NEURAL DATA COMPRESSION FOR PHYSICS PLASMA SIMULATION

Jong Choi, Qian Gong, David Pugmire & Scott Klasky
Oak Ridge National Laboratory, Oak Ridge, TN, USA

Michael Churchill, Seung-Hoe Ku & CS Chang
Princeton Plasma Physics Laboratory, Princeton, NJ, USA

Jaemoon Lee, Anand Rangarajan & Sanjay Ranka
Department of Computer Information Science and Engineering
University of Florida, Gainesville, FL, USA

ABSTRACT

We present a VAE-based data compression method, called VAe Physics Optimized Reduction (VAPOR), to compress scientific data while preserving physics constraints. VAPOR is based on Vector Quantized Variational Auto Encoder (VQ-VAE) and extended with physics-informed optimization functions and refinement layers, focusing on compressing and reconstructing scientific data with minimum loss of information under physics constraints. We demonstrate VAPOR by using outputs from XGC, a massively parallel fusion simulation code running on the largest supercomputers. Key features of VAPOR are three-fold; i) find a reduced representation of physics data, ii) reconstruct the data with a minimum loss, iii) preserve physics information (e.g., mass, energy, moment conservation).

We discuss challenges in XGC 5D data reconstruction and present our initial experiences and results on how we construct Deep Neural Network (DNN) of VAPOR to optimize the reconstruction quality of XGC data and integrate XGC's physics constraints.

1 INTRODUCTION

As data volume grows at an exceeding rate, several floating-point lossy data compressors, such as ZFP (Lindstrom, 2014), SZ (Di & Cappello, 2016), and MGARD (Ainsworth et al., 2019), have been actively developed and applied in many science applications. Recently, researchers have started looking at deep learning-based methods for data compression as well. While most lossy compression methods are broadly based on numerical solutions for regression, interpolation, and decompositions, data compression with deep learning is mostly based on developing a generative process for a given data set. A family of generative deep learning models, such as Variational Autoencoders (VAE) (Kingma & Welling, 2013) and Generative Adversarial Network (GAN) (Goodfellow et al., 2014), has been developed to reconstruct whole images or augment missing parts (Welander et al., 2018; Yoon et al., 2018; Ganguli et al., 2019).

We have been developing a VAE-based data compression method, called VAe Physics Optimized Reduction (VAPOR), with a scientific dataset from XGC, a fusion simulation code (Chang & Ku, 2008; Ku et al., 2009). VAPOR is based on Vector Quantized Variational Auto Encoder (VQ-VAE) (Razavi et al., 2019), focusing on compressing XGC's five-dimensional (5D) particle distribution data as well as preserving physics constraints. Key features of VAPOR are three-fold; i) find a reduced representation of physics data, ii) reconstruct the data with a minimum loss, iii) preserve physics information (e.g., mass, energy, moment conservation).

We will discuss challenges in the data reconstruction of the XGC 5D particle distribution function data and present our initial experiences and results on how we construct Deep Neural Network

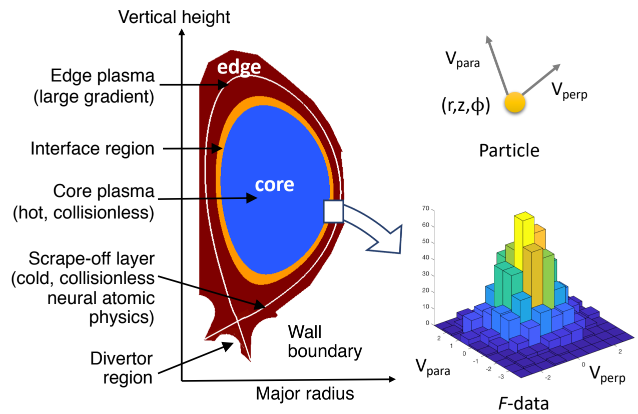


Figure 1: XGC, a massively parallel fusion simulation code, runs on the largest supercomputers around the world. It is based on a particle-in-cell approach. Regularly writing particle distribution data, called F-data, ranging from a few GBs up to TBs, is challenging due to the I/O bottleneck problem in many high-performance computing systems.

(DNN) for VAPOR to optimize the reconstruction quality of XGC 5D data and integrate XGC's physics constraints, and share performance results.

Related work: Multiple GAN-based work, such as SRGAN (Nagano & Kikuta, 2018) and ESRGAN (Wang et al., 2018), focus on generating photo-realistic, high-resolution synthetic data. In contrast, we focus on reconstructing scientific data based on physics quantities.

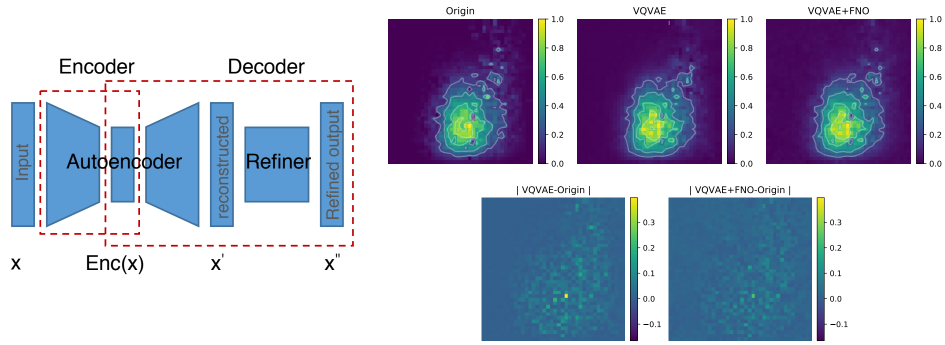
Adversarial super-resolution for climate data (Stengel et al., 2020) explores generating high-resolution climate data based on multi-level generative adversarial networks. The work focuses on generating scientific data by preserving probability densities around, such as, kinetic energy spectrum, velocity gradients, etc. Our work focus on data compression and reduction for fusion data with fusion physics constraints.

2 XGC AND PARTICLE DISTRIBUTION DATA

XGC (Chang & Ku, 2008; Ku et al., 2009) is a gyrokinetic, particle-in-cell code used to simulate fusion reactor designs, focusing on the multi-scale physics at the edge of the fusion plasmas. XGC is massively parallel, requiring supercomputers such as Summit in the Oak Ridge National Laboratory to run, and generates petabytes of data for large-scale runs. Virtual particles are time-advanced or "pushed" according to governing gyrokinetic equations of motion, and then interpolated or "gathered" to discrete mesh points, for solving electric and magnetic field equations which factor into the particle motion in subsequent steps. Particles are regularly histogrammed to form particle distribution functions, F, which is five-dimensional (2D velocity, 3D space) (Figure. 1) and ranges in size from a few GBs up to a few TBs per each iteration in a large run. Due to I/O bottlenecks on the supercomputer, often various reductions are applied to F to form physically relevant quantities to output, such as density (n) and temperature (T), instead of writing out the entire F data. But because these always involve a loss of information, methods to quickly and efficiently reduce or compress this F data (and in the future the particle data itself) are desireable, to enable scientists to capture richer physics from these simulations.

3 VAPOR

To address the output bottleneck challenge in XGC's F-data writing, we consider an auto-encoder-based lossy data compression approach where we can build a data-driven generative model to reconstruct the original data with customized error criteria. One of the key ideas is how to build a data-driven learning process to find the generative process of XGC's F-data and develop error criteria to meet XGC's physics constraints.



(a) The VAPOR architecture.

(b) An example output.

Figure 2: (a) VAPOR is based on a variational auto-encoder (VAE) architecture attached with a refinement layer to overcome VAE's generalization (blurring) effect. We save only the encoded data ($\mathrm{Enc}(\mathrm{x})$) for compression. Then, for reconstruction, we apply the decoder followed by the refiner. The objective is to minimize the differences between the input x and the reconstructed output x'' as well as the loss in physics constraints. We use a deterministic version of VAE, VQ-VAE, and the Fourier Neural Operator (FNO) for the refiner. (b) An example shows how VQVAE can regenerate an XGC output, and the FNO refiner makes further improvement closer to the original data.

We have been developing a VAE-based data compression method called VAe Physics Optimized Reduction (VAPOR), aiming at learning nontrivial data distribution in an unsupervised fashion and creating a DNN encoder to find reduced or compressed data representations. In VAPOR, we extend VAE with custom error functions and refinement processes to obtain the finer reconstructed output. Figure 2 shows the VAPOR's neural network architecture and an example output. In the following, we discuss our design decisions.

Vector quantization VAE: Using Gaussian prior distributions followed by Gaussian reparameterization is a well-known practice in VAE. However, recent reports (van den Oord et al., 2017; Razavi et al., 2019) demonstrate improved quality based on vector quantization (VQ). The deterministic nature of the VQ method is also favorable in data compression-reconstruction for its reproducibility.

Loss function: We devise flexible, physics-informed loss functions, improving the accuracy of reconstructed data and maintaining physics constraints XGC imposed in the reconstruction. In short, we define the following composite loss function for training:

$$\mathcal{L} = \alpha\mathcal{L}_{\text{recon}} + \beta\mathcal{L}_{\text{VQ}} + \gamma\mathcal{L}_{\text{physics}} \tag{1}$$

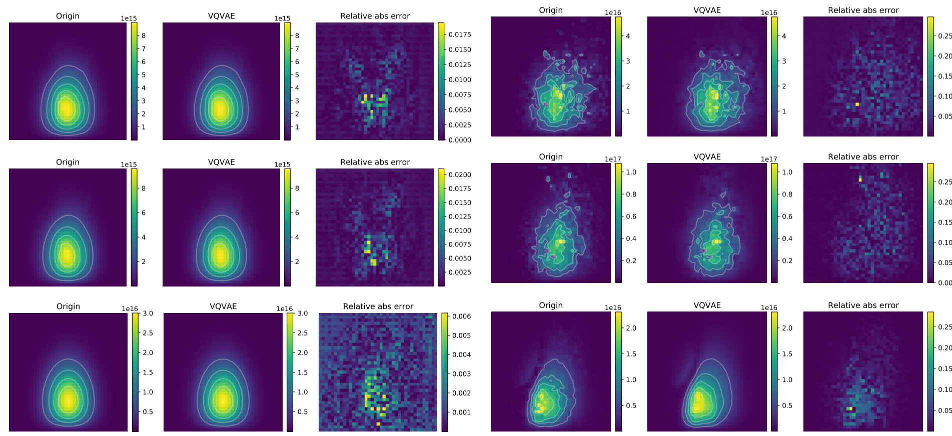
where $\mathcal{L}_{\text{recon}}$ represents reconstruction errors based on the mean squared error measurement, $\mathcal{L}_{\text{VQ}}$ represents the loss in the vector quantization, and $\mathcal{L}_{\text{physics}}$ represents physics quantity loss. We want to preserve mass, momentum, and energy derived from F-data after reconstruction. More details can be found in Ku et al. (2009); Miller et al. (2021). α, β, and γ are the weight control parameters.

Refinement: We observe VAE's generalization (blurring) effect. Inspired by super- and multi-level resolution approaches in (Ledig et al., 2017; Stengel et al., 2020), we add a refiner layer to improve point-wise accuracy. After testing with a few candidates, such as GAN and SR-GAN (Nagano & Kikuta, 2018), we find the Fourier Neural Operator (FNO) (Li et al., 2020) works well with XGC data. Fourier transformation helps to improve recovering detailed features in the data. A more detailed quantitative/qualitative study is scheduled for the next work.

4 Experiment results

We demonstrate the performance of VAPOR. We use about 131,160 examples of XGC F-data from a DIII-D[1] simulation. The full simulation consists of over 600 timesteps. We choose a single

[1]DIII-D is an experimental fusion device operated by General Atomics: `https://www.ga.com/magnetic-fusion/diii-d`



(a) Well-trained (b) Inferior

Figure 3: Examples of well-trained and inferior cases generated from VAPOR. The left-most column images are original, and the middle images show the reconstructed output from VAPOR. The right column images display the relative absolute error between the two.

Table 1: Data size and compression data size.

Type	Original Size	Raw Encoded Data	Compressed Encoded data*	Compressed Model Size*	ZFP compression
XGC F-data	1.6 GB	200 MB	54 MB	-	1.3 GB
VAPOR Model**	98 MB	-	-	91 MB	-

* We apply Gzip to compress further the raw encoded data and the VAPOR model itself.

** VAPOR model will be reusable as it integrated into the compression/reconstruction software.

timestep where F-data shows turbulent behavior, making it more challenging to compress. Each F-data is a 2D histogram array of sizes of 39-by-39. We designed VAPOR and FNO with the number of layers of 41 and 21 respectively, which contains about 1.8 and 2.4 million parameters to optimize. In Table 1, we summarize the size of the XGC F-data, the size of encoded data generated by VAPOR, and the size of VAPOR model itself. As common in many lossy compression methods (such as SZ and MGARD), we demonstrate how we can further reduce the size of encoded data and VAPOR model parameters by using Gzip. We achieved 8x reduction with VAPOR encoding and about 29.4x reduction after compressing the encoded data with Gzip. To compare the size by using other lossy compression method, we chose ZFP and observed only 1.2x compression ratio. More detailed comparison study remains a future work.

Figure 3 shows a few examples showing the best and worst cases generated by VAPOR. One of the challenges in compressing XGC F-data is to maintain detailed features after reconstruction. While VAPOR works well for the F-data in the core area shown in Figure 3(a), F-data near the edge area (Figure 3(b)) are challenging due to its complex features.

5 Conclusion

We present a VAE-based data compression method, called VAe Physics Optimized Reduction (VAPOR), to compress scientific data while preserving physics constraints. Based on Vector Quantized Variational Auto Encoder (VQ-VAE), we extend VAPOR with custom optimization functions to integrate physics constraints. In addition, we add an extra refinement layer based on Fourier Neural Operator (FNO) to improve point-wise accuracy by overcomeing the blurring generalization effect common in GAN/VAE-based methods.

We discuss design decisions in VAPOR and demonstrate VAPOR by using real-world scientific outputs from XGC, a massively parallel fusion simulation code running on DOE supercomputers.

REFERENCES

Mark Ainsworth, Ozan Tugluk, Ben Whitney, and Scott Klasky. Multilevel techniques for compression and reduction of scientific data-quantitative control of accuracy in derived quantities. *SIAM Journal on Scientific Computing*, 41(4):A2146–A2171, 2019.

CS Chang and S Ku. Spontaneous rotation sources in a quiescent tokamak edge plasma. *Physics of Plasmas (1994-present)*, 15(6):062510, 2008.

Sheng Di and Franck Cappello. Fast error-bounded lossy hpc data compression with sz. In *2016 ieee international parallel and distributed processing symposium (ipdps)*, pp. 730–739. IEEE, 2016.

Swetava Ganguli, Pedro Garzon, and Noa Glaser. Geogan: A conditional gan with reconstruction and style loss to generate standard layer of maps from satellite images. *arXiv preprint arXiv:1902.05611*, 2019.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

S Ku, CS Chang, and PH Diamond. Full-f gyrokinetic particle simulation of centrally heated global itg turbulence from magnetic axis to edge pedestal top in a realistic tokamak geometry. *Nuclear Fusion*, 49(11):115021, 2009.

Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Peter Lindstrom. Fixed-rate compressed floating-point arrays. *IEEE transactions on visualization and computer graphics*, 20(12):2674–2683, 2014.

Marco Andres Miller, Randy Michael Churchill, A Dener, CS Chang, T Munson, and R Hager. Encoder–decoder neural network for solving the nonlinear fokker–planck–landau collision operator in xgc. *Journal of Plasma Physics*, 87(2), 2021.

Yudai Nagano and Yohei Kikuta. Srgan for super-resolving low-resolution food images. In *Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management*, pp. 33–37, 2018.

Ali Razavi, Aaron van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with vq-vae-2. *arXiv preprint arXiv:1906.00446*, 2019.

Karen Stengel, Andrew Glaws, Dylan Hettinger, and Ryan N King. Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29): 16805–16815, 2020.

Aaron van den Oord, Oriol Vinyals, et al. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, pp. 6306–6315, 2017.

Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 0–0, 2018.

Per Welander, Simon Karlsson, and Anders Eklund. Generative adversarial networks for image-to-image translation on multi-contrast mr images-a comparison of cyclegan and unit. *arXiv preprint arXiv:1806.07777*, 2018.

Jinsung Yoon, James Jordon, and Mihaela Schaar. Gain: Missing data imputation using generative adversarial nets. In *International Conference on Machine Learning*, pp. 5689–5698. PMLR, 2018.