СУПЕРКОМПЬЮТЕРНАЯ ОТРАСЛЬ В ПРЕДДВЕРИИ EXASCALE

РОЛЬ РОССИИ В МИРОВОЙ ЭКОСИСТЕМЕ

МГУ им. М. В. Ломоносова, 4 декабря 2012 г.

Андрей Слепухин
Главный системный архитектор
andrey.slepuhin@t-platforms.ru

Требования к Exascale-технологиям

	2012 (BG/Q)	2018-2020 (Exascale)	Относительно 2012
System peak	20 Pflops	1 Eflops	$O(10^2)$
Power	8.6 MW	~20 MW	
System memory	1.6 PB	32-64 PB	O(10)
Node performance	205 Gflops	1.2 or 15 Tflops	$O(10)$-$O(10^2)$
Node memory BW	42.6 GB/s	2-4 TB/s	$O(10^3)$
Node concurrency	64 threads	$O(10^3)$ or $O(10^4)$	$O(10^2)$-$O(10^3)$
Node interconnect BW	20 GB/s	200-400 GB/s	O(10)
System size (nodes)	98304	$O(10^5)$ or $O(10^6)$	$O(10)$-$O(10^2)$
Total concurrency	5.97 M	$O(10^9)$	$O(10^3)$
MTTI	4 days	O(<1 day)	-O(10)

Направления развития технологий

- Интеграция компонентов
- Память и интерконнект
- Оптические коммуникации
- Co-design

Интеграция компонентов

- Интеграция процессора и интерконнекта
- Интеграция процессора и памяти
- К 2020 году 1 узел ≈ 1 чип (возможно с дополнительной памятью большого объема)
- Перспективы развития:
  - 2015-2016:
    - Интеграция интерконнекта в процессор
    - Интеграция процессора и памяти в одном корпусе (system-in-a-package)
  - 2017-2018
    - 3D-stacking процессора и памяти

Память и интерконнект

- Затраты энергии на выполнение арифметических операций на порядки меньше затрат энергии на доступ к данным

	2011	2020
DP FMADD flop	100 pJ	10 pJ
DP DRAM read	4800 pJ	1920 pJ
Local Interconnect	7500 pJ	2500 pJ
Cross System	9000 pJ	3500 pJ

- В Exascale-системе до 50% энергопотребления может приходиться на интерконнект
  - Разработка энергоэффективных методов передачи данных
  - Разработка новых топологий и алгоритмов маршрутизации
- Новые энергоэффективные технологии памяти (HMC, Wide I/O)
- Постоянная память (non-volatile memory)

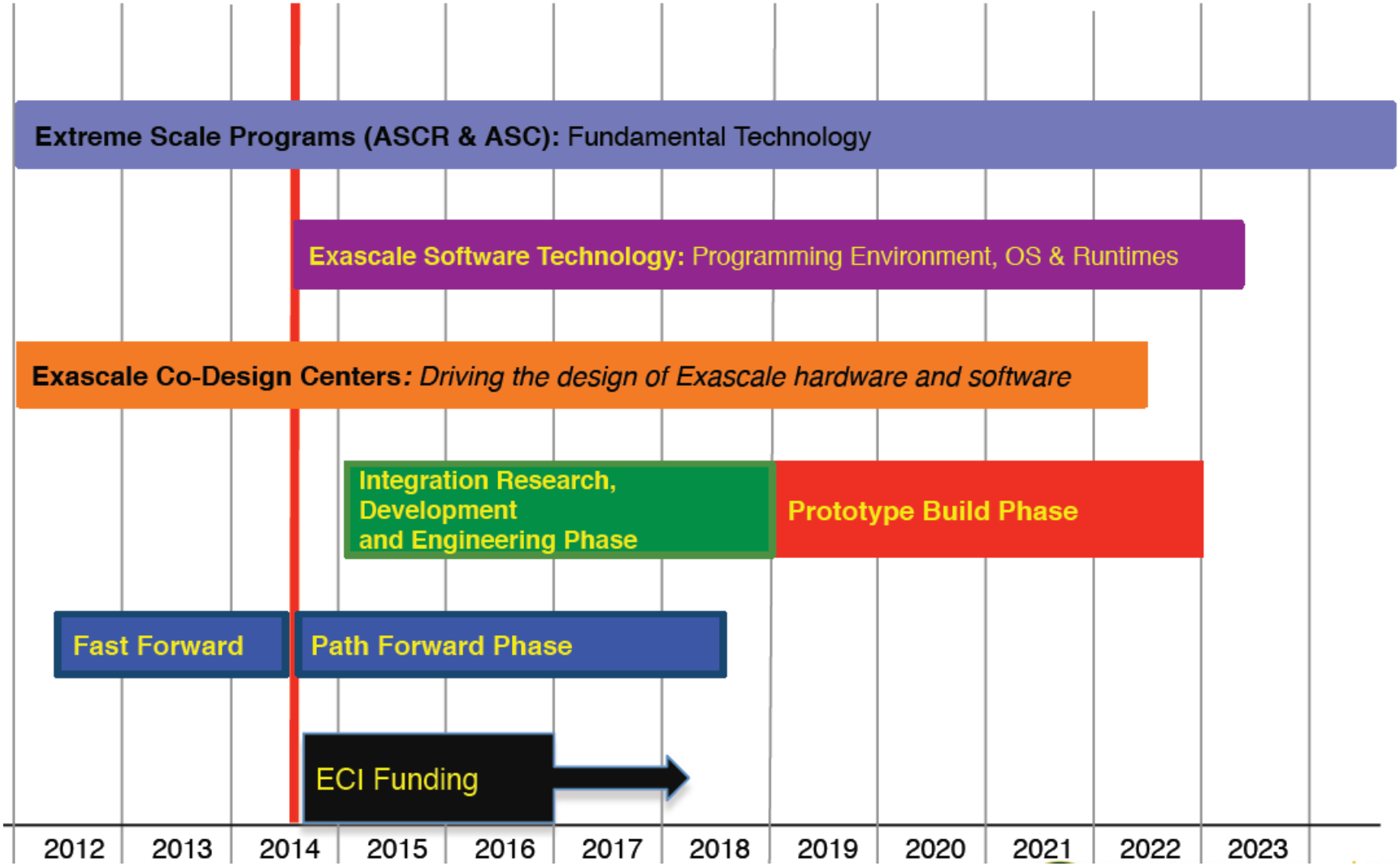
Оптические технологии

- Возможности высокоскоростной передачи данных по медным проводникам ограничены
  - 40Gbit/sec => несколько сантиметров по плате, ≈1 м по кабелю
- Оптическая передача данных обеспечивает намного лучшую плотность и энергоэффективность
- Перспективы развития:
  - Интеграция оптических трансиверов в одном модуле с логическими устройствами (опытные образцы есть уже сейчас): 2015-2016
  - Интеграция оптических волноводов на печатную плату: 2017-2018
  - Кремниевая фотоника, оптические коммуникации внутри микросхемы: после 2018

Co-design

- Exascale-суперкомпьютер будет представлять собой сложную иерархическую систему
  - Иерархия вычислительных мощностей
  - Иерархия данных
  - Иерархия интерконнекта
- Иерархия «железа» накладывается на новые классы задач:
  - Мультидисциплинарные и мультидоменные задачи
  - Нерегулярные шаблоны коммуникаций
  - Локальность данных жизненно необходима
- Невозможно построить универсальную систему, обеспечивающую эффективное решение всех классов задач
- Дизайн системы должен учитывать появление новых математических методов и алгоритмов, а разработка новых алгоритмов должна учитывать аппаратные особенности и ограничения – co-design!

Стратегия DOE в области Exascale



«Т-Платформы» и Exascale

- Система стоечного уровня интеграции A-Class
  - Единая базовая инфраструктура на ближайшие 3-5 лет
  - До 50 Pflops в 2013, до 200 Pflops в 2015-2016
  - Унифицированное решение для узлов на базе CPU, CPU+GPU, CPU+MIC
  - Энергоэффективные решения по электропитанию и охлаждению
  - Различные топологии интерконнекта
  - Выделенные сети для доступа к данным и управления/мониторинга

«Т-Платформы» и Exascale

- Исследования в области микроэлектроники
  - Дочерняя компания «Байкал Электроникс», специализирующаяся на разработках в области микроэлектроники
  - Исследования и разработки в области создания «систем на чипе» и интерконнекта
  - Кооперация с ведущими российскими и зарубежными специалистами
  - Первые продукты – 2015 год
- Кооперация с российскими разработчиками
  - МГУ, Томский ГУ, НИИ Квант, МЦСТ: совместные разработки в области суперкомпьютерных архитектур и аппаратных решений на базе отечественных компонентов

«Т-Платформы» и Exascale

- Международная коллаборация
  - Разработка и поставка прототипа PRACE в суперкомпьютерный центр CSC (Финляндия)
    - Совместный проект с суперкомпьютерными центрами CSC, SARA (Нидерланды) и CSCS (Швейцария)
    - Цель проекта: валидация энергоэффективных технологий, адаптация и анализ производительности ПО для гибридных архитектур
  - Разработка и поставка специализированной вычислительной системы для проекта QPACE II
    - Совместный проект с университетами Регенсбурга и Вупперталя (Германия)
    - Цель проекта: создание вычислительной системы с наилучшим соотношением производительность/ватт для решения задач квантовой хромодинамики
  - Поставка прототипа мультипетафлопсной системы в исследовательский центр Jülich (Германия)
    - Участники коллаборации: «Т-Платформы», Jülich, ParTec, Intel, Mellanox
    - Цель проекта: отработка программных решений в области управления и мониторинга большими суперкомпьютерными системами, исследования в области масштабируемости коммуникаций, обеспечения целостности данных и отказоустойчивости приложений

Роль России в Exascale

- Россия на сегодняшний день имеет достаточный опыт в области разработки суперкомпьютерных систем на базе готовых микроэлектронных компонентов, необходимо сосредоточить усилия по выводу отечественных продуктов на широкий мировой рынок и по их дальнейшему развитию
- Решение фундаментальных проблем Exascale в области компонентной базы и технологий, а также подготовка к созданию системы экзафлопсного уровня полностью внутри России требует интеграции всех имеющихся ресурсов и выделенного государственного финансирования
- В текущей ситуации особо важную роль играет международная, в первую очередь европейская коллаборация