

Supercomputació al CSUC

D. Tur¹, C. Gomollon², A. Gil³

¹ David Tur (Eivissa, 1977) és doctor en Nanoestructures i Nanotecnologies per la Universitat degli Studi di Milano-Bicocca (2010).

² Cristian Gomollon, llicència en física per la UAB i tècnic d'aplicacions científiques al CSUC.

³ Alfred Gil (Barcelona, 1971) és doctor en Química Teòrica i Computacional per la Universitat Rovira i Virgili (2007).

Resum

Aquest article descriu el servei de supercomputació que proporciona el Consorci de Serveis Universitaris de Catalunya (CSUC) a la comunitat acadèmica i de recerca. Repassa breument les activitats del Consorci, els projectes i usuaris que utilitzen aquest servei i el tipus de simulacions científiques que s'hi porten a terme. A més, es fa un repàs dels projectes de física i dels nous reptes en supercomputació al Consorci.

Doi: <http://dx.doi.org/10.2436/20.2001.01.1>

1 Introducció

El Consorci de Serveis Universitaris de Catalunya (CSUC) té com a objectiu compartir o mancomunat serveis acadèmics, científics i de gestió de les seves entitats consorciades, per aconseguir millorar-ne l'eficàcia i l'eficiència potenciant les sinergies i les economies d'escala. Entre l'ampli ventall de serveis que ofereix destaca el servei de supercomputació, que permet, entre d'altres, fer prediccions climatològiques precises i fiables, fabricar nous productes amb materials més resistents i flexibles, conèixer l'origen de l'univers i de la vida, o predir l'impacte de l'ésser humà en el medi ambient.

El CSUC disposa de supercomputadors amb diferents arquitectures i de nombrosos programari especialitzat per satisfer les necessitats ben diverses de càlcul científic dels usuaris. També els dóna suport tècnic i científic, de manera que constitueix un servei integral de supercomputació que a més d'oferir la infraestructura també dóna assistència tant tècnica com científica.

2 CSUC, nou consorci de serveis universitaris

El CSUC va néixer el gener de 2014 com a conjunció de les trajectòries del Centre de Serveis Científics i Acadèmics de Catalunya (CESCA), antic Centre de Supercomputació de Catalunya, que al llarg de 22 anys havia prestat serveis relacionats amb les infraestructures TIC, i del Consorci de Biblioteques Universitàries de Catalunya (CBUC), que ho havia fet al llarg de 18 anys en l'àmbit de les biblioteques.

Aquest Consorci està format per deu universitats (UB, UAB, UPC, UPF, UdL, UdG, URV, URL i UVic-UCC) i la Generalitat de Catalunya. Té com a missió detectar, dissenyar i executar de manera col·laborativa projectes compartits que permetin a les universitats catalanes augmentar la seva eficiència. Perquè les universitats catalanes poden ser molt

més eficients a base de compartir serveis i infraestructures sense que això els hagi de ser cap obstacle per preservar la seva personalitat, mantenir la seva autonomia, perseguir l'excel·lència i continuar competint entre si.

2.1 Breu història

El 10 de desembre de 2013 es va aprovar la creació del Consorci de Serveis Universitaris de Catalunya (CSUC) per impulsar l'eficiència en la gestió de les universitats catalanes a través de la cooperació i la coordinació amb l'objectiu fonamental de compartir o mancomunat serveis acadèmics, científics, bibliotecaris, de transferència de coneixement i de gestió de les entitats consorciades per potenciar sinergies.

Aquest Consorci, creat a iniciativa de la Generalitat de Catalunya, fusionava dos consorcis que, fins aleshores, havien treballat promovent i oferint serveis consorciats en els àmbits TIC (CESCA) i bibliotecari (CBUC), alhora que incorporava noves línies d'activitat, com ara les compres conjuntes, i qualsevol àmbit en què la col·laboració i la compartició poguessin beneficiar el sistema universitari i de recerca.

A principis dels noranta, la Generalitat de Catalunya havia impulsat la creació del CESCA, aleshores sota la denominació de Centre de Supercomputació de Catalunya, a través de la llavors Fundació Catalana per a la Recerca i amb la col·laboració de les universitats i el CSIC. Aquest Centre ha estat reconegut com a instal·lació científica i tècnica singular (ICTS, 2000) i ha rebut la Placa Narcís Monturiol (2001).

El CESCA va ser creat amb l'objectiu de compartir una infraestructura de càlcul d'altres prestacions que la comunitat científicotècnica necessitava per desenvolupar la recerca en diverses àrees de coneixement, com ara la química teòrica, la meteorologia o l'astrofísica. Aquest servei es va començar a prestar amb un IBM 3090 i un Cray X-MP. Més tard, i per al camp de les ciències de la vida, es va crear un servei de cerca de farmacòfors (actual Servei de Disseny de Fàrmacs).

Els seus serveis es van estendre a les comunicacions, amb la creació de l'Anella Científica l'any 1993, una xarxa de transmissió de dades d'alta velocitat que va ser pionera a Europa. L'any 1999 es va crear el Punt Neutre d'Internet a Catalunya (CATNIX) per afavorir l'intercanvi de tràfic entre els diferents proveïdors d'Internet.

L'any 2001 es va estrènyer la col·laboració amb el CBUC perquè van posar en marxa conjuntament Tesis Doctorals en Xarxa (www.tdx.cat), el primer de diversos repositoris digitals creats per incrementar l'accessibilitat i la visibilitat dels continguts de les institucions que hi participen.

Per emfatitzar el suport i els serveis TIC que s'oferien a les universitats catalanes, l'any 2009 es va impulsar una sèrie de serveis comuns d'administració electrònica: plataforma de vot, de registre telemàtic d'entrada i sortida, d'arxiu i preservació de documents digitals, etc.

D'altra banda, el 1996, després de diversos anys de treball conjunt de les biblioteques universitàries i amb la finalitat inicial de posar en marxa el Catàleg Col·lectiu de les Universitats de Catalunya (CCUC), es va crear el CBUC. Aquest consorci estava integrat per les universitats públiques i per la Biblioteca de Catalunya. Des de la seva creació, però fins i tot en els treballs previs de preparació del catàleg, que es remunten a principis dels noranta, el CBUC havia treballat per millorar els serveis bibliotecaris a través de la cooperació, no només amb les institucions consorciades, sinó també amb nombrosos membres associats i col·laboradors que han participat i s'han beneficiat dels seus serveis.

Immediatament després de la seva creació, es va considerar que era possible i beneficiós organitzar un programa de préstec interbibliotecari. En aquest sentit, i des de l'any 1997, el CBUC havia promogut acords que han permès als usuaris de les biblioteques de les institucions que n'han format part poder accedir a documents que no es troben a les seves biblioteques. Els bons resultats dels programes inicials van animar a iniciar-ne de nous, com ara compres conjuntes d'equipament, formació i *benchmarking*. En particular, es va crear la Biblioteca Digital de Catalunya, un servei que reuneix tant el conjunt d'informació electrònica contractada conjuntament, a partir del 1999, com els dipòsits d'accés obert.

2.2 Els serveis del CSUC

Avui dia el CSUC ofereix un ampli ventall de serveis per a la universitat i la recerca, basant-se en diverses línies d'activitat: els sistemes per a càlcul científic, tant acadèmic com industrial; les xarxes de comunicacions (l'Anella Científica i el CATNIX); els portals i repositoris per a informació universitària (TDX, RECERCAT, RACO, MDX, etc.) i per a l'emmagatzematge de dades; el suport per millorar els serveis bibliotecaris a través de la cooperació (CCUC, PUC, BDC, MDC, etc.); l'administració electrònica (certi-

ficació digital, registre d'entrada i sortida, votació, etc.) i la consorciació de serveis per millorar i racionalitzar l'ús de recursos i enfortir el sistema universitari i de recerca.

Pel que fa al càlcul científic, el CSUC ofereix serveis de supercomputació, tant a investigadors acadèmics com al teixit industrial, amb un ampli ventall de programari científic disponible. Gestiona el Servei de Disseny de Fàrmacs (SDF), que ajuda els laboratoris farmacèutics i grups de recerca a trobar nous medicaments més eficientment.

En comunicacions, el CSUC gestiona l'Anella Científica, xarxa d'altres prestacions que connecta universitats, centres de recerca i altres entitats vinculades a l'R+D+I. Allotja el node de RedIRIS a Catalunya, que connecta l'Anella Científica amb la resta de xarxes acadèmiques i de recerca. També gestiona el Punt Neutre d'Internet a Catalunya (CATNIX), que facilita les comunicacions a Internet gràcies a l'intercanvi local de tràfic.

El Consorci proporciona accés a webs institucionals, universitaris o relacionats amb la societat de la informació, i a repositoris digitals, entre els quals hi ha TDX, que recull tesis doctorals; RECERCAT, amb documents de recerca; RACO, amb revistes científiques, culturals i erudites; PADICAT, amb patrimoni digital català; MDX, amb material docent; MDC, amb col·leccions del patrimoni català, i CALAIX, amb documents i arxius d'interès cultural.

També contribueix a millorar els serveis bibliotecaris a través de la cooperació, tant mitjançant diversos programes i serveis com a través de repositoris i d'altres activitats com ara la formació i el *benchmarking*. Entre d'altres, gestiona el Catàleg Col·lectiu de les Universitats de Catalunya, el programa de préstec interbibliotecari i la Biblioteca Digital de Catalunya.

El CSUC ofereix també serveis d'administració electrònica, nascuts del treball conjunt amb l'ACUP, com ara el registre electrònic d'entrada i sortida (e-registre), el vot electrònic (e-vot) i l'arxiu digital (e-arxiu), la signatura electrònica (e-signatura), la interoperabilitat, a més del Servei de Certificació Digital.

Amb referència a la consorciació de serveis i les compres conjuntes, el CSUC millora i racionalitza l'ús dels recursos i ajuda a enfortir el sistema universitari i de recerca. Homogeneïtza aplicacions, plataformes i tecnologies, i aglutina la contractació de recursos i serveis bàsics, com ara energia elèctrica, gas, impressió, telefonia mòbil, gestors documentals, centres de processament de dades, entre d'altres, per facilitar-ne la gestió i les economies d'escala. Potencia les sinergies per millorar l'eficàcia i eficiència de serveis comuns.

3 Supercomputació al CSUC

Els serveis relacionats amb la supercomputació per a investigadors tant acadèmics com industrials són un eix

important dins el Consorci, que ofereix un servei integral que possibilita un ús compartit dels recursos de càlcul per part del sistema universitari i de recerca de Catalunya.

A més d'un conjunt de supercomputadors d'altres prestacions i programari específic per a la investigació, aquest servei també inclou el suport científic a l'usuari, una característica singular que suposa un tret diferencial, ja que no es proporciona en cap altre centre de càlcul ni empresa privada. Actualment, un equip d'enginyers i doctors donen assistència a una cinquantena de grups de recerca amb investigadors de totes les universitats catalanes.

Aquest suport inclou l'acompanyament de l'usuari en tot el procés d'instal·lació, optimització i posada en producció del programari de càlcul, així com la resolució de dubtes sobre la metodologia que cal aplicar en les simulacions. El fet de disposar d'interlocutors amb la mateixa formació que l'usuari aporta una gran agilitat al procés d'adaptació a la rutina de funcionament del Consorci. Com a conseqüència directa, els grups de recerca poden dedicar més temps a la investigació, sense haver-se de preocupar dels detalls més tècnics o metodològics.

3.1 Les simulacions computacionals: tercer paradigma científic

Les simulacions fetes amb computadors d'altres prestacions han esdevingut una part fonamental pel progrés científic i una eina imprescindible per a moltes branques de l'enginyeria [1]. Hi ha un gran nombre de problemes de l'enginyeria o de les ciències on els experiments són impossibles, perillosos o simplement extremadament cars; l'estalvi en recursos materials i temporals que permeten aquestes simulacions, per comparació als experiments i els dissenys clàssics, ha provocat un gran augment en la demanda dels serveis de computació d'altres prestacions.

Les simulacions computacionals proporcionen una eina útil per a l'investigador a l'hora de decidir si un experiment és físicament factible. Es parla comunament de les simulacions computacionals com a tercer paradigma científic després de la teoria i els experiments. De fet, actualment ja es comença a parlar del quart paradigma [2], que prové de l'explosió de l'ús de dades massives per extreure'n resultats.

Les aplicacions de la supercomputació abasten camps tan diversos com ara estudis de canvi climàtic, eficiència i seguretat a les centrals nuclears, desenvolupament d'energies renovables i disseny en materials avançats, entre d'altres.

3.2 Física computacional al CSUC

La física computacional és la ciència que s'encarrega d'estudiar i implementar algorismes numèrics per tal de resoldre problemes de física pels quals coneixem ja una teoria qualitativa que els descriu. Per tant, la física computacional és la branca de la física que se centra en l'elaboració

de models per ordinador de sistemes que no poden ser tractats analíticament.

Tal com s'ha mencionat anteriorment, la ciència moderna es fonamenta, d'una banda, en una teoria o un conjunt de teories que descriuen el fenomen i, d'altra banda, en els experiments i les simulacions que validen o donen suport a la teoria.

Aquesta disciplina comprèn pràcticament la totalitat de les teories físiques, des de la mecànica clàssica de Newton fins a la més moderna teoria de partícules o la cromodinàmica quàntica. No obstant això, no és trivial classificar les diferents branques de la física computacional ja que, en major o menor grau, es dóna certa transversalitat en les disciplines a l'hora de resoldre un problema.

Per exemple, es podria voler simular de forma realista el comportament d'un metall mesoscòpic que té certa temperatura. El fet que sigui mesoscòpic implica que s'ha de modelitzar com un conjunt de partícules puntuals amb les propietats característiques del material. També, pel fet de tenir temperatura, aquestes partícules han d'oscil·lar al voltant del seu punt d'equilibri i els seus electrons han de cenyir-se a la distribució de Fermi-Dirac: interaccionen constantment amb els seus fotons i donen lloc, finalment, a un model no ideal de cos negre.

Si es repassen tots els fenòmens físics que han intervingut per donar el resultat final d'aquest model, es comprova que en el moviment dels àtoms ha intervingut la mecànica newtoniana; en el dels electrons, la mecànica quàntica i, en el comportament dels fotons, la teoria quàntica de camps. Si s'hagués intentat resoldre de manera analítica, hauria estat pràcticament impossible obtenir característiques realistes d'aquest material. Aquí és on es fa evident la transversalitat de la física computacional.

3.2.1 Branques de la física computacional

Tot i l'estreta interrelació que es dóna en la física computacional per resoldre un sistema «realista», es poden distingir certes branques principals:

Física estadística

Estudia el moviment i la interacció de grans conjunts de partícules per obtenir dades macroscòpiques mitjançant les seves propietats microscòpiques. Dins d'aquesta branca es troba la termodinàmica, la matèria condensada o el plasma. Hi ha diferents paradigmes per atacar el problema: la dinàmica molecular, els mètodes de Montecarlo o les aproximacions típiques de la termodinàmica.

Física de l'estat sòlid

A diferència de la física estadística —que s'interessa pel comportament «mecànic» dels àtoms que formen un de-

terminat cos—, la física de l'estat sòlid se centra en l'estudi de l'entorn generat per aquests àtoms i tracta de determinar les interaccions entre la xarxa d'àtoms i els electrons (o fotons) que les orbiten. Bàsicament es pretén predir les característiques elèctriques i òptiques d'un determinat material tot coneixent la seva microestructura.

Química física

Com el seu nom indica pretén resoldre problemes químics mitjançant la resolució de l'equació de Schrödinger d'un conjunt concret d'àtoms. També fa servir alguns mètodes de l'estat sòlid i la física estadística per obtenir una descripció precisa de com es comporten aquestes molècules quan interaccionen entre elles o quan es modifiquen les seves condicions ambientals (proteïnes en dissolució, per exemple).

Física de medis continus

Bàsicament modelitza tots els sistemes d'infinits graus de llibertat (com ara sòlids deformables i fluids) com un únic cos. Es pot pensar en aquesta disciplina com una aproximació macroscòpica de la física estadística, ja que s'ignora l'atomicitat i les propietats microscòpiques del material i es modelitza com un volum o superfície contínua amb propietats macroscòpiques. La tècnica de resolució més freqüent per a aquest tipus de problemes és la d'elements finits, consistent a discretitzar la totalitat del cos en regions infinitesimals (però macroscòpiques) i aplicar les lleis de la mecànica (tensor d'esforços, Navier-Stokes, Burnett, etc.) sobre cada regió.

Astronomia i astrofísica

En l'àmbit de l'astronomia, la física computacional proporciona als científics eines per tractar la quantitat massiva de dades que envien els telescopis, cosa que permet analitzar i obtenir resultats útils en un temps raonable. Un exemple és el projecte Gaia, que pretén conformar un mapa 3D de la nostra galàxia. També a nivell cosmològic, permet validar teories com el Big Bang, la inflació o la dinàmica interna de cossos celestes mitjançant la solució de les equacions d'Einstein.

Teoria de camps

Un cas típic de problema (no trivial) sense solució analítica és la teoria de camps. Són moltes les raons per les quals es necessita la solució d'una equació de camp, per exemple en enginyeria d'antenes de telecomunicacions és necessari obtenir la geometria d'antena que genera un determinat camp electromagnètic, o, al contrari, quina forma tindrà l'ona generada per una antena de geometria do-

nada. Llevat que l'antena o el camp siguin molt simètrics, la resolució analítica és quasi impossible i fa necessari el càlcul numèric. Tampoc es pot obviar la teoria quàntica de camps que permet predir la dinàmica de les partícules subatòmiques en els col·lisionadors de partícules.

Sistemes complexos i no lineals

Tot i que la majoria de problemes físics queden ben descrits per equacions lineals no acoblades, la natura, en essència, és no lineal i acoblada. Exemples típics són la climatologia o física atmosfèrica on la seva dinàmica és resultat d'un conjunt de complexos fenòmens acoblats i fortament dependents de les condicions inicials, cosa que dona com a resultat el moviment de masses d'aire a les capes altes de l'atmosfera, la generació d'huracans, tempestes, corrents de superfície, gradients tèrmics i el clima en general.

3.2 Recursos de supercomputació al CSUC

Actualment el CSUC ofereix tres equips de càlcul d'altas prestacions, dels quals destaca per la seva arquitectura i capacitat un SGI Altix UV1000, anomenat *pirineus*.

La memòria compartida i els 6 TB de memòria, que equivalen a uns 1.500 ordinadors portàtils junts, són un recurs singular que no s'ofereix a cap altre centre de càlcul a l'Estat. L'arquitectura dels computadors que hi ha habitualment a la resta de centres (memòria distribuïda), no permet encabir segons quin tipus de simulacions, considerades com a simulacions d'alta demanda de memòria.

Així, al CSUC, els usuaris gaudeixen d'una infraestructura enfocada a la física computacional que consisteix en dos clústers de càlcul i aquesta màquina de memòria compartida amb 6 TB de memòria.

Sobre aquestes màquines s'executa programari científic de tot tipus, com ara Gaussian, enfocat a la química; Fluent, per a dinàmica de fluids o NAMD, per a dinàmica molecular. Totes aquestes eines, com s'ha comentat anteriorment, disposen d'un servei de suport a l'usuari que els ajuda a acabar correctament les tasques.

4 Projectes i usuaris

Diferents grups de recerca d'arreu de l'Estat utilitzen els serveis de supercomputació del CSUC per avançar en la seva recerca. El Consorci dona servei a universitats i centres de recerca en les àrees següents: física, química teòrica, modelització biomolecular, ciències de la Terra, mètodes numèrics en enginyeria i informàtica.

Actualment el CSUC té 44 projectes de recerca acadèmics de diferents institucions, com ara les universitats de Barcelona, Autònoma de Barcelona, Politècnica de Catalunya, de Girona, Rovira i Virgili, Internacional de Catalunya

i de les Illes Balears, i centres com l'Institut Català d'Investigació Química, el Consell Superior d'Investigacions Científiques, l'Institut de Ciència de Materials de Barcelona-CSIC i el Centre Internacional de Mètodes Numèrics en Enginyeria, entre d'altres.

El 2013, el projecte de recerca de més consum computacional va ser el liderat pel Dr. Juan José Novoa, de la UB, amb el títol *Enginyeria de cristalls moleculars d'interès tecnològic*, que va usar un 14,8 % del total d'hores computacionals (HC).

Aquest projecte té com a principal objectiu establir la metodologia necessària per portar a terme el disseny de cristalls moleculars amb propietats magnètiques, conductores de l'electricitat o superconductors. Per això, s'estudien les interaccions intermoleculars que es produeixen en aquests per mètodes *ab initio* i s'analitza com és possible fer-les més fortes i/o més direccionals.

El segon projecte va ser el liderat per la Dra. Núria López, de l'ICIQ, *Catàlisi de processos químics i química supramolecular*, amb un 10,8 %; i el tercer projecte, *Modelització molecular de sistemes amb metalls de transició*, del Dr. Agustí Lledós, de la UAB, amb un 8,5 % d'hores computacionals consumides.

El quart projecte de més ús computacional va ser *Anàlisi de densitats mono i bioelectròniques: nous desenvolupaments metodològics i aplicacions* del Dr. Miquel Duran i Miquel Solà, de la UdG, amb un 8,2 %; el cinquè projecte va ser *Reconeixement molecular* del Dr. Xavier Luque, de la UB, amb un 8 %; el sisè, *Estructura i propietats dels polímers: càlculs electrònics, simulacions atomístiques i models coarse-grained*, del Dr. Carlos Alemán, de la UPC, amb un 7,2 %; el setè projecte de recerca de més consum va ser *Models computacionals en materials d'interès tecnològic: de les nanopartícules a la catàlisi* del Dr. Francesc Illas, de la UB, amb un 7,1%; el vuitè va ser *Estudis teòrics d'estructura i reactivitat química. Aplicacions en sistemes d'interès biològic* de la Dra. Mariona Sodupe, de la UAB, amb un 6,8 %; el novè, *Estant els mètodes dinàmics cap a noves aplicacions en química i biologia* del Dr. Josep Maria Lluch, de la UAB, amb un 4,8 %, i, finalment, el desè projecte de més ús computacional va ser *Dinàmica de reaccions químiques elementals*, del Dr. Antoni Aguilar, de la UB, amb un 2,9 % del total d'hores computacionals consumides.

El CSUC, a més, també col·labora amb projectes d'R+D+I no acadèmics aportant-hi supercomputadors, programari especialitzat i suport professional.

Alguns dels projectes d'ús industrial en què participa la unitat de càlcul d'altres prestacions del CSUC són:

La Formula Student Spain, organitzada per la Sociedad de Técnicos de Automoción (STA), reuneix anualment universitaris d'enginyeria de tot el món perquè participin en el disseny, el desenvolupament, la construcció i la conducció d'un monoplaça de competició. Aquesta iniciativa té el suport de la indústria de l'automòbil, la participació d'enginyers d'alt nivell i el suport institucional de diverses entitats,

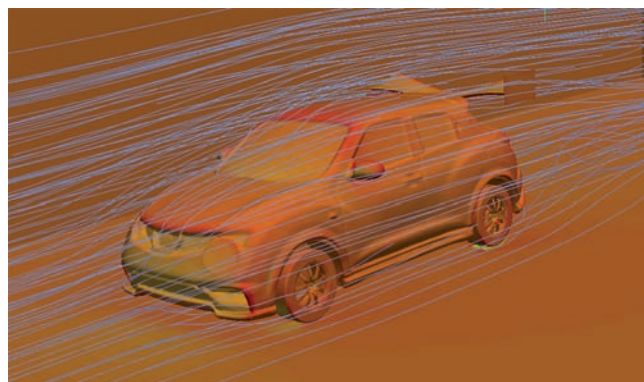


Figura 1: Simulació de fluidodinàmica realitzada amb OpenFoam per minimitzar la resistència a l'aire i optimitzar la sustentació del vehicle.

entre les quals es troba el CSUC, que ofereix als equips espanyols que participen a la competició accés al seu maquinari per fer simulacions del comportament dels prototips.

Relacionat amb l'anterior, també participa en el projecte Beyond Formula Student, que s'ha basat en el disseny i la construcció del cotxe de competició Nissan Juke Nismo que ha competit en les 24 h de Barcelona d'Automobilisme-Trofeu Fermí Vélez, la cursa que va tenir lloc al mes de setembre al Circuit de Barcelona-Catalunya. En aquesta col·laboració, el CSUC ha ofert també el seu maquinari de supercomputació per fer simulacions d'aerodinàmica que han ajudat a confeccionar i perfeccionar aquest cotxe de competició.

Un altre, que es desenvolupa en les pàgines següents, és el projecte NUMEXAS (Numerical Methods and Tools for Key Exascale Computing Challenges in Engineering and Applied Sciences), que té per objectiu desenvolupar, implementar i validar la següent generació de mètodes numèrics que permetran resoldre de



Figura 2: El CSUC ha participat en l'equip B_FS desenvolupant les millores aerodinàmiques usant eines avançades de simulació computacional.



Figura 3: El Nissan Juke, de l'equip Beyond Formula Student (B_FS), durant la seva participació en les 24 hores de resistència a Montmeló.

forma rutinària problemes de classe exaescala equivalents a milions de nuclis de computació en enginyeria i les ciències aplicades.

Un altre projecte, que va iniciar-se el 2010, va ser Supercalculus, que tenia com a objectiu crear un demostrador d'aplicacions de càlcul amb necessitats extremes de computació. Així, el projecte incloïa el desenvolupament i la implementació de solucions computacionals que encara no són a l'abast de la indústria a causa dels seus grans requisits.

4.1 Els projectes de física al CSUC

La majoria dels usuaris del Consorci tenen projectes científics relacionats directament o indirectament amb la física. Des de l'any 1991, la branca científica majoritària en els càlculs computacionals al Consorci ha estat la química física i, de fet, el programari més usat (amb un ús superior al 50 % del total d'hores de càlcul) és Gaussian [3], un paquet de programari computacionals que s'usa principalment per resoldre problemes en aquest àmbit.

4.1.1 Simulació numèrica de fronts costaners nocturns a la conca de la Mediterrània

Les ciències del clima constitueixen una altra branca de la física computacional que més recursos ha consumit. Aquest és l'àmbit d'estudi del grup de recerca liderat pels Drs. Jordi Mazón i David Pino del Departament de Física Aplicada de la UPC.

L'estudi del clima i la dinàmica atmosfèrica sempre han estat un terreny propici per a l'ús de la supercomputació. Les simulacions que es duen a terme es basen en equacions no lineals i acoblades, fet que els confereix un comportament eminentment caòtic. Els models que prediuen l'evolució del

clima i dels fenòmens atmosfèrics es basen en complicades interaccions entre diferents branques de la física.

Des de la dinàmica de fluids, per predir el moviment de masses d'aire, fins a la física estadística, per estimar la probabilitat de precipitacions a una hora i lloc determinats. Aquests models, per obtenir la seva «relativament» alta precisió (a una setmana vista), necessiten mesures acurades. Per exemple, dels gradients tèrmics i de pressió, especialment en les regions costaneres, on són particularment alts respecte a l'interior.

Aquests gradients tèrmics, que generen els anomenats *fronts costaners*, tenen una forta influència sobre la freqüència de precipitacions a les conques dels litorals. Els resultats s'han obtingut mitjançant els models Weather Research and Forecasting (WRF) i MM5, com expliquen els Drs. Jordi Mazón i David Pino en el seu article que s'exposa en aquest número de la revista, resultats que s'han obtingut de simulacions fetes amb el maquinari de supercomputació del CSUC.

Aquest estudi és un exemple paradigmàtic de la importància que tenen les simulacions computacionals en el món real, ja que permeten predir els efectes de fenòmens naturals per prevenir-ne d'aquesta manera les possibles conseqüències. Així, en aquest cas s'estudien les freqüències de precipitacions associades al balanç hídric global en zones costaneres del Mediterrani on hi ha greus problemes amb les reserves hídriques, a conseqüència de la gran pressió demogràfica que suporten moltes d'aquestes zones.

4.1.2 Simulació de defectes en l'estructura atòmica de metalls i aliatges

Un altre camp on s'usa àmpliament la física computacional és en les ciències dels materials i més concretament en la caracterització de materials mitjançant eines de dinàmica molecular. Com és ben conegut en la ciència de materials, la microestructura (organització interna dels àtoms) determina les característiques macroscòpiques del material (duresa, flexibilitat, impureses, etc.), de manera que seria una bona opció poder determinar la resposta d'aquesta microestructura en produir-li alteracions diverses.

L'eina computacional típica per a aquests propòsits és la dinàmica molecular (MD, de les seves sigles en anglès). Aquest mètode tracta senzillament de modelitzar el material com un conjunt de molts cossos esfèrics (àtoms) que interaccionen mútuament sota la influència de certes interaccions i forces preestablertes. De la mitjana del comportament individual es poden obtenir les característiques mecàniques, tèrmiques, etc., bàsicament qualsevol magnitud que pugui ser «derivada» del moviment d'aquestes partícules.

La Dra. Anna Serra, en el seu article, ens introdueix en el món dels assajos de materials modelitzats amb MD, amb l'objectiu de determinar la influència dels defectes

de l'estructura cristal·lina sobre les propietats mecàniques de certs metalls i aliatges. S'ha centrat principalment en assaigs plàstics de microindentació i irradiació.

4.1.3 Simulacions de Montecarlo del transport de la radiació per a aplicacions en medicina i protecció radiològica

Encara que la dinàmica molecular és una eina molt precisa per caracteritzar sistemes estadístics, a vegades la complexitat intrínseca del sistema la fa inadequada o, directament, inaplicable. En aquestes ocasions, cal abordar el problema des d'una perspectiva diferent i, en comptes de simular la pròpia interacció i determinar-ne els efectes, s'usen mètodes estadístics.

Així, per exemple, amb el mètode de Montecarlo se simula directament l'efecte individual de cada part constituent del sistema i se n'estima l'efecte en la mitjana. Aquest mètode treballa amb la probabilitat que es doni un o un conjunt d'efectes mitjançant l'ús de nombres aleatoris (en rigor, pseudoaleatoris) i estima el comportament col·lectiu del sistema. Un camp on és típica la utilització d'aquesta metodologia és en la simulació de processos radioactius, ja que aquests fenòmens obeeixen una llei de probabilitats ben coneguda.

Precisament aquesta és la metodologia seguida pels Drs. Ramon Casanovas, Maria Cros i Marçal Salvadó en el seu article, on proposen un conjunt de dispositius i protocols per millorar la fiabilitat en l'estimació de les dosis, efectives, rebudes en pacients humans, així com en el medi ambient.

4.1.4 Gaia: la Galàxia en un petabyte

Una altra faceta de la supercomputació que cada cop està prenent més importància és, a banda del càlcul, el tractament de grans quantitats de dades. No és estrany que bases de dades de bioinformàtics, per exemple els arbres filogenètics, o resultats de simulacions produeixin terabytes de dades que s'han de processar. Si s'haguessin de processar per tal de cercar en aquest oceà de resultats amb un ordinador d'escriptori es perdria tot el temps que s'ha usat per fer càlculs amb els supercomputadors. Per tant, és lògic que el processament d'aquesta gran quantitat de dades obtingudes amb els càlculs es faci també mitjançant grans màquines de computació.

Aquesta ingent quantitat d'informació serà el resultat de la missió Gaia, com expliquen els Drs. Eduard Masana, Xavier Luri i Jordi Portell en el seu article. Per fer-se una idea del que aquesta gran quantitat d'informació pot suposar, la capacitat d'un processador actual pot moure, en el millor dels casos, de l'ordre d'1 GBps. Els algorismes per tractar-les s'han estat desenvolupant durant aquests últims anys que han precedit l'inici de la

missió amb l'ajut de supercomputadors, entre d'altres, els del CSUC.

Gaia és un dels projectes paradigmàtics per al CSUC amb qui col·labora en el processament de dades des de l'any 1998 i, anteriorment, amb el seu predecessor, Hipparcos. Aquest projecte és una mostra de la importància dels serveis que s'ofereixen al Consorci, que serveixen com a eina en projectes de gran envergadura científica com l'esmentat, que té per propòsit científic desvelar la història de la formació i l'evolució de la Via Làctia. Un equip nombrós d'astrònoms i enginyers de la UB i de l'IEEC hi treballen des de fa més de quinze anys.

5 Els reptes en computació científica al CSUC

El Departament de Càlcul i Aplicacions del CSUC participa en projectes europeus relacionats amb els dos reptes més importants que té en un futur proper la supercomputació.

El primer desafiament científic que té la computació és trobar la forma de construir màquines més potents que siguin a la vegada sostenibles energèticament. A més, s'ha de dissenyar en paral·lel una nova generació de codis i algorismes per dur a terme les simulacions en aquesta nova família de computadors. Aquest és un dels objectius del projecte NUMEXAS [4].

L'altre repte és la gestió de la ingent quantitat de dades que avui dia es generen amb els programaris científics, com ara els que s'utilitzen en camps com la biologia clínica o la genètica. Per al seu tractament també s'estan desenvolupant, en el marc del projecte GENIUS [5], nous paradigmes de programació i formes d'ús del maquinari disponible, que conflueixen amb l'actualment omnipresent model de dades massives (*big data*).

5.1 La supercomputació de camí a l'exaescala

Les simulacions amb supercomputadors s'han convertit en una eina imprescindible per a moltes branques de la ciència, així com per a un nombre cada cop més gran de processos industrials [6]. Això ha fet que el càlcul d'altres prestacions s'hagi convertit en una prioritat per als governs de les principals potències econòmiques mundials, que l'estan potenciant amb molts esforços i recursos econòmics per mantenir o millorar la seva competitivitat.

La potència de càlcul dels computadors es mesura habitualment en flop/s (nombre d'operacions de coma flotant per segon). Les màquines més potents avui dia es troben a la petaescala (capaces d'executar 10^{15} flop/s) i les regions més potents del món estan lluitant per ser les primeres a arribar a l'exaescala (10^{18} flop/s). Aquesta competició per arribar a ser els primers a disposar d'una màquina a l'exaescala pot comparar-se a la cursa espacial que hi va haver els anys seixanta entre la Unió Soviètica

i els Estats Units. En aquesta ocasió, la cursa està liderada principalment pel Japó i els Estats Units, seguits per la Xina i Europa.

Aquesta nova generació de supercomputadors a l'era de l'exaescala permetran solucionar un gran nombre de reptes científics [7] que fa uns anys eren impensables de resoldre. Aquestes màquines seran una realitat al voltant de l'any 2020, però perquè siguin possibles s'han de resoldre moltes qüestions de caràcter tècnic i lògic relacionades amb els nous tipus d'arquitectures que estan emergent i les famílies d'algoritmes i programari que vindran [8].

5.1.1 Noves arquitectures híbrides de càlcul

La coneguda llei de Moore [9] diu que les millores tecnològiques introduïdes al mercat, principalment en les tecnologies dels semiconductors, fan que el rendiment dels processadors es dobli aproximadament cada divuit mesos. Però aquesta llei pot estar arribant a la seva fi si no hi ha un canvi de paradigma en el disseny dels computadors usats actualment.

Els supercomputadors actuals estan formats bàsicament per multitud de nodes de CPU amb múltiples nuclis. Aquesta arquitectura difícilment podrà escalar en un factor 1.000 com requeriria el salt a l'exaescala per un motiu ben senzill, la inabordable despesa energètica.

El supercomputador més potent que existeix actualment és el Tianhe2, que amb una potència de 33 Pflop/s requereix una potència de 17 MW. Això vol dir que, seguint una evolució lineal, per arribar a aconseguir Eflop/s es necessitaria una potència energètica de l'ordre d'alguns GW, és a dir, caldria una petita estació de generació elèctrica només per fer funcionar el supercomputador. La conseqüència directa d'aquest fet és que s'han de desenvolupar, i de fet ja estan apareixent, noves arquitectures de computació que bàsicament es divideixen en dos tipus.

El primer, processadors amb multitud de nuclis (GPU, Xeon Phi, etc.) que actuen com a acceleradors i que aporten potència amb un baix consum energètic. I, el segon, un nombre massiu de processadors de tipus ARM (Advanced Rise Machine) de molt baix consum energètic, similars als que s'utilitzen per a les tauletes tàctils i dispositius mòbils intel·ligents.

En qualsevol cas, els futurs supercomputadors estaran formats d'alguns d'aquests nous tipus de processadors, combinats entre ells o amb CPU clàssiques, i per aquesta raó es parla d'arquitectures heterogènies o híbrides.

5.1.2 El projecte NUMEXAS

Aquest salt en la supercomputació no es pot aconseguir només amb l'ampliació de codis existents, sinó que requereix el desenvolupament de nous paradigmes de mètodes numèrics específics per explotar al màxim les capacitats

intrínseques d'aquestes infraestructures de computació que contenen milions de nuclis.

La Unió Europea en el seu programa marc per a la recerca i la innovació té un apartat especial dedicat a la recerca de les noves arquitectures i programari necessari per aconseguir arribar a calcular a l'exaescala [10].

Un d'aquests projectes és NUMEXAS, que té una durada de tres anys, està liderat pel CIMNE i hi participen, a més del CSUC, l'empresa Quantech Arz i les universitats Leibniz Universität Hannover i la National Technical University of Athens.

El projecte es va posar en marxa l'octubre de 2013, i té l'objectiu de desenvolupar, implementar i validar una nova generació de mètodes numèrics que es puguin executar en les noves arquitectures de càlcul a l'exaescala, perquè els problemes més complicats en enginyeria i les ciències aplicades es puguin resoldre de manera rutinària.

El principal resultat de NUMEXAS serà un nou conjunt de mètodes numèrics i codis que permetran a la indústria, al govern i al món acadèmic resoldre problemes en l'enginyeria i ciències aplicades amb l'eficiència i la facilitat d'ús dels codis actuals d'última generació.

La principal tasca del CSUC consisteix a proveir els *partenaires* de les eines de computació necessàries per fer les seves tasques (tant de maquinari com de programari específic) i d'altra banda està involucrat en les feines de *benchmarking* i optimització de les diferents arquitectures dels codis generats pels desenvolupadors.

5.2 Tractament de dades massives (*big data*)

El gran volum de dades que s'està generant actualment, provinent de diverses fonts, i entre les quals hi ha el programari científic, ha provocat l'aparició d'un nou paradigma per al tractament de dades. És el que es coneix com el model de dades massives, un terme aplicat a conjunts de dades que superen la capacitat del programari habitual per ser capturats, gestionats i processats en un temps raonable. Així, malgrat que el terme només faci menció a la grandària de les dades, aquesta no és l'única variable important en aquest conjunt de dades. Quan es parla d'aquesta tecnologia també és molt important la velocitat en què es poden processar les dades i la seva varietat.

El problema principal del tractament de grans volums de dades es redueix al fet que, malgrat que la capacitat dels discos ha augmentat enormement en els darrers anys, la velocitat de transferència de dades no ho ha fet al mateix ritme. Això pot provocar que la lectura de dades d'un disc pugui trigar diverses hores. La solució passa per llegir diversos discos alhora, en paral·lel, cosa que proporciona una velocitat d'accés a les dades més elevada.

Aquesta solució pot comportar dos problemes. El primer és que com més discos s'afegeixen en paral·lel més probabilitat hi ha que es produeixi un error de maquinari.

El segon és que en moltes de les anàlisis que es fan és habitual que es vulgui combinar dades de diferents discos, i això no és una tasca trivial. Una solució a aquests dos tipus de problemàtiques és la que ens proporciona Hadoop. Ofereix, d'una banda, un sistema d'emmagatzematge fiable mitjançant HDFS (Hadoop Distributed FileSystem) i, de l'altra, un sistema d'anàlisi de dades distribuïdes a través del model de programació MapReduce.

5.2.1 El projecte GENIUS

L'objectiu final del projecte GENIUS és incrementar l'impacte de la missió astromètrica Gaia, considerada com una de les pedres angulars de l'Agència Espacial Europea (ESA), i que suposarà un avanç cabdal en el camp de l'astrofísica. El satèl·lit Gaia, el llançament del qual va ser el 19 de desembre de 2013, té l'objectiu de produir el mapa tridimensional de la Via Làctia més precís i complet fins avui.

L'ESA va delegar el processament de dades de Gaia en un consorci paneuropeu anomenat DPAC (Data Processing and Analysis Consortium) que ja està treballant en la seva implementació. El resultat final serà un catàleg i un arxíu de dades que contindran més de mil milions d'objectes.

El sistema d'arxíu que contindrà aquestes dades és al Centre Europeu d'Astronomia Espacial (ESAC), i servirà com a base per a l'explotació científica de les dades de Gaia. El fet de posar a disposició de la comunitat científica les dades incloses a l'arxíu comporta una tasca prèvia de definició, implementació i operació de l'arxíu. Amb aquest objectiu es va crear una nova unitat de coordinació (CU9) dins de DPAC. GENIUS està pensat per fer contribucions significatives dins d'aquest context, basant-se en els principis següents:

- Dissenyar l'arxíu segons les necessitats de la comunitat d'usuaris.
- Oferir eines d'explotació per maximitzar el retorn científic.
- Assegurar la qualitat dels continguts de l'arxíu i la interoperabilitat amb els arxius astronòmics existents i futurs (ESAC, ESO, etc.).
- Cooperar amb les altres dues úniques missions astromètriques al món, nanoJASMINE i Jasmine (Japó).
- Facilitar les activitats acadèmiques i de divulgació per fomentar l'interès públic en la ciència en general i l'astronomia en particular.

GENIUS encaixa a la perfecció en activitats ja existents de Gaia, explotant les diferents sinergies creades arran dels desenvolupaments en curs. Els seus membres participen activament en aquestes tasques i aporten un coneixement profund de la missió, així com experiència en àrees clau de desenvolupament.

Aquestes sinergies s'estenen, d'altra banda, gràcies a la col·laboració amb l'equip de científics experts i desenvolupadors de programari de DPAC, els quals treballen des de l'any 2006 en la definició de la millor manera de processar les dades de la missió Gaia. A més, també s'han establert llaços de cooperació amb els equips de treball de les missions astromètriques japoneses esmentades.

5.2.2 Participació del CSUC

El CSUC participa dins el projecte GENIUS en tres àrees diferenciades. La primera consisteix a oferir un catàleg simulat que permeti omplir la base de dades amb dades realistes amb l'objectiu de comprovar i desenvolupar prototipus de l'arxíu i les eines associades. D'aquesta manera, un cop arribi la primera entrega de les dades de Gaia, els productes del projecte estaran disponibles a temps per al seu ús en l'arxíu allotjat a l'ESAC. Els treballs duts a terme pel simulador de Gaia s'han executat en els supercomputadors del CSUC, que, a banda de proporcionar el maquinari adient, també han aportat el servei de suport tècnic per aconseguir un ús més òptim de les seves instal·lacions.

La segona àrea de col·laboració del CSUC s'emmarca en el desenvolupament i la implementació d'un portal comunitari que inclourà informació sobre el projecte, documents, enllaços a referències, alertes, notícies, guies, etc. A més, el CSUC proporciona la infraestructura bàsica per acollir el portal, incloent el manteniment necessari conseqüent, tant de programari com de maquinari.

Finalment, la tercera àrea de col·laboració se centra a obtenir nova informació sobre les dades recollides per la missió Gaia, fent servir metodologies i eines adients per al tractament de quantitats massives de dades. Aquí és on entra en acció el model de dades massives abans mencionat, en què s'interrogarà la immensa base de dades que resultarà ser el catàleg de Gaia. Aquest interrogatori posarà en evidència relacions entre les dades que n'aportarà una nova visió.

El CSUC ha contribuït en aquesta fase amb el disseny i posterior desplegament d'un prototip de clúster de Hadoop dins de les seves instal·lacions. Aquest clúster consta de 16 nodes de computació, corrent sobre sistema operatiu CentOS 6.5 i on s'ha instal·lat el CDH 5.0, la distribució de codi lliure de Cloudera que inclou Apache Hadoop i altres projectes relacionats. Aquest pilot ha permès de fer diferents proves sobre el funcionament d'aquest nou paradigma i comprovar les eines que Cloudera aporta per executar treballs altament distribuïts en aquest entorn. Tot i el fet que aquest clúster continuarà en producció a mig termini, ja s'està treballant en el disseny d'un de nou, amb una arquitectura de computadors més moderna i que ofereix unes prestacions més elevades.

Bibliografia

- [1] R. ROSNER et al., *The Opportunities and Challenges of Exascale Computing*, Office of Science, U. S. Department of Energy, 2010.
- [2] J. CHEN, A. CHOUDHARY, S. FELDMAN, B. HENDRICKSON, C.R. JOHNSON, R. MOUNT, V. SARKAR, V. WHITE, D. WILLIAMS. *Synergistic Challenges in Data- Intensive Science and Exascale Computing*, DOE ASCAC Data Subcommittee Report, Office of Science, Department of Energy, March, 2013.
- [3] Gaussian 09, Revision D.01, M. J. FRISCH et al., Gaussian, Inc., Wallingford CT, 2009.
- [4] www.numexas.eu.
- [5] genius-project.csuc.cat.
- [6] C.S. GLOTZER et al., *WTEC Panel Report on International Assessment of Research and Development in Simulation Based Engineering and Science*. World Technology Evaluation Center (WTEC.org), 2009.
- [7] P.M.C. DAVID, L. BROWN, *Scientific Grand Challenges: Crosscutting Technologies for Computing at the Exascale*, U. S. Department of Energy, Washington D.C., 2010.
- [8] J. DONGARRA, P. BECKMAN, T. MOORE, P. AERTS, G. ALOISION and JEAN-CLAUDE, The International Exascale Software Roadmap, *International Journal of High Performance Computer Applications*, **25(1)**, 2011.
- [9] MOORE GORDON, E. Cramming more components onto integrated circuits, *Electronics Magazine*, p. 4, 1965.
- [10] <http://exascale-projects.eu>.