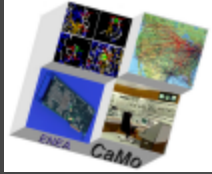


Modelli e Architetture



(1) Inevitabilità epistemologica del computer

(2) Modelli fisici e modelli numerici

(3) Modelli numerici e architetture di calcolo

Inevitabilità del computer (1)

La scienza negli ultimi secoli ha lavorato in un'ottica riduzionista ("de-construire") sotto ipotesi quali:

- (1) Scomponibilità della realtà in "effetti" elementari*
- (2) Sovrapposizione degli effetti*

*Questa destrutturizzazione ha consentito di ideare modelli esplicativi e predittivi di ciascun ""effetto"" (o di molti di essi). Tale capacità costituisce **la misura della conoscenza** dei fenomeni.*

Inevitabilità del computer (2)

Modello

Insieme di assunzioni ipotetiche sulla struttura non osservabile di un sistema, stabilite a partire dalle loro proprietà empiriche.

(Vocabolario Treccani)

Inevitabilità del computer (3)

I modelli numerici costituiscono la forma più alta (nel senso estetico del termine) di "sistema esperto".

(1) Sintetico : racchiude con un'alta densità informativa tutte le proprietà del fenomeno che rappresenta.

(2) Esplicativo: consente di accedere ai meccanismi che generano gli effetti a partire dalle cause.

(3) Predittivo: consente di valutare il comportamento in condizioni complesse o fittizie ("gedanken experiment")

Inevitabilità del computer (4)

*La scienza dei prossimi anni effettuerà un'operazione "inversa" rispetto al "de-costruttivismo; essa cercherà di **ri-costruire** la realtà attraverso la composizione dei modelli elementari delle sue singole componenti.*

I modelli elementari possono anche essere semplicemente la rappresentazione della realtà su piccole scale (spazio-temporali) e, dunque, essere in grado di generare il comportamento del sistema reale anche su scale maggiori (caso del protein folding).

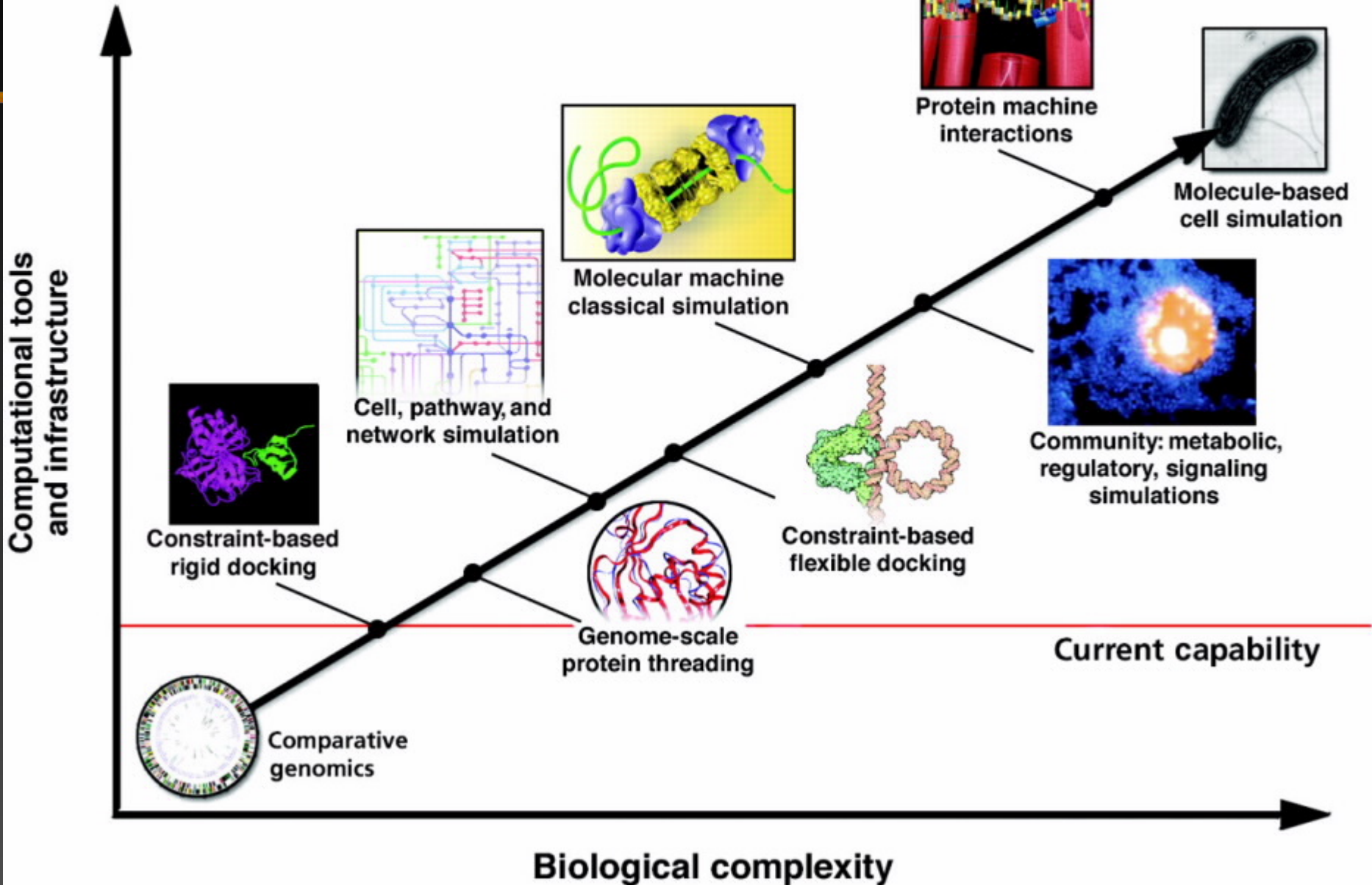
Inevitabilità del computer (5)

Il computer è lo strumento essenziale per operare questa operazione di ri-costruzione della realtà a partire dalle sue componenti elementari.

I sistemi di interesse (dall'infinitamente piccolo -proteine- all'infinitamente grande -clima, spazio -) non consentono rappresentazioni "analogiche" (nella maggior parte vi sono leggi di scala che lo impediscono); in questo senso non esistono che modelli numerici di questi sistemi tali che su di essi possano essere eseguite le stesse "operazioni di misura delle quantità osservabili" e i risultati comparati direttamente con le osservazioni sui sistemi reali.

Genomes to Life Computing

Frazier et al., Science **300** (2003) 290



Esempio di modello (1)

Potenziale di interazione tra due atomi di un sistema metallico:

$$V(r_{ij}) = A_{ij} \exp(-p(r_{ij}/r_0 - 1)) - \frac{C_{ij}}{r_{ij}^2} \exp(-q(r_{ij}/r_0 - 1))$$

F. Cleri, V. Rosato, Phys. Rev. B **48** (1993) 22

Esempio di modello (2)

Esempio di proteoma (*Bacillus halodurans*)

...

>gi|15612564|ref|NP_240867.1| initiation of chromosome replication [*Bacillus halodurans*]
MENIHDLWERALKSMEKKVSKPSFETWLKQTKANSIEDSTIIITAPNEFARDWLEKHYDELISETIDDLT
GVRLYPKFVIPTSQLDEPFVEQELKKPMKQPPAQNGEMPNNMLNDKYTFDTFVIGSGNRFAHAASLAVAE
APAKAYNPLFIYGGVGLGKTHLMHAIGHYVMDHNPNAKVVYLSSEKFTNEFINAIRDNKAVNFRNKYRNV
DVLLIDDIQFLAGKEQTQEEFFHTFNALHEDNKQIVISSDRPPKEIPTLEDRLRSRFEWGLITDITPPDL
ETRIAILRKKAKAENLDIPNEVMLYIANQIDTNIRELEGALIRVVAYSSLINQDMNADLAAEALKDIIPN
AKPRVLTITDIQKTVGEYFHVKLEDFKAKKRTKSVAFPRQIAMYLSRELTDASLPKIGSEFGGRDHTTVI
HAHEKISKLLSTDQELQDKIQDISDKLRS

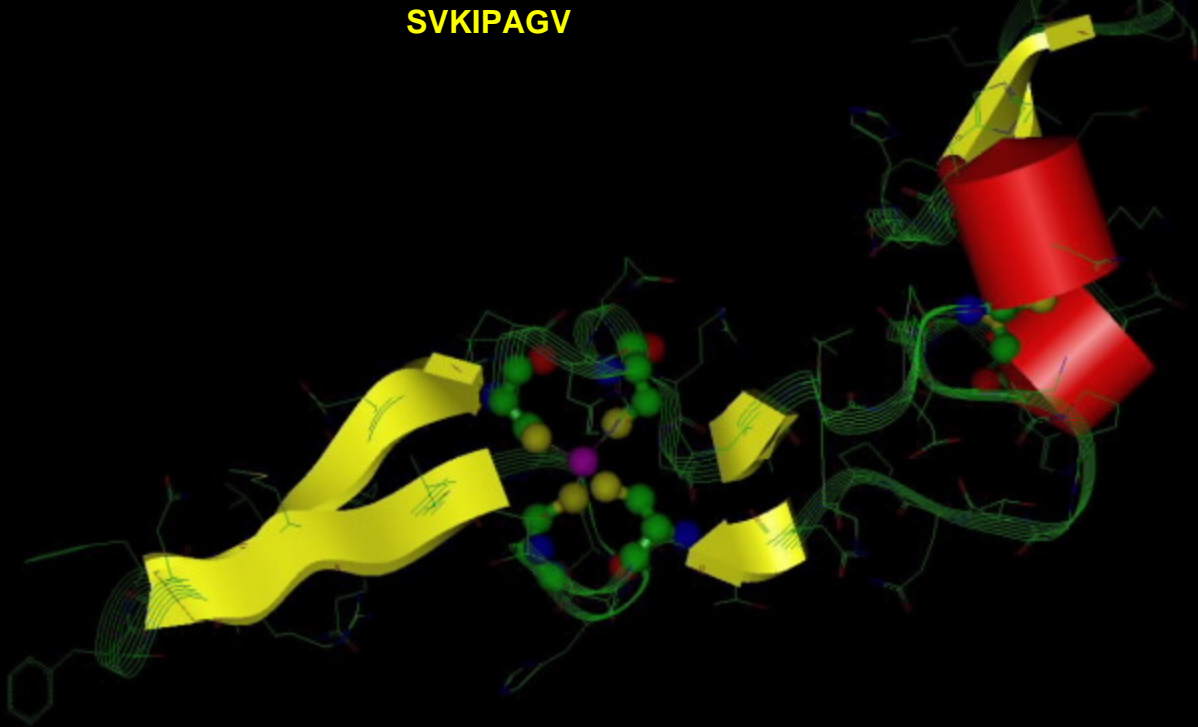
>gi|15612565|ref|NP_240868.1| DNA polymerase III (beta subunit) [*Bacillus halodurans*]
MHFVIDRDIFVQNVNHVSKAVSSRTTIPILTGIKIVADHEGVTLTGSDSDISIETFIPLLEEGDRQNVEVK
QEGSIVLQAKVFAEIVKKLPEQEIEIHVQDSFVTTIRSGSSVFNLNGLDPDEYPRLPVLEEDHVFRLPQK
ILKDIIRQTVFAVSTQETRPVLTGVNFEIEDGILTCTATDSHRLAMRKVPVEKNDDELQFSNVVIPGKSL
NELSKILDENEELLDIVVTENQTLFKLKNMLFFSRLLEGKYPVTKNMIPKEAKTSFAVHTKAFLQTLERA
LLLSREGKNQVINLKTLDGQVVEVTAITPEIGKVTENVATQGLEGEELRISFNGKNVIDALKVVDSSESIH
IAFTGAMSPFVLSPTDHDQSLHLFSPVRTY

.....

I modelli sono "collegabili" nello spazio dei "gradi di libertà"

Un'altra attività di enorme rilievo scientifico è quella di passare dalla sequenza degli aminoacidi alla forma tridimensionale della proteina (che dice molto sulla sua funzione)

MGKIDGKRAARGADLRYNMELTLEEAVRGVTKEIRIPTLEECDVCHGSGAK
PGTQPQT
CPTCHGSGQVQMRQGFFAVQQTCPHCQGRGTLIKDPCNKCHGHGRVERSKTL
SVKIPAGV



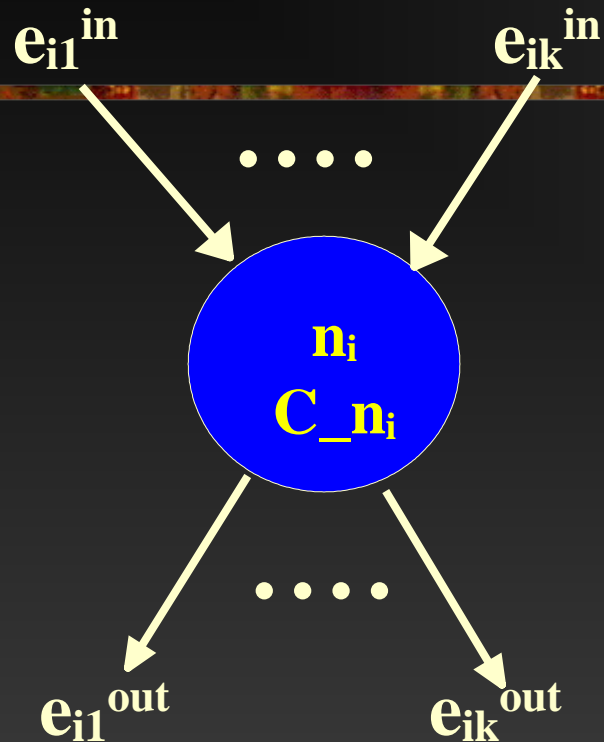
Modelli e architetture di calcolo (1)

La simulazione dei I modelli consiste, in genere, nella predizione di

- (1) uno stato "stazionario" a partire da una qualunque condizione iniziale***
- (2) Uno stato dinamico di equilibrio sul quale effettuare una serie di misure (medie temporali)***
- (3) Uno stato dinamico transitorio***
- (4) Uno stato frutto della ottimizzazione di una qualche funzione di costo***
- (5) Un path che collega uno o più stati delle variabili***

In genere i modelli sono espressi da equazioni differenziali (nello spazio, nel tempo, in entrambi) successivamente discretizzati oppure legati alla determinazione di funzioni (o funzionali) incogniti sulla base di processi di ottimizzazione vincolata.

Un algoritmo puo' essere rappresentato da un flusso di dati organizzati gerarchicamente I cui elementi n_i sono caratterizzati da una complessità C_{n_i} e da un set di comunicazioni I/O $e_{ij}^{in,out}$

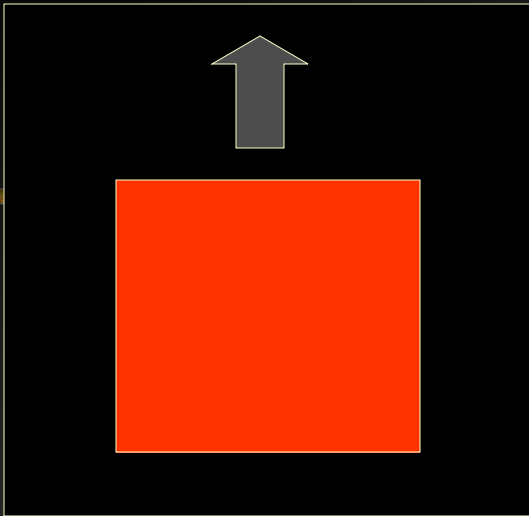


$$g_t = \frac{\text{\# computation operations}}{\text{\# bytes of I/O data}}$$

Una piattaforma computazionale é
descrivibile come una (o un
insieme) di macchine caratterizzate
da differenti proprietà quali

$$g_m = \frac{\text{node peak comput. speed}}{\text{communication bandwidth}}$$

Una singola piattaforma può essere vista a scale
diverse come caratterizzata da diversi valori di g_m



Definizione di Granularità

$$G = \frac{g_t}{g_m}$$

G misura l'influenza dell'overhead delle comunicazioni sulle performance complessive del sistema.

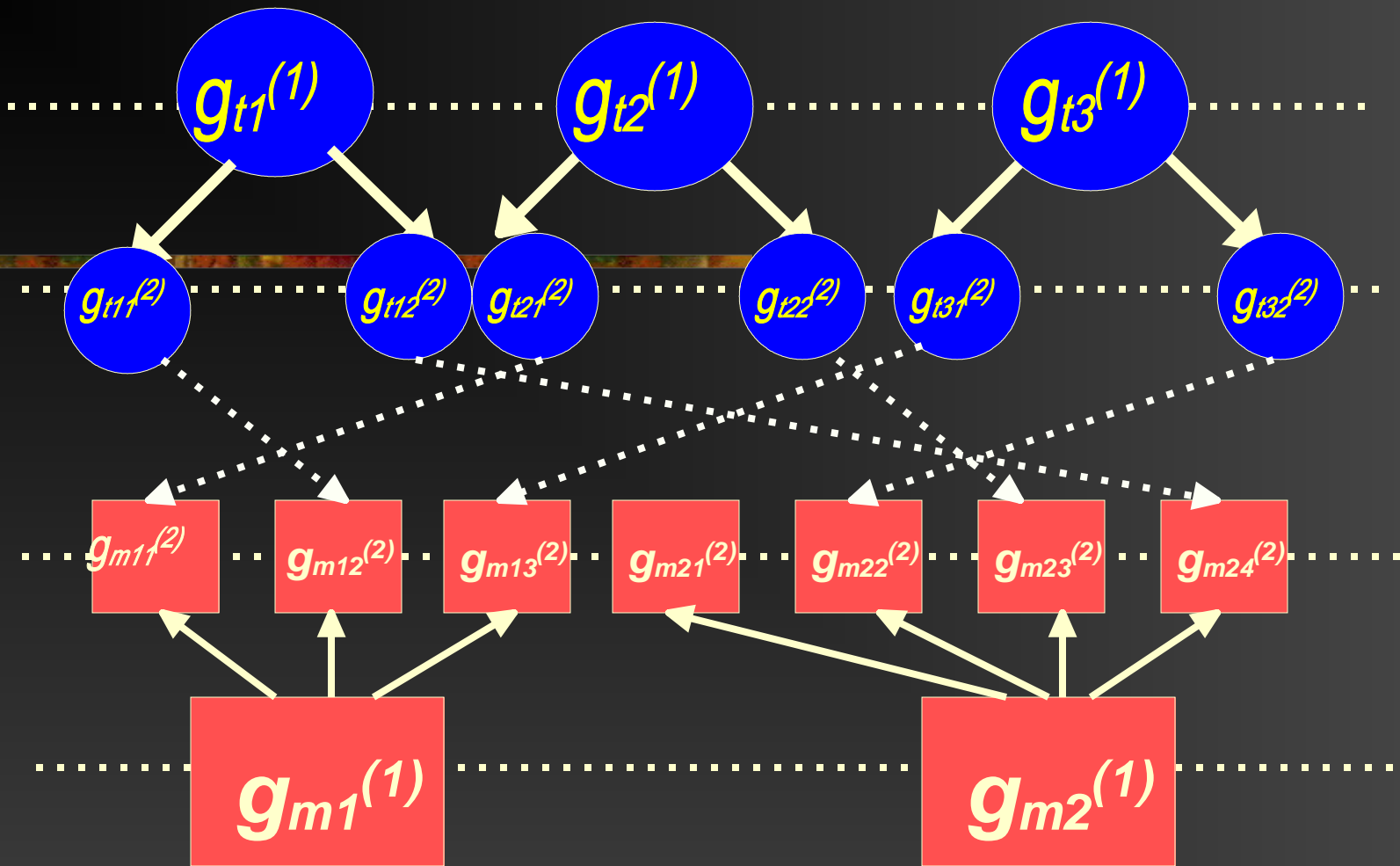
$$Eff = \frac{\text{Computations execution time}}{\text{Total time}}$$

$$Eff = \frac{1}{1 + \frac{1}{G}} = \text{Max} \left[1, \frac{1}{G} \right]$$

$$g_t > \frac{1}{g_m}$$

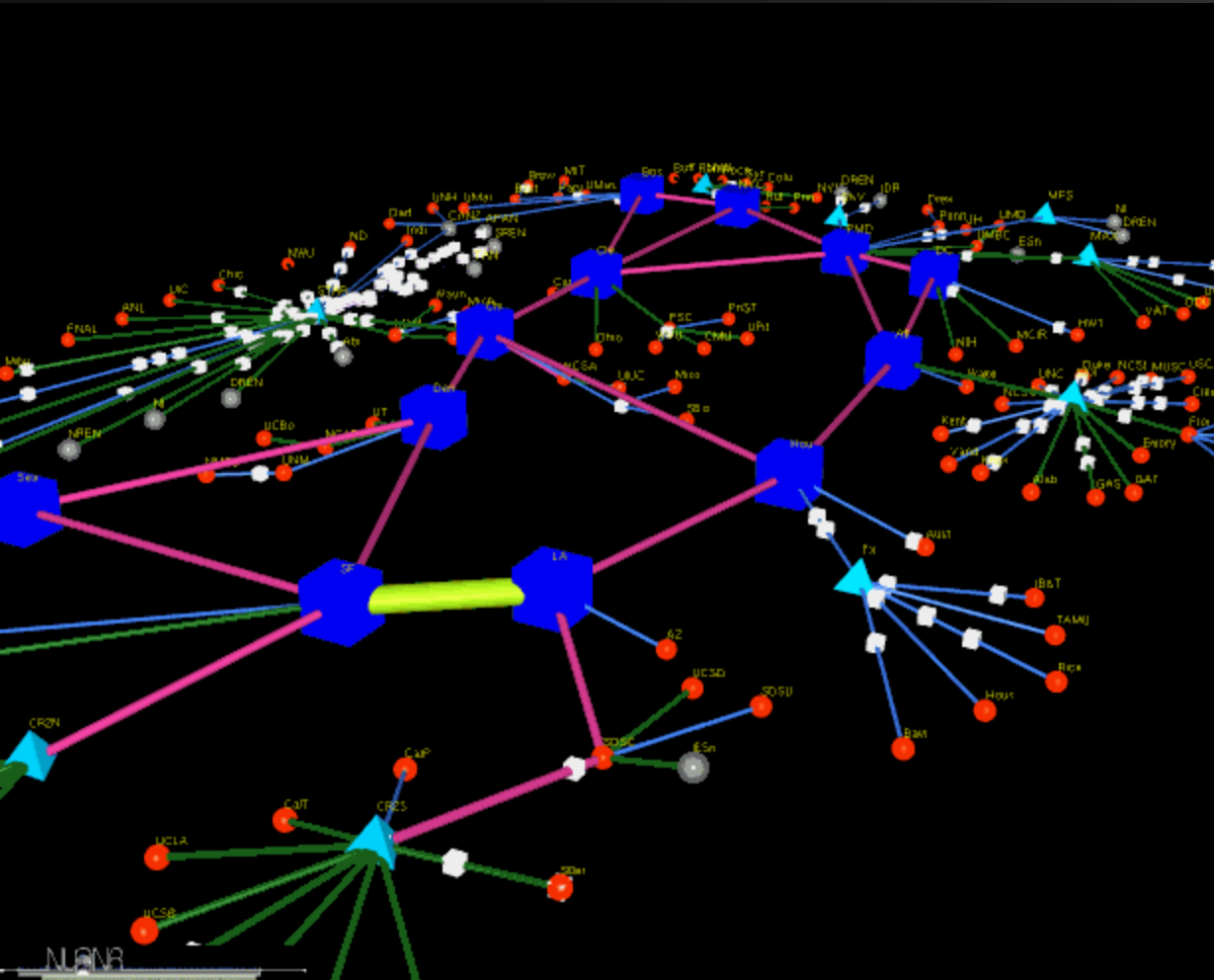
Questo risultato definisce il rapporto tra granularità della piattaforma e quello della computazione. E' necessario, per rendere massima l'efficienza della implementazione del task sull'architettura di calcolo, rispettare questa disuguaglianza.

Le piattaforme "eterogenee" attorno al cui concetto il nostro gruppo sta lavorando da anni, consentono di avere a disposizione il più ampio spettro di g_m possibili



$$g_{tij}^{(k)} \quad g_{mij}^{(k)} : g_{mij}^{(k)} < g_{tij}^{(k)}$$

World-wide computing farms



Una GRID è un insieme di piattaforme di calcolo distribuite su un'ampia regione che possono eseguire parti diverse dello stesso codice oppure "repliche" dello stesso codice con dati diversi.

United Devices Anthrax Research Project Results

In just 24 days, 3.57 billion molecules were scanned against one of the protein components that contributes to the anthrax bacteria's lethal processes. Of those molecules, 376,064 molecules were found to be potential candidates for development into new anti-anthrax drugs, with 12,000 looking very promising.

In a ceremony in Washington DC on March 8, 2002, Oxford University researchers gave the US and British government two compact disks containing the identity of those molecules.

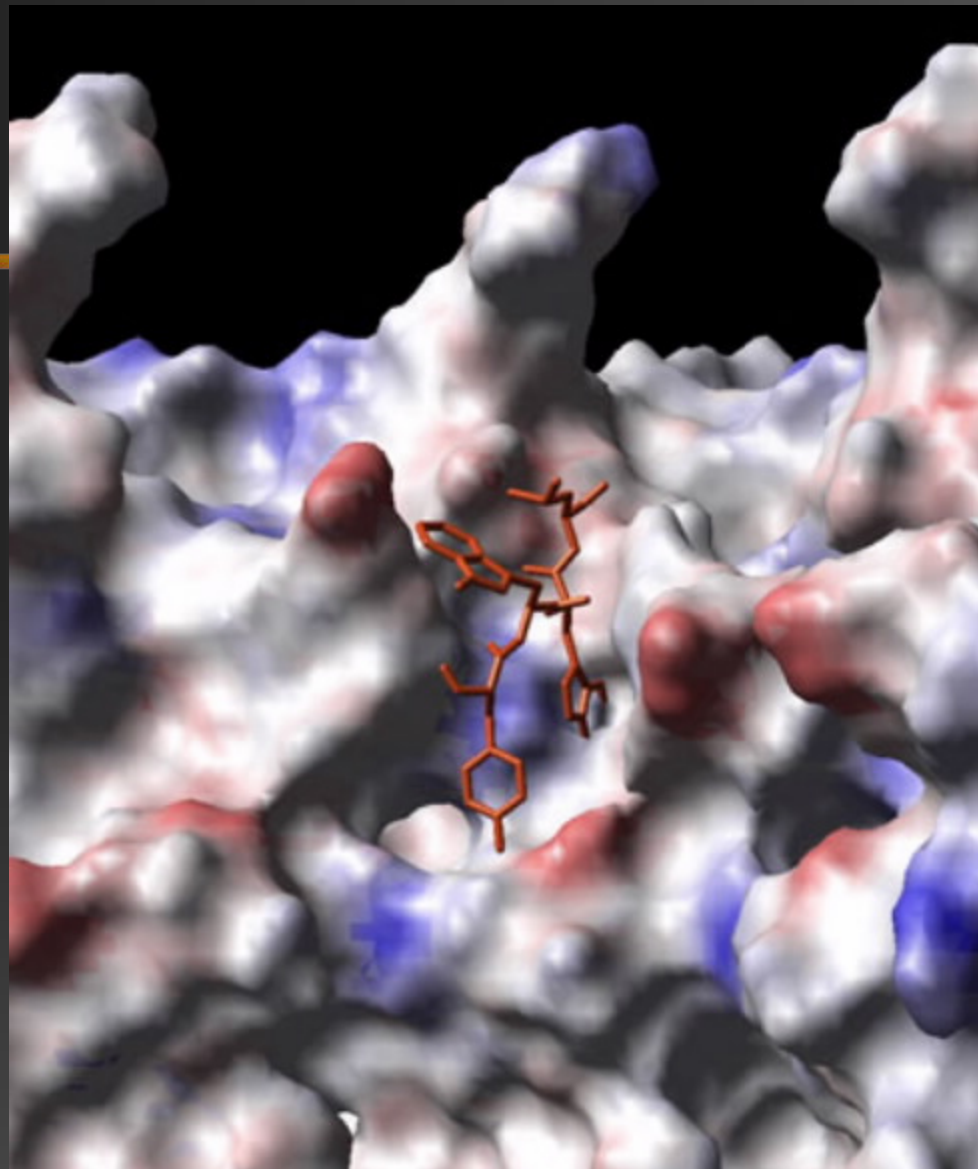
Dr. Anna Johnson-Winnegar, deputy assistant defense secretary for chemical and biological defense, said the information would be shared with the National Institutes of Health. "The Department of Defense is very happy to accept this contribution and looks forward to enabling technology advances through molecular design studies based on the structures of the compounds that have been identified," said Johnson-Winnegar.

Dr. Graham Richards, the lead research on the Oxford University research team, indicated that he believes this process which took 24 days could reduce drug development time from the typical 10 years to as little as 2 to 6 years.

Quick Facts

- 5,436 Compute Years were consumed (47,621,948 hours).
- During this project, enough computers joined to calculate more than 60 trillion floating point operations every second (60 teraflops).
- If the fastest 20 supercomputers in the world were all used at the same time to run these calculations, it would take almost twice as long to get these results.*

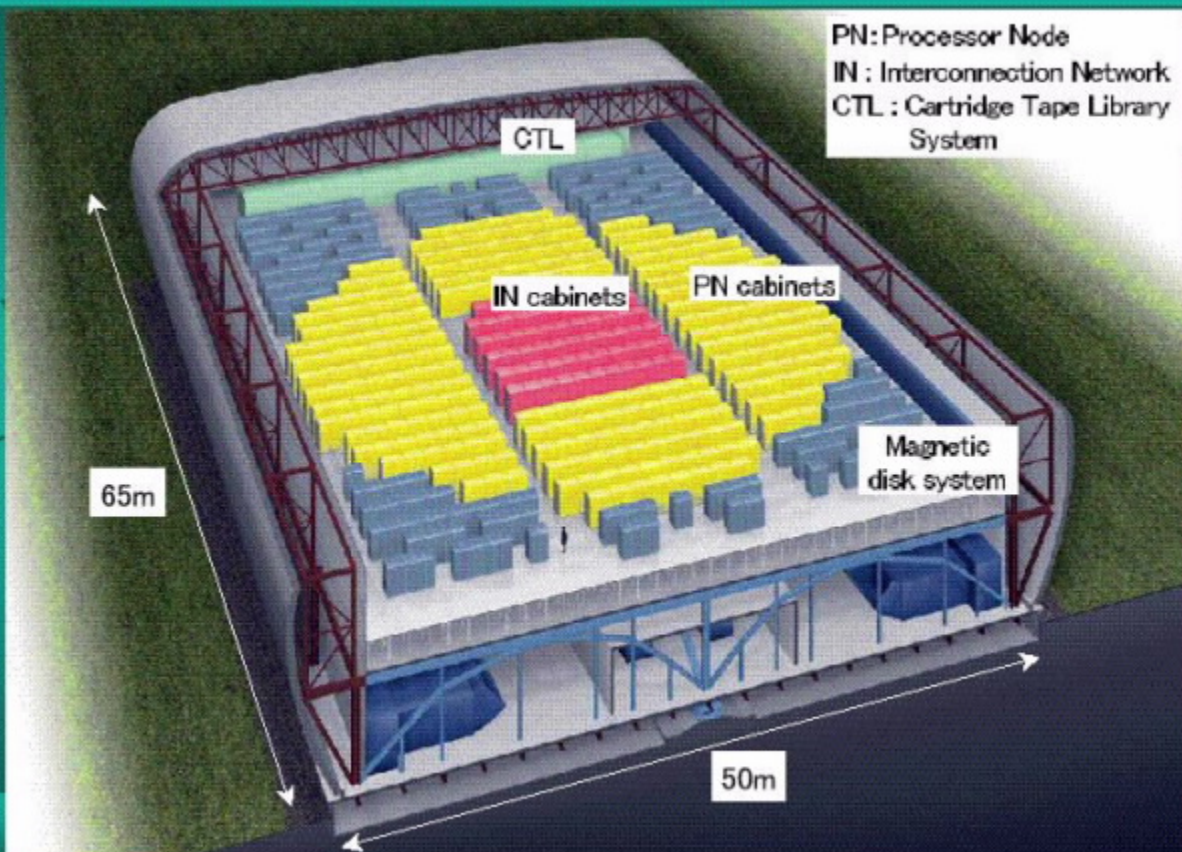
*Based on the measured sustained rate as listed on top500.org as of Nov, 2001



Terascale Computers



Bird's-eye View of the Earth Simulator Building



Earth Simulator
5192 processors
~ 40 Tflops

(1 Tflops=
1000 Miliardi op/sec)

Conclusioni (1)

L'approccio computazionale è divenuto un nuovo paradigma della conoscenza scientifica (ri-costruzione).

I modelli numerici sono un indispensabile strumento per simulare sistemi non riproducibili attraverso modelli analogici.

Esiste un accoppiamento ideale tra architetture di calcolo e modelli numerici. Questo accoppiamento consente di massimizzare l'efficienza funzionale delle piattaforme di calcolo per svolgere l'esecuzione dei codici di simulazione.

Conclusioni (2)

La riflessione fatta su questi temi ha prodotto, da un lato, le esperienze fatte negli ultimi 10 anni con i progetti Genesi (1 e 2) e, successivamente, l'ideazione dello spin-off "Ylichron".

Il progetto su cui si basa Ylichron è proprio quello della progettazione e realizzazione di piattaforme di calcolo nelle quali, sia il sw che l'hw, possano essere il più efficacemente "accoppiati" e resi massimamente efficienti, nel senso descritto.