

HECToR Phase II

Progress Overview

April 2009 – Oct 2010

Dai Jenkins, EPSRC



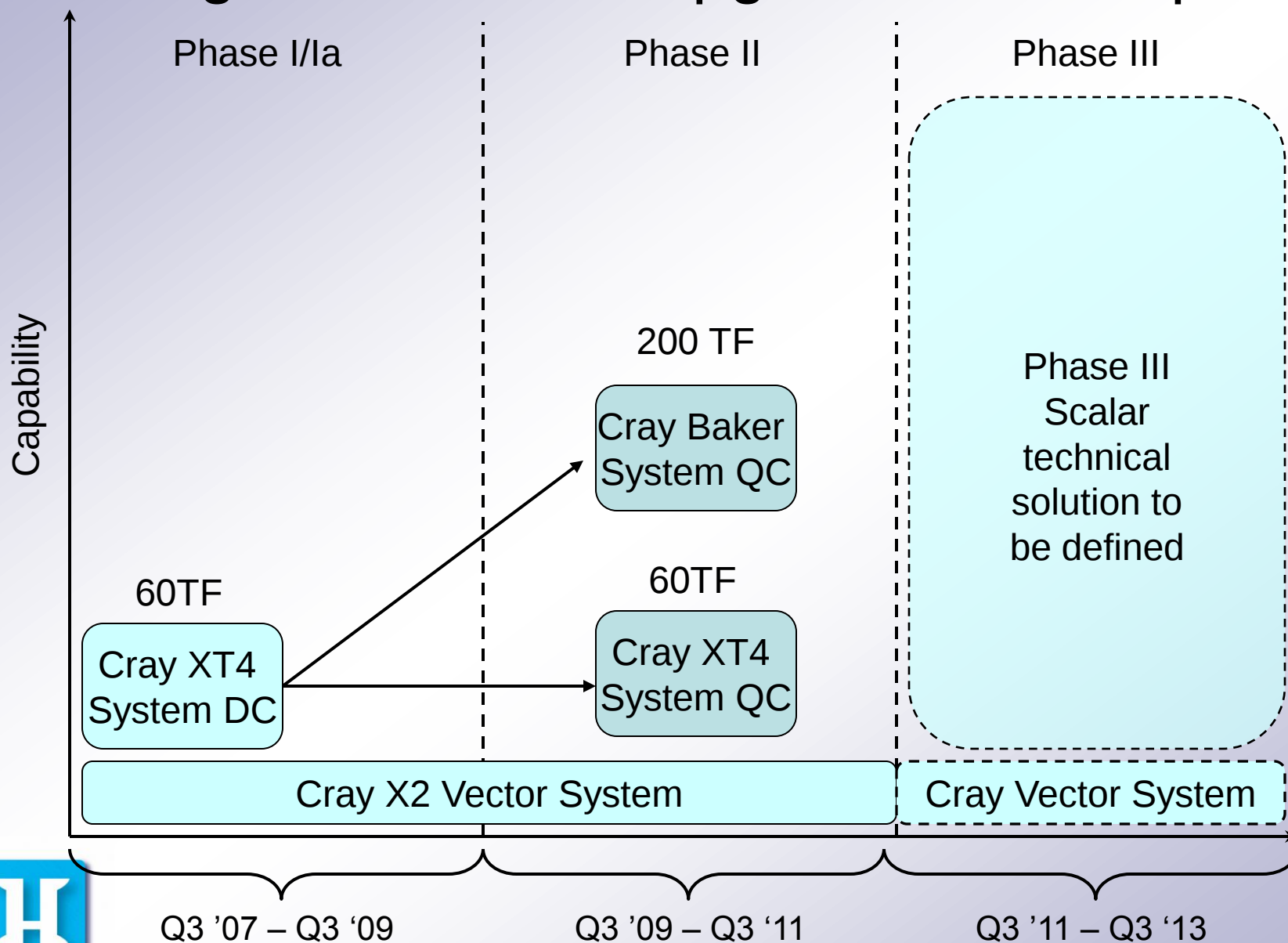
Overview

- Original Upgrade Path for Phase II;
- The need for change - context;
- Revised service upgrade path;
- Advantages and disadvantages vs in service experience;
- Summary
- Questions

Original system and upgrade path

- **Phase I - Q3 2007**
 - 60 cabinet Cray XT4 system;
 - 60 TF peak system performance;
 - 33.2 TB system memory;
- **Phase Ia - Q1 2008**
 - 1 cabinet Cray X2 Vector system;
 - 2 TF peak system performance;
- **Phase II - Q3 2009**
 - Resize of XT4 system to 16 Cabinets 60 TF
 - 20 cabinet Cray “Baker” system 200 TF
 - Retain X2 Vector System at 2 TF

Original HECToR Upgrade Roadmap



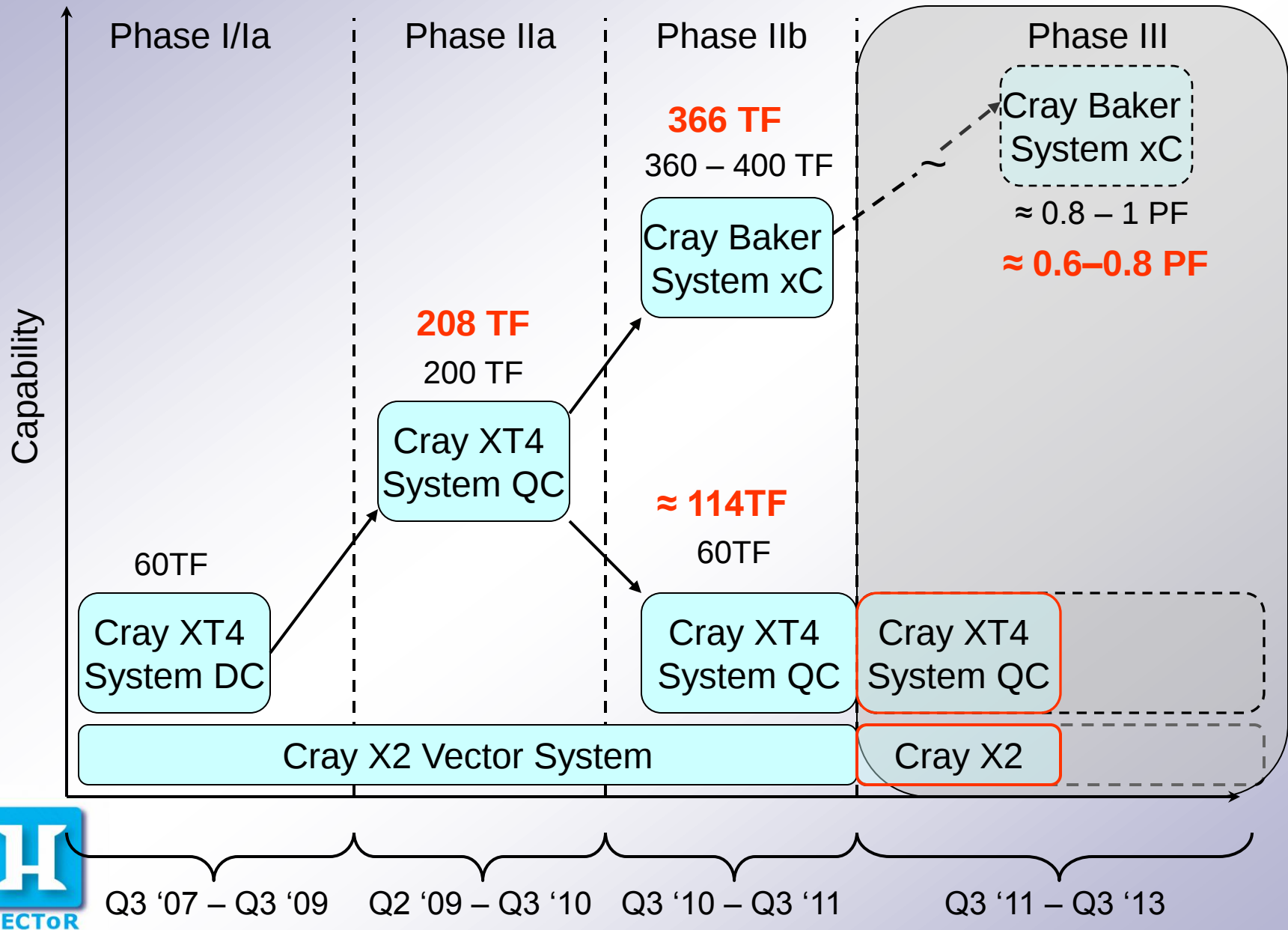
Why change the upgrade path?

- **Technical:**
 - Delays in AMD processor roadmap and resulting in delays in Cray's own technical roadmap;
- **Contractual:**
 - Cray committed to provide a viable and agreed upgrade solution in Q3 2009;
- **User considerations:**
 - Increasing number of grants requesting HECToR in Phase II timeframe;
 - Probable lack of capacity in original Phase II timeframe to meet user demand if Phase I system were extended;
 - High utilisation, high turnaround times detract from capability focus of the service

Revised Upgrade Path

- **Phase II system – delivered in two phases:**
 - ***Phase IIa Q2 '09/'10***
 - All 60 XT4 cabinets upgraded to Quad Core processors ;
 - Overall memory increase to provide 2GB/core;
 - 208 TF Peak
 - ***Phase IIb Q4 '09/'10***
 - XT4 system re-sized to 16 cabinets (60 TF, 2 GB/core);
 - Cray Baker system installed 22 cabinets (360 TF)
- **X2 Vector system retained but not expanded**
- **Archival Solution – Install and accept Q3 '09/'10**
- **External Services (esfs) – Install and accept Q4 '09/'10**
- **Outline options for Phase III circa Q1/2 10/11**

Revised HECToR Upgrade Roadmap



Advantages of Revised Path Apr' 2009

- **Increase in overall memory;**
 - Memory/CPU increase from 6 to 8 GB;
- **Increase in peak performance for 2011-'12;**
 - 413 TF versus 260 TF in original roadmap (aggregate);
- **Value for money;**
 - Order of 4x increase, 10% Phase I cost
 - Progressive transition from DC > QC > Multi/Many-core;
- **Significant energy savings during the lifetime of the phase;**
 - New cooling technology
- **Cost effective route to a possible 1 PF machine in Phase III**

Disadvantages of proposed route Apr '09

- **Memory per core physically limited to 8 GB/CPU, reduction per core from 3 GB to 2 GB;**
 - Possible need for node de-population in memory intensive applications;
 - This trend is not unique and is likely to continue as systems transition from multi-many core;
- **Quad core clock speed of Phase IIa less than that of Phase I system, 2.3 GHz versus 2.8 GHz;**
 - Higher scientific throughput but at the expense of shorter end to end completion times for individual jobs;
 - Again, likely to be an artifact of continuing move to higher core counts;
- **Current Interconnect system will be used in Phase IIa;**
 - No increased performance for communications intensive applications;
 - Current interconnect still efficient and will be replaced with Cray next generation interconnect in Phase IIb system;

Summary

- Phase IIa and Phase IIb delivered to time;
- Performance in line with expectations and in excess of original upgrade path;
- Archival system successful;
- External systems solution – still issues;
- Phase III options – tabled