Суперкомпьютер IBM Roadrunner

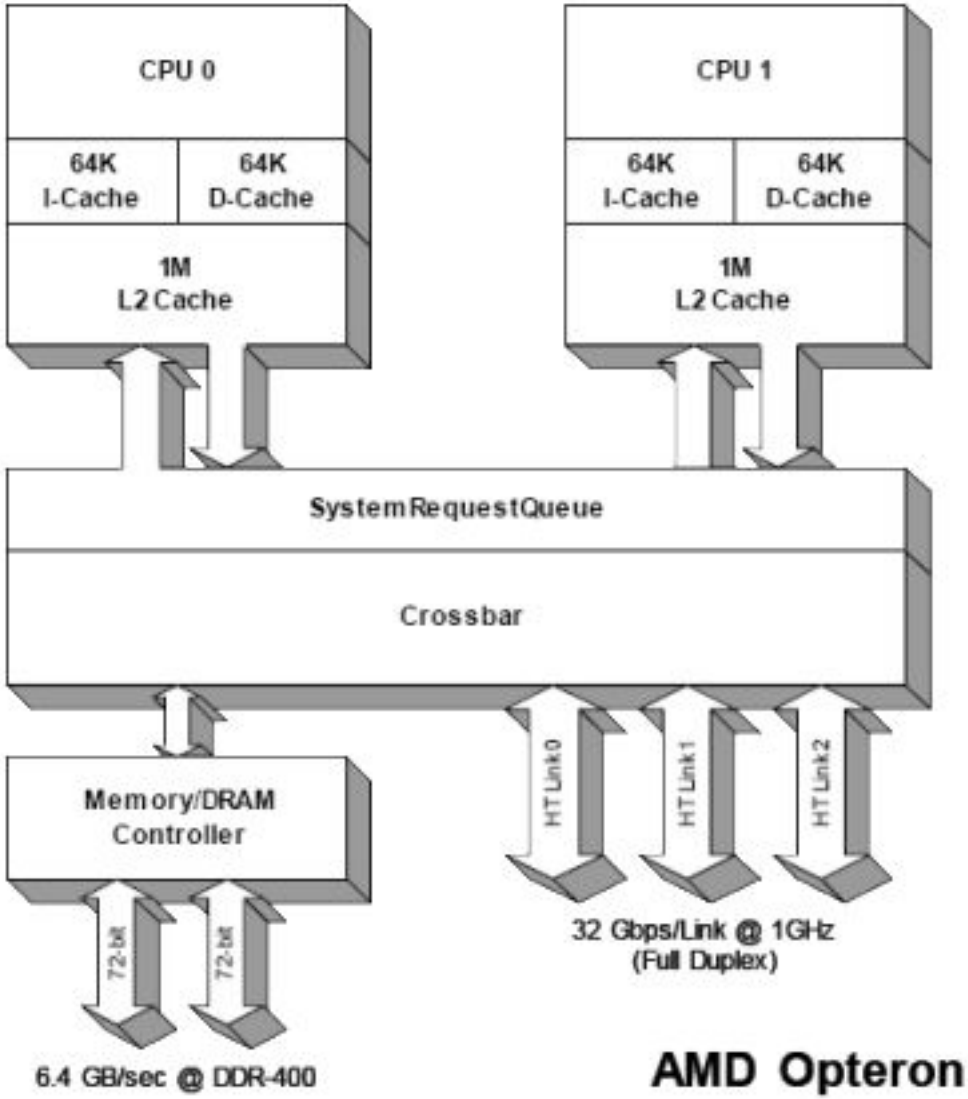
Общая информация

- Roadrunner – суперкомпьютер, построенный IBM для Лос-Аламосской национальной лаборатории в Нью-Мексико (США).
- 25 мая 2008 года он достиг 1,026 петафопс и занял первое место в рейтинге ТОП500 суперкомпьютеров, поддерживая производительность в 1,0 петафлопс в бенчмарке LINPACK
- Roadrunner занимал примерно 296 серверные стойки площадью 560 $м^2$ и начал работать в 2008 году.
- Он был списан 31 марта 2013 года.
- Министерство энергетики США использовало Roadrunner для расчёта старения ядерных материалов и анализа безопасности и надёжности ядерного арсенала США. Также он использовался для научных, финансовых, транспортных и аэрокосмических расчётов.

Строение

- Суперкомпьютер IBM Roadrunner
  - Узел Connected Unit (CU)
    - Модуль TriBlade
      - Блейд-сервер IBM BladeCenter LS21
        - Процессор AMD Opteron
      - Блейд-сервер IBM BladeCenter QS22
        - Процессор IBM PowerXCell 8i

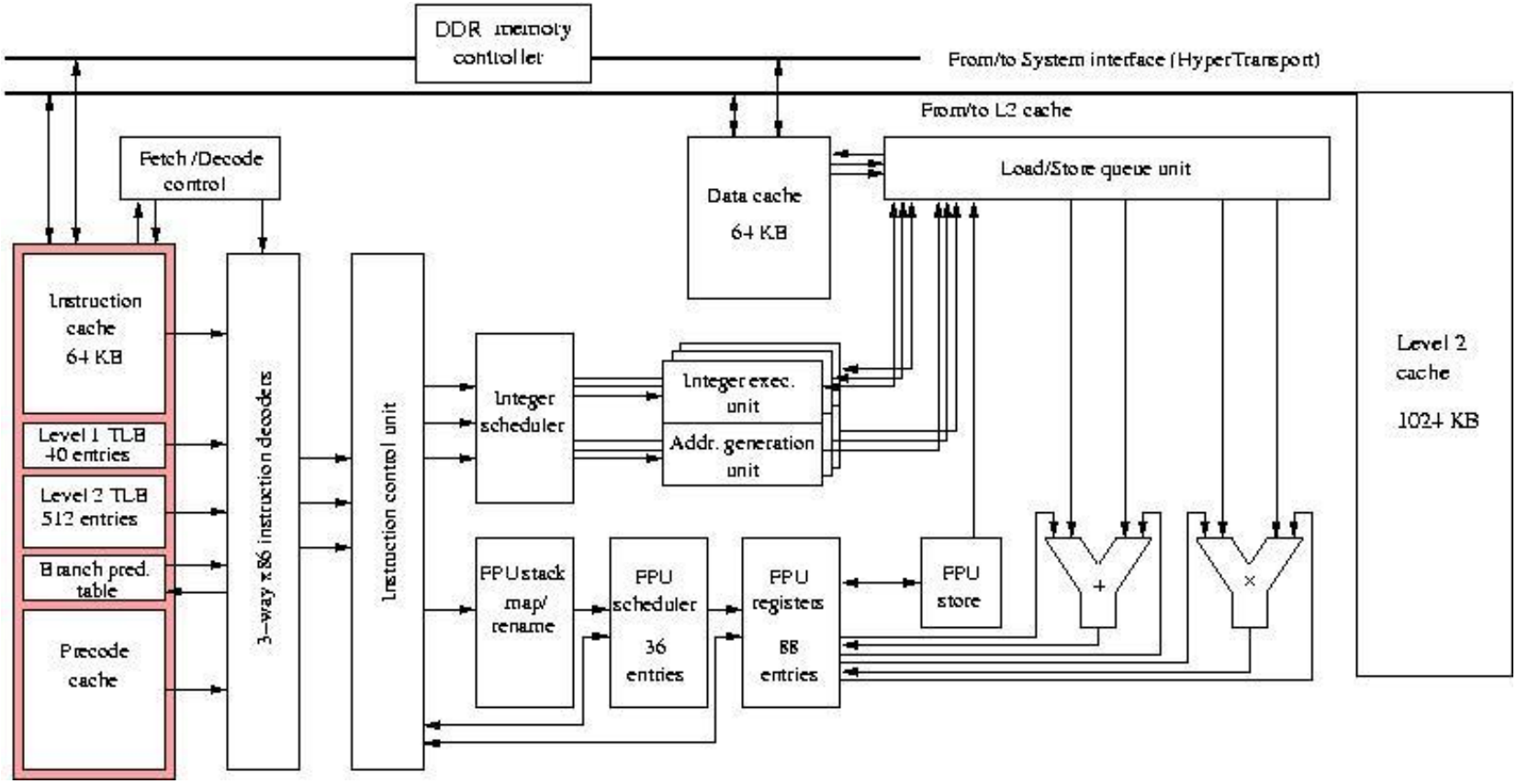
Процессор AMD Opteron



- Архитектура AMD64 (x86-64)
- Процессор SMP
- Многопроцессорная система NUMA

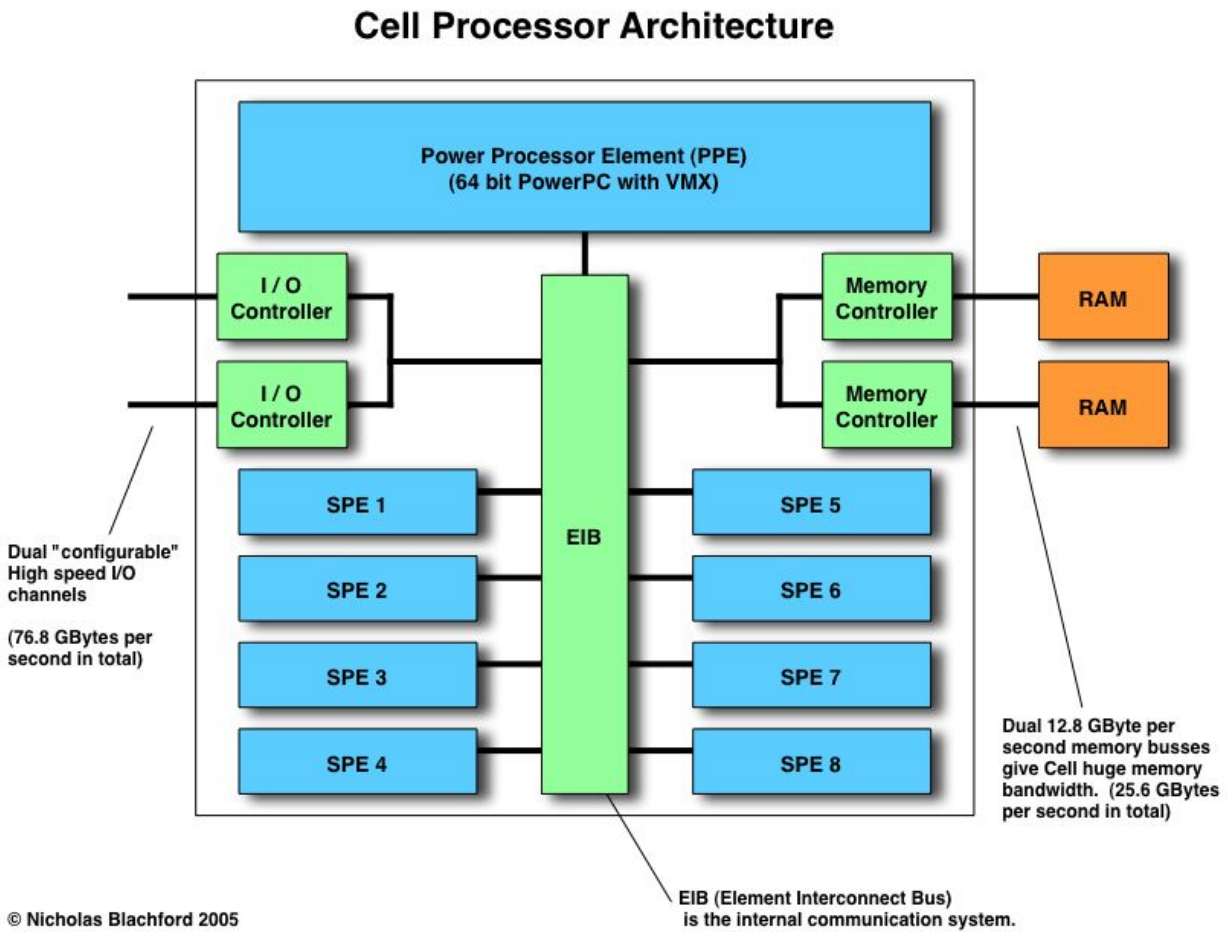
P_{SMP} (AMD Opteron Core) = $I_{pv\ 64}$[2 Core (AMD Opteron Core), U (Crossbar), CtrM (Memory/DRAM Controller), 3 U (HT Link)]

Ядро процессора AMD Opteron



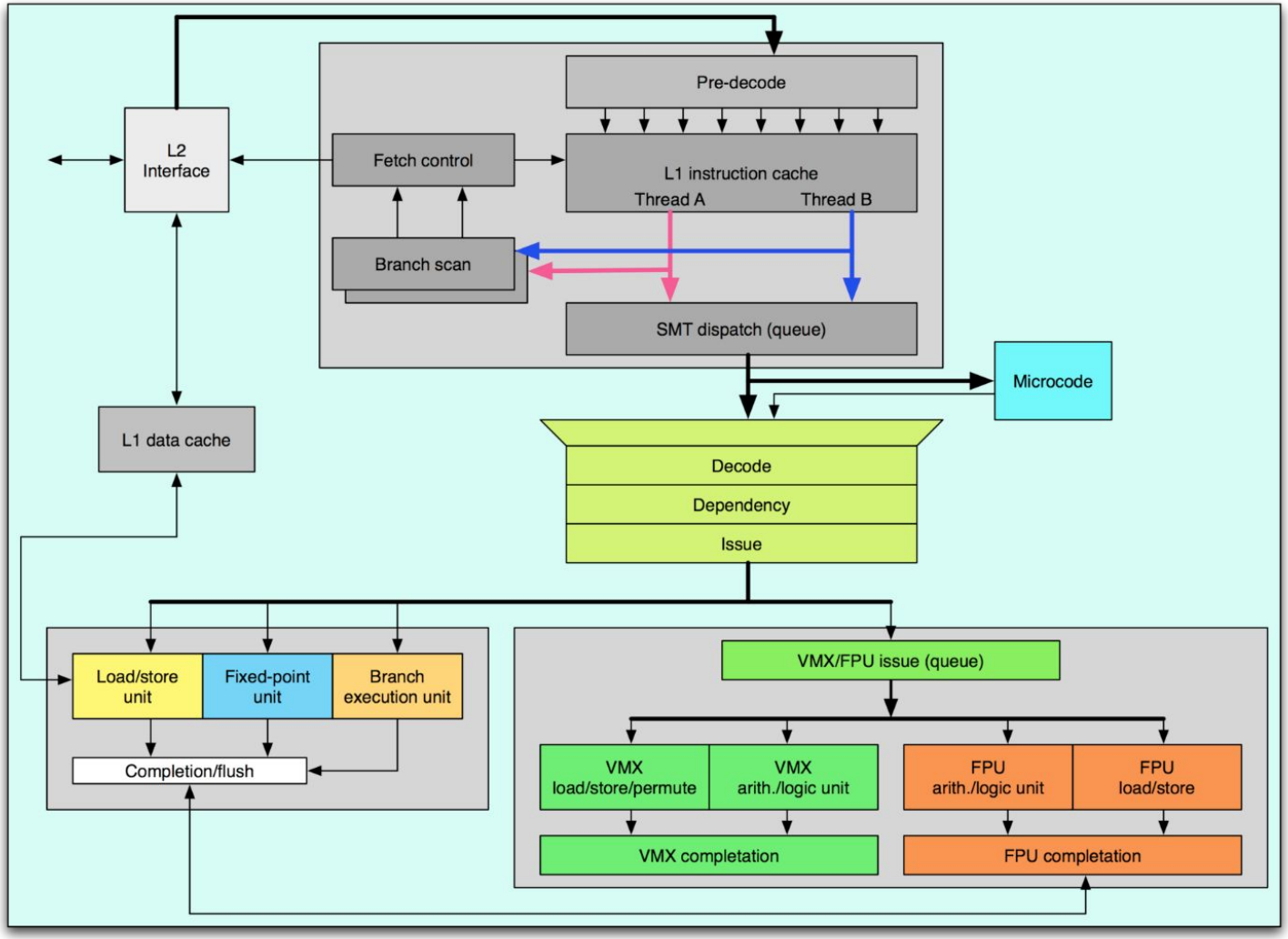
Core (AMD Opteron Core) = $\{Rg_{64}, B, F, Cshi1_{64KB}, Cshd1_{64KB}, Csh2_{1024KB}\}$

Процессор IBM PowerXCell 8i



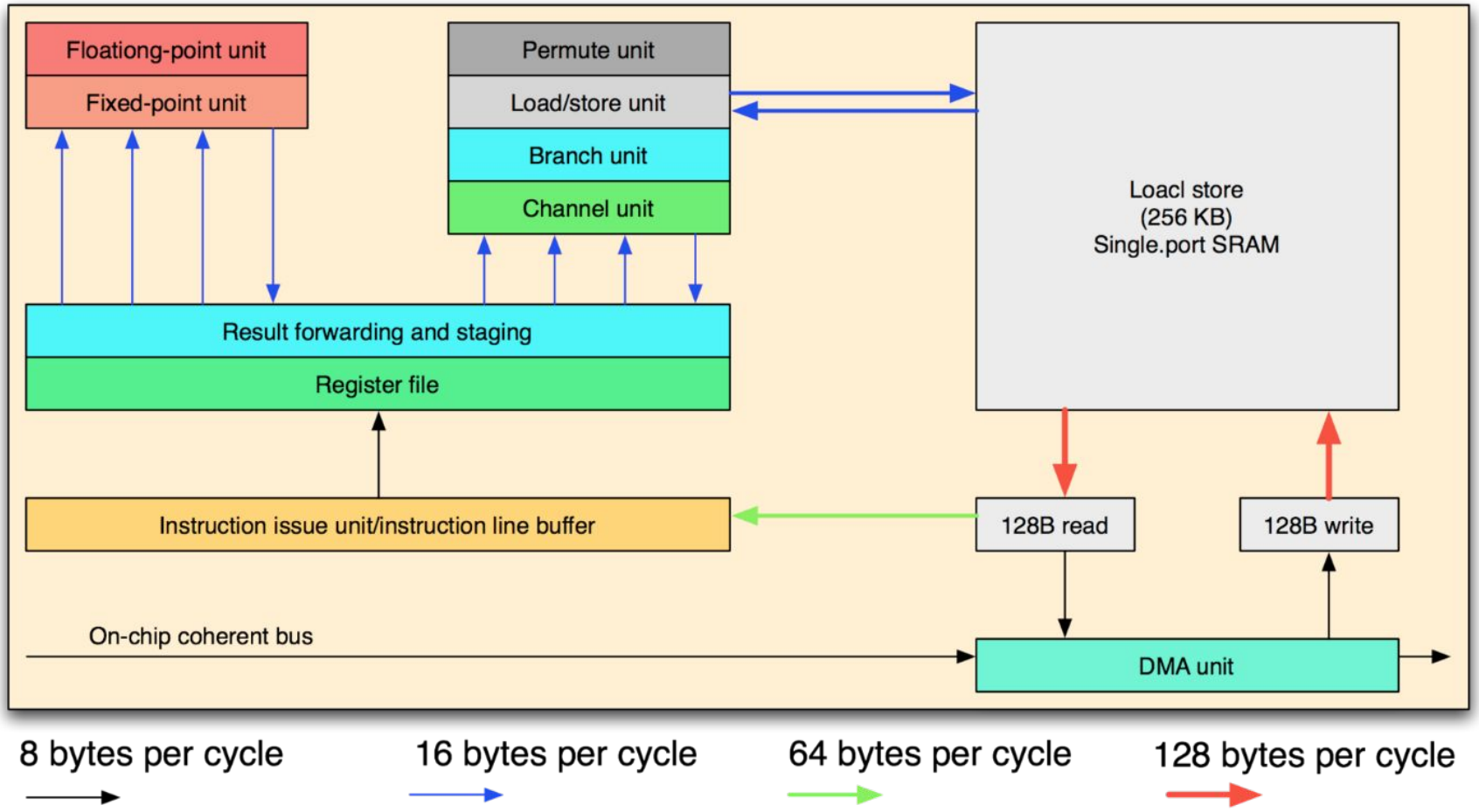
- 1 PPE (PowerPC Processor Element)
- 8 SPE (Synergistic Processor Element)
- Совместно используют память

IBM PowerXCell 8i PPE



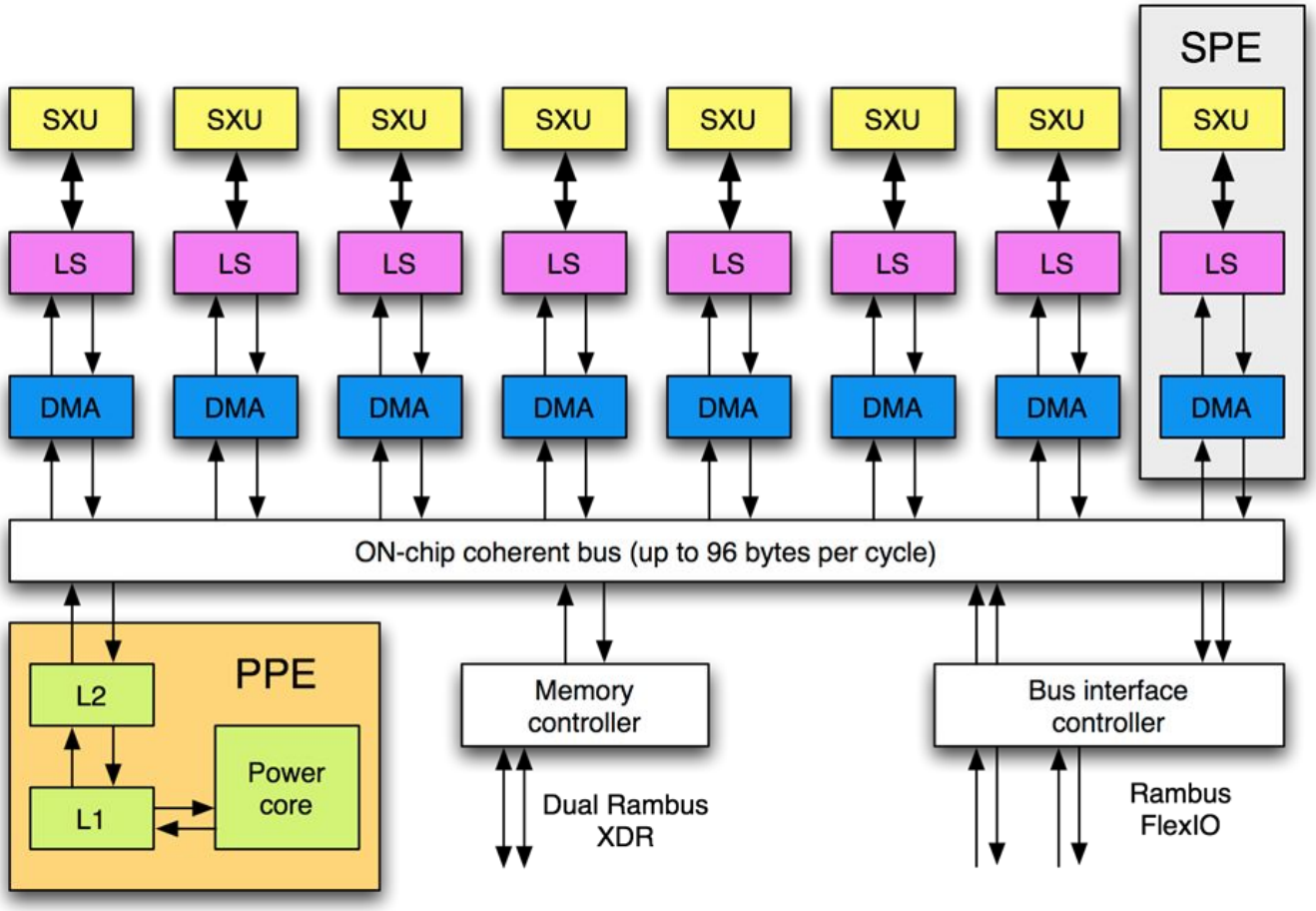
Core (IBM PowerXCell 8i PPE) = $\{Rg_{64}, B, F, Cshi1_{32KB}, Cshd1_{32KB}, Csh2_{512KB}\}$

IBM PowerXCell 8i SPE



Core (IBM PowerXCell 8i SPE) = $\{Rg_{64}, B, F, M_{256KB\ (SRAM)}, U(DMA)\}$

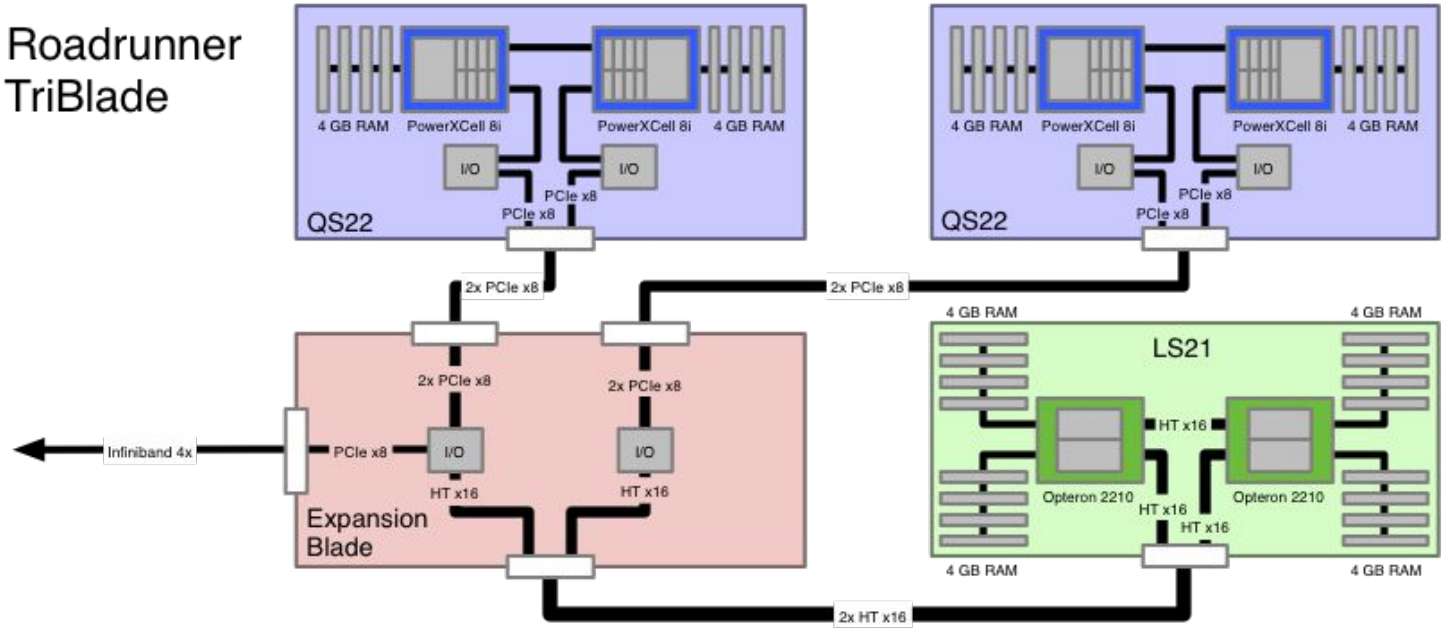
IBM PowerXCell 8i



- NUMA, так как SPE работает с LS, а не с общей памятью, но имеет DMA

P_{NUMA} (IBM PowerXCell 8i) = $I_{pv\ 64}$[Core (IBM PowerXCell 8i PPE), 8 Core (IBM PowerXCell 8i SPE), U (Element Interconnect Bus), 2 CtrM (Memory Controller), 2 U (I/O Controller)]

Модуль TriBlade

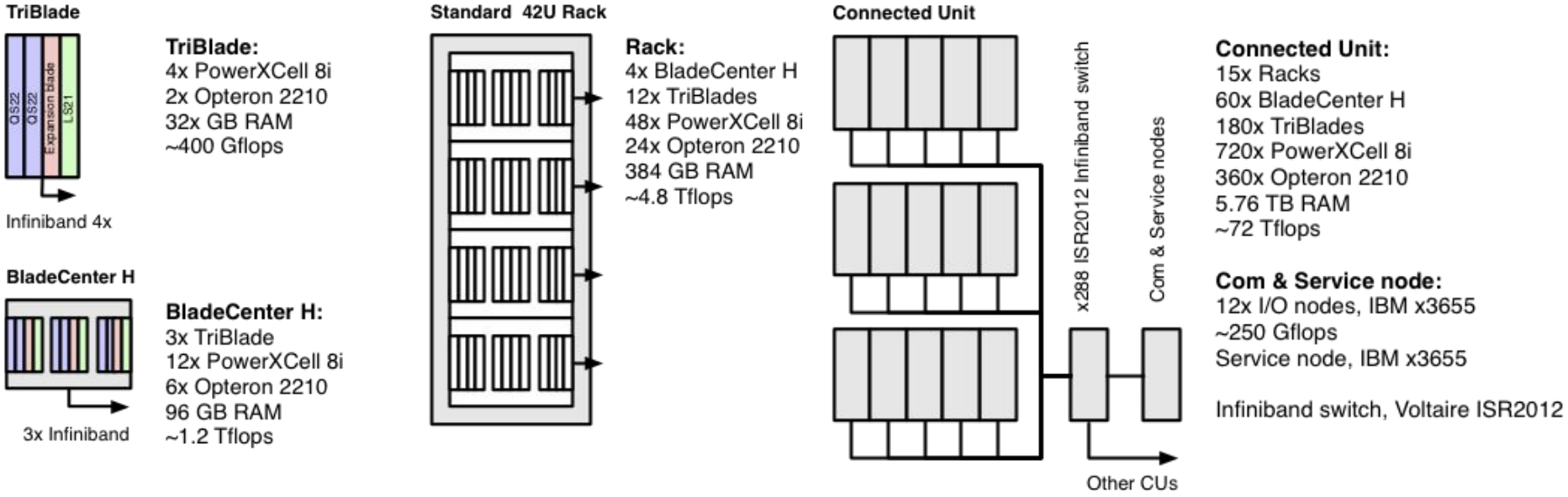


Bld_{NUMA} (IBM BladeCenter LS21) = {2 $M_{4GB\ (DDR\ SDRAM)}$ ^{200MHz} <- 6.4 GB/s -> P_{SMP} (AMD Opteron Core) <- Hyper Transport x16 32Gb/s 1GHz-> P_{SMP} (AMD Opteron Core) <- 6.4 GB/s -> 2 $M_{4GB\ (DDR\ SDRAM)}$ ^{200MHz}}

Bld_{NUMA} (IBM BladeCenter QS22) = {$M_{4GB\ (DDR2\ SDRAM)}$ ^{400MHz} <- 6.4 GB/s -> P_{NUMA} (IBM PowerXCell 8i) <- -> P_{NUMA} (IBM PowerXCell 8i) <- 6.4 GB/s -> $M_{4GB\ (DDR2\ SDRAM)}$ ^{400MHz}}

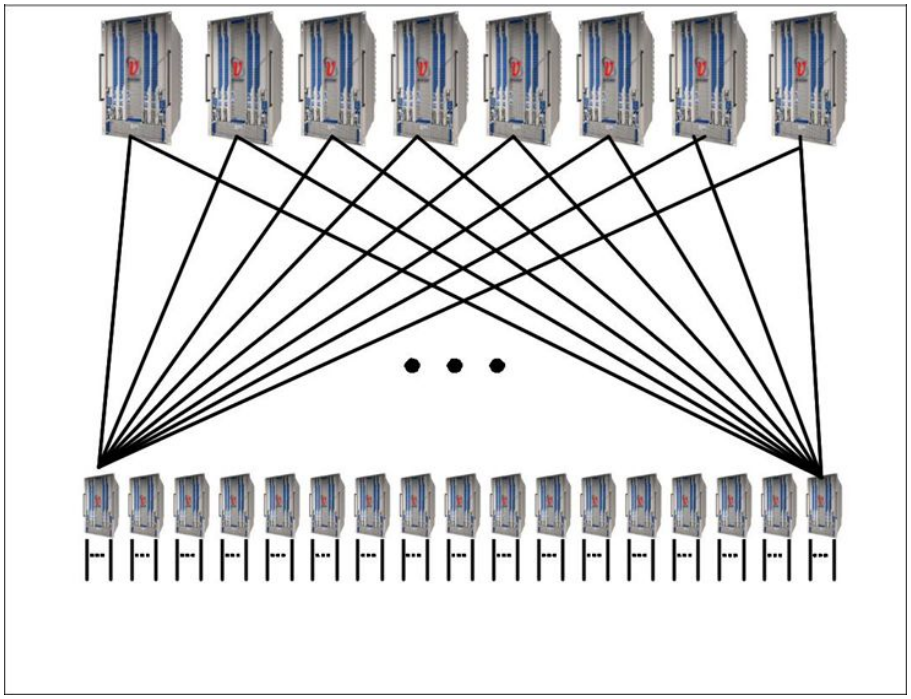
$Node_{NUMA}$ (TriBlade) = {Bld_{NUMA} (IBM BladeCenter LS21) <- 2x Hyper Transport x16 32Gb/s 1GHz -> Bld (Expansion Blade) <- 4x PCIe x8 32Gb/2 -> 2 Bld_{NUMA} (IBM BladeCenter QS22)}

Узел Connected Unit (CU)



$Node_{ClusterHPC}$ (Connected Unit) = < Switch (Voltaire ISR2012 $_{2GB/s}$) >
[180 $Node_{NUMA}$ (TriBlade), I/O, SS]

Суперкомпьютер IBM Roadrunner



- 17 узлов CU (планировалось 18)
- 8 коммутаторов второго этапа Infiniband ISR2012 (6 необходимо)
- 12 восходящих линий до каждого (96 всего)

$Super_{ClusterHPC}$ (IBM Roadrunner) = < 8 Switch (Voltaire ISR2012 $_{2GB/s}$) > [17 $Node_{ClusterHPC}$ (Connected Unit)]

Структурная нотация

Core (AMD Opteron Core) = $\{Rg_{64}$, B, F, $Cshi1_{64KB}$, $Cshd1_{64KB}$, $Csh2_{1024KB}\}$

P_{SMP} (AMD Opteron Core) = $I_{pv\ 64}$[2 Core (AMD Opteron Core), U (Crossbar), CtrM (Memory/DRAM Controller), 3 U (HT Link)]

Bld_{NUMA} (IBM BladeCenter LS21) = $\{2\ M_{4GB\ (DDR\ SDRAM)}^{200MHz}$ <- 6.4 GB/s -> P_{SMP} (AMD Opteron Core) <- Hyper Transport x16 32Gb/s 1GHz-> P_{SMP} (AMD Opteron Core) <- 6.4 GB/s -> 2 $M_{4GB\ (DDR\ SDRAM)}^{200MHz}\}$

Core (IBM PowerXCell 8i PPE) = $\{Rg_{64}$, B, F, $Cshi1_{32KB}$, $Cshd1_{32KB}$, $Csh2_{512KB}\}$

Core (IBM PowerXCell 8i SPE) = $\{Rg_{64}$, B, F, $M_{256KB\ (SRAM)}$, U(DMA)$\}$

P_{NUMA} (IBM PowerXCell 8i) = $I_{pv\ 64}$[Core (IBM PowerXCell 8i PPE), 8 Core (IBM PowerXCell 8i SPE), U (Element Interconnect Bus), 2 CtrM (Memory Controller), 2 U (I/O Controller)]

Bld_{NUMA} (IBM BladeCenter QS22) = $\{M_{4GB\ (DDR2\ SDRAM)}^{400MHz}$ <- 6.4 GB/s -> P_{NUMA} (IBM PowerXCell 8i) <-> P_{NUMA} (IBM PowerXCell 8i) <- 6.4 GB/s -> $M_{4GB\ (DDR2\ SDRAM)}^{400MHz}\}$

$Node_{NUMA}$ (TriBlade) = {Bld_{NUMA} (IBM BladeCenter LS21) <- 2x Hyper Transport x16 32Gb/s 1GHz -> Bld (Expansion Blade) <- 4x PCIe x8 32Gb/2 -> 2 Bld_{NUMA} (IBM BladeCenter QS22)}

$Node_{ClusterHPC}$ (Connected Unit) = < Switch (Voltaire ISR2012 $_{2GB/s}$) > [180 $Node_{NUMA}$ (TriBlade), I/O, SS]

$Super_{ClusterHPC}$ (IBM Roadrunner) = < 8 Switch (Voltaire ISR2012 $_{2GB/s}$) > [17 $Node_{ClusterHPC}$ (Connected Unit)]

Расчёт пиковой производительности

AMD Opteron = 2 FLOP/такт * 1,8 ГГц * 2 ядра = 7,2 GFLOPS

IBM BladeCenter LS21 = 2 процессора * 7,2 GFLOPS (AMD Opteron) = 14,4 GFLOPS

IBM PowerXCell 8i PPE = 2 FLOP/такт * 3,2 ГГц = 6,4 GFLOPS

IBM PowerXCell 8i SPE = 2 FLOP/такт * 2 слова * 3,2 ГГц = 12,8 GFLOPS

IBM PowerXCell 8i = 6,4 GFLOPS (PPE) + 8 * 12,8 GFLOPS (SPE) = 108,8 GFLOPS

IBM BladeCenter QS22 = 2 процессора * 108,8 GFLOPS (PowerXCell 8i) = 217,6 GFLOPS

TriBlade = 14,4 GFLOPS (LS21) + 2 * 217,6 GFLOPS (QS22) = 449,6 GFLOPS

Connected Unit = 180 * 449,6 GFLOPS (TriBlade) = 80,928 TFLOPS

IBM Roadrunner = 17 * 80,928 TFLOPS (Connected Unit) = 1,375776 PFLOPS

Значение в рейтинге ТОП500 = 1375,78 TFLOPS

Спасибо за внимание