

# Retours séminaire ARISTOTE

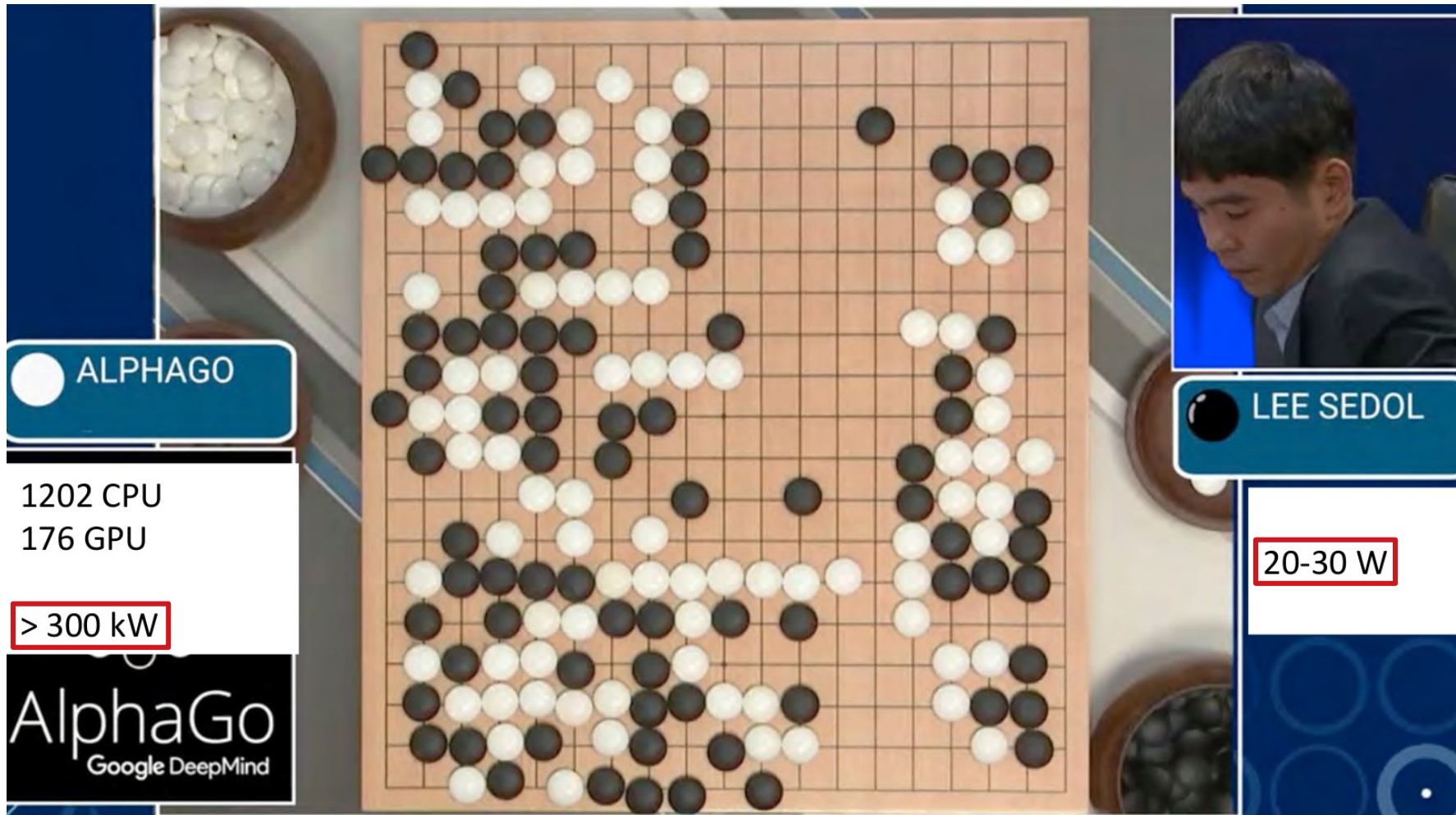
## « Réinventer l'informatique »

Hadrien Grasland  
 LAL – Orsay  
 07/01/2019

# Donne du temps CPU à un programmeur...

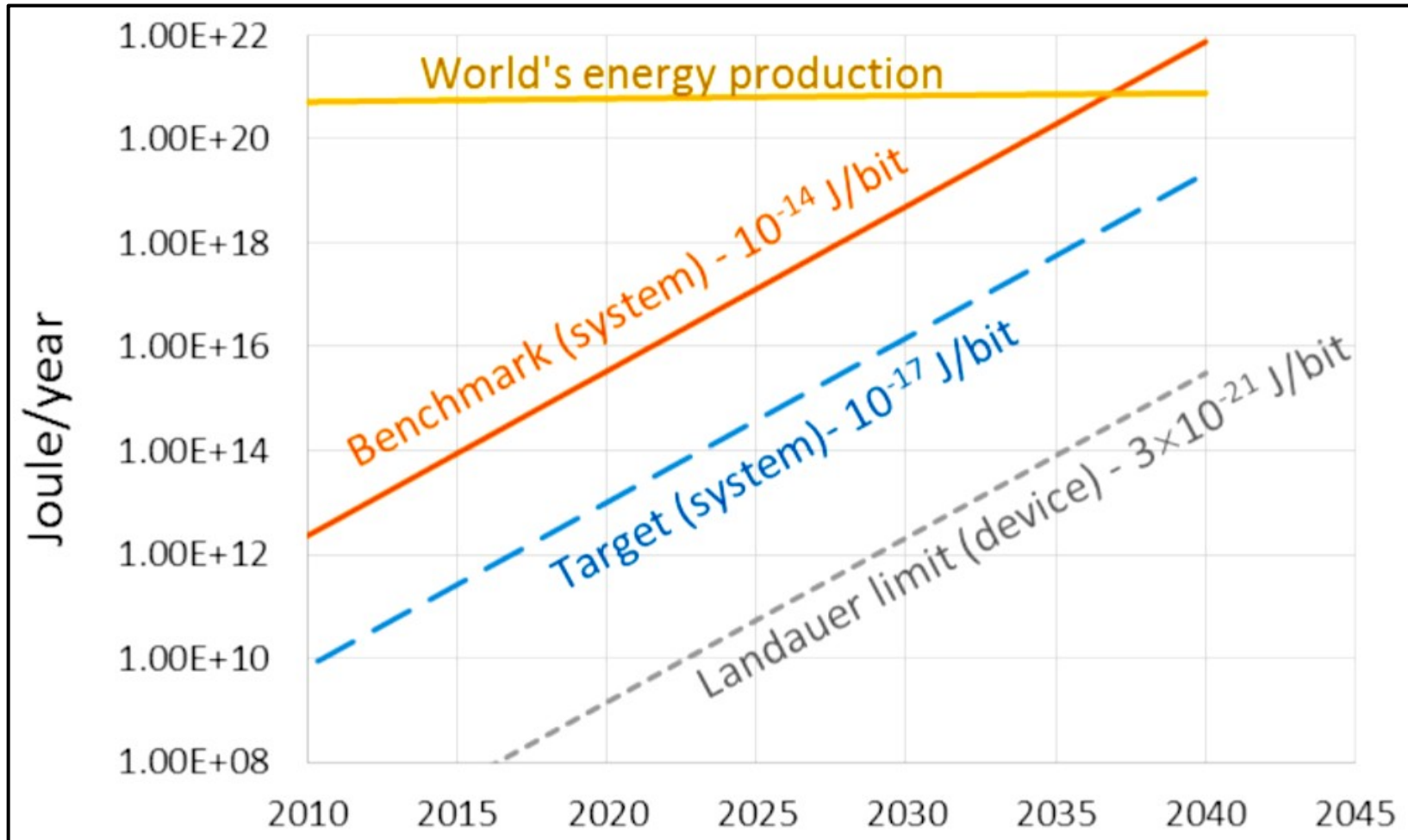


# Donne du temps CPU à un programmeur...



Source : « Fin de l'ère 'logique binaire-architecture Von Neumann-CMOS' : tapage ou véritable 'nouveau monde' ? », Alain Cappy, 12/2018

# Donne du temps CPU à un programmeur...



Source : Rapport « Rebooting the IT revolution : A Call to Action », SIA et SRC, 09/2015

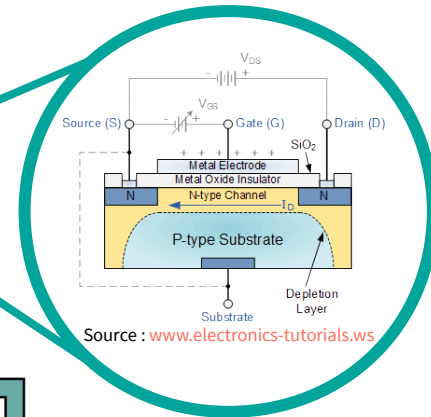
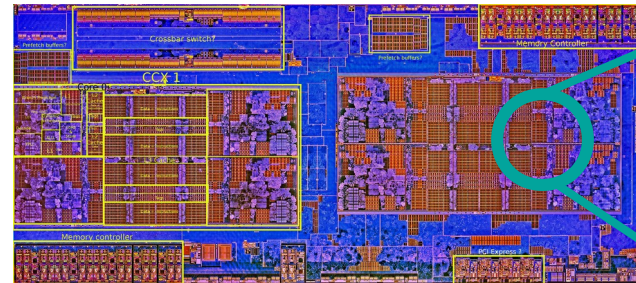
# Au programme

- Le trio de choc binaire – CMOS – Von Neumann
- Pourquoi certains pensent qu'il atteint ses limites
- Des idées pour dépasser ces dernières

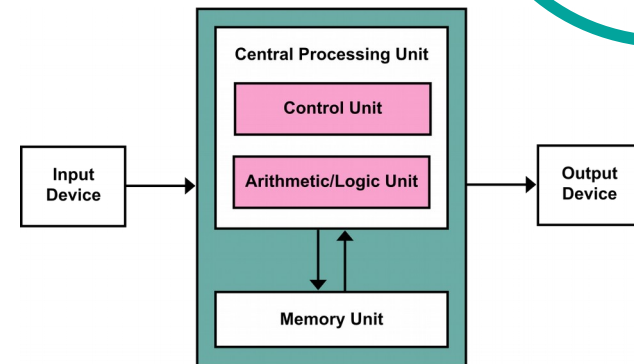
# Fondements du matériel moderne

- Au coeur d'un ordinateur actuel :

- Des transistors MOS...



- ...organisés selon une architecture Von Neumann...



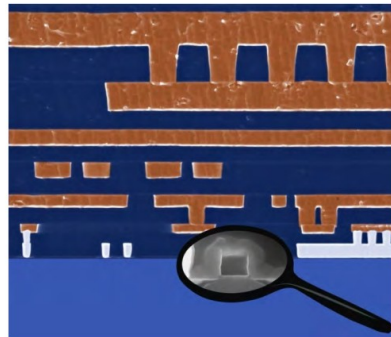
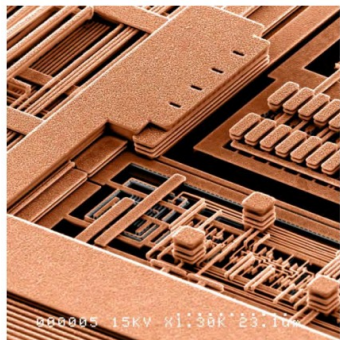
- ...pour calculer de l'algèbre de Boole



# Efficacité énergétique des calculs

## Coût de transfert de charge

- Transition 0-1 :  $E = CV^2/2$
- C prop. *longueur du fil*
- Transistor :  $V > V_{thr} \sim 1V$



Source : « Fin de l'ère 'logique binaire-architecture Von Neumann-CMOS' : tapage ou véritable 'nouveau monde' ? », Alain Cappy, 12/2018

Parameter (scale factor = a)	Classic Scaling	Current Scaling
Dimensions	1/a	1/a
Voltage	1/a	1
Current	1/a	1/a
Capacitance	1/a	> 1/a
Power/Circuit	1/a <sup>2</sup>	1/a
Power Density	1	a
Delay/Circuit	1/a	~1

Source : « Les challenges posés par les systèmes de calcul pour les applications cyber-physiques cognitives », Marc Duranton, 12/2018

## Rétrécir les transistors ?

- Loi d'échelle pas idéale
- Coût en augmentation...

# Faut-il revoir les fondamentaux ?

- **La technologie CMOS atteint ses limites**
  - Difficile de rétrécir davantage les transistors
  - Difficile de gérer le flux de chaleur qui en résulte
- **Von Neumann ne l'exploite pas très efficacement**
  - Séparation données-calcul → fils longs → dissipation
- **Les évolutions itératives sont peu convaincantes**
  - Rien de comparable au transistor CMOS à l'horizon
  - Pas de remplaçant à la charge électrique non plus



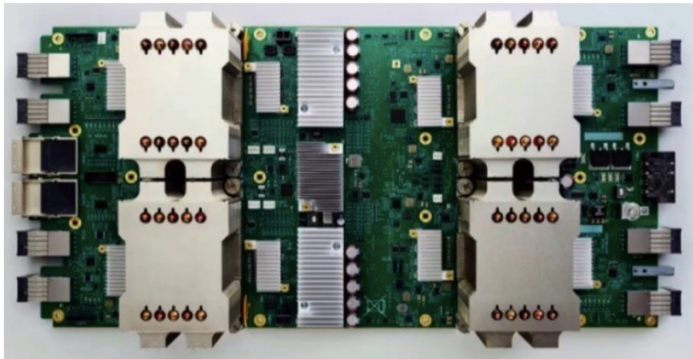
# Approches plus révolutionnaires

- Spécialiser le matériel
- Renoncer au déterminisme
- Rapprocher le calcul des données
- Améliorer la complexité algorithmique

# Pourquoi spécialiser ?

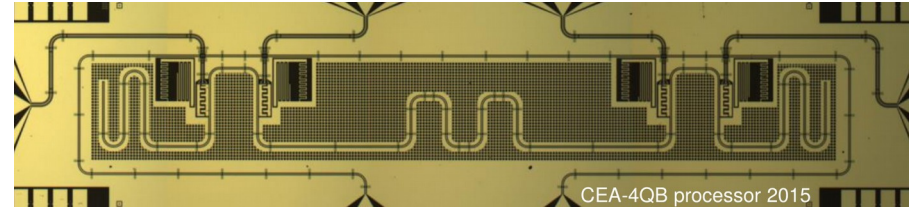
## Utile même dans le cadre Von Neumann

Type of device	Energy / Operation
CPU	1690 pJ
GPU	140 pJ
Fixed function	10 pJ



Google's TPU2 : training and inference in a **180 teraflops<sub>16</sub>** board (over 200W per TPU2 chip according to the size of the heat sink)

Source : « Les challenges posés par les systèmes de calcul pour les applications cyber-physiques cognitives », Marc Duranton, 12/2018

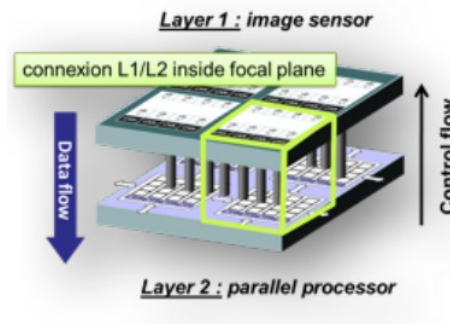
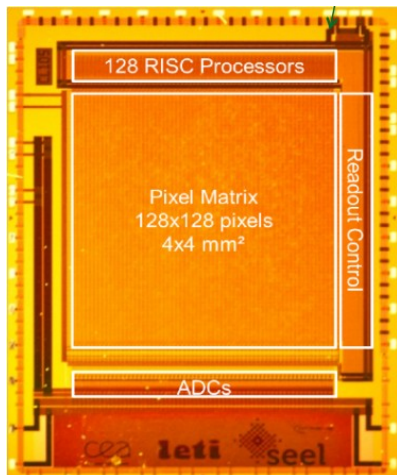


## Nécessaire pour le dépasser

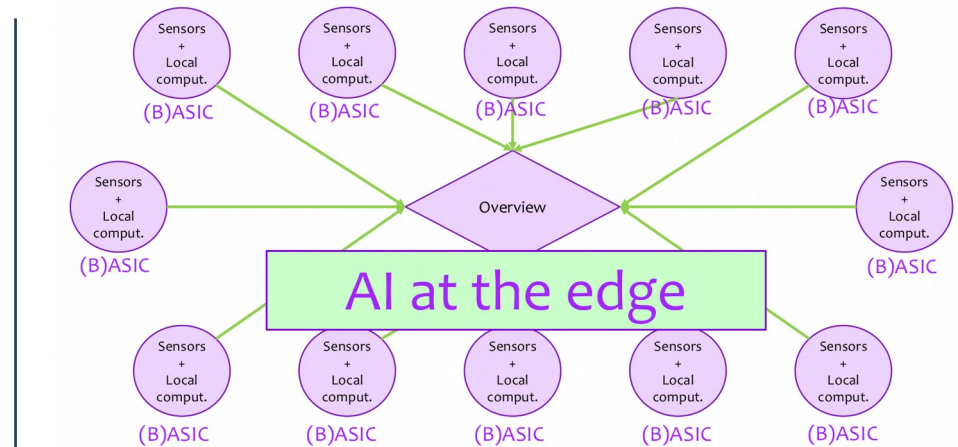
- Calcul dans les capteurs
- Fonctions optiques
- Approches bayésiennes
- Circuits neuromorphiques
- Optim. & info. quantique

# Calcul dans les capteurs

- Communiquer coûte cher ? Traitons à la source !



Source : « Les challenges posés par les systèmes de calcul pour les applications cyber-physiques cognitives », Marc Duranton, 12/2018

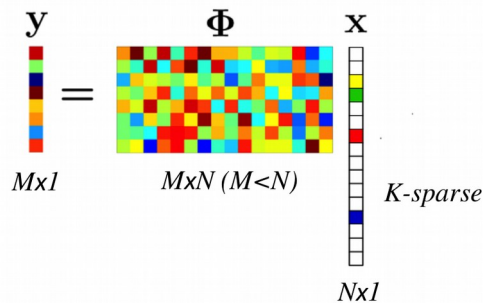


Source : « Bayes, de la cellule à la machine », Pierre Bessière, 12/2018

- Une idée bien connue de la physique des 2 infinis
  - *Trigger*, traitement de données *online*, satellites...

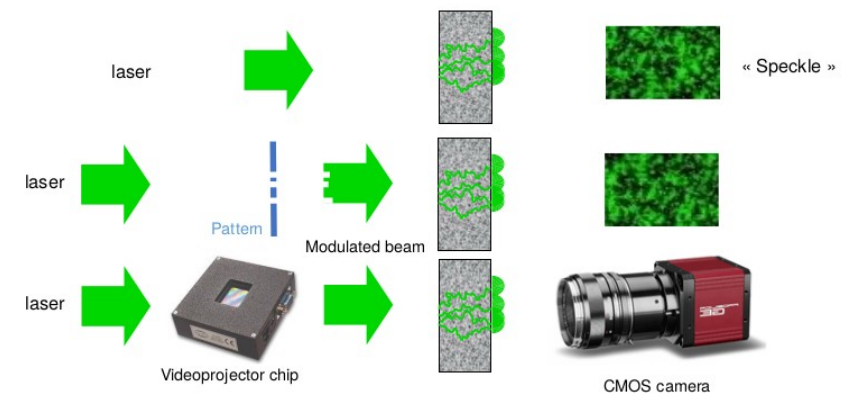
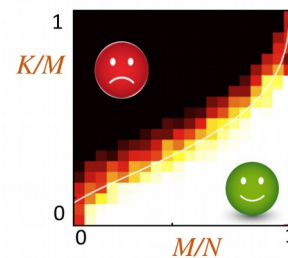
# Fonctions optiques

- Parfois, l'analogique va plus vite que le numérique
  - FFT =  $O(N \times \log(N))$  ops vs Fourier optique =  $O(1)$  ops
- Ici, on fait de la compression par diffusion...



Can one recover x from y ?

YES with tractable algorithms for right values of  $N, M, K$

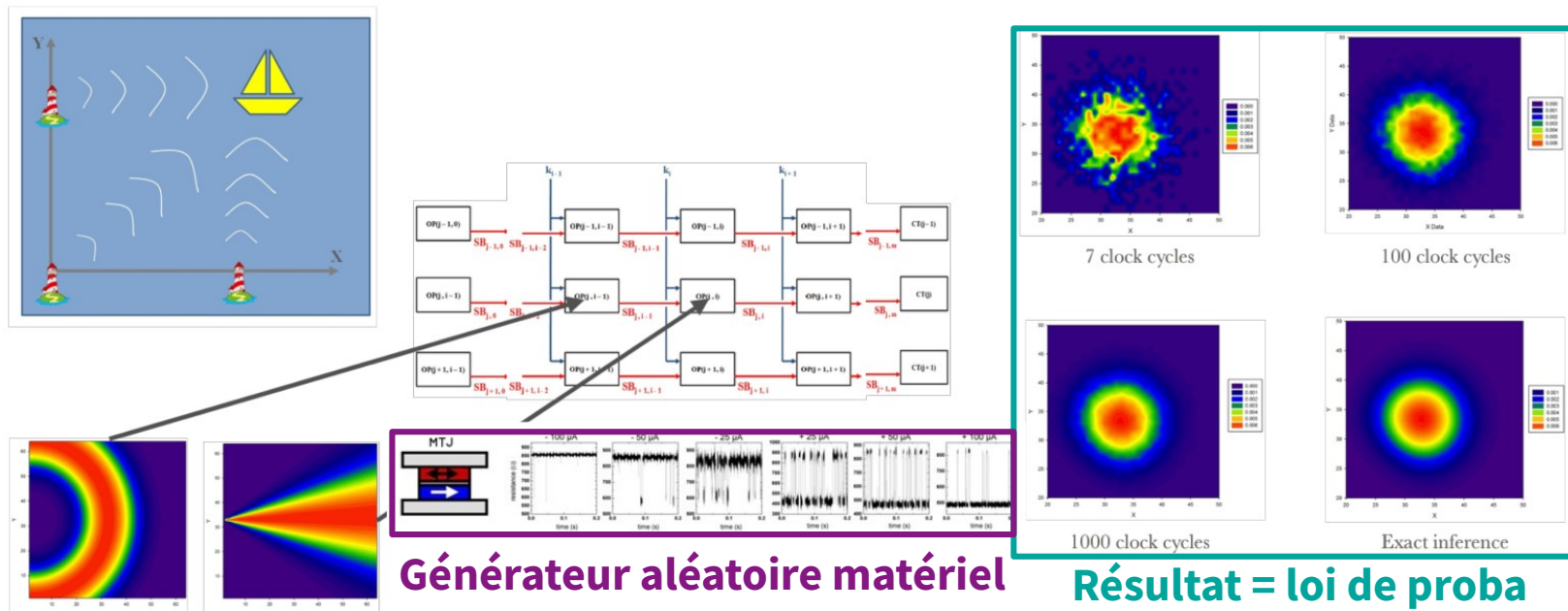


Source : « Utiliser la lumière pour éclairer l'avenir de l'intelligence artificielle », Igor Carron, 12/2018

- ...pour réduire l'info dans un réseau de neurones

# Approche bayésienne

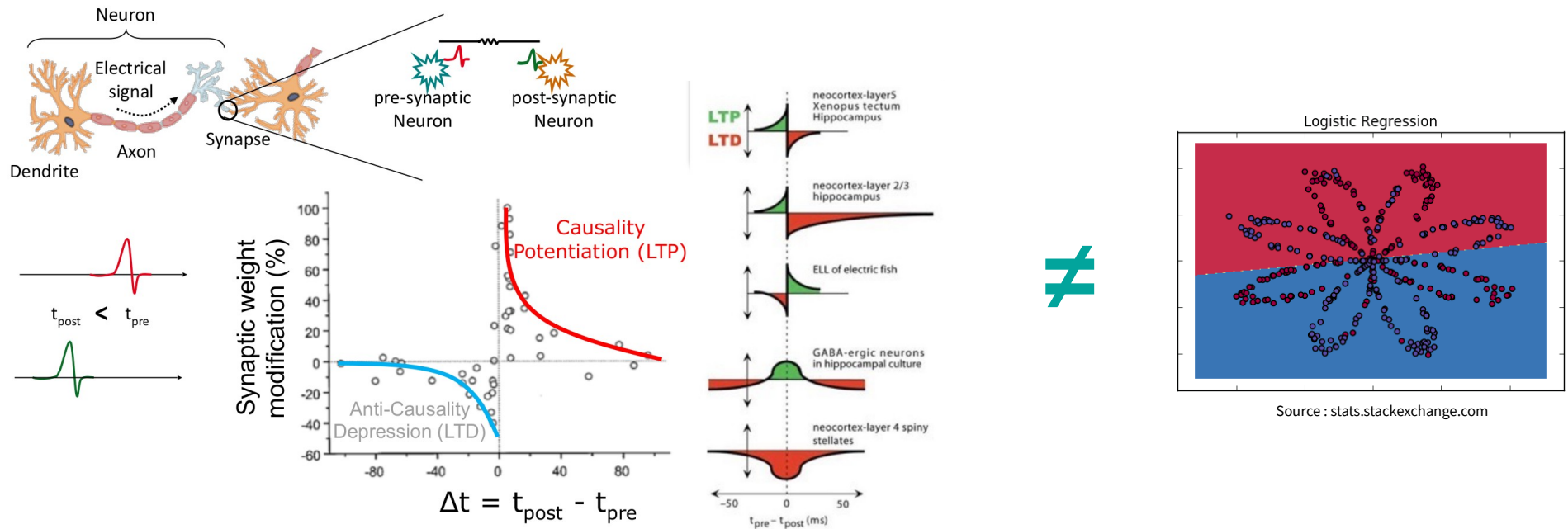
- De nombreux calculs sont d'ordre statistique
- Proposition : Matériel pour le calcul probabiliste



Source : « Bayes, de la cellule à la machine », Pierre Bessière, 12/2018

# Parlons biomimétisme

- « Réseau de neurones » ? Une imposture !
  - Le cerveau n'est pas une pile de régressions logistiques...

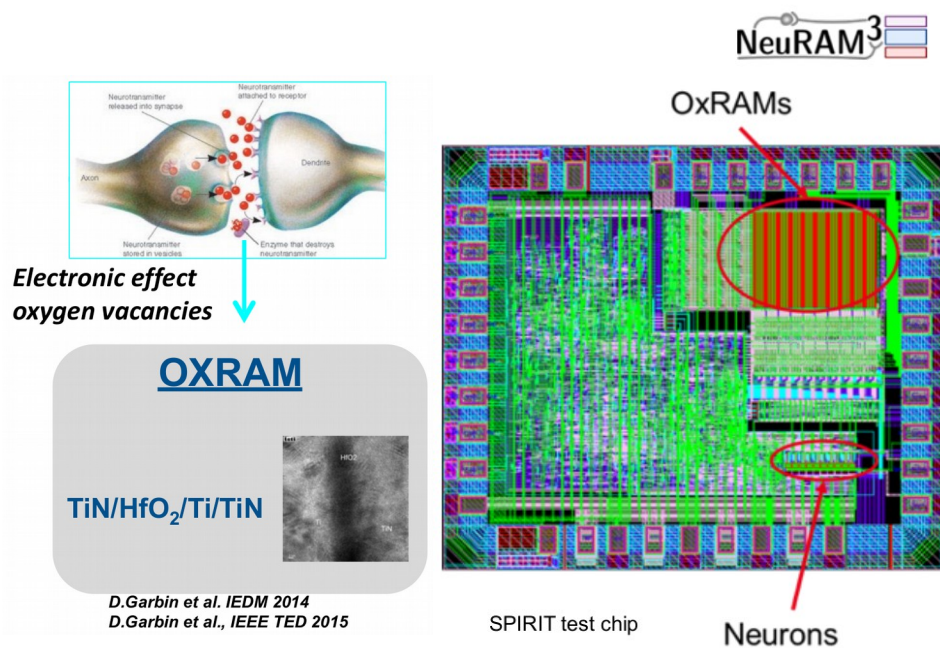


Source : « Les challenges posés par les systèmes de calcul pour les applications cyber-physiques cognitives », Marc Duranton, 12/2018

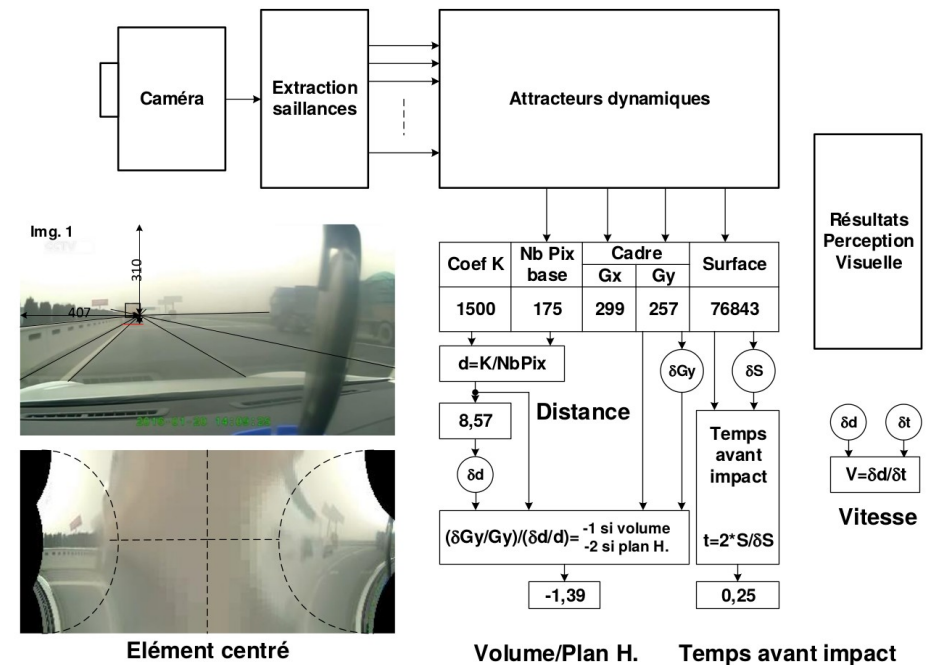
# Circuits neuromorphiques / bio-inspirés

- Certains tentent donc de mieux approximer

Au niveau du neurone...



...ou de la structure globale



Source : « Les challenges posés par les systèmes de calcul pour les applications cyber-physiques cognitives », Marc Duranton, 12/2018

Source : « Vers une IA organique intégrée », Patrick Pirim, 12/2018

# Calculs quantiques

- Il existe différents types d'ordinateurs quantiques
- **Recuit quantique**  $\approx$  ASIC de recuit simulé
  - Réseau Ising artificiel + champ magnétique + effet tunnel
  - Un algo inspiré par la physique qui revient aux sources...
- **Logique quantique**  $\approx$  Algèbre de Boole généralisée
  - Applications nombreuses (pour du post-Von Neumann)
  - Mais beaucoup plus difficile à réaliser...

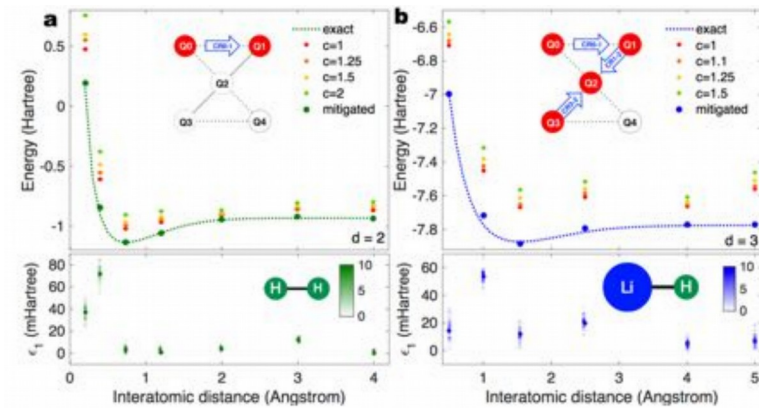


# Bases de logique quantique

- En physique, un bit = un système à deux niveaux
- Leurs propriétés quantiques sont étonnantes
  - **Superposition** → « 2 états simultanés », mesure aléatoire
  - **Intrication** →  $\text{Proba}(\{01, 10\}) = 1/2$ ,  $\text{Proba}(\{00, 11\}) = 0$
  - **Mesure projective** → Manipuler un bit influence un autre
- Une partie de la logique fonctionne en quantique
  - Opérations réversibles : Hadamard, CNOT...
  - ...mais ni copie d'état, ni instructions de contrôle !

# Pourquoi c'est intéressant

- Meilleure complexité algo. sur certains problèmes
  - Chimie quantique
  - Problèmes d'optimisation
  - Factorisation d'entiers



Source : « Quantum Computing with superconducting qubits: Applications in Chemistry », I. Tavernelli, 12/2018

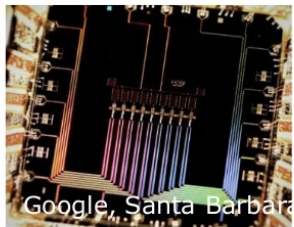
- Lié à la possibilité d'explorer de nombreux états
  - On parle parfois de « parallélisme quantique »
  - L'exploiter nécessite des algorithmes adaptés !

# Pourquoi c'est difficile

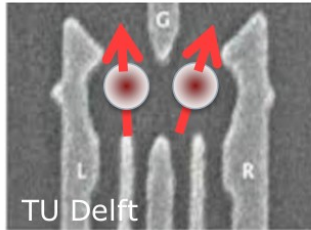
- **Capacités dépendant du nombre de bits intriqués**
  - ~60 qubits : On ne sait plus simuler sur un ordi classique
  - ~200 qubits : Simulation de molécules intéressantes
  - ~2000 qubits : La cryptographie a du soucis à se faire
- **Et ça, c'est sans compter les erreurs...**
  - Un état quantique est très fragile (effets de décohérence)
  - Correction Steane : 7x qubits, corrige erreur HW ~ 0,01 %
  - Surface codes : ~1000x\* qubits, corrige erreur HW ~ 1 %

\* J'ai vu des chiffres allant de 100 à 30k, de la biblio est nécessaire pour clarifier la situation

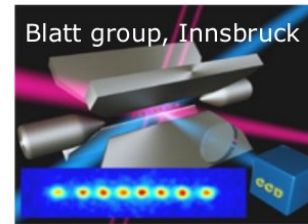
# Etat de l'art du matériel



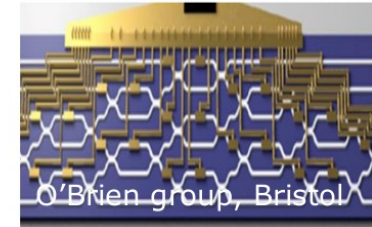
Google, Santa Barbara  
**Superconductors**



TU Delft  
**Silicon**



Blatt group, Innsbruck  
**Trapped ions**



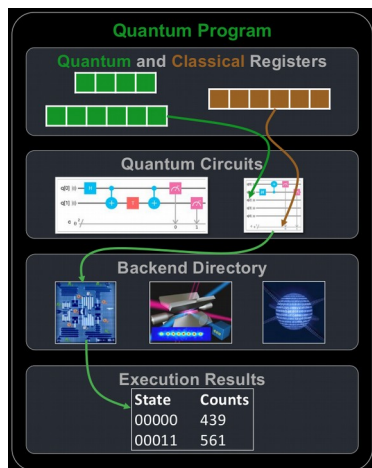
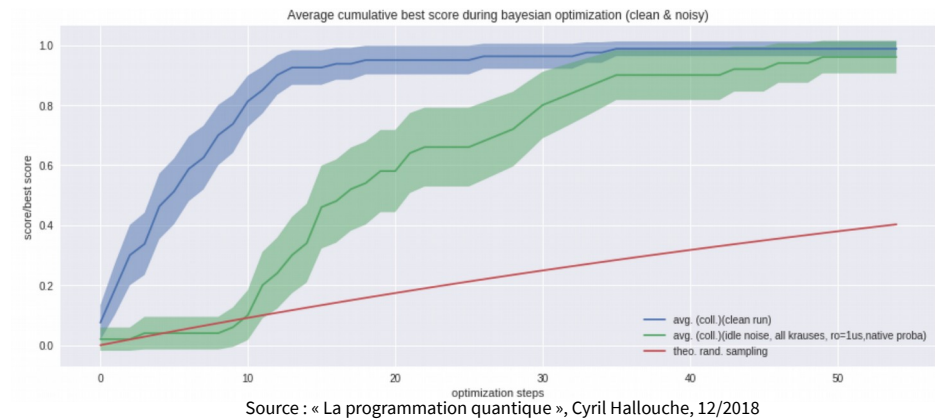
O'Brien group, Bristol  
**Photon**

<b>Nb of entangled qubits</b>	20	2	20	10
<b>Size</b>	$(100\mu\text{m})^2$	$(100\text{nm})^2$	$(1\text{mm})^2$	$(1\text{mm})^2$
<b>Fidelity</b>	~99.9%	~98%	99.99%	50% (measurement /generation) 98% (one, two-qubit gate)
<b>Speed</b>	100 ns	1 $\mu\text{s}$	100 $\mu\text{s}$	1 ms

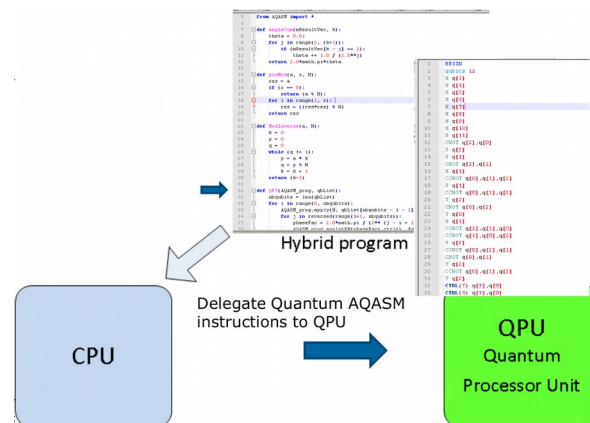
Source : « Vers un traitement de l'information quantique à grande échelle basée sur les spins électroniques dans les semi-conducteurs », Tristan Meunier, 12/2018

# Perspectives à court terme

- Il existe des algorithmes qui tolèrent les erreurs
  - Potentiels chimiques
  - Optimisation QAOA
  - OK avec ~100 qubits bruités



Source : « Quantum Computing with superconducting qubits: Applications in Chemistry », I. Tavernelli, 12/2018



Source : « La programmation quantique », Cyril Hallouche, 12/2018

## Ce qui est en bonne voie

- Systèmes bruités « NISQ »\*
- Modèle de programmation
- Simulateurs « cloud »

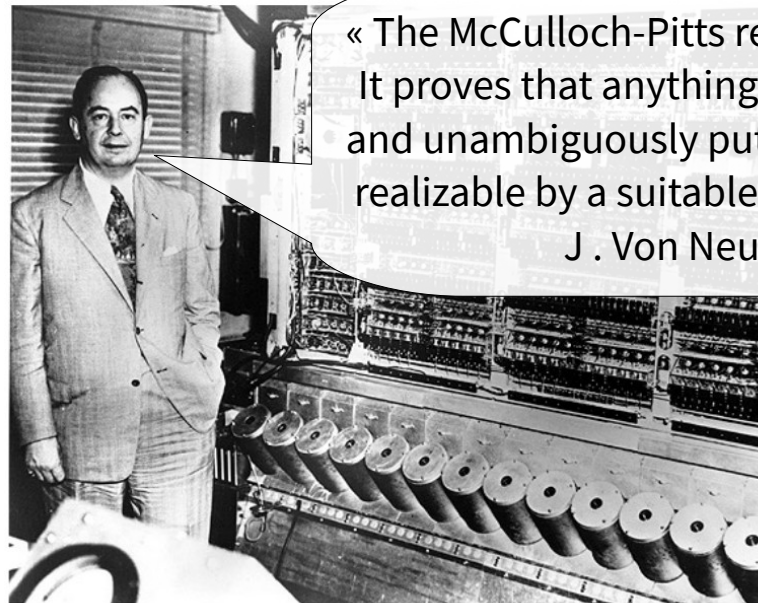
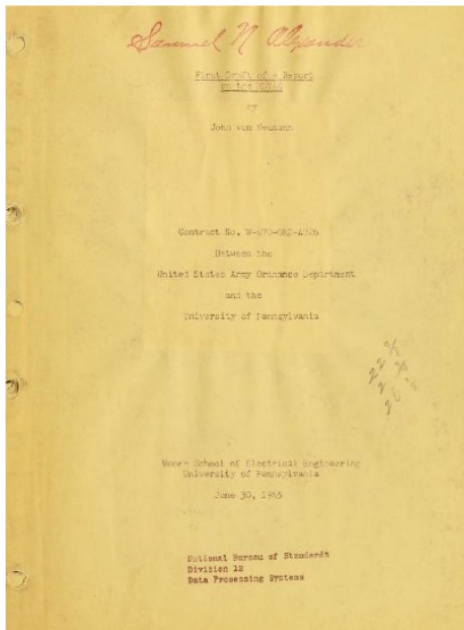
\* Noisy Intermediate Scale Quantum (Computing)

# Conclusion

- **Il y a comme un problème d'efficacité énergétique**
  - Cette fois, la techno CMOS ne va pas le régler pour nous
  - Surtout pas avec une architecture Von Neumann...
- **Différentes pistes à l'étude, leurs points communs :**
  - Spécialisées dans un domaine restreint (coprocesseurs)
  - Résultats stochastiques et entachés d'erreur
  - Rapprochement calcul + mémoire

# Le mot de la fin

In “First Draft of a Report on the EDVAC,” the first published description of a stored-program binary computing machine - the modern computer, John von Neumann suggested modelling the computer after Pitts and McCulloch’s neural networks.



« The McCulloch-Pitts result puts an end to this. It proves that anything that can be completely and unambiguously put into words is ipso facto realizable by a suitable finite neural network. »

J . Von Neumann, 1951

**Des questions ?**