Scientific Machine Learning and HPC-AI technology convergence

Atos

Some of the most well-known examples of the use of machine learning technics in science applications are the detection and classification of gravitational-waves signals from LIGO and Virgo in astrophysics[1], the recent DeepMind Alpha-Fold2 capabilities outperforming classical methods in protein folding[2] or the winning team of the Gordon Bell 2020 with the Deep Potential Molecular Dynamics[3] which is opening new breakthroughs in the drug design process and could speed up future pandemic response efforts. Beyond these key examples, the convergence between HPC and AI is natural where DL-based surrogate modelling is more and more widely applied in research and recent advances in physics-informed neural networks such as HNN[4] bring physical properties and constraints to neural networks loss functions opening a great path towards a new generation of simulation.

Atos Centers of Excellence – combining data sciences and industry expertise

In Atos, we built a dedicated approach to support the scientific community and Industries by bringing data science and HPC expertise through the Atos Centers of Excellence. Each center is oriented towards a specific domain where our experts and our customers can jointly bring innovations and technologies with the support of some of our partners. Some of the first Atos Centers of Excellence are dedicated to weather forecast & climate changes[5] and life sciences[6]. In such advanced domains, applying innovation means creating a strong AI research support, defined through specific programs.

Those programs aim to enhance the state of the art of machine learning model for science applications as well as exploring coupling capabilities between simulation and AI model inference including the orchestration of AI augmented HPC workflows to brings such these development in production at scale. Some of the use cases we are addressing are about surrogating modelling and dimensionality reduction on CFD applications or data assimilation and deep learning for chaotic systems.
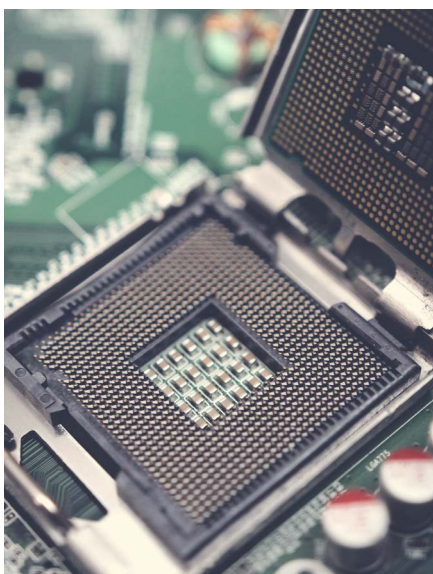
AI for science applications – more options to come!

In addition to this approach, AI into science applications also means adapting technical architectures of such converged systems. Today most of the HPC applications benefit from SIMD acceleration through GPU technologies, acceleration that AI benefit as well. GPU architectures evolved over the time by increasing the part of their silicon dedicated to AI processing to boost this convergence[7] but the maturity of dedicated AI chips such as the Graphcore IPU or Intel Habana is a new potential keystone for science thanks to the performance gap these technologies could provide.

Nevertheless, the current adoption limitation of dedicated AI technologies is due to the low level of maturity of code hybridization between AI and HPC application which is not sufficient enough to embed dedicated AI acceleration. Hardware acceleration for HPC platforms implies support both HPC applications and AI capacities which is today a key limiting factor for the wide adoption of dedicated AI chips due to the associated software stack not able to naturally support legacy codes with associated programming model (MPI/OpenMP) as AI chipmakers are building their software environment on AI frameworks with dedicated back-end compilers and runtimes. Despite this challenge, promising examples of using AI chips for HPC applications have been recently published, for example on Graphcore IPU but at the cost of an algorithm reformulation or a kernel focus which naturally exposed a tensor representation in the code to be accelerated by dedicated hardware[8].

Large-scale AI infrastructure brings competitive advantage for HPC-AI convergence

HPC-AI convergence is also often seen via the use of HPC knowledge to strengthen AI. As deep learning model complexity continue to growth over years, especially related to natural language processing where the computational costs of transformers models such as GPT-3 gathers 15B parameters network to be trained and a training computing cost around ~1023Flops[9-10], dedicated large scale AI infrastructures becomes crucial for enterprise competitiveness. Architecture design differs from HPC-AI converged platforms by a different tradeoff between general purpose processing and AI hardware acceleration oriented towards more AI acceleration with an increased interconnect capabilities per node.

Turnkey solution – putting AI solution at your fingertips!

In Atos, we bring decades of experience in large scale computing to make available to any industries scalable AI solutions to support AI business growth and reach a new level of performance with AI turnkey solutions integrating AI technologies at scale, management software, data science platform, AI and HPC knowledge in addition of ML-based software to enhance HPC operations addressing data movement, energy consumption, resources allocation and predictive maintenance to globally enhance HPC operations excellence and total cost of ownership by learning behaviors from each production system.

To conclude, artificial intelligence applied to science application is a promising field of research that will bring great scientific breakthroughs in the future and the Atos HPC-AI team is proud to support this lead-edge research topic.

Dr Cédric Bourrasset
Atos AI Distinguished Expert
ATOS BDS HPC - AI Sales Operations
& AI Product Manager
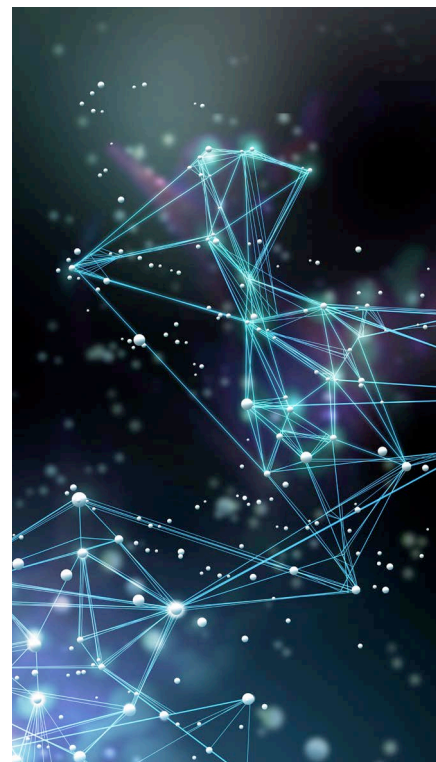
References:

[1] Enhancing gravitational-wave science with machine learning, Elena Cuoco et al - https://iopscience.iop.org/article/10.1088/2632-2153/abb93a

[2] AlphaFold: a solution to a 50-year-old grand challenge in biology, Deepmind Blog, https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology

[3] GordonBell 2020: "Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning." https://awards.acm.org/bell

[4] Hamiltonian Neural Networks - Sam Greydanus, Misko Dzamba, Jason Yosinski - https://arxiv.org/abs/1906.01563

[5] Atos and ECMWF launch Center of Excellence in Weather & Climate Modelling to support researchers with HPC, AI and quantum capabilities. https://atos.net/en/2020/press-release_2020_10_05/atos-and-ecmwf-launch-center-of-excellence-in-weather-climate-modelling-to-support-researchers-with-hpc-ai-and-quantum-capabilities

[6] Atos establishes Life Sciences Centre of Excellence at the Wellcome Genome Campus https://atos.net/en-gb/lp/lifesciences-coe

[7] NVIDIA TensorCore - https://www.nvidia.com/fr-fr/data-center/tensor-cores/

[8] Using the Graphcore IPU for traditional HPC applications, Thorben Louw, Simon McIntosh-Smith, Dept. of Computer Science University of Bristol Bristol, United Kingdom https://hubs.li/H0Fc8Z50

[9] The cost if training NLP Models - A concise overview - https://arxiv.org/pdf/2004.08900.pdf

[10] OpenAI's GPT-3 Language Model: A Technical Overview - https://lambdalabs.com/blog/demystifying-gpt-3/

About Atos

Atos is a global leader in digital transformation with 105,000 employees and annual revenue of over € 11 billion. European number one in cybersecurity, cloud and high performance computing, the Group provides tailored end-to-end solutions for all industries in 71 countries. A pioneer in decarbonization services and products, Atos is committed to a secure and decarbonized digital for its clients. Atos operates under the brands Atos and Atos|Syntel. Atos is a SE (Societas Europaea), listed on the CAC40 Paris stock index.

The **purpose of Atos** is to help design the future of the information space. Its expertise and services support the development of knowledge, education and research in a multicultural approach and contribute to the development of scientific and technological excellence. Across the world, the Group enables its customers and employees, and members of societies at large to live, work and develop sustainably, in a safe and secure information space.

Find out more about us
atos.net
atos.net/career

Let's start a discussion together

For more information: hpc@atos.net