UNIVERSIDAD PONTIFICIA ICAI ICADE COMILLAS MADRID

upcomillas es

upcomillas es

10-High Performance Computing

Advanced Computing Tools for Applied Research

(*Herramientas Computacionales Avanzadas para la Investigación Aplicada*)

Rafael Palacios, Fernando de Cuadra

MRE

Advanced Computing Tools for Applied Research
Contents

Implementing computational tools

1

Overview

Definition of HPC

- High-Performance Computing uses powerful computers to solve advanced computation problems
  - Supercomputers
  - Computer clusters
  - Computer grids

Top 500

- The fastest computers in the world
  (source Nov/2014: http://www.top500.org/)

RANK	SITE	SYSTEM	CORES	RMAX (TFLOP/S)	RPEAK (TFLOP/S)	POWER (KW)
1	National Super Computer Center in Guangzhou China	**Tianhe-2 (MilkyWay-2)** - TH-IVB-FEP Cluster, Intel Xeon E5-2692 12C 2.200GHz, TH Express-2, Intel Xeon Phi 31S1P NUDT	3,120,000	33,862.7	54,902.4	17,808
2	DOE/SC/Oak Ridge National Laboratory United States	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560,640	17,590.0	27,112.5	8,209
3	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1,572,864	17,173.2	20,132.7	7,890
4	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705,024	10,510.0	11,280.4	12,660
5	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786,432	8,586.6	10,066.3	3,945

Top 500

- The fastest computers in the world
  (source Nov/2012: http://www.top500.org/)

Rank	Site	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
1	DOE/SC/Oak Ridge National Laboratory United States	**Titan** - Cray XK7 , Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x Cray Inc.	560640	17590.0	27112.5	8209
2	DOE/NNSA/LLNL United States	**Sequoia** - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom IBM	1572864	16324.8	20132.7	7890
3	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect Fujitsu	705024	10510.0	11280.4	12660
4	DOE/SC/Argonne National Laboratory United States	**Mira** - BlueGene/Q, Power BQC 16C 1.60GHz, Custom IBM	786432	8162.4	10066.3	3945
5	Forschungszentrum Juelich (FZJ) Germany	**JUQUEEN** - BlueGene/Q, Power BQC 16C 1.600GHz, Custom Interconnect IBM	393216	4141.2	5033.2	1970

Top 500

- The fastest computers in the world
  (source **Nov/2010**: http://www.top500.org/)

Rank	Site	Computer/Year Vendor	Cores	R_{max} (Tflops)	R_{peak} (Tflops)	Power (kW)
1	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz 6C, NVIDIA GPU, FT-1000 8C / 2010 NUDT	186368	2566.00	4701.00	4040.00
2	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz / 2009 Cray Inc.	224162	1759.00	2331.00	6950.60
3	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU / 2010 Dawning	120640	1271.00	2984.30	2580.00
4	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows / 2010 NEC/HP	73278	1192.00	2287.63	1398.61
5	DOE/SC/LBNL/NERSC United States	Hopper - Cray XE6 12-core 2.1 GHz / 2010 Cray Inc.	153408	1054.00	1288.63	2910.00
6	Commissariat a l'Energie Atomique (CEA) France	Tera-100 - Bull bullx super-node S6010/S6030 / 2010 Bull SA	138368	1050.00	1254.55	4590.00

Top 500

- The fastest computers in the world
  (source **Nov/2010**: http://www.top500.org/)

				Tflops		kW
Rank	Site	Computer/Year Vendor	Cores	R_{max}	R_{peak}	Power
1	National Supercomputing Center	Tianhe-1A - NUDT TH MPP, X5670 2.93Ghz , NVIDIA GPU, FT-1000 8C / 2010 UDT	186368	2566.00	4701.00	4040.00
2	United States	guar - Cray XT5-HE Opteron 6-core 2.6 GHz 009 Cray Inc.	224162	1759.00	2331.00	6950.60
3	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU / 2010 Dawning	120640	1271.00	2984.30	2580.00
4	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP ProLiant SL390s G7 Xeon 6C X5670, Nvidia GPU, Linux/Windows / 2010 NEC/HP	73278	1192.00	2287.63	1398.61
5	DOE/SC/LBNL/NERSC United States		153408	1054.00	1288.63	2910.00
6	Commissariat a l'Energie Atomique (CEA) France		138368	1050.00	1254.55	4590.00

2012 becomes 8th

2012 – DOE
20x computing power
½ electric power

Top 500 (SPAIN)

2010

Rank	System	Procs	Memory(GB)	Rmax (GFlops)	Rpeak (GFlops)	Vendor
118	*MareNostrum* BladeCenter JS21 Cluster, PPC 970, 2.3 GHz, Myrinet	10240	20480	63830	94208	IBM
296	Cluster Platform 3000 BL2x220, E54xx 3.0 Ghz, Infiniband	4864	N/A	38292.4	48640	HP
338	Cluster Platform 3000 BL460c G1, Xeon L5410 2.33 GHz, GigE	7392	N/A	36390.6	68893.4	HP

2012

36	Barcelona Supercomputing Center Spain	**MareNostrum** - iDataPlex DX360M4, Xeon E5-2670 8C 2.600GHz, Infiniband FDR IBM	33664	636.9	700.2	699
290	Barcelona Supercomputing Center Spain	Bullx B505, Xeon E5649 6C 2.53GHz, Infiniband QDR, NVIDIA 2090 Bull SA	5544	103.2	182.9	82

2014

57	Barcelona Supercomputing Center Spain	**MareNostrum** - iDataPlex DX360M4, Xeon E5-2670 8C 2.600GHz, Infiniband FDR IBM	48,896	925.1	1,017.0	1,016

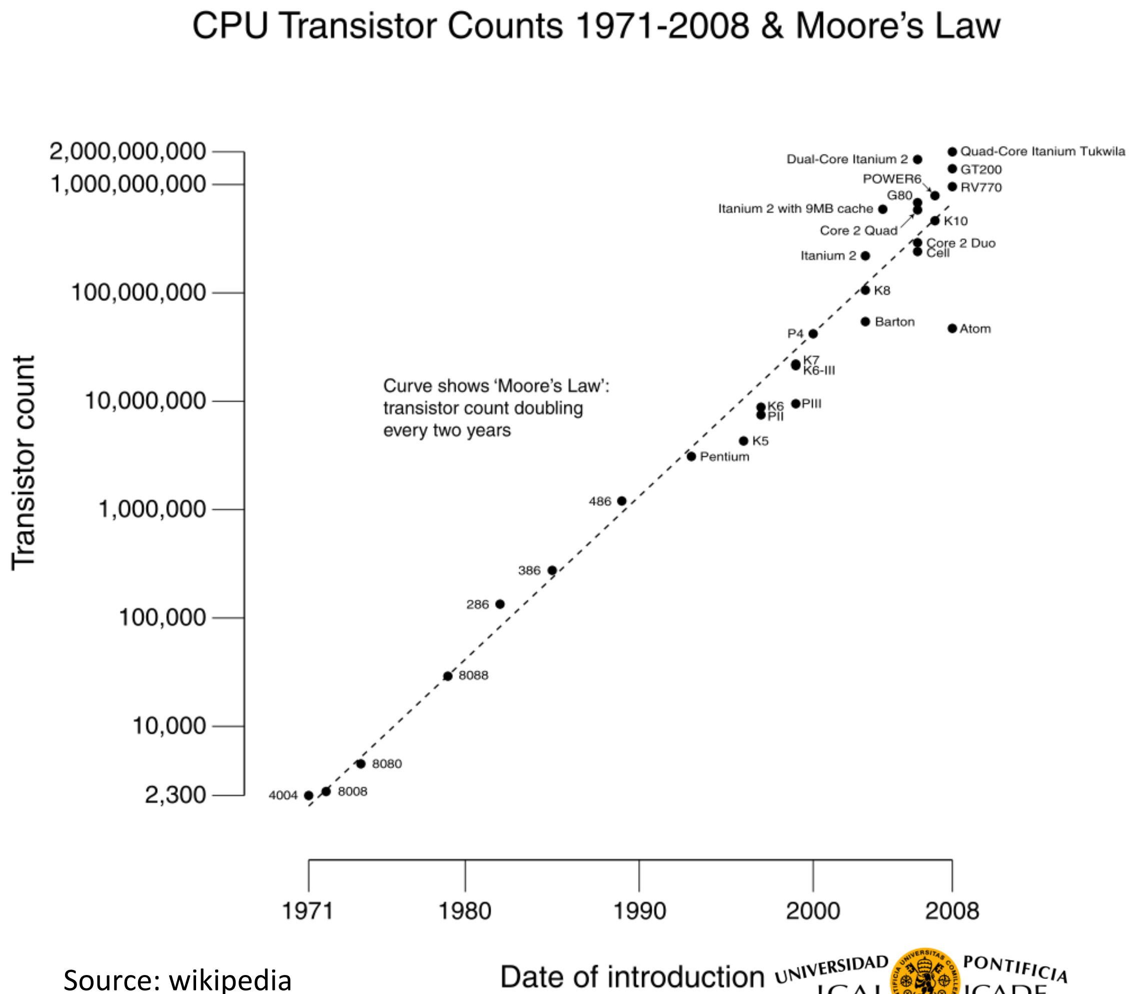
TFLOP/s

High performance in standard computers

- Moore's Law (1970)

"The number of transistors doubles every 2 years"

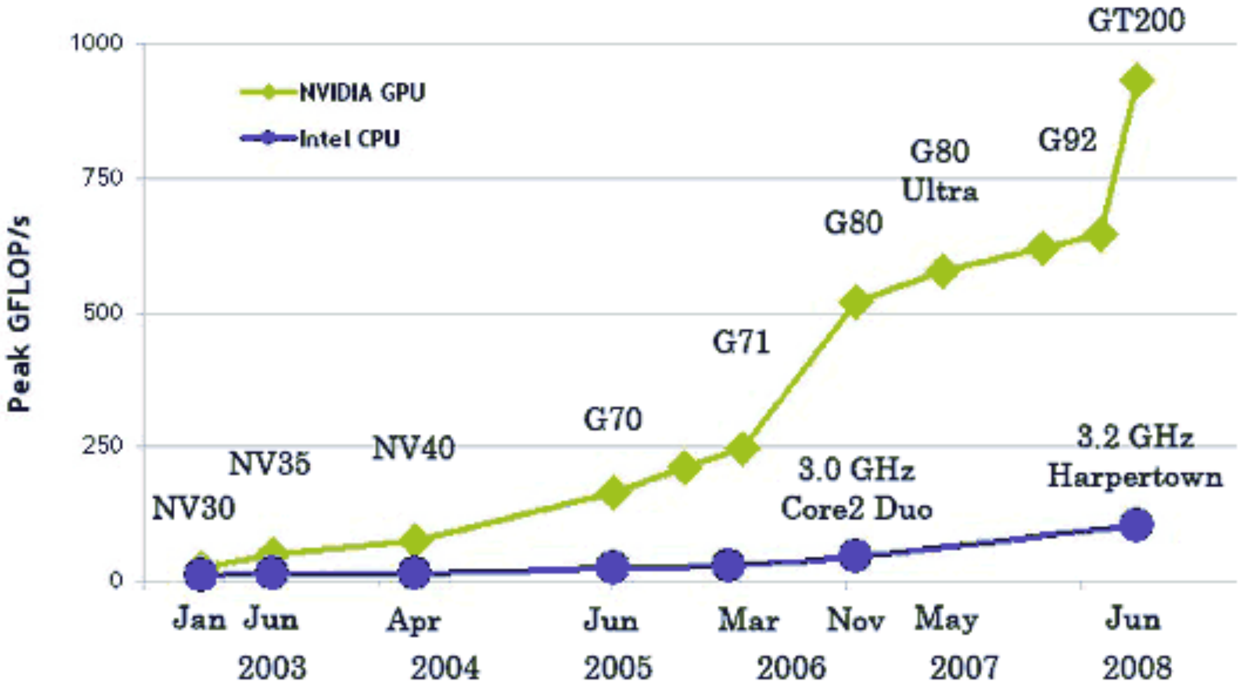
An Osborne Executive portable computer, from 1982, and an iPhone, released 2007. The Executive weighs 100 times as much, has nearly 500 times the volume, cost 10 times as much, and has a 100th the clock frequency of the iPhone.



Source: wikipedia

High performance in standard computers

- Parallel processing capabilities
  - Multi-core: 2 or 4
  - Multi-processor: 1, 2 or 4(server)
  - Graphical Processor Units (GPU)



Source: nvidia.com

Parallel processing methods

- Classification according to processing method and data management

	Single data	Multiple data
Single Instruction	SISD	SIMD or SPMD
Multiple Instruction	MIDS	MIMD

Parallel processing methods

- Classification according to processing method and data management

	Single data	Multiple data
Single Instruction	SISD	SIMD or SPMD
Multiple Instruction	MIDS	MIMD

SISD → Normal computer

Parallel processing methods

- Classification according to processing method and data management

	Single data	Multiple data
Single Instruction	SISD	SIMD or SPMD
Multiple Instruction	MIDS	MIMD

Today's video cards
Simulation

Parallel processing methods

- Classification according to processing method and data management

	Single data	Multiple data
Single Instruction	SISD	SIMD or SPMD
Multiple Instruction	MIDS	MIMD

No use

Parallel processing methods

- Classification according to processing method and data management

	Single data	Multiple data
Single Instruction	SISD	SIMD or SPMD
Multiple Instruction	MIDS	MIMD

→ MIMD:

Multi-processors with distributed memory
Grid systems

UNIVERSIDAD PONTIFICIA ICAI ICADE COMILLAS MADRID

Distributed vs Parallel

- Distributed systems
  - Resources are separated
  - Could be local network or Internet
  - Better reliability

Grid systems
Cloud computing
Amazon Elastic Compute Cloud (EC2)

- Parallel systems
  - One piece
  - Several processors in one server, or several servers in a fast local network
  - Better locality

Multicore system
Supercomputer

2

Parallel performance

Amdahl's Law (1967)

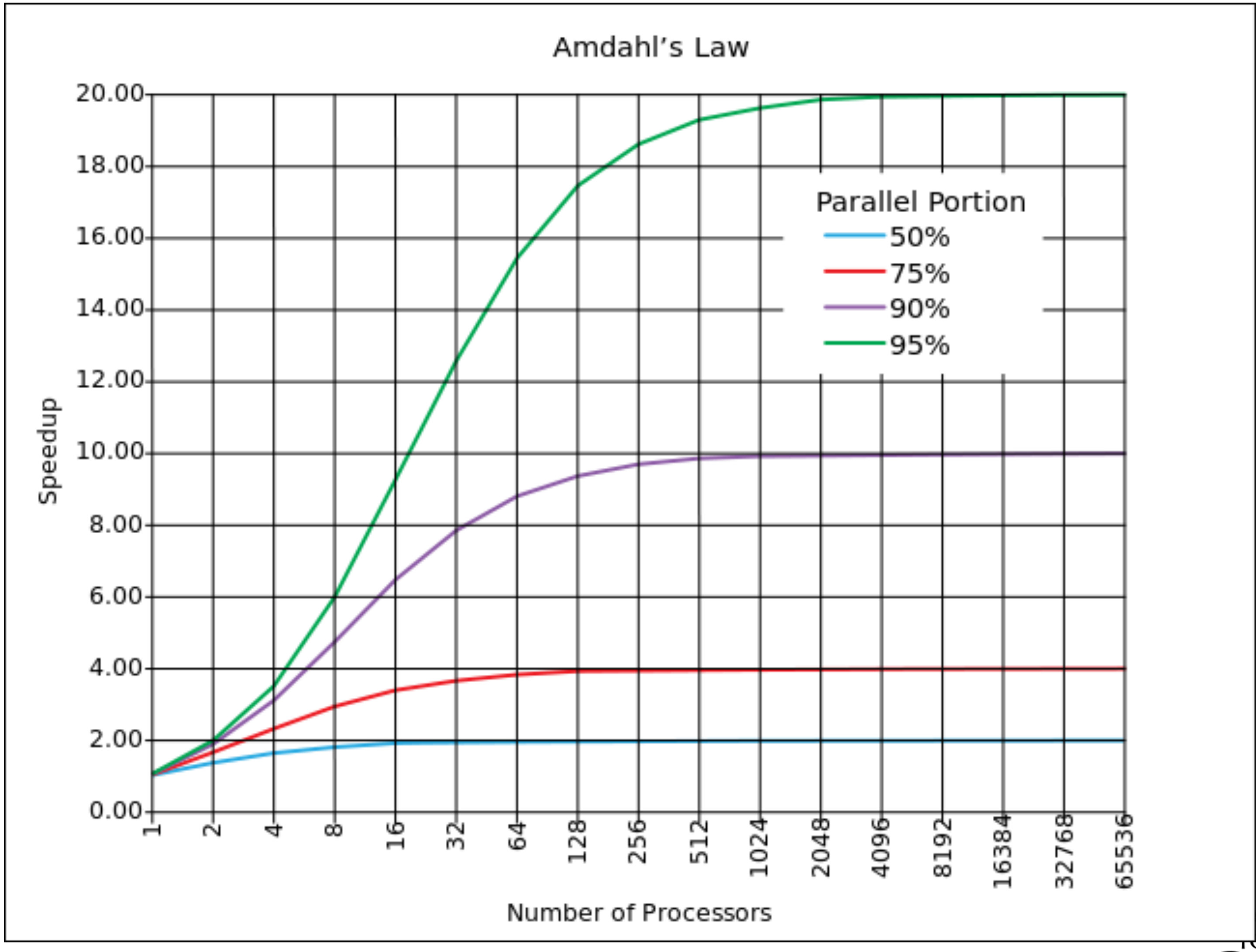
- If 1/s of a computation is sequential, then the maximum performance improvement by parallelism is at most a factor of s.

$$T_{sequential} = \frac{1}{s}T_{sequential} + (1-\frac{1}{s})T_{sequential}$$

$$T_{parallel} = \frac{1}{s}T_{sequential} + \frac{(1-\frac{1}{s})T_{sequential}}{\#\ processors}$$

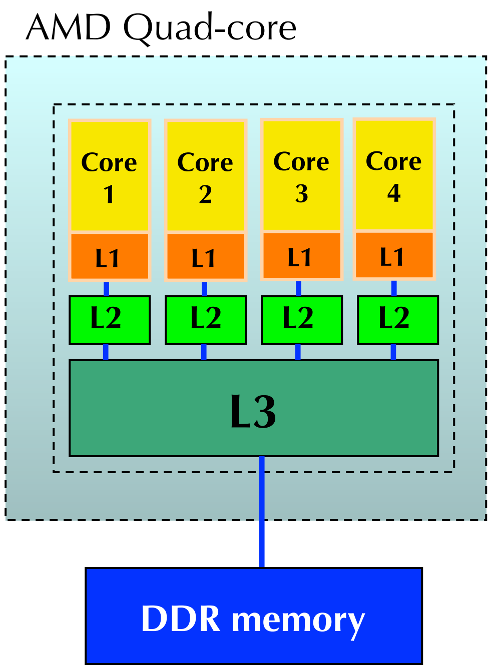
$$speedup = \frac{T_{sequential}}{T_{parallel}} = s$$ (For infinite #processors)

Amdahl's Law (1967)



Real-life performance

- Again, It's all about locality!



Latency	Transfer-rate
0.5 ns	6000 GB/s
1 ns	
5 ns	1000 GB/s
60 - 100 ns local remote	50 GB/s

Real-life performance

- Memory access is a bottleneck

- In the old days:
  - Read memory: 1 CPU cycle
  - FLOP: several cycles

- Today's computers
  - Read memory: up to 1000 cycles
    (due to DRAM latency compared with CPU speed)
  - FLOP: one cycle

3

Parallel programming

Parallel Programming

- Typical parallel programming languages are:
  - **MPI** (Message Passing Interface)
    - Distributed memory virtual machine
    - Evolution of PVM (parallel virtual machine)
  - **OpenMP** (Open Multi-Processing)
    - Similar to MP
    - Available for multiple standard programming languages
  - **CUDA** (Compute Unified Device Architecture)
    - Library for NVIDIA graphic cards
    - Available for multiple standard programming languages

Cloud computing

- Advantages of cloud computing:
  - You don’t need to maintain hardware infrastructure
  - Cost depends on CPU usage
  - Reliable, Scalable, Easy to start, Easy to use

- **Cloud computing providers**
  - Amazon Elastic Compute Cloud
  - Microsoft Azure
  - Salesforce
  - Google
  - IBM

4

Parallel Matlab

Parallel Matlab

- There are three main methods to implement parallel applications in Matlab:
  - Matlab official toolboxes
  - pMatlab
  - StarP

Parallel Matlab: Toolboxes

- Parallel Computing Toolbox
  - up to 8 cores in a single computer
  - compatible with DC server
- MATLAB Distributed Computing Server
  - For clusters and grid systems
  - Installs on the supercomputer
  - Is accessed by clients using Parallel Computing Toolbox
- Support for Amazon's EC2
  - http://www.mathworks.com/programs/techkits/ec2_paper.html

Parallel Matlab: Toolboxes

- Parallel Computing Toolbox example

```
matlabpool(4) % 4 cores (5 Matlab processes can be seen)

sz=2000000; %16MB vector
v=rand(sz,2);
res=zeros(sz,2); %for memory allocation
tic
parfor i=1:2
        res(:,i)=sort(v(:,i));  %each iteration independent of the
previous
end
toc

matlabpool close  %back to normal (when you are done)
```

Warning: Memory allocation and data transfer may be more time consuming that the computation itself.

Parallel Matlab: pMatlab

- Developed at Lincoln Laboratory
- MIT student Jeremy Kepner
- Based on MPI
- Freeware

Example:

```
unzip pMatlab
cd pMatlab
startup.m
cd examples
cd mandelbrot
run pMandelbrot:
  eval(pRun('pMandelbrot',1{}));  %1 core
  eval(pRun('pMandelbrot',4{}));  %4 cores
```

Parallel Matlab: starP

- Set of extensions for parallel programming in Matlab
- Now belongs to Microsoft

Without Star-p

```
>> clear all
>> a = rand(10000);
>> b = rand(10000);
>> tic; c = a * b; toc;
Elapsed time is 223.620259 seconds.
```

With Star-p

```
>> clear all
>> a = rand(10000*p);
>> b = rand(10000*p);
>> tic; c = a * b; toc;
Elapsed time is 27.684885 seconds.
```

Instituto de Investigación Tecnológica
C/ Santa Cruz de Marcenado, nº 26
28015 Madrid
Tel +34 91 542 28 00
Fax + 34 91 542 31 76
info@iit.upcomillas.es

www.iit.upcomillas.es