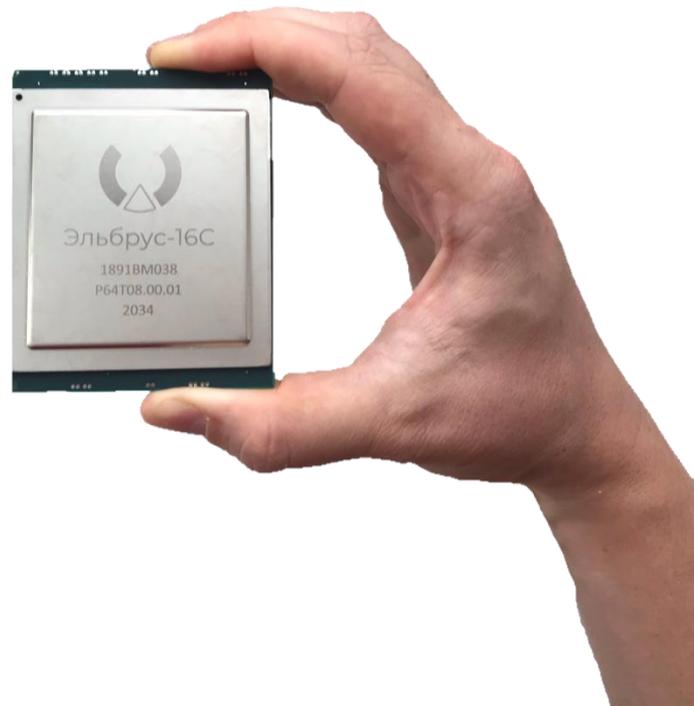




СИСТЕМА-НА-КРИСТАЛЛЕ ЭЛЬБРУС-16С: ОСНОВНЫЕ ПРОЕКТНЫЕ РЕШЕНИЯ

Алексей Кожин

Заместитель главного конструктора
АО МЦСТ



МЭС-2020, 24 ноября 2020

Цель:

- **Разработка универсального микропроцессора на 16нм для высокопроизводительных серверов, СХД, НРС**

Задачи:

- Увеличить однопоточную производительность
- Увеличить число процессорных ядер
- Реализовать аппаратную поддержку виртуализации
- Оптимизировать иерархию кэш-памяти для серверных задач
- Повысить пропускную способность подсистемы памяти и подсистемы ввода-вывода
- Разработать новую сеть-на-кристалле
- Улучшить технологию энергосбережения
- Улучшить технологии отказоустойчивости и повышения выхода годных микропроцессоров

Параллельная энергоэффективная архитектура

- 25-48 оп. за такт за счет явного аппаратного параллелизма операций
- Высокая однопоточная производительность при малом энергопотреблении
- Распараллеливание оптимизирующим компилятором
- Сбалансированное соотношение производительности процессора и памяти

Эффективная двоичная совместимость с Intel x86, x86-64

- Базируется на аппаратно-поддержанной технологии динамической двоичной компиляции
- Исполнение Windows XP, 7, 10, Linux, QNX
- Исполнение приложений в кодах кодах x86/x86-64
- Производительность до 80% от нативной (в кодах Эльбруса)

Технология безопасных вычислений

- Аппаратная защита логической структуры памяти
- Гарантированное обнаружение атак, нарушающих структуру памяти
- Повышение надежности программ
- Повышение скорости отладки программ



МИКРОПРОЦЕССОРЫ ЭЛЬБРУС



**Эльбрус
130нм**

1-ое поколение

- 5 GFLOPS
- 1 ядро
- 300 МГц

2Н07



**Эльбрус-2С+
90нм**

2-ое поколение

- 28 GFFLOPS
- 2 ядра + 4 DSP
- 500 МГц

1Н11



**Эльбрус-4С
65нм**

3-ое поколение

- 50 GFLOPS
- 4 ядра
- 800 МГц
- 8 MB L2

2Н13



**Эльбрус-8С
28нм**

4-ое поколение

- 250 GFLOPS
- 8 ядер
- 1.3 ГГц
- 16 MB L3

2Н15



**Эльбрус-8СВ
28нм**

5-ое поколение

- 0.58 TFLOPS
- 8 ядер
- 1.5 ГГц
- SIMD-128
- 16 MB L3

2Н18



**Эльбрус-16С
16нм**

6-ое поколение

- 1.5 TFLOPS
- 16 ядер
- 2.0 ГГц
- Виртуализация
- 16 MB L2, 32 MB L3
- SoC

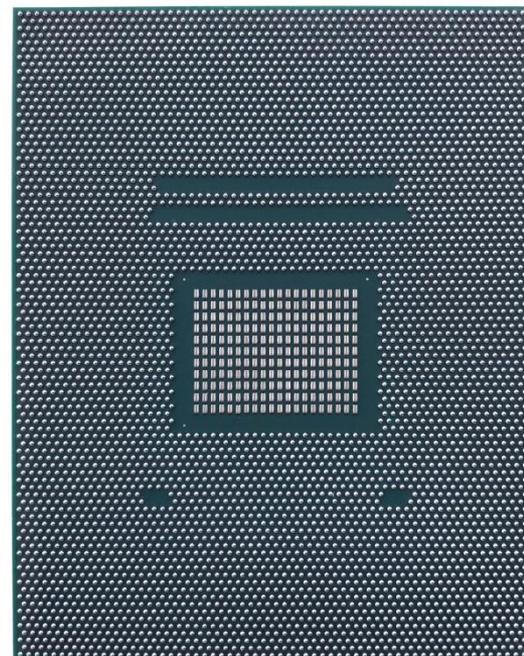
2Н21

ЭЛЬБРУС-16С

- 1.5 TFLOPS (FP32) / 0.75 TFLOPS(FP64)
- 16 ядер Эльбрус v6 @2 ГГц (SIMD-128, FMA, ...)
- Виртуализация
- Новая подсистема кэш-памяти, L3 кэш 32 МВ
- Распределенная сеть-на-кристалле
- Система-на-Кристалле
 - 8 каналов памяти DDR4-3200*
 - 32 PCIe 3.0 lanes
 - SATA, ETH, USB, Legacy IO
 - 3 канала межпроцессорных обменов x16 @12Gbs*
- Многопроцессорность (2S-4S конфигурации)
- Встроенные средства Power Management, RAS
- 16 нм FinFET, 618 мм² (25.3x24.4), ~12В транзисторов
- Мощность не более 130 Вт

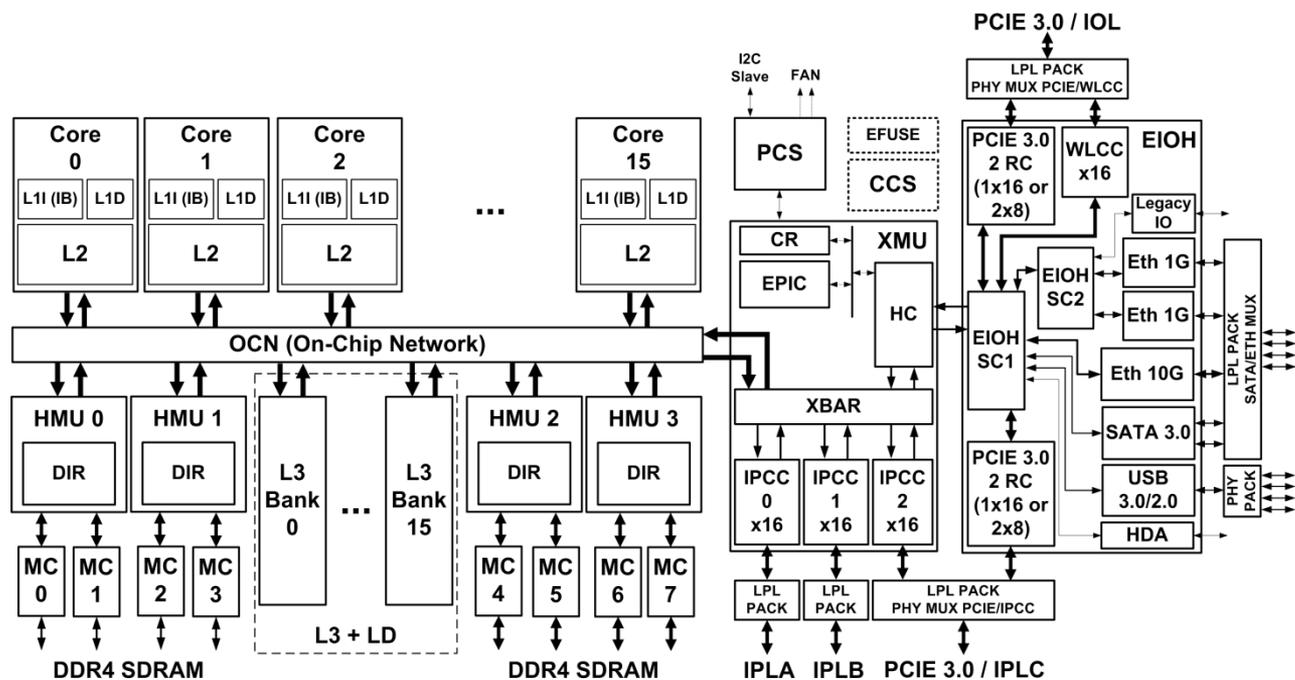
* Достижимость пиковых значений определяется конфигурацией и характеристиками платы





Корпус 63x78 мм, HFСВGA 4804

Структурная схема



LD – локальный справочник

DIR – глобальный справочник

HMU – Home Memory Unit

MC – Memory Controller

OCN – сеть-на-кристалле

XMU – eXternal Memory Unit

HC – Host-контроллер

CR – блок конфигурационных регистров

EPIC – Elbrus Programmable Interrupt Controller

PCS – Power Control System

CCS – Clock Control System

EIOH – Embedded IO-Hub

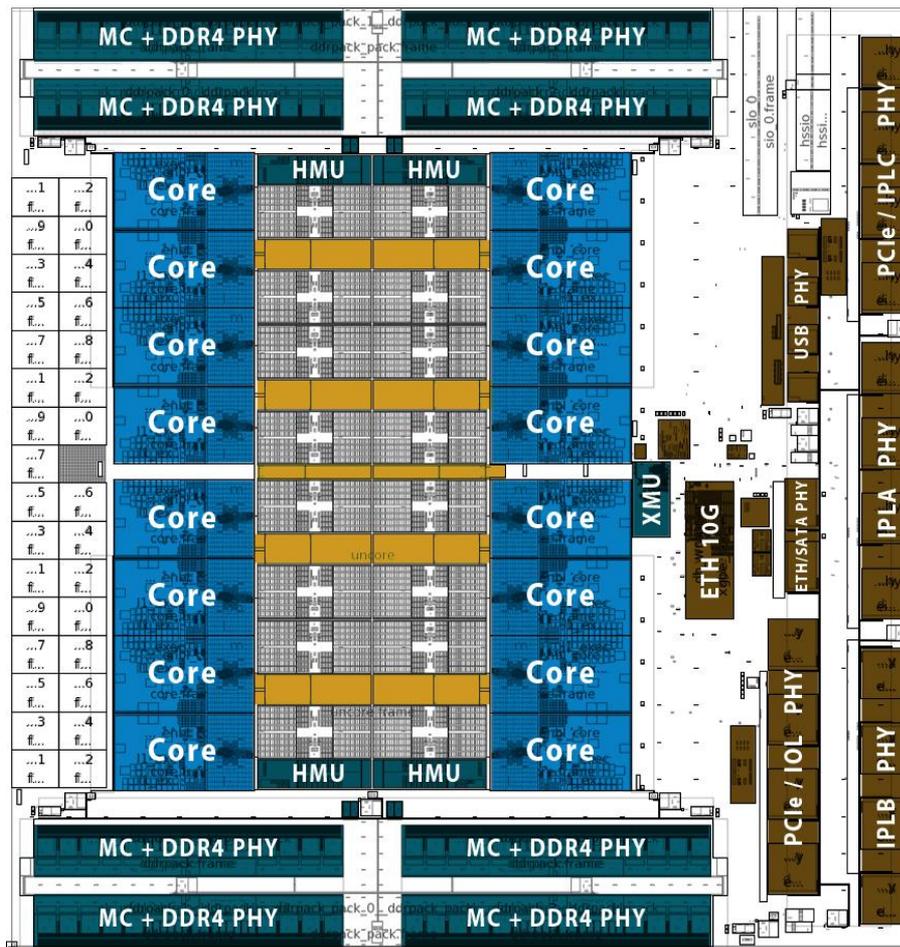
IPCC – контроллер межпроцессорного линка

WLCC – контроллер линка ввода/вывода

IPL – межпроцессорный линк (A, B, C)

IOL – линк ввода-вывода (КПИ-2)

ТОПОЛОГИЧЕСКАЯ СХЕМА



Core – процессорное ядро с частной кэш-памятью L1I, L1D, L2

L3 – общая кэш-память третьего уровня со встроенным локальным справочником LD

HMU – Home Memory Unit со встроенным глобальным справочником DIR

OCN – сеть-на-кристалле

MC – Memory Controller

DDR4 PHY – физуровень DDR4

XMU – eXternal Memory Unit

PCS – Power Control System

CCS – Clock Control System

EIOH – Embedded IO-Hub

IPCC – контроллер межпроцессорного линка

WLCC – контроллер линка ввода/вывода

ETH 10G – контроллер Ethernet 10G

IPL PHY – физуровень межпроцессорного линка (A, B)

PCIe/IPLC PHY – физуровень PCIe 3.0/межпроцессорного линка C

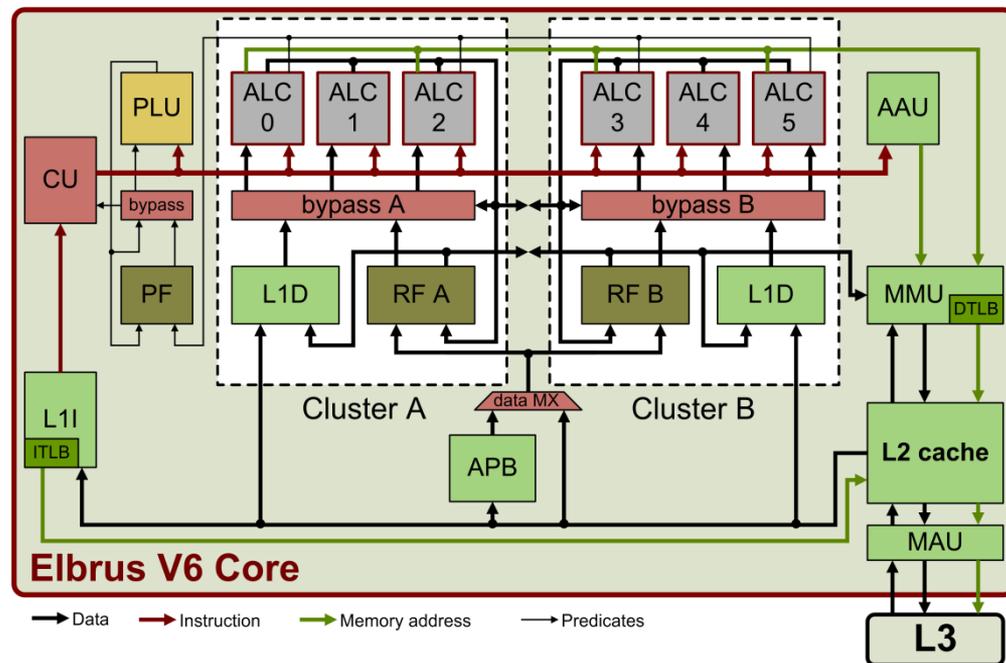
PCIe/IOL PHY – физуровень PCIe 3.0/линка ввода-вывода КПИ-2

- Расширение разрядности физического адреса с 40 до 48 бит
- Аппаратная поддержка виртуализации
 - Вычислительные ресурсы
 - Память
 - Периферия
 - Межпроцессорные и внешние прерывания
- Эффективная реализация невыровненных обращений в память
- Введение операции FMA (Fused Multiply-Add по стандарту IEEE-754-2008)
- Введение оптимизированных барьерных операций
- Введение дополнительных упакованных операций SIMD-128
- Аппаратная поддержка динамических оптимизаций
- Оптимизация операций доступа в память под новую подсистему кэш-памяти
 - Управление размещением в кэш-памяти, Prefetch for Write, Clear Cache Line
- Новый контроллер прерываний EPIC (Elbrus Programmable Interrupt Controller)



ПРОЦЕССОРНОЕ ЯДРО

- Архитектура Эльбрус V6
 - 25/48 операций за такт
- Повышение тактовой частоты до 2 ГГц
 - Редизайн микроархитектуры ядра
 - Custom-блоки (RF, L1D)
- Операции FMA во всех 6 каналах
- Оптимизация AAU / APB
- Увеличение размера L2 кэша с 512 КБ до 1 МБ
- Увеличение размера MAU (Memory Access Unit)
 - для чтений с 48 до 80
 - для записей с 32 до 40
- Производительность 96 GFLOPS (FP32) / 48 GFLOPS (FP64)



ПОДСИСТЕМА КЭШ-ПАМЯТИ

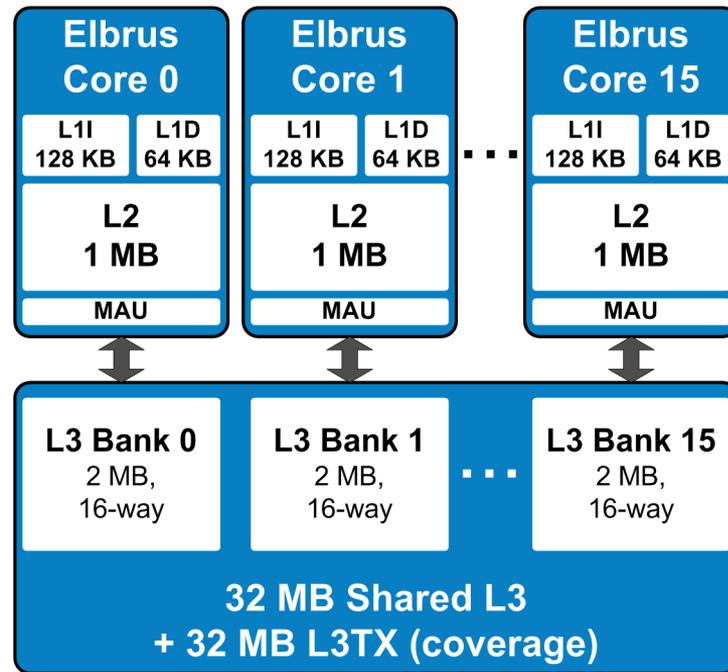
- **Кэш-память ядра:**

- L1I: 128 KB, 256B line, 4-way
- L1D: 64 KB, 32B line, 4-way
- L2: 1 MB, 64B line, 4-way, неинклюзивная

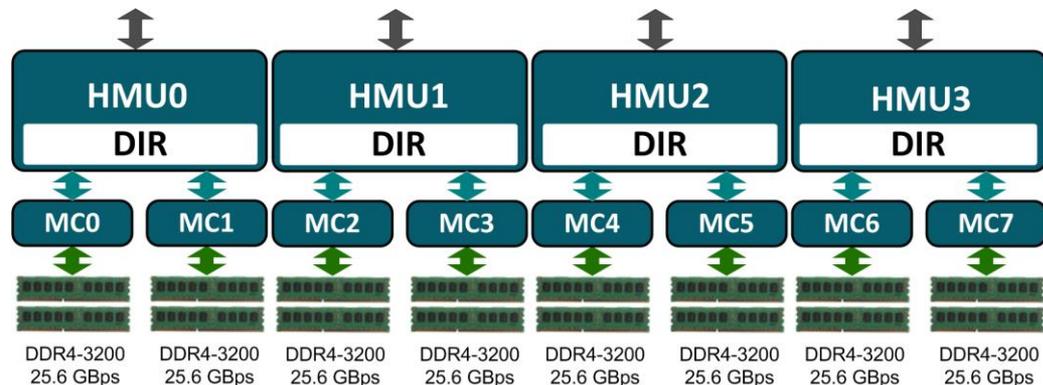
- **Кэш-память L3, общая для всех ядер:**

- 32 MB, 64B line, 16-way (data), неинклюзивная
- Инклюзивный локальный справочник, 32-way (tags)
- 16 независимых банков (S-NUCA)
 - По умолчанию разделение по 9,8,7,6 битам адреса
- Байпасирование данных
- Cache or Victim Mode
- Адаптивная политика замещения
- DCA: Direct Cache Access по DMA
- QoS с программным управлением
 - Way Partitioning, 32 QoS ID, Way Mask для кода/данных для каждого QoS ID

- **Суммарный объем кэш-памяти 51 MB**



- 4 блока НМУ (Home Memory Unit)
 - Сериализация запросов в память и поддержка когерентности
 - Глобальный справочник DIR
- 8 контроллеров памяти МС (Memory Controller)
 - Увеличен объем реестра запросов
 - Адаптивные фильтры планировщика



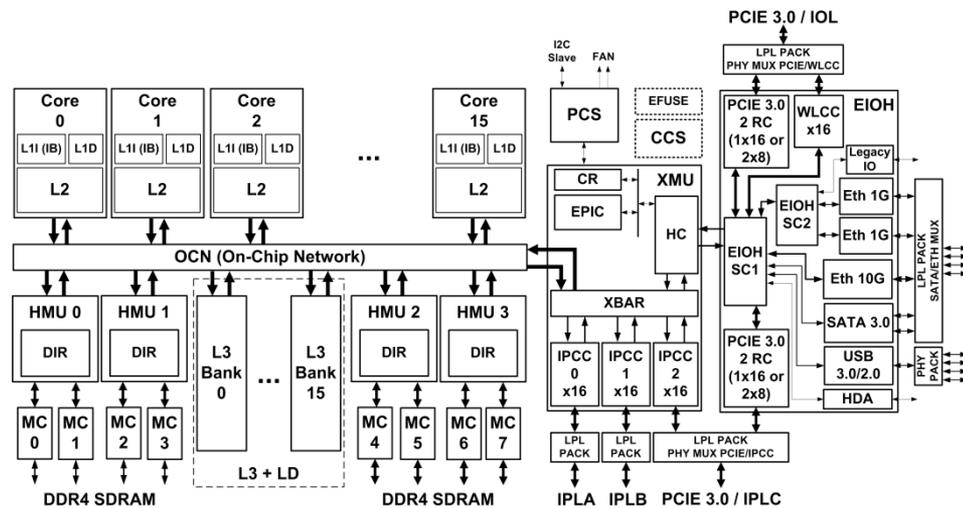
- 8 каналов памяти, 16 слотов DIMM
- Теоретическая пропускная способность до 200 Гбайт/с (DDR4-3200*)
- Объем – до 4 TB на процессор, до 16 TB в 4-процессорной системе
- Поддержка модулей UDIMM, RDIMM, LRDIMM и 3DS

* Достижимость пиковых значений определяется конфигурацией и характеристиками платы

ПОДДЕРЖКА КОГЕРЕНТНОСТИ ДАННЫХ

Два уровня сериализации

- Общий L3 кэш (16 запросов/такт)
 - Сериализация кэшируемых запросов от ядер
- Устройства доступа в память и ввод-вывод
 - Сериализация запросов в оперативную память (HMU, 4 запроса/такт)
 - Сериализация запросов в ввод-вывод (НС, 1 запрос за такт)
- Оптимальный трафик обращений



Двухуровневый протокол когерентности

- Локальный справочник LD (L3 кэш) – 25 состояний
 - L3 кэш – MOESI+: Modified, Owned, Exclusive, Shared, Invalid + Reserved, Bypassed, Deleted
 - Кэш-память ядра – MOESI
- Глобальный справочник DIR (HMU) – MOSI
 - Уменьшение когерентного трафика и времени доступа в память

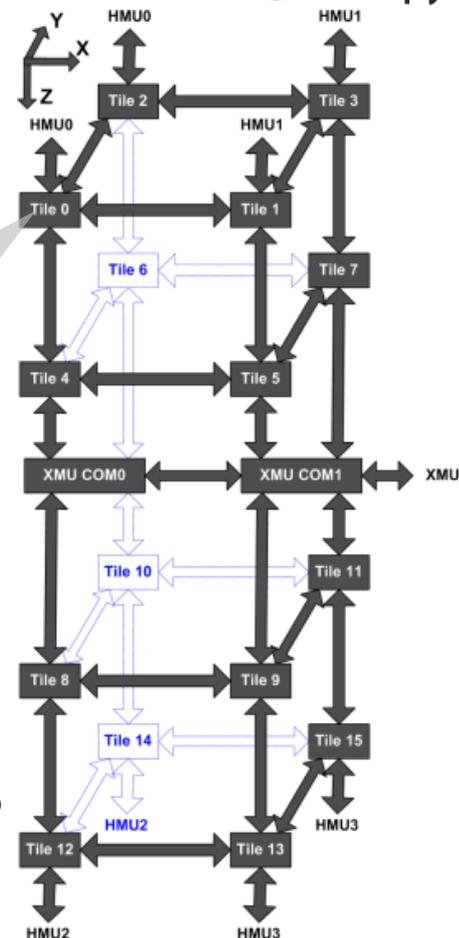
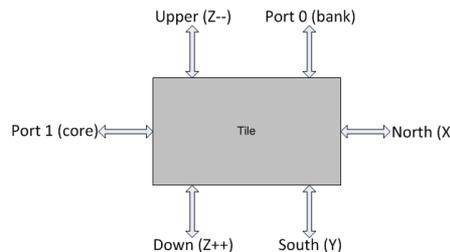
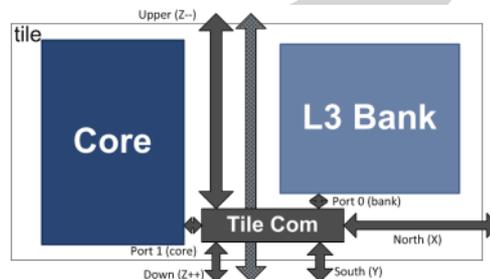
СЕТЬ-НА-КРИСТАЛЛЕ

Elbrus System Protocol (ESP)

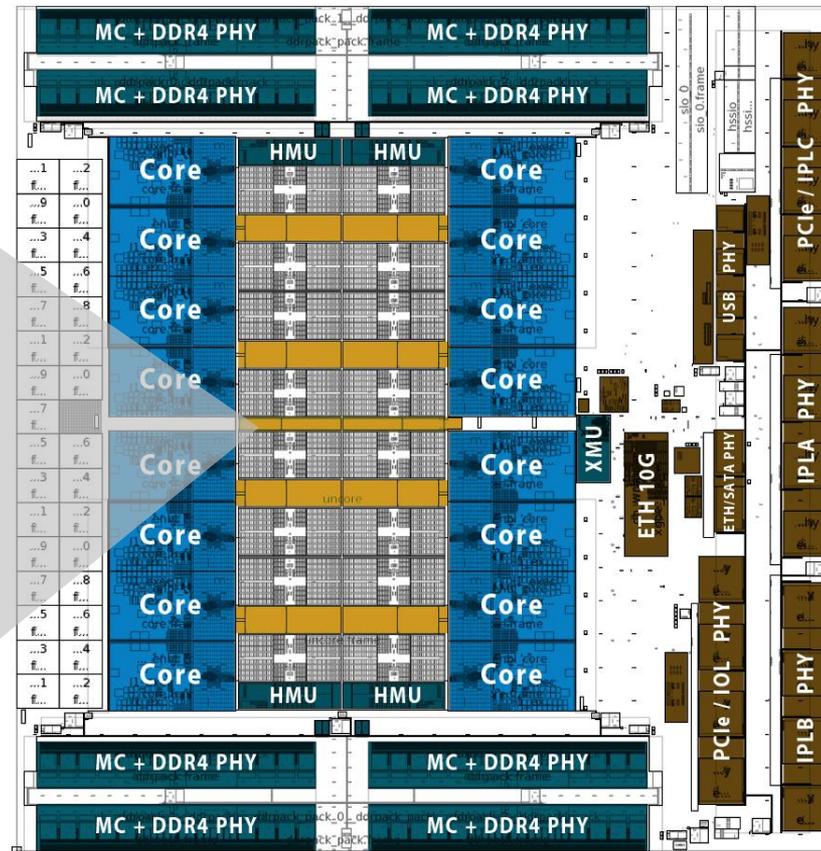
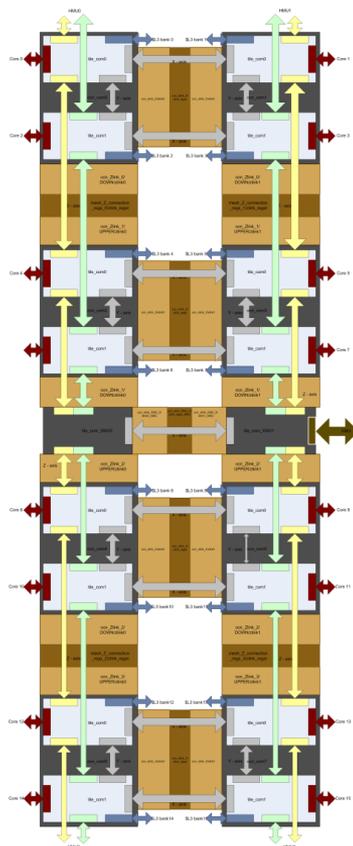
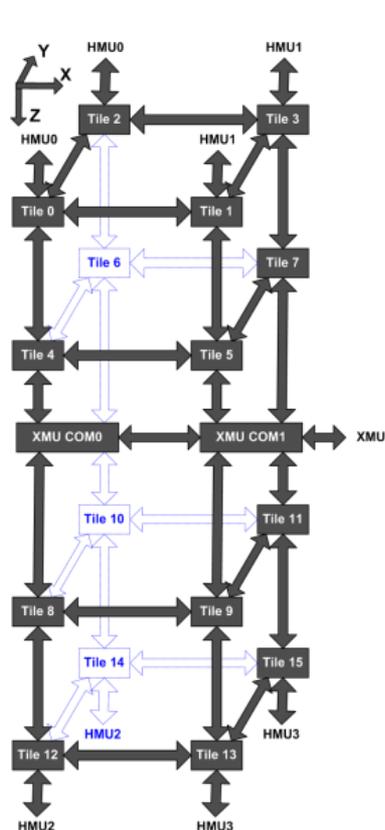
- Протокольный уровень (Protocol Layer)
 - MOESI+, >50 системных операций (base+hints, Clear Cache Line, Evicted, Write-Back Clean, Atomics)
- Транзакционный уровень (Transaction Layer)
 - Несколько уровней сериализации, байпасирование
- Транспортный уровень (Transport Layer)
 - Единый формат сообщений, 14 логических каналов, QoS
- Сетевой уровень (Network Layer)
 - Unicast и Multicast, сквозная маршрутизация пакетов
- Канальный уровень (Link Layer)
 - Физические и виртуальные каналы, credit-based flow control, multi-flit packets
- Физический уровень (Physical Layer)

On-Chip-Network, OCN

- 16 тайлов (ядро + банк L3 кэша) + 4 HMU + XMU
- Топология 2d-torus-mesh 4x5
- Статическая маршрутизация Z-X-Y
- 7 физических каналов



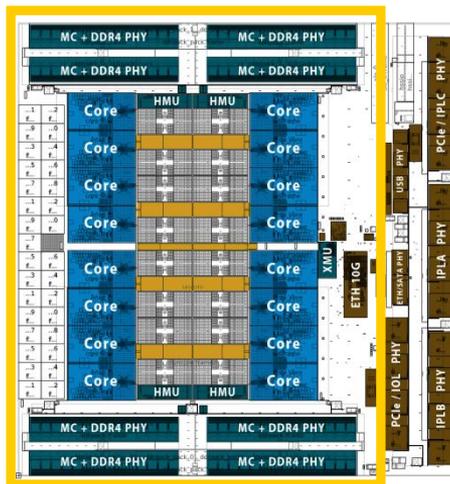
СЕТЬ-НА-КРИСТАЛЛЕ



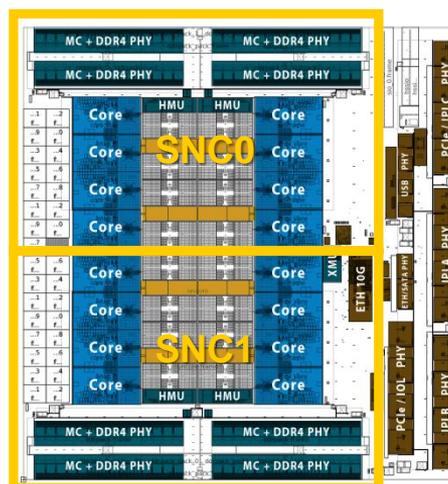
Пропускная способность: суммарная до 2 ТБ/с, bisectional bandwidth до 512 ГБ/с

СЕТЬ-НА-КРИСТАЛЛЕ

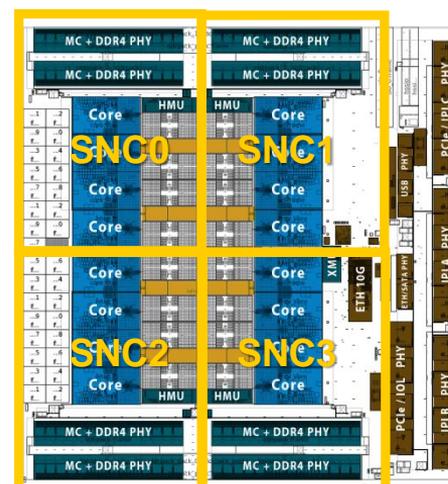
- Промежуточная буферизация пакетов, credit-based flow control, multi-flit packets
- Unicast & Multicast, поддержка работы мимо L3 кэша, Quality-of-Service
- Поддержка отключения банков L3, HMU, MC
- Настраиваемый интерливинг L3, HMU, MC
- Настраиваемые хэш-маски интерливинга для равномерной загрузки L3 кэша и сети
- Sub-NUMA Clustering



No Clustering



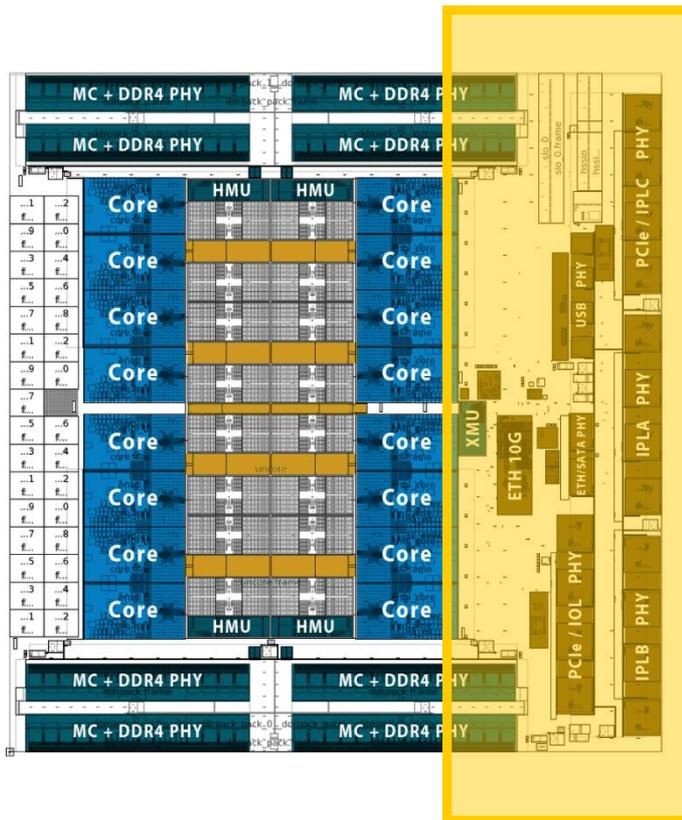
2 Clusters



4 Clusters

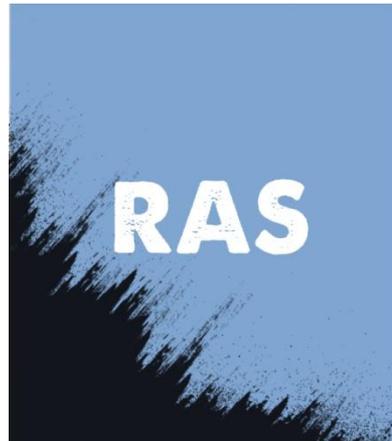
- **4 PCIe 3.0 RC, 32 линии**
- **1 блок PCIe/IOL x16**
 - 1x16 PCIe 3.0
 - 1x16 IOL
 - 2x8 PCIe 3.0
 - 1x8 PCIe 3.0 + 1x8 IOL
- **1 блок PCIe/IPL x16**
 - 1x16 PCIe 3.0
 - 1x16 IPL 12G*
 - 2x8 PCIe 3.0
 - 1x8 PCIe 3.0 + 1x8 IPL 12G*
- **2x16 IPL 12G***
- **Мультилинк (2/3 IPL)**

* Достижимость пиковых значений определяется конфигурацией и характеристиками платы



- **2 порта SATA 3.0**
- **2 порта SATA / Ethernet**
 - 2x SATA
 - 1x1G (SGMII)+1x10G (BASE-KR)
 - 2x1G (SGMII)
 - 1x1G (SGMII)+1x2.5G (SGMII)
 - 1x2.5G (SGMII)+1x10G (BASE-KR)
 - 2x2.5G (SGMII)
- **4 порта USB 3.0/2.0**
- Низкоскоростные интерфейсы:
 - I2C (5+1 порт),
 - SPI (на 4 устройства),
 - RS-232 (2 порта),
 - GPIO (16 входов/выходов),
 - INTEL HDA

- **Встроенные средства самотестирования**
 - Проверка на старте, защита от старения
 - BIST с column redundancy (2), Don't Use bits, последовательный запуск
 - EFUSE для хранения конфигурации каждого экземпляра процессора (256 строк x 21 бит)
- **Кэш-память ядра**
 - Четность и ECC SECDED
- **L3 кэш**
 - ECC SECDED для памяти тэгов и данных
 - Demand & Patrol Scrubbing для тэгов и данных
 - Dynamic Bad Line Disabling
- **Сеть-на-кристалле**
 - Четность для всех пакетов и интерфейсов всех устройств
- **Оперативная память**
 - ECC SECDED
 - Patrol Scrubbing
 - Повтор выдачи последовательности команд после обнаружения ошибок чётности на адресно-командной шине и ошибок CRC для данных записи



- Первый отечественный универсальный микропроцессор на технологии 16нм
- Первый отечественный универсальный микропроцессор с производительностью более 1 TFLOPS
- Первая высокопроизводительная СпК с архитектурой Эльбрус
- Спроектирован для серверного применения
- Получены инженерные образцы, идёт тестирование



Эльбрус-16С

```

top - 17:56:27 up 29 min, 1 user, load average: 0,00, 0,00, 0,00
Tasks: 180 total, 1 running, 179 sleeping, 0 stopped, 0 zombie
%Cpu0 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu1 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu2 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu3 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu4 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu5 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu6 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu7 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu8 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu9 :  0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu10 : 0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu11 : 0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu12 : 0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu13 : 0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu14 : 0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
%Cpu15 : 0,0 us,  0,0 sy,  0,0 ni,100,0 id,  0,0 wa,  0,0 hi,  0,0 si,  0,0 st
MiB Mem : 63803,7 total, 62846,1 free, 618,3 used, 339,3 buff/cache
MiB Swap:  0,0 total,  0,0 free,  0,0 used. 62645,2 avail Mem

```

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
1	root	20	0	5548	2376	2256	S	0,0	0,0	0:35.40	init
2	root	20	0	0	0	0	S	0,0	0,0	0:00.00	kthreadd

```

e16c-stand-01 /export # lscpu
Архитектура: e2k
Порядок байт: Little Endian
CPU(s): 16
On-line CPU(s) list: 0-15
Thread(s) per core: 1
Ядер на сокет: 16
Сокетов: 1
NUMA node(s): 1
ID производителя: Elbrus-MCST
Семейство ЦПУ: 6
Модель: 0
Имя модели: E16C
CPU MHz: 2000.452
VogoMIPS: 4000.07
L1d cache: 64K
L1i cache: 128K
L2 cache: 1024K
L3 cache: 32768K
NUMA node0 CPU(s): 0-15
e16c-stand-01 /export # █

```

БЛАГОДАРЮ ЗА ВНИМАНИЕ



Alexey.S.Kozhin@mcst.ru, +7(916)592-95-75