IICT supercomputing infrastructure
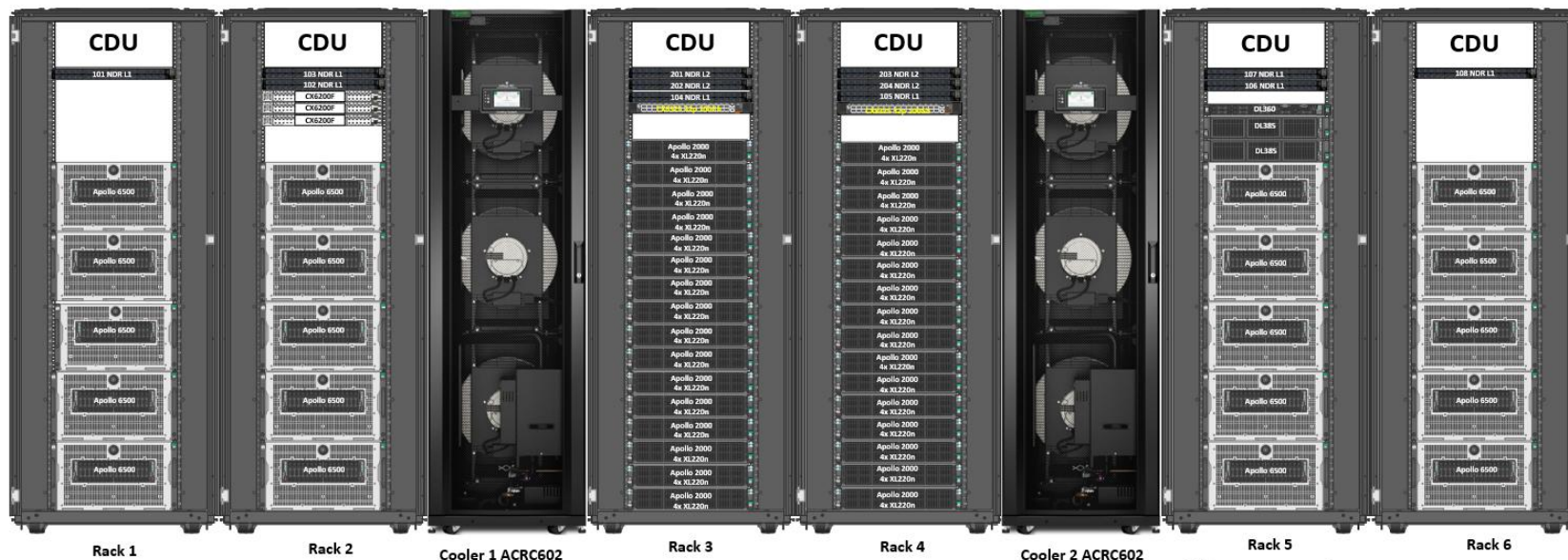
E. Atanassov
IICT-BAS
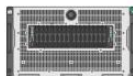
HEMUS positioning within the datacenter

- The supercomputer HEMUS has the natural focus within the datacenter.

- Infiniband connection with the storage system with capacity of more than 6 Petabytes

- Building monitoring system (BMS)

- UPS + diesel generator for power

- Infiniband connection with Avitohol

General view

	Консумирана ел. мощност, kW	Излъчвана топлина към система за директно течно охлаждане, kW	Излъчвана топлина към въздуха, kW
Шкаф 1	31.5	24.6	6.9
Шкаф 2	34.0	24.6	9.4
Шкаф 3	43.0	24.9	18.1
Шкаф 4	43.0	24.9	18.1
Шкаф 5	33.8	24.6	9.2
Шкаф 6	32.9	24.6	8.3
ОБЩО:	218.2	148.2	70

Таблица с прогнозни стойности на излъчвана топлина

Легенда:

Сървърно шаси Apollo 6500 Gen10 Plus с 1 бр. сървър XL675d Gen 10 Plus с универсални многоядрени процесории и графични ускорители

Сървърно шаси Apollo 2400 Gen10 Plus с 4 бр. сървъри XL220n Gen 10 Plus с универсални многоядрени процесори

Обслужващ сървър ProLiant DL385 Gen 10 Plus V2

Сървър за автоматизирана централизирана система за мониторинг HPE ProLiant DL360 Gen10 Plus

Комутатор за автоматизирана централизирана система за мониторинг Aruba 6200F 48G 4SFP+

Средно производителен мрежов комутатор Aruba 8325-32C 32-port 100G QSFP+/QSFP28

Високопроизводителен мрежов комутатор NVIDIA QUANTUM-2 QM9700

CDU Разпределител на охлаждащ агент Cooling Distribution Unit (CDU)

Hardware and software features

- Two subsystems – pure CPU-based, GPU-based

- SUSE HPC Linux distribution

- Intel One API

- HPE Ezmeral, IBM Storage Scale (GPFS)

GPU-based subsystem

|	
Servers	20 x HPE ProLiant XL675d Gen10 Plus
CPUs	2 x AMD EPYC 7742 (2.25GHz/64-core/225W) 64 cores/128 HT (128 cores / 256 HT per node)
GPU	8 x NVIDIA A100 TENSOR CORE 40GB SXM Peak 64-bit floating point performance - 19,5 Tflops Memory bandwidth - 1,555GB/s Interconnectivity within node NVLink: 600GB/s (8 x 20 =160 in total, 160x40 GB = 6,25 TB RAM)
RAM	16x32 GB = 512 GB DDR4-3200 MHz per node
Total CPUs	2 560 cores / 5 120 HT
Total RAM	10 TB
Interconnect	Non-blocking HDR Infiniband
Bandwidth	200 Gbps
Storage	2 x 480GB SATA SSD - 960 GB 4 x 3.84TB NVMe Gen3 SFF SC U.3 SSD = 15.36 TB Total: 16.30 TB per node 20 x 16.30 TB = 320 TB total

CPU based subsystem

Servers	128 x HPE ProLiant XL220n Gen10 Plus
CPUs	2xIntel Xeon-Platinum 8352Y 2.2GHz 32-cores/64 HT (64 cores / 128HT per node)
RAM	16x16GB = 256 GB DDR4-3200 MHz per node
Total CPUs	8 192 cores/16 384 HT
Total RAM	32 TB
Interconnect	Non-blocking NDR Infiniband
Bandwidth	200 Gbps
Storage	2 x 480GB NVMe Gen3 SSD = 960 GB per node Total: 120 TB
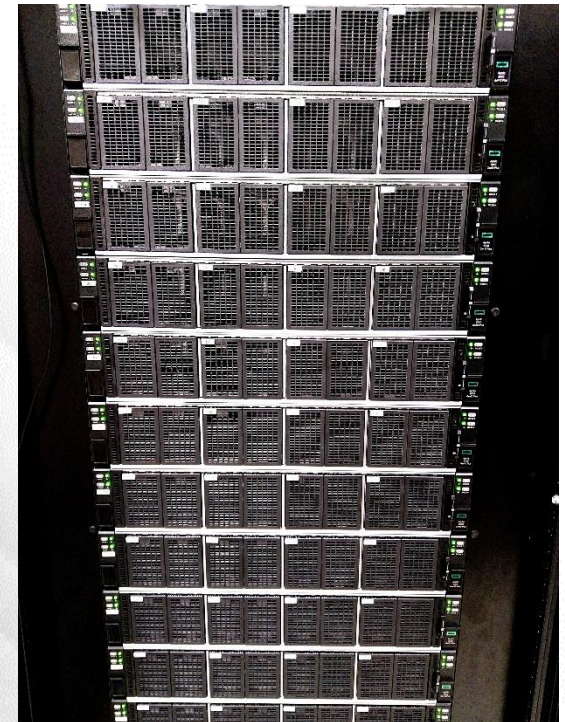
Infiniband interconnection
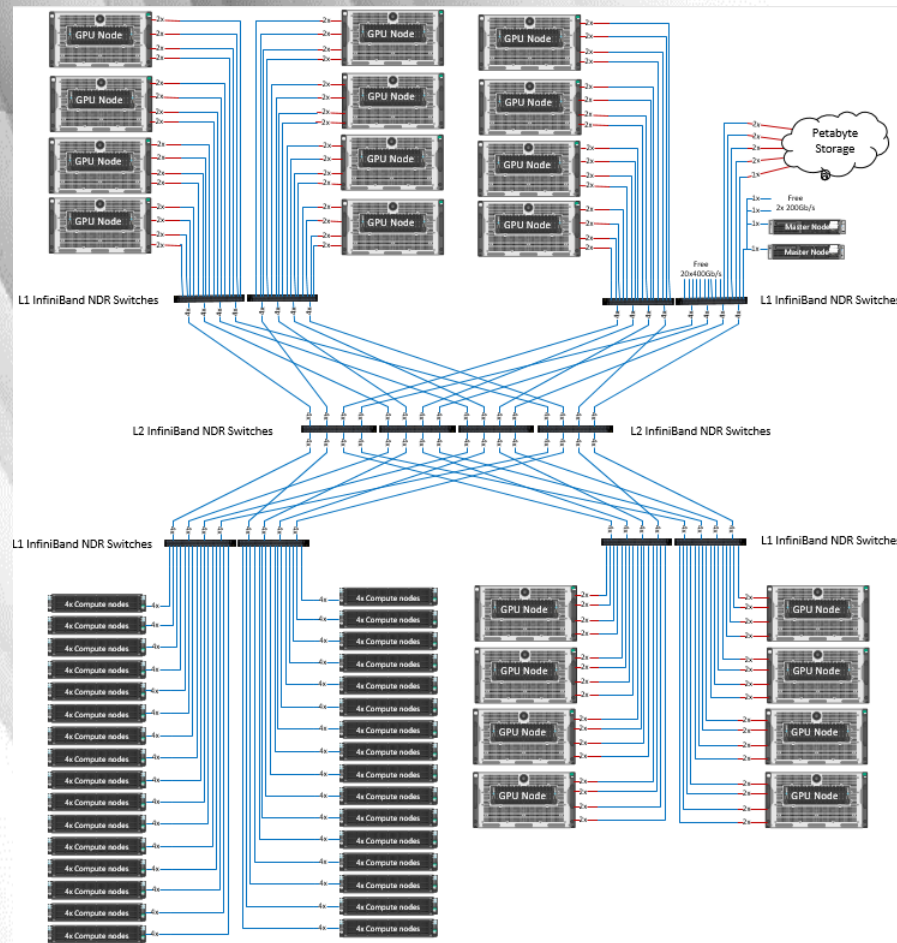
System for data storage and processing

- Access to more than 5 PB of HDD/SSD storage
- Servers with large amounts of RAM for in-memory data processing
  - 4 x Intel Xeon Gold 6238L 2,1GHz, 22 cores;
  - 3 TB RAM per server
  - 2 x 480 GB SSDs
- Melanox InfiniBand HDR 40-port QSFP56 managed, 200 Gbps per port

Linpack test results and discussion

- 360th place in Top500 list November 2024

- 2.53 PFlop/s Linpack performance

- 3.21 Pflop/s theoretical peak performance (of the gpu subsystem only, otherwise 0.5 more from the cpu subsystem)

- The configuration of the grid in the HPL.dat file was important, 160 cards map to 8x20.

- Usually cpu-based systems benefit from `more square' grids.

- GPU cards draw maximum power - 400W each card – 64 kW just for the cards.

HPCG test results and discussion

- Much less parameters could be changed in the input.
- Less systems provided such tests (the efficiency with respect to the theoretical peak performance is smaller)
- 109th place with 36.42 Teraflops
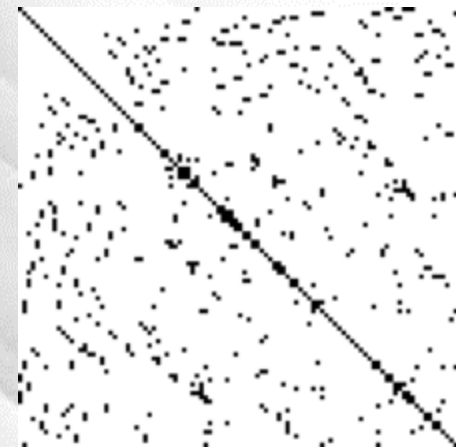
Mixed precision test

- The test is at https://hpl-mxp.org/
- Much fewer systems provided test results.
- The test measures acceleration of the computations done in mixed precision (with refinement) vs double precision Linpack.
- We again used NVIDIA binary named HPL-AI, because of the connection with AI workloads, where smaller precisions are used (even smaller than single precision).
- 19.76s 15.37 Pflops   9.12s   10.52 Pflops with refinement
- Speedup vs double precision Linpack – 4 times.
- NVIDIA A100 exceptional performance in machine learning tasks:
  - Half Precision (FP16) – 78 TFLOPS
  - Brain Floating Point (BF16) – 39 TFLOPS
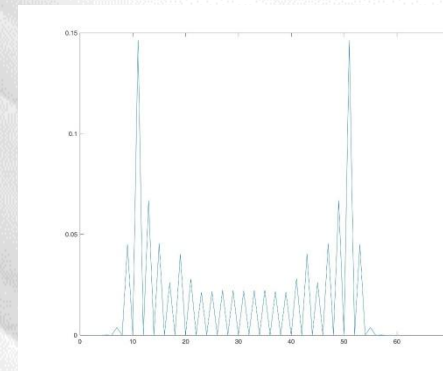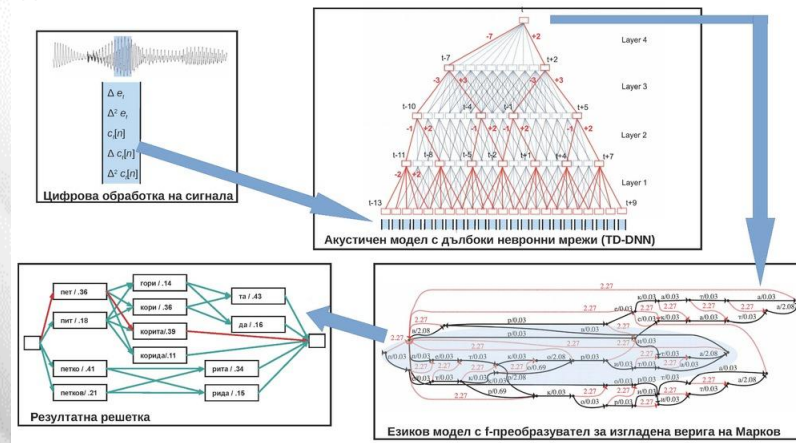
Quantum random walk modelling

Efficiency

- HEMUS achieves >2.5 Pflop/s with less power than Avitohol
- Efficiency of measured linpack performance vs theoretical peak performance

Name	HP Linpack	Theoretical	Efficiency
Avitohol	0.26	0.41	63%
Discoverer	4.52	5.94	76%
HEMUS	2.53	3.21	79%

Avitohol supercomputer

Peak Performance CPU	50 Tflop/s
Pear Performance Accelerators	362 Tflop/s
Total Peak Performance	412 Tflop/s
Real Measured Performance	264 Tflop/s
Max Power	250 kW

The system consists of 4 dual racks of type HP MCS 200. Each one provides power and cooling for up to 50 kW of equipment, cooled by water. Last on Top500 List on 389 place (Nov 2015)
http://www.top500.org/system/178609

Avitohol supercomputer

Avitohol supercomputer

Hardware features

Servers	150 x Dual CPU HP ProLiant SL250s Gen8
CPUs	Intel Xeon E5-2650v2 2.6GHz – 8 cores /16 HT
RAM	64 GB per node
Coprocessors	300 x Intel Xeon Phi 7120P(x86) - 61 cores / 244 HT
Total CPUs	2400 cores/4800 HT + 18300 cores/ 73200 HT
Total RAM	14400 GB (9600 + 4800)
Disk Storage	100 TB
Interconnect	Non-blocking FDR Infiniband
Latency	1.1 μs
Bandwidth	56 Gbps

System for data storage and processing

- Access to more than 5 PB of HDD/SSD storage

- Servers with large amounts of RAM for in-memory data processing
  - 4 x Intel Xeon Gold 6238L 2,1GHz, 22 cores;
  - 3 TB RAM per server
  - 2 x 480 GB SSDs

- Melanox InfiniBand HDR 40-port QSFP56 managed, 200 Gbps per port

GPU computing servers

- Fujitsu Primergy RX 2540 M4 with 2 CPUs and 1 GPU, interconnected with InfiniBand
- GPU NVIDIA Tesla V100 PCIe 32 GB
- CPU Intel(R) Xeon(R) Gold 5118 CPU @ 2.30GHz – 12 cores
- 128 GB RAM
- 2x800GB SSD
- 3*12 TB HDD
- Suitable for machine learning, intensive processing, visualisation

Thanks