

Arquitecturas Paralelas I

Computação Paralela



João Luís Ferreira Sobral

www.di.uminho.pt/~jls

jls@...

<http://gec.di.uminho.pt/lesi/ap1>

Computação paralela

Resultados da Aprendizagem

- Conceber, implementar e otimizar aplicações que executem de forma eficiente em arquiteturas multi-processador (multi-core e multi-thread) e/ou arquiteturas distribuídas (clusters)

Programa (resumido)

- **Arquitectura dos sistemas de computação**
 - sistemas de memória partilhada e sistemas de memória distribuída.
 - introdução à GRID (rede mundial de sistema de computação).
- **Modelos de programação paralela e linguagens**
 - processos comunicantes e objectos activos.
 - distribuição de componentes pelos recursos computacionais
 - mecanismos de especificação de concorrência/paralelismo
- **Projecto de aplicações paralelas**
 - gestão eficiente da distribuição de componentes; medição, análise e optimização do desempenho
 - relação custo/benefício de concorrência/paralelismo e sua quantificação (métricas);
 - adaptação da granularidade da computação/comunicação;
 - estratégias de distribuição da carga de dados/computação/comunicação.
- **Análise de ambientes comerciais/tecnologias**
 - Software de gestão de "GRIDs" (PBS)
 - Desenvolvimento de aplicações empresariais em *clusters* de máquinas (J2EE em ambiente de *clustering*)

Computação paralela

Planeamento

Aula	Matéria Teórica	Tecnologia
1 (19/09)	- Introdução às arquitecturas paralelas/Grid	-
2 (26/09)	- Revisão de conceitos de programação	Programação em Java
3 (3/10)	- <i>Middleware</i> para suporte a distribuição	Java RMI/.Net Remoting (em C#)
4 (10/10)	- Programação OO concorrente	<i>Threads</i> em C#/ Java
5 (17/10)	- Metodologias de desenvolvimento de aplicações concorrentes/distribuídas/paralelas	Desenvolvimento de algoritmos em C# + Remoting / Java
6 (24/10)	- Medição e optimização de desempenho	Análise de algoritmos em Java
7 (31/10)	- Desenvolvimento de aplicações paralelas baseadas em passagem de mensagens	Programação em MPI (<i>Message Passing Interface</i>) e C++
8 (7/11)	-OpenMP	Exercícios
9 (14/11)	- Análise de sistemas de <i>Grid</i>	<i>PDS</i> , introdução à utilização de <i>cluster Search</i>
10 (21/11)	- Aplicações empresariais em <i>Clusters</i>	<i>J2EE</i>
11 (28/11)	- Framework JaSkel	Exercícios
12 (5/12)	Projecto PPC-VM (Aspectos para paralelização)	Exercícios
13 (12/12)	Folga (Projecto – parte 1)	
14 (19/12)	Folga (Projecto – parte 2)	-

Computação Paralela

Avaliação

- ❑ Um trabalho prático (desenvolvimento aplicação para um ambiente de *cluster*)

Pré-requisitos

- ❑ Conhecimentos de Java e de C

Bibliografia (Base)

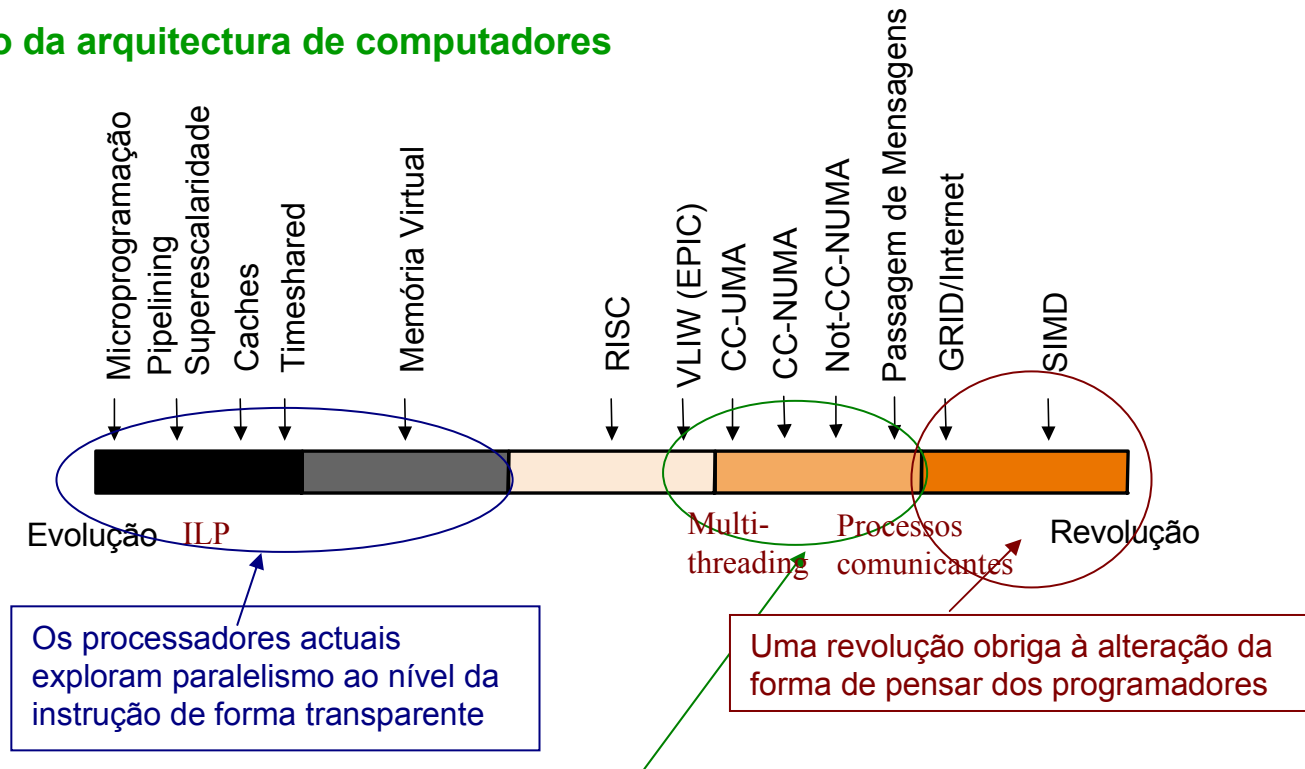
- ❑ Acetatos
- ❑ M. Quinn. *Parallel programming in C with C nad OpenMP*, McGraw Hill, 2003
- ❑ I. Foster. *Designing and Building Parallel Programs*, Addison-Wesley, 1995.
- ❑ D. Lea. *Concurrent Programming in Java: Design Principles and Patterns*, Addison-Wesley, 2000.
- ❑ G. Andrews. *Multithreaded, Parallel and Distributed Programming*, Addison-Wesley, 2000.

Bibliografia (Adicional)

- ❑ R. Gerber, A. Binstock. *Programming with Hyper-Threading Technology*, Intel Press, 2004.
- ❑ F. Buschmann, D. Schmidt, M. Stal, H. Rohnert, *Pattern-oriented Software Architecture Vol 2: Patterns for Concurrent and Networked Objects*, John Wiley and Sons Ltd, 2000.
- ❑ T. Christopher, G. Thiruvathukal, *High-performance Java platform computing*, Prentice Hall, 2000.
- ❑ I. Foster, C. Kesselman (ed). *The Grid2*, Morgan Kaufmann, 2004.

Computação Paralela

Futuro da arquitectura de computadores



A ênfase de computação paralela é na programação deste tipo de arquitecturas

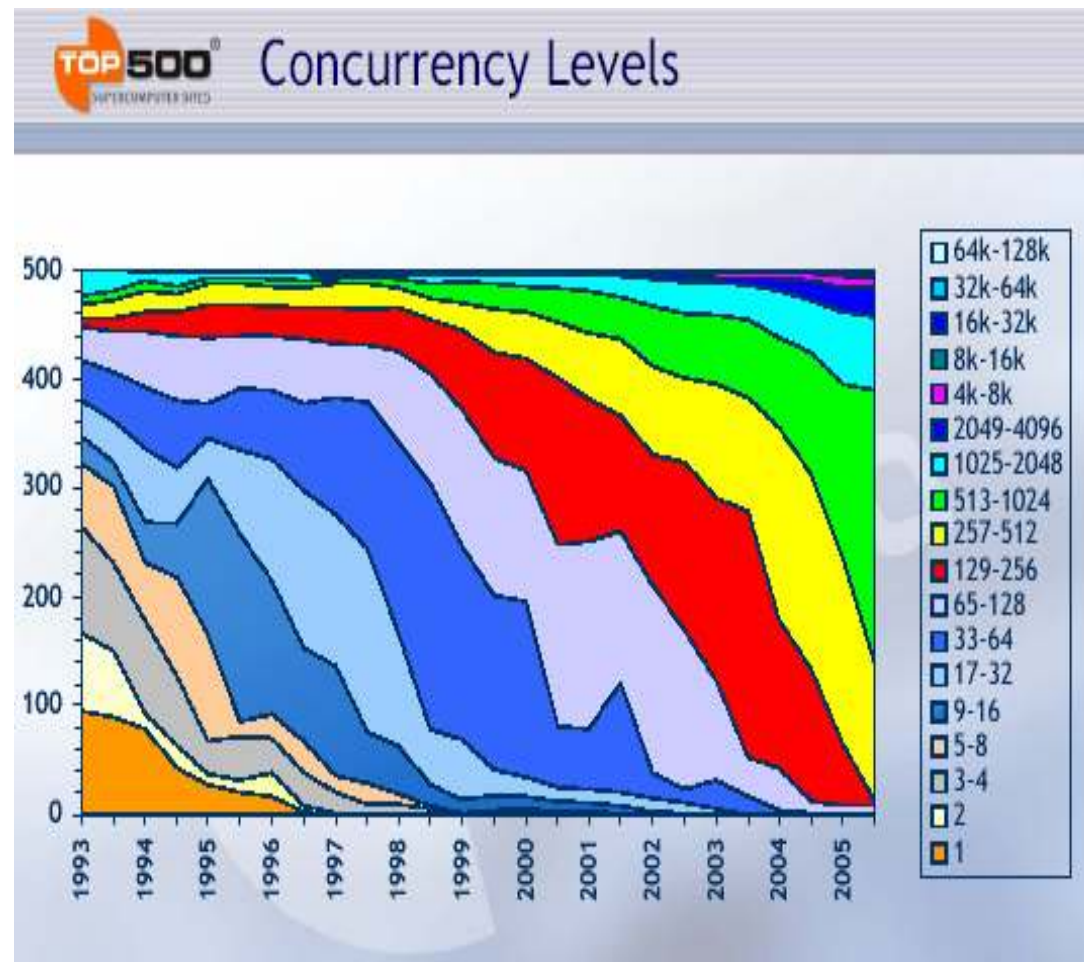
Computação Paralela

www.top500.org (Nov//2005)

	Manufacturer	Computer	Rmax [TF/s]	Installation Site	Country	Year	#Proc
1	IBM	BlueGene/L eServer Blue Gene	280.6	DOE/NNSA/LLNL	USA	2005	131072
2	IBM	BGW eServer Blue Gene	91.29	IBM Thomas Watson	USA	2005	40960
3	IBM	ASC Purple eServer pSeries p575	63.39	DOE/NNSA/LLNL	USA	2005	10240
4	SGI	Columbia Altix, Infiniband	51.87	NASA Ames	USA	2004	10160
5	Dell	Thunderbird	38.27	Sandia	USA	2005	8000
6	Cray	Red Storm Cray XT3	36.19	Sandia	USA	2005	10880
7	NEC	Earth-Simulator	35.86	Earth Simulator Center	Japan	2002	5120
8	IBM	MareNostrum BladeCenter JS20, Myrinet	27.91	Barcelona Supercomputer Center	Spain	2005	4800
9	IBM	eServer Blue Gene	27.45	ASTRON University Groningen	Netherlands	2005	12288
10	Cray	Jaguar Cray XT3	20.53	Oak Ridge National Lab	USA	2005	5200

Computação Paralela

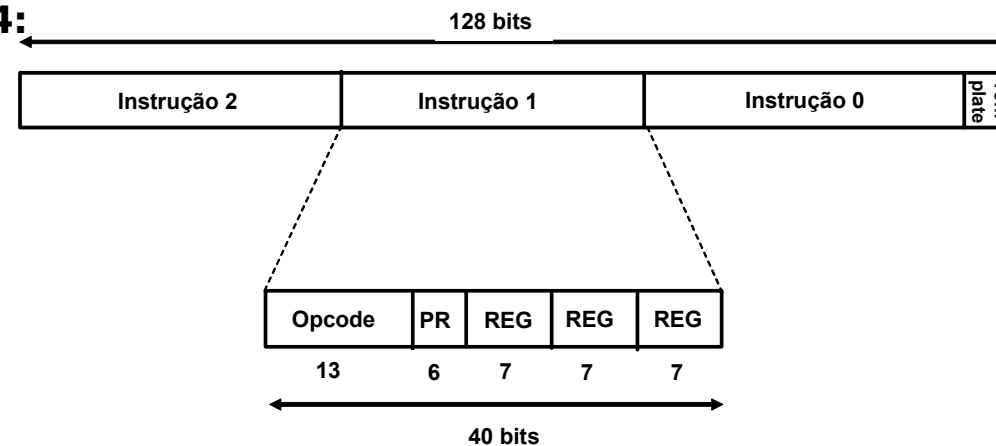
www.top500.org (Nov//2005)



Computação Paralela

VLIW (*Very Long Instruction Word*)

- ❑ O escalonamento dinâmico incrementa de forma considerável a complexidade do Hardware.
- ❑ VLIW efectua um escalonamento estático, sendo o compilador responsável por indicar as instruções que podem ser realizadas em paralelo.
=> tem-se revelado um fracasso
- ❑ O formato de instrução indica as operações que são realizadas em paralelo por cada unidade funcional.
- ❑ Exemplo IA-64:



Computação Paralela

VLIW (cont.)

- **Limitações de VLIW**
 - O código gerado tende a ser de maior dimensão, porque é necessário inserir *nop* nos campos da instrução não preenchidos.
 - Compatibilidade de código entre gerações dos mesmo processador uma vez que tende a expor a arquitectura interna do processador
 - É mais penalizado com *stalls* que o escalonamento dinâmico

- **EPIC – IA-64 / Itanium**
 - 64 registos de inteiros + 64 registos FP, ambos com 64 bits
 - 3 instruções em 128 bits (LIW?)
 - menos bits que VLIW clássico, produzindo código mais compacto
 - possibilidade de ligação entre os vários grupos de instruções
 - Verificação de dependências em HW => compatibilidade de código

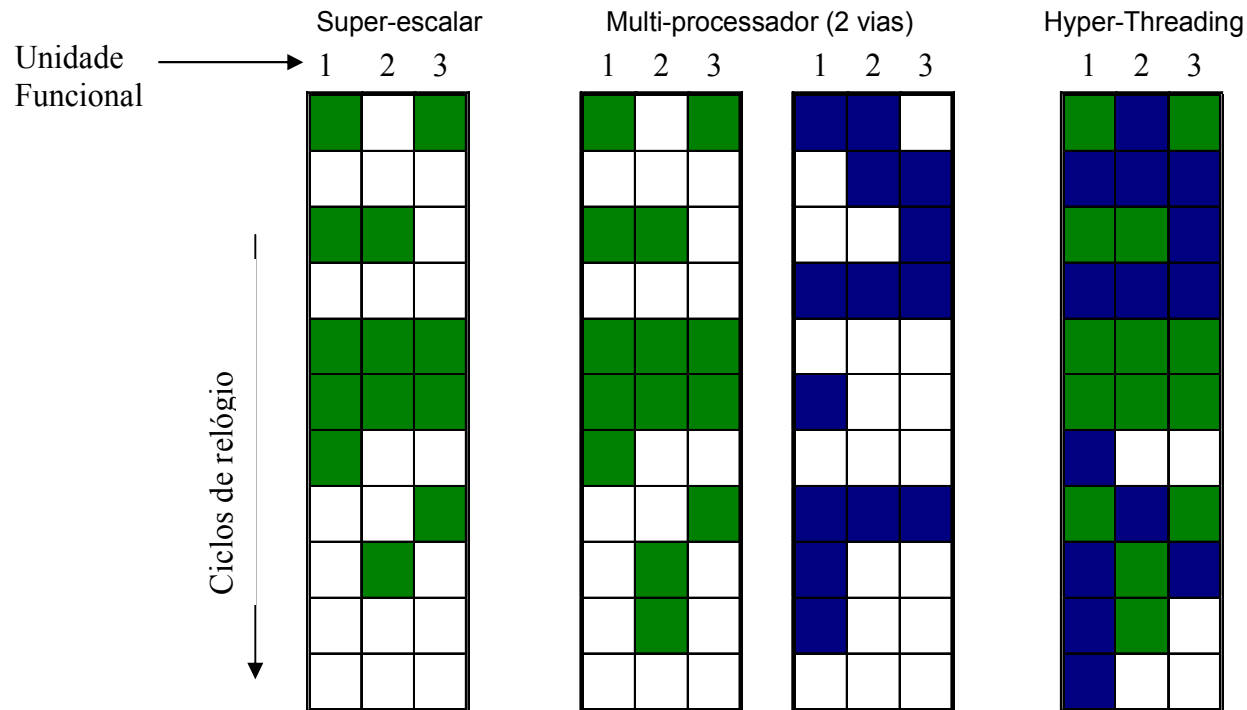
Computação Paralela

Hyper-Threading (Intel)

- ❑ **A adição de mais unidades funcionais nos processadores actuais não conduz a ganhos significativos de desempenho.**
- ❑ **O processador tem a capacidade de simular vários processadores (CC-NUMA) ou pode executar vários fios de execução em simultâneo.**
 - **baseia-se mais no paralelismo ao nível dos fios de execução ou dos processos e menos paralelismo ao nível da instrução.**
- ❑ **Implica um aumento de <5% de transístores por processador e pode conduzir a ganhos no desempenho até 35% (Xeon actuais, com grau 2).**
- ❑ **Os ganhos de desempenho são proporcionados por uma utilização mais eficiente dos recursos**

Computação Paralela

Hyper-Threading (cont.)



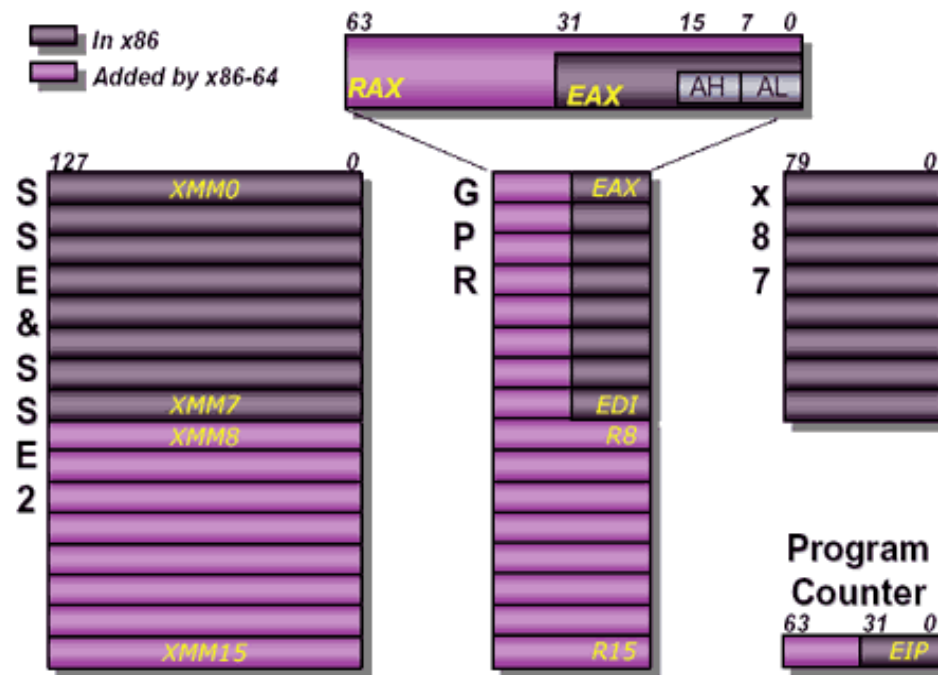
1 processador = 2 processadores ?

Não porque parte dos recursos do processador não são duplicados (i.é., são divididos pelos vários fios de execução) (*caches*, registros internos, *buffers* internos, etc.)

Computação Paralela

Hammer (AMD com 64 bits)

- Estende o IA-32 para 64 bits:

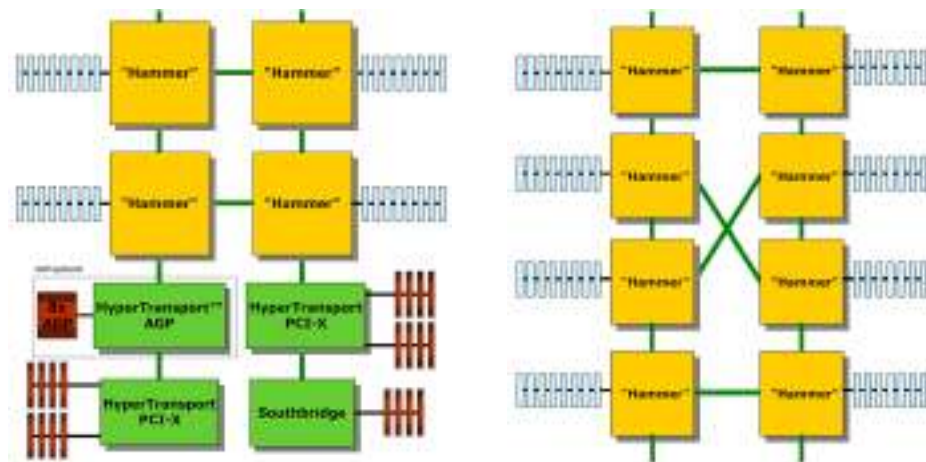


- A *pipeline* possui mais dois estágios (12 contra 10) que a arquitectura K7 (Athlon XP)

Computação Paralela

Hammer (cont.)

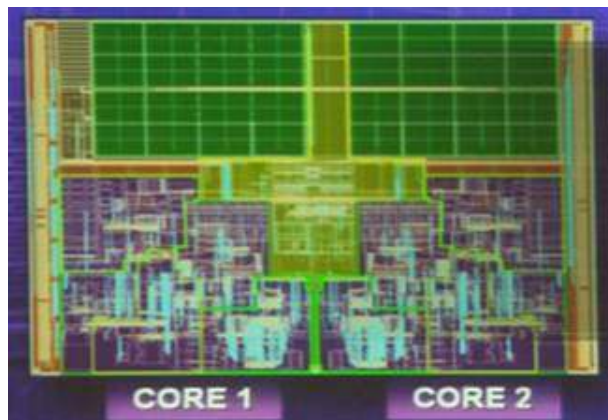
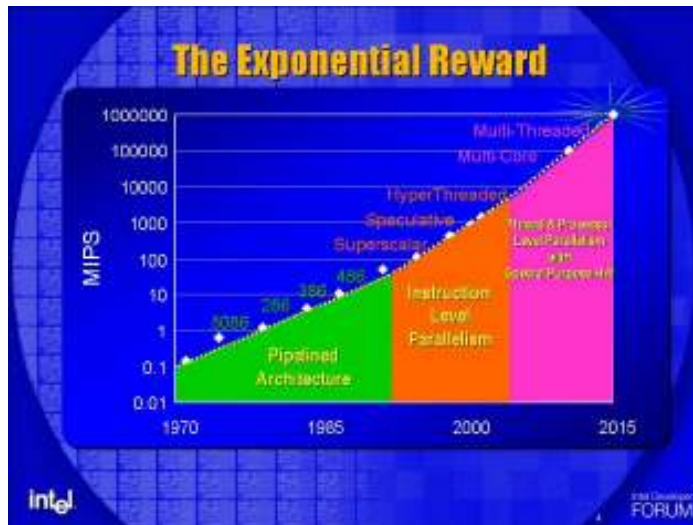
- ❑ Integra o controlador de memória do próprio chip e 2 ligações externas (Hyper Transport) para ligação a outros processadores.
- ❑ O Hyper Transport permite a ligação até 8 processadores sem recurso a componentes adicionais:



- ❑ A configuração multiprocessador assemelha-se mais ao tipo CC-NUMA, uma vez que cada processador possui o seu controlador de memória e a largura de banda disponível para a memória local é superior à disponibilizada pelo Hyper Transport, embora a velocidade do Hyper Transport escale com a velocidade do processador

Computação Paralela

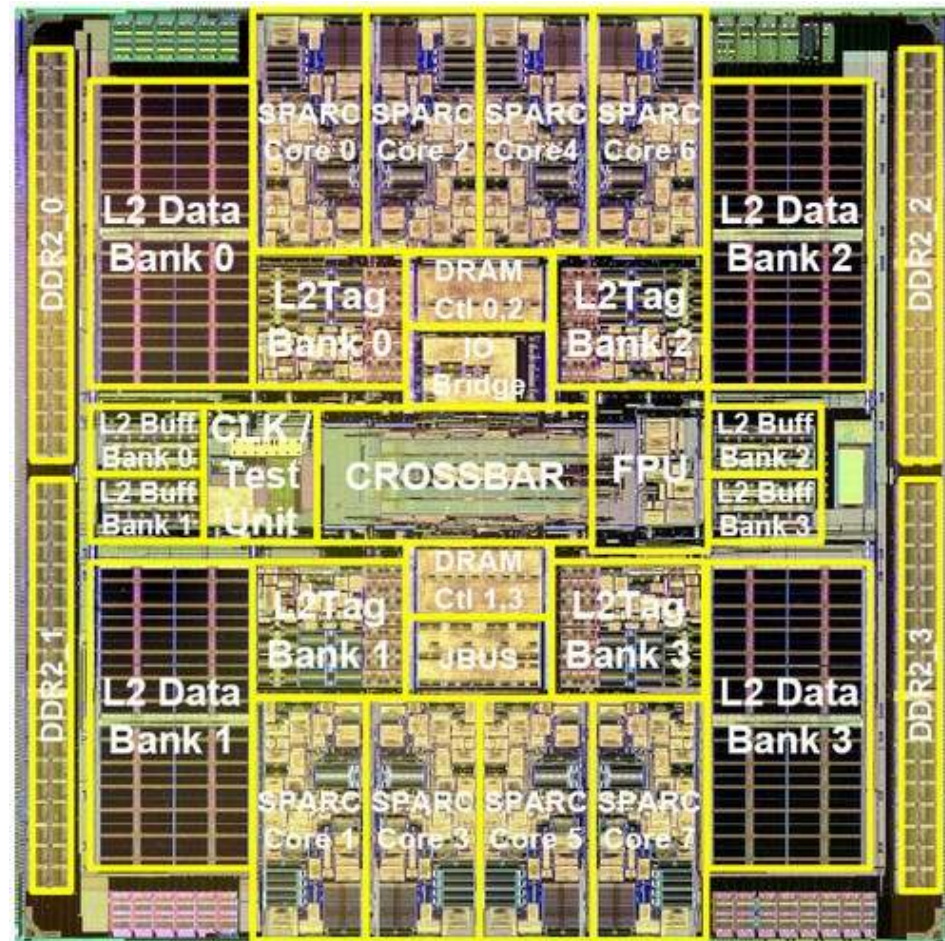
Yonah (centrino dual-core)



Computação Paralela

UltraSparc T1

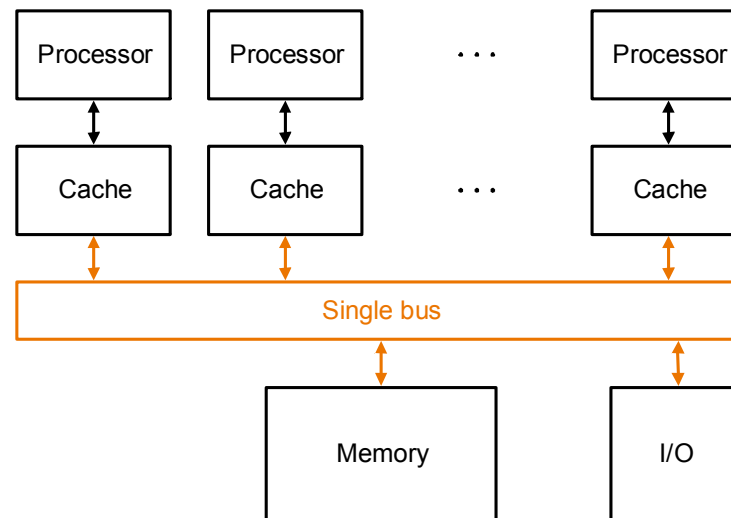
- 8 cores
- 32 *threads* simultâneas (4/core)
- 4 *caches* L2



Computação Paralela

Memória partilhada centralizada (CC-UMA, CC-NUMA)

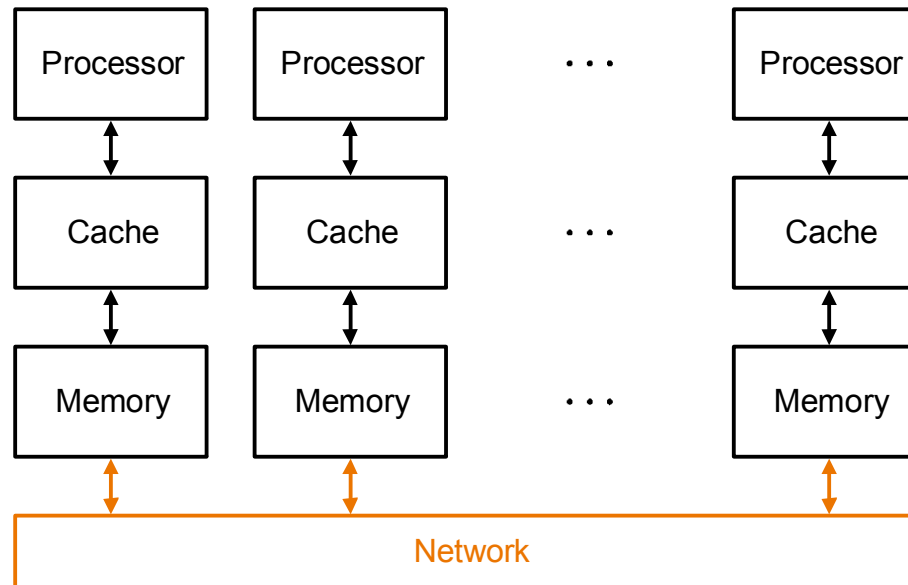
- ❑ Vários processadores partilham um barramento de acesso à memória
- ❑ As *caches* de cada processador contribuem para reduzir o tráfego no barramento e a latência dos acessos à memória
- ❑ um valor pode estar replicado em vários sítios => são necessários mecanismos para assegurar a coesão entre as *caches* dos vários processadores e a memória
- ❑ A largura de banda de acesso à memória é partilhada pelos vários processadores => limitação à escalabilidade deste tipo de arquitectura:



Computação Paralela

Memória distribuída – processadores conectados por uma rede de interligação (CC-NUMA, passagem de mensagens)

- ❑ Os sistemas interligados por um barramento tendem a limitar o número de processadores que efectivamente podem ser ligados
- ❑ A alternativa reside na utilização de uma rede dedicada à interligação dos vários processadores, possuindo cada processador a sua memória dedicada



Computação Paralela

Cluster de máquinas

- ❑ **Constituídos pode HW "normal", interligados por uma rede de alta velocidade (Gbit/s com muito baixa latência)**
- ❑ **Cada nodo de processamento pode ser uma máquina de memória partilhada com vários processadores**
- ❑ **Cada nodo possui uma cópia do SO**
- ❑ **Alta disponibilidade: quando falha uma máquina basta substituí-la por outra**
- ❑ **Custos de administração na ordem do número de máquinas**
- ❑ **Exemplo (Cluster investigação da Universidade do Minho - 2006):**
 - **8 nodos com SAN FC 3TB**
 - **8 nodos com NVIDIA 7800 GTX**

CPU	Processador	Freq. Relógio	Memória Actual/Máxima	Comunicação	Nodos	Máximo de nodos
8	AMD Athlon MP	1,5 GHz	4GB/16 GB	250 MB/s, 10us (2,0 Gbit/s)	2-way	4
14	Intel Xeon	3,2 GHz	10 GB / 112 GB	250 MB/s, 10us (2,0 Gbit/s)	2-way	7
96	Intel Xeon	3,0 GHz 2,0 GHz	96GB / 768GB	1 GB/s, 3us (10 Gbit/s)	2-way / 4-way	48

Computação paralela : Cluster Search

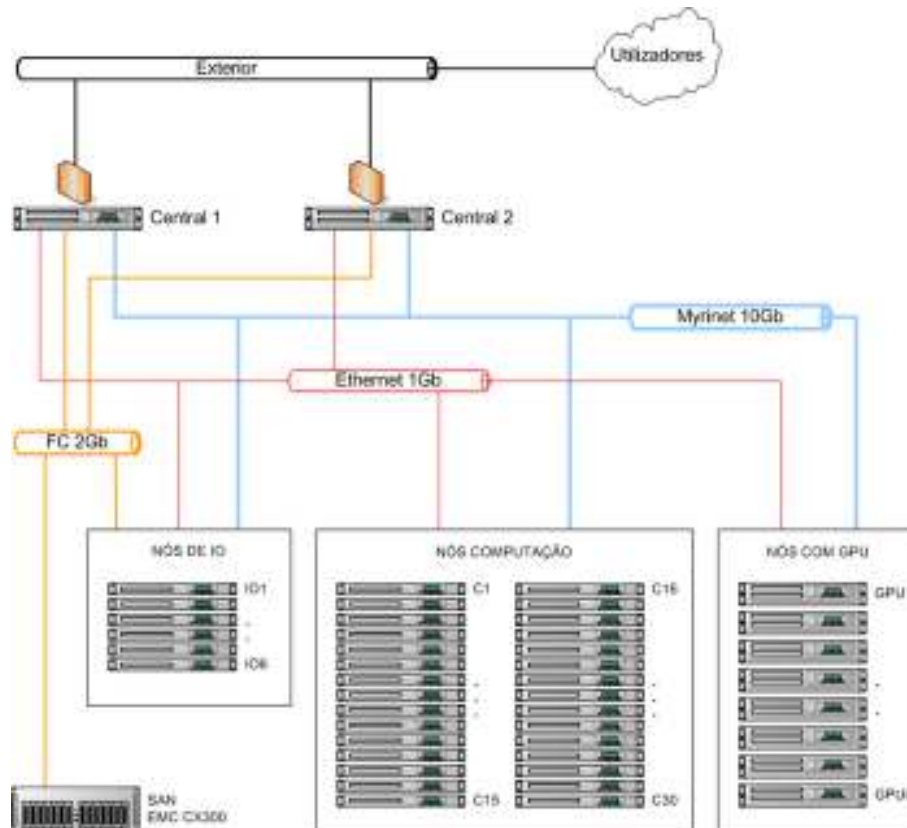
Services and Advanced Research Computing with HTC/HPC clusters

- ❑ 30 nós duplo Xeon
3.2GHz, 2 GB RAM
- ❑ 8 nós duplo Xeon-dual core
2.0 GHz, 4 GB RAM
- ❑ rede 10Gbit/s Myrinet
- ❑ armazenamento central SAN
com 3TB
- ❑ Total de 92 cores,
92 GB RAM !



Computação paralela : Cluster Search

Services and Advanced Research Computing with HTC/HPC clusters



```
[jls@search ~]$ showq
ACTIVE JOBS-----
JOBNAME          USERNAME        STATE  PROC  REMAINING          STARTTIME
11019            mrocha          Running  1     4:19:20  Mon Sep 18 23:31:46
10976            hpinto          Running  8    1:23:58:47  Tue Sep 19 13:11:13
10977            hpinto          Running  8    2:16:11:29  Mon Sep 18 21:23:55
10818            cris            Running  1    8:07:46:43  Tue Sep 19 12:59:09
10938            mrocha          Running  1    9:20:19:42  Tue Sep 12 17:32:08
10939            mrocha          Running  1    9:20:20:23  Tue Sep 12 17:32:49
10940            mrocha          Running  1    9:20:20:25  Tue Sep 12 17:32:51
10941            mrocha          Running  1    9:20:20:31  Tue Sep 12 17:32:57
10996            rbrito          Running  1   10:05:00:16  Sun Sep 17 06:12:42
10984            rbrito          Running  2   11:09:10:29  Mon Sep 18 10:22:55
11004            exp             Running  2   13:15:55:09  Mon Sep 18 05:07:35
10942            mrocha          Running  1   14:00:20:39  Tue Sep 12 17:33:05
10943            mrocha          Running  1   14:00:20:41  Tue Sep 12 17:33:07
10944            mrocha          Running  1   14:00:20:43  Tue Sep 12 17:33:09
11026            mrocha          Running  1   20:16:25:28  Mon Sep 18 23:37:54
11027            mrocha          Running  1   20:16:26:35  Mon Sep 18 23:39:01
11028            mrocha          Running  1   20:16:27:28  Mon Sep 18 23:39:54
11029            mrocha          Running  1   20:16:28:45  Mon Sep 18 23:41:11
11030            mrocha          Running  1   20:16:29:40  Mon Sep 18 23:42:06
11031            mrocha          Running  1   20:16:30:55  Mon Sep 18 23:43:21
11032            mrocha          Running  1   20:16:32:03  Mon Sep 18 23:44:29
11033            mrocha          Running  1   20:16:32:27  Mon Sep 18 23:44:53
11034            mrocha          Running  1   20:16:32:32  Mon Sep 18 23:44:58
10577            nmama          Running  1   54:08:55:04  Mon Aug 21 14:07:30
10578            nmama          Running  1   64:13:33:41  Thu Aug 31 18:46:07
10579            nmama          Running  1   64:13:33:41  Thu Aug 31 18:46:07
10580            nmama          Running  1   64:13:33:44  Thu Aug 31 18:46:10
10581            nmama          Running  1   75:05:35:10  Mon Sep 11 10:47:36
```

28 Active Jobs 44 of 60 Processors Active (73.33%)
23 of 30 Nodes Active (76.67%)

```
IDLE JOBS-----
JOBNAME          USERNAME        STATE  PROC  WCLIMIT          QUEUETIME
0 Idle Jobs

BLOCKED JOBS-----
JOBNAME          USERNAME        STATE  PROC  WCLIMIT          QUEUETIME
10819            cris            Idle   1    8:08:00:00  Wed Aug 30 01:24:02
10820            cris            Idle   1    8:08:00:00  Wed Aug 30 01:24:02
10972            rbrito          Idle   4    4:04:00:00  Fri Sep 15 09:58:25
```

Total Jobs: 31 Active Jobs: 28 Idle Jobs: 0 Blocked Jobs: 3

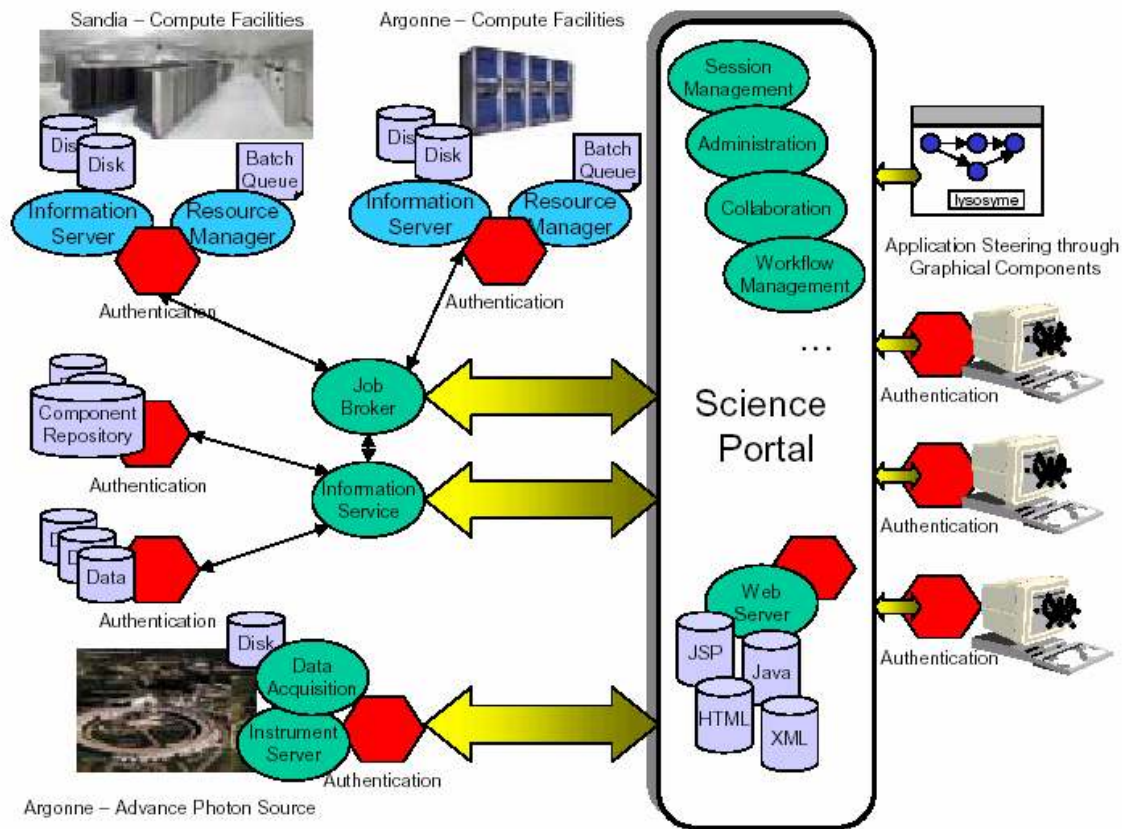
Computação Paralela

GRID

- ❑ **Pode ser a próxima geração da internet**
- ❑ **"A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities.", Foster 1998.**
- ❑ **"coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations.", Foster and AI, 2000.**
- ❑ **A partilha de recursos é altamente controlada, com os fornecedores e os utilizadores dos recursos a definirem claramente o que é partilhado, que está autorizada a utilizar cada recurso e em que condições a partilha ocorre.**
- ❑ **a coordenação de recursos não deve ser centralizada (os vários recursos encontram-se em domínios diferentes)**
- ❑ **deve utilizar protocolos standard e abertos**
- ❑ **deve fornecer vários níveis, não triviais, de serviço**
- ❑ **requer novos paradigmas de programação? (os recursos dinâmicos, altamente heterogéneos e sujeitos a grandes variações de desempenho)**
- ❑ **pode tornar, a longo prazo, os supercomputadores dispensáveis?**

Computação paralela

Aplicações de GRID



Computação paralela

- **Evolução das arquitecturas de computadores (Grid)**
 - **CERN LCG**



Site Reports	GIIS Host	totalCPU	freeCPU	runJob	waitJob	seAvail TB	seUsed TB
CERN-PROD	prod-bdii.cern.ch	18228	251	1153	0	931.40	465.66
INFN-T1	ce03-lcg.cr.cnaf.infn.it	2826	1905	259	4444	93.55	46.62
FZK-LCG2	lcg-gridka-ce.fzk.de	2656	1408	12	0	45.56	11.24
IN2P3-CC	cclcgip01.in2p3.fr	1581	1406	171	4	16.98	0.62
USCMS-FNAL-WC1	cmslsgce.fnal.gov	965	163	1584	9294	127.49	29.06
RAL-LCG2	site-bdii.gridpp.rl.ac.uk	858	0	99	13	52.66	60.30
SARA-LISA	mu9.matrix.sara.nl	714	132	2	0	8.44	12.16
QMUL-eScience	ce01.esc.qmul.ac.uk	574	341	111	0	13.28	3.65
TORONTO-LCG2	bigmac-lcg- ce.physics.utoronto.ca	491	69	31	3	1.12	0.37
UKI-LT2-IC-LeSC	mars- ce.mars.lesc.doc.ic.ac.uk	400	338	94	0	0.01	0.00
UKI-NORTHGRID-MAN-HEP	ce01.tier2.hep.manchester.ac. uk	400	289	111	0	0	0
IN2P3-LPC	clrlcgce02.in2p3.fr	360	240	120	1	2.71	0.46
INFN-CATANIA	grid012.ct.infn.it	338	68	148	0	2.67	0.86
INFN-LNL-2	t2-ce-02.lnl.infn.it	316	188	64	3	0.72	1.23
Taiwan-LCG2	lcg00125.grid.sinica.edu.tw	284	201	82	3	10.54	13.22
DESY-HH	grid-giis.desy.de	284	150	145	29	8.93	3.44
scotgrid-gla	ce1-gla.scotgrid.ac.uk	262	262	0	0	3.58	0.39
CYFRONET-LCG2	zeus02.cyf-kr.edu.pl	260	206	0	48884	0.35	1.64
WARSAW-EGEE	ce.polgrid.pl	256	9	228	1	3.05	0
...
		totalCPU	freeCPU	runJob	waitJob	seAvail TB	seUsed TB
		39411	12247	6955	17199 5	3433.62	1712.81

Computação paralela

Iniciativa nacional GRID

(www.gridcomputing.pt)

- "apoiar o desenvolvimento de estruturas de suporte à computação distribuída para a partilha de recursos na resolução de problemas complexos com necessidade de processamento intensivo de dados"
- "assegurar o desenvolvimento de competências e capacidades nacionais de interesse estratégico para a evolução deste tipo de computação distribuída"



Computação paralela

Tecnologias para desenvolvimento de aplicações em GRID

