

Povídání na téma

SUPERPOČÍTAČE DNES A ZÍTRA (aneb krátký náhled na SC)



Co je to vlastně SC?

Výpočetní systém, který určuje hranici maximálního možného výpočetního výkonu...

...v dané době samozřejmě.

Jak poznat onu maximální hranici v konkrétní době?

www.top500.org

Historie výkonu SC



Kategorizace SC

Podle architektury SC

- konstelace (constellation)
- masivně paralelní počítač (MPP)
 - klastr (cluster)

Podle architektury CPU

- vektorové
- skalární

Masivně paralelní počítače

(massive parallel processing - MPP)

- počítač s globálně sdílenou pamětí dvěmi a více procesory
- dělení na systémy se symetrickým (SMP) a nesymetrickým (NUMA/ccNUMA) přístupem do paměti
- paralelizace pomocí systému vláken (OpenMP), možnost autoparalelizace

Výhody – jednoduchá paralelizace i správa

Nevýhody – cena, často proprietární řešení

Například Cray XT5 (1. místo), IBM Blue Gene P (4. místo).

Klastr

(Cluster)

- většinou homogenní soubor vzájemně propojených počítačů, tzv. uzlů (node)
- paralelizace pomocí zasílání zpráv (message-passing – MPI)
- potřeba kvalitního propojení (interconnect) s velkou propustností a malými latencemi

Výhody – cena, dostupnost

Nevýhody – nároky na interconnect (topologie), složitější paralelizace, správa

Například RoadRunner (2. místo), Tianhe-1 (5. místo).

Konstelace

(constellation)

- cluster “menších” MPP propojených do clusteru (min. 16 a více CPU na uzel)
- paralelizace pomocí kombinace MPI a OpenMP

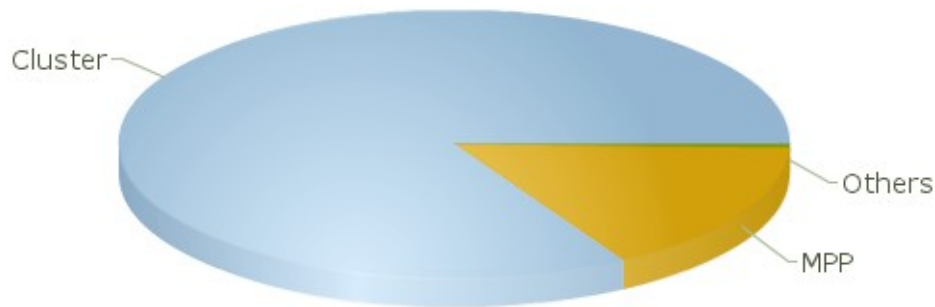
Výhody – menší závislost na propojení uzlů, možnost paralelizace v rámci uzlu

Nevýhody – složitost paralelizace (MPI vs OpenMP), složitost plánování úloh, poměr cena/výkon

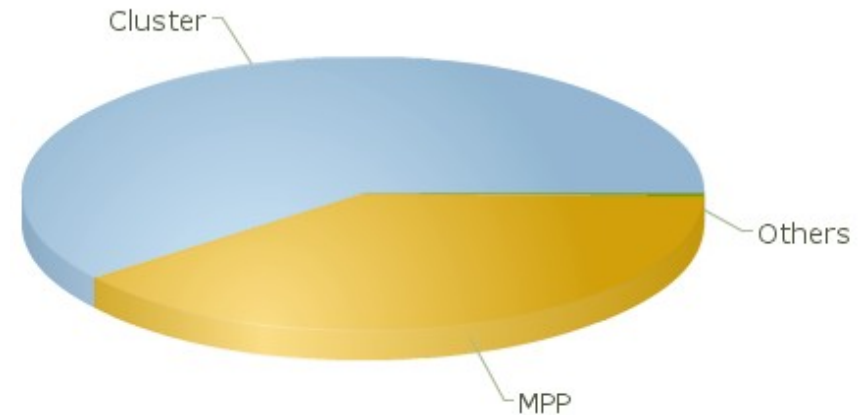
Tera-10 (86. místo) a CEA-CCRT-Platin (118. místo) – BULL

Zastoupení podle architektury SC

Architecture / Systems
November 2009



Architecture / Performance
November 2009



Architektura	Počet	Proc. podíl
Konstelace	2	0.40 %
Mas. Par. Sys.	81	16.20 %
Clustery	417	83.40 %

Výkon Rmax	Výkon Rpeak
94970	112947
10939467	13614091
16943065	27223084

Údaje z listopadu 2009 TOP500

Architektury CPU v SC

Dominantní (99%) – skalární architektura (superskalární, VLIW, RISC)

Intel EM64T (79%)

IBM Power (10%)

AMD x86_64 (8%)

IA-64 (1%)

SPARC a další (1%)

Minoritní (1%) – vektorová architektura
NEC SX-9 – Earth simulator (31. místo)

Akcelerátory

Možný doplněk clusterů, MPP i konstelací.

GP-GPU akcelerátory (grafické karty)
FPGA (programovatelná hradlová pole)
Proprietární (např. vektorové X2 pro CRAY XT)

Vysoký výpočetní výkon, ale spíše single precision (mění se), omezená práce s pamětí (vhodné pro některé typy úloh), dobrá cena.

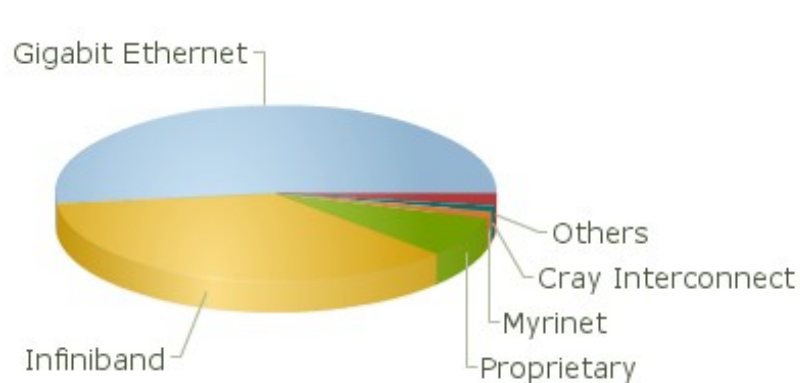
Propojení

(interconnect)

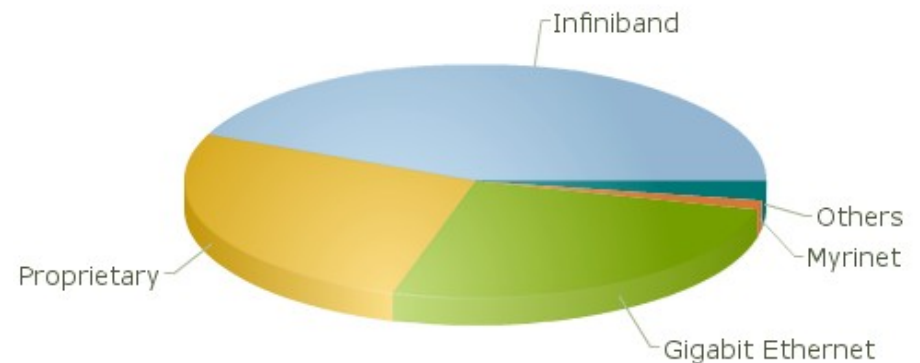
- v rámci NUMA systémů – mezi jednotlivými CPU (HyperTransport, QPI, NUMALINK, Cray IC)
- v rámci clusterů – mezi jednotlivými uzly (Ethernet, Infiniband, Myrinet, Quadrics)

Zastoupení IC

Interconnect Family / Systems
November 2009



Interconnect Family / Performance
November 2009



Ethernet (52%) – Gigabit (51%) a 10G (0,2%)

Infiniband (36%) – 27%, DDR (6%), DDR 4x (2%), QDR ()

Myrinet (1,4%)

Cray IC (1%)

Proprietární (7,4%)

Vlastnosti IC

Ethernet – levný, běžný, otevřený, limit rychlosti a linek (STP)

Infiniband – otevřený, poměrně levný, rychlý (BW i latence)

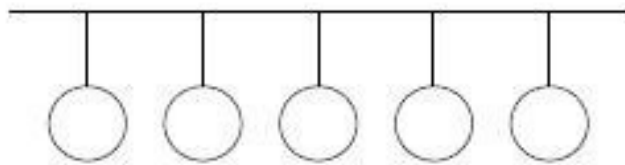
Cray IC – proprietární, vázáný na Cray XT

Myrinet – dneš již starý, pomalý (BW), proprietární

Quadrics – zastaralý, uzavřený, pomalý

Základní topologie

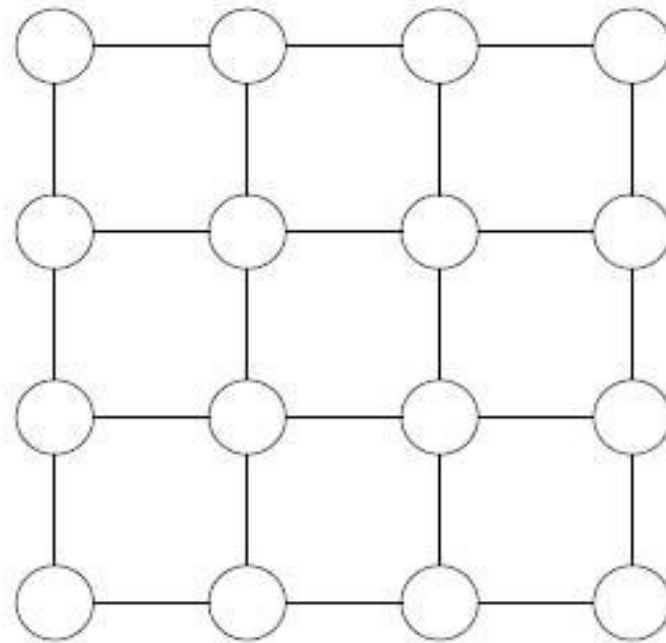
shared bus



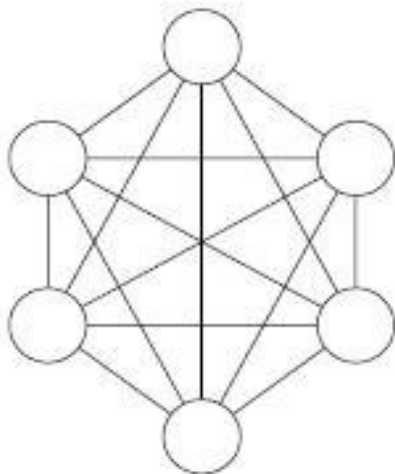
line



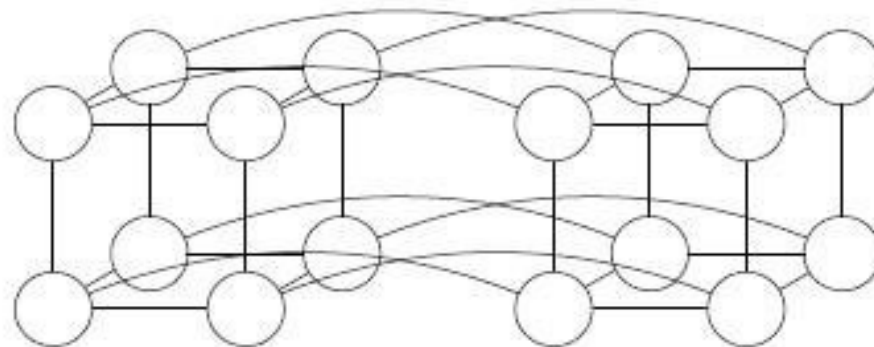
2-D mesh



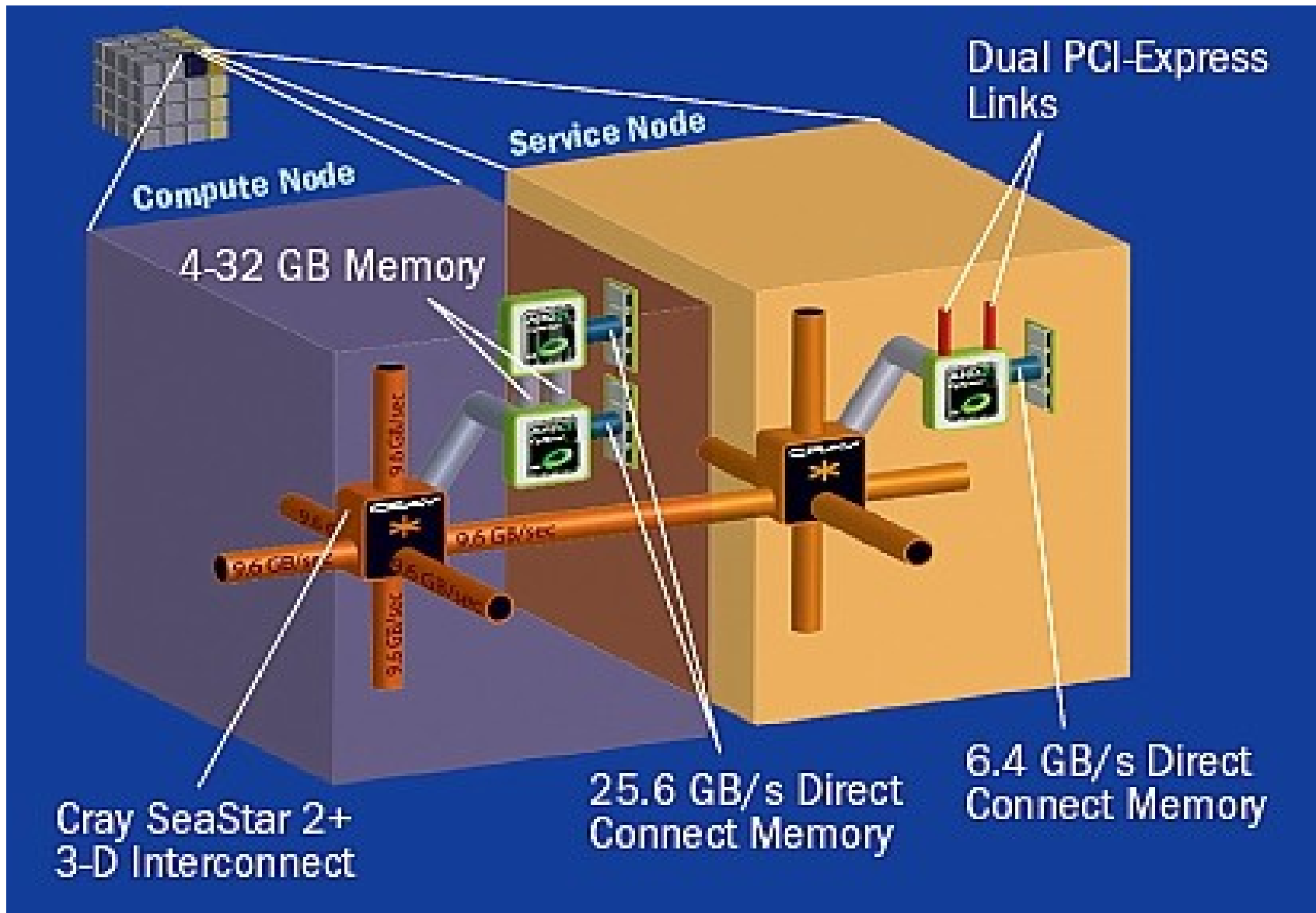
fully connected



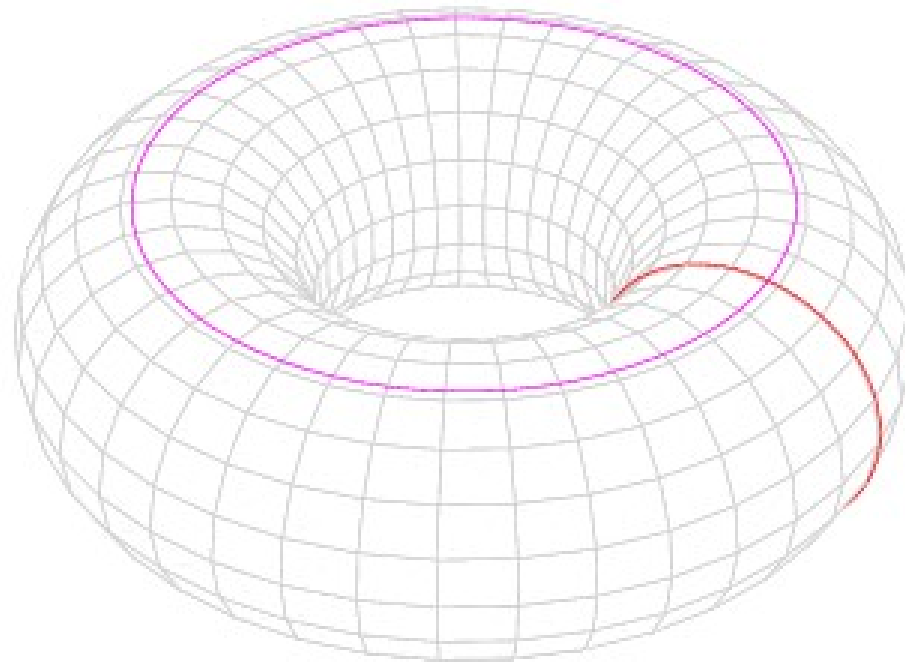
hypercube



3D-Mesh

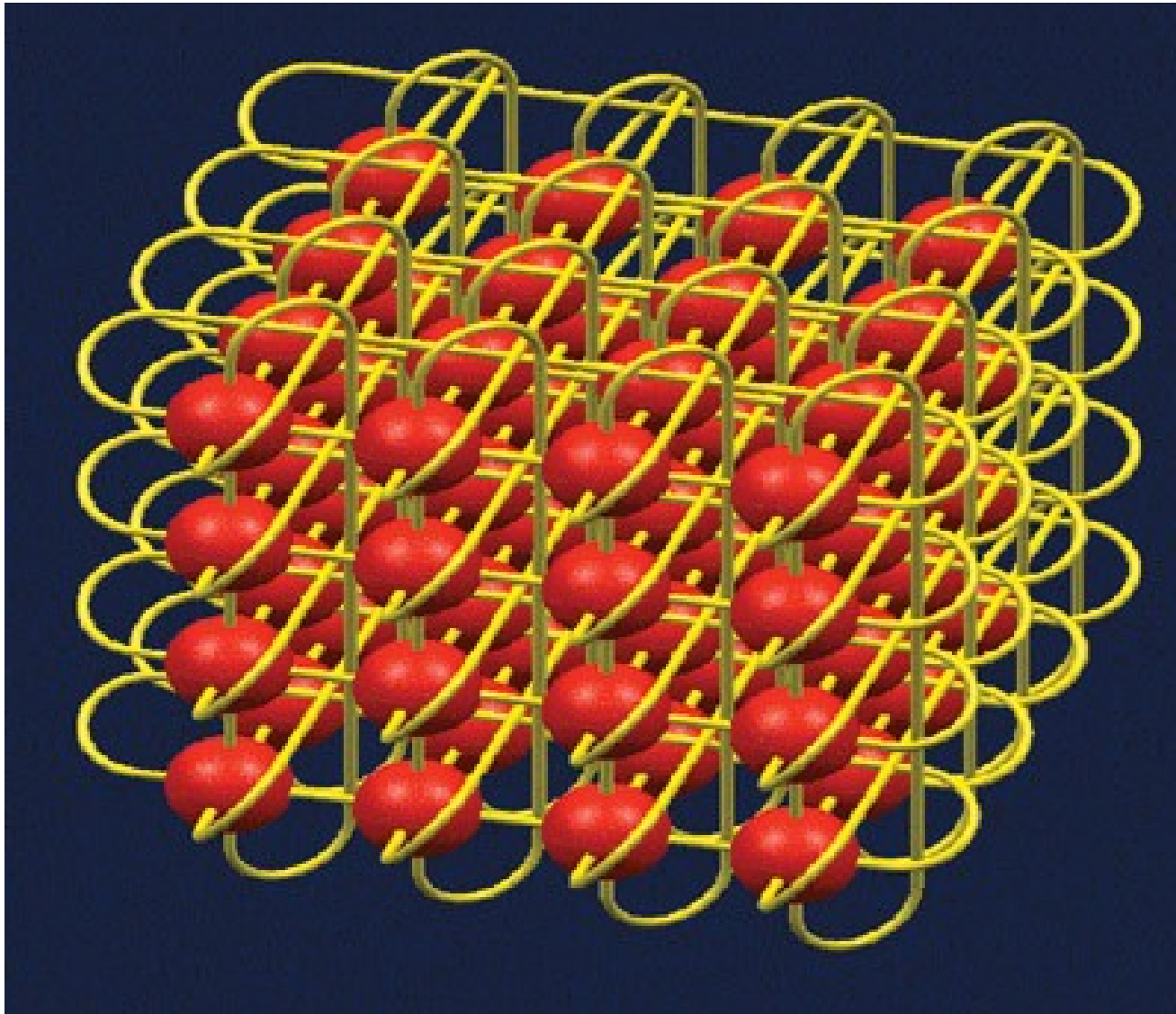


Torus



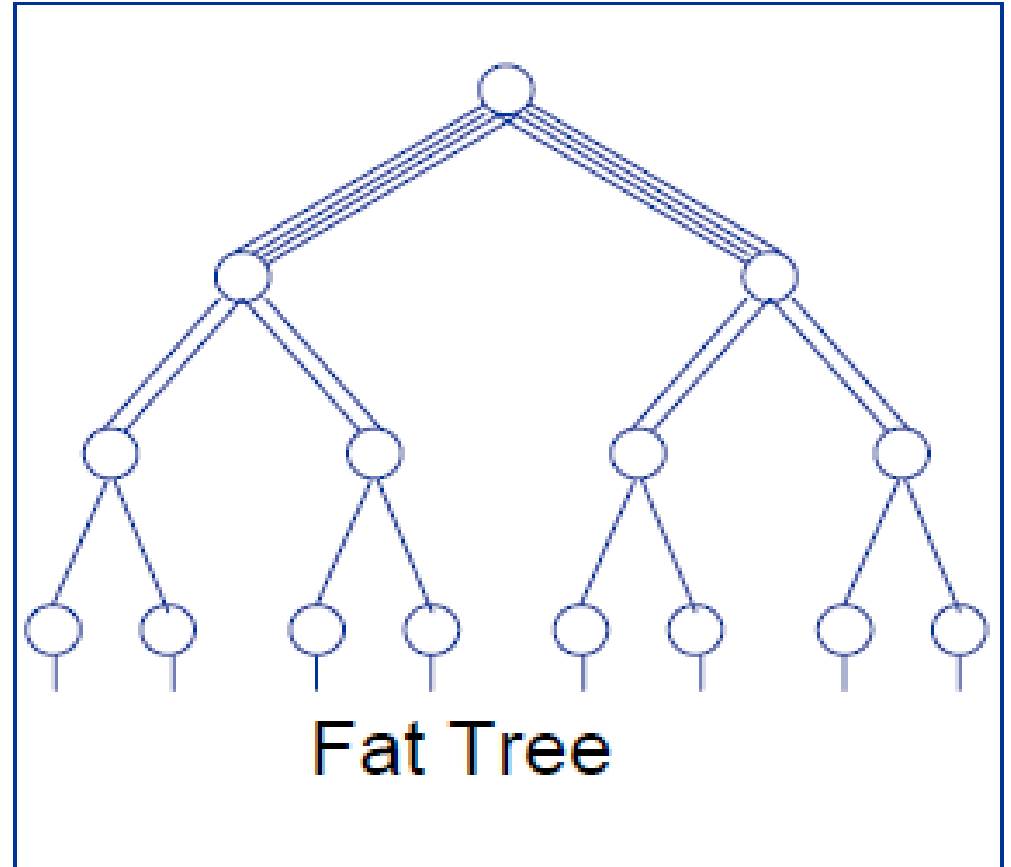
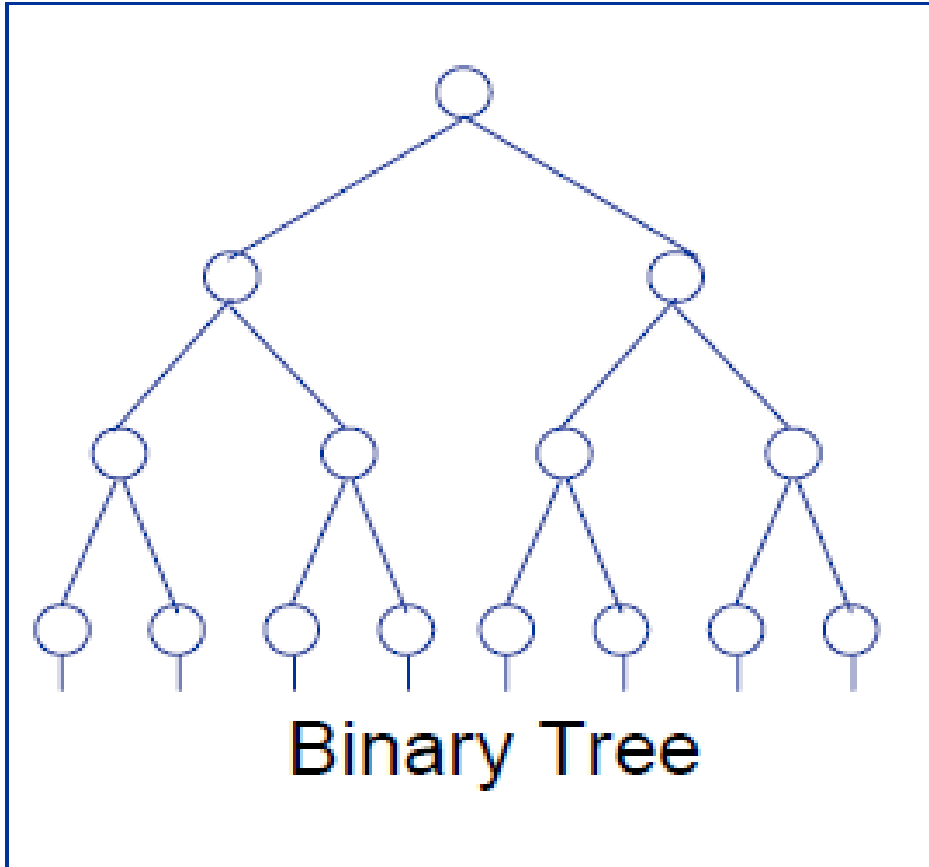
Cray XT – úplná topologie

3D-Torus



IBM Blue Gene

CLOS (Fat-Tree)



Software

Operační systém – Linux (89%), UNIX (5%),
smíšené (5%), Windows (1%)

Paralelní souborový systém – LUSTRE, GPFS,
Panasas

Plánovač – Platform LFS, PBS/Torque, SUN
SGE, MOAB

Jak se stanovuje výkon SC?

FLOP/s – počet operací s plovoucí čárkou za 1 s.
(konkrétně se jedná o 64bitové číslo s plovoucí čárkou a operaci sčítání nebo násobení)

Dnes se používá spíše TFLOP/s (10^{12} FLOP/s).

R_{peak} – maximální teoretický výkon pro jedno jádro: $PO \cdot FR$
(PO – počet operací za cyklus, FR – frekvence CPU)

R_{max} – výkon naměřený v High Performance Linpack (HPL) benchmarku

Linpack – program na měření výpočetního výkonu, pomocí řešení husté soustavy lineárních rovnic

Nejvýkonější SC

Pořadí	Jméno - Země Počítač/Rok	Výrobce	Jádra	Rmax	Rpeak	Energie (kW)
1	Jaguar - USA Cray XT5-HE Opteron Six Core 2.6 GHz / 2009	Cray Inc.	224162	1759.00	2331.00	6950.60
2	Roadrunner - USA BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, IB / 2009	IBM	122400	1042.00	1375.78	2345.50
3	Kraken - USA XT5 - Cray XT5-HE Opteron Six Core 2.6 GHz / 2009	Cray Inc.	98928	831.70	1028.85	
4	JUGENE - Německo Blue Gene/P Solution / 2009	IBM	294912	825.50	1002.70	2268.00
5	Tianhe-1 - NUDT TH-1 Cluster Xeon E5540/E5450, ATI Radeon HD 4870 2, Infiniband / 2009 Čína	NUDT	71680	563.10	1206.19	

Nejčastější úlohy SC

Název	Počet	Procent. podíl
Nespecifikováno	117	23.40 %
Výzkum	78	15.60 %
Finance	47	9.40 %
Informační služby	34	6.80 %
Geofyzika	31	6.20 %
Logistické služby	30	6.00 %
Polovodiče	22	4.40 %
Služby zpracování informací	18	3.60 %

Příklady úloh SC

- kvantová mechanika
 - předpověď počasí
 - výzkum klimatických změn
- molekulární modelování (výpočty struktury a vlastností chemických sloučenin, biologických makromolekul, polymerů a krystalů)
- fyzikální simulace (proudění kapalin – aerodynamický tunel, detonace jaderných zbraní, výzkum jaderné fúze)

Budoucnost SC

EXA Scale Computing

Nové technologie CPU – integrace GPU do jader CPU, lepší vektorové jednotky a více jader...

Problémy s propustností paměti proti CPU + další I/O problémy.

Energetická náročnost.

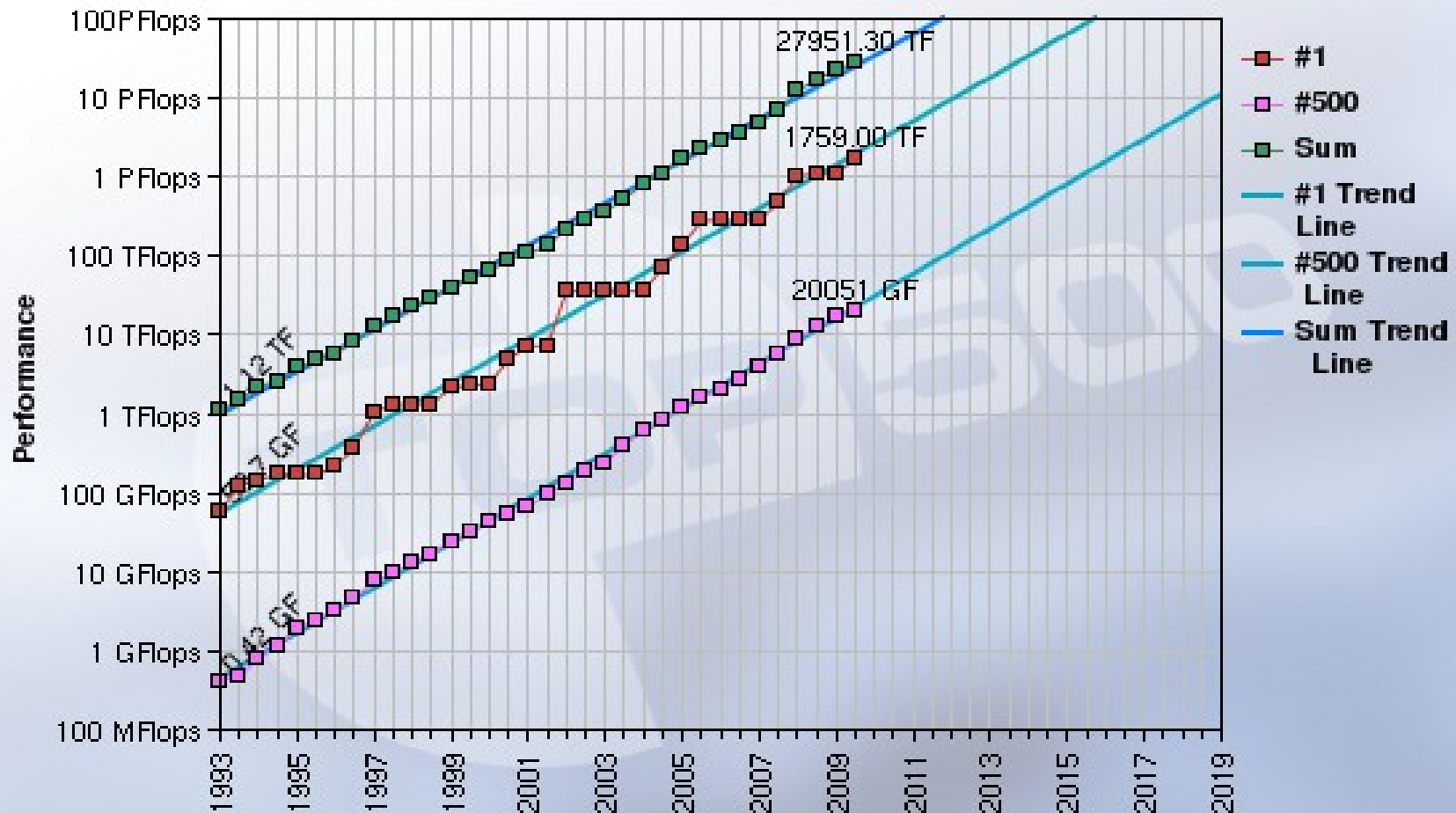
Nasazení technologie memristoru.

Optické přenosy a uchovávání dat (plně optické kruhové přepínače).

Predikce výkonu SC



Projected Performance Development



Situace v ČR

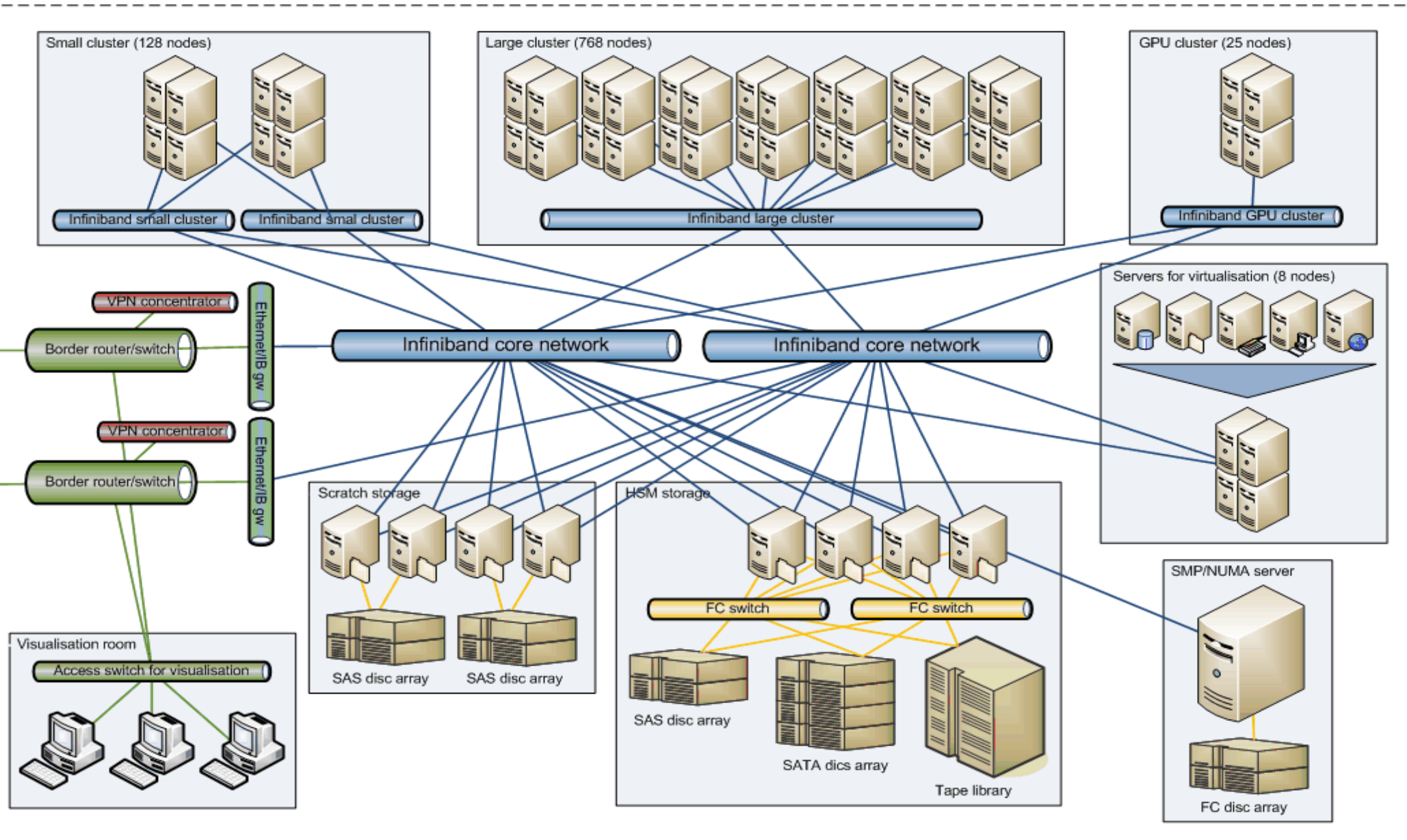
V současnosti neexistuje žádný SC (jak je známe z TOP500).

Do budoucna plánován SC v rámci IT4Innovations.

Předpokládaný výkon přes 1 PFLOP/s.
(rok 2015)

Příklad SCC

(supercomputing centre)



Zdroje informací

www.top500.org
www.hpcwire.com
www.wikipedia.org
www.google.com

Děkuji za pozornost

Otázky?