

# Monitoring de la consommation énergétique des centres de calcul

Mathilde JAY

# Présentation

- Étudiante en doctorat sur la consommation énergétique de l'intelligence artificielle à l'université de Grenoble Alpes
- En deuxième année
- Encadrée par
  - Denis Trystram, Prof. ENSIMAG
  - Laurent Lefevre, CR Inria
- Équipes
  - Inria DataMove au LIG (UGA)
  - Inria Avalon au LIP (ENS Lyon)
  - MIAI Edge Intelligence
- Article « an experimental comparison of software-based power meters »

Pour faire le TP, je me suis basée sur le travail de

- Sylvain Bouveret, MdC ENSIMAG
  - Cours Numérique responsable : <https://recherche.noiraudes.net/ecoinfo/numres/ressources/TP/04-mesure-conso.html>
- Vladimir Ostapenco, Doctorant à l'ENS de Lyon
  - M2 ENS de Lyon

On utilisera les machines de Grid'5000, une plateforme expérimentale partagée et dédiée aux chercheurs :

<https://www.grid5000.fr/>

# Programme

- Context : Notions d'énergie
  - Définition
  - Sources d'énergie
  - Ordre de grandeurs
  - Énergie consommée par le numérique
  - Ouverture sur l'impact environnemental
- Comment monitorer la consommation d'énergie d'un serveur ?
- Partie TP : Monitorer l'énergie en pratique

# Définition de l'énergie

D'après Wikipedia :

“L'énergie est un concept relié à ceux d'**action**, de **force** et de **durée** : la mise en œuvre d'une action nécessite de maintenir une certaine force pendant une durée suffisante, pour vaincre les inerties et résistances qui s'opposent à ce changement.

[...]

En science physique, l'énergie, mesurée en **joules** dans le système international, est une mesure de la capacité d'un système à **modifier un état**, à produire un travail entraînant un mouvement, un rayonnement électromagnétique ou de la chaleur.”

$$\text{Énergie (Joules)} = \text{Puissance (Watt)} \times \text{Durée (Secondes)}$$

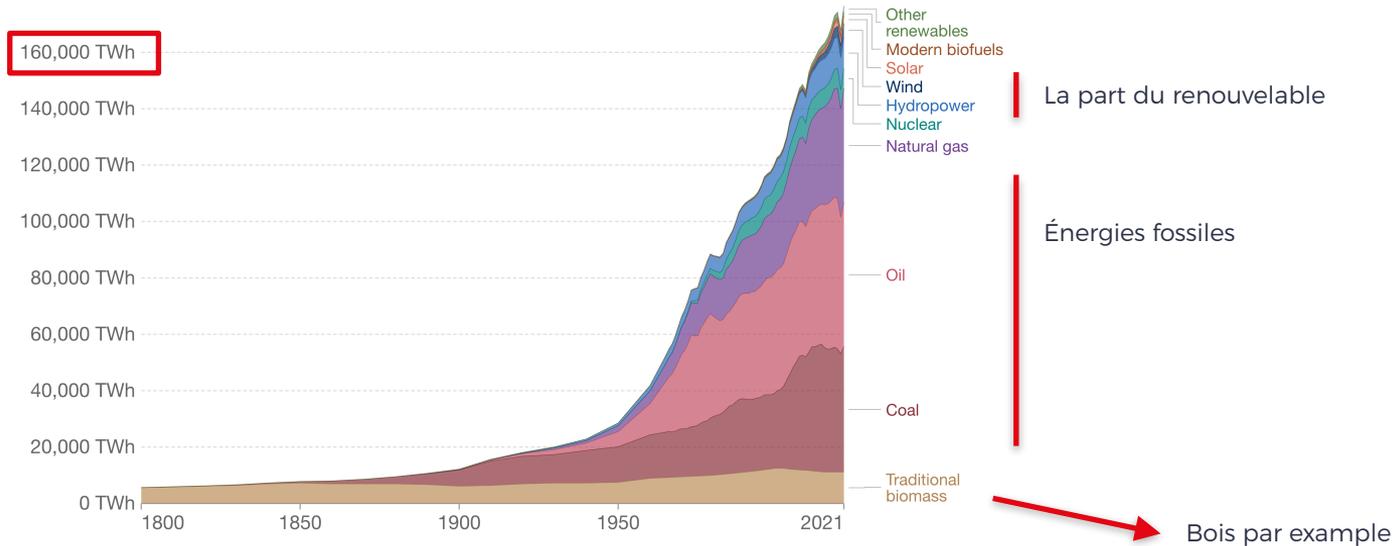
Autre unité très utilisée : kWh

# Sources d'énergie

## Global primary energy consumption by source

Primary energy is calculated based on the 'substitution method' which takes account of the inefficiencies in fossil fuel production by converting non-fossil energy into the energy inputs required if they had the same conversion losses as fossil fuels.

Our World  
in Data



Source: Our World in Data based on Vaclav Smil (2017) and BP Statistical Review of World Energy

OurWorldInData.org/energy • CC BY

# Exemples pour appréhender des ordres de grandeurs

	Puissance	Durée pour atteindre 1 kWh
Sèche-cheveux	2 kW	30 minutes
Ampoules LED	4 W	10,4 jours
Voiture Renault ZOE	100 kW	36 secondes
4 moteurs d'avion	34 MW	0,1 secondes
Macbook Air	Batterie de 50 Wh	20 décharges
Pile AA alcaline	Batterie de 3,75 Wh	267 piles

# Exemples pour appréhender des ordres de grandeurs

HOW MUCH ENERGY DOES IT TAKE  
TO TOAST A SLICE OF BREAD?

# Numérique : définitions

**TDP** : Thermal Design Power

« Quantité maximale de chaleur générée par une puce ou un composant informatique que le système de refroidissement d'un ordinateur est conçu pour dissiper sous n'importe quelle charge de travail. »

Bonne approximation de la **puissance** du composant à **utilisation maximale**.

**PUE** : Power Usage Effectiveness

$$\text{PUE} = \frac{\text{Énergie **totale** consommée}}{\text{Énergie consommée par l'**infrastructure informatique**}}$$

Évalue l'efficacité énergétique des centres de données, en particulier du système de **refroidissement**.

Autre indicateurs : WUE (pour l'eau)

# Exemples dans le numérique : TDP

	TDP
CPU	100 W
GPU H100 PCIe	350 W
GPU H100 SXM	700 W
TPU v4	200 W



Et pour un serveur complet ?

- Serveur cloud : quelques centaines de watt
- Serveur IA avec 8 GPUs : Quelques milliers de watt

# Exemples dans le numérique : PUE

Moyenne mondiale : **1,55**

(D'après <https://uptimeinstitute.com/about-ui/press-releases/2022-global-data-center-survey-reveals-strong-industry-growth>)

Moyenne chez les GAFAM : **1,1**

(D'après eux)

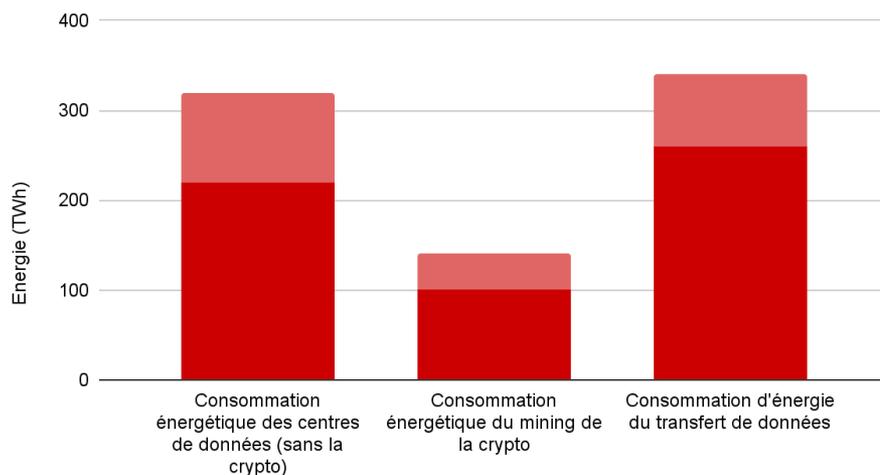
	PUE
Google	1,06-1,10
Meta	1,09
Amazon	1.07–1.15
Microsoft	1,18

# Énergie consommée par les centres de données

Par rapport au total de  
160 000 TWh

Centre de données +  
crypto + transfert des  
données  
= **0,4 %** de l'énergie  
produite mondialement

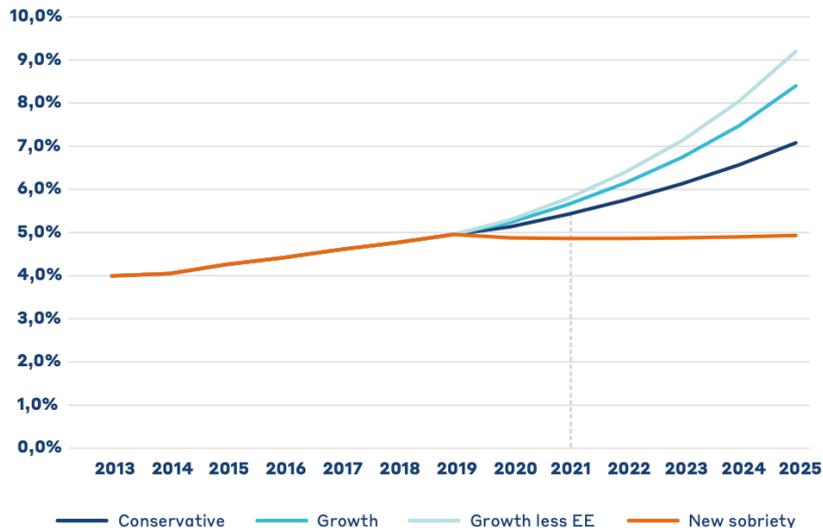
En 2021



D'après l'agence internationale de l'énergie :  
<https://www.iea.org/reports/data-centres-and-data-transmission-networks>

# Énergie consommée par le numérique

Part du numérique dans la consommation d'énergie primaire mondiale



Estimation faite par le Shift Project (2021)

Prend en compte tous les équipements numériques donc

- Centres de donnée
- Infrastructures réseaux
- Terminaux utilisateurs
  - Smartphones
  - Ordinateurs fixes ou portables

Par rapport au graphe précédant : **10 fois plus important** dans la part dans la consommation d'énergie primaire mondiale

[https://theshiftproject.org/wp-content/uploads/2021/03/Note-danalyse\\_Numerique-et-5G\\_30-mars-2021.pdf](https://theshiftproject.org/wp-content/uploads/2021/03/Note-danalyse_Numerique-et-5G_30-mars-2021.pdf)

# Ouverture sur l'impact environnemental



# Source d'énergie du numérique : l'électricité

Par rapport au total de 160 000 TWh

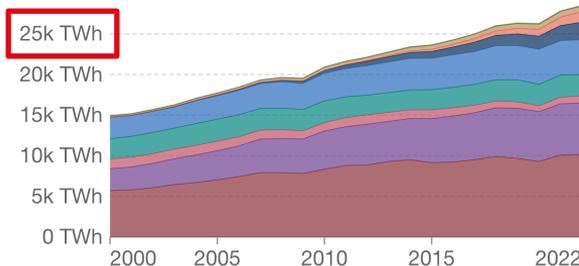
Électricité = **15 %** énergie produite mondiale

## Electricity production by source

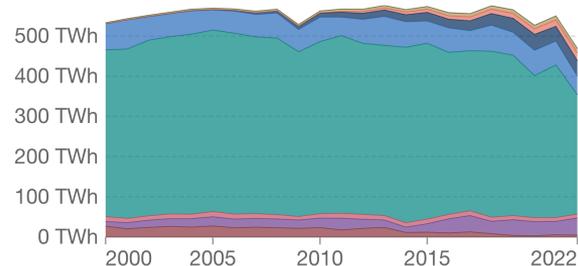


Other renewables Bioenergy Solar Wind Hydropower Nuclear Oil Gas Coal

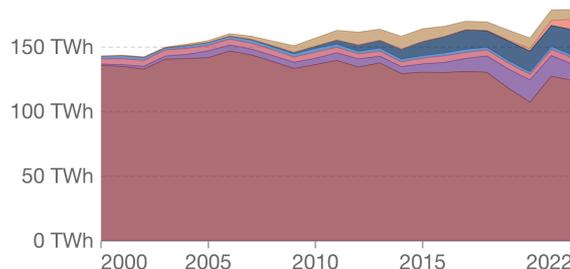
### World



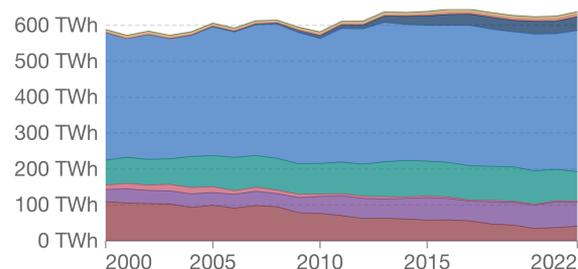
### France



### Poland



### Canada



Source: Our World in Data based on BP Statistical Review of World Energy (2022); Ember (2023)  
Note: 'Other renewables' includes waste, geothermal, wave and tidal.

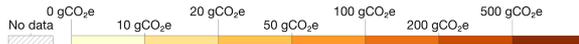
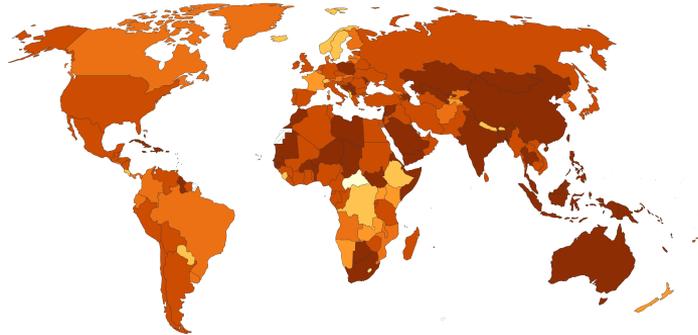
OurWorldInData.org/energy • CC BY

# Évaluer les émissions CO2 équivalent

## Carbon intensity of electricity, 2022

Carbon intensity is measured in grams of carbon dioxide-equivalents<sup>1</sup> emitted per kilowatt-hour of electricity.

Our World  
in Data



Source: Ember Climate (from various sources including the European Environment Agency and EIA)

OurWorldInData.org/energy • CC BY

1. **Carbon dioxide-equivalents (CO<sub>2</sub>eq):** Carbon dioxide is the most important greenhouse gas, but not the only one. To capture all greenhouse gas emissions, researchers express them in 'carbon dioxide-equivalents' (CO<sub>2</sub>eq). This takes all greenhouse gases into account, not just CO<sub>2</sub>. To express all greenhouse gases in carbon dioxide-equivalents (CO<sub>2</sub>eq), each one is weighted by its global warming potential (GWP) value. GWP measures the amount of warming a gas creates compared to CO<sub>2</sub>. CO<sub>2</sub> is given a GWP value of one. If a gas had a GWP of 10 then one kilogram of that gas would generate ten times the warming effect as one kilogram of CO<sub>2</sub>. Carbon dioxide-equivalents are calculated for each gas by multiplying the mass of emissions of a specific greenhouse gas by its GWP factor. This warming can be stated over different timescales. To calculate CO<sub>2</sub>eq over 100 years, we'd multiply each gas by its GWP over a 100-year timescale (GWP100). Total greenhouse gas emissions – measured in CO<sub>2</sub>eq – are then calculated by summing each gas' CO<sub>2</sub>eq value.

Émissions CO2 équivalent (CO2equ)

=

Énergie consommée (kWh) x Intensité carbone

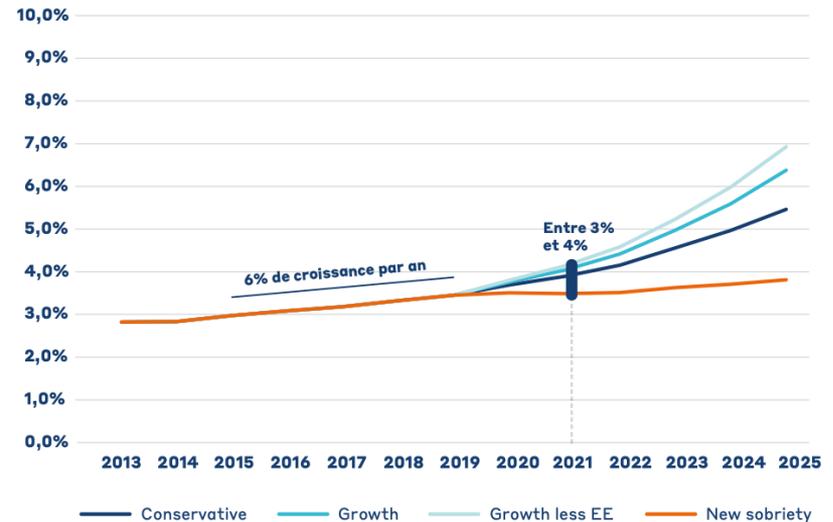
# Impact du numérique

D'après l'agence internationale de l'énergie,  
pour les centres de données :

En 2020

- 300 Millions de tCO<sub>2</sub>eq en 2020
  - Effets directs et indirects
- **0,6%** des émissions CO<sub>2</sub> globales

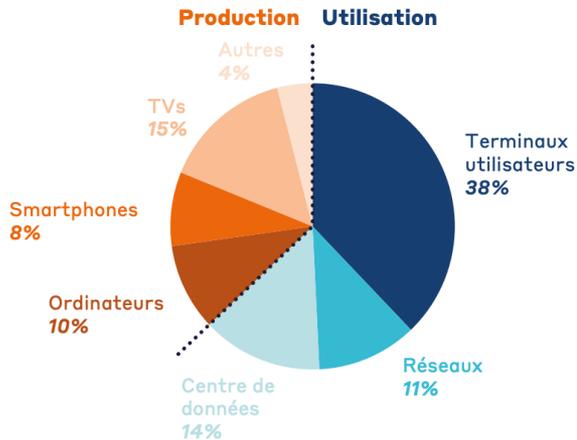
Part du numérique dans les émissions de GES mondiales



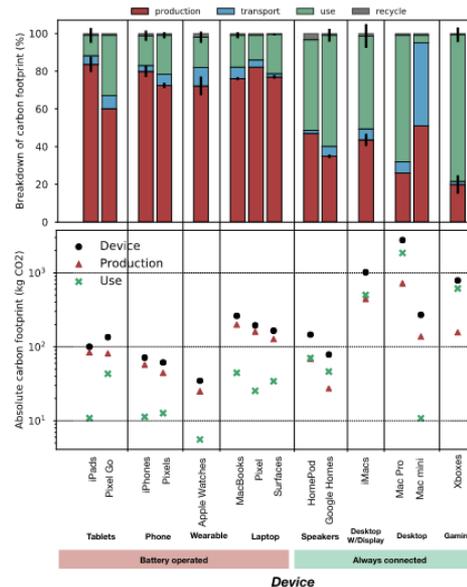
[https://theshiftproject.org/wp-content/uploads/2021/03/Note-danalyse\\_Numerique-et-5G\\_30-mars-2021.pdf](https://theshiftproject.org/wp-content/uploads/2021/03/Note-danalyse_Numerique-et-5G_30-mars-2021.pdf)

# Analyse par cycle de vie

Distribution de l'empreinte carbone du numérique mondial par poste en 2019



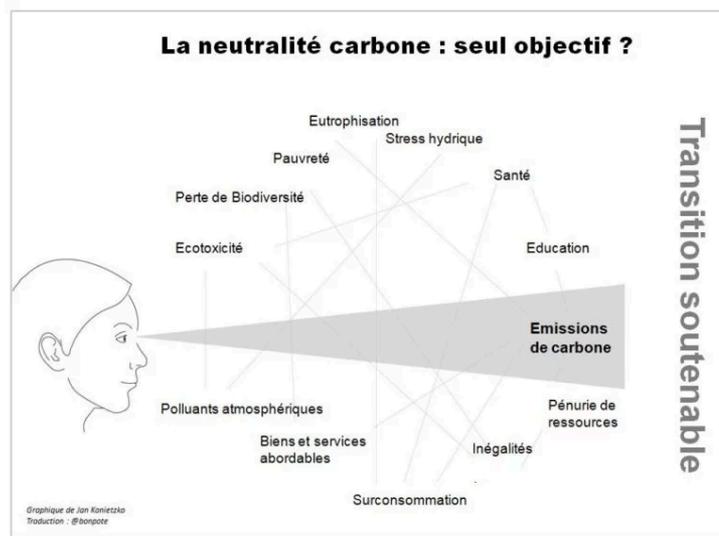
D'après le shift project



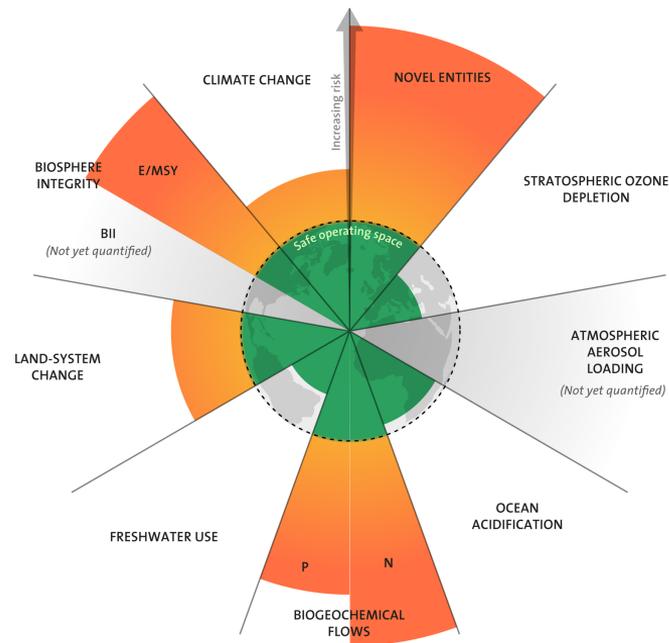
U. Gupta et al., "Chasing Carbon: The Elusive Environmental Footprint of Computing," arXiv:2011.02839 [cs], Oct. 2020, Accessed: Oct. 15, 2021. [Online]. Available: <http://arxiv.org/abs/2011.02839>



# Tout n'est pas mesurable - ou ne doit pas être mesuré



<https://bonpote.com/la-5eme-limite-planetaire-vient-detre-officiellement-franchie-et-tout-le-monde-sen-fout/>



<https://www.stockholmresilience.org/research/research-news/2022-01-18-safe-planetary-boundary-for-pollutants-including-plastics-exceeded-say-researchers.html>

# Monitorer la consommation énergétique d'un centre de données

- Wattmètre / carte de management
- Wattmètres logiciels
- Estimation

# Wattmètres

- PDU (Power Distribution Unit)
- Beaucoup de choix !

Wattmètres internes au server

- BMC (Baseboard management controller)
- PowerMon2, PowerInsight, ...



- + Méthode **fiable** - utilisée depuis longtemps
- + **Compatible** avec n'importe quelle machine

- Mesure la consommation **globale** du server (au niveau de la prise électrique)
- Fréquence et précision de mesure **aléatoire**
- Difficile à **mettre en place** à grande échelle
  - Coût, main d'oeuvre
  - Collection des données à automatiser

# Interfaces internes

+ **Disponibles** sur la quasi-totalité des serveurs récents (Intel, Nvidia, AMD)

+ **Bonne** précision

- Nombreuses études de qualité
- Mise à jour instantanée

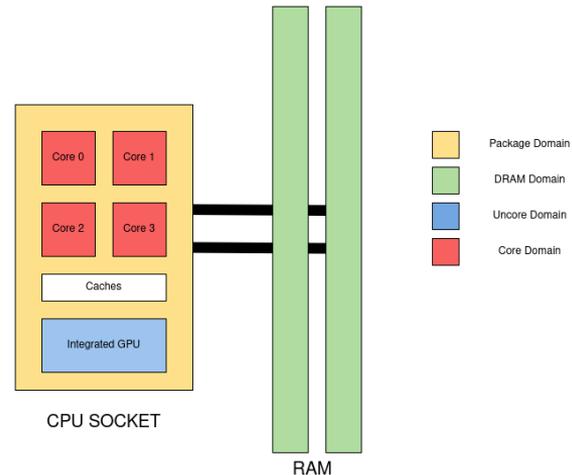
• Donne la consommation **uniquement** du CPU, de la DRAM et du GPU

- Utilisation directe relativement **compliquée**

- Besoin d'accès **privilegiés** à la machine

Beaucoup d'outils sont basés sur ces interfaces et les rendent plus facile à utiliser.

- Intel RAPL (Running Average Power Limit)
  - Pour les CPUs Intel
- Nvidia NVML (NVIDIA Management Library)
  - Pour les GPUs Nvidia



# Modèles de consommation

Estimation à partir des caractéristiques de la tâche effectuée (temps d'exécution, ressources utilisées)

$$\begin{array}{c} \text{Énergie} \\ = \\ \text{TDP} \times \text{Utilisation} \times \text{Temps d'exécution} \end{array}$$

- + **Facile** à comprendre et à utiliser
- + **Compatible** avec n'importe quelle machine
- Bonne précision
  - si l'utilisation des ressources est constante
  - Si bonne connaissance de la tâche
  - Pas de garanties
- Donne la consommation **uniquement** du CPU et du GPU

# Compléments

## Reproductibilité

Pour ne pas perturber l'énergie consommé, il faut être encore plus strict que normal :

- Fixer la fréquence des processeurs
- Attendre que la température soit revenue à la normale avant de relancer les expériences
- Utiliser des images minimalistes

## Cloud computing

Il faut faire un estimation par processus pour isoler les VMs

- De nombreuses méthodes
  - Utilisation
  - Registres
- Pas de source de vérité

# Partie pratique

- Utiliser des outils de monitoring d'énergie
- Visualiser quels composants sont inclus dans les interfaces internes, les BMC, les wattmètres
- Étudier l'impact de la fréquence des composants
- Manipuler des valeurs et faire des calculs d'énergie et de CO2 équivalent