CESGA

díxitos

NOVAS DO CENTRO DE SUPERCOMPUTACIÓN DE GALICIA

XULLO 2008

COMPUTATIONAL SCIENCE

CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

XUNTA DE GALICIA
CONSELLERÍA DE INNOVACIÓN, E INDUSTRIA

FEDER
FONDO EUROPEO DE DESENVOLVEMENTO REXIONAL

dixitos

Fundación Centro Tecnolóxico de Supercomputación de Galicia

DIRECCIÓN
Javier García Tobío

COORDINACIÓN
Fernando Bouzas Sierra

REDACCIÓN
María Piñeiro
Sandra Rodríguez (ipuntocom)

COLABORACIÓN
Andrés Gómez Tato
Carlos Fernández Sánchez
Ignacio López Cabido

PORTADA E MAQUETACIÓN
Yolanda González

FOTOMECÁNICA E IMPRESIÓN
Artes Gráficas Litonor

DEPÓSITO LEGAL
C-1604-1998
ISNN
1139-563X

EDITA
FUNDACIÓN CESGA
Avenida de Vigo s/n (Campus Sur)
15705 Santiago de Compostela
Teléfono 981 569 810
Fax 981 594 616
dixitos@cesga.es

CONTIDOS

De esq. a drta.: Salustiano Mato, Presidente do CESGA, Carlos Martínez, Secretario de Estado de Investigación, Emilio Perez touriño,

CESGA FINISTERRAE C

Os pasados dias 12 e 13 de xuño o CESGA presentou oficialmente o supercom sucederon en diferentes localizacións.

CESGA Finis Terrae Computational Science Conference.

Máis de oitenta investigadores usuarios dos servizos do Centro de Supercomputación de Galicia (CESGA), xunto con outros investigadores e xestores de infraestruturas de toda España e Europa, asistiron á FinisTerrae Computational Science Conference, a xornada intensiva de traballo que tivo lugar o 12 de xuño, con anterioridade aos actos de presentación do novo supercomputador FINIS TERRAE. Técnicos do CESGA, usuarios que participaron nos retos lanzados contra o ordenador no seu período de probas e expertos internacionais, presentaron diversos relatorios en torno á ciencia computacional e ás posibilidades do novo supercomputador.

Durante a sesión, responsables do CESGA explicaron o plan estratéxico do centro para o período 2008-2012, centrado na súa evolución cara ao futuro "Computational Science Research Centre", coñecido tamén como C2SRC.

Os obxectivos do plan estratéxico son continuar mellorando a resolución das necesidades diarias de supercomputación dos usuarios; contar con diferentes sistemas operativos baseados en software aberto que poidan correr tanto cos códigos novos como cos antigos

Presidente da Xunta de Galicia, Fernando Blanco, Conselleiro de Industria, autoridades e invitados durante a visita ás instalacións do Finis Terrae no CESGA.

OMPUTATIONALSCIENCECONFERENCE2008

putador Finis Terrae. Autoridades, xestores, usuarios e público xeral participaron nas xornadas científicas e actos que se

que seguen sendo eficientes; complementar a oferta existente, en lugar de duplicala; facer simulación numérica aplicada á industria; reducir o tempo de espera dos usuarios e optimizar tanto a "capability" como a "capacity" da instalación.

En definitiva, as actuacións previstas para os próximos catro anos diríxense a mellorar a competitividade da comunidade científica; captar investimentos do sector privado; atraer talento do exterior e reter o local; e contribuir á excelencia na ciencia computacional.

O futuro C2SRC

O CESGA afronta este mesmo ano o seu proceso de evolución ao C2SRC, para o que está prevista a incorporación de ata 140 investigadores e tecnólogos ate 2012. Para captar estes expertos cóntase coa colaboración das universidades galegas e do CSIC e aproveitarase o programa IMAN da Xunta de Galicia, dirixido á captación de investigadores de prestixio, ademais de lanzar convocatorias abertas.

A creación do novo centro de excelencia ten a súa orixe na combinación de varios factores: a decisión do Goberno galego de facer fincapé en cinco áreas determinadas de investigación entre as que se atopa a ciencia computacional, a decisión do Ministerio de Ciencia e Innovación de instalar en Galicia unha nova infraestrutura científico-técnica singular e a decisión do Consejo Superior de Investigaciones Científicas (CSIC) de crear unha infraestrutura de cálculo avanzada para a comunidade investigadora aproveitando os quince anos de experiencia do CESGA.

En materia de ciencia computacional, priorizaranse catro áreas de interese estratéxica para Galicia, que presentan unha grande demanda de cálculo e son singulares: Ciencias da Vida, Nanotecnoloxía, Ciencias do Mar e Novas Enerxías, ás que se suman a Supercomputación e a Simulación Numérica como áreas propias.

"CONSTELACIÓN", A NOVA SEDE DO CESGA, SAE A CONCURSO

A data límite para presentar as propostas é o 26 de setembro de 2008.

O concurso seleccionará a mellor idea para posteriormente elaborar o Proxecto Técnico-Básico e de Execución, e a Dirección de Obra para o novo edificio do CESGA, cun orzamento de 22 millóns de euros, e unha superficie superior a 16.000 m² no Parque Empresarial de Santiago de Compostela "A Sionlla".

Será un complexo arquitectónico, con todas as instalacións e infraestruturas necesarias para as dúas areas de traballo do CESGA: O Centro de Investigación en Ciencia Computacional e o Centro de Servizos, Innovación e Xestión.

No primeiro desenvolverase a actividade científica do persoal investigador, con 70 postos de traballo para investigadores, 40 para investigadores sénior e 12 visitantes. O Centro de Servizos aloxará e xestionará as infraestruturas de supercomputación e comunicacións e outros servizos, con mais de 100 postos de traballo e 800 m² de sala de supercomputadores.

Toda a información referente ao concurso na páxina web do Colexio de Arquitectos de Galicia: www.coag.es

FINISTERRAE's ARCHITECTURE

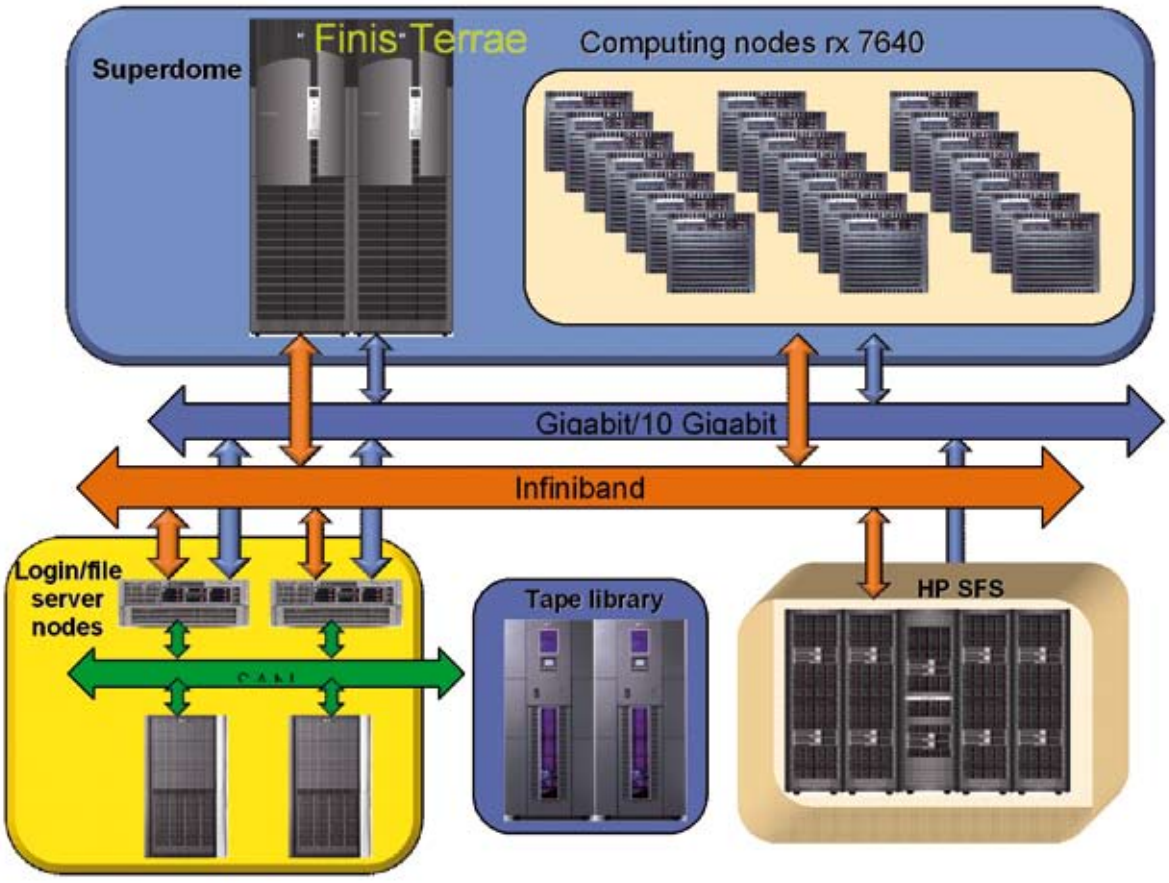
by Ignacio López Cabido, Technical Deputy Director, CESGA

Finis Terrae is a complete solution intended to solve the problems of the community of researchers using CESGA. One fundamental aspect has been taken into account in its design, the coexistence of great computational challenges with other kinds of simulation, perhaps more common, but important as well. On the one hand, the system allows solving huge problems that require thousands of processors and enormous amounts of memory of up to 20 TeraBytes. On the other hand, many times the user's application does not allow a scaling to a significant number of processors, no more than 4 or 8 in many cases, but it is necessary to deal with a large amount of data under the paradigm of shared memory.

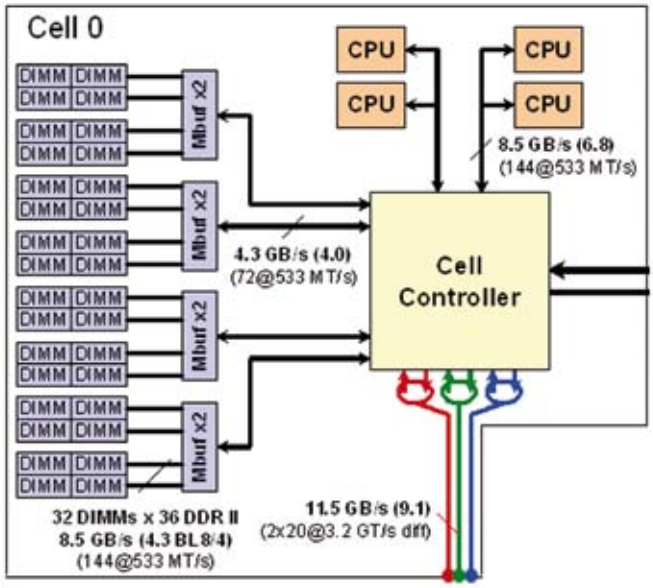
Finis Terrae is made up of the following blocks:
- rx7640 Computing Nodes
- Superdome Computing Nodes
- Infiniband Interconnection Network
- HP-SFS Storage System
- Login nodes/file servers
- Tape Library

The computing nodes are the elements intended to solve the mathematical operations of the simulations and there are two kinds of them in Finis Terrae: Rx7640 and Superdomes.Both kinds of nodes have the same architecture and from the user's point of view, the only difference between them is the size of the node. The HP Rx7640 are servers with 16 processor cores and 128 GB memory, while the main HP Superdome node has 128 processor cores and 1TB memory.The system also includes other two 64-processor Superdomes, which provide an additional 384 GB of main memory.

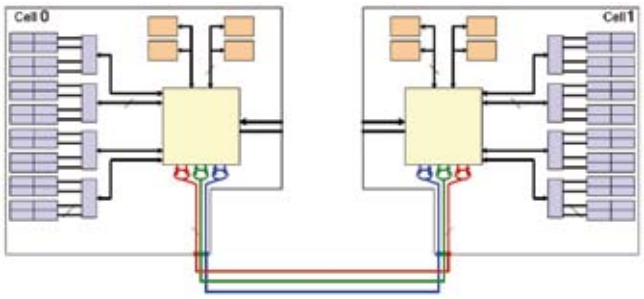
Internally, each of the nodes is constituted by the union of cells as shown on the image. In the case of the Rx7640 nodes there are two cells, while the Superdome has 16. Each cell has 4 last-generation 1.6 GHz Itanium Montvale processors, each of which incorporates two processor cores and 18 MB cache. The cell also provides a 64 GB memory block. One-cell processors access the local memory of their cell through the 'cell controller' and there exists an interconnection of the different cell controllers that allows each processor to access directly the remote cell memory as well, practically at the same speed, but with a slightly greater latency. The architecture of the system is thus ccNUMA: within each computing node, each processor accesses



directly all the memory, although the access time is slightly different depending on whether we are inside the local cell or we are accessing a remote one. The operating system is in charge of arranging each processor's affinity to its cell's memory, trying to put the data accessed by each processor in their local cell.

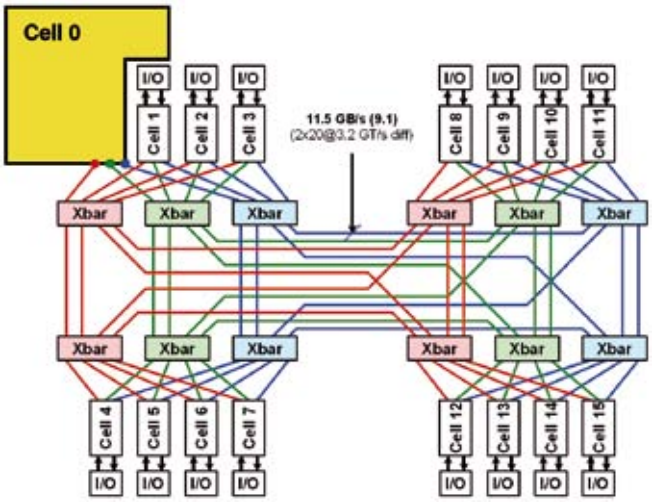


The cells are linked by three different pathways. In the case of the Superdome, the 16 cells are interconnected by means of a network of Crossbar type switches in a two-level architecture.



The different computing nodes can work with independent users' problems, that is, each user work is executed within one single node, but of course for work requiring the utilisation of more processors or more memory than that available in one node, therefore the system has an Infiniband communications network. This network has as its most important characteristics the 16 Gbps high interconnection band width, as well as a very low latency, that for the user is 6 microseconds. Each of the computing and storage nodes is connected directly to a large central switch, so every pair of system elements can always communicate at the same speed.

Another important element in the Finis Terrae solution is the HP-SFS storage system. It is a parallel storage system, made up of a total amount of 20 servers that have 72 cabins connected to 864 SATA type disks. The system acts as a great 216 Terabytes hard disk, shared by all the computing nodes, in a way that the information of this system can be accessed from any of them and at a high speed. Parallel supercomputing systems need to treat and store great amounts of information, and it is only possible to do this at a high speed with parallel storage systems such as this one, based on Lustre free/libre software.



The computing nodes are reserved for the users' problems execution, so the system has some login nodes, which are the servers to which users connect and from which they can send and check the state of their works, etc. There exists a redunded configuration that also provides access to the rest of CESGA's storage cabins, where users' permanent data are normally located, available in the home directory.

The system has, at the same time, a tape library that provides a storage capacity of 2,2 Petabytes of information, in its 1424 slots of LTO4 type tapes. The library has 12 reading units that allow a high transference speed, of almost 3 Gigabytes/seg.

Finally, we have to say that all the system components work using Linux operating system, in its SuSe version.

FINISTERRAE USER'S GUIDE

by Carlos Fernández Sánchez, Systems Administrator, CESGA.

The new user's guide of the supercomputer Finis Terrae is at the users' disposal on CESGA's web page (www.cesga.es, under the Computing section). In this article we will try to present some basic concepts for the optimum use of Finis Terrae as an introduction for users.

Users with an account at CESGA have access to the new server Finis Terrae at the moment. Given the complementarity between this system and the SVG cluster, users can choose which one to utilise depending on the computational requirements, as image 2 shows.

When users connect to the system, they access the login servers, from which they can send jobs to the queue system, check the state of their jobs and do file moving operations. In order to be able to compile applications and run tests with few computational demands, it is necessary to use the command compute in order to access the computing nodes in interactive mode.

In order to store and process information, there are different storage spaces (filesystems) available for users, not only to store data, but also to use as temporal directory during the execution of jobs (scratch). The home directory should be used only to store critical information, for its storage capacity per user is very limited, while the HP-SFS system allows storing a great amount of information (Terabytes per user) and it is accessible with a large bandwidth from all the nodes. Therefore, it is the storage system that should be used for parallel computing simulations.

The way to send and check the state of the jobs in Finis Terrae uses the same commands and parameters as in the other computing servers available at CESGA, except for the new way of sending parallel jobs, depending on the parallelization scheme they use (MPI, OpenMP or hybrid). Specifically, to send purely MPI jobs, the option –pe mpi number_of_processors must be used, and for OpenMP parallel jobs the option num_proc=number_processors should be used. If the parallelization scheme is hybrid (OpenMP and MPI), both options should be combined and the total number of processors dedicated to the job will be the result of multiplying the values introduced in the fields num_proc and –pe mpi. In image 1 and in the user's guide more information is provided about how the queue system places the parallel processes depending on the combination of options used. In the processes layout default scheme (mpi option), the system tries to place the greatest number of processes in the same node. If the option mpi_rr is used, the processes are distributed in all the nodes available, placing the smallest number of processes per node.

The queue system automatically assigns the priority of the jobs, depending on the previous utilization of the system by the user and of the resources required. The execution of parallel jobs with a great number of processors and high memory demands are also priorized, given that for those simulations with low parallelization levels that do not require more than 4GB of memory, the SVG server can be used.

Regarding the maximum resource limits that can be applied for through the queue system, it is important to highlight that they can be surpassed through the special resource application form, available on the web (in the Special Resources section), with which the prioritization of those jobs that require an immediate ending period for special circumstances is also possible.

Regarding future improvements, we could emphasize the upgrade of the HP-SFS software, planned for July 15, as well as verification tests of the nodes where the jobs are going to be executed, in order to guarantee that they do not show any problem and that the performance obtained will be correct. Another planned improvement consists in completely isolating the jobs of the different users by means of processing environments closed in a way that we could guarantee that there are no interferences among them and that the resources

FINIS TERRAE	SVGD
Capability computing	Throughput and Capacity computing
Parallel jobs (>4 ... 2528 cores)	Sequential jobs (up to 8 cores)
Huge memory (>4... 19TB)	Low-medium memory (<8GB)
Huge parallel scratch (>50... 390,000GB)	Low-single node scratch (<100GB)

Shared storage: /COMPARTIDO /sfs
Linux O. S.
Grid Engine Batch Scheduler

Fig 1: Complementarity of Systems SVG and Finis Terrae

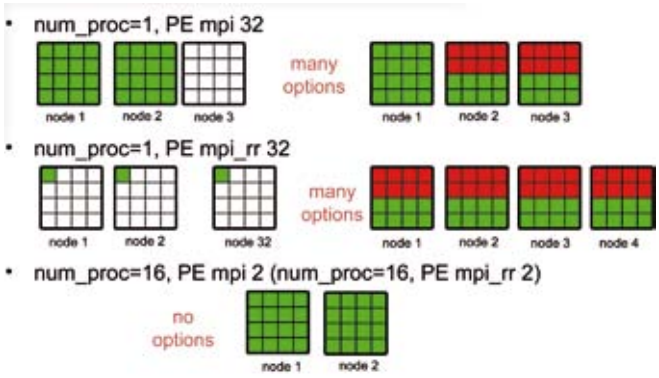


Fig 2: Distribution of MPI processes on the nodes of Finis Terrae depending on the queue system options

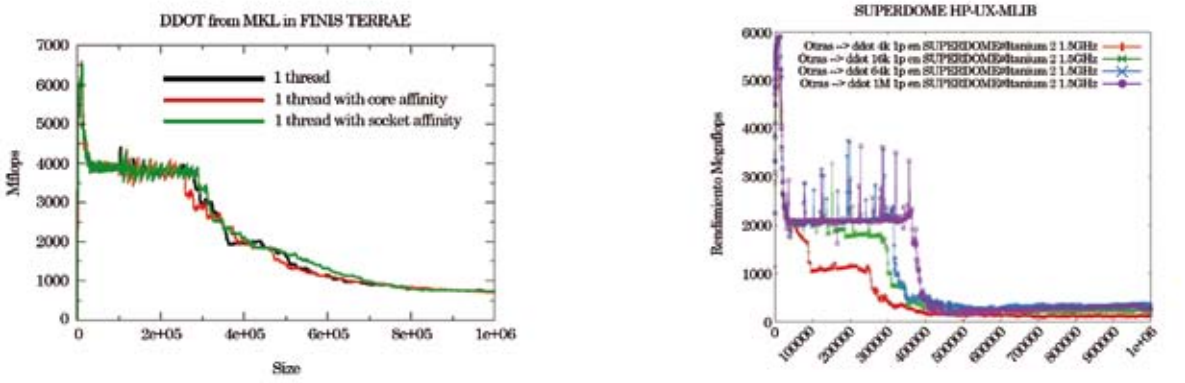
FINISTERRAE TOOLS AND APPLICATIONS

by Andrés Gómez Tato, Applications and Projects Dept. Manager, CESGA

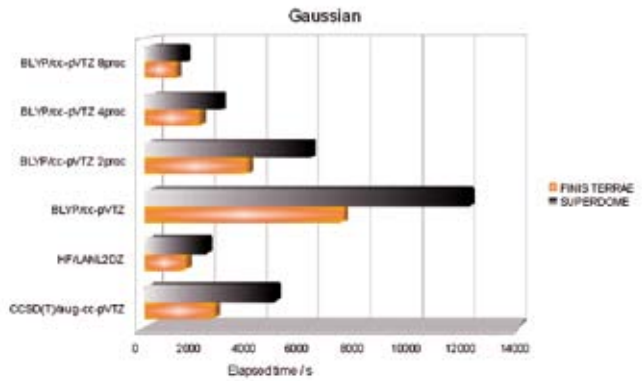
The installation of Finis Terrae meant a change in the way of executing applications at CESGA. When Finis Terrae came into operation, the modules system was incorporated for the configuration of the user's scope according to the application or applications that are to be used. This system automatically configures the field variables needed for a proper execution of the application, facilitating the management of the great number of applications installed. Other important changes were the choice of the ELMER free/libre software as the partial substitute of ANSYS multiphysics simulation software and the change to Intel´s compilers and development and analysis tools. Among these tools, we can find CMKL Math libraries that include BLAS, LAPACK or ScaLapack.

The performance of these libraries as well as that of general use applications was analysed during the tests run previous to bringing FT into operation. In the CMKL libraries tests, the substantial improvement of the output obtained was confirmed (see image 1), practically doubling the number of operations per second in relation to what was obtained in the same test with the HP Superdome, despite the fact that the clock frequency was increased only by 0,1 GHz. This improvement in the output is also noticeable in the applications that were tested, such as Gaussian 03, where the execution times for the tested methods are substantially lower to those obtained with the HP Superdome (see image 2). This improvement is also substantial when models that need a great amount of disk are executed thanks to the SFS storage system (see image 3).

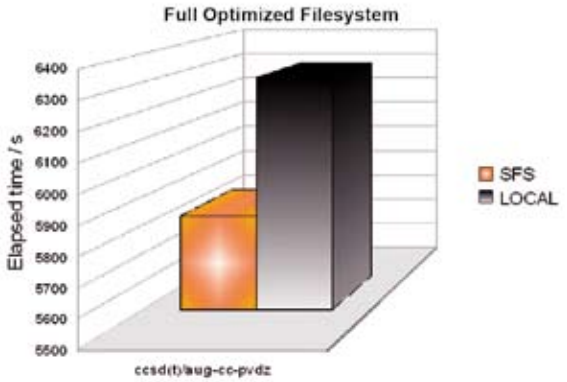
The Applications and Projects Dept. run these tests within the scope of their work to support CESGA´s users. This task includes, among other services, the installation, adaptation and compilation of scientific applications; the help in the parallelization of users' programmes, in the selection of scientific libraries or in the utilization of development tools; and the development of small scientific computing or data analysis applications.



Img 1: Comparison between the output of the Finis Terrae and the HP Superdome for the DDOT function. The significant increase in the output between both systems can be noticed.



Img 2: Comparison between the execution times for different Gaussian 03 calculus methods in HP Superdome and Finis Terrae. It can be clearly noticed that the Finis Terrae notably improves the execution times in all cases.



Img 3: Comparison of the execution times when the SFS or the local storage are used.

Esq: Leif Laaksonen, Presidente do e-IRG.
Drta: Christian Saguez, Presidente de Ter@tec

PRACE e TER@TEC: o futuro está en compartir recursos

A Computational Science Conference 2008 contou coa presenza de representantes dos proxectos PRACE e Ter@tec, dúas iniciativas europeas baseadas na colaboración e o compartimento de recursos de HPC para mellorar as capacidades e a competitividade da comunidade investigadora e a industria.

O Partnership for Advanced Computing in Europe (PRACE) foi presentado polo Director de Colaboración do CSC Finlandés (Finnish IT Centre por Science), Leif Laaksonen, que explicou que o proxecto ten como obxectivo crear un servizo paneuropeo de High Performance Computing ao servizo da comunidade científica da UE.

Segundo Laaksonen, a supercomputación implicou un cambio importante na ciencia e a enxeñaría ao longo da última década e os retos médicos, o cambio climático e outros desafíos de futuro precisan de importantes recursos de simulación. Dado que Europa está por detrás dos Estados Unidos en capacidade de cálculo e que a folla de ruta da European Strategy Forum on Research infraestrutures (ESPRI) xa recoñecía en 2006 a necesidade de mellorar a potencialidade da UE en HPC, a iniciativa para poñer en marcha o PRACE foi recoñecida como un proxecto estratéxico do 7º Programa Marco na súa fase preparatoria, que durará ata 2010, e que comezou cun "call for projects".

PRACE parte da base de que a estreita cooperación entre os centros de supercomputación locais, rexionais e estatais facilitará o acceso a recursos computacionais de grande envergadura para científicos e enxeñeiros das universidades e da industria. Por agora adheríronse a este consorcio de HPC 16 países europeos, entre os que se atopa España.

Laaksonen expuxo que a clave do éxito do proxecto son os desenvolvementos técnicos necesarios para permitir a funcionalidade dunha infraestrutura de supercomputación distribuída, a escalabilidade e optimización do software de aplicacións e a avaliación de prototipos dos futuros computadores.

O obxectivo de PRACE para o ano 2010 é instalar un sistema petaflop/s, é dicir, con capacidade para realizar mil billóns de operacións por segundo. A nivel xurídico, os seus promotores propuxéronse constituír a infraestrutura de investigación baixo unha única entidade xurídica, segundo explicou o representante do consorcio. "Estamos a falar dunha comunidade de colaboración virtual, que é ao que todo indica que nos diriximos no futuro en todos os ámbitos".

Cooperación ciencia-industria

O presidente de Ter@tec, Christian Saguez, analizou as liñas xerais do que os seus promotores denominaron 'Ecosistema Europeo da HPC'', explicando que Ter@tec foi creado para promover o uso da simulación numérica e do cálculo de alto rendemento a través da cooperación científica entre investigadores e industria.

Saguez coincidiu en que a simulación numérica é estratéxica para os retos científicos, pero tamén para a creación e o desenvolvemento de novos produtos e servizos por parte da industria. Neste contexto naceu Ter@tec en 2005, coa intención de contribuír ao desenvolvemento do coñecemento, "pero tamén ao desenvolvemento económico", puntualizou o seu presidente para subliñar o compoñente empresarial do cluster, no que se inxectaron 50 millóns de euros nos últimos anos.

Actualmente, Ter@tec é unha organización sen ánimo de lucro con 53 membros, entre empresas de informática, industrias usuarias e universidades e centros de investigación. A súa capacidade actual é de 50 teraflops e está previsto aumentala a 300 en 2009, co obxectivo de 1 petaflop/s para 2010. Ademais, está asociado ás principais iniciativas internacionais para a creación de consorcios de supercomputación, como PRACE e outras de Xapón e os Estados Unidos.

No futuro de Ter@atec, segundo expuxo Saguez, aparece un grande complexo dedicado á supercomputación de 15.000 metros cadrados preparado para acoller un millar de investigadores e enxeñeiros en 2010. No centro situarase un Very Large Computer Center de 2.600 metros cadrados, arredor do que está previsto configurar un auténtico campus cunha incubadora para empresas start-up, un instituto de formación e dous laboratorios dedicados a Arquitectura e Sistema e ao Deseño de Sistemas Complexos.

Mais información en
www.prace-project.eu e en **www.teractec.eu**

Necesidades Computacionais para a Ciencia de Materiais

A simulación en materiais aumenta as demandas de supercomputación

As necesidades de supercomputación na área da Ciencia de Materiais estanse incrementando de forma moi notoria, sobre todo no ámbito da simulación de tipo atomístico. As principais demandas dos investigadores deste campo de coñecemento foron expostas na Finisterrae Computational Science Conference por Pablo Ordejón, profesor de investigación e group leader de Teoría e Simulación do Centro de Nanociencia e Nanotecnoloxía (CIN2) do CSIC, que explicou que estas necesidades son importantes desde o punto de vista do equipamento de HPC, pero tamén do soporte de programación, cada vez máis relevante para sacar partido aos supercomputadores masivamente paralelos.

Os elementos básicos da simulación no campo dos novos materiais son, desde o punto de vista deste investigador e desenvolvedor, un modelo de interacción, un algoritmo de simulación e ferramentas de análises.

Ordejón expuxo que na súa área é importante contar con infraestruturas de cálculo cada vez máis rápidas e potentes, pero puxo ao mesmo nivel contar con soporte profesional para que o investigador poida centrarse nos aspectos científicos do proxecto en lugar de nos computacionais.

Cada vez máis presentes

O group leader do CIN2 xustificou a súa afirmación de que os novos materiais son un tema de investigación moi importante sinalando que a súa presenza en todas as facetas das nosas vidas vai en aumento. "Están desde nos nosos computadores portátiles ata nos protectores solares", asegurou. Neste sentido, indicou que o 26% das horas de cálculo da Rede Española de Supercomputación (RES) consumidas polos investigadores das áreas de Química e de Física e Enxeñería están relacionadas con materiais, o que supón 15,6 millóns de horas ao ano.

Á súa vez, destacou o imprescindible papel que a simulación cumpre no desenvolvemento da Ciencia de Materiais indicando que a experimentación non permite chegar a todos os aspectos básicos. Segundo Ordejón, a simulación é un virtual lab que chega onde a experimentación non pode, permite recrear condicións impensables a escala real e, o máis importante, fai posible crear materiais con propiedades predefinidas mediante a predición e o deseño. En todo caso, quixo deixar claro que a simulación non substitúe en ningún caso á experimentación, senón que é un complemento.

O representante do CIN2 considera que os principais retos da simulación orientada ao campo dos materiais son a integración das escalas de tempo e lonxitude e as simulacións mecánico-cuánticas.

A nivel de necesidades, destacou os avances metodolóxicos, interfaces gráficas para input/output, ferramentas de visualización para análises, infraestruturas GRID, a mellora da eficiencia dos códigos de paralelización para o hardware existente e, para rematar, ferramentas de programación e paralelización que resulten eficientes nas novas máquinas e accesibles para investigadores non especializados en computación.

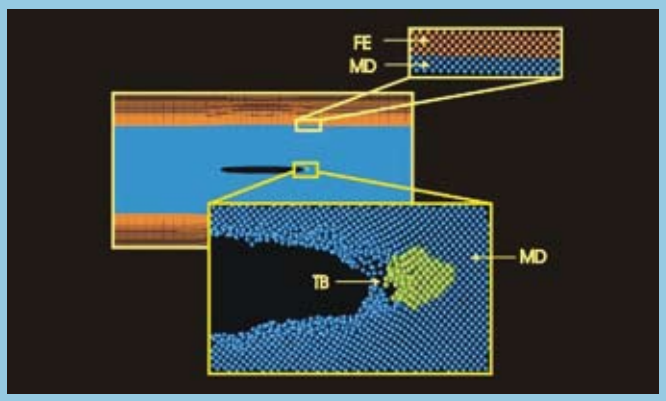


Fig. 1: MÉTODOS MULTIESCALA. Propagación de gretas en Silicio, (Prof. E. Kaxiras, Harvard).

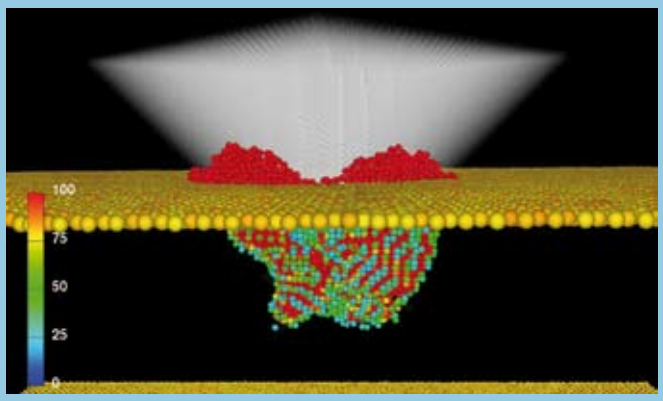


Fig. 2: POTENCIAIS INTERATOMICOS. Nanoindentation. Priya Vashishta et ao. (U. Southern California)

Rede Galega de HPC

Ramón Doallo, Coordinador da Rede Galega de Computación de Altas Prestacións (Rede GHPC) presentou aos participantes na conferencia os obxectivos e proxectos en marcha desta rede temática fundada en 2007 a través dun proxecto financiado pola Xunta de Galicia. Constituída na actualidade por once grupos de investigación do Sistema Universitario Galego, co CESGA como provedor de infraestruturas HPC e de soporte técnico, nun futuro pretende ampliarse cara a outros grupos de investigación, centros tecnolóxicos e empresas de sectores estratéxicos de Galicia que sexan potenciais usuarios de HPC.

O obxectivo fundamental é establecer colaboracións interdisciplinares en ciencia computacional con requirimentos de supercomputación e difundir as tecnoloxías para o aproveitamento eficiente dos recursos de HPC dispoñibles polos grupos de investigación que integran a rede. A finalidade última destas colaboracións é acadar produción científica conxunta, e o establecemento dunha base sólida de colaboración para concorrer a convocatorias de proxectos de investigación ou redes de excelencia, principalmente da UE.

A formación de persoal investigador predoutoral e/ou postdoutoral para especializalo na aplicación de tecnoloxías HPC ao seu campo de investigación é outra das prioridades establecidas, así como o fortalecemento da base de usuarios de diferentes áreas científicas con requirimentos de tecnoloxías HPC e recursos proporcionados polo CESGA (e en particular o FinisTerrae).

Ver: “Díxitos”, Febreiro 2008, páxina 10.
http://ghpc.cesga.es

Rede de e-Ciencia de España

O Coordinador Científico da Rede de e-Ciencia de España, Vicente Hernández, presentou esta rede como unha iniciativa para fomentar a e-Ciencia desde a perspectiva do Grid e da computación. O primeiro paso que se deu foi a elaboración dun libro branco para detectar necesidades e intereses neste ámbito en España. O resultado foi moi amplo, xa que o estudo englobou desde a Astronomía, a Biomedicina, a Enxeñería de Materiais e a Ciencia da Terra ata a Física, a Química Computacional e un longo etcétera.

A creación dun Programa Nacional de e-Ciencia vén motivada pola necesidade de coordinación global das actividades que xa estaban levando a cabo os investigadores españois en colaboración con proxectos internacionais a través de RedIris e GEANT. Tívose en consideración así mesmo a demanda de desenvolvemento de ferramentas comúns e a necesidade de fácil acceso aos recursos de investigación.

A iniciativa foi aprobada en decembro de 2007 e na actualidade a rede integra máis de 700 investigadores de 68 grupos correspondentes a máis de 40 institucións.

Os obxectivos da rede son a promoción e o desenvolvemento da e-Ciencia en España e a coordinación das infraestruturas españolas desde o punto de vista dos investimentos necesarios, a xestión e o soporte a usuarios. Como fins complementarios Hernández sinalou a intención de que a rede convértase no interlocutor español da e-Ciencia na Rede europea, promover a colaboración con Portugal e Latino América a través do IberGrid, e con outros países e transferir os resultados da Rede española.

Catro áreas

A Rede consta de catro áreas operativas: aplicacións, infraestruturas Grid, infraestruturas de supercomputación e middleware. Ao redor destas áreas pivota a súa estrutura, cun grupo de traballo por cada unha delas.

Polo que respecta á área de infraestruturas Grid, Hernández explicou que se está traballando en establecer unha iniciativa estatal no contexto do Grid europeo coa contribución dos grupos participantes dispostos a compartir os seus recursos de computación. Ata o momento, asegurou, xa mostraron o seu interese en participar 18 centros.

Na área de infraestruturas de supercomputación estase tratando de lograr unha maior cooperación entre a Rede Española de Supercomputación e os centros autonómicos, co obxectivo de que está colaboración se traslade tamén á área Grid.

O grupo de traballo da área de middleware está centrado en definir proxectos estratéxicos para desenvolver novas aplicacións e infraestruturas.

Finalmente, na área de aplicacións a prioridade é consolidar as organizacións virtuais que existen e achegarlles novos socios.

Catro éxitos para catro grandes retos

A mellor demostración de que o Finis Terrae é unha infraestrutura de cálculo a grande escala preparada para facer fronte a importantes requirimentos quedou patente nos seus primeiros meses de operatividade no CESGA, cando durante a súa posta a punto conseguiu dar soporte á resolución de catro grandes retos científicos.

A relevancia destes fitos e a súa representatividade acerca da capacidade do Finis Terrae xustificaron a súa importante presenza na Computational Science Conference 2008. A segunda parte da xornada correu a cargo dos responsables destes catro proxectos, aos que precedeu Aurelio Rodríguez, técnico senior de Aplicacións do CESGA, encargado de presentalos explicando por que foron seleccionados e en qué aspectos representaron un reto para o novo supercomputador.

Os criterios de selección que se seguiron para elixir estas catro iniciativas foron que procedían de campos científicos distintos e de grupos de institucións diferentes e que implicaban unha elevada demanda computacional. Como conclusión, Rodríguez puxo de manifesto que neste momento están a correr novos retos no Finis Terrae. Sobre os xa realizados, explicou que demostraron que a máquina é usable dende o primeiro momento e que apenas se requiriron adaptacións de código nin de reprogramación. Por último, destacou que estes traballos puxeron de manifesto que se logrou reducir o tempo de solución nun amplo espectro de problemas.

RETO 1: "SEPARACIÓN DE FASES"

· Física do Estado Sólido: Deseño de Materiais.
· Departamento de Química Física, Instituto de Investigacións Tecnolóxicas, USC

· Wien2k: cálculos de estrutura electrónica de sólidos utilizando a teoría da densidade (DFT)

· Dous níveis de paralelismo:
  · "Coarse grain": k puntos
  · "Fine grain": Scalapack

Reto interesante porque presentaba dous níveis de paralelismo e requiría bastante memoria de proceso.

RETO 2: "DISTRIBUCIÓN DE PUNTOS NA ESFERA BIDIMENSIONAL"

· I-MATH: Matemática Aplicada (Teoría Potencial e Métodos Numéricos)
· GRUPO VARIDIS: Universidad Politécnica de Cataluña
· Estimación de puntos de Fekete
· Varios paradigmas de paralelización:

· MPI · MPI/OpenMP
· OpenMP · Alto Rendemento

Reto interesante porque permitía practicamente todos os paradigmas de paralelización (MPI, Open MP, MPI/Open MP e High Throughput)

RETO 3: "COMPATIBILIDADE ELECTROMAGNÉTICA"

· Deseño de Estruturas Complexas
· GRUPO HEMCUVE, UVIGO, Universidad de Extremadura
· HEmCUVE++: cálculos electromagnéticos sobre a base de métodos FAST MULTIPOLE
· Varios paradigmas de paralelización:

· MPI · MPI/OpenMP

Reto interesante porque implicaba varios paradigmas de paralelización e unha alta demanda de memoria por proceso que puxo a proba ao Finis Terrae no balance entre o tempo requirido de CPU e a demanda de memoria

RETO 4: "COMPRENDER A MAIORÍA DAS ESTRELAS MASIVAS DO UNIVERSO CON ALGORITMOS XENÉTICOS"

· Astrofísica Moleculare de Infravermello
· Instituto de Estructura de la Materia, CSIC
· Algoritmos Xenéticos: PIKAIA, multimodal, problemas de optimización / FASTWIND
· Esquema MPI "Master" "Slave"

· "Master Task" encargouse das operacións relacionadas con GA
· "Slave Task" para realizar os modelos de cálculo
· 7.000 modelos posibles

Reto interesante porque supuxo un novo esquema de paralelización

Separación de fases: o cuarto problema físico da década

O gasto computacional deste proxecto supuxo o uso 384 cores (24 nodos de 16 cores cada un), cun consumo de memoria de 2-4 GB por estrutura e un espazo en disco de 7-15 GB por estrutura.

Alberto Piñeiro (esq.) e Daniel Baldomir (drta.) do Departamento de Física da USC.

A separación de fases é o cuarto problema físico máis importante da década segundo a American Physical Society (APS). Nos últimos meses, un equipo de investigadores do Departamento de Física da Universidade de Santiago de Compostela (USC) logrou acadar unha conclusión de grande relevancia que achega luz a este problema. Grazas a unha investigación soportada nos cálculos realizados no Finis Terrae, conclúese que a separación de fases magnética por si soa non basta para explicar o problema, senón que para que exista é necesaria unha transición de fases química.

Unha fase (estado da materia) posúe características físicas e químicas relativamente homoxéneas e pode constar dun ou varios compostos. Non obstante, cando as propiedades doutro ou máis compostos difiren en tal grao que deixan de ser compatibles, entón hai separación de fases.

O fenómeno concreto que estes investigadores están a estudar consiste na aparición de distintas fases magnéticas e electrónicas coexistentes nun mesmo material sólido cristalino que se produce por estar este nas proximidades dunha 'transición' magnética. En particular, centran os seus traballos nas manganitas (óxidos magnéticos), co obxectivo de dar cunha explicación ao feito de que nelas a separación de fases se produza en condicións de presión e temperatura constantes.

Aplicacións tecnolóxicas

Segundo expuxo o representante deste grupo, o investigador Alberto Piñeiro, durante a súa intervención na Computational Science Conference, este problema cobrou importancia a partir do descubrimento da magnetorresistencia xigante ou colosal, efecto mecánico-cuántico descuberto por Peter Grünberg e Albert Fert hai vinte anos e que lles valeu o Premio Nobel de Física en 2007. A súa aplicación á informática doméstica en 1997 como base de novos discos duros espertou o interese por resolver esta incógnita, coa vista posta en abrir as portas a unha maior condutividade e, polo tanto, eficiencia, destes dispositivos.

En xeral, as aplicacións tecnolóxicas dos materiais estudados neste reto son o almacenamento de carga eléctrica, o transporte electrónico, os sensores magnéticos e a optimización da tecnoloxía das baterías de dispositivos como os teléfonos móbiles e os ordenadores portátiles.

Tras realizar os primeiros cálculos, o equipo comprobou que a separación de fases por si soa con isomagnetismos non é estable. A continuación, probouse coa inhomoxeneidade química a través de varias configuracións químicas e puídose observar que tampouco era estable. Así pois, concluíuse que un estado intermedio logrado a través dun fenómeno químico consistente en concentracións locais de dopado si explicaba a separación de fases nestes materiais e que isto implica a creación de zonas ricas en ocos e en electróns dentro do cristal.

Alta demanda de computación

O representante do equipo de investigadores que realizou o reto científico co apoio do Finis Terrae explicou que para realizar este traballo necesitarían un cluster típico formado por 20 máquinas durante máis dun ano, pero que co supercomputador do CESGA conseguiron darlle solución en pouco máis dun mes.

Como método computacional, por ser necesaria unha alta precisión, utilizouse a Teoría funcional da densidade, desenvolvida polo Premio Nobel Walter Kohn. Segundo precisou Piñeiro, o programa ao que se recorreu ten máis de 1.200 usuarios en todo o mundo, incluíndo empresas, e é aplicable a multitude de sistemas diferentes, entre os que se atopan os sólidos cristalinos.

José Manuel Gesto Beiroa, do Grupo VARIDIS da UPC, presentando os resultados do reto computacional.

Sétimo problema de Smale

distribución de puntos nunha esfera bidimensional

No ano 2000 o matemático Steve Smale elaborou unha lista dos 18 enigmas máis importantes da súa disciplina que foran quedando sen resolver ao longo da Historia e expúxoos como os grandes retos das Matemáticas cara ao século XXI. O sétimo é o problema dos puntos Fekete aplicado a unha 2-esfera, que fai referencia á distribución de puntos nunha esfera bidimensional.

O reto, que foi executado no Finis Terrae durante a segunda quincena de febreiro de 2008, foi exposto na Computational Science Conference por José Manuel Gesto, que explicou que este traballo contribuíu decisivamente a establecer a plausibilidad dunha solución probabilística positiva ao sétimo problema de Smale.

O cálculo realizado no CESGA grazas ao Finis Terrae abre así a porta á resolución efectiva dun enigma exposto desde fai un século e que terá aplicacións na industria a partir do seu desenvolvemento en Física, Nanotecnoloxía, Bioloxía e Métodos numéricos, entre outras, segundo explicou Gesto.

Segundo Gesto, a cantidade de información recompilada durante as dúas semanas de traballo no CESGA permitiu deseñar modelos teóricos compatibles coas hipóteses do custo polinómico previsto, o que senta as bases para unha resposta ao problema.

O problema dos puntos de Fekete busca determinar a posición dun certo número de puntos sobre un obxecto, de maneira que a enerxía potencial producida pola interacción dos devanditos puntos sexa mínima. Cantas máis partículas ou puntos se inclúan, máis complexo faise o equilibrio.

En termos precisos, o problema consiste en minimizar, baixo restricións xerais, funcionais de enerxía potencial dependentes das distancias relativas entre N puntos ou partículas. En particular, os investigadores do equipo que se desprazou a Galicia centráronse en desenvolver un algoritmo capaz de obter en tempo polinómico en N unha boa estimación do mínimo absoluto asociado á enerxía logarítmica de N puntos na 2-esfera -o tipo de obxecto concreto ao que fai referencia o problema de Smale-, fórmula que puxeron a proba no Finis Terrae.

Pero o seu algoritmo vai máis aló, dando solucións para unha ampla gama de xeometrías e diferentes tipos de interacción entre as partículas. Para demostralo, aplicárono a obxectos como plátanos, mazás ou poliedros. Ademais, puideron comprobar que a súa fórmula posibilita que os tempos de cálculo para obter estas configuracións non sexan moi elevados. Os matemáticos da Universidade Politécnica de Cataluña, consumiron 350.000 horas de cálculo durante a súa estancia no CESGA.

Experimentos a gran escala

"Chegamos ao CESGA con case 9 millóns de datos recompilados nun ano enteiro grazas a un cluster e saímos con 60 millóns ao cabo de quince días". Con estas palabras resumiu José Manuel Gesto o aproveitamento de cálculo que obtiveron do Finis Terrae durante a súa intervención.

O seu traballo no novo supercomputador centrouse, segundo o representante do equipo investigador, en aproveitar o seu potencial para facer experimentos a gran escala para determinar a complexidade computacional do problema de Fekete (concretamente o custo computacional dun mínimo local e a robustez do seu algoritmo) e conseguir a maior cantidade de datos para resolver o 7º problema de Smale.

Electromagnética computacional

O grupo HEMCUVE ++ de electromagnetismo computacional logrou resolver 30 millóns de incógnitas durante a fase de probas do Finis Terrae cunha marca de resposta asociada á utilización dun algoritmo alternativo máis eficiente a nivel de escalabilidade.

Luis Landesa (Universidad de Extremadura) e Fernando Obelleiro (Universidade de Vigo) nas instalacións de Finis Terrae no CESGA

A marca mundial de resolución de incógnitas está situado en 42 millóns e con anterioridade a marca correspondía á mesma cifra alcanzada polo grupo de investigadores que o conseguiron no Cesga, aínda que neste caso acadaronse nunha semana, moito menos tempo que o investido polos seus predecesores.

Fernando Obelleiro, da Universidade de Vigo, durante o seu relatorio na C.S.C. sinalou que o seu obxectivo non era romper a marca mundial, senón comprobar que a súa fórmula permitía mellorar o time-to-solution. Con algunhas melloras realizadas xa desde a culminación do seu reto no Finis Terrae, aseguran que poderían acadar, agora si, unha nova marca con 150 millóns de incógnitas resoltas.

Fast Multipole Method

Este equipo mixto de investigadores Universidade de Vigo-Universidad de Extremadura desenvolveu unha ferramenta de supercomputación electromagnética que permite a resolución de problemas de compatibilidade electromagnética en grandes estruturas dotadas de sistemas de radio, como buques, avións ou vehículos terrestres.

Obelleiro explicou, que de entre todos os métodos de análise electromagnética o dos momentos (MM) é o máis empregado pola comunidade científica en electromagnetismo computacional, xa que posibilita o cálculo rigoroso das correntes eléctricas e/ou magnéticas inducidas sobre as superficies da estrutura.

Tradicionalmente, a excesiva complexidade computacional do método dos momentos impediu a súa aplicación á resolución de problemas de gran tamaño eléctrico. Con todo, o desenvolvemento do algoritmo denominado Fast Multipole Method (FMM) permitiu rebaixar o tempo e, xa que logo, o custo de computación, o que permitiu a aplicación do método dos momentos á resolución de problemas de gran envergadura.

Segundo Obelleiro, nos últimos tempos a comunidade científica vén utilizando preferentemente a variante multinivel do FMM porque presenta vantaxes computacionais fronte á mononivel. Con todo, no seu grupo decidiron centrarse nesta última en vista dos problemas de escalabilidade que tiña a máis xeralizada. “O algoritmo mononivel compensa a súa menor eficiencia computacional cunha mellor escalabilidade, característica que permitiu aproveitar mellor a arquitectura do Finis Terrae. Rompemos o límite do multinivel, establecido en 32 procesadores, chegando a usar 1.024”, explicou o representante do HEMCUVE ++.

A utilización intensiva das capacidades de memoria e de comunicacións do Finis Terrae permitiu ao código desenvolvido polo equipo en base ao algoritmo FMM mononivel acadar os 30 millóns de incógnitas nun tempo moi baixo.

Acento nas Comunicacións:

As características do reto eran, por unha banda, o uso intensivo de recursos (memoria, comunicacións e CPU), e, por outro, a necesidade de centos de GB e de centos de procesadores. Os obxectivos que se propuxeron foron probar o código que desenvolveran, avaliar as capacidades do Finis Terrae, analizar un problema electromagnético formado por millóns de incógnitas e establecer as condicións adecuadas para alcanzar a marca mundial.

Os investigadores do HEMCUVE ++ esforzáronse por sacar o máximo rendemento ao Finis Terrae adaptando o seu método de traballo ás especificacións técnicas do supercomputador. Segundo Obelleiro, “esta é unha máquina estupenda para este tipo de retos. demostrou que ten unha arquitectura perfecta porque a escalabilidade non presenta desfasamentos”. A clave está na interconexión dos nodos, é dicir, nas comunicacións, que permiten que o tempo de solución sexa moi eficiente.

Estrelas masivas: entendelas usando algoritmos xenéticos

Pódense estudar fenómenos tan complexos e afastados como as estrelas aplicando a lóxica da selección natural?

Un grupo do Instituto de Estrutura da Materia do CSIC, en colaboración con varios investigadores tinerfeños, alemáns e holandeses, abordou o estudo das estrelas masivas utilizando a lóxica dos xenes no que foi o cuarto reto científico realizado ata agora con cargo á capacidade de cálculo do Finis Terrae, o único desenvolvido na súa fase de produción efectiva.

O astrofísico Francisco Najarro explicou o proxecto en relación ao que cualificou como "enorme potencial" dun computador como o Finis Terrae para a análise de grandes mostras de estrelas masivas na nosa Galaxia e noutras do universo local. O reto realizado no CESGA está vinculado a un proxecto do Gran Telescopio de Canarias para observar cúmulos deste tipo de estrelas e analizalas mediante algoritmos xenéticos co obxectivo de poder xerar modelos que as caractericen.

Dez veces máis grandes que o Sol

Najarro explicou que as estrelas masivas son as que ao formarse teñen unha masa dez veces maior que a do Sol -e, polo tanto, unha temperatura moito máis elevada- e unha luminosidade 100.000 veces superior á do Sol. Son moi escasas e viven moi pouco tempo (tres ou catro millóns de anos fronte aos máis de 1.000 millóns que vivirá o Sol), pero considéranse moi importantes polo seu papel na evolución das galaxias, o nacemento das supernovas, as explosións de raios gamma, a orixe dos buratos negros e a recente re-ionización do Universo.

Ata hai pouco o método máis seguido para o estudo deste tipo de estrelas era a observación directa a través da espectroscopia cuantitativa, que achega datos sobre o estado, a evolución, a masa, etc. coas limitacións asociadas que iso leva, xa que é moi laborioso e non permite tratar con grandes mostras. O grupo coordinado por Najarro abordou a análise deste tipo de estrelas mediante algoritmos xenéticos, técnicas informáticas que imitan o funcionamento da evolución biolóxica para resolver problemas en disciplinas moi variadas.

Evolución das poboacións

Os algoritmos xenéticos naceron nos anos 70 como unha das liñas máis prometedoras da intelixencia artificial. Operan mediante simulación por computador, por iso é polo que se recorreu á capacidade de cálculo do Finis Terrae para lograr resultados moi fiables en canto á xeración de modelos de análise.
Najarro expuxo o esquema polo que se abordou o estudo, combinando aleatoriamente solucións posibles para producir diversidade e seleccionando as solucións máis aptas, que se volven a recombinar e avaliar sucesivamente a modo de xeracións ata que se consegue a mellor solución.

CASE DESCRIPTION

About 10 parameters coded on 2 digits each: gene length ≈ 20
Model running in 30 min, 50 processors: population size ≈ 100

Random search would take <1.E8 generations> for 1 digit accuracy
Need a high selection pressure to be "fast"

Initial population: each gene value present <pop/base=10 times>
Without mutation all genes are the same after ≈ 50 generations

Mutation needs to be high enough to keep gene diversity
Mutation rate needs to be low enough to allow gene convergence
Uniform mutation --> genetic based adaptative mutation

Mutation rate needs to be low enough to maintain evolution
Full generation replacement --> Steady state replace worst

Fgn= Fgn-1 x Psel x (1-Pmut)+Pmut x pop/base

De esq. a drta. : José Manuel Fernández Labastida, Secretario Gral. de Política Científica do Ministerio de Ciencia e Innovación, Fernando Blanco, Conselleiro de Innovación e Industria; Salustiano Mato, Director Xeral de I+D e Presidente do CESGA; Carlos Martínez, Secretario de Estado de Investigación; Santiago Cortés, Presidente de HP España, Rafael Rodrigo, Presidente do CSIC; e Javier G. Tobío, Director Xerente do CESGA, durante a visita institucional ao Finis Terrae.

CESGA presenta

Superada a fase de probas e coas expectativas do seu período de arranque e axustes amplamente cumpridas, o supercomputador Finis Terrae presentouse en sociedade cunha apertada axenda que durante dous días concentrou xornadas de traballo e actos institucionais. Tras o CESGA Finis Terrae Computational Science Conference 2008, celebrada na mañá do 12 de xuño, a tarde dedicouse integramente aos actos institucionais, nos que medios de comunicación, usuarios, investigadores e autoridades tiveron a oportunidade de coñecer as novas instalacións e os proxectos inmediatos.

Unha vez que os usuarios e demais investigadores chegados de toda España terminaron a visita ás instalacións do Finis Terrae coa que se puxo fin á Conference, chegou a quenda das autoridades. A delegación institucional estaba encabezada polo presidente da Xunta de Galicia, Emilio Pérez Touriño; e contou coa presenza do secretario de Estado de Investigación, Carlos Martínez; os conselleiros de Innovación e Industria e de Educación da Xunta de Galicia, Fernando X. Blanco e Laura Sánchez Piñón; o secretario xeral de Política Científica e Tecnoloxía do Ministerio de Ciencia e Innovación, José Manuel Fernández de Labastida; o presidente do Consello Superior de Investigacións Científicas (CSIC), Rafael Rodrigo; o presidente do CESGA, e Director Xeral de I+D+i da Xunta, Salustiano Mato de la Iglesia, o Director Xeral de Planificación e Coordinación do Ministerio de Ciencia e Innovación, Juan José Moreno, o presidente de HP España, Santiago Cortés e o director do Área Sector Público de Intel-España, Norberto Mateos, entre outros.

Tras unha breve explicación sobre as capacidades e a estrutura do Finis Terrae a cargo de Salustiano Mato, o grupo fixo un percorrido pola sala de 300 metros cadrados que ocupan o novo supercomputador e os equipos anteriores que se conservaron. O director do CESGA, Javier García Tobío, foi o encargado de mostrar as instalacións á delegación e responder ás preguntas dos asistentes.

De esq. a drta.: Martínez, Blanco, Rodrigo, Tobío e Mato co Presidente da Xunta de Galicia, Emilio Pérez Touriño, noutro momento da visita institucional ao Finis Terrae.

FINISTERRAE

Os medios de comunicación tiveron ocasión de participar tamén na visita.

Salustiano Mato, Director Xeral de I+D+i da Xunta de Galicia e Presidente do Consello de Administración e da Fundación CESGA:
"O Finis Terrae é unha gran achega ao potencial que España pode demostrar no campo da supercomputación e ese é un xeito fantástico de colocar a Galicia no mapa do Mundo".
"Contar con infraestruturas científico-tecnolóxicas como Finis Terrae é fundamental para obter un crecemento económico estrutural que poña á sociedade a salvo das crises e que só virá da man da investigación e a innovación".
"O CESGA foi unha gran iniciativa desde os seus inicios e agora é o momento de recoñecer o labor de todos os que iniciaron esa aventura. Hoxe, co Finis Terrae, gaña o Sistema de Ciencia e Tecnoloxía, gana Galicia e gañamos todos".

Presentación Oficial do FINISTERRAE

Na tarde do 12 de xuño, tivo lugar no Centro Galego de Arte Contemporánea (CGAC), o acto oficial de presentación do Finis Terrae ao que acudiron máis de 200 invitados. Logo dos discursos oficiais, nos que todos coincidiron en sinalar a nova infraestrutura de cálculo como un fito na potencia de supercomputación española e europea, celebrouse un cóctel na terraza do edificio.

Ademais de parte da delegación da Xunta de Galicia e o Goberno central que previamente visitara o CESGA, estiveron presentes os reitores das tres Universidades galegas -Senén Barro (Santiago de Compostela), José María Barja (A Coruña) e Alberto Gago (Vigo)- e vicerreitores das Universidades de León ou Valencia; a conselleira de Economía e vicepresidenta segunda da Xunta de Estremadura, Francisca Chávez; o director xeral do Centro de Supercomputación de Castela e León, Luís Muñoz; o director do Centro Informático Científico de Andalucía, Juan Antonio Ortega; a directora do Centro de Supercomputación do Parque Científico de Murcia, María Eugenia Requena; a Fundación para o Fomento en Asturias da Investigación Científica Aplicada e a Tecnoloxía, Ángeles Álvarez; o director de Informática do Corpo Nacional de Policía, José Luís Díez Aguado; o director de Informática da Dirección Xeral da Garda Civil, Ricardo Llorente; e outros representantes de Universidades, Centros de Supercomputación e Tecnolóxicos, empresas e Administracións, ademais de numerosos directivos de HP e Intel. As firmas de todos os presentes recolléronse nunha lona xigante na que se solicitou un desexo de futuro para o Finis Terrae.

De esq. a drta.: Norberto Mateos, Director Área Sector Público de Intel; Rafael Rodrigo, Presidente do CSIC; Carlos Martínez, Secretario de Estado do Ministerio de Ciencia e Innovación; Salustiano Mato, Director Xeral de I+D+i da Xunta de Galicia e Presidente do CESGA, e Santiago Cortés, Presidente de HP España.

Carlos Martínez, Secretario de Estado do Ministerio de Ciencia e Innovación:

"Finis Terrae abre unha nova etapa para a ciencia, a comunidade científica, Galicia e España. Trátase dunha infraestrutura única e singular que supón un fito en Europa pola súa capacidade de cálculo e de computación compartida".

"Esta novidade tecnolóxica vai permitir dar solución a problemas de diferentes áreas que necesitaban da capacidade do novo supercomputador para resolvelos e emprender outros novos aínda máis ambiciosos".

"Quero anunciar hoxe aquí a nova conexión entre a Rede Iris española e a rede portuguesa a través de fibra escura no marco de IberGrid, que terá o seu primeiro tramo entre Tui e Santiago a finais de ano e que co tempo estenderase tamén a Francia".

Rafael Rodrigo, Presidente do CSIC:

"Poucos casos hai tan paradigmáticos de contribución á vertebración do Sistema Español de I+D+i como o CESGA, no que a colaboración entre diferentes administracións mantívose no tempo aínda sendo de diferentes cores políticas".

"Finisterrae abre unha nova etapa para os investigadores do CSIC e das Universidades, á vez que supón unha oportunidade para pór a disposición de moitos outros en España recursos únicos en toda Europa".

Santiago Cortés, Presidente HP España:

"Debemos asegurarnos de que as autoestradas globais da innovación pasen polo noso país. Por iso é tan importante a infraestrutura que hoxe presentamos".

"As redes globais da innovación han de aproveitar o potencial de todos os investigadores e profesionais. Por iso HP está promovendo proxectos nos que os expertos da empresa traballan cos expertos da Universidade e de centros como o CESGA, co obxectivo de crear valor engadido".

Norberto Mateos, Director Área Sector Público de Intel:

"Este proxecto é estratéxico para Intel, que demostra unha vez máis o seu compromiso co avance da I+D+i en Galicia coa súa achega a unha infraestrutura nacida da colaboración entre o ámbito público e o privado, condición fundamental para o avance da Sociedade da Información".

"O Finisterrae é un claro exemplo da filosofía de accesibilidade para moitos investigadores a recursos de computación antes restrinxidos a uns poucos".

Javier García Tobío, Director Xerente do CESGA:

"Hoxe o CESGA ten, grazas ao Finisterrae, unha capacidade 10.000 veces superior á do seu primeiro supercomputador, o que nos permitiu que os 100 grupos de investigación que foron os usuarios iniciais convertéronse en 645".

"O CESGA apostou polo desenvolvemento da e-ciencia, por iso a colaboración e a cooperación están enraizadas a nosa filosofía de actuación. O desenvolvemento da ciencia, como motor do benestar social, non debe ter fronteiras".

HP e Intel revelan o seu roadmap

Medio centenar de representantes de empresas, centros tecnolóxicos, universidades e administracións de toda España asistiron o 13 de xuño á xornada de presentación dos plans estratéxicos do CESGA e dos seus dous socios tecnolóxicos no Finisterrae: HP e Intel. Con esta sesión, celebrada no salón de actos do Instituto de Investigaciones Agrobiológicas do CSIC, e unha visita guiada para os asistentes pechouse o programa de traballo organizado ao redor da inauguración oficial da nova infraestrutura de cálculo galega.

O director do CESGA, Javier García Tobío, presentou o plan estratéxico do centro para o período 2008-2012 precisando que o novo supercomputador foi un fito, ?pero nós xa estamos traballando no futuro?, en referencia a que xa se están definindo as liñas mestras do plan estratéxico que se porá en marcha logo de 2012.

Asegurou que a pesar de multiplicar a capacidade de cálculo do centro por 10.000 desde a súa creación ata a actualidade, "sempre imos por detrás do que a comunidade investigadora demanda, de aí o noso compromiso coa mellora continua e sempre pensando en clave de futuro".

O reto do Finisterrae

A posta en marcha do novo supercomputador do CESGA supuxo un reto para HP, segundo explicou o seu director da área HPC, Isidro Cano. Concretamente, implicou deseñar novos produtos en exclusiva; ampliar a corrente eléctrica da súa fábrica en Alemaña; coordinar equipos de enxeñería, supply chain, fábrica, preventa, etc; e "ser os primeiros no mundo en moitas cousas e crer niso".

Sobre o TOP500, (ranking semestral dos supercomputadores máis potentes do mundo) Cano asegurou que HP conta co maior número de rexistros en total e xustificou a súa aposta por esta área afirmando que o mercado de HPC xa supón o 9% do mercado mundial en informática.

Segundo Cano, os retos da supercomputación son: o crecemento, "porque non hai suficientes recursos para atender aos usuarios"; os custos, xa que "as novas tecnoloxías prometen aforros ligados ao retorno do investimento"; e os riscos, "polo que é importante contar con equipos que teñan experiencia".

Hewlett-Packard inviste 3.600 millóns de euros en I+D cada ano. Actualmente están investigando en chips de 256 cores, arquitecturas avanzadas con mellores interconexións, programas de software para optimizar os multinúcleos, dense computing (maior rendemento a menor prezo e en menos espazo), aceleradores matemáticos aplicados ao mercado do cálculo, redución do cableado en instalacións grandes e eficiencia en consumo e refrixeración. Un dos seus principais obxectivos é potenciar os instrumentos que crearon para fomentar a colaboración entre os seus clientes, como a HP-CAST IBERIA, asociación de usuarios de supercomputación de HP de España. Ademais, puxeron en marcha a Fundación I+E para empresas estranxeiras que teñen centros de I+D en España co obxectivo de que cada vez sexan máis. Tamén recordou que teñen aberto o seu programa Innovation Research, a través do que o HP Lab propón temas de investigación que lle interesan, convoca unha selección de ideas e financia as que máis se achegan aos seus obxectivos.

Do Itanium ao Kittson

O director de tecnoloxía de Intel, Antonino Albarrán, quixo destacar o compromiso da súa empresa coa supercomputación e co I+D. Os principais fitos aos que fixo referencia no ámbito da investigación e o desenvolvemento en High Performance Computing (HPC) por parte

De esq. a drta.: Norberto Mateos, Director Área Sector Público de Intel-España; Santiago Cortés, Presidente HP España; Salustiano Mato, Presidente do CESGA e Javier G. Tobío, Director Xerente do CESGA, durante a rolda de prensa para medios internacionais celebrada ao comezo da xornada.

de Intel foron a redución do consumo, o investimento en novas interconexións, a localización de centros e plataformas de referencia en supercomputación en todo o mundo e o aumento do número de expertos neste campo no seu persoal.

Albarrán precisou que o 70% dos sistemas presentes no TOP500 levan procesadores Intel e que este índice vai en aumento. "A supercomputación é un área estratéxica para Intel na que seguiremos investindo".

O representante de Intel recordou que tras os petaflops virán os exaflops, os zetaflops e a continuación os yotaflops, deixando claro que "isto non se para e nós tampouco imos facelo porque hai disciplinas como a Xenómica e a Medicina que o van a necesitar".

Sobre o futuro dos seus procesadores, Albarrán explicou que a xeración actual de sistemas para supercomputadores, os Itanium (implantado no Finisterrae), están sendo mellorados para lograr un maior rendemento, máis robustez e mellor eficiencia enerxética. Estas optimizacións materializaranse a finais deste ano nunha actualización que se chamará Tukwilla, á que en 2011 seguirá Poulson, con máis procesadores por núcleo. Aínda sen data prevista de saída ao mercado as melloras continuarán con Kittson, a novena xeración de procesadores Itanium.

Os investimentos de Intel no campo da HPC estanse centrando en aumentar a capacidade de fabricación; en seguir avanzando na Lei de Moore (cofundador de Intel, que formulou unha lei segundo a cal aproximadamente cada dous anos duplícase o número de transistores nun circuíto integrado); en integrar o prototipo de 1 teraflop nun só chip; en terabytes de ancho de banda; e en comunicacións ópticas dentro do propio procesador.

Salustiano Mato, Presidente do CESGA e Santiago Cortés, Presidente de HP España, durante a xornada.

O CESGA crea grupos de traballo para atender as suxestións dos usuarios

Grupos de traballo de usuarios e membros do equipo técnico de CESGA comezarán a traballar en breve para dar solucións ás propostas de mellora xurdidas da última enquisa de satisfacción elaborada a partir das consultas realizadas aos responsables dos grupos de investigación que efectúan operacións de cálculo e almacenamento no CESGA.

A posta en marcha destes grupos mixtos foi a principal decisión adoptada no marco da xornada de traballo celebrada a finais do mes de xuño para que os coordinadores de área coñecesen e analizasen en equipo os resultados da enquisa e propuxesen melloras operativas baseándose nos problemas identificados. Unha vez presentadas as conclusións do estudo de satisfacción ao equipo técnico do CESGA, analizáronse os problemas identificados e se estableceron prioridades en torno ás cuestións máis significativas.

A celebración da sesión e a decisión de crear estes catro grupos iniciais de traballo inscríbense no marco da política de mellora continua e a vocación de priorizar o servizo ao usuario do CESGA e responden ao compromiso coa calidade derivado da obtención, hai tres anos, da ISO 9001:2000.

A conclusión xeral do estudo de satisfacción dos usuarios, é a percepción moi positiva da calidade e do servizo. A taxa de resposta foi moi elevada, o que se interpretou polos expertos encargados da enquisa, como un indicador do vínculo de confianza que se estableceu entre o centro e os seus usuarios.

A valoración positiva das canles de comunicación revela un peso elevado do factor persoal, principal vía de contacto entre os usuarios e os técnicos do centro. Esta valoración baséase no esforzo de atención personalizada que realiza o centro. Non obstante, os usuarios máis intensivos demandan avanzar en servizos electrónicos de comunicación que superen as demoras de resposta asociadas ao e-mail, como por exemplo sistemas de mensaxería instantánea, posibilidade que será estudada.

Sistema de colas

A percepción xeral dos usuarios sobre a información e a xestión do sistema de colas é de transparencia por parte do centro. As demandas de mellora diríxense a optimizar o sistema de xestión nos puntos nos que se poida evitar a variabilidade no ritmo e a saturación de colas. Os usuarios perciben ademais unha boa capacidade e tempo de resposta ante incidencias, aínda que demandan maior formación e que o centro fomente as colaboracións entre eles.

Diversas áreas temáticas centrarán o traballo dos grupos mixtos técnicos-usuarios, adiantándose xa na sesión durante a que se decidiu crealos, posibles actuacións estratéxicas por parte dos técnicos que maior capacidade de actuación teñen en cada un dos ámbitos de referencia. No futuro, constituiranse os grupos dende unha perspectiva que transcenda os muros do centro para ter a posibilidade de atopar solucións dentro e fóra.

Papers on Computational Science

TER@TEC: an European Eco-System For High Performance Computing

by Christian Saguez, President of TER@TEC

High Performance Computing is one of the most important areas for Science and Industry in terms of capacity of innovation and competitiveness. In order to federate the main industrial and academic actors and to be at the best worldwide level in terms of scientific and technological research and economical development, the European initiative TERATEC was launched in mid 2005. After two years of development, TERATEC is now able to propose all the infrastructures for the future European Petascale Systems and to welcome research laboratories and companies in a new campus located in the south of Paris. Large collaborative research projects, especially with the French Cluster for Competitiveness System@tic, in the area of Architecture, Networking and Application Software have been launched. In this paper, we present the motivation and the objectives of TERATEC.

1 – High Performance Computing: a European priority

Developing capacities of innovation and increasing competitiveness are for Europe two main strategic challenges both for Science and Technology and for Industry and Services. Two characteristics of the new environment of this century are: firstly, that we have to control larger and larger amounts of numerical information; and secondly, that the systems that we have to analyse, design, manage, are more and more complex. In this context, modelling as well as simulation technologies and tools, and high performance computing systems are the key components to address to these challenges.

The following examples illustrate these general considerations:

i) The design and the life cycle management of new products, systems or services

The objective is now to design new planes or new cars, using only virtual prototypes in a collaborative way, as it was been done for the new Falcon 7X plane of Dassault Aviation. The situation is similar for global systems such as transportation systems, retail systems, multimedia networks or financial or insurance products.

ii) The study of interaction problems in Biology and the design of new molecules. The objective is to build C.A.D software for new molecules and medicines by using data mining techniques and dynamic simulation of large amounts of information.

iii) The study of new innovative materials. Materials science is one of the major challenges for research and industry. We have to perform large multi-scale and multi physics simulation from ab initio simulation to continuous mechanics simulation.

iv) The indexation and spreading of large multimedia databases, including texts, images, videos… Domains such as virtual libraries, access to multimedia documents… are challenges for our society and also a crucial instrument for the promotion of our artistic, scientific and technological cultures.

v) Sustainable growth, environmental simulation, global change study in climatology, risk analysis and management... are other important examples for which HPC Simulations are fundamental in order to understand, propose and validate original solutions.

In order to address these problems and many others, the availability of High Performance Computing facilities (Hardware and Software) is fundamental. New innovative capacities, technological breakthroughs, decrease of financial costs and design time, or differentiation from competitors are some of the advantages immediately reachable with HPC. Therefore, HPC is probably the most important element for industry and services competitiveness. It strongly impacts all the challenges for the 21st century society, such as Energy, Environment, Sustainable Growth, Health, Risk analysis, etc.

HPC **must be**, as in USA and in Japan, **one of the priorities of Europe**. If not, our economical and scientific development and our independence will be surely compromised.

2 - TERATEC – Objectives and present situation

Taking into account the previous analysis, the TERATEC association, following an idea of CEA, was founded in 2005 as an European cluster dedicated to high-performance computing, in order to:
+ create, in a single location, all necessary infrastructures for operating petascale computers and for welcoming laboratories and industrial companies.

+ give access to the most powerful HPC Systems (Hardwares – Softwares – Networks) to a large community of users, in particular SMEs.

+ federate the major HPC actors, academic institutions and industry companies, technology and service providers, end-users ...

+ initiate and coordinate important collaborative R&D projects in order to gather all the competences to build and manage HPC Systems and related software.

+ promote HPC technologies, especially for SMEs, and to incubate new start-up companies.

+ participate in learning activities for future engineers and long life education courses.

After two years of activities, the first results are extremely positive, prove the efficiency of the model and confirm the needs of the HPC communities:

1 – The available computing power is now more than 100 Teraflops (150 Teraflops in 2008) and we expect at least 2 Petaflops with a 10 Petaflops forecast by the end of 2009.

"HPC is now of a strategic importance, but it should turn to be a high priority in Europe"

2 – The number of members of the association is rapidly growing with participation of providers and end-users, of research institutions and industry and services companies.

3 – We started building all the infrastructures for petascale systems (building of 6500 m2) and new office buildings (~ 15 000 m2) for laboratories and companies. All these infrastructures will be available before the end of 2009.

4 – We launched numerous large collaborative R&D projects, especially in strong connection with the cluster competitiveness System@tic and with important financial support from national and local administrations for a global budget of more than 50M€

POPS

This aim of the POPS project is to develop a new generation of high-power Petaflops-capacity applications, and to validate the associated computing servers.

The POPS project enable companies and research centers to start preparing now for tomorrow's parallel applications, so they can benefit from the power that future supercomputers will offer as soon as they become available. The POPS project therefore offers enterprises particularly small to medium-sized businesses and the research partners involved in the project, the opportunity to look several years ahead.

POPS is including the following features:

- Hardware, Architecture, Integration and Performances

- Software Development and Tools

- High Performance Computing Applications

- Information Modeling – Data management and processing

- Petaflopic Simulation

POPS project is leaded by BULL and gathers 20 partners from industry and research in the frame of competitively cluster System@tic.

EHPOC

This project EHPOC is dedicated to the development of platforms and numerical tools (algorithms…) for global optimization of multiphysics and multiscale complex systems. EHPOC is including the following features:

- industrialization and customization of the numerical tools performance computing applications for example, the pre and post-processing tools

- development of multidisciplinary optimization tools

- introduction of the uncertainties inside the simulations

- optimized links between CAD and CAE

- customization of the numerical tools to the future multi-core and massively parallel architectures

The project is coordinated and managed by CS Communication & Systems (IT services company involved in HPC) in the frame of competitively cluster System@tic.

The project includes real-life demonstrators. These demonstrators will be focused on one hand onto advanced materials applications (durability and aging, multi-materials assemblies, computer-aided design of new nano-structured artificial materials) and on the other hand onto grand challenges high performance computing applications (ultra large-scale CFD, structural mechanics, crash, plasma , etc… applications). The main partners of the project are the members of the Teratec association involved in high performance computing (industrial partners, ISVs and research laboratories)

CARRIOCAS

CARRIOCAS studies and implements an ultra high bit rate optical fiber network able to answerr the scientific and industrial needs in interactive numerical simulations on remote supercomputers and in treatment of very large volumes of distant data. The objective is to elaborate the best technical and economical solutions to adapt ultra high bit rate telecom networks to the requirements of these applications, that is the adaptation of the high bit rate transport physical layers, of the protocols and algorithms in order to dynamically manage the resource allocation and to guarantee the expected quality of service.

Project results are validated on an experimental network with bit rate per wavelength up to 40 Gb/s, interconnecting in its initial phase four major locations in the south west of Paris (including Ter@tec computing center in Bruyères le Châtel). It will enable in particular to evaluate two applications developed by the project: Treatment of massive volumes of data via a distributed file system and remote collaborative high resolution visualisation involving a 20 Mpixel picture wall. It will allow the industrial and academic partners of the project, and later on others interested partners in domains such as energy, avionics, pharmaceutical, finance, etc to confirm the expected benefits of high bit rate connectivity for high performance distributed applications.

CARRIOCAS project, leaded by Alcatel-Lucent, gathers 25 partners in the frame of competitivity cluster system@tic.

"TERATEC has proved to be, two years after its creation, the success of its eco-system model"

5 – We participate in the European Project PRACE under GENCI (Grand Equipement National de Calcul Intensif) coordination.

6 – We are associated with three master degrees specialised in HPC, in cooperation with the most prestigious education institutions in Paris.

7 – Every year, we organise a European HPC forum (conferences – symposia and exhibition). The next one will be held June 3-4 near Paris at Evry (France) and we participate in a lot of dedicated forums or congresses.

8 – We had important discussions with our colleagues from USA and Japan, and with the major European initiatives.

We expect to contribute actively to the HPC activities in Europe with such activities, creating an ecosystem strongly involved in the Research and Economical challenges.

Europe's background in Applied Mathematics and Computer Sciences is worldwide recognized. HPC is one of the key areas for competitiveness and innovation, both for scientific challenges and for industry and services. We need to transform these facts into successes. **HPC is now of a strategic importance, but it should turn to be a high priority in Europe**, as it is in USA and in Japan, for it is fundamental in all the major challenges of the 21st century. In order to achieve that, important R&D activities with adequate financial support are necessary, both for basic technologies (hardware – software) and for applications. We also need to have access to the most powerful systems in terms of computing power, storage capacities and networking.

TERATEC has proved to be, two years after its creation, the success of its eco-system model. It is ready to contribute actively with other European initiatives to achieve the necessary success of HPC in Europe. It will propose the best computing facilities, it will continue launching initiatives for studying major scientific and industrial challenges and it will participate to the development of the HPC economy.

HPC is a need and a chance for Europe.

A Forward Look exercise on computational sciences infrastructures

by Simone Meloni, Scientific Officer of the ESF Forward Look on "European Computational Science: The "Lincei Initiative": from computers to scientific excellence".

In the past decades, most of the European countries have established an infrastructure for computational sciences, that is scientific research conducted by means of computers. So far, these infrastructures consisted of large computer facilities with some limited technical support for porting scientific codes on a specific machine. The situation is quite similar outside Europe. Scientists and funding agencies of these countries are now reflecting on the adequacy of such a kind of infrastructure for supporting computational sciences of the next decade. The European Science Foundation, in collaboration with several national institutions and European networks, has promoted and funded a foresight study named "Forward Look on European Computational Science: "the Lincei Initiative": from Computers to Scientific Excellence". The aim of this Forward Look is to analyse the current state of the computational infrastructures in European countries and propose recommendations for its evolution in the coming 10 to 20 years on the basis of the foreseen scientific and technological evolution.

Seven workshops involving European top researchers in materials sciences, nanotechnology, astrophysics, chemistry, engineering, climatology, biology, physics, meteorology and other disciplines have been organised. Representatives of supercomputer centres and major ICT companies have also participated to these workshops. It emerged that multi-disciplinary multi-scale simulations will be the major scientific challenge of the next decade. In multi-disciplinary multi-scale simulations, techniques developed in several disciplines and models at different scales in space and time are combined for simulating real complex systems at an affordable computational cost. A typical perspective offered by multi-disciplinary multi-scale approach is that of the simulation of a living cell. In this case, very accurate atomistic simulations are needed to simulate chemical reactions, like production of proteins in ribosomes. At the same time, continuum models can be used to simulate the effect of solvent (water) in regulatory processes, like in proton diffusion across cell membranes. All these information can be combined into a model of the life cycle of cells under different condition without carrying out any real (expensive and, sometimes, ethically problematic) experiment. It is also expected that experimental data will be directly used in computer simulations in the future. Indeed, this is already done, for example, in climatology, meteorology and astrophysics. At the same time, since computational sciences are still rather young, new methods and algorithms will also continue to emerge within each single discipline and time/space scale.

On the same foot, experimental and simulation data need to be made accessible to the scientific community at large, not only to the members of the scientific institution were data have been produced. For example, the Virtual Observatory is a first attempt to make theoretical and experimental astrophysics data available worldwide using standard metadata, data and file format. This will help interoperability with simulation codes.
Unfortunately, current scientific application software and archives seem not adequate to meet these challenges. The main reason of this inadequacy is the model of development, maintenance and support. In fact, these tasks are actually carried out by scientists themselves. This often results in software that is not flexible and unable to interoperate with codes other codes. At the same time, often this software is not able to exploit the performances of best-in-class supercomputers. The result is that typically scientific codes have a short life cycle. In addition, the time invested for developing and maintaining scientific codes and archives is not rewarding for the scientific career and this makes this task, and the hard work of acquiring the needed IT expertise, little appealing.
In addition, scientific codes are not able to exploit performance of modern supercomputers, which are expected to reach petaflops in few years. The situation is expected to become even more complex with the advent of "exotic" architectures, such as GPU (Graphic Processing Unit), for scientific and technical computing.

On the basis of this analysis, this Forward Look proposes to make management of scientific codes and data archives part of the (future) European Computational Infrastructure. Of course, the initial stage of software development has to remain part of research activity. So, while implementing the initial version of an algorithm for demonstrating its accuracy or effectiveness is still a research duty, the inclusion of its generalised version into community codes is a technical task that should be accomplished by professionals.

In order to achieve these objectives, this Forward Look envisages an extension of the current infrastructure based on three pillars:

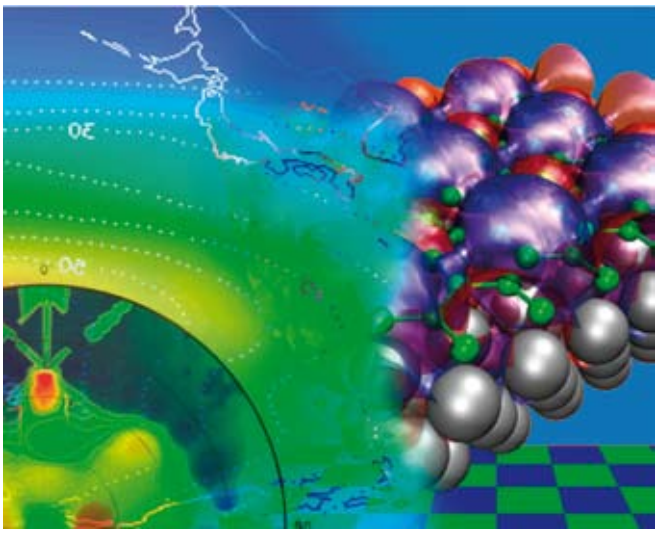
- standardisation
- development and maintenance of codes and archives
- training and support

Some standardisation is needed as it is the lack of standards that makes difficult code interoperability among them and with experimental data. Design and development of well engineered scientific codes should be zcarried out by software engineers and IT experts. This will improve flexibility by improving modularity, a key factor as scientific codes are in continuous evolution with new algorithms and models being inserted every week. Thanks to a modular design, scientists could easily insert the initial version of their novel algorithm into codes that already provide the other modules needed for the complete simulation. Modularity will also improve performance. For example, it would be possible to optimize an algorithm for specific computer architecture, adopting new programming languages if needed. Analogous considerations hold true also for data. A proper use of such complex software and archives requires regular courses and day-by-day support. So far, the scientists, sometimes receiving some support from scientific networks, have provided these services voluntarily. However, this is insufficient and, in practice, the presence of a proper support structure will be one of the major promoting factors in the adoption of computational tools by the scientific community at large, including experimentalists, and by industrial partners.
Of course, all this should be coordinated at a European level, as science is already organised at this level.

In conclusion, European scientists feel that the computational infrastructures need some restructuring, complementing the current infras-

"... standardisation is needed as it is the lack of standards that makes difficult code interoperability"

tructures, oriented toward the provision of hardware, with a second leg oriented toward design, development, maintenance and support of scientific application software and data archives. If such an infrastructure would be in place, not only the scientific community would unlock its power but also collaboration between academic institutions and private companies would be improved.



A picture "summarizing" Computational Sciences.

Computer -Aided Drug Discover Efficiency Requirements in the U

Anna Maria Bianucci[1], Emo Chiellini[2*] and Ilaria Massarelli[2]

[1] Department of Pharmaceutical Sciences,University of Pisa, Via Bonanno Pisano, 6 56126 Pisa - Italy
E-mail: bianucci@farm.unipi.it

[2] Laboratory of Bioactive Polymeric Materials for Biomedical and Environmental Applications (BIOlab) - UdR INSTM, Department of Chemistry, University of Pisa, Via Vecchia Livornese 1291, 56122 S. Piero a Grado, Pisa - Italy
E-mail: emochie@dcci.unipi.it

Abstract

A brief survey of the Computer–Aided Molecular Design applied to drug discovery (CADD), with specific consideration on the efficiency requirements in the use of software tools, is reported.
The focal point in CADD, also implying the prediction of Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) of a drug candidate, is approached by a combination of structure-based and ligand-based (QSAR) methodologies.

1. Introduction

Computer-aided molecular design is being currently exploited for the discovery of new drugs. It starts now being used for handling a wide range of other biomedical issues. One of the most challenging fields, where it finds valuable applications, is represented by the prediction of absorption, distribution, metabolism, excretion and toxicity (ADMET) properties, that is gaining more and more place as a crucial branch of drug design. Indeed, R&D costs are increasing together with high failure rates of clinical candidates, which makes clear that undesirable drug properties and toxicity strongly contribute to pipeline attrition. Costly failure may be solely prevented by early identifying and removing undesirable leads. That can be achieved by enriching discovery programs with computational tools for ADMET prediction.

2. Computational tools most commonly exploited in Computer-Aided Drug Discovery (CADD)

Several different approaches, both of structure-based and ligand-based type, are being currently used with the aim of getting a direct insight on ligand-macromolecule interactions or developing Quantitative Structure-Activity Relationship (QSAR) studies of ligands. It is worth to briefly recall here that the so-called structure-based approaches make use of the three-dimensional (3D) structure of the macromolecular target, which has to be known with sufficient accuracy. Molecular modelling (1) and molecular docking (2) techniques combined with geometry optimization methods enable direct analysis of interactions between ligands and their target proteins with details at a molecular level. Such a type of approach allows predicting binding affinities between partners but doesn't account for most of the complex pharmacological profiles of biologically active molecules. The

so-called ligand-based (QSAR) approaches work by correlating many different molecular descriptors, which reflect chemical and structural properties of compounds, to the already known biological activity indicators of interest (both affinities or ADMET properties). QSAR approaches can make use of classical mathematical equations (3), solved by various regression methods, or can use Artificial Intelligence tools (4). A physical meaning may be only attributed to the simplest molecular descriptors, but often a QSAR model may be seen as a "black box" enabling the prediction of molecular properties arising from a sum of many different contributions, which are not directly amenable to molecular structures. That happens mostly when the model refers to in vivo properties.

On the side of resources optimization, computational techniques of particular interest seem to be provided by classification methods. They started being developed in the late seventies/early eighties, but only recently they started finding quite a wide application in QSAR (5). They work by associating molecular features to few nominal classes related to a biological indicator of interest (for example, Yes/Not or Yes/Uncertain/Not, with regard to a wanted or unwanted property).

The search for optimal compromises between fastness and accuracy, among different approaches, is required by the drug discovery process, when developing predictive models, as optimal choices strongly impact on the use of financial and human resources.

3. Problems encountered in CADD and possible solutions

Optimally performing predictive QSAR models must obviously be highly reliable, but, at the same time, they must be as less demanding as possible in terms of computational effort. Indeed the number of virtual molecules to be screened for each project may be over thousands and thousands. The preliminary selection of molecular descriptors exploited for building QSAR models turns out to be crucial. Taking into account their statistical significance is just one of the many issues to be considered. Very fast QSAR models are obtained by only using the so-called 2D type molecular descriptors, to which atom types and atom connectivities mostly contribute, while very accurate models are obtained by the so-called 3D-type descriptors, to which properties, depending upon the 3D molecular arrangement, do mostly contribute. Based on previously acquired experience, the ratio between CPU times required by a medium performance personal computer (MPPC) for calculating a ten of descriptors, referring to a molecule of about 50 atoms, is expected to be a few tens of seconds.

According to some of the above considerations, one would expect that QSAR models, based on 3D type structural information about the known ligands, were always more successful that the ones obtained by only using 2D type information. However, this concept may be deceptive, as handling 3D structural information requires that molecular structures are optimized, so that biologically active conformations are properly simulated. The most commonly followed protocol, which implies a simple search for low-energy conformations, did not appear to lead to reliable results in many cases; at the same time it may also be quite demanding in terms of computer time (the CPU time required by a MPPC for quantum-chemical optimization of a molecules comprising about 50 atoms is of the order of minutes or a few tens of minutes).

The problem of optimizing structures of molecules interacting with a given macromolecular target may be solved, instead, by docking the molecules of interest within a proper 3D model of the active site of such a target. Docking a pool of ligands into the proper macromolecular target requires that the target (3D) structure is known with good accuracy from experimental data (usually x-ray) or from molecular modelling. When the target 3D structure is obtained by means of molecular modelling, the computational effort may be significant, as Molecular Dynamics (MD) simulations are often required. Just for giving an idea of it, the CPU time required when performing a 100 ps MD simulation on a protein comprising about 5000 atoms, computed by of the AMBER (6) software with implicit solvent, is a few tens of hours, when running on a IBM SP Cluster 1600 (512 processors, 64 nodes). Moreover the target active-site structure may be affected by the interaction with each single ligand (induced-fit phenomena), that makes the required calculations even more complex and demanding in terms of computer time. Indeed most of the docking software currently available doesn't allow automated flexible docking of the target macromolecule (with the exception of some prototypes of computer programs), so time-consuming MD runs, for each ligand-macromolecule complex, need to be performed.

As a last consideration, it may be worth mentioning that, due to the wide diversity shown by many molecular libraries involved in the ADMET predictions, the combination of different strategies, summarized below, appears to lead to a successful use of tools developed "in silico". The whole QSAR problem should be split into a classification problem with a limited use of resources, and a series of sub-problems more accurately analyzed, with larger use of resources. The classification task enables rapid screening of wide diverse libraries just on the basis of a few nominal classes (see paragraph 2), while the subproblems are subsequently afforded by highly accurate local QSAR models only used for screening focused libraries.

4. Use of a combined (structure- and ligand-based) approach: an example of ADMET study

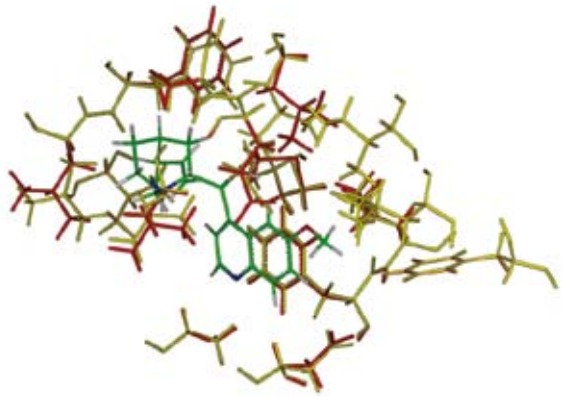
An example of combined approach (of structure- and ligand-based type), which has been successfully applied to ADMET studies, concerns the search for optimal QSAR models enabling the prediction of CYP2D6 inhibition properties of New Chemical Entities. CYP2D6 is one of the most relevant isoforms of Cytochrome P450, which is responsible for the metabolism of a huge number of drugs. The initial work was planned with the aim of estimating the predictive power of QSAR models obtained by using 2D only, or both 2D and 3D molecular descriptors, for estimating CYP2D6 inhibition properties of drug candidates. The ultimate target was to find the best compromise between accuracy and required computational effort (7).

In the specific case of CYP2D6, previous structure-based studies had shown that induced-fit phenomena arise during the interaction between the enzyme isoform and its ligands (8) (See Figure 1) . The conformations of few tens of known (structurally diverse) inhibitors, selected for building predictive QSAR models, were optimized into slightly different arrangements of the enzyme active site, previously fitted with a few structurally representative compounds.

"Molecular modelling and molecular docking techniques combined with geometry optimization methods enable direct analysis of interactions between ligands and their target proteins with details at a molecular level"

Then 2D and 3D molecular descriptors, including the quantum-chemical and thermodynamic ones, were computed and processed by a classical equation-based approach which exploits multi-linear regression. Several QSAR models were obtained and validated by applying rigorous statistical criteria (9) (See Figure 2). The comparison among different models, developed at different level of accuracy, showed that, while the QSAR models only based on 2D molecular descriptors appeared to be quite poor in treating the system of interest, the models based on 3D descriptors turned out to possess a quite good predictive power (10).

Figure 1. An example of ligand-induced fit. Superimposition of CYP2D6 binding site optimized in two different conditions: (a) (yellow) unliganded form and (b) (red) in the presence of an inhibitor. The inhibitor is colored by atom type.

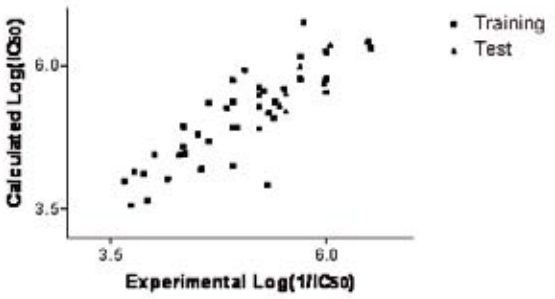


5. Conclusions

In order to be able of giving a valuable contribution to Drug Discovery, the "in silico" tools must be very efficient, i.e. they must be, at the same time, very fast and very accurate. In view of achieving high efficiency in performing CADD (including the development of computational models for ADMET predictions), a combination of structure- and ligand-based approaches may need to be exploited, at least when the 3D structure of the target macromolecule is known from experimental data or may be modelled with sufficient accuracy. Moreover, when widely diverse molecular libraries are involved in the project at hand, the winning strategy may arise from splitting the whole problem into a preliminary classification task and other tasks implying data analysis, made through very accurate local QSAR models.

Acknowledgment
The Authors wish to thank Dr. Marilena Saraceno for helpful assistance in the finalization of the manuscript.

Figure 2. Plot of calculated versus experimental values of the selected biological indicator (–Log IC50 of CYP2D6 inhibition) obtained by a well performing QSAR model.



6. References

(1) InsightII package, Accelrys, San Diego, USA; SYBYL7, molecular modeling software packages, Tripos Associates, Inc., St. Louis.
(2) Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J., "Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function", J. Computational Chemistry, 1998, 19, 1639-1662.
(3) (a) Katritzky, A. R., Lobanov, V. S., Karelson, M., (1994) "CODESSA: Reference Manual; Version 2", University of Florida.; (b) Zamora, I.; Oprea, T.; Cruciani, G.; Pastor, M.; Ungell, A. L. "Surface descriptors for protein-ligand affinity prediction", J. Med. Chem. 2003, 46, (1) 25-33.
(4) Witten, Ian. H., Frank, E. (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
(5) Platt, J. C., "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines", Microsoft Research, Technical Report MSR-TR-98-14, April 21, 1998.
(6) Case, D. A, Cheatham, T. E. III, Darden T., Gohlke, H., Luo, R., Merz, K. M. Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R., "The AMBER biomolecular simulation programs"., J. Computat. Chem. 26, 1668-1688 (2005).
(7) Marilena Saraceno, PhD Thesis – University of Pisa (Italy).
(8) Saraceno, M., Coi, A., Bianucci, A. M. "Molecular Modelling of human CYP2D6 and molecular docking of a series of ajmalicine- and quinidine-like inhibitors", IJBIOMAC, in press.
(9) Golbraikh, A., Tropsha, A., "Beware of q2!", J. Mol. Graphics Mod. 2002, 20, 269-276.
(10) Saraceno, M., Bianucci, A. M., "A Search of optimal fastness/accuracy compromises in developing predictive models for estimating CYP2D6 inhibition properties of drug candidates", Work submitted for presentation at the IXth Tetrahedron Symposium, Challenges in Organic and Bioorganic Chemistry - July 22-25, 2008 • Berkeley, CA, USA.

Climate modelling and supercomputing: WACCM at CESGA

Juan A. Añel[1],*, Luis Gimeno[1], L. de la Torre[1], Rolando R. García[2]

1 *Facultade de Ciencias de Ourense, Universidade de Vigo, Ourense, Spain*
2 National Center for Atmospheric Research, Boulder, Colorado, USA

**corresponding author: j.anhel@uvigo.es*

In the last couple of decades numerical modeling has become one of the main tools for the study of climate. The phenomenon of global warming and the implicit need of knowing how it will evolve in the next several decades makes climate models an indispensable tool to deal with this subject and to obtain responses for the questions formulated by researchers, policymakers and the general public.

Climate models are conceptually straightforward from a theoretical point of view. In essence, a climate model is a description of the climate system in terms of basic physical laws. The system is represented by a set of coupled equations corresponding to the laws that govern atmospheric dynamics and thermodynamics, which in turn are solved by means of a suitable computer code. Climate models can be very simple or very complicated, depending on how much detail about the climate system they include, and the computational resources required for solution can be extremely high for comprehensive models. However, even in this case, a climate model remains a simplification of the real world. Here we discuss a project developed at CESGA during the past year: the implementation and use of a state-of-the-art climate model which encompasses the entire atmosphere up to an altitude greater than 100 km. The implementation discussed is relevant for research on the tropopause and the stratosphere, two of the main topics in current climate research. This model in question is the Whole Atmosphere Community Climate Model (WACCM), developed by the National Center for Atmospheric Research (NCAR) of the USA. WACCM has been developed and improved since 1999 and the first results from interactive calculations were published in 2003.

What makes WACCM different from most of the climate models are mainly two features: first, WACCM has a deep vertical domain, with 66 vertical levels from the ground to the model top, around 140 km above the Earth's surface. The code is flexible in that it allows the number of vertical levels to be adjusted according to the research application. Second, WACCM includes a detailed chemical mechanism, adapted from NCAR's Model of Ozone and Related Chemical Tracers (MOZART). In contrast, most climate models have their top around 30-40 km, covering only the troposphere and part of the lower stratosphere, and include only basic chemistry. The inclusion of detailed, interactive chemistry in a climate model requires unusually

large computational resources, but it is very desirable feature when one wants to study the upper atmosphere and obtain the most representative results. In general terms the current standard configuration of WACCM is:

• horizontal grid : $\Delta\varphi \times \Delta\lambda = 1.9^o \times 2.5^o$ ($\Delta\varphi$ = latitude; $\Delta\lambda$ = longitude);
• 66 vertical levels;

This configuration has variable vertical resolution of about 1.2 km in the troposphere and lower stratosphere, decreasing to about 3 km in the thermosphere. This standard configuration is normally implemented on the Blueice supercomputer at NCAR, which has among its main characteristics a propietary AIX system (IBM Unix) on Power 5+ processors with a 1.9-GHz clock cycle. Running on 128 Blueice cpus (8 x 16 cpu nodes), the computational performance is about 2.7 years of simulated climate per wallclock day. That is, 24 hours of continuous calculation produces 2.7 years of simulated climate. The volume of data output is approximately 25 GB per simulated year for this standard configuration. For testing purposes, WACCM version 3 was installed in CESGA's SuperDome HP-UX system (propietary HP Unix, running on Itanium2 processors with a 1.5-GHz clock cycle). The implementation was carried out in collaboration with the personnel from CESGA and NCAR. The configuration of WACCM used in the SuperDome tests was constrained by the computational limitations of this machine both as regards cpu clock rate and number of processors available. Under these conditions the configuration of the model was:

• horizontal grid : $\Delta\varphi \times \Delta\lambda = 4^\circ \times 5^\circ$;
• 66 vertical levels (configured as in the $1.9^\circ \times 2.5^\circ$ version);

Diagnostic simulations performed with this implementation of WACCM on the SuperDome produced realistic climate results. Moreover, a simulation of five complete years was carried out in order to obtain a small subset of climatic statistics. Obviously, the complete results cannot be reproduced here, but Figure 1 is an example from one of the diagnostic simulations. The computational performance obtained with the above implementation of WACCM on SuperDome was 1.6 years of simulation per wallclock day using 32 cpu's. If we had run WACCM at high resolution ($\Delta\varphi \times \Delta\lambda = 1.9^o \times 2.5^o$) on the SuperDome using the same number of cpus (32), we would have obtained performance no better than 0.4 years of simulation per wallclock day. Working with 64 cpu's a scaling in the performance not larger that about 85% was expected. Aditionally, from the computational point of view, WACCM can take advantage of symmetric multi-processing (SMP) and simultaneous multithreading (SMT), as well as hybrid parallelization techniques using shared memory with Open Multi-Processing (OpenMP) within a node and distributed memory across nodes with Message Passing Interface (MPI). These techniques allow a substantial improvement in computational performance.

Our projected research using WACCM at CESGA includes plans to:

• build WACCM at the new CESGA's supecomputing facilities, namely to migrate the model to Finis Terrae (free GNU/Linux system on Itanium2 processors). A similar configuration has been tested at the Columbia supercomputer of the National Aeronautics and Space Administration (NASA), Ames Research Center;
• obtain a basic subset of climate statistics and diagnostic simulations in order to evaluate whether WACCM on Finis Terrae produces a realistic climate;
• work with two different configurations of WACCM, the standard and a new one with high vertical resolution around the tropopause;
• compare the trend of different tropopause variables with observations and simulate their evolution under future climate change scenarios;
• compute climate scenarios for 50-100 years with doubled CO2 concentrations;
• study possible variations in the stratospheric polar vortex and model its vertical structure.

"Diagnostic simulations performed with this implementation of WWCCM on the SuperDome produced realistic climate results"

Acknowledgements
The authors would like to thank to José Carlos Mouriño from CESGA and Stacy Walters from NCAR for their work on the implementation of WACCM on the SuperDome. We would like to thank CESGA for providing the necessary facilities and computational resources for this project.

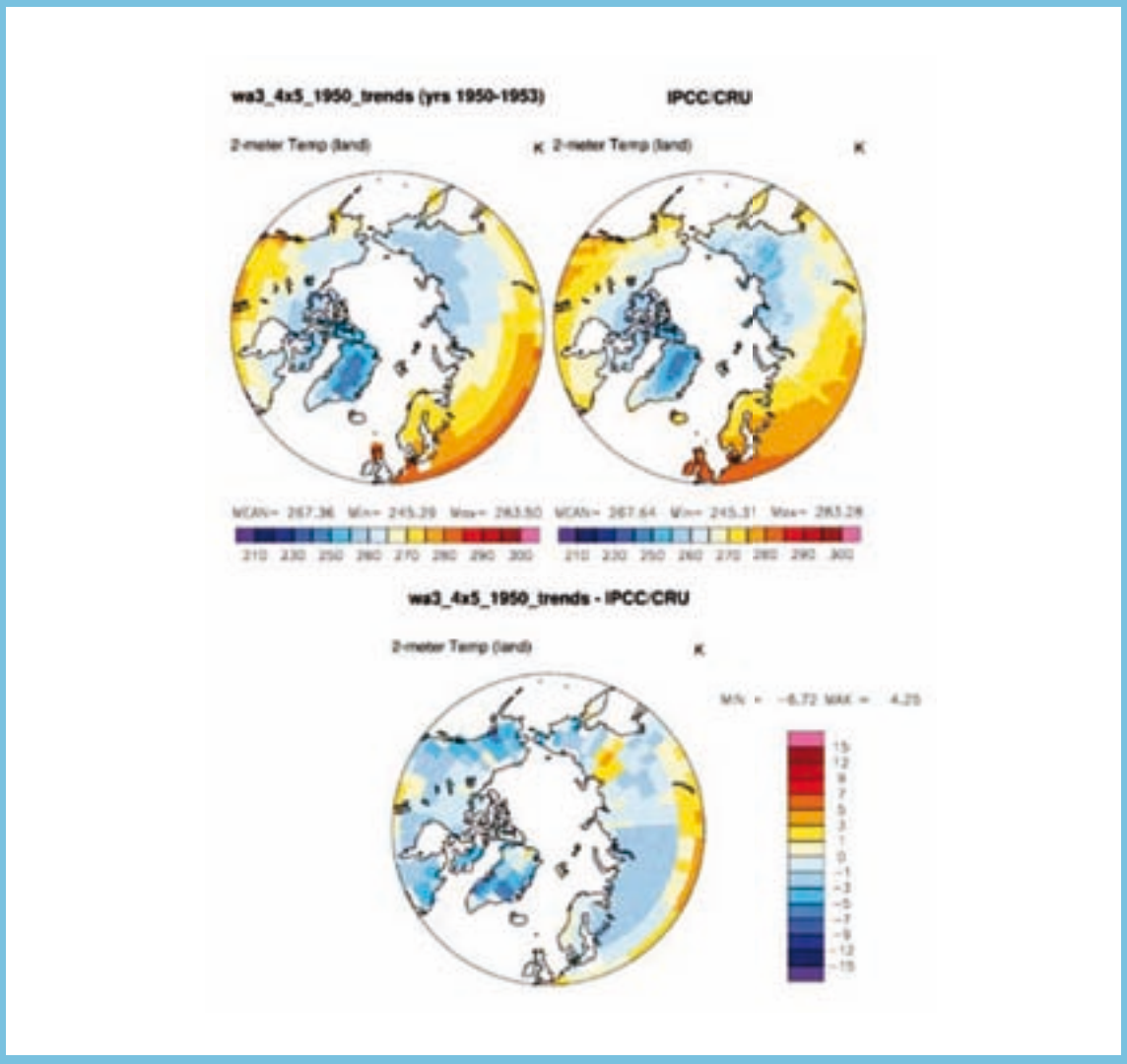


Figure 1. Comparison of the Northern Hemisphere polar plots corresponding to the 2 meters temperature trend on land for the period 1950-1953. Upper left: results using WACCM on the CESGA's SuperDome with $\Delta\varphi \times \Delta\lambda = 4^{\circ} \times 5^{\circ}$.
Upper right: observations obtained from the Intergovernamental Panel on Climate Change (IPCC) and the Climate Research Unit of the University of East Anglia (CRU). Lower: results of substracting the values from the observations from those obtained with WACCM.

Structural, dynamic, and energetic characterization on cyclodextrin-based supramolecular systems

Ángel Piñeiro[a,†], Alfredo Amigo[b], and Pilar Brocos[b]

[a]*Departamento de Fisicoquímica, Facultad de Química, Universidad Nacional Autónoma de México (UNAM), Cd. Universitaria, México D.F. 04510, Mexico*

[b]*Departamento de Física Aplicada, Facultad de Física, Universidad de Santiago de Compostela (USC), E-15782 Santiago de Compostela, Spain*

[†]*Present address at USC: fangel@usc.es, famigo@usc.es, and fapili@usc.es*

Cyclodextrins (CD) are cyclic oligosaccharides with the shape of a hollow truncated cone (see Fig. 1) that have a hydrophobic cavity able to host non-polar molecular groups. Aside from their well-known utility for encapsulation and transport purposes (http://en.wikipedia.org/wiki/Cyclodextrin), their structural features make them suitable as mobile pieces of molecular machines amongst other supramolecular assemblies. Our main goal is to design CD-based supramolecular systems with special properties that could be controlled by external variables such as temperature and electric fields, by pH changes, or by the presence of co-solutes. That kind of systems could be employed to design new materials or to build biological sensors, among other potential applications. Experimental results like thermodynamic or mechanical parameters, as well as microscopy observations at different resolutions, are complemented and better understood by means of molecular dynamics simulations at atomic level.

Recently we have discovered that short time molecular dynamics simulations suffice to observe reversible and spontaneous rupture-formation processes of supramolecular complexes consisting of one or two cyclodextrins and one or two surfactant molecules (see Fig. 2). Using specific algorithms, the same technique can be employed to determine the dependence of Gibbs energy change on some given reaction coordinate such as an angle or a distance, which entails getting a map of reaction probabilities as a function of the coordinate values. Up to now, we have considered complexes involving native CDs as hosts and either sodium dodecyl sulfate (SDS) or octyl-β-D-glucopyranoside as illustrative cases of ionic and non-ionic guests, respectively. Thus, the populations of the different feasible stoichiometries were assessed, the structures of the corresponding cyclodextrin-surfactant complexes were investigated, and the energy of these systems is being characterized. These computational results complement very well those obtained by our collaborators from experimental techniques like calorimetry or surface tension measurements.

Experimental results suggested that taking the abovementioned complexes as building blocks, films are spontaneously formed at aqueous solution/air interfaces. By combining Brewster angle and atomic force microscopy with molecular dynamics simulations of these systems, it was proved that those films consist of nanotubular structures formed either by self-assembled empty cyclodextrin dimers (CD_2) or by CD_2 with an ionic surfactant molecule in the cavity of the structure. It was proved that the mechanical properties of such films can be controlled by several external parameters and, thus, new cyclodextrin-based materials with specific properties could be designed.

Some of the previous results have already been published (see below) but much work is still in progress. Other projects are also being considered. For instance, modified CDs have been employed to bind proteins to gold electrodes although no much information at atomic level has been obtained from these systems. Molecular dynamics simulations could help to understand how this binding process affects the folding state of proteins. Besides, interaction between proteins (or polypeptides) and lipid molecules in solution or at aqueous solution-air interfaces has also been studied by several experimental techniques. In this context, we are starting a project to study the adsorption of apolipoproteins (proteins involved in the transport of cholesterol through the bloodstream) to solution-air interfaces in the presence of several lipids and other co-solutes.

Our computational work is always designed under the principle that it should complement or/and help to understand experimental results. This is possible in many cases because classical simulations at atomic level can manage several thousands of particles at time scales of up to microseconds with the current computational facilities. Trajectories are generated using the GROMACS package and they are analyzed with molecular viewers such as RasMol, VMD, and PyMOL, with tools from the GROMACS package and with programs specifically developed in C, perl, awk, and UNIX code. Parallelization of the processes is usually needed to obtain long trajectories and high-end computing is required for this purpose. Currently, computer clusters of thousands of processors with architecture designed for efficient parallelization are maintained by many institutions and, in some cases, even some groups have clusters of these characteristics for their individual use. In our case, the HP Superdome and the SVG clusters, maintained by the Centro de Supercomputación de Galicia (CESGA), were employed.

Collaborating Organizations/ Researchers

Norma Díaz-Vergara, Xavier Banquy, Jorge Hernández-Pascacio, Cristina Garza, Edgar Tovar, Abel García, Salvador Ramos, Rolando Castillo, Silvia Pérez-Casas, and Miguel Costas from UNAM (Mexico); Alessandra Villa from J.W. Goethe Univ. (Germany), and Alan E. Mark from Univ. of Queensland (Australia).

Grants

The cyclodextrin-based complexes project was financially supported by CONACYT (grants 41328-Q and J49811-Q), PAPIIT-UNAM (grant IN105107), and MEC (grant FIS2007-63479). P.B. thanks USC and Xunta de Galicia for supporting her with research stay grants under the programs “Axudas Propias 2006” and “PGDIT-INCITE 2007”, respectively.

Acknowledgements

Thanks to the developers of essential free software like GROMACS (http://www.gromacs.org/), RasMol (http://www.umass.edu/microbio/rasmol/), VMD (http://www.ks.uiuc.edu/Research/vmd/), and PyMOL (http://pymol.sourceforge.net/). Technical support from Systems and

"New cyclodextrin-based materials with specific properties could be designed"

Applications areas of CESGA is also gratefully acknowledged.

Scientific Publications

• Cyclodextrin-Based Self-Assembled Nanotubes at the Water/Air Interface. The Journal of Physical Chemistry B, 111 (2007), 12625-12630.

• On the Characterization of Host-Guest Complexes: Surface Tension, Calorimetry, and Molecular Dynamics of Cyclodextrins with a Non-ionic Surfactant. The Journal of Physical Chemistry B, 111 (2007), 4383-4392.

Three more articles derived from this project are currently in preparation and others are being planned.

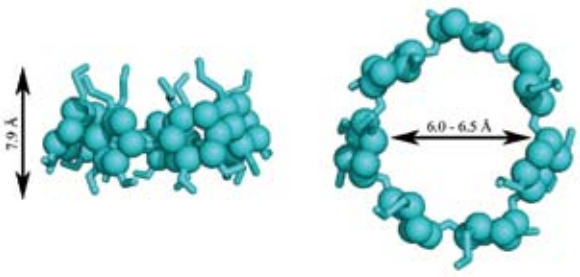


Figure 1.- Two views of the β-CD structure.

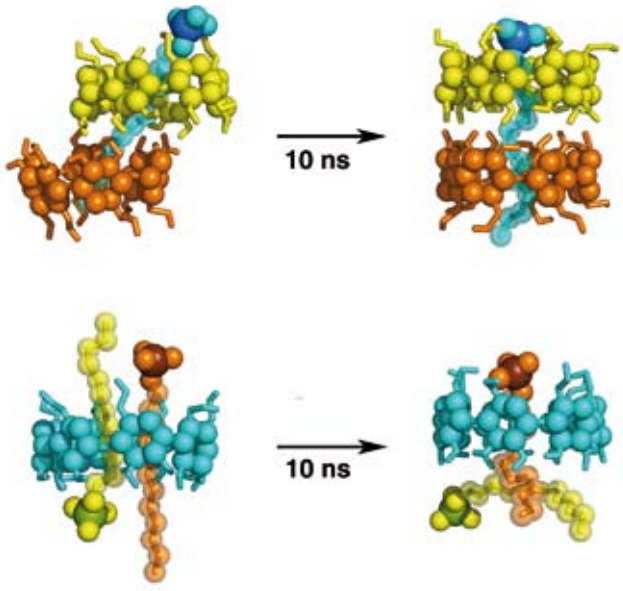


Figure 2.- Proposed initial conformation for a α-CD_2SDS_1 (left-top) and a β-CD_1SDS_2 complex (left-bottom), together with the structures obtained after 10 ns of molecular dynamics simulation.

A couple of atomistic studies on the Cementitious Calcium Si

Jorge S. Dolado [1,2] (*), H. Manzano [1], A. Ayuela [3], Jan Hamaekers [4], Michael Griebel [4]

[1] LABEIN-Tecnalia, Parque Tecnológico de Bizkaia, Derio, Spain.

[2] Nanostructured and Eco-efficient Materials for Construction Unit, Associated Unit LABEIN-Tecnalia/CSIC, Derio, Spain.

[3] Unidad de Física de Materiales, Centro Mixto CSIC-UPV/EHU ; and Donostia Internacional Physics Center. PO Box 1072. San Sebastián/ Donostia. Spain.

[4] Institut for Numerical Simulation, University of Bonn, Wegelerstraße 6, D-53115 Bonn, Germany.

** Corresponding author. E-mail: jsanchez@labein.es*

Abstract

This work aims to highlight the importance of atomistic computational approaches to describe the structure and properties of most important hydration product present in cementitious materials, the C-S-H gel. To this end two recent studies will be presented, along with a brief discussion about the future goals and perspectives on this topic.

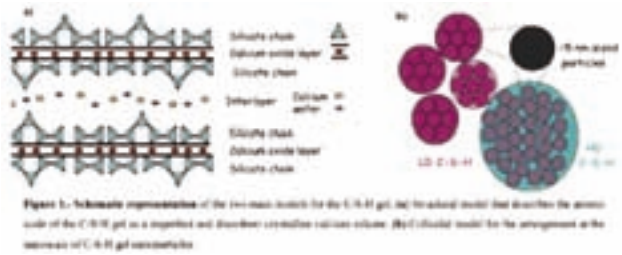
I. Introduction

Cement-based materials, such as mortars, fibre-reinforced cementitious composites, and concrete are the most widely used construction materials in the world. They are indeed complex composite materials with truly multi-scale internal structures that evolve over centuries: Just by pouring cement powder to water, a myriad of chemical reactions take place to form a rigid, complex and porous structure that is called cement paste. This cement paste is a multiphase material which evolves over time. Basically it can be viewed as a composite in which calcium hydroxide crystals (portlandite), (sulpho)aluminate hydrates, and non-reacted cement powders are embedded into an amorphous nanostructured hydration product, the so-called C-S-H gel (Calcium silicate hydrate) [1].

Calcium silicate hydrate (C-S-H) gel is definitely the most important hydration product of cement based materials. It constitutes about 60-70% of the fully hydrated cement paste and is responsible for most of the properties of cement-based materials. From a compositional point of view, C-S-H gel is often characterized by its Ca/Si ratio, which typically ranges from 0.7 to 2.3. This variability in composition explains why, though intensively studied by techniques like XRD, SEM, TEM, NMR, etc [2], many features of C-S-H gel remain unknown. Our incomplete picture of C-S-H gel relies on two complementary visions, each one oriented to explain different sizes and phenomena (see Fig. 1): On the one hand, we count with the works that have centred their analysis on the recognition of the structural features of C-S-H gel at the molecular level. Much of our knowledge at this level has been gained from structural comparisons with crystalline calcium silicate hydrates. In fact, several models [3-8] have been proposed so far that draw structural analogies with distorted tobermorite (T) and jennite (J) crystals. From these models, C-S-H gels can be approximately

licate Hydrate (C-S-H) Gel

viewed as layered structures, in which calcium oxide sheets are ribbed on either side with silicate chains, and free calcium ions and water molecules are present in the interlayer space (Figure 1(a)). On the other hand, other segment of works has paid attention to the recognition of the colloidal properties of C-S-H gel [9,10] .These models basically consider that the C-S-H gel is composed of approximately 5nm-sized rounded particles, which, in turn, aggregate to form High-density (HD) C-S-H or Low-density (LD) C-S-H depending on a packing factor (Fig.1 (b)).



The existence of two varieties of C-S-H gel (the LD- and HD-ones), is of paramount importance from a technological point of view. Thanks to nanoindentation experiments [11, 12] we know that both varieties of C-S-H exhibit a dissimilar mechanical/bearing capacity, which is even more pronounced in their resistance to degradation processes [13]. Akin to what happens with the cholesterol we can speak about “good” and and “bad” C-S-H gels. In our case the “good” one is HD C-S-H, being the LD-C-S-H the “bad”one.

This said, the question that arises is straightforward: Can we promote the appearance of HD-varieties over the LD-ones? To the best of our knowledge the answer to this question doubtlessly represents the most important technological challenge for cementitious materials. Unfortunately, our knowledge on C-S-H gel is too vague at present so as to realistically face such a challenge. Prior to everything, we need to bridge the gap between the structural and colloidal models of C-S-H gel. To this end, it is clear that an ineludible prerequisite is to obtain further knowledge on the short-range ordering of C-S-H particles from which the colloidal models start; namely we need to change our current picture of these basic C-S-H bricks (based on ~5-nm sized particles) by an improved version which accurately accounts for their actual morphology and chemistry.

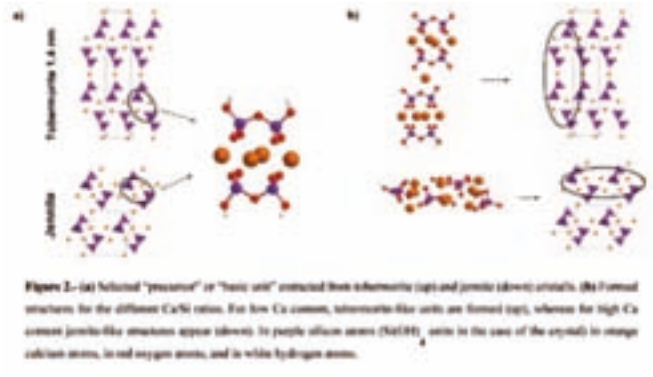
In this scenario, it is clear that any hint on the structure of C-S-H gels that can be given by modelling and numerical simulations would be tremendously beneficial. Although atomistic simulations have been barely applied to deal with C-S-H gel [14-21] there is lately a great agreement that they represent an unmatched opportunity to unravel the smallest features of C-S-H gel [17]. This paper aims precisely to explain two recent studies of our computational research program in this direction, as well as to highlight the most promising possibilities and challenges of the computational schemes for the near future.

II. Two examples of our activity in this area

Example 1: Formation of small T-like and J-like pieces

Our first example is the work made by Manzano et al. [22] where the underlying mechanisms that govern the formation of small pieces of C-S-H gel were analyzed by means of ab-initio calculations at the Hartree-Fock level (this is, solving Schrödinger equation). Differently to previous studies which rested on perfect tobermorite or jennite structures [14, 15, 17, 20, 21], in this ab-initio approach we did not impose any structural model. By taking as starting point an adequate guess for a precursor (P) of tobermorite-like and jennite-like pieces (see Fig 2a) composed of two silicate dimers and a single calcium oxide layer, we studied the self-assembly process of these basic precursors under different Ca/Si ratios.

Interestingly, the formation of T-like and J-like structures turned to be a natural self-assembly process of these basic precursors under different Ca/Si ratios. Two dissimilar and competing growth mechanisms were identified to rule the mentioned self-assembly processes, dipole-dipole interactions and bond formation. The first one dominates at low Ca/Si ratios and leads to the formation of T2-like units, whereas the second one is the leading mechanism at high Ca/Si ratios, and gives rise to the appearance of J2-like units (see Fig 2b).

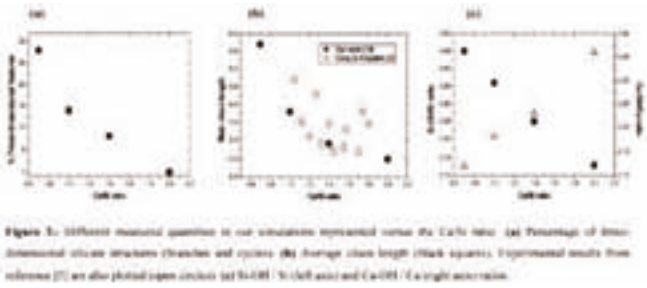


Example 2: A Molecular Dynamic study on the formation of C-S-H structures.

Our second example is a recent work [23] in which we studied the formation of C-S-H structures by means of a Molecular Dynamic study of the polymerization of silicic acids (Si(OH)4) in presence of solvated calcium ions (Ca(OH)2 4 H2O). By changing the number of the solvated calcium ions present in our simulation unit cell, we have analyzed the number and types of siloxane bonds (Si-O-Si) (i.e. the connectivity of the silicate chains) along with the number of Si-OH and Ca-OH bonds as a function of the Ca/Si ratio.
We have found that the appearance of three-dimensional networks clearly decreases when the amount of calcium ions is raised (see Fig. 3a). This result is in clear agreement with the experimental evidence [5, 24] that shows that the linear forms are dominant. We also analy-

"The appearance of three-dimensional networks clearly decreases when the amount of calcium ions is raised"

zed the main chain length (MCL) of the silicate chains as a function of the Ca/Si ratio. We compared our results (solid squares) with the experimental data obtained by Cong and Kirkpatrick [5] through 29Si NMR experiments (open circles). As can be seen from Fig 3b, our results are in reasonable agreement with the experimental values and the trend is reproduced that shorter chains are formed when the Ca/Si ratio increases [5, 24]. Finally we counted the amount of Si-OH and Ca-OH bonds. These bonds are important fingerprints for assessing the presence of both tobermorite and jennite-like features. No Ca-OH bond is present in perfect tobermorite crystals, whereas no Si-OH exists in perfect jennite crystals [25, 26]. The results obtained from our simulations are in Fig 3c. Basically our simulations suggest that tobermorite and jennite features are appropriate structural models at low and high Ca/Si ratios respectively.



According to our Molecular Dynamic study, we can suggest the following structural explanation of C-S-H gel: At low Ca/Si ratios cementitious C-S-H gel could be seen as mixtures of long polymerized (pentamers and longer chains) 1.4-nm tobermorite and jennite structures, whereas at high Ca/Si ratios they seem to be composed of short (dimers) 1.4-nm tobermorite and jennite pieces. In between there is a gradual evolution from long to short chains, and from tobermorite-like to jennite-like features.

III. Computational methods and resources

The above studies intend to illustrate how the atomistic simulations are nowadays providing an unmatched description of cementitious nanostructure. We must say, however, that the prize to pay for this accurate picture is the high cost in terms of computational resources. The atomistic simulations are very high demanding computer-wise. Consequently, their description would be clearly unattainable without counting with efficient computation codes along with superb computational infrastructures. A brief description of the methods and resources employed in the two examples shown above is summarized in the following:

When searching for the mechanisms that lead the formation of Tobemorite and Jennite pieces (Example 1) the computational approach rested on Hartree-Fock calculations. They were performed under the GAMESS [27] and GAUSSIAN [28] codes on the HPC 320 and Superdome clusters at the Supercomputing Center of Galicia (CESGA).

The Molecular Dynamic study of C-S-H gel (Example 2) was carried out with help of the molecular dynamic software package TREMOLO [29, 30], which is a load-balance distributed memory parallel code. All the MD simulations of this study were performed on the Himalaya cluster at the University of Bonn.

IV. Conclusions and future challenges

Thanks to the ever-growing current computational resources we are obtaining an unprecedented description of the smallest bricks which make up cementitious structures by employing very high demanding (computational-wise) atomistic simulations. Our feeling is that we are on the verge of identifying the basic C-S-H bricks and bridging the existing gap between the structural and colloidal vision of C-S-H gel. Some works are already pointing toward such a direction.

To obtain a comprehensive computational (and not computational) description of the aggregation and self-assembly of the C-S-H bricks to form C-S-H gel (either the LD or HD variety) is our following challenge. The employment of computational techniques based on "brute force", such as the atomistic simulations, is hardly a wise strategy to deal with so big systems. It is known that the C-S-H gel assemblages occupy large spatial regions (above 100nm3) [1], so any realistic computational description should employ unit cells of at least the same dimensions. The volume of the unit cells and the number of atoms (about 1000 times larger than those we are currently employing to study the basic C-S-H brick) make suck kind of calculations be clearly unfeasible nowadays. Since it is a piece of evidence that any improvement of our computational resources & capacities would enable a faster achievement of our technological dream (i.e. the promotion of HD-C-S-H varieties over the LD-ones!), any action in such a direction would be strongly beneficial.

Nevertheless, the development of more efficient computational algorithms and schemes is something that is also mandatory. In this sense, it is clear that our ultimate computational challenge is, in fact, to provide smooth handshakes and relationships between existing computational methods, to cross over different spatial and temporary scales and eventually lead to truly multi-scale descriptions. Only by this procedure, together with a close feedback with experimental data, will our technological dream (i.e. the promotion of cementious structures mainly made of HD C-S-H varieties) be accomplished. It is worth mentioning that this strategy is indeed the one recently followed by the CODICE project (COmputationally Driven design of Innovative CEment-based materials), a collaborative initiative that has been recently successful in the FP7 program [31]. This challenging project aims to develop efficient multi-scale computational schemes to guide the design of new, stronger and much more durable cementitious materials, and represents an excellent framework so that our "cement-based dreams" come really true.

Acknowledgements
This work has been supported by the Basque Government (NANOMATERIALES and HIBRICEM projects), the Spanish Government

"... we are on the verge of identifying the basic C-S-H bricks and bridging the existing gap between the structural and colloidal vision of C-S-H gel"

(project MONACEM (Ref-MAT2005-03890)), and, in parts, by the Schwerpunktprogramm 1165 and the Sonderforschungsbereich 611 of the Deutsche Forschungsgemeinschaft.

References

z[1] H. F. W. Taylor, Cement Chemistry, 2nd ed., Thomas Telford, London, (1997).
[2] V. S. Ramachandran and J. Beaudoin (eds.), Handbook of Analytical Techniques in Concrete Science and Technology, Building Materials Series (William Andrew Publishing/Noyes Publications, Norwich, New York, 2001.)
[3] H. F. Taylor, Proposed structure for calcium silicate hydrate gel, J. Am. Ceram. Soc. 73 [6] 464-467 (1986).
[4] I. G. Richardson and G. W. Groves, Models for the composition and structure of calcium silicate hydrate (C-S-H) gel in hardened tricalcium silicate pastes, Cem Concr Research, 22 1001-1010 (1992).
[5] X. Cong and R. J. Kirkpatrick, 29Si MAS NMR study of the structure of calcium silicate hydrate, Adv. Cem. Based Mater, 3 144-146 (1996).
[6] A. Nonat and X. Lecoq, The structure, stoichiometry and properties of C-S-H prepared by C3S hydration under controlled conditions, pp 197-207 in Nuclear Magnetic Resonance Spectroscopy of Cement-Based Materials. Edited by P. Colombet, A. R. Grimmer, H. Zanni and P. Sozzani. Springer, Berlin (1998).
[7] J. J. Chen, J. T. Thomas, H. F. Taylor, and H. M. Jennings, Solubility and structure of calcium silicate hydrate, Cem ConcrResearch, 34 1499-1519 (2004).
[8] I. G. Richardson, Tobermorite/jennite and tobermorite/calcium hydroxide-based models for the structure of C-S-H: Applicability to hardened pastes of tricalcium silicate, b-dicalcium silicate, Portland cement, and blends of Portland cement with blast-furnace, metakaolin, or silica fume, Cem Concr Research, 34 1733-1777 (2004).
[9] H. M. Jennings, A model for microstructure of calcium silicate hydrate in cement paste, Cem Concr Research, 30 101-116 (2000).
[10] P. D. Tennis and H. M. Jennings, A model for two types of calcium silicate hydrate in the microstructure of Portland cement pastes, Cem Concr Research 30 855-863 (2000).
[11] G. Constantinides and F-J. Ulm, The efect of two types C-S-H gel on the elasticity of cement-based materials: Results from nanoindentations and micromechanicalmodeling. Cem Concr Research, 34, 67-80 (2004).
[12] G. Constantinides and F-J Ulm, The nanogranular nature of C–S–H, Journal of the Mechanics and Physics of Solids, 55, 64-90, (2007).
[13] H. M. Jennings, J. Thomas, J. S. Gevrenov, G. Constantinides and F-J. Ulm. A multitechnique investigation of the nanoporosity of cement paste, Cem Concr Research, 37 (3), 329-336 (2007).
[14] P. Faucon, J. M. Delaye, J. Virlet, J. F. Jacquinot and Adenot, Study of the structural properties of the C-S-H (I) by molecular dynamics simulation, Cem Con. Research, 27 (10), 1581-1590, (1997).
15] S. Kashihara, S. Yamanaka, T. Inoune, T. Komatsu and H. Toyoshima.Quantum mechanical determination of the Al-substituted site in tobermorite, J. Am. Ceram. Soc. 11, 3023-3026, (1994).
[16] M. Rahman, S. Nagasahi and S. Tanaka, Ab-initio calculations for the substitution of Al(OH)4-and SO4-2 with Si tetrahedra, J. Nuc. Sci. Tech-nol 38 (7), 533-541 (2001).
[7] R. J.-M. Pellenq, N. Lequeux and H. van Damme, Engineering the bonding scheme in C–S–H: The iono-covalent framework, 38, 159-174 (2008).
[18] R. J.-M. Pellenq, J. M. Caillol, and A. Delville, Electrostatic attraction between two charged surfaces: A (N, V, T) Monte Carlo simulation, J. Phys. Chem. B, 101 [42] 8584–4 (1997).
[19] A. Delville and R. J.-M. Pellenq, "Electrostatic Attraction and/or Repulsion Between Charged Colloids: A (NVT) Monte-Carlo Study," Mol. Simulation, 24 1–24 (2000).
[20] A. Gmira, M. Zabat, R. J.-M. Pellenq, and H. Van Damme, "Microscopic physical basis of the poromechanical behavior of cement-based materials," Mater. Struct./Concrete Sci. Eng. 37, 3–14 (2004).
[21] A. Kalinitchev, J. Wang, and R. J. Kirkpatrick, Molecular dynamics modeling of the structure, dynamics and energetics of mineral-water interfaces: Applications to Cement Materials, Cem Concr Research 37, 337–47 (2007).
[22] H. Manzano, A. Ayuela, and J. S. Dolado, On the formation of cementitious C-S-H nanoparticles, Journal of Computer-Aided Materials Design, 14 [1] 45-51 (2007).
[23] J. S. Dolado, M. Griebel, and J. Hamaekers, A molecular dynamics study of cementitious Calcium Silicate Hydrates (C-S-H) gels, J. Am. Ceram. Soc. 90, 3938 (2007).
[24] I. G. Richardson, The nature of C-S-H in hardened cements, Cem Con Research, 29, 1131-1147, (1999).
[25] E. Bonaccorsi and S. Merlino, The Crystal Structure of Tobermorite 14 Å (Plombierite), a C-S-H Phase, J Am Ceram Soc, 88 505-512 (2005).
[26] E. Bonaccorsi, S. Merlino, and H. F. W. Taylor, The crystal structure of jennite, Ca9Si6O18(OH)6 • 8H2O, Cem Concr Research, 34 1481-1488 (2004).
[27] GAMESS – General Atomic and Molecular Electronic Structure System. Schmidt, M.W, Baldridge, K.K., Boatz, J.A., Elbert, S.T., Gordon, M.S., Jensen, J.J., Koseki, S., Matsunaga, N., Nguyen, K.A., Su.,Windus, T.L., Dupuis, M.,Montgomery, J.A.: J. Comput. Chem. 14, 1347–1363 (1993)
[28] Gaussian 03, Revision B.04, Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery, J.A., Vreven, Jr., T.,Kudin, K.N., Burant, J.C.,Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H.M.X., Hratchian, E.P., Cross, J.B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev,O.,Austin,A. J., Cammi, R., Pomelli, C.,Ochterski, J.W.,Ayala, P.Y.,Morokuma, K., Voth,G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen,W.,Wong, M.W., Gonzalez, C., Pople, J.A., Gaussian, Inc., Pittsburgh PA (2003).
[29] M. Griebel, S. Knapek and G. Zumbusch. Numerical Simu-lation

"New cyclodextrin-based materials with specific properties could be designed"

in Molecular Dynamics. Springer, Berlin, Heidelberg, 2007.
[30] M. Griebel and J. Hamaekers,. 409-454 in Handbook of Theoretical and Computational Nanotechnology, vol. 9, chap. 8. Edited by M. Rieth and W. Schommers. American Scientific Publishers, 2006.
[31] After being favourable evaluated, the CODICE proposal is, at present, under negotiation with the Commission services. Leaded by LABEIN, CODICE is composed of seven groups more: The Delft University of Technology, The University of Bonn, The Eduardo Torroja Institute for Construction Sciences, the Advanced Concrete and Masonry Centre, Italcementi CTG, Bikain and BASF.

Computational studies on peptides and aromatic compo

Bernardo Herradón,* Mercedes Alonso, Carlos Miranda

Instituto de Química Orgánica General, CSIC, Juan de la Cierva 3, 28006 Madrid, Spain

** Fax: +34915644853; E-mail: herradon@iqog.csic.es*

Project goals and objective

The main objective of the project is to understand the structure and properties of peptide derivatives and aromatic compounds (arenes)

Project description

Our group investigates the synthesis, structure, biological activity and properties of peptides, arene, and derivatives. Our work combines experimental and computational approaches. The computational research is done with two purposes: to help to understand experimental findings and to predict future results. Some computational aspects we investigate are:
1) Toxicology.
2) Conformational analysis of peptide-scaffold hybrids and intermolecular interactions.
3) Structure and biological activity of inhibitors of proteases.

The relevance of this research topic arises from the fact that, separately, both peptides and arenes constitute some of the most important kinds of compounds in organic chemistry. The combination of both structural features in the same molecules (peptide-arene hybrids, a kind of peptide-scaffold hybrids: compounds having peptides chains linked to a non-peptidic scaffold) have provide compounds with interesting properties.

Methods

The computational methods can be divided in three classes.
1) Methods based on molecular mechanics. They include conformational searches using a variety of strategies (random search, systematic search, simulated annealing) and molecular dynamics simulations. Thee results are useful in the field of conformational analysis and provide reliable starting structures for ab initio calculations.
2) Methods based on quantum chemistry. They include ab initio calculations at both Hartree-Fock (HF) and density functional theory (DFT) levels. These results provide geometrical data and their corresponding accurate energies. Additionally, we can compute a variety of properties, including electronic, electrostatic, magnetic, and chemical. Overall, these properties can help to rationalize biological activity and technological properties.
3) Statistical analysis of data. We used neural networks, especially self-organizing maps (SOMs). This methodology allows us to manage a huge amount of data and represent them in a two-dimensional graphic, which permits a straightforward identification of patterns.

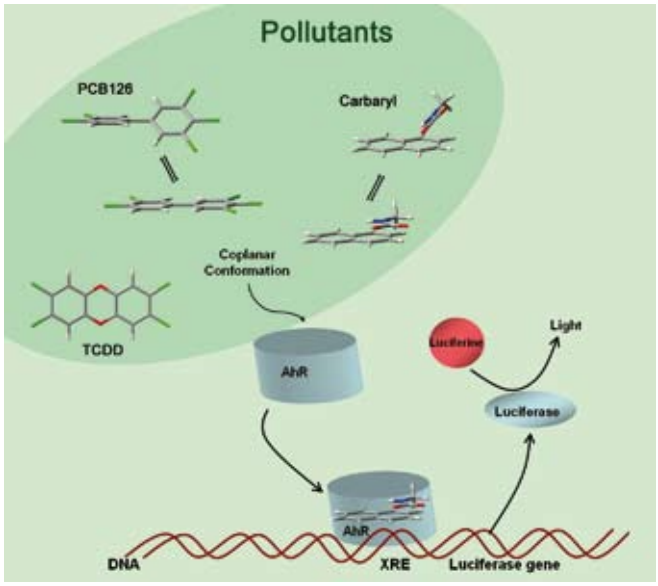
Computing Resources Used

The calculations at CESGA have been carried out in the COMPAQ HPC 320 supercomputer using GAUSSIAN 03 suite of programs.

Results

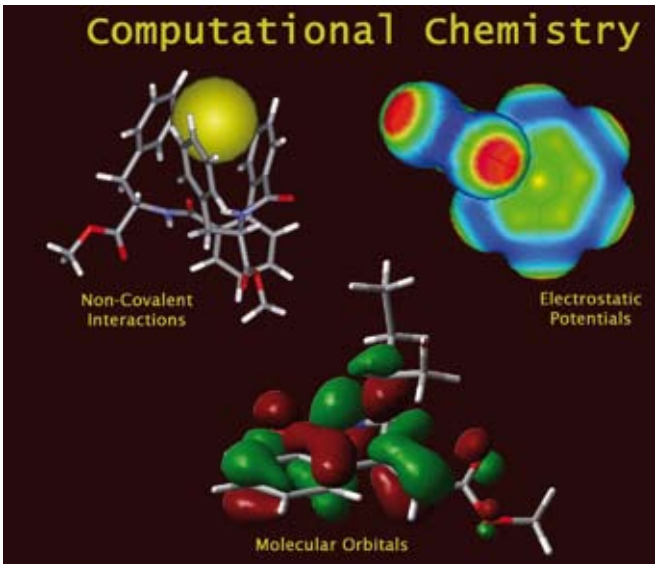
A summary of results in the three topics is indicated below.

1) Computational toxicology.
We study organic compounds with toxic effects, especially on activators of aryl hydrocarbon receptor (AhR), estrogen receptor (ER), and P450 cytochromes. The studied compounds are polycyclic arenes, polychlorinated biphenyls (PCBs), polychlorinated dibenzo-p-dioxins (PCDDs), polychlorinated dibenzofurans (PCDFs), flavonoids, aromatic heterocycles, imidazoles, and related compounds.

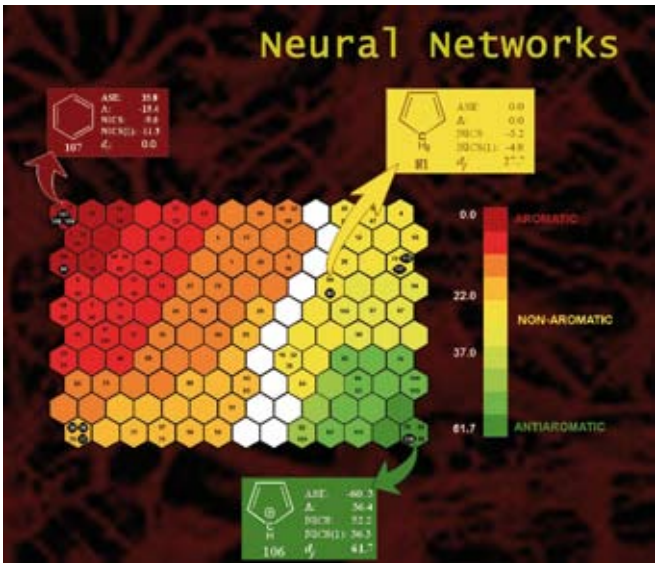


The objective of this research is to deepen on the factors responsible of the toxicity of organic compounds; which, in turn, can serve as remediation of their toxicity and to advance in the design of compounds with pharmacological activity.

2) Conformational analysis of peptide-scaffold hybrids and intermolecular interactions.
Research in this area deals with static and dynamical aspects of aromatic compounds, carbocycles, heterocycles, and peptides. To achieve this objective, we use complementarily computational and experimental approaches. Some issues we are investigating computationally are:
- Influence of the interactions of aromatic rings and hydrogen bonds on the crystal packing and other supramolecular structures.
- Computational studies on aromatic compounds, including electrostatic and electronic properties and aromaticity; as well as their relationships with biological properties (toxicological and protease inhibition).



- Conformational analysis of peptides, carbocycles, and carbohydrates.
- Application of neural networks in organic chemistry, with a special attention to the classification of aromatic compounds and quantification of aromaticity.



3) Inhibitors of proteases: Calpain.
Proteases or peptidases catalyze the hydrolysis of peptide bonds. They are classified in five classes based on their mechanism for catalysis: serine, threonine, cysteine, aspartate and metallo proteasas. The proteolytic reactions are essential in diverse biological processes and, accordingly, they require a strict regulation. Calpains are a family of Ca2+-activated cysteine protease with an active metabolic role. These enzymes catalyze the hydrolysis of a wide variety of proteins involved in signal transduction, apoptosis, cell cycle regulation, cell

"The relevance of this research topic arises from the fact that, separately, both peptides and arenes constitute some of the most important kinds of compounds in organic chemistry"

motility, and so on. Overactivation of calpain is implicated in several degenerative diseases.
Our contribution to the area of calpain inhibition can be summarized by the fact that we have synthesized and evaluated over 500 inhibitors, some of them having IC50 in the nanomolar and picomolar ranges. The structure of the inhibitors is diverse: biaryls, heterocycles, and peptide-scaffold hybrids having aromatic fragments. The experimental results are complemented by computational modelling which provide data on the structural features responsible of the biological activity and help in the design of future generations of inhibitors.

Future work

The future work will continue the current research, and some aspects we are going to be engaged include:
1) To assess the role of intermolecular interaction in the supramolecular structure of arenes, peptides and peptide-scaffold hybrids.
2) To establish structure-activity relationships for calpain inhibition
3) To learn on the mechanism of activation of the aryl hydrocarbon receptor.
4) To model dynamical properties of peptide-scaffold hybrids, including the crystallization process.

Acknowledgements

Financial support from the Spanish Ministry of Education through the grants CTQ2004-01978 and CTQ2007-64891/BQU is acknowledged. We thank Dr. J. M. Navas (INIA) for fruitful collaborative research along the years on activators of AhR. We also thank the Centro de Supercomputación de Galicia (CESGA) for computational time.

Scientific publications derived from this work

Articles in peer-reviewed journals:
1) M. Alonso, S. Casado, C. Miranda, J. V. Tarazona, J. M. Navas, B. Herradón. Decabromobiphenyl (PBB-209) activates the aryl hydrocarbon receptor while decachlorobiphenyl (PCB-209) is inactive. Experimental evidences and computational rationalization of the different behavior of some halogenated biphenyls. Chem. Res. Toxicol. 2008, in press.
2) M. Alonso, B. Herradón. Neural Networks as a Tool to Classify Compounds According to Aromaticity Criteria. Che. Eur. J., 2007, 13, 3913-3923.
3) S. Casado, M. Alonso, B. Herradón, J. V Tarazona, J. M. Navas. Activation of the aryl hydrocarbon receptor by carbaryl: computational evidence of the ability of carbaryl to assume a planar conformation. Environm. Toxicol. Chem. 2006, 25, 3141-3147.

Communications and lectures in congress:
1) M. Alonso, B. Herradón. Towards a quantitative scale of aromaticity. XXXI Reunión Bienal de la Real Sociedad Española de Química, Toledo (Spain), September 9-14, 2007. Invited lecture.
2) M. Alonso, C. Miranda, B. Herradón. Comprendiendo el comportamiento molecular de los híbridos péptido-biarilo, un nuevo tipo de inhibidor no-electrófilo de calpaina. XXXI Reunión Bienal de la Real Sociedad Española de Química, Toledo (Spain), September 9-14, 2007. Communication.
3) B. Herradón, M. Alonso, R. Chicharro, M. T. Mazo, C. Miranda, A. Montero, V. J. Arán. Híbridos péptido-molde: Síntesis, estructura y propiedades. X Encuentro Peptídico Ibérico (EPI). Zaragoza, February 1-4, 2006, Oral communication.
4) R. Chicharro, M. Alonso, C. Miranda, M. T. Mazo, V. J. Arán, B. Herradón. Peptide Heterocyclic Hybrids: Synthesis, Structure and Properties. 4th Spanish~Portuguese~Japanes Organic Chemistry Symposium (4SPJ-OCS). Santiago de Compostela (Spain), September 8-11, 2006. Communication.
5) M. Alonso, C. Miranda, A. Montero, B. Herradón. Experimental and Computacional Studies on Aromatic Compounds: Synthesis, Structure and Properties. 4th Spanish~Portuguese~Japanes Organic Chemistry Symposium (4SPJ-OCS). Santiago de Compostela (Spain), September 8-11, 2006. Invited lecture.
6) A. Jos, H. Segner, B. Herradón, G. Repetto, J. M. Navas. Induction of EROD activity by 1-phenylimidazole and β-naphthoflavone in rainbow trout cultured hepatocytes: A comparative study. INVITOX 2006. 14th International Workshop on in vitro toxicology. Ostende (Belgium), Octobe 2-5, 2006. Communication

Master´s thesis:
1) M. Alonso. Estudios computacionales de compuestos aromáticos. Universidad Complutense, Madrid (Spain), 2005.

To be or not to be… in water

Gonzalo Jiménez-Osés, Francisco Corzana, Jesús H. Busto, Marisa García de Luis, Jesús M. Peregrina and Alberto Avenoza

Departamento de Química, Universidad de La Rioja, UA-CSIC, Madre de Dios 51, E-26006 Logroño, Spain.
gonzalo.jimenez@unirioja.es

Project description.

The protein glycoslylation process is one of the principal co-translational and post-translational modification steps in the synthesis of membrane and secreted proteins, and the majority of proteins synthesized in the rough endoplasmatic reticulum undergo glycosylation. The resulting molecules of this biological reaction, i.e. glycoproteins, are involved in fundamental biological processes such as inflammation, immune response, cell-cell communication, cell growth, cell adhesion, and antifreeze activity, among others. In this context, mucins —a particular class of glycoproteins— have attracted a great interest in therapeutic approaches, especially for the development of vaccines for cancer treatment. On the other hand, it has been described that in Nature, glycosylation has a profound organizational effect on the structure of the underlying peptide backbone, forcing it to adopt unusual conformations. Logically, to discern how glycoproteins interact with their biological targets it is essential to improve our understanding of the mechanisms that allow the carbohydrate to modify the conformational equilibrium of the peptide backbone.

Therefore, a large number of attempts have been made to elucidate the active forms of these systems by using a wide range of experimental and theoretical methodologies. A common mistake in these approaches is the dedication of a huge experimental and computational effort to analyze the conformational space of the isolated biomolecules, neglecting other important factors that could play a key role to define the true structure. Undoubtedly, one of these factors is the solvent, and particularly water, to simulate the physiological conditions. However, how does water interact with glycoproteins? How far does it do? These are some of the main questions that we have been trying to elucidate in the last years.

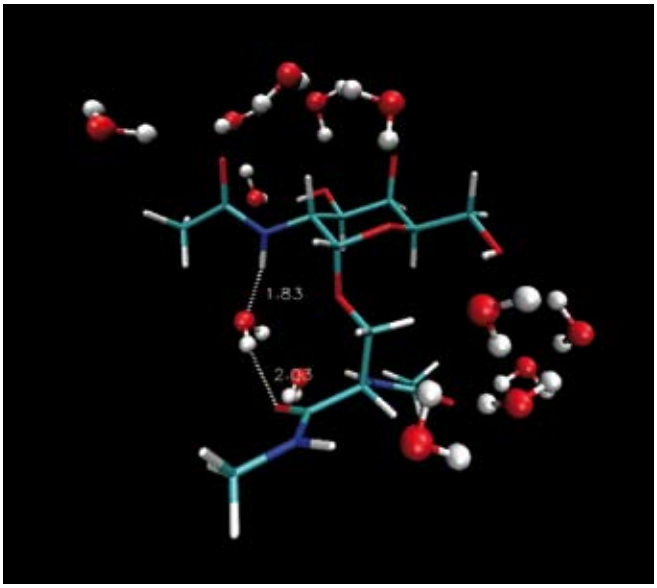
Methods.

In the course of our borderline investigations we have been required to apply a number of multidisciplinary techniques such as fine organic synthesis, multidimensional NMR, X-ray diffraction, molecular dynamics (MD) simulations, and quantum mechanics (QM) calculations. A plethora of new organic compounds have been prepared and characterized to build useful models of small glycopeptides, being our main targets the glycosylated amino acids serine and threonine (Ser/Thr). The core of our research is, therefore, to analyze the changes that take place in the structure of these compounds in aqueous solution when carbohydrates like α-D-N-acetylgalactosamine (α-D-GalNAc) and β-D-glucose (β-D-Glc) are attached to their hydroxyl groups.

Computational methods and resources. Once the target compounds are synthesized and purified, NOE and multidimensional NMR experiments provide a large amount of structural information that has to be adequately interpreted with the assistance of diverse computational tools such as MD and QM calculations. In many cases, experimental data can be reproduced only if water is explicitly introduced in the simulations, so the computational cost of the calculations grows exponentially. As an example, a typical model for the MD simulations contains more than 3000 atoms including explicit solvation, being the standard runtime at the nanosecond scale. On the other hand, model glycopeptides together with their first hydration shell model include over 100 atoms for routine QM calculations. The extremely high flexibility of these systems adds an unavoidable extra complication to the conformational analysis and very often leads to an overflow on the available computational resources.

Results.

In contrast to common belief, the structural study of the mucin-type simplest model glycopeptide, α-*O*-GalNAc-L-Ser diamide, reveals that intramolecular hydrogen bonds involving sugar and peptide residues are very weak, and consequently, they cannot determine the defined geometry of this type of molecule. Our experimental and theoretical results point toward the existence of water pockets/bridges between the sugar and the peptide moieties and indicate that the surrounding water molecules are essential to keep the defined conformation (Figure 1). Interestingly, a different behavior has been observed for the glycosidic linkages of D-GalNAc-Ser and D-GalNAc-Thr motifs, allowing the carbohydrate moiety to adopt a completely different orientation. In addition, the water pockets found in α-D-GalNAc-Thr differ from those obtained for its serine analogue. This fact could be related to the different capability that the two model glycopeptides have to structure the surrounding water, which could explain the loss of activity of antifreeze glycopeptides when the Thr is replaced by the Ser.



"... results point toward the existence of water pockets/bridges between the sugar and the peptide moieties..."

Future work.

Once we have analyzed the conformational preferences of small glycopeptides and the influence of the surrounding water on them, we plan to extend these studies to more complex models in order to get closer to those biosystems present in Nature. In a parallel way, we are interested in testing the influence that slight structural modifications made on the amino acid moieties could have in the global structure of synthetic unnatural glycopeptides. The in-deep study of the conformational preferences in water of these compounds will require a great experimental effort and highly demanding computational resources to accurately carry out MD and QM simulations. Therefore, high-end computing would be a very valuable tool to explore new frontiers in our investigations.

Scientific publications derived from this work:
1. Corzana, F.; Busto, J. H.; Engelsen, S. B.; Jiménez-Barbero, J.; Asensio, J. L.; Peregrina, J. M.; Avenoza, A. Chem. Eur. J. 2006, 12, 7864-7871.
2. Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. J. Am. Chem. Soc. 2006, 128, 14640–14648.
3. Corzana, F.; Busto, J. H.; Jiménez-Osés, G.; García de Luis, M.; Asensio, J. L.; Jiménez-Barbero, J.; Peregrina, J. M.; Avenoza, A. J. Am. Chem. Soc. 2007, 129, 9458–9467.

Collaborating institutions and researchers: Jesús Jiménez-Barbero (Centro de Investigaciones Biológicas, CSIC, Spain), Juan L. Asensio (Instituto de Química Orgánica General, CSIC, Spain), SØren B. Engelsen (The Royal Veterinary and Agricultural University, Denmark).
Acknowledgment. We thank the Ministerio de Educación y Ciencia and FEDER (projects CTQ2005-06235/BQU, CTQ2006-05825 and Ramón y Cajal contracts of F.C. and J.H.B.), the Universidad de La Rioja (project API-05/B01 and grants of G.J.-O. and M.G.L.), and the Gobierno de La Rioja (ANGI-2004/03 and ANGI-2005/01 projects). We also gratefully thank CESGA for computer support.

Environment and Roles of the Active Site of Heme Perox

D. R. Ramos, M. V. García and J. A. Santaballa
Chemical Reactivity & Photoreactivity Group. Dept. of Physical Chemistry & Chemical Eng. I, University of A Coruña, Alejandro de la Sota 1, E-15008 A Coruña, Spain.

Project Description

Peroxidases are heme-containing enzymes that reduce H2O2 to water and oxidize a wide range of substrates. The neutrophil protein myeloperoxidase has been implicated in promoting tissue damage in numerous inflammatory diseases, and plays a central role in the antimicrobial activity of the human innate immune system. However, their oxidative mechanisms are not fully understood. We have recently initiated this project to extensively investigate the processes undergone by these enzymes, and namely human myeloperoxidase. Among other techniques, electronic structure calculations have revealed essential to obtain more precise molecular information to establish reaction mechanisms.
Empirical studies (structural and geometrical elucidation, kinetics, identification of intermediates, etc) on peroxidase catalysis yield very important but incomplete information about the processes involving their reactivity. Combination of experimental and computational tools are systematically applied in order to support the existence of possible intermediates or transition structures and the nature of some elusive reaction products, reveal the effect of the precise peptide structure –particularly the interaction of the protein with the heme and the environment of the active site– on the specific oxidative reactions taking place, and finally determine their detailed reaction mechanisms.
The proper elucidation of the concerned mechanisms is of great importance to understand the multiple processes and diseases related to peroxidase activity, opening the door to additional relevant research either in biochemistry or in the more applied medical and pharmaceutical fields.

Computational Details

All peroxidases share the same prosthetic group (ferriprotoporphyrin IX) and the same proximal ligand (the imidazole side chain of a His residue). Therefore, their active site was modelled by using only porphine and imidazole adjacent to the core iron atom (Fig. 1).
Density functional calculations were performed with the Gaussian 98 suite of programs, while the latest 03 version was used to compute PCM solvation. Structures were optimized using the DFT/UB3LYP/6-31G* computational level. Other basis sets successfully employed to similar systems, such as LANL2DZ associated to an effective core potential for Fe and heavier triple-ζ 6-311G, were applied to previously optimized geometries and their suitability was also tested. Different spin multiplicities were considered in all calculations carried out.
The effect of the protein on heme stabilization was roughly modelled simulating bulk water with a polarisable continuum model (PCM). Despite the fact that this method reproduces neither the bulk protein interaction on the active site nor the direct covalent bonding between peptide and heme, it adequately accounts for the effect of the water molecules, placed nearby in the distal cavity, on the highly reactive Fe center where the study is focused.

"...electronic structure calculations have revealed essential to obtain more precise molecular information to establish reaction mechanisms"

All calculations were performed in a cluster HP Integrity Superdome. Resources consumed by every single calculation were as follows: 7 Gb of memory, 50 Gb of disk space, by using two parallel 1.5 GHz Itanium 2 processors.
It should be taken into account that the particular reaction taking place depends very much on the specific nature of enzyme and substrate when dealing with peroxidase reactivity. The origin of these differences could be related to subtle changes in the protein structure; mostly, the direct interaction between peptide and heme through some covalent linkages.
The use of all-electron computational levels seems essential to correctly reproduce these effects, then requiring a noticeable amount of resources, which increase exponentially with further considered atoms. Thus, the use of improved, closer to reality, models demands resources not yet available at CESGA supercomputing center.

Results

Some important conclusions have already emerged, like the detailed molecular structures of several known intermediates, the assignation of compound-I observed pKa to the oxoferryl protonation –whereas in other cases it must be ascribed to distal His– or the presence of proposed complex compound-I-Cl. Preliminary data also confirm that the latter represents a critical intermediate for the chlorination process; its fragmentation yields the oxidizing agent HOCl.
Observed effect of [Cl-] on chlorination rate was explained with calculated thermodynamic and kinetic values for the chlorination reaction of Fe(III)-native and compounds I and II. Computational data support the existence of the postulated complex with hydroperoxyl radical compound-II-HO2; besides, they showed that superoxide radical can be released upon deprotonation, which can be attributed to the compound-II conversion observed at pH ≈ 9 by resonance Raman spectroscopy.
Current research focus on testing the validity of a direct chlorination mechanism, via an oxidizing enzymatic complex, and studying the effect of molecular structure on the redox potential of different peroxidases, or even different complexes of a particular enzyme.

Acknowledgements

Dr. Paul G. Furtmüller and Prof. Dr. Christian Obinger from the Division of Biochemistry at the BOKU-University of Natural Resources and Applied Life Sciences in Vienna (Austria) have been essential collaborators for performing the experimental studies, as well as quite helpful in the evaluation of the reliability of computational results.
This research was funded by the former Spanish MCyT (BQU2002-02948), by the Spanish MEC (CTQ2004-00534/BQU), the Xunta de Galicia (PGIDT03 PX1C10303PN), and the Universidade da Coruña. D.R. Ramos thanks the former Spanish MECyD for an FPU grant and the Fundación Segundo Gil Dávila for a research grant. We are indebted to CESGA for computer resources and technical support.

Scientific Publications derived from this work:

Pure experimental and joint empirical-computational results have been published to date; these data are available in the following papers:
• "Myeloperoxidase-catalyzed taurine chlorination: initial versus equilibrium rate"
D.R. Ramos, M.V. García, M. Canle L., J.A. Santaballa, P.G. Furtmüller, C. Obinger; Arch. Biochem. Biophys. 466 (2007) 221-233
• "Myeloperoxidase-catalyzed chlorination: the quest for the active species"
D.R. Ramos, M.V. García, M. Canle L., J.A. Santaballa, P.G. Furtmüller, C. Obinger; J. Inorg. Biochem. (2008) In press
Already published results represent the basis for the subsequent manuscripts, currently in preparation, dealing with the computational approach. Those results to be published have been advanced as communications at many international conferences.

Figure 1. Model of the active site of heme peroxidases. Fe atom is surrounded by porphine and the imidazole of proximal His.

Highly efficient simulations of semiconductor devices

N. Seoane, A. J. GarcíaLoureiro and M. Aldegunde

Dept. of Electronics and Computer Science, University of Santiago de Compostela

Edificio Monte da Condesa, Campus Sur, 15782 Santiago de Compostela (Spain)

Introduction

Numerical simulation tools are essential in semiconductor device research, allowing the study, explanation or even prediction of some phenomena arising in semiconductor devices due to the shrinking of characteristic dimensions. Their impact on the industry is also very important due to the reduction in the trial and error iterations required to obtain the final device design.
The accuracy of the simulation results is proportional to the complexity of the simulation technique and therefore to its computational cost. Two of the most popular simulation approaches describing the behaviour of semiconductor devices are the drift-diffusion (D-D) method and the Monte Carlo (MC) technique to solve the full Boltzmann transport equation.
The main objectives of our research group are related to the 3D parallel simulation of several kinds of semiconductor devices currently under development and research, such as HEMTs[1], IF heterostructure MOSFETs[2], Si MOSFETs[3] (bulk, double gate, TriGate, nanowire ...) and HBTs among others. One of our main research topics is the study of the impact of several sources of intrinsic parameter fluctuations on the performance and reliability of the devices [1, 2]. To accomplish these objectives we use different simulation methodologies including D-D and MC simulations based on 3D finite element solvers [4, 5].

Simulation methodology

Regardless of the chosen simulation technique, the problem can be reduced to the resolution of a set of equations that describe the physical process. These equations are discretised and solved numerically throughout an iterative process. Initially, the Poisson equation is solved to obtain the electrostatic potential. After that, this value is used to obtain the concentrations of charge carriers, either solving the carrier continuity equations in the D-D method or the Boltzmann transport equation (BTE) with a MC method. The Poisson equation and the carrier continuity equations are discretised using a finite element method on unstructured tetrahedral meshes. In the D-D method, this results in coupled non-linear systems of equations, which are first decoupled, using a Gummel iterative method, and then linearised, using the Newton-Raphson procedure. So, basically, the computational kernel of the simulation process is the solution of linear systems of equations. In three-dimensional self-consistent MC simulations, the solution of the linear systems associated to the Poisson equation are also the most demanding part of the whole process.
Depending on the size of the problem and the device characteristics, a MC simulation may take several days. On the other hand, to study the impact of different sources of intrinsic parameter fluctuations in the device behaviour it is necessary to carry out statistical studies. Therefore, the cost of the 3D simulation increase by the size of the statistical sample. These problems call for the use of high performance computing. Hence, both the 3D D-D and the MC simulators are fully parallelised using the MPI standard library.
To solve in parallel the linear systems, domain decomposition methods have been employed, assigning a fraction of the system to each processor. The local linear systems in each subdomain are solved using iterative methods, combining preconditioners, such as incomplete LU factorisations, with Krylov solvers, such as the BiCGSTAB or the FGMRES methods.

Results

Numerical results presented in this work have been obtained in an HP Superdome Cluster [6] formed by two HP Integrity Superdome servers, each with 64 Itanium2 1.5 GHz, 6 MB cache processors.
Using the 3D D-D simulator, to analyse the impact of several sources of intrinsic parameter fluctuations in the behaviour of a semiconductor device a proper calibration of the simulator is required. Therefore, simulated I_D-V_G characteristics are compared to the measured data from real devices as well as to simulation results obtained from a Monte Carlo device simulator H2F/MC. For instance, Figure 1 compares the I_D-V_G characteristics at a drain bias of 0.1 and 1.0 V, for a 120 nm gate length PHEMT, and Figure 2 shows an example of the 3D electron concentration in equilibrium for the same device.
The tetrahedral mesh of a HEMT device divided in three subdomains is shown in Figure 3.

Using a 10 nm gate length DG thin body Si MOSFET the impact of random dopants on the I_D-V_G characteristics has been studied. Each random dopant distribution produces a unique potential profile, an example of which is shown in Figure 4.

Using the superdome cluster, the parallel performance of the 3D MC device code has been analysed. This study has been carried out by simulating a DG MOSFET over a small period of 20 fs. Figure 5 represents the parallel efficiency (circles) and speed-up (crosses) of the MC simulator. The ideal values of linear scaling are also shown for comparison (dashed lines). Results show that the obtained values are over the linear scaling limit. This effect is due to the efficiency of the Poisson solver and it starts to vanish for a high number of processors.

" ... the computational kernel of the simulation process is the solution of linear systems of equations"

Figure 1

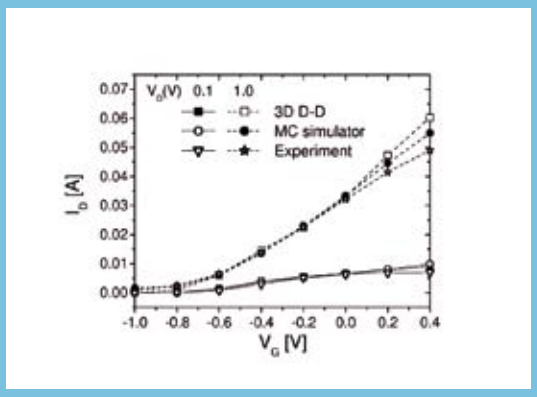


Figure 2

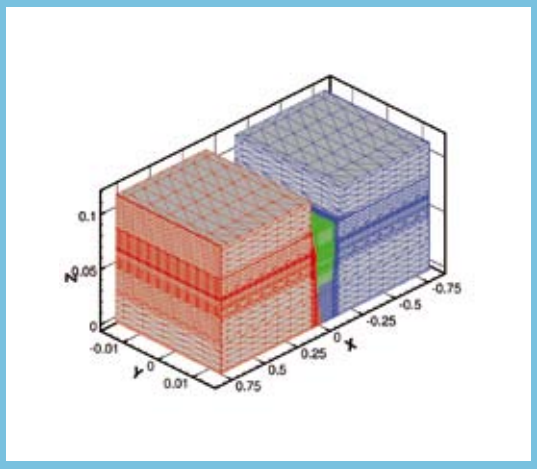


Figure 3

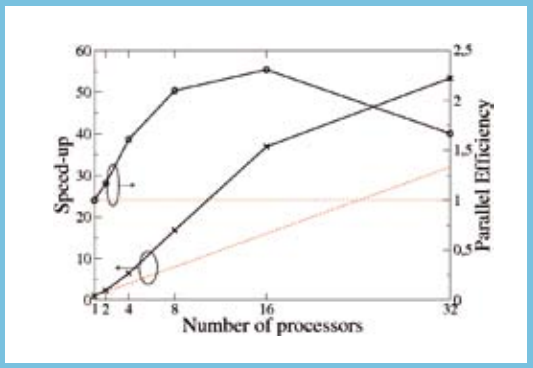


Figure 4

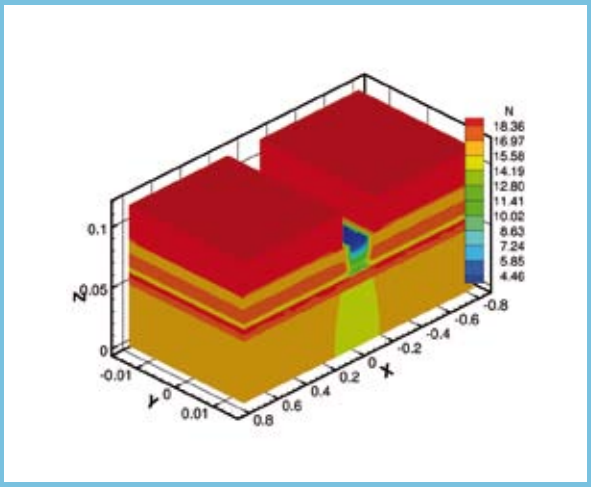


Figure 5

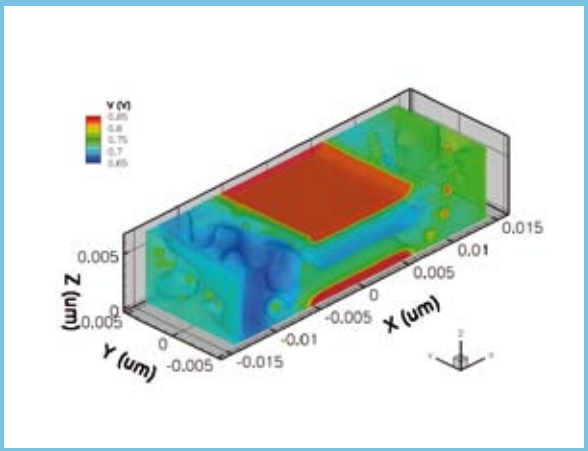


Fig. 1: I_D-V_G characteristics at a low drain bias of 0.1 V and a high drains bias of 1.0 V for the 120 mm PHEMT with $In_{0.2}Ga_{0.8}As$ channel

Fig. 2: Electron concentration on a logarithmic scale for the 120 mm gate length PHEMT

Fig. 3: Example of a tetrahedral mesh of a HEMT device divided in three subdomains

Fig. 4: Electrostatic potential for a drain bias of 0.05 V and a gate bias of 1.1 V for one configuration of the dopants located at atomic positions

Fig. 5: Parallel efficiency and speed-up of the Finite Element MC simulator

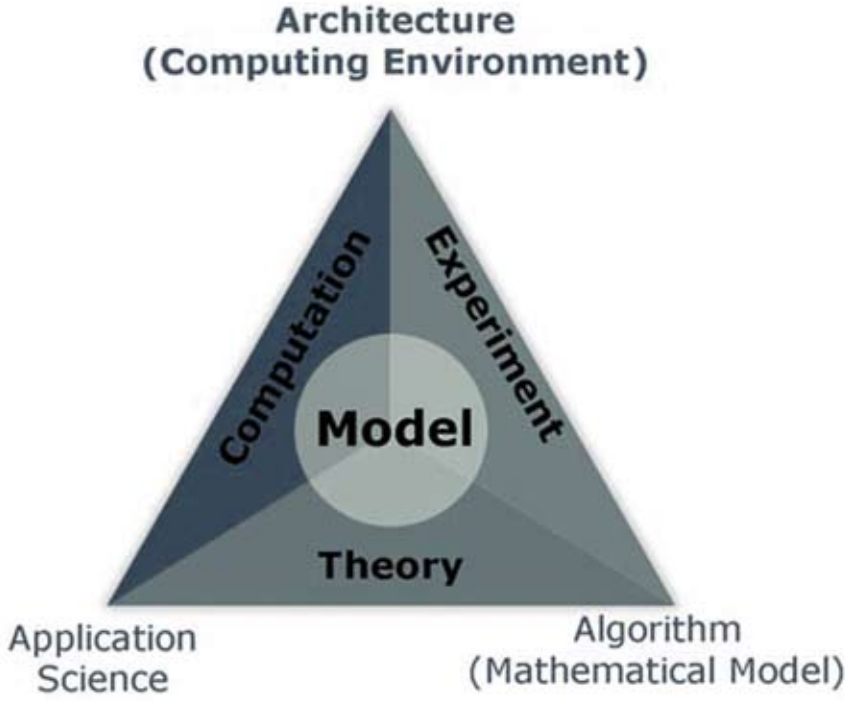
This work was supported by the Spanish Government (CICYT) under the project TIN2004-07797-C02.
Authors would like to thank CESGA (Galician Supercomputing Center) for providing access to the HP Superdome system.

References

[1] N. Seoane, A. J. García-Loureiro, K. Kalna, A. Asenov. Impact of intrinsic parameter fluctuations on the performance of HEMTs studied with a 3D parallel drift-diffusion simulator. Solid–State Electron., 51 (2007), pp. 481-488.
[2] N. Seoane, A. J. García-Loureiro, K. Kalna, A. Asenov. Threshold voltage variability in nanoscaled $In_{0.75}Ga_{0.25}As$ implant free MOSFET. Proc. 12th Int. Workshop on Computational Electronics, (2007).
[3] M. Aldegunde, A. J. García-Loureiro, K. Kalna. 3D parallel finite element Monte Carlo simulations of nano-MOSFETs. Proc. 6th Int. Conference on Large-Scale Scientific Computations, (2007).
[4] M. Aldegunde, J. J. Pombo, A. J. García-Loureiro. Modified octree mesh generation for Manhattan type structures with narrow layers applied to semiconductor devices. Int. J. Numer. Model.-Electron. Netw. Device Fields, 16 (2006), pp. 473-489.
[5] N. Seoane, A. J. García-Loureiro. Study of parallel numerical methods for semiconductor device simulation. Int. J. Numer. Model.-Electron. Netw. Device Fields, 19 (2006), pp. 15-32.
[6] Galician Supercomputing Center, http://www.cesga.es

Acknowledgements

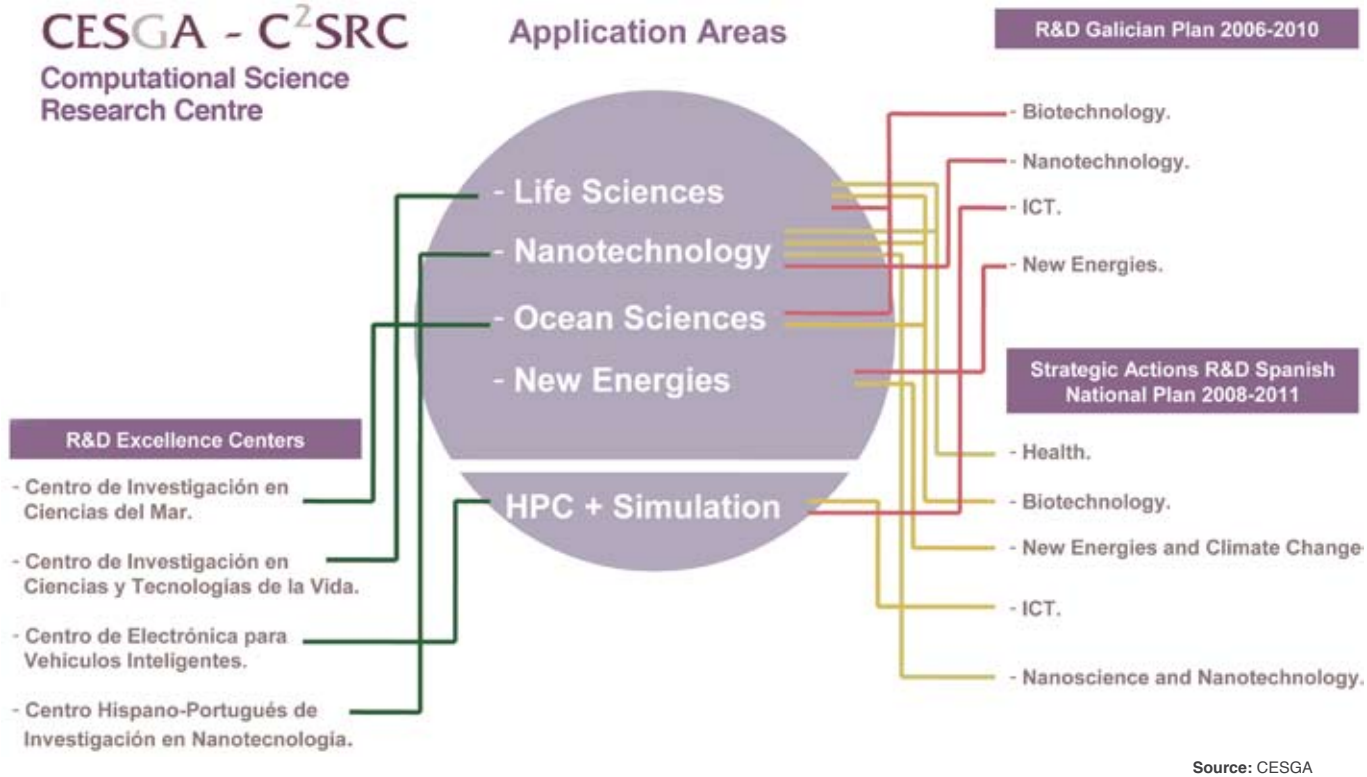
COMPUTATIONAL SCIENCE PARADIGM



Source: Dept. of Computational Science
National University of Singapore

CESGA - C²SRC

Computational Science
Research Centre



Source: CESGA