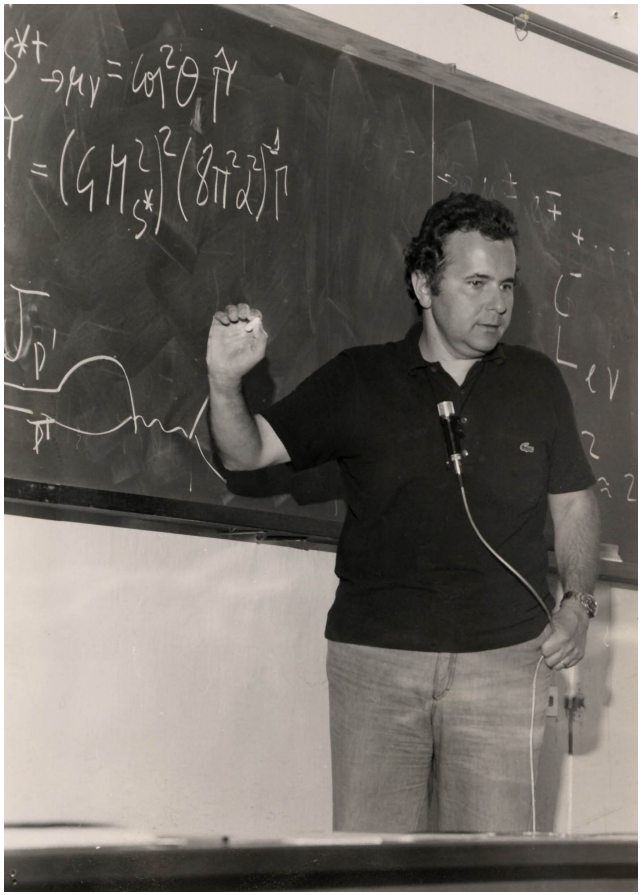
“Facile, forse anche possibile”: Venti anni di calcolo dedicato per la fisica teorica

R. (lele) Tripiccione
Dipartimento di Fisica, Universita' di Ferrara
Ferrara (Italy)

tripiccione@fe.infn.it

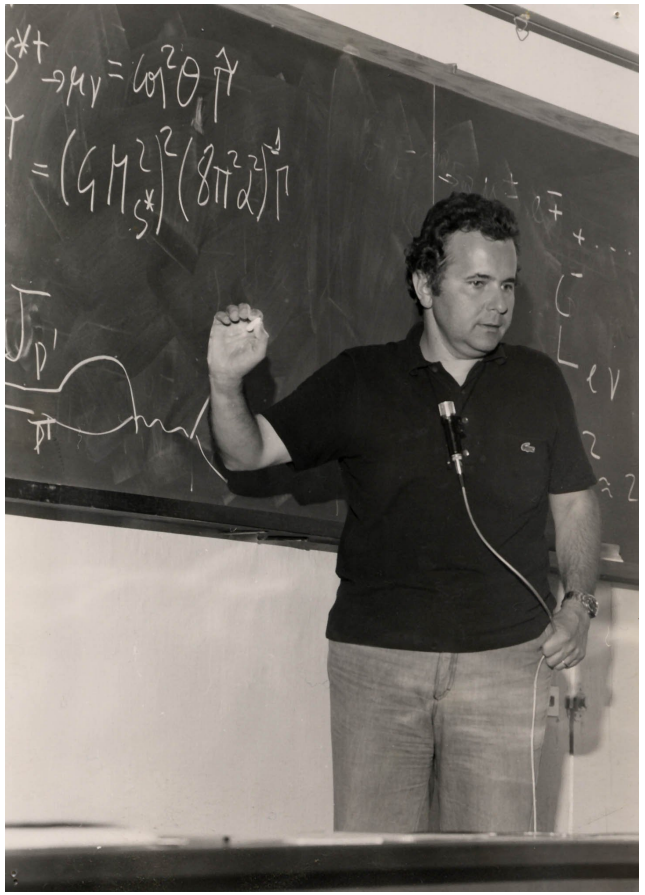
La CEP prima della CEP …
Pisa, 12 novembre 2011

In memoria di Nicola Cabibbo



1935 - 2010

In memoria di Nicola Cabibbo e Roby Marchesini



1935 – 2010

1968 - 2011

Schema del talk

La storia di una (piccola) saga in cui …

… degli outsider rispetto al mondo del calcolo hanno costruito e fatto funzionare (per circa 20 anni) degli strumenti di calcolo scientifico

- *con l' obbiettivo di risolvere un piccolo numero di problemi fisici molto importanti*

- *utilizzando tecnologie relativamente modeste*
- *ma modellando fortemente la struttura sullo specifico problema*

- *perche' l' industria dei computer non era interessa a questa particolare nicchia*

Schema del talk

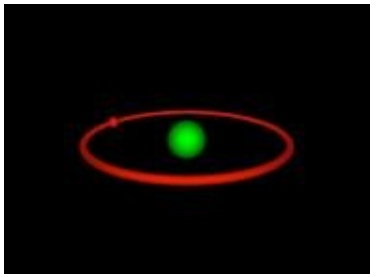
Diverso quindi – negli obbiettivi – dalla CEP

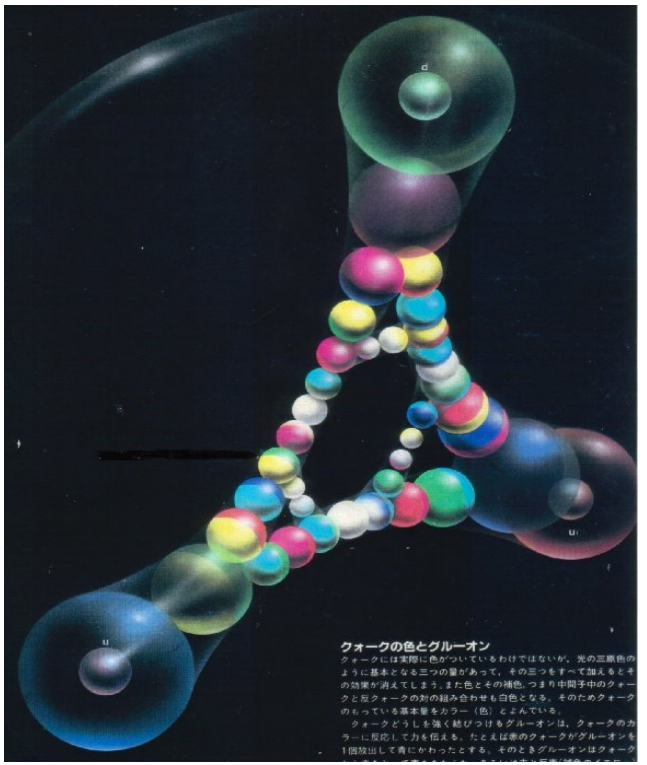
Ma affine come spirito: accettare una sfida di grosso calibro e non aver paura a portarla avanti

Vent anni dopo: Cosa abbiamo imparato?

Spesso la gente da buoni consigli
Se non puo dare il cattivo esempio

Lattice Quantum Chromo-Dynamics (LQCD)





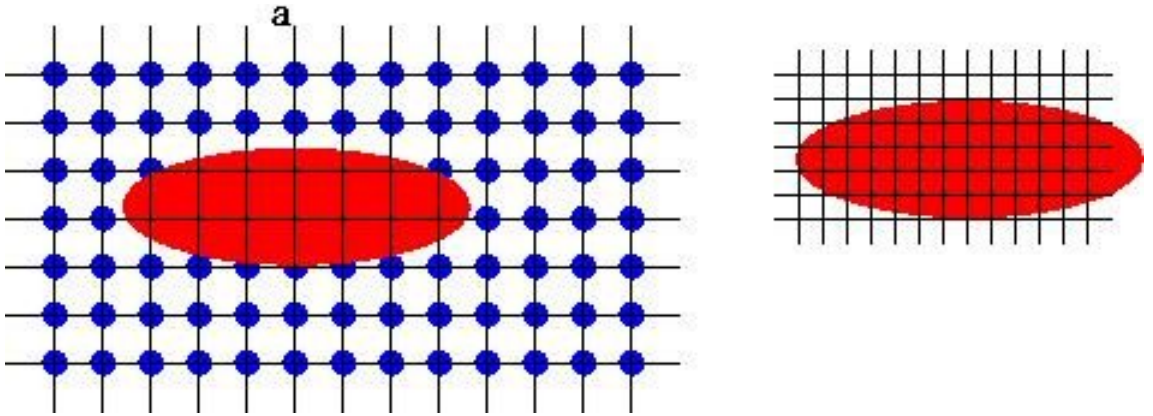
Unfortunately it is not yet known whether the quarks in Quantum Chromodynamics actually form the required bound states. To establish whether these bound states exist one must solve a strong coupling problem and present methods for solving field theories don't work for strong coupling.

K. Wilson, Cargese Lectures, 1976

LQCD algorithms are conceptually very much similar to spin systems (just much more complex). The computational cost is huge:

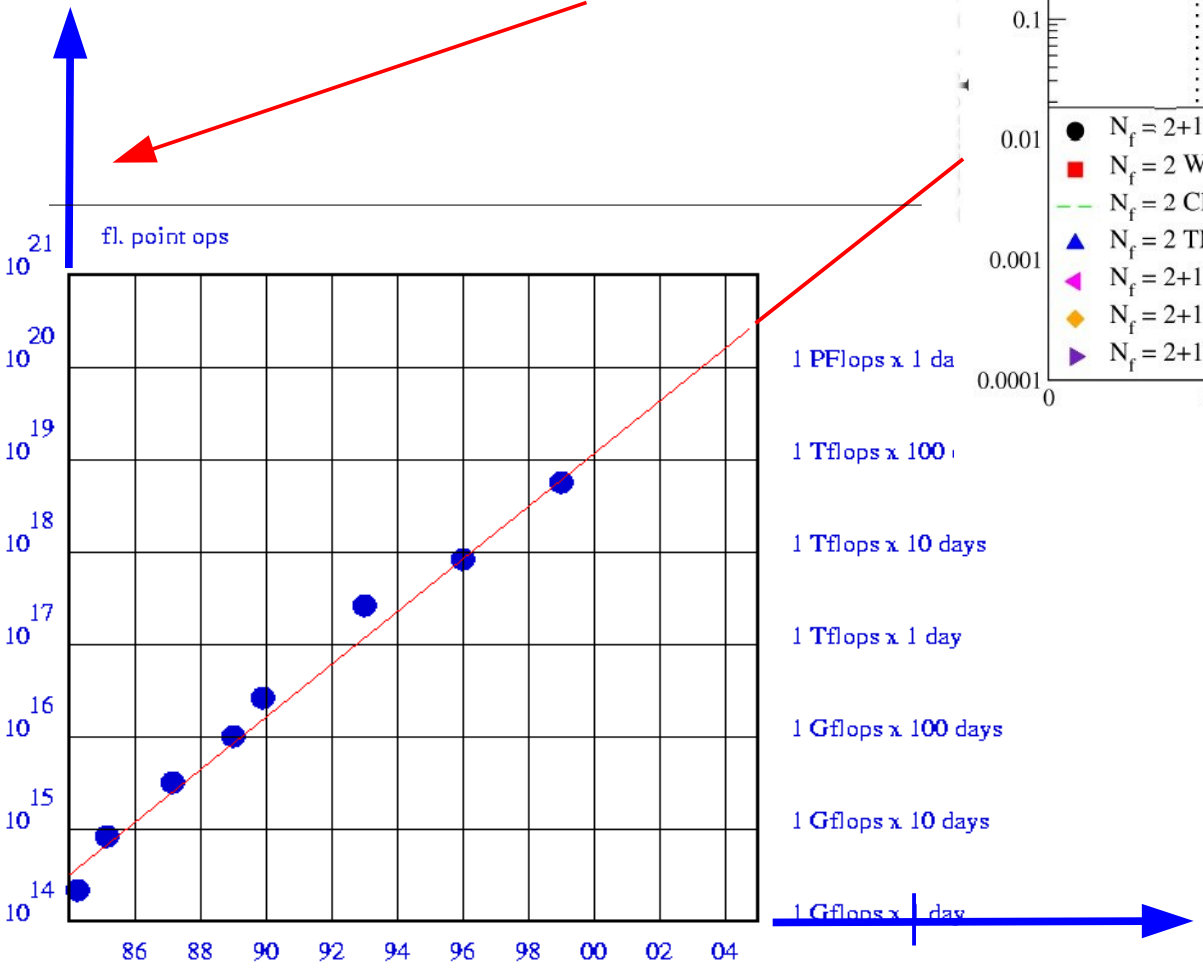
$$N_{flop} \sim L^{5\ldots6} \times (1/a)^{6\ldots7} \times (1/m_q)^{1\ldots2}$$

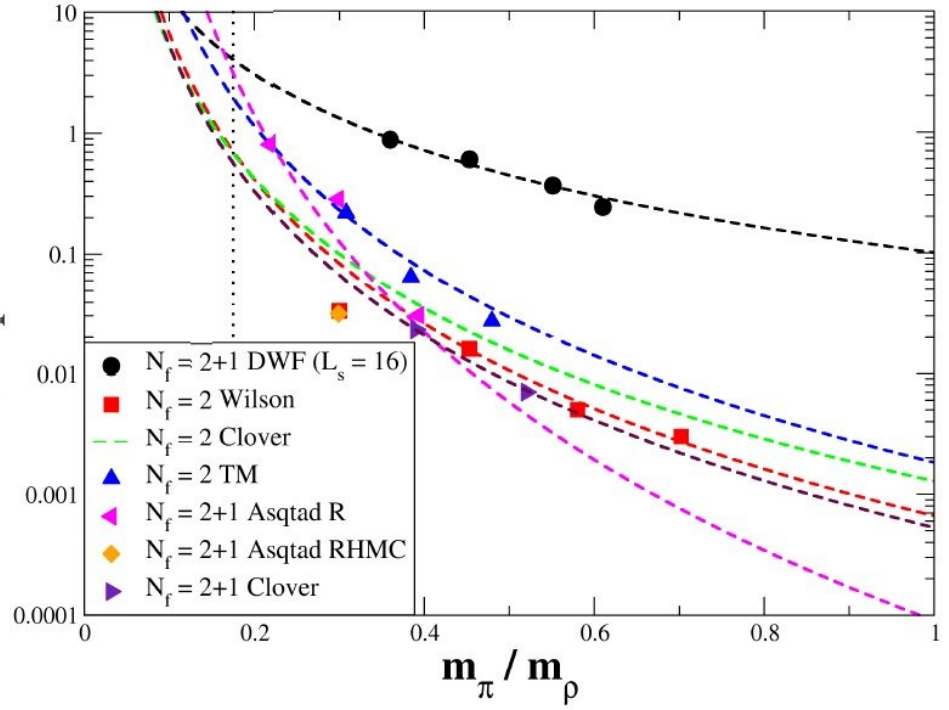
… as one tries to take into account all relevant scales of the problem.



Two different views of the same problem

0.5 Pflops x 100 d ~ 100 Tflops x 500 d





L = 2 fm
a = 0.08 fm
1000 configs

What is the key feature?

The most relevant common feature of this (and many similar problems) is that they have a very large degree of parallelism.

It is all too natural to try to identify, expose and exploit as large a fraction of this parallelism as possible ...

... and in these cases it is very-very easy to do so (almost in principle) because other features help as well:

- very simple algorithmic structures
- easily predicted patterns of memory access
- almost automatic load balancing

One has to be optimistic ...

In most cases, computing accurate predictions of the behaviour of a complex physical system is hopeless, unless numerical techniques are used

One has to be optimistic ...

In most cases, computing accurate predictions of the behaviour of a complex physical system is hopeless, unless numerical techniques are used

However Nature has been friendly to us

One has to be optimistic ...

In most cases, computing accurate predictions of the behaviour of a complex physical system is hopeless, unless numerical techniques are used

However Nature has been friendly to us

the (simple) physics laws behind the behaviour of computers make it relatively easy to build machines able to simulate complex physics!!!!

Two cornerstones of physics-friendly architectures

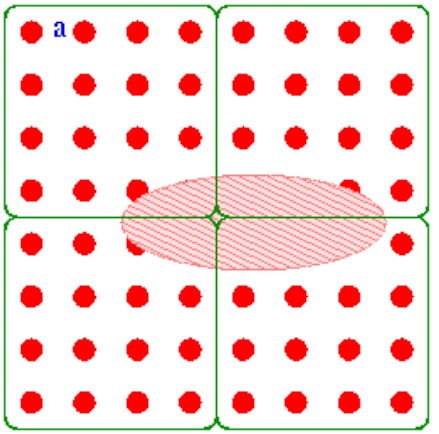
1) Parallel computing is available in (almost) all cases ...
... and parallel computing is the physics sponsored way to compute:

The basic building block is the transistor
Industry learns to build smaller and smaller transistors. As $\lambda \to 0$
obviously $N \propto 1/\lambda^2$ *but speed scales less favourably* $\tau \propto \lambda$

Trade rule: perform more and more things in parallel
rather than a fixed number of things faster and faster

Two cornerstones of physics-friendly architectures

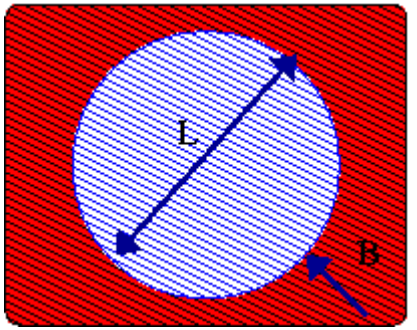
2) We are (often) interested in modeling "local" systems:
This has to go over to the computer structure ->
Keep data close in space to where it is processed



Failure to do so will asymptotically bring a data bottleneck:

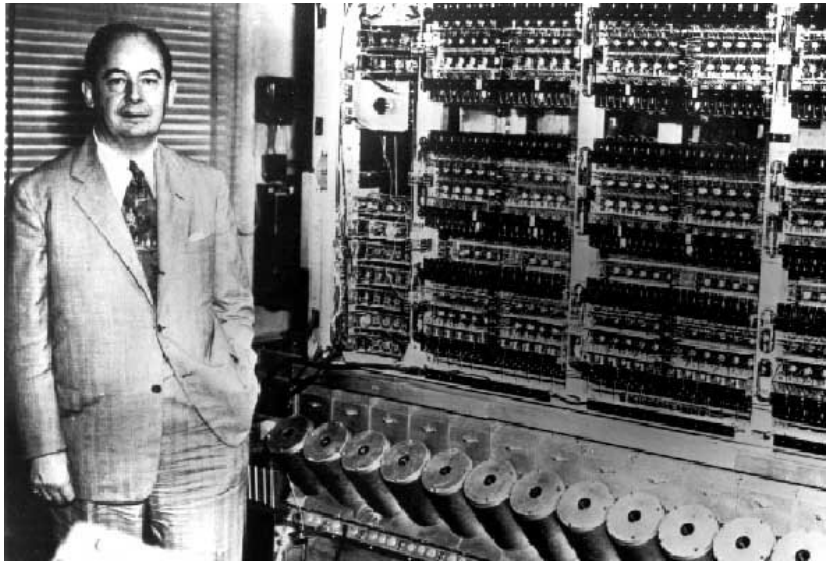
$$B(L) \propto L$$

$$P(L) \propto L^2$$



A historical question:

The guy who invented computer(s-models) made his model a physics-friendly beast???

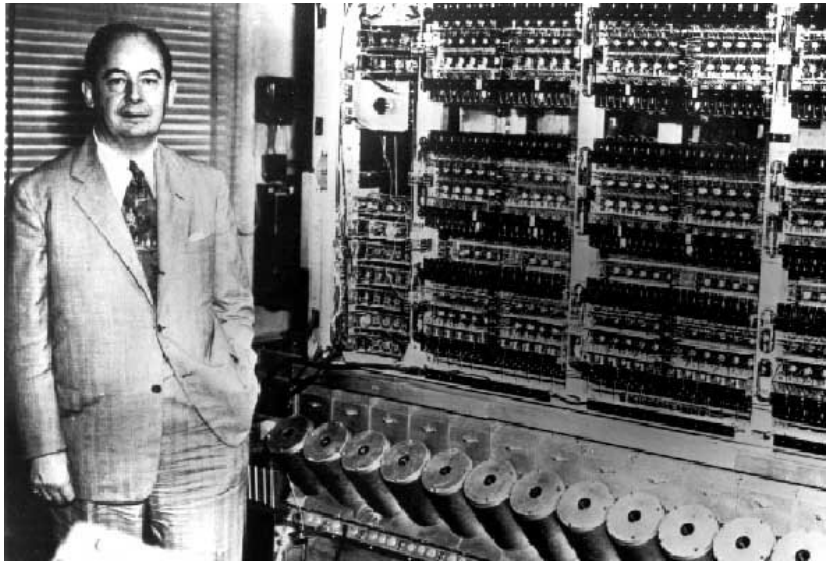


The Answer is: NO!

Doing things one after the other (serially)
Keeping data storage and data processing separated (in principle and practice)

A historical question:

The guy who invented computer-(models) made his model a physics-friendly beast???



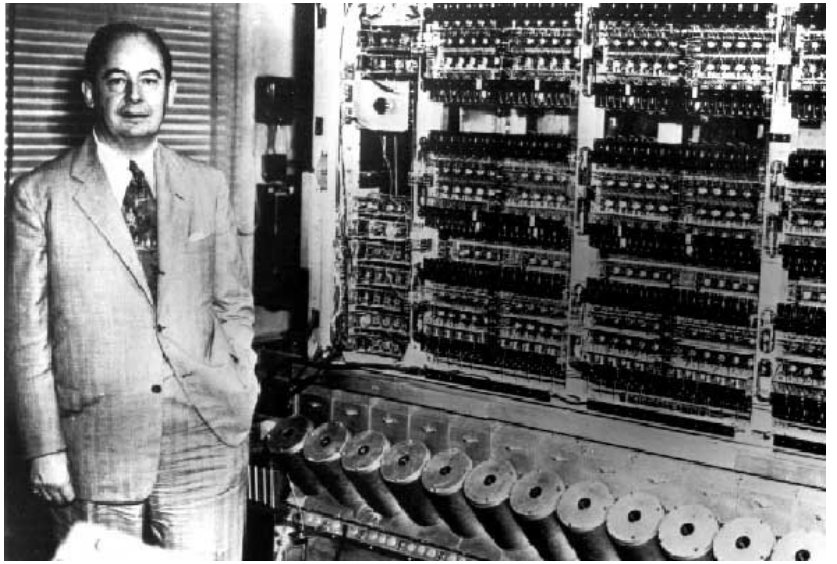
The Answer: NO!

Doing things one after the other (serially)
Keeping data storage and data processing separated (in principle and practice)
are the cornerstones of the famous von Neumann model of computing

Q: So was Von Neumann wrong?

A historical question:

The guy who invented computer-(models) made his model a physics-friendly beast???



The Answer: NO!

Doing things one after the other (serially)
Keeping data storage and data processing separated (in principle and practice)
are the cornerstones of the famous von Neumann model of computing

Q: So was Von Neumann wrong?

A: No, he was interested in the $P \to 1, \tau \to \infty$ *regime*
today we are approaching the $P \to \infty, \tau \to 0$ *regime .....*

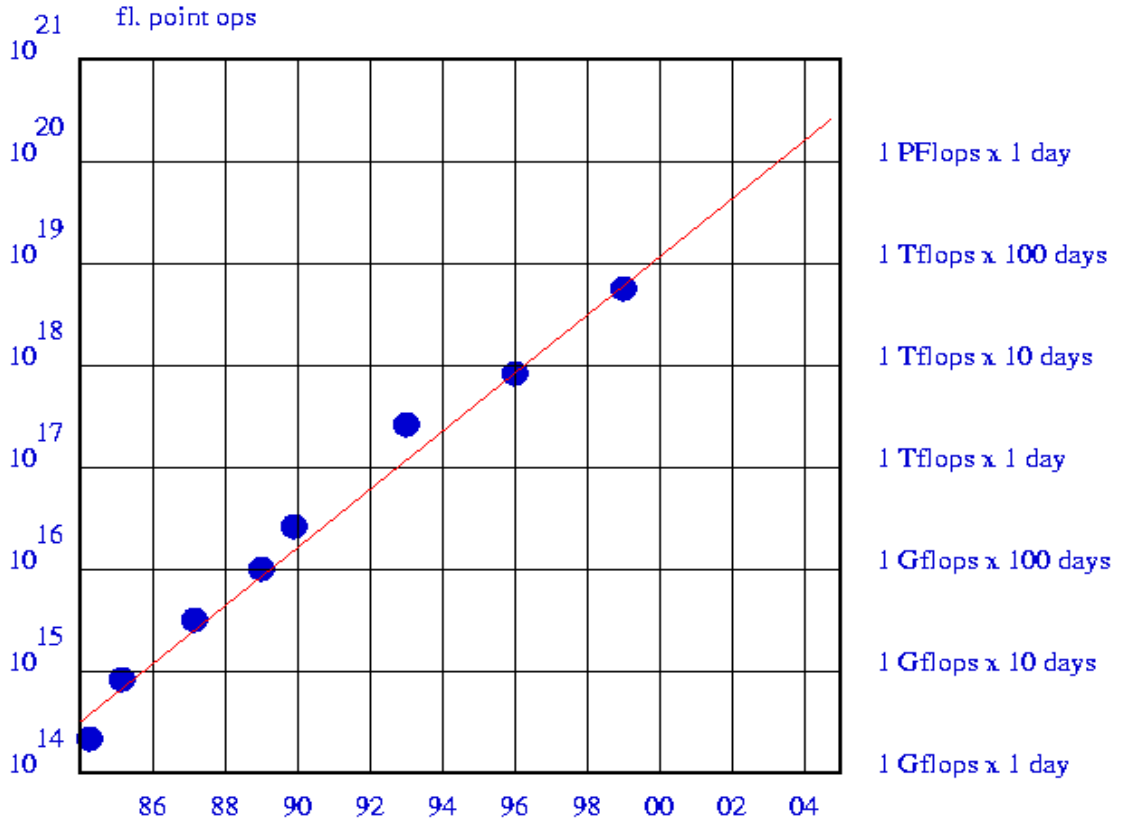
Physics "unfriendly" processors .....

Surprising results (G. Bilardi et al.)

			Opteron – quad
die size(mm2)		0	258.0
clock (GHz)		0	2.5
technol. (nm)		0	45.0
fps possible		0	320
100% fp (TF)		5	0.80
real peak (TF)		1	0.04
fraction		3	**0.05**

A small but lively community has worked for some 20 years, building LQCD-optimized number cruncher that over the years have given physicist the compute cycles they needed (in spite of very poor budgets) ....

"E' facile, forse anche possibile!" (G. Parisi, circa 1986)



Basically several generations of two big projects:

Columbia University in the US

The APE project in Europe

- *Bologna*
- *Ferrara*
- *Pisa*
- *Padova*
- *Roma*

\+ *Bielefeld - DESY - Orsay - Swansea*

Basically several generation of two big projects:

Columbia University in the US
The APE project in Europe

Bits of history(1)

1979: The early pioneers:
the Caltech Ising machine (D. Toussant, G. Fox, C. Seitz)

circa 1985:

APE (16 nodes, 1 Gflops)
Columbia (~ 1 Gflops)
GF11 (IBM/Yorktown)

1990 - 1995:

APE100 (500 – 1000 nodes, 50 – 100 Gflops)
Columbia2 (also about 100 Gflops)

Bits of history(2)

1995 – 2000:

APEmille (1.8 Tflops installed)

QCDSP (1 + 1 Tflops at Columbia & Broohhaven)

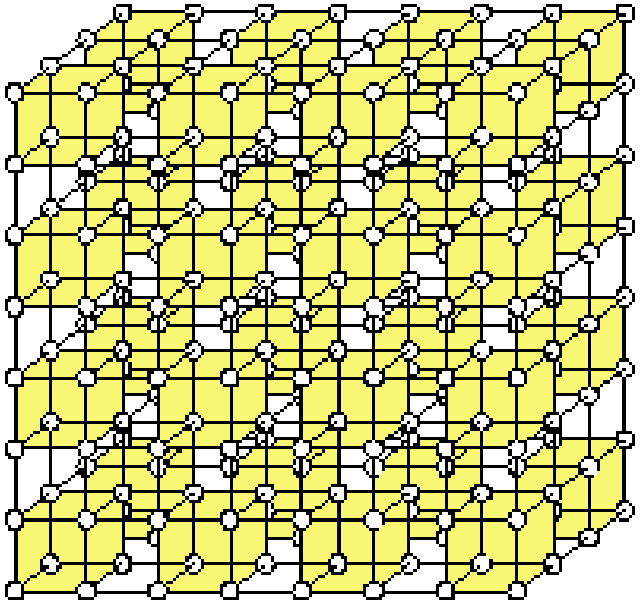
CP-PACS (Tsukuba + Hitachi, 600 Gflops)

2000 – 2005:

ApeNEXT (15 Tflops installed)

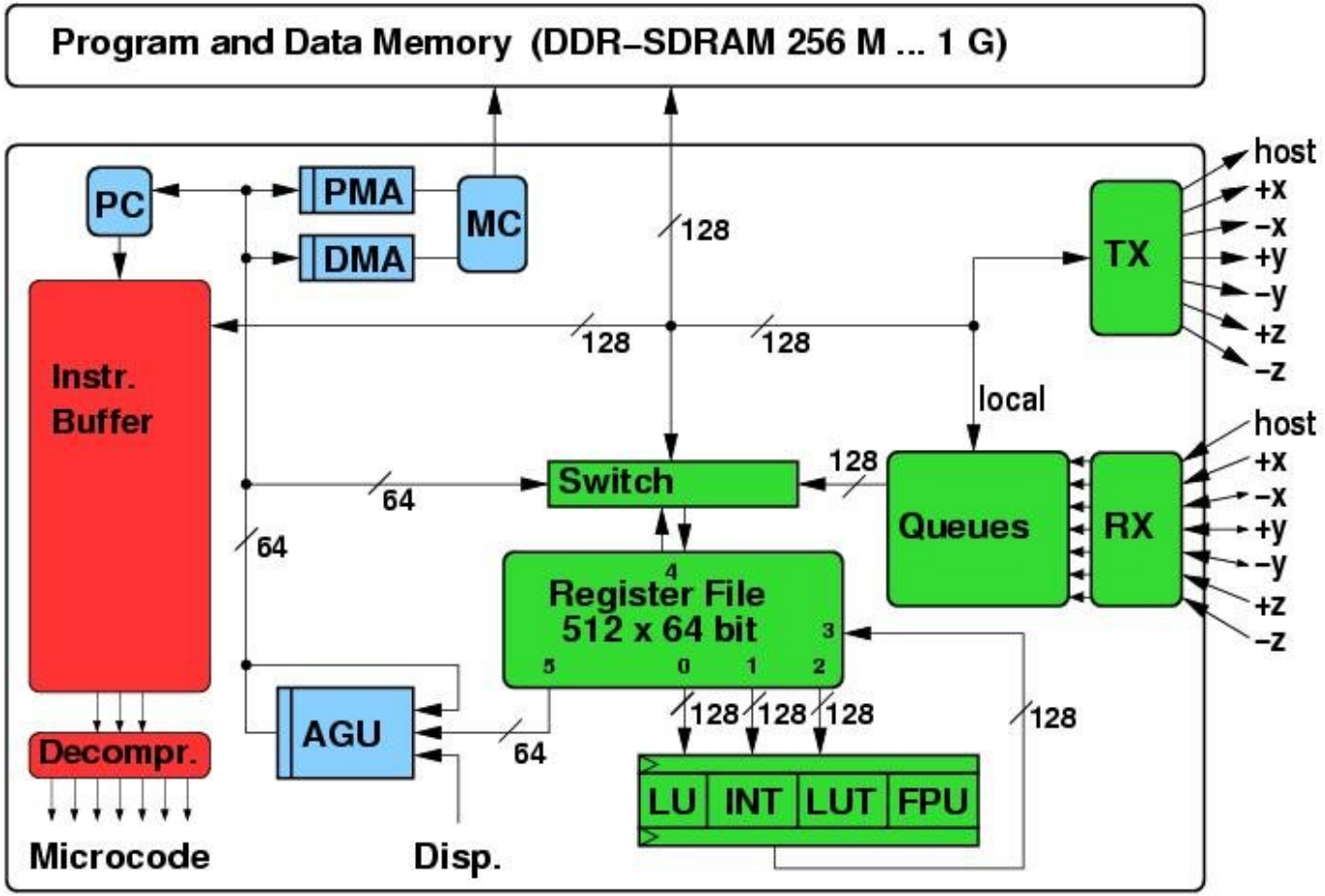
QCDOC (Columbia + Brookhaven + IBM/Yorktown)

ApeNEXT: the global structure



- *The core of apeNEXT is a 3-D array of processing elements. The physical lattice is divided in equal size partitions on all processors. The fourth+ dimension(s) is fully contained inside each processor*

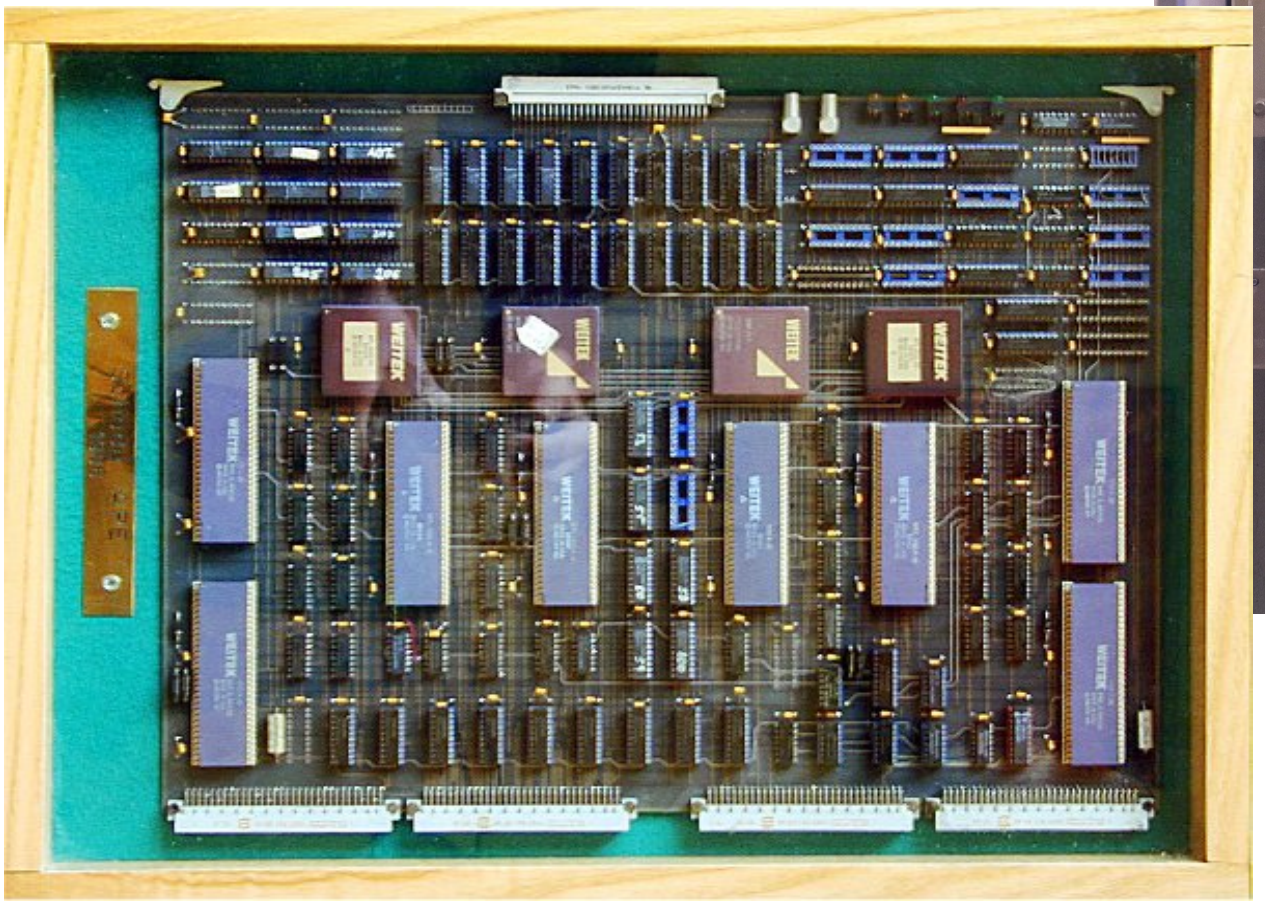
ApeNEXT: the node processor



1.6 GFlops @ 4.2 W →
400 Mflops / W
~ #4 in the Green500 list

*The apeNEXT processor (so called J&T) contains all the functional blocks needed for efficient computation, memory access and communication with other nodes (**and nothing else**)*

A saga starting in 1985 .....



My daddy said we looked ridiculous,
but, boy, we broke some hearts!

(from "I was only joking", Rod Stewart)

.... and ending in ~ 2006

That community went essentially unnoticed outside theoretical physics

Till it had a minor but not negligible role in a small revolution that happened just a few years ago ....

A small revolution happened in the early 2000s...

In the early 2000s, computers companies started to build physics-friendly computers (even if only for specific market niches):

- Blue-Gene

- The Cell Processor

- GPUs

- (FPGAs)

Dai diamanti non nasce niente,
dal letame nascono i fiori ....

The Blue Gene revolution ...

i) very large 3D meshes of simple, relatively low performance distributed-memory processors (largely inspired by earlier LQCD application-driven number-cruncher)

ii) you better learn to adapt your algos / programs to this specific architecture ...

carries the BigBlue brand...

Physics-friendly at the largest (system) scale

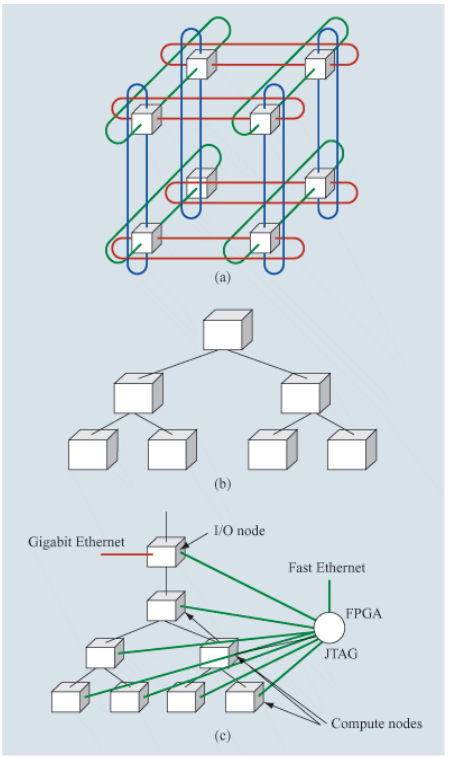


Figure 4

(a) Three-dimensional torus. (b) Global collective network. (c) Blue Gene/L control system network and Gigabit Ethernet networks.

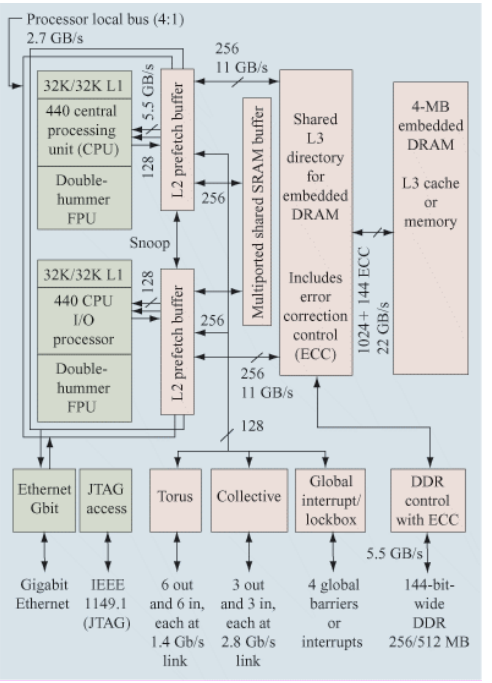
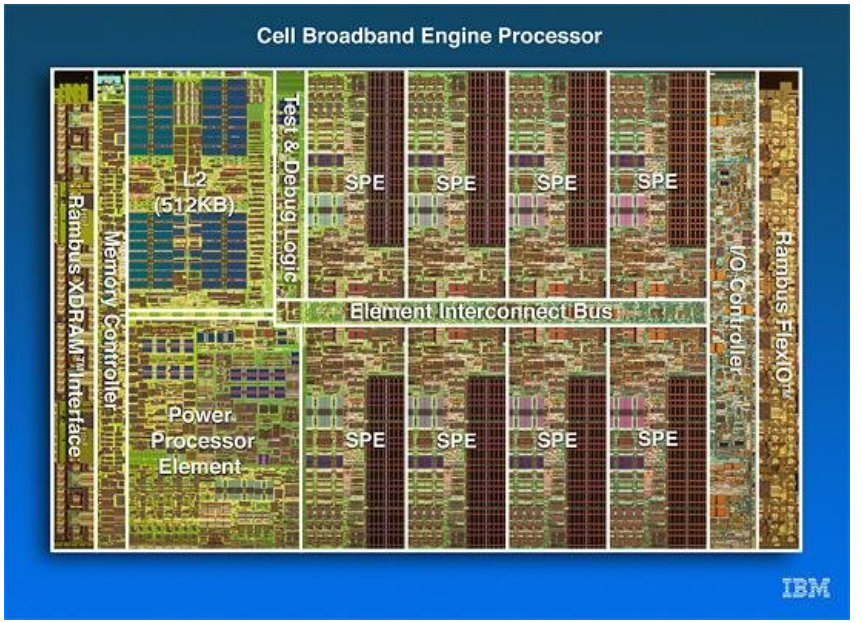


Figure 5

Blue Gene/L compute (BLC) chip architecture. Green shading indicates off-the-shelf cores. ©2002 IEEE. Reprinted with permission from G. Almasi et al., "Cellular Supercomputing with System-on-a-Chip," *Digest of Technical Papers*, 2002 IEEE International Solid-State Circuits Conference.

The cell processor

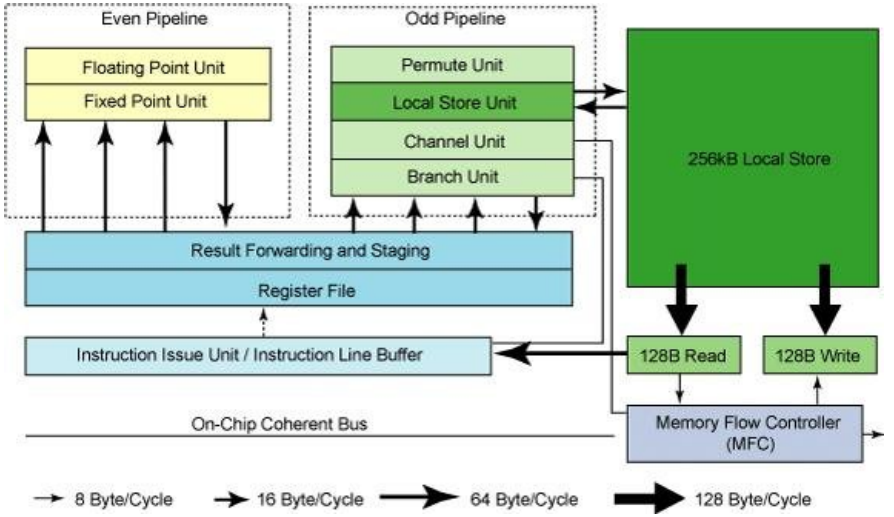
Physics-friendly architecture at the processor scale....



3 – 4 Ghz x 8 cores x 8 ops →
256 Gflops (100 Gflops DP)

25 Gbyte/sec mem. Bandwidth
300 Gbyte/sec internal bandwidth
75 Gbyte/sec fast IO

256 Kbyte local store (each core)
Here is the problem.....

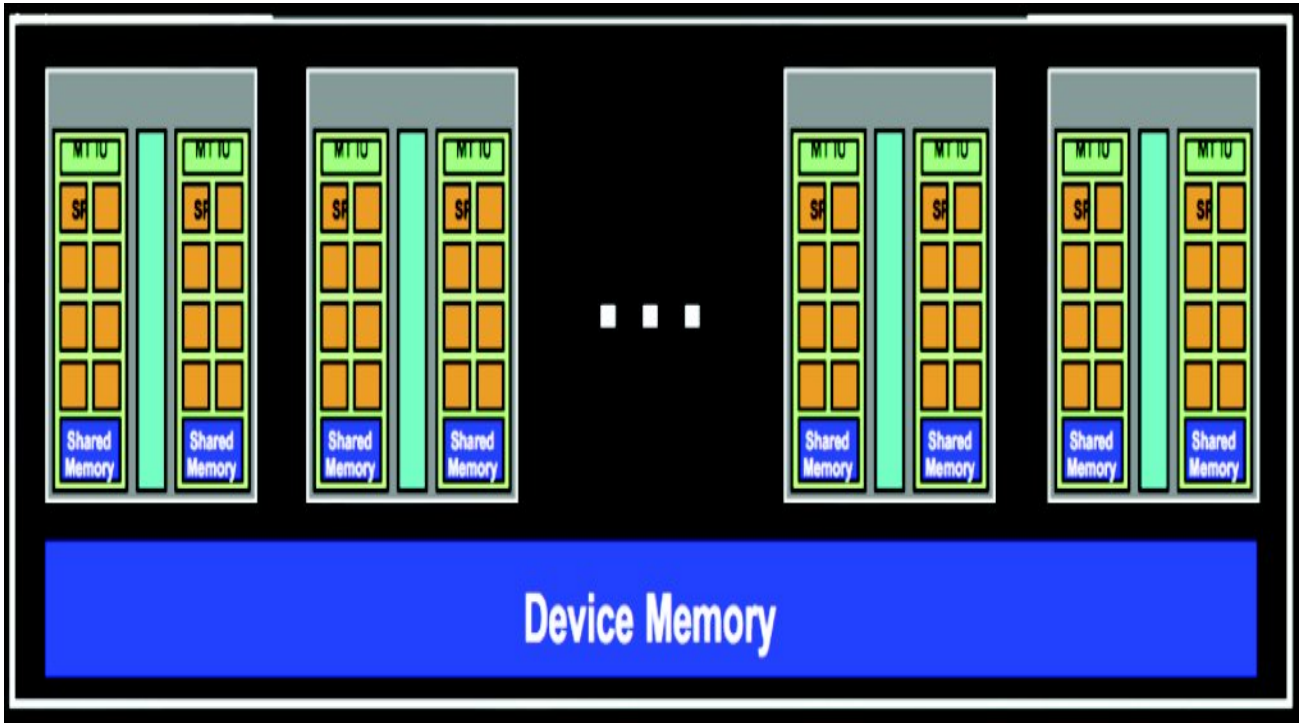


An even more radical approach: GPUs

The key idea is to pack a very large number of small processors....

Typical figures:

240 processing cores ----> 930 Gflops ----> 2 ops/core/clock



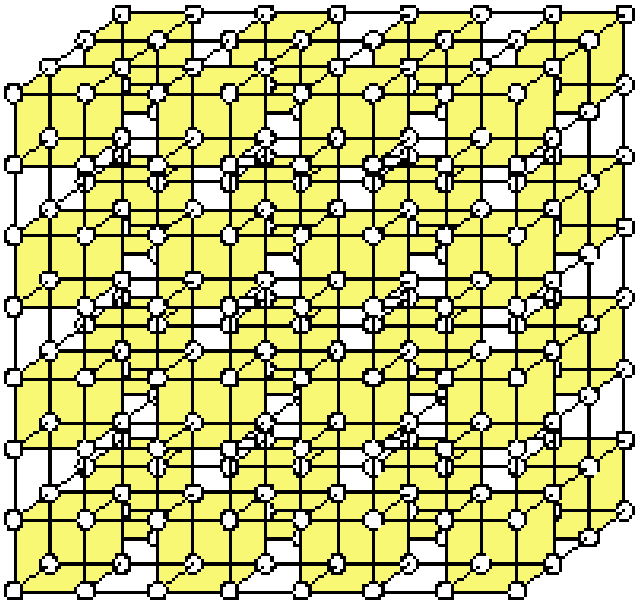
So, everything OK, now???

	GPU	Cell	Opteron – quad
die size(mm2)	576.0	221.0	258.0
clock (GHz)	1.3	3.0	2.5
technol. (nm)	65.0	45.0	45.0
fps possible	1300	250	320
100% fp (TF)	1.70	0.75	0.80
real peak (TF)	0.9	0.1	0.04
fraction	**0.55**	**0.13**	**0.05**

Well, much better than before !!!!!!

The APE legacy: QPACE

Remember what we did?!?!?.



Now, replace our cute little old apeNEXT processors with the big and powerful Cell ones

QPACE

The QPACE projects bets on the guy at right in the previous slide

A Cell based 3-d system

Each processor ~ 100 Gfs (DP) ---> 30 Gfs sustained

Network balanced at 1 Gbyte/sec for each link

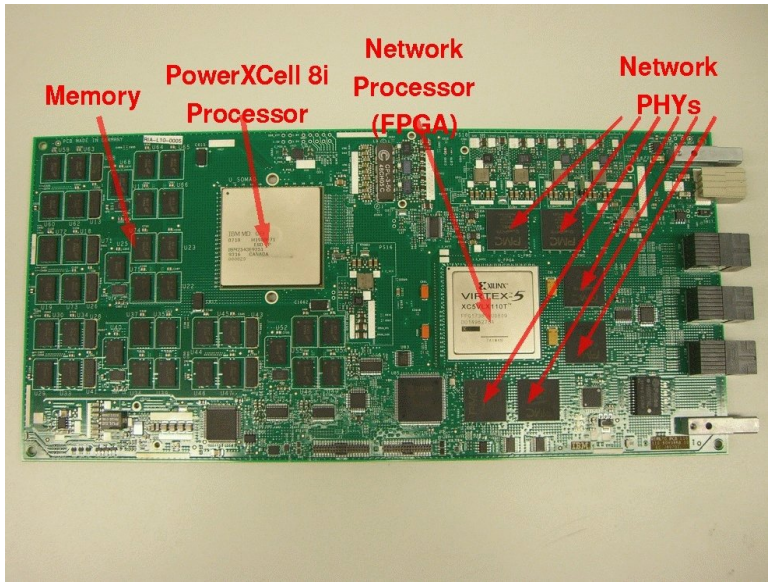
Collaboration of :
Regensburg – Wuppertal – Juelich (Germany)

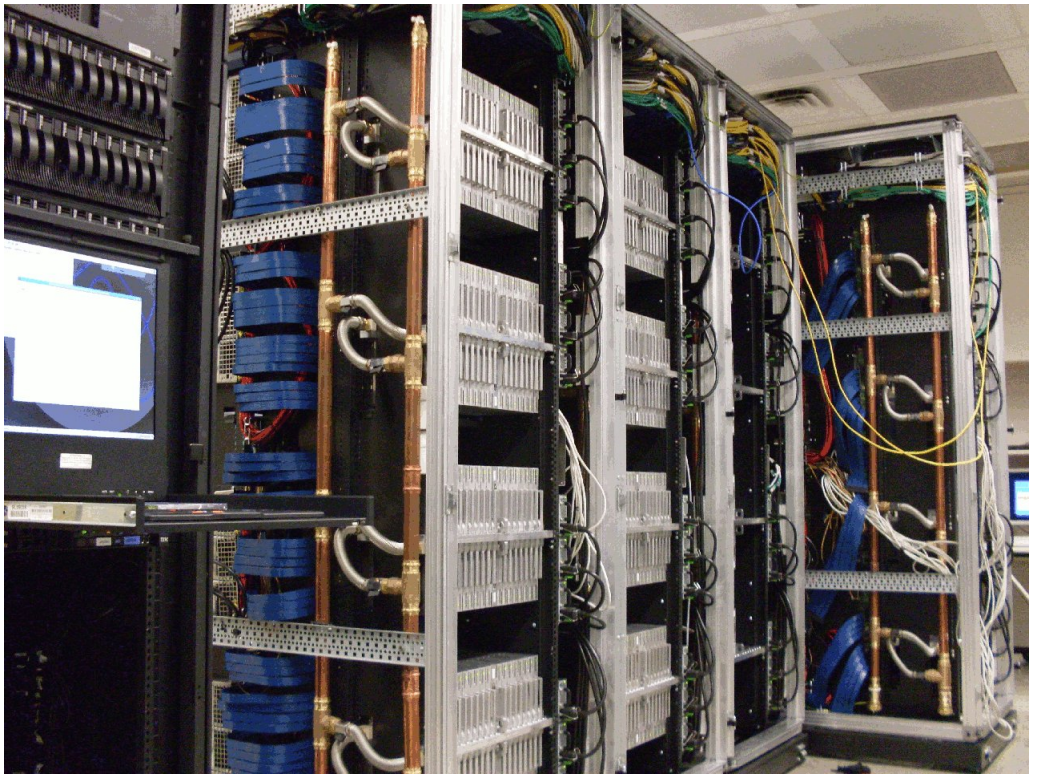
Ferrara – Milan (Italy)

IBM / Boeblingen (main industrial partner)
Eurotech (additional industrial partners)
Knurr

QPACE

A very quick development and bring up cycle:





QPACE

3 machines installed

512 nodes (2048 cores)
2 Gbyte / node main memory

12 x 1Gbyte / sec communication links
1 microsecond link latency

Some 150 Tflops overall peak performance

Surprisingly cheap: XXX Euro for the whole project

QPACE:

Rewarding results ....

Entry #1, #2, #3 of the Green500 list for (at least …) two times

The Green500 List - June 2010

Listed below are the June 2010 The Green500's energy-efficient supercomputers ranked from 1 to 100.

SUPERMICRO

Green500 Rank	MFLOPS/W	Site*	Compute	
1	773.38	Forschungszentrum Juelich (FZJ)	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.54
1	773.38	Universitaet Regensburg	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.54
1	773.38	Universitaet Wuppertal	QPACE SFB TR Cluster, PowerXCell 8i, 3.2 GHz, 3D-Torus	57.54
4	492.64	National Supercomputing Centre in Shenzhen (NSCS)	Dawning Nebulae, TC3600 blade CB60-G2 cluster, Intel Xeon 5650/ nVidia C2050, Infiniband	2580
5	458.33	DOE/NNSA/LANL	BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Infiniband	276
5	458.33	IBM Poughkeepsie Benchmarking Center	BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Infiniband	138
7	444.25	DOE/NNSA/LANL	BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband	2345.5
8	431.88	Institute of Process Engineering, Chinese Academy of Sciences	Mole-8.5 Cluster Xeon L5520 2.26 Ghz, nVidia Tesla, Infiniband	480
9	418.47	Mississippi State University	iDataPlex, Xeon X56xx 6C 2.8 GHz, Infiniband	72

SUPERMICRO
Proud Sponsor of The Green500

Vent' anni dopo?

Cosa e' rimasto di tutto questo, vent' anni dopo?

<u>Buone notizie:</u>

Un riconoscimento internazionale della qualita' dei risultati di f fisica resi possibili da queste iniziative

Una generazione di giovani a loro agio tra fisica e computer

Qualche timido approccio allo studio del computer come "sistema fisico"

Vent' anni dopo?

Cosa e' rimasto di tutto questo, vent' anni dopo?

<u>Cattive notizie:</u>

Mentre negli Stati Uniti alcune persone chiave di Columbia University (Al Gara, J. Sexton, P. Boyle) inventavano e costruivano Blue Gene.

Alcuni svariati e svariatamente maldestri tentativi di collaborazione con l' Industria (di cui non parlo in questo intervento) non hanno prodotto nessun risultato significativo.

E' mai possibile -- corpo di un cane --
Che le avventure in questo reame ...