Comparaison expérimentale de logiciels de mesure d'énergie : côté GPU

Mathilde Jay, Denis Trystram, Laurent Lefèvre

▶ **To cite this version:**

Mathilde Jay, Denis Trystram, Laurent Lefèvre. Comparaison expérimentale de logiciels de mesure d'énergie : côté GPU. Compas 2022, Jul 2022, Amiens, France. hal-03945744v1

HAL Id: hal-03945744

https://inria.hal.science/hal-03945744v1

Submitted on 18 Jan 2023 (v1), last revised 20 Mar 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

Comparaison expérimentale de logiciels de mesure d'énergie : côté GPU

Mathilde Jay[1], Laurent Lefevre[2], Denis Trystram[1]
[1]Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, MIAI
[2]Univ Lyon, EnsL, UCBL, CNRS, Inria, LIP

Contexte

Suite aux accords de Paris et à la neutralité carbone à l'horizon 2050, de nombreux développeurs cherchent à estimer l'**impact environnemental** de leur logiciel ou application. Le coût environnemental de l'**intelligence artificielle** (IA) reste peu connu [1][2].

L'entraînement de modèle d'IA demande des calculs intensifs et l'utilisation de **Graphic Processing Units (GPUs)**. Mesurer leur consommation énergétique nécessite des outils adaptés.

Dans cette étude, nous **comparons quantitativement et qualitativement** une sélection de **logiciels de mesure d'énergie** de programmes utilisant des GPUs.

Environnement de test

- Cluster **Gemini** de **Grid5000** [10]
  - 8 GPUs Nvidia Tesla V100-SXM2-32GB (32 Go)
  - 2 CPUs Intel Xeon E5-2698 v4 (Broadwell, 2.20GHz, 20 cœurs/CPU)
  - 512 Go de mémoire
- **Wattmètres / BMC** (Board Management Controller)
  - Mesure directe de la consommation totale de la machine
- **Benchmarks NAS** implémentés en CUDA [9] :
  - **EP** (Embarrassingly Parallel) : GPU
  - **LU** (Lower-Upper Gauss-Seidel solver) : GPU, RAM
  - **MG** (Multi-Grid) : RAM, CPU, GPU

Logiciels testés

- Outils en ligne
  - **GA** : Green Algorithm [7]
  - **MCI** : ML CO2 Impact [8]
- Librairies Python
  - **CC** : Code Carbon [3]
  - **CT** : Carbon Tracker [4]
  - **EIT** : Experiment Impact Tracker [5]
- Logiciels
  - **ES** : Energy Scope [6]

Méthodes d'estimation

Interfaces internes

- Estimation basée sur des mesures et/ou des compteurs de performance
- Calibration à la fabrication du processeur
- **RAPL** (Running Average Power Limit) : CPU Intel
- **NVML** (Nvidia Management Library) : GPU Nvidia

Modèle d'estimation

- **TDP** (Thermal Design Power) : Puissance du processeur à utilisation maximale
- Énergie = TDP x utilisation moyenne x temps d'execution

Classification proposée

	Mesures (références)		Logiciel	Libraries Python			Outils en ligne	
	Wattmètre	**BMC**	**ES**	**CC**	**EIT**	**CT**	**GA**	**MCI**
Développement								
Année			2021	2020	2020	2020	2021	2019
Environnement								
Cible	Machine totale	Machine totale	CPUs, GPUs, Memory	CPUs, GPUs, Memory	CPUs, GPUs, Memory, Processus	CPUs, GPUs, Memory	CPUs, GPUs, Memory	CPUs, GPUs
Compatibilité	Tout	Tout	RAPL, NVML	RAPL, NVML	RAPL, NVML	RAPL, NVML	Tout	Tout
Fonctionnement								
Modèle énergie	Mesure	Mesure	RAPL, NVML	RAPL, NVML	RAPL, NVML, processus	RAPL, NVML	TDP	TDP
Fréquence acquisition (Hz)	50	1/5	10	1/15	1	1/10		
Temps réel	Yes	Yes	Yes	Non	Non	Par epoch	Non	Non
Donnée sortie	Série temporelle	Série temporelle	Série temporelle	Énergie total	Énergie total	Énergie total	Énergie total	Énergie total
Accessibilité								
Disponibilité du code source			Non	Oui	Oui	Oui	Oui	Oui
Diponibilité en ligne	Non	Non	Non	Non	Non	Non	Oui	Oui
Facilité d'installation et d'utilisation	Mauvais	Mauvais	Assez bon	Plutôt bon	Plutôt bon	Bon	Pas d'installation	Pas d'installation
Qualité de la documentation	Bon	Bon	Assez bon	Assez bon	Assez bon	Bon	Assez bon	Assez bon
Configurabilité	Mauvais	Mauvais	Bon	Mauvais	Plutôt bon	Mauvais	Bon	Plutôt bon

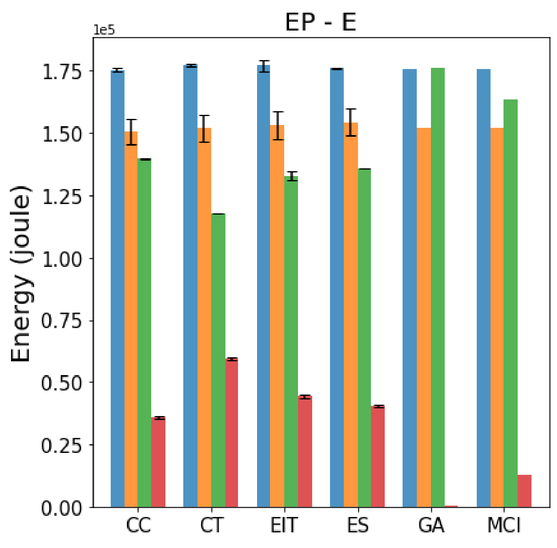
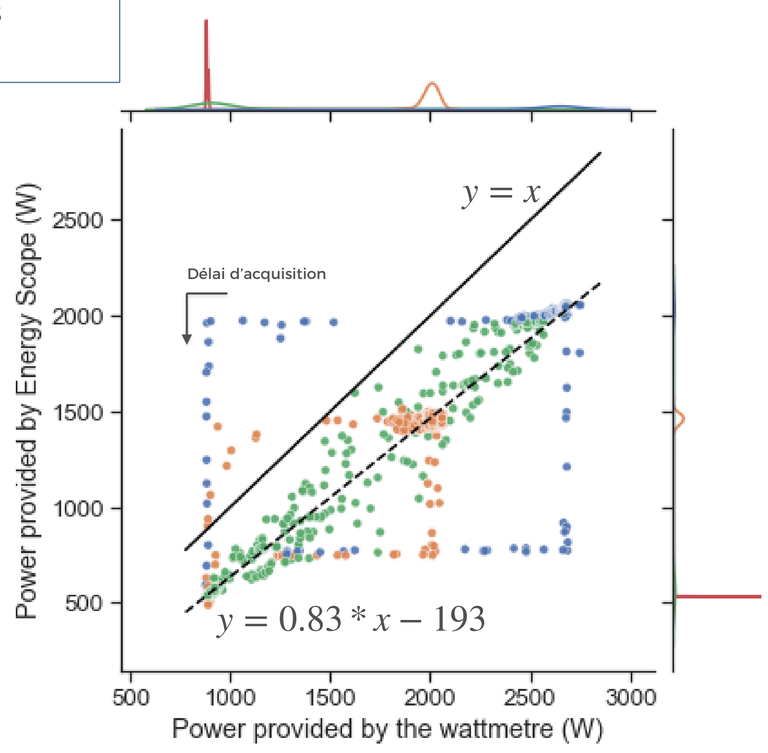


Figure 1. Énergie consommée (10^5 Joule) par les outils testés sur les trois benchmarks EP, LU, MG en comparaison avec les appareils de mesure.



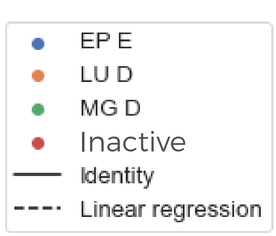


Figure 2. Rapport entre la puissance fournie par Energy Scope et celle mesurée par un wattmètre pour les trois benchmarks EP, LU, MG.

Analyse experimentale

Outils en ligne

- Conditions pour estimation correcte
  - Connaître le matériel et leur utilisation moyenne
  - Avoir une charge stable
- ML CO2 Impact est **moins précis** que Green Algorithm (Figure 1)

Librairies Python

- Code Carbon et Energy Scope : **écarts** les plus faibles (Figure 1)
- Experiment Impact Tracker
  - Isole le processus du programme pour une estimation plus précise
  - **Variabilité** quand intensif en mémoire

Logiciels

- Focus sur **Energy scope** (Figure 2)
  - Bonne corrélation avec le wattmètre : 0.98
  - L'écart semble augmenter avec la puissance mesurée

Surcharge en énergie due aux outils **inférieure à 4%**

Contact

Mathilde Jay
- Équipe DataMove, LIG, UGA
- Équipe Avalon, LIP, ENS Lyon
- Chaire Edge Intelligence, MIAI

Email: mathilde.jay@univ-grenoble-alpes.fr

Collaboration avec Vladimir Ostapenco

(1) Strubell, Emma, Ananya Ganesh, and Andrew McCallum. "Energy and Policy Considerations for Deep Learning in NLP." ArXiv:1906.02243 [Cs], June 5, 2019. http://arxiv.org/abs/1906.02243.
(2) Bannour, Nesrine, Sahar Ghannay, Aurélie Névéol, and Anne-Laure Ligozat. "Evaluating the Carbon Footprint of NLP Methods: A Survey and Analysis of Existing Tools." In Proceedings of the Second Workshop on Simple and Efficient Natural Language Processing, 11–21. Virtual: Association for Computational Linguistics, 2021. https://aclanthology.org/2021.sustainlp-1.2.
(3) Victor Schmidt, Kamal Goyal, Aditya Joshi, Boris Feld, Liam Conell, Nikolas Laskaris, Doug Blank, Jonathan Wilson, Sorelle Friedler, and Sasha Luccioni. "CodeCarbon: Estimate and Track Carbon Emissions from Machine Learning Computing." Zenodo, 2021. https://github.com/mlco2/codecarbon.
(4) Anthony, Lasse F. Wolff, Benjamin Kanding, and Raghavendra Selvan. "Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models." ArXiv:2007.03051 [Cs, Eess, Stat], July 6, 2020. http://arxiv.org/abs/2007.03051.
(5) Henderson, Peter, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. "Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning." ArXiv:2002.05651 [Cs], January 31, 2020. http://arxiv.org/abs/2002.05651.
(6) Mathieu, Hervé. Energy Scope, 2020. https://sed-bso.gitlabpages.inria.fr/datacenter/energy_scope.html.
(7) Lannelongue, Loïc, Jason Grealey, and Michael Inouye. "Green Algorithms: Quantifying the Carbon Footprint of Computation." Advanced Science 8, no. 12 (2021): 2100707. https://doi.org/10.1002/advs.202100707.
(8) ML CO2 IMPACT, n.d. https://mlco2.github.io/impact/?#compute.
(9) Araujo, Gabriell, Dalvan Griebler, Dinei A. Rockenbach, Marco Danelutto, and Luiz G. Fernandes. "NAS Parallel Benchmarks with CUDA and Beyond." Software: Practice and Experience n/a, no. n/a. Accessed February 4, 2022. https://doi.org/10.1002/spe.3056.
(10) Balouek, Daniel, Alexandra Carpen-Amarie, Ghislain Charrier, Frédéric Desprez, Emmanuel Jeannot, Emmanuel Jeanvoine, Adrien Lebre, et al. "Adding Virtualization Capabilities to Grid'5000," n.d., 24.

© Poster Template by Genigraphics® 1.800.790.4001 www.genigraphics.com