

C2115

Praktický úvod do superpočítání

I. lekce

Petr Kulhánek, Jakub Štěpán

kulhanek@chemi.muni.cz

Národní centrum pro výzkum biomolekul, Přírodovědecká fakulta,
Masarykova univerzita, Kotlářská 2, CZ-61137 Brno

- **Historie, využití a budoucnost výpočetní techniky**
- **Příklady reálných problémů**
projekty z Laboratoře výpočetní chemie
- **Přehled výpočetních center ČR**
MetaCentrum, CERIT-SC, IT4 Innovation
- **Zahraniční výpočetní centra**
centra dostupná pro zájemce z ČR, Top500
- **Cvičení**

Historie

http://en.wikipedia.org/wiki/History_of_computing_hardware

1800 počátky dřevných štítků

1946 ENIAC

1947 objev tranzistoru

1971 Intel 4004 (4 bit)

1974 Intel 8080 (8 bit)

1976 Intel 8086 (16 bit)

1985 Intel 80386 (32 bit)

2001 IA-64 (64 bit)

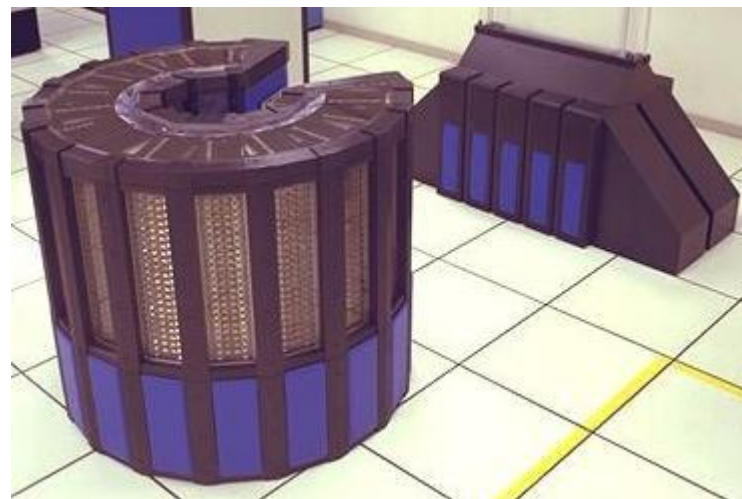
2003 AMD64/EM64T (64 bit)

**2010 Intel Core i7 980X: @3,33 GHz
(6C/12T, Turbo@3,46 GHz): 109 GFLOPS**



<http://www.root.cz>

Pavel Tišnovský, Unixové vykopávky



1985 Cray 2 1,9 GFLOPS

proprietární vektorové CPU

zdroj: wikipedia.org, intel.com

Využití výpočetní techniky

Výpočetní technika (počítače) zasáhla do všech odvětví lidské činnosti a stala se nedílnou součástí našich životů. Dopomohl k tomu především **bouřlivý vývoj** v posledních 20 letech. Výpočetní techniku používáme pro zábavu, k zpracovávání a konzumování informací.

Výpočetní technika (a hlavně superpočítače) se využívají k **řešení numericky náročných** problémů jako jsou:

- simulace počasí, klimatologických a geologických změn (šíření záplav, vln tsunami, zemětřesení)
- návrh nových materiálů a léčiv
- modelování ekonomického vývoje
- **vědeckotechnické výpočty** (**chemie**, fyzika, matematika)
- vojenské účely (simulace jaderných zbraní)
-



Masivní využití GPGPU ...

Simulátor lidského mozku:

<http://www.humanbrainproject.eu/>

... kvantové počítání, ...

Skupina výpočetní chemie

přehled řešených projektů

Skupina výpočetní chemie



prof. RNDr. Jaroslav Koča, DrSc.

- 1 profesor
- 5 výzkumných asistentů
- 2 post-doc studenti
- 11 doktorských studentů
- 7 bakalářských a magisterských studentů



Mgr. Zdeněk Kříž, Ph.D.

E-mail: zdenek.kriz@ceitec.muni.cz
Expertise: QM, MD, Docking



RNDr. Petr Kulhánek, Ph.D.

E-mail: petr.kulhanek@ceitec.muni.cz
Expertise: QM, QM/MM, MD, Free Energy



Mgr. Martin Prokop, Ph.D.

E-mail: martin.prokop@ceitec.muni.cz
Expertise: Software dev, Docking



RNDr. Radka Svobodová, Ph.D.

E-mail: radka.svobodova@ceitec.muni.cz
Expertise: Chemo and Bioinformatics



RNDr. Robert Vácha, Ph.D.

E-mail: robert.vacha@ceitec.muni.cz
Expertise: MD, MC, Coarse Grain, Free Energy



Mgr. Stanislav Kozmon, Ph.D.

E-mail: stano@chemi.muni.cz
Expertise: QM, QM/MM

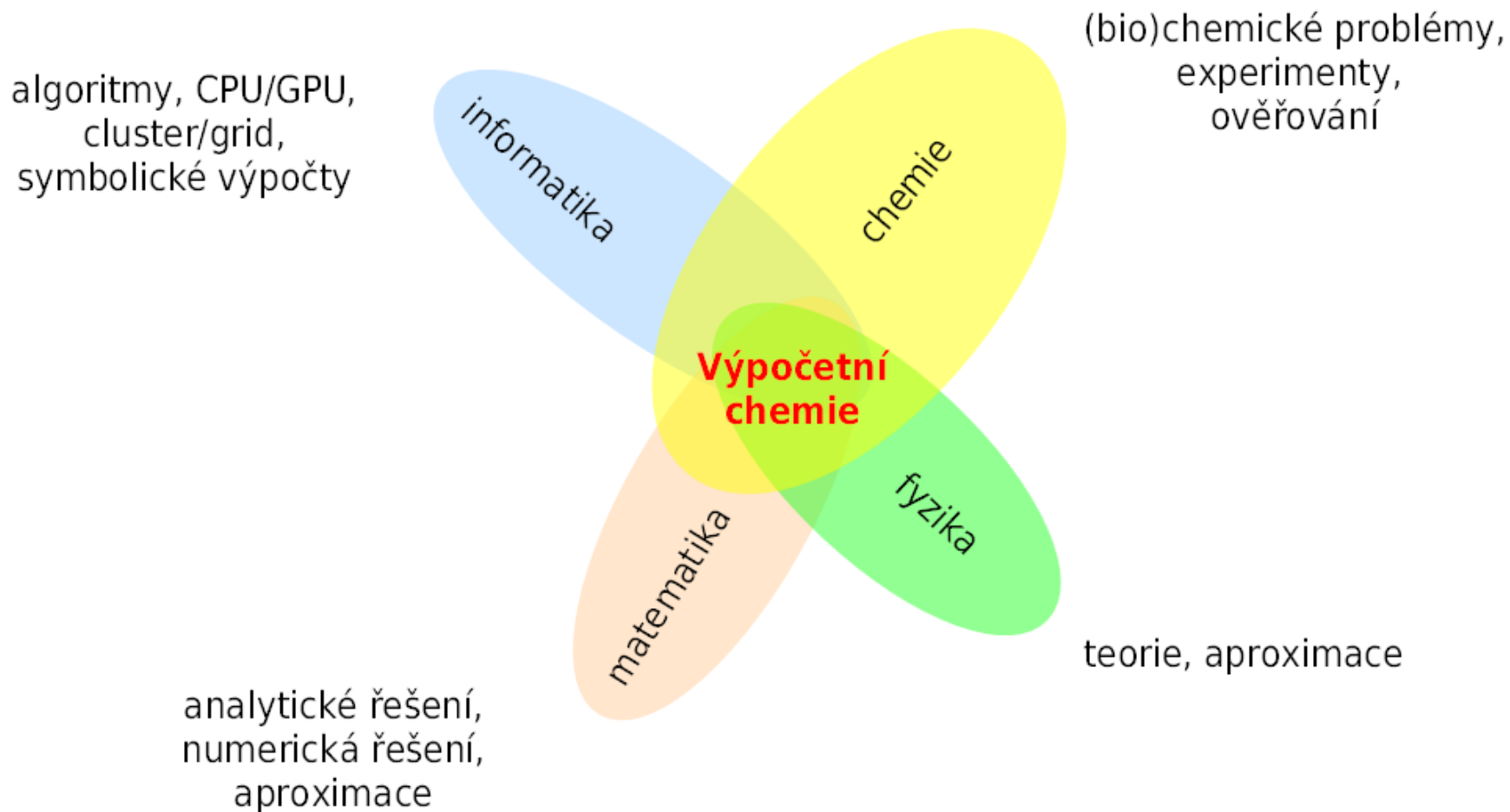
Výpočetní chemie

Výpočetní chemie (Computational Chemistry, počítačová chemie)

je odvětví chemie, které využívá **počítačů** při řešení **chemických problémů**. Používá výsledků **teoretické chemie** implementované do výkonných **počítačových programů** určených k výpočtům struktury, vlastností a reaktivity molekul a pevných látek. I když její výsledky doplňují informace získané chemickými experimenty, v určitých případech může předpovědět doposud nepozorované chemické jevy. Výpočetní chemie je široce používaná v **návrhu nových léčiv a materiálů**.

www.wikipedia.org

Interdisciplinární obor



Co děláme a nabízíme ...

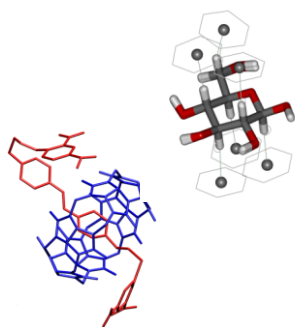
- **expertíza v oblastech výpočetní chemie**

molekulárně dynamické simulace, dokování, chemo a bioinformatika, vazebné energie, kvantová mechanika, hrubozrnné modely ...

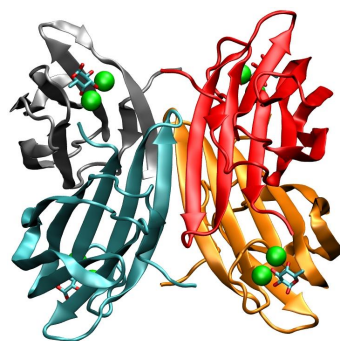
- **vývoj výpočetních nástrojů, postupů a softwaru**

in silico návrh léčiv a mutantních proteinů, výpočty volných energií, rychlé výpočty nábojů, virtuální screening, predikce tunelů a kavit v biomolekulách ...

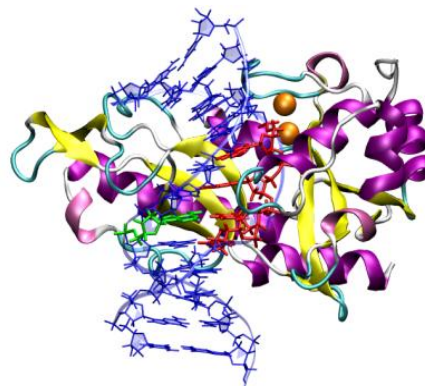
- **studium systémů různých velikostí**



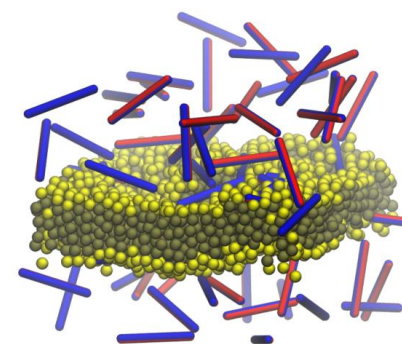
malé komplexy



lectiny

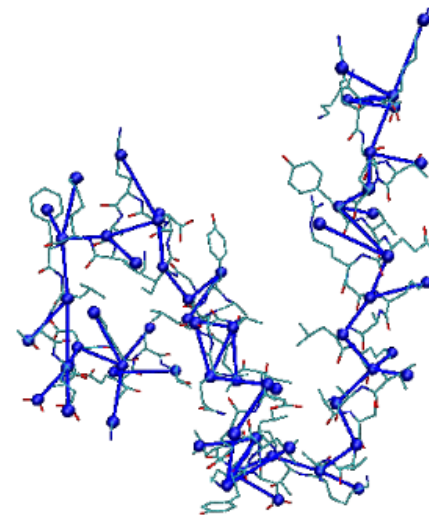
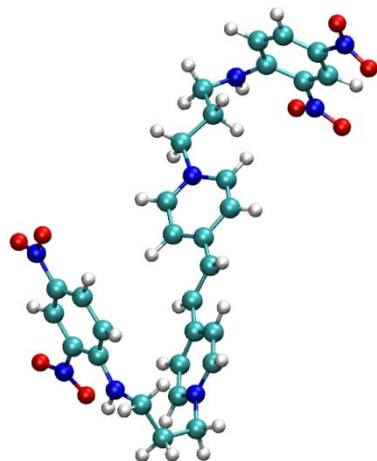
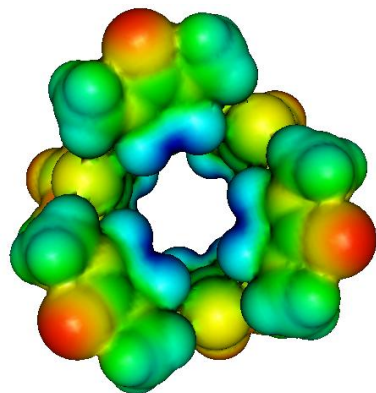


enzymy



velké biomolekulární komplexy

Úrovně teorie



Kvantová mechanika

Molekulová mechanika

***Coarse-grained* mechanika**

atomové rozlišení

bead resolution

reaktivita

konformační pohyby

pohyb domén, folding

až 1'000 atomů *

až 1'000'000 atomů *

až 1'000'000 beads *

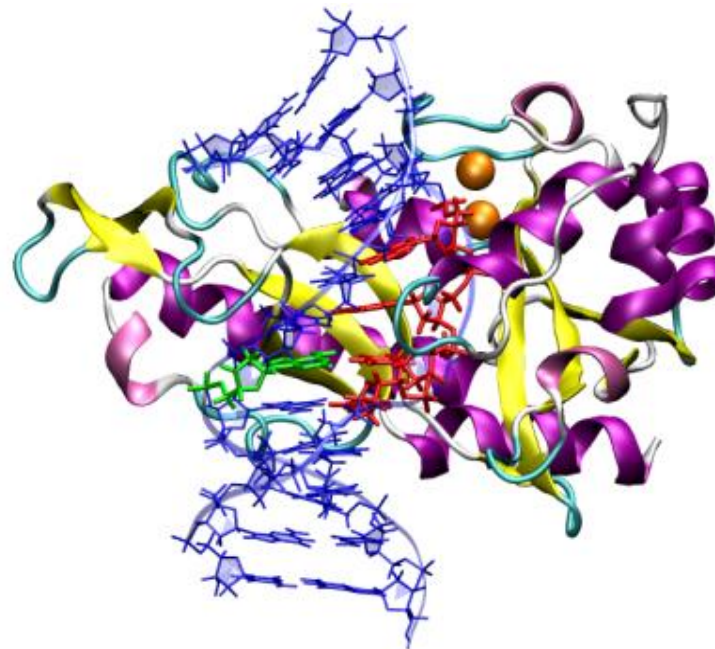
až 100 ps *

až 1 μ s *

až ms *

Projekty

Studium (bio)molekulárních systémů



Kvantově chemické výpočty

časově nezávislá Schrödingerova rovnice

$$\hat{H} \psi_k(\mathbf{r}) = E_k \psi_k(\mathbf{r})$$

Formální škálování	Metody			
	HF	CI metody	MP metody	CC metody
$N^4 \rightarrow N^2 \rightarrow N^1$	HF, DFT			
N^5			MP2	CC2 (iterativní)
N^6		CISD	MP3, MP4(SDQ)	CCSD (iterativní)
N^7			MP4	CCSD(T), CC3 (iterativní)
N^8		CISDT	MP5	CCSDT
N^9			MP6	
N^{10}		CISDTQ	MP7	CCSDTQ (iterativní)

Škálování, časová náročnost: http://en.wikipedia.org/wiki/Time_complexity

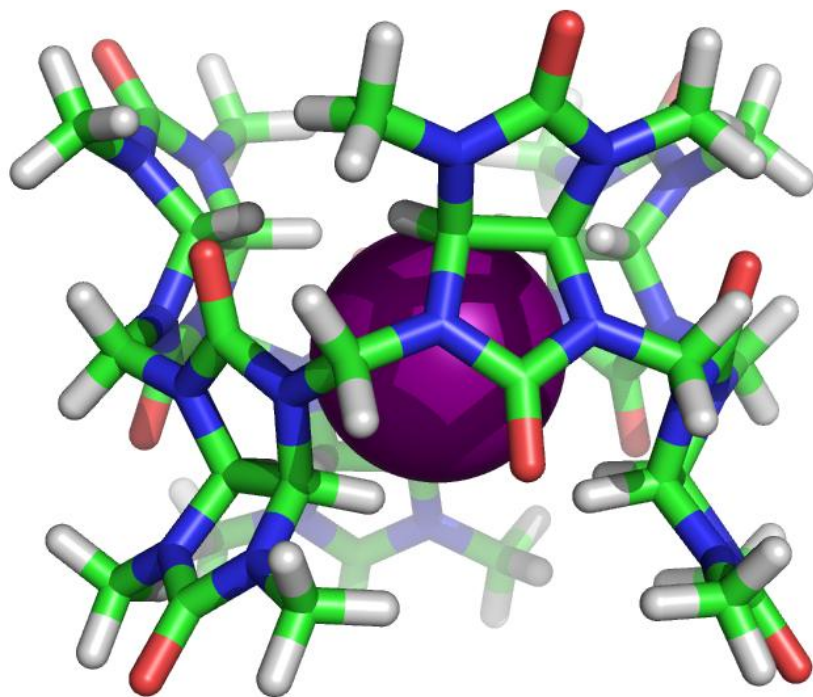
HF - Hartreeho–Fockova metoda, DFT - teorie funkcionálu hustoty,

CI - metody konfigurační interakce, MP - Møllerova–Plessetova poruchová teorie,

CC - metoda vázaných klastrů, N - počet bázeových funkcí

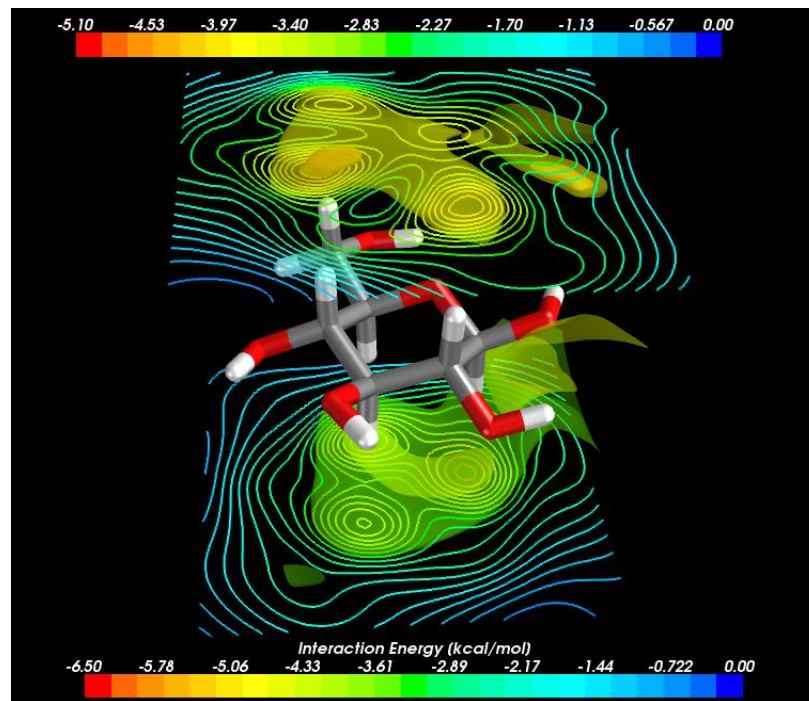
Jensen, F. Introduction to computational chemistry; 2nd ed.; John Wiley & Sons: Chichester, England; Hoboken, NJ, 2007.

Kvantově chemické výpočty



supramolekulární komplexy

vazebné schopnosti sacharidů



Molekulová mechanika

Schrödingerova rovnice => kvantově mechanický pohled

$$H_a \Psi(r) = E(R) \Psi(r)$$

aproximace využívající klasickou fyziku
neuvažuje se explicitní pohyb elektronů
(pohyb je implicitně zahrnut v empirických parametrech)

$$E(R) = \underbrace{E_{bonds} + E_{angles} + E_{torsions}}_{\text{vazebné příspěvky}} + \underbrace{E_{el} + E_{vdw}}_{\text{nevazebné příspěvky}}$$

Klasická fyzika => mechanický pohled

vazebné příspěvky

nevazebné příspěvky

Formální škálování: $N^2 \rightarrow N \log_2 N$

N - počet atomů

Molekulová dynamika

$$-\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}} = \mathbf{F} \quad \boxed{\mathbf{F}_i = m_i \mathbf{a}_i} \quad \mathbf{a}_i = \frac{d^2 \mathbf{r}_i}{dt^2}$$

II. Newtonův pohybový zákon (zákon síly)

$$-\frac{\partial E(\mathbf{R})}{\partial \mathbf{R}} = m_i \frac{d^2 \mathbf{r}_i}{dt^2}$$

soustava diferenciálních rovnic druhého řádu
vyžaduje numerické řešení



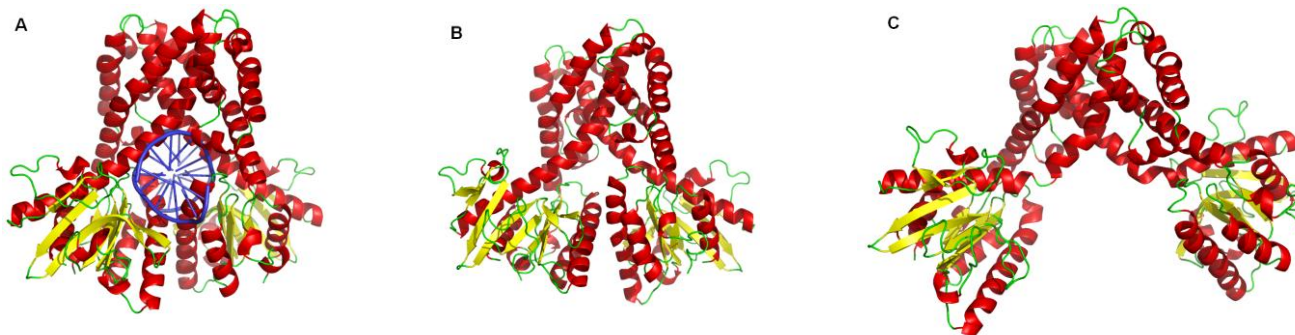
diskretizace molekulárního pohybu v krátkých
časových intervalech

dáno nejrychleším pohybem (vibrace vazeb) \longrightarrow **1 fs** typický integrační krok

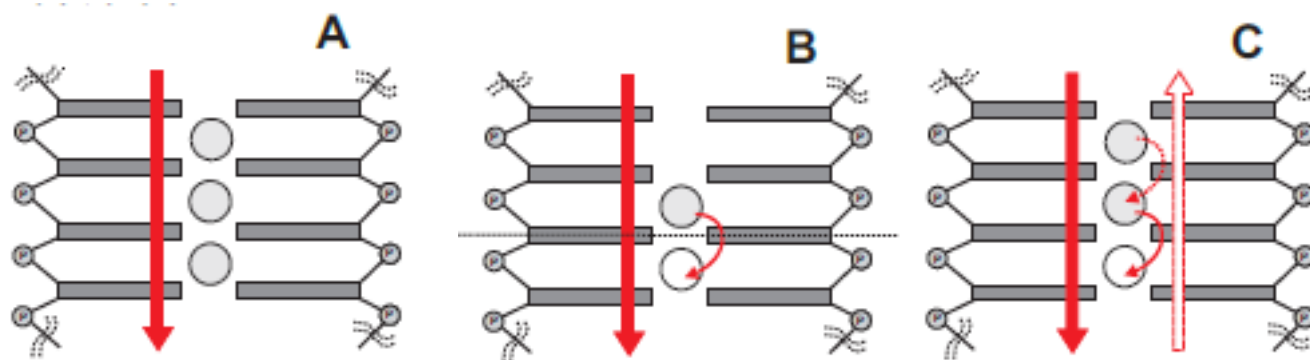
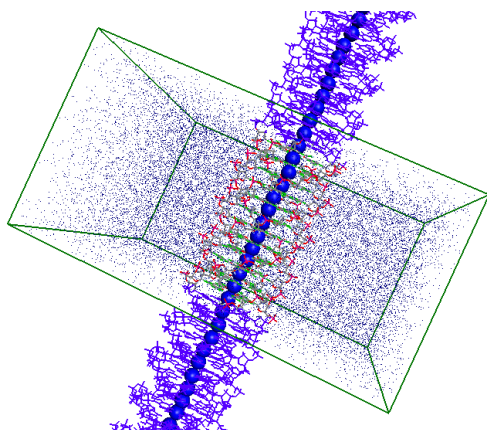
Nedokonalosti v integraci se odstraňují použitím **termostatů** a **barostatů**, které zároveň zajišťují požadované podmínky simulace.

Konformační pohyby a transport

- Konformační pohyby a transport jsou důležité pro funkci bimolekulárních systémů

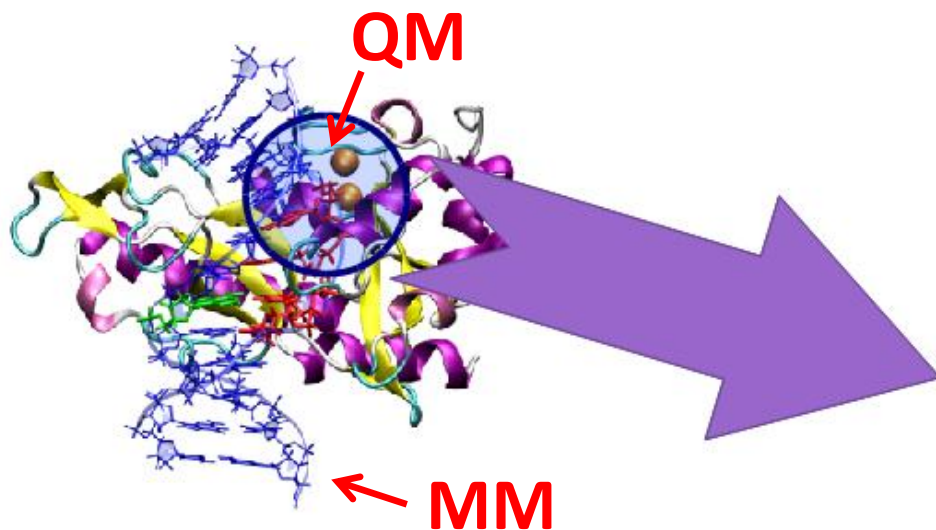


otevírání endonukleasy BsoBI

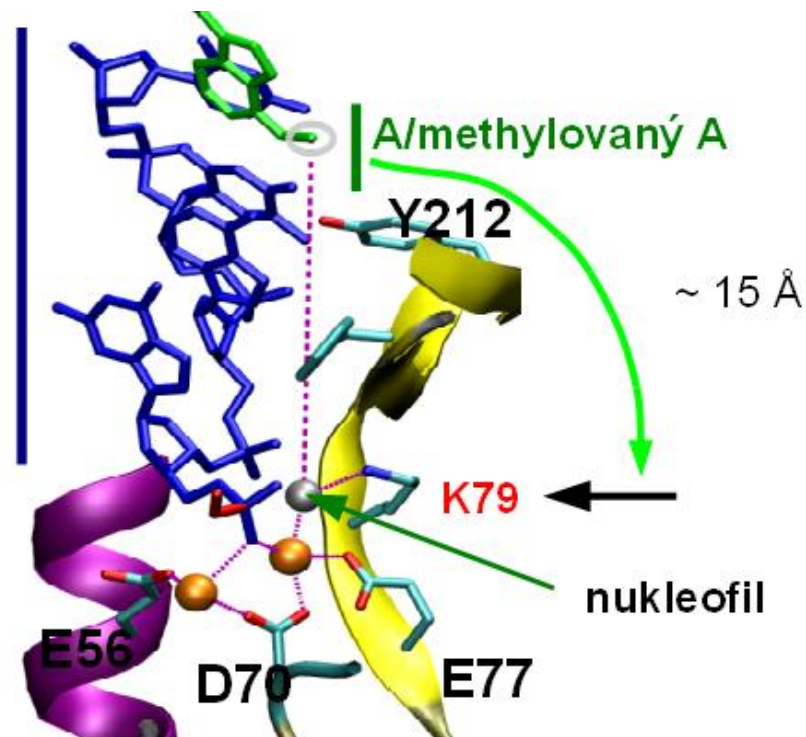


transport ionů v DNA kvadruplexech

Reakční mechanismy enzymů QM/MM



Je součástí opravných mechanismů poškozené DNA v bakteriích.

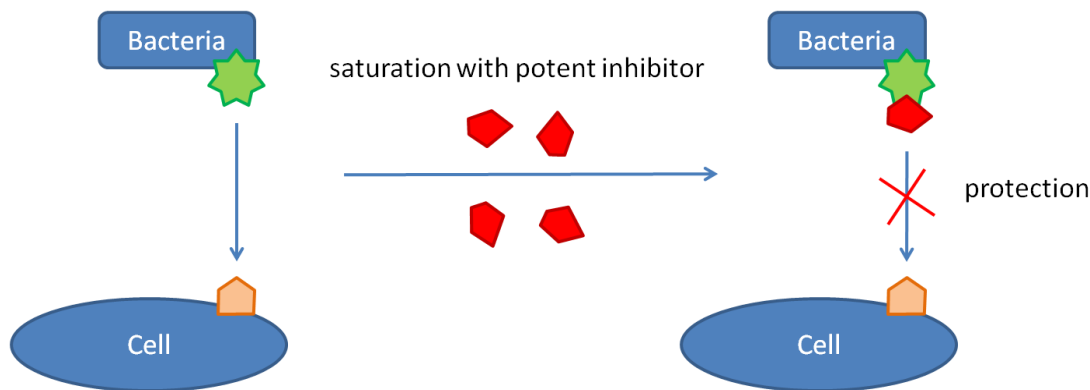


Katalyzovaná reakce – hydrolýza fosfodiesterové vazby

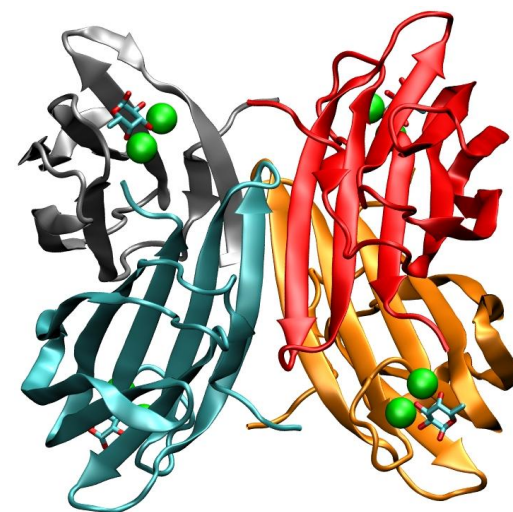


Virtuální screening

Motivace

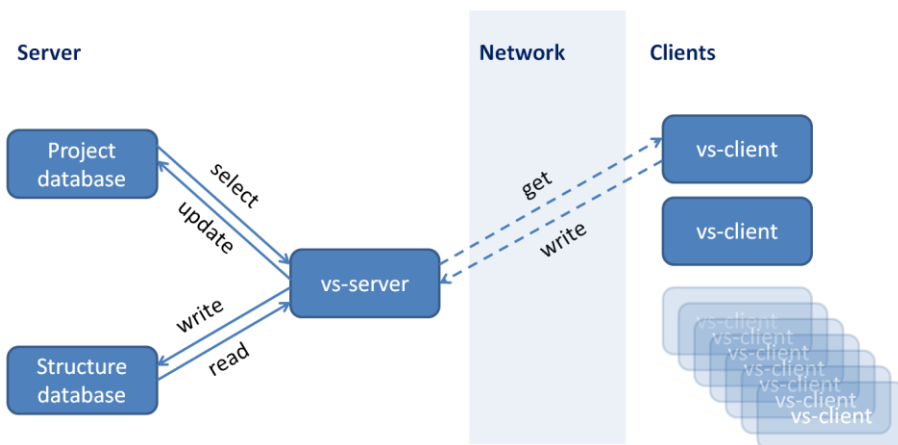


Cíl



Lectin PA-III

Software nástroje



Výkonnost

- Autodock Vina
- heterogenní výpočetní zdroje
- jeden doking cca 1-10 minut/ligand
- ca **900 souběžně běžících úloh**
- rychlost hledání cca **250 000 ligandů za den**

Kontakty

Laboratoř výpočetní chemie

Národní centrum pro výzkum biomolekul, UKB, PAVILON A4

<http://lcc.ncbr.muni.cz>



Semináře LCC skupiny každý čtvrtek v 10 hodin v místnosti 2.11/A4.

Výpočetní centra v ČR

Cílem projektu **MetaCentrum** je provoz a koordinace distribuované výpočetní infrastruktury a datových úložišť a odpovídajícího podpůrného prostředí v České republice jako součást pan-evropské infrastruktury budované v rámci projektu EGI Inspire.

Vytvoření virtuálního pracovního prostředí MetaCentrum přispívá k podstatně efektivnějšímu využití instalované techniky, umožňuje využití dostupných výpočetních zdrojů pro řešení velmi náročných výpočetních úloh, jejichž zvládnutí je nad možností samostatného pracoviště v ČR.

MetaCentrum je aktivita sdružení **CESNET**.

Centrum CERIT-SC (CERIT Scientific Cloud) je národním centrem poskytujícím flexibilní úložné a výpočetní kapacity a související služby, včetně podpory jejich experimentálního využití. Současně centrum provádí výzkum a vývoj v oblasti flexibilních e-infrastruktur a spolupracuje na výzkumných aktivitách svých uživatelů.

Centrum CERIT-SC vzniká transformací Superpočítačového centra Brno (SCB), které je součástí Ústavu výpočetní techniky (ÚVT) Masarykovy univerzity (MU).

Centrum CERIT-SC bude plánuje do roku 2013 poskytovat více než 3500 jader a cca 3,5 PB úložného prostoru. Pořizované výpočetní a úložné kapacity jsou zprovozňovány postupně a jsou dostupné prostřednictvím Národní gridové infrastruktury, cloudová rozhraní a další experimentální rozhraní.

Sdružuje výpočetní zdroje poskytované MetaCentrem, projektem CERIT-SC a dalšími partnery.

- Národní gridová infrastruktura
 - OS Debian
 - ca **2500 CPU** jader
 - **CEITEC/NCBR vlastní zdroje cca 850** CPU jader
 - 3 x 100 TB úložných diskových polí
 - cca **3 TB na uživatele**
-
- Účet může získat student libovolné vysoké školy ČR.
 - Přístup není vázán na konkrétní projekt a je udělen na 1 rok.
 - Prodloužení přístupu je podmíněno odevzdáním výroční zprávy.

Sdružuje výpočetní zdroje poskytované MetaCentrem, projektem CERIT-SC a dalšími partnery.

- Národní gridová infrastruktura
- OS Debian
- ca **2500 CPU** jader
- **CEITEC/NCBR vlastní zdroje cca 850** CPU jader
- 3 x 100 TB úložných diskových polí
- cca **3 TB** na uživatele

**Prakticky: infrastruktura,
klastry, datové úložiště,
zadávání úloh**

- Účet může získat student libovolné vysoké školy ČR.
- Přístup není vázán na konkrétní projekt a je udělen na 1 rok.
- Prodloužení přístupu je podmíněno odevzdáním výroční zprávy.

IT4Innovations je projekt, jehož cílem je vybudovat národní centrum excelentního výzkumu v oblasti informačních technologií. Součástí projektu bude pořízení velmi výkonného superpočítače, který by měl být uveden do provozu okolo roku 2014.

Základem centra bude **computing** (počítání), který je formulován do tří vzájemně propojených klíčových oblastí výzkumu:

- **IT4People (Information Technology for People)** – výzkum zaměřený na zlepšení kvality života společnosti prostřednictvím moderních informačních technologií.
- **SC4Industry (Supercomputing for Industry)** – superpočítačové výpočty pro řešení průmyslových problémů, modelování v oblasti přírodních věd a nanotechnologií (tvarové optimalizace, návrh materiálů, biomechanické simulace, ...).
- **Theory4IT (Theory for Information Technology)** - oblast zaměřená do základního výzkumu, a to především na rozvoj nových netradičních výpočetních metod (dolování znalostí, teorie mravenišť).

Projekt společně připravuje pět subjektů: Vysoká škola báňská-Technická univerzita Ostrava, Ostravská univerzita v Ostravě, Slezská univerzita v Opavě, Vysoké učení technické v Brně a Ústav geoniky AV ČR.

Výpočetní centra v zahraničí

Project Types:

- **Multi-year Access** is available to major European projects or infrastructures that can benefit from PRACE resources and for which Project Access is not appropriate.
- **Project Access** is intended for individual researchers and research groups including multi-national research groups and has a one year duration. Calls for Proposals for Project Access are issued twice yearly (February and September).
- **Preparatory Access** is intended for resource use required to prepare proposals for Project Access. Applications for Preparatory Access are accepted at any time.

PRACE - členové

Austria: JKU - Johannes Kepler University of Linz

Belgium: DGO6-SPW - Direction générale opérationnelle de l'Économie, de l'Emploi et de la Recherche – Service Public de Wallonie

Bulgaria: NCSA - Executive agency "Electronic communication networks and information systems"

Cyprus: CaSToRC – Computation-based Science and Technology Research Center, The Cyprus Institute

Czech Republic: VŠB - Technical University of Ostrava

Denmark: DeIC - Danish e-Infrastructure Cooperation

Finland: CSC - IT Center for Science Ltd.

France: GENCI - Grand Equipement National de Calcul Intensif

Germany: GCS - GAUSS Centre for Supercomputing e.V

Greece: GRNET - Greek Research and Technology Network S.A.

Hungary: NIIFI - National Information Infrastructure Development Institute

Ireland: ICHEC - Irish Centre for High-End Computing

Israel: IUCC - Inter-University Computation Center

Italy: CINECA - Consorzio Interuniversitario

Norway: SIGMA – UNINETT Sigma AS – The Norwegian Metacenter for Computational Science

The Netherlands: SURFSARA: SARA Computing and Networking Services

Poland: PSNC – Instytut Chemii Bioorganicznej Pan – Institute of Bioorganic Chemistry – Poznan Supercomputing and Networking Center

Portugal: Universidade de Coimbra

Serbia: IPB - Institute of Physics Belgrade

Slovenia: ULFME - University of Ljubljana, Faculty of Mechanical Engineering

Spain: BSC – Barcelona Supercomputing Center – Centro Nacional de Supercomputación

Sweden: Vetenskapsrådet – Swedish Research Council

Switzerland: ETH – Eidgenössische Technische Hochschule Zürich – Swiss Federal Institute of Technology, Zürich

Turkey: UYBHM – Ulusal Yuksek Basarimli Hesaplama Merkezi, Istanbul Technical University – National Center for High Performance Computing

UK: EPSRC – The Engineering and Physical Sciences Research Council

TOP500

<http://www.top500.org/>

TOP500 je projekt, který udržuje seznam 500 nejrychlejších počítačů na světě.

TOP500 benchmark

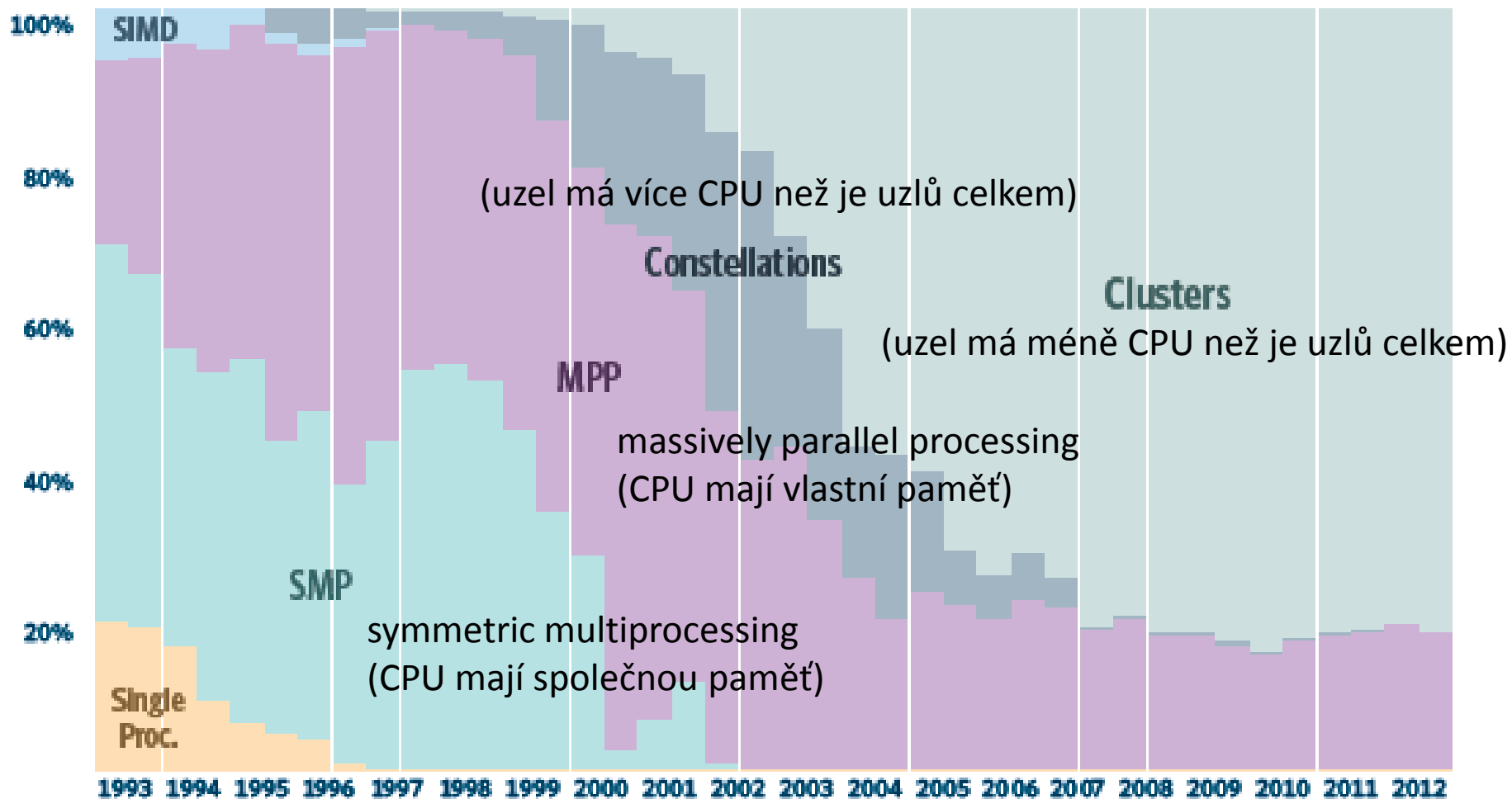
Our simple **TOP500** approach does not define “supercomputer” as such, but we use a benchmark to rank systems and to decide on whether or not they qualify for the TOP500 list. The benchmark we decided on was **Linpack**, which means that systems are ranked only by their ability to solve a set of linear equations, $Ax = b$, using a **dense random matrix A**.

Listopad 2012

	NAME	SPECS	SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	TITAN	Cray XK7, Operon 6274 16C 2.2 GHz + Nvidia Kepler GPU, Custom Interconnect	DOE/OS/ORNL	USA	560,640	17.6	8.3
2	SEQUOIA	IBM BlueGene/Q, Power BQC 16C 1.60 GHz, Custom Interconnect	DOE/NNSA/LLNL	USA	1,572,864	16.3	7.9
3	K COMPUTER	Fujitsu SPARC64 VIIIfx 2.0GHz, Custom Interconnect	RIKEN AICS	Japan	705,024	10.5	12.7
4	MIRA	IBM BlueGene/Q, Power BQC 16C 1.60 GHz, Custom Interconnect	DOE/OS/ANL	USA	786,432	8.16	3.95
5	JUQUEEN	IBM BlueGene/Q, Power BQC 16C 1.60 GHz, Custom Interconnect	Forschungszentrum Jülich	Germany	393,216	4.14	1.97

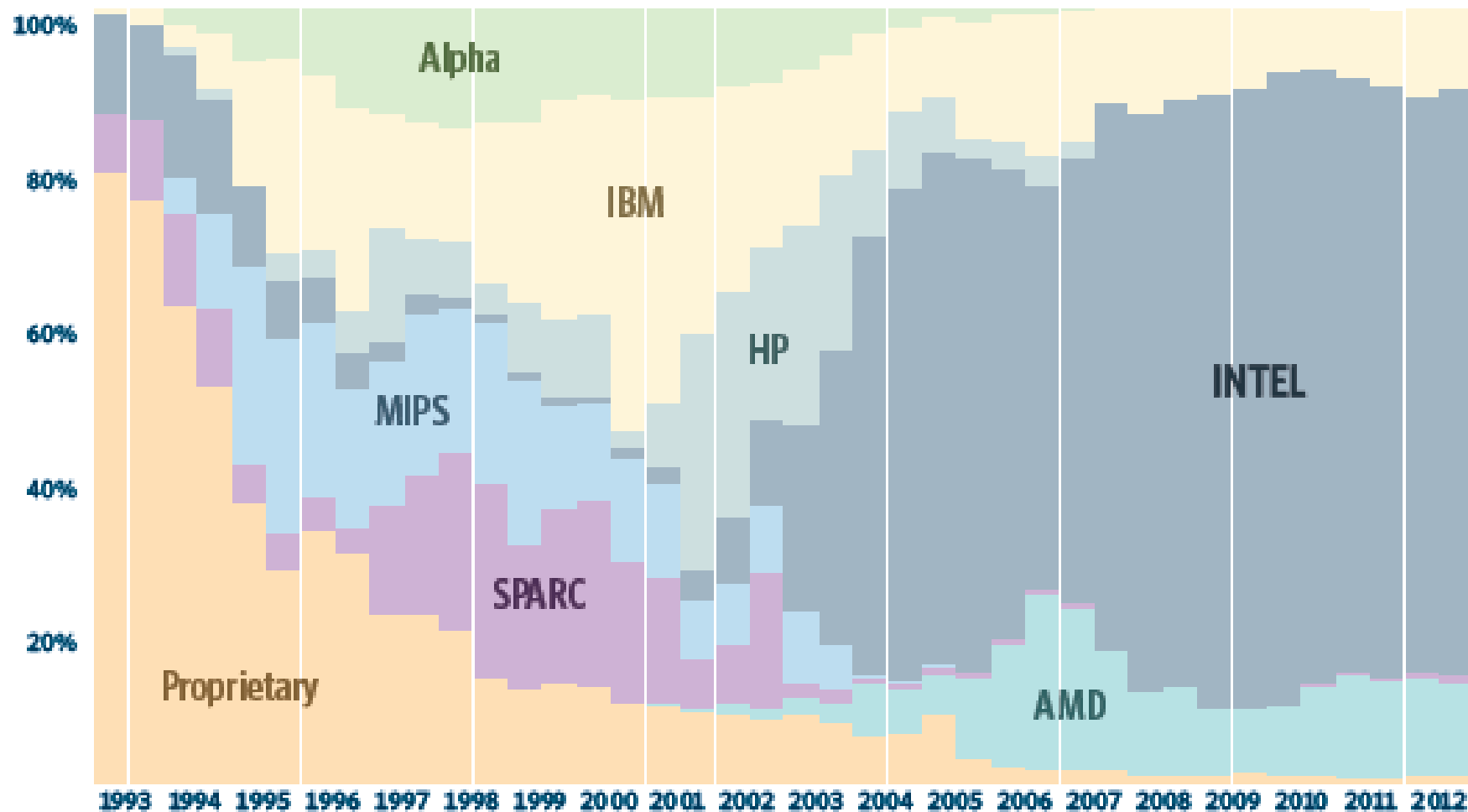
TOP500 – Topologie

ARCHITECTURES



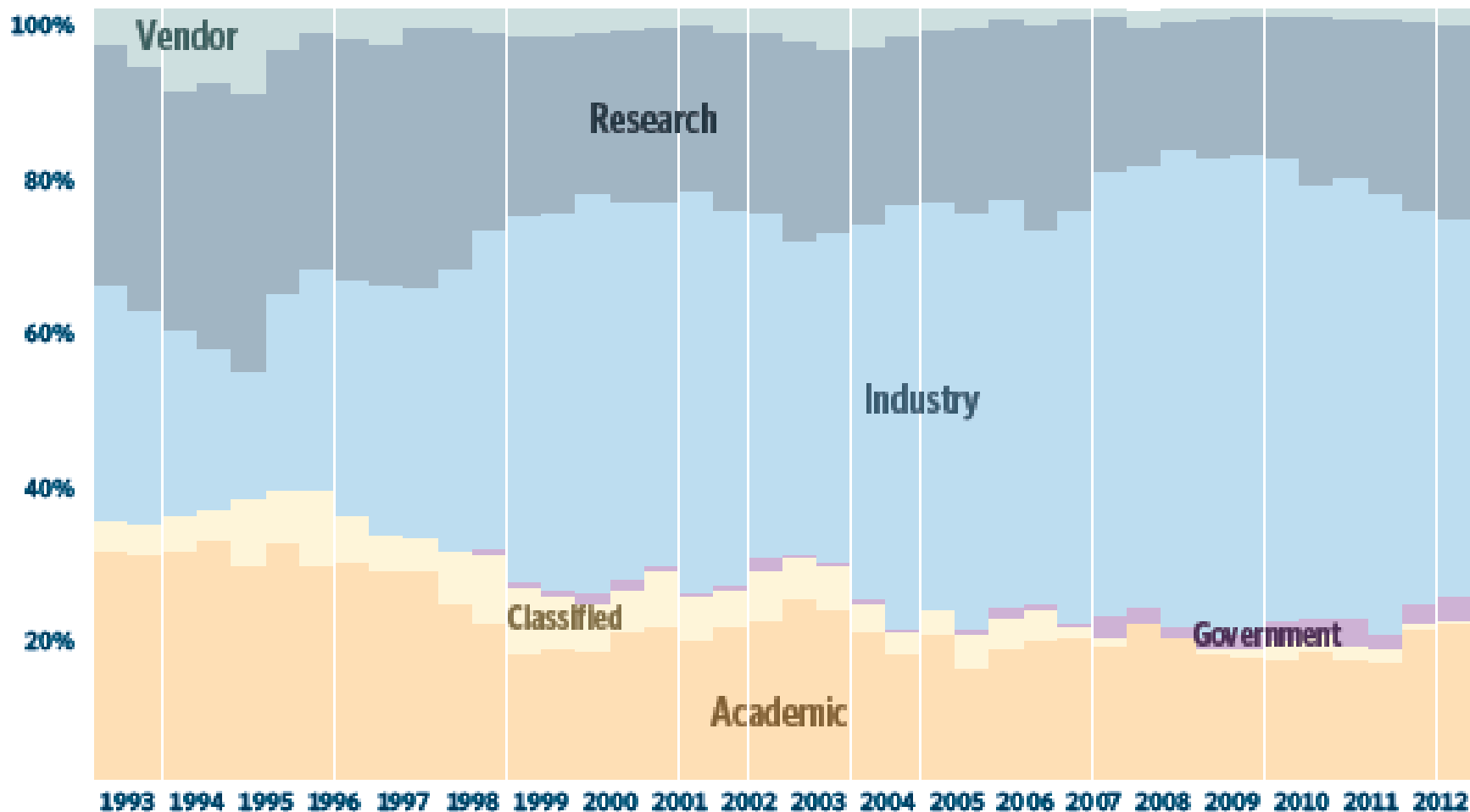
TOP500 – CPU architektura

CHIP TECHNOLOGY



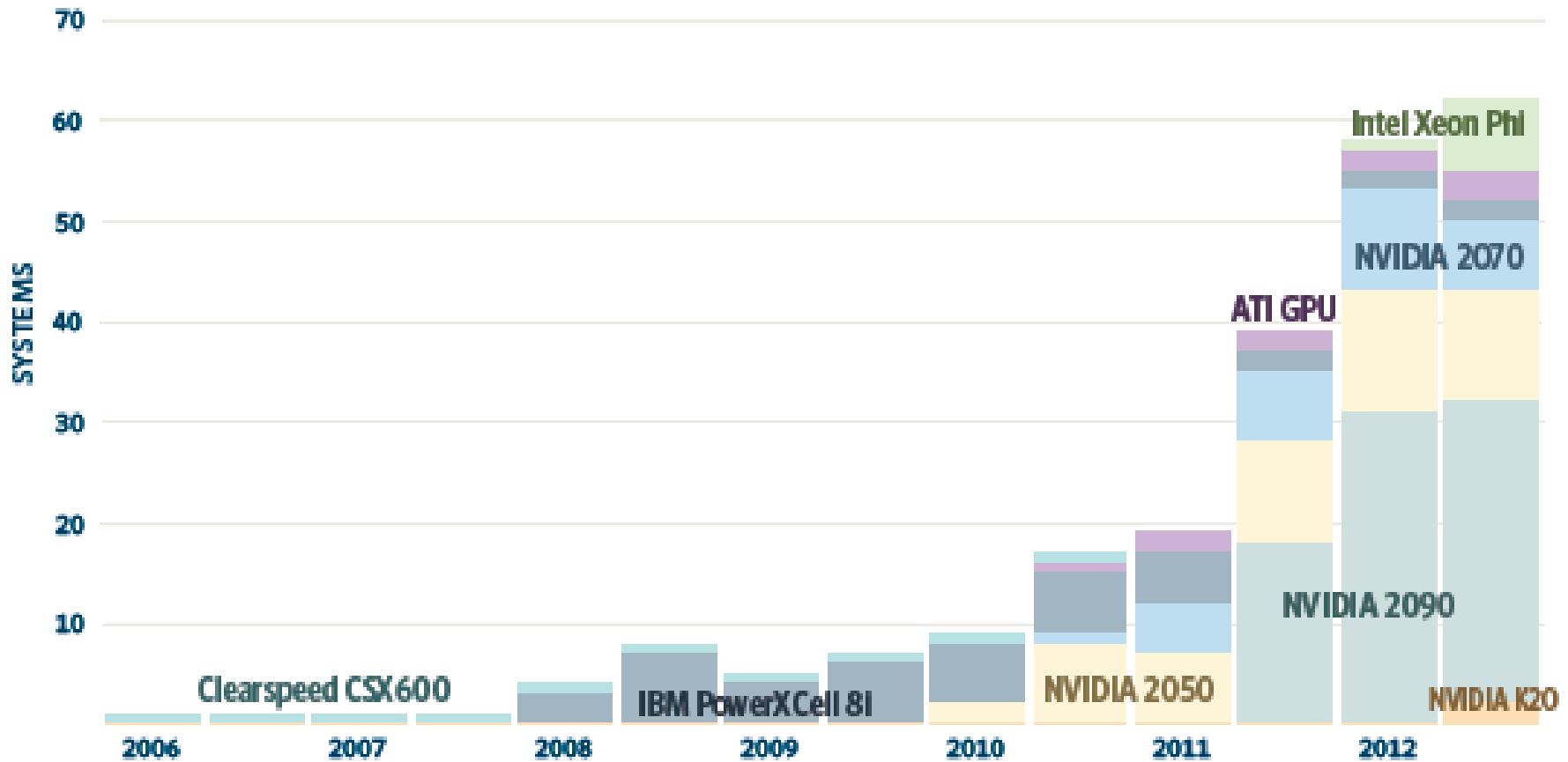
TOP500 – Typ použití

INSTALLATION TYPE



TOP500 – Akcelerátory/Koprocesory

ACCELERATORS/CO-PROCESSORS



K – computer, 3. místo

<http://www.youtube.com/watch?v=UJPslu90aTc>

pustit video

Cvičení LI.1

1. Kolikrát se prodlouží výpočet energie molekuly benzenu kvantově chemickou metodou CCSD(T) pokud změníme použitou bázi z aug-cc-pVDZ na aug-cc-pVTZ? Počet bázových funkcí je 192 pro aug-cc-pVDZ a 414 pro aug-cc-VTZ.
2. Kolik dní bude trvat simulace fragmentu celulosy po dobu 1 μ s ve vodním boxu o celkovém počtu 408609 atomů na jedné grafické kartě typu GTX680 za NPT podmínek? Pro řešení použijte data poskytnutá zde:
<http://ambermd.org/gpus/benchmarks.htm#Benchmarks>
3. V roce 2009 činila průměrná spotřeba energie na jednoho obyvatele domácnosti v ČR 1398 kWh (statistika Energetického regulačního úřadu). Kolik lidí by pokrylo svoji roční spotřebu z energie, kterou spotřebovává nejvýkonnější počítač světa Titan s výpočetním výkonem cca 17 PFLOPS?