Using ML training computations for grid stability in 2050

Scott Evans
Machine Learning Lab
GE Research
Niskayuna, NY, USA
evans@ge.com

Tapan Shah
Machine Learning Lab
GE Research
San Ramon, CA, USA
tapan.shah@ge.com

I. Introduction

In a study published by French power management [1], a zero-carbon regime will be achieved by reducing energy consumption by 40% and multiplying the energy production from renewable sources by 4. Increased integration of renewable sources will lead to an increased vulnerability of the power grids thereby leading to increased occurrences of local and global blackouts which can have disastrous social and economic consequence [7]. The dynamic nature of wind and solar energy means that the grid can either be under-utilized or over-utilized, the later can lead to outages and thermal damage of grid components.

In order to stabilize the grid, there has been a push to controllable loads such as battery storage [8], vehicle-to-grid [8] and hydrogen production [9]. In this paper, we propose the use of ML training jobs as a controllable load to stabilize grid.

There has been prior work in using compute-centers as controllable load for efficient grid operations [10], [4] where the demand response requirements are met by taking advantage of variable power consumption in compute-centers. However, our main idea focuses on using a specific type of compute, machine learning (ML), to act as controllable loads. As illustrated in Figure 1, there has been a rapid growth in the amount of compute required by training tasks of machine learning. For example, the demand for ML training has increase by 150% per year at Facebook [6].

We propose that the increasing popularity as well as specific properties of ML, especially deep learning training, makes it an ideal candidate as controllable loads.

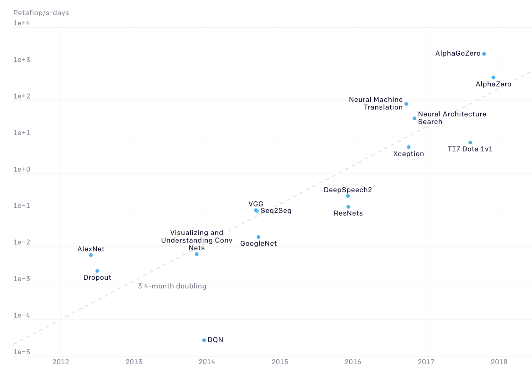

Fig. 1: The total amount of compute, in petaflop/s-days to train certain landmark models. This figure was reproduced from [2]. A petaflop/s-day (pfs-day) consists of performing 10^{15} neural net operations per second for one day, or a total of about $a0^{20}$ operations.

II. ML training for grid stability: An information theoretic perspective

Many researchers consider machine learning as compression. In fact, in [5], the author expresses "Why unify information theory and machine learning? Because they are two sides of the same coin.". In another paper [3], the authors introduce the concept of "Entropy Economy" which expresses machine learning as information work, drawing a parallel between other kinds of thermodynamic work. Using this as a backdrop, we compare ML training (information work) against other competitors like battery storage and hydrogen conversion.

Battery storage converts excess electrical energy into chemical energy. Hydrogen converter converts excess electrical energy into nuclear energy. On the other hand, ML training converts electrical energy into information. Each of the three entails entropy transfer. The concept is further illustrated in [3].

III. How will this work?

Though this work is at a very nascent stage, we explain two scenarios which illustrates our proposed idea. In Figure 2, substations are represented by circle while compute-centers are represented by squares. The colors red, yellow and green indicate high, medium and low loads, respectively. Black lines indicate electrical flow while blue lines indicate data flow. Thick lines represent high transfer (electrical or data) and thin lines represent low transfer. Below are 2 illustrative examples.

At time 1, the grid state is such that substation 2 is heavily loaded while substation 1 has light load and substation 3 has medium load. At time 2, our proposed idea will stabilize the grid by moving ML training jobs from compute-centers B to compute-centers A and C to reduce the energy burden at substation 2. Moving these jobs stabilizes the grid by levelling the load among the substations

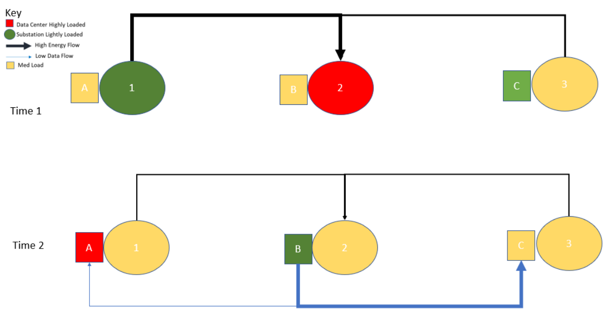

Fig. 2: Grid Stabilization through moving ML jobs.

In another scenario shown in Figure 3, rather than moving load, we modulate the quality of the ML training to stabilize the grid. At time 1, the compute-centers are executing their ML jobs to achieve maximum quality. In order to reduce the load on substation 1, the ML training at compute-centers 2 is performed at a lower quality thereby stabilizing the grid.

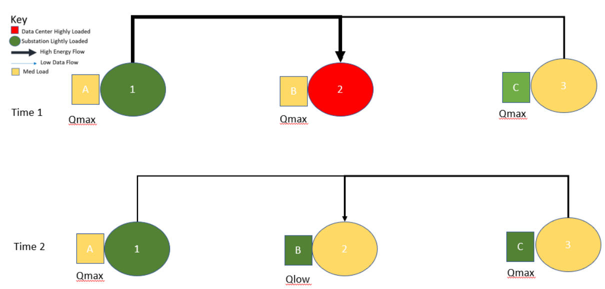

Fig. 3: Grid Stabilization by modulating quality of ML jobs.

A. Conclusion

In the above section, we illustrated using very simple examples of how ML jobs can provide the flexibility to act as controllable loads. The key advantages of the proposed idea are as follows.

1) Transferring information bits is more efficient as compared to transferring electrical power to achieve grid stability. In fact, as an added benefit, it was shown in [11], that load migration within the existing data center capacity has the potential to reduce 113–239 KtCO2e per year of GHG emissions.
2) Machine learning as a computational load has multiple tunable knobs to work in a dynamic environment. For example, a deep learning model can be split and trained at multiple locations, trained at different levels of accuracy by changing the number of epochs etc.
3) Batteries suffer from continual charge leakage which reduces the efficiency of energy conversion. On the other hand, machine learning models, once built, can be used repetitively until it requires an update.

As part of next steps, we intend to work with a grid operator to design an optimization framework which can suitably modulate the machine learning loads to stabilize the grid.

References

[1] Futurs énergétiques 2050 : les scénarios de mix de production à l'étude permettant d'atteindre la neutralité carbone à l'horizon 2050.

[2] AI and Compute, May 2018.

[3] Scott Evans, Tapan Shah, and Acahlesh Pandey. The Entropy Economy: A New Paradigm for Carbon Reduction and Energy Efficiency for the Age of AI. In *2021 55th Asilomar Conference on Signals, Systems, and Computers*, pages 1372–1377, October 2021. ISSN: 2576-2303.

[4] Peicong Luo, Xiaoying Wang, Hailong Jin, Yuling Li, and Xuejiao Yang. Smart-Grid-Aware Load Regulation of Multiple Datacenters towards the Variable Generation of Renewable Energy. *Applied Sciences*, 9(3):518, January 2019. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[5] David J C MacKay. Information Theory, Inference, and Learning Algorithms. page 640.

[6] Maxim Naumov, John Kim, Dheevatsa Mudigere, Srinivas Sridharan, Xiaodong Wang, Whitney Zhao, Serhat Yilmaz, Changkyu Kim, Hector Yuen, Mustafa Ozdal, Krishnakumar Nair, Isabel Gao, Bor-Yiing Su, Jiyan Yang, and Mikhail Smelyanskiy. Deep Learning Training in Facebook Data Centers: Design of Scale-up and Scale-out Systems, August 2020. arXiv:2003.09518 [cs].

[7] Gaëtan Serré, Eva Boguslawski, Benjamin Donnot, Adrien Pavão, Isabelle Guyon, and Antoine Marot. Reinforcement learning for Energies of the future and carbon neutrality: a Challenge Design, July 2022. arXiv:2207.10330 [cs].

[8] Jingshuang Shen, Chuanwen Jiang, and Bosong Li. Controllable Load Management Approaches in Smart Grids. *Energies*, 8(10):11187–11202, October 2015. Number: 10 Publisher: Multidisciplinary Digital Publishing Institute.

[9] Dai Wang, Matteo Muratori, Joshua Eichman, Max Wei, Samveg Saxena, and Cong Zhang. Quantifying the flexibility of hydrogen production systems to support large-scale renewable energy integration. *Journal of Power Sources*, 399:383–391, September 2018.

[10] Rui Wang, Nagarajan Kandasamy, Chika Nwankpa, and David R. Kaeli. Datacenters as Controllable Load Resources in the Electricity Market. In *2013 IEEE 33rd International Conference on Distributed Computing Systems*, pages 176–185, July 2013. ISSN: 1063-6927.

[11] Jiajia Zheng, Andrew A. Chien, and Sangwon Suh. Mitigating Curtailment and Carbon Emissions through Load Migration between Data Centers. *Joule*, 4(10):2208–2222, October 2020. Publisher: Elsevier.