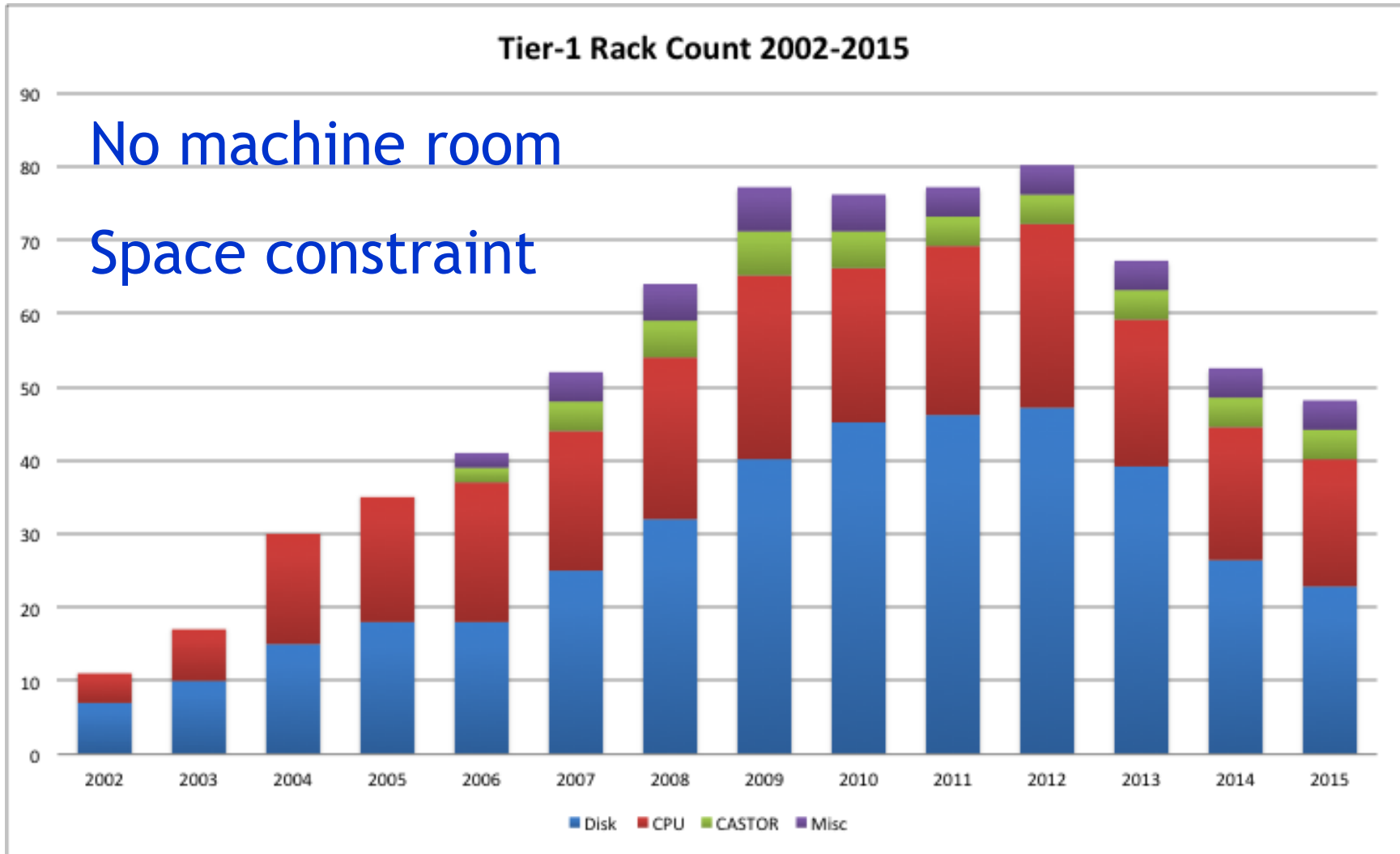
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
■ CPU	340	149	399	145	104	496	344	291	390	273	231
■ Disk	446	170	301	209	500	903	765	843	391	605	506

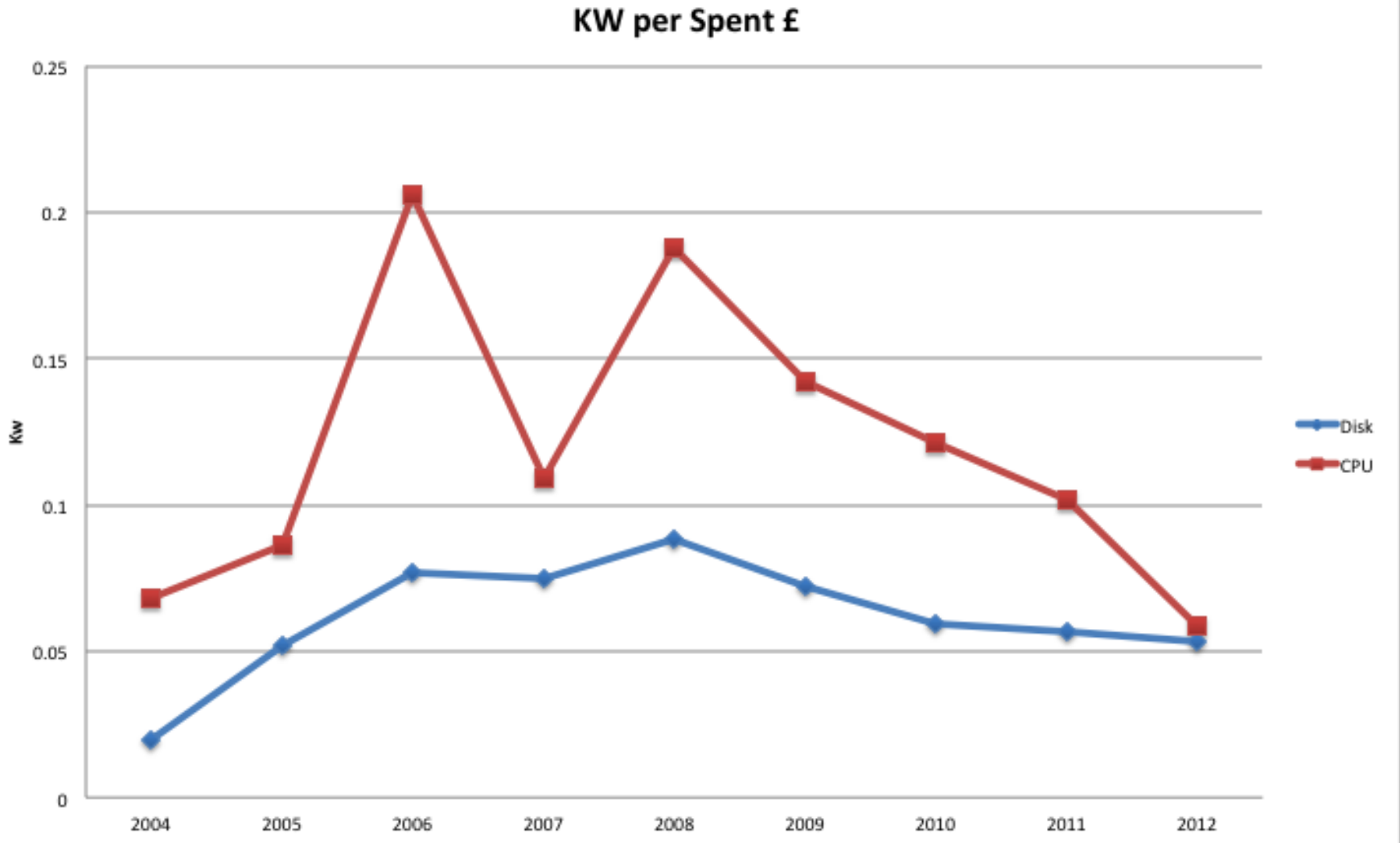
25 Sep



Role	Effort
Tier-1 share of machine room support staff (excl robot) manage machine room (SCD cost)	1.4 FTE
Estates staff support electricity and cooling	??????
Electricity charge paid by STFC “Estates”	

- Tier-1 continues to be largest player in machine room, although other scientific services getting large
- STFC Scientific Computing Department (not GRIDPP) pay directly the Tier-1’s share of machine room costs.
- SCD confirm they will continue to fund in GRIDPP5.





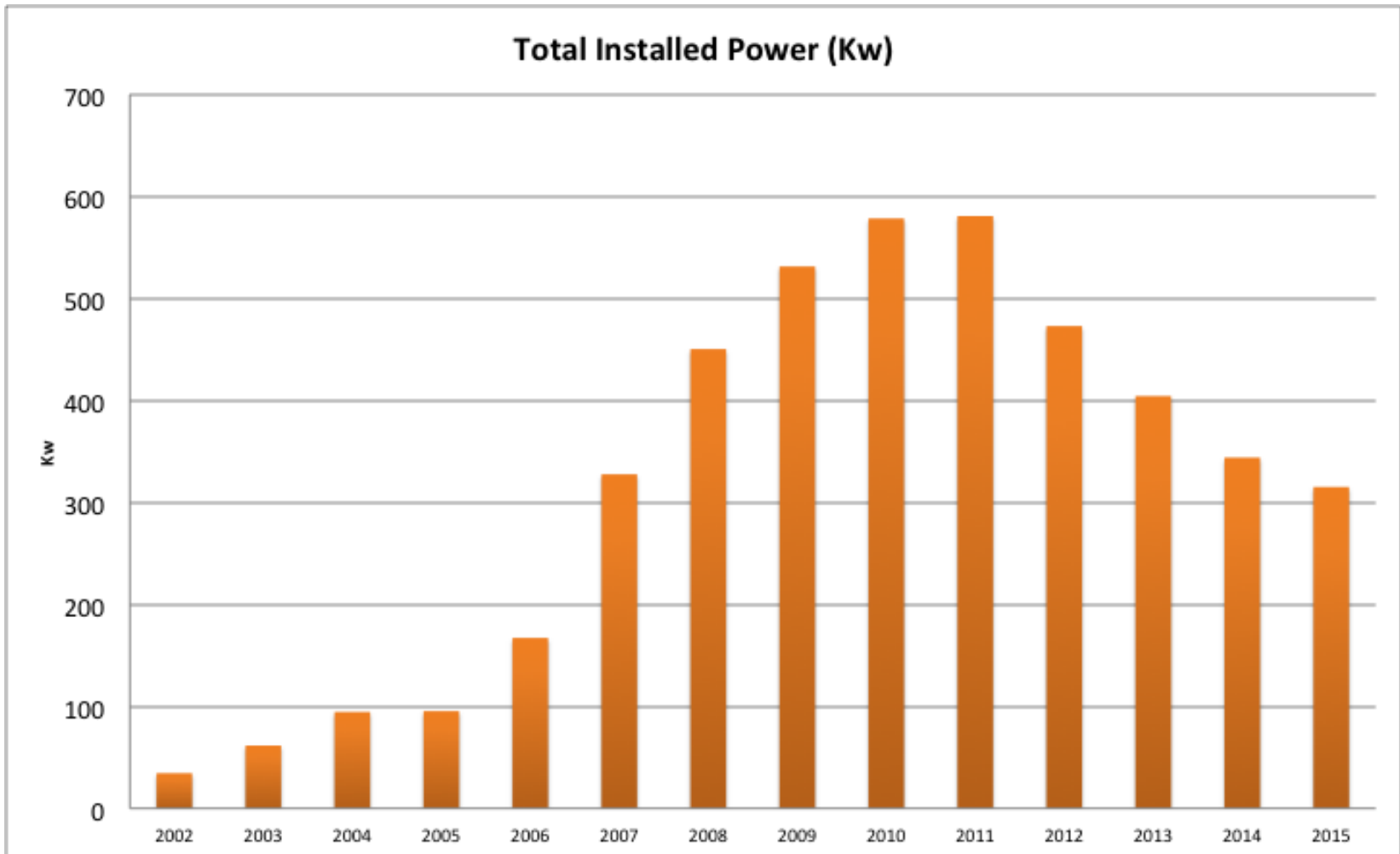
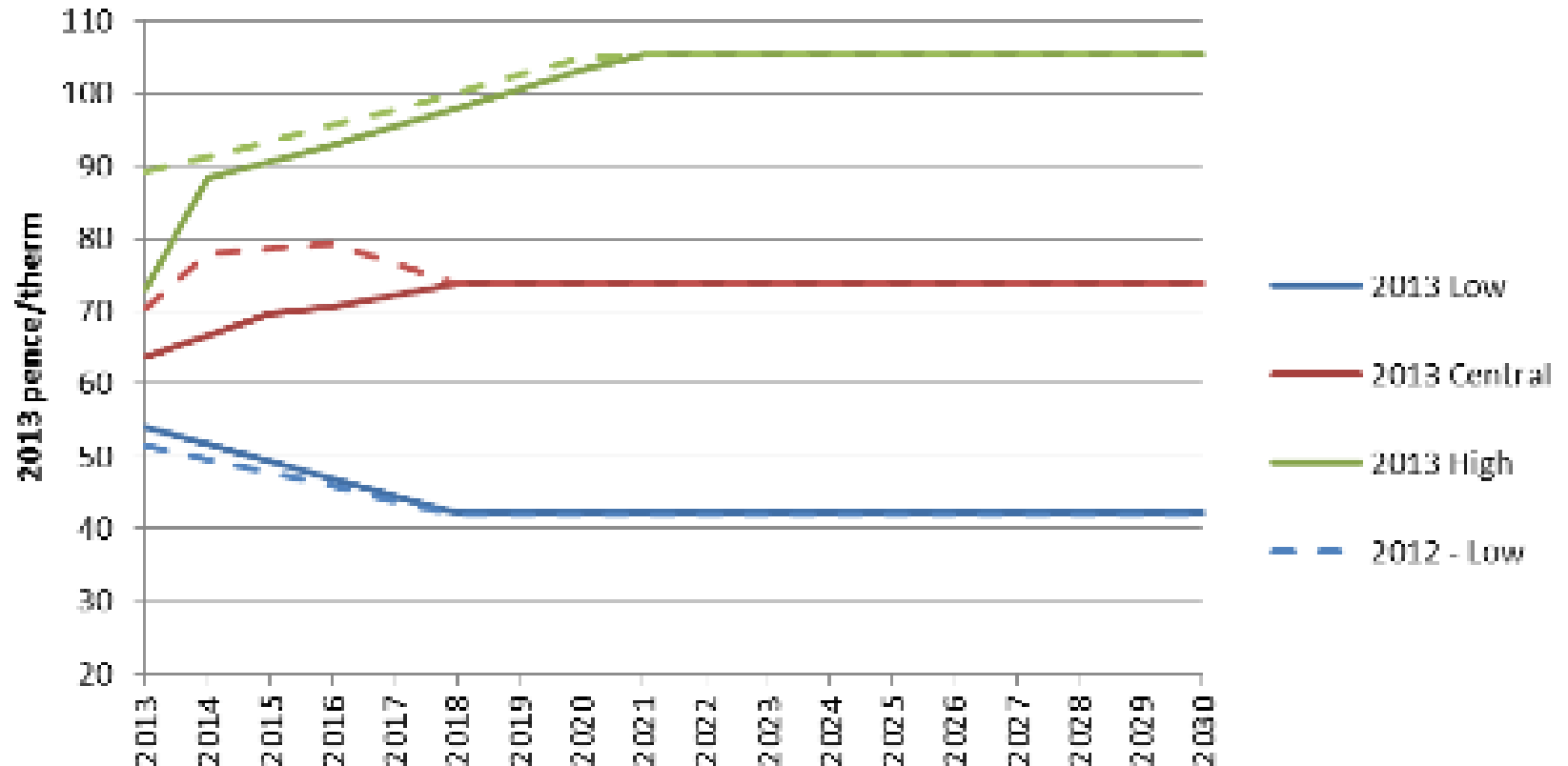
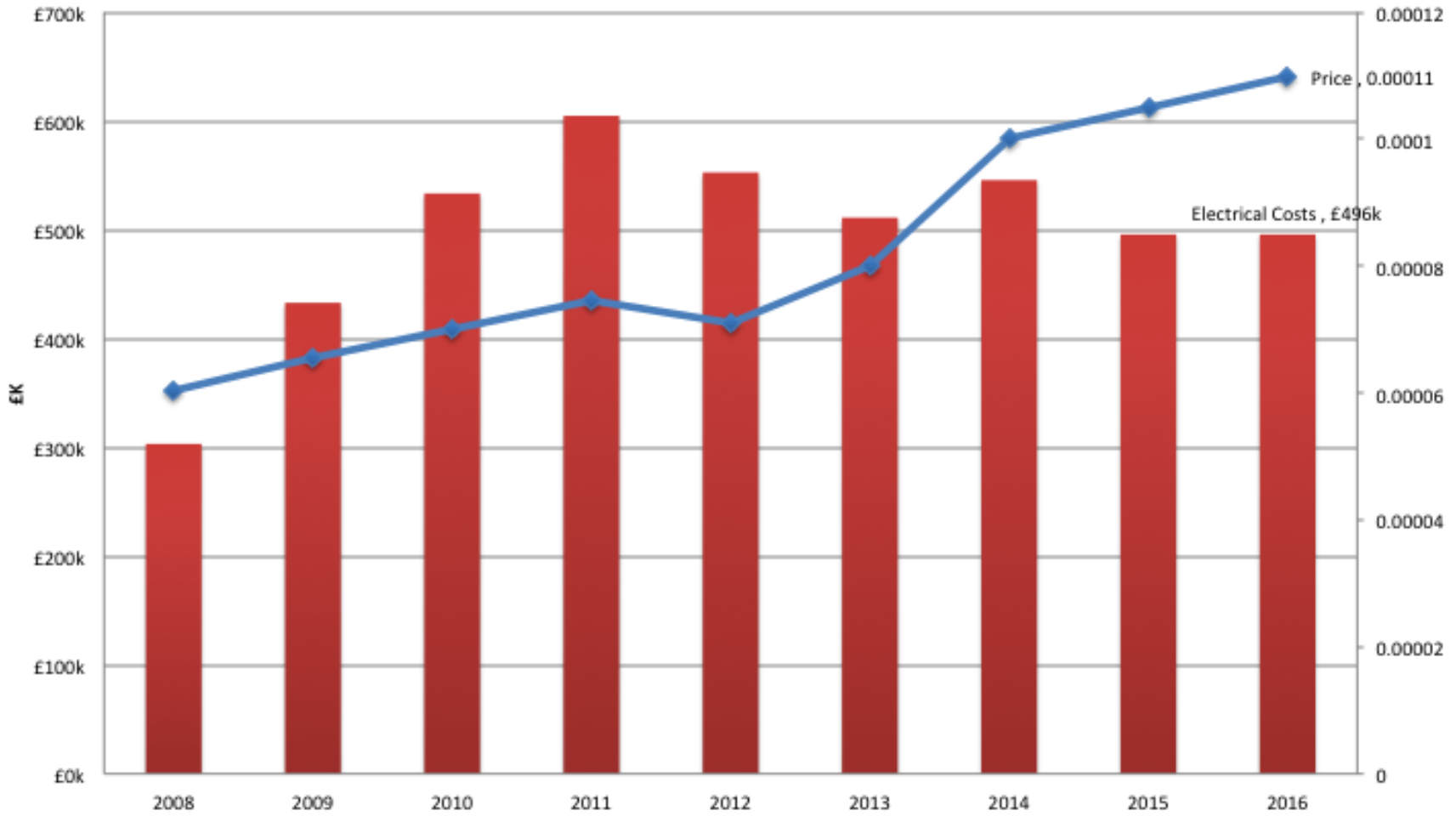


Figure 3: Comparison of 2013 DECC gas projections with those from 2012



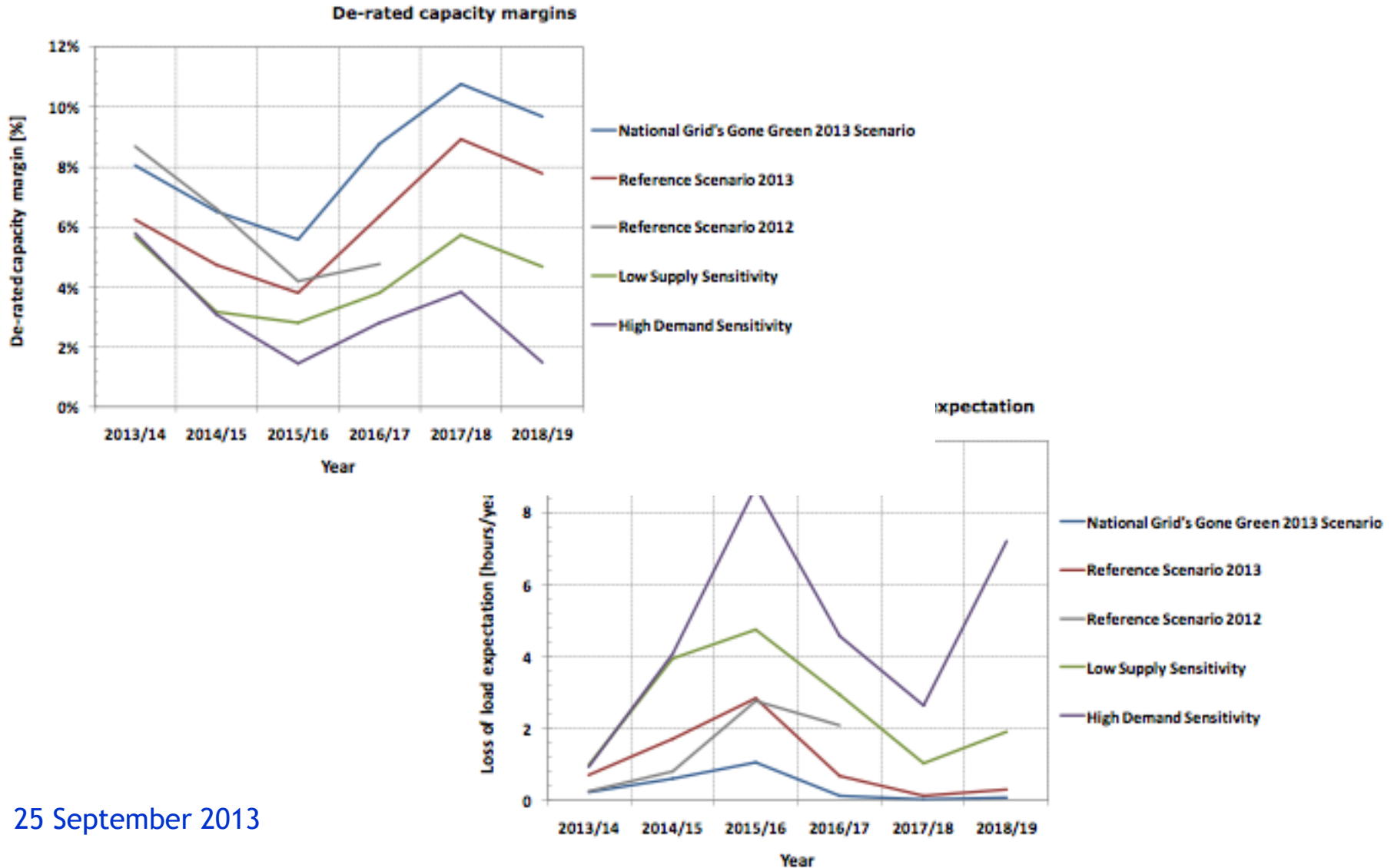


Electrical Charge (red) v Electricity Price (Blue)





1. De-rated capacity margins



Role	Effort
Production Manager	1.0 FTE
Production Team (2 staff)	1.9 FTE

- Production manager role - communications, intervention management and planning, continuous service improvement.
- Daily Admin on Duty (AoD) role rotates in team. Monitor service, manage service exceptions, triage tickets.
- Team run monitoring systems (eg Nagios/Ganglia)
- Team run batch farm operations, worker node upgrades.
- Effort much in demand from other parts of SCD - could reduce but would degrade GRIDPP service



- One primary on-call + 4(3) experts at any time guarantee response within 2 hours (10 mins in practice).
- Additional staff informally available to team if needed
- Team on duty 365 days per year, 250 calls (many pre-emptive), 99% handled remotely.
- During holidays team (3) do routine checks each day

FY	2011	2012	2013
# Out of hours exceptions	228	172	163
Total exception handling cost	£22K	£18K	£13K
Cost per exception	£95	£105	£82
Team standing cost	£20K	£20K	£20K
Hardware (mobiles,dongles,etc)	£5K	£5K	£5K
Total Cost	£46K	£43K	£38K



Role	Effort
Grid Team Lead	1.0 FTE
System admin/Software integrators	3 FTE
Closely integrate with experiment support	

- Not a good name any more - need a better one!
- The **G**eneralised **R**esearch **I**nfrastructure **D**eployment Team
- Respond to VO requirements track evolving technology (see Ian’s talk), eg Cloud
- Run national services (eg FTS, LFC, 3D, BDII, CVMFS, WMS)
- Lead early deployment site for CVMFS and FTS
- Maintain internal middleware, Ces, GlExec, Argus, batch..
- Manage and develop configuration management system

- We have seen a couple of teams at critical level, badly stretched at times following staff departure/re-deployment.
 - Fabric team were able to keep service running but not progressing deployment
 - CASTOR team struggling to fix grumbling faults and progress 2.1.13
 - Database team reduced below critical level (emergency contractor)
- Recruitments have now plugged these holes but we know where the edge is. Can live on it for a while but not indefinitely
- However early 2010 was a pessimistic time to plan GRIDPP4 at the Tier-1. We believe we can run the GRIDPP5 Tier-1 on less effort than GRIDPP4
 - We'd been through difficult times with CASTOR
 - Were just experiencing stretch from data taking
 - Had not yet fully reaped benefits of professionalisation/improvement