

Chapitre 1

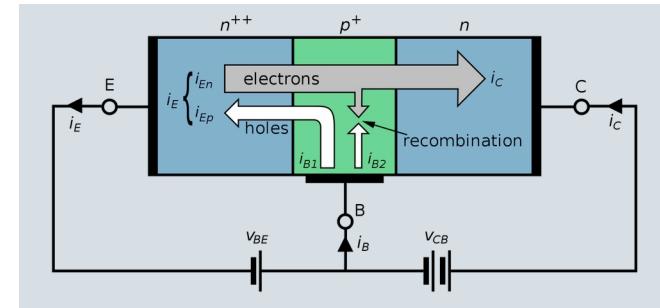
# Diversité des architectures processeur



### Composants élémentaires

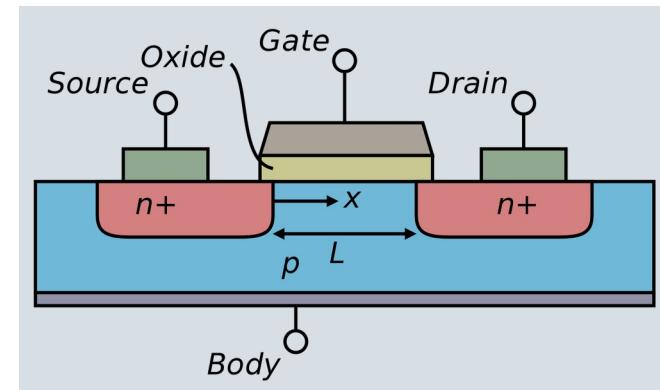
1947: Invention du **Transistor à Jonction Bipolaire (BJT)**

Bardeen, Schokley et Brattain (Bell labs), lauréats du Prix Nobel



1958/1959: Crédit des **Circuits Intégrés (IC)**

Texas Instruments (*hybrid IC*), puis Fairchild (*true monolithic IC*)



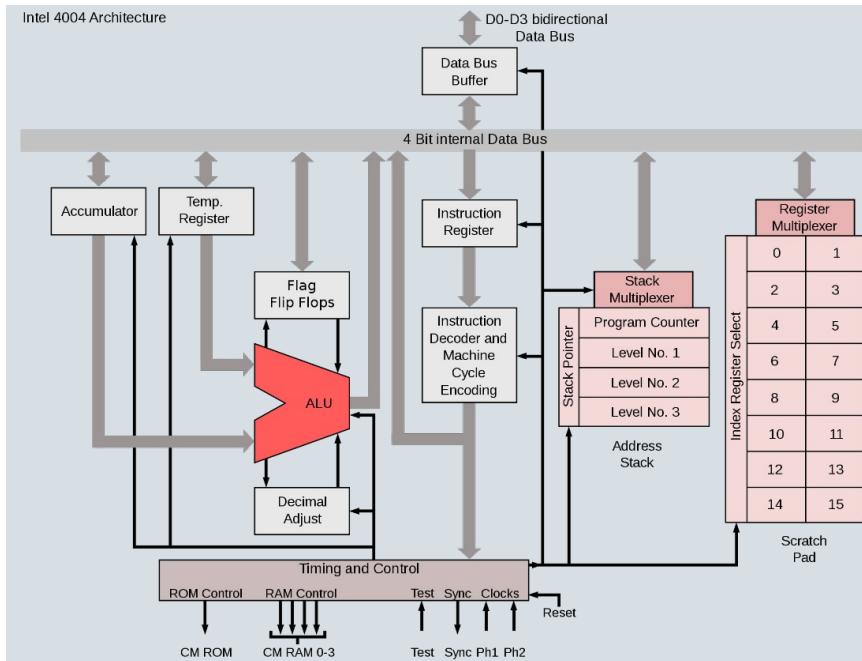
1960: Invention du **Transistor MOS à effet de champ**

Mohammed Atalla et Dawon Kahng

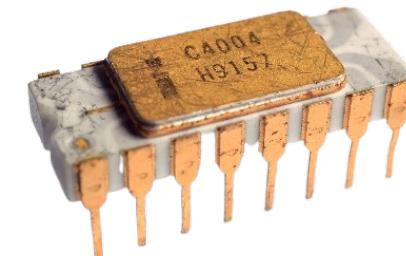
## Histoire du processeur

Le premier processeur commercial est le 4004 d'Intel, apparu en 1971.

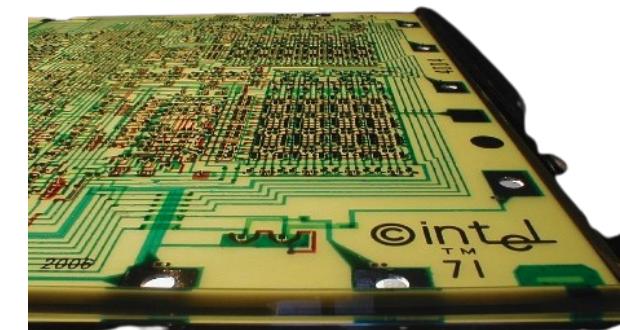
Avec 2300 transistors gravés en  $10\text{ }\mu\text{m}$ , c'est un processeur 4-bit fonctionnant à 740 kHz (90 kIPS ou kilo-Instruction Per Second). Son boîtier est un DIL-16.



Intel 4004 integrated circuit



Intel 4004 die

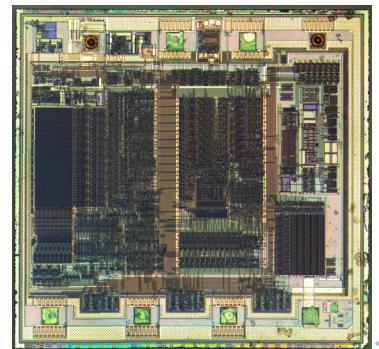
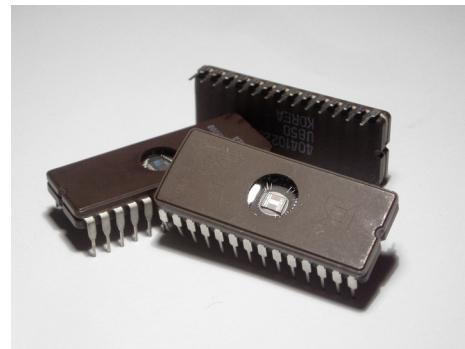
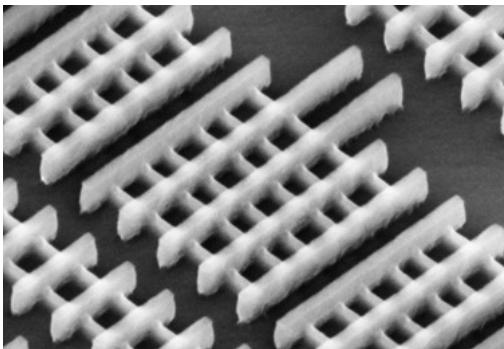
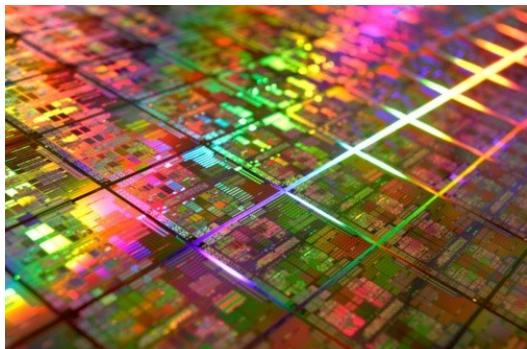
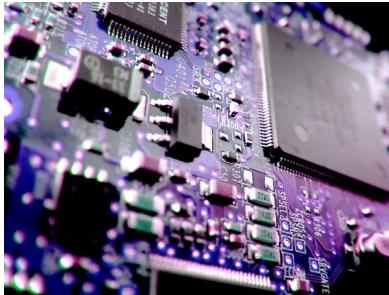


Intel 4004 architecture

## Histoire du processeur

Depuis, les processeurs ont évolués suivant un processus de sélection naturelle.

Ceux répondant à des besoins spécifiques se sont développés et améliorés, tandis que d'autres ont disparu des marchés et laboratoires de recherche.



## Histoire du processeur

Comme pour le domaine du vivant, le processus d'évolution sera toujours en cours.  
De nouvelles architectures de processeur pourraient apparaître dans le futur proche !



En attendant, regardons les principales architectures utilisées aujourd'hui.

## Architectures généralistes

*Processeurs de contrôle*

**MCU**

Micro  
Controller  
Unit

**AP**

Application  
Processor

**GPP**

General  
Purpose  
Processor

**CONTROL**



## Architectures hybrides

**SoC / SoB**

System  
on  
Chip / Board

- FPGA-AP
- FPGA-MCU
- GPP-GPU
- AP
- MCU-analog

## Architectures spécialisées

*Coprocesseurs ou processeurs de calcul*

**FPGA**

Field  
Programmable  
Gate Array

**DSP**

Digital  
Signal  
Processor

**(GP) GPU**

Graphics  
Processing  
Unit

General  
Purpose  
GPU

**CALCULUS**

## Architectures généralistes

*Processeurs de contrôle*

**MCU**

Micro  
Controller  
Unit

**AP**

Application  
Processor

**GPP**

General  
Purpose  
Processor

**CPU**

## Architectures hybrides

**SoC / SoB**

System  
on  
Chip / Board

- FPGA-AP
- FPGA-MCU
- GPP-GPU
- AP
- MCU-analog

## Architectures spécialisées

*Coprocesseurs ou processeurs de calcul*

**DSP**

Digital  
Signal  
Processor

**(GP) GPU**

Graphics  
Processing  
Unit

General  
Purpose  
GPU

**LOGIC**

**CPU**



# - MCU -

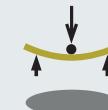
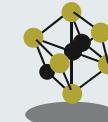
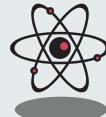
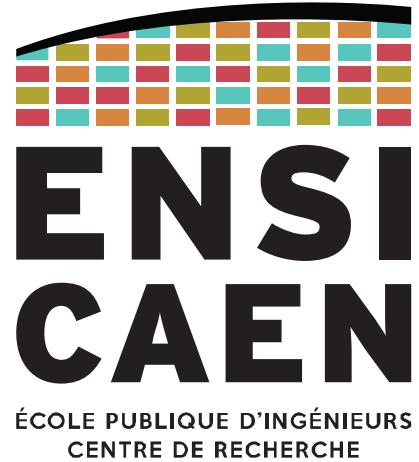
# MICROCONTROLLER UNIT

Applications

Architectures

Fabricants et produits

Parts de marché



## Applications

Les micro-contrôleurs (MCU, *Microcontroller Units*) sont les processeurs les plus répandus dans notre environnement.

De près ou de loin, nous utilisons environ 200 processeurs par jour !



Les micro-contrôleurs sont des processus dédiés à la supervision des systèmes électroniques. Ils contrôlent leur environnement via leurs interfaces et leur firmware embarqué développé pour une application spécifique.

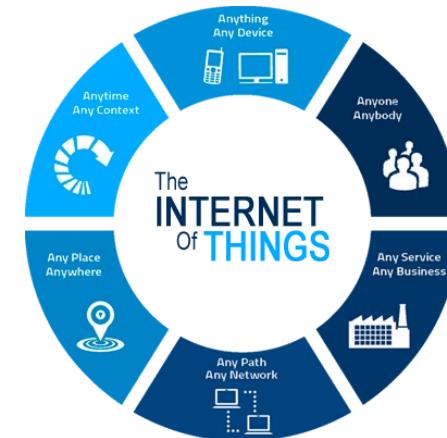
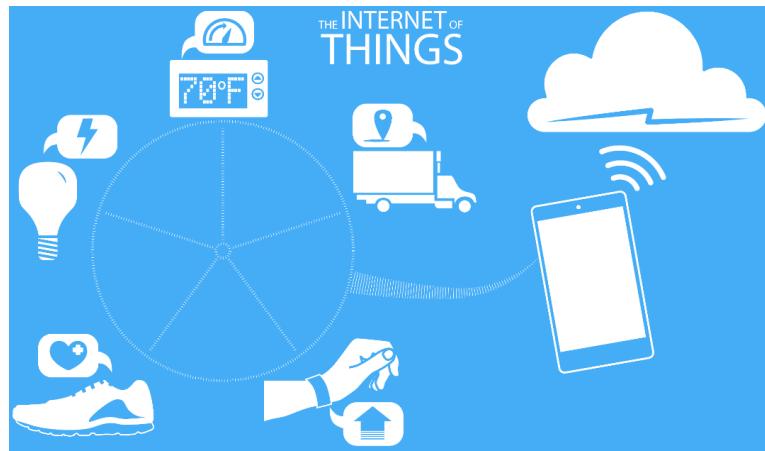
Ils ciblent des marchés où les applications sont faible coût, faible consommation, faible encombrement et gros volumes de production.



## Applications

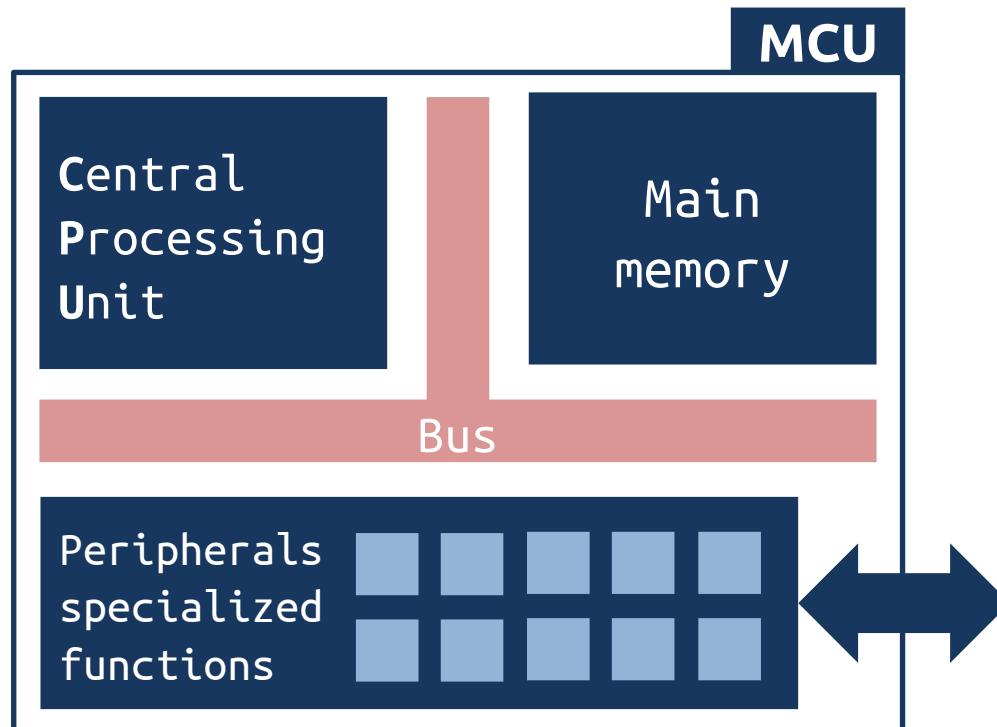
L'un des marchés phares actuels des MCU est celui des objets connectés (**IoT ou Internet of Things**). L'IoT représente l'extension d'Internet à des objets et lieux du monde physique. Il est considéré comme la troisième évolution d'Internet et, à ce titre, a été baptisé « Web 3.0 ».

Avec 3,6 milliards de connexions actives en 2015, 11,7 milliards en 2020 et 30 milliards prévues en 2025, l'IoT représentait 18 % des MCU en 2019 et 29 % en 2025.



Ces processeurs sont des systèmes numériques intégrés sur puce.

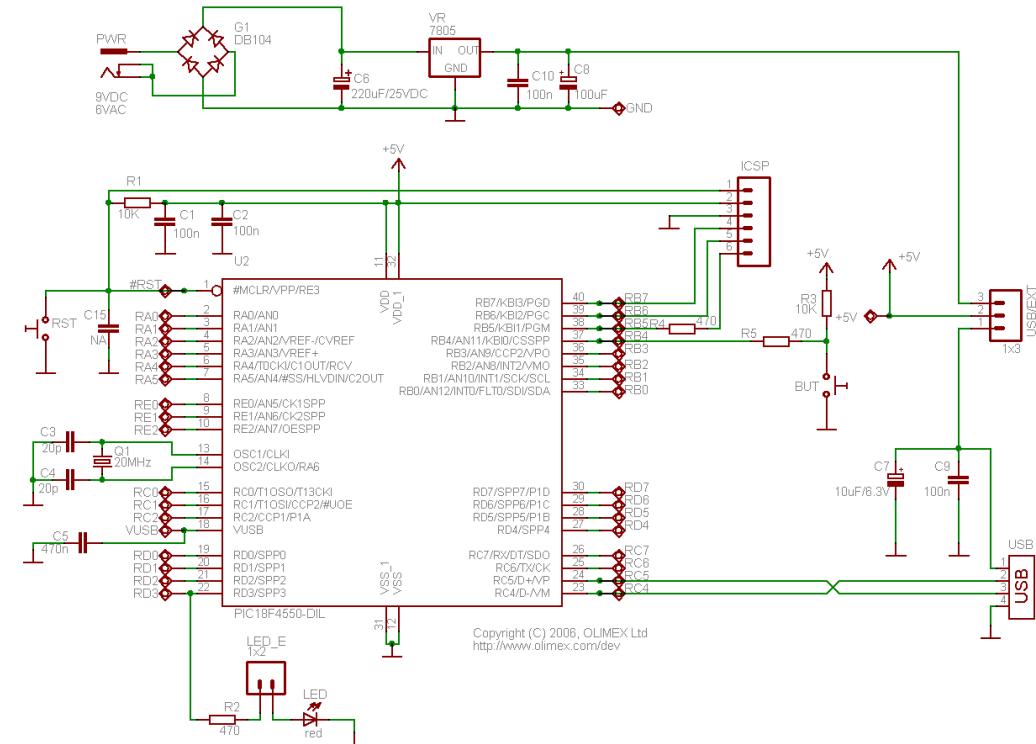
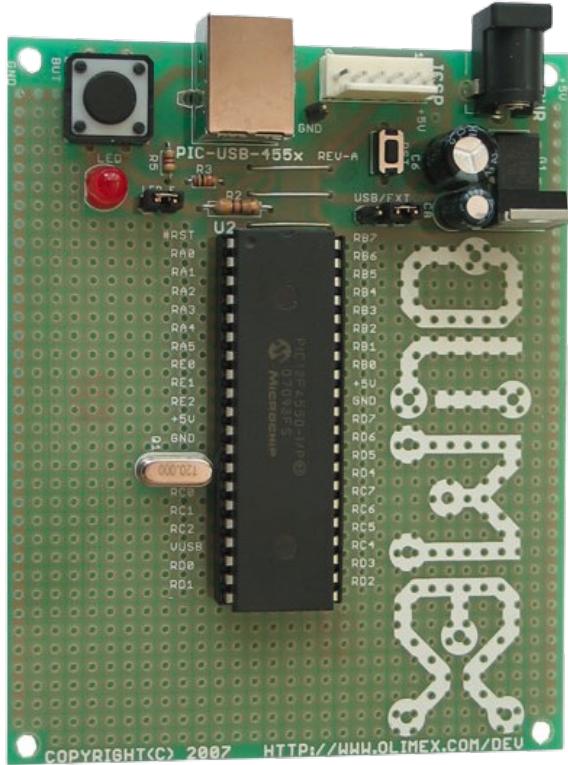
Ils sont pensés pour être autonomes (pas besoin de RAM, de HDD, ...).



## Schéma et carte

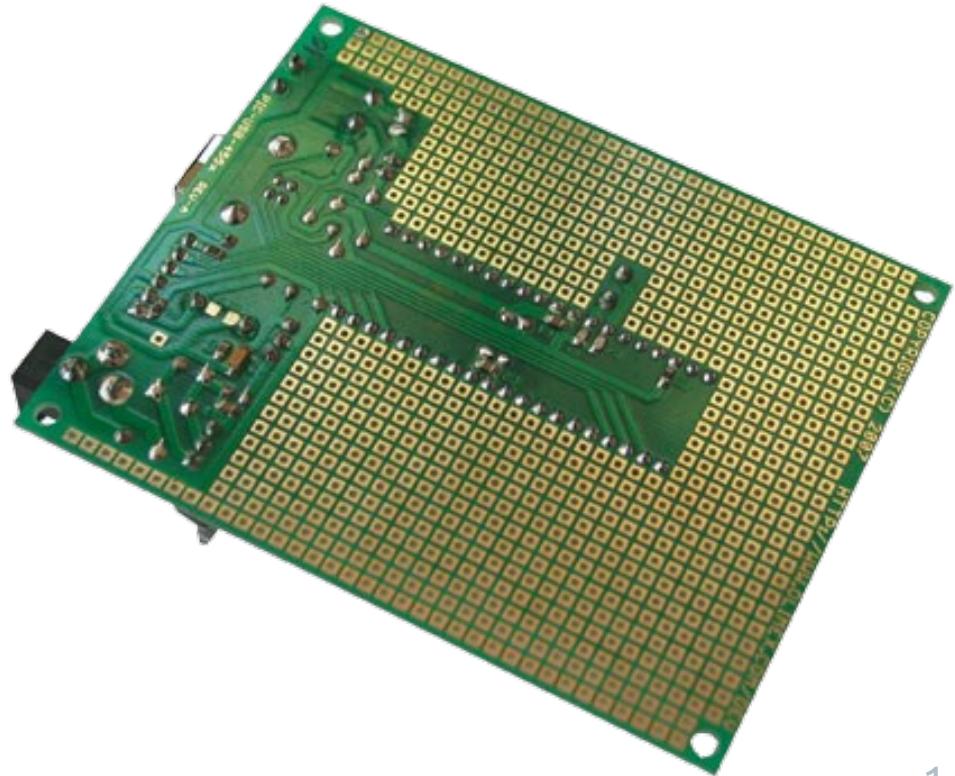
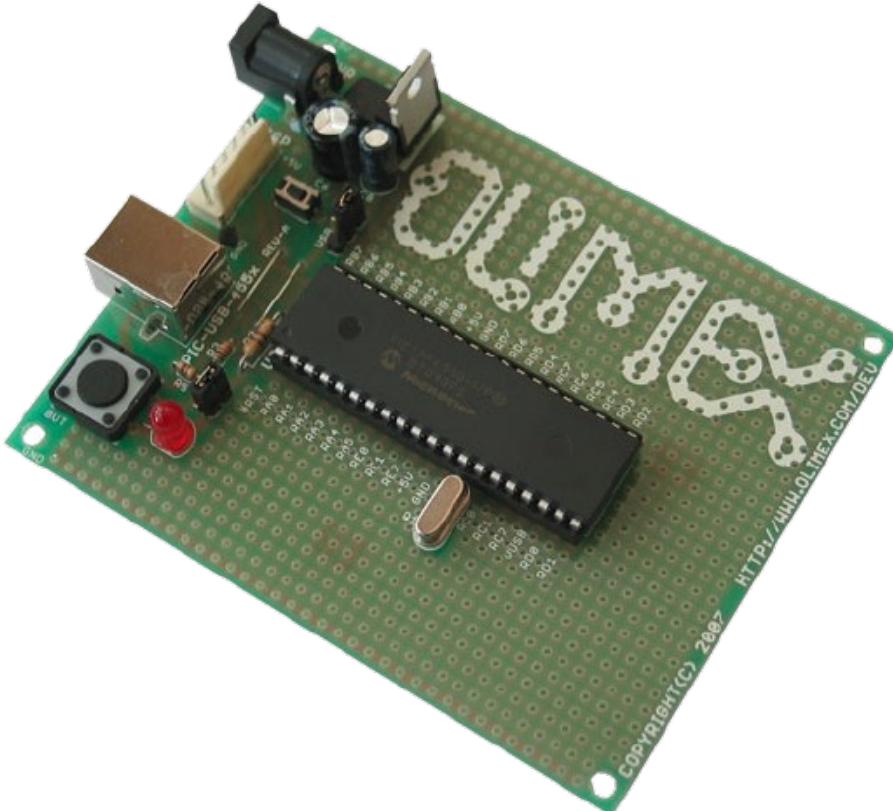
Exemple de schéma utilisant un PIC18 de Microchip.

Olimex PIC-USB-4550 board.



## Schéma et carte

Exercice : repérez les composants du schéma précédent sur les photos ci-dessous.

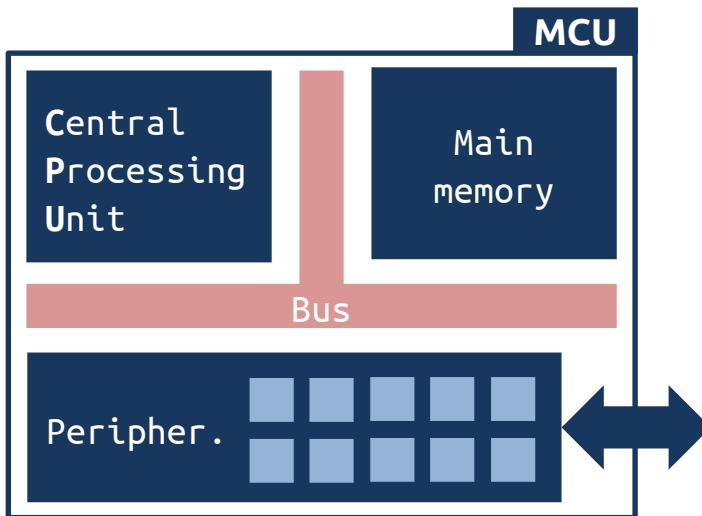




## Familles de MCU

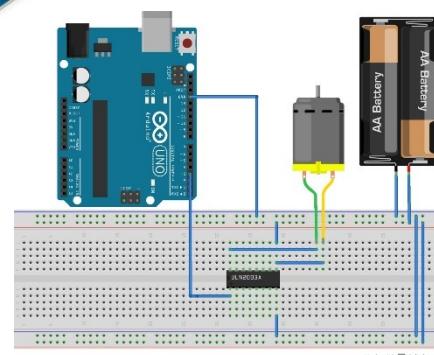
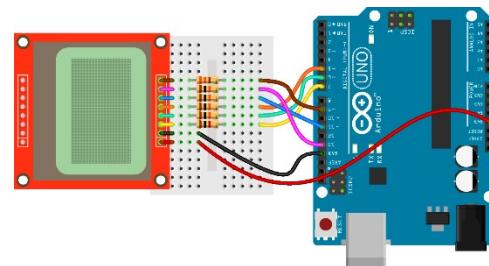
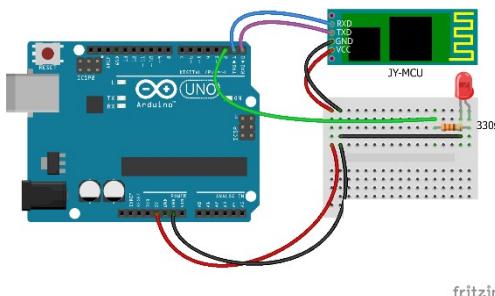
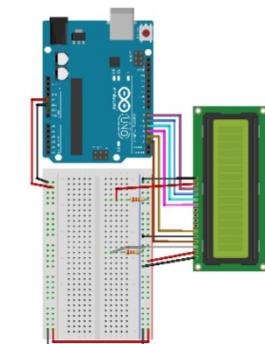
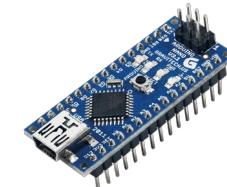
Il existe un très grand nombre de solutions MCU chez différents fondeurs, permettant de résoudre un cahier des charges.

Les MCU d'une même famille sont caractérisés par le même CPU et bus associés. Le **jeu d'instructions (ISA, Instruction Set Architecture)** et donc les outils de compilation sont similaires. Ce qui différencie les MCU d'une même famille sera le jeu de périphériques associés et les ressources mémoire disponibles.



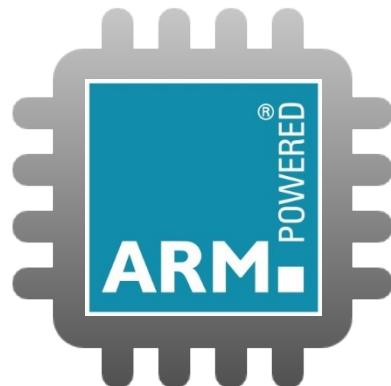
## Projet Arduino

Sûrement le plus populaire des projets électroniques basés sur un MCU, il reste déprécié en enseignements ingénieurs pour son côté trop *friendly/maker* et sa non-application aux marchés en sortie d'école.



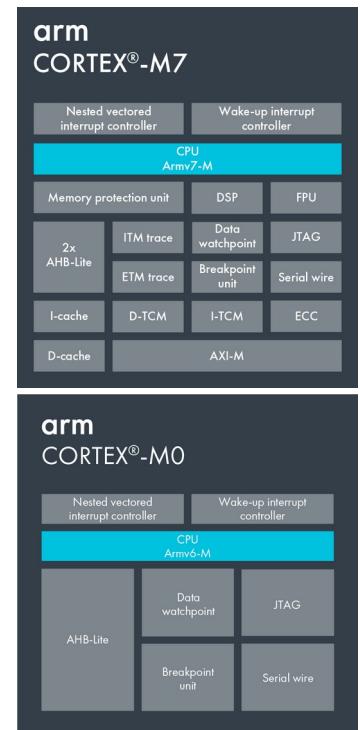
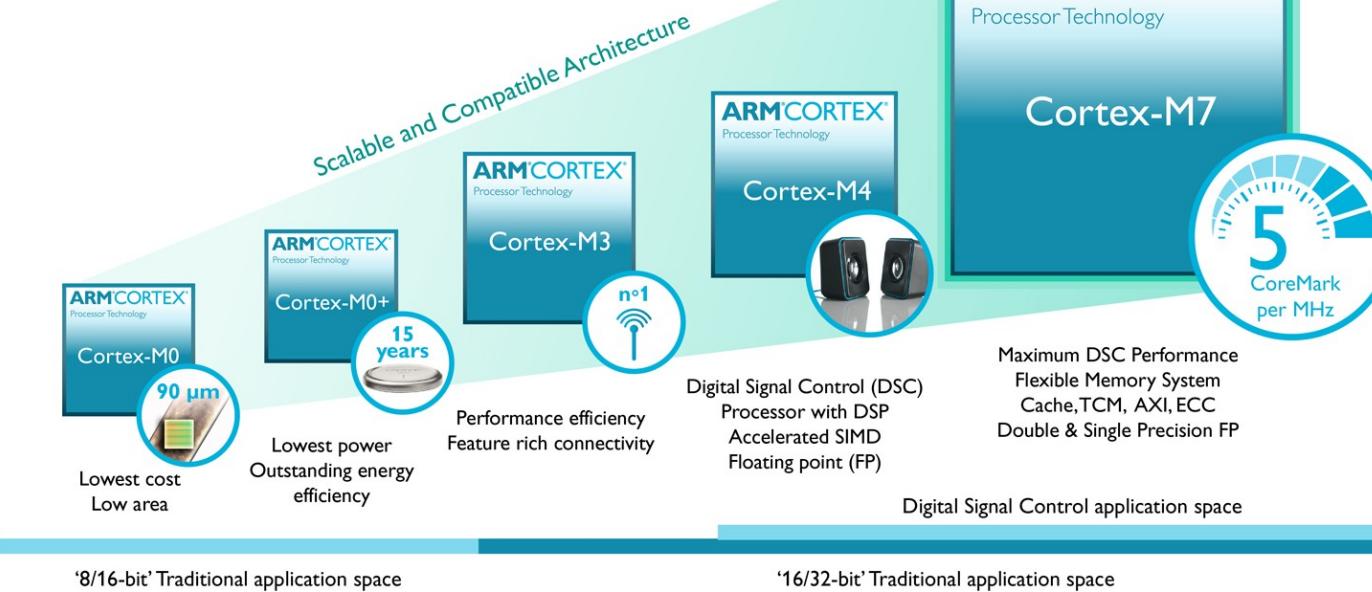
Même si le marché des MCU reste concurrentiel, la grande majorité des fondeurs de MCU (STMicro, Renesas, Texas Instruments, NXP, ...) utilisent des architectures CPU similaires, toutes proposées par la société ARM : la famille des **Cortex-M**.

Cela garanti un accès à des outils de développement, bibliothèques et services logiciels fiables, pouvant être libres et open-source (IP / *Graphical* / USB / Bluetooth, *stack*, RTOS, ...).



ARM propose la série des processeurs Cortex-M, où M signifie MCU.

Cette série comporte toute une famille de cœurs pour MCU adaptée à un large choix d'application.



Observons à titre d'illustration les gammes des STM32, qui sont des MCU 32-bits basés sur un cœur ARM Cortex-M.

Ils sont proposés par la société STMicroelectronics, société franco-italienne et principal fondeur européen.



## STMicroelectronics

Common core peripherals and architecture:

Communication peripherals:  
USART, SPI, I<sup>2</sup>C

Multiple general-purpose timers

Integrated reset and brown-out warning

Multiple DMA

2x watchdogs  
Real-time clock

Integrated regulator PLL and clock circuit

External memory interface (FSMC)

Up to 3x 12-bit DAC

Up to 4x 12-bit ADC (Up to 5 MSPS)

Main oscillator and 32 kHz oscillator

Low-speed and high-speed internal RC oscillators

-40 to +85 °C and up to 105 °C operating temperature range

Low voltage 2.0 to 3.6 V or 1.65/1.7 to 3.6 V (depending on series)

Temperature sensor

STM32 F4 series - High performance with DSP (STM32F405/415/407/417)

168 MHz Cortex-M4 with DSP and FPU	Up to 192-Kbyte SRAM	Up to 1-Mbyte Flash	2x USB 2.0 OTG FS/HS	3-phase MC timer	2x CAN 2.0B	SDIO 2x I <sup>2</sup> S audio Camera IF	Ethernet IEEE 1588	Crypto/ hash processor and RNG
------------------------------------	----------------------	---------------------	----------------------	------------------	-------------	--	--------------------	--------------------------------



STM32 F3 series - Mixed-signal with DSP (STM32F302/303/313/372/373/383)

72 MHz Cortex-M4 with DSP and FPU	Up to 48-Kbyte SRAM & CCM-SRAM	Up to 256-Kbyte Flash	USB 2.0 FS	2x 3-phase MC timer (144 MHz)	CAN 2.0B	Up to 7x comparator	3x 16-bit ΣΔ ADC	4x PGA
-----------------------------------	--------------------------------	-----------------------	------------	-------------------------------	----------	---------------------	------------------	--------



STM32 F2 series - High performance (STM32F205/215/207/217)

120 MHz Cortex-M3 CPU	Up to 128-Kbyte SRAM	Up to 1-Mbyte Flash	2x USB 2.0 OTG FS/HS	3-phase MC timer	2x CAN 2.0B	SDIO 2x I <sup>2</sup> S audio Camera IF	Ethernet IEEE 1588	Crypto/ hash processor and RNG
-----------------------	----------------------	---------------------	----------------------	------------------	-------------	--	--------------------	--------------------------------



STM32 F1 series - Mainstream - 5 product lines (STM32F100/101/102/103 and 105/107)

Up to 72 MHz Cortex-M3 CPU	Up to 96-Kbyte SRAM	Up to 1-Mbyte Flash	USB 2.0 OTG FS	3-phase MC timer	Up to 2x CAN 2.0B	SDIO 2x I <sup>2</sup> S audio	Ethernet IEEE 1588	
----------------------------	---------------------	---------------------	----------------	------------------	-------------------	--------------------------------	--------------------	--



STM32 F0 series – Entry level (STM32F050/051)

48 MHz Cortex-M0 CPU	Up to 12-Kbyte SRAM	Up to 128-Kbyte Flash	3-phase MC timer	Comparator	CEC			
----------------------	---------------------	-----------------------	------------------	------------	-----	--	--	--



STM32 L1 series - Ultra-low-power (STM32L151/152/162)

32 MHz Cortex-M3 CPU	Up to 48-Kbyte SRAM	Up to 384-Kbyte Flash	USB FS device	Up to 12-Kbyte EEPROM	LCD 8x40 4x44	Comparator	BOR MSI VScal	AES 128-bit
----------------------	---------------------	-----------------------	---------------	-----------------------	---------------	------------	---------------	-------------

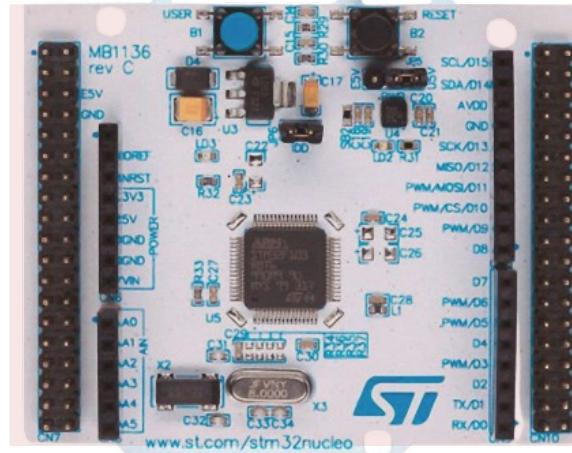
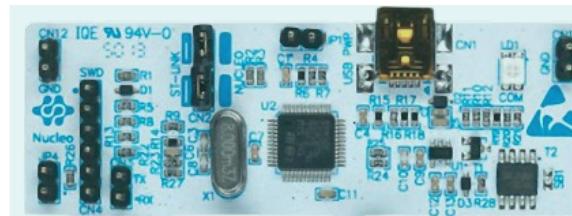
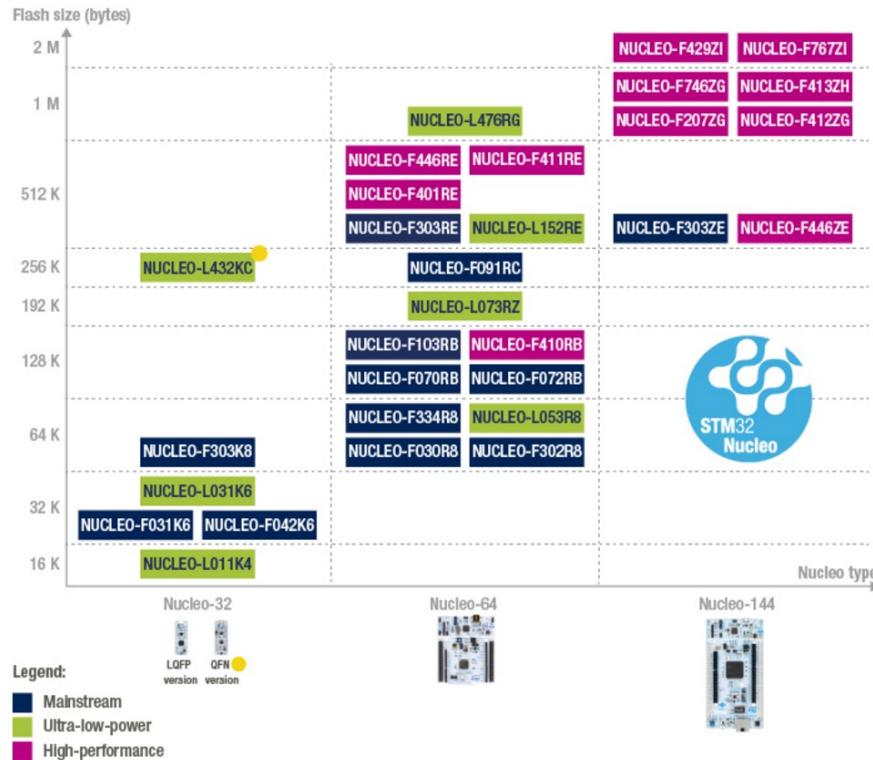


STM32 W series - Wireless (STM32W108)

24 MHz Cortex-M3 CPU	Up to 16-Kbyte SRAM	Up to 256-Kbyte Flash	2.4 GHz IEEE 802.15.4 Transceiver	Lower MAC Digital baseband	AES 128-bit	
----------------------	---------------------	-----------------------	-----------------------------------	----------------------------	-------------	--



Le projet Nucleo propose des maquettes d'évaluation à bas coût utilisant des solutions MCU et outils de développement de l'industrie ( $\approx 10$  €).



Nucleo-64

- Power supply
- Programmer (JTAG emulator)
- Target MCU
- Switch and LED
- External ports
- Shields connectors
- Arduino shield connectors

Observons les résultats d'une étude de marché réalisée chaque année.



## 2019 Embedded Markets Study

Integrating IoT and Advanced Technology Designs,  
Application Development & Processing Environments

March 2019

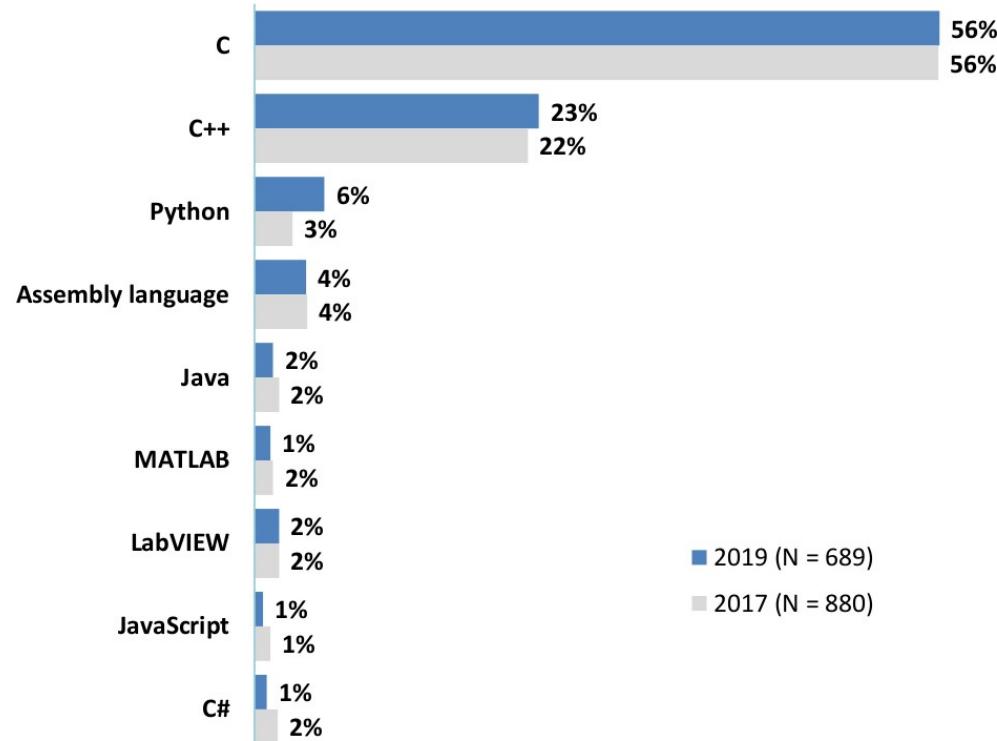
Presented By: **EE Times** **embedded**

© 2019 AspenCore All Rights Reserved

## Acteurs et marchés

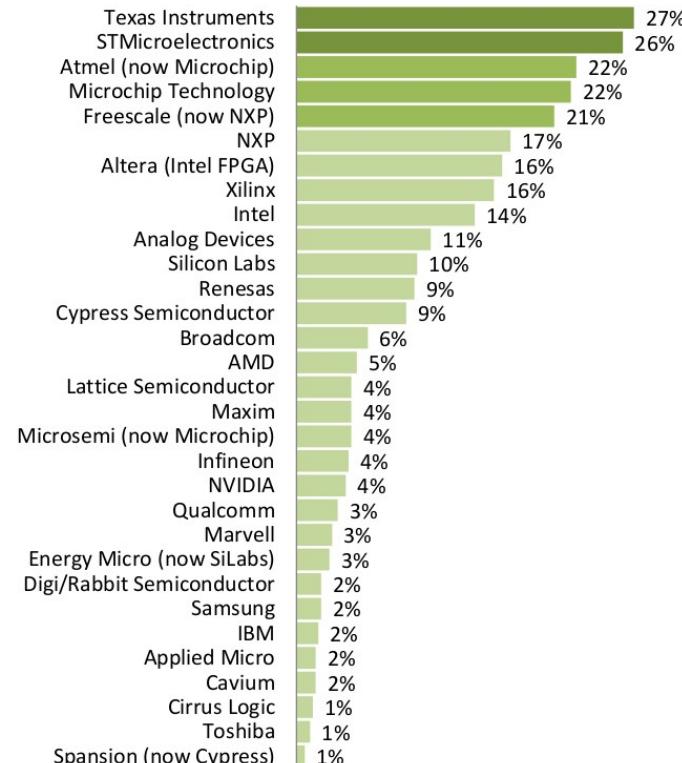


*My current embedded project is programmed mostly in:*



## Acteurs et marchés

Please select the processor vendors you are currently using.



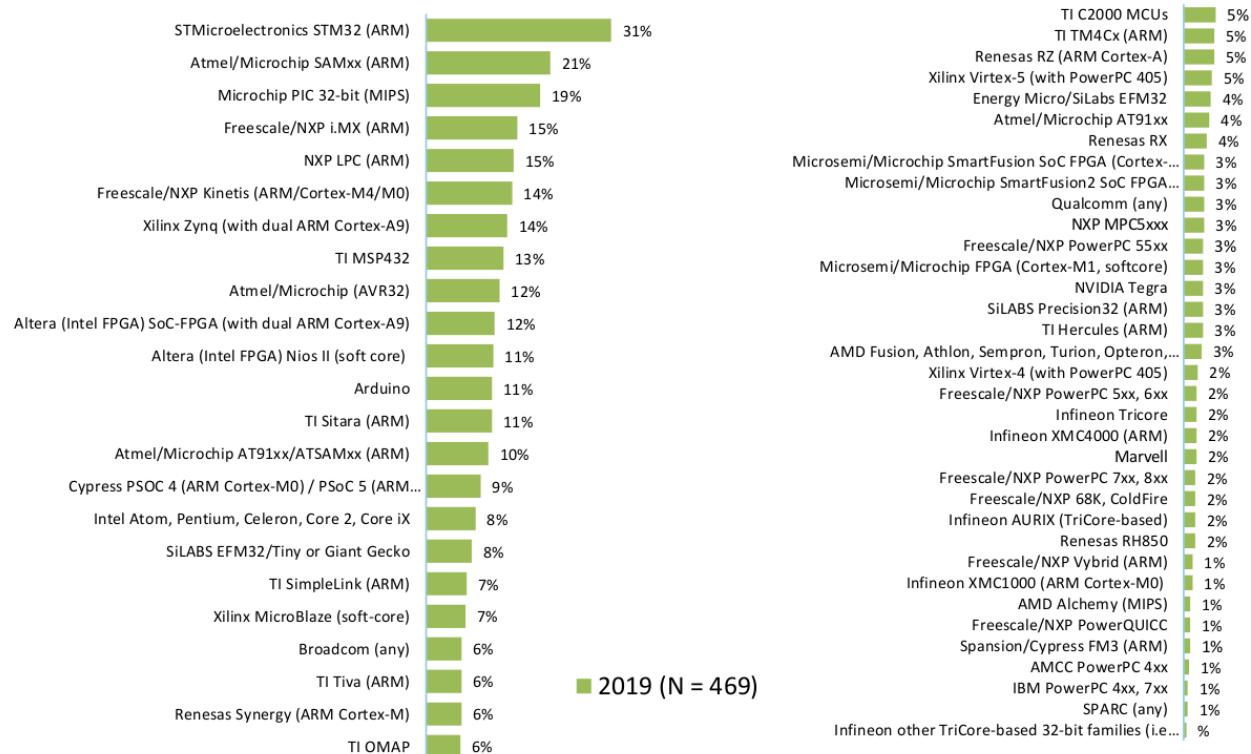
2019 (N = 458)

**Top Four Brands by Region:**  
**Americas:** TI, Microchip, STMicro, Atmel  
**EMEA:** STMicro, NXP, TI, Atmel  
**APAC:** TI, Atmel, Freescale, STMicro

## Acteurs et marchés



Which of the following 32-bit chip families would you consider for your next embedded project?

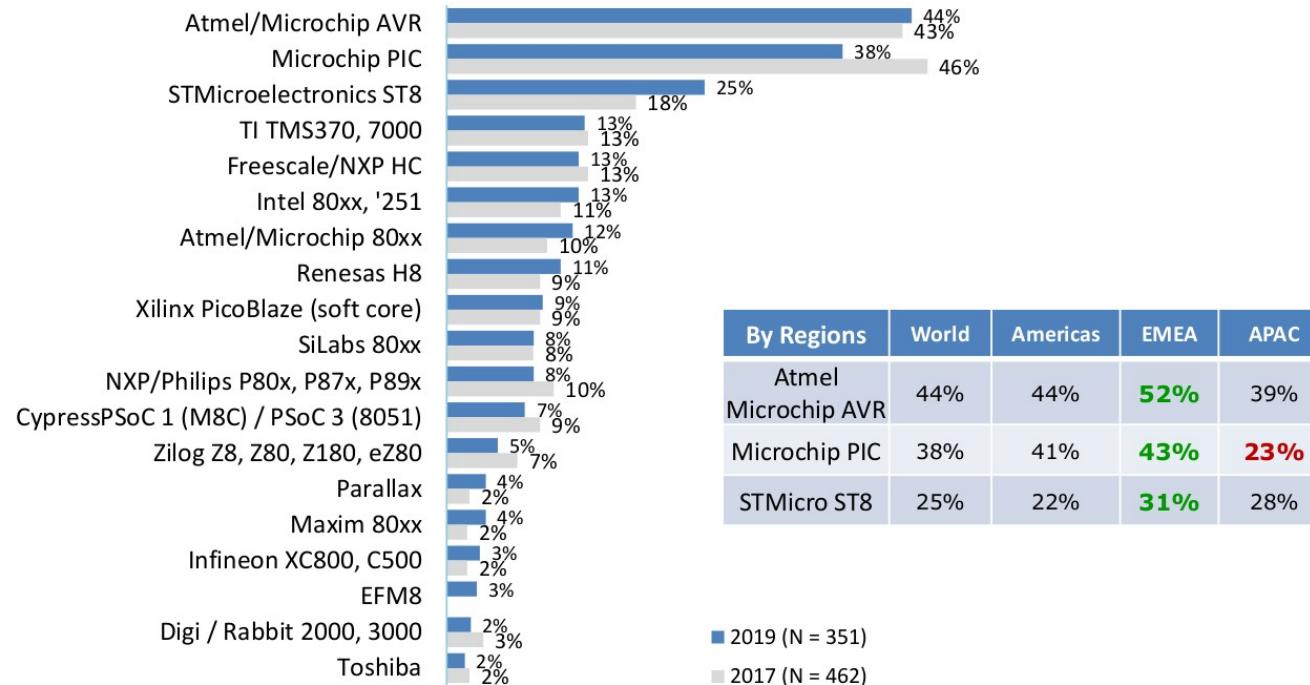


■ 2019 (N = 469)

## Acteurs et marchés



Which of the following 8-bit chip families would you consider for your next embedded project?



# - GPP -

# GENERAL PURPOSE PROCESSOR

Applications

Architecture

Carte mère

Processeur superscalaire



## Applications

Les **General Purpose Processors (GPP)** possèdent une architecture CPU complexe leur offrant une **grande polyvalence**, notamment à l'exécution de code faiblement optimisé.

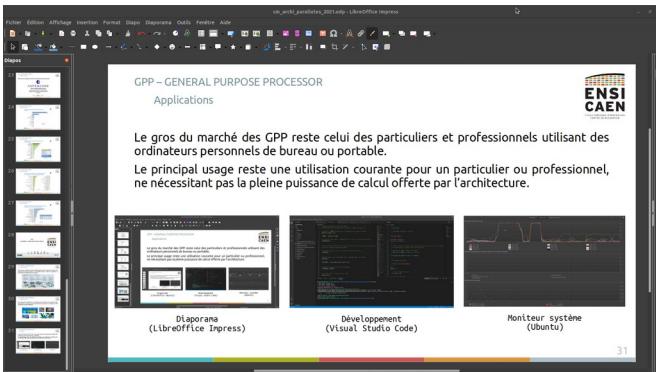
Il s'agit par exemple de programmes de contrôle offrant un code séquentiel avec un grand nombre de tests et d'appels de fonctions. Codes difficiles à accélérer.

```
444     prev = NULL;
445     for (mpnt = oldmm->mmap; mpnt; mpnt = mpnt->vm_next) {
446         struct file *file;
447
448         if (mpnt->vm_flags & VM_DONTCOPY) {
449             vm_stat_account(mm, mpnt->vm_flags, -vma_pages(mpnt));
450             continue;
451         }
452         charge = 0;
453         if (mpnt->vm_flags & VM_ACCOUNT) {
454             unsigned long len = vma_pages(mpnt);
455
456             if (security_vm_enough_memory_mm(oldmm, len)) /* sic */
457                 goto fail_nomem;
458             charge = len;
459         }
460         tmp = kmem_cache_alloc(vm_area_cachep, GFP_KERNEL);
461         if (!tmp)
462             goto fail_nomem;
463         *tmp = *mpnt;
464         INIT_LIST_HEAD(&tmp->anon_vma_chain);
465         retval = vma_dup_policy(mpnt, tmp);
466         if (retval)
467             goto fail_nomem_policy;
```

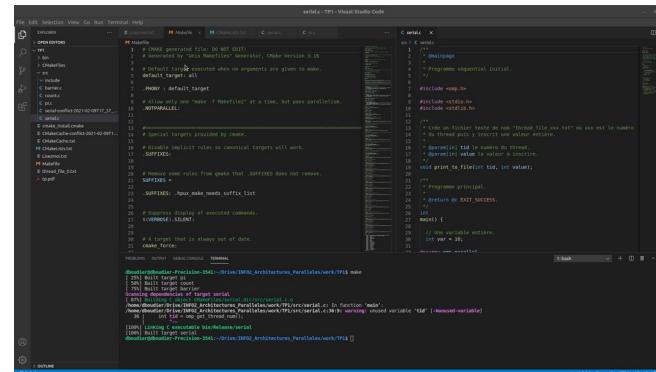
## Applications

Le gros du marché des GPP reste celui des particuliers et professionnels utilisant des ordinateurs personnels de bureau ou portable.

Le principal usage reste une utilisation courante pour un particulier ou professionnel, ne nécessitant pas la pleine puissance de calcul offerte par l'architecture.



Diaporama  
(LibreOffice Impress)



Développement  
(Visual Studio Code)

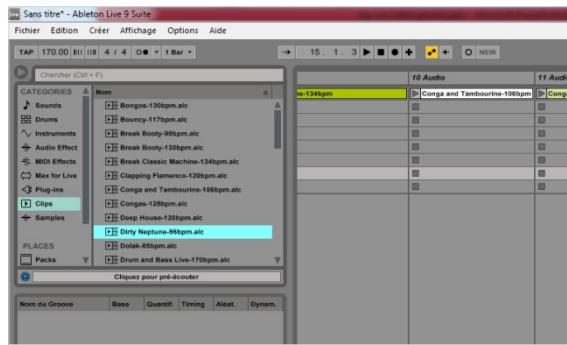


Moniteur système  
(Ubuntu)

## Applications

On peut également citer les applications de traitement du son, de traitement d'image, de traitement du signal, de développement logiciel ou de montages de médias.

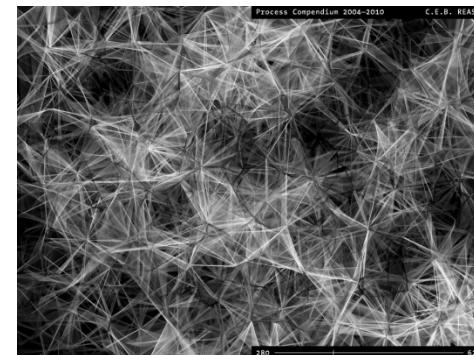
Celles-ci sont plus contraignantes au regard des ressources et exploitent souvent le plein potentiel du matériel.



Montage audio (Ableton)



Traitement du son



Traitement d'image

## Applications

Les applications industrielles sont également un terrain historique des GPP.

Ils sont typiquement rencontrés sur des tâches de contrôle ou des fonctions de calculs spécialisés. Ce marché tend à utiliser des solutions intégrées (AP, SoC, DSP, FPGA).



Radar GM400  
(Thalès)



Rafale  
(Dassault)

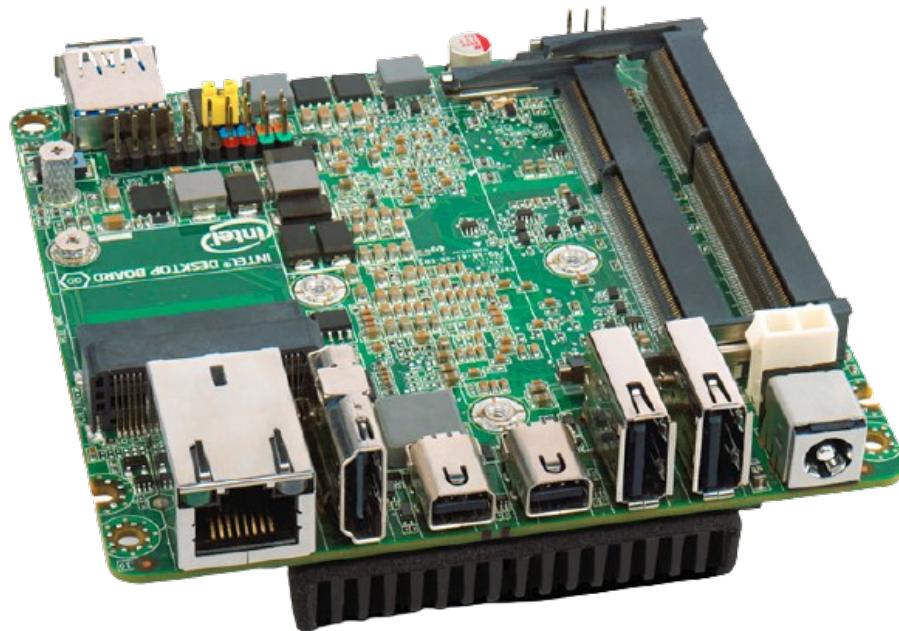


Borne automatique  
Box j200

## Applications

Notons que les GPP peuvent également être exploités par des applications rattachées au domaine des systèmes embarqués.

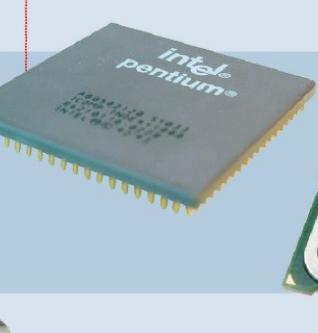
Voici par exemple la carte mère NUC Core i5 de Intel.



## Architectures Intel

Observons les architectures phares d'Intel, leader actuel et historique du marché des GPP (*General Purpose Processor*) ou MPU (*MicroProcessor Unit*) mais également du marché des semi-conducteurs au sens large.

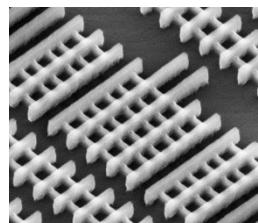
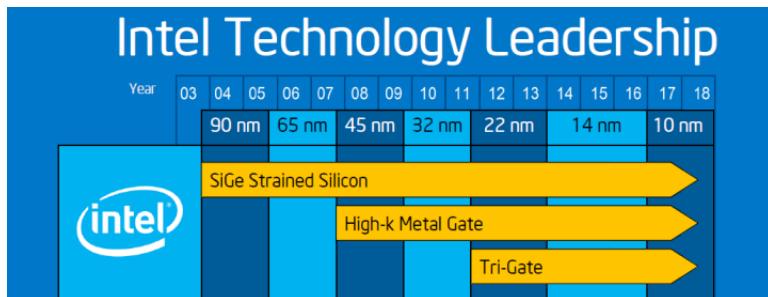
### 40 ANS DE COURSE À L'INNOVATION

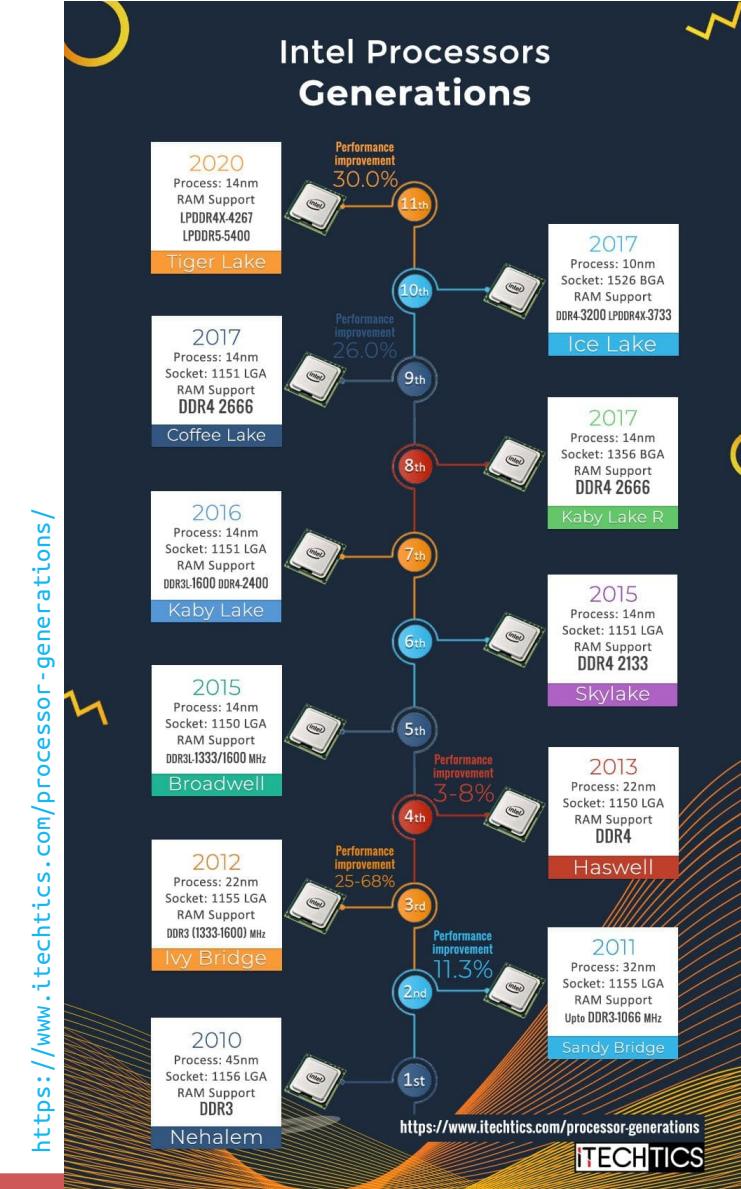
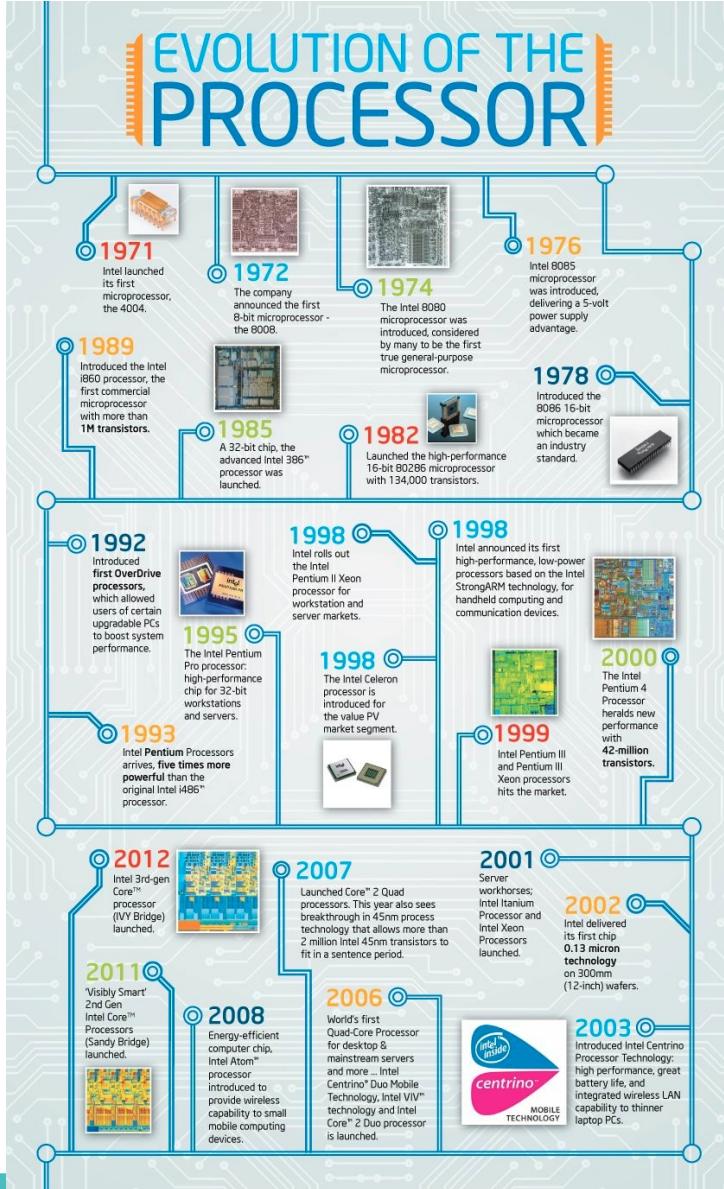
1971	1981	1993	2006	2012
<b>PROCESSEUR 4004 D'INTEL</b> Nombre de transistors : <b>2.300</b> Puissance : <b>108 kilohertz</b> <b>10 microns</b>  	<b>PROCESSEUR 8088</b> Introduit dans les PC d'IBM Nombre de transistors : <b>29.000</b> Puissance : <b>5 megahertz</b> <b>3 microns</b>  	<b>PENTIUM</b> Nombre de transistors : <b>3.1 millions</b> Puissance : <b>66 megahertz</b> <b>0,8 micron</b>  	<b>INTEL CORE 2 DUO</b> Nombre de transistors : <b>291 millions</b> Puissance : <b>2.93 gigahertz</b> <b>65 nanomètres</b>  	<b>PROCESSEURS IVY BRIDGE</b> Nombre de transistors : <b>1.400 millions (3D)</b> Puissance non communiquée <b>22 nanomètres</b>  

IDÉ / SOURCE ET PHOTOS : INTEL

## Architectures Intel

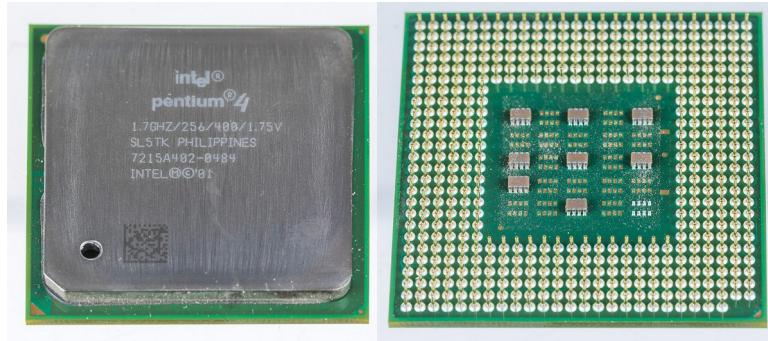
Les architectures GPP phares à notre époque sont les familles Core i3/i5/i7 de Intel. Mais prudence, il existe un grand nombre d'autres architectures et fondeurs de GPP ciblant divers marchés différents.





## Architectures Intel

4004	(1971)	Processeur 4 bit
8008	(1972)	Processeur 8 bit
<b>8086</b>	(1978)	Processeur 16 bit
<b>80386</b>	(1985)	Processeur 32 bit
Pentium	(1993)	Processeur 32 bit
<b>Pentium 4</b>	(2000)	Processeur 32 bit
Core 2 Duo	(2006)	Proc. 32/64 bit
		→ Apparition du multi-core chez Intel
		→ Naissance de l'ISA x86-64 (calé sur celui d'AMD), rétro-compatible x86-32 et x86-16 !
Core	(2008)	12 générations se succèdent jusqu'à aujourd'hui (2022)



→ Premier CPU x86 (ISA x86-16)

→ ISA x86-32, rétro-compatible x86-16

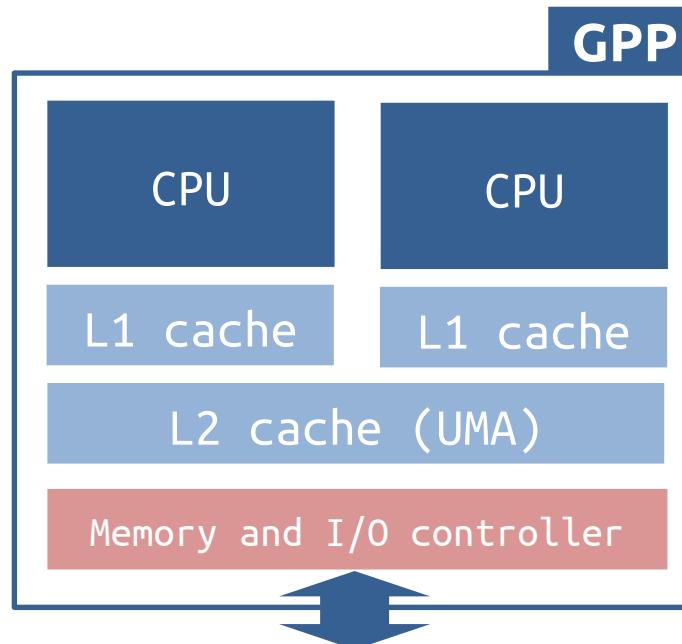
→ Premier superscalaire commercialisé

→ 2 cœurs logiques (2 threads)

## Architecture

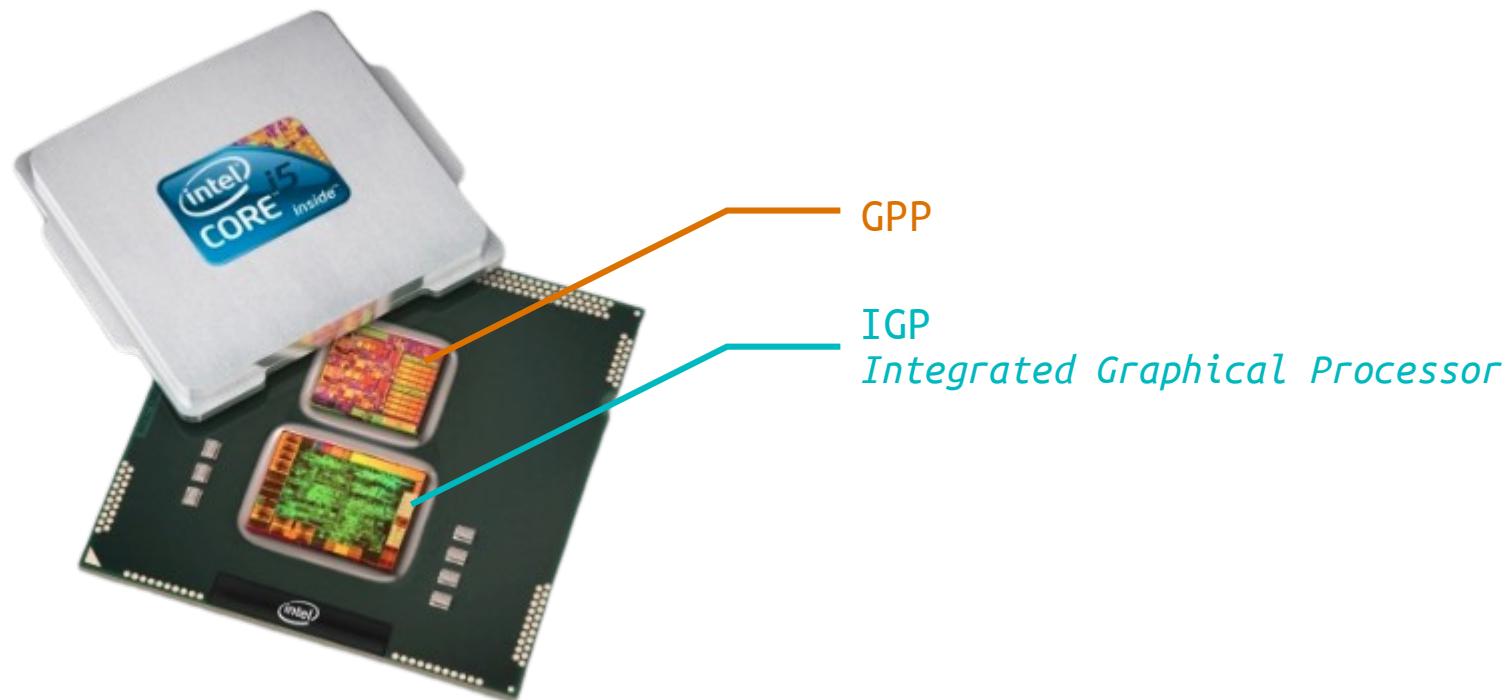
Processeur de traitement nu, dépourvu de mémoire principale.

Il embarque un ou plusieurs CPU (architecture homogène) mariés avec leurs caches, possède un modèle mémoire uniforme (UMA) et embarque un contrôleur d'interfaces.



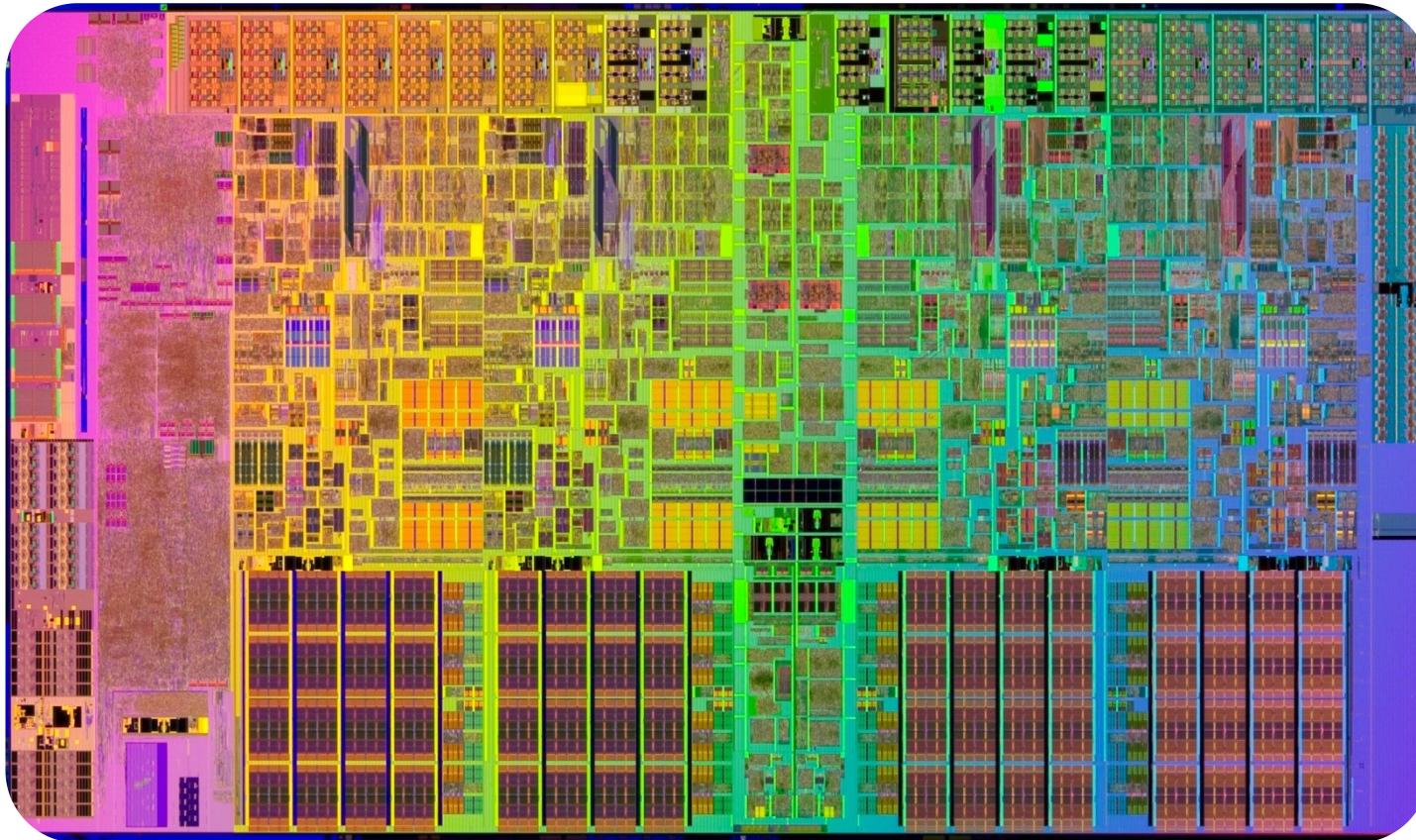
Exemple : Intel Core i5

Exemple de la famille Core i5 de Intel.



# GPP – GENERAL PURPOSE PROCESSOR

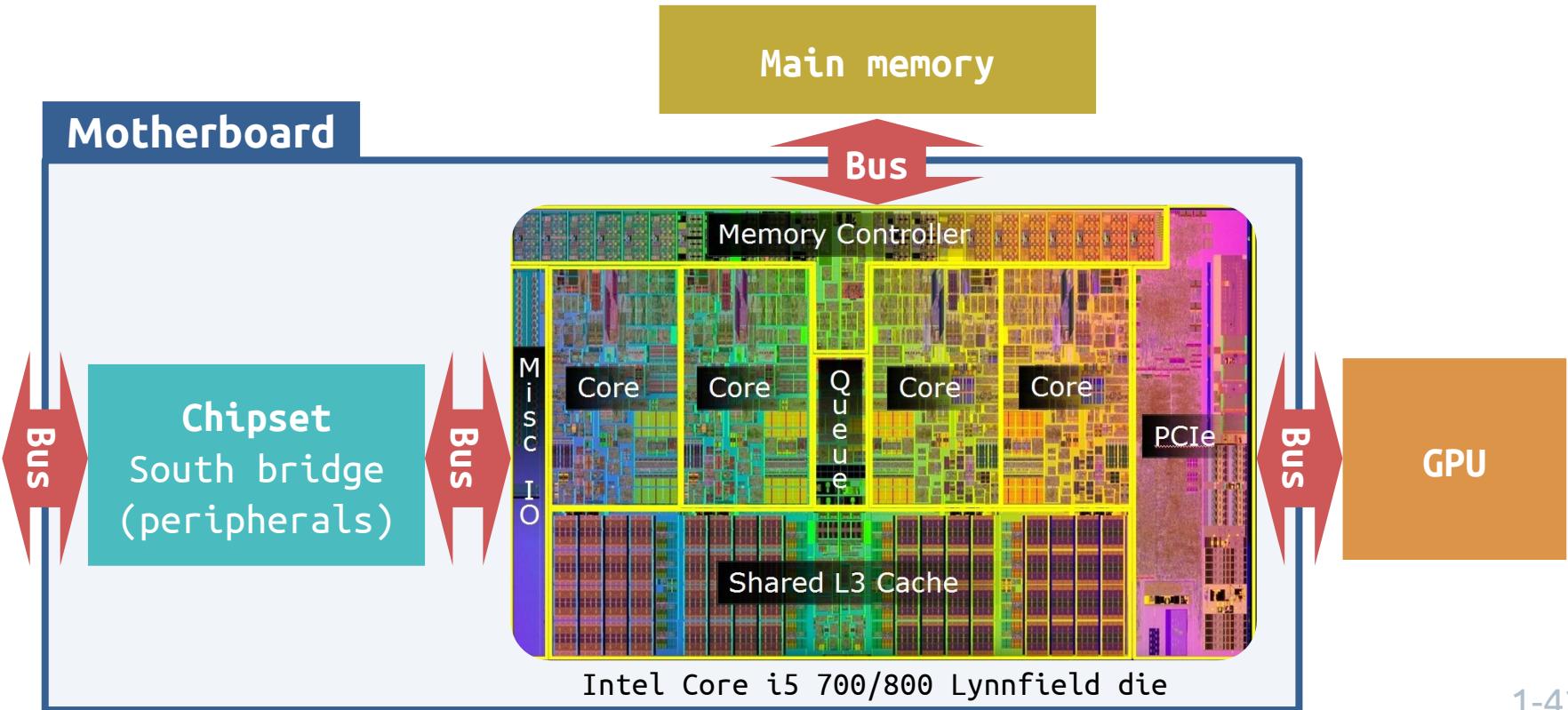
Exemple : Intel Core i5



Intel Core i5 700/800 Lynnfield die

Exemple : Intel Core i5

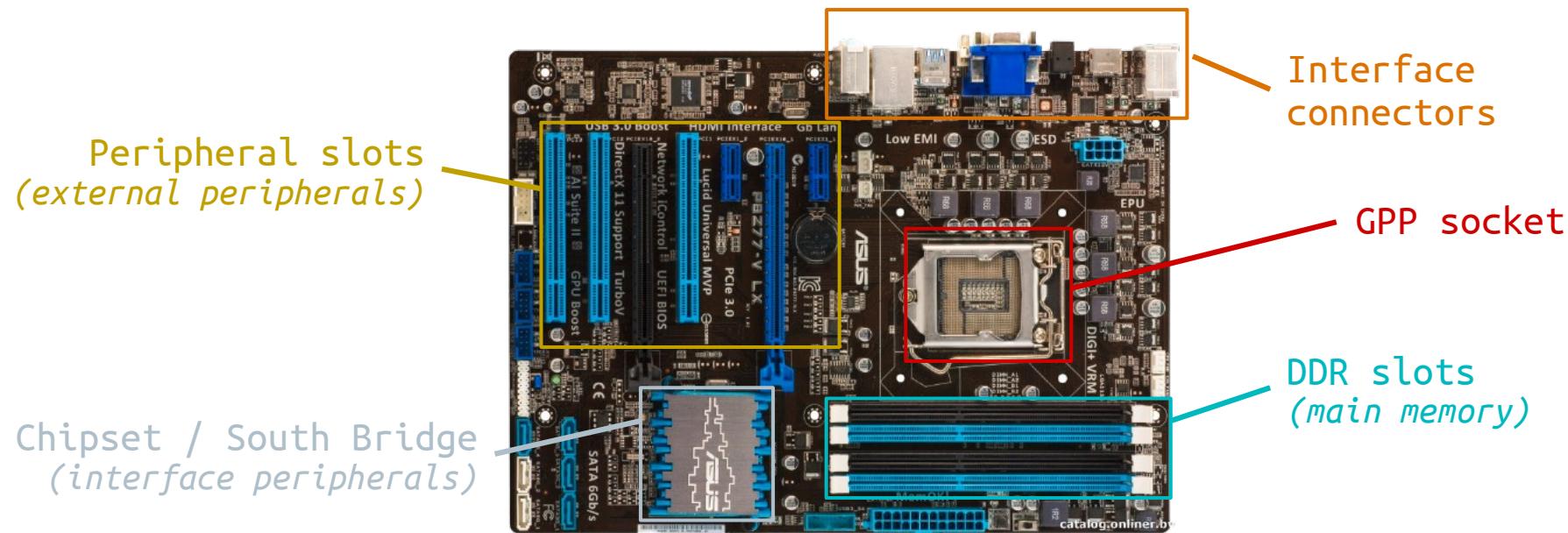
## Intégration dans le système (carte mère)



## Carte mère

Un GPP doit forcément être porté sur une carte mère avec mémoire principale et périphériques d'interfaces externes déportés.

Exemple de carte mère ASUS, n°2 du marché mondial en 2016.

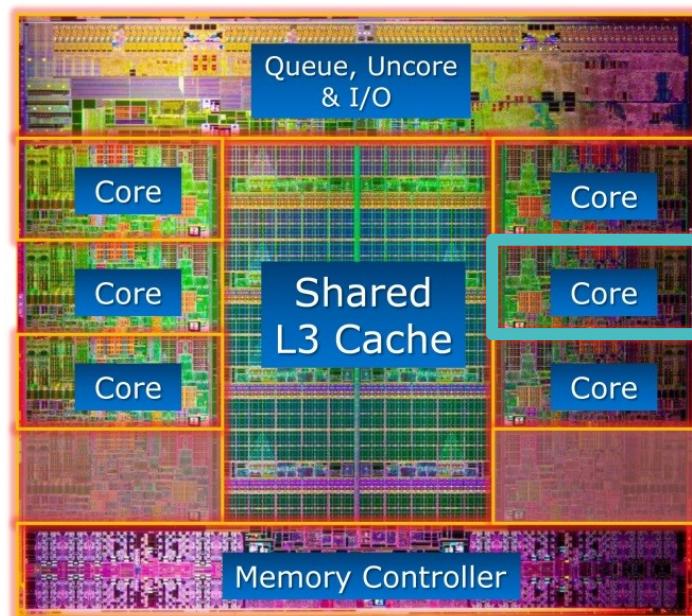


Les GPP possèdent un CPU dit **superscalaire**. Les processeurs possédant ce type de pipeline CPU se caractérisent le plus souvent par le déploiement des mécanismes d'accélération matériels suivants :

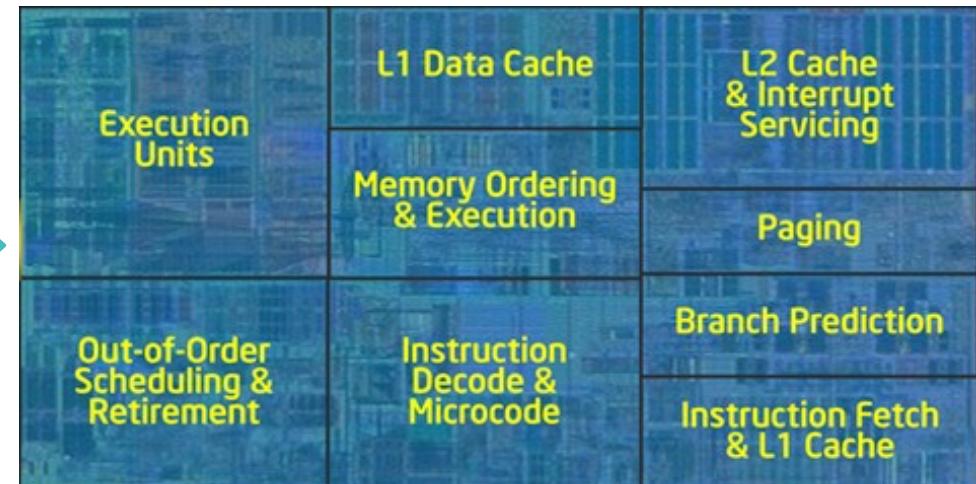
- **Étage d'exécution *Out Of Order*** : Exécution des instructions dans le désordre. Ordonnanceur matériel gérant les dépendances fonctionnelles et sur les données, étages de renommage des registres (résultats intermédiaires) et de réordonnancement
- **Étage de prédiction au branchement**
- **Étage d'exécution *RISC-like***, même si l'ISA est CISC

## Architecture superscalaire

Die d'un CPU de la génération Sandy Bridge de Intel, illustré pour un Core i7.



Intel Core i7



Sandy Bridge CPU/Core

## Architecture superscalaire

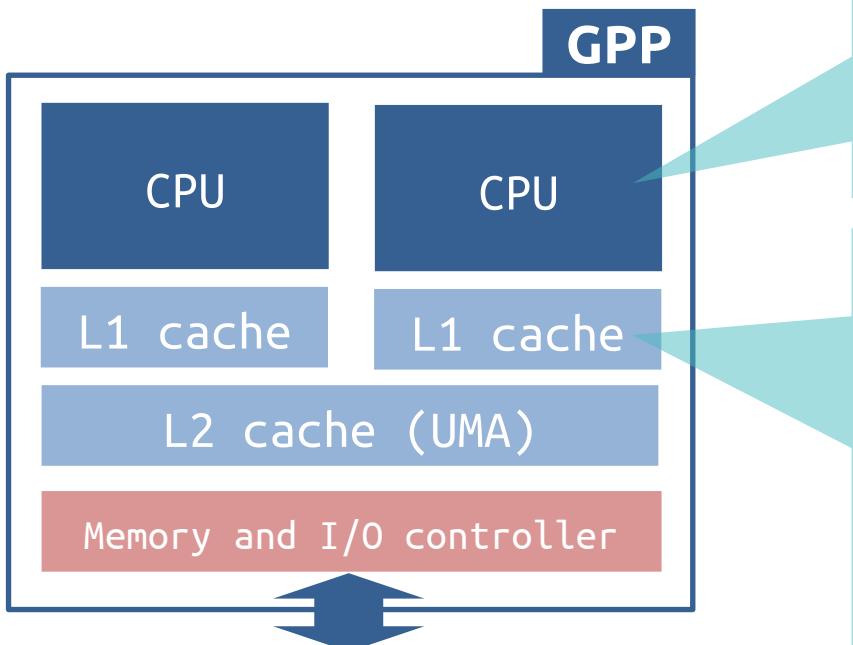
Attention, cette grande polyvalence et complexité matérielle se paye par un manque de déterminisme voire de performance à l'exécution sur des traitements algorithmiques spécifiques.

**Les GPP offrent un ratio performance de calcul ramené au coût et au Watt peu intéressant.**

Ils sont pensés pour porter un OS (*Operating System*) évolué et exécuter du code applicatif. Prenons les exemples des applications de traitement du son, traitement d'image, traitement vidéo, traitement d'antenne ... pour lesquels ils ne sont pas spécialisés.



À savoir



### CPU superscalaire

- exécution Out Of Order
- prédiction de branchement
- non déterministe
- mauvais ratio (puissance calcul) / (Watt x Coût)

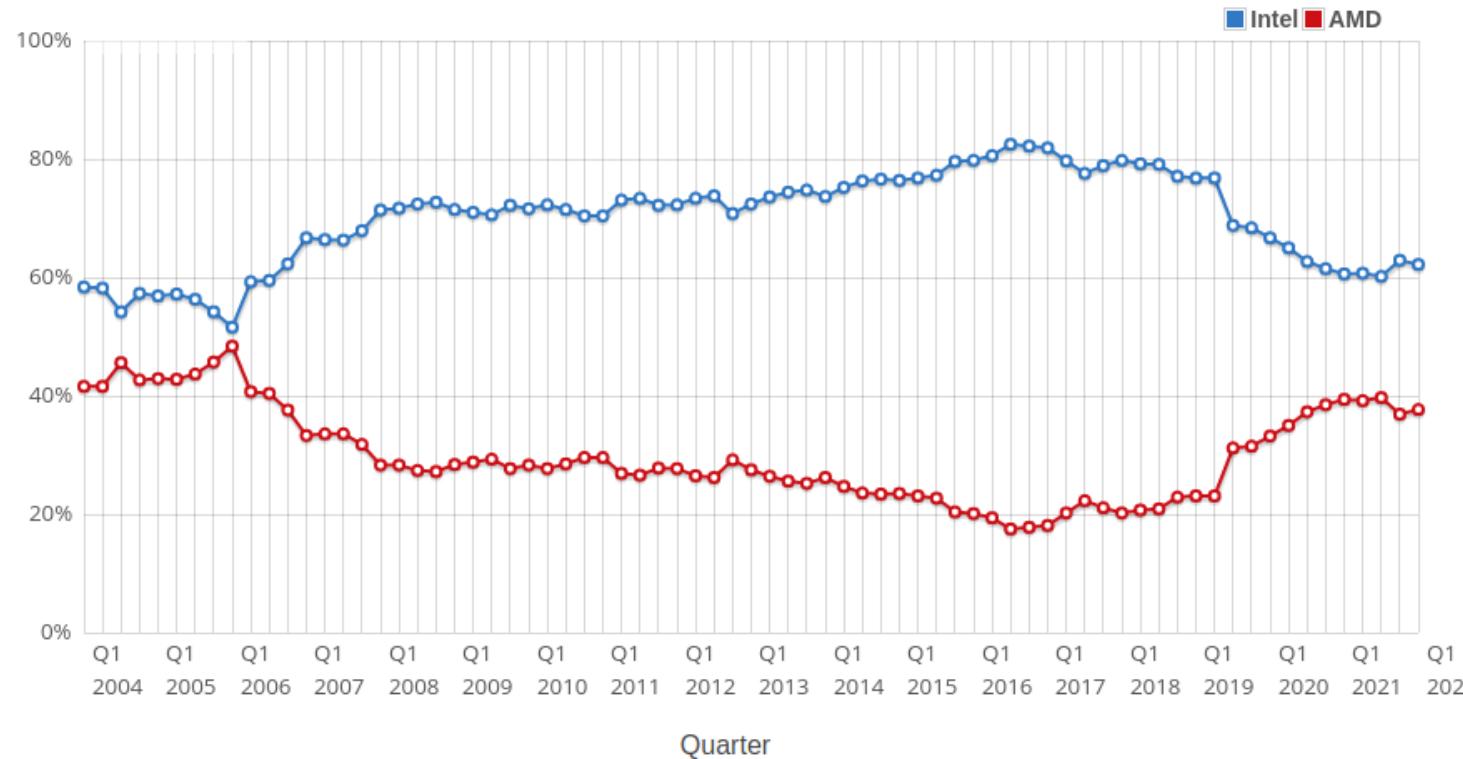
### Mémoire

- Modèle mémoire uniforme (UMA)
- Cache processeur
  - Technologies de transfert rapides
  - Copies d'informations depuis la mémoire principale (DATA ou INST.)
  - Intelligence déportée dans les contrôleurs de caches (LRU)
  - Non déterministe

## Parts de marché : Intel vs. AMD

## AMD vs Intel Market Share (All CPUs)

Last updated on the 4th of January 2022



- AP -  
**APPLICATION PROCESSOR**

Applications

Architecture

Solution Qualcomm

Solution ARM



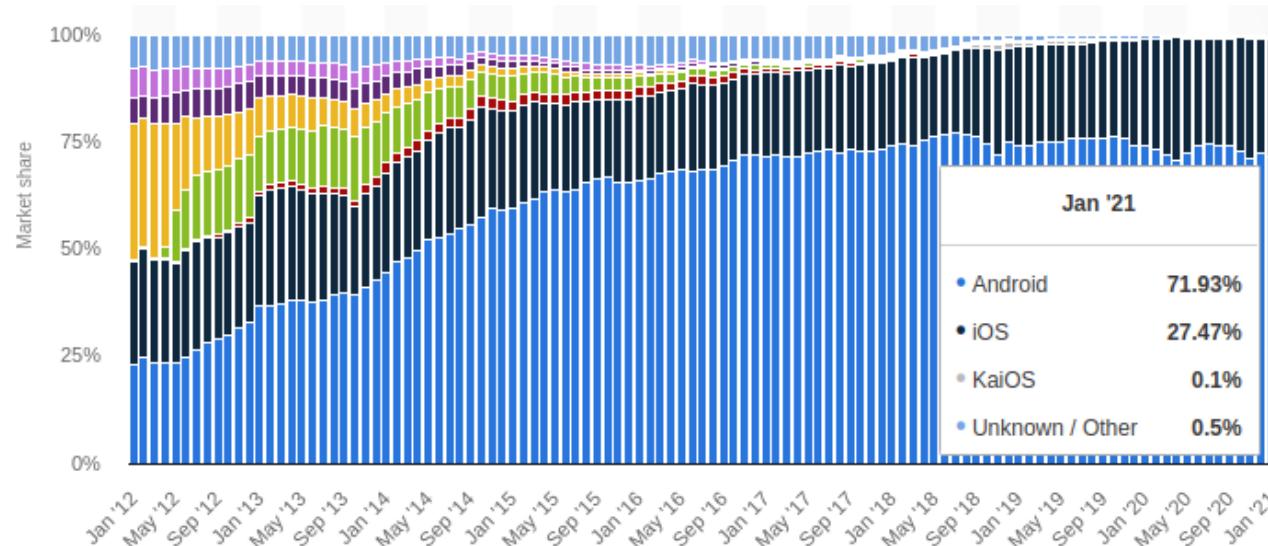
Le marché des AP (*Application Processor*), processeurs riches en fonctionnalités et services matériels de type SoC (*System on Chip*), reste un marché récent qui a vu son envole avec celui des terminaux mobiles (smartphone, phablette et tablette).



## Applications

Le principal marché des AP en terme de parts reste donc celui des terminaux mobiles.

Ce marché voit une utilisation écrasante du système d'exploitation Android en 2016, système basé sur un noyau Linux.



Source:  
Statista 2021

- Android ● iOS ● KaiOS ● Windows Phone ● Series 40 (Nokia)\* ● Symbian OS\* ● Samsung
- BlackBerry OS ● Unknown / Other

Néanmoins les processeurs applications sont très rencontrés dans les systèmes embarqués au sens large, tous domaines confondus : *consumer*, défense, transport ...

Ces systèmes embarquent généralement un OS et une interface graphique.



Freebox Revolution



Télévision 4K X94C Sony



Tablette Cook  
(fait à Caen par EOLANE)

Dans la majorité des cas, ces processeurs sont exploités par des systèmes évolués.

Sur ce marché les systèmes GNU/Linux (très souvent customisés) règnent en maîtres.



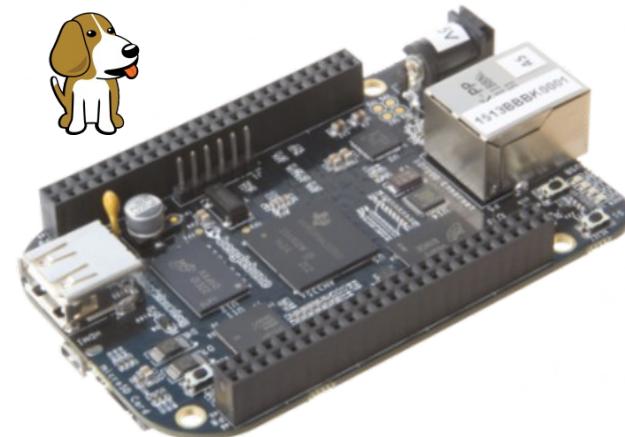
Exemple de plateforme industrielle durcie EOLANE (Français n°2 Européen) travaillant autour de SoC/AP iMX6 proposé par Freescale sur système GNU/Linux.



Voici les deux plateformes non-durcies à bas coût qui dominent le marché :  
les projets **Raspberry Pi** et **Beaglebone** (SoC AM335x TI).

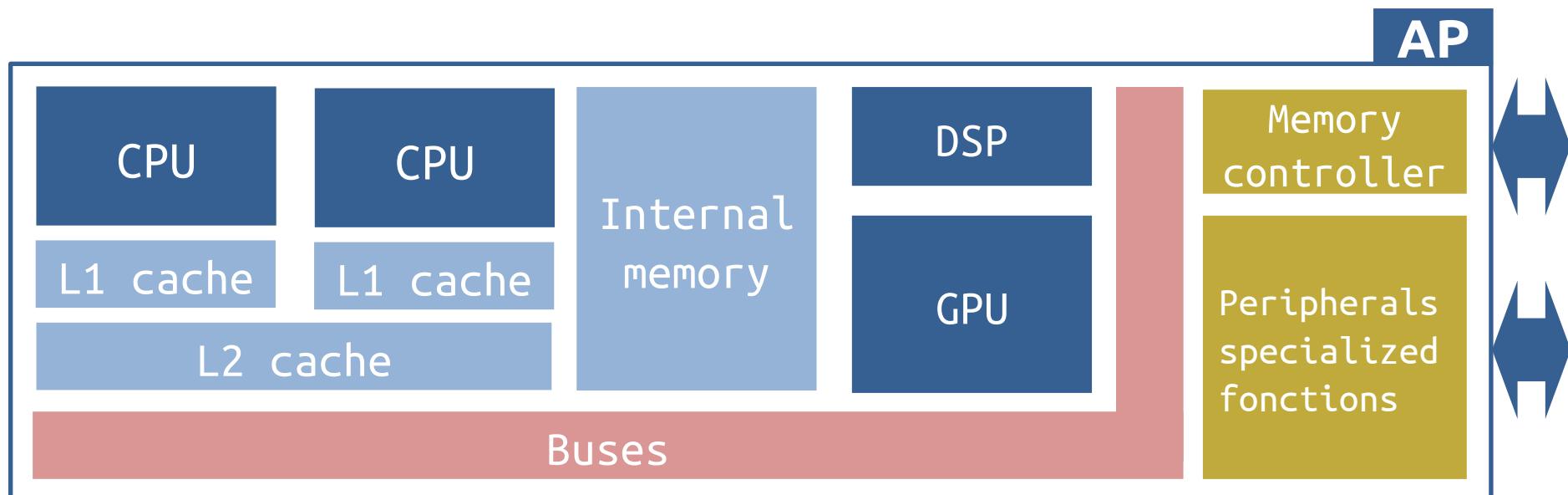
Ces solutions sont également basées sur des systèmes GNU/Linux

Elles sont très rencontrées durant les phases de prototypage ou en milieu universitaire,  
mais ne peuvent être industrialisées. Néanmoins des versions durcies existent.



Les AP sont des systèmes numériques complets intégrés dans une puce (architecture hétérogène).

Néanmoins, la mémoire principale doit être ajoutée en externe.

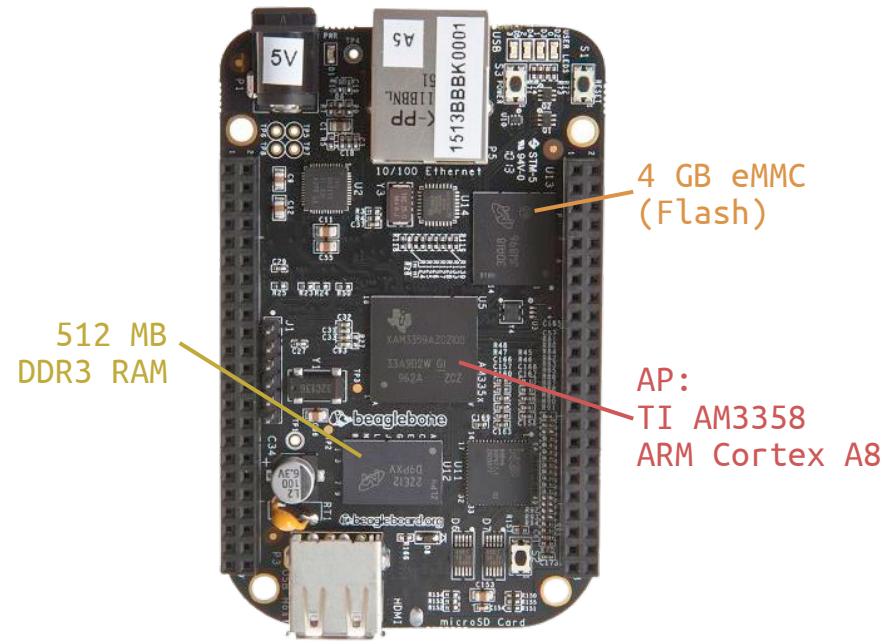


## Architecture

Un **processeur application** embarque toujours un voire plusieurs CPU généralistes superscalaires. Ils sont dédiés à l'exécution du ou des systèmes d'exploitation évolués (virtualisés ou réels) ainsi que des applicatifs.

Un **AP** contient également une voire plusieurs fonctions spécialisées de calcul (GPU, DSP, crypto ...), un jeu de périphériques évolués complet et une mémoire interne ne permettant pas d'accueillir le système (*bootloader*).

Par conséquent, une **mémoire principale** (DDR volatile) et une mémoire non-volatile de **stockage de masse** (MMC, eMMC, SDCard ...) externes doivent lui être ajoutées.

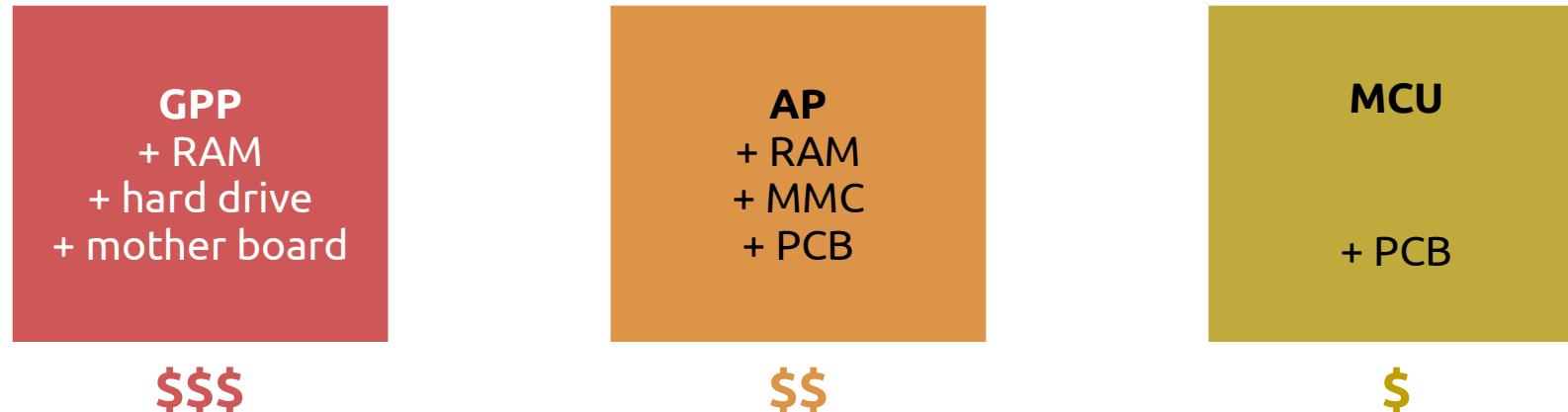


<https://beagleboard.org/black>

## Comparaison des processeurs généralistes

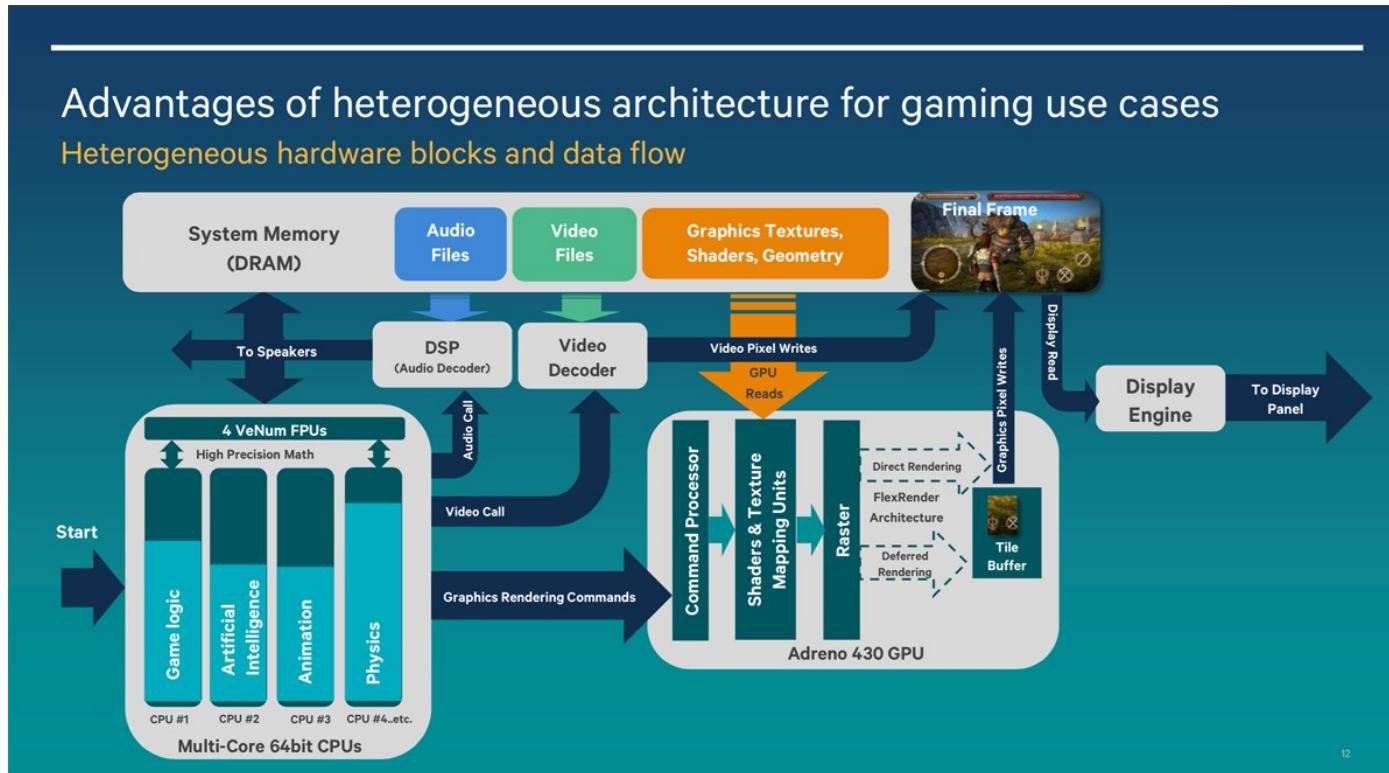
Contrairement aux MCU embarquant tous les services matériels sur la puce afin de contrôler un système (*on chip*), les AP exigent un coût unitaire non négligeable et restent dépréciés pour les applications à faible coût et fort volume.

Ils sont alors utilisés si il y a nécessité d'une interface et/ou de connectivités évoluées dans l'application.



## Architecture

Observons l'intérêt d'une architecture hétérogène pour une application aux jeux vidéos



## Solution Qualcomm Snapdragon

Le leader du marché en terme de part de marché est Qualcomm, grâce à sa famille Snapdragon dédiée au marché des terminaux mobiles.



## Solution Qualcomm Snapdragon

## Fonctions matérielles de l'architecture interne de la famille Qualcomm Snapdragon 810

## Introducing the Snapdragon 810 Processor

Advanced Graphics & Compute with the Adreno 430 – the best GPU Qualcomm Technologies' has ever made

4K primary & external display support with ecoPix and TruPalette and 3:1 pixel compression

Mobile industry's FIRST announced multi-channel 4G LTE SoC supporting Category 9 Carrier Aggregation

**Location**  
GPS, GLONASS, Beidou, Galileo Satellites

**Cortex-A57 & Cortex-A53 CPUs**

**Memory**  
LPDDR4 UFS2.0

**FIRST Announced ARM®v8-A/64-bit using Cortex®-A57+ Cortex®-A53**

**Mobile industry's FIRST announced dual channel 1600 MHz LPDDR4 memory**  
**Qualcomm Technologies' FIRST UFS 2.0 Support**

**Display Engine**  
4K, Miracast, picture enhancement

**Qualcomm® Hexagon™ DSP**  
Ultra Low Power Sensor Engine

**Greatly improved power management for DSP/Sensor Engine, Low Power Snapdragon Voice Activation (SVA), 12-channel surround sound decode**

**Modem**  
4<sup>th</sup> gen CAT 9 LTE Up to 3x20MHz CA

**USB 3.0**

**Dual ISPs**  
14bit Depth Assist Local Tone Mapping Hybrid Auto Focus with Phase Detect

**Multimedia Processing**  
4K Encode/Decode Snapdragon Voice Activation Gestures Studio Access Security

**Qualcomm Technologies' FIRST hardware implementation of 4K HEVC/H.265 video encode. HEVC designed to deliver up to 50% better video compression**

Not drawn to scale.

Qualcomm Adreno and Qualcomm Hexagon are pro

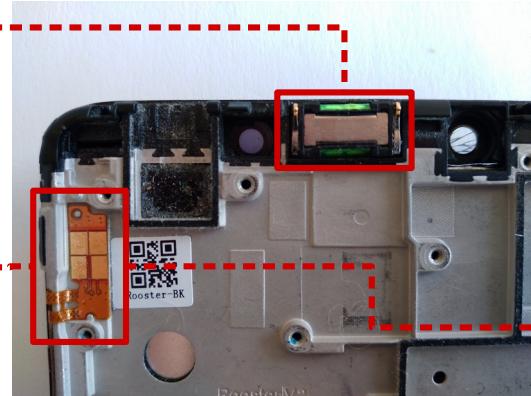
Qualcomm Technologies' FIRST 14-bit Dual ISP for highest quality, depth enabled photography. Up to 21MP for main camera with depth assist, phase detect, for sharper dual camera user experiences

# AP – APPLICATION PROCESSOR

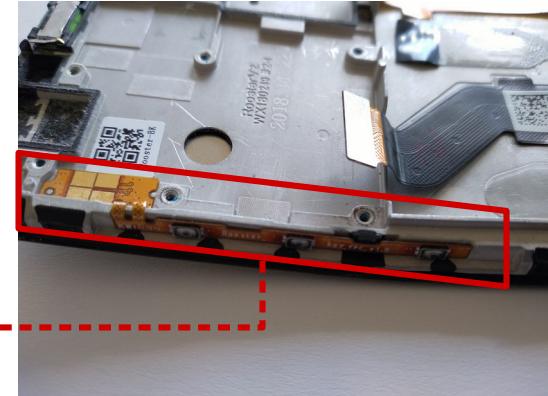
Exemple smartphone (Nokia 3.1 Plus, 2018)



Speaker



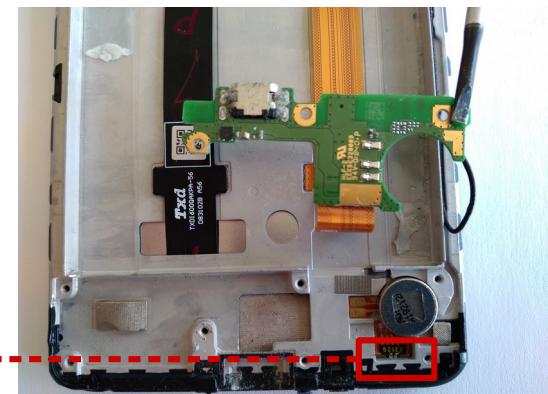
Switches connector



Battery  
3400 mAh



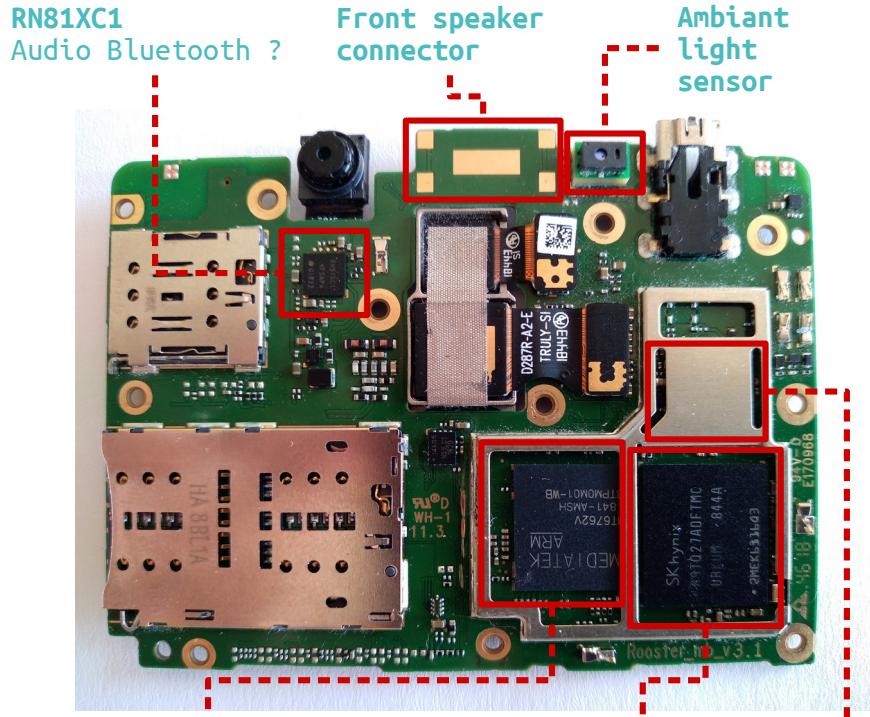
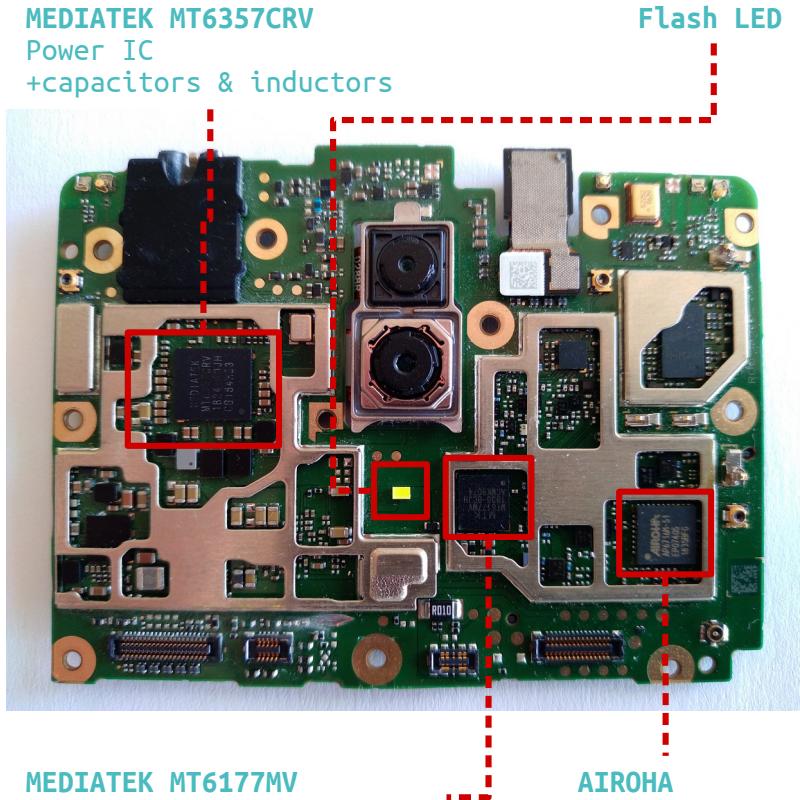
Loudspeaker



Microphone

# AP – APPLICATION PROCESSOR

Exemple smartphone (Nokia 3.1 Plus, 2018)



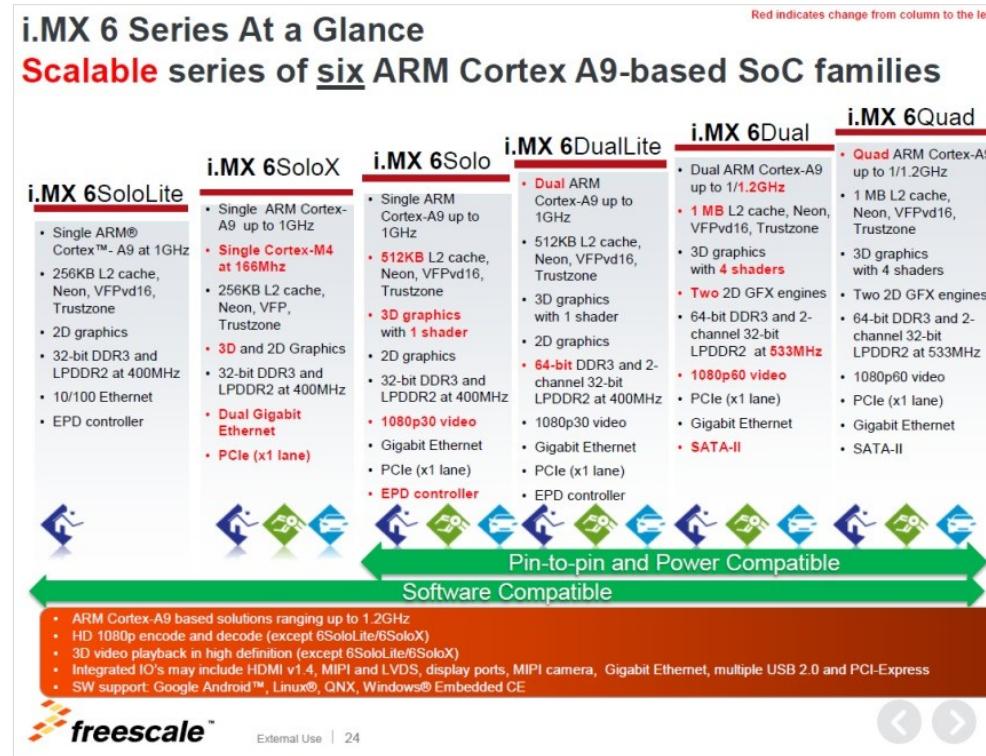
**MEDIATEK MT6762V (Helio P22)**  
Qualcomm SDM439 Snapdragon (2018)  
ARM v8-A (64-bit), Cortex-A53  
8 cores, 2 GHz, 12 nm  
2-core GPU, DSP

**SK hynix**  
H9TQ27ADFTMC  
32 GB Flash  
Nand eMMC

**RAM ?**  
2 GB  
LPDDR3  
(Low-Power)

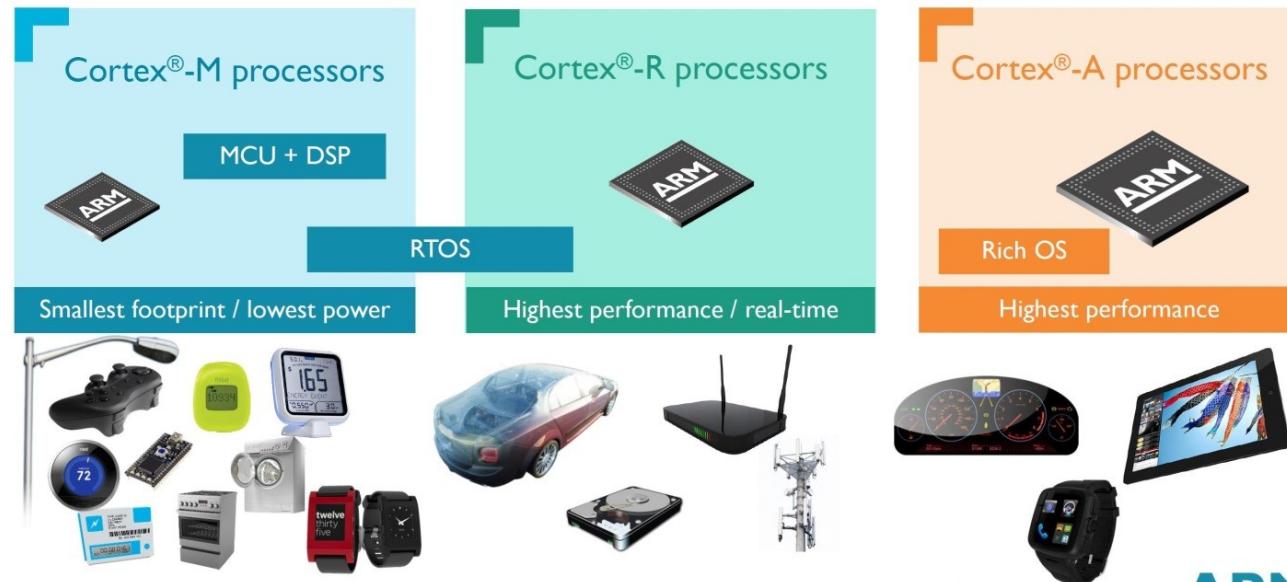
Les deux leaders du marché hors terminaux mobiles sont Texas Instruments et Freescale, deux fondeurs offrant de larges communautés d'utilisateurs.

Observons la famille i.MX6 de Freescale :



Hors marché des terminaux mobiles, sur le marché de l'embarqué les architectures Cortex-A de ARM sont également reines. Le « A » signifie *Application*.

### ARM® Cortex® Processors across the Embedded Market



# - GPU -

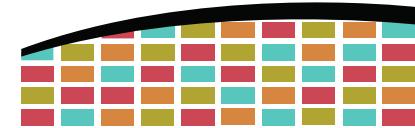
# GRAPHICS PROCESSING UNIT

Applications

Architecture

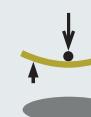
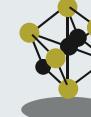
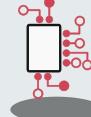
Produits Nvidia

Marchés



# ENSI CAEN

ÉCOLE PUBLIQUE D'INGÉNIEURS  
CENTRE DE RECHERCHE



## Applications

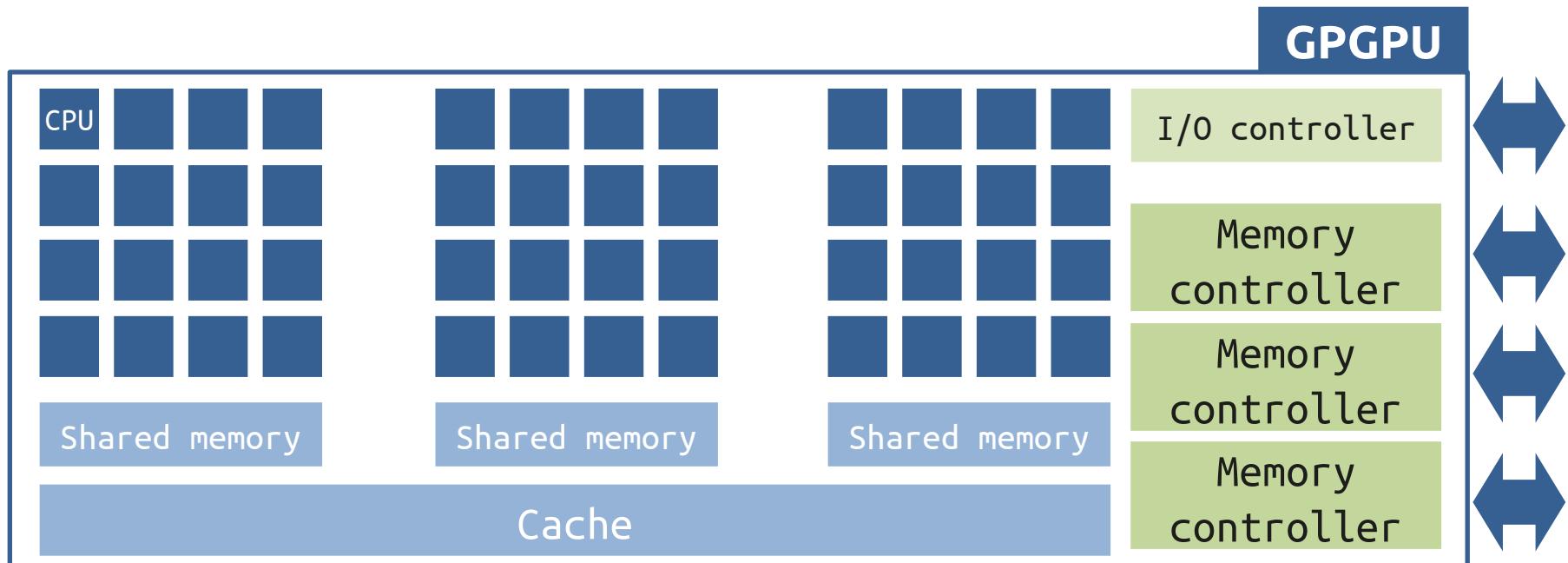
Les **GPU (*Graphics Processing Unit*)** sont des **coprocesseurs** de traitement spécialisés pour le calcul intensif.

Depuis quelques années, nous parlons de GPGPU (*General Purpose GPU*), GPU dédié au calcul massif au sens large. Les applications sont multiples : finance, recherche et sciences, imagerie médicale, jeux vidéos ...



## Architecture

Les GPU possèdent un modèle mémoire réparti non uniforme de type NUMA (*Non Uniform Memory Access*), permettant un clonage des données à traiter et un parallélisme d'exécution. Ils intègrent une architecture massivement parallèle.



## Solution Nvidia : la carte Tesla P100

Observons le potentiel de la carte Tesla P100 proposée par Nvidia courant 2016 et dédiée aux data center les plus avancés du moment. Basée sur GPU GP100.



#### SPECIFICATIONS

GPU Architecture	NVIDIA Pascal
NVIDIA CUDA® Cores	3584
Double-Precision Performance	5.3 TeraFLOPS
Single-Precision Performance	10.6 TeraFLOPS
Half-Precision Performance	21.2 TeraFLOPS
GPU Memory	16 GB CoWoS HBM2
Memory Bandwidth	732 GB/s
Interconnect	NVIDIA NVLink
Max Power Consumption	300 W
ECC	Native support with no capacity or performance overhead
Thermal Solution	Passive
Form Factor	SXM2
Compute APIs	NVIDIA CUDA, DirectCompute, OpenCL™, OpenACC

## Solution Nvidia : architecture Pascal



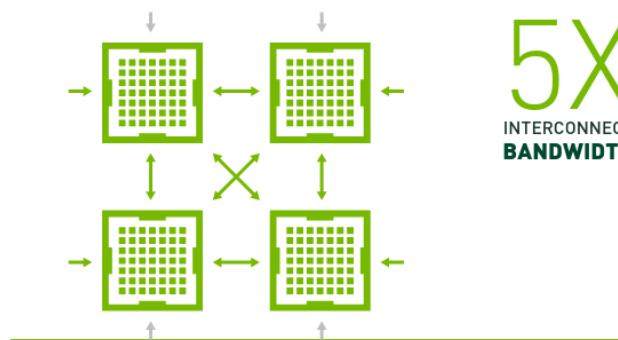
LEAP IN NEURAL NETWORK TRAINING PERFORMANCE  
WITH NEW **NVIDIA PASCAL ARCHITECTURE**



FABRICATED WITH **16 NANOMETER FINFET**  
FOR UNPRECEDENTED ENERGY EFFICIENCY



WITH **CoWoS® WITH HBM2** COMPARED TO  
NVIDIA MAXWELL™ ARCHITECTURE FOR  
BIG DATA WORKLOADS



WITH **NVIDIA NVLINK™** FOR MAXIMUM  
APPLICATION SCALABILITY

# GPU – GRAPHICS PROCESSING UNITS

## Solution Nvidia : Architecture du GPU GP100



## Le GPU Nvidia GP100 en chiffres

- 6 Graphics Processing Clusters
- 30 Texture Processing Clusters (5 / GPC)
- 60 Streaming Multiprocessors (2 / TPC)
- 3840 single precision cores (64 / SM)
- 1920 double precision unit (32 / SM)
- 240 texture units (4 / SM)
- 8 memory controllers
  - $8 \times 512 \text{ KB} = 4096 \text{ KB}$  L2 cache
  - 4 pairs that control HBM2 DRAM

Tesla Products	Tesla K40	Tesla M40	Tesla P100
<b>GPU</b>	GK110 (Kepler)	GM200 (Maxwell)	GP100 (Pascal)
<b>SMs</b>	15	24	56
<b>TPCs</b>	15	24	28
<b>FP32 CUDA Cores / SM</b>	192	128	64
<b>FP32 CUDA Cores / GPU</b>	2880	3072	3584
<b>FP64 CUDA Cores / SM</b>	64	4	32
<b>FP64 CUDA Cores / GPU</b>	960	96	1792
<b>Base Clock</b>	745 MHz	948 MHz	1328 MHz
<b>GPU Boost Clock</b>	810/875 MHz	1114 MHz	1480 MHz
<b>Peak FP32 GFLOPs<sup>1</sup></b>	5040	6840	10600
<b>Peak FP64 GFLOPs<sup>1</sup></b>	1680	210	5300
<b>Texture Units</b>	240	192	224
<b>Memory Interface</b>	384-bit GDDR5	384-bit GDDR5	4096-bit HBM2
<b>Memory Size</b>	Up to 12 GB	Up to 24 GB	16 GB
<b>L2 Cache Size</b>	1536 KB	3072 KB	4096 KB
<b>Register File Size / SM</b>	256 KB	256 KB	256 KB
<b>Register File Size / GPU</b>	3840 KB	6144 KB	14336 KB
<b>TDP</b>	235 Watts	250 Watts	300 Watts
<b>Transistors</b>	7.1 billion	8 billion	15.3 billion
<b>GPU Die Size</b>	551 mm <sup>2</sup>	601 mm <sup>2</sup>	610 mm <sup>2</sup>
<b>Manufacturing Process</b>	28-nm	28-nm	16-nm FinFET

<sup>1</sup> The GFLOPS in this chart are based on GPU Boost Clocks.

Note : la carte Tesla P100 exploite 56 SM sur les 60 SM disponibles dans le GP100.

## Solution Nvidia : Architecture du GPU GP100

Les GPU intègrent un grand nombre de CPU à pipeline classique mais avec des EU vectorielles SIMD.

EU = Execution Unit

SIMD = Single Instruction Multiple Data

GPC = Graphics Processing Cluster

TCP = Texture Processing Cluster

SM = Streaming Multiprocessor  
(multithreaded processor)

Warp = thread of SIMD instructions

DP = Double Precision

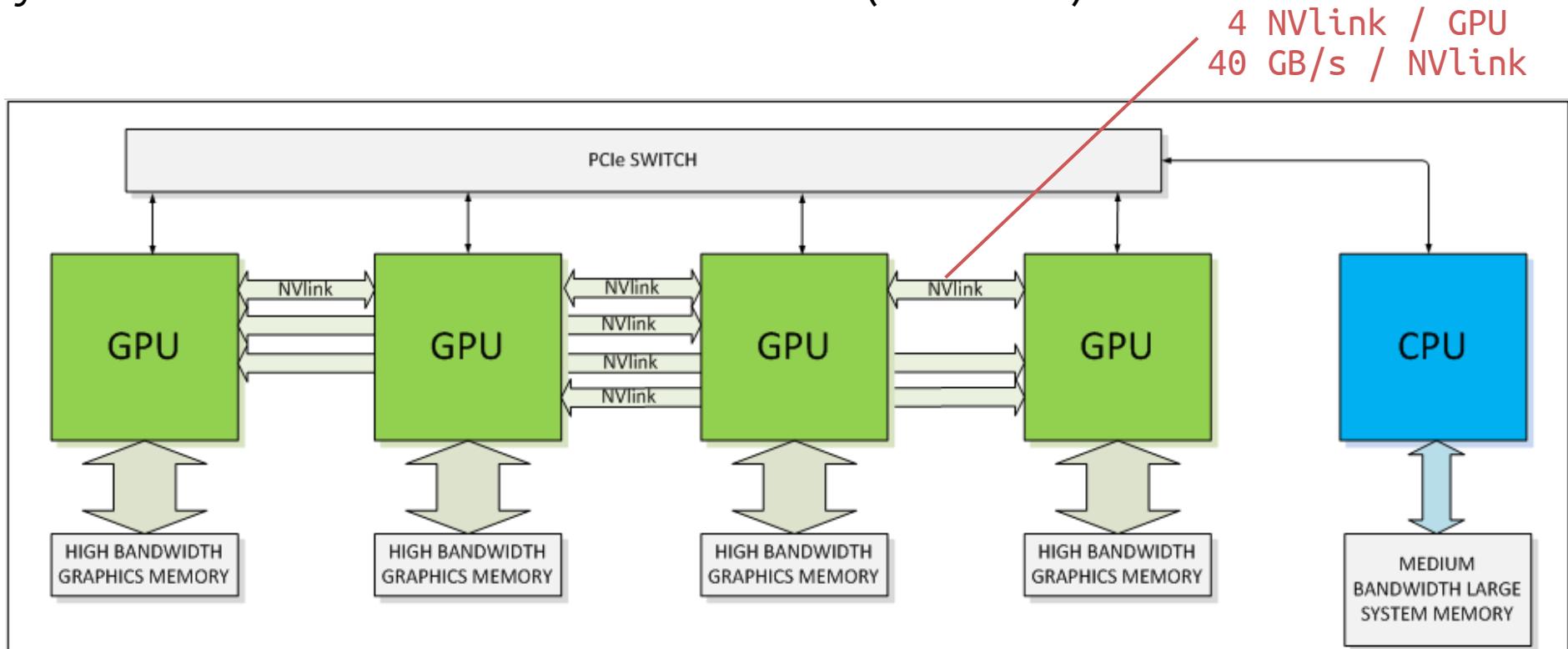
LD/ST = Load/Store

SFU = Special Function Unit

Tex = Texture

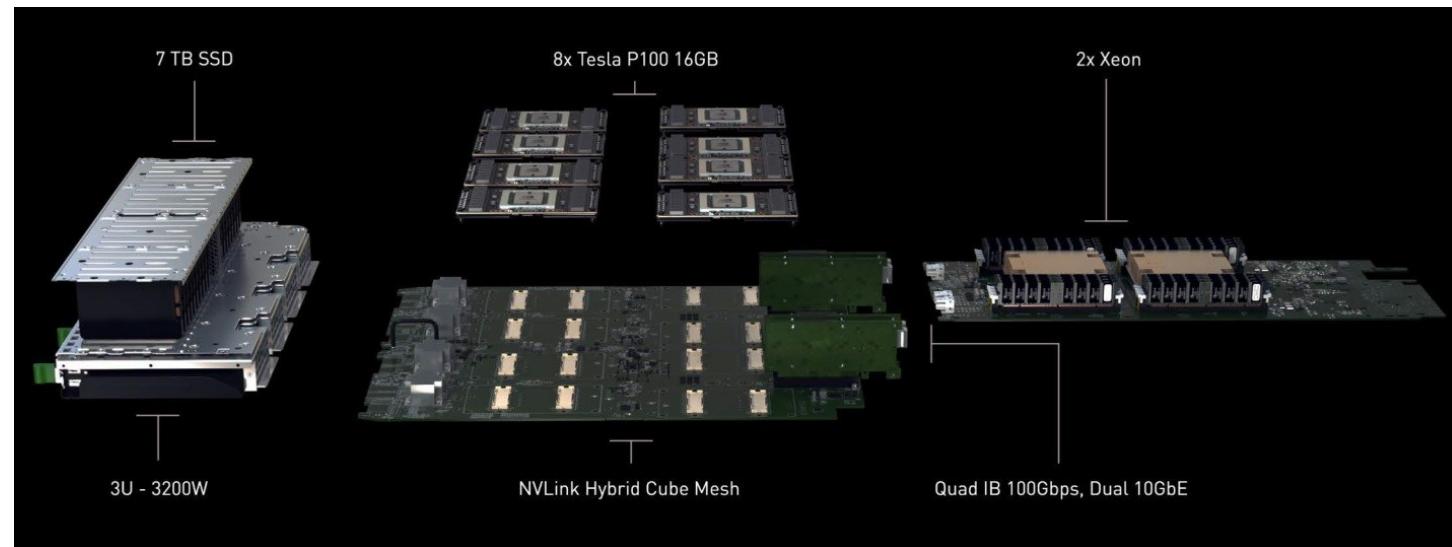
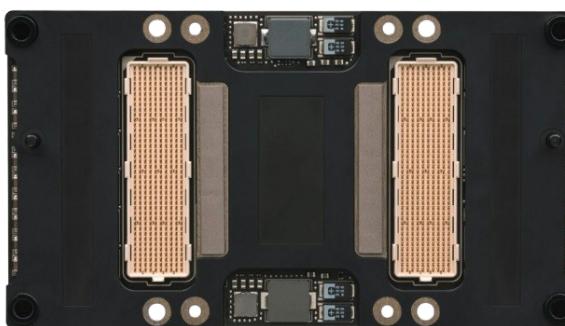
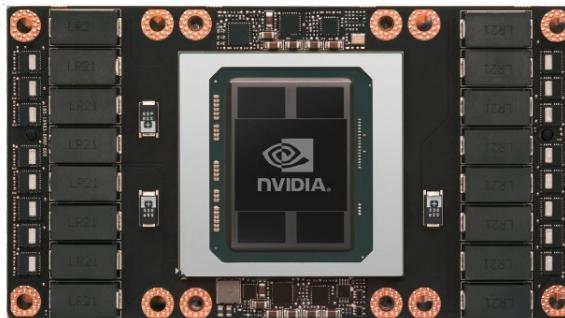


## Système d'interconnexion et de communication (Tesla P100)



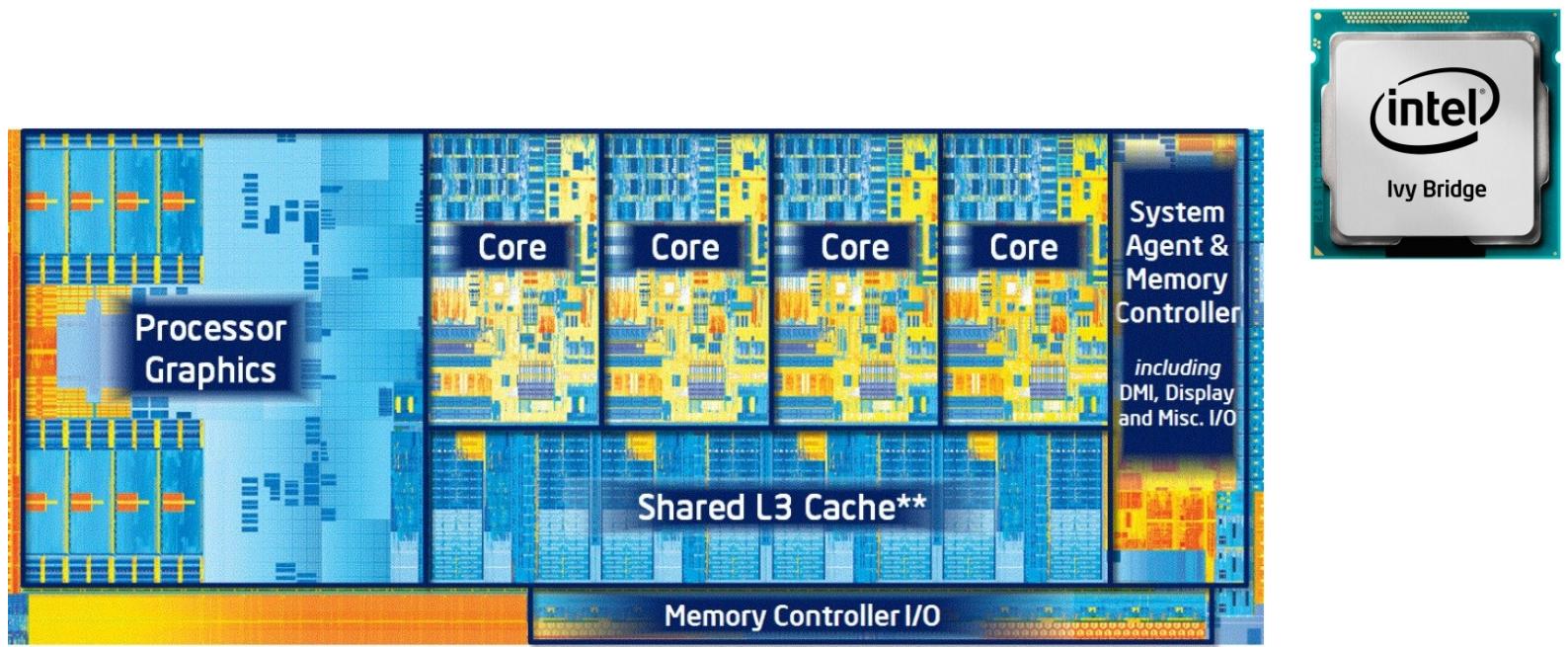
## Solution Nvidia : exemple d'application

Exemple d'application utilisant la carte Nvidia Tesla P100.



## Marchés

Le leader incontesté du marché des GPU/IGP en terme de part est Intel grâce aux coprocesseurs graphiques IGP (*Integrated Graphics Unit*) intégrés dans une grande partie de leurs gammes processeurs GPP (plus de 70% en 2016).



Toutefois, le leader des solutions hautes performances externes est l'américain Nvidia.



Tesla K20C



Tesla P100



**NVIDIA.**®

# - DSP -

# DIGITAL SIGNAL PROCESSOR

Applications

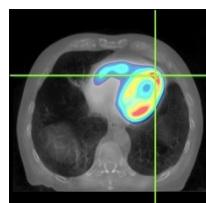
Architecture

Texas Instruments

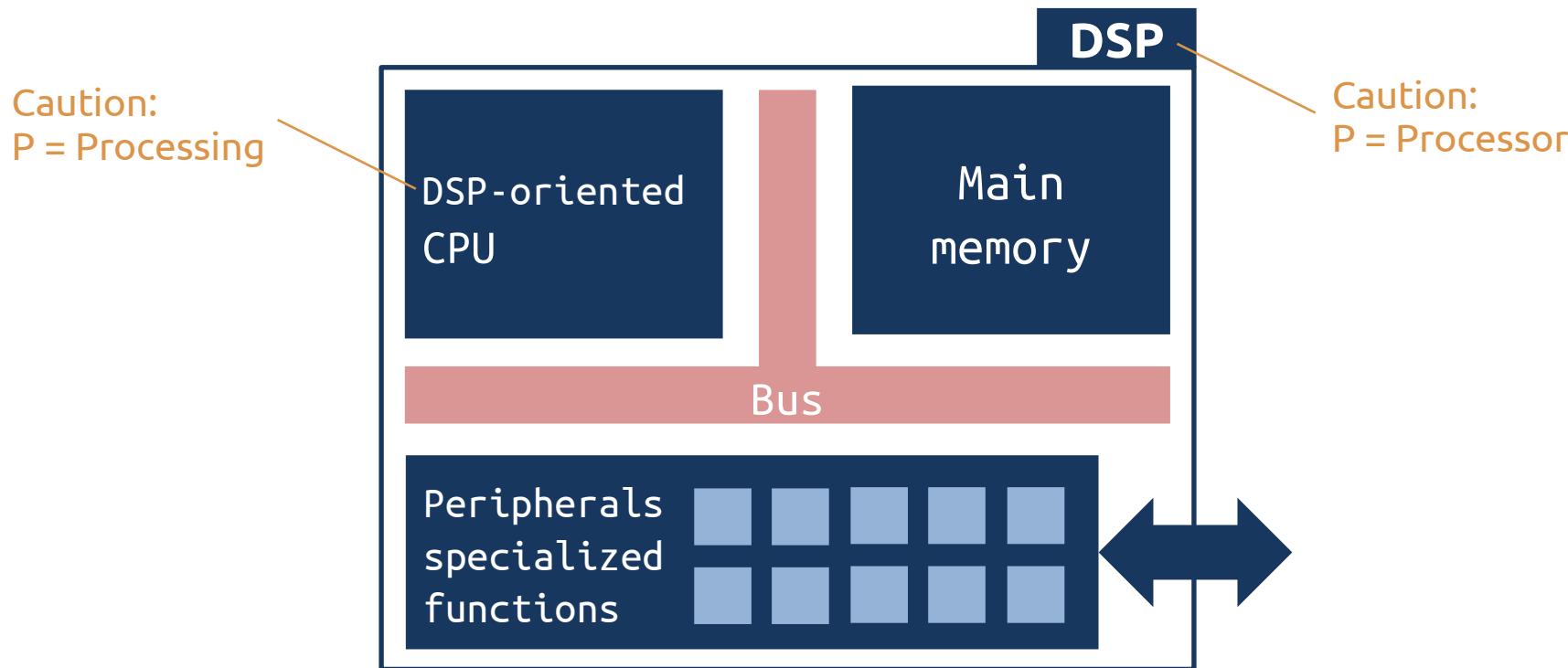


## Applications

Les **DSP (Digital Signal Processor)** sont dédiés aux applications impliquant du Traitement Numérique du Signal (TNS ou DSP ou *Digital Signal Processing*).



Les DSP sont très proches des MCU : ce sont des systèmes autonomes.  
Leur CPU est néanmoins spécialisé pour le calcul numérique.



## Architecture

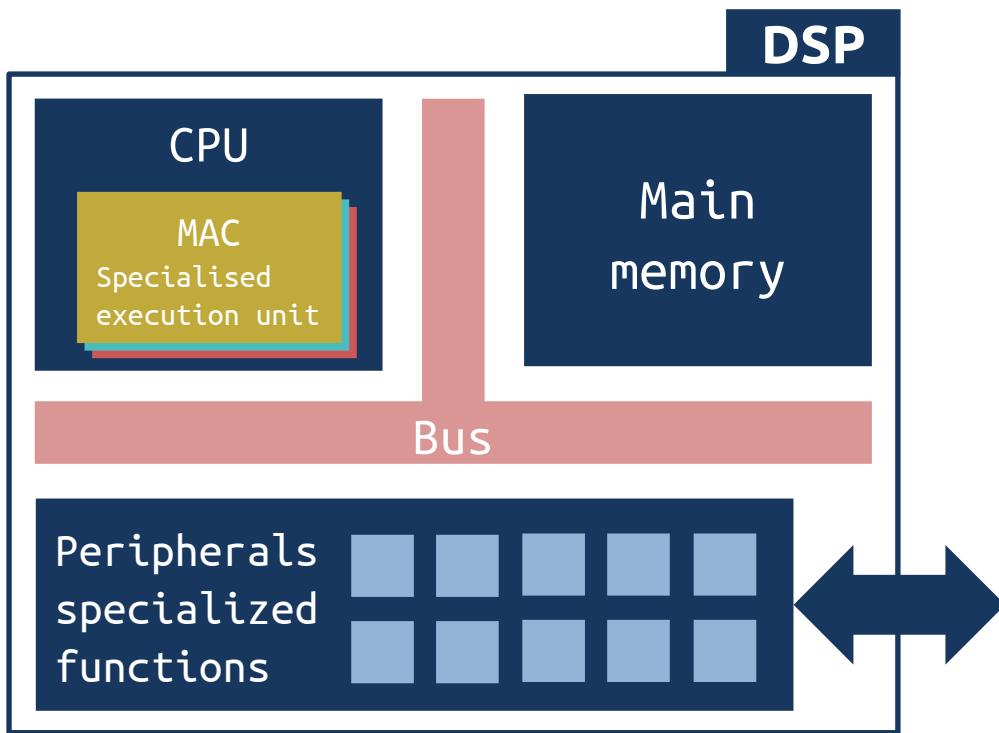
Leur CPU possède des extensions d'instructions et unités d'exécution dédiées au calcul de **MAC (Multiply Accumulate)** ou **SOP (Som Of Products)**. Il s'agit des opérations élémentaires rencontrées dans tout algorithme de Traitement Numérique du Signal.

Expansion of the Danielson-Lanczos Lemma to 8 terms:

$$\begin{aligned}
 F(n) = & \sum_{k=0}^{N/8-1} x(8k)e^{\frac{-j2\pi kn}{8}} + W_N^n \sum_{k=0}^{N/8-1} x(8k+4)e^{\frac{-j2\pi kn}{8}} + \\
 & W_{\frac{N}{2}}^n \sum_{k=0}^{N/8-1} x(8k+2)e^{\frac{-j2\pi kn}{8}} + W_{\frac{N}{2}}^n W_{\frac{N}{4}}^n \sum_{k=0}^{N/8-1} x(8k+6)e^{\frac{-j2\pi kn}{8}} + \\
 & W_N^n \sum_{k=0}^{N/8-1} x(8k+1)e^{\frac{-j2\pi kn}{8}} + W_N^n W_{\frac{N}{4}}^n \sum_{k=0}^{N/8-1} x(8k+5)e^{\frac{-j2\pi kn}{8}} + \\
 & W_N^n W_{\frac{N}{2}}^n \sum_{k=0}^{N/8-1} x(8k+3)e^{\frac{-j2\pi kn}{8}} + W_N^n W_{\frac{N}{2}}^n W_{\frac{N}{4}}^n \sum_{k=0}^{N/8-1} x(8k+7)e^{\frac{-j2\pi kn}{8}}
 \end{aligned}$$

## Architecture

CPU avec unités d'exécution dédiées au calcul de **MAC** ou **SOP**. Le jeu d'instructions (ISA) dispose d'instructions spécifiques pour exploiter ces EU.



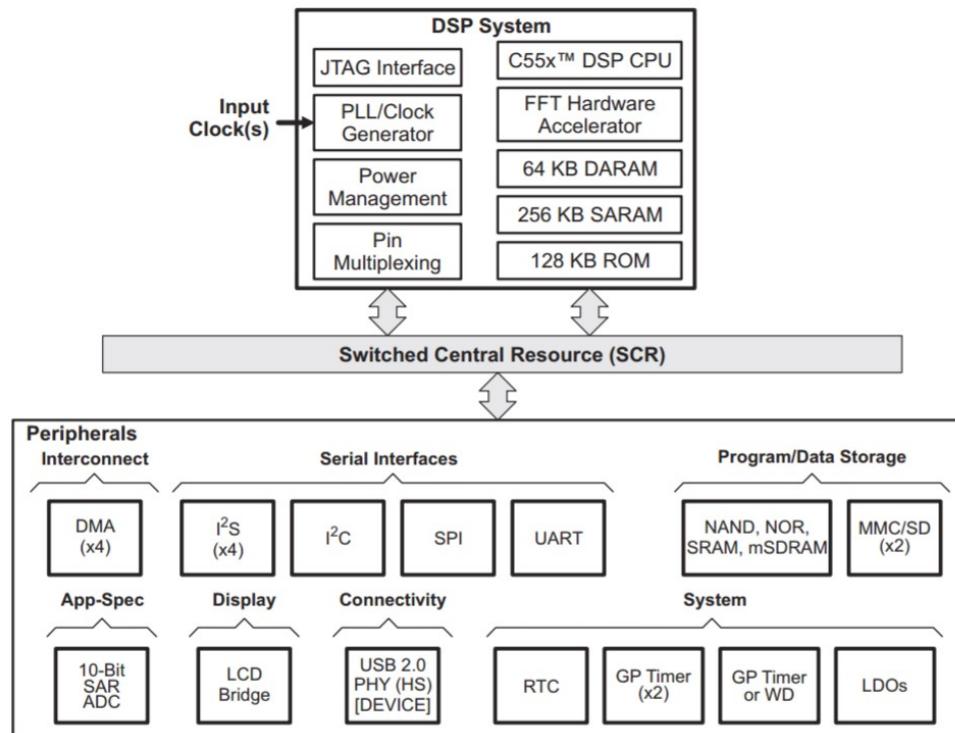
MAC = SOP

MAC : Multiply-Accumulate  
SOP : Som of Products

ISA : Instruction Set Architecture  
EU : Execution Unit

## Texas Instruments : C5500

Observons la solution C5500 DSP proposée par Texas Instruments, l'une des solutions phares du fondeur américain.



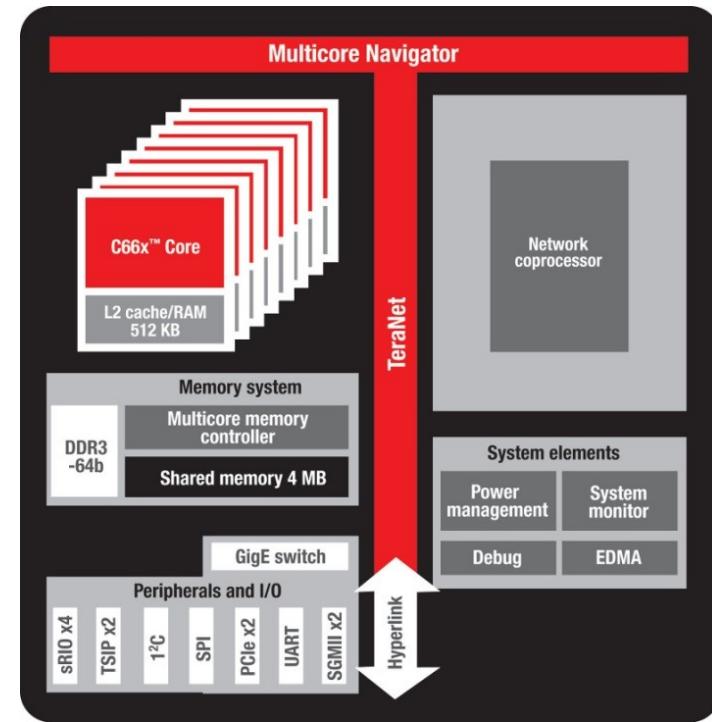
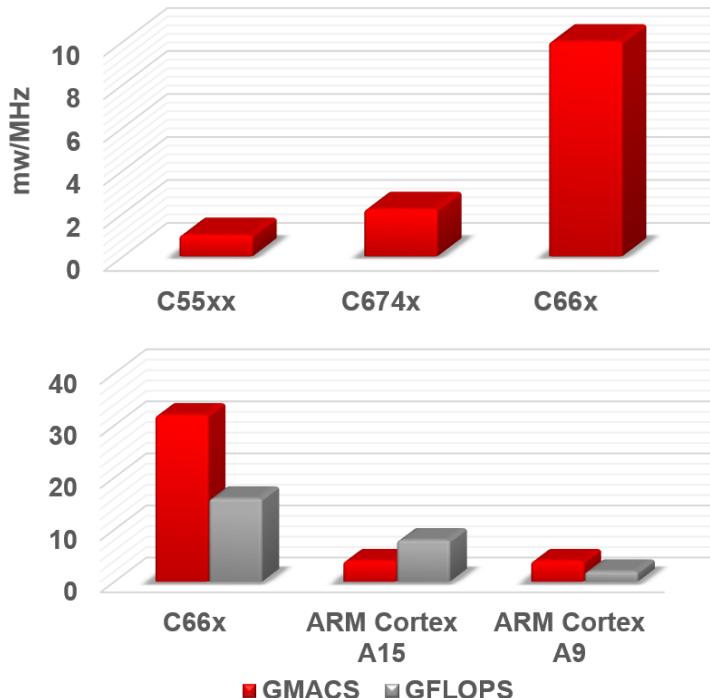
Étudions cet extrait de sa datasheet,  
le résumé des spécifications du CPU.

### 1.1 Features

- CORE:
  - High-Performance, Low-Power, TMS320C55x Fixed-Point Digital Signal Processor
    - 20-, 10-ns Instruction Cycle Time
    - 50-, 100-MHz Clock Rate
    - One or Two Instructions Executed per Cycle
    - Dual Multiply-and-Accumulate Units (Up to 200 Million Multiply-Accumulates per Second [MMACS])
    - Two Arithmetic and Logic Units (ALUs)
    - Three Internal Data and Operand Read Buses and Two Internal Data and Operand Write Buses
    - Software-Compatible with C55x Devices
    - Industrial Temperature Devices Available
  - 320KB of Zero-Wait State On-Chip RAM, Composed of:
    - 64KB of Dual-Access RAM (DARAM), 8 Blocks of 4K x 16-Bit
    - 256KB of Single-Access RAM (SARAM), 32 Blocks of 4K x 16-Bit
  - 128KB of Zero Wait-State On-Chip ROM (4 Blocks of 16K x 16-Bit)
  - Tightly Coupled FFT Hardware Accelerator

## Exemple Texas Instruments : C6600

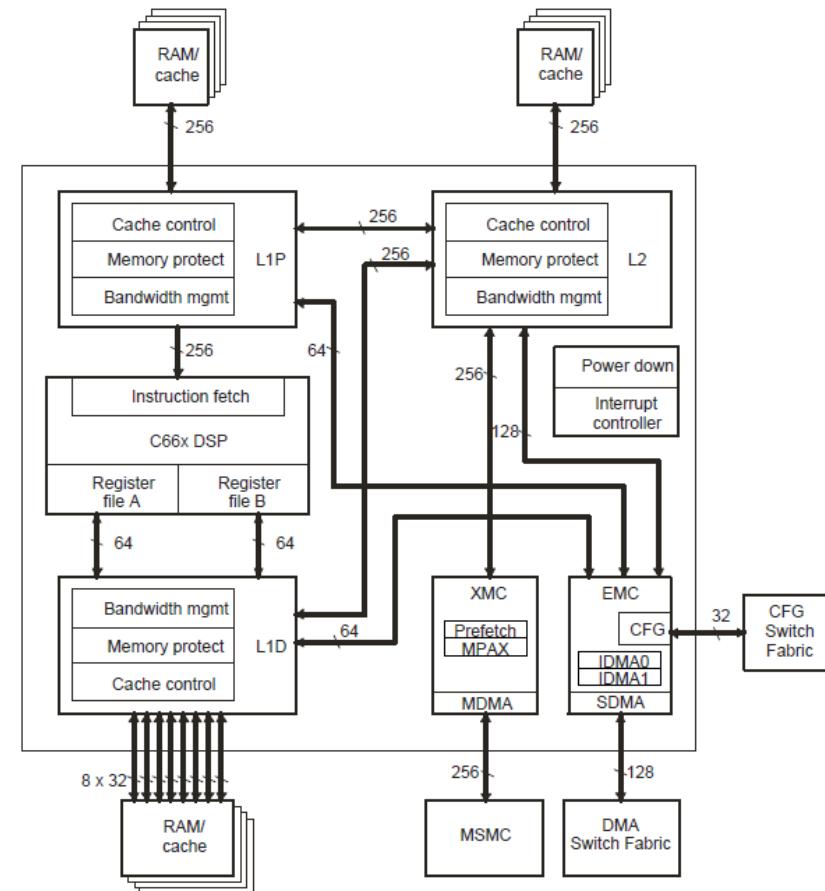
Passons maintenant à la gamme Keystone C6600 proposée par Texas Instruments. Cette architecture DSP est l'une des plus performantes du marché.



Cœur (CorePac) C6600 développé par TI.

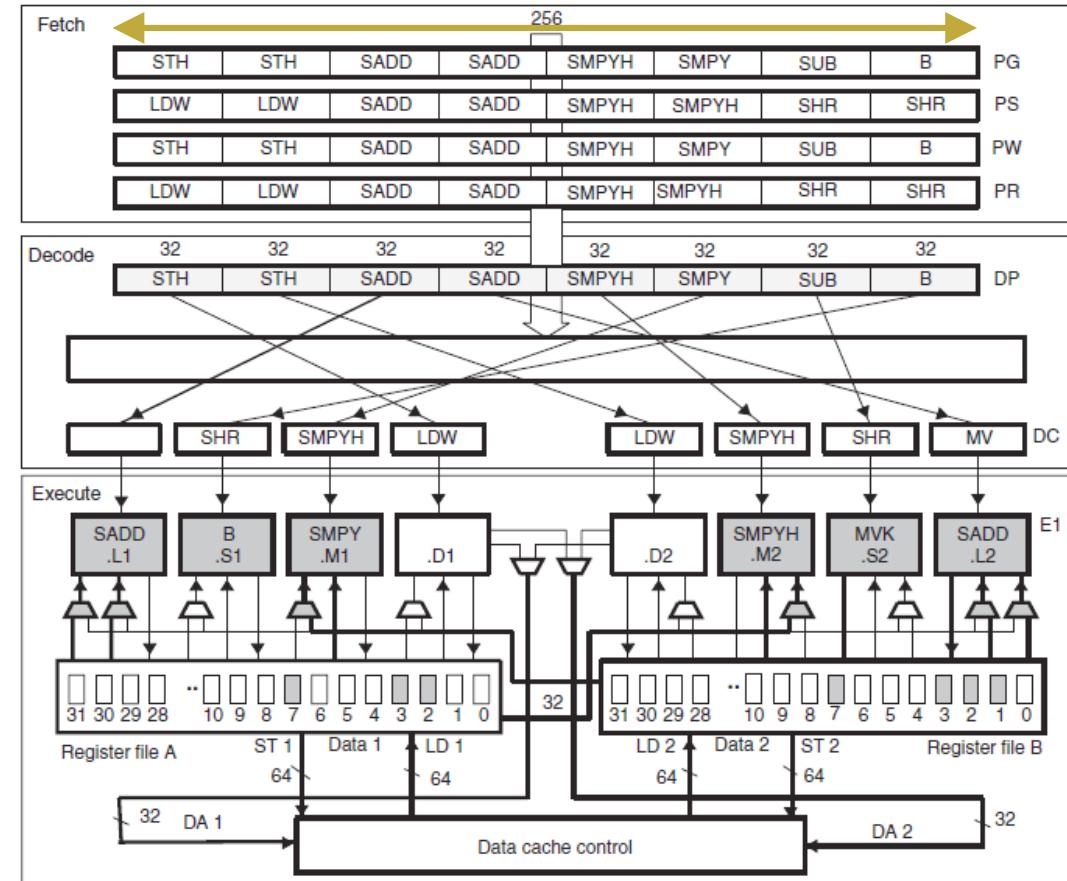
Hiérarchie mémoire configurable en cache ou SRAM adressable sans perte de bande passante.

Modèle UMA ou NUMA configurable pour chaque cœur.



Cœur C6600 avec :

- pipeline matériel VLIW (*Very Long Instruction Word*) à 14 étages
- pipeline logiciel d'une largeur maximale de 8 instructions.



Ces DSP sont pensés aussi bien pour un usage en parallèle que pour être chaînés afin d'encaisser des chaînes des traitements profondes (reliés en *daisy-chain*).



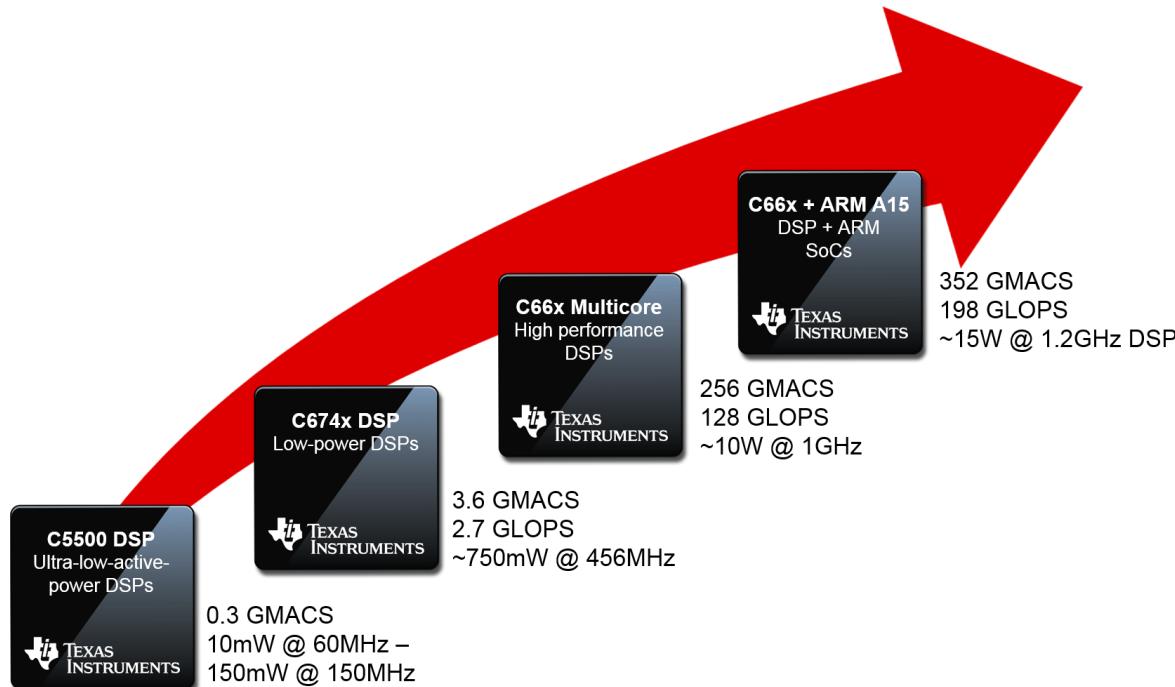
Avantage du *daisy-chain* :



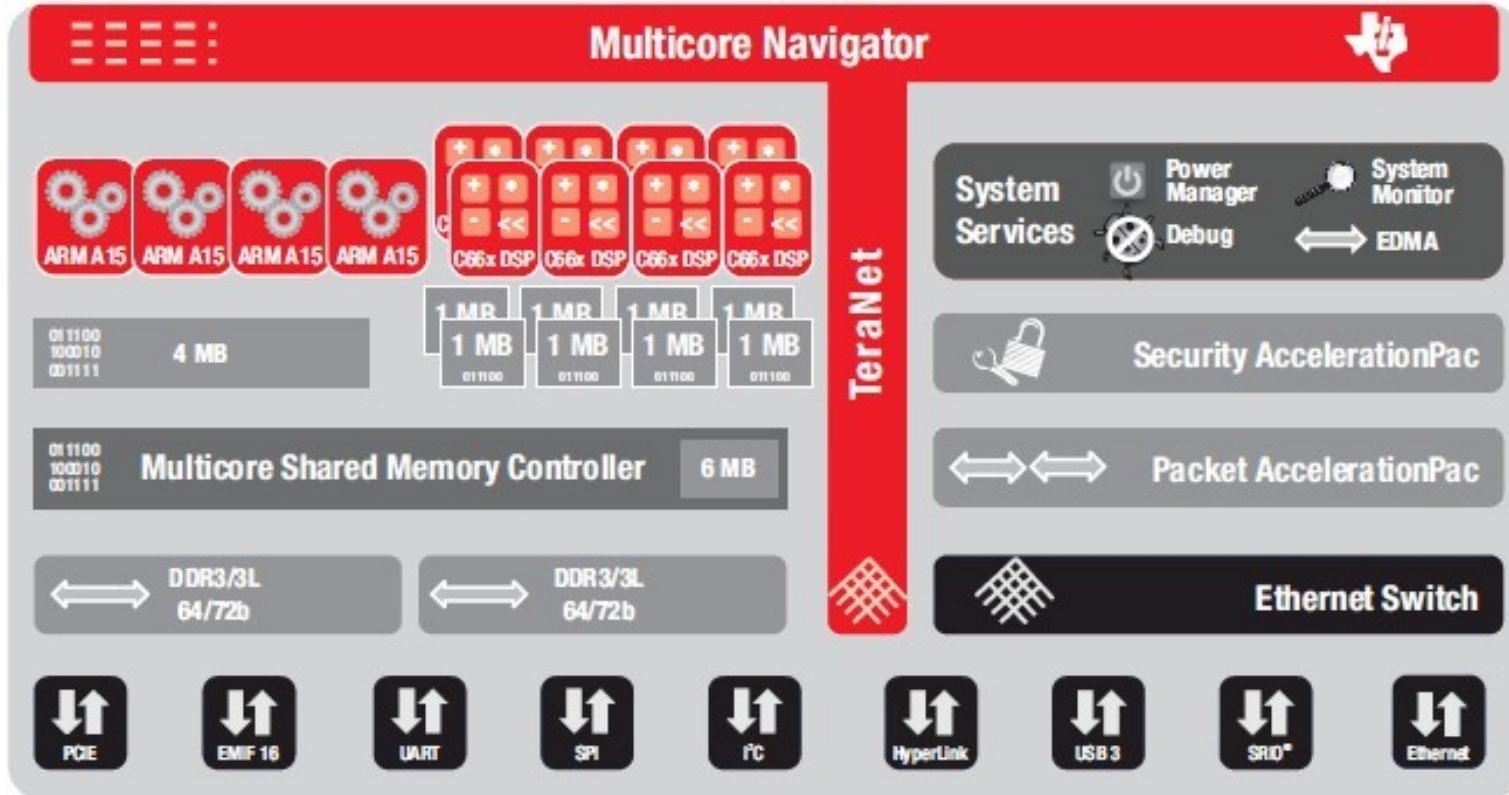
## Exemple Texas Instruments : Keystone II

Mais ce n'est pas tout, TI propose la gamme Keystone II. Il s'agit de SoC de type AP spécialisés pour les applications DSP.

Principales applications visées : les télécommunications.



## Exemple Texas Instruments : Keystone II



## Acteurs du marché

Le leader historique et actuel incontesté du marché est Texas Instruments.  
TI a été la première société à proposer des processeurs DSP en 1982.



## Acteurs du marché

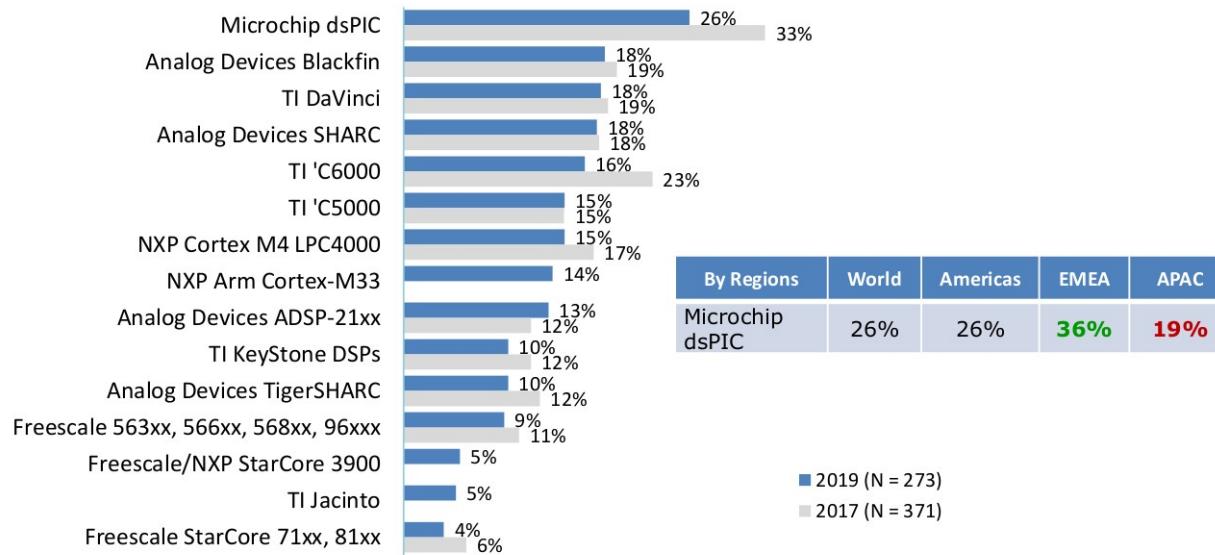
Voici la gamme de processeurs proposée par Texas Instruments.

Microcontrollers (MCUs)		ARM®-based Processors			Digital Signal Processors		
16-bit Ultra Low Power MCU	32-bit Real-Time MCU	32-bit ARM MCU	32-bit ARM Processors for Performance Applications	Application Processors	Singlecore DSP	Multicore DSP	Ultra Low Power DSP
<ul style="list-style-type: none"> <li>• MSP430™ </li> <li>• C2000™</li> </ul>	<ul style="list-style-type: none"> <li>• TMS570 Cortex® R4</li> <li>• RM4 Cortex® R4F</li> <li>• TMS470M Cortex® M3 Automotive</li> </ul>	<ul style="list-style-type: none"> <li>• Sitara™ Cortex A and ARM9</li> <li>• KeyStone Cortex® A15 and Cortex® A15 + DSP</li> </ul>	<ul style="list-style-type: none"> <li>• OMAP™ Processors</li> <li>• DaVinci™ Video Processors</li> </ul>	<ul style="list-style-type: none"> <li>• C6000™ Power Optimized</li> </ul>	<ul style="list-style-type: none"> <li>• KeyStone Multicore DSP+ARM</li> <li>• C6000™ Multicore</li> </ul>	<ul style="list-style-type: none"> <li>• C5000™ </li> </ul>	

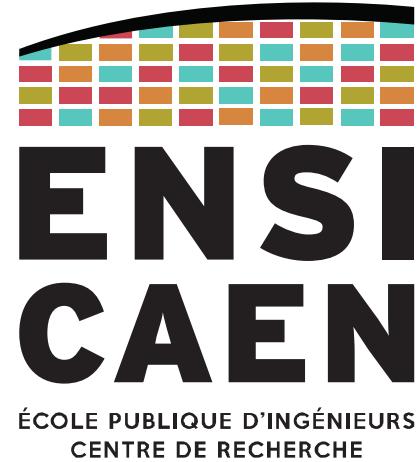
## Acteurs du marché



Which of the following DSP chip families would you consider for your next embedded project?

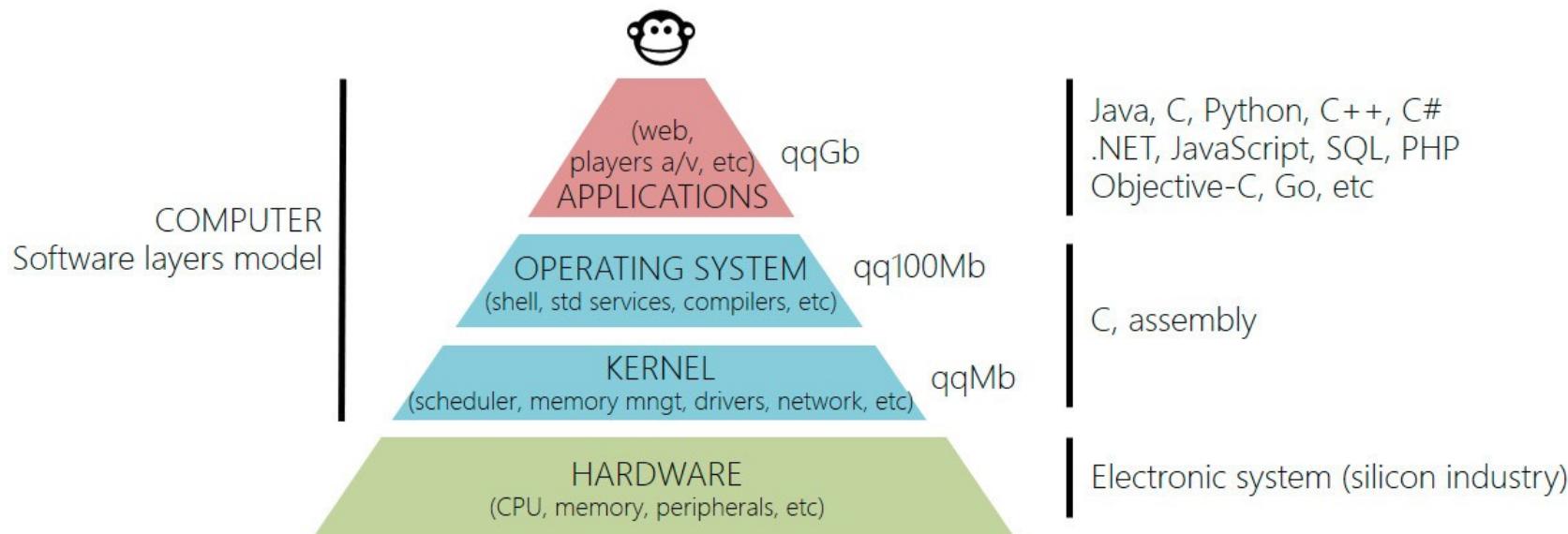


# CHOISIR UN PROCESSSEUR



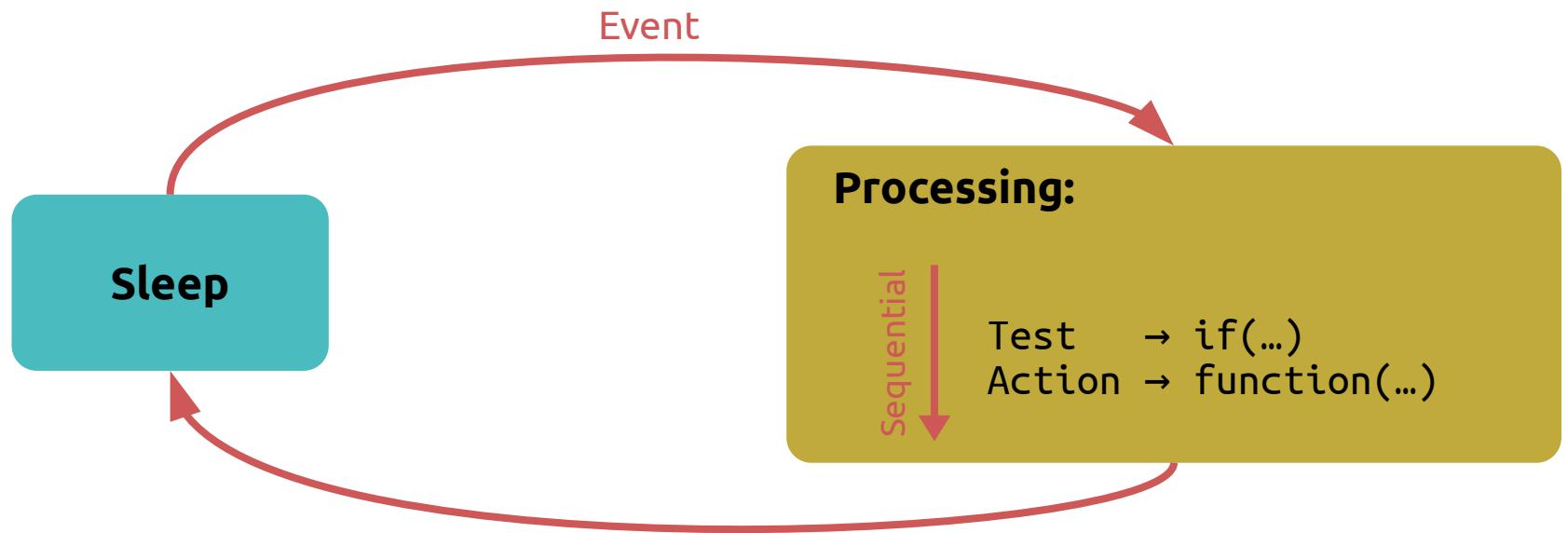
Système = Hardware + Application

Le système (composé du matériel et du logiciel) est développé dans l'optique de répondre à des spécifications ou exigences bien définies.



Si application == supervision

Dans 90 % des cas, le traitement logiciel consiste à effectuer une **simple supervision**.



→ Dans ces cas, opter pour un MCU, AP ou GPP

En fonction du niveau de complexité des interfaces (réseau, IHM, ...) et de l'application (mono-/multi-tâches, ...)

Si application == algorithme

Mais de temps en temps, la fonction à exécuter peut être un **algorithme**, c'est-à-dire l'application d'un traitement pour une certaine quantité de données (informations).



Exemples d'algorithme : recherche, tri, traitement numérique du signal (audio, radar, communication, ...)

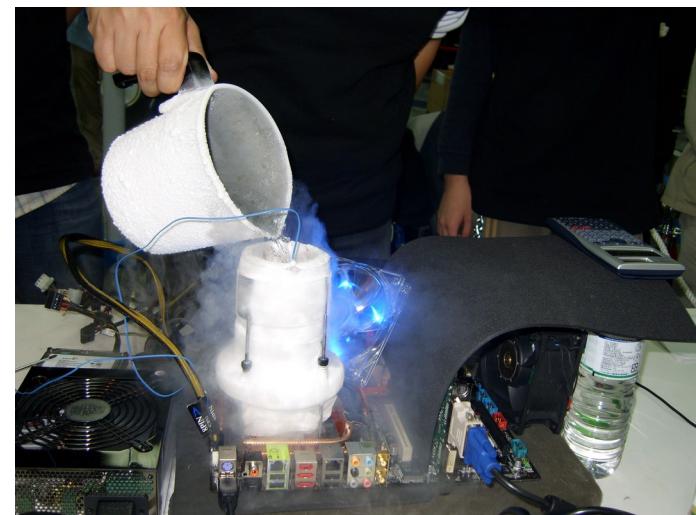
Si application == algorithme

Le premier choix de processeur devrait toujours être un processeur généraliste.

Cependant, quand ses performances ne permettent pas de répondre aux spécifications, il est sage de basculer vers une architecture spécialiste pour :

- Réduire le temps de traitement
- Réduire la taille du code et/ou son empreinte mémoire

Passer d'une architecture généraliste vers une architecture spécialiste devrait toujours être justifié par des mesures.



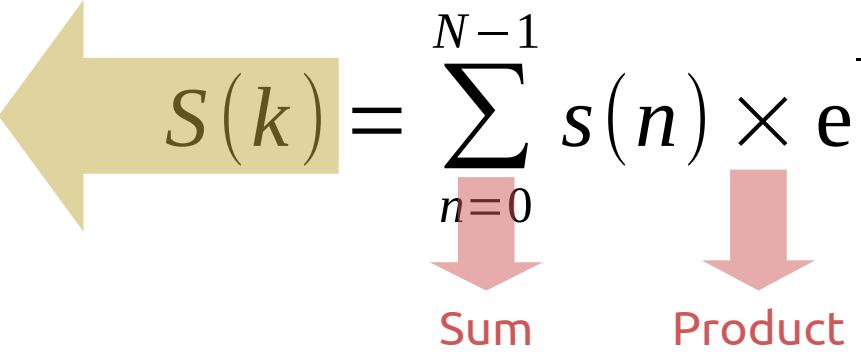
Prenons par exemple l'algorithme de DFT (*Discrete Fourier Transform*) :

$$S(k) = \sum_{n=0}^{N-1} s(n) \times e^{-j 2 \pi k \frac{n}{N}}$$

For ONE frequency sample

Sum      Product

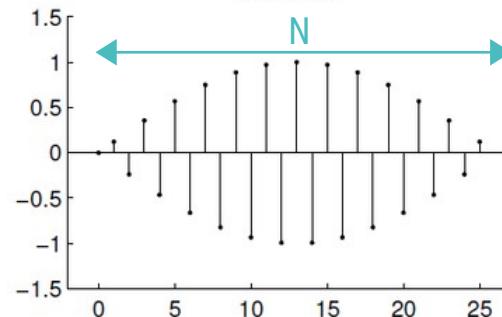
} Sum Of Product (SOP)  
Multiply-Accumulate (MAC)



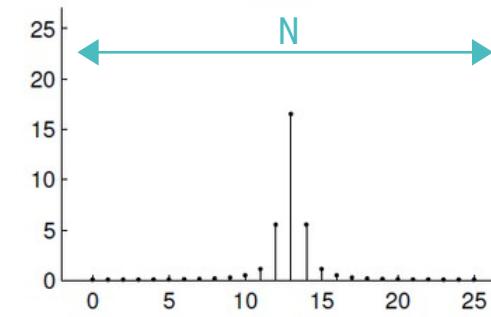
- Chaque produit est indépendant des autres
  - → Traitement en parallèle possible !
- Chaque échantillon de fréquence est indépendant des autres
  - → Traitement en parallèle possible !

## Exemple d'algorithme

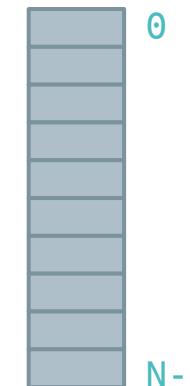
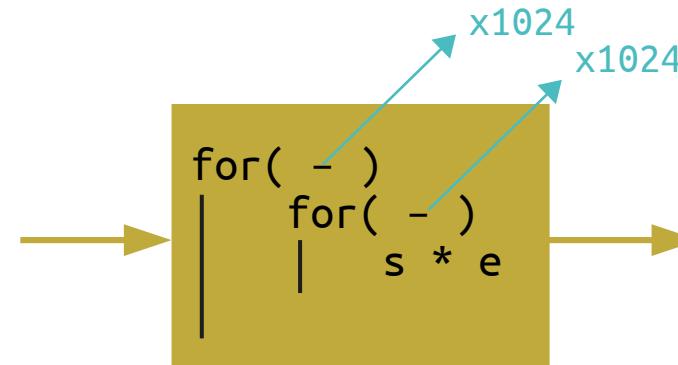
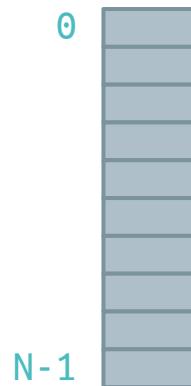
MATH



$$\sum \sum s \times e$$



SOFT



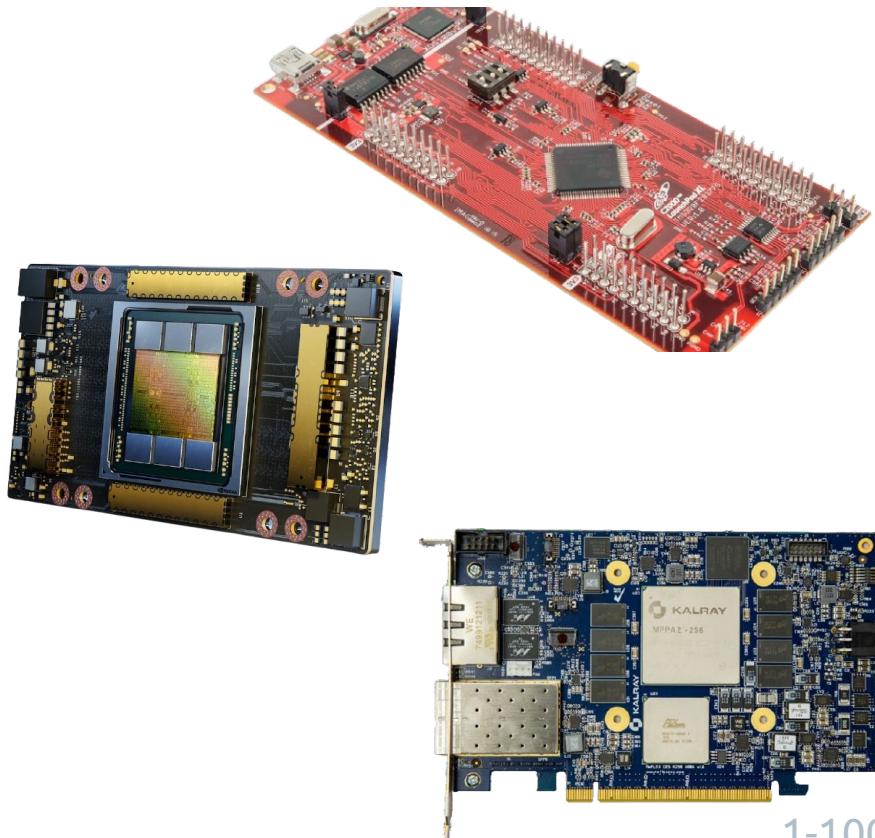
## Choix de l'architecture spécialiste

Choisissons un CPU en fonction des besoins :

**DSP** : faible consommation, faible coût,  
développement bas niveau (C, asm)

**GPU** : hautes performances, prix fort,  
développement haut niveau (C++, OpenMP,  
Cuda, ...), gros potentiel de parallélisme.

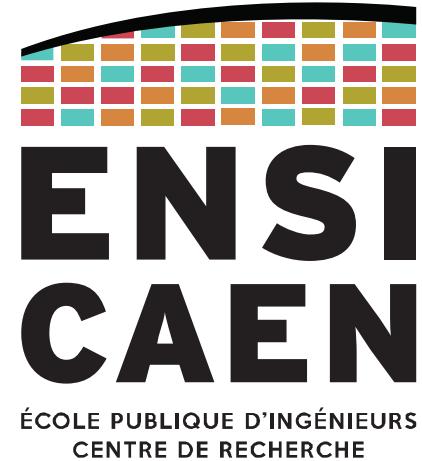
**MPPA** : *Massively Parallel Processor Array*, pas  
encore répandu, mais gros potentiel (répartir  
les coeurs pour des parties d'algorithmes).



# MODÈLES D'EXÉCUTION

Classification des processeurs selon leur modèle d'exécution

SISD – SIMD – MISD – MIMD

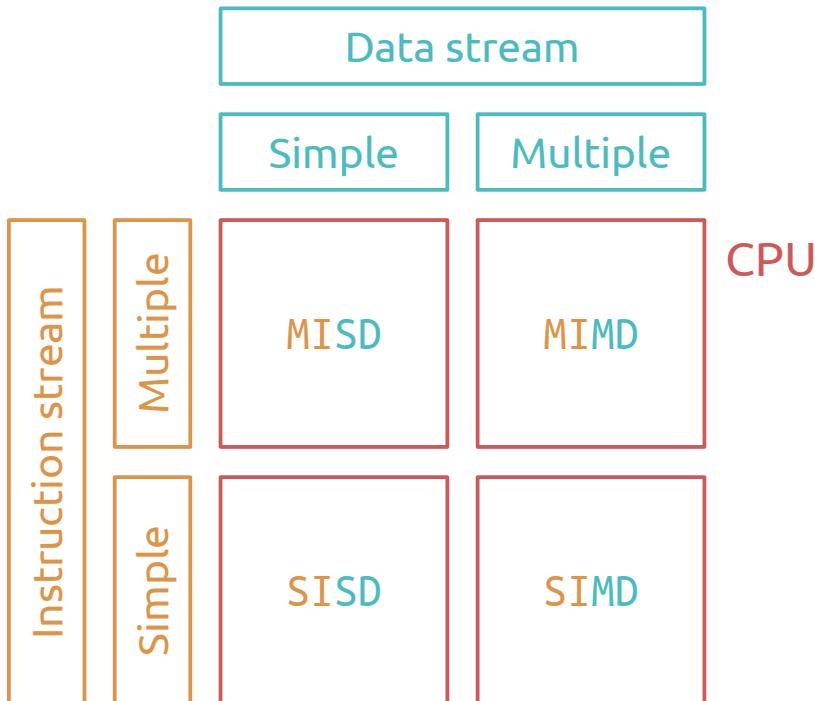


Les prochaines diapos concernent des termes que vous entendrez régulièrement. Elles n'ont pas pour vocation de donner des explications approfondies mais de vous donner une première idée sur leur définition.



## Classification de Flynn

## Classification de Flynn (1972)



*Simple data stream*: chaque opérande ne contient qu'une seule donnée (une case mémoire par opérande).

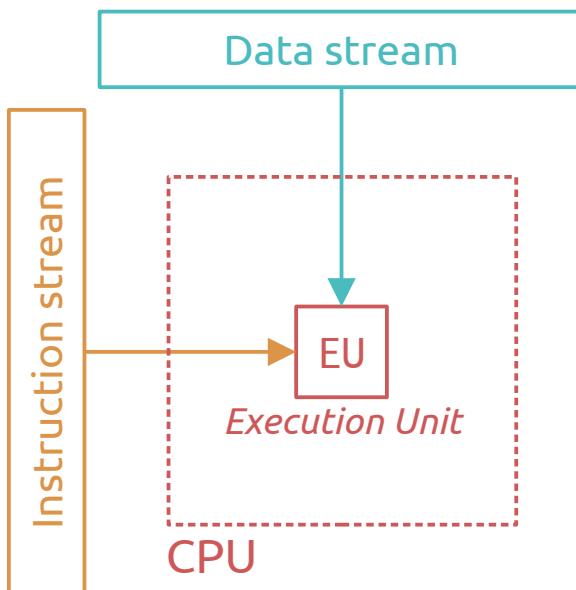
*Multiple data streams*: chaque opérande contient plusieurs données (un tableau fixe par opérande).

*Single instruction stream*: le CPU peut exécuter une instruction à la fois (exécution séquentielle).

*Multiple instruction streams*: le CPI peut exécuter plusieurs instructions simultanément, soit en utilisant le parallélisme de données (ex : boucle *forall*) ou le parallélisme de contrôle (ex : sections parallèles).



## SISD – Single Instruction stream, Single Data stream



Le processeur exécute une instruction à la fois, chaque instruction traitant une donnée.

Architecture mono-processeur classique :

- Architecture Von Neumann
  - Micro-contrôleurs, anciennes générations de processeurs
  - Processeur séquentiel (pas de parallélisme)
- **Processeur scalaire**
  - Une seule donnée (une seule case mémoire) par opérande

## SISD – Single Instruction stream, Single Data stream

*Example: TI C6600 assembly language  
Adding two floats*

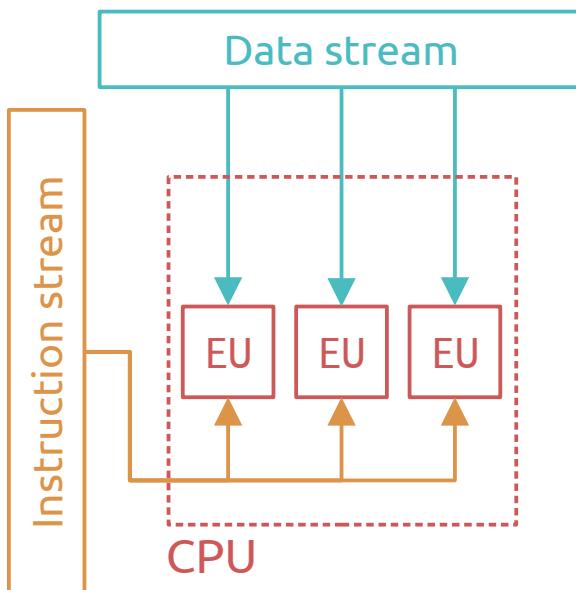
```
; Single Precision ADD  
ADDSP      A17, A5, A5
```

```
; Result:  
; A5 = A5 + A17
```

*Example canonical C:  
Adding two floats*

```
float a, b ;  
  
// Initialising a and b ...  
  
a = a + b ;
```

## SIMD – Single Instruction stream, Multiple Data streams



La même instruction est exécutée par plusieurs EU, chaque EU traite sa propre donnée. Ainsi le CPU exécute la même instruction sur plusieurs données.

Archi. parallèle avec unité de contrôle centralisée :

→ Processeur vectoriel

→ GPU

→ Jeu d'instructions SSE et AVR d'Intel pour x86

SSE = Streaming SIMD Extension (SSE, SSE2, SSE3, SSE4)

AVR = Advanced Vector Extensions (AVX, AVX2, AVX512)

## Classification de Flynn

**SIMD – Single Instruction stream, Multiple Data streams**

*Example: TI C6600 assembly language  
Adding two couples of floats*

```
; Dual ADD Single Precision
DADDSP    A21:A20, A25:A24, A25:A24
```

```
; Result:
; A25 = A25 + A21
; A24 = A24 + A20
```

; Just like the SSE for Intel, the C6600  
; DSP has a C extension (C functions)  
; for vectorial instructions

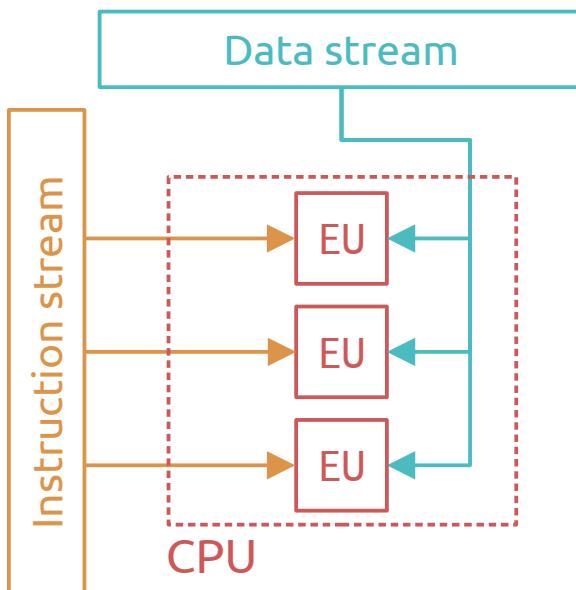
*Example: x86 SSE C, adding four couples of floats*

```
float A[N], B[N], C[N] ;
for( int i = 0 ; i < N ; i += 4 ) {
    __m128 reg_b = _mm_load_ps( &B[i] );
    __m128 reg_c = _mm_load_ps( &C[i] );
    __m128 reg_a = _mm_add_ps( reg_b , reg_c ) ;
    _mm_store_pd( &A[i] , reg_a );
}
```

Lanes per type in a 128-bit SIMD register

int8x16	s0	s1	s2	s3	s4	s5	s6	s7	s8	s9	s10	s11	s12	s13	s14	s15
int16x8	s0	s1	s2	s3	s4	s5	s6	s7								
int32x4 / float32x4	s0 (x)				s1 (y)				s2 (z)				s3 (w)			
float64x2	s0 (x)								s1 (y)							

## MISD – Multiple Instruction streams, Single Data stream



Chaque UE exécute une instruction propre, chaque EU traite une seule donnée.

Peu d'applications en pratique :

→ redondance de code (détection d'erreur d'exécution)

→ **Processeur VLIW** (*Very Long Instruction Word*)

Ex. C66xx Texas Instruments DSP

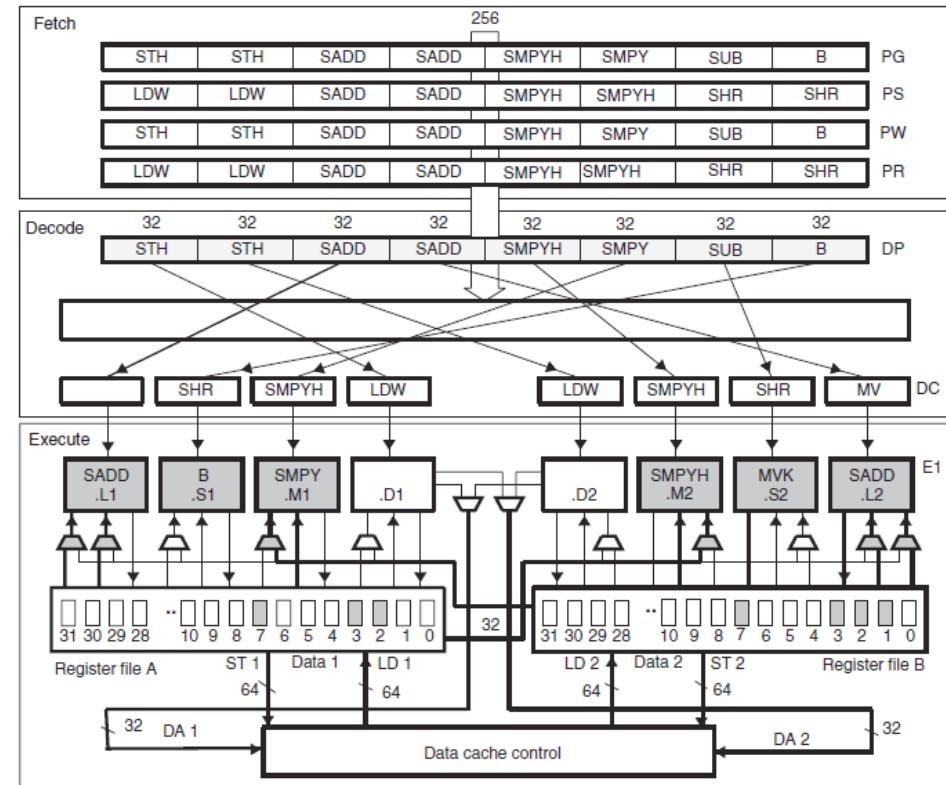
## MISD – Multiple Instruction streams, Single Data stream

*Example: TI C6600 assembly language  
Simultaneously adding and multiplying*

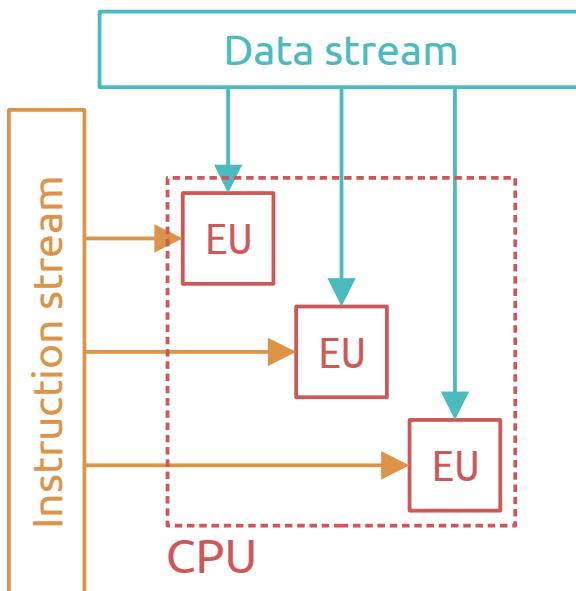
```
; ADD Single Precision
; MULTIPLY Single Precision
ADDSP      A3, A9, A3
|| MPYSP    B3, B9, B3
```

; The pipes (||) explicitly indicate that  
; instructions must be executed in parallel  
; (use of software pipeline)

; Result
; A3 = A9 + A3
; B3 = B9 + B3



## MIMD – Multiple Instruction streams, Multiple Data streams



Chaque EU exécute son propre lots d'instructions (les EU peuvent être groupées), en agissant chacun sur des données différentes.

Archi parallèle avec unités de contrôle indépendantes

→ Processeur super-scalaire

→ N'importe quel GPP moderne : x86-x64 (CISC), Cortex-A (RISC)

→ Inclut le SPMD (*Single Program, Multiple Data*)

## MIMD – Multiple Instruction streams, Multiple Data streams

*Example: TI C6600 assembly language  
Simultaneously adding and multiplying two different couples of data*

```
; Dual ADD Single Precision
; Dual SUBSTRACT Single Precision
    DADDSP    A21:A20, A25:A24, A25:A24
||  DSUBSP    B25:B24, B23:B22, B23:B22
```

; The pipes (||) explicitly indicate that  
; instructions must be executed in parallel  
; (use of software pipeline)

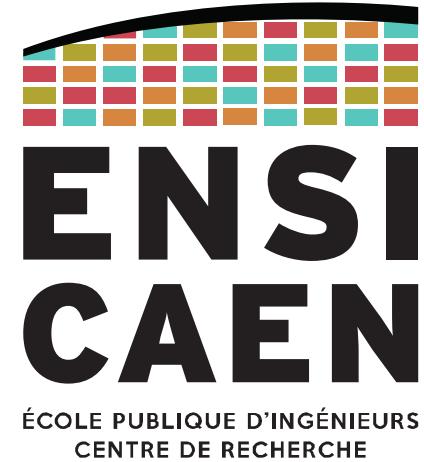
```
; Result
; A25 = A25 + A21
; A24 = A24 + A20
; B23 = B25 - B23
; B22 = B24 - B22
```

*Example: C and OpenMP  
Parallelisation of for loop*

```
#pragma omp parallel reduction(+:acc)
{
    #pragma omp for schedule(static)
    for( k = 0; k < size; k ++ )
    {
        acc += A[i * size + k] * x[k];
    }
}
```

# TECHNOS À SUIVRE

NPU - MPPA

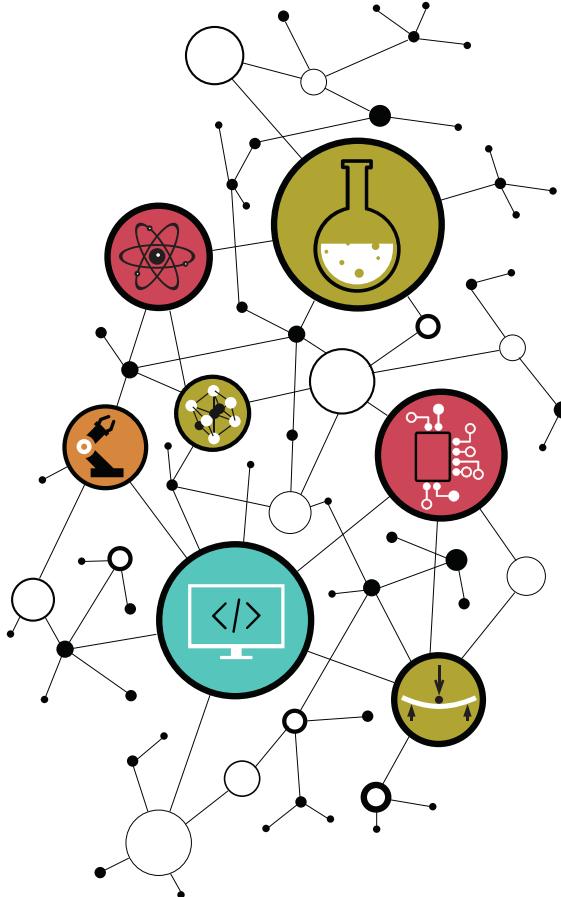




Section en cours de construction

De toute façon on a rarement le temps  
d'arriver jusqu'ici !

## CONTACT



Dimitri Boudier – PRAG ENSICAEN  
[dimitri.boudier@ensicaen.fr](mailto:dimitri.boudier@ensicaen.fr)

Avec l'aide précieuse de :

- Hugo Descoubes (PRAG ENSICAEN)



Except where otherwise noted, this work is licensed under  
<https://creativecommons.org/licenses/by-nc-sa/4.0/>