

# hpc focus



天河

天河

精

精



5.

VÝROČIE  
SUPERPOČÍTAČA  
AUREL





Mgr. Lukáš Demovič, PhD.  
Riaditeľ VS SAV

## Úvodné slovo riaditeľa

Vážení čitatelia,

**19. októbra** je to presne **5 rokov**, čo bol do ostrej prevádzky spustený najvýkonnejší slovenský superpočítač Aurel. Za ten čas sa toho udialo veľa - v spolupráci s našimi dodávateľmi, firmami Datalan a IBM, sme vychytali "detské choroby" Aurela a zorganizovali niekoľko školení pre jeho používateľov, ktoré odučili priamo vývojári superpočítača z amerického Poughkeepsie. Úspešne sme sa etablovali na európskej úrovni prostredníctvom nášho členstva v organizácii PRACE, vďaka ktorému majú slovenskí používatelia prístup k najvýkonnejším HPC zariadeniam v Európe a celému radu aktivít a podujatí, ktoré táto organizácia usporiada.

Ako je tiež známe, vyskytli sa aj problémy, ktoré súviseli hlavne s nedostatkom finančných prostriedkov na prevádzku Aurela, čo v konečnom dôsledku viedlo k dočasnému odstaveniu časti zariadenia - Aurel však ako celok nebol nikdy vypnutý z prevádzky, a naši používatelia na ňom mohli počítať za nezmenených podmienok. Dlhodobým problémom boli aj odlišné názory vtedajšieho vedenia Výpočtového strediska a predsedníctva SAV na prevádzkovanie Aurela, ktoré napokon vyvrcholili odvolaním

riaditeľa VS z funkcie. Pri pohľade na výstupy projektu však musí byť každému zrejmé, že ide o jasný úspech - veď viac ako 170 registrovaných používateľov spomedzi slovenských vedcov ukázalo, že superpočítač vedia nielen vyťažiť, ale hlavne využiť na zmysluplnú prácu, o čom svedčí vyše 85 %-ná vyťaženosť a takmer 90 publikácií v zahraničných recenzovaných vedeckých časopisoch, ale aj neustály záujem nových používateľov o používateľský účet.

Môžete si povedať, že 5 rokov nie je tak veľa, ale vo svete IT (a zvlášť v HPC) sú to celé veky, a tak ako časom zastaráva každý prístroj, pre superpočítač to platí dvojnásobne. Vývoj ide dopredu neuveriteľným tempom a dnešné HPC zariadenia sú oproti Aurelovi nielen výkonnejšie, ale hlavne energeticky efektívnejšie. Menia sa tiež požiadavky používateľov, pribúdajú vedné odbory, ktoré majú diametrálne iné nároky ako tie, na ktoré bol navrhnutý Aurel. Je preto najvyšší čas začať riešiť otázku pokračovania projektu SIVVP a hľadať podporu a financovanie superpočítača ďalšej generácie. Ako riaditeľ Výpočtového strediska SAV pre to hodlám spraviť maximum. Dovolím si tiež tvrdiť, že v tejto otázke sme za jedno aj s vedením SAV, pozitívne signály sa ukazujú aj pri jednaniach s decíznou sférou. Preto dúfam, že v ďalšom vydaní HPC Focus-u vám už budem môcť podrobnejšie predstaviť koncepciu ďalšieho smerovania projektu.

Pri tejto slávnostnej príležitosti by som však chcel aj poďakovať bývalému riaditeľovi VS SAV Ing. Tomášovi Lackovi za množstvo obetavej práce, ktorú pri vytváraní a implementácii projektu SIVVP vynaložil, ale aj partnerom z Ústavu informatiky SAV, Ústavu experimentálnej fyziky SAV, Slovenskej technickej univerzity v Bratislave, Žilinskej univerzity v Žiline, Univerzity Matej Bela v B. Bystrici a Technickej univerzity v Košiciach. Ďalej by som chcel poďakovať svojim kolegom z oddelenia VVT, ktorí sa starajú o chod Aurela a o používateľskú podporu, ako aj pracovníkom z dodávateľských firiem Datalan a IBM pomáhajúcim nám riešiť rôzne technické problémy, ktoré sa pri prevádzke vyskytli. Veľká vďaka tiež patrí vedeniu SAV za podporu projektu, ako aj férovú a konštruktívnu diskusiu o jeho pokračovaní. Najviac si však ceníme podporu a záujem našich používateľov z ústavov SAV a slovenských univerzít, bez ktorých by naše úsilie nemalo zmysel.

Vám, našim čitateľom, by som chcel popriať príjemné čítanie tohtoročného vydania HPC Focus-u, a ak máte hlbší záujem o HPC, neváhajte nás kedykoľvek kontaktovať - sme tu pre Vás!

*Lukáš Demovič*



Strana 06 - 17

## VVP

Článok pána prof. Ing. Ivan Plander, DrSc. o masívnom paralelizme a vysokovýkonnom počítaní.

Strana 18 - 35

## APLIKÁCIE VVP

Články oslovených užívateľov superpočítača Aurel: Ing. Marek Macák, PhD.; RNDr. Magdalona Majeková, PhD. a Mgr. Michal Novotný.

Strana 36 - 47

## PRACE

Stať, v ktorej nájdete bližšie informácie o programe H2020 organizácie PRACE, ktorá od svojho počiatku organizuje medzinárodné dvojmesačné stáže s názvom Summer of HPC, a rozhovory s jeho účastníkmi.

46 Terabit/sec  
Optical Backplane

01

W/P

# IVAN PLANDER

## Masívny paralelizmus a vysokovýkonné počítanie

**P**aralelné vysokovýkonné počítanie je úzko spojené so superpočítačmi. Superpočítače sú vysokovýkonné počítače založené na paralelných architektúrach. Podľa Flynnovej taxonómie do kategórie paralelných architektúr spadajú počítačové architektúry **SIMD** (Single Instruction- Multiple Data) a **MIMD** (Multiple Instruction-Multiple Data). Prevládajúci štýl paralelného počítania reprezentuje počítačová architektúra **SPMD** (Single Program-Multiple Data), kde všetky procesory používajú rovnaký program a každý procesor má svoje vlastné dáta. Počítačová architektúra SIMD, často implementovaná v superpočítačoch, umožňuje obrovské výpočtové rýchlosti, vďaka masívnemu paralelizmu. Metódy riešenia vedecko-technických problémov sú súčasťou výpočtovej vedy (computational science). Matematické modely týchto problémov sú spravidla opísané systémami lineárnych a nelineárnych rovníc alebo systémami obyčajných alebo parciálnych diferenciálnych rovníc, stochastickými reláciami a pod. Riešenie takých modelov v uzavretej forme zväčša nie je možné, používať sa musia numerické metódy a ich spracovanie na superpočítačoch. Paralelné architektúry počítačov a paralelné algoritmy výpočtov sú náplňou teórie vysokovýkonného počítania.

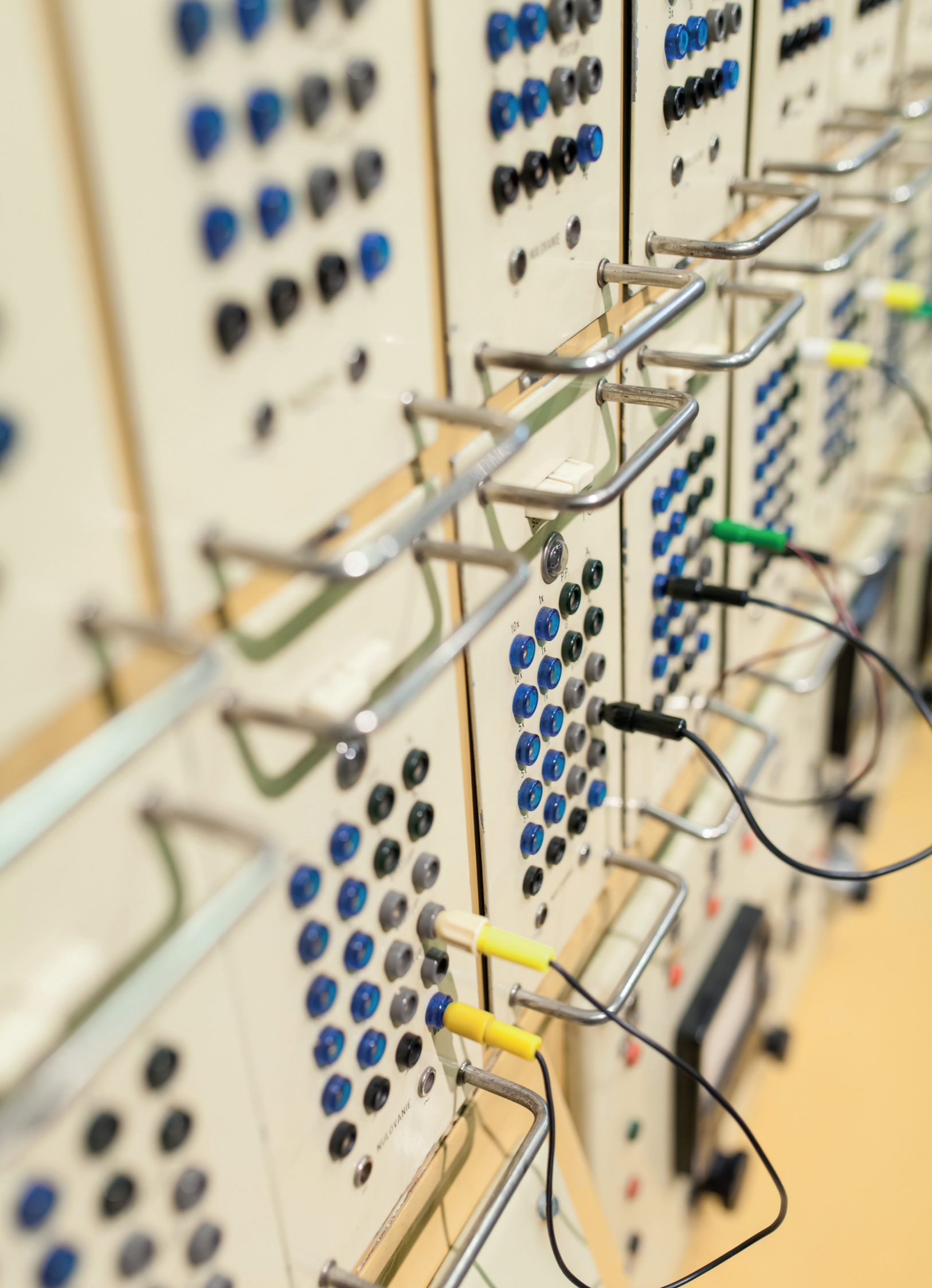
Počítače a informatika sa rozvíjajú tromi hlavnými smermi: Rozvojom súčiastkovej základne a technológií, rozvojom počítačových architektúr a aplikovaním umelej inteligencie.

Paralelné počítačové architektúry sú reprezentované tromi druhmi paralelných systémov:

- » superpočítačmi,
- » počítačovými klastrami,
- » počítačovými sieťami.

Pre vedecko-technické výpočty najdôležitejšie sú superpočítače založené na vysokom paralelizme tesne viazaných procesorov. Počítačové architektúry sa dnes prudko rozvíjajú a diverzifikujú. V minulosti väčšina programov sa spracúvala na konvenčných jedno-jadrových procesoroch, umožňujúcich vykonávať jedno vlákno v čase. V poslednom čase tieto jednoduché CPU sa nahrádzajú mnoho-jadrovými CPU obsahujúcimi akcelerátory, ktoré sú schopné vysokovýkonného univerzálneho počítania. Za ostatných niekoľko rokov sa objavila nová trieda systémov vysokovýkonného počítania **HPC** (High Performance Computing). Tieto systémy používajú nekonvenčné procesorové architektúry, ako napr. bunkové procesory (Cell Processors) alebo programovateľné polia resp. grafické procesorové jednotky (**GPU** - Graphics Processing Units) - pre „náročné“ výpočty a konvenčné CPU - zväčša pre nie-výpočtovo-intenzív-





ne úlohy, ako sú vstupy/výstupy, komunikácie a pod.

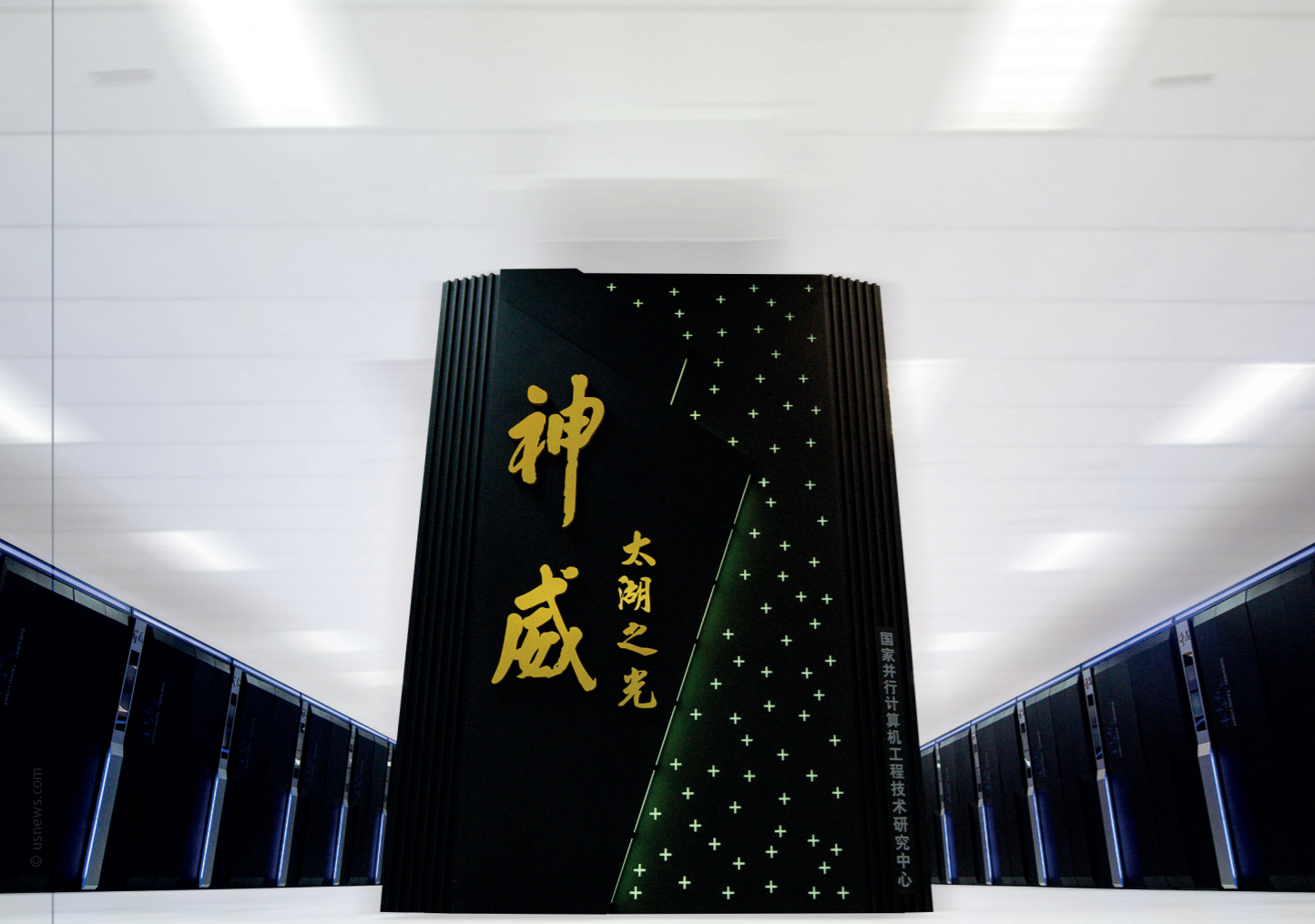
Počítačové klastery poskytujú porovnateľný výpočtový výkon ako superpočítače pri podstatne nižších nákladoch. Významnú úlohu vo vývoji počítačových klastrov, okrem operačného systému UNIX, zohral vývoj softvéru **PVM** (Parallel Virtual Machine). Tento otvorený zdrojový softvér umožnil tvorbu virtuálneho superpočítača – v podstate vysokovýkonného počítačového klastera, vytvoreného z ľubovoľných navzájom prepojených počítačov PC. Heterogénne clustery stojace na vrchole tohto modelu dosiahli celkový výkon porovnateľný s priepustnosťou aj najdrahších superpočítačov, pri podstatne nižšej cene. klastre typu „beowulf“ je schopný vykonávať tesne viazané paralelné vysokovýkonné výpočty na úrovni superpočítača. Gridová štruktúra a gridová paradigma sa často prirovnáva energetickej sieti. Grid umožňuje bezproblémovo virtualizovať zdroje tak, že tieto môžu poskytovať používateľovi pripojenému na Grid prístup k prakticky neobmedzeným výpočtovým cyklom a dátovým pamätiam. Infraštruktúra určujúca, ktoré počítače, pamäte, siete a ostatné zdroje sa používajú v danej aplikácii, sú zväčša pred používateľom skryté podobne, ako jednotlivci pri energetickej sieti zvyčajne nevidia ktorá elektrárňa, ktorý generátor, transformátor resp. iná súčasť systému sú požitá, keď svoj spotrebič pripoja na sieť.

Gridové štruktúry a snahy o vývoj ich aplikácií sa odvíjali od vývoja midlvéru, najmä prostredníctvom multidisciplinárnych tímov. Jedna z najrýchlejšie sa rozvíjajúcich aplikačných oblastí v Gridovom počítaní je biológia, výpočtová biológia, bioinformatika, výpočtové neurovedy a ďalšie oblasti, obsahujúce Gridové technológie, ako spôsob prístupu k spracovaniu, zberu a dolovania dát. Grid poskytuje dôležitú platformu pre realizáciu zdrojovo-intenzívnych vedecko-technických aplikácií, nákladovo efektívnejšie. Fyzikálne aplikácie sú ďalšou rýchle sa rozvíjajúcou triedou Gridových aplikácií. Príkladom je lineárny akcelerátor – Large Hadron Collider v Cerne (Ženeva), ktorý generuje obrovské množstvo dát, pričom každá fyzikálna udalosť môže byť spracovaná nezávisle a vyúsťuje do  $10^{18}$  - cestného paralelizmu. Nie menej dôležité sú komerčné aplikácie.

Súčasný vedecko-výpočtový a komunikačný pokrok v počítačovom hardvéri a softvéri spôsobujú dramatické zmeny a zrýchlenia vo všetkých oblastiach počítania. Výpočtová zložitosť vo vede, technike a ekonomike rýchlo narastá.



**Dr. h.c. prof. Ing. Ivan Plander, DrSc., je jeden zo zakladateľov Ústavu technickej kybernetiky SAV (1956), autor prvého analógového počítača na Slovensku (SAV, 1957), autor a hlavný koordinátor projektu počítača RPP-16 a iniciátor výskumu v oblasti umelej inteligencie a robotiky v SAV.**



Zložitosť simulácií spracúvaných na superpočítačových systémoch vyžaduje vysokú rýchlosť spracovania, vysokú úroveň paralelizmu (systémy s tisícami a miliónmi procesorových elementov).

Mnohé vedecké a ekonomické aplikácie vyžadujú zrýchlenie operácií lineárnej algebry pre riešenie veľkých simulačných problémov, ako je napr. násobenie alebo inverzia matíc rozmeru 100 000 x 100 000 a pod.) .

V ostatných rokoch sa objavila nová trieda vysokovýkonných superpočítačových systémov. Tieto systémy používajú nekonvenčné procesorové architektúry – ako sú bunkové akcelerátory alebo grafické procesorové jednotky (GPU) – pre ťažké výpočty a konvenčné centrálné spracovateľské jed-

notky – pre zväčša nie výpočtovo intenzívne úlohy, ako sú vstupy/výstupy, komunikácie apod. Pre vedecké výpočty boli vyvinuté univerzálne grafické procesorové jednotky – GPGPU. Grafické procesorové jednotky GPU po mnoho rokov boli integrálnou časťou väčšiny osobných počítačov a hracích konzol.

Snaha po dosiahnutí realistickejších zobrazení hier vyžadovala vývoj GPU z jednoduchých akcelerátorov 2D určených pre aplikácie založené na grafike prechod na extrémne výkonné jednotky potrebné pre hry 3D. Výpočtový výkon moderných GPU viedol k explózii záujmu o ich využitie nie len v grafike, ale aj v numericky intenzívnom počítaní. Tento záujem je demonštrovaný výrobou univerzálnych koprocesorov GPU alebo GPGPU takými výrobcami ako NVIDIA (Tesla2011) alebo AMD (Radeon 2011) a pod. Akceptovanie procesorov GPGPU komunitou vysokovýkonného počítania vidieť z faktu, že štyri špičkové superpočítače z desiatich v zozname Top 500 z júna 2017 využívajú koprocesory GPGPU (Tabuľka vpravo).

Globálne možno povedať, že grafická procesorová jednotka GPU je akcelerátor, niekedy nazývaný koprocesor, navrhnutý pôvodne na vykonávanie špeciálnych grafických operácií rýchlejšie než štandardné CPU. Obsahuje jeden alebo viac mikročipov obsahujúcich obmedzený počet algoritmov vektorových alebo rastrových.

## TOP 10 NAJVÝKONNEJŠÍCH SUPERPOČÍTAČOV

PORADOVÉ ČÍSLO	UMIESTENIE	SYSTÉM	POČET JADIER	R <sup>MAX</sup> PETAFL0P/S	R <sup>PEAK</sup> PETAFL0P/S	PRÍKON MW
1	National Super Computing Centre in Wuxi ČÍNA	Sunway TaihuLight Sunway MPP	10 649 600	93,0	125,4	15,4
2	National Super Computer Centre in Guangzhou ČÍNA	Tianhe-2 (Milky Way-2) Intel Xeon E5	3 120 000	33,9	54,9	17,8
3	Swiss National Supercomputing Centre (CSCS) SWITZERLAND	PizDaint Cray XC50 NVIDIA P100 Cray Inc.	361 760	19,6	25,9	2,3
4	DOE/SC/Oak Ridge National Laboratory USA	Titan-Cray XK7 NVIDIA K20x Cray Inc.	560 840	17,6	27,1	8,2
5	DOE/NNSA/LLNL USA	Sequoia BlueGene/Q IBM	1 572 864	17,2	20,1	7,9
6	DOE/SC/LBLN/NERSC USA	Cori-CrayXC40 Intel Xeon Cray Inc.	622,3	14,0	27,9	3,9
7	Joint Center for Advanced HPC JAPAN	Oakfoest PACS-PRIMARY CX 1640 M1 Fujitsu	556,1	13,6	24,9	2,7
8	RIKEN Advanced Institute for Computational Science JAPAN	K computer SPARC64 VIIIfx Fujitsu	705,0	10,5	11,3	12,7
9	Argonne National Laboratory USA	Mira BlueGene/Q IBM	786,4 185 000	8,5 5,6	10	3,9
10	DOE/NNSA/SNL USA	Trinity Cray XC40 Cray Inc.	301,1	8,1	11,1	4,2

Zoznam špičkových superpočítačov Top 500 z júna 2017.  
(Ranking of Supercomputers according to the LINPACK benchmarks).



Grafické procesorové jednotky GPU boli vyvinuté v priebehu rokov 1970-tych a 1980-tych, avšak iba v rokoch 1990-tych sa začali brať vážne. V tom istom čase sa tieto akcelerátory objavili v hracích konzolách, kde umožňovali oveľa realistickejšie hry vyvinutím 3D GPU. Moderné grafické karty umožňujú prúdové spracovanie (zreťazenie lebo pipeline). Viacero stupňov zreťazenia dáva možnosť vytvárať tiež dátový paralelizmus. To znamená, že každý geometrický vrchol alebo obrazový element môže byť spracovaný nezávisle od ostatných, ale používať tie isté algoritmy, inými slovami používajú známy prístup SIMD. Dátový paralelný prístup SIMD sa potom vyvinul do zložitejšieho modelu SPMD.

Moderné GPU obsahujú desiatky až stovky procesorových jednotiek, pričom každá má ďalšie aritmeticko-logické jednotky ALU, umožňujúce efektívnejšie využitie SIMD-ovej architektúry. Skutočnosť, že moderné GPU obsahujú stovky ALU má za následok, že majú spracovateľský výkon rádo-vo Tflop/s, (napr. NVIDIA GPU GeForce GTX580M – NVIDIA 2011). Tento výkon prirodzene viedol k záujmu použiť GPU pre výpočtovo intenzívne problémy mimo tradičnej grafickej oblasti.

Mnohé vedecké aplikácie vyžadujú zrýchlenie lineárnych algebraických operácií. Napr. zrýchlenie násobenia matíc možno dosiahnuť použitím algoritmu štiepenia. Ak uvažujeme násobenie matíc

$$C = A \cdot B,$$

kde  $A$  je matica ( $m \times k$ ),  $B$  je matica ( $k \times n$ ). Maticu  $A$  možno rozdeliť na stĺpcové vektory tvaru ( $r+1$ ) matíc

$$A = \begin{pmatrix} A_0 \\ A_1 \\ \vdots \\ A_r \end{pmatrix}$$

kde každý vstup  $A_i$  je matica ( $p_i \times k$ )

$$A = \sum_i p_i = m.$$

V praxi, všetky  $p_i$  sú rovnaké.

Podobným spôsobom možno rozdeliť maticu  $B$  na riadkové vektory, ktorými sú matice ( $s+1$ ), čiže

$$B = (B_0, B_1, \dots, B_s),$$

kde každá  $B_j$  je matica ( $k \times q_j$ ) a

$$\sum_j q_j = n.$$

Tu opäť všetky  $q_j$  budú rovnaké. Potom vektorový súčin týchto dvoch vektorov možno napísať v tvare  $C = A \cdot B$

$$A = \begin{pmatrix} A_0 \\ A_1 \\ \vdots \\ A_r \end{pmatrix} \cdot (B_0, B_1, \dots, B_s) = \begin{pmatrix} A_0 B_0 & A_0 B_1 & \dots & A_0 B_s \\ A_1 B_0 & A_1 B_1 & \dots & A_1 B_s \\ \vdots & \vdots & \ddots & \vdots \\ A_r B_0 & A_r B_1 & \dots & A_r B_s \end{pmatrix}$$

Každý jednotlivý element  $C_{ij} = A_i \cdot B_j$  je matica ( $p_i \times q_j$ ) a môže sa vypočítať nezávisle. Zovšeobecnením tohto prístupu na plnú implementáciu pre superpočítač dostaneme maticové násobenie SIMD. Hodnoty  $p_i$  a  $q_j$  možno zvoliť tak, že každé pod-násobenie vyhovuje veľkosti pamäte superpočítača. Každé násobenie sa vykonáva na jednom procesore GPU a sčítanie na procesore CPU. To znamená, že každý uzol výpočtu využíva model SPMD. Na superpočítači s modelom SPMD analogicky možno vykonať všetky matematické operácie.

V Slovenskej Republike prostredníctvom projektu SIVVP (Slovenská Infraštruktúra pre Vysokovýkonné Počítanie) máme k dispozícii vysokovýkonný superpočítač IBM 775 AUREL s výpočtovým výkonom 130 Tflop/s, ktorý je umiestnený vo Výpočtovom stredisku SAV. Superpočítač AUREL už dnes vykonáva dôležité vedecko-technické výpočty v oblasti kvantovo-chemických štúdií, seizmologických vlnových polí, simulácie priebehu požiarov a šírenia dymu, teórie elektrónovej štruktúry látok a rad ďalších.

Programovanie a využívanie tohto systému vyžaduje odborníkov schopných programovať a zabezpečovať realizáciu zložitých numerických výpočtov. Vyškoliť ich možno na pracoviskách vysokých škôl a SAV, za predpokladu, že budú mať voľný prístup k tomuto superpočítačovému prostrediu. Zo skúsenosti je známe, že len dostatočný počet progra-



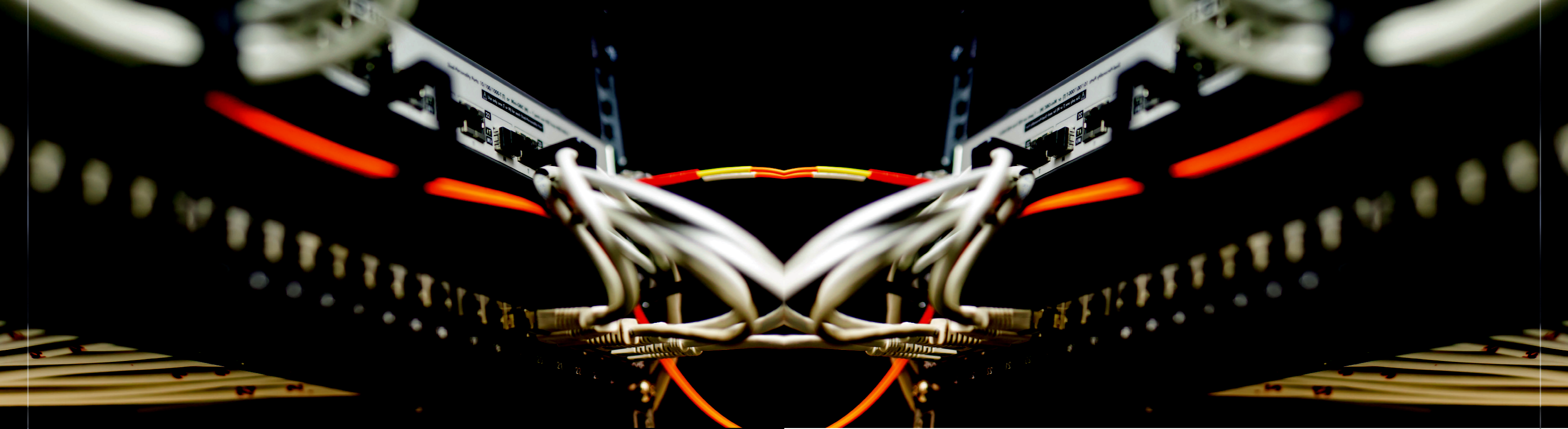
mátorov a pracovníkov z aplikačnej sféry môže zabezpečiť efektívne využívanie superpočítačového prostredku. Za tým účelom by bolo vhodné na Slovensku vybudovať inštitúciu, najlepšie Národné superpočítačové centrum (NSC), ktoré by svojim technickým a programovým vybavením a personálnym obsadením bolo schopné zabezpečiť realizáciu superpočítačových numerických výpočtov na vlastnom výpočtovom prostredí a bolo by celoštátnym garantom realizácie rozvoja vysokovýkonného počítania pre vedu a techniku. Takéto superpočítačové centrá má už viacero členských štátov EÚ. Najznámejšie je EPCC (Edinburgh Parallel Computing Centre) v Edinburgu. Superpočítačové centrá poskytujú bezplatne služby pre univerzity a ich výskum, vývoj a školenie, čím si vytvárajú potenciál používateľov a vlastných programátorov, pretože ich finančné zdroje by neboli postačujúce na úhradu potrebného strojového času na superpočítači. V zmysle projektu SIVVP na Národné Superpočítačové Centrum by boli napojené odborné centrá pre vedecko-technické a ekonomické počítanie na univerzitách a SAV. Sú to: Doprava v Žiline; hutníctvo, fyzika a veterinárstvo v Košiciach; astronómia v Starej Lesnej; medicína, elektrotechnika, nukleárna fyzika v Bratislave. Uvedené centrá by pritom zabezpečovali ďalšie funkčné oblasti: Výskum a vývoj metód vysokovýkonného počítania, školenie využívania superpočítačov na Slovensku, nepretržitú technickú prevádzku a systémovú obsluhu superpočítačov, dátovú bezpečnosť a medzinárodnú spoluprácu s národnými centrami ostatných členských štátov EÚ.

Vytvorenie slovenského superpočítačového centra by zaradilo Slovenskú republiku do širokej rodiny používateľov superpočítačového počítania, v ktorej sa už napr. Slovinsko, Česká republika a Poľsko dávnejšie nachádzajú a tým by sa zrovnoprávnila slovenská vedecko-výskumná a akademická obec s jej zahraničnými partnermi. Vyrovnanie sa Slovenska okolitým štátom v inštalovanej vysokovýkonnej výpočtovej technike by zabezpečilo rovnaké možnosti pre slovenský vedecký výskum, držať krok s výskumom v susedných štátoch a celej Európe.

**DR. H.C. PROF. ING. IVAN PLANDER, DRSC.,  
JE ZAKLADATEĽ A PRVÝ REKTOR  
TRENČIANSKEJ UNIVERZITY ALEXANDRA DUBČEKA (1997)  
A V SÚČASNOSTI PROFESOR A VEDEC  
V OBLASTI APLIKOVANEJ INFORMATIKY  
A PARALELNÝCH POČÍTAČOVÝCH SYSTÉMOV.**

02

# APLIKÁCIE W/P



Vysokovýkonné  
počítanie  
na katedre matematiky  
a  
deskriptívnej geometrie  
na  
**Svf STU v Bratislave**

PROF. RNDR. KAROL MIKULA, DRSC.  
ING. ROBERT ČUNDERLÍK, PHD.  
ING. MAREK MACÁK, PHD.  
ING. ZUZANA MINARECHOVÁ, PHD.  
ING. ROBERT ŠPIR, PHD.

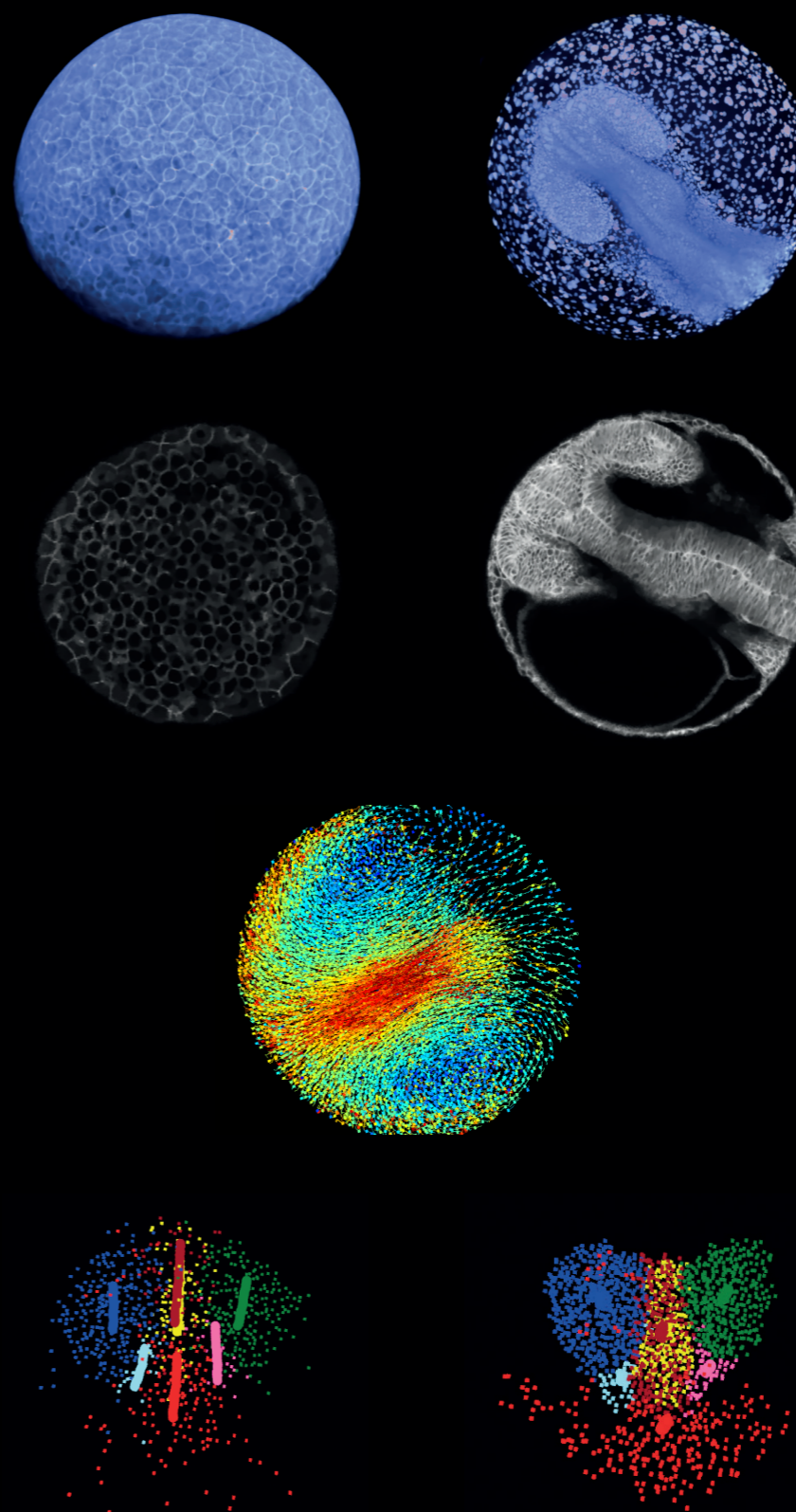
Katedra matematiky a deskriptívnej geometrie  
Stavebná fakulta STU v Bratislave  
Radlinského 11  
810 05 Bratislava

**N**a našej katedre sa formoval tím pre HPC výpočty už dlhé roky pod vedením **prof. RNDr. Karola Mikulu, PhD.** a vďaka jeho kontaktom a usilovnej práci sa zaradil medzi svetovú elitu v úlohách, ktoré rieši.

## 1

**Spracovanie rozsiahleho množstva biomedicínskych dát**

Pri našich aplikáciách HPC výpočtov na klastri sa venujeme spracovaniu dlhých sérií 3D biomedicínskych dát. Tieto dáta získavame od francúzskych biológov z **CNRS - CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE**, s ktorými na tomto projekte spolupracujeme. Jedná sa o trojrozmerné mikroskopické skeny skorého štádia bunkového vývoja rôznych organizmov, ako napríklad ryбка danio pruhované (zebrička), morský ježko, alebo zajac. Tieto snímky zachytávajú veľmi skoré štádia vývoja, približne od 2 hodín po oplodnení vajíčka až po 15 hodín vývoja v približne minútových intervaloch. Takto nám vznikne niečo ako 4D video. Naše algoritmy, ktorými spracúvame tieto dáta, sa týkajú najmä oblasti spracovania obrazu. Vstupné dáta najprv filtrujeme na odstránenie šumu metódou geodesic mean curvature flow, a potom z nich získavame súradnice bunkových jadri metódou level-set center detection. Obe tieto metódy pracujú na samostatných 3D dátach, avšak už tu používame paralelné výpočty na zrýchlenie. Naša výpočtovo najnáročnejšia aplikácia je automatické sledovanie a rekonštrukcia pohybu buniek v celom dátovom súbore. Tu si zo série 3D dát vytvoríme 4D súbor, v ktorom následne na základe získaných súradníc bunkových jadri konštruujeme 4D segmentáciu, predstavujúcu pohyb buniek v čase a priestore. Následným výpočtom funkcie vzdialenosti od počiatočných buniek vo vnútri tejto segmentácie získame takzvané potenciálové pole, ktoré nám dokáže aproximovať pohyby jednotlivých buniek v dátovom súbore a môžeme z neho získať trajektórie jednotlivých buniek. Všetky tieto výpočty sú paralelizované pomocou rôznych rozhraní, ako napríklad OpenMP alebo MPI, čím nám umožňujú optimálnu distribúciu výpočtov na viacero výpočtových uzlov klastra, lepšie využitie väčšieho množstva operačnej pamäte a zrýchlenie výpočtov z niekoľko dní na niekoľko hodín. **Naše výsledky** dosiahnuté v spolupráci s francúzskymi kolegami boli v roku 2016 publikované v prestížnom časopise *Nature Communications* zo skupiny časopisov *Nature*.

**OBRÁZOK 1:**

Na prvých štyroch obrázkoch sú zobrazené mikroskopické snímky raného bunkového vývoja ryбки zebričky. Na týchto snímkach máme zvlášť zachytené bunkové jadrá a bunkové membrány. Na piatom obrázku sú zobrazené zafarbené bunkové jadrá, kde farba zodpovedá ich rýchlosti (červené rýchle a modré pomalé bunky). Na obrázkoch vo štvrtom rade je zobrazená vizualizácia formovania jednotlivých, farebne odlíšených orgánov živočicha aj so zobrazením stredných trajektórií jednotlivých bunkových populácií.



**Ing. Marek Macák,  
PhD. pôsobí ako  
vysokoškolský  
pedagóg na  
Stavebnej fakulte  
Slovenskej  
technickej  
univerzity  
v Bratislave.**

## 2

### Riešenie geodetickej okrajovej úlohy

Medzi základné úlohy geodézie patrí určovanie veľkosti Zeme, jej tvaru a tiažového poľa. Tiažové modely Zeme sa dajú vypočítať rôznymi spôsobmi a na našej katedre sa venujeme ich určovaniu riešením parciálnych diferenciálnych rovníc numerickými metódami ako sú metóda hraničných elementov, metóda konečných prvkov a metóda konečných objemov.

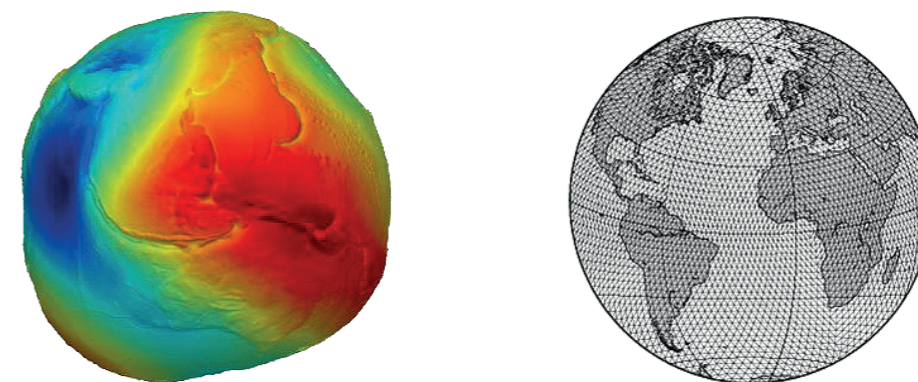
Kedže hranicou našej výpočtovej oblasti je zemský povrch, ktorý je členitý, na jeho aproximáciu používame trojuholníky. Aby sme dosiahli čo najpodrobnejšie výsledky, je potrebné vytvoriť jemnú sieť s rozlíšením napr. jedna minúta (približne 110 km na rovníku). Takéto delenie potom zodpovedá 21600x10800 neznámym hodnotám na zemskom povrchu. Pri našich výpočtoch používame ako vstupné údaje dáta generované zo satelitov a pozemných meraní.

Jednou z aplikácií nášho výskumu je určovanie konštanty  $W_0$  – referenčnej hodnoty tiažového potenciálu, ktorá je základom pri definícii a tvorbe svetového výškového systému. Tento problém súvisí s tým, že morská voda jednotlivých morí a oceánov má v rôznych oblastiach rôznu teplotu, slanosť a obsah iných látok. To má za následok, že jednotlivé lokálne výškové systémy odvodené od morskej hladiny sú medzi sebou vzájomne posunuté. Napr. v Európe máme dodnes platných viac než 12 výškových systémov a rozdiely medzi nimi dosahujú až decimetre a rovnaké problémy sú i na iných kontinentoch či ostrovoch. Napríklad Nový Zéland ako ostrovná krajina má až 13 lokálnych výškových systémov, ktoré sú medzi sebou niekoľko centimetrov posunuté. Aby sa zjednotili lokálne výškové systémy a zadefinoval sa svetový výškový systém, je potrebné zvoliť si jednu základnú referenčnú plochu. Najvhodnejšou plochou je taká, ktorá prechádza stredom ustálenej hladiny mora. Takáto plocha je potom definovaná spomínanou konštantnou hodnotou tiažového potenciálu  $W_0$  a nazýva sa geoid.

V roku 2011 sa Medzinárodná geodetická asociácia (IAG) rozhodla zostaviť tím vedcov, ktorí sa pod vedením Dr. Laury Sánchezovej z Technickej univerzity v Mníchove podieľal na výpočte novej hodnoty konštanty  $W_0$ .

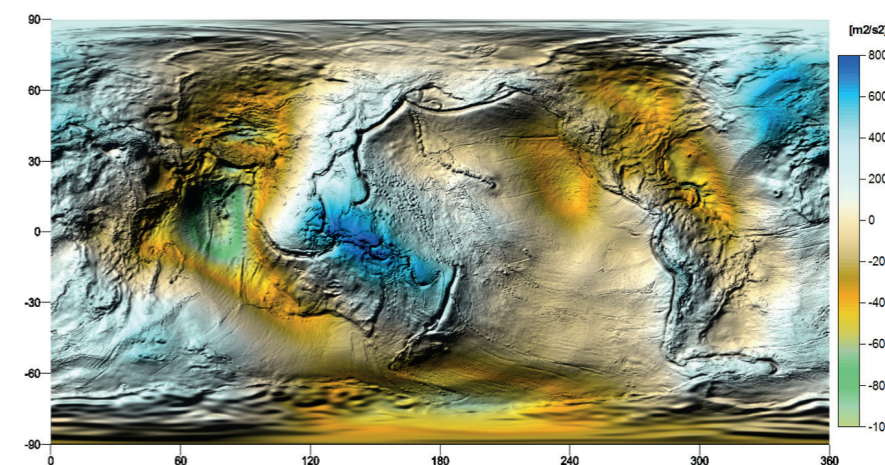
Jednou zo štyroch skupín vedeckého tímu sa stala aj naša pracovná skupina.

Nová hodnota konštanty  $W_0 = 62\,636\,853,4 \text{ m}^2/\text{s}^2$  sa oficiálne prijala na 26. valnom zhromaždení Svetovej únie geodetov a geofyzikov v Prahe v lete 2015.



**OBRÁZOK 2:**

Vľavo: Zobrazenie fyzikálneho tvaru zeme (Geoidu).  
Vpravo: Triangulácia - ilustrácia výpočtovej siete.



**OBRÁZOK 3:**

Poruchový potenciál je priamym výsledkom našich výpočtov a je definovaný ako rozdiel skutočného potenciálu a normálneho tiažového potenciálu. V porovnaní so skutočným potenciálom Zeme je tým väčší, čím väčšie sú jeho odchylky od normálneho telesa, ideálnej gule. Poruchový potenciál je fyzikálna veličina, pomocou ktorej sa určujú geometrické charakteristiky anomálneho, a tým aj skutočného tiažového poľa, t. j. výšku geoidu, alebo výškovú anomáliu.



### 3

## Výpočtový klaster

Za účelom zvládnutia i takýchto výpočtov bol na katedre budovaný paralelný počítačový klaster od roku 2005, ktorý beží pod operačným systémom GNU/Linux. Klaster je konfigurovaný ako 8 GB RAM/jadro, čo je najoptimálnejšia konfigurácia pre naše výpočty, s prístupom na 40 TB zdieľaného úložného priestoru. Rýchla komunikácia v rámci klastra je zabezpečená ethernetovou sieťou s rýchlosťou 1 Gbit/s. V súčasnosti má klaster 160 jadier a je navrhnutý tak, aby sa v budúcnosti dal ľahko rozšíriť (ak na to budú prostriedky).

Vybudovaním Aurela sme získali nový nástroj na výpočty, ktoré u nás na katedre klaster nezvládal. I za cenu dlhšieho času vo fronte sme sa dopracovali k výsledkom.

Vybudovaním Aurela sme získali nový nástroj na výpočty, ktoré u nás na katedre klaster nezvládal. I za cenu dlhšieho času vo fronte sme sa dopracovali k výsledkom. Aurel sa pre nás stal účinným nástrojom, ktorý vďaka svojej veľkosti posunul naše výpočty o krok ďalej.

# Magdaléna Májeková Yoel Rodriguez

Možnosti HPC pre membránové proteíny  
– príklad  $\text{Ca}^{2+}$ -ATP-ázy  
sarkoplazmatického retikula

MAGDALÉNA MÁJEKOVÁ

Ústav experimentálnej farmakológie a toxikológie SAV, Dúbravská cesta 9, 841 04 Bratislava, SR

YOEL RODRIGUEZ

Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, USA; Department of Natural Sciences, Eugenio María de Hostos Community College, 500 Grand Concourse, Bronx, New York 10451, USA

**P**ri výskume fyziologických a patologických procesov v organizme je často dôležité čo najpresnejšie modelovať ich dôležitú zložku – proteíny. Dĺžka proteínového reťazca v eukaryotických bunkách (a teda aj u človeka) býva v priemere niekoľko sto aminokyselín, aj keď môže dosahovať aj hodnoty viac ako 20 000 (Brocchieri et al., 2005). Čo sa týka okolitého prostredia, proteíny možno rozdeliť zhruba do dvoch skupín: cytoplazmatické (nachádzajúce sa viac-menej vo vodnom prostredí) a membránové (zakotvené v biologickej membráne). Práve membránové proteíny predstavujú aktuálnu výzvu pre počítačové modelovanie.

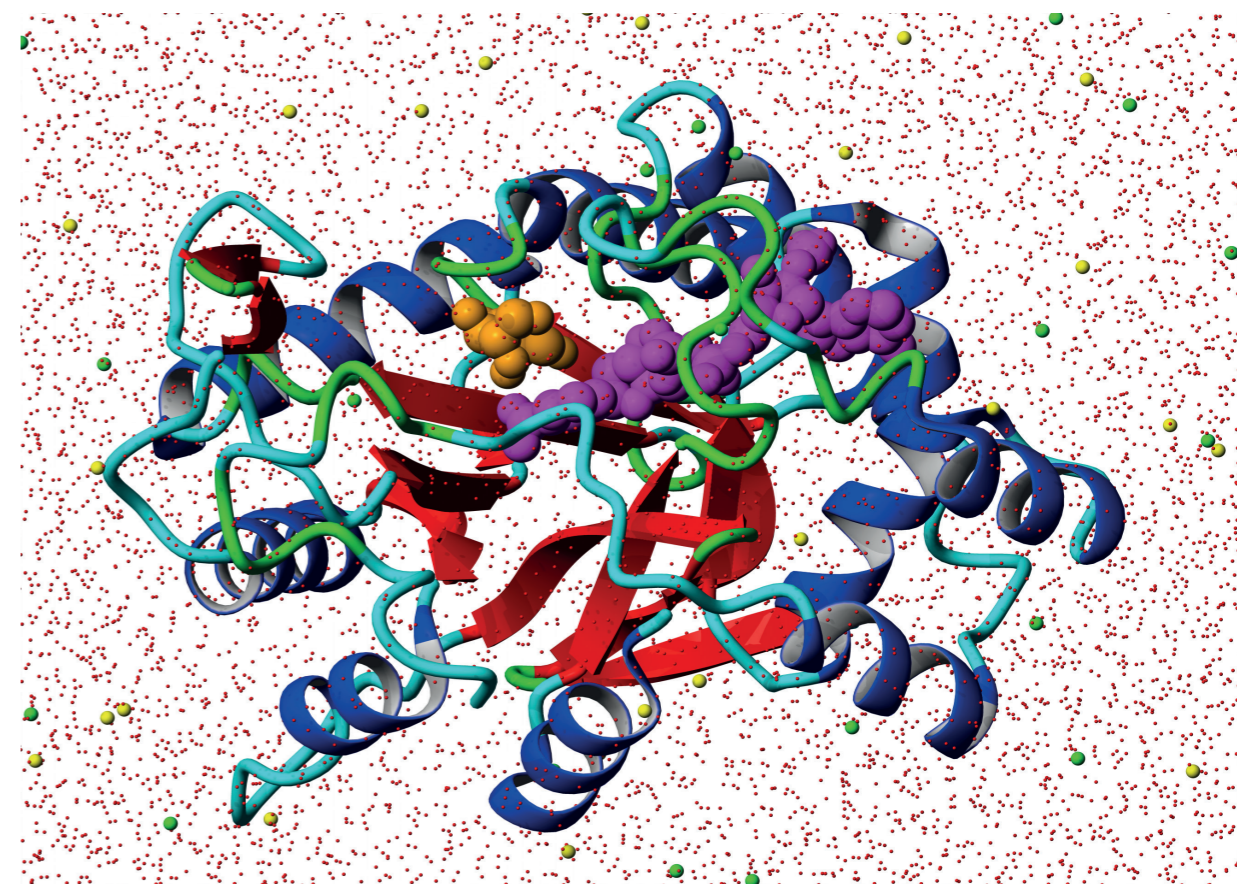
Prvým dôvodom je ich veľkosť. Samotné proteíny inkorporované do membrán bývajú spravidla omnoho väčšie, než proteíny cytoplazmatické, keďže objemom musia dosiahnuť tri oddelené fázy. Kým na počítačový experiment s proteínom v cytoplazmatickom prostredí stačí okolitá voda so základnými iónmi a niekoľkými v modeli zahrnutými malými molekulami (na Obr. 1 je ako príklad znázornený enzým aldózareduktáza s kofaktorom a substrátom, umiestnené vo vodnom prostredí s iónmi), pri počítačovej simulácii membránového

proteínu je potrebné okrem vodného prostredia zahrnúť aj fyzikálne reálne chovanie lipidickej dvojvrstvy, ktorá tvorí základ biologickej membrány. Základné stavebné jednotky membrány sú omnoho väčšie a zložitejšie ako molekula vody, naviac je každá membrána zložená z niekoľkých druhov lipidov. Z toho vyplýva ďalší bod obťažnosti pre modelové výpočty, ktorým je komplexita, čiže zložitost' modelovaného systému.

Ďalší problém predstavuje reálny časový úsek, ktorý by mal byť pokrytý počítačovým experimentom. Priemerný relaxačný čas vody v reak-

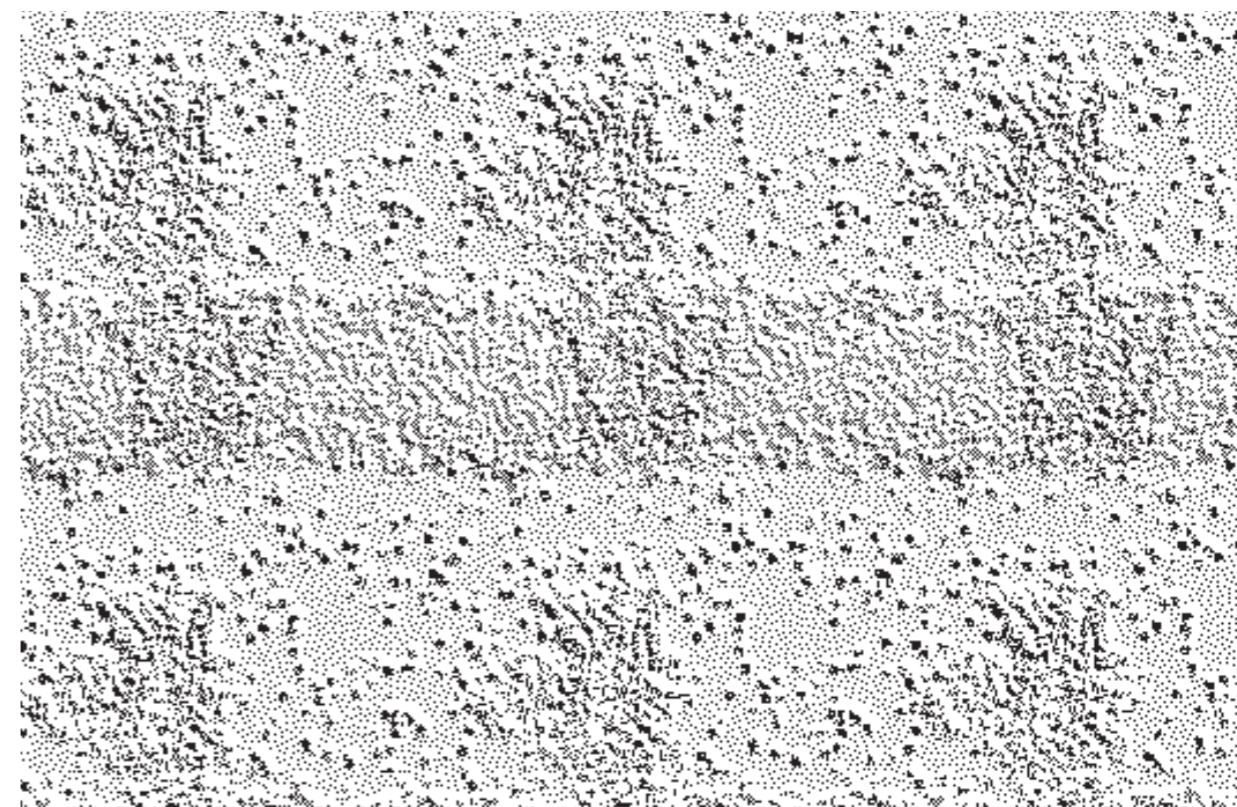
Obrázok 1:

Enzým aldózareduktáza s kofaktorom NADPH a substrátom.



Obrázok 2:

Periodický obraz proteínu SERCA v membráne.





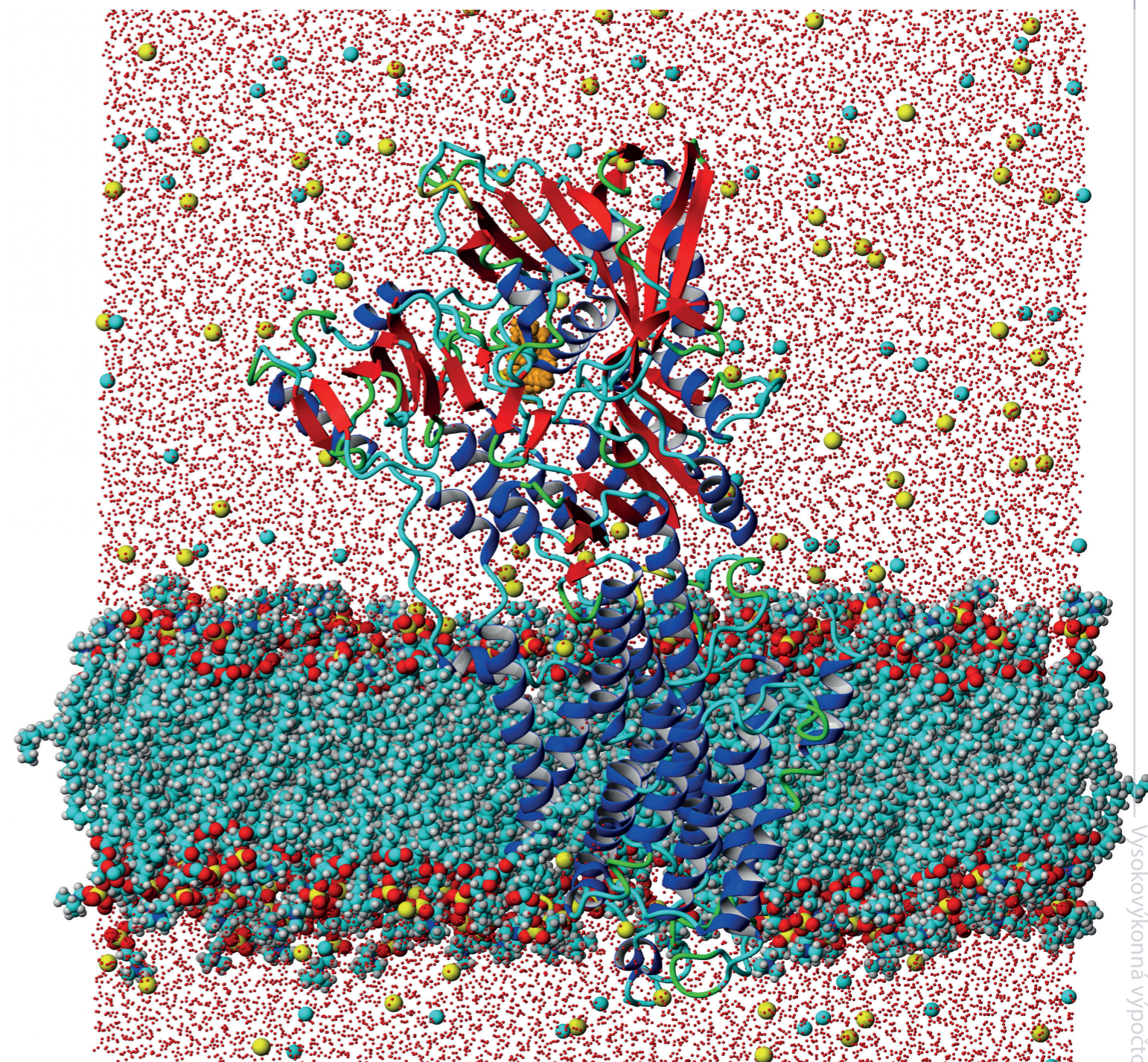


**RNDr. Magdaléna Májeková, PhD.**  
pôsobí ako vedúci  
vedecký pracovník  
v Ústave  
experimentálnej  
farmakológie  
a toxikológie  
SAV v Bratislave.

cii na rozpúšťanú molekulu s hydrofóbnym povrchom (aký majú čiastočne aj proteíny) sa pohybuje v rádoch ps (pozri napr. Comez et al. 2014, Muntean et al. 2012). Relaxačné časy súvisiace so štrukturálnymi zmenami proteínov sa pohybujú rádovo od nanosekundy až po mikrosekundy, v závislosti od veľkosti a typu konformačnej zmeny. Najdlhší čas je potrebný na zmeny týkajúce sa hlavného reťazca proteínu (Shaw et al. 2010; Khodadadi et al. 2015). V prípade zložitých proteínových komplexov, ako sú napr. receptory (napr. ryanodínový, glutamátový), sa navyše môže jednať o viacero konformačných zmien uskutočnených za sebou a tomu odpovedajúco rastú aj nároky na čas simulovaného experimentu.

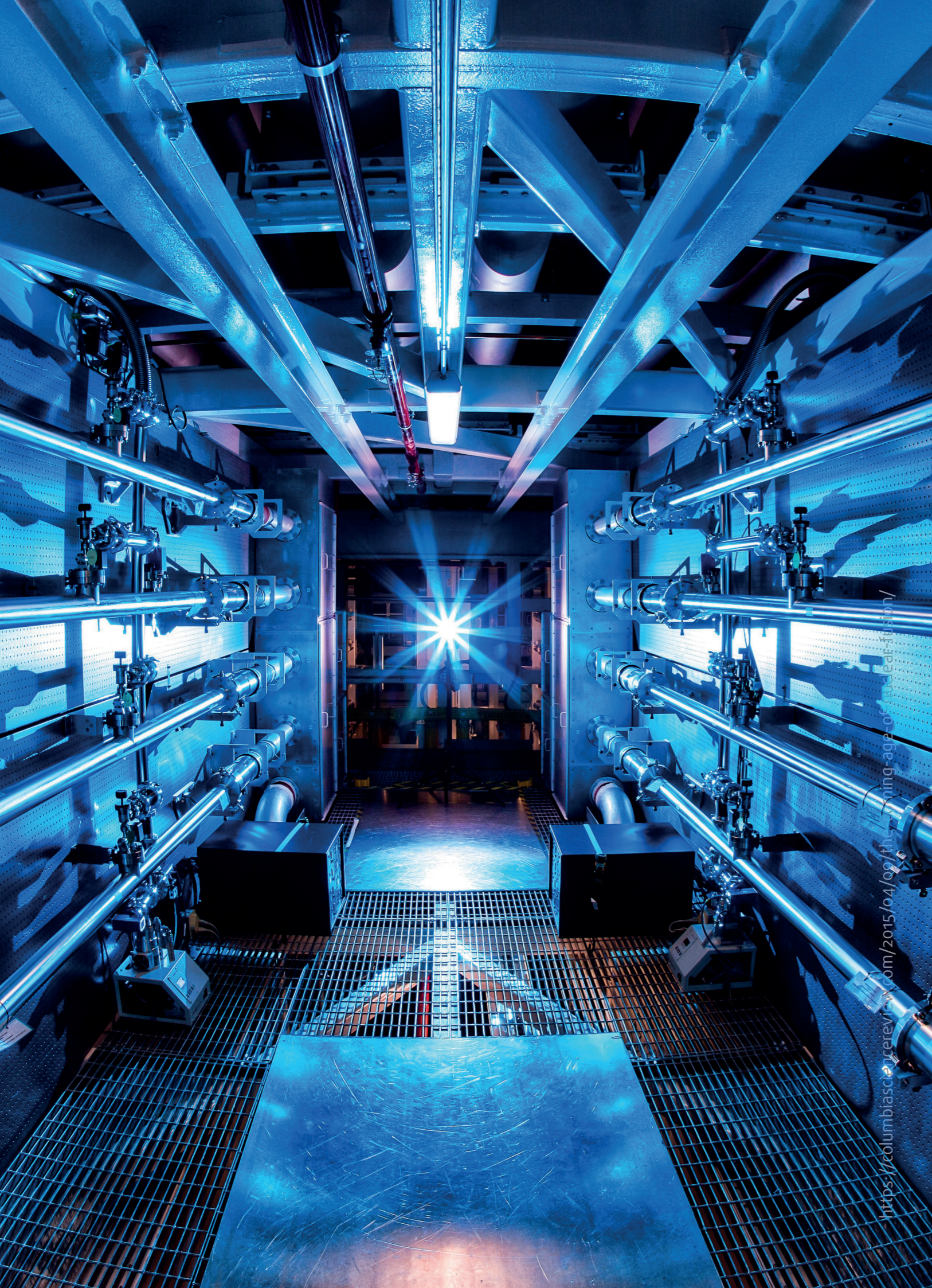
Kolegyne z oddelenia biochemickej farmakológie UEFT SAV L. Horáková, J. Lomenová a P. Rezbáriková sa už dlhší čas venujú experimentálnemu štúdiu štruktúry a funkcie SERCA - vápnikovej pumpy sarkoplazmatického retikula (SR). Pred niekoľkými rokmi sa tento smer rozšíril o teoretické štúdium izolovaného proteínu (M. Májeková) a v predchádzajúcom roku sa štúdium posunulo do roviny realistického modelu s membránou ponorenou do vodného prostredia s iónmi (Y. Rodriguez). Fyziologická úloha proteínu SERCA spočíva v pumpovaní dvojmocných iónov vápnika z cytoplazmy (vnútorného prostredia bunky) do lumenu SR (vnútorného prostredia SR), čo pomáha udržiavať homeostázu vápnika v bunke. Koncentrácia vápnika hrá dôležitú úlohu pri mnohých biologických procesoch, ako napr. svalová kontrakcia, génová expresia, sekrécia inzulínu a pod. Jej narušenie môže napomáhať vzniku závažných chronických ochorení, ako sú kardiovaskulárne, neurodegeneratívne alebo svalové ochorenia, chronický zápal, rakovina a cukrovka.

Pomocou experimentov boli objavené látky, ktoré dokážu urýchľovať, spomalovať, prípadne celkom zastaviť činnosť vápnikovej pumpy. Porozumenie mechanizmu týchto javov nám môže pomôcť navrhovať účinné látky schopné ovplyvniť činnosť SERCA a tým aj spomaliť alebo zamedziť rozvoj chorobných procesov. Prof. Rodriguez, ktorý sa k nám dostal vďaka štipendiu udelenému Fulbrightovou nadáciou, vypracoval model SERCA pre počítačovú simuláciu pomocou metódy molekulovej dynamiky (Obr. 3). Pomocou tohto modelu chceme zistiť kľúčové mechanizmy inhibičného a aktivačného pôsobenia rôznych ligandov – malých molekúl. Tie nám potom pomôžu robiť cieľový dizajn nových modulátorov.



**Obrázok 3:**

Prierez modelu SERCA1 zabudovaného do biologickej membrány a obklopeného vodným prostredím s rozpustenými iónmi  $\text{Na}^+$  a  $\text{Cl}^-$ . V cytozolickej časti proteínu je slabými medzimolekulovými interakciami naviazaný rutín (oranžová farba). Stav po 100 ns simulácii.



# Michal Novotný

Materiály na báze molybdénu s využitím v jadrovej fúzii.

Rozvoj priemyslu a vývoj nových technológií sa v posledných dvoch storočiach neustále urýchlil. Čas od objavu nového fenoménu, až po jeho praktické využitie sa v mnohých prípadoch skrátil z desaťročí na niekoľko rokov. Jeden z prírodných fenoménov s obrovským potenciálom pre ľudstvo však stále zostáva nevyužitý. Je ním jadrová fúzia. Od jej teoretických počiatkov v tridsiatych rokoch až dodnes, si praktické prevedenie tejto technológie vyžaduje výskum v rôznych odvetviach. Počnúc jadrovou fyzikou a výskumom samotnej fúznej reakcie, cez fyziku plazmy a dizajn reaktora, po rozsiahly vývoj nových materiálov využiteľných vo fúzných zariadeniach. Limitujúcim faktorom pre väčšinu zo spomenutých oblastí výskumu fúzných reaktorov sú materiály, ktoré sú momentálne k dispozícii. Mnohé z tých ktoré sa bežne používajú majú buď nevhodné vlastnosti alebo je ich výroba príliš nákladná pre ekonomické priemyselné využitie. Výskum a vývoj nových materiálov konvenčným spôsobom dokáže byť veľmi zdĺhavý a finančne náročný nakoľko je nutné nový materiál najskôr synteticky pripraviť a následne otestovať, pričom nemusí mať predpokladané vlastnosti. V tejto fáze je obzvlášť sľubné použiť metódy teoretickej a počítačovej chémie a materiálového dizajnu.

Čo bolo pred dvadsiatimi rokmi nepredstaviteľné sa dnes stáva bežnou praxou. Vďaka rozvoju HPC a s tým spojenými metódami počítačovej chémie, vieme dnes simulovať realistické systémy (desiatky až stovky atómov) a tak za krátky čas a minimálne finančné prostriedky dodať spoľahlivý návrh materiálov ktoré majú požadované vlastnosti. Počítačová chémia je neodvratne závislá a limitovaná HPC infraštruktúrou, ktorú má k dispozícii, keďže jej potenciál ju využívať je neobmedzený.

Jadrová fúzia prebieha pri teplotách  $15 \cdot 10^7$  K. Na dosiahnutie takto vysokej teploty sa využíva kombinácia troch mechanizmov. Prvý je ohrev plazmy pomocou elektromagnetických



**Mgr. Michal Novotný** pôsobí ako PhD študent na Katedre fyzikálnej a teoretickej chémie na Prírodovedeckej univerzite UK v Bratislave.

vín s frekvenciami 40 až 55 MHz (Ion Cyklotron Resonance Heating, ICRH), druhý využíva frekvencie okolo 170 GHz (Electron Cyclotron Resonance Heating ECRH). Tretí spôsob ohrevu spočíva v generovaní záporných iónov deutéria, ich urýchlení na 1 MeV a ich následnom

alebo inkorporovaného do vnútra povrchu. Výstupná práca je energia, ktorú je potrebné dodať materiálu, aby sa z neho vyrazil elektrón. Hodnota výstupnej práce určuje efektivitu, s akou bude dochádzať ku konverzii atómov vodíka na vodíkové ióny a tak určuje aj efektivitu

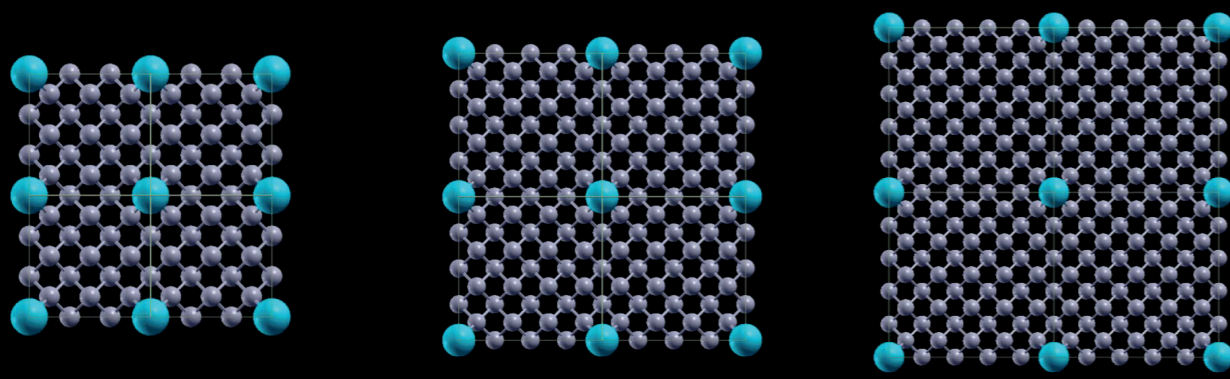
produkuje experiment v počítačových simuláciách.

Aby boli naše modeli čo najbližšie ku skutočným systémom, je nutné v našich simuláciách zahrnúť čo najväčší počet častíc. Tu sme však limitovaný hardvérom, ktorý máme k dispozícii. V tuhých

počty je možné paralelizovať pre jednotlivé elektrónové pásy, ako aj pre dané k-body.

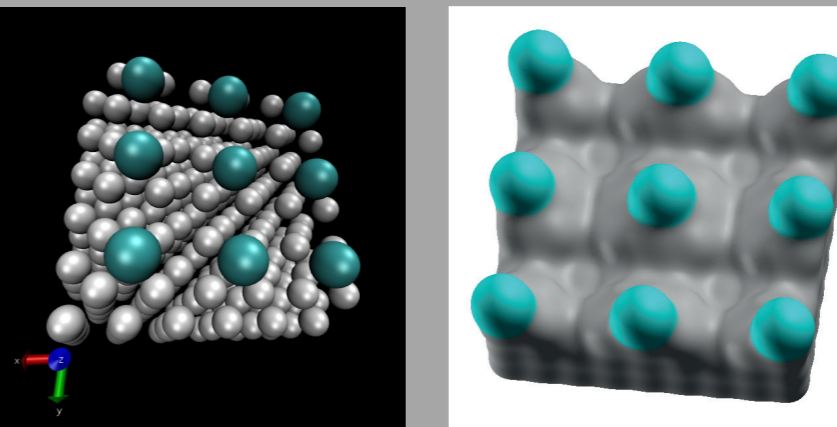
Naše simulácie prebiehali v troch etapách. Prvá zahŕňala štrukturálnu relaxáciu základných buniek povrchov. V týchto výpočtoch systém obsahoval rádovo 55 atómov, 15 k-bodov a 464 pásov.

etapa boli výpočty samotnej výstupnej práce, ktoré si vyžadovali iba relaxáciu elektrónovej štruktúry. Pre horeuvedené povrchy bol výpočtový čas na 4 nódoch približne 24 hodín. Celkovo sme takto analyzovali 20 rôznych povrchov.



**Obr. 1:**

Rôzne pokrytia povrchu molybdénu céziom.



**Obr. 2:**

Povrch molybdénu pokrytý céziom.

injektovaní do plazmy (Beam Injector, NBI). Spoločne poskytujú tieto zdroje výkon 50 MW ktorý je potrebný na „zapálenie“ plazmy. V našej skupine sme sa zamerali na výskum materiálov na generovanie záporne nabitých iónov deutéria do NBI. Ide o materiály pozostávajúce z molybdénu, wolfrámu alebo európie, na ktorých povrchu je naviazané malé množstvo cézia. V tejto fáze nášho výskumu nás zaujíma zmena vo výstupnej práci elektrónov spôsobená prítomnosťou cézia na povrchu

NBI systému ohrevu plazmy. Z experimentálneho výskumu uskutočneného na wolfráme pokrytom céziom je známe, že pri istom pokrytí povrchu nastáva minimum výstupnej práce, ktoré je ešte nižšie ako výstupná práca cézia (výstupná práca wolfrámu je 4.55 eV, cézia 2.14 eV a v minime 1.82 eV). Dôvod, prečo k tomu dochádza nie je z experimentu zjavný. Aby sme pochopili tento fenomén a vedeli ho následne využiť pri iných materiáloch, naša skupina čiastočne re-

látkach sa elektróny usporiadajú do takzvaných elektrónových pásov veľmi podobne ako sa usporiadajú do orbitálov v atóme. Na rozdiel od molekulových výpočtov, kde sa rieši Schrödingerova rovnica pre jednotlivé elektróny a ich orbitály, pri tuhých látkach sa rieši pre elektrónové pásy ktoré sa následne naplňajú (populujú) elektrónmi na základe ich tepelnej energetickej distribúcie. Ďalšou osobitosťou simulácií tuhej fázy je numerické vzorkovanie priestoru pomocou k-bodov. Naše vý-

Tieto výpočty prebiehali na superpočítači Aurel na 8 nódoch (256 jadrách) a trvali približne 79 hodín celkového času („wall clock“). V druhej etape sme zo zrelaxovaných štruktúr zostavovali povrchy s požadovaným pokrytím cézia a nechali znovu celú štruktúru zrelaxovať. Tieto simulácie obsahovali rádovo 200 atómov 9 k-bodov a 1700 pásov. Keďže jednotlivé segmenty povrchu boli predrelaxované, takáto simulácia na 8 nódoch trvala približne 50 hodín. Finálna

Je zrejme, že aj pre takýto, relatívne menší projekt je potrebná veľká výpočtová kapacita, ktorú je pre väčšinu pracovísk nemožné či už získať alebo prevádzkovať. Superpočítač Aurel, výpočtový klaster v Žiline a iné prostriedky získané z projektov SI-VVP predstavujú pre nás esenciálny nástroj na to, aby sme vedeli naďalej produkovať kvalitné vedecké výsledky na svetovej úrovni.

03

# PRACE



# PRACE

## SUMMER OF HPC

na

## VÝPOČTOVOM STREDISKU SAV

Michal Pitoňák

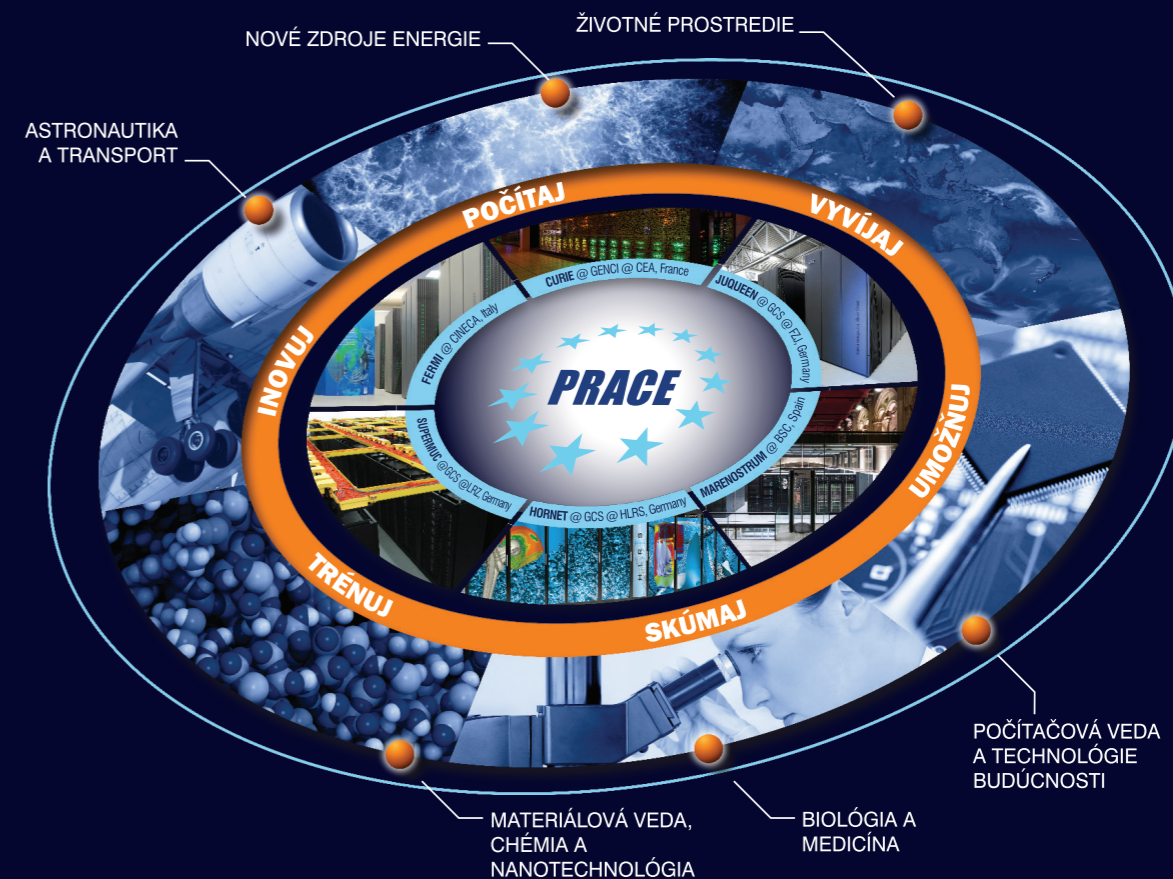
# H2020

**program PRACE**  
(Partnership for Advanced

Computing in Europe), ktorého je Výpočtové stredisko SAV od roku 2015 partnerom, organizuje od svojho počiatku medzinárodné dvojmesačné stáže s názvom Summer of HPC (High-performance Computing), <http://summerofhpc.prace-ri.eu>. V rámci tohto programu je každoročne, v čase letných prázdnin, rozdelených (okolo) dvadsať vybraných študentov medzi superpočítačové pracoviská - partnerské organizácie PRACE, kde sa podieľajú na výskumných projektoch súvisiacich s HPC.

Študenti v žiadosti o účasť predkladajú svoje životopisy, demonštrujú motiváciu zúčastniť sa na programe a svoje programátorské zručnosti. Uvádzajú taktiež tri preferenčné projekty zo zoznamu dostupných projektov pre daný ročník, na ktorých by chceli pracovať. Vybraní uchádzači získavajú štipen-

dium, ktoré bezvýhradne pokrýva všetky súvisiace náklady počas stáže. Program Summer of HPC začína týždňovým kurzom paralelného programovania, práce v HPC prostredí, vizualizácie (veľkých) dát, pod., ktorý sa koná na jednej z hostujúcich inštitúcií. Po úvodnom týždni odchádzajú študenti na cieľové pracoviská, kde sa už venujú prideleným projektom. Na konci stáže študenti vypracujú záverečnú správu, v ktorej v krátkosti sumarizujú výstupy ich projektov, a pripravujú krátke, päťminútové videá prezentujúce projekty širokej verejnosti



(<https://youtube.com>, hľadať napr. „PRACE Summer of HPC 2017 presentations“). Počas leta sú účastníci motivovaní dokumentovať a zdieľať priebeh riešenia projektu na oficiálnom blogu programu Summer of HPC a sociálnych sieťach tak, aby upúťali záujem o HPC čo najväčšieho počtu ich rovesníkov. Odmenou za ich popularizačné aktivity je finančne honorovaná cena „Best HPC ambassador award“, ktorá je spolu s „Best visualization award“ každoročne udelená víťazom v príslušnej kategórii.

Zo Slovenska sa doposiaľ Summer of HPC zúčastnil iba jeden študent, Mgr. Ján Hreha, PhD študent fyziky na Fakulte matematiky, fyziky a informatiky Univerzity Komenského v Bratislave, ktorý strávil leto 2015 na inštitúte NIIF

(National Information Infrastructure Development Programme) v Budapešti prácou na projekte „Bringing Hybrid Architecture’s power for Atomistic Simulations“.

Výpočtové stredisko SAV sa od počiatku svojho členstva v PRACE podieľa na organizácii Summer of HPC a hostí každoročne dvoch študentov. V roku 2016 to bol Oisín Benson z Írska a Katerina Galata z Grécka. Oisín pracoval pod vedením doc. Mgr. Michala Pitoňáka, Phd. na téme „Quantum Chemistry in Spark“, Katerina pod vedením prof. RNDr. Jozefa Nogu, DrSc., na téme „Calculating Nanotubes“. V tomto roku, 2017, to boli študenti Adrián Rodríguez-Bazaga zo Španielska, pracujúci pod vedením doc. Pitoňáka na téme „Apache Spark: Are Big Data tools applicable in HPC?“, a Andreas Neophytou z Grécka, pracujúci na projekte „Calculation of nanotubes by utilizing the helical symmetry properties“ pod vedením prof. Nogu. Rozhovory s minulými účastníkmi Summer of HPC na Výpočtovom stredisku SAV sú publikované samostatne v tomto čísle HPC Focus.

Začiatkom roka 2018 bude na internetových stránkach PRACE (<http://www.prace-ri.eu>) opäť zverejnená výzva na účasť na projekte Summer of HPC 2018. Chceli by sme vyzvať všetkých potenciálnych záujemcov, či už z radov študentov vyšších ročníkov vysokých škôl alebo doktorandov, aby využili túto možnosť získať množstvo nových informácií o HPC na špičkových európskych superpočítačových pracoviskách.

# SUMMER of HPC 2016

2-mesačná stáž  
vo VS SAV

Katerina Galata  
Oisín Benson

**Povedzte nám niečo o sebe. Odkiaľ pochádzate a aké máte akademické zázemie?**

**KG:** Volám sa Katerina Galata a som z Atén v Grécku. Som chemická inžinierka v národnej technickej univerzite v Aténach a momentálne začínam posledný rok svojho magisterského štúdia. Tiež teraz pracujem na svojej diplomovke pod dohľadom Prof. Dorosa Theodorou.

**OB:** Som z Dublinu v Írsku. Som šachista a bývalý kapitán tímu Trinity Armstrong (Leinster Division 1). Mám rád vysoko strategické hry a sci-fi, hlavne Diasporu autora Grega Egansa. Práve som ukončil bakalárske štúdium v odbore teoretická fyzika na vysokej škole Trinity v Dubline.

**Ako ste sa dozvedeli o PRACE Summer of HPC? Prečo ste sa rozhodli zúčastniť?**

**KG:** Môj školiteľ mi odporučil aby som sa uchádzala o tento program, lebo by to pre mňa bola veľká príležitosť. Naučila by som sa veľa nových vecí a získala by som skúsenosti v odbore, v ktorom mám záujem pracovať aj naďalej. Takže som prešla zoznam tém projektov a zistila som že mnohé sú naozaj zaujímavé. Rozhodla som sa prihlásiť na projekty, ktoré sa mi zdali zaujímavé a blízke k mojej špecializácii (samozrejme že som musela spĺňať všetky

podmienky).

**OB:** O SoHPC (Summer of HPC) som už počul viackrát. Poznal som viacerých študentov z iných ročníkov, ktorý ho už absolvovali, vrátane môjho šachového spoluhráča Anthonyho, ktorý bol veľmi spokojný so svojím pobytom v Českej republike. Odporúčali mi to aj výskumníci v ICHEC (Irish Centre for High-End Computing) a profesori, ktorý učili programovanie počas môjho štúdia. Zo všetkých mojich predmetov sa mi najviac páčili tie, ktoré súviseli s programovaním. Implementovanie algoritmov ako Ising Model a zbieranie ich výsledkov je veľmi uspokojujúce. PRACE SoHPC sa zdalo byť vynikajúcou príležitosťou získať skúsenosti s úlohami a nástrojmi vrcholového výskumu a zároveň šancu chvíľu žiť v inej krajine.

**Prečo ste sa rozhodli prísť do našej organizácie a čo iné ste mali ešte na výber?**

**KG:** Rozhodla som sa prísť do SAV, lebo sa mi veľmi páčila téma na ktorú som sa prihlásila. Jej názov je „Výpočet nanotrubic využitím vlastností špirálovej symetrie“. Mala som v tomto odbore dobré teoretické vzdelanie a spĺňala som všetky podmienky. Tiež som sa prihlásila na program „Písať španielsky“, ale program v SAV bola moja prvá voľba a dúfala som, že ma sem vyberú.

**OB:** Obidva projekty, ktoré sa mi najviac páčili súviseli s big data. Predchádzajúce leto som strávil vo firme Workday, kde som pracoval s dátami pomocou prostredí Hive a R a zároveň som nazbieral skúsenosti s programovacím prostredím Scala. Keď som pracoval v tejto firme, vychádzali aj ďalšie nové nástroje ako napríklad Spark. Projekt Michala Pitoňáka sa mi páčil, lebo mi umožňoval stavať na mojich predchádzajúcich skúsenostiach so Scalou a naučiť sa pracovať s novým nástrojom na spracovanie údajov Spark. Navyše som mal veľké skúsenosti s kvantovou chémiou z predmetov o teórii zhustenej hmoty v rámci môjho štúdia, takže som mohol viac času stráviť učením sa pracovať so Sparkom.

**Aké boli vaše zadania? Boli vaši školitelia nápomocní pri riešení úloh?**

**KG:** V tomto projekte som musela paralelizovať niektoré rutiny z existujúceho kódu. Toto bolo celkom zložité, lebo ako každý vie, je vždy ťažké porozumieť tomu, čo robí iný programátor, a prečo to robí. Teraz si predstavte, že tento kód bol napísaný niekoľkými programátormi v priebehu niekoľko rokov.



**Katerina Galata je chemická inžinierka v Národnej technickej univerzite v Aténach a momentálne začína posledný rok svojho magisterského štúdia.**



**Oisín Benson  
ukončil  
bakalárske  
štúdium v odbore  
teoretická fyzika  
na vysokej škole  
Trinity v Dubline.**

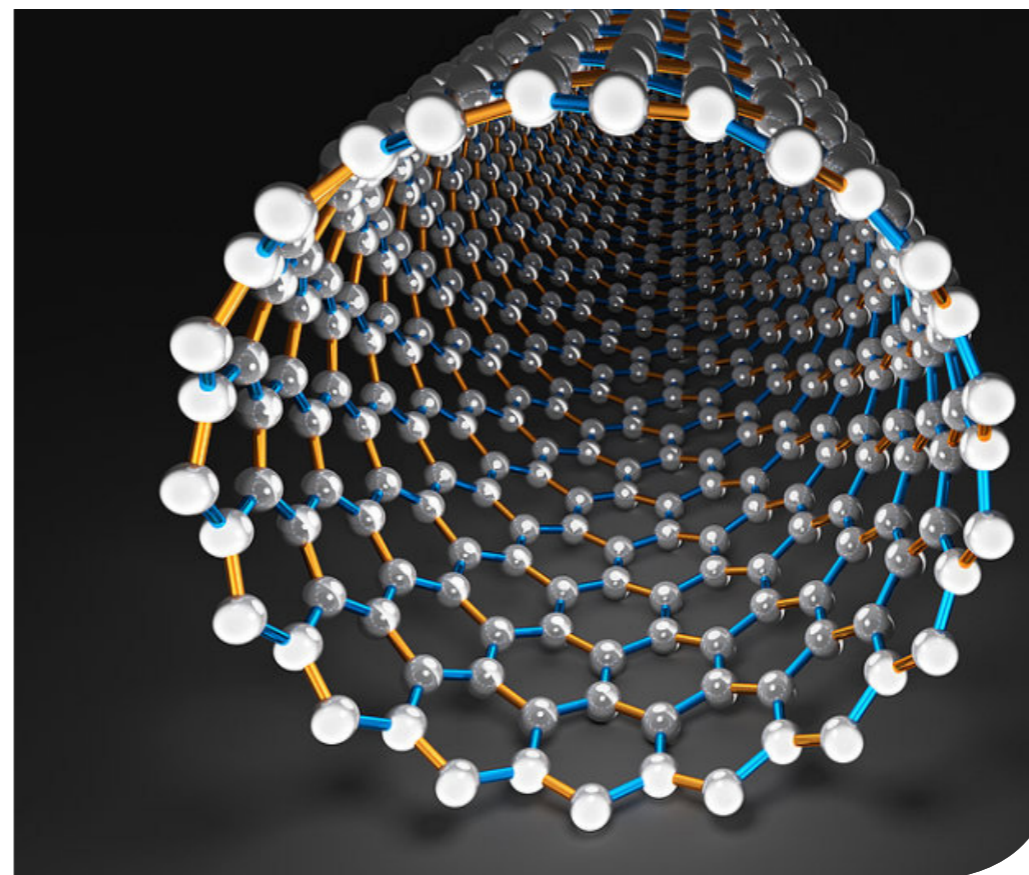
Našťastie som pracovala len s niektorými časťami tohto kódu. Prof. Noga mi poskytol potrebné inštrukcie a odporúčania ako všetko spraviť čo najlepšie a najjednoduchšie. Keď som mala ťažkosti, pomáhal a viedol ma Lukáš Demovič. Vďaka nemu som získala schopnosti a zručnosti s MPI. Tiež mi ukázal ako pracovať s výpočtovými klastrami ako s tým v Žiline. Na tak zložitom projekte som ešte nikdy som nerobila a som im obom veľmi vďačná.

**OB:** Mojou úlohou bolo prispôsobiť knižnicu libint C pre metódu Hartree-Fock v kvantovej chémii pre Spark, spustiť záťažové testy a optimalizovať ju. Najprv sme ale mali implementovať tento kód v multiplatformovom jazyku Scala, ktorý používa mnoho funkčných metód a transformácií, na ktoré sa spolieha aj Spark. Lukáš Demovič mi pomohol nastaviť veľa potrebných nástrojov a ušetril mi veľa času. Michal potom načrtnol plán ďalšej činnosti a odporučil mi užitočné príklady iných implementácií metódy Hartree-Fock. Nevýhodou používania týchto nových metód bolo, že som všetky svoje otázky ohľadom Scaly a Sparku musel smerovať na Michala. Lukáš mi tiež pomohol s odstraňovaním chýb, aj keď s týmito nástrojmi nemá skúsenosti.

**Boli ste spokojní s výsledkami, ktoré ste dosiahli? Sú vaše výsledky uplatniteľné do budúcnosti?**

**KG:** Cieľom projektu bolo vypočítať vlastnosti nanotrubic pomocou špirálovej symetrie. Implementácia je založená na formulácii v dvojrozmernom recipročnom priestore, kde jeden rozmer je spojitý a jeden diskretný. Na výpočet pásových štruktúr sa používajú metódy kvantovej chémie nezávislých častíc, ako Hartree-Fock, DFT alebo jednoduchý post Hartree-Fock MP2. MPI paralelizácia umožňuje veľmi presné výpočty pásových štruktúr nanotrubic na viacerých nódoch s rozdelenou pamäťou. K dispozícii teraz máme nové možnosti veľmi presných výpočtov energií a pásových štruktúr s dobrou schopnosťou predvídať a pásovou topológiou, ktorej interpretácia je omnoho transparentnejšia ako pri jedno priestorových prístupoch. Výsledky umožnia používať tento projekt ako štandardný nástroj na in silico dizajn. Kód ale stále potrebuje viac paralelizácie a hlavne optimalizácie.

**OB:** Aj keď sme nanešťastie nestihli tolko, koľko sme plánovali, podarilo sa nám významne pokročiť v riešení problému. Najprv som prepísal C implementáciu libint v Scale pomocou jej výpočtovej knižnice Breeze. Nástroj Breeze sme zvolili lebo jeho operácie sú základom Sparkovej knižnice strojového učenia Mlib. Keď sa nám ju podarilo spustiť pre



malé bázy, začali sme ju adoptovať pre Spark. To znamenalo, že sme začali viac používať funkčnú časť jazyku Scala, ktorý natívne obsahuje mnoho dátových transformácií, ktoré sú aj súčasťou Sparku. Nakoniec sa nám podarilo vytvoriť verziu, ktorá fungovala pre molekulu H<sub>2</sub>O v malej báze. Zatiaľ sa nám nepodarilo vyriešiť to, že kód produkuje rozdielne energie pre väčšie bázy. Michal a jeho Phd študent Miro na tom budú ďalej pracovať a isto sa im podarí dosiahnuť, aby kód škáloval správne. Algoritmus by potom mal fungovať správne aj pre veľké bázy. Aj keď by som to rád dokončil osobne, som spokojný s pokrokom, ktorý sme dosiahli a s nástrojmi, ktoré som sa naučil používať a určite využijem vo svojej budúcej kariére.

**Ako sa Vám páčilo v Bratislave a u nás vo Výpočtovom stredisku SAV?**

**KG:** No... z toho čo som videla má Slovensko úžasne krásnu prírodu. A život v Bratislave je uvoľnený, čo mi pomohlo, lebo prvýkrát v živote som musela celé leto pracovať. Ďalšou výhodou je, že sa všade dá dostať pešo za krátku

dobu. Bratislava je tiež blízko ďalších Európskych hlavných miest, kam sa dá cestovať cez víkendy. Vo výpočtovom stredisku sa mi páčilo, všetci tam boli milí a veselí a snažili sa, aby som sa cítila dobre. Povodili nás s Oisínom aj po Bratislave. Lukáš mi na leto dokonca požičal svoj bicykel. Nemôžem sa na nič sťažovať a som veľmi vďačná, lebo byť cudzincom v novom meste môže byť ťažké, hlavne ak nerozprávate jazykom domácich.

**OB:** V Bratislave sa mi páčilo. Je veľmi energická a plná mladých ľudí. Nechýbalo mi šedivé Írske leto a bolo príjemné ísť von bez veľkého miňania. Bratislava má ešte jednu výhodu, keďže susedí s ďalšími veľkými Európskymi mestami. Je jednoduché spraviť si krátky výlet do Viedne alebo Budapešti. Všetci v práci boli ku mne a Katerine veľmi milí a letná opekačka bola vynikajúcim ukončením nášho výskumného pobytu. V práci som spoznal ďalšieho silného šachového hráča, Jozefa Federiča, s ktorým sme odohrali niekoľko rýchlych partíí. Celkovo sa mi toto leto veľmi páčilo, veľa som sa naučil, našiel som nových kamarátov a navštívil som niekoľko skvelých miest.

**Ďakujeme za rozhovor a prajeme vám do budúcnosti veľa nielen pracovných úspechov.**

# SUMMER of HPC 2017

Adrián Rodríguez-Bazaga  
Andreas Neophytou

**Poprosím vás na začiatok o krátke predstavenie. Kde a aký odbor ste študovali a aká je vaša špecializácia?**

**ARB:** Volám sa Adrián Rodríguez-Bazaga a som španielsky počítačový vedec špecializujúci sa na algoritmy, umelú inteligenciu a výpočtovú teóriu. Momentálne študujem magisterský program v odbore inovácia a výskum špecializovaný na strojové učenie a dátovú vedu.

Pracujem ako výskumník v oddelení vied o živote Barcelonskej univerzity, v superpočítačovom centre. Spolupracujem na výskumnom projekte v odbore dátovej vedy a spracovania prirodzených jazykov. Snažím sa analyzovať evolúciu bioinformatiky v poslednom storočí s použitím novátorských postupov ktoré vyžadujú techniky strojového učenia ako latentnú Dirichlet Allokáciu a zoskupovacie techniky ako spektrálne zoskupovanie.

**AN:** Volám sa Andreas Neophytou a pochádzam z Birminghamu v Anglicku, kde študujem prírodné vedy. Mojou špecializáciou je chémia a biochémia a ako vedľajší smer študujem informatiku, ktorá mi otvorila bránu k HPC a vedeckým

výpočtom. Aktuálne som študentom magisterského štúdia na University of Birmingham a pracujem na projekte výskumu energií kryštálových štruktúr. V rámci projektu sa venujem aplikácii optimalizačných metód na semiempirické modely malých organických molekúl za účelom nájdenia ich energeticky najnižších štruktúr.

**Ako ste sa dozvedeli o akcii Summer of HPC organizované PRACE a čo vás viedlo k rozhodnutiu zapojiť sa a vybrať si práve nás - Výpočtové stredisko SAV?**

**ARB:** O Summer of HPC som sa dozvedel lebo mi o ňom

povedal učiteľ z mojej predchádzajúcej university a rozhodol som sa zúčastniť, lebo to bola vynikajúca príležitosť získať skúsenosti s vysokovýkonným počítaním. Vybral som si Slovenskú akadémiu vied, lebo jeden z projektov vyžadoval znalosti a používanie nástrojov, ktoré už poznám, ako Apache Spark.

**AN:** O programe som sa dozvedel od kolegu, doktoranda, na univerzite. Nakoľko boli moje skúsenosti s HPC minimálne, využil som túto možnosť získať skúsenosti „z prvej ruky“, či už z oblasti práce so superpočítačmi ako aj paralelným programovaním. Vybral som si Výpočtové stredisko SAV kvôli projektu, ktorý predkladalo. Bol to jediný projekt, ktorý zahŕňal aplikovanie HPC na problémy súvisiace s chémiou.

**Môžete nás bližšie oboznámiť s priebehom projektu?**

**ARB:** Môj projekt postupoval celkom dobre, podarilo sa nám dosiahnuť niekoľko cieľov a objaviť nové spôsoby ako pokračovať v ďalšom výskume. V spolupráci pokračujeme aj po skončení SoHPC a náš konečný cieľ je spolu napísať vedecký článok.

**AN:** Mojou úlohou bolo implementovať MPI paralelizáciu v presných výpočtoch pásovej štruktúry nanotrúbiek pre výpočtové systémy typu „klastery“, t.j. na samostatné výpočtové uzly. Výstup projektu bol v zásade v súlade so zadaním, napriek veľkosti a zložitosti paralelizovaného programu, ktoré predčili moje obavy a očakávania.

**Aká úloha vám bola pridelená? Mali ste komplikácie so zadaním, alebo sa vám ju bez problémov podarilo vyriešiť?**

**ARB:** Naším hlavným cieľom bolo skombinovať výhody pôvodných MPI prístupov s nástrojmi na spravovanie veľkých dát. Snažili sme sa prekonať rozpor medzi HPC a veľkými dátami a zistiť, či sa bežne používané nástroje na správu veľkých dát dajú použiť aj pri riešení HPC úloh porovnaním dvoch výpočtových rámcov implementovaných na komoditných klastrových architektúrach. Porovnávali sme Apache Spark Big Data rámec implementovaný v programovacom jazyku Scala s tradičnými prístupmi ako používanie modelu distribuovanej pamäte s MPI na distribuovanom súborovom systéme HDFS (Hadoop Distributed File System) a natívnymi C knižnicami, ktoré využívajú výhody tohto súborového systému. Zvolili sme K-stredový zoskupovací algoritmus, ktorý bol spustený na dátových množinách rôznych veľkostí a porovnávali sme



**Adrián Rodríguez-Bazaga je študentom magisterského programu na Barcelonskej univerzite špecializujúci sa na algoritmy, umelú inteligenciu a výpočtovú teóriu.**

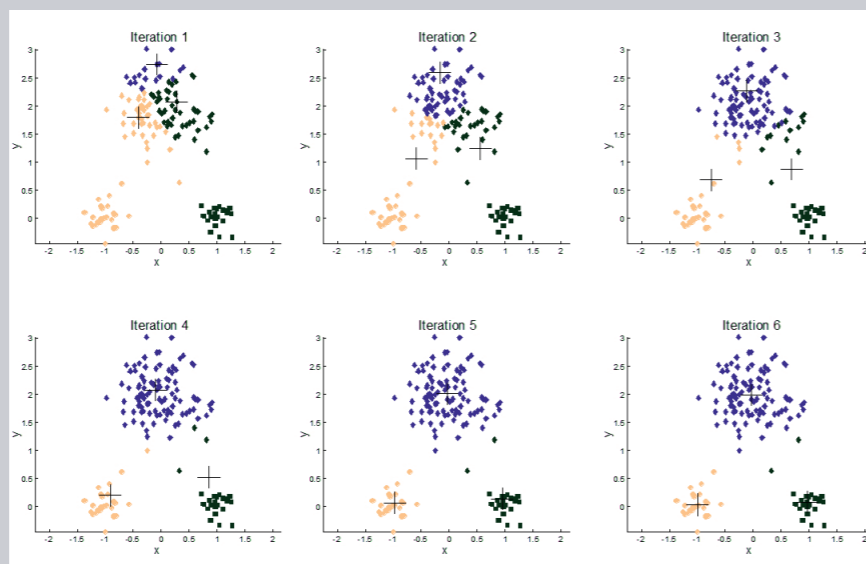


výpočtový čas a odolnosť voči poruchám oboch postupov.

AN: Zadanie projektu bolo od začiatku jasné, čo však neznamená, že projekt ako taký,

AN: Mojím školiteľom bol prof. Jozef Noga. Počas projektu bol extrémne nápomocný a pochybujem, že by som bol schopný projekt dotiahnuť do úspešného konca bez jeho pomoci. Nápomocnými boli aj ostatní kolegovia z Výpočtového strediska SAV, ktorí mi pomáhali nie len s projektom, ale mi spríjemňovali celý letný pobyt.

**OBRÁZOK:**  
Znázornenie jednotlivých iterácií K-means algoritmu na hľadanie centier klastrov s použitím modelových dát.



bol jednoduchý! Napriek tomu som sa s ním vysporiadal, aj keď som bol občas v pochybnostiach.

**Kto boli vaši školitelia? Boli nápomocní pri riešení?**

ARB: Mojím školiteľom bol doc. Mgr. Michal Pitoňák, PhD. a naozaj mi veľmi pomohol s mojou prácou. Poskytol mi mnoho nových informácií, snažil sa odpovedať na všetky moje pochybnosti a dával pozor, aby sa projekt vyvíjal správne. Je to nielen vynikajúci výskumník, ale aj skvelý človek, s ktorým momentálne pracujem na rozšírení výsledkov, ktoré sme dosiahli počas SoHPC.

**Uplatníte do budúcnosti výsledky a informácie, ktoré ste získali počas školenia?**

ARB: Vedomosti získané počas SoHPC samozrejme využívam v každodennej práci, lebo riešim veľmi zložité úlohy s pomocou HPC technológií. Vďaka SoHPC som s HPC prostredím získal množstvo užitočných skúseností.

AN: Samozrejme! Ihneď po návrate som využil novonadobudnuté vedomosti o MPI paralelizme pri programovaní, ktoré je súčasťou projektu, na ktorom aktuálne pracujem.

**Ako by ste celkovo zhodnotili vaše pôsobenie vo VS SAV počas dvojmesačnej stáže.**

ARB: Vo všeobecnosti bol môj pobyt veľmi dobrý. Samozrejme všade dobre, doma najlepšie, ale nemôžem sa sťažovať. Ľudia boli veľmi milí a som veľmi spokojný so svojim výberom destinácie.

AN: Pobyt na Výpočtovom stredisku SAV som si naozaj užil. Ak by som ho mal ohodnotiť bodovo, dal by som mu osem bodov z desiatich (dva body som strhol kvôli daždivému

počasí!). Naozaj neviem, čo by bolo potrebné vylepšiť na zvýšenie kvality pobytu.

**Bola to vaša prvá návšteva Slovenska, Bratislavy?**

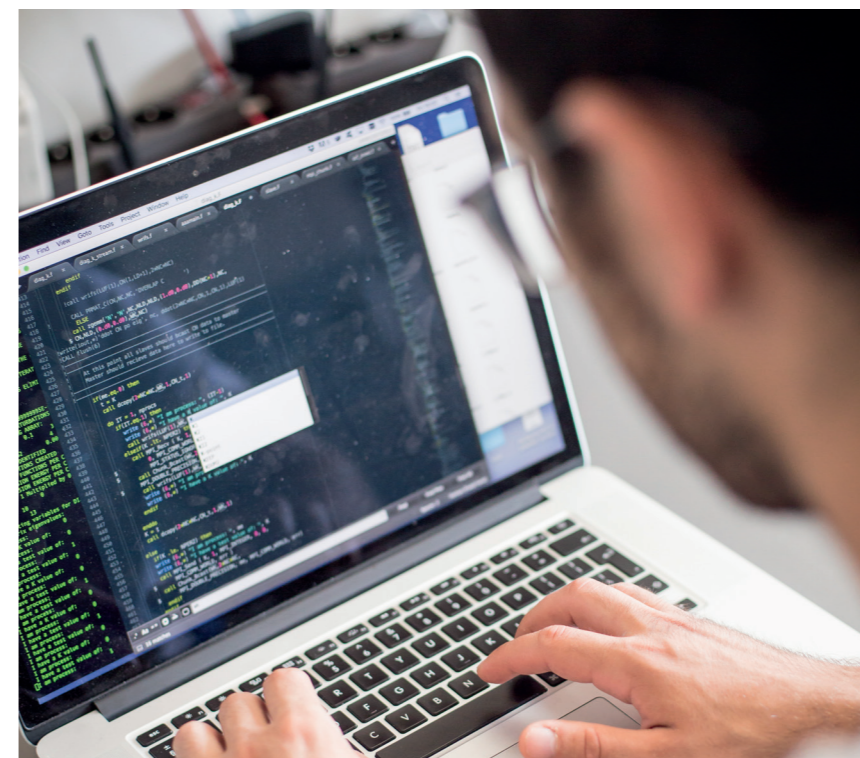
ARB: Áno, prvýkrát v živote som bol v Bratislave, dokonca aj vo východnej Európe. Je to veľmi pekné mesto s množstvom historických monumentov.

AN: Bola to moja prvá návšteva Slovenska a veľmi som si to tu užil. Ľudia sú priateľskí, zažil som veľa zaujímavých rozhovorov s „cudzincami“ cestou autobusom do práce.

**Podarilo sa vám navštíviť nejaké pamiatky? Vrátili by ste sa ešte naspäť, ale už ako turisti?**

ARB: Navštívili sme niekoľko zaujímavých miest ako Bratislavský zámok, prezidentský palác a jeho záhradu. Tiež si pamätám sochu Čumila v starom meste, bola veľmi vtipná. V Bratislave je ešte veľa iných zaujímavých miest a rád by som sa jedného dňa vrátil, aby som si ich pozrel aspoň ako turista.

AN: Počas pobytu som mal možnosť navštíviť niekoľko zaujímavých miest v Bratislave ako aj zavítať do susedných veľkomiest (Viedeň a Budapešť). Vrcholom pobytu, z tohto hľadiska, bol výlet na lodkách na Malom Dunaji s kolegami z Výpočtového strediska SAV. Určite plánujem Slovensko opäť navštíviť, nakoľko sa mi na mnohé zaujímavé miesta, ako napríklad Vysoké Tatry, dostať nepodarilo.



**Andreas Neophytou** študuje chémiu a biochémiu a ako vedľajší smer informatiku. Aktuálne je študentom magisterského štúdia na Birminghamskej univerzite, kde pracuje na projekte výskumu energií kryštálových štruktúr.

## RIADITEĽ VS SAV

Lukáš Demovič

## ZODPOVEDNÝ REDAKTOR

### A KOORDINÁTOR

Gabriela Obadalová

## GRAFIKA A DTP

Gabriela Obadalová

## PREKLAD

Michal Pitoňák

Miloslav Valčo

## FOTOGRAF

Pavol Novák

## TLAČ

FOART, s.r.o.

Prešovská 45

821 02 Bratislava

Slovenská republika

Tel.: +421 (0)2/207 821 18

E-mail: foart@foart.sk

[www.foart.sk](http://www.foart.sk)

## ADRESA

Výpočtové stredisko

Slovenskej akadémie vied

Dúbravská cesta 9

845 35 Bratislava

Slovenská republika

Tel.: +421 (0)2/3229 3111

Fax: +421 (0)2/3229 3103

E-mail: vssav@savba.sk

[www.vs.sav.sk](http://www.vs.sav.sk)

Texty neprešli jazykovou korektúrou.

Výpočtové stredisko  
Slovenskej akadémie vied

Dúbravská cesta 9  
845 35 Bratislava  
Slovenská republika

[www.vs.sav.sk](http://www.vs.sav.sk)

