НЕ ФЛОПСОМ ЕДИНЫМ… О ПАМЯТИ В СУПЕРКОМПЬЮТЕРАХ И РАЗВИТИИ ЕЕ ТЕХНОЛОГИЙ

28.09.2015

Андрей Слепухин
andrey.slepuhin@t-platforms.ru

Барьеры памяти в суперкомпьютинге

- Барьер #1: Задержки при обращении к памяти
  - Решается с помощью иерархии процессорных кэшей
- Барьер #2: Уменьшение удельной пропускной способности памяти (по отношению к производительности процессора)
- Барьер #3: Памяти нужно много… очень много…
  - Объем памяти должен расти как минимум вместе с производительностью процессора
  - Количество узлов в системе хочется минимизировать
  - Больше локальной памяти – меньше энергопотребление для решения задачи (joules-to-result)

Эволюция пропускной способности памяти: CPU

Processor	Launch	#cores	Flops/cycle	Peak DP FP/board, Gflops	Memory type	Peak RAM BW/board, GB/s	Byte/flop
Intel Xeon X5472 (Harpertown)	2007	4	4	96	DDR2-667 FBDIMM	21	0.22
Intel Xeon X5570 (Nehalem)	2009	4	4	94	DDR3-1333 RDIMM	64	0.68
Intel Xeon E5-2670 (Sandybridge)	2012	8	8	333	DDR3-1600 RDIMM	102.4	0.31
Intel Xeon E5-2698 v3 (Haswell)	2014	16	16	1178	DDR4-2133 RDIMM	136	0.12
Intel Skylake	2017?	28?	32	4000?	DDR4-2400?	230.4?	0.06?

- Intel Skylake: 6 каналов DDR4, более 3000 контактов!
- DDR4 плохо подходит на роль кандидата для памяти в будущих суперкомпьютерах

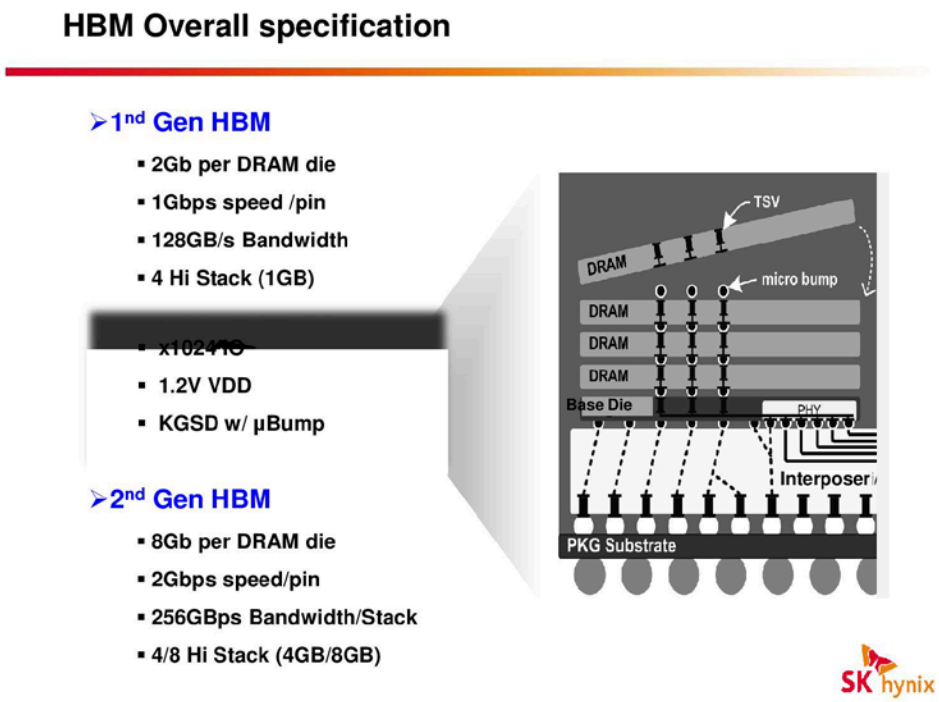
Эволюция пропускной способности памяти: GPGPU/ACC

Processor	Launch	Peak DP FP/board, Gflops	Memory type	Peak RAM BW/board, GB/s	Byte/flop
Nvidia C1060 (Tesla)	2008	78	GDDR3/1600MHz	102.4	1.31
Nvidia M2090 (Fermi)	2011	666	GDDR5/1848MHz	177	0.27
Nvidia K40 (Kepler)	2013	1430	GDDR5/2600MHz	288	0.20
Nvidia K80 (Kepler)	2014	2910	GDDR5/2505MHz	480	0.16
Intel Xeon Phi 7120P	2013	1208	GDDR5/2750MHz	352	0.29

- Пропускная способность памяти GDDR5 лимитируется размерами микросхемы и частотами
- Высокие частоты памяти затрудняют разводку платы и лимитируют объем памяти

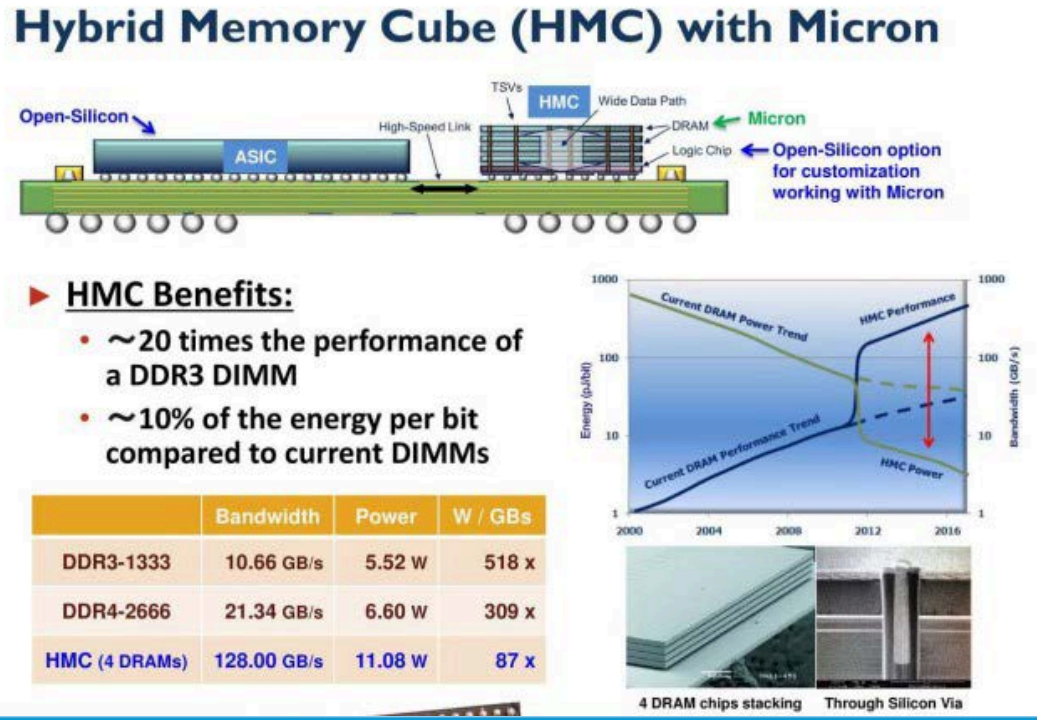
Новые типы памяти с высокой пропускной способностью

- HBM (High-Bandwidth Memory)
  - Интеграция непосредственно на подложку микросхемы рядом с процессором/ускорителем позволяет использовать очень широкую шину (1024 бит) и значительно уменьшить энергопотребление
  - Стандарт JEDEC (2013)
  - Первый продукт – GPU AMD Radeon R9 Fury X, пропускная способность памяти – 512GB/s при частоте всего 500MHz
  - Следующее поколение GPU от Nvidia (Pascal) будет использовать HBM
  - HBM в CPU?
    - Возможно AMD выпустит APU на базе своей новой архитектуры Zen



Новые типы памяти с высокой пропускной способностью

- Hybrid Memory Cube (HMC)
  - Основное преимущество – ***сериализованный*** интерфейс с возможностью построения цепочек чипов позволяет увеличивать не только пропускную способность, но и объем памяти
  - Потенциальная пропускная способность ***одной*** микросхемы HMC – 240GB/s (480GB/s bidirectional)
  - Применение
    - Intel Knights Landing – MCDRAM на одной подложке с микросхемой процессора является вариацией HMC
    - Процессор Fujitsu SPARC64 XIfx и суперкомпьютер PRIMEHPC FX100



Зачем нужно много памяти?

- Перемещение данных – дорого!

	2011	2018
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

  - *Источник: Jack Dongarra, 2012*
- DP DRAM read DDR4 – 3264pJ, HMC – 1024pJ
  - *Источник: Xilinx, 2015*
- Нужно уменьшить энергопотребление ⇒ локальность данных ⇒ «толстые» узлы

Где взять лишние гигабайты?

- DRAM – дорого, существенные ограничения по количеству каналов и модулей на канал
- NAND Flash – медленная и ненадежная
- Нужна новая память, желательно энергонезависимая

- Попытка #1: Micron&Intel 3D Xpoint
  - Примерно в 10 раз более плотная, чем DRAM
  - Намного быстрее, чем NAND Flash
  - Можно использоваться в виде стандартных модулей DDR4

- Что дальше?
  - Потенциально следует ожидать появления нового типа энергонезависимой памяти в районе 2018 года, с выходом на массовый рынок после 2020 года

Возможная иерархия памяти для экзафлопной машины

- Производительность узла: ~20-30Tflops (процессор 1-2Tflops + ускорители)
- Память ускорителей: одной подложке с кристаллом (HBM или HMC), 64GB, ~2TB/s
- Пямять процессора:
  - HMC, 0.25-0.5TB, 1TB/s
  - NVRAM, 2-3TB
    - Идеальный вариант – NVRAM с интерфейсом HMC и той же пропускной способностью

- На текущий момент самым критическим местом представляется интерконнект между процессором и ускорителями, его производительность должна быть сбалансирована с памятью процессора

СПАСИБО!