

De Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten is zich de jongste jaren steeds meer gaan toeleggen op het analyseren van maatschappelijk relevante fenomenen en vraagstukken, die gekenmerkt worden door een hoog actualiteitsgehalte. Zo werden al standpunten ingenomen in verband met de vergrijzing, de financiering van de sociale zekerheid, de demografische evolutie en het gebruik van bibliometrie in de menswetenschappen.



Koninklijke Vlaamse Academie van België  
voor Wetenschappen en Kunsten

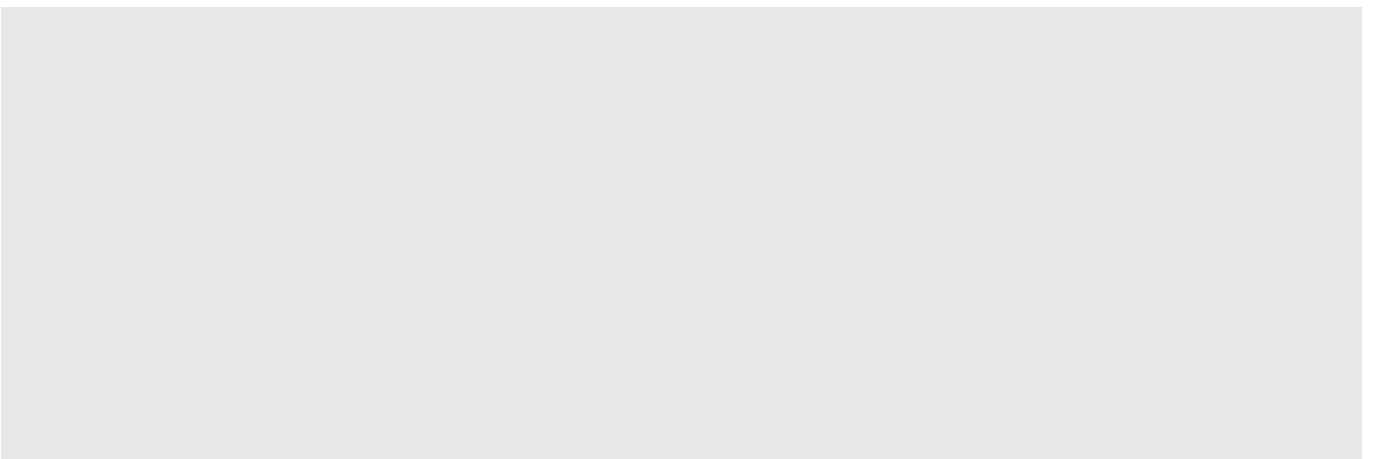
Hertogsstraat 1  
1000 Brussel  
Tel. 02 550 23 23  
Fax 02 550 23 25  
[www.kvab.be](http://www.kvab.be)  
[info@kvab.be](mailto:info@kvab.be)

STANDPUNTEN

High Performance Computing  
(HPC) in Vlaanderen



## High Performance Computing (HPC) in Vlaanderen



Standpunten  
van  
de Koninklijke  
Vlaamse Academie  
van België  
voor  
Wetenschappen  
en Kunsten

Standpunt, nummer 5

---



Koninklijke Vlaamse Academie van België  
voor Wetenschappen en Kunsten

Hertogsstraat 1  
1000 Brussel  
Tel. 02 550 23 23  
Fax 02 550 23 25  
[www.kvab.be](http://www.kvab.be)  
[info@kvab.be](mailto:info@kvab.be)



High Performance Computing (HPC)  
In Vlaanderen



Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

© Copyright 2008 KVAB  
D/2008/0455/06

No part of this book may be reproduced in any form, by print, photo print, microfilm or any other means without written permission from the publisher.



## INHOUD

MANAGEMENT SAMENVATTING .....	7
<b>1. INLEIDING</b>	
1.1. Werkwijze .....	9
<b>2. OVERZICHT VAN HPC-GEBONDEN BAANBREKEND ONDERZOEK .....</b>	<b>10</b>
2.1. De rol van HPC in Wetenschap & Technologie .....	10
2.2. Voorbeelden van Research Challenges .....	11
2.2.1. Biologie en Biotechnologie .....	11
2.2.2. Klimaat en Global Change .....	11
2.2.3. Microelektronica .....	11
2.2.4. Astrofysica en Kosmologie .....	12
2.3. HPC-toepassingen in Vlaanderen .....	15
2.3.1. Materiaalonderzoek en Nanotechnologie .....	15
2.3.2. Moleculaire Biologie en Biotechnologie, Genetica en Fysiologie ..	15
2.3.3. Engineering .....	15
2.3.4. Moleculaire Chemie .....	15
2.3.5. Toegepaste Wiskunde en Mathematische Algoritmen .....	16
2.3.6. Astrofysica en Kosmologie .....	16
2.3.7. Subatomaire Fysica .....	16
2.3.8. Turbulentie en Verbranding .....	16
2.3.9. Milieustudies .....	16
2.3.10. Fundamentele Fysica .....	17
2.3.11. Taalkunde .....	17
<b>3. OVERZICHT VAN BESTAANDE HPC-INFRASTRUCTUUR, WERELDWIJD, IN EUROPA EN IN VLAANDEREN .....</b>	<b>17</b>
3.1. HPC op wereldvlak .....	17
3.1.1. De top-10 HPC-installaties – November 2006 .....	18
3.2. HPC in Europa .....	18
3.2.1. Top HPC-systemen in Europa (November 2006) .....	18
3.2.2. HPC-centra in vergelijkbare landen .....	19
3.3. HPC in Vlaanderen .....	22
3.3.1. De HPC-cluster van UAntwerpen .....	22
3.3.2. De HPC-cluster van KULeuven .....	22
3.3.3. De HPC-cluster van UGent .....	22
<b>4. DE EUROPESE CONTEXT .....</b>	<b>23</b>
4.1. DEISA .....	23
4.2. EGEE (Enabling Grids for E-sciencE) .....	23
4.3. De Europese HPC-visie .....	24



5. HPC-BEHOEFTEN IN VLAANDEREN . . . . .	25
5.1. Voor een HPC-infrastructuur in Vlaanderen . . . . .	26
5.2. Budgettaire en structurele implicaties . . . . .	27
6. SLOTBESCHOUWINGEN . . . . .	28
BIJLAGE 1: SAMENSTELLING VAN DE KVAB WERKGROEP HPC . . . . .	30
BIJLAGE 2: ENQUETE FORMULIER . . . . .	31
BIJLAGE 3: SAMENVATTINGEN VAN DE ENQUÊTE RESULTATEN IN VLAANDEREN . . .	33
Bijlage 3a: Materiaalonderzoek en Nanotechnologie . . . . .	33
Bijlage 3b: Moleculaire Biologie en Biotechnologie, Genetica en Fysiologie . .	37
Bijlage 3c: Engineering . . . . .	40
Bijlage 3d: Moleculaire Chemie . . . . .	47
Bijlage 3e: Toegepaste Wiskunde en Mathematische Algoritmen . . . . .	51
Bijlage 3f: Astrofysica en Kosmologie . . . . .	54
Bijlage 3g: Subatomaire Fysica . . . . .	56
Bijlage 3h: Turbulentie en Verbranding . . . . .	59
Bijlage 3i: Milieustudies . . . . .	63
Bijlage 3j: Fundamentele fysica . . . . .	64
BIJLAGE 4: DE TOP500 HPC LIJST VAN NOVEMBER 2007 . . . . .	67
Summary . . . . .	71

**ADVIES VAN DE KVAB**  
**over**  
**HIGH PERFORMANCE COMPUTING (HPC)**  
**IN VLAANDEREN**

**Management samenvatting**

De Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten (KVAB), Klasse van de Natuurwetenschappen, heeft het initiatief genomen een advies uit te brengen over de behoeften in Vlaanderen aan een dynamische High Performance Computing (HPC) strategie.

High Performance Computing (HPC) is gericht op de aanwending van grootschalige computercapaciteit voor de studie en het oplossen van wetenschappelijke, industriële en maatschappelijke problemen.

De motivatie voor dit initiatief vertrok van de vaststelling dat Vlaanderen, net als Federaal België, onvoldoende voorbereid is om een plaats in te nemen in de systematische en grootschalige aanwending van HPC. Hierdoor ontstaat een ernstig risico voor Vlaanderen om achterop te geraken op enerzijds het internationaal wetenschappelijk forum en anderzijds in de ontwikkeling van een potentieel voor wetenschappelijke, industriële en economische vernieuwing.

De KVAB HPC-werkgroep heeft rekening gehouden met de Beleidsnota van Minister F. Moerman, Vlaanderen i2010: 'Tijd voor een Digitale Stroomversnelling in de Innovatieketen', en heeft de aanbevelingen over de behoefte aan digitale infrastructuur in Vlaanderen, als richtlijnen voor haar activiteiten overgenomen.

De Werkgroep besliste een diepgaande enquête uit te voeren bij alle wetenschappelijke instellingen in Vlaanderen, om de behoeften en wensen van de Vlaamse onderzoeksgemeenschap voor HPC in kaart te brengen.

De informatie uit de enquête werd in lijn gebracht met de wereldwijde impact van HPC op wetenschappelijke innovatie en baanbrekend onderzoek.

Een overzicht van de bestaande HPC-infrastructuren wereldwijd, in Europa en in de omliggende landen wordt ook gegeven. Kenmerkend voor de HPC-organisatie in alle landen is een piramidale structuur, met aan de basis HPC-installaties in lokale universiteiten van kleine schaal (Niveau 1) met daar bovenop regionale HPC-centra met een hogere capaciteit van de orde van 10 000 à 50 000 processoren of cores (Niveau 2). Het derde niveau kan slechts bereikt worden op Europees vlak (Niveau 3).

De meeste Europese HPC-installaties zijn aan elkaar verbonden door een hogesnelheids Europees netwerk, o.a. binnen de projecten DEISA en EGEE.

**Op basis van de verzamelde informatie wordt de oprichting van een HPC-centrum in Vlaanderen, op het Niveau 2, sterk aanbevolen.**

Dit mag echter niet gebeuren ten koste van de bestaande computerparken op Niveau 1 binnen de universiteiten en onderzoeksinstellingen, die verder ontwikkeld moeten worden om een stevige basis te vormen waarop de piramidale structuur berust.

Naast het uitvoeren van de eigenlijke grootschalige HPC-projecten moet een HPC-centrum een belangrijke rol vervullen ter ondersteuning van de gebruikers. Er moet hulp geboden worden aan de gebruikers voor taken zoals parallelisering, implementatieproblemen, software engineering, pre- en postprocessing.

De budgettaire implicaties situeren zich rond een investering van de orde van 10 M€, met een jaarlijks werkingsbudget van 5 M€, rekening houdende met een hardware afschrijving over 3 jaar en de financiering van een ondersteuningsploeg van een vijftiental medewerkers.

Als bijkomende spin-off van een Vlaams regionaal HPC-centrum mag vermeld worden dat het ook een belangrijke aansporing zou betekenen voor talrijke rand-

ontwikkelingen. Onder andere: de impact op de ontwikkeling van toegepaste wiskunde, mathematische algoritmen, beeldverwerking en visualisatie; de impact op de ontwikkeling van software engineering; de stimulans voor technologie van hoge snelheid netwerken; nieuwe HPC-gerichte ontwikkelingen in sectoren zoals Taalkunde of Financiën; nieuwe mogelijkheden voor de farmaceutische industrie bij de ontwikkeling van geneesmiddelen van de toekomst.

**Het is de overtuiging van de KVAB, dat een initiatief rond een sterk regionaal HPC-centrum in Vlaanderen, een uitzonderlijk sterke impuls zou vormen voor de verdere uitbreiding van de wetenschappelijke excellentie en baanbrekend onderzoek in alle sectoren.**

**Het zou ook een belangrijke ondersteuning betekenen voor het potentieel tot innovatie in Vlaanderen.**

**ADVIES VAN DE KVAB**  
**over**  
**HIGH PERFORMANCE COMPUTING (HPC)**  
**IN VLAANDEREN**

## **1. Inleiding**

De Klasse van de Natuurwetenschappen van de KVAB besliste op haar vergadering van oktober 2006, een advies uit te brengen over de implicaties van HIGH PERFORMANCE COMPUTING (HPC) IN VLAANDEREN, alsmede zich uit te spreken over aan te bevelen acties.

De motivatie voor dit initiatief werd als volgt geformuleerd:

- High Performance Computing (HPC) is gericht op de aanwending van grootschalige computercapaciteit voor de studie en het oplossen van wetenschappelijke en industriële problemen.
- Het huidige potentieel, met rekencapaciteiten tot 100 Teraflops ( $10^{14}$  bewerkingen/sec), opent de weg tot nieuw grensverleggend onderzoek en ontdekkingen, in alle sectoren van wetenschap en technologie.
- Vlaanderen, zowel als Federaal België, zijn onvoldoende voorbereid om een plaats in te nemen in de systematische en grootschalige aanwending van HPC in wetenschappelijk onderzoek en industriële toepassingen.
- Ten gevolge daarvan is Vlaanderen ook weinig aanwezig bij Europese initiatieven rond een Europees HPC-netwerk.
- Hierdoor ontstaat een ernstig risico voor Vlaanderen om achterop te geraken op enerzijds het internationaal wetenschappelijk forum en anderzijds op de ontwikkeling van een potentieel voor wetenschappelijke, industriële en economische vernieuwing.

### **1.1. Werkwijze**

Een Werkgroep bestaande uit leden van de Klasse van de Natuurwetenschappen werd samengesteld. Er werd gezorgd voor een brede vertegenwoordiging van de belangrijkste wetenschappelijke sectoren waar HPC een rol kan spelen en waarin Vlaamse wetenschappers actief zijn.

De samenstelling van de KVAB HPC-werkgroep is te vinden in Bijlage 1.

De KVAB HPC-werkgroep werd door Confrater J. Vandewalle op de hoogte gebracht van de belangrijke beleidsnota van Minister F. Moerman, Vlaanderen i2010 'Tijd voor een Digitale Stroomversnelling in de Innovatieketen'.

Dit document benadrukt de behoefte aan digitale infrastructuur in Vlaanderen.

Uit dit document werden de volgende aanbevelingen als richtlijnen voor de activiteiten van de Werkgroep overgenomen (blz. 40):

*'Er moet worden uitgemaakt hoe die digitale infrastructuur een verschil kan maken voor het Vlaamse onderzoekslandschap. Er moet worden nagegaan*

- *wat de huidige status is in Vlaanderen,*
- *wat de noden zijn,*
- *waar we naartoe willen,*
- *welke middelen hiervoor nodig zijn en*
- *welke strategie gevolgd moet worden om het doel te bereiken'.*

De Werkgroep besliste als volgt te werk te gaan:

1. Een diepgaande enquête uit te voeren bij alle wetenschappelijke instellingen in Vlaanderen, om de behoeften en wensen van de Vlaamse onderzoeksgemeenschap voor HPC in kaart te brengen. De Werkgroep wenste hiermee uit te maken
  - hoe een hoogstaande HPC-infrastructuur een verschil kan maken voor het Vlaamse onderzoekslandschap.
  - wat de huidige status is in Vlaanderen,
  - wat de noden zijn,
  - welke middelen hiervoor nodig zijn

Een vragenlijst (zie Bijlage 2) werd opgesteld en werd verspreid naar alle universitaire en onderzoeksinstituten in Vlaanderen. De vragenlijst werd verstuurd naar onderzoekers met bestaande of potentiële belangstelling voor HPC-gebaseerd onderzoek, in alle wetenschappelijke gebieden die beïnvloed kunnen worden door HPC-ontwikkelingen.

2. De antwoorden van de enquête werden voor de verschillende onderzoeksdomeinen samengevat in Bijlage 3.

De werkzaamheden van de HPC-werkgroep hebben zich toegespitst op de volgende punten, met het doel een overzicht te geven van de internationale en Vlaamse situaties in verband met HPC, om hiermee de basis te leggen voor aanbevelingen naar de overheid.

**DEEL I:** Overzicht van het baanbrekend onderzoek dat gevoerd kan worden met behulp van HPC

- wereldwijd innovatief onderzoek met behulp van HPC, met indeling naar de verschillende wetenschappelijke disciplines
- evaluatie van de toestand in Vlaanderen, met nadruk op het potentieel tot 'wereld-top' onderzoek, en weerslag van HPC op de wetenschappelijke vooruitgang in Vlaanderen.

**DEEL II:** Overzicht van bestaande en geplande HPC-infrastructuur, wereldwijd, in Europa, in Vlaanderen en in de omliggende landen

- Voor de bestaande HPC-centra werden drie centra geselecteerd als referentie
  - CSC in Finland
  - CSCS in Zwitserland
  - SARA in Nederland

**DEEL III:** HPC-behoefte in Vlaanderen

- omvang van een eventuele HPC-infrastructuur in Vlaanderen
- budgettaire en structurele implicaties

**DEEL IV:** Slotbeschouwingen

## **2. Overzicht van HPC-gebonden baanbrekend onderzoek**

### ***2.1. De rol van HPC in Wetenschap & Technologie***

In het verleden werd wetenschappelijk onderzoek nagenoeg uitsluitend ondersteund door experimentele en theoretische analyse.

Vandaag vormen HPC-gebaseerde simulaties een derde weg voor de ontdekking van nieuwe wetten en voor de uitbreiding van onze kennis in wetenschap en technologie.

De indrukwekkende kracht van HPC opent onderzoekswegen die ontoegankelijk zijn voor de andere methoden, wanneer hun complexiteit de analyse onmogelijk maakt en experimenten te beperkt blijken.

- Systemen kunnen gesimuleerd worden om fysische wetten te ontdekken voor welke er geen theoretische vergelijkingen bestaan.
- Bestaande vergelijkingen, die onaangetaast bleven tot voor kort, kunnen opgelost worden

Computersimulaties vormen een essentieel element in de hedendaagse wetenschappelijke ontwikkeling en belangrijke wetenschappelijke sectoren zijn vandaag grotendeels afhankelijk van de evolutie naar grootschalige computercapaciteit, o.a.

- fundamentele fysica (kosmologie, kwantumfysica en elementaire deeltjes), kernfysica,
- moleculaire chemie, farmacologie, moleculaire biologie en biotechnologie, fysiologie,
- turbulentie en verbranding,
- geologie en kennis van de aarde, klimaatvoorspelling en 'global change',
- materiaalonderzoek en nanotechnologie,
- energie en fusie
- en talrijke andere .....

Op industrieel vlak, wordt het ontwerp van vele complexe technische systemen zoals in de automobiellindustrie, meer en meer geëvalueerd op een **virtueel prototype**, d.w.z. door simulatie van een model op een computerinfrastructuur of HPC, eerder dan met een fysisch prototype. De kortere evaluatietijd en ontwikkelingstijd, het flexibel combineren met de optimalisatie en de gereduceerde kosten, zijn de voornaamste redenen voor deze verschuiving naar virtuele prototypes.

## 2.2. Voorbeelden van Research Challenges

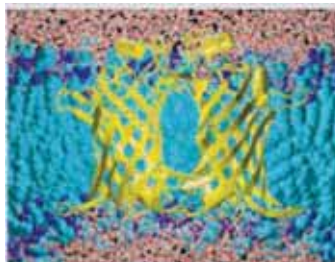
De volgende voorbeelden zijn illustratief voor het potentieel van HPC in enkele representatieve sectoren van wetenschappelijke onderzoek, op internationaal vlak.

### 2.2.1. Biologie en Biotechnologie

De HPC-behoefte voor de modelisatie in biofysica en biochemie reiken veel verder dan de huidige mogelijkheden. Met computers in de 100 Teraflops\* klasse, worden belangrijke vorderingen verwacht in moleculaire dynamica technieken voor de simulatie van biologische systemen.

Het voorbeeld hierboven behandelt een belangrijke biologische vraag, waarbij de simulatie van twee enzymen vergeleken worden: (i) acetylcholinesterase (AChE), een centraal enzym van het zenuwstelsel en (ii) Outer Membrane Phospholipase A (OMPLA), een bacterieel enzym kenmerkend voor pathogenese.

De simulatie toont aan dat de twee enzymen gelijksoortige actieve sites hebben<sup>1</sup>.

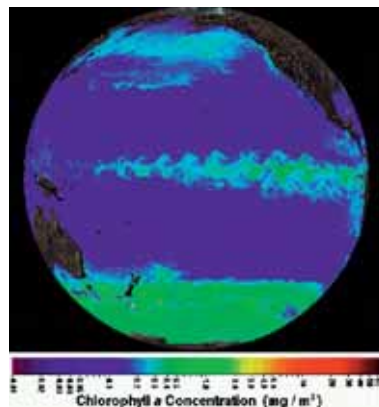


### 2.2.2. Klimaat en Global Change

Voorspellingen van klimaat evolutie en broeikast effecten zijn volledig afhankelijk van de beschikbare HPC-rekencapaciteit, om de complexe fysica van de wisselwerking tussen aarde, atmosfeer, oceanen en zonnestraling te modelleren.

Ramingen<sup>2</sup> van de behoeften voor de beschrijving van de ozonlaag, duiden aan dat een rekenkracht van de orde van 100 Teraflops noodzakelijk is

Een ander voorbeeld, uit dezelfde referentie<sup>2</sup>, bevat een HPC-schatting voor een hogere resolutie bij de klimaatvoorspelling en de invloed van ondermeer de bewolking. Om een resolutie van 3 km te bereiken, is een rekencapaciteit van de orde van 10 Petaflops vereist. De bijgaande figuur toont een simulatie van de invloed van biogeochemische processen op de chlorofylconcentratie in de Stille Oceaan, als gevolg van de toename van het fytoplankton.



### 2.2.3. Microelektronica<sup>3</sup>

Defecten in halfgeleiders hebben een dynamisch gedrag afhankelijk van o.a. temperatuur op faseveranderingen. Onzuiverheden beïnvloeden de beweging van

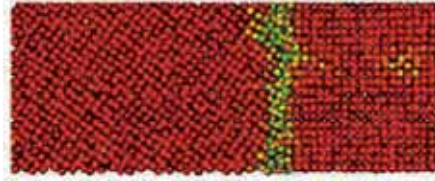
\* Flops staat voor 'Floating point operations per second' en is een maat voor het aantal aritmetische bewerkingen (zoals optellingen) uitgevoerd door de computer. Een Gflops is één miljard ( $10^9$  Flops) bewerkingen per seconde; een Teraflops of TFlops is duizend Gflops ( $10^{12}$  Flops) en een Petaflops is duizend Teraflops ( $10^{15}$  Flops).

<sup>1</sup> Stimulating simulated biology.

<http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/research/esci/biosimgrid.pdf>

<sup>2</sup> Frontier Simulations for the Climate of the 21st Century. Jean-Claude ANDRE. CERFACS, Toulouse (France) <http://www.cerfacs.fr>

interfaces, bij zeer lage concentraties, op het niveau van enkele ppm. De figuur hiernaast toont een snapshot van de oriëntatie van de atomen in de interface en in de twee kristallen, verkregen met behulp van een HPC-simulatie.



#### 2.2.4. Astrofysica en Kosmologie

De modellering die de vorming van materie in sterren, sterrenclusters en grotere eenheden beschrijft uitgaande van de basiswetten van de fysica, leidt tot zeer rekenintensieve simulaties waarvoor supercomputer-clusters onontbeerlijk zijn.

De figuur hiernaast toont een snapshot van de simulatie van een Type Ia supernovae ontploffing<sup>4</sup>. Dit vereist, naast de modelisering van turbulentie en deflagratie, ook de effecten van transport door neutrinostraling. De vergelijking van de driedimensionale radiatiespectra met waarnemingen is essentieel voor de bepaling van kosmologische parameters.

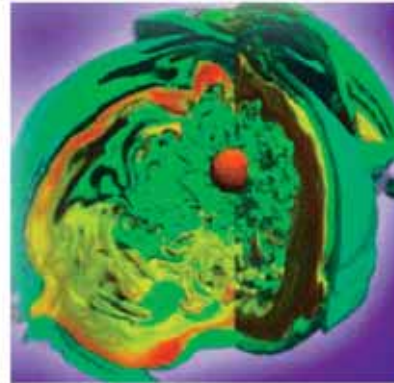
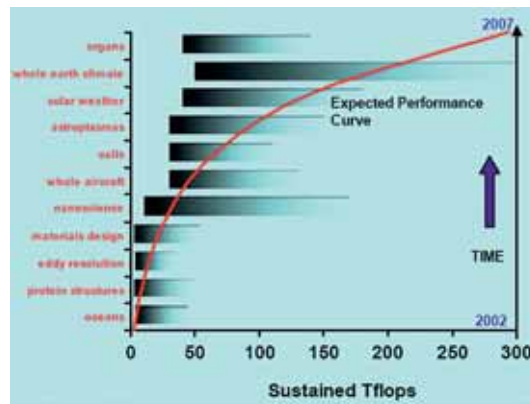


Figure 5. Snapshot from a stellar explosion simulation. Capturing the complex, turbulent dynamics in a supernova environment is a challenge for computational astrophysicists and visualization experts alike.

De visie op de HPC-behoeften om kritische problemen voor wetenschap, milieu en technologie te kunnen behandelen in de toekomst, wordt samengevat in de volgende figuur<sup>5</sup>.



De volgende tabel 2.1 geeft een overzicht per discipline van de behoeften om een bepaald niveau van kennis te bereiken, op het Europees niveau van een PetaFlops capaciteit. Dit recent verslag<sup>6</sup> van HET, de High-Performance Computing in Europe Taskforce, bevat een uitgebreide beschrijving van de onderwerpen vermeld in deze Tabel 2.1

<sup>3</sup> Opportunities for discovery: theory and computation in basic energy sciences. U.S. Department of Energy. Subcommittee on theory and computation of the basic energy sciences advisory committee. January 2005.

<sup>4</sup> A Science-Based Case For Large-Scale Simulation, Vol 2. Office of Science, U.S. Department of Energy, September 2004.

<sup>5</sup> A Strategic Framework for High End Computing –UK. <http://www.epsrc.ac.uk/ResearchFunding/FacilitiesAndServices/HighPerformanceComputing/HPCstrategy/2006StrategicFramework.htm>

<sup>6</sup> The Scientific Case for a European Super Computing Infrastructure: Petascale Computing in Europe (2006). Report from the High-Performance computing in Europe Taskforce (HET), 95 pages. <http://www.hpcineuropetaskforce.eu/deliverables>

Tabel 2.1. The Challenges and Outcomes in Science and Engineering to be addressed through Petascale HPC Provision.

Area	Application	Science Challenges & Potential Outcomes
<b>WEATHER, CLIMATOLOGY AND EARTH SCIENCES</b>	Climate change	Quantify uncertainties on the degree of warming and the likely impacts by increasing the capability and complexity of 'whole earth system' models that represent in ever-increasing realism and detail the scenarios for our future climate.
	Oceanography and Marine Forecasting	Build the most efficient modeling and prediction systems to study, understand and predict ocean properties and variations at all scales, and develop economically relevant applications to inform policy and develop services for government and industry.
	Meteorology, Hydrology and Air Quality	Predict weather and flood events with high socio-economic and environmental impact within a few days. Understand and predict the quality of air at the earth's surface; development of advanced real-time forecasting systems for allowing early enough warning and practical mitigation in the case of pollution crisis.
	Earth Sciences	Sciences Challenges span a wide range of disciplines and have significant scientific and social implications, such as the mitigation of seismic hazards, treaty verification for nuclear weapons, and increased discovery of economically recoverable petroleum resources and monitoring of waste disposal. Increased computing capability will make it increasingly possible to address the issues of resolution, complexity, duration, confidence and certainty, and to resolve explicitly phenomena that were previously parameterized, and will lead to operational applications in other European centers, national centers and in industry.
<b>ASTROPHYSICS, HEP AND PLASMA PHYSICS</b>	Astrophysics	Deal with systems and structures which span a large range of different length and time scales; almost always non-linear coupled systems of ordinary and partial differential equations have to be integrated, in 3 spatial dimensions and explicitly in time, with rather complex material functions as input. Grand challenges range from the formation of stars and planets to questions concerning the origin and the evolution of the Universe as a whole. Evaluate the huge amount of data expected from future space experiments such as the European Planck Surveyor satellite.
	Elementary Particle Physics	Quantum field theories like QCD (quantum chromodynamics) are the topic of intense theoretical and experimental research by a large and truly international community involving large European centers like CERN and DESY. This research not only promise to yield a much deeper understanding of the standard model of elementary particles and the forces between them, as well as nuclear forces, but is also expected to discover hints for a yet unknown physics beyond the standard model.
	Plasma physics	The science and technology challenge raised by the construction of the magnetic confinement thermonuclear fusion reactor ITER calls for a major theory and modeling activity. Both the success of the experiment and its safety rely on such simulators. The quest to realize thermonuclear fusion by magnetically confining a high temperature plasma poses some of the computationally most challenging problems of nonlinear physics.
<b>MATERIALS SCIENCE, CHEMISTRY AND NANOSCIENCE</b>	Understanding Complex Materials	The determination of electronic and transport properties central to many devices in the electronic industry and hence progress the understanding of technologically relevant materials. Simulations of nucleation, growth, self-assembly and polymerization central to the design and performance of many diverse materials e.g., rubbers, paints, fuels, detergents, functional organic materials, cosmetics and food. Multiscale descriptions of the mechanical properties of materials to determine the relation between process, conditions of use and composition e.g., in nuclear energy production. Such simulations are central to the prediction of the lifetime of high performance materials in energy technology, such as high-efficiency gas-turbines
	Understanding Complex Chemistry	Catalysis is a major challenge in the chemistry of complex materials, with many applications in industrial chemistry. The knowledge of atmospheric chemistry is crucial for environmental prediction and protection (clean air). Improving the knowledge of chemical processing (from soft chemistry including polymers to the atomistic description of combustion) would improve the durability of chemicals. Supra molecular assemblies open new possibilities for the extraction of heavy elements from spent nuclear fuels. In biochemistry, a vast number of reactions taking place in the human body (for example) are not understood in any detail. A key step in the development of the clean fuels of the future requires the realistic treatment of supported catalytic nanoparticles
	Nanoscience	The advance of faster information processing or the development of new generations of processors requires the shrinking of devices, which leads inevitably towards nanoelectronics. Moreover, many new devices, such as nanomotors can be envisioned, which will require simulation of mechanical properties at the nanolevel. Composite high performance materials in the fields e.g. adhesion and coatings will require an atomistic based description of nanorheology, nanofluidics and nanotribology. As an example the description of the complex magnetic and mechano-optical properties of nanodevices components is only feasible only on systems in the Petaflops range.



Area	Application	Science Challenges & Potential Outcomes
LIFE SCIENCES	Systems Biology	The use of increasingly sophisticated models to represent the entire behavior of cells, tissues, and organs, or to evaluate degradation routes predicting the final excretion product of any drug in any organism. To position Europe in the next 4 years to host the first "in silico" cell.
	Chromatine Dynamics	The organization of DNA in nucleosomes largely modifies the accessibility of transcription factors recognition sites playing then a key role in the regulation of gene function. The understanding of nucleosome dynamics, positioning, phasing, formation and disruption or modifications induced by chemical modifications, or by changes in the environment will be crucial to understand the mechanism of gene regulation mediated by chromatine modeling.
	Large Scale Protein Dynamics	The study of large conformational changes in proteins. Major challenges appear in the simulation of protein misfolding, unfolding and refolding (a key element for the understanding of prion-originated pathologies).
	Protein Association & aggregation	One of the greatest challenges is the simulation of crowded "not in the cell" protein environments. To be able to represent "in silico" the formation of the different protein complexes associated with a signaling pathway opens the door to a better understanding of cellular function and to the generation of new drugs able to interfere in protein-protein interactions.
	Supramolecular Systems	The correct representation of protein machines is still out of range of European groups using current simulation protocols and computers. The challenge will be to analyze systematically how several of these machines work e.g., ribosome, topoisomerases, polymerases.
	Medicine	Genome sequencing, massive genotyping studies are providing massive volumes of information e.g. the simulation of the determinants triggering the development of multigenic-based diseases and the prediction of secondary effects related to bad metabolism of drugs in certain segments of population, or to the interaction of drugs with macromolecules others than their original targets.
ENGINEERING	Complete Helicopter Simulation	The European helicopter industry has a strong tradition of innovation in technology and design. Computational Fluid Dynamics (CFD) based simulations of aerodynamics, aero-acoustics and coupling with dynamics of rotorcraft already play a central role and will have to be improved further in the design loop.
	Biomedical Flows	Biomedical fluid mechanics can improve healthcare in many areas, with intensive research efforts in the field of the human circulatory system, the artificial heart or heart valve prostheses, the respiratory system with nose flow and the upper and lower airways, and the human balance system. Although experiments have significantly improved the understanding in the field, numerous questions, the answers of which need a high resolution of the flow field, of the surrounding tissue, or of their interactions, require a detailed numerical analysis of the biomedical problem.
	GasTurbines & Internal Combustion Engines	Scientific challenges in gas turbines or piston engines are numerous. First, a large range of physical scales should be considered from fast chemical reaction characteristics (reaction zone thicknesses of about tens of millimeters, 10-6 s) pressure wave propagation (sound speed) up to burner scales (tens of centimeters, 10-2 s resident times) or system scales (meters for gas turbines).
	Forest Fires	The development of reliable numerical tools able to model and predict fire evolution is critically important in terms of safety and protection ("numerical fire simulator"), fire fighting and could help in real time disaster management. The social impact is very important and is concerned with land, buildings, human and animal life, agriculture, tourism and the economy.
	Green Aircraft	ACARE 2020 provides the politically agreed targets for an acceptable maximum impact of air traffic on people and the environment, while at the same time allowing the constantly increasing amount of air travel. The goals deal with a considerable reduction of exhaust gas and noise. Air traffic will increase by a factor of 3, accidents are expected to go down by 80%. Passenger expense should drop (50%) and flights become largely weather independent. The "Green Aircraft" is the answer of the airframe as well as engine manufacturing industry. However, it is only by a far more productive high quality numerical simulation and optimization capability that such a challenging development will be possible. It will be indispensable to be able to compute the real aircraft in operation, including all relevant multi-disciplinary interaction.
	Virtual Power Plant	Safe production of high quality and cost effective energy is one of the major concerns of Utilities. Several challenges must be faced, amongst which are extending the lifespan of power plants to 60 years, guaranteeing the optimum fuel use and better managing waste. These challenges demand access to Petascale machines to perform advanced simulations along with a new generation of codes and simulation platforms.

***Uit de voorgaande voorbeelden en Tabel 2.1. blijkt duidelijk dat een gebrek aan reken capaciteit de Vlaamse wetenschappelijke gemeenschap gewoon buitenspel zet in de internationale competitie voor wereldwijd excellentie-niveau.***

Een strategie waarbij de wetenschappelijke ontwikkeling gekoppeld wordt aan een beschikbare rekenkracht, met een geleidelijke evolutie naar de Petaflops schaal is bijgevolg hoogst wenselijk. Er moet evenwel rekening gehouden worden met de noodzakelijke bijdrage van computerwetenschappers, mathematici, enz. voor de optimalisatie van de software programma's gericht op massieve parallelle architectuur.

De onderzoeksteams in een bepaalde wetenschappelijke sector moeten beroep kunnen doen op computerexperts met de verantwoordelijkheid voor de optimalisatie van de computeromgeving en de ondersteuning naar het behalen van maximale performantie van de beschikbare configuratie.

### **2.3 HPC-toepassingen in Vlaanderen**

De hierna volgende informatie werd samengesteld op basis van de enqueteresultaten binnen de Vlaamse onderzoeksinstituten. Deze resultaten zijn weergegeven in Bijlage 3, op basis van de ontvangen commentaren, om een overzicht te geven van de verschillende visies en standpunten.

#### 2.3.1. Materiaalonderzoek en Nanotechnologie

Er zijn belangrijke activiteiten in Vlaanderen in deze sector, met innovatief fundamenteel en toegepast onderzoek; zie Bijlage 3a.

Courante HPC-aanwendungen worden hoofdzakelijk uitgevoerd op de lokale clusters van de onderzoeksinstituten. Ze zijn echter beperkt van omvang. Duidelijke plannen bestaan voor simulaties op grotere schaal gekoppeld aan baanbrekende ontwikkelingen, indien een hoge capaciteit HPC-infrastructuur ter beschikking zou staan.

Een belangrijke vraag naar ondersteuning op het vlak van software, numerieke aspecten, parallelisatie en visualisatie software wordt naar voor geschoven.

#### 2.3.2. Moleculaire Biologie en Biotechnologie, Genetica en Fysiologie

Alhoewel grensverleggend onderzoek uitgevoerd wordt in deze sector, (zie Bijlage 3b), is de aanwending van HPC nog niet voldoende uitgebuit, naar het voorbeeld van de evolutie op internationaal vlak; zie het voorbeeld van Finland in Figuur 3.3. Nochtans toont de enquête aan, dat er een significante vraag bestaat om baanbrekend onderzoek te verrichten met behulp van HPC.

#### 2.3.3. Engineering

Praktisch alle onderzoekscomponenten binnen de wereld van engineering zijn genoodzaakt beroep te doen op simulaties; zie Bijlage 3c.

Dit wordt ook gestaafd door het recent verslag van de Amerikaanse National Science Foundation, gericht op Simulation Based Engineering Science (SBES)<sup>7</sup>.

Een recent initiatief, FLAMAC, dient ook vermeld te worden; zie Bijlage 3a.

Alhoewel er reeds uitgebreid gebruik gemaakt wordt van bestaande clusters, is de vraag naar HPC op het niveau van 100 TeraFlops bijzonder groot.

#### 2.3.4. Moleculaire Chemie

Moleculaire en kwantumchemie zijn bijzonder afhankelijk van HPC-simulaties. Gevorderd onderzoek met behulp van beperkte bestaande middelen is ruim verspreid over de Vlaamse universiteiten; zie Bijlage 3d.

<sup>7</sup> Revolutionizing Engineering Science through Simulation (2006). A Report of the National Science Foundation. Blue Ribbon Panel on Simulation-Based Engineering Science. [http://www.nsf.gov/pubs/reports/sbes\\_final\\_report.pdf](http://www.nsf.gov/pubs/reports/sbes_final_report.pdf)

Een zeer ruime vraag naar hoogstaande HPC-infrastructuur wordt door de betrokkenen uitgedrukt en deze wetenschappelijke sector zou ongetwijfeld een belangrijke gebruiker zijn van een centrale HPC-eenheid.

Een sterke vraag van deze sector gaat uit naar ondersteuning op het vlak van software, parallelisatie, software optimalisatie en eenvoudige toegangsprocedures.

#### 2.3.5. Toegepaste Wiskunde en Mathematische Algoritmen

De ontwikkeling van nieuwe mathematische algoritmen is een noodzaak voor doeltreffende HPC-toepassingen in alle domeinen. Daarnaast zijn er ook talrijke problemen in de zuivere als in de toegepaste wiskunde die gebruik maken van grote rekenclusters, waaronder de ontwikkeling van nieuwe statistische modellen voor het analyseren van complexe gegevens, verbanden tussen groepentheorie en meetkunde (zuivere Wiskunde), machine learning algoritmen voor gedistribueerde identificatie.:

Een uitbreiding van de huidige reken- en geheugencapaciteit is zeer wenselijk.

De ondersteuning van simulatie-gericht onderzoek door mathematici is van groot belang en het is verheugend vast te stellen dat een significant deel van deze gemeenschap op onze universiteiten, reeds betrokken is bij numerieke methodologie en toepassingen; zie Bijlage 3e. Deze steun is gekenmerkt door een grote verscheidenheid en biedt de garantie dat een bredere ontwikkeling van HPC op de gewenste steun van de wiskundige gemeenschap zal kunnen rekenen.

#### 2.3.6. Astrofysica en Kosmologie

De belangrijkste onderzoeksgebieden in Vlaanderen die grote rekencapaciteiten vereisen, zijn gericht op de studie van de structuur, evolutie en dynamica van sterrenstelsels; modelleren van astrofysische plasma's; structuur, evolutie en samenstelling van sterren; simulaties voor wat betreft de beginfase van het ontstaan van het universum; zie Bijlage 3f.

Wat betreft ondersteuning, lijken de sterrenkundigen er vrij gerust in te zijn dat zij grotere apparatuur aankunnen.

#### 2.2.7. Subatomaire Fysica

Hiermee wordt verwezen naar de studie van de allerkleinste elementen in de opbouw van de materie, waarbij het gebied zich spreidt over het oneindig kleine (deeltjesfysica, kernfysica, atoomfysica) tot de schaal van het oneindig grote (kosmologie, galaxieën,..), waarbij zowel theoretici als experimentatoren veelvuldig beroep doen op HPC; zie Bijlage 3g.

De complexiteit van de fundamentele wetten van de materie en de theorieën over de oorsprong van het heelal, kunnen slechts met behulp van grootschalige HPC-simulaties benaderd worden.

#### 2.3.8. Turbulentie en Verbranding

De directe simulatie van turbulentie is doorslaggevend voor talrijke milieu- en industriële toepassingen. Dit geldt o.a. voor voorspellingen van pollutieverspreiding, rookgassen bij branden, geluidsbronnen van machines en beheersing van uitstootgassen van branders. De HPC-vereisten voor directe simulaties van turbulentie zijn onbegrensd en worden ook gedeeltelijk opgenomen onder de rubriek Engineering, punt 2.3.3. Dit wordt ook bevestigd door de gestelde behoeften van de Vlaamse onderzoeksgroepen; zie Bijlage 3h.

#### 2.3.9. Milieustudies

De enquête werd beperkt tot de activiteiten in het VITO van ondermeer "airborne en spaceborne remote sensing" ter ondersteuning van het rampenbeleid, het 3D fotogrammetrisch verwerken van geografische data via digitale weg, en toepassingen in zowel milieustudies als de veiligheidsproblematiek. Ook de geïntegreerde

omgevingstudies komen aan bod met ondermeer de modellering van de luchtkwaliteit, de meteorologie, de omgevingsrisico's, het ondersteunen van het beleid in verband met rivieren, kustzones, steden en regio's, en ecosystemen. Wegens de complexiteit, de grote verscheidenheid en de omvang van de materie is een uitbreiding van de huidige reken-en geheugencapaciteit wenselijk. Zie Bijlage 3i.

#### 2.3.10. Fundamentele Fysica

Het belang en de noodzaak van numerieke simulaties zullen de in de komende jaren nog drastisch toenemen. Het space exploration programma Aurora van ESA, b.v., staat of valt met degelijke ruimteweersvoorspellingen: realistische simulaties zullen een rekenkrachtcapaciteit van 100-200 Tflops ( $50 \times$  meer dan thans beschikbaar) vereisen, om met buitenlandse onderzoeksgroepen in competitie te kunnen treden. De quantumfysica van gecondenseerde materie (o.m. nanostructuren) wordt momenteel beperkt door de beschikbare reken capaciteit (grootte van het systeem dat gemodelleerd kan worden). Ook de 'traditionele' theoretische fysica plant om de numerieke component van het onderzoek meer uit te bouwen. De betrokken groepen ontwikkelen doorgaans hun computercodes zelf en maken ze parallel, en/of hebben samenwerking met wiskundigen/informatici voor ondersteuning op het gebied van algoritmen en software, vooral voor intensieve parallelisatie. Zie Bijlage 3j.

#### 2.3.11. Taalkunde

Alhoewel geen enquêtegegevens voorhanden zijn, vormt taalkunde een nieuw gebied waar nieuwe onderzoeksdomeinen behoefte hebben aan HPC.

Hierbij kan vermeld worden: machineren van natuurlijke taal; studie van algoritmen om taalkundige informatie automatisch uit een tekst te extraheren; gebruik van natuurlijke taalverwerkingsprogramma's om toe te passen op "text mining".

### **3. Overzicht van bestaande HPC-infrastructuur, wereldwijd, in Europa en in Vlaanderen**

Dit overzicht heeft tot doel de werking en financiële basis van representatieve HPC-centra in kaart te brengen. Vervolgens worden in §3.2.2. drie centra uit naburige landen besproken als referentie voor wat in Vlaanderen overwogen zou kunnen worden.

#### **3.1 HPC op wereldvlak**

De rangschikking van de 500 grootste HPC-installaties wereldwijd worden jaarlijks gepubliceerd; <http://www.top500.org/>.

De tabel 3.1 geeft de lijst aan van de top-10 installaties wereldwijd, geactualiseerd in November 2006. De laatste kolom verwijst naar de gemiddelde rekensnelheid uitgedrukt in aantal TeraFlops of Tflops, dit is in duizend miljard bewerkingen per seconde ( $10^{12}$  Flops).

Wij wensen hierbij de aandacht van de lezer te vestigen op de bijzonder snelle evolutie van de HPC capaciteit wereldwijd, op basis van een vergelijking met de nieuwe rangschikking verschenen in November 2007<sup>8</sup>. Deze laatste is te vinden in Bijlage 4.

<sup>8</sup> Bij het ter perse gaan, is de nieuwe rangschikking van november 2007 verschenen. Bijlage 4 bevat de nieuwe lijst van de top-10 HPC-installaties wereldwijd en de nieuwe rangschikking van de belangrijkste HPC-installaties in Europa. De commentaar van de uitgever wordt ook samengevat in deze Bijlage 4.

### 3.1.1. De top-10 HPC-installaties – November 2006

Tabel 3.1: Rangschikking van de top-10 HPC-installaties wereldwijd (November 2006)

Rang	Plaats	Computer	Aantal Processoren	Snelheid in Teraflops
1	DOE/NNSA/LLNL United States	BlueGene/L - eServer Blue Gene Solution IBM	<b>131 072</b>	<b>280</b>
2	NNSA/Sandia National Laboratories United States	Red Storm - Sandia/ Cray Red Storm, Opteron 2.4 GHz dual core Cray Inc.	<b>26 544</b>	<b>101</b>
3	IBM Thomas J. Watson Research Center United States	BGW - eServer Blue Gene Solution IBM	<b>40 960</b>	<b>91</b>
4	DOE/NNSA/LLNL United States	ASC Purple - eServer pSeries p5 575 1.9 GHz IBM	<b>12 208</b>	<b>76</b>
5	Barcelona Supercomputing Center Spain	MareNostrum - BladeCenter JS21 Cluster, PPC 970, 2.3 GHz, Myrinet IBM	<b>10 240</b>	<b>63</b>
6	NNSA/Sandia National Laboratories United States	Thunderbird - PowerEdge 1850, 3.6 GHz, Infiniband Dell	<b>9024</b>	<b>53</b>
7	Commissariat a l'Energie Atomique (CEA) France	Tera-10 - NovaScale 5160, Itanium2 1.6 GHz, Bull SA	<b>9968</b>	<b>53</b>
8	NASA/Ames Research Center/NAS United States	Columbia - SGI Altix 1.5 GHz, Voltaire Infiniband SGI	<b>10 160</b>	<b>52</b>
9	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME Grid Cluster - Sun Fire x4600 Cluster, Opteron 2.4/2.6 GHz NEC/Sun	<b>11 088</b>	<b>47</b>
10	Oak Ridge National Laboratory United States	Jaguar - Cray XT3, 2.6 GHz dual Core Cray Inc.	<b>10 424</b>	<b>43</b>

### 3.2 HPC in Europa

Een overzicht van de belangrijkste Europese HPC-installaties biedt een nuttige achtergrond voor de evaluatie van de Belgische en Vlaamse positionering op HPC-vlak. Er wordt vervolgens ingegaan op drie representatieve HPC-centra in Finland, Zwitserland en Nederland, die als referentie kunnen dienen voor een Vlaamse strategie.

#### 3.2.1. Top HPC-systemen in Europa (November 2006)

Tabel 3.2 toont de lijst van de belangrijkste Europese HPC-installaties, geactualiseerd in November 2006.

Binnen Europa, komen na Spanje de omliggende landen in aanmerking met hoogstaande installaties. De eerste kolom geeft de plaats aan binnen de 500-lijst, terwijl de 4de kolom het aantal processoren aangeeft en de laatste kolom is de gemiddelde rekensnelheid in Teraflops ( $10^{12}$  bewerkingen/s).

België komt in deze lijst voor met één enkele HPC-installatie, in een niet verder vernoemde bank, op plaats 162! Verder in de lijst is er geen enkele vermelding van installaties in België of Vlaanderen binnen de eerste 500 HPC-installaties.

Tabel 3.2.: Rangschikking van de belangrijkste HPC-installaties in Europa (November 2006)

Nr	Plaats	Land	Aantal Processoren	Snelheid in Teraflops
5	Barcelona HPC	Spain	10 240	63
7	CEA	France	9968	53
13	Forschungszentrum Juelich (FZJ)	Germany	16 384	37
15	Atomic Weapons Establishment	UK	7812	32,5
17	ASTRON/University Groningen	Netherlands	12 288	27
18	Leibniz Rechenzentrum	Germany	4096	24
20	Cambridge University	United Kingdom	2340	18
22	Ecole Polytechnique Federale de Lausanne	Switzerland	8192	18
34	CeSViMa - BSC	Spain	2408	14
38	ECMWF	United Kingdom	2240	13,9
43	HPCx	United Kingdom	2560	12,9
49	Technische Universitaet Dresden	Germany	2048	11,9
50	Scientific Supercomputing Center Karlsruhe	Germany	3000	11,9
60	BMW Sauber F1	Switzerland	1024	9,6
61	EDF R&D	France	4096	9,4
72	HWW/Universitaet Stuttgart	Germany	576	8,9
80	CSC (Center for Scientific Computing)	Finland	2048	8,2
84	SCS S.r.l.	Italy	1024	8
89	Statoil	Norway	1024	7,5
94	Swiss Scientific Computing Center (CSCS)	Switzerland	1664	7,2
106	Technische Universitaet Dresden	Germany	2584	6,2
112	Norwegian University of Science and Technology	Norway	960	6
119	Forschungszentrum Juelich (FZJ)	Germany	1312	5,6
124	Universität Erlangen	Germany	728	5,4
128	KTH - Royal Institute of Technology	Sweden	886	5
135	Adam Opel AG	Germany	720	4,8
159	Max-Planck-Gesellschaft MPI/IPP	Germany	688	4,6
162	Bank (K)	Belgium	1300	4,5

### 3.2.2. HPC-centra in vergelijkbare landen

Uit de bestaande HPC-centra werden drie centra geselecteerd als referentie  
 CSC in Finland  
 CSCS in Zwitserland  
 SARA in Nederland

Om een idee te krijgen over de omvang van de relevante investeringen kan een vergelijking met de omvang en/of de budgetten van de drie referentielanden gemaakt worden.

#### **CSC in Finland**

Het CSC HPC-centrum in Finland is nr 80 in Tabel 3.2.:(<http://www.csc.fi>). CSC maakt deel uit van het Finse nationaal researchsysteem and levert HPC-diensten voor onderzoek en onderwijs aan universiteiten, onderzoeksinstituten en bedrijven, die als klanten beschouwd worden. CSC levert diensten in software ontwikkeling, GRID toepassingen, IT infrastructures en netwerken. De geleverde diensten dekken alle stappen van de ontwikkeling van een computermodel tot de selectie van numerieke methodes, gebruik van software en van databanken, en dit op alle gebieden van engineering tot biowetenschap.

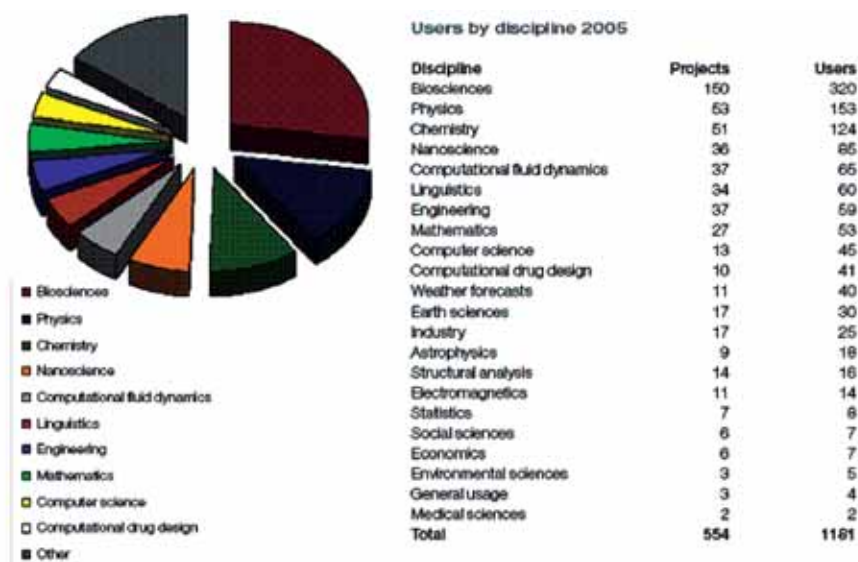
In 2005 kwamen de grootste HPC-toepassingen uit de disciplines nano-wetenschappen, fysica, chemie, biowetenschappen, Computational Fluid Dynamics (CFD) en astrofysica.

Het CSC opereert als een vzw met de organisatie van een naamloze vennootschap en meldt de volgende cijfers voor het boekjaar 2005. De omzet was 15,1 M€ waarvan 7,1 M€ voor personeel en een operationele winst van 95 000 €; de R&D kosten bedragen 6,7% van de omzet.

CSC heeft een staf van 140 medewerkers, die dan wel gedeeltelijk ook voor andere taken (buiten HPC) wordt ingezet (zij spelen bvb ook een belangrijke rol in de Finse telecommunicatie).

Figuur 3.3. uit het CSC jaarverslag 2005<sup>9</sup> toont de indeling van de HPC-gebruikers tussen de verschillende wetenschappelijke disciplines in Finland.

Figuur 3.3: Indeling van de HPC-gebruikers tussen de verschillende wetenschappelijke disciplines in Finland.



### CSCS in Zwitserland

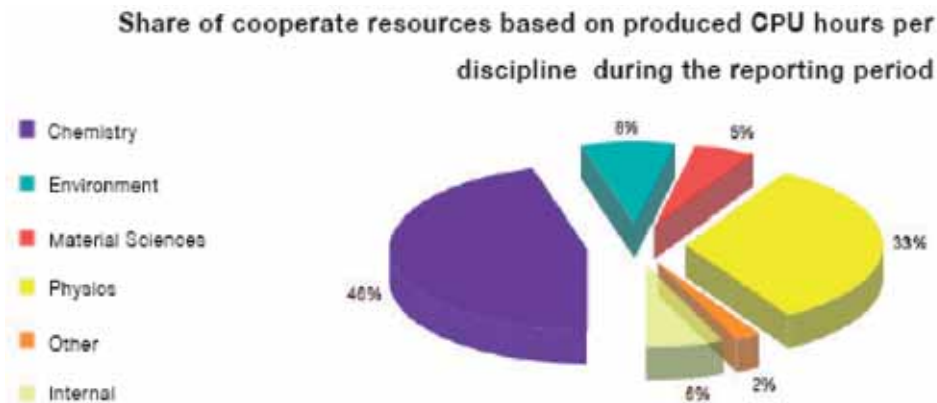
CSCS is een nationaal HPC-centrum en is nr 94 in Tabel 3.2. Het CSCS komt slechts op de derde plaats in Zwitserland, na de meer recente HPC-installaties van de Ecole Polytechnique Fédérale de Lausanne (EPFL), die als een regionale installatie beschouwd kan worden, en de supercomputer van het F1 bedrijf BMW-Sauber. Het CSCS heeft een jaarlijks budget (2005) van meer dan 11 M€: grosso modo de helft voor (vervangings) apparatuur en de andere helft voor werking en een personeelsstaf van circa 30 medewerkers.

De distributie van de onderzoeksdomeinen op basis van het aantal gebruikte CPU uren in 2005 is weergegeven in Figuur 3.4, uit het CSCS jaarverslag 2005<sup>10</sup>

<sup>9</sup> CSC, The Finnish IT Center For Science | Annual Report 2005. <http://www.csc.fi/english/csc/publications/reports>

<sup>10</sup> CSCS Swiss National Supercomputing Centre. Annual Report 2005. <http://www.cscs.ch/publications/annualr/annualreport05.pdf>.

Figuur 3.4: Distributie van de HPC-gebruikers in 2005 volgens de disciplines in Zwitserland. CSCS jaarverslag 2005.



**SARA in Nederland: (<http://www.sara.nl>)**

SARA Reken- en Netwerkdiensten is een geavanceerd ICT-diensten centrum dat sinds ruim 30 jaar een compleet pakket aan diensten en producten levert op het gebied van high performance computing & visualisatie, high performance netwerking en infrastructuurdiensten. Onder de klanten van SARA bevinden zich wetenschappelijke en educatieve instellingen, overheidsinstellingen en zakelijke bedrijven.

Sinds 1985 verzorgt SARA de huisvesting van de nationale supercomputer-dienstverlening. In opdracht van de stichting Nationale Computerfaciliteiten (NCF) beheert SARA de Nationale Supercomputer (Teras/Aster). Via NCF maken Nederlandse onderzoekers gebruik van de reken capaciteit van deze systemen. Ook beheert SARA het Nationaal Rekencluster. Voorts levert SARA aan (semi-)overheidsinstellingen en bedrijven diensten en producten op het gebied van systeembeheer, netwerken, informatie en expertise, die aansluiten bij de kennis en ervaring in de genoemde 'high-end activiteiten'. De levering van deze diensten en producten wordt gekenmerkt door een aantrekkelijke prijsprestatie verhouding.

SARA is begonnen als een HPC-centrum, maar ook hier zijn in de loop van de jaren tal van andere taken bijgekomen. Zo is SARA wereldwijd een zeer belangrijk internet-knooppunt; SARA is één van de ankerpunten waar binnenkort de Large Hadron Collider (LHC)-gegevens vanuit CERN in Genève zullen binnenkomen en verwerkt worden. Er is ook een wereldtop-niveau 3D-visualisatiecentrum ontwikkeld door SARA, alsmede een service voor massale data-opslag en co-locatie van apparatuur voor o.a. industrie. SARA beschikt over een waaier aan clusters, waaronder het Nationaal Cluster (LISA) met een piekprestatie van 8,5 TeraFlops<sup>11</sup>.

Organisatie:

De aanvraag voor rekentijd en voor ondersteuning door specialisten bij het optimaliseren van wetenschappelijke software wordt beheerd door de stichting Nationale Computerfaciliteiten NCF. Deze onafhankelijke stichting (NCF) wordt gefinancierd door de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO). NCF werd opgericht in 1990 met het doel om strategisch wetenschappelijk onderzoek mogelijk te maken en te stimuleren door geavanceerde ICT-faciliteiten beschikbaar te stellen voor excellente onderzoekers. In de organisatie van NWO wordt NCF als een apart wetenschapsgebied gezien, op gelijke voet met gebieden als Exacte Wetenschappen.

SARA draagt er in een belangrijke mate toe bij dat Nederland wereldwijd op de computationele kaart staat. Voor de eigenlijke HPC heeft SARA trouwens onlangs een

<sup>11</sup> Zie [http://www.nwo.nl/nwohome.nsf/pages/NWOA\\_4YAHHF](http://www.nwo.nl/nwohome.nsf/pages/NWOA_4YAHHF) voor een beschrijving.



tweede locatie geopend in Almere, omdat de ruimtenoden de mogelijkheden van de eerste site in Amsterdam ruimschoots te boven ging. Dit alles suggereert dat SARA en Nederland als geheel (met zijn 16 miljoen inwoners) iets te groot is om als adequaat model te dienen.

Het feit dat SARA nu al spontaan in twee stukken gesplitst is, suggereert dat Vlaanderen (met 6 miljoen inwoners) een adequate grootte heeft om een centrale HPC-eenheid uit te bouwen.

Het budget van SARA wordt niet vermeld in het 2005 jaarverslag<sup>12</sup>, maar SARA beschikt over een staf die in 2005 100 medewerkers bedroeg, met een geplande uitbreiding naar 150 in 2009.

### **3.3. HPC in Vlaanderen**

De lokale HPC-clusters van de UAntwerpen, UGent en KULeuven worden kort beschreven. Deze installaties, weliswaar van beperkte omvang op internationaal vlak, spelen een belangrijke rol bij de ondersteuning van vooruitstrevend onderzoek.

#### 3.3.1. De HPC-cluster van UAntwerpen

De UAntwerpen cluster kan opgesplitst worden in rekennodes en "back-end" (het management gedeelte, de servers):

rekennodes: 256 Sun Fire V20z: dual AMD Opteron 250, 73 GB lokale schijf. Hiervan zijn 192 nodes uitgerust met 4 GB intern geheugen en onderling verbonden via Gigabit Ethernet. De overige 64 nodes zijn uitgerust met 8 GB intern geheugen en onderling verbonden via Myrinet. Het besturingssysteem is Suse.

back-end: 2 Sun Fire V440: 4 UltraSparc IIIi en 16 GB intern geheugen. De centrale opslag gebeurt op 2 Sun StorEdge 3510 FC Array, samen goed voor bijna 6 TB. Het besturingssysteem is Solaris.

#### 3.3.2. De HPC-cluster van KULeuven

De HPC-infrastructuur (een 'computer cluster') bestond in 2005 uit 177 'rekenknopen' met telkens minstens 2 processoren (AMD Opteron) en 2 GB RAM geheugen. Een van de 177 knopen bevatte 8 processoren, een andere knoop was een zgn. 'dual-core 4-way', en nog drie andere knopen waren 'dual core dual servers'. De grote meerderheid van de nieuwe nodes werden onderling verbonden met het zeer snelle Infiniband interconnectie netwerk, waardoor sommige berekeningen die uitgevoerd worden op meerdere nodes veel sneller uitgevoerd kunnen worden. Bovendien bevatten acht knopen 16 GB RAM geheugen zodat het volledige systeem in totaal 372 rekenprocessoren bezat en meer dan 500 GB RAM geheugen. Dit HPC-systeem had een theoretische piek-performantie van 1,8 Teraflops in dubbele precisie, 3,6 Teraflops in 'single precision' en een totale opslagcapaciteit van meer dan 30 TB.

In September 2006 werd de cluster uitgebreid met 120 extra rekennodes. Het betreft zgn. 'two way dual core AMD Opteron275' nodes (dus in praktijk met 4 'compute cores' of rekenkernen per node) met elk 4 GB RAM geheugen en onderling verbonden met het zeer snelle 'Infiniband interconnect'. De huidige HPC-infrastructuur bestaat dus uit 297 nodes met in totaal 852 AMD Opteron cores waarvan er 752 via een snelle Infiniband switch met elkaar verbonden zijn. Deze cluster heeft 1,14 TB RAM geheugen en een theoretisch piek-performantie van ongeveer 4 Teraflops (3912 Gflops)!

#### 3.3.3. De HPC-cluster van UGent

In de vakgroep Informatietechnologie van de UGent is er een cluster met 222 CPU's (vervat in een 80-tal working nodes, waarvan een aantal met 2 CPU's en een aantal met

<sup>12</sup> [http://www.sara.nl/aboutsara/pdf/SARA\\_jaarverslag\\_2005.pdf](http://www.sara.nl/aboutsara/pdf/SARA_jaarverslag_2005.pdf)

4 CPU's) geïnstalleerd die verbonden is aan BEGrid (meer info te vinden op <http://gridmon.atlantis.UGent.be/ganglia/>). Dit was een investering van de Vlaamse overheid en de Provincie O-Vlaanderen. Er komen dit jaar nog een aantal nodes bij zodat de installatie ongeveer 300 CPU's zal bevatten.

#### 4. De Europese context

Talrijke initiatieven werden gelanceerd tijdens de twee laatste Kader Programma's van de Europese Commissie, gericht op HPC-structuren en behoeften. Alhoewel HPC-methodologie en simulatie niet aan bod kwamen in FP6 en ook niet als prioriteit beschouwd worden binnen FP7, wordt echter een grote aandacht besteed aan de uitbouw van een Europese infrastructuur voor grootschalige HPC, tot op de 'Petaflops' ( $10^{15}$  bewerkingen/s) schaal.

Een brede waaier aan initiatieven op Europees vlak worden ondersteund ter bevordering van HPC-communicatie en GRID computing. Enkele van de vooraanstaande initiatieven zijn DEISA; (Distributed European Infrastructure for Supercomputing Applications<sup>13</sup>); EGEE, (Enabling Grids for E-science).

##### 4.1 DEISA

DEISA is een consortium van de belangrijkste nationale supercomputerinstallaties met een Europese 'schaal' en dimensie. DEISA werd ingevoerd tijdens FP6 als een onderzoeksinfrastructuur om wetenschappelijke ontdekkingen mogelijk te maken over een breed spectrum van wetenschap en technologie, door een versterking van de Europese capaciteit in HPC.

Dit wordt in de hand gewerkt door een diepgaande integratie van nationale 'high-end' platformen, met elkaar verbonden door een hoge snelheid netwerk en ondersteund door innovatieve systemen en GRID technologie.

DEISA is een structuur 'boven' de nationale capaciteit die rekenniveaus toelaat op een verhoogde schaal, door de koppeling van meerdere installaties.

Het DEISA netwerk telt de volgende 11 leden:

1. Institut du Développement et des Ressources en Informatique Scientifique (IDRIS-CNRS), Orsay, France.
2. Forschungszentrum Juelich (FZJ), Zentralinstitut fuer Angewandte Mathematik (ZAM), Juelich, Germany.
3. Rechenzentrum Garching of the Max Planck Society (RZG), Garching, Germany.
4. Consorzio Interuniversitario (CINECA), Bologna, Italy.
5. Edinburgh Parallel Computing Centre (EPCC), Edinburgh, UK
6. Finnish Information Technology Centre for Science (CSC), Espoo, Finland.
7. SARA Computing and Networking Services, Amsterdam, The Netherlands.
8. European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK.
9. Leibniz Computing Centre of the Bavarian Academy of Sciences and Humanities,
10. Barcelona Supercomputing Center (BSC), Barcelona, Spain
11. High Performance Computing Center Stuttgart, Stuttgart, Germany.

##### 4.2 EGEE (Enabling Grids for E-science)<sup>14</sup>

EGEE is een GRID infrastructuur voor 'e-Science' die de mogelijkheid geeft aan ingenieurs en wetenschappers van meer dan 90 instellingen permanent toegang te krijgen tot een gewenste reken capaciteit.

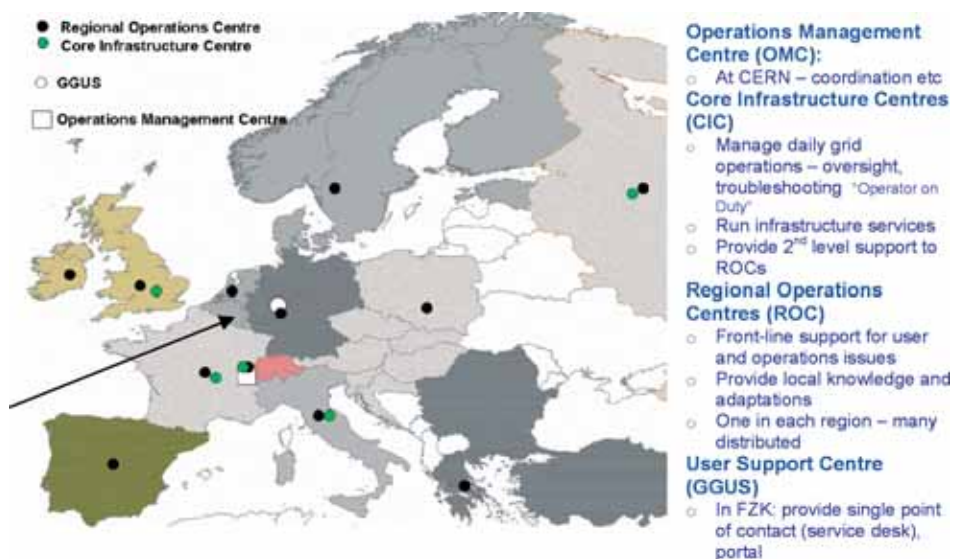
De EGEE grid beschikt over 20 000 CPU, beschikbaar 24/24 uur met 20 000 gelijklopende jobs, verdeeld over 5 Petabytes geheugen. De EGEE infrastructuur is ideaal

<sup>13</sup> <https://www.deisa.org/>

<sup>14</sup> Erwin Laure (2006). EGEE Production Grid Infrastructures Enabling eScience. INFOS-RI-508833, CESNET conference. <http://www.eu-egee.org>

voor wetenschappelijk onderzoek, waar de traditionele IT communicatie ontoereikend is.

De volgende figuur toont de regionale centra verbonden aan de EGEE operationele structuur.



**Wat hier bijzonder opvalt, is dat uit alle omringende landen, België het enige land is zonder HPC regionaal operationeel centrum.**

#### 4.3 De Europese HPC-visie

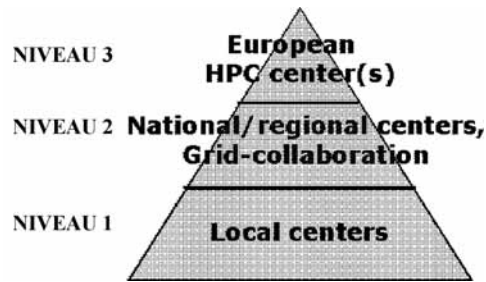
De Europese visie berust op een structuur op drie niveaus<sup>15</sup>. Deze visie is in zeer ruime mate bekrachtigd door de HPC-organisatie in alle Europese landen.

Deze piramidale structuur bevat drie niveaus:

- NIVEAU 1: lokale rekencentra op het niveau van lokale universiteiten, instituten of onderzoekslaboratoria.
- NIVEAU 2: het intermediair Niveau is gevormd door nationale of regionale rekencentra met voldoende reken capaciteit om de HPC-gebruikers toegang te geven tot een internationaal niveau, en een instapticket te geven tot de rekencentra van het Europees (Niveau 1) niveau.
- NIVEAU 3: een HPC-capaciteit op Europees niveau bestemd om de hoogst mogelijke reken capaciteit ter beschikking te stellen van de top researchgroepen in Europa. Dit moet wetenschappelijke toepassingen mogelijk maken die de nationale HPC-capaciteit overtreffen. De Europese doelstelling is het Petaflops (duizend Teraflops) niveau te bereiken tegen 2009.

Deze Europese visie benadrukt het belang van de grid-interconnectie tussen de verschillende niveaus van de piramide, en een eventuele opening naar een sterke integratie.

<sup>15</sup> P.J.C. Aerts, K. Koski, F. Lozano (2007). A sustainable High Performance Computing Ecosystem for Europe. HPC in Europe Taskforce (HET) <http://www.hpcineuropetaskforce.eu/>



Omgezet naar rekencapaciteit, kan de volgende hiërarchie als referentie genomen worden.

	Rekencapaciteit	Aantal Processoren (Cores)
NIVEAU 3	500-1000 Teraflops	100 000-500 000
NIVEAU 2	0-200 Teraflops	10 000-50 000
NIVEAU 1	1-25 Teraflops	250-5 000

De visie voor Vlaanderen die hier naar voren gebracht wordt, situeert zich binnen dit schema.

## 5. HPC-behoeften in Vlaanderen

De enquête van de HPC-werkgroep van de KVAB toont aan dat de computationele wetenschappen zich in Vlaanderen moeten behelpen met een infrastructuur die duidelijk minder sterk ontwikkeld is dan in veel andere, vergelijkbare landen of regio's.

Er zijn weliswaar een aantal centra uitgebouwd aan de verschillende universiteiten, maar er wordt op gewezen dat deze centra vele malen minder krachtig zijn dan doorsnee buitenlandse faciliteiten.

De vergelijking tussen Vlaanderen (6 miljoen inwoners) en enkele (min of meer) analoge landen zoals Nederland (16 miljoen), Zwitserland (7 miljoen) en Finland (5 miljoen) is zeer sprekend. In elk van die drie landen is er een centrale computerfaciliteit uitgebouwd, die sterker tot veel sterker is dan in Vlaanderen, waar de grootste cluster zich momenteel bevindt aan de KULeuven. Toch kunnen de drie bedoelde centra, SARA in Nederland, CSCS in Zwitserland, en CSC in Finland in een internationale context niet echt groot genoemd worden: zij zijn "medium range HPC", van het 'Niveau 2' van het Europees model (sectie 4.3.). Tegelijk is de globale computationele infrastructuur, buiten de nationale centra, in elk van die drie landen, ook duidelijk sterker uitgebouwd dan in Vlaanderen.

De oprichting van een Vlaams centraal computersysteem zal op zich zelf de ICT-nood niet volledig kunnen opvangen en kanaliseren, en zullen de huidige, bestaande, kleinere faciliteiten zeker niet overbodig maken. In Zwitserland bijvoorbeeld, waar nochtans ook heel wat kleinere faciliteiten actief zijn, wijst een audit-commissie in 2005 op een te zwakke uitbouw van het Zwitserse HPC-hinterland, waardoor de centrale faciliteit dreigt te verstikken (to get cluttered) door een overvloed aan te veel kleinere jobs. Vaak gebeurt het ook dat, met het oog op sommige zeer specifieke taken, een aantal kleinere, maar speciaal daartoe ontworpen (dedicated) computerarchitecturen een oplossing kunnen brengen voor problemen die niet of bezwaarlijk centraal kunnen aangepakt worden. (er zijn voorbeelden in Nederland, VK, Zwitserland,...). Met andere woorden, één geïsoleerde grote HPC-installatie functioneert nooit in een computerwoestijn.

Het betekent wel dat Vlaanderen momenteel een zeer ernstige HPC achterstand heeft opgelopen, die alleen door nieuwe, en zeer significante investeringen kan worden geredieerd. Een van de redenen waarom de vereiste inspanning nu, in 2007, zo aanzienlijk is, is dat de diverse overheden (Belgisch, Vlaams of universitair) vroegere aanvragen in de loop van de vorige decennia steeds halfslachtig en minimalistisch hebben afgehandeld. De hoger vermelde centra in de drie referentielanden functioneren daarentegen alle drie reeds meer dan tien of twintig jaar, en hebben intussen een vrij grote graad van maturiteit bereikt.

Er moet rekening mee gehouden worden dat HPC eigenlijk een kip-en-ei karakter heeft. Momenteel zijn er in Vlaanderen weinig of geen HPC-projecten in geologische prospectie, in elementaire-deeltjesfysica, etc ...wellicht ten dele omdat er toch geen HPC-faciliteit beschikbaar is. Indien van in het begin de nieuwe structuur draait rond bijv. biofysische toepassingen, dan zal het potentieel voor die andere disciplines niet echt vooruit geholpen worden. Na verloop van tijd zullen ongetwijfeld ook projecten opduiken uit andere richtingen, zoals humane en biomedische wetenschappen.

Tegelijk wordt, nagenoeg door alle geraadpleegde onderzoeksgroepen, gevraagd naar een sterke ondersteuning en betere toegankelijkheid van de bestaande systemen, in de vorm van hulp bij het paralleliseren, hulp bij de implementatie van de software, aanpassing van de wachtlijnen, verkorting van de wachttijden, optimalisatie van de architectuur.

Het staat vast, zoals bevestigd ook door de organisatie van de HPC-centra in het buitenland, dat een regionaal HPC-centrum een sterke rol te vervullen heeft bij de ondersteuning van nieuw baanbrekend HPC-gericht onderzoek. Dit vereist, naast de hardware infrastructuur [met inbegrip van efficiënte storage (grote databases) en geheugen, snelle I/O, hoge-snelheid netwerken], een ploeg van experts in numerieke optimaliseren van fysische modellen, ontwikkeling van nieuwe parallele toepassingen, analyse van grote stambomen, software engineering en visualisatie.

### **5.1 Voor een HPC-infrastructuur in Vlaanderen**

**Een synthese van de Europese toestand met de collectieve informatie uit de enquête binnen de Vlaamse wetenschappelijke wereld, pleit duidelijk voor het invullen van een HPC-capaciteit op een schaal van Niveau 2 in Vlaanderen.**

**Op deze basis, zou een HPC-centrum in Vlaanderen een omvang moeten hebben van minimum 10 000 processoren (cores).**

**Dit centrum zou rechtstreeks door de Vlaamse overheid beheerd en gesubsidieerd worden.**

**Een regionaal Vlaams HPC-centrum moet berusten op vier essentiële componenten:**

1. Er is nood aan een versterking van de bestaande computerparken binnen de universiteiten die het Niveau 1 vormen waarop de piramidale structuur moet berusten.
  - a. Elke computationeel intensieve eenheid heeft nood aan een lokale, (min of meer) kleinschalige faciliteit waar het normale werk kan gebeuren. Het voordeel van deze lokale faciliteiten is de flexibiliteit van architectuur en queuing systemen, de beheerbaarheid, de sociale controle, de mogelijkheid om te leren en te experimenteren met weinig risico op rampscenario's.
  - b. Het behoud en de uitbreiding van deze lokale parken is vitaal voor het overleven van computationele groepen. De verwervingskanalen voor deze basis-computer apparatuur blijven best – zoals dat nu het geval is - het FWO, de universitaire onderzoeksraden, het IWT, etc.
2. Met deze 'Niveau 1' basis, blijven echter veel mogelijkheden volledig buiten het bereik van de Vlaamse onderzoekers, namelijk problemen waarvoor massief parallel werk nodig is, met een computercapaciteit die ongeveer één tot twee grootte-orden sterker is dan de lokale, kleinschalige machines van Niveau 1. Om dit potentieel ook voor Vlaamse onderzoekers te openen, wordt best een nieuw HPC-centrum opgericht op Vlaams niveau.

3. Naast het uitvoeren van de eigenlijke grootschalige HPC-projecten moet het nieuwe centrum een belangrijke rol vervullen ter ondersteuning van de gebruikers.
  - a. In de eerste plaats, moet er aan de gebruikers (klein- of grootschalig) hulp geboden worden bij parallelisering, implementatieproblemen, software engineering, pre- en postprocessing, ..., al dan niet tegen betaling.
  - b. De vereiste ondersteuning, wat betreft personeel, voor de operatie en het onderhoud van een grootschalig HPC-centrum vereist een centraal beheer door een onafhankelijk orgaan. Een verdeling van de verantwoordelijkheid over bepaalde delen van het systeem, kan alleen maar zorgen voor bijkomende technische problemen bij het verzekeren van de uniformiteit en van een coherente werking ervan.
  - c. Het is van primordiaal belang de toegangsmodaliteiten tot dit centrum zeer goed te regelen. In elk geval moet er een technische screening komen om te verhinderen dat er jobs gedraaid worden die in feite beter op de lokale machines lopen. Verder moet er een vlotte toegankelijkheid gewaarborgd worden voor de echt grote projecten, maar die tegelijk een monopolisering van de faciliteit verhindert door een te beperkt aantal groepen.
  - d. Ook kan het centrum fungeren als een ontmoetingsplaats en discussieforum voor interdisciplinair onderzoek, gekatalyseerd door computationele gelijklopendheid.
  - e. Er kan ook gedacht worden aan een adviesfunctie voor de aankoop en de werking van de machines in de kleinschalige, lokale faciliteiten. Wat onrechtstreeks kan leiden tot een betere besteding van de huidige middelen.
4. Het HPC-centrum moet verbonden worden met alle 'Niveau 1', universitaire HPC-installaties met hoge snelheid netwerken. Dit regionaal netwerk, moet aangesloten worden in een breder Europees netwerk. Hiermee zal het regionaal centrum kunnen fungeren als ankerpunt om deel te nemen aan ruimere samenwerkingsverbanden, zoals DEISA, EGEE en andere GRID georiënteerde toepassingen, waarin met behulp van snelle communicatie tussen een aantal bestaande HPC-centra (tijdelijk) een zeer grote Europese supercomputer tot stand gebracht kan worden.

## ***5.2 Budgettaire en structurele implicaties***

**Rekening houdend met de buitenlandse voorbeelden stelt de KVAB HPC-Werkgroep een initiële investering van 10 M€ voor, en een personeelsbezetting die op kruissnelheid kan oplopen tot een paar tientallen mensen.**

### **Hardware en investeringskosten:**

- De gemiddelde kost per core is ongeveer 500 € en een installatie van rond tienduizend cores met inbegrip van de netwerkkosten, **zou kunnen oplopen tot 5 M€ als initiële investering.**

**De investering in algemene infrastructuur, met inbegrip van de gebouwen en kantoren wordt geraamd op 5 M€.**

### **Afschrijvingskosten:**

Met een afschrijving van de HPC-hardware over 3 jaar, is een jaarlijks budget te voorzien van de orde van 2 M€.

### **Werking en personeelskosten:**

De kosten voor een staf van een vijftiental medewerkers, kunnen oplopen tot 1,5 M€/jaar. Hierbij moeten kosten voorzien worden voor onderhoud, elektriciteitsverbruik, software en algemene werking van een centrum met een initiële staf van een 15-tal medewerkers: schatting ongeveer 1,5 M€.

**Een jaarlijks budget van 5 M€ is bijgevolg te voorzien.**

Het is essentieel dat dit personeel een permanente functie heeft en rechtstreeks door de Vlaamse overheid betaald wordt, en dat het niet op het budget van de universiteiten komt. Het gaat eigenlijk om een service die de Vlaamse overheid beschikbaar zou stellen, onafhankelijk van de evaluaties en de kwaliteit van de projecten. In die zin is het een gok op de toekomst, maar het potentieel van ICT in het wetenschappelijk onderzoek zou hier voldoende garanties moeten bieden. Tenslotte, moet het centrale HPC-centrum ook uitgebouwd worden tot een voor de Vlaamse bedrijven nuttig en onmisbaar instrument.

Een soepele toegang tot CPU aanvragen moet voorzien worden door een eenvoudige selectieprocedure en minimale bureaucratie.

#### Bijkomende Opties:

- de toegekende CPU kredieten moeten ook de kosten voorzien voor de ondersteuning op het vlak van software, parallelisatie, netwerken, enz.. De operationele leefbaarheid zou gegarandeerd kunnen worden door de kosten voor de ondersteuning van de HPC-gebruikers te dekken door de toegekende CPU kredieten.
- Een bijkomende inbreng zou kunnen komen door gedeeltelijk toegang te geven aan industriële HPC toepassingen. Dit zou ook leiden tot een aanvullende bron van inkomsten.

## **6. Slotbeschouwingen**

- De studie van de KVAB, gestaafd door een analyse van de internationale en Europese toestand en door een enquête bij de Vlaamse onderzoeksgemeenschap, leidt tot de sterke aanbeveling voor de oprichting van een HPC-centrum in Vlaanderen, op een schaal van een Niveau 2 structuur.
- De universitaire HPC-centra moeten tegelijkertijd verder uitgebouwd worden en versterkt. Ze moeten een stevige Niveau 1 basis vormen waarop de piramidale structuur moet berusten.
- Een HPC-centrum in Vlaanderen moet verbonden worden met de universitaire centra enerzijds en met de Europese HPC-centra anderzijds, met behulp van hoge snelheid netwerken.
- Naast de hardware basis, heeft de overgrote meerderheid van de geraadpleegde universitaire vorsers gepleit voor een belangrijke menselijke infrastructuur, met als doel het gebruik en de toegang tot de HPC-rekencapaciteiten te begeleiden op alle vlakken.
- Als bijkomende spin-off van een regionaal HPC-centrum, mag vermeld worden dat het ook een belangrijke stimulatie zou betekenen voor talrijke randontwikkelingen. Onder andere, moet vermeld worden:
  - De impact op de ontwikkeling van toegepaste wiskunde, mathematische algoritmen, beeldverwerking en visualisatie
  - Impact op de ontwikkeling van software engineering
  - Noodzaak voor technologie van hoge snelheid netwerken
  - Nieuwe HPC-gerichte ontwikkelingen in sectoren zoals Taalkunde of Financiën

Deze randontwikkelingen vormen een belangrijk onderdeel van de HPC-gebonden nieuwe toepassingen en technologie.

- Een initieel investeringsbudget van de orde van 10 M€ zou toelaten een HPC-centrum op te richten met een capaciteit van rond de tienduizend processoren (cores).
- Afschrijvingskosten over drie jaar, een initiële staf van 15 personen, en werkingskosten zouden oplopen tot een jaarlijks budget van 5 M€.
- Het voorstel van een HPC-centrum in Vlaanderen van Niveau 2, ligt in de lijn van de objectieven uitgestippeld in de beleidsnota van Minister F. Moerman, Vlaanderen i2010 'Tijd voor een Digitale Stroomversnelling in de Innovatieketen'.

**Het is de overtuiging van de KVAB, dat een initiatief rond een sterk regionaal HPC-centrum in Vlaanderen, een uitzonderlijk sterke impuls zou vormen voor de verdere uitbreiding van de wetenschappelijke excellentie en baanbrekend onderzoek in alle sectoren.**

**Het zou ook een belangrijke ondersteuning betekenen voor het potentieel tot innovatie in Vlaanderen.**



## **Bijlage 1: Samenstelling van de KVAB Werkgroep HPC**

De Werkgroep werd samengesteld uit leden van de Klasse van de Natuurwetenschappen.

### **Voorzitter van de Werkgroep**

Prof. Charles Hirsch      Computational Fluid Dynamics (CFD), Turbulentie, verbranding

### **Leden**

Prof. Yvan Bruynseraede      Materiaalonderzoek en nanotechnologie  
Prof. Annie Cuyt\*      Computerwetenschappen  
Prof. Norbert De Kimpe      Moleculaire chemie, farmacologie  
Prof. Renaat Gijbels      Moleculaire chemie, farmacologie  
Prof. Kris Heyde      Subatomaire fysica en fusie  
Prof. Dirk Inzé      Moleculaire biologie en biotechnologie, genetica, fysiologie  
Prof. Joseph Thas      Toegepaste wiskundige en mathematische algoritmen  
Prof. Luc Vanquickenborne      Fundamentele fysica (kosmologie, kwantumfysica en elementaire deeltjes)  
Prof. Joos Vandewalle      Engineering (methodologie, ontwerp, visualisatie, beeldverwerking, optimalisatie, telecommunicatie, cryptografie)  
Prof. Christoffel Waelkens      Fundamentele fysica (kosmologie, kwantumfysica)

\*Om de contacten met het initiatief Vlaanderen i2010 van Minister F. Moerman te verzorgen, werd Prof. Annie Cuyt van de Universiteit Antwerpen (UAntwerpen), na de eerste vergadering, aanvaard als lid van de HPC-Werkgroep.

Prof. Annie Cuyt is verantwoordelijk voor HPC aan de UAntwerpen, en is ook lid van de werkgroep HPC, binnen het initiatief Vlaanderen i2010.

## **Bijlage 2: Enquete formulier**

Werkgroep HIGH PERFORMANCE COMPUTING IN VLAANDEREN

Geachte Collega's,

De Klasse van de Natuurwetenschappen van de Koninklijke Vlaamse Academie van België voor Wetenschappen En Kunsten heeft een Werkgroep opgericht met als opgave een advies uit te brengen over de implicaties van HIGH PERFORMANCE COMPUTING (HPC) IN VLAANDEREN, en zich uit te spreken over de aan te bevelen acties.

### **INLEIDING**

High Performance Computing (HPC) is gericht op de aanwending van grootschalige computercapaciteit voor het oplossen van wetenschappelijke en industriële problemen.

Het huidige potentieel, met rekencapaciteiten tot 100 Teraflops (10<sup>14</sup> bewerkingen/sec), opent de weg tot nieuw grensverleggend onderzoek en ontdekkingen, in alle sectoren van de wetenschap en technologie.

Vlaanderen heeft tot nu toe slechts beperkte inspanningen gedaan om een plaats in te nemen in de systematische en grootschalige aanwending van HPC in wetenschappelijk onderzoek en industriële toepassingen.

Hiermee ontstaat een ernstig risico voor Vlaanderen achteruit te geraken op internationaal wetenschappelijk vlak en op het potentieel tot innovatie en industriële en economische vernieuwing.

Het staat reeds vast dat talrijke wetenschappelijke sectoren wereldwijd beïnvloed worden door HPC. Onder andere:

- fundamentele fysica (kosmologie, kwantumfysica en elementaire deeltjes)
- kernfysica en fusie
- moleculaire chemie en farmacologie,
- moleculaire biologie, biotechnologie, genetica en fysiologie
- turbulentie en verbranding
- geologie en kennis van de aarde, klimaatvoorspelling en 'global change'
- materiaal onderzoek en nanotechnologie
- toegepaste wiskundige en mathematische algoritmen
- software engineering methodologie, visualisatie en beeldverwerking, optimalisatie en hoge snelheid netwerken

De Werkgroep heeft beslist een enquête te ondernemen bij alle wetenschappelijke instellingen in Vlaanderen, om de behoeften en wensen van de Vlaamse onderzoeksgemeenschap voor HPC in kaart te brengen.

De Werkgroep wenst uit te maken

- hoe een hoogstaande HPC-infrastructuur een verschil kan maken voor het Vlaamse onderzoekslandschap
- wat de noden zijn
- welke middelen hiervoor nodig zijn

Om deze actie te ondersteunen, zouden wij u willen verzoeken de volgende vragen te beantwoorden, niet later dan 15 november 2006.

Mogen wij u ook verzoeken deze vragenlijst naar geïnteresseerden binnen uw onderzoeksgemeenschap, die niet door deze e-mail aangesproken werden, verder te willen doorsturen.

Met dank bij voorbaat voor uw medewerking,

Hoogachtend,

**Vragenlijst**

1. Beschrijf uw onderzoeksgebied  
.....
2. Beschrijf hoe HPC van nut kan zijn voor uw onderzoek
  - A. Reeds ondernomen HPC-gerichte toepassingen. Beschrijf het onderzoek objectief met vermelding van de maximaal gebruikte rekencapaciteit.  
.....  
.....
  - B. Plannen voor de komende 5 jaren voor HPC-gebruik. Beschrijf het onderzoeksobjectief met vermelding van de maximaal gewenste rekencapaciteit  
.....  
.....
3. Beschrijf het baanbrekend onderzoek dat door uw groep gevoerd zou kunnen worden met behulp van een hoogstaande HPC-infrastructuur van de orde van 100 Teraflops  
.....  
.....
4. Beschrijf de gewenste, of noodzakelijke, ondersteuning op het vlak van algoritmieken en software engineering (meer bepaald voor intensieve parallelisatie) om een hoge capaciteit HPC-infrastructuur maximaal uit te baten voor uw wetenschappelijke doelstellingen  
.....
- . Suggesties  
.....

## **Bijlage 3: Samenvattingen van de enquête resultaten in Vlaanderen**

### ***Bijlage 3a: Materiaalonderzoek en Nanotechnologie***

Prof. Yvan Bruynseraede

#### **Belangrijkste Research Richtingen**

- Metallische nanostructuren: structurele, elektronische en magnetische eigenschappen. (K. Temst, KULeuven)
- Halfgeleiderfysica: defectspectroscopie, foto- en elektroluminescente materialen, elektronen magnetische resonantie, metaal/halfgeleidersystemen. (P. Clauws, D. Poelman, F. Callens, R. Vanmeirhaeghe, UGent)
- Materiaalonderzoek, twee-fasestromingen, massatransport in elektrochemische systemen. (J. Fransaer, KULeuven)
- Depositie van deklagen: kathodeprocessen, simulatie van de plasmaprocessen. Ontwikkeling van numerieke modellen voor de beschrijving van plasma's/gasontladingen, de interactie tussen plasma/laser en vaste stof oppervlak. (A. Bogaerts, UAntwerpen)
- Gebruik van "High Throughput Methodologieën" voor de materiaalindustrie: ontwikkelen van combinatorische methoden. (J. Paul, FLAMAC)
- Multi-scale modelling van de plastische vervorming van metalen: eindige-elementen berekeningen van metaalomvormingsprocessen, polykristaldeforlatie-modellen, dislocatiepatronen, effect van twee-fazedeeltjes, reacties tussen individuele dislocaties, klassieke en nieuwe bronnen van dislocaties. (P. Van Houtte, KULeuven)
- Studie van elektronische en structurele eigenschappen van materialen: computationele technieken, ab initio elektronische structuurberekeningen. (D. Lamoen, UAntwerpen)
- Moleculaire dynamica simulaties van de groei van structuren met silicium- en germaniumlagen. Ab initio berekeningen en moleculaire dynamica: elektronische bandstructuur, fononspectra en elektron-fononinteractie in kwantumdraden met uiteenlopende afmetingen, gelaagde halfgeleiderstructuren en koolstofnanobuizen. (M. Van Rossum, IMEC)
- Eigenschappen van vaste stoffen onderzocht en geanalyseerd door kwantum-simulaties; "ab initio" of "first-principles" berekeningen. (S. Cottenier, KULeuven)

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen.*

- Moleculaire-dynamica simulaties werden uitgevoerd, waarvoor computertijd ter beschikking gesteld werd op de IBM 3090/600 VF computerfaciliteit van de KULeuven. (K. Temst, KULeuven)
- Simulaties van de turbulente stroming rond roterende cilinderelektrodes uitgevoerd op de cluster van de KULeuven. Starten van grootschalige simulaties met meer dan 200 processoren. (J. Fransaer, KULeuven)
- Eindige-elementsimulaties van plastische vormgevingsprocessen voor anisotrope materialen. op de HPC van de KULeuven. (P. Van Houtte, KULeuven)
- Berekeningen op snelle PCs of op de supercomputer-cluster van de UAntwerpen (Sun Microsystems Belgium, met 256 computers), die elk uitgerust zijn met twee 64 bit AMD processoren. Plannen om enkele van onze programma's te paralleliseren. (A. Bogaerts, UAntwerpen)
- Parallelberekeningen werden uitgevoerd met 10-100 processoren en 2 GB RAM geheugen aan het Max-Planck-Instituut te Stuttgart en het FZ Jülich. (D. Lamoen, UAntwerpen).
- Ab-initio-berekeningen van elektronische bandstructuur en fononspectra zijn uitgevoerd voor lange, dunne kwantumdraden. Maximale rekencapaciteit die ter beschikking staat: PC cluster bestaande uit 24 Pentium IV (3 GHz) processoren. Iedere processor heeft 2 Gb geheugen. Totale hardeschijfruimte: 450 Gb. De

computers zijn via een gigabit-ethernet switch met elkaar verbonden. (M. Van Rossum, IMEC)

- Typische berekeningen zijn voor eenheidscellen van 20-100 atomen op een tiental parallel gekoppelde pc's. Maximale rekencapaciteit die ter beschikking staat: PC-cluster bestaande uit 24 Pentium IV. (3 GHz) processoren. Iedere processor heeft 2 Gb geheugen. Totale hardeschijfruimte: 450 Gb. De computers zijn via een gigabit-ethernet-switch met elkaar verbonden. (S. Cottenier, KULeuven)

*b) Plannen voor de komende 5 jaren voor HPC-gebruik.*

- Grootschalige micromagnetische simulaties. (K. Temst, KULeuven)
- Simulaties uitbreiden naar de stroomdistributies op roterende cilinderelektrodes met 100 of meer processoren. (J. Fransaer, KULeuven)
- Verder ontwikkelen van numerieke modellen met snellere computers: laser ablatie processen beschrijven op langere tijdschaal, nl. ms-regime; simuleren van de afzetting van nanogestructureerde filmen. (A. Bogaerts, UAntwerpen)
- Begrijpen van materiaal eigenschappen op atomaire schaal mbv modellerings-technieken. (J. Paul, FLAMAC)
- Onderzoek van ternaire en quaternaire halfgeleider structuren dmv. "first principles" elektronische structuurberekeningen. Nood aan HPC-systemen met grotere geheugencapaciteit. (D. Lamoen, UAntwerpen)
- Uitbreiding van de ab initio berekeningen, de zelf-consistente berekeningen en de kwantumtransportsimulaties tot structuren met kleinere symmetrie en mesoscopische afmetingen (2 tot 100 nm) . Nood aan 100 processoren met optische verbinding ("infiniband switch") en 4 GB geheugen per processor. (M. Van Rossum, IMEC)
- Nood aan het voltijdse equivalent van 20-40 cpu's per project of onderzoeker. (S. Cottenier, KULeuven)

*c) Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*

- Driedimensionale micromagnetische simulaties van magnetische nanostructuren. (K. Temst, KULeuven)
- De invloed van turbulent massatransport op (elektro)chemische processen simuleren (J. Fransaer, KULeuven)
- Om tot een volledige simulatie van het depositieproces te komen moeten de verschillende onderdelen van het huidig lopend onderzoek met elkaar gekoppeld worden: Kinetic Monte Carlo simulaties en Particle-In-Cell simulaties. (D. Depla, UGent)
- Het ontwerp en de ontwikkeling van materialen versnellen en deze strategie combineren met de aanwezig kennis binnen de materiaal industrie in Vlaanderen: platform voor pragmatisch modelleren in materiaal onderzoek. (J. Paul, FLAMAC).
- Realistische simulaties aan multilaagstructuren waarbij voldoende atomaire lagen in rekening worden gebracht (~ enkele honderden atomen) en waarbij ook de structuur (i.e. atoomposities) volledig gerelaxeerd kan worden. (D. Lamoen, UAntwerpen)
- Informatie over nieuwe typen halfgeleidercomponenten mbv ab-initio-berekeningen en transportsimulaties. (M. Van Rossum, IMEC)
- Het toekomstideaal is dat materialen met nieuwe gewenste eigenschappen eerst volledig computationeel ontworpen worden, voordat deze in een labo gesynthetiseerd worden: 'computational materials design'. (S. Cottenier, KULeuven)

*d) Gewenste ondersteuning*

- Er zou een adequate ondersteuning moeten komen voor de technische en numerische aspecten van de HPC-infrastructuur. (K. Temst, KULeuven)
- Onderlinge samenwerking is essentieel voor het slagen van bepaalde simulaties. (J. Fransaer, KULeuven)
- Een vlotte technische ondersteuning op vlak van het operating systeem is absoluut nodig. Al te dikwijls blijkt, dat een programma dat in principe reeds op een

parallelcomputer of een clustercomputer zou moeten kunnen lopen (wat elders lukt, bijvoorbeeld bij de software leverancier), dit niet wil doen op onze eigen installatie. (P. Van Houtte, KULeuven)

- Een software ondersteuning zou uitermate gewenst zijn. Wij wensen computer-programma's te ontwikkelen voor beschrijving van chemische en fysische processen, maar deze programma's zijn verre van geoptimaliseerd, omdat we hiervoor te weinig expertise bezitten. (A. Bogaerts, UAntwerpen)
- Software ondersteuning is cruciaal. Gebruikte codes optimaal laten renderen of zonder problemen laten werken op verschillende computer platformen is geen trivialiteit. (D. Lamoen, UAntwerpen)
- Gebruik van een HP-programmeertaal om de mogelijkheden tot parallelisering en vectorisering maximaal uit te buiten (bijvoorbeeld HPFortran 90/95). Gebruik van een bijbehorende, flexibele debugger; aankoop van specifieke, numerieke bibliotheken. (M. Van Rossum, IMEC)
- Transparant systeem, waarop standaard wetenschappelijke software als efficiënte compilers, bibliotheken en visualisatietools aanwezig zijn. De meest nuttige vorm van ondersteuning is het kunnen raadplegen van systeemexperts die deze hulpmiddelen vlot kunnen toepassen op mijn problemen. (S. Cottenier, KULeuven)

### **Aanbevelingen en Verwachtingen**

- Ik steun ten volle acties in deze richting: aankoop van performante toestellen; adequate technische ondersteuning van de apparatuur; gebruik van de reeds bestaande Belgische expertise; bouw een eventueel Belgisch HPC-centrum of overweeg samenwerking met buitenlandse centra. (K. Temst, KULeuven)
- Investeer niet enkel in hardware maar ook in ondersteuning. Maak ook gebruik van de reeds bestaande expertise aan de Belgische universiteiten en bouw een eventueel Belgisch HPC-centrum uit aan een Belgische universiteit of overweeg samenwerking met buitenlandse centra. (D. Lamoen, UAntwerpen)
- Een krachtige supercomputer is een mooi ding, maar die zit steeds in een merkwaardig spanningsveld. Met veel geld kan er een systeem opgezet worden dat zo goed is dat iedereen er gebruik van wil maken, met als gevolg lange wachttijden. Waarna mensen terugkeren naar hun eigen, individuele systemen omdat het daar misschien wel wat trager gaat, maar ze uiteindelijk toch sneller geholpen zijn. Het lijkt me essentieel dat er vooraf goed nagedacht wordt welke rol de 'Vlaamse supercomputer' in dit spanningsveld wil spelen, dat er onderzocht wordt welke capaciteit daarvoor dan wel nodig is, dat er vervolgens nagegaan wordt of de financiële middelen er zijn om voor die capaciteit te zorgen, en of die middelen ook op langere termijn gegarandeerd zijn om het systeem in de gewenste positie op de markt te houden. Het schrijven van projecten om toegang te krijgen tot grote systemen is een energie-opslopende bezigheid die het nut van deze systemen soms tegenwerkt. Als een systeem met Vlaams belastingsgeld gefinancierd wordt, dan moet een laagdrempelige instap voor Vlaamse onderzoekers ook verzekerd zijn. (S. Cottenier, KULeuven)

### **FLAMAC**

Begin 2005 werd het Flanders Materials Centre (Flamac) opgericht door drie bedrijven met belangrijke activiteiten in onderzoek en ontwikkeling in Vlaanderen (Agfa, Arcelor en Umicore) tesamen met Agoria Vlaanderen.

Financiële steun werd voorzien door de drie bedrijven alsook de Vlaamse regering.

Het doel van Flamac is een competentiecentrum te worden in "High Throughput Methodologieën" om het onderzoek in de materiaalindustrie te ondersteunen. Op deze manier zal innovatief onderzoek naar nieuwe materialen en formuleringen, hun eigenschappen en toepassingen versneld en verruimd worden.

Hiervoor zal Flamac combinatorische methoden ontwikkelen en aanwenden. Het streeft naar gemeenschappelijk, multi-disciplinair onderzoek tussen bedrijven, universiteiten of hoge scholen en onderzoekscentra.

Deze competentie van Flamac rond “High Throughput Methodologieën” moet zowel KMO’s alsook grotere bedrijven toelaten om de ontwikkeling van nieuwe producten en processen te stimuleren alsook hun time-to-market voor nieuwe materialen te verkorten.

High Throughput Methodologieën (HTM) worden aangewend om de ontwikkeling van nieuwe materialen en formuleringen te versnellen. HTM maken gebruik van een brede waaier aan high tech experimentele technieken en software middelen die toelaten om testen snel en in parallel uit te voeren. Dit resulteert in een verhoogde productiviteit en een kortere time-to-market.

De eerste focus van Flamac situeert zich in de opbouw van de nodige kennis rond combinatorische technieken en high throughput methodologieën. De traditionele manier van materiaalonderzoek gebeurt door experimenten één-na-één uit te voeren. In het onderzoek naar nieuwe materialen spelen echter verschillende – vaak samenhangende – parameters een rol, zodat traditionele procedures niet meer voldoende zijn.

Combinatorische strategieën lossen dit dilemma op door automatisatie, miniaturisatie, en belangrijker, experimenten in parallel uit te voeren. Door gebruik te maken van deze strategieën kunnen honderden experimenten simultaan, of snel achter elkaar, uitgevoerd worden. Op deze manier wordt de tijd voor de ontwikkeling van een nieuw product drastisch versneld. Huidige robotica in combinatie met snelle screening-methoden, miniaturisatie van testen, en aanwenden van data mining en statistische en visualisatie technieken, laten toe om op korte termijn een grote hoeveelheid bruikbare data te vergaren en zodoende een betere relatie te kunnen leggen tussen verschillende syntheseparameters van nieuwe materialen of formuleringen en hun eigenschappen.

### **Open Structuur voor Gemeenschappelijk Onderzoek**

Flamac staat open voor bedrijven, onderzoekscentra en academische laboratoria. Het centrum werd opgestart in samenwerking met verschillende onderzoeksgroepen. Het centrum is opgebouwd als een vereniging zonder winstoogmerk, waarbij nieuwe leden zich kunnen aansluiten. Oorspronkelijk is het onderzoekscentrum een initiatief van Agfa, Arcelor, Umicore en Agoria Vlaanderen, met ondersteuning van de Vlaamse regering.

Flamac zal een centrum worden waar zowel industriële alsook universitaire zich ontmoeten en werken rond dezelfde competenties.

Flamac wil dan ook een plaats zijn waar collectief onderzoek gebeurt rond “High Throughput Methodologieën”. Onderzoek zal vaak project gedreven zijn voor derde partijen en kan gesteund worden door excellentiecentra in onderzoeksinstituten en universiteiten, welke elke hun specifieke kennis en competentie inbrengen naast de competentie opgebouwd in Flamac.

Elk van de huidige industriële partners ondersteunt een project om Flamac op te starten in een periode van 3 jaar.

**Belangrijkste Onderzoeksrichtingen**

- De gedetailleerde kwalitatieve en semi-kwalitatieve analyse van het functionele eiwitcomplement van het genoom (proteoom). (K. Gevaerts, UGent/VIB)
- De comparatieve analyse van genomen en het ontrafieren van de evolutie van genomen. Om dit te kunnen doen moeten de geninhouden van complete genomen (we spreken hier meestal over enkele tienduizenden genen) met elkaar vergeleken worden. Dit gebeurt door gebruik te maken van wiskundige algoritmen die genen en hun sequenties optimaal met elkaar proberen te vergelijken. Dit vereist enorme computertijden. Om hieraan nu reeds ten dele tegemoet te kunnen komen, worden momenteel reeds meer en meer computersystemen gebruikt waar algoritmes in hardware geïmplementeerd zijn, omdat de software implementaties op de beschikbare systemen niet meer voldoen. (Y. Van de Peer, UGent/VIB)
- Het ontwikkelen en gebruik van nieuwe bioinformatica methoden om genomen te bestuderen. Het bepalen van de genoomsequentie van een organisme verloopt steeds sneller en wordt aldoor goedkoper. Het is de verwachting dat er in de komende jaren het aantal genoomsequenties exponentieel zal stijgen. Nieuwe methoden (bv. 454 sequentie bepaling Roche) laten toe om genomen zeer snel opnieuw te sequencen waardoor polymorfismen kunnen opgespoord worden. Met meer reken capaciteit zullen we in staat zijn om predicties in genomisch DNA uit te breiden naar alle beschikbare genomen en de predicties te optimaliseren via genetisch programmeren, hetgeen niet 1 genomische "run" zou moeten uitvoeren per query, maar tienduizenden, snel toenemende hoeveelheid DNA sequenties vereist een toenemende reken capaciteit om bijvoorbeeld nieuwe regulatorische elementen op te sporen in het niet-coderend DNA, dat het leeuwendeel van het genoom uitmaakt. (B. De Strooper, KULeuven/VIB)
- De gecoördineerde regulatie van genexpressie vormt de basis van de adaptieve respons van een cel aan zich steeds veranderende cellulaire condities. De moleculaire mechanismen, die deze gecoördineerde genregulatie aansturen en bepalen of de transcriptie van een gen onder bepaalde omstandigheden wordt geactiveerd dan wel afgeremd, kunnen worden opgedeeld in verschillende niveaus. Het laagste niveau bestaat uit individuele transcriptiefactor bindingsites (TFBS), dit zijn korte specifieke DNA sequentie fragmenten (5-25 bp) gelegen in niet-eiwit coderend DNA in de nabijheid van het gen onder controle, waarop transcriptiefactoren (TF) kunnen binden om op die manier een respons te bewerkstelligen. Men neemt aan dat een optimale respons van een cel ten opzichte van nieuwe leefomstandigheden wordt bekomen, wanneer optimale integratie van verschillende signalen ter hoogte van de regulatie van genexpressie plaats grijpt. Dit tweede niveau van transcriptieregulatie wordt bekomen wanneer specifieke collecties van TFBS georganiseerd worden in modules waarbij de interacties van de TF met elkaar en met hun TFBS een fijnregeling waarborgen van de transcriptionele respons van het gen, onder controle van deze module, ten opzichte van een gegeven impuls. Ondanks de toenemende erkenning dat modules, eerder dan individuele TFBS, instaan voor de transcriptionele regulatie in hogere eukaryoten is nog relatief weinig inspanning geleverd voor de ontwikkeling van computationele benaderingen voor de predictie van modules. Wij ontwikkelen een methode die zich niet beperkt tot enerzijds de identificatie van minimale modules van potentieel samenwerkende TFBS paren of anderzijds strikte regels oplegt met betrekking tot de vereiste afstand tussen TFBS binnen een module, aangezien kan worden verwacht. (F. Van Roy, UGent/VIB)
- De doelstelling van systeembioïologie is het uitwerken van een model voor het complexe netwerk van interacties tussen de verschillende componenten van de levende cel. Gesofisticeerde wiskundige procedures zijn nodig om de grote experimentele data sets van deze netwerken te analyseren. Het uiteindelijke doel van systeem biologie is het begrijpen van de fundamentele werking van de levende cel en het organisme en het blootleggen van de onderliggende genetische



netwerken en regulatiesystemen. Voor het reverse engineeren van genetische regulatienetwerken worden optimalisatietechnieken gebruikt, gaande van "simulated annealing" tot "ensemble learning", waarbij honderden of duizenden netwerken met elk tienduizenden nodes en honderdduizenden interacties in parallel worden berekend. In de toekomst zal ook comparatieve analyse van dergelijke netwerken in verschillende organismen steeds belangrijker worden, bijvoorbeeld om te achterhalen hoe deze regulatienetwerken zijn geëvolueerd en welke mechanismen aanleiding geven tot geobserveerde emergente eigenschappen van moleculaire netwerken zoals modulariteit, plasticiteit en robuustheid bij perturbatie. Dergelijke comparatieve en evolutionaire netwerkanalyses zijn bijzonder rekenintensief, en het aantal mogelijke vergelijkingen neemt meer dan exponentieel toe met het aantal (honderden tot duizenden) organismen waarvoor voldoende data beschikbaar is. Dit leidt tot een onvoorstelbare combinatoriële explosie die steeds snellere en krachtigere computerinfrastructuur vraagt. (S. Maere, Y. Van de Peer, D. Inzé, UGent/VIB)

- Genetisch onderzoek naar de oorzaken van aandoeningen van het zenuwstelsel. Onderzoek mogelijk betrokken genen en pathways. Optimalisatie van bestaande experimentele en theoretische technieken die hierbij gebruikt worden. (C. Van Broeckhoven, UAntwerpen/VIB)
- Linkage analyses van genome wide scans in grote pedigrees: Associatie analyses; zoeken van onbekende sequenties in sequentie databanken (BLAST); checken van primers tegen sequentie databanken; genome wide sequentie analyse (opzetten van resequencing experimenten); genome wide sequentie vergelijkingen. Voornamelijk uitgevoerd op CalcUA.
- Computationale neurowetenschappen: Onderzoeksrichtingen, berekeningen en simulaties: moleculaire simulatie van signaaltransductie netwerken in neuronen (reactie-diffusie systemen in realistische morfologie); cellulaire simulatie van vuurgedrag van grote neuronmodellen (compartimentele modellen met actief membraan); netwerk simulatie van grote neurale netwerken (conductantie gebaseerde modellen van neuronen). (E. De Schutter, UAntwerpen)

#### **Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops**

- HPC zal toelaten om meer parameters in de zoekopdrachten die de massaspectroscopische data verwerken in te voeren waardoor er mogelijks onverwachte eiwitmodificaties kunnen opgespoord worden. (K. Gevaerts, UGent/VIB)
- De comparatieve analyse van genomen en het ontraferelen van de evolutie van genomen. Om dit te kunnen doen moeten de geninhouden van complete genomen (we spreken hier meestal over enkele tienduizenden genen) met elkaar vergeleken worden. Dit gebeurt door gebruik te maken van wiskundige algoritmen die genen en hun sequenties optimaal met elkaar proberen te vergelijken. Dit vereist enorme computertijden. Om hieraan nu reeds ten dele tegemoet te kunnen komen, worden momenteel reeds meer en meer computersystemen gebruikt waar algoritmes in hardware geïmplementeerd zijn, omdat de software implementaties op de beschikbare systemen niet meer voldoen. (Y. Van de Peer, UGent/VIB)
- De snel toenemende hoeveelheid DNA sequenties vereist een toenemende rekencapaciteit om bijvoorbeeld nieuwe regulatorische elementen op te sporen in het niet-coderend DNA, dat het leeuwendeel van het genoom uitmaakt. Tevens is bijzonder veel rekencapaciteit nodig om grote genomen met elkaar te vergelijken teneinde de structuur en evolutie van de genomen te doorgronden. (B. De Strooper, KULeuven/VIB; Y. Van de Peer, UGent/VIB)
- Genoomwijde microarrays (zogenaamde tilingarrays) laten toe om na te gaan welk DNA (coderend als niet-coderend) in het genoom overgeschreven wordt in een transcript. Deze micro-arrays bevatten veel meer features dan de huidige "gen" microarrays (enkel coderende sequenties) en de analyse van dergelijke arrays vereist een grote rekencapaciteit. (D. Inzé, UGent/VIB)
- Methodes voor de reconstructie van genetische netwerken zijn in toenemende mate afhankelijk van statistiek en automatisering van leer algoritmes voor patroon herkenning. Deze methodes zijn erg gerelateerd aan het veld van datamining, waarin op een geautomatiseerde manier patronen en relaties worden gezocht in

grote hoeveelheden gegevens. Bepaalde nieuwe manieren van data integratie worden geconfronteerd met het probleem dat de mining algoritmes te maken hebben met leer-problemen die NP-hard zijn (niet-deterministische polynomiaaltijd hard) of moeilijker, dus een belangrijk onderdeel van dit vakgebied is algoritmes te ontwikkelen die de oplossing benaderen. De benaderbare oplossingen zullen aanmerkelijk nauwkeuriger zijn wanneer een aanzienlijke hoeveelheid rekenkracht beschikbaar is. (M. Kuiper, UGent/VIB)

- De ontwikkeling van methodes om 'fuzzy' interacties tussen TFBS te modelleren. De idee is om virtuele structuren te construeren die alle associaties tussen TFBS voor elk van de sets van differentieel gereguleerde genen weergeven onder de vorm van afstanden. Door toepassing van het 'distance difference matrix' (DDM) concept uit de structurele biologie, gevolgd door 'multidimensional scaling' (MDS) zijn we in staat de TFBS associaties die betrokken zijn in de regulatie van de differentieële genexpressie te scheiden van de willekeurige associaties die optreden met vals positieve TFBS predicties. Het eindresultaat is een DDM-MDS plot die op een visueel overzichtelijke manier informatie geeft over het betrokken regulatorisch netwerk, meer bepaald de verschillende modules aanwezig in de promotoren van de differentieel gereguleerde genen in combinatie met indicaties over de sterkte van de associaties tussen de individuele TFBS. Na uitvoering van het DDM-MDS protocol, berekenen we: (1) de afstand tussen de oorsprong van de MDS plot en elke TFBS op de plot. Deze afstand zal dienen als een score om de mate waarin elke TFBS is oververtegenwoordigd in elke dataset te kwantificeren en (2) de afstanden tussen alle TFBS op de plot. Deze afstanden zijn een score voor de mate van de associaties tussen de verschillende TFBS. Vervolgens schatten we een P-waarde voor deze scores. Hiertoe definiëren we de nulhypothese dat een TFBS niet participeert in associaties met andere TFBSs. Dit vereist specificatie van een probabilistisch model dat wordt gebruikt om de waarschijnlijkheid te schatten dat dit model een afstand genereert die minstens even groot is als de waargenomen afstand. Willekeur kan worden geïmplementeerd in het nulmodel door ofwel permutatie van de promotorsequenties of door het nemen van willekeurig gekozen genoom- of promotorsequenties. De DDM-MDS procedure wordt toegepast op verschillende willekeurig samengestelde sets en de resulterende afstanden voor elke geprojecteerde TFBS tot de oorsprong van de MDS plot en tussen alle TFBS worden genoteerd. De P-waarde van een echte afstand kan dan worden berekend uit de fractie van het aantal keer dat een afstand van een TFBS t.o.v. de oorsprong of t.o.v. een andere TFBS in de willekeurig samengestelde promotor set groter is dan de echte afstanden. Het spreekt vanzelf dat deze laatste stappen computationeel zeer intensief zijn aangezien typisch 5000 tot 10.000 randomisaties worden gebruikt voor de berekening van de P-waarden. (F. Van Roy, UGent/VIB)
- De complexiteit van cellulaire systemen vereist de ontwikkeling van mathematische modellen om de complexiteit te visualiseren, te begrijpen en op termijn voorspellingen te doen hoe we een biologisch systeem kunnen beïnvloeden om een gewenst effect (bijvoorbeeld het bestrijden van een ziektebeeld) te bekomen. Systeembioologie maakt gebruik grote hoeveelheden data, bijvoorbeeld genomwijd (transcriptome, metabool, proteoom, interactoom, localisoom,...). Daarenboven worden deze gegevens bijvoorbeeld verzameld op zeer veel tijdspunten en na verschillende genetische (mutaties) of chemische perturbaties. Het spreekt voor zich dat het gebruik van deze enorme hoeveelheid gegevens voor het berekenen van de het moleculaire netwerk dat cellulaire processen reguleert zeer veel computertijd opsloopt. De mathematische modellen die een biologisch proces beschrijven kunnen gevisualiseerd worden in natuurgetrouwe dynamische modellen. (S. Maere, Y. Van de Peer, Dirk Inzé, UGent/VIB)

### **Bijlage 3c: Engineering**

Prof. Joos Vandewalle

#### **Belangrijkste Research Richtingen**

- Computersimulaties en ontwikkeling van parallele, efficiënte methoden om korrelgroei in aanwezigheid van tweede-fasedeeltjes op een snelle en betrouwbare manier te simuleren, oplossen van grootschalige, complexe simulatieproblemen in wetenschap en techniek, snelle solvers voor partiële differentiaalvergelijkingen, numerieke methodes voor niet-lineaire en dynamische systemen en simulatie van het macroscopisch gedrag van microscopische modellen. (D. De Schreye, P. Dutré, P. Verbaeten, KULeuven, Computerwetenschappen)
- Modelleren van problemen uit diverse applicatiegebieden met behulp van neurale netwerken, genetische algoritmen, genetisch programmeren, lerende classificier systemen en support vector machines. (D. Aeyels, G. De Cooman, UGent, Elektrische energie, Systemen en Automatisering (EESA))
- Modelleren van problemen uit diverse applicatiegebieden met behulp van neurale netwerken, genetische algoritmen, genetisch programmeren, lerende classificier systemen en support vector machines; onderzoek van geavanceerde modelleringstechnieken, EM simulatie, oplossen van modelleringproblemen zoals het betrouwbaar doorsturen van data over een computernetwerk, inverse problemen (vormreconstructie), ontwikkelen van gevalideerde numerieke bibliotheken, financiële modellen. (W. Hendrickx, UAntwerpen, Wiskunde en Informatica)
- Mathematische modellen van systemen, klassificatie en gegevensontginning, bioinformatica, signaalverwerking. (M. Moonen, Y. Moreau, J. Suyskens, B. Preneel, KULeuven, Elektrotechniek - SCD)
- Computatieve chemie en medicijnontwerp. (Molmo Services BVBA, Turnhout)
- Simulatie en optimalisering van chemische reactoren: op grote schaal; toegepaste productieprocessen dmv Warmte- massa- en impulstransport in chemische reactoren of ovens simuleren met CFD-codes (Computational Fluid Dynamics), theoretische berekeningen van moleculaire systemen met als doel fysische en chemische eigenschappen te voorspellen en te verklaren op microscopische basis, de elektronische structuur van materialen, onderzoek van nieuwe katalysatoren, ontwikkeling van gefunctionaliseerde materialen, chemische reacties bij solventen. (G. Heynderickx, Ugent, Lab. Petrochemische Techniek)
- Multi-scale modelling van de plastische vervorming van metalen: eindige-elementen berekeningen van metaalvormingsprocessen, polykristaldeformatie-modellen, dislocatie- patronen, effect van twee-faze deeltjes, reacties tussen individuele dislocaties, klassieke en nieuwe bronnen van dislocaties. (D. De Schreye, P. Dutré, P. Verbaeten, KULeuven, Computerwetenschappen)
- Beeldverwerking met complexe algoritmen om hun inhoud te interpreteren: Inverse problemen en gegevensanalyse in computervisie; Inverse problemen in elektromagnetische data gerelateerd aan tomografische reconstructie van verborgen voorwerpen of fouten; Inverse problemen in infrarood thermografie; Verwerking en analyse van multimodale gegevens (video en audio); Moleculaire dynamicsimulaties van de groei van structuren met silicium- en germaniumlagen. (J. Cornelis, VUbrussel, ETRO)
- Ontwikkeling van data acquisitie software, beeld reconstructie software en beeldverwerking software. (Medische Instrumentatie DNTK/WE)
- Monte-Carlosimulaties voor het realistisch modelleren van nucleair medische scanners. (K. Deboschere, J. Van Campenhout, UGent, Elektronica en Informatie-systemen - MEDISIP)
- Musculoskeletale biomechanica: studie van de sterkte (mechanische eigenschappen in het algemeen) van biologische weefsels (beenderen en spieren) en gecorreleerd met de structuur. (H. Van Brussel, S. Swevers, D. Vandepitte, J. Vander Sloten, KULeuven, Mechanica)
- Gedistribueerde energieopwekking, vermogenelektronica, power quality, elektriciteitsmarkten en economische energie aspecten, robuustheid en betrouwbaarheid, energie & ICT-infrastructuur, elektromagnetische materiaalverwerking, modellering en optimalisatie. (R. Belmans, KULeuven, Elektrotechniek Electa)

- Hydrodynamica (golven en stromingen) + sedimenttransport (niet cohesief & cohesief) in kustgebieden en estuaria. (G. Degrande, J. Monbaliu, E. Toorman, KULeuven, Bouwkunde)
- Modelbouw en simulatie van hogere organismen op verschillende schalen van ruimte en tijd (multiscale modellering), incl. Transport van metabole gassen en water in vruchten (Reconstructie van (micro)tomografische beelden, Metabole netwerken, Computational Fluid Dynamics, Luchstrooming in koelruimten, ovens enz., Microfluidics, Discrete-elementen methode, Stroming van korrelige materialen, Micro-mechanica van plantaardig weefsel). (D. Berckmans, KULeuven, BIOSYST - MeBioS)
- Ontwerp van fotonische geïntegreerde schakelingen bevat diverse componenten en steunt op geavanceerde CAD-tools, die de volledige vergelijkingen van Maxwell in twee of drie dimensies moeten oplossen. Informatietechnologie: nanofotonica, fotonisch geïntegreerde chips. Communicatienetwerken en gedistribueerde software. (P. Bienstman, D. Botteldooren, D. De Zutter, I. Lemahieu, UGent, Informatie Technologie)
- Onderzoek van geavanceerde modelleringstechnieken, EM-simulatie, oplossen van modelleringsproblemen zoals het betrouwbaar doorsturen van data over een computernetwerk, inverse problemen (vormreconstructie), ontwikkelen van gevalideerde numerieke bibliotheken, financiële modellen. (A. Cuyt, UAntwerpen)

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen.*

- 2D-simulaties van korrelgroei op een node met 16GB werkgeheugen
- 3D-simulaties van korrelgroei in dunne filmen op een node met 16GB werkgeheugen
- 3D-simulaties van korrelgroei op 100 nodes met elk 2 tot 4GB werkgeheugen (lopend onderzoek). (D. De Schreye, P. Dutré, P. Verbaeten, KULeuven, Computerwetenschappen)
- Aerothermodynamisch gedrag van het European Experimental Reentry Testbed (EXPERT) Vehicle, ontwikkeld door ESA met een eindige-volume code, op een ongestructureerd rooster, geïmplementeerd binnen het COOLFluid object-georiënteerd framework voor hoog performante parallele simulaties. 3D Navier-Stokes vergelijkingen, aangevuld met chemische reactievergelijkingen voor 5 stoffen uiterst fijn discretisatie rooster (met 100 miljoen onbekenden) een computercluster met 480 CPUs. (D. De Schreye, P. Dutré, P. Verbaeten, KULeuven, Computerwetenschappen)
- Simulatie van een volledige PET-scanner voor de design van nieuwe scanners, en om beter de gemeten prestaties van het systeem te kunnen begrijpen, en iteratieve reconstructie met een cluster van ongeveer 50 PC's - zeer hoog dimensionale microarray data analyse en ordinale classificatie van slijtagegraad en kwaliteit van tapijstalen op basis van beeldmateriaal met IBM BladeCenter chassis, one master node, IBM X335, 40 dual CPU (80 CPU's) computing blades, Linux RH3 and Sun Gridware, Network Appliance NAS storage. (Medische instrumentatie DNTK/WE)
- Docking algoritme (Abida) wordt ontwikkeld en getest voor de potentiaalvelden van een enzyme (1973 atomen, 8048 QM basisfuncties) met een quantum mechanische ab initio berekening. Dit vereist 3000 cpu-uren op 62 Itanium-2 cpu's van de 'Aster' machine van het SARA in Amsterdam (dit probleem schaalte goed tot deze systeemgrootte) met gemiddeld 0,3 van de 2,2 Teraflops capaciteit van deze machine. In 2004 analyse van een zeer grote hoeveelheid klinisch geïsoleerde HIV-protease varianten (meer dan 3000) in silico op resistentie tegen het AIDS-medicijn amprenavir in twee weken op de 216 cpu Xeon cluster van J&J in Raritan NJ (plm 4,5 Teraflops). (Molmo Services BVBA - Turnhout)
- Kwantumchemische berekeningen en van CFD codes de berekeningen uitgevoerd op een cluster met 24 nodes met 48 CPU's tot 3,0 GHz elk. Ongeveer 2/3 van de clustercapaciteit gaat naar kwantumchemische berekeningen, het overige derde naar CFD berekeningen met eigen ontwikkelde codes. (G. Heynderickx, Ugent, Lab. Petrochemische Techniek)

- Organische reacties die zich afspelen in solventen, waarvoor een groot deel van het solvent expliciet in rekening moet gebracht worden met 2 Altix rekenunits met elk 12 Itanium processoren en 2GB RAM geheugen per processor en recent daarnaast 5 Power 5 IBM machines met elk 8 processoren en 4GB RAM geheugen per processor. (V. Van Speybroeck, M. Waroquier, UGent, VVS- CMM)
  - Beeldverwerking: niet-lineaire least-squares solver verplaatsinggebaseerde niet-lineaire eindig elementenmodel toegepast voor de 3D vervormingschatting van een menselijk gezicht. De verwerking van 1 beeld uit de sequentie neemt nu een halve dag in beslag op een 2Ghz processor voor een eenvoudige gezichtsmodel structuur met 850 punten. Om een realistische gezichtsbeweging te verkrijgen, zou het netwerk van punten moeten worden verfijnd wat zal leiden tot hogere rekenbelasting. Driedimensionale tomografische reconstructie van de ondergrond om de posities, vorm, en materiële parameters van ondergrondse voorwerpen en/of fouten te schatten uit data geleverd door een grondradar of andere elektromagnetische sensoren. De oplossing van de hieraan gerelateerde inverse problemen vergt de oplossing of verscheidene driedimensionale voorwaartse problemen die elk 2 uren vereisen op een standaardmachine met 2GHz processorsnelheid. (J. Cornelis, VU Brussel, ETRO)
  - Realistisch modelleren en simuleren van patient acquisities op SPECT en PET scanners door middel van bovenvermelde GATE-software. Daartoe werd in 2003 een 37 nodes tellende cluster aangekocht (per node telkens twee 2,4 GHz processoren), m.a.w. in totaal dus een 75-tal CPUs. Deze cluster werd aangewend voor GATE Monte-Carlobaseerd onderzoek naar detector-modellering, evaluatie van reconstructie- en correctietechnieken, validatie van theoretische mathematica alsook voor respons-karakterisering en voor simulatie van niet-meetbare fenomenen zoals septale penetratie van collimatoren. (K. Debusschere, J. Van Campenhout, UGent, ELIS-MEDISIP)
  - Bioinformatica toepassingen vereisen de HPC-infrastructuur voor toepassingen zoals text mining (16 miljoen abstracts in PubMed), analyse van grote datasets, in het bijzonder na datafusie van verschillende experimenten en databronnen, en de zoektocht naar regulatorische elementen in het genoom. Bijvoorbeeld, in twee studies betreffende de analyse van regulatorische elementen, uitgevoerd in 2006 werd gebruik gemaakt van 5000 cpu-dagen rekentijd. (Y. Moreau, KULeuven, Elektrotechniek -Bioinformatica).
- b) *Plannen voor de komende 5 jaren voor HPC-gebruik.*
- Grotere simulaties met een lage volumefractie aan deeltjes met een systeem van  $512 \times 512 \times 512$  onbekenden gebruikt worden om op het einde van een simulatie nog voldoende korrels over te houden voor een statistisch relevant resultaat. Nood aan meer 4Gb-node
  - Model uitbreiden: toevoeging van oriëntatie-afhankelijkheid van de eigenschappen van de korrelgrenzen, diffusie van opgeloste onzuiverheden, faseformaties en het effect van elastische spanningen op de evolutie van de microstructuur. Elke simulatie vereist ongeveer 100 nodes van 4GB. Om de invloed van de verschillende invoerparameters na te gaan moeten meerdere simulaties uitgevoerd worden. Indien er meer dan 100 nodes zijn, kunnen meerdere simulaties tegelijk lopen.
  - Uitbreiding naar multischaal-simulatie, waarbij microscopische modellen (bv. moleculaire dynamica, modellen voor de evolutie van de microstructuur in heterogene materialen) uitgerekend worden, al dan niet gekoppeld aan macroscopische modellen (differentiaalvergelijkingen). Deze berekeningen vergen een zeer grote computercapaciteit (rekenwerk, geheugen) door de zeer fijne ruimten-tijdschalen, en door de stochastische aspecten. Daarnaast grootschalige optimalisatieproblemen met partiële differentiaalvergelijkingen als beperkingen. In het domein van de onzekerheidspropagatie ontwikkelen van numerieke technieken die toelaten om het effect te berekenen van onzekere of stochastische parameters op de oplossing van een wiskundig model (een partiële differentiaalvergelijking) (stochastische en fuzzy eindige-elementen methodes). (D. De Schreye, P. Dutré, P. Verbaeten, KULeuven, Computerwetenschappen)

- Capaciteit van ongeveer 1 Teraflop is zeker ruim voldoende voor PET scanner simulatie. (Medische instrumentatie DNTK/WE)
- Capaciteit nodig voor het lopende Abida-project voor validatie tests voor verschillende enzymsystemen in de orde van 2 Teraflops gedurende 1 week per modelsysteem, ofwel  $10 \times 7 \times 2 = 140$  Teraflops dagen. De ontwikkeling van meer modellen voor de voorspelling van resistentie tegen een tiental vaak voorgeschreven medicijnen van klinische relevante HIV mutanten zou  $10 \times 14 \times 4 = 640$  Teraflops dagen bedragen. De volledige 3D In-Silico geometrie screening van grote virtuele databases naar schatting 700 Teraflops dagen kosten voor 10M stoffen. Zulke 'multiple-low energy conformers' kunnen gebruikt worden om zeer snel een hoogwaardige in-silico screening van kandidaatstoffen voor nieuw geïdentificeerde targets uit de biotech industrie toe te laten. Het soort problemen (bijvoorbeeld proteïne-proteïne interacties, homologie modellen en dynamica optimalisatie van GPCR's) is beperkt door de hoeveelheid beschikbare rekenkracht. Aangezien deze methoden zeer schaalbaar ofwel 'embarrassingly parallel' zijn, kunnen deze problemen wel in een paar dagen tot een week op een 100 Teraflops machine te behandelen. (Molmo Services BVBA, Turnhout)
- Kwantumchemische berekeningen op zwaardere atomen of grotere moleculen, maar ook de toepassing van andere technieken zoals moleculaire dynamica (Car-Parinello,...). CFD berekenen voor geometrieën met behulp van grids met kleinere cellen en multifasesystemen in 2 of zelfs 3 dimensies. De uitbreiding van de bestaande capaciteit met de best beschikbare technologie zoals core 2 technologie of zelfs core 4 technologie. Een alternatief is het zogenaamde 'distributed computing'. (G. Heynderickx, Ugent, Lab. Petrochemische Techniek)
- Wetenschappelijke competitiviteit vereist aanschaf van een HPC-unit met 12 tot 24 processoren en minstens 4 tot 8 GB RAM geheugen per processor. (V. Van Speybroeck, M. Waroquier, UGent, VVS- CMM)
- Voor het modelleren van effecten op het elektriciteitsnet die verschillende tijds-grootteordes omvatten (omwille van de snelschakelende vermogenelektronica en variabiliteit van de opwekking), hoge resolutie 3D-modellen voor elektromagnetisch materiaalverwerken, tegelijk simuleren van ICT en energienetwerken, etc. er amper geparalleliseerde software ter beschikking die compatibel is met clusters of grids middleware. Er is nood aan krachtiger werkstations (ordegrootte Teraflop) met groot werkgeheugen (ordegrootte tientallen GB) en bijbehorende opslag (ordegrootte tientallen GB). (R. Belmans, KULeuven, Elektrotechniek Electa)
- Voor human observer studies wordt een infrastructuur uitgebouwd door de bestaande infrastructuur uit te breiden tot een werkbare hoeveelheid CPUs (grootte-orde hondertallen) of gelijkwaardig door participatie in GRID-initiatieven. (K. Deboschere, J. Van Campenhout, UGent, ELIS-MEDISIP)
- Voor parallelle implementatie van een 3D-stromingsmodel, spectraal golfmodel en sedimenttransportmodel zullen vermoedelijk 50 tot 100 CPU's nodig zijn om een naar huidige normen typische simulatie voor een echt kustgebied binnen een redelijke tijd af te handelen (minder dan 8 uur). (G. Degrande, J. Monbaliu, E. Toorman, KULeuven, Bouwkunde)
- Multischaal gas- en watertransport in plantaardige systemen van 10-6-10-1 m: stromingsproblemen met 10<sup>7</sup>-10<sup>8</sup> eindige elementen/volumes. Discrete-elementensimulatie van plantaardige systemen. (D. Berckmans, KULeuven, BIOSYST - MeBioS)
- Spraakherkenning met machines met 100 GB werkgeheugen vereisen voor parameterschatting en herkenning in de grootte-orde van 100 dagen/aantal processoren per experiment. (D. Van Compernelle, H. Van Hamme, P. Wambacq, KULeuven, Elektrotechniek - PSI)
- Verder onderzoek i.v.m. netwerkplanning en simulatie. Typisch een 100-tal CPU's. Geheugen is geen groot probleem. (P. Bienstman, D. Botteldooren, D. De Zutter, I. Lemahieu, UGent, INTEC)
- geavanceerde CAD-tools, die de volledige vergelijkingen van Maxwell in twee of drie dimensies moeten oplossen. bij het uitvoeren van 3D-simulaties is de vereiste rekenkracht aanzienlijk, die wellicht met grid computing kan aangepakt worden.

Vereisten 30 tal processoren en geheugenvereisten van 512 tot 1024 Mb per processor, occasioneel meer. Software onmiddellijk inzetbaar op het grid platform (Linux). (UGent)

- De beschikbare hoeveelheid sequentiedata stijgt sneller dan de wet van Moore. Ook de hoeveelheid andere biologische/bioinformatische data neemt exponentieel toe (o.m. de literatuur, alsook de hoeveelheid high-throughput experimenten en de hoeveelheid data die gegenereerd wordt per experiment). Verder zijn er meer complexe (en rekenintensieve) analysemethoden nodig om de huidige data te analyseren. Tenslotte verwachten we dat bioinformatica meer en meer een intrinsiek deel zal uitmaken van het experimenteel onderzoek (i.e. bioinformatici in elk experimenteel labo), dus we verwachten een sterke expansie van het domein. Om deze redenen verwachten we dat de nood aan HPC in de bioinformatica de volgende 5 jaren bijzonder sterk zal toenemen, sterker zelfs dan in andere domeinen. (Y. Moreau, KULeuven, Elektrotechniek -Bioinformatica)

c) *Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*

- Limiet korrelgrootte voor lage volumefracties is niet bekend en experimenteel moeilijk te bepalen. Deze resultaten kunnen gebruikt worden bij de ontwikkeling van nieuwe materialen om hun korrelgrootte te controleren. De limietkorrelgrootte heeft een grote invloed op de macroscopische eigenschappen van het materiaal. De huidige simulaties zijn gebaseerd op vereenvoudigde modellen en er zijn nog grote verschillen tussen experimentele data en simulatieresultaten. Door oriëntatie-afhankelijkheid en diffusie van atomen in rekening te brengen zullen de simulatieresultaten nauwkeuriger zijn. 100 Teraflops systeem zou toelaten om meer realistische modellen te simuleren met hogere nauwkeurigheid of bij de studie van stochastische aspecten of in een optimalisatieprocedure. (D. De Schreye, P. Dutré, P. Verbaeten, KULeuven, Computerwetenschappen)
- Ontwikkelen van de gereedschappen waarmee over tien jaar een groot farmaceutisch bedrijf zal werken, net zoals eerder ontwikkelde 1 Teraflops algoritmen nu bij elk groot bedrijf op de bestaande hardware inzetbaar is. HIV mutanten-screening zou een belangrijke bijdrage tot 'computational personalized medicine' kunnen leveren. (Molmo Services BVBA, Turnhout)
- Op het gebied van kwantumchemische berekeningen kunnen reactiepadanalyses gedetailleerder uitgevoerd worden en zijn dus minder vereenvoudigende aannames vereist. Zadelpunten in energieoppervlakken, die transitietoestanden voorstellen van elementaire stappen, kunnen sneller en exacter bepaald worden. Een reactiepadanalyse zoals deze voor de hydrogenering van benzeen, kan met behulp van HPC uitgevoerd worden in een termijn van minder dan één jaar.
- Voor CFD berekeningen opent HPC-perspectieven naar de gedetailleerde beschrijving van multifasesystemen in het algemeen en gas-vloeistofsystemen in het bijzonder. De hoge graad aan discontinuïteit in een gas-vloeistofstelsel, in het bijzonder bij koken, condenseren of verdampen, maakt de beschrijving en optimalisatie van de optredende stromingstypes mogelijk. (G. Heynderickx, Ugent, Lab. Petrochemische Techniek)
- Een accurate beschrijving van industriële heterogene katalyse toepassingen. Ontwerp van gefunctionaliseerde nanomaterialen die naast de actieve chemische sites ook nog eens voorzien zijn van grote poriën om de diffusielimiteringen naar de chemische sites te omzeilen, zoals het ontwerp van nanogestructureerde aluminosilicaten. Daarnaast de accurate beschrijving van reacties die zich afspelen in solventen. De organische reacties die beoogd worden zijn heel divers gaande van belangrijke homogene katalysatoren die gebruikt worden binnen de farmaceutische industrie als een efficiënte ontwikkeling van synthese routes voor moleculen met bio-activiteit. In de meeste recente literatuur wordt het solvent niet accuraat gemodelleerd en ook de bepaling van chemische kinetiek verdient aandacht. (V. Van Speybroeck, M. Waroquier, UGent, VVS- CMM)
- Realistische tomografische reconstructie en geautomatiseerde classificatie van ondergrondse voorwerpen of fouten gebaseerd op een bibliotheek; 3D stads modellering. (J. Cornelis, VU Brussel, ETRO)

- Bij iteratieve beeldreconstructie wordt het mogelijk om een patient-specifieke reconstructie alsook individuele dosimetrie van een pretherapeutische speurstof uit te voeren. Op die manier kan potentieel het aantal vals-positieven verlagen en/of het algemeen patientcomfort in radionuclidetherapie verhogen. (K. Debosschere, J. Van Campenhout, UGent, ELIS-MEDISIP)
- Betere schatting van de invloed van klimaatscenarios op de fysische processen van de stromingen en golven en sedimentatiedynamica door een studie van de gekoppelde systemen voor de studie van kustgebieden die tot de meest kwetsbare gebieden behoren (G. Degrande, J. Monbaliu, E. Toorman, KULeuven, Bouwkunde)
- Het modelleren van de fysiologie (biofysica en biochemie) van volledige organismen of macroscopische onderdelen hiervan (bv. vruchten) zou een belangrijke doorbraak betekenen in computational biology en systems biology. (D. Berckmans, KULeuven, BIOSYST - MeBioS)

*d) Gewenste ondersteuning*

- Rechtstreekse toegang tot knopen voor optimaal gebruik (zelf kunnen bepalen op welke knoop een bepaalde deelopdracht uitgevoerd zal worden)
- Begeleiding en ondersteuning door een administratorteam met ervaring in parallel programmeren en wetenschappelijk rekenen op de HPC-infrastructuur. (D. De Schreye, P. Dutré, P. Verbaeten, KULeuven, Computerwetenschappen)
- De aanbieder van HPC zal de nodige informatie voor het correct gebruiken van de infrastructuur aan alle betrokken onderzoekers meedelen. (D. Aeyels, G. De Cooman, UGent, Elektrische energie, Systemen en Automatisering (EESA))
- 'Audit' van de rekefficiëntie van deze codes door een onafhankelijke expert een meerwaarde kunnen bieden. (G. Heynderickx, Ugent, Lab. Petrochemische Techniek)
- Nodige ondersteuning om de computercodes op een zo efficiënt mogelijke manier te compileren en uit te voeren; taken voor personeel dat heel erg thuis is in de verschillende architecturen en compilers en parallelisatie.
- Intensieve interactie tussen systeembeheerders en gebruikers. Voor moleculaire modelleringstoepassingen is het wenselijk dat één van de systeembeheerders ervaring heeft met het werken met pakketten eigen aan het onderzoeksveld. Zo niet is het rendement van dergelijke infrastructuur zeer twijfelachtig. (V. Van Speybroeck, M. Waroquier, UGent, VVS- CMM)
- Transparante en snelle procedure voor de aanvraag van rekentijd (Molmo Services BVBA, Turnhout)
- Software-ingenieur moet het omzetten van de prototypeprogramma's in matlab naar C++ op een efficiënte manier bijstaan. Er is tevens een behoefte aan codeoptimalisering .
- Nood aan software die parallelisatie vergemakkelijkt zoals een Parallele compiler.
- Nood aan goed ontworpen softwarepakketten rond optimalisering, numerieke methodes hoofdzakelijk om Partieel Differentiaal vergelijkingen op te lossen De huidige commerciële modellerings- en simulatiesoftware is niet geparalleliseerd, en moet HPC-ready gemaakt worden. Het zelf blijven ontwikkelen van eigen parallelle software gaat echter ten koste van het core-onderzoek. (J. Cornelis, VUBrussel, ETRO)
- Logistieke ondersteuning professionaliseren van deze submittie- en recollectie-brokering. (K. Debosschere, J. Van Campenhout, UGent, ELIS-MEDISIP)
- Ondersteuning bij overdracht van bestaande codes (zowel eigen ontwikkeling of ontwikkeling door andere wetenschappelijke instellingen en al dan niet reeds geparalleliseerd) naar HPC-infrastructuur is absoluut noodzakelijk. Veel 'kleine' onderzoekslabo's hebben niet de nodige expertise in huis om dit zelf te doen. Dergelijke expertise is ook maar voor beperkte tijd echt nodig zoals bij implementatie/optimalisatie van nieuwe modellen of modules. (G. Degrande, J. Monbaliu, E. Toorman, KULeuven, Bouwkunde)
- Ondersteuning is nodig voor het koppelen van Comsol Multiphysics met snelle, parallelle solvers (b.v., AMG). (D. Berckmans, KULeuven, BIOSYST - MeBioS)
- Goede documentatie, cursussen en debuggers/tools voor MPI-type libraries zouden de parallelisatie-effort aanzienlijk versnellen. (D. Van Compernelle, H. Van Hamme, P. Wambacq, KULeuven, Elektrotechniek - PSI)



### **Aanbevelingen en Verwachtingen**

- Nood aan uitgebreide informatiesessies en lessenreeksen over parallel programmeren en wetenschappelijk rekenen, en workshops waarin problemen en oplossingen uitgewisseld en besproken worden, en uitgebreide uitleg en een overzicht van commando's op een website (wikipedia) die regelmatig geupdate worden. (D. De Schreye, P. Dutré, P. Verbaeten, KULeuven, Computerwetenschappen)
- Een transparante en snelle procedure voor de aanvraag van rekentijd. We hebben soms de ervaring dat de tijd die nodig is om rekentijd toegekend te krijgen niet in verhouding staat tot de toegekende rekentijd zelf.
- Suggestie om de configuratie van het systeem op gezette tijden (bijv. om de x maanden) te wisselen tussen single-multiple system images (en natuurlijk de projectplanning daarop aan te passen). De HPC-systeembouwers software gereedschappen zijn beschikbaar voor een soepele configuratie wisseling. (D. Aeyels, G. De Cooman, UGent, Elektrische energie, Systemen en Automatisering (EESA))
- Niet alleen is een krachtige HPC-omgeving noodzakelijk, maar het is minstens even belangrijk dat deze gemakkelijk bruikbaar is voor niet HPC-specialisten. (R. Belmans, KULeuven, Elektrotechniek Electa)
- In kaart brengen van de verschillende lokale, Vlaamse, Belgische, Europese en mondiale GRID-initiatieven en een inventarisering maken van de betrokken Vlaamse universiteiten en onderzoekscentra om aldus te komen tot een betere collaboratie. Dit vooral voor groepen die HPC aanwenden eerder dan ze als onderzoeksdomein te hebben. (K. Deboschere, J. Van Campenhout, UGent, Elektronica en Informatiesystemen - MEDISIP)
- HPC en GRID-computing kan niet omschreven worden door alleen maar "aantal Teraflops". Beschikbaar RAM geheugen en bandbreedte minstens even belangrijk. (D. Van Compennolle, H. Van Hamme, P. Wambacq, KULeuven, Elektrotechniek - PSI)

### **Bijlage 3d: Moleculaire Chemie**

Prof. Norbert De Kimpe en Prof. Renaat Gijbels

#### **Research Richtingen**

- Theoretische berekeningen van moleculaire systemen met als doel fysische en chemische eigenschappen te voorspellen en te verklaren op moleculaire basis. Modellen met hoge accuratesse (M. Waroquier, V. Van Speybroeck, UGent, K. K. Pierloot et al., KULeuven).
- Moleculaire modellering, theoretische chemie en quantumchemie (QC), meer bepaald theoretische en toegepaste aspecten van moleculaire computationele en conceptuele Density Functional Theory, DFT. (P. Geerlings, VU Brussel).
- Zeer nauwkeurige (onzekerheid <1 kJ/mol) ab initio computationele thermochemie; ontwikkeling van nieuwe functionalen in DFT met bijzondere nadruk op reactiebarrières; toepassingen van DFT voor reactiemechanismen; berekening van moleculaire krachtvelden voorbij de harmonische benadering (J. Martin, vroeger U Hasselt en U Antwerpen, thans Weizmann Institute of Science).
- Diverse QC-methoden, waaronder multiconfiguratie methoden en DFT toegepast op grote moleculaire systemen (grote rekencapaciteit nodig!); Algebraic Diagrammatic Construction scheme (ADC(3)) voor de studie van ionisatieverschijnselen en -spectroscopie. (J.P. François, M. Deleuze, U Hasselt).
- Ontwikkelen van methodologieën en bijhorende software voor QC-studie van grote moleculen (relevant voor farmacie en biochemie), clusters van moleculen en atomen en van moleculaire kristallen. (C. Van Alsenoy, U Antwerpen)
- Studie van zwak-gebonden complexen tussen Ar atomen en kleine moleculen in cryovoorwaarden. (W. Herrebout, U Antwerpen)
- Modelleren van de elektronenstructuur van materialen; onderzoek van nieuwe katalysatoren; ontwikkelen van gefunctionaliseerde materialen; chemische reacties die zich afspelen in solventen. (M. Waroquier, V. Van Speybroeck, UGent).
- Chemische reactiekinetiek en heterogene katalyse, vooral adequate simulatie en optimalisering van op grote schaal toegepaste productieprocessen. (J. Thybaut, UGent)
- Berekening en interpretatie van descriptoren van reactiviteit van heterogene katalysatoren; berekening van transitietoestanden en reactiepaden van heterogeen gekatalyseerde reacties; berekening van conformatie en activiteit van geadsorbeerde proteïnen. (R. Schoonheydt, KULeuven i.s.m. UGent en VU Brussel)
- Contract onderzoek op gebied van computationele chemie en medicijnontwerp; ontwikkeling van HPC-toepassingen in chemie/farmacie. (MolMo Services bvba, Turnhout).
- Softcomputing technieken, die meer robuuste en betrouwbare resultaten leveren. (L. Boullart, UGent)

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen*

- QC-berekeningen op een computercluster met totale rekencapaciteit van ca 45 Gigaflops. (K. Pierloot, KULeuven).
- Grote moleculaire systemen vergen grote rekencapaciteit (totaal thans 1200Gb schijfruimte, 30 Gb geheugen beschikbaar). ADC(3) schema voor ionisatieverschijnselen en -spectroscopie is computationeel veeleisend, laagste eigenwaarden en corresponderende vectoren b.v. vergen berekening van  $10^4 \times 10^4$  matrices. (J.P. François, M. Deleuze, U Hasselt).
- Verschillende jobs QCI(SDT) en MP4 berekeningen aan S2N2CO liepen 2/3 maanden op 1 processor van CalcUA maar konden niet met succes beëindigd worden wegens problemen met storage en CPU tijd. (C. Van Alsenoy, U Antwerpen).
- Berekening van zeer exacte interactie-energieën tussen moleculen met behulp van ab initio rekenmethoden op CalcUA gaf zeer goede overeenkomst met experimentele cryospectroscopie. (W. Herrebout, U Antwerpen)

- J. Martin (Weizmann Institute of Science) geeft gedetailleerde uitleg, die ook in Vlaanderen zeer relevant is, over de eigen Linux farm, voor 2 types berekeningen: (a) met zeer hoge eisen aan I/O bandbreedte (op maat gemaakte machine, 4 dual-core Opteron CPU's, 16 Gb RAM geheugen, 2 Terabyte tijdelijke disk, met speciale I/O controllers), en (b) CPU intensieve berekeningen. (dual Xeon machines, eerlang te vervangen door quad-core Intel CPUs)
- De aanwezige reeds grote reken capaciteit (2 Altix rekenunits met elk 12 Itanium processoren en 2 GB RAM geheugen per processor wordt als te klein ervaren (M. Waroquier, V. Van Speybroeck, UGent)
- In silico analyse van klinisch geïsoleerde HIV-protease varianten op resistentie tegen AIDS op 216 CPU Xeon cluster van J&J in Raritan NJ, USA (ca 4,5 Teraflops). (MolMo, Turnhout).
- Analyse van zeer hoog-dimensionele microarray data i.s.m. de vakgroep Moleculaire Genetica, en klassificatie van tapijstalen voor het bepalen van de slijtagegraad, IBM Blade-Center chassis, IBM X335, 40 dual CPU computing blades; Linux RH3, Sun Gridware. (L. Boullart, UGent)

*b) Plannen voor de komende 5 jaren voor HPC-gebruik*

- Algemeen: Computatieve chemie is een domein met een 'moving horizon'. Men wil enerzijds redelijke nauwkeurigheden op steeds grotere systemen (met steeds meer vraag naar biomoleculen, in het bijzonder enzymen, en nanotech toepassingen). Aan het andere eind van de schaal, voor kleine moleculen, wil men steeds grotere nauwkeurigheden.
- Studie van steeds grotere moleculaire systemen, beter overeenkomend met chemisch realistische systemen; toepassingen in katalyse, biomoleculen, ook de tijdsevolutie van moleculaire systemen zal een grotere rol gaan spelen. (P. Geerlings, VU Brussel)
- Steeds accuratere berekeningen aan kleine moleculen. (P. Geerlings, VU Brussel)
- Aanpassing van het in-house ontwikkelde programmapakket BRABO waaronder implementatie van parallelle processing, bv MPI, om routinematig state-of-the art berekeningen te kunnen doen aan moleculen met 1000 atomen. (C. Van Alsenoy, UAntwerpen)
- Realistische theoretische modellen met predictief karakter vereisen aanschaf van een HPC-unit met 12-24 processoren, en min. 4-8 Gb RAM geheugen per processor. (M. Waroquier, UGent).
- QC en CFD berekeningen van multifasesystemen in 2 of 3 dimensies. (J. Thybaut, UGent).
- Ontwikkeling van meer modellen ter voorspelling van resistentie tegen ca 10 medicijnen van klinisch relevante HIV mutanten zou 640 Teraflops dagen bedragen; dit project staat sinds 2004 in de wachtkamer tot voldoende capaciteit beschikbaar is. (MolMo, Turnhout)
- Volledige 3D in silico geometrie-screening van grote virtuele databases (dit zou 700 Teraflops dagen kosten) (screening van kandidaatstoffen voor nieuw geïdentificeerde targets uit biotech industrie). (MolMo, Turnhout)
- Studie van proteïne-proteïne interacties zijn beperkt door de hoeveelheid beschikbare rekenkracht. (MolMo, Turnhout)
- Computatieve modellen voor de studie van EEG-signalen en lokale veldpotentialen van diepe breinstimulatie. (L. Boullart, UGent)

*c) Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*

- Berekening van grote systemen en belangrijke vergroting van de dimensie van de systemen; ook uitvoeren van moleculaire dynamica met langere simulatietijden ten einde een goede kans te maken op het observeren van chemisch interessante fenomenen. (P. Geerlings, VU Brussel)
- Berekeningen op ADC(4) niveau van ionisatiefenomenen. Voor welke systemen deze code zal beschikbaar zijn is nog niet te voorspellen. (J.P. François, M. Deleuze, U Hasselt)
- Gesofisticeerde post-Hartree-Fock methoden, zoals MP4(SDQ) e.a., toegepast op stoffen zoals 5-oxo-1,3,2,4-dithiazool e.a. eisen erg veel van de computatieve

infrastructuur. Zaken als disk space (bv 700 GB/file) en snelheid (bv 11 maanden rekentijd en langer) zijn de beperkende factoren in dit onderzoek. Elke aanzienlijke uitbreiding van de hardware kan dit onderzoek een stap verder brengen. (C. Van Alsenoy, UAntwerpen)

- Bij simulaties van reacties aan dit soort verbindingen met Carr-Parinello moleculaire dynamica komen de huidige computationele beperkingen naar boven (per simulatie verschillende weken CPU tijd vereist). Ook hier is uitbreiding van de hardware een vooruitgang. (C. Van Alsenoy, UAntwerpen)
- Berekening van NMR chemical shifts voor grote biologisch actieve moleculen zoals peptiden vereist eveneens hoogstaande HPC-infrastructuur. (C. Van Alsenoy, UAntwerpen)
- Realistische beschrijving van solute-solvent, solvent-solvent en solute-solute interacties zowel via Monte Carlo als Moleculaire Dynamica berekeningen, in cryogene Ar oplossingen. (W. Herrebout, UAntwerpen)
- Conformatie van geadsorbeerde proteïnen en hun dynamica aan oppervlakken alsook reactiepad en transitietoestanden van heterogeen gekatalyseerde reacties met zeolietkristallen als katalysator, dit via Carr-Parinello, VASP. (R. Schoonheydt, KULeuven i.s.m. UGent en VUBrussel)
- Realistische voorspellingen aan macroscopische fenomenen, ontwerp van nanogestructureerde aluminosilicaten, beschrijving van reacties in solventen, en ontwikkeling van de synthese van moleculen met bioactiviteit vergen alle massieve rekentijd. (M. Waroquier, UGent)
- Gedetailleerde reactie-pad analyse, en gedetailleerde beschrijving van multifase-systemen (gas/vloeistof). (J. Thybaut, UGent)
- Ontwikkelen van de tools waarmee over 10 jaar de farmaceutische industrie zal werken. (MoMo, Turnhout)
- Bronlocalisatie van EEG signalen voor een brein-computer interface. (L. Boullart, UGent)
- Ordinale klassificatie van tapijstalen door analyse van 3D-laserscanner data en visuele beelden in real time. (L. Boullart, UGent)
- Baanbrekend werk is binnen het domein van de onderzoeksgroep vaak niet zozeer een kwestie van rekensnelheid dan wel van "kort op de bal spelen"; hiertoe is een kleinschaliger park, met speciale aandacht voor de particuliere noden van elke individuele gebruiker, onontbeerlijk. (K. Pierloot, A. Ceulemans, G. Hendrickx, M. T. Nguyen, H. Vansweevelt, KULeuven)

d) *Gewenste ondersteuning*

- Het doorvoeren van intensieve parallelisatie is vooralsnog lastig met commerciële codes die vaak in de QC gebruikt worden. Deze codes zijn tot op grote hoogte reeds geparallelliseerd. Het paralleliseren van een eigen ontwikkelde code is hoog specialistisch werk waarvoor ondersteuning zou moeten kunnen aangeboden worden. (P. Geerlings, VUBrussel)
- Hulp bij implementatie van standaard parallelisatie software (MPI) en lineaire algebra (eigenwaarden/eigenvectoren) is zeer gewenst. (C. Van Alsenoy, UAntwerpen)
- J. Martin (Weizmann) werkt uitsluitend met SMP (shared memory parallelisme) via OpenMP of Global Arrays; zijn ervaringen met MPI of PVM zijn erg teleurstellend.
- Recente ervaringen binnen de CalcUA gebruikersgroep tonen aan dat vele chemici zelf niet in staat zijn om door hen geschreven codes te optimaliseren en optimaal gebruik te maken van de verschillende parallelisatie omgevingen. Begeleiding van de onderzoekers is dus zeker belangrijk. (W. Herrebout, UAntwerpen)
- Nodige ondersteuning vereist. Intensieve interactie tussen systeembeheerders en -gebruikers is essentieel. Voor moleculaire modelleringstoepassingen is het zelfs wenselijk dat één van de systeembeheerders ervaring heeft met het werken met pakketten eigen aan de discipline. Zoniet is het rendement van dergelijke infrastructuur zeer twijfelachtig. (M. Waroquier, UGent)
- 'Audit' van de rekefficiëntie van zelf ontwikkelde computercodes door onafhankelijk expert is gewenst. (J. Thybaut, UGent)

- De nodige informatie over het gebruik van de infrastructuur aan de onderzoekers meedelen. (L. Boullart, UGent)
- Beperkte behoefte voor MolMo, Turnhout.

#### **Aanbevelingen en verwachtingen**

- Naast de kostprijs van een HPC-systeem, moet rekening gehouden worden met het onderhoud van het systeem (hard- en software), personeelsomkadering, eventuele omzetting van de momenteel gebruikte software: wie zal dit doen? Leden van de onderzoeksgroep of personeel van het HPC-systeem? Zullen de universiteiten die er gebruik van maken financieel moeten bijdragen? Zullen er zoals in het verleden FWO rekenkredieten ter beschikking gesteld worden? (J.P. François, UHasselt)
- De afdeling Fysische Chemie en Kwantumchemie beschikt over een computer-cluster (ca 45 Gigaflops). Hoewel het voorstel voor HPC-infrastructuur van ca 100 Teraflops aantrekkelijk uitziend, verkiest de groep te blijven investeren in de verdere uitbouw van hun (relatief) kleinschaliger computerpark na ervaring met super-computerparken (verlies aan flexibiliteit en alertie (snelheid van disk access, centrale beheerregels)). Wanneer een investering op grote schaal voorzien wordt, moet er terdege rekening mee gehouden worden dat een eenmalige impuls-financiering niet volstaat, en dat er grote recurrente kosten moeten worden begroot qua personeel, aangepaste airco-ruimte, en regelmatige update-vereisten. Indien deze kosten worden doorgeschoven ten laste van de gebruikers is het gevaar reëel dat het computationeel onderzoek in Vlaanderen daarvan nadeel zal ondervinden. (K. Pierloot et al., KULeuven)
- Er is de laatste jaren een trend naar "grid-computing". Voor het doorvoeren van parallelisatie valt nog steeds een cluster configuratie te verkiezen. (P. Geerlings, VUBrussel)
- Ooit waren grote shared-memory machines het aangewezen platform voor bepaalde specifieke HPC-toepassingen (zoals CFD). De kostprijs per gigaflop van deze 'boutique technology' is echter enorm veel hoger dan die van een cluster, en recente ontwikkelingen (goedkope 64-bitprocessors, relatief goedkope RAM, betaalbare low-latency netwerktechnologieën zoals Myrinet of Infiniband...) hebben clusters veel CFD vriendelijker gemaakt. Om die reden wordt het steeds moeilijker de aanschaf van peperdure shared-memory machines te rechtvaardigen als een veel grotere geaggregeerde performantie voor hetzelfde geld met een cluster kan bereikt worden. In het bijzonder de beschikbaarheid van relatief goedkope clustereenheden met 8 verwerkingseenheden (dankzij multicore-CPUs zoals de Intel Clovertown en zeer binnenkort zijn AMD-tegenhanger) biedt een zeer aantrekkelijk compromis. (J. Martin, Weizmann).
- Transparante en snelle procedure voor de aanvraag van rekestijd, alsook wisselen van configuratie van single naar multiple system images (kleine clusters versus ganse apparatuur). (MolMo, Turnhout).

### **Bijlage 3e: Toegepaste Wiskunde en Mathematische Algoritmen**

Prof. Joseph. Thas

#### **Research Richtingen**

- Ontwikkeling van nieuwe statistische modellen voor het analyseren van complexe gegevens. (K. Bogaerts, KULeuven)
- Zuivere Wiskunde: verbanden tussen groepentheorie en meetkunde (constructie van verschillende meetkonden voor een gegeven groep), berekening van Ramsey-getallen (G. Brinkmann, UGent; P. Cara, VUBrussel)
- Similariteits- en inclusiematen en hun toepassing binnen de numerieke taxonomie, Computertoepassingen voor taxonomie en identificatie van microorganismen,
- Gedistribueerde informatiesystemen: StrainInfo.net bioportal, (P. Dawyndt, UGent)
- Toepassing van Bayesiaanse mengmodellen voor multilocus sequentieanalyse (in het genus Burkholderia), (P. Dawyndt, UGent).
- Toepassing van machine learning algoritmen voor gedistribueerde identificatie. (P. Dawyndt, UGent)
- Incidentiemeetkunde: studie van (eindige) projectieve ruimten, polaire ruimten en veralgemeende vierhoeken (voornamelijk het zoeken naar bepaalde deelstructuren). (J. De Beule, UGent)
- Interdisciplinair onderzoek in de levenswetenschappen: relatie tussen eiwit functie en dynamica en hoe deze relatie aan de grondslag ligt van menselijke ziektes (Alzheimer, Parkinson, kanker). Interdisciplinaire aanpak die in silico modellering combineert met zowel in vitro biophysische studies als met basis cell biologische observaties. (T. Lenaerts, VUBrussel/VIB)
- Plasma astrofysica: (schok-)golven, instabiliteiten, verhitting en versnelling van plasma's op aarde, in het zonnestelsel en ver daarbuiten. Opstellen van wiskundige modellen die de dynamica en de interactie van plasma's met magnetische velden beschrijven. De wiskundige modellen gebruiken zowel de magnetohydro-dynamica als de kinetische theorie, of een combinatie van beide. (S. Poedts, KULeuven)
- Theoretische scheikunde: substructuren van fullerenen. (G. Brinkmann, UGent)

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen*

- Statistische modellen: simulaties om de nieuw ontwikkelde methodes te testen op hun performantie. Verbruikte rekencapaciteit op ongeveer 1 jaar: 5 jaar CPU time met meer dan 7000 jobs. (K. Bogaerts, KULeuven)
- Berekenen van de deelgroepentrialies van gegeven groepen. Experimenten uitgevoerd op een kleine lokale cluster van 16 computers aan de VUBrussel. Grote hindernis is geheueentekort. (P. Cara, VUBrussel).
- Er werd een gemeenschappelijke rekencluster opgezet door de Vakgroep Toegepaste Wiskunde en Informatica, de onderzoeksgroep Kennisgebaseerde systemen en het Laboratorium voor Microbiologie (UGent). De architectuur van deze cluster is gebaseerd op INTEL Blade Server technologie: een Linux cluster met 14 nodes, elk met 4 Dual-Core Intel Xeon CPU 3GHz processoren en 1 Gb RAM aan boord. (P. Dawyndt, B. De Baets, P. Vandamme, P. De Vos, UGent)
- Incidentiemeetkunde: zoeken naar bepaalde deelstructuren via exhaustieve methoden. Hardware die gebruikt wordt bestaat grosso modo uit een dertigtal processoren, aan 3 GHz elk. (J. De Beule, UGent)
- Eiwitonderzoek: een belangrijke beperking op de voortgang van het huidige onderzoek is het gebrek aan voldoende rekencapaciteit. (T. Lenaerts, VIB-VUBrussel)
- Plasma astrofysica (CPA): zelf ontwikkelde computerprogramma's. COOLFluid (Computational Object Oriented Library for Fluid Dynamics), ontwikkeld door CPA en von Karman-Instituut, geschreven in C++. Typische runs: VAC: enkele uren tot enkele dagen op 48 CPU's, AMRVAC: 4 dagen op 8-way node. Er wordt momenteel

een 3D VAC-simulatie gepland die 10 dagen zal duren op 480 CPU's. (S. Poedts, KULeuven)

- Combinatorische wiskunde en problemen uit theoretische scheikunde: rekencluster met 100 machines in Bielefeld, met meer dan 500 machines in Australië. Allemaal linux met automountsysteem. (G. Brinkmann, UGent).

*b) Plannen voor de komende 5 jaren voor HPC-gebruik*

- Statistische modellen: simulaties om de nieuw ontwikkelde methodes te testen op hun performantie. (K. Bogaerts, KULeuven)
- Enumeratie problemen in de combinatorische wiskunde (codeertheorie, grafentheorie, incidentiemeetkunde,...). Reken capaciteit is niet kritisch, maar de benodigde geheugenruimte moet liefst groter zijn dan 6 GB per processor. Reeds experimenten uitgevoerd op BEgrid met 128 processoren (technische drempel zeer hoog). Voor bepaalde berekeningen wordt ook een cluster met meer dan 500 machines in Australië gebruikt. (P. Cara, VUBrussel; G. Brinkmann, UGent)
- Computertoepassingen: verdubbeling van de bestaande reken capaciteit zou moeten volstaan voor de volgende 5 jaar (1/1000 van de voorgestelde capaciteit) (P. Dawyndt, UGent)
- Incidentiemeetkunde: een 200 tal processoren voldoende om de volgende stap te zetten. Werkgeheugen belangrijker dan processoren: 4 Gb per processor is absoluut minimum. (J. De Beule, UGent)
- SWITCH lab heeft nood aan cluster met gemiddeld 10 nodes per onderzoeker. Momenteel maken zij gebruik van clusters van anderen (NORDIC, EMBL, CALCUA). (T. Lenaerts, VUBrussel/VIB)
- Ruimteweervoorspellingen (essentieel voor AURORA programma van ESA): rekenkracht moet minstens 50 keer groter worden in de volgende 5 jaren, dwz. 100-200 Teraflops. (S. Poedts, KULeuven)

*c) Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*

- Verfijnen van de klassieke bio-informatica technieken vanuit statistisch oogpunt. (K. Bogaerts, KULeuven)
- Berekenen van volledige lijsten van meetkunden voor de meeste sporadische enkelvoudige groepen. (P. Cara, VUBrussel)
- Bepaling van de sequentie specificiteit van amyloidosis. Grote verbetering van de voorspelbaarheid van amyloidose (het fenomeen ligt aan de basis van Amyotrophic Lateral Sclerosis, cataract en prion ziektes)
- Modelleren van de sequentie specificiteit van chaperones (T. Lenaerts, VUBrussel/VIB)
- Beste (meest geavanceerde) numerieke simulaties van coronale massa-ejecties en CME schokgolven ter wereld. (S. Poedts, KULeuven)
- In rekening brengen van deeltjes-versnelling en multicomponent plasma's waardoor de simulaties nog realistischer worden (S. Poedts, KULeuven)

*d) Gewenste ondersteuning*

- Grote nood aan ondersteuning van parallelisatie. Een gebruiksvriendelijke "middleware" zou moeten ontwikkeld worden. Ook load balancing en de distributie van processen zou vereenvoudigd moeten worden. Onderhoudskosten en energieverbruik van een HPC-installatie mag niet onderschat worden. (P. Cara, VUBrussel)
- Opzet misschien te hoog gegrepen als er niet voldoende technische en budgettaire ondersteuning is. Op vlak van algoritmieken lijkt het ondenkbaar dat men op gecentraliseerde wijze de hedendaagse onderzoeker ondersteuning kan bieden. (P. Dawyndt, UGent)

**Aanbevelingen en verwachtingen**

- Er zijn reeds ervaringen met BEgrid. Misschien kunnen die ook helpen bij het inschatten van de noden wanneer men wil paralleliseren. De technische obstakels blijken groter dan verwacht. (P. Cara, VUBrussel)
- Hele problematiek rond toegang, beschikbare capaciteit, toegelaten programma-tuur, confidentialiteit van gegevens, rechtmatig gebruik, mag niet over het hoofd

gezien worden. BEgrid mag men niet uit het oog verliezen. Reeds jaren doet men daar inspanningen om via massale samenwerking reken capaciteit te genereren, en zijn ondertussen de verschillende bottlenecks genoegzaam gekend. Het lijkt dan ook niet verstandig een tweede gelijkaardig initiatief op te starten. Het is belangrijk om de gebruikers van een cluster dicht genoeg bij elkaar te brengen, zodat zij voldoende ervaringen kunnen uitwisselen. De beschikbaarheid van veel rekenkracht lijdt immers snel tot laksheid. Voor ons volstaat de eigen rekencluster (verhoogde capaciteit weegt niet op tegen praktische complicaties, echte werk zit in de man-power (algoritmen) niet in de machine-power). (P. Dawyndt, UGent)

- Goed management van HPC-systeem is cruciaal. Aanwenden van een systeem mag niet ingewikkelder zijn dan het aanwenden van software waarmee de gebruiker vertrouwd is. (J. De Beule, UGent)
- Een grote centrale computercluster is zonder meer nodig in Vlaanderen. Ik zou de bestaande infrastructuur van de KULeuven (4Teraflops) uitbreiden om de investeringen maximaal te laten renderen. Het is belangrijk één grote cluster te bouwen en niet 3 of 4 kleinere. (S. Poedts, KULeuven)
- Gebruik de computers op een verstandige manier (efficiënt: allemaal linux met een automountsysteem). (G. Brinkmann, UGent)



### **Bijlage 3f: Astrofysica en Kosmologie**

Prof. Christoffel Waelkens

#### **Belangrijkste Research Richtingen**

- Structuur, evolutie en dynamica van sterrenstelsels.
- Modelleren van astrofysische plasma's.
- Structuur, evolutie en samenstelling van sterren.

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen.*

Bijdragen van M. Baes, S. De Rijcke (UGent);

C. Aerts, L. Decin, R. Keppens, S. Poedts (KULeuven)

- Monte Carlo stralingsoverdracht simulaties van sterrenstelsels, circumstellaire stofschijven en actieve kernen van melkwegstelsels. Met een eigen cluster van 15 processoren met telkens 2 GB RAM geheugen.
- Kosmologische simulaties van de vorming van dwergmelkwegstelsels en ab initio berekeningen van hun observationele eigenschappen. Met een eigen cluster van 16 processoren met telkens 2 GB RAM geheugen.
- Berekeningen van 2D, 2,5D en 3D magnetohydrodynamische codes ontwikkeld in samenwerking met het von-Karman-Instituut en het FOM (NL), zowel in een parallelle versie als in een versie met adaptieve roosterverfijning. De berekeningen vereisen typisch enkele dagen op de 48 CPUs of met de 8-way node van de KULeuven HPC-cluster.
- Modelleren van tijdreeksen van data van pulserende sterren. Automatische classificatie van veranderlijke sterren uit steekproeven van tijdsreeksen.
- Stralingstransfert codes voor uitgebreide steratmosferen.

##### *b) Plannen voor de komende 5 jaren voor HPC-gebruik.*

- Monte Carlo stralingsoverdrachtssimulaties van sterrenstelsels, circumstellaire stofschijven en actieve kernen van melkwegstelsels, in volledig vrije 3D configuraties, met afwijkingen van thermisch evenwicht. Koppeling van stralingstransfert en hydrodynamische simulaties. Vereist een geheugen van de orde van 100GB en grotere kloksnelheid.
- Kosmologische simulaties van de vorming van dwergmelkwegstelsels en ab initio berekeningen van hun observationele eigenschappen uitbreiden door de programma's te optimaliseren en de huidige cluster uit te breiden met de 36 processoren van een andere onderzoeksgroep.
- Ruimtelijke toepassingen van de magneto-hydrodynamische simulaties van de zonnewind, met zijn instabiliteiten en de interactie met de magnetosfeer van de aarde. Maatschappelijk relevante voorspellingen vereisen een rekenkracht van 100-200 Teraflops (i.e. bijna twee grootteordes boven hetgeen thans beschikbaar is).
- Automatische classificatie van veranderlijke sterren uit steekproeven van tijdsreeksen voor de data van de satellieten CoRoT (2007, weken rekentijd op een 50 Gflops cluster) en Gaia (2012, maanden rekentijd op een 250 Gflops cluster).
- Modelleren van multispectrale (Herschel, APEX, ALMA) data van de massa-uitstromingen van sterren in late evolutiestadia. Modelleren van de sterrenwinden van sterren van grote massa. Uitbreiding van huidige capaciteit van 350 Gflops tot 3,5 Teraflops is belangrijk om de incorporatie van fysische effecten en de grid-dichtheid te verfijnen.

##### *c) Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*

- Een dergelijke infrastructuur biedt de mogelijkheid van zelfconsistent modelleren van de verdeling en samenstelling van het interstellair midden in sterrenstelsels.
- Kosmologische simulaties met een deeltjesmassa die voldoende klein is, zodat – in tegenstelling met de Millenium Survey – ook relevante voorspellingen zouden volgen voor de kleinere sterrenstelsels, die in aantal dominant zijn in het heelal.

- Wat MHD-simulaties betreft: de claim is dat de Europese rekentechnieken superieur zijn, maar dat de onderzoekers in de VS over veel sterkere apparatuur beschikken. Met andere woorden: schaalvergroting is een sleutel tot een leidende rol in onderzoek rond ruimteweer.
  - Snelle analyse en interpretatie van data met satellieten als CoRoT (asterseismologie) en Herschel (stereolutie, circumstellaire media) wordt mogelijk voor een prijs die hoe dan ook veel kleiner is dan de kostprijs van die satellieten.
- d) *Gewenste ondersteuning*
- De sterrenkundigen lijken er vrij gerust in te zijn dat zij grotere apparatuur aankunnen!

### **Bijlage 3g: Subatomaire Fysica**

Prof. Kris Heyde

- Hoge-energiefysica (nucleon structuur, deeltjesfysica, astro-deeltjesfysica)
- Experimentele subatomaire fysica: kernfysica
- Modelling van veeldeeltjessystemen
- Modelling van golven in tokamak plasmas
- Onderdeel van fusieonderzoek (groep Toegepaste Fysica van UGent, in samenwerking met KMS, Brussel) is ook in dit overzicht opgenomen (met \* gemarkeerd)

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen.*

- Simulaties in het IIHE (<http://w3.iihe.ac.be>) voor scheiding signaal-achtergrond (CMS) in deeltjesfysica op een cluster met  $\approx 50$  CPUs. Uitgebreid met een grid infrastructuur bestaande uit 120 CPUs. Het grid is operationeel sedert de zomer 2005. Voor het IceCube experiment werd een cluster van 25 CPUs gebruikt, (C. De Clercq, W. Van Doninck, S. Tavernier; UAntwerpen: E. De Wolf, VU Brussel).
- Gebruik van een OpenPBS Linux PC-cluster bestaande uit een tiental batch nodes door de onderzoeksgroep in Gent voor analyse van HERMES en IceCube experimenten. De GRID-infrastructuur van de Universiteit Gent bestaat momenteel uit ongeveer 54 dual-core 1,6-2,8 GHz nodes met een totale opslagcapaciteit van 2Tb. (D. Ryckbosch, UGent)
- Sporadisch gebruik van HPC door experimentele subatomaire kernfysica groep van de KULeuven. (M. Huyse, P. Van Duppen, N. Severijns, G. Neyens)
- Gebruik van een 10 processor-cluster in Gent voor Monte-Carlo simulaties in veeldeeltjessystemen. Voor de meest nauwkeurige berekeningen wordt momenteel nog beroep gedaan op een cluster van meer dan 100 processoren in het buitenland (ETH Zürich, UC Berkeley). (S. Rombouts, UGent)

##### *b) Plannen voor de komende 5 jaren voor HPC-gebruik*

- In het kader van het CMS experiment wordt wereldwijd een systeem uitgebouwd, gebaseerd op een grid infrastructuur, dat gebruikt zal worden voor de reconstructie van de gegevens en de productie van de nodige simulaties. Men voorziet dat CMS per jaar 1 miljoen Gb aan gegevens zal opnemen. In dit grid systeem voorziet men enkele grote centra voor de eerste (lage niveau) verwerking en daarop aangesloten een grid systeem per deelnemend land, TIER-2 genaamd. In België zal de TIER-2 (T2B) tussen 2006 en 2010 gradueel opgebouwd worden tot een capaciteit van 2,7 Msi2k CPU eenheden (180 PCs met dual CPU, 3,6 GHz, 2Gb ram) en 700 Tb schijfgeheugen. Het systeem zal geografisch verdeeld zijn over de UCL (1/3 van de capaciteit) en de ULB-VU Brussel (2/3 van de capaciteit). Op dit moment wordt de mogelijkheid onderzocht om in Brussel de TIER-2 in het ULB-VU Brussel rekencentrum te plaatsen. De uitrusting van de eerste fase wordt 0,7 Msi2k reken capaciteit en 250 Tb schijfgeheugen voor de T2B, wat overeenkomt met een derde van de uiteindelijke infrastructuur. Gezien de natuur en modulariteit van het grid systeem bestaat de intentie om het ULB-VU Brussel computersysteem uit te breiden met 20% zodat ook IceCube er gebruik van kan maken. De huidige Gentse grid infrastructuur is een onderdeel van de BEGgridstructuur met voldoende reken capaciteit voor huidige analyses op HERMES en IceCube. (C. De Clercq, W. Van Doninck, S. Tavernier, VU Brussel; E. De Wolf, UAntwerpen: D. Ryckbosch, UGent)
- Aan de KULeuven wordt gedacht aan simuleren van detectoren (electronische segmentatie, effectieve efficiëntie, optimaliseren opstellingen) wat noodzakelijk is voor toekomstige meer grootschalige experimenten. (M. Huyse, P. Van Duppen, N. Severijns, G. Neyens)
- De ontwikkelde kwantum-Monte-Carlo-methodes kunnen hun capaciteit ten volle benut zien met een cluster die kan beschikken over minstens een honderd nodes met elk enkele gigabytes geheugen. Dankzij internationale samenwerkingen

slagen wij er momenteel in om aan deze behoeftes te voldoen maar dit zal niet voldoende zijn naar de toekomst toe. (J. Ryckebusch, S. Rombouts, UGent)

\* Meer realistische modellering van golfpatronen in tokomaks (bv. JET): nood aan oplossen van groot set integro-differentiaalvergelijkingen. Verhittingsprobleem modelleren (zie SELFO code van Alfven Lab., Stockholm). Oplossen via technieken van adaptieve gridding of gebruik van wavelet analyses vereisen HPC. (T. Cattaert, UGent en KMS)

c) *Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*

- Een 100 Teraflops systeem zou de noden van de Vlaamse onderzoeksgroepen in deeltjesfysica en astro-particle fysica ruimschoots overtreffen. In de eerste plaats zou een dergelijk systeem niet alleen de minimale capaciteit voor een Belgisch CMS Tier-2 gridcentrum garanderen, maar zou het eventueel zelfs mogelijk worden om een volwaardig CMS Tier-1 centrum in België op te richten. In een dergelijk geval zou het potentieel van de Belgische instituten binnen de CMS collaboratie enorm verhogen. België zou hiermee kunnen instaan voor de verwerking van een zeer significante fractie van de binnenkomende CMS meetgegevens en zou daarnaast een van de belangrijkste centra voor verdere CMS en astro-particle fysica-analyse kunnen worden. (VU Brussel, UAntwerpen en UGent groepen)
- Een machine met capaciteit van 100 Teraflops is vanzelfsprekend bestemd voor gedeeld gebruik. Maar dan nog zou een benutting van 1% voor modellering van veeldeeltjesystemen neerkomen op 1 Teraflop, wat zeker zal toelaten om baanbrekend onderzoekswerk te verrichten in o.a. de studie van asymmetrische kwantumsystemen met toepassing in bvb. neutronsterren, verdeling van quarks in nucleonen, large-scale schillenmodelberekeningen (LSM). Een 100-Teraflops machine zou toelaten om driedimensionale systemen met voldoende deeltjes te simuleren. (UGent groepen).

\* De specifieke bijdrage tot het ICRH-onderzoek (ion-cyclotron resonant heating) betreft het aanwenden van wavelet analyse in de poloïdale richting, wat zal toelaten extra te focussen op de golfkoppelingsregio's waar de zeer korte golflengten voorkomen. Het huidige plan is de berekeningen uit te voeren op de JUMP-supercomputer van het Forschungszentrum Juelich in Duitsland. Het zou erg nuttig zijn als Vlaanderen straks over een eigen HPC-infrastructuur kan beschikken. (UGent en KMS groep)

d) *Gewenste ondersteuning*

- De middleware die nodig is om het GRID systeem (deeltjesfysica) te doen werken wordt ontwikkeld in de internationale EGEE samenwerking, en in het CERN. Vlaanderen neemt daaraan actief deel. Een vereiste hierbij is wel dat een dergelijk HPC-systeem permanent de recentste ontwikkelingen binnen dit LCG model zou volgen. Er is wel een noodzaak om competente mensen aan de onderzoekscentra te kunnen behouden en aanwerven.
- Monte-Carlo-methodes zijn van natuur uit zeer goed paralliseerbaar. De belangrijkste voorwaarde is te kunnen beschikken over gestandaardiseerde libraries voor parallelisatie, zoals bijvoorbeeld MPICH. Een tweede aspect is een goed en eenvoudig procesbeheersysteem (queing), zoals bijvoorbeeld PBS. Performantie en eenvoud zijn hierbij twee belangrijke aspecten. Het laatste vooral omdat onderzoek gebeurt in een universitaire omgeving waarbij steeds weer nieuwe mensen opgeleid moeten worden. Een goede documentatie (bij voorkeur on-line) is dan ook onontbeerlijk om zulk een infrastructuur ten volle te laten renderen.
- Het is nuttig een "backbone" te creëren die gebruikers toelaat zich zo snel mogelijk toe te spitsen op hun eigen probleem zonder al te veel tijd te moeten besteden aan taken die gemeenschappelijk zijn voor ieder supercomputing project. Optimaal hierbij is een groep van software engineers nodig die zich full time kunnen bezig houden met het opzetten en onderhouden van software die voor een breed spectrum van gebruikers nuttig/nodig is.

e) *Suggesties*

- De GRID computer infrastructuur die nu door deeltjesfysici ontwikkeld en geïmplementeerd wordt kan eveneens open staan voor andere gebruikers.
- Gezien de grootschaligheid van een dergelijk HPC-systeem, de vereiste ondersteuning qua personeel voor de operatie en het onderhoud ervan en de veelzijdigheid van de toepassingen, lijkt een centraal beheer door een onafhankelijk orgaan binnen België onontbeerlijk. Een verdeling van de verantwoordelijkheid over bepaalde delen van het systeem, kan alleen maar zorgen voor bijkomende technische problemen bij het verzekeren van de uniformiteit en van een coherente werking ervan.
- Eventueel kan een spreiding van een dergelijk systeem over verschillende fysische locaties overwogen worden. Dit zou b.v. kunnen garanderen dat er bij technische problemen op een bepaalde site misschien toch een gedeelte van het systeem op andere plaatsen operationeel kan blijven.
- Een centraal beheer door een onafhankelijk orgaan zou er bovendien voor zorgen dat het wetenschappelijk personeel dat van dit systeem gebruik wil maken slechts een minimum aan tijd en moeite dient te investeren om zijn toepassingen uit te voeren; dit in tegenstelling tot de huidige situatie waarbij een steeds grotere fractie van de tijd voor wetenschappelijk onderzoek verloren gaat aan het uitbouwen en onderhouden van de nodige computing infrastructuur.
- Met het oog op de optimale rendabiliteit van zulk een investering zijn eenvoudige, gestandaardiseerde en toegankelijke procedures zeer belangrijk. Indien administratieve en bestuurlijke procedures het te omslachtig maken om toegang te bekomen tot zulke infrastructuur, zullen veel onderzoekers aarzelen om er gebruik van te maken. Om 'high performance' te leveren met zulke computerinfrastructuur moet ook de organisatorische omkadering optimaal en efficiënt zijn.

### **Bijlage 3h: Turbulentie en Verbranding**

Prof. Charles Hirsch

#### **Research richtingen**

- Het onderzoek van de groep omvat ontwikkeling van rekenmethodes voor stromingsberekening, turbulentiemodellering en modellering van turbulente verbranding. Op het gebied van turbulentiemodellering gaat de aandacht vooral naar hybride RANS/LES methodes en LES (dynamische en multi-scale methodes). Voor turbulente verbranding wordt gestreefd naar de koppeling van hybride RANS/LES en LES methodes met PDF methodes (Monte Carlo berekeningen). Ook de koppeling van LES met CMC is een onderzoeksdomein. De stap naar het simuleren van spray combustion is gepland voor de nabije toekomst. Daarenboven wordt dit onderzoek uitgebreid naar het domein van brand (simuleren van vlamuitbreiding op vaste oppervlakken door de koppeling van CFD in de gasfase met pyrolysemodellering voor de degradatie van het vaste materiaal). Er wordt computationeel onderzoek verricht op stroming in complexe bewegende geometriën. Het onderzoek betreft hier vooral het manipuleren van een bewegend rekenrooster. De rekenkracht is nodig voor de stromingsberekeningen. Methodes worden ontwikkeld voor koppelen van stromingsberekeningen en structuurberekeningen. (E. Dick, UGent)
- Numerische Stromingsmechanica – Computational Fluid Dynamics. Toepassingen in: Lucht- en Ruimtevaart; Turbomachines en propulsie; Industriële en milieu-stromingen; Biological fluid dynamics; Fundamenteel onderzoek op turbulentie (Large Eddy Simulation, Direct Numerical Simulation; Fundamenteel onderzoek in aërothermodynamica van atmosferische reentrystromingen (ruimtevaart)
- Rooster generatie: Hybride nietgestructureerde roosters, integratie met CAD. (H. Deconinck, VKI)
- De onderzoeksgroep Stromingsmechanica en Thermodynamica, vakgroep Toegepaste Mechanica richt zich wat betreft numeriek onderzoek op Computational Fluid Dynamics met focus op Large Eddy Simulatie (LES) en toepassingen hiervan in turbulente verbranding, aeroakoestiek, biologische stromingen en windenergie. (C. Lacor, VU Brussel)

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen.*

- De hierboven beschreven simulatietechnieken zijn zeer rekenintensief. Niettegenstaande hybride RANS/LES methodes computationeel veel voordeliger zijn dan LES of DNS, vereisen ze nog steeds veel rekenkracht en geheugen gezien het altijd gaat om 3-dimensionele en tijdsafhankelijke problemen. In de onderzoeksgroep wordt sinds enkele jaren intensief gerekend met hybride RANS/LES. Typische rekenroosters hebben tussen de 500 000 en 2 miljoen roosterpunten. De grootste kost komt vooral van het tijdsafhankelijke karakter, en de statistische uitmiddeling van de resultaten (vooral van hogere orde statistieken). Typisch wordt per berekening tussen 8 en 16 Intel Xeon 3GHz cpu's (L2 cache = 1MB of 2MB, FSB=800MHz) in parallel gebruikt. Het nodige geheugen voor een typische berekening per cpu is rond de 200Mb. Rekentijden voor de meest recente case bedroegen ongeveer 10 dagen op 16 processoren. Meestal worden verschillende van die berekeningen tegelertijd uitgevoerd, om de invloed van randvoorwaarden, roosterresolutie, verschillende turbulentiemodellen en variatie van de modelconstanten na te gaan. Voor een testcase worden dan gemiddeld tussen de 24 en 32 processoren tegelijk gebruikt. Voor de turbulente niet-voorgemengde verbranding wordt totnogtoe met voorgeschreven PDF-methoden gewerkt (met a priori tabellering van de chemie), zodat dezelfde grootteordes gelden. Voor het simuleren van brand is voornamelijk de geometrie momenteel een beperkende factor: er zijn heel veel cellen nodig voor een LES-berekening van een brand in een realistisch gebouw. (E. Dick, UGent)

- Bij het onderzoek op stromingen in bewegende complexe geometriën heeft een typisch rekenrooster ongeveer 2 miljoen cellen. De bewegende delen van het rooster leggen momenteel een extra beperking op het maximaal aantal processoren dat gebruikt kan worden, in casu een maximum van 6 tot 8 processoren. Het geheugen nodig per processor voor een stromingsberekening bedraagt rond 500 Mb. Indien ook turbulentie en/of meerdere fasen uitgerekend worden kan dit sterk stijgen tot meer dan één Gb. De rekentijden voor een enkele stromingsberekening op 6 processoren kunnen variëren van een week tot twee maand naargelang het doel van de berekening.(E.Dick, UGent).
  - Op dit ogenblik wordt routinematig gerekend op de parallel computing cluster van het VKI. Deze bestaat uit 85 PC's Pentium4, 3,2GHz, met elk 2 GB geheugen. Daarnaast wordt gebruik gemaakt van de cluster op de KULeuven (500 processoren).
  - Alle in-house ontwikkelde software voor stromingsmechanische problemen wordt nu ontworpen en gebruikt op parallele rekenplatformen:
    - COOLFluid, object georiënteerd multiphysics pakket,
    - THOR compressible flow solver for aëronautische toepassingen
    - MIOMA: LES solver voor Cartesische roosters en immersed boundary technieken
    - SFELES: Spectral Finite Element solver voor LES, tweefase Lagrangian bubble tracking
  - Daarnaast worden op het VKI ook een aantal commerciële (of niet in-house ontwikkelde) parallele solver pakketten gebruikt voor allerlei toepassingen, o.a.
    - Fluent: Onderzoek in Industriële toepassingen, warmteoverdracht, multiphase flow
    - Fine Turbo (NUMECA) voor turbomachines. (H. Deconinck, VKI)
  - LES toepassingen vergen een grote reken capaciteit. Enkele concrete toepassingen: LES simulatie van stroming rond cilinder (aeroakoestiek), LES simulatie van SANDIA D vlam (turbulent niet-voorgemengde verbranding), LES simulatie van stroming in bovenste luchtwegen. De onderzoeksgroep beschikt over een cluster van 16 Itanium processoren (1.5 GHz) elk met 2G RAM geheugen (totaal 32 G RAM) en een cluster van 12 Opteron dual-processoren 880 elk met 8G RAM (totaal 96G RAM) geheugen. Gezien het groot aantal gebruikers binnen de groep (12-tal onderzoekers) werden in de meeste simulaties slechts 4 of 8 processoren gebruikt. Ook wordt gebruikt gemaakt van de Hydra cluster van het VU Brussel/ULB reken centrum die uit gelijkaardige Opteron processoren bestaat. Nadeel hier zijn de soms lange wachttijden. (C. Lacor, VU Brussel)
- b) Plannen voor de komende 5 jaren voor HPC-gebruik*
- In de toekomst zullen berekeningen van het type dat nu uitgevoerd wordt uitgebreid naar complexere stromingen met warmte- en massatransfer, en naar stromingen met verbranding. Vooral voor de stromingen met verbranding is veel rekenkracht vereist om naast de stroming nog relatief dure Monte Carlo berekeningen uit te voeren, vereist voor de gebruikte getransporteerde PDF methodiek. Daarbij komt dat verbranding met eindige kinetiek wordt gemodelleerd, met het gebruik van in situ adaptieve tabellering. Hiervoor zijn grote geheugens nodig (orde 4GByte en meer), alsook een groot aantal processoren om de partikels (gebruikt in de Lagrangiaanse Monte Carlo techniek) te verdelen. Typisch zijn minstens enkele honderden partikels per cel nodig. De kostprijs van de berekeningen van het turbulente stromingsveld blijft ongeveer dezelfde. Een substantiële uitbreiding van de reken capaciteit is onontbeerlijk. Typische berekeningen van collega's op 64 processoren duren ongeveer een maand. Uitgebreide rekenkracht is ook gewenst bij het uitvoeren van zogenaamde numerieke experimenten, waar DNS of zeer fijn geresolveerde LES gebruikt worden om referentiedata te genereren voor modelleringsdoeleinden. (E. Dick, UGent)
  - Verdere ontwikkeling van COOLFluid tot een volwaardig multiphysics pakket voor gekoppelde problemen met toepassing in aero-elasticiteit; aeroakoestiek; chemically reacting flows (aerothermisch combustion); tweefasestromingen;

- elektrochemisch massatransport; vliegtuig aërodynamica: ontwikkeling van hogere orde Residu Distributie schema's op niet gestructureerde roosters; Aerothermodynamica voor ruimtevaart; Studie van Microgasturbines. (H. Deconinck, VKI)
- Er wordt een explosie van de behoeften in rekencapaciteit tijdens de komende twee jaar verwacht. Het is duidelijk dat om competitief te zijn met buitenlandse groepen een zeer krachtig parallel rekenplatform zal nodig zijn, berekeningen op 500 tot 1000 processoren zijn geen uitzondering meer in deze branche. (H. Deconinck, VKI)
  - Verdere LES toepassingen in turbulente verbranding met focus op de zgn. Conditional Moment Closure modellen. LES toepassingen in aeroakoestiek: caviteiten, fans, uitlaatdempers van wagens. LES toepassingen biologische stromingen: stroming in bovenste luchtwegen m.i.b. van aërosol deeltjes. Verder ook simulaties van stromingen waar het niet-deterministische karakter van bepaalde gegevens (bijv. inlaatvoorwaarden, geometrie) of variabelen (viscositeit, constanten in turbulentiemodellen) automatisch in rekening wordt gebracht. Dergelijke simulaties vragen ongeveer 128 dagen continu rekenen op 1 processor. Een cluster van 64 tot 128 processoren zou toelaten dit te beperken tot ongeveer 1 à 2 dagen wat gewenst is bij onderzoek teneinde nieuwe modellen efficiënt te kunnen uittesten. Gezien de verschillende toepassingen binnen de onderzoeksgroep zou een cluster van 256 à 512 processoren, continu ter beschikking van de groep, gewenst zijn. (C. Lacor, VU Brussel)
- c) *Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*
- Op gebied van turbulentiemodellering zou met een uitgebreidere rekenkracht het onderzoek zich meer toespitsen op het genereren van referentiedata door DNS of zeer fijne LES berekeningen. Bovendien kan met uitgebreidere rekenapparatuur onderzoek gedaan worden naar specifieke fysische interacties, om rigoureuzere modellering toe te laten. Dit type onderzoek werd binnen onze onderzoeksgroep tot nu toe steeds vermeden wegens computationeel onhaalbaar. Het kalibreren van parameters aan de hand van benchmark stromingen, zoals courant bij RANS modellering, is bij LES en hybride RANS/LES modellering veel minder bruikbaar, omdat bij iedere kleine aanpassing van een parameter opnieuw de ganse LES berekening moet uitgevoerd worden. Subgridmodellering vergt dus meer fysisch inzicht en een ietwat gesofisticeerdere manier van modellering, en hiervoor zijn zulke numerieke experimenten onmisbaar. Ook subgridmodellering van complexere fluida, zoals interactie van bellen of vaste deeltjes met turbulentie, turbulente verbranding en plasma's, vergen hoog geresolveerde numerieke experimenten ter verduidelijking van de fysische interacties. Tenslotte zou grotere rekenkracht toelaten om de ontwikkelde modellen te gebruiken in toepassingen die veel dichter bij de realiteit van industriële stromingsproblemen staan dan de ietwat academische testgevallen die nu uitgerekend worden.
  - Op gebied van turbulente verbranding is de stap nog niet gezet naar de koppeling van LES met PDF (of FDF), wegens gebrek aan rekenkracht. Het is duidelijk dat de toekomst in het onderzoek zich in dit domein bevindt. Er is in het bijzonder een sterke toename aan interesse in vlammen met swirl, wat steevast gepaard gaat met grootschalige unsteadiness die niet kan gecapteerd worden met de RANS-techniek. Hierin zou baanbrekend onderzoek kunnen verricht worden in het domein van de modellering van interactie tussen turbulentie en verbranding (micro-mengmodellen, met nadruk op een betere modellering van de conditionele scalaire dissipatiesnelheid). Ook zou de stap kunnen gezet worden naar het simuleren van spray combustion. Hiervoor is de koppeling LES-FDF ideaal, maar de huidige rekenkracht laat deze stap niet toe. Ook hier zal baanbrekend werk kunnen verricht worden in het domein van de ontwikkeling van micro-mengmodellen, gebaseerd op een betere modellering van de conditionele dissipatiesnelheid (wat hier nog complexer is door de verdamping van druppels). Ook in het domein van brand is het noodzakelijk om de grootschalige unsteadiness te resolveren, zodat het gebruik van LES zich opdringt. Hier ligt de beperking voornamelijk in de geometrie: er zijn heel veel rekencellen nodig om een betrouwbare LES-berekening te doen van brand in een realistisch gebouw. (E. Dick, UGent)



- Een platform met dergelijke capaciteiten laat toe om belangrijke nieuwe toepassingen aan te pakken: turbulentie aërodynamische stromingen bij hoog Reynoldsgetal via Large Eddy Simulation voor toepassing op high lift configuraties (landing en opstijgen van een vliegtuig); aerothermodynamische studies voor ruimtetuigen bij hun terugkeer in de atmosfeer: thermisch en chemisch niet evenwicht (ESA programma's); simulatie van de stroming in de ICP (Inductive Coupled Plasma) windtunnel PLASMATRON op het VKI, d.i. Navier-Stokes gekoppeld met de Maxwell vergelijkingen.; Simulatie van complexe 3D stromingen in elektrochemische reactoren met behulp van het MITREM model (Multi Ion Transport Electrochemical Model) voor verdunde oplossingen, met inbegrip van gasontwikkeling, warmteoverdracht en massatransport, zowel voor plating (elektrochemische afzetting) als ECM (Electro-Chemical Machining); Simulatie van complexe Space Weather stromingen: invloed van de zonnwind en Coronale Massa ejecties op het magnetisch veld van de aarde. (H. Deconinck, VKI)
- Analyse en studie van realistische en industriële configuraties (dwz complexe geometrieën met de bijhorende roosters van verschillende miljoenen punten en aan realistische Reynolds getallen van orde 10<sup>7</sup>) op basis van LES en eventueel DNS worden mogelijk. Een betere design wordt mogelijk door bijv. de verbrandingsprocessen (bijvoorbeeld verbrandingsmotoren wagens, vliegtuigmotoren) te optimaliseren of de aëroakoestische eigenschappen (bijv. volledig vliegtuig bij landing) te verbeteren op basis van simulaties. Ook kunnen meer complexe modellen ingevoerd en getest worden, bijv. verbranding. (C. Lacor, VUBrussel)

*d) Gewenste ondersteuning*

- De parallellisatie van de gebruikte code gebeurde reeds binnen de vakgroep, of is reeds geïmplementeerd in de gebruikte commerciële code. (E. Dick, UGent)
- Wat VKI gaat het voornamelijk om het ter beschikking stellen van de reken infrastructuur.. Op het gebied van software ontwikkeling en parallellisatie is de nodige expertise voorhanden. (H. Deconinck, VKI)
- Hoewel de parallellisatie van onze software binnen de vakgroep zelf werd ingevoerd is ondersteuning door specialisten wenselijk, zeker indien men naar massieve parallellisatie (1000 proc en meer) gaat. Problemen zijn onder meer een efficiënte communicatie en een goede, en liefst geautomatiseerde load balancing. Dit laatste is zeker noodzakelijk bij simulaties met dynamic adaptation waar het rooster dynamisch aangepast wordt gedurende de simulaties. (C. Lacor, VUBrussel)

### **Bijlage 3i: Milieustudies**

Prof. Yvan Bruynseraede

#### **Belangrijkste Research Richtingen**

Remote sensing: hoge-resolutie afstandswaarneming mbv "airborne" en "unmanned air vehicle" systemen (UAV). Ondersteuning van het rampenbeleid, 3D fotogrammetrisch verwerken en in kaart brengen van geografische data via digitale weg, en HPC-gebruik voor de studie van het milieu en de veiligheid. Geïntegreerde studies van de omgeving door het modelleren van de luchtkwaliteit en de milieufactoren. (F.De Schutter, VITO).

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen*

- Airborne remote sensing (pre-operationele fase). Demonstratie van hoge-resolutie "remote sensing" vanuit UAV-systemen. Samenwerken in het ontwikkelingsproces van een "Sensor Web" systeem. Gebruikte "hardware": cluster bestaande uit 20 dual processoren Intel-XEON machines (3,2 GHz) en een opslagsysteem van 45 TB in RAID5.
- Spaceborne remote sensing (operationele fase). Met de lage-resolutie "remote sensing" is de "disk-access" de voornaamste hinderpaal in het operationele scenario. Gebruikte "hardware": cluster bestaande uit 10 dual processoren Intel-XEON machines (3,2 GHz) en een opslagsysteem van 75 TB HP EVA- en NAS-systemen in RAID-5. Noodzakelijke "hardware": linux cluster bestaande uit 4 CPU's met een capaciteit van 8TB in RAID5. (F.De Schutter, VITO).

##### *b) Plannen voor de komende 5 jaren voor HCP-gebruik*

- Airborne en spaceborne remote sensing: focus op "real-time delivery" van beelden afkomstig van hoge resolutie beeldsystemen. Nodige computer infrastructuur: zie beschrijving in a).
- Automatisering van "image processing workflows"
- Opzetten van de GRID-omgeving. (F.De Schutter, VITO)

##### *c) Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*

- Introductie van onzekerheid/fout voortplanting technologie (bootstrapping, resampling en Monte Carlo technieken)
- Voorspellen van de regionale landbouwverankering en landbouwproductie steunende op een beeld- klassificatie- systeem
- Opstellen van verbeterde omgevingsmodellen voor epidemiologische studies van de gezondheid
- Opstarten van ingewikkelde assimilatie algorithmen in 3D-luchtkwaliteit modellen. (F.De Schutter, VITO).

##### *d) Gewenste ondersteuning*

- Noodzaak om de configuratie en het onderhoud van de GRID-omgeving te ondersteunen. (F. De Schutter, VITO).

### **Bijlage 3j: Fundamentele fysica**

Prof. Luc Vanquickenborne en Prof. Renaat Gijbels

#### **Research richtingen**

- Onderzoek naar (schok)golven, instabiliteiten, en versnelling van plasma's (geïoniseerde gassen) op aarde (thermonucleaire fusieplasma's e.a.), in het zonnestelsel (corona van de zon, zonnewind, magnetosfeer, ionosfeer,...) en ver daarbuiten (accretieschijven, astrofysische jets,...). Er worden wiskundige modellen opgesteld die de dynamica en de interactie van plasma's met magnetische velden beschrijven. Deze modellen worden toegepast en door-gerekend en de resultaten ervan worden teruggekoppeld naar de onderzochte plasmafenomenen om ze verklaren en te voorspellen, en vergeleken met observationele data om de modellen waar nodig te verbeteren. De wiskundige modellen gebruiken zowel de magnetohydro-dynamica (MHD, vloeistofbenadering) als de kinetische theorie, of een combinatie van beide. (S. Poedts, Wiskunde/Centrum Plasma Astrofysica, CPA, KULeuven)
- Materiaalonderzoek en nanowetenschap en -technologie; fundamentele fysica (quantumfysica) (F. Peeters, Condensed Matter Theory, CMT, UAntwerpen). Onderwerpen die bestudeerd worden en het type van computationele problemen: (1) Nanogestructureerde supergeleiders (oplossen van 2 gekoppelde niet-lineaire 3D differentieelvergelijkingen à supergeleidende dichtheid en de vectorpotentiaal); (2) Nanosystemen bestaande uit halfgeleiders: a) quantum dots: elektron correlaties in veel-deeltjes systemen; b) quantum wires: ab initio berekeningen (diagonalizeren van zeer grote matrices; berekeningen van stress en strain); c) correlaties in klassieke systemen, bv colloïden, dusty plasma's (Monte Carlo, Moleculaire Dynamica)
- Mathematische en theoretische fysica, kanstheorie, studie van dynamische systemen, statistische mechanica (C. Maes, Theoretische Fysica, KULeuven)

#### **Niveau verwachte HPC-infrastructuur**

##### *a) Reeds ondernomen HPC-gerichte toepassingen*

- De onderzoeksgroep CPA, KULeuven gebruikt zelf ontwikkelde computer-programma's. COOLFluiD (Computational Object Oriented Library for Fluid Dynamics) wordt ontwikkeld in samenwerking met het von Karman Instituut (St. Genesius-Rode). Het betreft een state-of-the-art object geïoriënteerd framework, geschreven in C++. COOLFluiD omvat zowel eindige-volume, eindige-elementen- als residual distribution (fluctuation splitting) schema's die conservatief, upwind, en shockcapturing zijn en bovendien serieel en parallel (distributed memory machines) kunnen uitgevoerd worden, met expliciete en impliciete tijdsstappen. VAC (Versatile Advection Code) wordt ontwikkeld in samenwerking met het FOM-Instituut voor Plasmafysica 'Rijnhuizen' (NL). Dit software pakket bevat state-of-the-art eindige-volume-schema's en er bestaat zowel een parallelle versie van die op distributed memory machines (zoals de KULeuven-HPC-cluster), als een versie met adaptieve roosterverfijning die momenteel enkel op shared memory machines draait (zoals de 8-way node van de KULeuven-HPC-cluster). Typische runs: VAC: enkele uren tot enkele dagen op 48 CPU's. AMRVAC: 4 dagen op 8-way node. Er wordt ook een 3D VAC-simulatie gepland die 10 dagen zal duren op 480 CPU's.
- De onderzoeksgroep CMT, UAntwerpen heeft een eigen PC-cluster met 21 nodes met dual processoren, en toegang tot de UAntwerpen cluster CalcUA (256 nodes met dual processoren), waarvan zij de grootste gebruiker is (60% in oktober-november 2006); de onderzoeksgroep telt 2 ZAP, 1 visiting professor, 9 postdocs en 14 doctorandi.
- De onderzoeksgroep Theoretische Fysica, KULeuven voert alle numeriek werk uit op lokale PC-clusters.

*b) Plannen voor de komende 5 jaren voor HPC-gebruik*

- Het belang en de noodzaak van de numerieke simulaties zullen in de komende 5 jaren nog drastisch toenemen aangezien de modellen zullen ingeschakeld worden voor de verbetering van ruimteweersvoorspellingen. Het space exploration programma AURORA van ESA staat of valt met dergelijke voorspellingen. Er zullen dus nog heel wat simulaties moeten gebeuren en de event studies zullen nog veel realistischer moeten zijn. De nodige rekenkracht zal dan ook minstens VIJFTIG keer groter worden in de komende 5 jaren. De huidige rekenkracht die ter beschikking staat is 4 Tflops. Binnen 5 jaar wordt verwacht dat 100-200 Tflops zullen nodig zijn (S. Poedts, CPA, KULeuven).
- De onderzoeksgroep CMT (F. Peeters, UAntwerpen) plant in de volgende 5 jaren ab-initio berekeningen aan nanosystemen, simulatie van de groei van materialen, voorspelling van de eigenschappen van materialen. Te veel rekencapaciteit bestaat niet. De onderzoeksonderwerpen worden o.a. bepaald (beperkt) door de beschikbare rekencapaciteit, bijv. de grootte van het systeem dat gemodelleerd kan worden.
- C. Maes, Theoretische Fysica KULeuven plant de numerieke component van zijn onderzoek ten gronde en uitbundig te verstevigen, maar kan er momenteel geen getal op plakken. Computerkracht is voldoende beschikbaar, maar hij is op zoek naar mankracht (expertise).

*c) Baanbrekend onderzoek met HPC-infrastructuur van 100 Teraflops*

- Met een machine van 100 Teraflops zou de groep CPA (S. Poedts, KULeuven) eindelijk kunnen concurreren met groepen in de USA. Hun rekentechnieken zijn beter maar ze worden over-powered door de Amerikanen (die met minder goede numerieke technieken toch betere simulaties kunnen maken doordat ze 10 tot 100 keer meer rekenkracht ter beschikking hebben). CPA zou dan de beste technieken met de hoogste resoluties kunnen combineren en dus de meest geavanceerde numerieke simulaties van coronale massa- ejecties en CME schokgolven ter wereld kunnen uitvoeren. Bovendien zouden dan de wiskundige modellen kunnen worden uitgebreid, en zouden bijvoorbeeld deeltjes-versnelling en multicomponent plasma's in rekening worden gebracht, waardoor de simulaties nog realistischer worden. Dit kan momenteel niet wegens gebrek aan rekenkracht. De geplande simulaties zijn overigens van strategisch belang. Het is duidelijk dat de groep die de beste simulaties kan uitvoeren, de beste ESA en EU contracten zal binnenhalen in de volgende decennia.
- Nano-supergeleiding: berekening van de invloed van de eindige afmeting van de supergeleider op de supergeleidende eigenschappen zou mogelijk zijn. Voorts voorspelling van de kritische parameters van een supergeleider op basis van 'ab initio' berekeningen. 'Ab initio' berekeningen van de elektronische eigenschappen van nano-gestructureerde halfgeleiders. (CMT, F. Peeters, UAntwerpen)
- Mits voldoende competente gebruikers zouden ingewikkelde biofysische processen kunnen gemodelleerd worden, en ter ondersteuning van theoretische ontwikkelingen (Theoretische Fysica, C.Maes, KULeuven).

*d) Gewenste ondersteuning*

- Het CPA (S. Poedts, KULeuven) ontwikkelt al zijn codes zelf, en maakt ze parallel. Op dat vlak is geen ondersteuning nodig. Parallellisatie en 'load balancing' behoren tot hun onderzoeksdomeinen en -expertise.
- CMT (F. Peeters, UAntwerpen) werkt reeds samen met dept.Wiskunde/Informatica van UAntwerpen voor ondersteuning op het gebied van algoritmen en software.
- Ondersteuning op het vlak van algoritmieken en software engineering, meer bepaald voor INTENSIEVE parallellisatie is gewenst om een hoge capaciteit HPC infrastructuur maximaal uit te baten voor de wetenschappelijke doelstellingen (C. Maes, KULeuven)

**Aanbevelingen en verwachtingen**

- Een grote centrale computercluster is zonder meer nodig in Vlaanderen. S. Poedts (KULeuven) zou de bestaande infrastructuur van de KULeuven willen uitbreiden om

de investeringen maximaal te laten renderen. Het is belangrijk één grote cluster te bouwen, en niet 3 of 4 kleinere aan de verschillende universiteiten.

- Ondersteuning is uiterst belangrijk voor een efficiënt gebruik van de HPC infrastructuur. Men mag zich niet blindstaren op het aantal Teraflops. Vlaanderen loopt achter op het gebied van computer modellering. B.v. de inbreng van Vlaamse onderzoekers bij CECAM-workshops is klein. (CECAM is een Europees forum voor de toepassing van numerieke technieken in verschillende gebieden van de wetenschappen. De CECAM workshops behandelen een brede waaier van topics uit chemie, fysica, biologie en materiaalonderzoek).(F.Peeters, UAntwerpen).
- C. Maes, KULeuven, heeft meer behoefte aan mensen die kunnen werken met de infrastructuur. Men moet voorzichtig zijn geen eerder wetenschappelijk triviale dingen te gaan inkleden in computationele hoogstandjes. Natuurlijk is er nood aan een efficiënte infrastructuur die moet groeien naar de behoeftes van de wetenschappelijke gemeenschap. Er is nood aan een cel voor computationele wetenschappen, om alles wijselijk te coördineren.

## **Bijlage 4: De TOP500 HPC lijst van November 2007**

### **Evolutie sinds November 2006**

Hierna een uittreksel uit de inleiding bij de laatste uitgave van de Top500 HPC lijst van November 2007. Het weerspiegelt de snelle evolutie van de rekencapaciteit wereldwijd.

*'The Top 10 shows five new and one substantially upgraded system with five of these changes placing at the top five positions.*

*The No. 1 position was again claimed by the BlueGene/L System, a joint development of IBM and the Department of Energy's (DOE) National Nuclear Security Administration (NNSA) and installed at DOE's Lawrence Livermore National Laboratory in Livermore, Calif. Although BlueGene/L has occupied the No. 1 position since November 2004, the current system has been significantly expanded and now achieves a Linpack benchmark performance of 478.2 TFlop/s ("teraflops" or trillions of calculations per second), compared to 280.6 TFlop/s six months ago before its upgrade.*

*At No. 2 is a brand-new first installation of a newer version of the same type of IBM system. It is a BlueGene/P system installed in Germany at the Forschungszentrum Juelich (FZJ) and it achieved performance of 167.3 TFlop/s.*

*The No. 3 system is not only new, but also the first system for a new supercomputing center, the New Mexico Computing Applications Center (NMCAC) in Rio Rancho, N.M. The system, built by SGI and based on the Altix ICE 8200 model, posted a speed of 126.9 TFlop/s.*

*For the first time ever, India placed a system in the Top 10. The Computational Research Laboratories, a wholly owned subsidiary of Tata Sons Ltd. in Pune, India, installed a Hewlett-Packard Cluster Platform 3000 BL460c system. They integrated this system with their own innovative routing technology and achieved 117.9 TFlop/s performance.*

*The No.5 system is also a new Hewlett-Packard Cluster Platform 3000 BL460c system and installed at a Swedish government agency. It was measured at 102.8 TFlop/s.*

*The last new system in the Top 10, placing No. 9, is a Cray XT4 system installed at the National Energy Research Scientific Computing Center (NERSC) at DOE's Lawrence Berkeley National Laboratory and was ranked based on a Linpack performance of 85.4 TFlop/s.*

### **Here are some highlights from the latest list:**

*The entry level to the list - No. 500 - moved up to 5.9 TFlop/s, running the Linpack benchmark, compared to 4.0 TFlop/s six months ago. The last system on the 30th list would have been listed at position 255 in the 29th TOP500 just six months ago. This turnover rate is above average.*

*Total combined performance of all 500 systems has grown to 6.97 PFlop/s (petaflops, or quadrillions of calculations per second) compared to 4.92 PFlop/s six months ago and 3.54 PFlop/s one year ago.*

*The entry point for the top 100 increased in six months from 9.29 TFlop/s to 12.97 TFlop/s.*

*Multi core processors are the dominant chip architecture. The most impressive growth showed the number of systems using the Intel Clovertown quad core chips which grew in six months from 19 to 102 systems. The majority of remaining systems uses dual core processors.*

### **Other trends of interest:**

*The U.S. is clearly the leading consumer of HPC systems with 284 of the 500 systems. The European share (149 systems - up from 127) is still rising and is again larger than the Asian share (58 - down from 72 systems).*

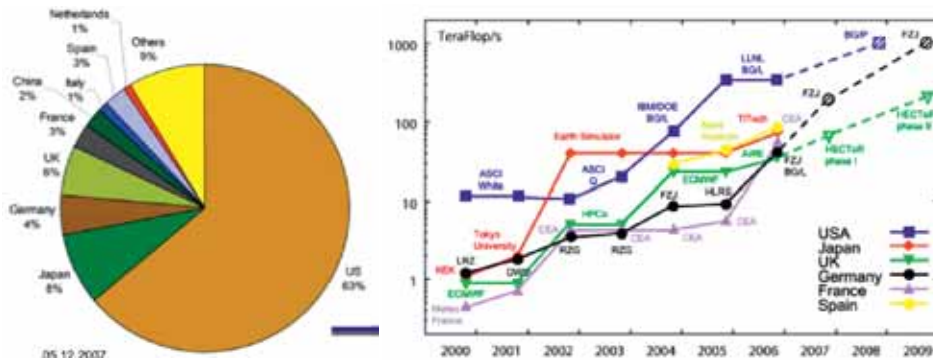
*Dominant countries in Asia are Japan with 20 systems (down from 23), Taiwan with 11 (up from 10), China with 10 systems (down from 13), and India with nine systems (up from eight).*

*In Europe, UK has established itself as the No. 1 with 48 systems (43 six months ago). Germany has to live with the No. 2 spot with 31 systems (24 six months ago)'.  
'*

**De top-10 HPC-installaties – November 2007**

Rank	Site	Processors	Rmax (GFlops)
1	DOE/NNSA/LLNL United States	212992	478200
2	Forschungszentrum Juelich (FZJ) Germany	65536	167300
3	SGI/New Mexico Computing Applications Center (NMCAC) United States	14336	126900
4	Computational Research Laboratories, TATA SONS India	14240	117900
5	Government Agency Sweden	13728	102800
6	NNSA/Sandia National Laboratories United States	26569	102200
7	Oak Ridge National Laboratory United States	23016	101700
8	IBM Thomas J. Watson Research Center United States	40960	91290
9	NERSC/LBNL United States	19320	85368
10	Stony Brook/BNL, New York Center for Computational Sciences United States	36864	82161

De volgende figuren tonen de verdeling van HPC capaciteit in de belangrijkste landen.



### Top HPC-systemen in Europa (November 2007)

Uit de nieuwe lijst van de belangrijkste installaties in Europa wordt, vergeleken met de 2006 rangschikking, de bijzonder snelle evolutie van de rekencapaciteit duidelijk,. Het Spaanse Barcelona Supercomputing Center is van de eerste plaats naar de derde plaats gedegradeerd, en komt wereldwijd niet meer voor binnen de eerste tien centra.

**Het enig Belgisch HPC centrum in de Top500 lijst is nu van Belgacom, en is nummer 315 wereldwijd.**

Rank	Site	Cores	Rmax (GFlop/s)
2	Forschungszentrum Juelich (FZJ) Germany	65536	167300
5	Government Agency Sweden	13728	102800
13	Barcelona Supercomputing Center Spain	10240	63830
15	Leibniz Rechenzentrum Germany	9728	56520
17	University of Edinburgh United Kingdom	11328	54648
19	Commissariat a l'Energie Atomique (CEA) France	9968	52840
23	National Supercomputer Centre (NSC) Sweden	6400	44460
26	Commissariat a l'Energie Atomique (CEA) France	7680	42130
28	Forschungszentrum Juelich (FZJ) Germany	16384	37330
33	Joint Supercomputer Center Russia	3760	33885.1
35	Atomic Weapons Establishment United Kingdom	7812	32498
37	ASTRON/University Groningen Netherlands	12288	27450
40	Max-Planck-Gesellschaft MPI/IPP Germany	8192	21910
48	CINECA Italy	5120	19910
51	University of Reading United Kingdom	3040	19039.5
53	EDF R&D France	8192	18665
54	Ecole Polytechnique Federale de Lausanne Switzerland	8192	18665
60	Cambridge University United Kingdom	2340	18270
67	ECMWF United Kingdom	2480	16202



<b>Rank</b>	<b>Site</b>	<b>Cores</b>	<b>Rmax (GFlop/s)</b>
68	ECMWF United Kingdom	2480	16202
71	CINECA Italy	10240	15760
83	University of Tromso Norway	5632	14471.3
84	Swiss Scientific Computing Center (CSCS) Switzerland	3328	14220
89	CeSViMa - BSC Spain	2408	14000
90	Financial Institution (U) United Kingdom	2768	13904.1
96	Financial Institution (D) Germany	2640	13380.6
97	Financial Institution (D) Germany	2640	13380.6
99	Financial Institution (D) Germany	2544	12980.4
100	Centro de Supercomp. de Galicia Spain	2400	12970
101	HPCx United Kingdom	2560	12940
104	Scientific Supercomputing Center Karlsruhe Germany	1616	12640
112	Technische Universitaet Dresden Germany	2048	12072
113	Financial Institution United Kingdom	1704	11887.2
...	.....	.....	.....
315	Belgacom Belgium	1044	7596.8

## Summary

The Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten (KVAB), Class of Sciences, has taken the initiative to deliver an advice on the role and needs for a dynamic strategy towards High Performance Computing (HPC) in Flanders.

The starting point of this initiative resides in the observation that in the present situation, Flanders, as well as Federal Belgium, is not well prepared to take position in the large scale applications of HPC. This leads to a serious risk of seeing Flanders be distanced on the worldwide scientific and technological scene, and in the way towards scientific and technological innovation.

The KVAB Task group decided to set up a large inquiry among all the scientific establishments in Flanders, to collect the HPC needs and wishes from its scientific community. This information has been evaluated in comparison with the worldwide impact of HPC on scientific innovation and breakthrough research.

In addition, a review of the existing HPC infrastructures worldwide, in Europe and in the surrounding countries has been performed. An essential characteristic of the HPC organization is a pyramidal structure, with a basic level of HPC infrastructures at local institutions (Tier 1), followed by regional HPC centers with a high capacity of the order of 10,000 to 50,000 processors or cores (Tier 2). The third level at the Petaflop scale can only be reached at the European level (Tier 3).

The most representative European HPC centers are connected by a high speed European network, in particular within the projects DEISA and EGEE.

**The creation of a Tier 2 HPC center in Flanders is strongly recommended, based on the collected information.**

This should not be realized at the expense of the existing Tier 1 facilities at the universities en research institutions, which should be further developed to establish a solid basis on which the pyramidal HPC structure relies.

In addition to the capacity to perform large scale HPC projects, it is expected that the HPC center will play an important role in support of the users, for tasks such as parallelization, implementation, software engineering, pre- and postprocessing.

The budgetary implications are estimated around 10M€, with an annual operational budget of 5M€, taking into account a depreciation period of 3 years and a support staff of the order of 15 coworkers.

Additional returns from a Flemish regional HPC center are expected through the stimulation of developments in the area of applied mathematics, algorithms, image processing and visualization, the development of software engineering, the stimulation for high speed networking technologies, new HPC applications towards linguistics, finances, and new opportunities for the pharmaceutical industries in the development of new medications.

**The KVAB is convinced that an initiative towards a regional HPC center in Flanders, would generate an exceptional impulse towards the development of scientific excellence and breakthrough research in all areas.**

