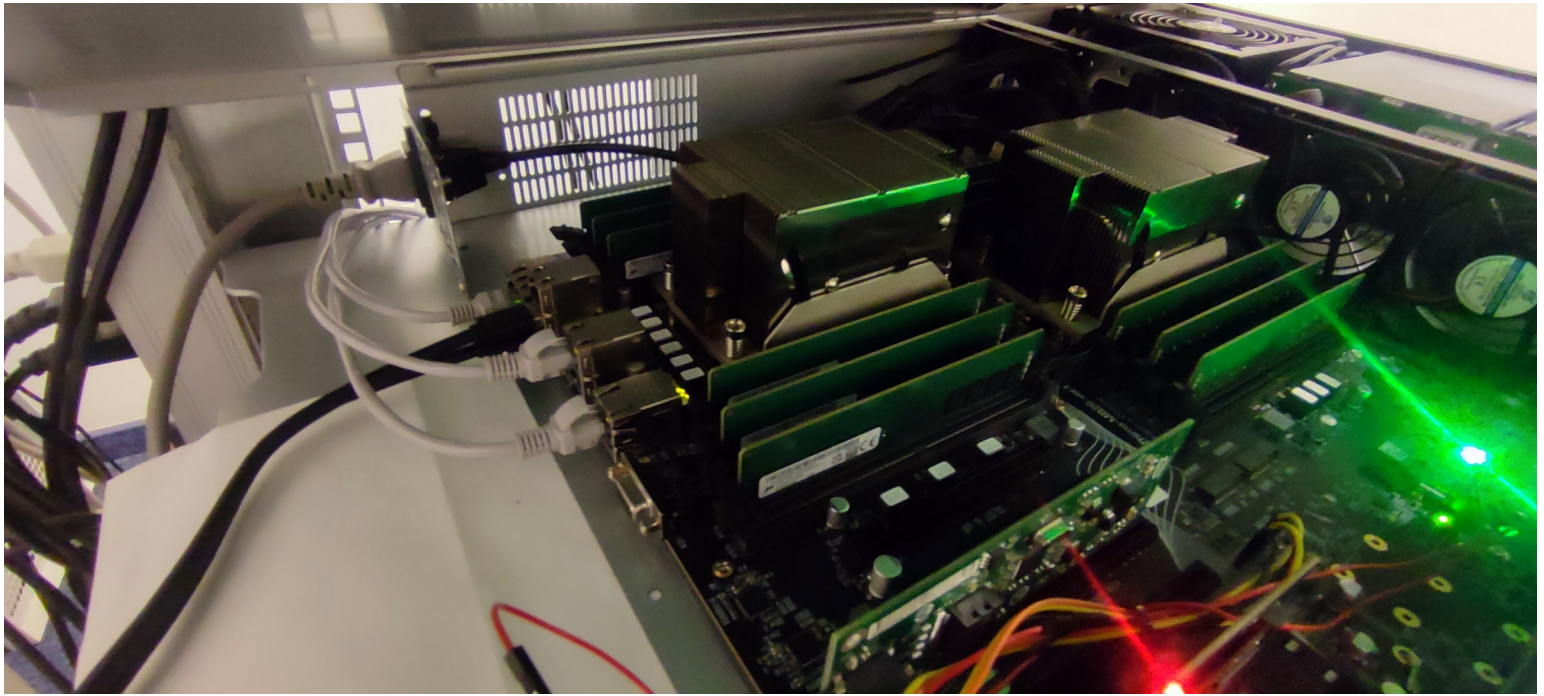
Benchmarking BE-S1000 SoC by Baikal Electronics



Maxim Kaurkin, Gregory Khrenov, Vladimir Gusev

Moscow, MSU, 2024

https://russianscdays.org/

План презентации

- Конфигурация оборудования
- Результаты HPL. Сравнение с *Intel Xeon 6230 и Kunpeng 920-4826*

- Результаты SPEC CPU 2017. Сравнение с *Intel Xeon 6230 и Kunpeng 920-4826*
- *Выводы*

Tests list

- **HPLinpack 2.3** http://www.netlib.org/benchmark/hpl/, http://top500.org

  The LINPACK Benchmarks are a measure of a system's floating-point computing power. They measure how fast a computer solves a dense n by n system of linear equations Ax = b, which is a common task in engineering. The latest version of these benchmarks is used to build the TOP500 list, ranking the world's most powerful supercomputers. Floating-point performance (double precision)

- **SPEC CPU® 2017** https://www.spec.org/cpu2017/results

  SPEC CPU 2017 focuses on compute intensive performance, which means these benchmarks emphasize the performance of the CPU, the memory hierarchy, including caches and main memory and C, C++, and Fortran compilers, including optimizers.

Microprocessors

Processor	Baikal BE-S1000	Kunpeng 920-4826	Intel® Xeon® Gold 6230
ARCH	ARM Cortex-A75	ARM-based TaiShan v110	X86_64
Release date	Q4'21	Q1'19	Q2'19
Lithography, nm	16 nm	7 nm	14 nm
Cores	48	48	20
Threads	48	48	40
Core base Frequency, GHz	2,5	2,6	2,1
GFLOPS perf., PEAK (DP)	480	499	1344
Cache memory	L2: 48x512 KB; L3: 12x2 MB; L4: 32 MB	L2: 48x512 KiB L3: 48 MiB	Intel® Smart Cache 27,5 MB
Max memory capacity	768 GB	1 TB	1 TB
Memory types	DDR4-3200	DDR4-2933	DDR4-2933
Max # of Memory Channels	6	8	6
PCI Express	Gen4	Gen4	Gen 3
TDP,W	120	150	125

Оборудование

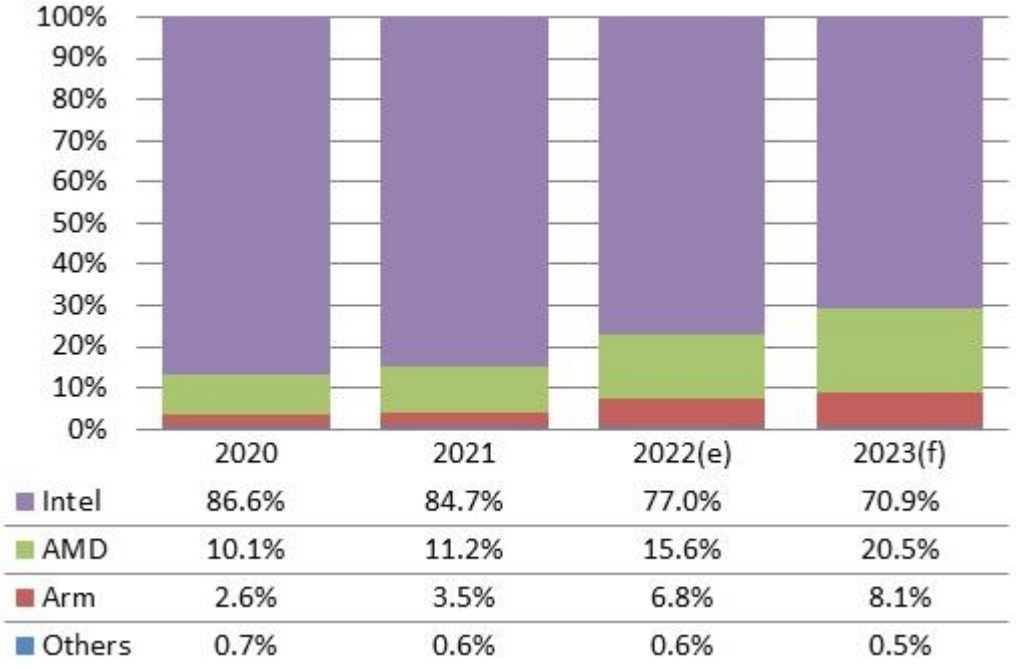
Board/Server	DBS-OV	MBS-2S	TaiShan 200 (Model 2280)	Supermicro SYS-6019P-MTR	Supermicro 2029TP-HTR
Microprocessor	BE-S1000		Kunpeng 920-4826	Intel® Xeon® Gold 6230	
Microprocessors #	1	2	2	2	2
Memory	6 x 64 GB DDR4-3200	12 x 16 GB DDR4-2400	16 x 32 GB DDR4-2933	8 x 64 GB DDR4-2933	12 x 32 GB DDR4-2933

MBS-2S hw&sw conf

Процессор	2 x Baikal-S1000 2.5 GHz, 930 mV core
Плата	mbs-2s
Память	2 x 6 x 16 GB Micron DDR4-3200 MT/s (скорость снижена до 2400)
Диск	Samsung 970 EVO PCIe Gen3x4 NVMe
CCIX conf	PCIe 2xGen4x8

OS	Debian bookworm GNU/Linux 5.4.244-baikal-arm64
Компилятор	GCC 12
Библиотека MPI	Open MPI 4.1.4
Линейной алгебры	Пробная версия Arm Performance Libraries 22.0.2

ARM64 servers

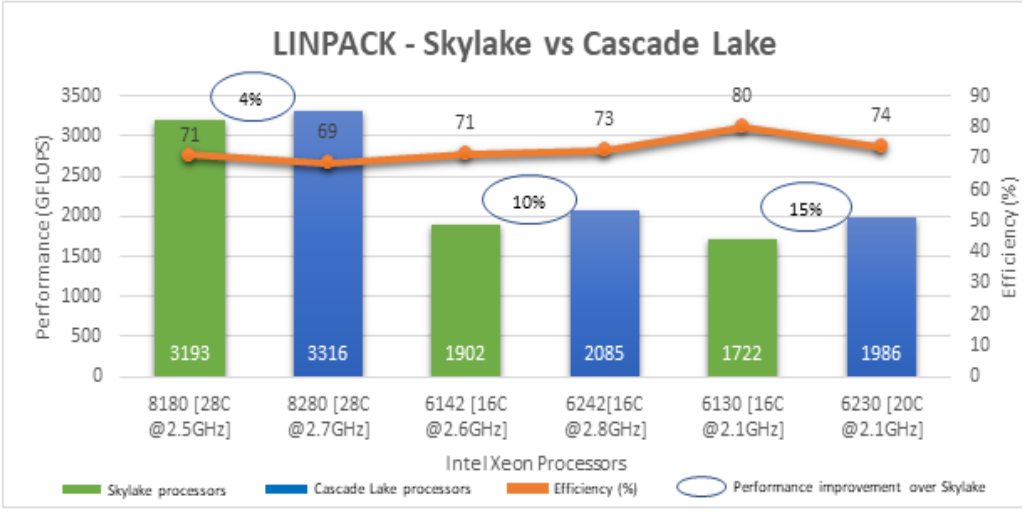


	2020	2021	2022(e)	2023(f)
Intel	86.6%	84.7%	77.0%	70.9%
AMD	10.1%	11.2%	15.6%	20.5%
Arm	2.6%	3.5%	6.8%	8.1%
Others	0.7%	0.6%	0.6%	0.5%

Server shipment share by CPU, 2020-2023. Source: DIGITIMES Research, February 2023.

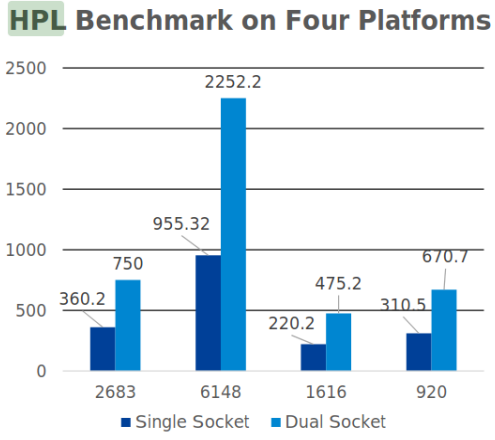
*In 2020, a system based on Arm processors took first place in the Top500 supercomputer ranking for the first time. This is the Japanese Fujitsu Fugaku, whose performance on the High Performance Linpack (**HPL**) test turned out to be **415.5 petaflops**. Fugaku is powered by Fujitsu's 48-core A64FX SoC. The system is installed at the RIKEN Center for Computational Science (R-CCS) in Kobe, Japan*

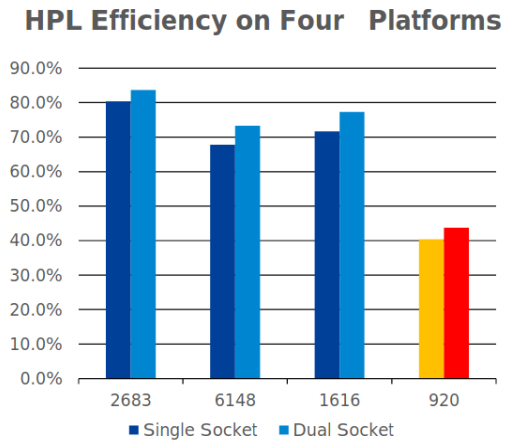
HPL



Float-point Arithmetic

- 41.1% Better than Hi1616, compared to a 165.3% increase from Haswell to Skylake in 3 years.
- HPL efficiency on Kunpeng 920 is around 40% compared to more than 70% on other chips.



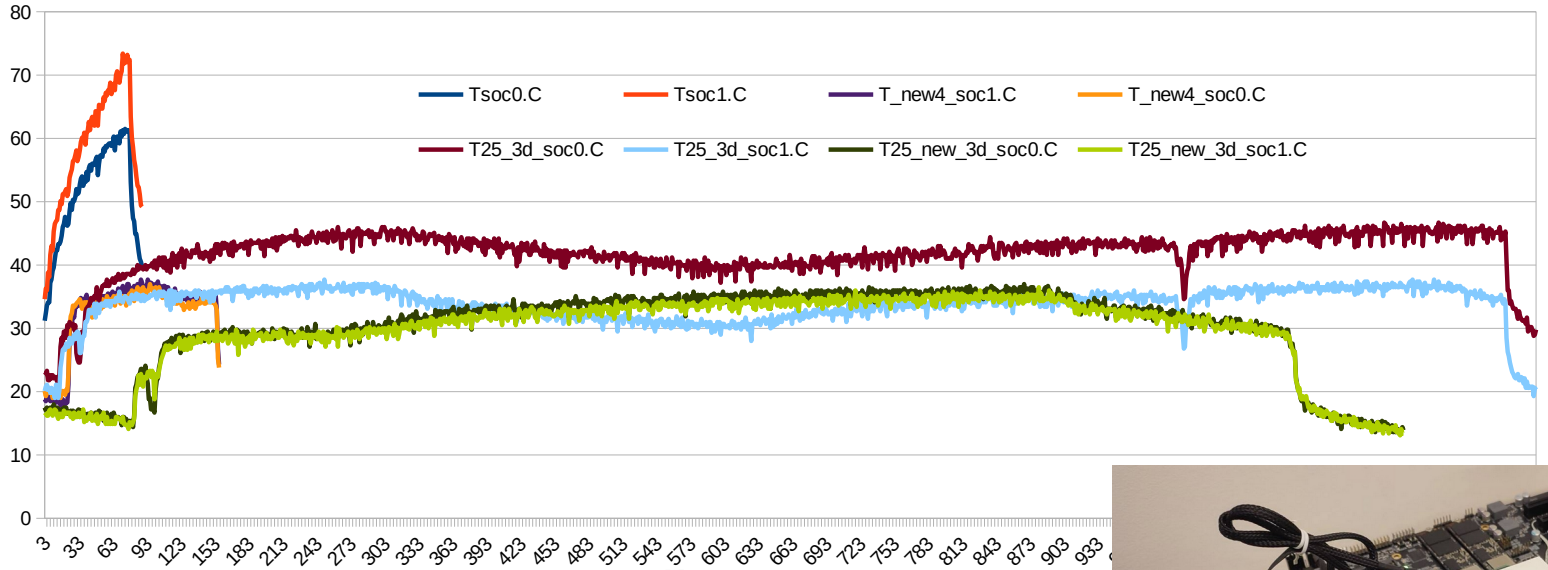


- LINPACK performance Dell with **Xeon Cascade Lake**[1] vs **Kunpeng 920**[2]
- *1) Dell, 2021. HPC synthetic benchmark performance using 2nd Generation Intel® Xeon® Scalable Processors – STREAM, HPL and HPCG*
- *2) Benchmarking Huawei ARM Multi-Core Processors for HPC workloads. Key Liao. Shanghai Jiao Tong University, 2019.*

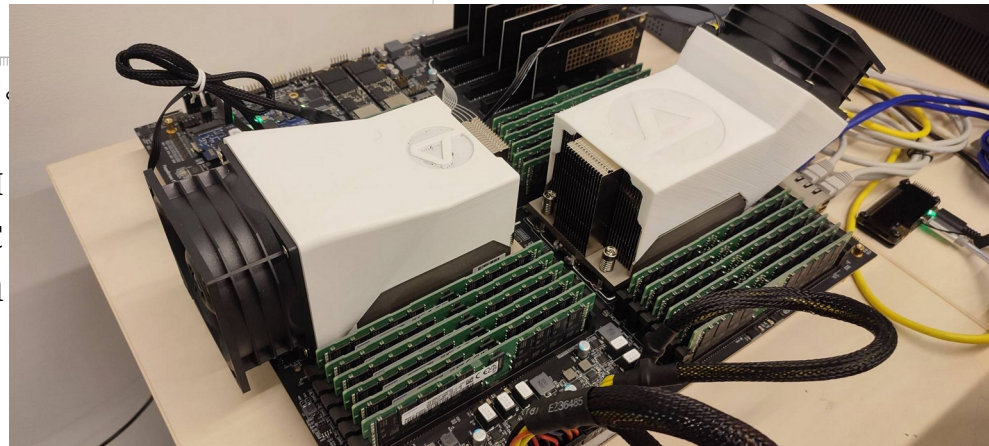
HPL

System	BE-S1000, 48 × 2.5GHz	BE-S1000, 2 SoC × 48 × 2.5GHz	Kunpeng 920-4826, 48 × 2.6GHz	Kunpeng 920-4826, 2 SoC × 48 × 2.6GHz	Xeon 6230 GCC	Xeon 6230 GCC, 2 chips
Mem	**6ch × 64GB DDR4 3200**	**2x6ch × 16GB DDR4 2400**	**8ch × 32GB DDR4 2933**	**2x8ch × 16GB DDR4 2933**	**4ch × 64GB DDR4 2933**	**2 x 4ch × 64GB DDR4 2933**
HPL result, all cores, GFLOPs	353	599	302	596	849	1700
PEAK PERF (DP), All cores, GFLOPs	480	960	499	998	1344	2688
Efficiency, %	74	62	60	60	63	63

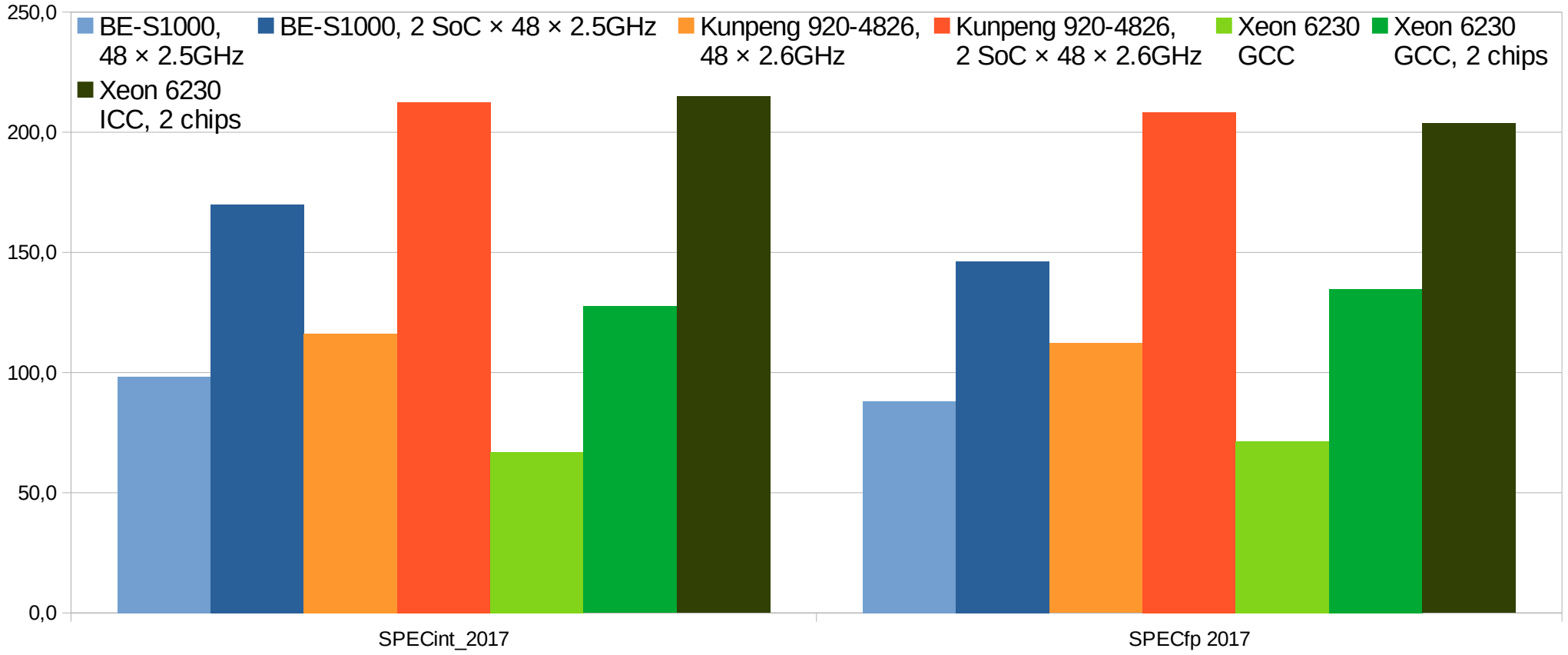
HPL Temp (mbs-2s A75 2.5 GHz, 930mV)



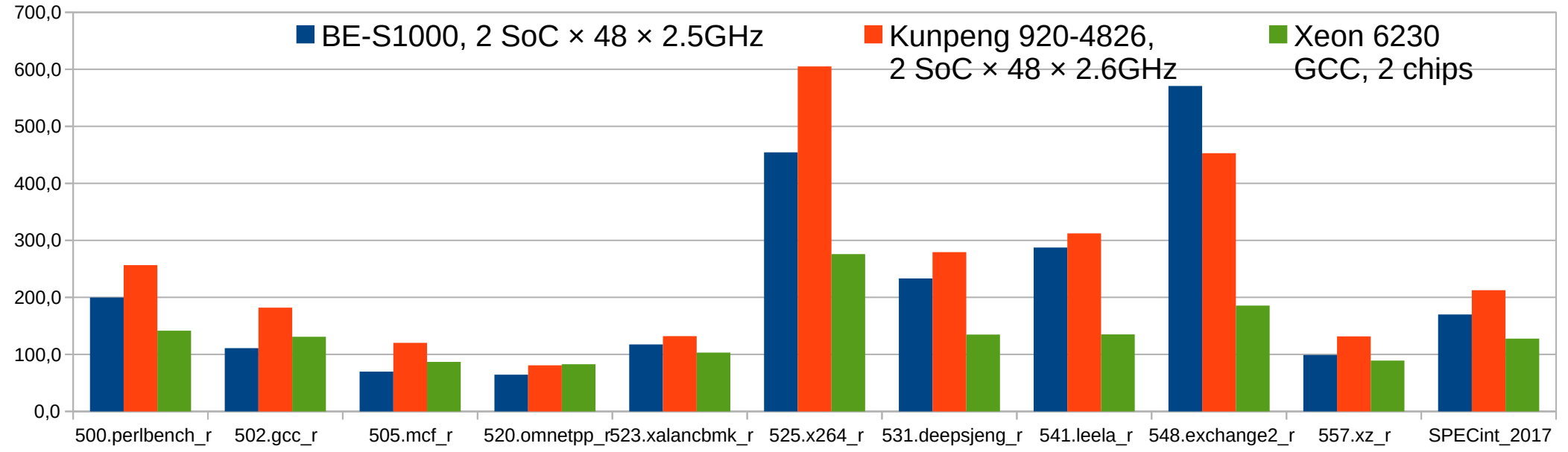
Использование **такой** → системы охлаждения достаточно для стабильной работы на частоте 2.5 ГГц на тесте HPL с использованием всех процессорных ядер на температуре до 50 °C



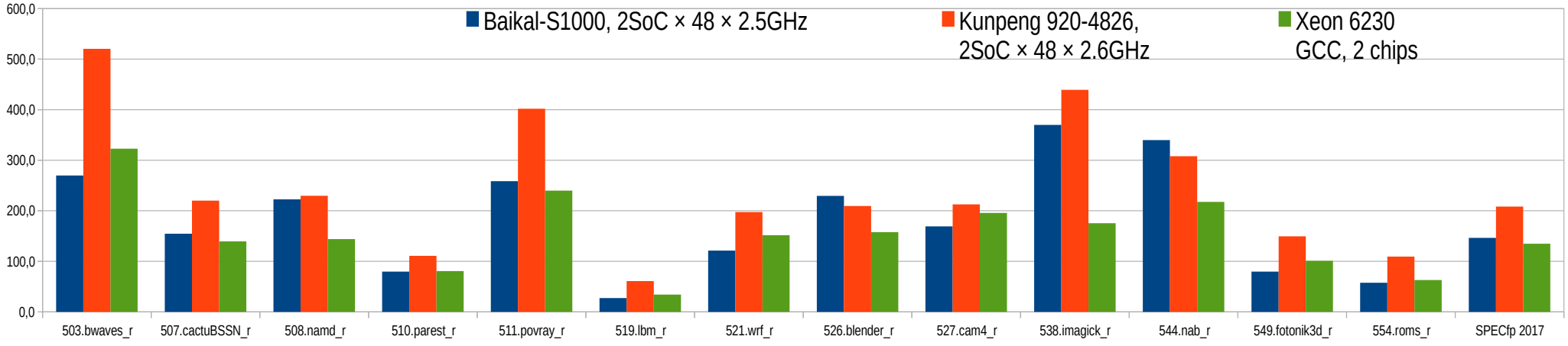
SPEC CPU 2017



SPEC CPU 2017 int_r



SPEC CPU 2017 fp_r



Выводы

- BE-S1000 показывает произвоизводительность сравнимую с Intel Xeon 6230 (2019) и Kunpeng 920-4826 (2019)
- Использование двухсокетной платы mbs-2s дает прирост производительности на `SPEC CPU 2017` и `HPL` на 80-100%, как и на Xeon или Kunpeng
- На `mbs-2s BE-S1000` работает на частоте 2.5 ГГц, при этом требуется поднятие напряжение до 0.93 В. Температура процессора ниже 50 °C под нагрузкой

Benchmarks BE-S1000

Спасибо за внимание!