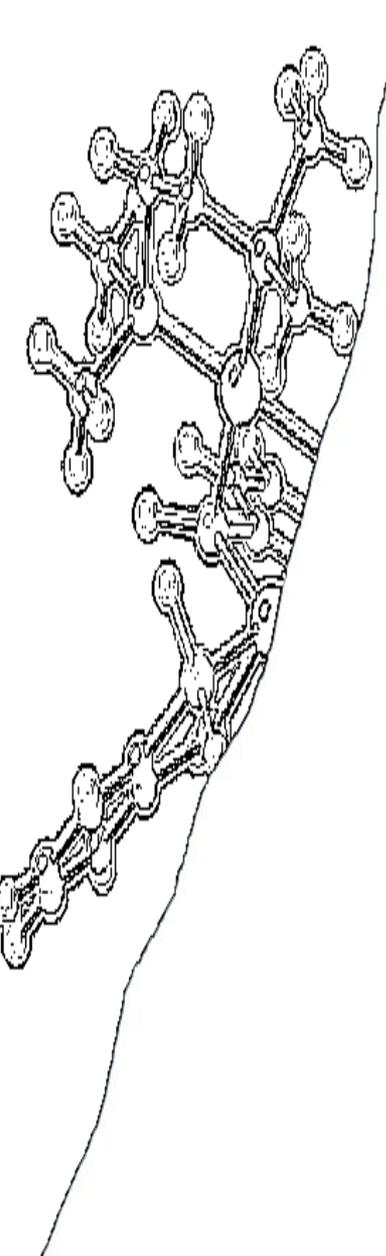


Códigos de Química Computacional en el CESGA (Gaussian, GAMESS, NWChem)

Aurelio Rodríguez,
Técnico de Aplicaciones
aurelio@cesga.es





✓ HARDWARE

- Componentes básicos
- Rendimiento de disco y memoria
- Uso de memoria y disco
- Uso óptimo del hardware
- Consideraciones sobre las colas



✓ SOFTWARE

- Gaussian, GAMESS, NWChem
- Direct vs. incore vs. disk
- Ejecución en paralelo
- Consideraciones acerca de los métodos: HF, DFT, MP2, CCSD, CCSD(T), CI



CCL

✓ GENERALIDADES

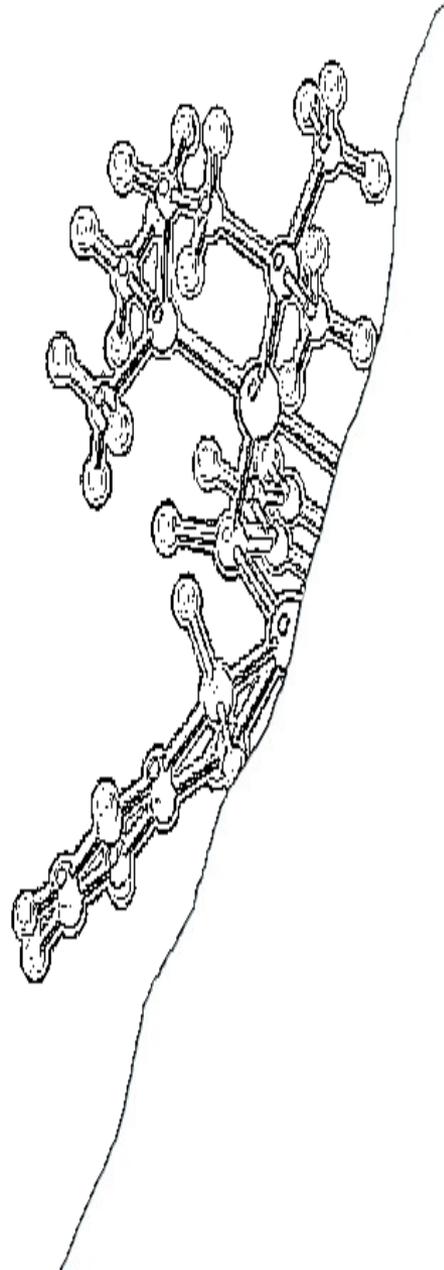
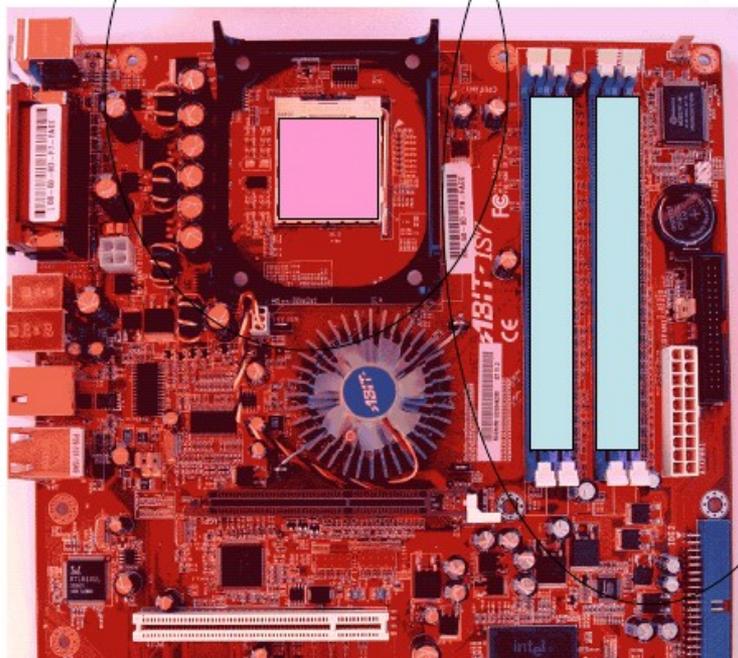
- FAQ



Basic Hardware Components

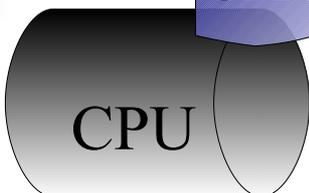
CPU Cache Memory

Disk





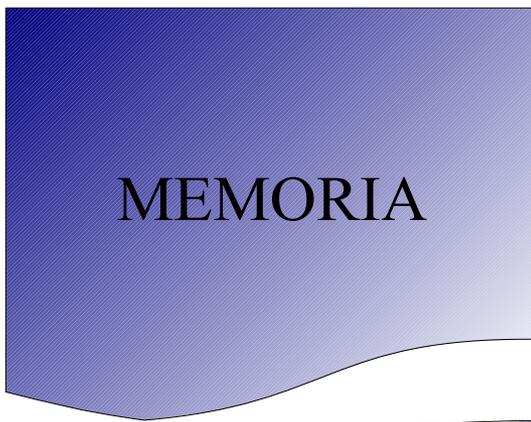
Tamaño: 4-12MB



CPU



CACHE



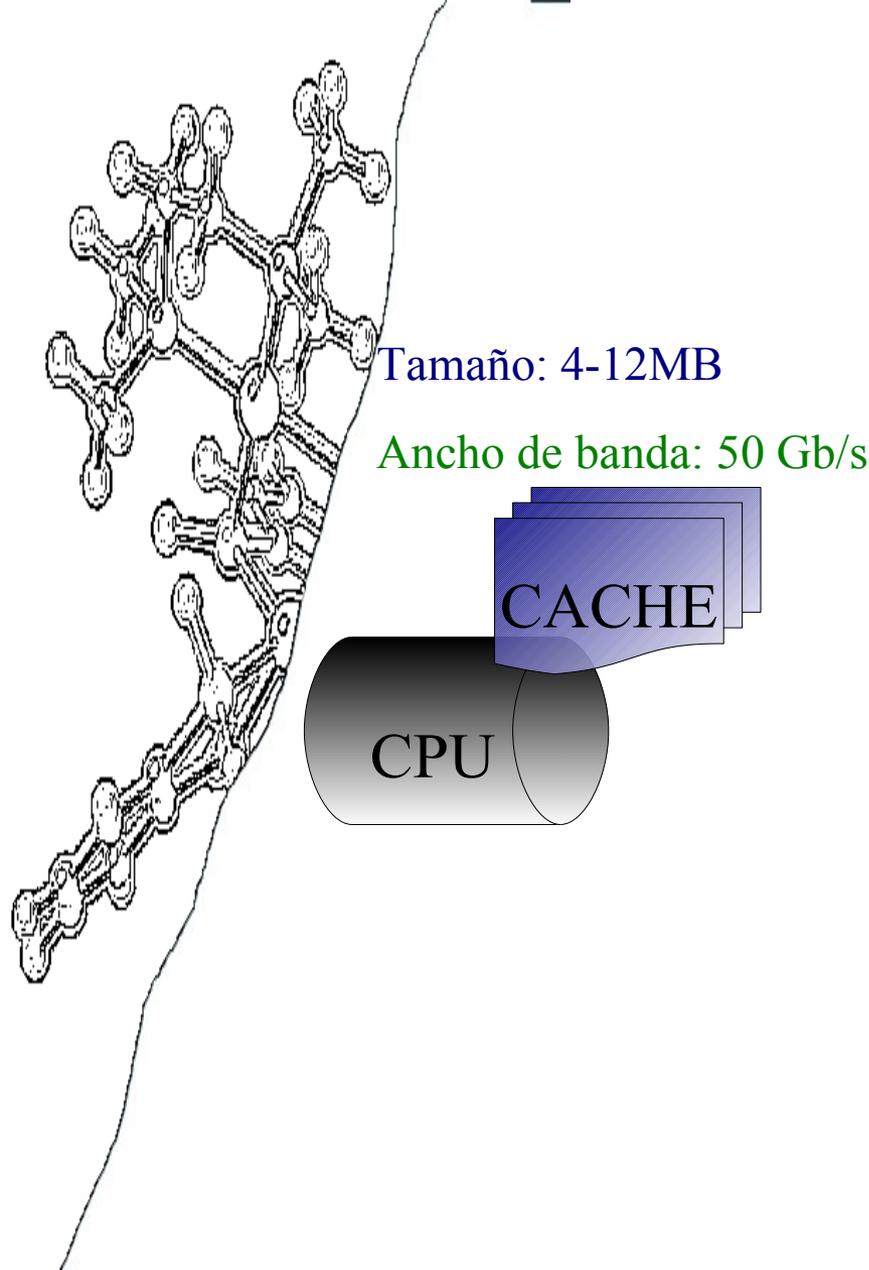
MEMORIA

Tamaño: 2-128GB



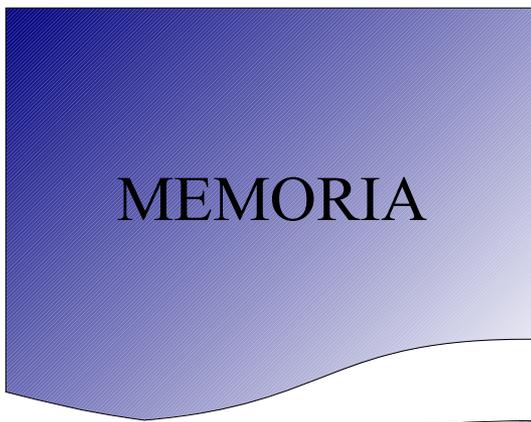
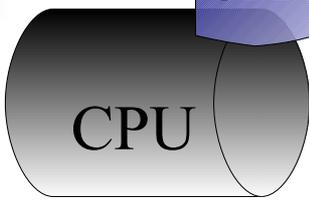
DISCO

Tamaño: 500 GB- muchos TB



Tamaño: 4-12MB

Ancho de banda: 50 Gb/s



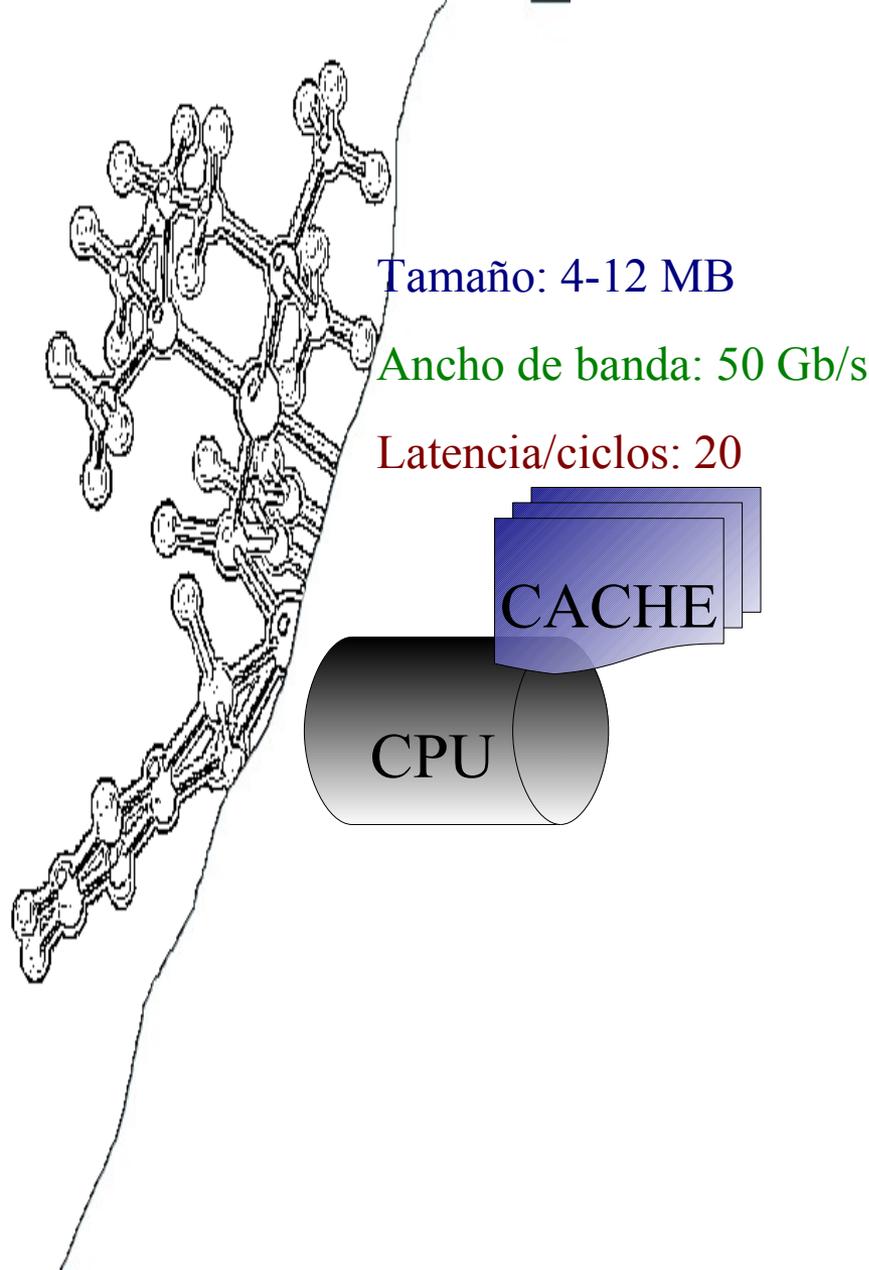
Tamaño: 2-128GB

Ancho de banda: 5 Gb/s



Tamaño: 500 GB- muchos TB

Ancho de banda: 250 MB/s



Tamaño: 4-12 MB

Ancho de banda: 50 Gb/s

Latencia/ciclos: 20

CACHE

CPU

MEMORIA

Tamaño: 2-128GB

Ancho de banda: 5 Gb/s

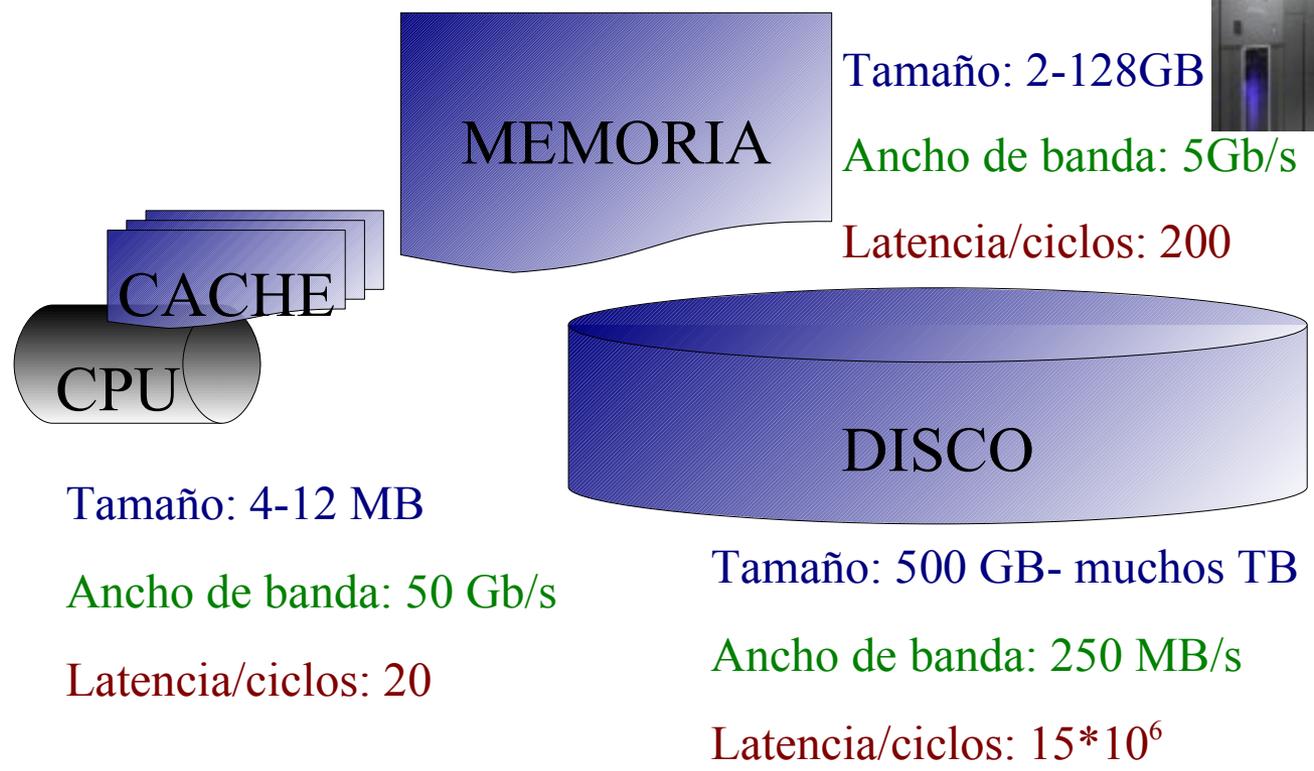
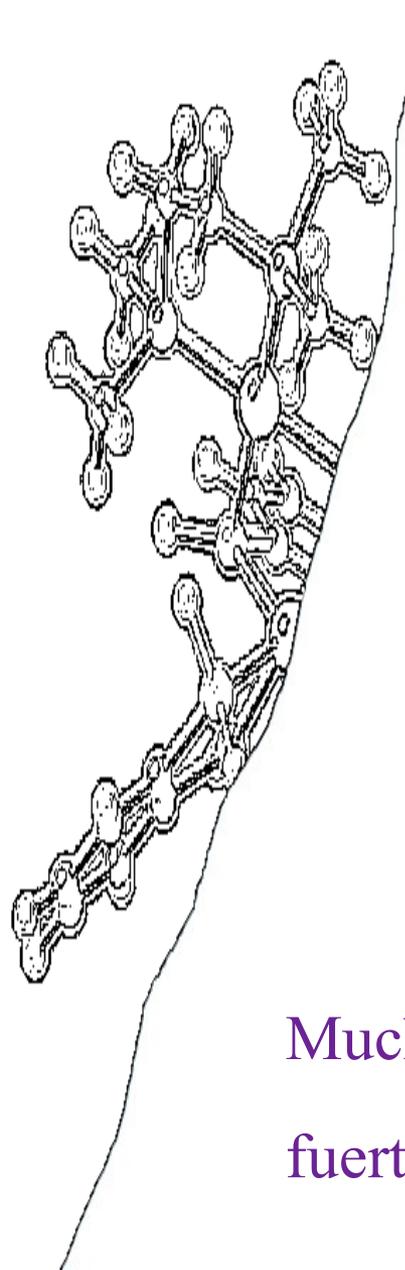
Latencia/ciclos: 200

DISCO

Tamaño: 500 GB- muchos TB

Ancho de banda: 250 MB/s

Latencia/ciclos: $15 \cdot 10^6$

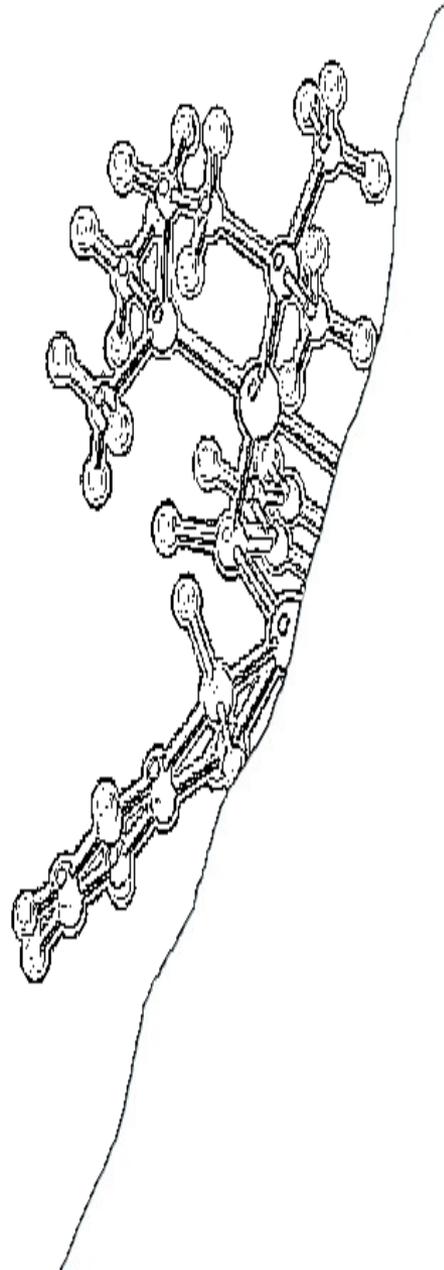


Muchos códigos permiten elegir algoritmos que dependen fuertemente de alguno de estos componentes



✓ Disco:

→ Cuando se ejecutan algoritmos que usan mucho espacio en disco, el rendimiento del disco puede afectar fácilmente en un factor muy grande el tiempo de ejecución.





Ejercicio con GAMESS:

Ejecuta GAMESS en el Finis Terrae:

- 1) con el valor por defecto de la variable de entorno \$TMPDIR
- 2) con \$TMPDIR fijada a \$HOMESFS/\$JOB_ID
- 3) con \$TMPDIR fijada a \$HOME/\$JOB_ID

Cuanto tarda cada caso?



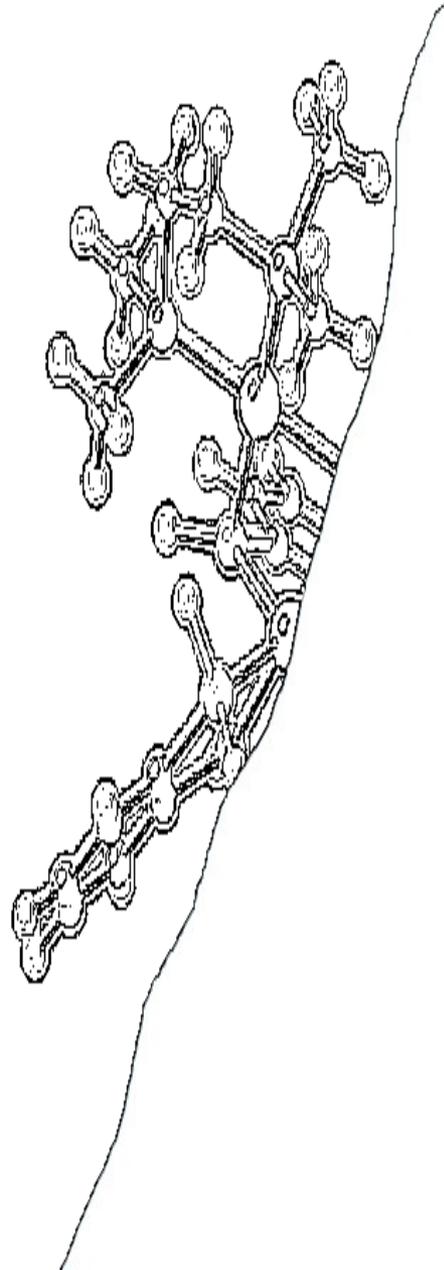


✓ Recomendaciones:

→ La configuración por defecto está pensada para ser óptima: La variable \$TMPDIR se fija al directorio con disco óptimo.

→ Si se requiere el uso de disco superior a 800GB contactar:

→ aplicaciones@cesga.es/sistemas@cesga.es





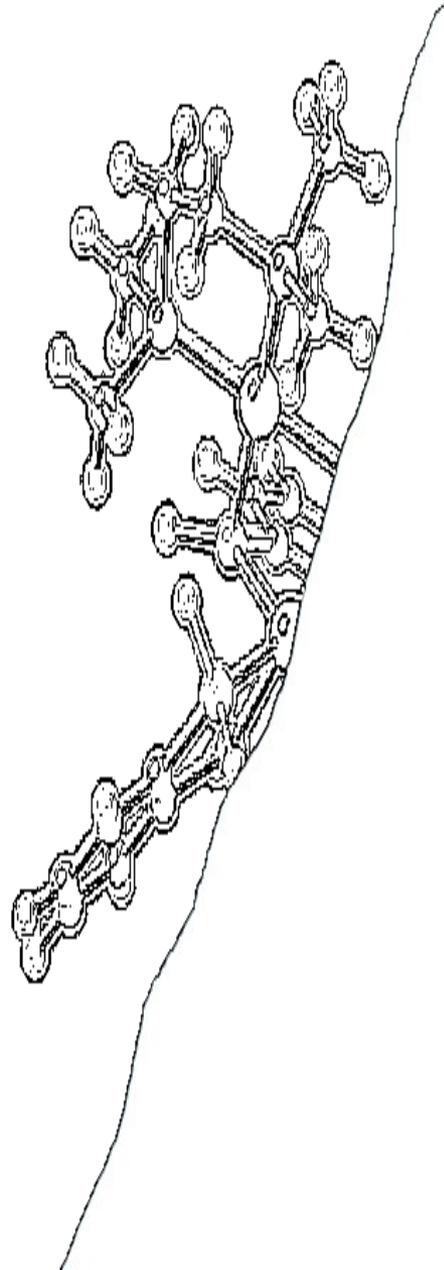
→ MEMORIA:

→ En general es recomendable usar tanta memoria como sea posible.

→ La mayoría de los algoritmos están diseñados para usar disco o recalcularse si los cálculos no caben en memoria.

→ Recalcular no suele ser recomendable (pero frecuentemente es mejor que escribir a disco)

OJO!!! con los cálculos paralelos





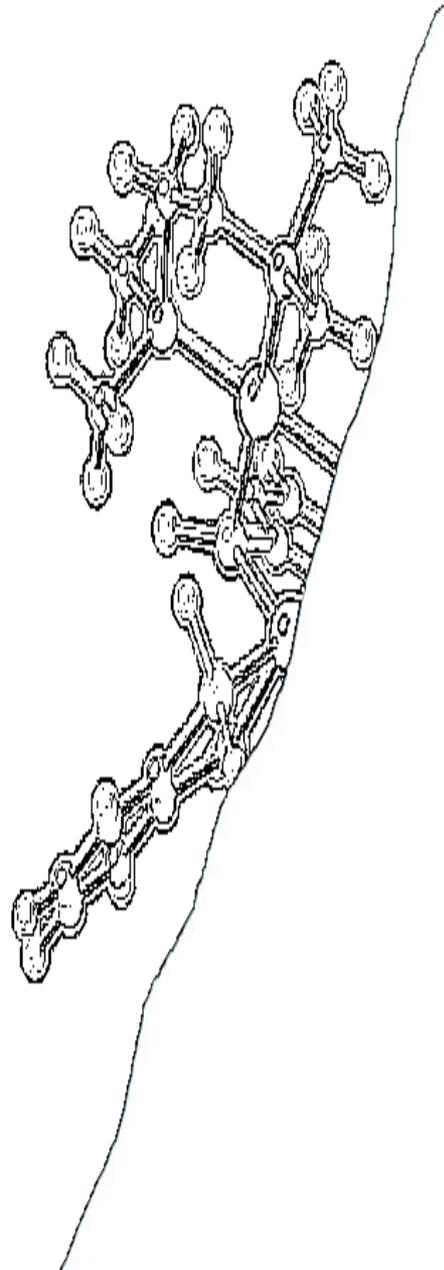
✓ Memoria disponible:

→ Finis Terrae:

- Nodos rx: 128 GB
- Superdome: 256 GB
- Superdome: 1 TB

→ SVGD:

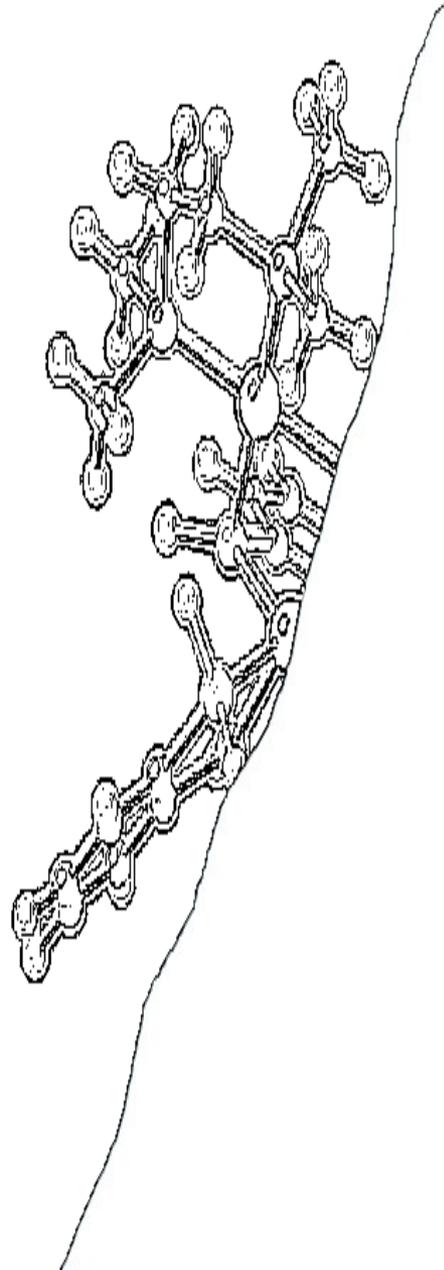
- Nodos 32: 1 GB
- Opteron: 4GB
- Blades: 4GB
- Blades: 8GB





Usa tanta memoria como te permita el programa

- Usa tanta memoria como puedas limitado por el hardware o por el programa que estás usando
- Asegúrate de que tu programa está haciendo uso de la memoria



GAMMESS
QUANTUM


NWChem
high performance computational chemistry software

 **CESGA**



Especificar Memoria en Gaussian

%mem=1200mb



Comando link0

#MP2/6-31G* opt

.....

.....

.....

qsub -l ...,s_vmem=1500M,... run.sh



Especificar memoria en GAMESS

```
$SYSTEM MWORDS=64 MEMDDI=128 $END
```

.....

.....

.....

(MEMDDI + MWORDS*NPROC)* 8 MB

```
qsub -l ...,s_vmem=2G,... run.sh
```





Especificar memoria en NWChem

memory total 1000 mb

memory heap 250 stack 250 global 500 mb

.....

Las rutinas individuales reservarán memoria. El programa abortará si encuentra una rutina que pide demasiada memoria.

.....

NWChem suele dar un mensaje de error muy descriptivo.

.....

Si estás ejecutando en paralelo adicionalmente seguirán muchos errores MPI

```
qsub -l ...,s_vmem=1500M,... run.sh
```



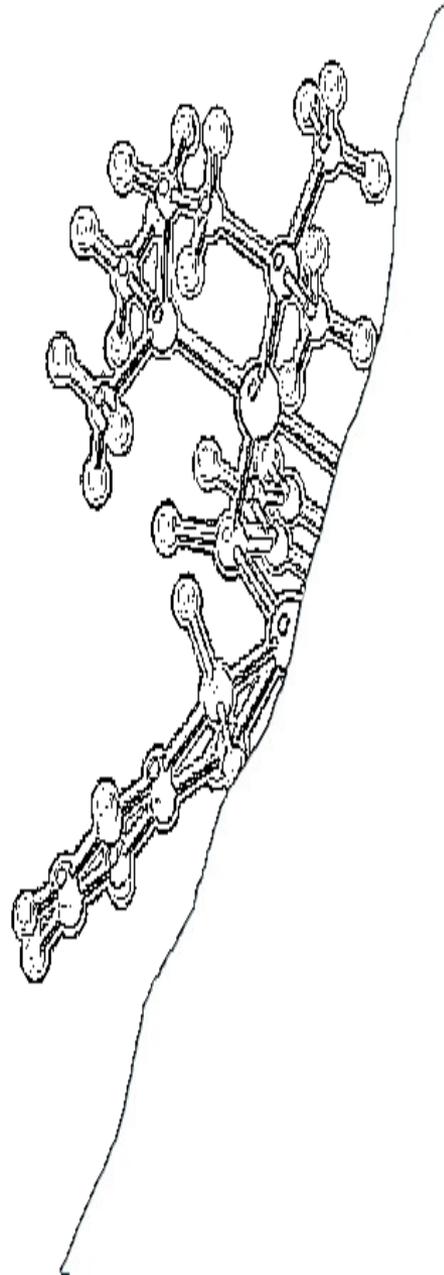
Ejercicio con Gaussian:

Ejecuta Gaussian 03 en el Finis Terrae:

- 1) con el método DFT usando el algoritmo incore
- 2) con el método DFT usando el algoritmo directo

¿Cuanto tarda cada caso?

¿Qué ocurre cuando se usan 2 procesadores?

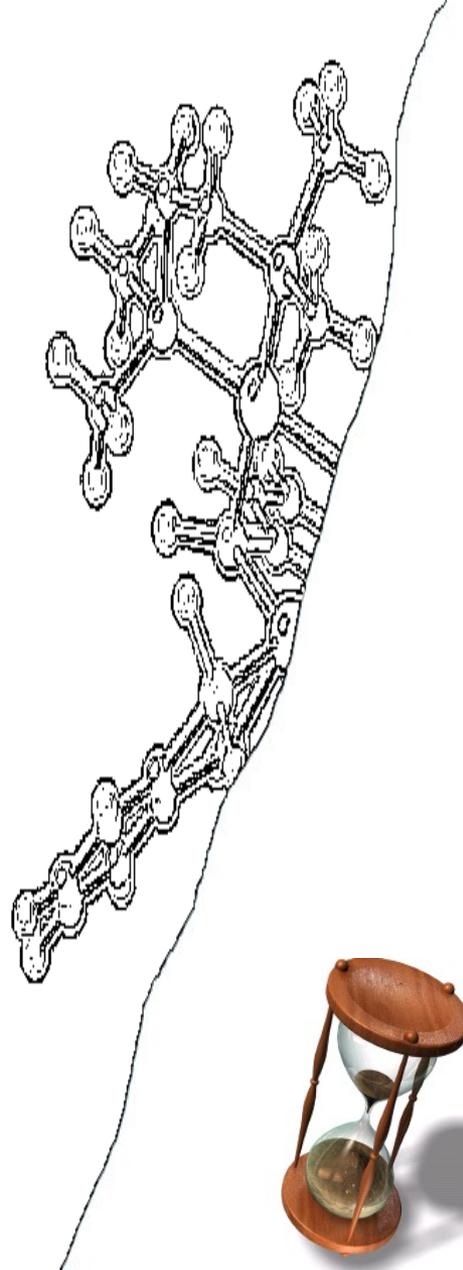




Consideraciones respecto a las colas

✓ Es muy conveniente ajustar la memoria que va a usar el caso a ejecutar:

- Mejor aprovechamiento del servidor
- Evita problemas en las aplicaciones
- Tener en cuenta las configuraciones de los servidores (**ratio memoria/procesador**) : FT no más de 8Gb por proceso



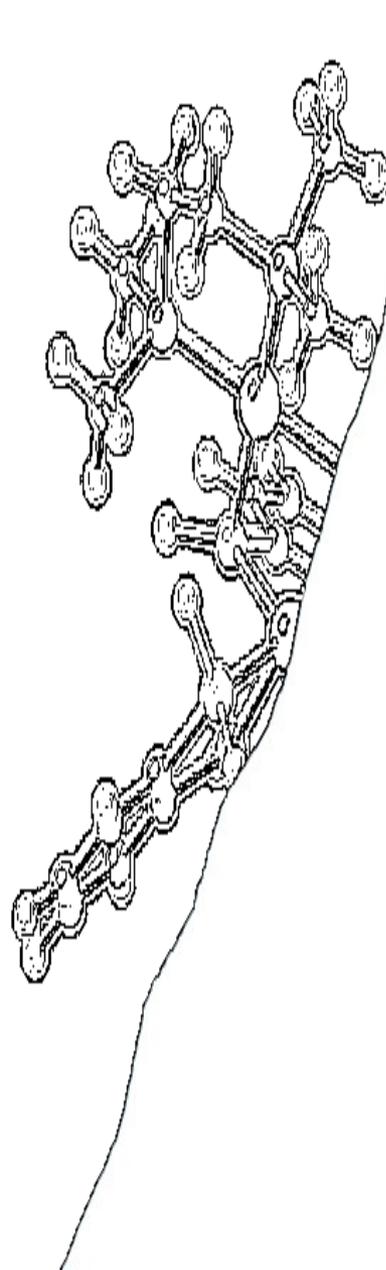


Uso óptimo del Hardware:

✓ El uso eficiente de la CPU y del CACHE depende enormemente de como fue compilado el programa:

→ Compiladores: gcc, Portland Group, Intel

→ Librerías: ATLAS, GOTO, ACML, Intel MKL ...





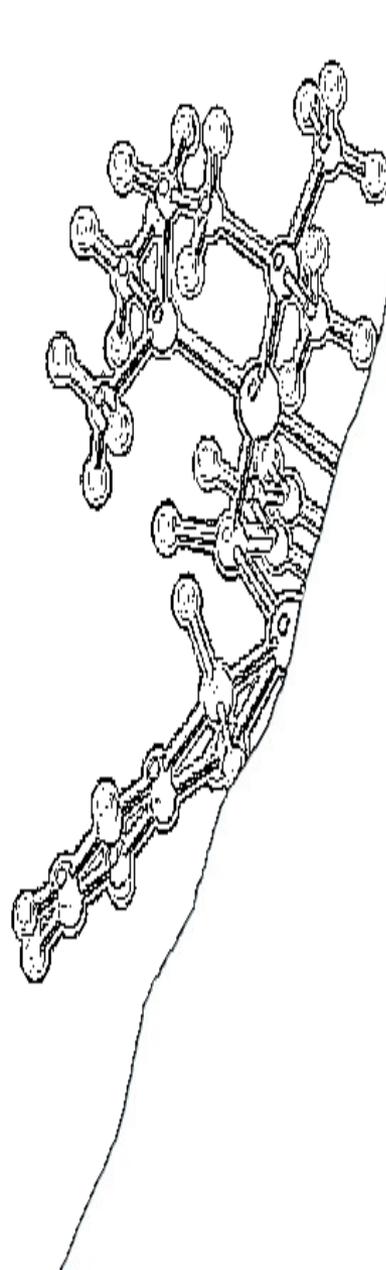
Uso óptimo del Hardware

✓ La mejor combinación de compilador y librería matemática dependerá de tu plataforma:

FT (ia64): Compiladores de INTEL + Intel MKL

✓ Compila y pasa los tests a el código tu mismo

✓ Mira si los vendedores o los desarrolladores ya conocen la mejor forma de compilación para tu plataforma



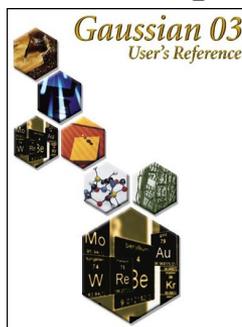


Gaussian 03:

Es un paquete de estructura electrónica capaz de predecir muchas propiedades de átomos, moléculas, sistemas reactivos, por ej:

- Energías moleculares
- Estructuras
- Frecuencias vibracionales
- Densidades electrónicas

Utilizando métodos ab initio, teoría de la funcional de la densidad, semi-empíricos, mecánica molecular y varios métodos híbridos



http://www.gaussian.com/tech_top_level.htm

http://www.gaussian.com/g_ur/m_eff.htm

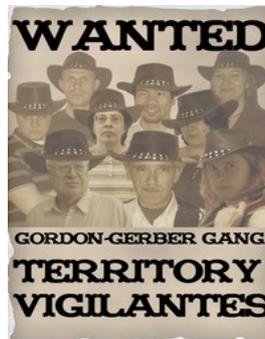
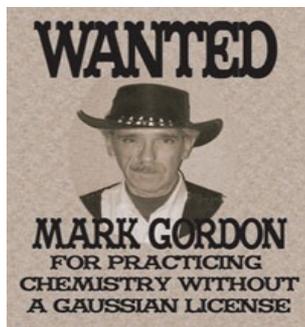


Games:

Es un programa para el modelado basado en la función de onda de la estructura electrónica de sistemas químicos de estructura electrónica.

GAMESS está mantenido por los miembros del “Mark Gordon's Quantum Theory Group”.

**General
Atomic and
Molecular
Electronic
Structure
System**



<http://www.msg.ameslab.gov/GAMESS/documentation.html>



GAMESS
QUANTUM THEORY GROUP



NWChem:

Paquete de estructura electrónica desarrollado en el Pacific Northwest National Laboratory y diseñado para obtener un gran rendimiento de su ejecución en paralelo

Presenta características bastante únicas como por ej:

Integración con python



<http://www.emsl.pnl.gov/docs/nwchem/doc/user/index.html>

Lista de correo:

<http://www.emsl.pnl.gov/docs/nwchem/support/support.html>



Gaussian/Gamess/NWChem

Comparativa [Gaussian 03](#) / [Gamess versión Sep. 7 2006 R4](#) / [NWChem 5.0](#) : Métodos y modelos implementados

- [Algoritmos fundamentales](#)
- [Energías](#)
- [Gradientes y optimizaciones de geometría](#)
- [Frecuencias y segundas derivadas](#)
- [Propiedades Moleculares](#)
- [Modelos de Solvatación](#)

Galego :: Español :: English centro de supercomputación de galicia

 buscar... Mapa Web Noticias Cursos Dixitos Empleo Concursos Contacto

Inicio » Descargas

Inicio
Cesga
Computación
Almacenamiento
Comunicaciones
e-learning
e-business
GIS
Software libre
Proyectos
Servicios
Descargas
FAQs

Buscar

Comparativa Gaussian 03 / Gamess versión Sep. 7 2006 R4 / NWChem 5.0 : Métodos y modelos implementados

Nombre: [DO_APL_CESGA_ProgramasQC_V1.pdf](#)

Tamaño: 77.42 KB

Descripción:

- Algoritmos fundamentales
- Energías
- Gradientes y optimizaciones de geometría
- Frecuencias y segundas derivadas
- Propiedades Moleculares
- Modelos de Solvatación

CESGA FINISTERRE
COMPUTATIONAL SCIENCE
CONFERENCE 2008

IBERGRID
May 12-14, 2008
Porto (Portugal)

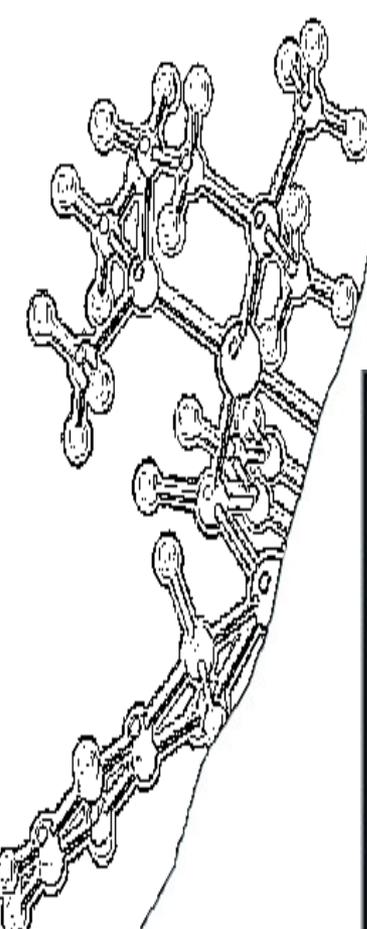
Registro

Nombre usuario

Contraseña



Gaussian/Gamess/NWChem



```

aurelio@fs001:~> module av g03

----- /opt/cesga/modules -----
g03/c1          g03/d2          g03/e1(default) g03/e1sfs
aurelio@fs001:~> module av gamess

----- /opt/cesga/modules -----
gamess/apr2008          gamess/mar2007(default)
aurelio@fs001:~> module av nwchem

----- /opt/cesga/modules -----
nwchem/5.1(default)
    
```



Cálculos HF o DFT



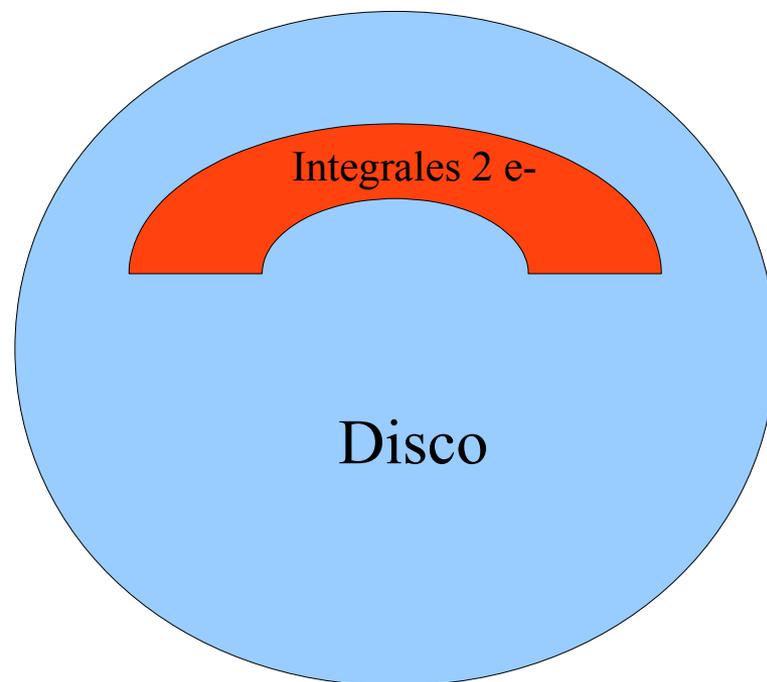
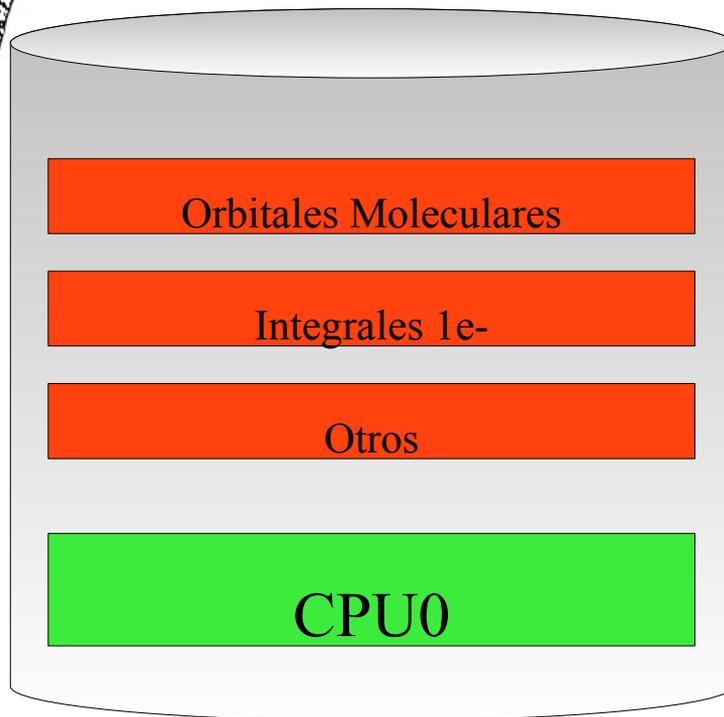
- ✓ Cálculos “in-core”
- o solo memoria:
 - Suficiente memoria para almacenar todo
 - Suele ser el método más rápido





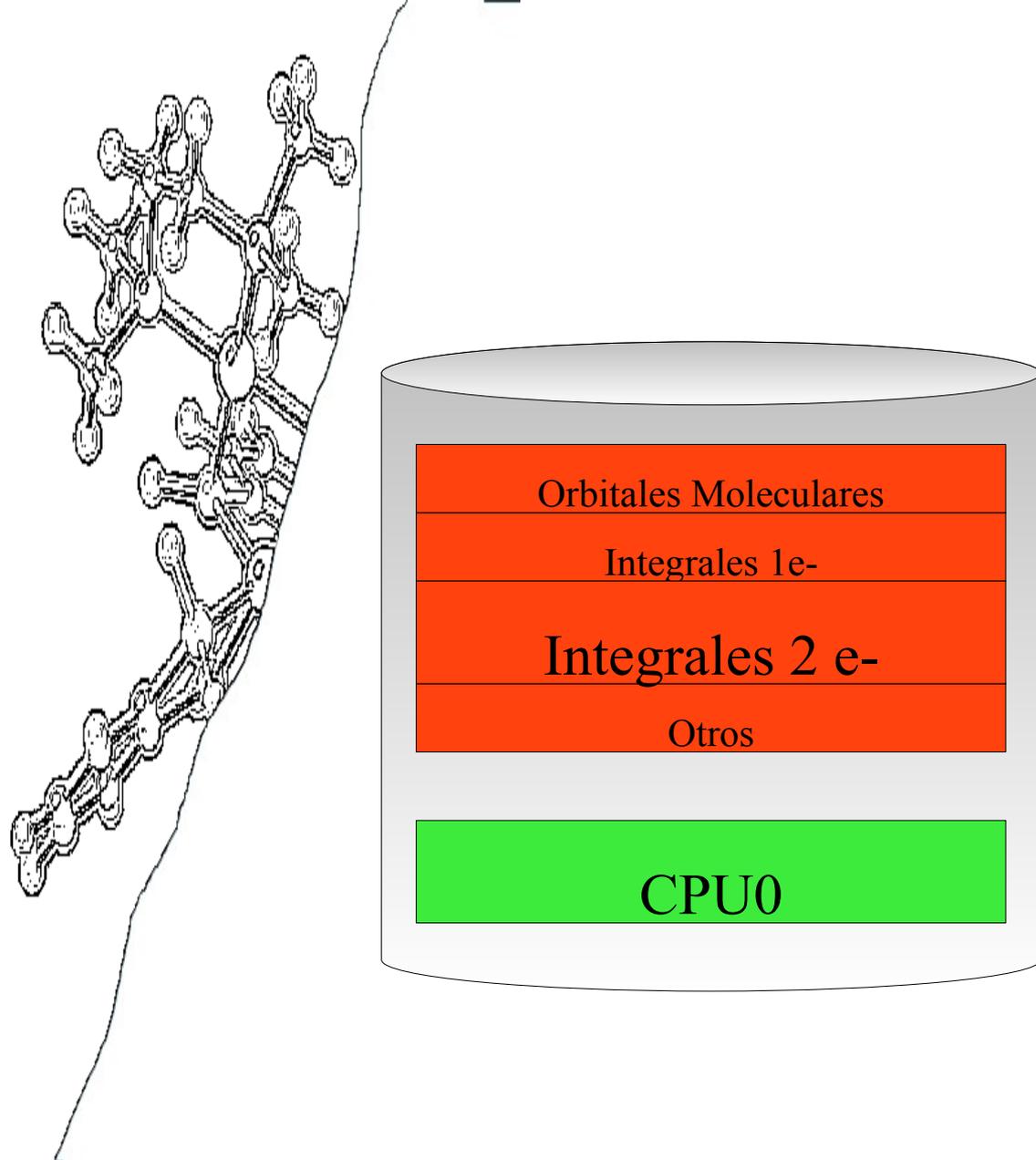
✓ Cálculos en disco:

→ Dependemos del rendimiento del sistema de ficheros





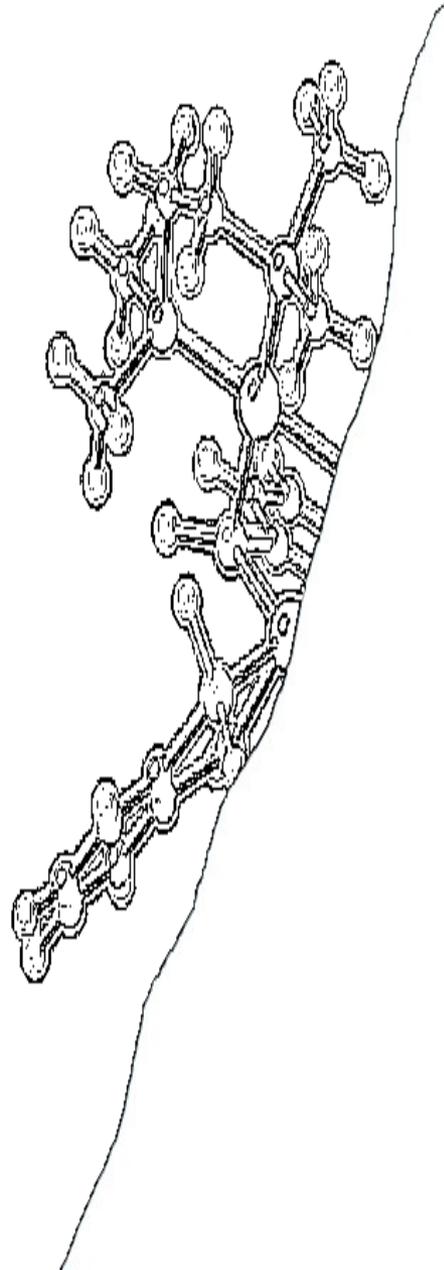
- ✓ Cálculos Directos:
- Cada vez que necesitamos algunas integrales 2 e- las recalculamos
 - Requiere mucha memoria





Directo vs. Incore vs. Disco

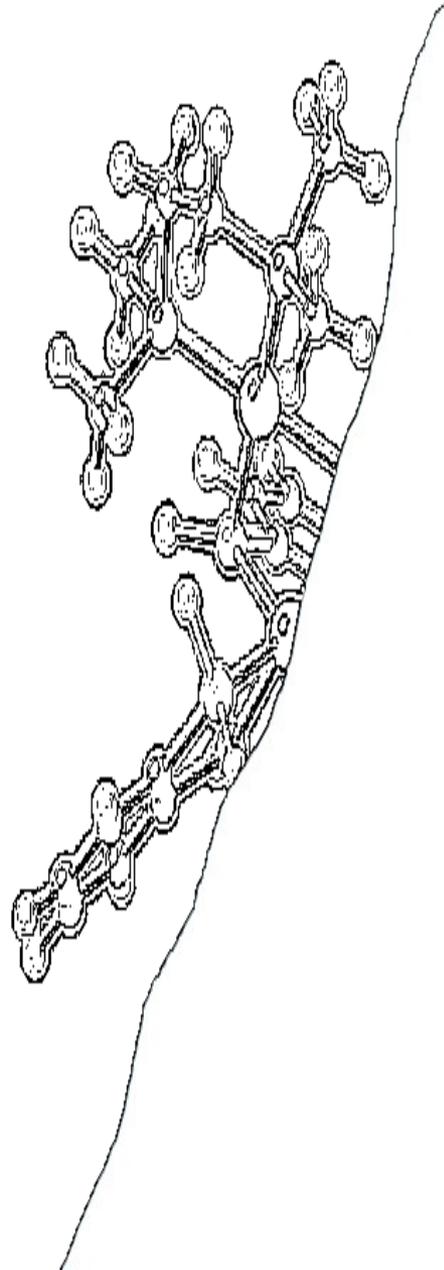
- ✓ En serie los cálculos in-core son los más rápidos pero frecuentemente requieren cantidades prohibitivas de memoria
- ✓ Son más frecuentes los cálculos directo o con almacenamiento en disco (semidirectos).





Directo vs. Incore vs. Disco

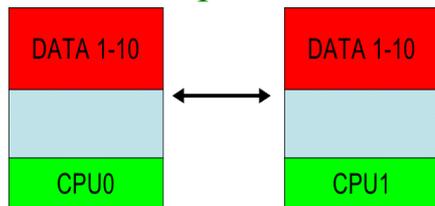
- ✓ Almacenar las integrales en disco requieren menos tiempo de CPU que los métodos directos pero para conseguir un buen tiempo final se necesitan discos muy rápidos.
- ✓ Más frecuentemente los cálculos directos o semidirectos dan mejor resultados comparado con almacenamiento completo en disco.



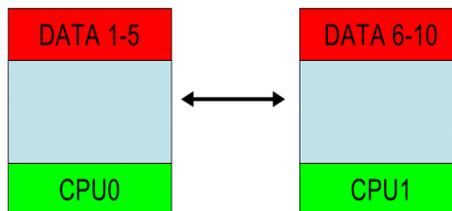
Uso de la memoria en programas paralelos

- ✓ Un programa paralelo tiene diferentes formas de distribuir el trabajo
- ✓ Dependiendo de la plataforma un método u otro de distribución funcionará mejor o peor

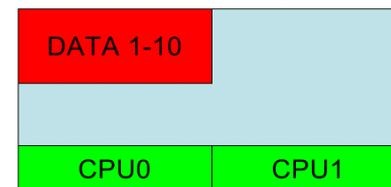
Datos replicados



Datos distribuidos



Memoria compartida





✓ Gaussian:

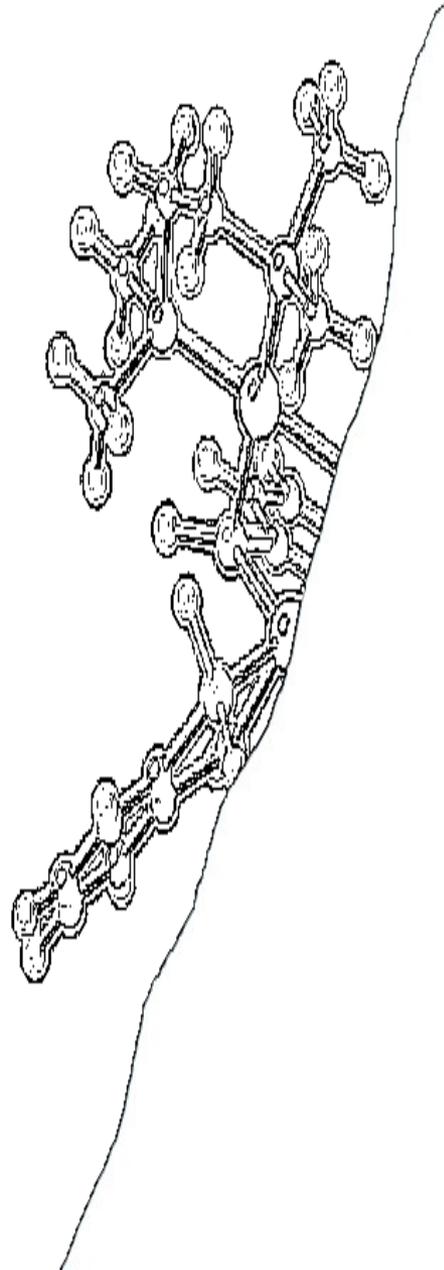
→ Memoria compartida (OpenMP)

✓ Gamess:

→ Datos distribuidos o replicados usando el “GAMESS distributed data interface (DDI)”

✓ NWChem:

→ Datos distribuidos o replicados usando el Global Arrays





Consideraciones acerca del método HF/DFT

✓ Los métodos HF y DFT involucran un único cálculo SCF para hallar la función de onda.

✓ En estos métodos el algoritmo SCF es muy importante:

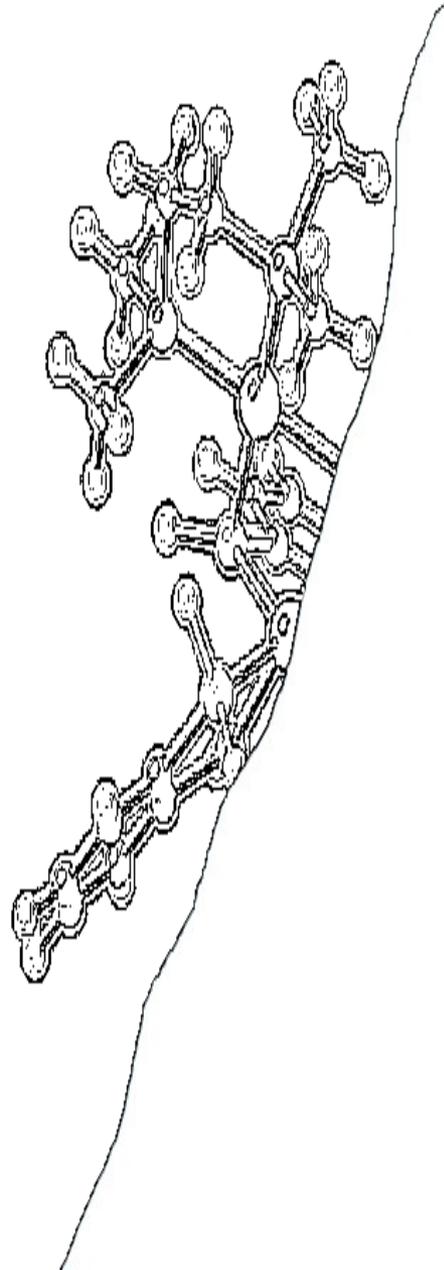
→ En Gaussian, `scf=xqc` suele ser la mejor opción. Tanto Gamess como Gaussian o NWChem tienen muchas opciones para los algoritmos SCF y “guesses” iniciales.





HF/DFT paralelo

- ✓ Gaussian:
 - Buen escalado hasta 4-8 procesadores
- ✓ Gamess:
 - Buen escalado hasta 8-16 procesadores
- ✓ NWChem:
 - Buen escalado hasta 256 procesadores





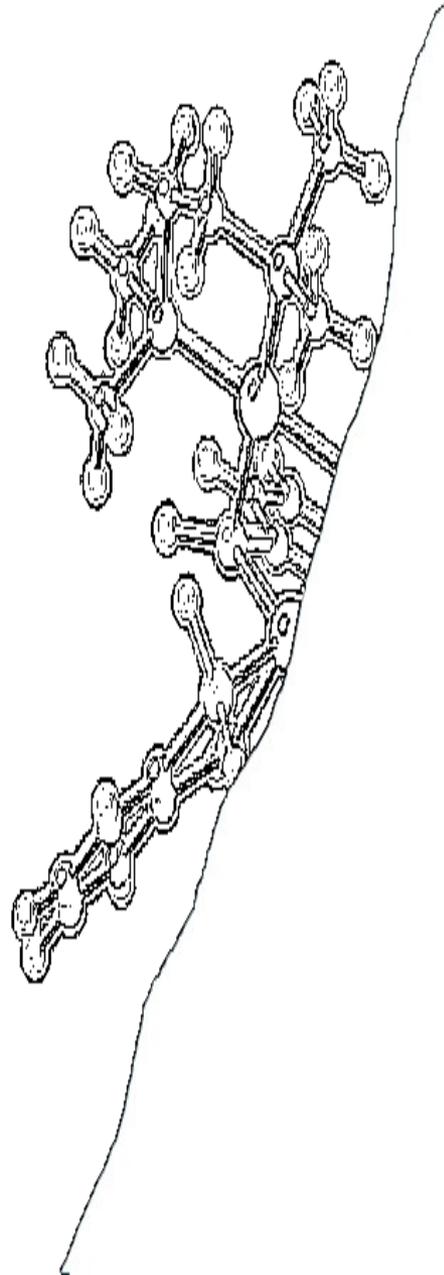
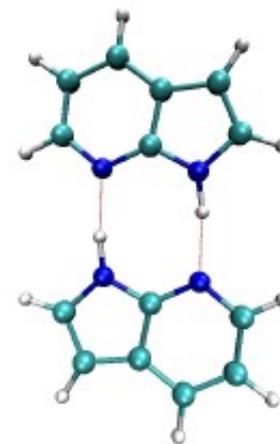
Ejercicio con NWChem:

Ejecuta NWChem en el Finis Terrae:

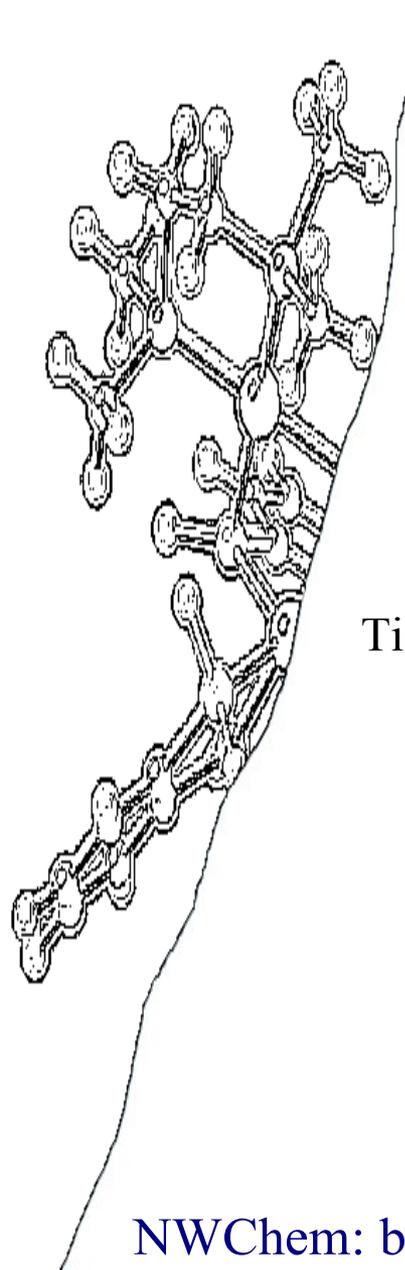
- 1) Ejecuta en serie
- 2) Ejecuta en paralelo con 2,4,8 y 16 procesadores

¿Cuanto tarda cada caso?

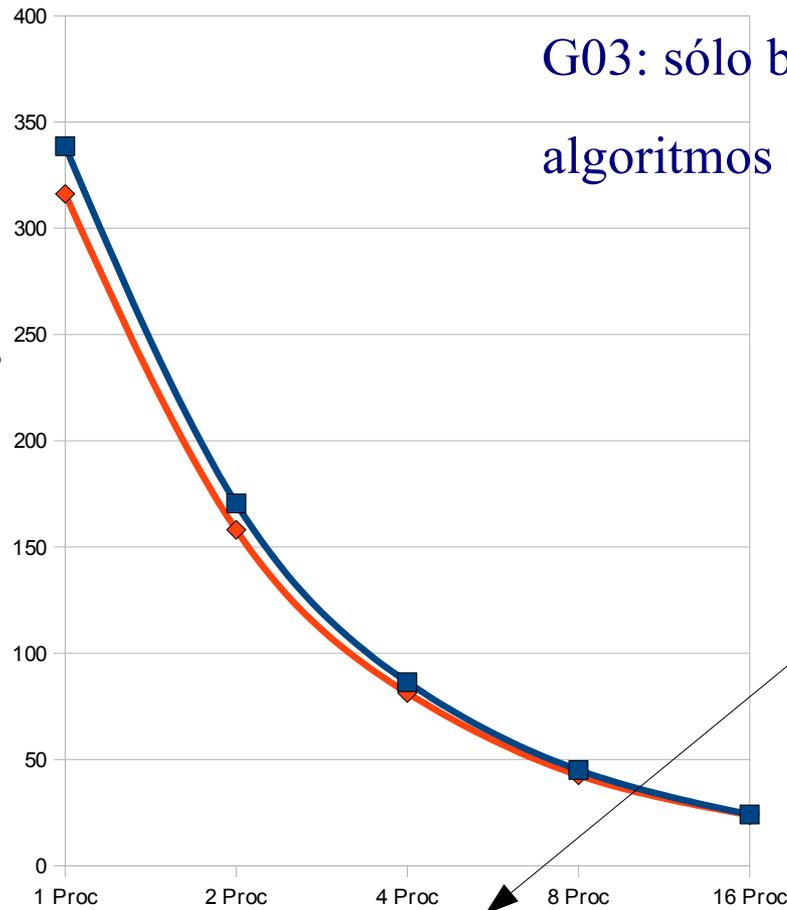
¿Cómo escala este programa?



Ejercicio con NWChem:



Tiempo/s



G03: sólo buen escalado con algoritmos directos

◆ G03
■ NWChem



NWChem: buen escalado en general pero no es recomendable para casos pequeños

Consideraciones acerca del método MP2

✓ Gaussian:

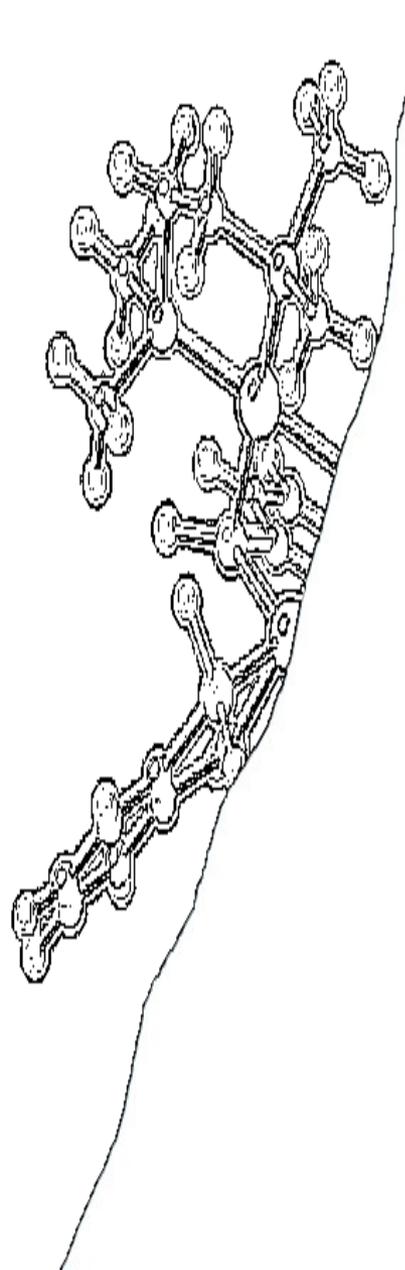
→ Buen escalado hasta 4 procesadores para energías y gradientes. Los gradientes analíticos se benefician mucho de una gran cantidad de memoria

✓ Gamess:

→ Buen escalado hasta 4 procesadores para energías y gradientes. Rendimiento muy dependiente de la memoria disponible

✓ NWChem:

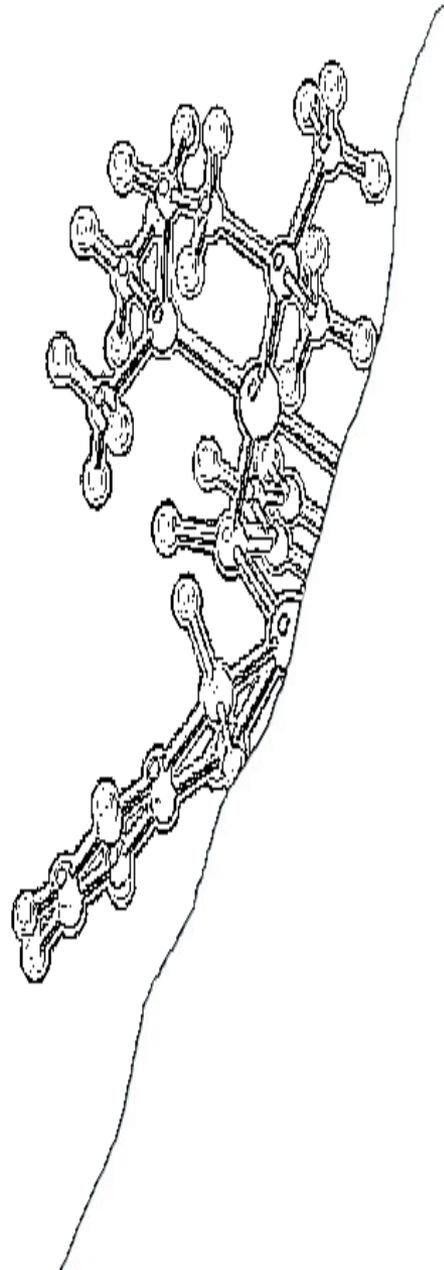
→ Buen escalado hasta 128 procesadores para energías y gradientes. Se debe chequear el fichero de output acerca de las estadísticas de memoria





Consideraciones acerca de los métodos QCISD y CCSD

- ✓ Gaussian: no está paralelizado pero... (librerías matemáticas). Los gradientes analíticos requieren un gran cantidad de disco
- ✓ Gamess: paralelizado para energías.
- ✓ NWChem: paralelizado para energías. Gradientes numéricos.



CCSD(T) paralelo

- ✓ Gaussian: no está paralelizado pero... (librerías matemáticas).
- ✓ Gamess: paralelizado para energías
- ✓ NWChem: paralelizado para energías



Interacción de configuraciones

- ✓ Gaussian: energías y gradientes paralelos
- ✓ Gamess: paralelizado para energías
- ✓ NWChem: paralelizado para energías



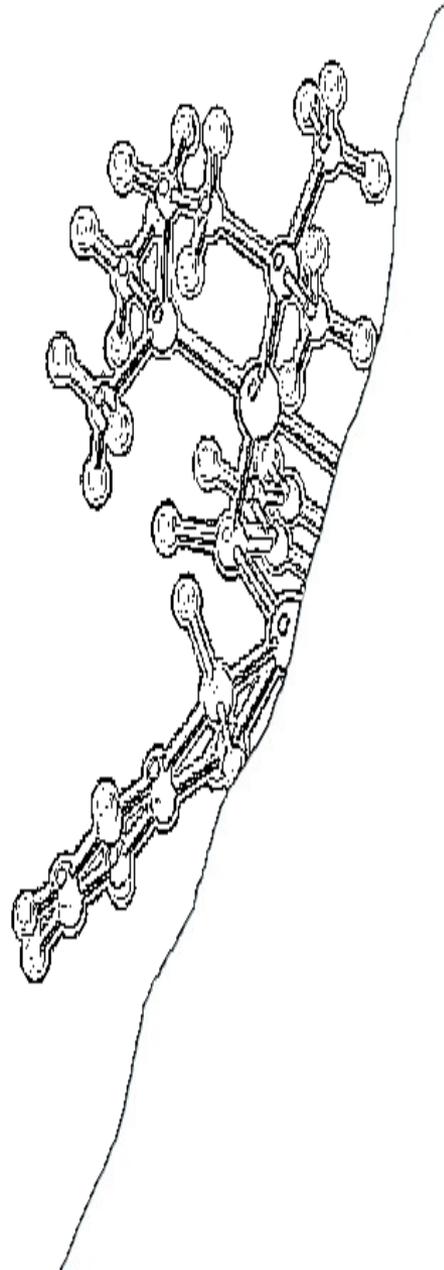
Optimización de Geometrías

✓ Minimizar el número de pasos es ideal para acelerar el proceso:

- Por ej el algoritmo de Berny en Gaussian funciona bastante bien

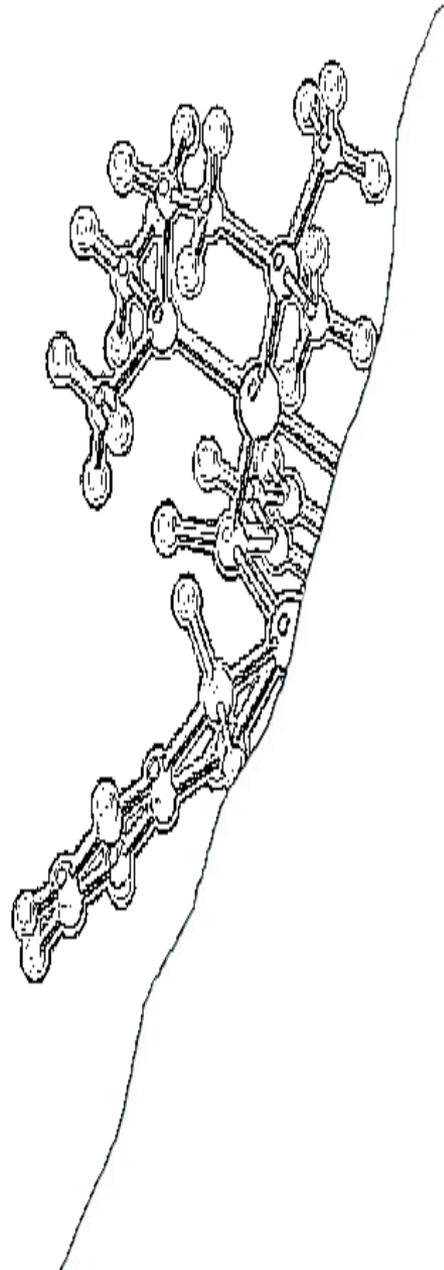
✓ El escalado paralelo está limitado por la paralelización de los gradientes y a veces de los hessianos.

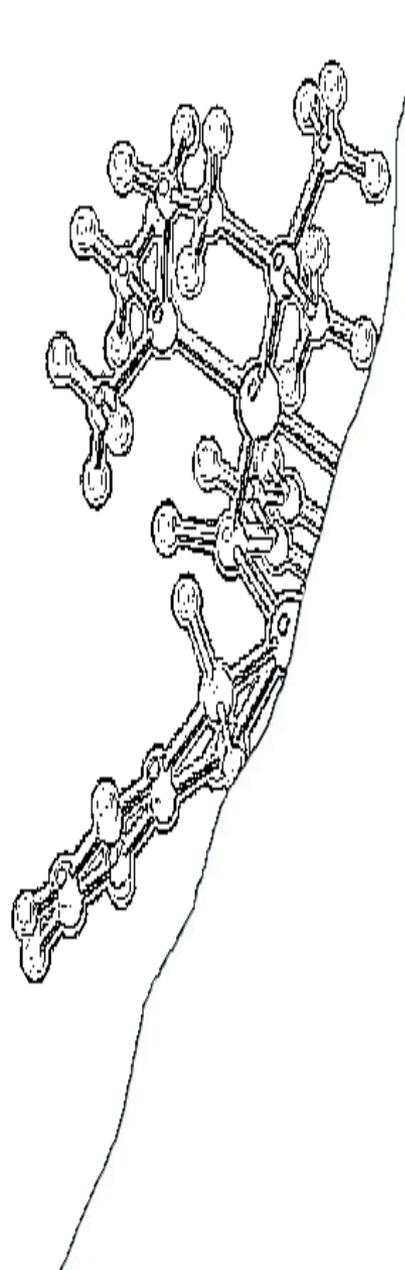
- Por ej en Gaussian `opt=calcFC` o `opt=calcall` hace que se calcule el hessiano en 1 o más pasos. (optimización de puntos de silla)



Generalidades/FAQ

- ✓ ¿Que servidor debo usar?
- ✓ ¿Debo usar las colas? ¿Qué recursos necesita mi trabajo?
- ✓ Checkpointing/Restart
- ✓ ¿Debo usar mpirun para ejecutar en paralelo?
- ✓ Otros programas...





✓ ¿Que servido debo usar?

→ **SVGD**

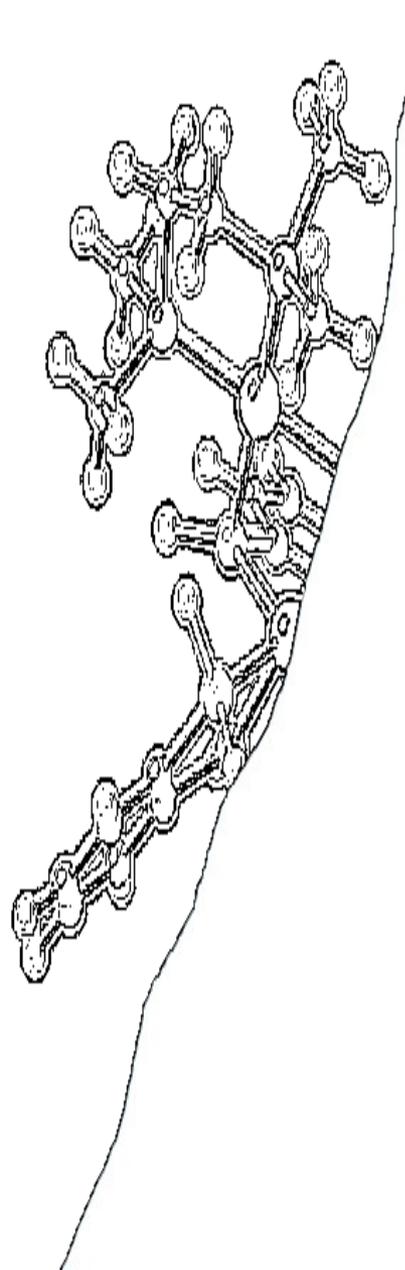
→ **Finis Terrae**

Casos que demanden:

Poca memoria

Poco disco

Adecuado para cálculos directos o pequeños



✓ ¿Que servido debo usar?

→ SVGD

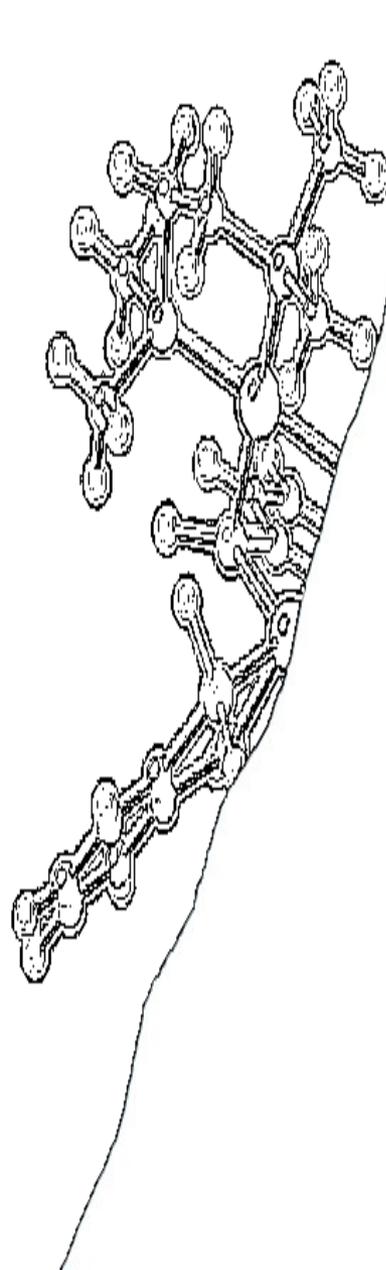
→ Finis Terrae

Casos que demanden:

Mucha memoria

Mucho disco

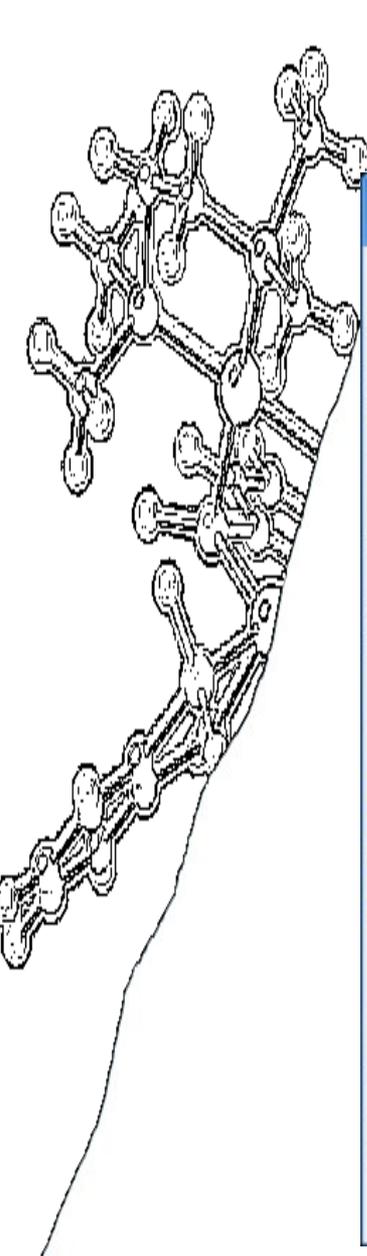
Adecuado para grandes cálculos



Method	Formal CPU	Formal Memory	Formal Disk	Actual CPU	Actual Disk
Conv. SCF	N^4	N^2	N^4	$N^{3.5}$	$N^{3.5}$
Incore SCF	N^4	N^4	-	N^4	N^2
Direct SCF	N^4	N^2	-	$N^{2.3}$	N^2
Conv. MP2	ON^4	N^2	N^4	ON^4	N^4
Dir MP2 SP	ON^4	OVN	-	O^2N^3	N^2
SD MP2 SP	ON^4	N^2	VN^2	O^2N^3	VN^2
Conv. MP2 Force	ON^4	N^2	N^4	ON^4	N^4
Dir MP2 Force	ON^4	N^3	-	O^2N^3	N^2
SD MP2 Force	ON^4	N^2	N^3	O^2N^3	N^3
MP3, CISD, QCISD	O^2N^4	N^2	N^4	O^2N^4	N^4
MP4, QCISD(T)	O^3V^4	N^2	N^4	O^3V^4	N^4

O: Número de orbitales ocupados V: Número de orbitales virtuales N: Número de funciones de base

g03mem fichero_de_input



```
aurelio@localhost:~ - Terminal - Konsole <2>
Sesión Editar Vista Marcadores Preferencias Ayuda
-----GAUSSIANO3 RESOURCES ESTIMATION-----CESGA-----
NPROC:  %nprocshared=1
ROUTE:
#p ump2/6-311+g(2df,p) test opt

NBasis:  128 basis functions

Minimum memory requirements:  F  6MW (%mem=7MW)
Method:  MP2

Conv.      CPU      Memory      Disk      Should Run On...
1864136    (%mem=8MW)  (MaxDisk=257MW)  SVG
SemiDir    364090      (%mem=8MW)  (MaxDisk=3MW)   SVG
Direct     364090      (%mem=8MW)  (MaxDisk=1MW)   SVG
InCore     (%mem=264MW)  HPC320/Superdome

/opt/cesga/g03/tests/com/test567.com link1-0000.com
-beta 0.1 aplicaciones@cesga.es-----CESGA-----
press enter to continue...
█
```

Restart/Checkpointing

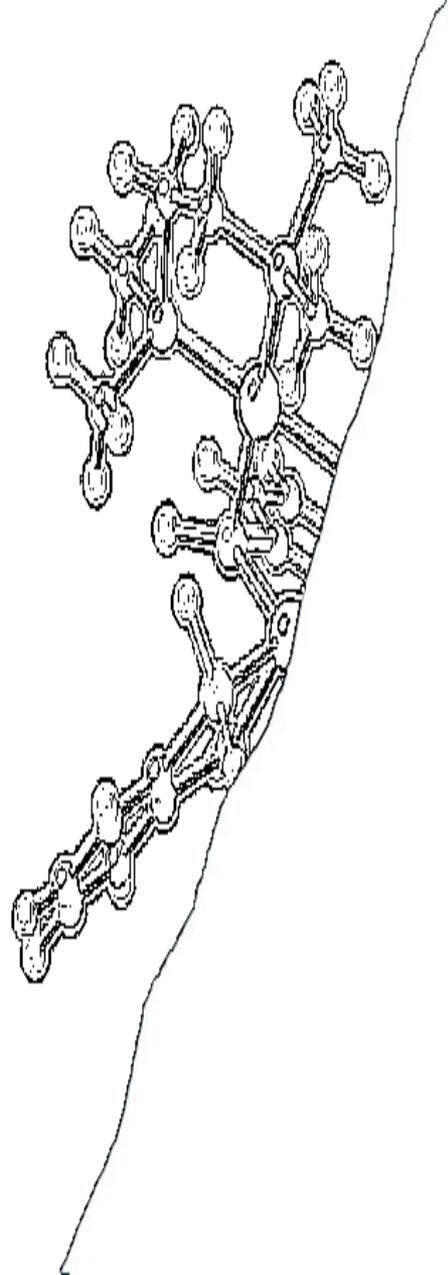
✓ Depende enormemente del caso:

→ Ficheros de checkpointing

→ Otros ficheros

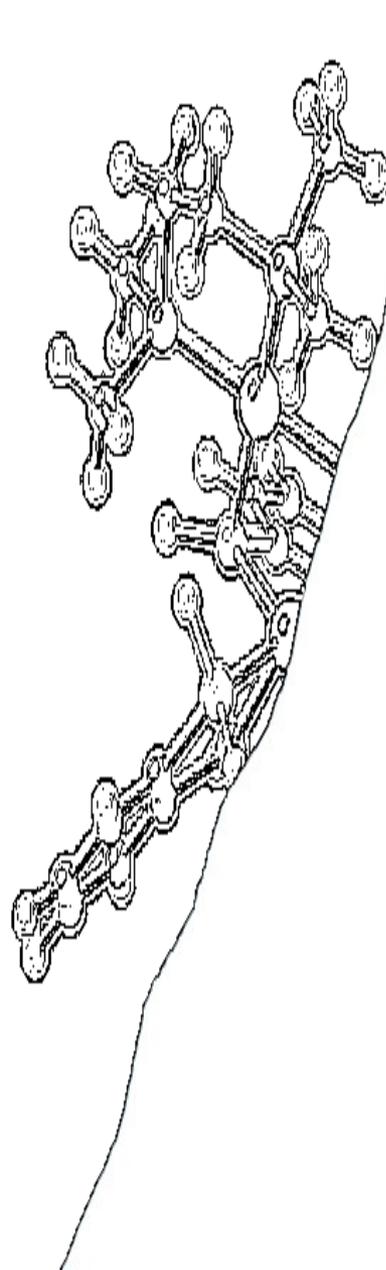
→ Database

→ No es posible



¿Debo usar mpirun para ejecutar en paralelo?

- ✓ Depende del programa: Memoria compartida o distribuida
- ✓ Dentro del CESGA en general no, ya que hay “wrappers” que se encargan de ello en el caso de ser necesario



Otros programas...



JAGUAR



COLUMBUS



DALTON



MOLCAS



MOLPRO

DALTON

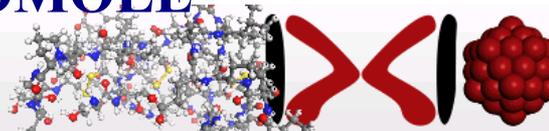


ORCA



TURBOMOLE

TURBOMOLE GmbH



Algunos programas más ...

aceticacid_g98.log - Infrared spectrum

Property	Value
File	
Atom	0
Number of Atoms	1
Bonds	1
Atom Bonds	0
Isolated	0
Coordinates	1
Energy	-14332.07648
Molar Mass	60.05136
Exact Mass	60.03129
Charge	0
Spin Multiplicity	1
Triplet	Yes
Radical	Yes
Preoptimized	Yes
Quick Printout	Yes

MOLEKEL

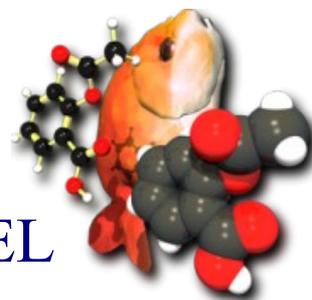
File Name: C2h2Big.log File Type: Gas
Atomic Orb.: Nothing Grid: Ok

Grid: -0.020 -0.010

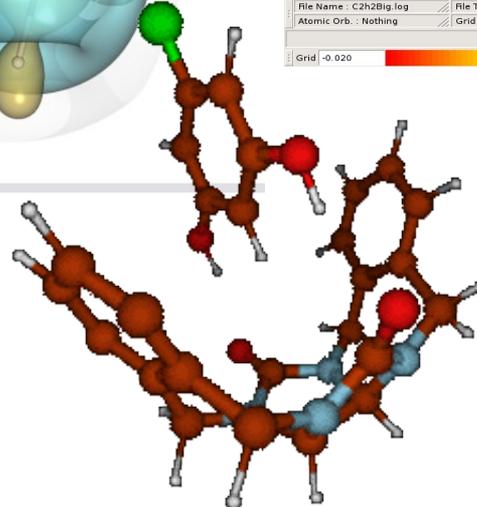
Fragment Selector

- Tyr
- Asp
- Gln
- His
- Leu
- Pro
- Trp
- Asn
- Glu

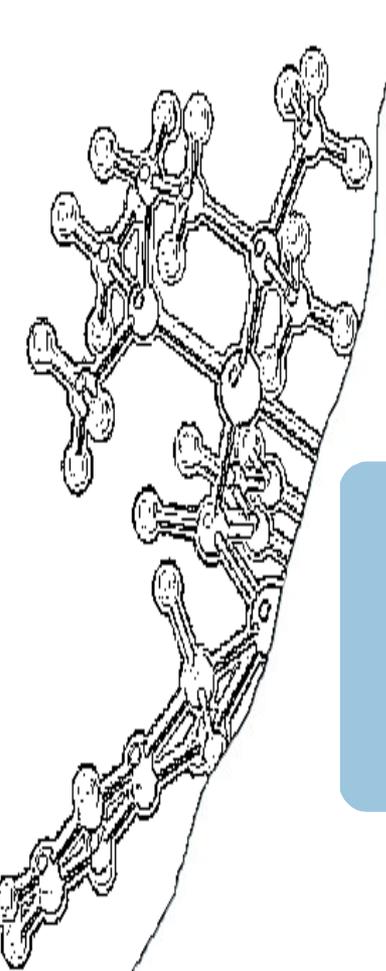
GABEDIT



OPENBABEL



MOLDEN



**GRACIAS POR VENIR A
ESTE CURSO!**



CONTACTO:

Aurelio Rodríguez

aurelio@cesga.es

<http://www.cesga.es>

