Telefónica
Investigación y Desarrollo Chile

High Performance Computing & Big Data (& Deep Learning)

Oscar Peredo A.
Telefónica I+D Chile
oscar.peredoa@telefonica.com

13 Abril 2016

ELEGIMOS TODO_

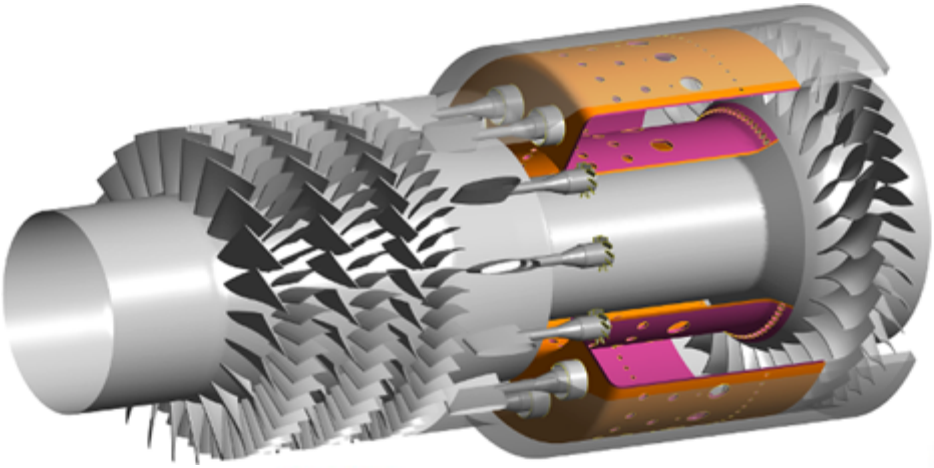
Contenidos

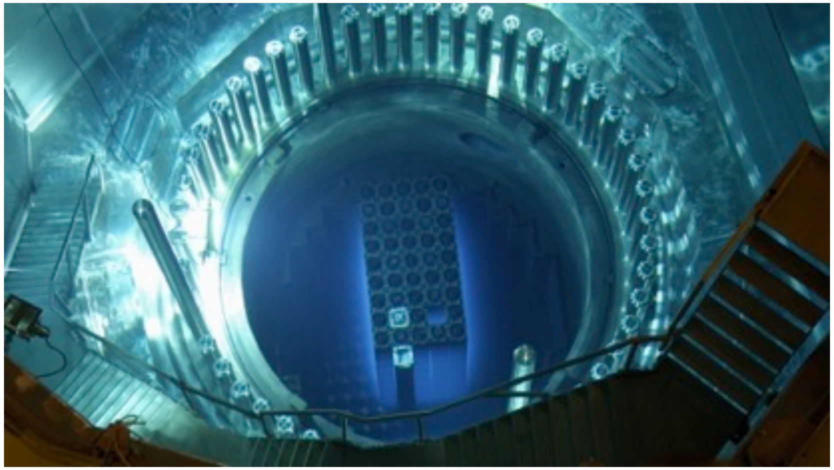
- ¿Que es Deep Learning?
- Historia
- Terminología
- Arquitecturas
- Theano
- Ejemplo MNIST
- Otros temas relacionados

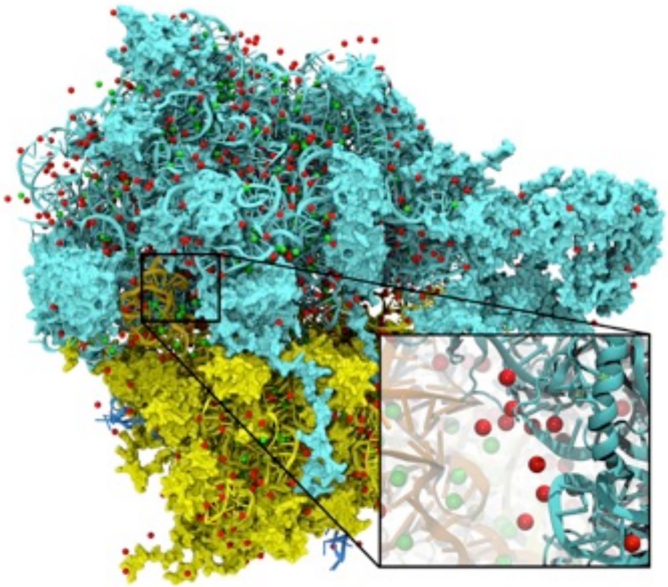
HPC y Big Data

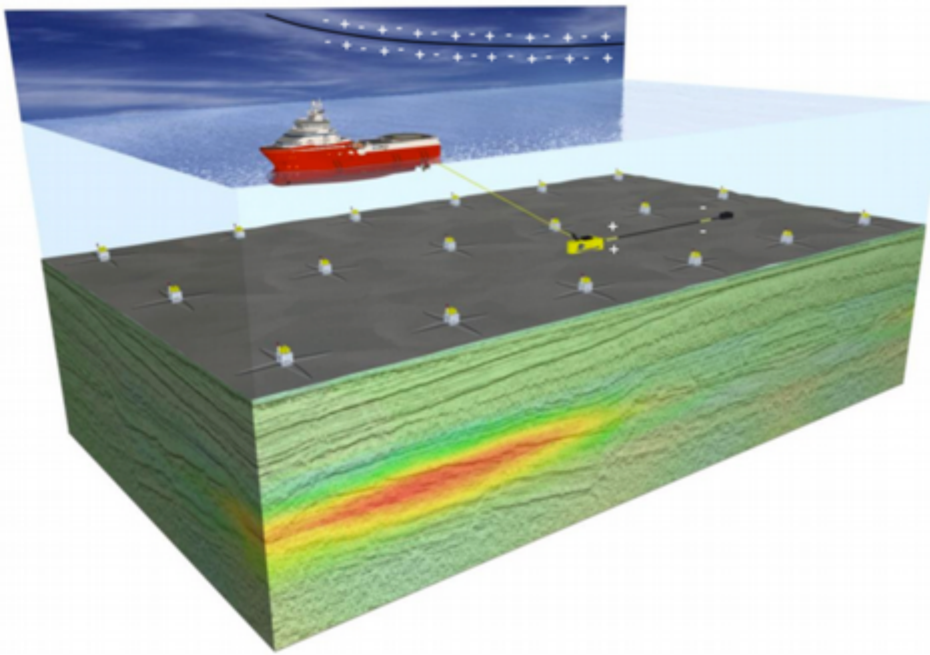
	HPC	Big Data
Desde	1960’s (Cray, CDC, IBM)	2000’s (Yahoo, Google)
Impulsores	Ciencia & Ingenieria	Negocios & Bases de datos
Apps	Memory & CPU-bounded	Memory & I/O-bounded
Lenguajes	C/C++, Fortran	Java, Scala
Herramientas	MPI, OpeMP, CUDA	Hadoop, MapReduce, Tez, Spark
Foco	Rendimiento, Velocidad	Tolerancia a fallos, Bajos costos

HPC apps

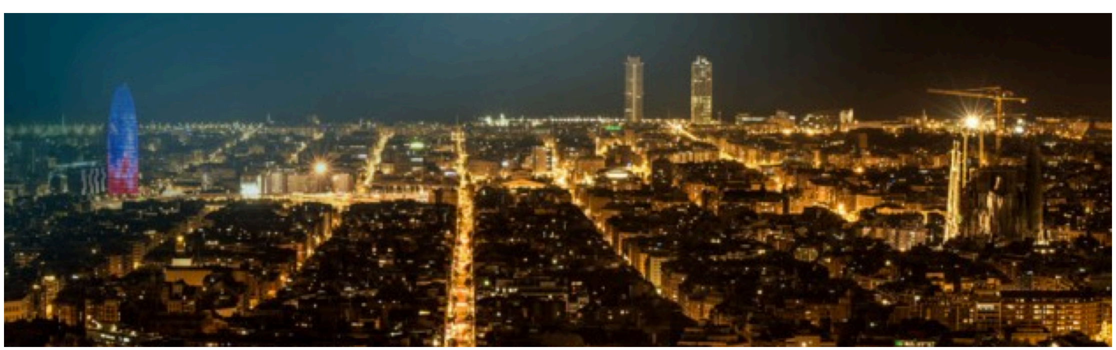




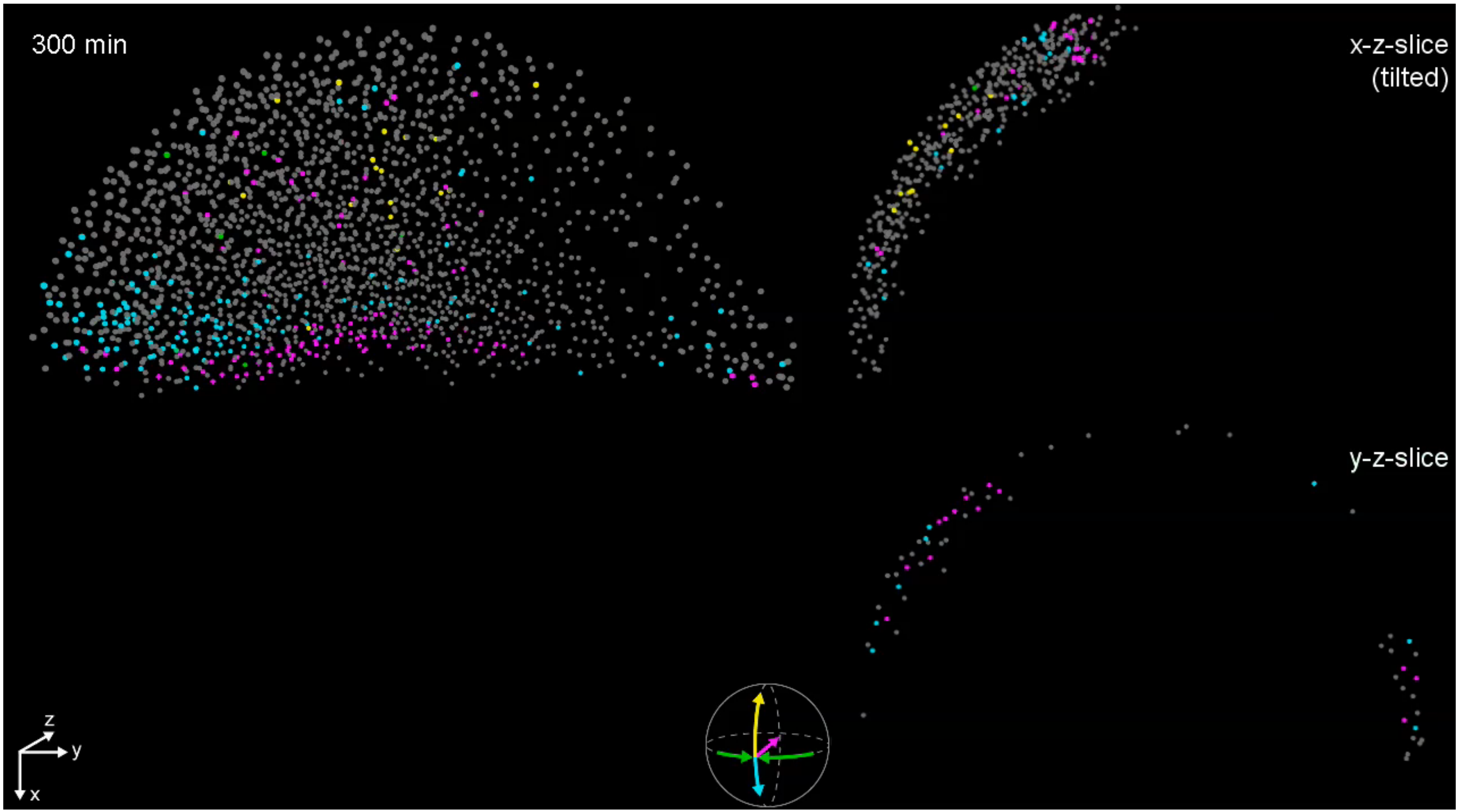




Big Data apps

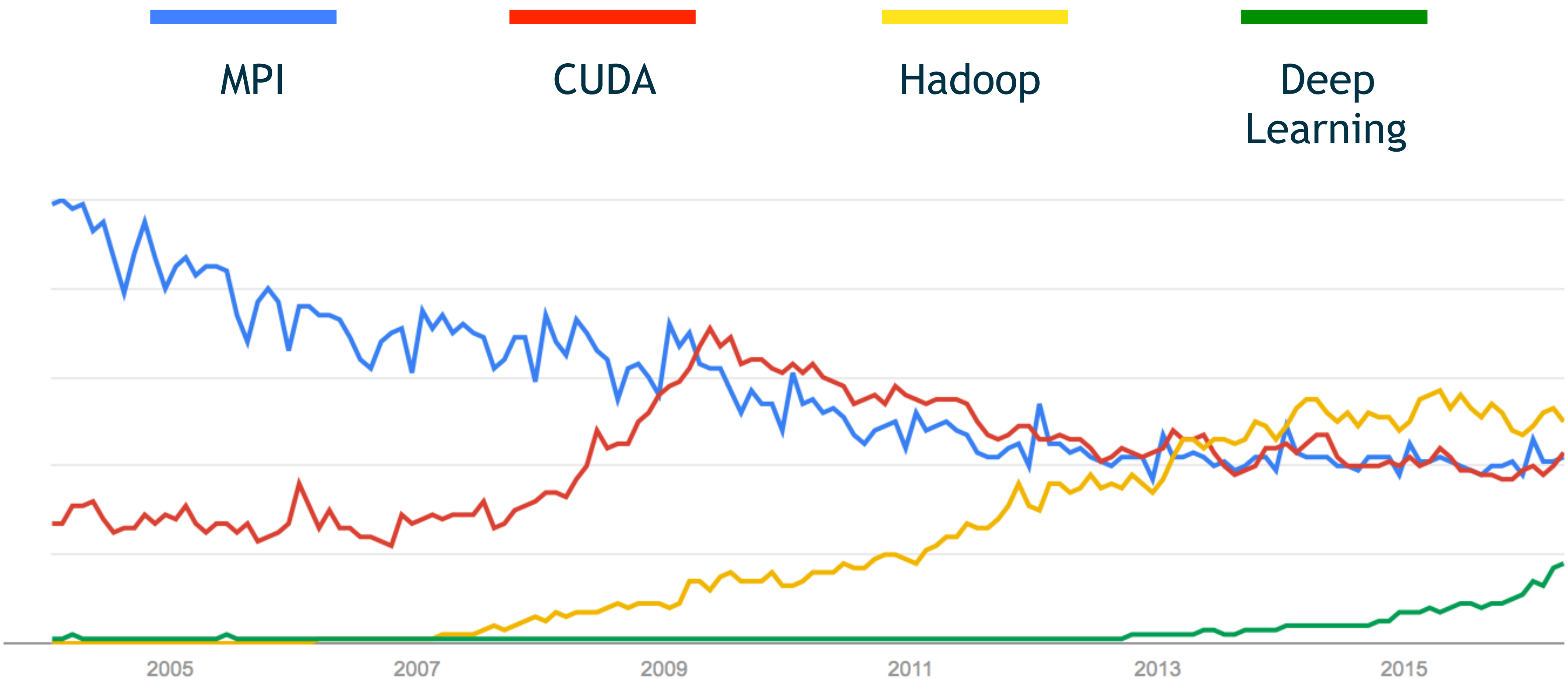


3.5TB~3500GB



http://www.embl.de/digitalembryo/fish.html#Movies

Google Trends



En un vistazo…

January 27, 2015

Three Ways Big Data and HPC Are Converging

Leo Reiter

Big data is becoming much more than just widespread distribution of cheap storage and cheap computation on commodity hardware. Big data analytics may soon become the new "killer app" for high performance computing (HPC).

There is more to big data than large amounts of information. It also pertains to massive distributed activities such as complex queries and computations (a.k.a analytics). In other words, deriving value through computation is just as "big" as the size of the data sets themselves. In fact, big data on HPC has already been coined by the analyst firm IDC – as High Performance Data Analysis.

HPC is well positioned to enable big data use cases through all three phases of typical workflows, including: data capture and filtering; analytics; and results visualization. In addition to the three phases, the speed of computation matters just as much as the scale. In order to unlock the full potential of big data, we have to pair it with "big compute," or HPC. Few industries can benefit as much from converged big data and HPC as the life sciences, where the data sets are enormous, the queries and comparisons intensive, and the visualizations complex.

Big Data Meets HPC

Fri, 03/07/2014 - 2:49pm

by Suzanne Tracy, Editor-in-Chief, Scientific Computing and HPC Source

Get the latest news in High Performance Computing, Informatics, Data Analysis Software and more - Sign up now!

Big Data Meets HPC

Addressing the major opportunities and challenges of the 21st century

High performance computing (HPC) has already contributed enormously to scientific innovation, industrial and economic competitiveness, national and regional security, and the quality of human life. This crucial role has been emphasized in recent years by U.S. and Russian presidents, as well as by senior officials in Europe and Asia.

Steve Conway, IDC VP HPC explains that, to date, most data-intensive HPC jobs in the government, academic and industrial sectors have involved the modeling and simulation of complex physical and quasi-physical systems. These systems range from product designs for cars, planes, golf clubs and pharmaceuticals, to subatomic particles, global weather and climate patterns, and the cosmos itself. However, he notes that from the start of the supercomputer era in the 1960s — and even earlier — an important subset of HPC jobs has involved analytics, attempts to uncover useful information and patterns in the data itself. For example, cryptography, one of the original scientific-technical computing applications, falls predominantly into this category.

Big Data Meets High Performance Computing

Intel® Enterprise Edition for Lustre* software and Hadoop combine to bring big data analytics to high performance computing configurations.

March 31, 2014

HPC and Big Data: A "Best of Both Worlds" Approach

Nicole Hemsoth

While they may share a number of similar, overarching challenges, data-intensive computing and high performance computing have some rather different considerations, particularly in terms of management, emphasis on performance, storage and data movement. Still, there is plenty of room for the two areas to merged, according to Indiana University's Dr. Geoffrey Fox.

Fox and his colleagues have been working on offering a basis of comparison for different data-intensive computing paradigms, including blending the "best of both worlds" via an MPI and Hadoop approach. "The success and evolution of the Apache big data stack [which includes Hadoop] into a widely deployed cluster computing framework yields many opportunities for traditional scientific applications," says the research team. They note that it's difficult to use both of the paradigms interoperably and the divergence of the two areas will likely continue, especially with so much support from the commercially-driven open source side of the big data stack.

Big Compute: The Collision of where HPC is Meeting the Challenges of Big Data

Fri, 03/07/2014 - 3:59pm

by Jason Stowe, Cycle Computing

Get the latest news in High Performance Computing, Informatics, Data Analysis Software and more - Sign up now!

En un vistazo…

International Supercomputing Conference 2016:

ISC High Performance

2016 FOCUS TOPICS

- Big Data and HPC Convergence
- Performance Modeling and Prediction
- Advanced Disaster Prediction and Mitigation
- State of the Art in Life Science & Computational Systems Biology
- Challenges in Extreme Engineering
- Exascale Architectures: Revolution versus Evolution
- Algorithms for Extreme Scale in Practice
- HPC in IoT and Robotics
- Deep Learning : Basics and Applications
- Connected Cars and Big Data

En un vistazo…

The White House

Office of the Press Secretary

For Immediate Release July 29, 2015

Executive Order -- Creating a National Strategic Computing Initiative

EXECUTIVE ORDER

\- - - - - - -

CREATING A NATIONAL STRATEGIC COMPUTING INITIATIVE

By the authority vested in me as President by the Constitution and the laws of the United States of America, and to maximize benefits of high-performance computing (HPC) research, development, and deployment, it is hereby ordered as follows:

Section 1. Policy. In order to maximize the benefits of HPC for economic competitiveness and scientific discovery, the United

artments, agencies, and offices
CI shall pursue five strategic

able exascale computing system
software capability to deliver
erformance of current 10 petaflop
lications representing government

the technology base used for
hat used for data analytic

ears, a viable path forward for
the limits of current
reached (the "post- Moore's Law

pability of an enduring national
a holistic approach that addresses
rking technology, workflow,
al algorithms and software,
evelopment.
-private collaboration to ensure
ch and development advances are,
between the United States
d academic sectors.

https://www.whitehouse.gov/the-press-office/2015/07/29/executive-order-creating-national-strategic-computing-initiative

En un vistazo…

The White House
Office of the Press Secretary

For Immediate Release

Executive Order -- National Strategic Initiative

EXECUTIV

- - - -

CREATING A NATIONAL STRAT

By the authority vested in me as Pr
the laws of the United States of An
of high-performance computing (H
deployment, it is hereby ordered as

Section 1. Policy. In order to maxi
economic competitiveness and sci

Sec. 2. Objectives. Executive departments, agencies, and offices (agencies) participating in the NSCI shall pursue five strategic objectives:

1. Accelerating delivery of a capable exascale computing system that integrates hardware and software capability to deliver approximately 100 times the performance of current 10 petaflop systems across a range of applications representing government needs.
2. Increasing coherence between the technology base used for modeling and simulation and that used for data analytic computing.
3. Establishing, over the next 15 years, a viable path forward for future HPC systems even after the limits of current semiconductor technology are reached (the "post- Moore's Law era").
4. Increasing the capacity and capability of an enduring national HPC ecosystem by employing a holistic approach that addresses relevant factors such as networking technology, workflow, downward scaling, foundational algorithms and software, accessibility, and workforce development.
5. Developing an enduring public-private collaboration to ensure that the benefits of the research and development advances are, to the greatest extent, shared between the United States Government and industrial and academic sectors.

https://www.whitehouse.gov/the-press-office/2015/07/29/executive-order-creating-national-strategic-computing-initiative

En un vistazo…

Advancing U.S. Leadership in High-Performance Computing

JULY 29, 2015 AT 4:09 PM ET BY THOMAS KALIL, JASON MILLER

Summary: Over the past 60 years, the United States has been a leader in the development and deployment of cutting-edge computing systems. Today, President Obama issued an Executive Order establishing the National Strategic Computing Initiative (NSCI) to ensure the United States continues leading in this field over the coming decades.

Over the past 60 years, the United States has been a leader in the development and deployment of cutting-edge computing systems. High-Performance Computing (HPC) systems, through their high levels of computing power and large amounts of storage capacity, have been and remain essential to economic competitiveness, scientific discovery, and national security.

Today, President Obama issued an Executive Order establishing the National Strategic Computing Initiative (NSCI) to ensure the United States continues leading in this field over the coming decades. This coordinated research, development, and deployment strategy will draw on the strengths of departments and agencies to move the Federal government into a position that sharpens, develops, and streamlines a wide range of new 21st century applications. It is designed to advance core technologies to solve difficult computational problems and foster increased use of the new capabilities in the public and private sectors.

n give clinicians tools to
t's disease, and to
o process large volumes
volume of data will
will shorten the time it

abyte storage capacity,
data analytics. This will
e coupled with actual
ods that require more
ficial intelligence to
e are also national
resistant vehicle

mands and emerging
ship for decades to
dress the pressing

EGIMOS TODO_

https://www.whitehouse.gov/blog/2015/07/29/advancing-us-leadership-high-performance-computing

En un vistazo...

the WHITE HOUSE PRESIDENT BARACK OBAMA

Contact Us ▸ Get Email Updates ▸

BRIEFING ROOM | ISSUES | THE ADMINISTRATION | PARTICIPATE | 1600 PENN

HOME · BLOG

The Precision Medicine Initiative (PMI) illustrates how data-driven HPC can give clinicians tools to better understand the complex biological mechanisms underlying a patient's disease, and to better predict the most effective treatments. Central to PMI is the ability to process large volumes of health and genomic data. As DNA sequencing technology improves, the volume of data will continue to increase and so too will the computational requirements. NSCI will shorten the time it takes to sequence samples and improve accuracy.

As NSCI drives forward these two goals of exaflop computing ability and exabyte storage capacity, it will also find ways to combine large-scale numerical computing with big data analytics. This will enable new forms of computation, including simulations of weather that are coupled with actual observations from weather satellites. It will also enable new analytic methods that require more extensive numerical processing, such as emerging techniques that use artificial intelligence to automatically learn new capabilities from large numbers of examples. There are also national security benefits, including using modeling and simulation to improve IED-resistant vehicle designs.

By strategically investing now, we can prepare for increasing computing demands and emerging technological challenges, building the foundation for sustained U.S. leadership for decades to come, while also expanding the role of high-performance computing to address the pressing challenges faced across many sectors.

Advancin

JULY 29, 2015 AT 4:09 P

Summary: Over t
cutting-edge com
National Strateg
over the coming

Over the past 60 y
of cutting-edge co
high levels of com
essential to econ

Today, President O
Initiative (NSCI) t
decades. This coo
strengths of depa
sharpens, develops, and streamlines a wide range of new 21st century applications. It is designed to advance core technologies to solve difficult computational problems and foster increased use of the new capabilities in the public and private sectors.

EGIMOS TODO_

https://www.whitehouse.gov/blog/2015/07/29/advancing-us-leadership-high-performance-computing

En un vistazo…

Desde la Casa Blanca también están impulsando la convergencia…

The White House
Office of the Press Secretary

For Immediate Release — July 29, 2015

Executive Order -- Creating a National Strategic Computing Initiative

EXECUTIVE ORDER

\- - - - - - -

CREATING A NATIONAL STRATEGIC COMPUTING INITIATIVE

By the authority vested in me as President by the Constitution and the laws of the United States of America, and to maximize benefits of high-performance computing (HPC) research, development, and deployment, it is hereby ordered as follows:

Section 1. Policy. In order to maximize the benefits of HPC for

THE WHITE HOUSE
WASHINGTON

http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/BDEC2_Convergence_Reed_Beckman.pdf

En un vistazo…

Desde la Casa Blanca también están impulsando la convergencia…

National Strategic Computing Initiative (NSCI)

White House Executive Order

1. Accelerating delivery of a capable exascale computing system that integrates hardware and software capability to deliver approximately 100 times the performance of current 10 petaflop systems across a range of applications representing government needs.
2. *Increasing coherence between the technology base used for modeling and simulation and that used for data analytic computing.*
3. Establishing, over the next 15 years, a viable path forward for future HPC systems even after the limits of current semiconductor technology are reached (the "post- Moore's Law era").
4. Increasing the capacity and capability of an enduring national HPC ecosystem by employing a holistic approach that addresses relevant factors such as networking technology, workflow, downward scaling, foundational algorithms and software, accessibility, and workforce development.
5. Developing an enduring public-private collaboration to ensure that the benefits of the research and development advances are, to the greatest extent, shared between the United States Government and industrial and academic sectors.

http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/BDEC2_Convergence_Reed_Beckman.pdf

En un vistazo…

Desde la Casa Blanca también están impulsando la convergencia…

NSCI workshop (October 2015)

Opening panels

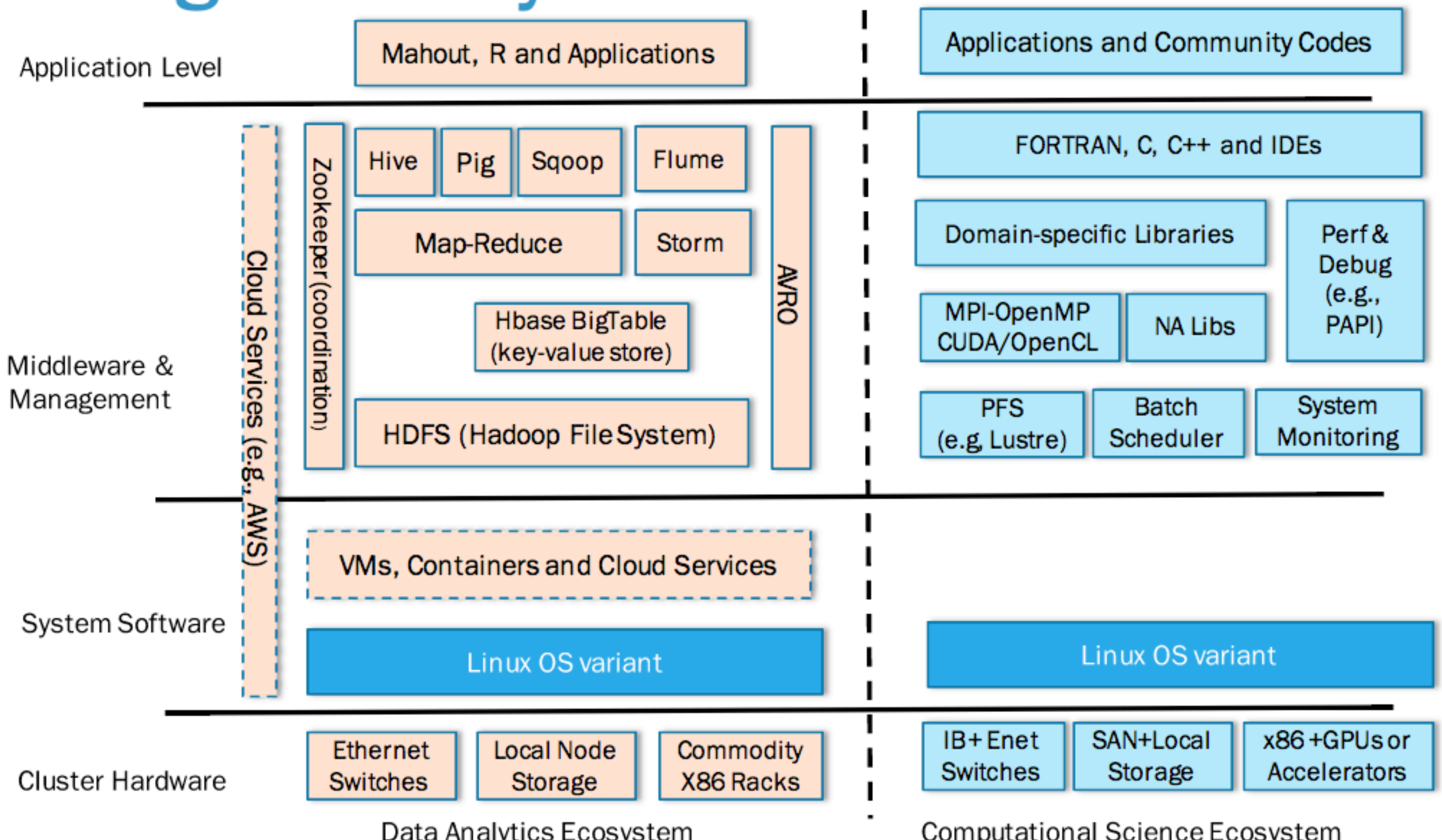
- Convergence of Data Analytics and Computationally-Intensive Computing

“*As NSCI drives forward these two goals of exaflop computing ability and exabyte storage capacity, it will also find ways to combine large-scale numerical computing with big data analytics.* This will enable new forms of computation, including simulations of weather that are coupled with actual observations from weather satellites. It will also enable new analytic methods that require more extensive numerical processing, such as emerging techniques that use artificial intelligence to automatically learn new capabilities from large numbers of examples.“

http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/BDEC2_Convergence_Reed_Beckman.pdf

http://www.exascale.org/bdec/sites/www.exascale.org.bdec/files/BDEC2_Convergence_Reed_Beckman.pdf

Divergent ecosystems

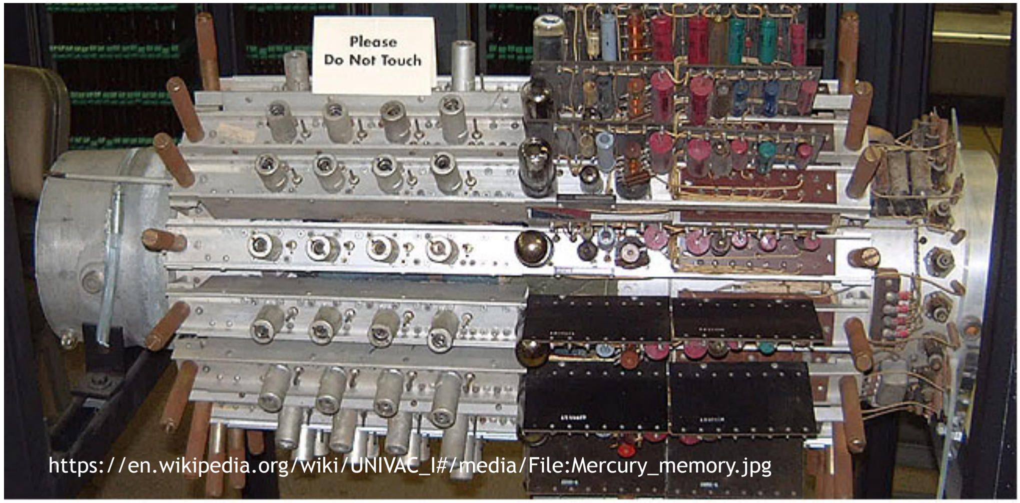


ELEGIMOS TODO_

Un poco de historia:
el hardware

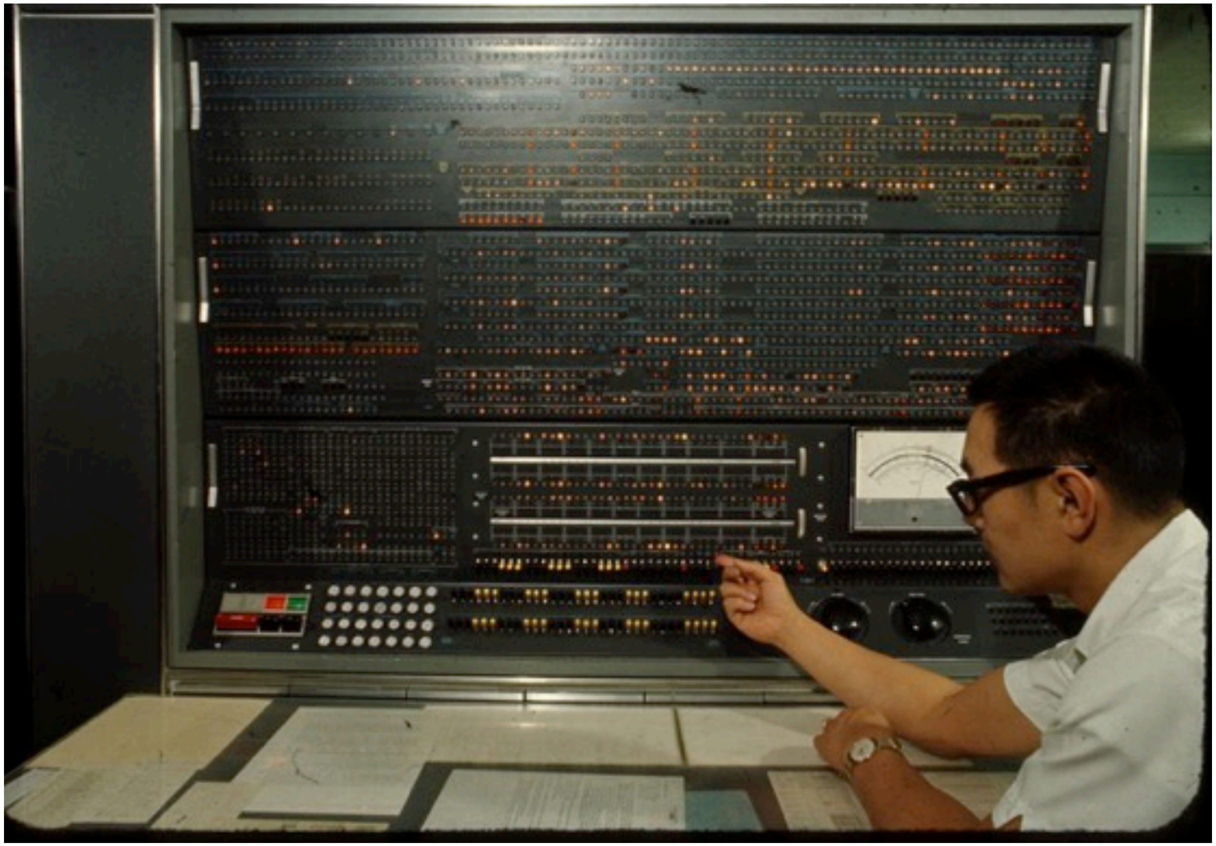
Los inicios…

- 1946: ENIAC
  - Armas termonucleares
  - Diseñado por Eckert & Mauchly
- 1951: MANIAC
  - Algoritmo de Monte Carlo y la bomba de hidrogeno
- 1956: UNIVAC I
  - 1er computador comercial producido en EEUU
  - También diseñado por Eckert y Mauchly



1961: IBM 7030 “Stretch”->**Predicción de saltos**

Problemas de performance limitaron su éxito, pero reaparecerá en los 80s



“Programmers such as Tad Kishi shown here, found this console to be very helpful in debugging their programs.”
http://www.computer-history.info/Page4.dir/pages/IBM.7030.Stretch.dir/

1964: CDC-6600 uno de los primeros supercomputadores comercializados -> *Superscalar processor*

https://en.wikipedia.org/wiki/CDC_6600

NEW REMOTE CALCULATOR brings high-powered computing to the individual scientist's desk

The Control Data Remote Calculator is an amazing new tool developed for you. It enables many scientists, mathematicians and statisticians to simultaneously use the CONTROL DATA® 6600—the world's most powerful computer—to get quick answers to problems stated in mathematical language. At the same time it frees the user from frustrating conventions, programming intermediaries and tedious waiting. This combination of power and convenience broadens your horizons—encourages experimentation, innovation and creativity.

Up to 2000 calculators can be installed remotely, via standard telephone channels, through the common-user dial network. Calculators can operate concurrently with normal processing and other types of remote terminals. This flexibility along with compact portability enables the calculator to be used for "homework." Whether at home or office, the user merely queries the computer from the calculator keyboard containing all conventional mathematical functions and symbols. Answers are shot back on the calculator's display panel.

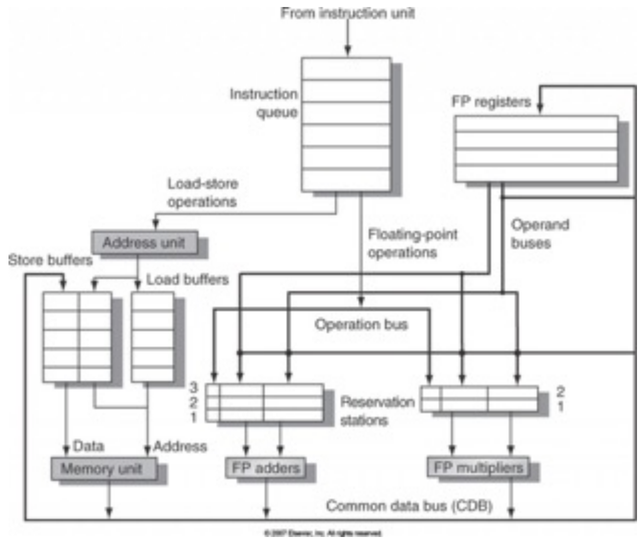
Powerful computer systems, and the means for people to get at them the moment they need to, account for today's preference for Control Data computer systems in scientific research and development. Talk over your requirements with your Control Data representative. Or write our Minneapolis address. Dept. G-115.

CONTROL DATA
CORPORATION

8100 34th AVE. SO., MINNEAPOLIS, MINNESOTA 55440

http://www.oldcalculatormuseum.com/a-cdc-11-65.html

1966: IBM's 360/91
-> *Ejecución fuera de orden*
Algoritmo de Tomasulo



https://en.wikipedia.org/wiki/Tomasulo_algorithm

https://commons.wikimedia.org/wiki/File:IBM_System_360_91.mh.jpg

https://en.wikipedia.org/wiki/IBM_System/360

1976: Cray-1
-> *Vectorización / SIMD* (8 registros de 64-bit)

http://www.cray.com/sites/default/files/resources/CrayTimeline.pdf

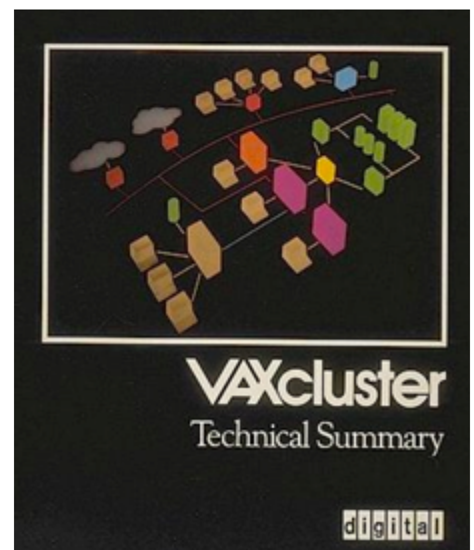
http://www.hermanotemblon.com/wp-images/cray.jpg

1984:

-La serie Rockwell Intl R65C** introdujo el ***dual core processor*** (MOS-based, core of Atari, NES, Commodore, … 8-bit revolution)

-DEC VAXcluster uno de los primeros ***clusters de computadores*** para procesamiento paralelo

https://upload.wikimedia.org/wikipedia/commons/1/14/MOS_6501_6502_Ad_Sept_1975.jpg



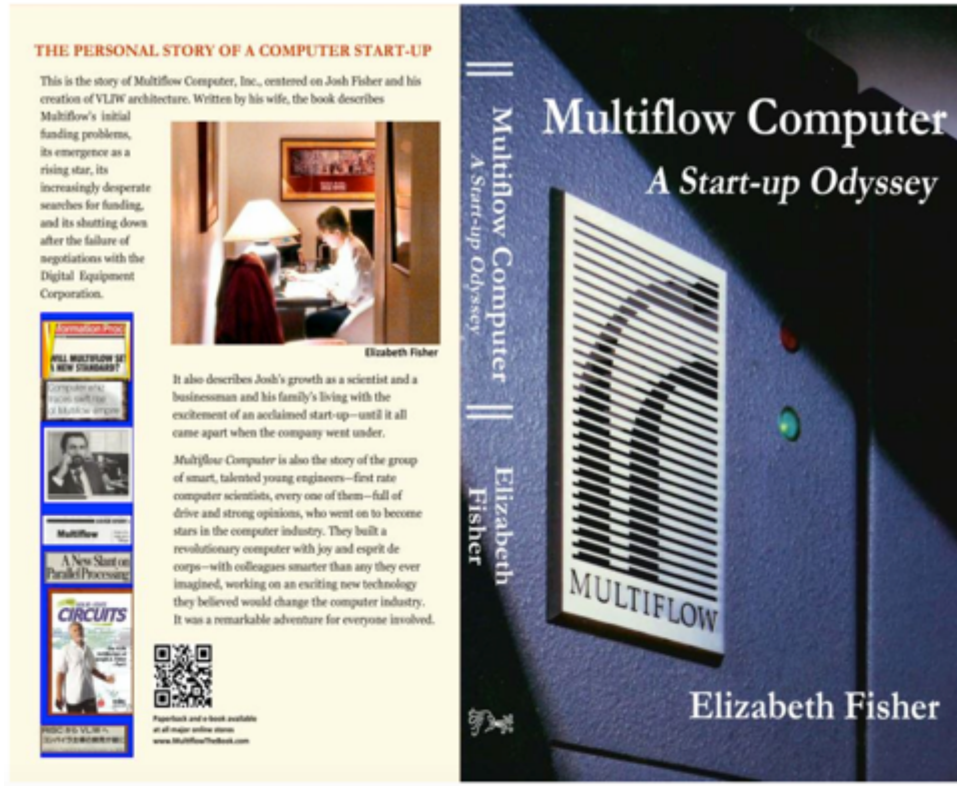
http://research.microsoft.com/en-us/um/people/gbell/digital/timeline/1983-1.htm

1984:
The Multiflow experience

VLIW → **1024-bit** arch

Muchas compañías adoptaron estas ideas decadas después

Ampliamente usado en SoC/Embedded archs



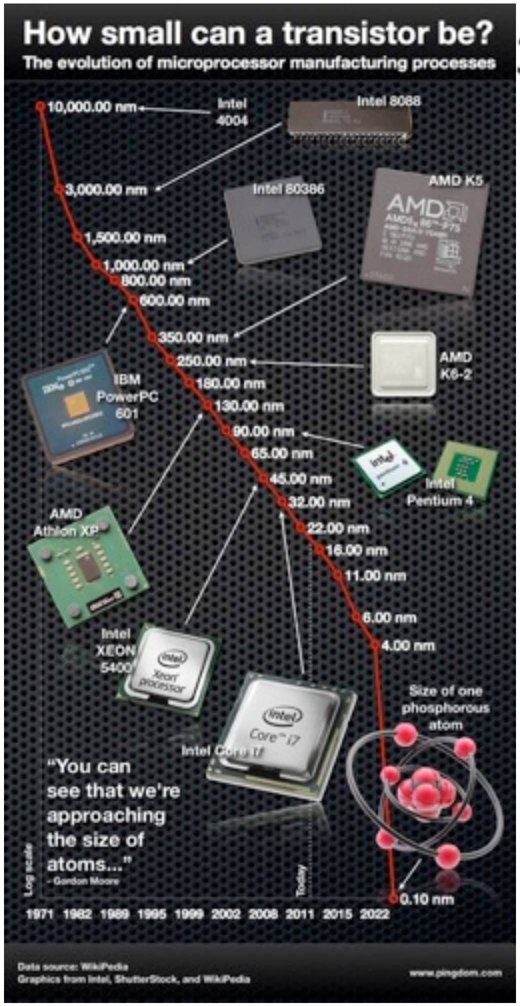
1984:
FPGAs: *Field-Programmable Gate Arrays*

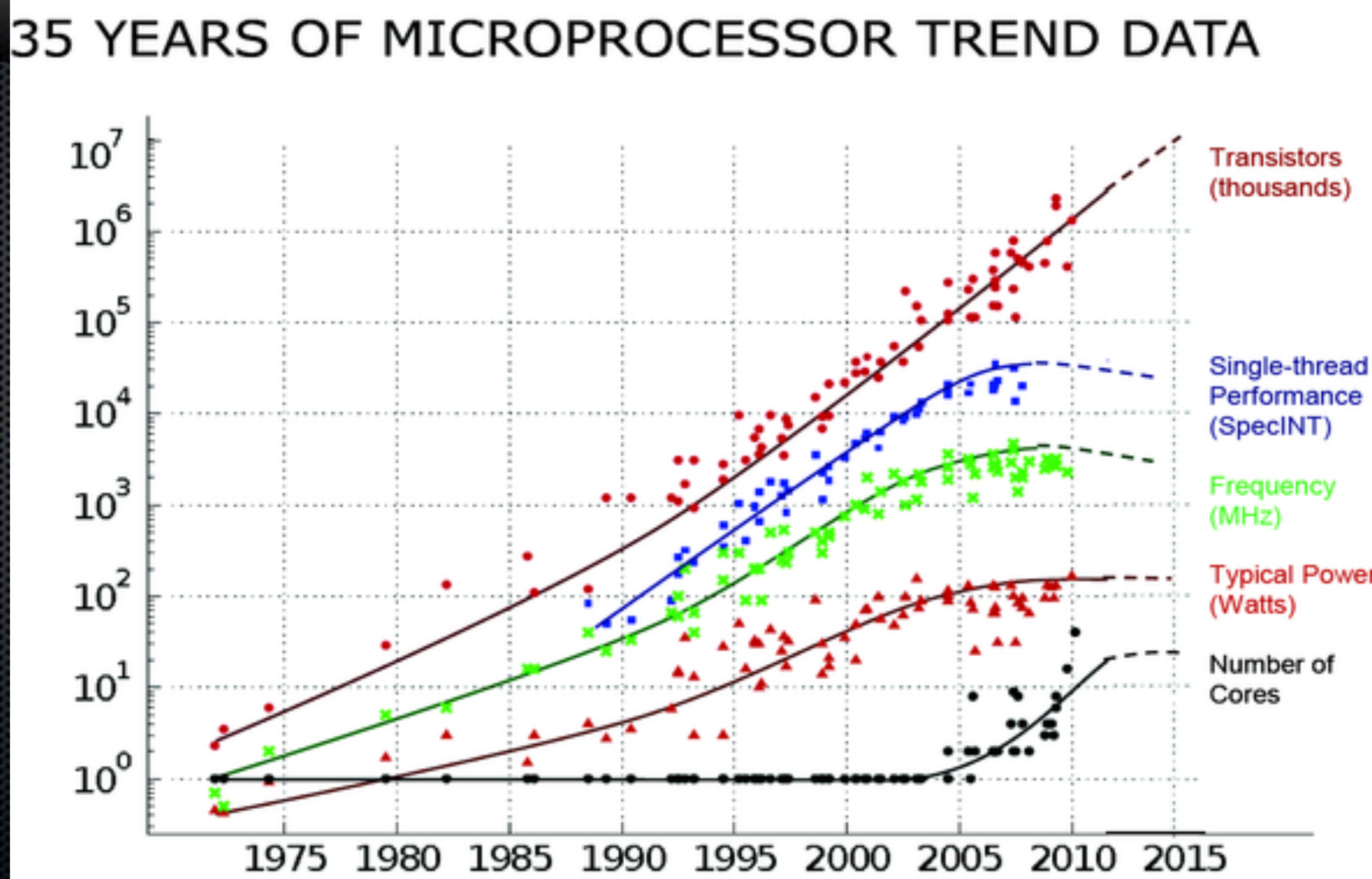
Inicios de la revolución *fabless*:
- foco en I+D
- sub-contratar la fabricación
- Ejemplo: TSMC

Empresas líderes: Altera y Xilinx

tsmc

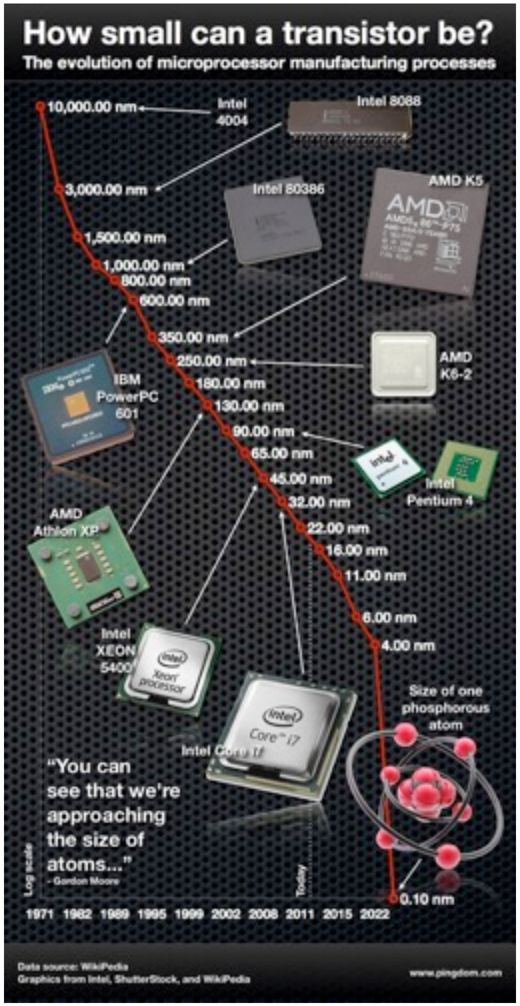
1985-hoy: ++Transistores

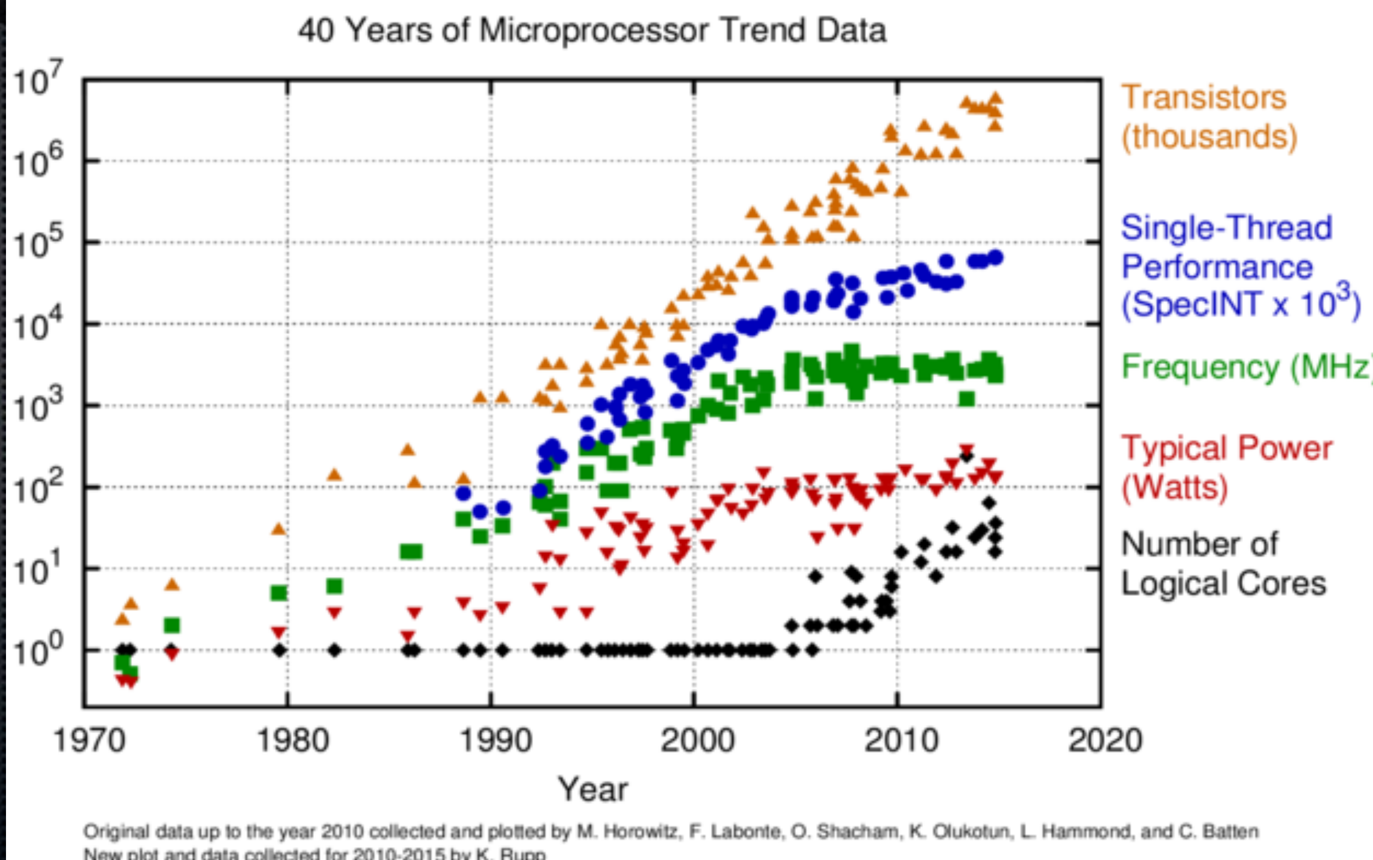




Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

1985-hoy: ++Transistores

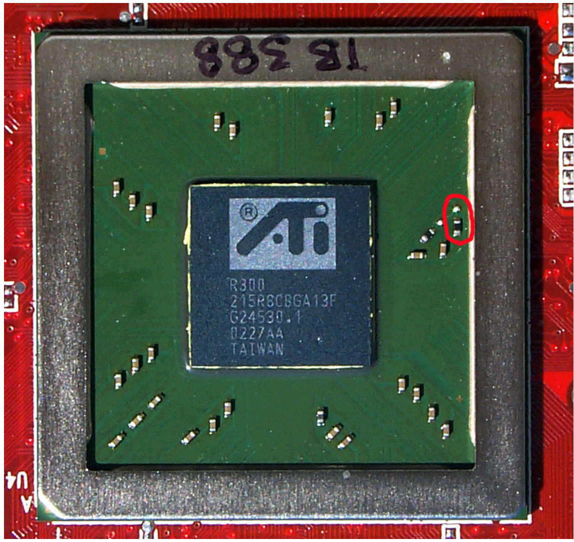




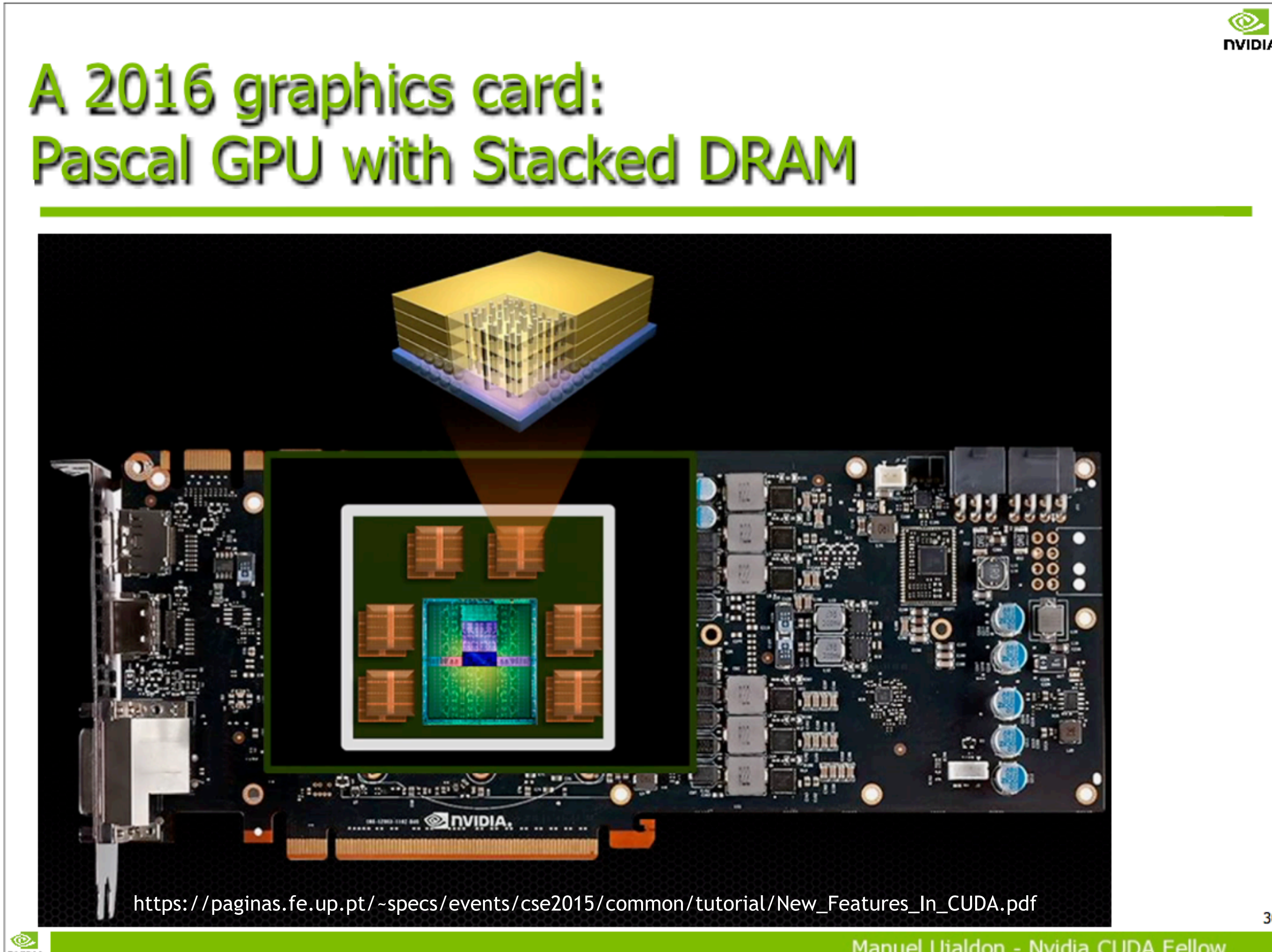
https://www.karlrupp.net/2015/06/40-years-of-microprocessor-trend-data/

2000-hoy: Programando en la GPU

- 1980~2000: *Fixed-function graphics pipelines*:
  - SGI OpenGL y Microsoft DirectX
- 2001: NVIDIA GeForce 3
  - shader extensions
  - seguida por la ATI Radeon 9700
- 2006: NVIDIA GeForce 8800
  - *Unified processor array*
- 2007: Lanzamiento de CUDA
- 2007 en adelante:
  - soporte Multi-GPU
  - nuevas microachs
  - soporte MPI-CUDA
  - memoria 3D…

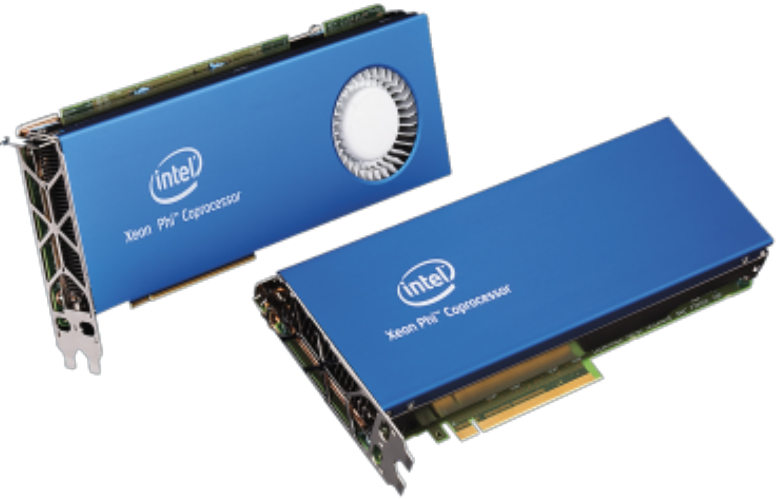


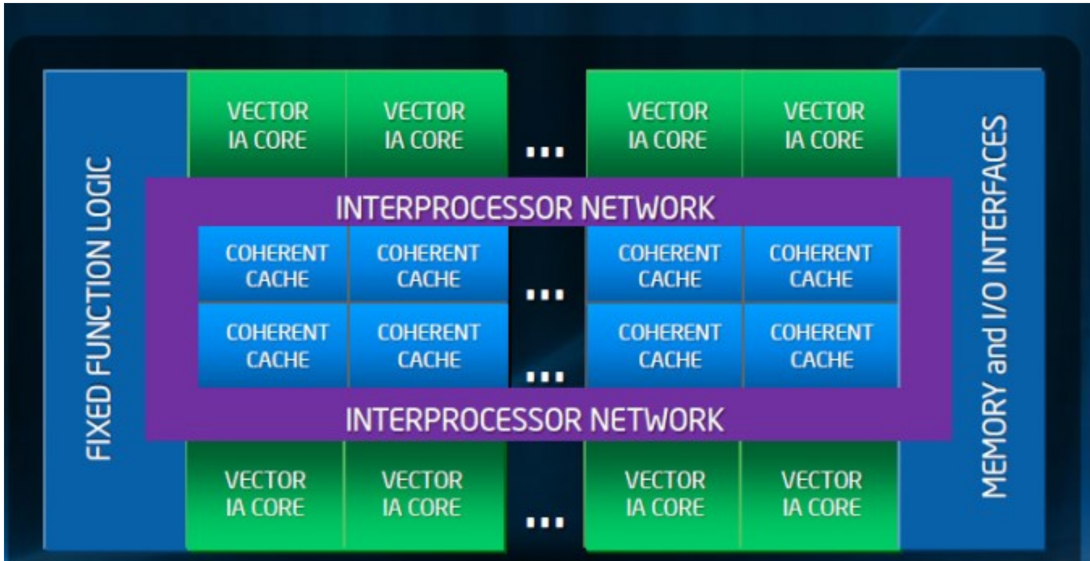
2000-hoy: Programando en la GPU



2006-hoy: Aterrizan los Co-Procesadores de Intel

- 2006: Larrabee microarquitectura (512-bit SIMD x86)
- 2010: Knights Ferry (prototipo Intel MIC)
- 2012: Knights Corner
  (*Intel Xeon Phi*, 1ª generación)
- 2013: Knights Landings (2ª generación)
- 2015: Knights Hill (3ª generación)





http://www.extremetech.com/extreme/133541-intels-64-core-champion-in-depth-on-xeon-phi

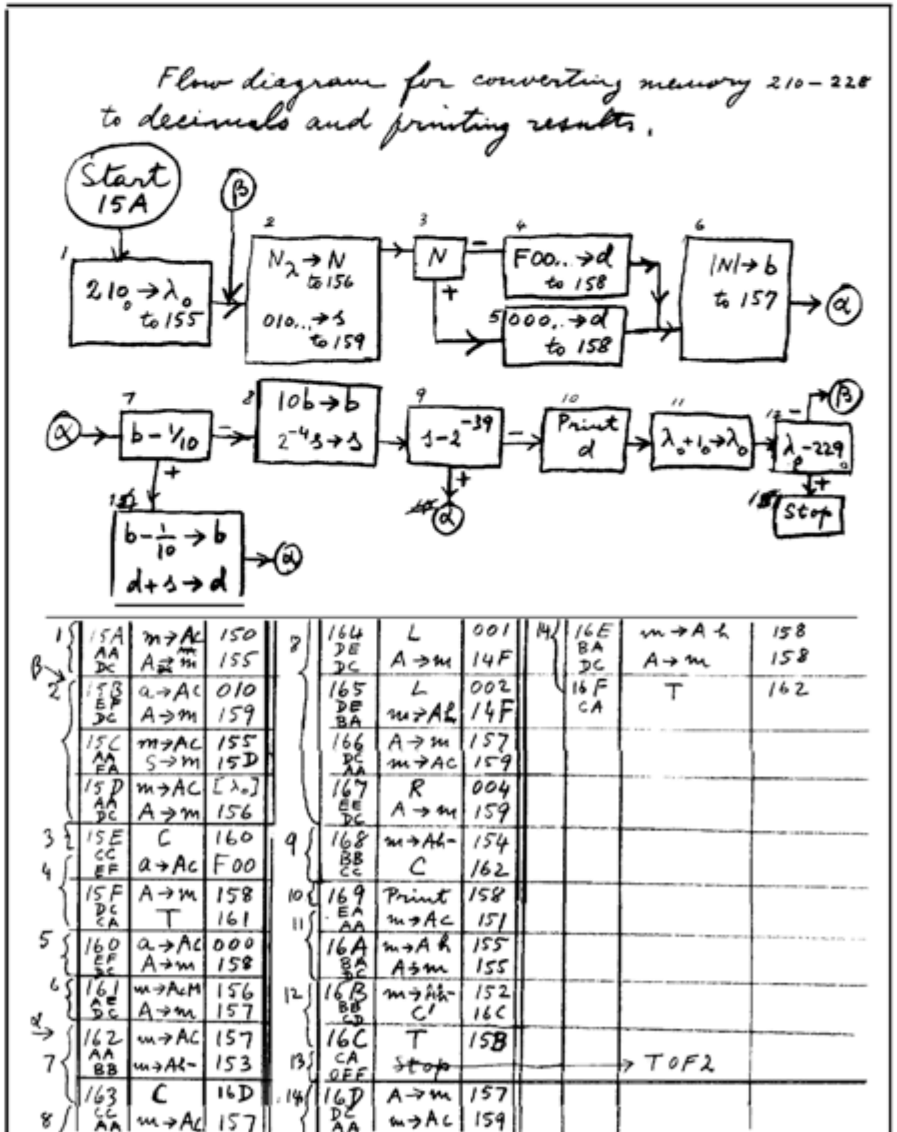
Supercomputadores en el mundo
http://top500.org/lists/2016/11/

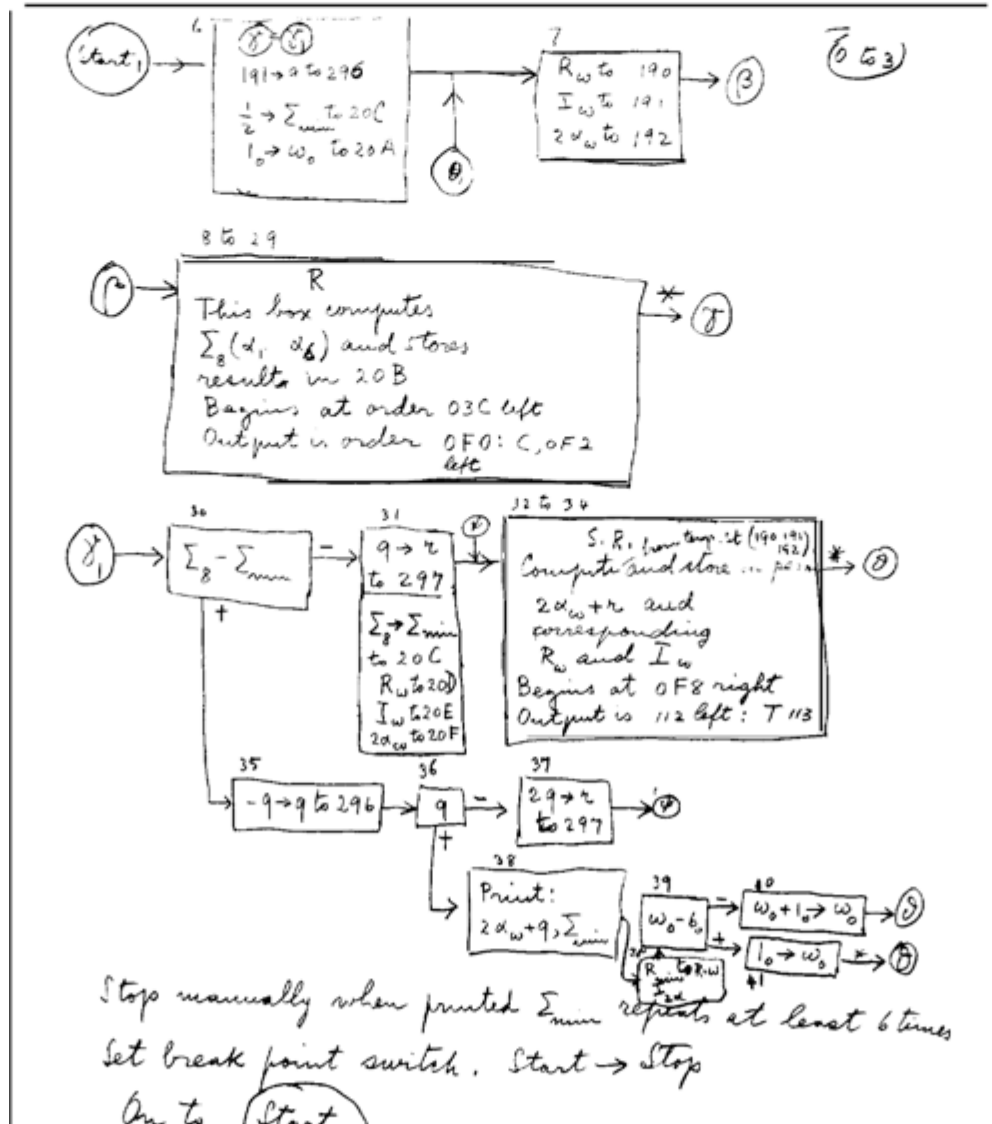
Tianhe-2 (MilkyWay-2)

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 ← Xeon Phi	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. ← NVIDIA K20x	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
6	DOE/NNSA/LANL/SNL United States	**Trinity** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.	301,056	8,100.9	11,078.9	
7	Swiss National Supercomputing Centre (CSCS) Switzerland	**Piz Daint** - Cray XC30, Xeon E5-2670 8C 2.600GHz, Aries interconnect , NVIDIA K20x Cray Inc. ← NVIDIA K20x	115,984	6,271.0	7,788.9	2,325
8	HLRS - Höchstleistungsrechenzentrum Stuttgart Germany	**Hazel Hen** - Cray XC40, Xeon E5-2680v3 12C 2.5GHz, Aries interconnect Cray Inc.	185,088	5,640.2	7,403.5	
9	King Abdullah University of Science and Technology Saudi Arabia	**Shaheen II** - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Aries interconnect Cray Inc.	196,608	5,537.0	7,235.2	2,834
10	Texas Advanced Computing Center/Univ. of Texas United States	**Stampede** - PowerEdge C8220, Xeon E5-2680 8C 2.700GHz, Infiniband FDR, Intel Xeon Phi SE10P Dell ← Xeon Phi	462,462	5,168.1	8,520.1	4,510

Un poco de historia:
el software

Los “buenos” tiempos antes de Fortran… (<1957)





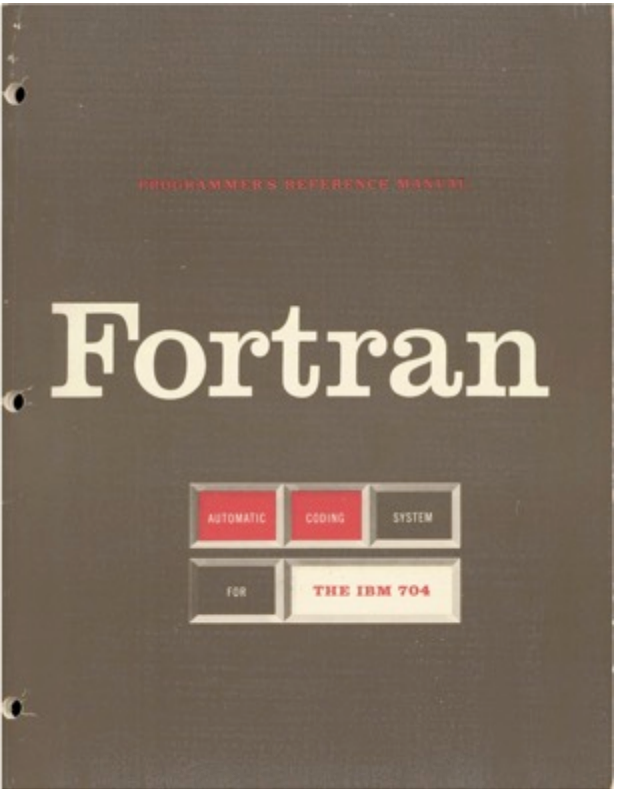
Monte Carlo method by Enrico Fermi circa 1954, http://library.lanl.gov/cgi-bin/getfile?00326886.pdf

Fig. 3. A subprogram written by Fermi for converting data in memory from hexadecimal to decimal form and printing the results.

Fig. 4. A subprogram written by Fermi for calculating phase shifts by finding a minimum chi-squared in a fit to the data.

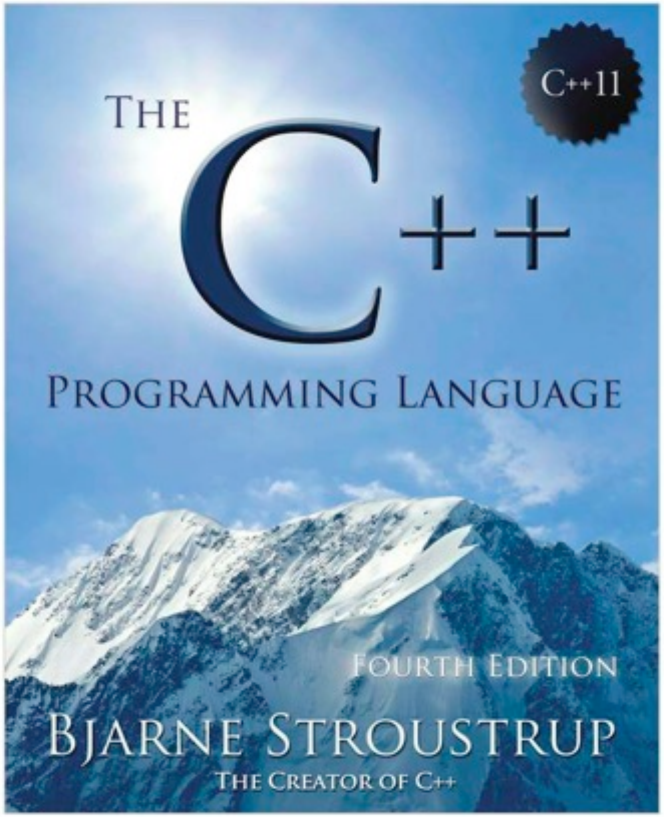
Y la era de la computación científica comienza…

1957



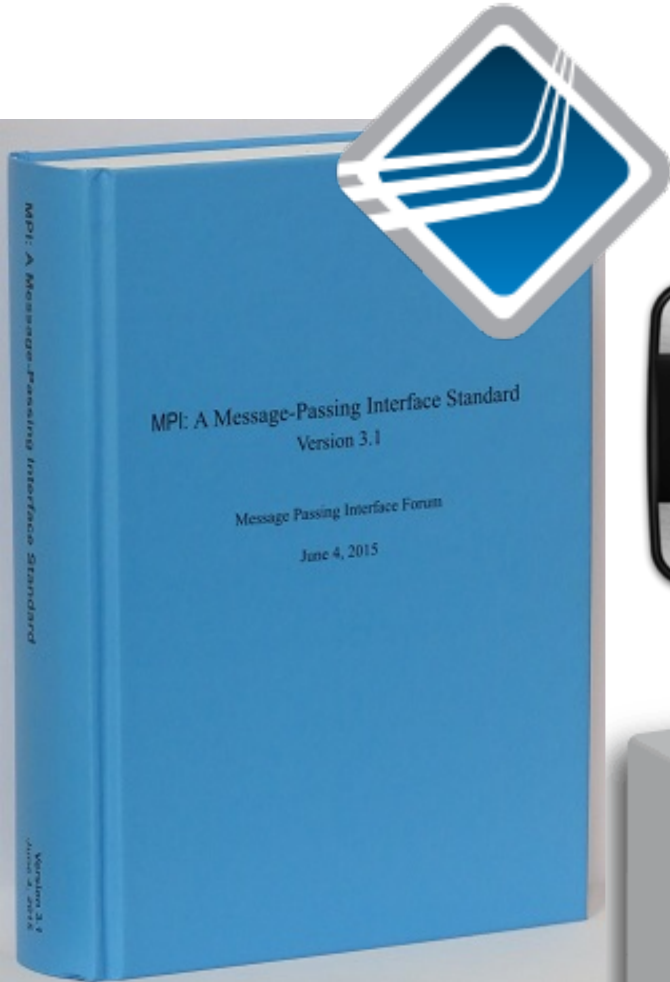
1972

1983



ELEGIMOS TODO_

El ecosistema crece...



Fortran, una historia de amor y odio

¿Usamos Fortran aún?

```
Oscars-MacBook-Pro:RAAE006W50009194 operedo$ ssh h[redacted].azurehdinsight.n
Authorized uses only. All activity may be monitored and reported.
hdiuser@perfsf-ssh.azurehdinsight.net's password:
Welcome to Ubuntu 14.04.4 LTS (GNU/Linux 3.19.0-58-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

  Get cloud support with Ubuntu Advantage Cloud Guest:
    http://www.ubuntu.com/business/services/cloud

15 packages can be updated.
6 updates are security updates.


Last login: Wed Apr 13 03:30:28 2016 from [redacted].movistar.cl
hdiuser@hn0-perfsf:~$ sudo apt-get install python-numpy
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following packages were automatically installed and are no longer required:
  linux-headers-3.16.0-57 linux-headers-3.16.0-57-generic
  linux-headers-3.16.0-59 linux-headers-3.16.0-59-generic
  linux-image-3.16.0-57-generic linux-image-3.16.0-59-generic
  linux-image-3.19.0-43-generic
Use 'apt-get autoremove' to remove them.
The following extra packages will be installed:
  libblas3 libgfortran3 liblapack3
Suggested packages:
  gfortran python-nose python-numpy-dbg python-numpy-doc
The following NEW packages will be installed:
  libblas3 libgfortran3 liblapack3 python-numpy
0 upgraded, 4 newly installed, 0 to remove and 18 not upgraded.
Need to get 3,797 kB of archives.
After this operation, 16.5 MB of additional disk space will be used.
Do you want to continue? [Y/n] n
```

Instalando numpy en una máquina Ubuntu 14.04

Dependencias: Librerías y Compilador para Fortran 77/90

¿Porqué Fortran es más rápido que C para operaciones numéricas?

Respuesta: ALIASING

https://people.freebsd.org/~lstewart/articles/cpumemory.pdf

"What every programmer should know about memory", Ulrich Drepper, Red Hat Inc. 2007

```
int main(){
        int y, *x;
        y=1;
        x=&y;
        printf("%d\n »,*x);
        y=2;
        printf("%d\n »,*x);
}
```

The rest of the complication here results from the fact that gcc is not very smart when it comes to optimizing array indexing. The introduction of the additional variables `rres`, `rmul1`, and `rmul2` optimizes the code by pulling common expressions out of the inner loops, as far down as possible. The default aliasing rules of the C and C++ languages do not help the compiler making these decisions (unless `restrict` is used, all pointer accesses are potential sources of aliasing). This is why Fortran is still a preferred language for numeric programming: it makes writing fast code easier.[29]

[29]In theory the `restrict` keyword introduced into the C language in the 1999 revision should solve the problem. Compilers have not caught up yet, though. The reason is mainly that too much incorrect code exists which would mislead the compiler and cause it to generate incorrect object code.

En 1972, E.W.Dijkstra escribió una carta (EDW340) con sus puntos de vista sobre la escena del software: (https://www.cs.utexas.edu/~EWD/ewd03xx/EWD340.PDF)

- La invención de la subrutina
- **FORTRAN**
- LISP
- ALGOL 60
- PL/1

The second major development on the software scene that I would like to mention is the birth of FORTRAN. At that time this was a project of great temerity and the people responsible for it deserve our great admiration. It would be absolutely unfair to blame them for shortcomings that only became apparent after a decade or so of extensive usage: groups with a successful look-ahead of ten years are quite rare! In retrospect we must rate FORTRAN as a successful coding technique, but with very few effective aids to conception, aids which are now so urgently needed that time has come to consider it out of date. The sooner we can forget that FORTRAN has ever existed, the better, for as a vehicle of thought it is no longer adequate: it wastes our brainpower, is too risky and therefore too expensive to use. FORTRAN's tragic fate has been its wide acceptance, mentally chaining thousands and thousands of programmers to our past mistakes. I pray daily that more of my fellow-programmers may find the means of freeing themselves from the curse of compatibility.

En 1972, E.W.Dijkstra escribió una carta (EDW340) con sus puntos de vista sobre la escena del software: (https://www.cs.utexas.edu/~EWD/ewd03xx/EWD340.PDF)

- La invención de la subrutina
- **FORTRAN**
- LISP
- ALGOL 60
- PL/1

The second major development on the software scene that I would like to mention is the birth of FORTRAN. At that time this was a project of great temerity and the people responsible for it deserve our great admiration. It would be absolutely unfair to blame them for shortcomings that only became apparent after a decade or so of extensive usage: groups with a successful look-ahead of ten years are quite rare! In retrospect we must rate FORTRAN as

date. The sooner we can forget that FORTRAN has ever existed, the better, for as a vehicle of thought it is no longer adequate: it wastes our brainpower, is too risky and therefore too expensive to use. FORTRAN's tragic fate has been its wide acceptance, mentally chaining thousands and thousands of programmers to our past mistakes. I pray daily that more of my fellow-programmers may find the means of freeing themselves from the curse of compatibility.

“The seven ages of FORTRAN”, Metcalf 2011

http://journal.info.unlp.edu.ar/journal/journal30/papers/JCST-Apr11-1.pdf

Abstract

When IBM's John Backus first developed the Fortran programming language, back in 1957, he certainly never dreamt that it would become a world-wide success and still be going strong many years later. Given the oft-repeated predictions of its imminent demise, starting around 1968, it is a surprise, even to some of its most devoted users, that this much-maligned language is not only still with us, but is being further developed for the demanding applications of the future. What has made this programming language succeed where most slip into oblivion?

...

...

languages remains an open question. However, there is every sign that Fortran continues to be used to tackle major scientific computing problems, and will long remain a living memorial to the early pioneers. Indeed, at a Workshop on Software in High-Energy Physics in 1982, I predicted that: *“Fortran is likely to remain into the next century as, at the very least, a special-purpose scientific and numerical language for large-scale, computing-intensive applications and, strengthened especially by its array capabilities, will be one of a small range of widely-used languages in general use”*. This turned out to be not too far from the truth!

...

“I don't know what the language of the year 2000 will look like, but I know it will be called Fortran.” —Tony Hoare, winner of the 1980 Turing Award, in 1982.

“The seven ages of FORTRAN”, Metcalf 2011

http://j … -1.pdf

years later. Given the oft-repeated predictions of its imminent demise, starting around 1968, it is a surprise, even to some of its most devoted users, that this much-maligned language is not only still with us, but is being further developed for the demanding applications of the future. What has made this

Physics in 1982, I predicted that: “*Fortran is likely to remain into the next century as, at the very least, a special-purpose scientific and numerical language for large-scale, computing-intensive applications and, strengthened especially by its array capabilities, will be one of a small range of widely-used languages in general use*”. This turned out to

When IBM's Fortran progr… certainly nev… world-wide su…

years later. Gi… imminent der… surprise, even… this much-mal… us, but is bein… applications c…

programming … oblivion?

…

…lowever, there … to be used to … lems, and will … early pioneers.

Indeed, at a workshop on Software in High-Energy

…rtran is likely … the very least, … ical language … applications … its array … nge of widely- … turned out to

“I don't know what the language of the year 2000 will look like, but I know it will be called Fortran.” —Tony Hoare, winner of the 1980 Turing Award, in 1982.

¿HPC y Deep Learning?

Componentes paralelizables

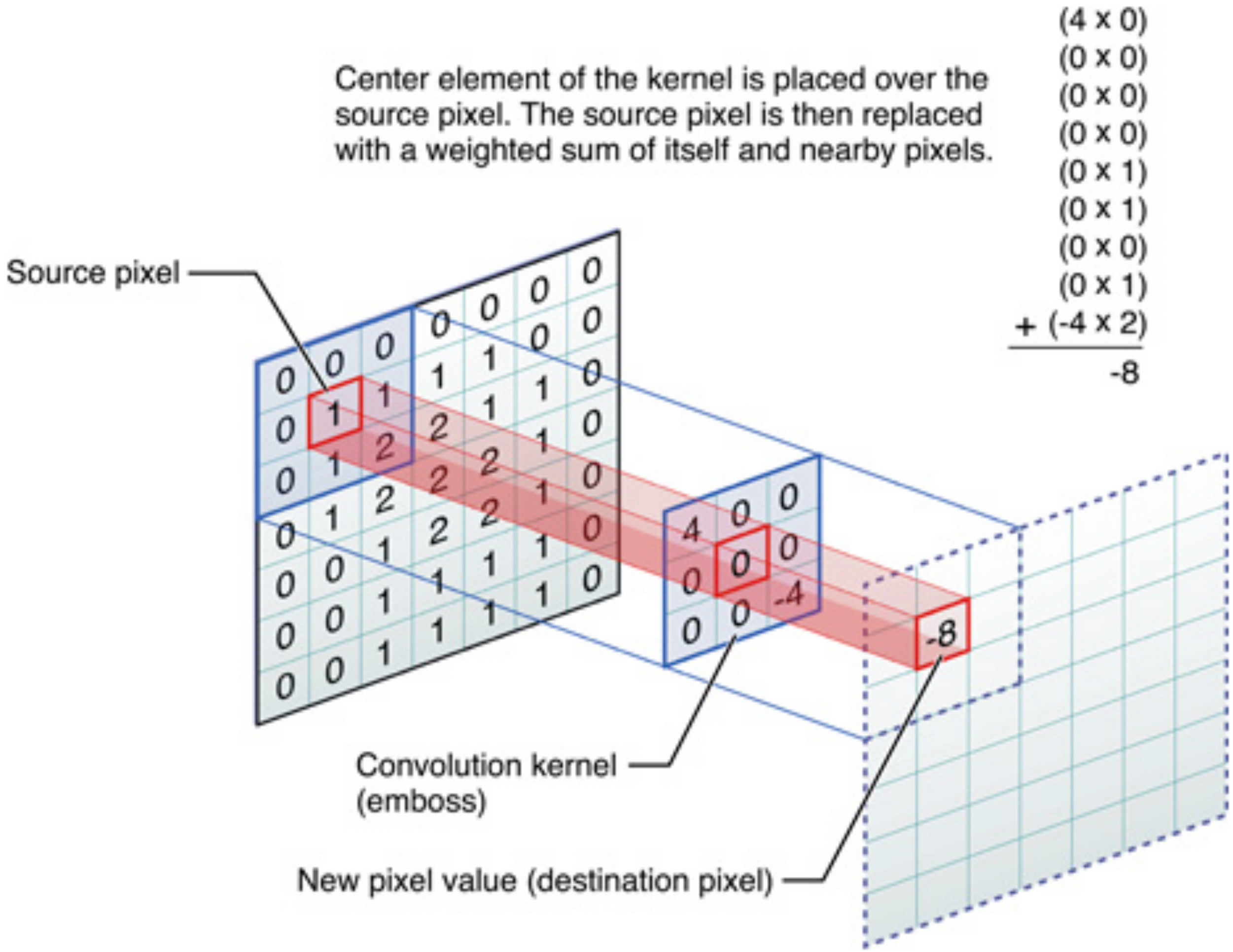
- Operaciones algebraicas (dense linear algebra)
  - `dgemm, sgemm` : multiplicación matríz-matríz
  - BLAS, cuBLAS, GotoBLAS, ATLAS
- Convolución 2D/3D
  - cuDNN, Theano, cuda-convnet
- Entrenamiento con mini-batches
  - Distribución de batches en multi-devices o multi-nodos

Componentes paralelizables

- Operaciones algebraicas (dense linear algebra)
  - `dgemm, sgemm` : multiplicación matríz-matríz
  - BLAS, cuBLAS, GotoBLAS, ATLAS
- Convolución 2D/3D
  - cuDNN, Theano, cuda-convnet
- Entrenamiento con mini-batches
  - Distribución de batches en multi-devices o multi-nodos

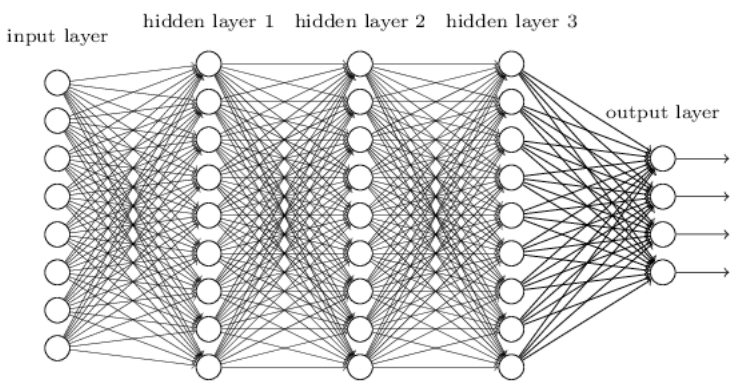
Convolución 2D/3D

Figure 2-4 Kernel convolution

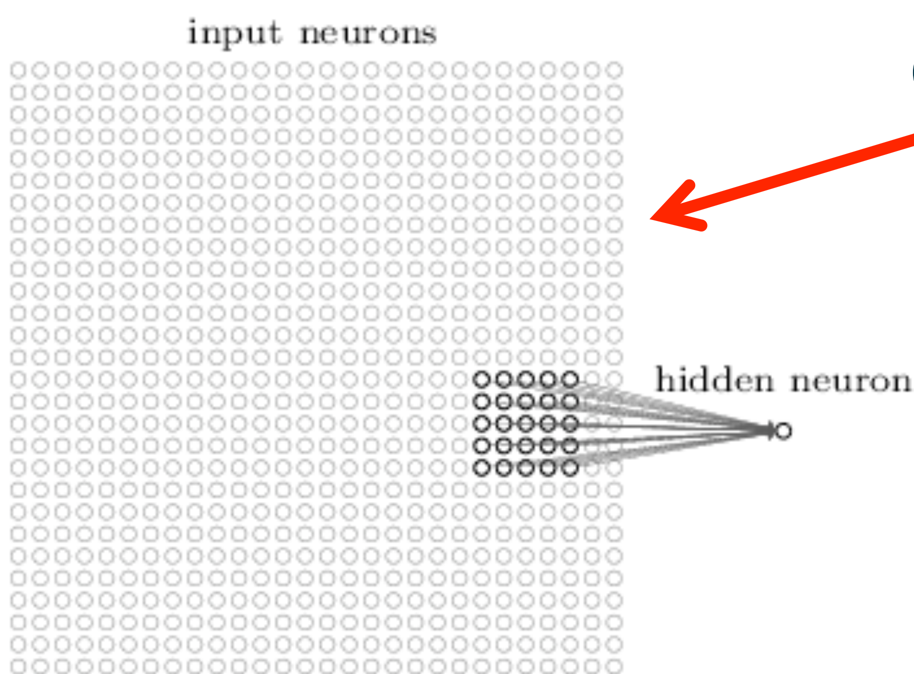


https://developer.apple.com/library/ios/documentation/Performance/Conceptual/vImage/ConvolutionOperations/ConvolutionOperations.html

Convolución en DL

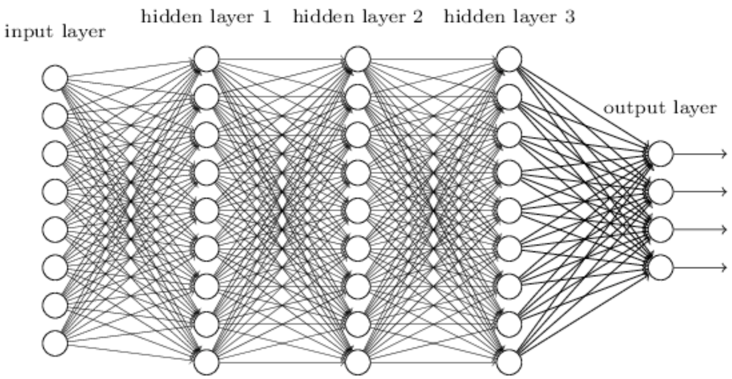


Redes Neuronales "clásicas"

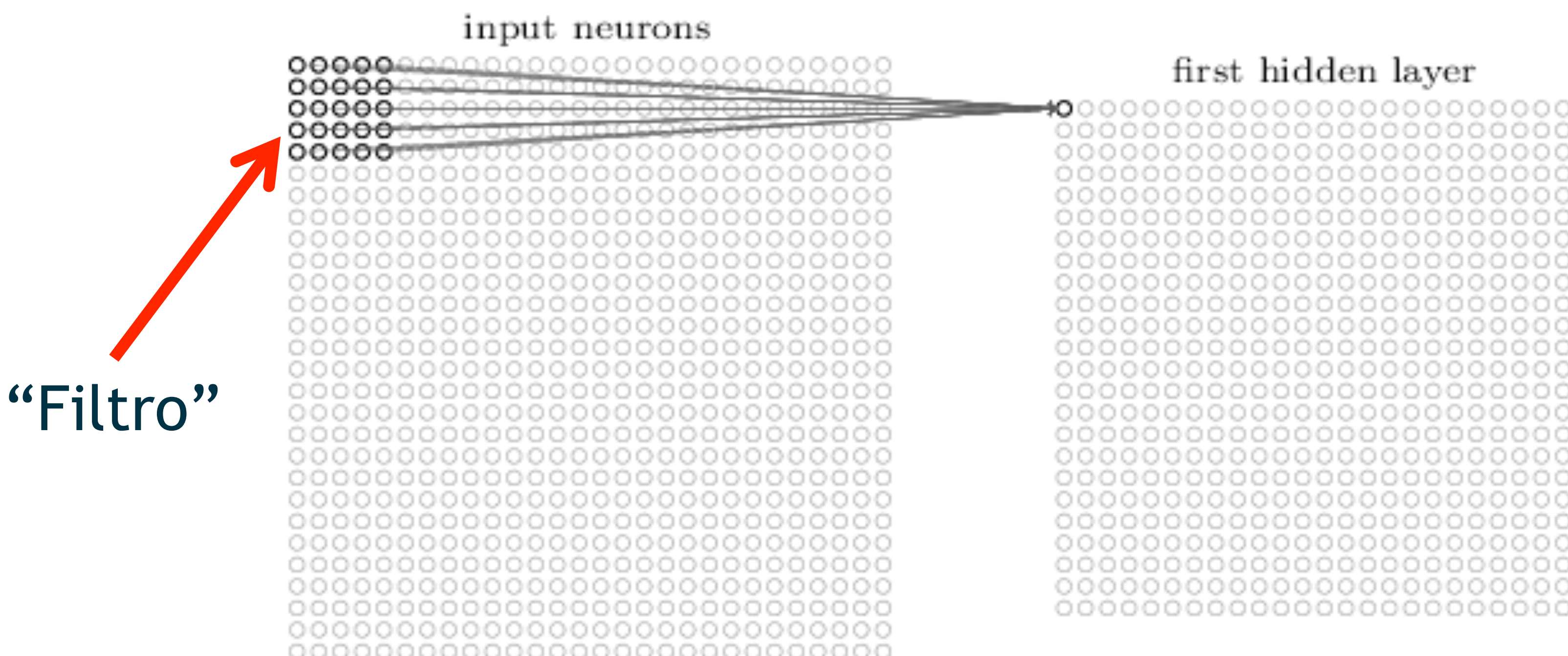


Redes Neuronales Convolucionales

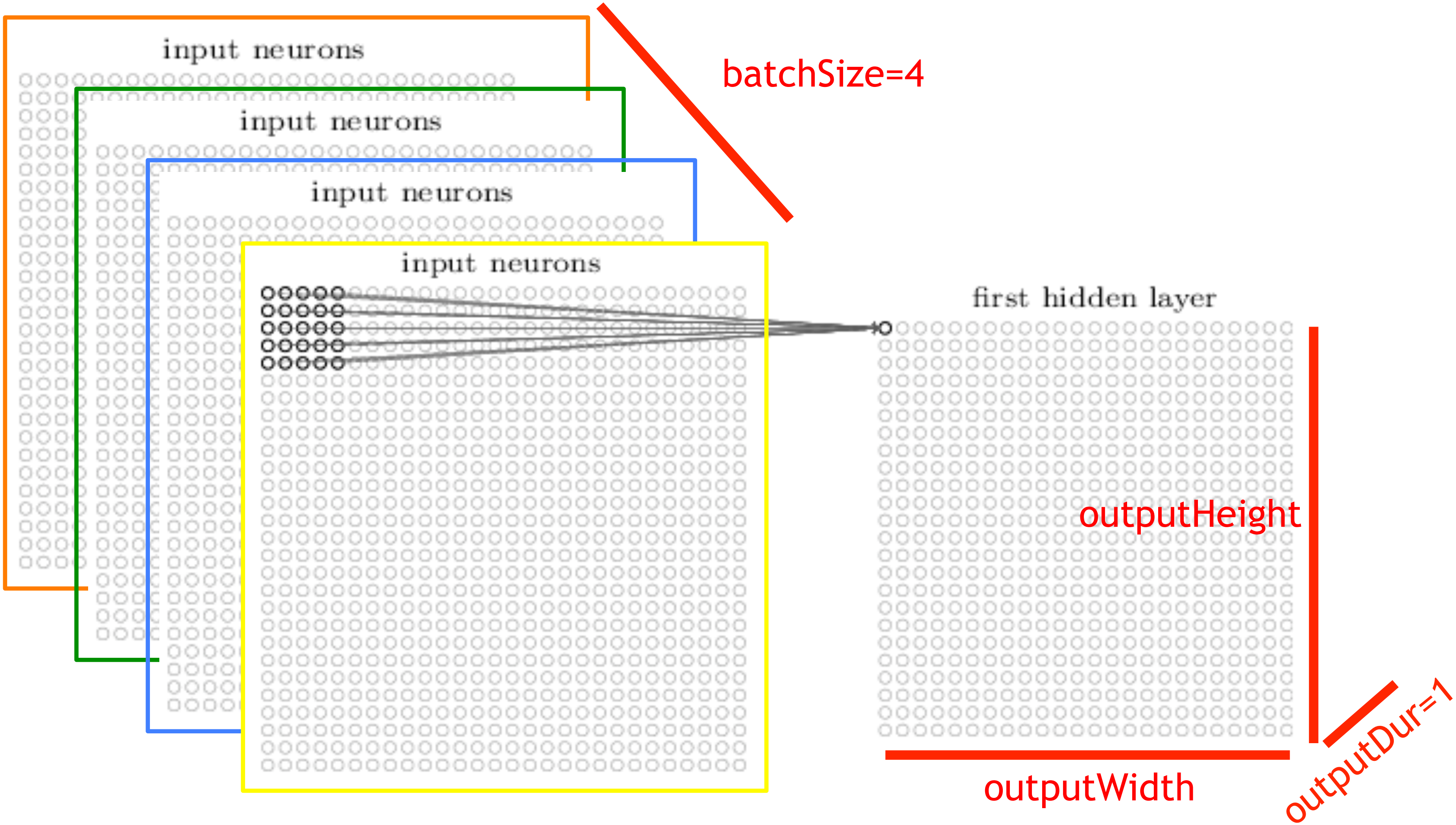
Convolución en DL



Se utiliza una convolución entre pixeles de una imagen y pesos de un "filtro"



Convolución en DL



stride=1 (desplazar filtro pixel-a-pixel)
outputChannels=3 (colores RGB)

Convolución DL + GPU (Theano CUDA)

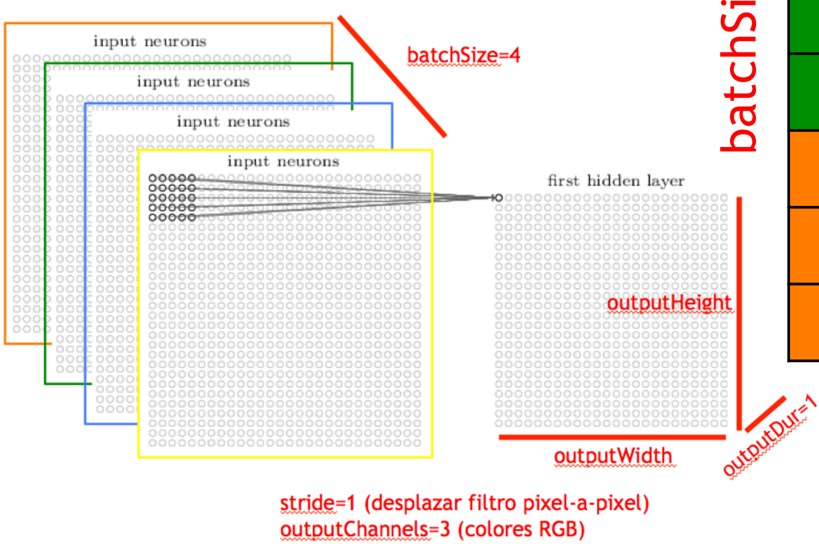
```
if(out_contiguous && !b_strided && (version==0||version==-1) && outputDur<=512 && !work_complete){
    //conv_rows_stack
    dim3 grid(outputHeight*outputWidth,batchSize*outputChannels);
    dim3 threads(outputDur);

    int shared_size=0;
        conv_rows_stack<<<grid, threads, shared_size>>>(
                CudaNdarray_DEV_DATA(%(V)s), CudaNdarray_DEV_DATA(%(W)s), CudaNdarray_DEV_DATA(%(b)s
                vidHeight, vidWidth, vidDur,
                filterHeight, filterWidth, filterDur,
                outputChannels, inputChannels,
                dr,dc,dt,
                vs3,vs2,vs1,vs4,vs0,
                ws3,ws2,ws1,ws4,ws0);

        CNDA_THREAD_SYNC;
        cudaError_t sts = cudaGetLastError();
        if (cudaSuccess == sts)
        {
            work_complete = true;
            if (verbose>1) printf("threads.x=%%i, threads.y=%%i, grid.x=%%i, grid.y=%%i, shared_size=
            if (verbose) printf("INFO: used 'conv_rows_stack' version\\n");
        }
        else
        {
            if (verbose) printf("threads.x=%%i, threads.y=%%i, grid.x=%%i, grid.y=%%i, shared_size=%%
            if (verbose) printf("ERROR: all implementations failed for GpuConv3D! (%%s)",cudaGetErro
            PyErr_Format(PyExc_RuntimeError, "ERROR: all implementations failed for GpuConv3D! (%%s)
                    cudaGetErrorString(sts));
            %(fail)s
        }
```

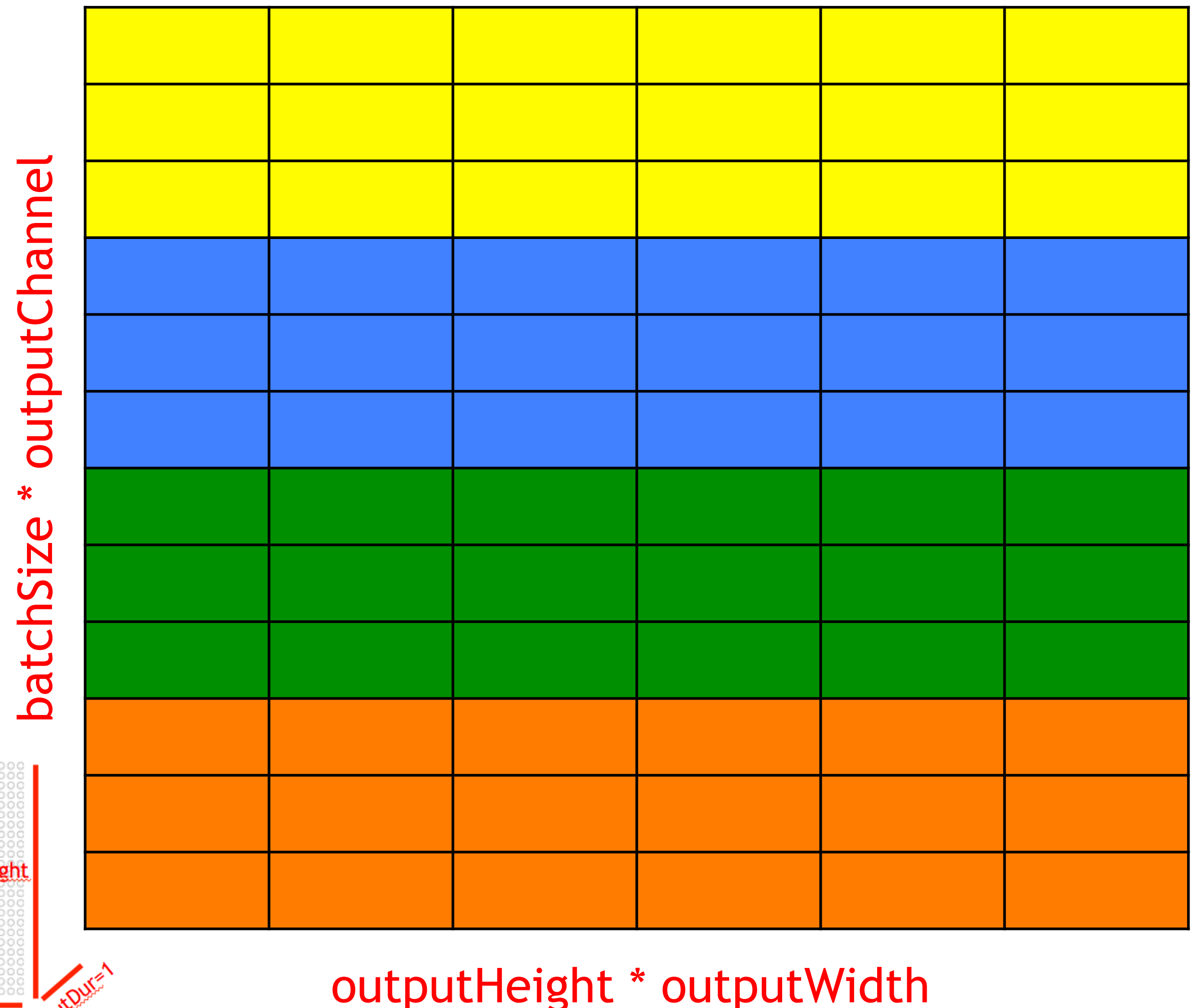
Theano/theano/sandbox/cuda/GpuConv3D.py

Convolución DL + GPU (Theano CUDA)

grid = dim3(outputHeight*outputWidth,batchSize*outputChannel)

Grilla de bloques 2D

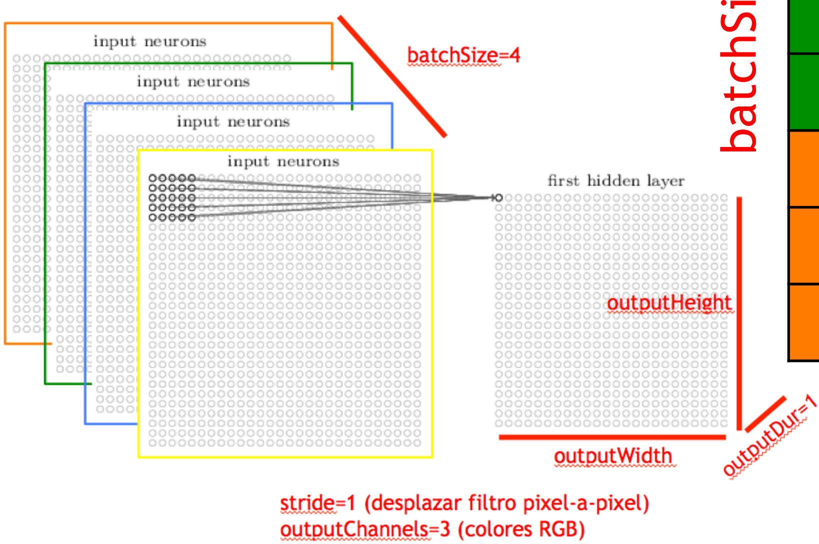


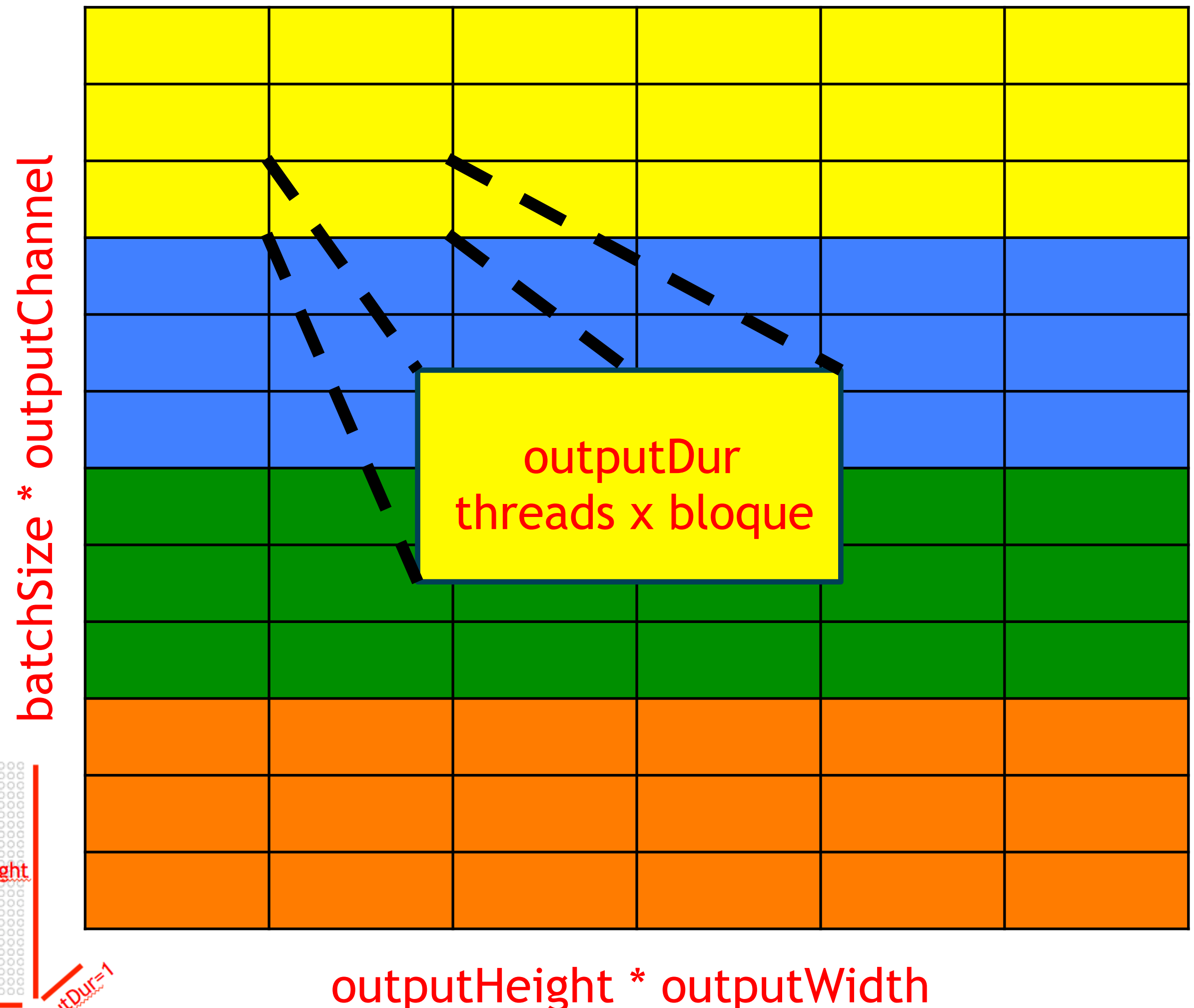


Convolución DL + GPU (Theano CUDA)

threads = dim3(outputDur)

Grilla de threads 1D



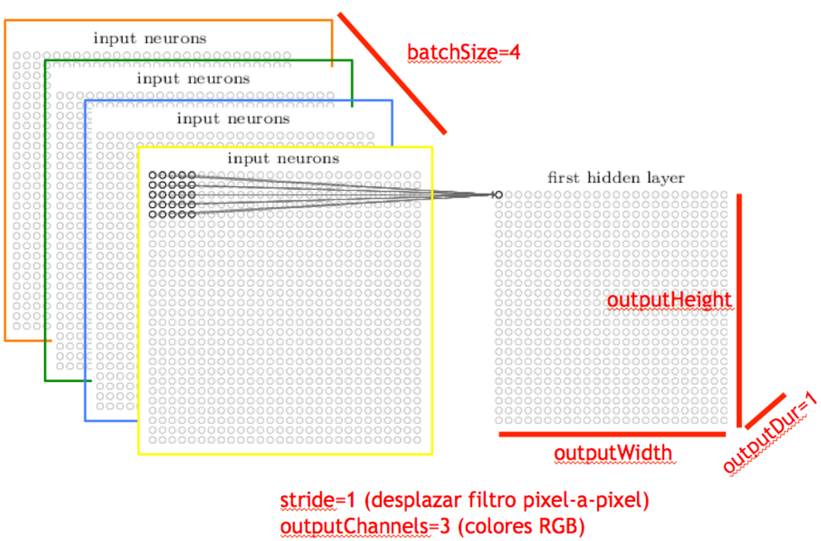


ELEGIMOS TODO_

Convolución DL + GPU (Theano CUDA)

Imagen 32 x 32 x1 pixeles
Filtro 4 x 4 x 1 pixeles
1 canal entrada
10 canales salida
128 imágenes por batch

Bloques: 1,076,480
Threads x bloque: 1



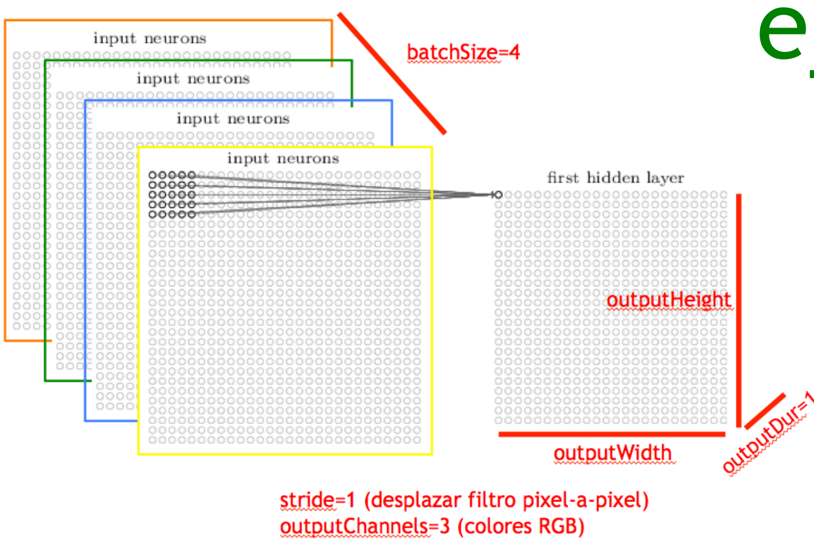
NVIDIA Tesla M2090
512 CUDA-cores
16 SM

Convolución DL + GPU (Theano CUDA)

Imagen 32 x 32 x1 pixeles
Filtro 4 x 4 x 1 pixeles
1 canal entrada
10 canales salida
128 imágenes por batch

Bloques: 1,076,480
Threads x bloque: 1

¿Se puede optimizar esta ejecución?



NVIDIA Tesla M2090
512 CUDA-cores
16 SM

Gracias!