Roadmap to Robust Science for High-throughput Applications: The Scientists' Perspective

M. Taufer[1], E. Deelman[2], R. Ferreira da Silva[2], T. Estrada[3], M. Hall[4]
[1]U. Tennessee Knoxville, [2]U. Southern California, [3]U. New Mexico, [4]U. Utah

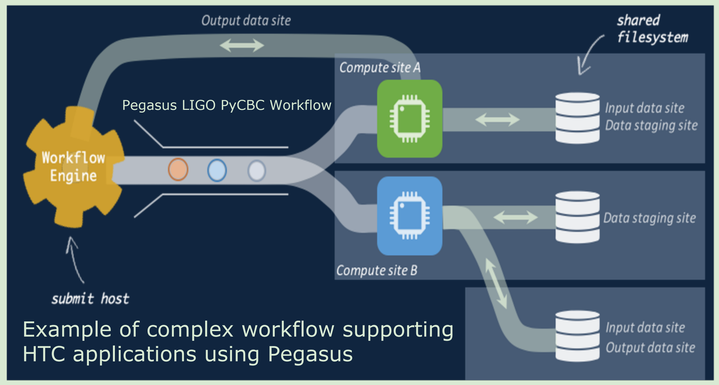
Our Mission

- Our **mission** is to **define a roadmap** for establishing a vibrant, multidisciplinary community that works together to design, implement, and use a set of solutions **for robust science applied to high-throughput applications**
- High throughput computing (HTC) delivers a large amount of computing capacity over a long period of time to accomplish a scientific or engineering task [1]
- **Data-driven applications** and **scientific simulations** are amenable to high throughput computing (HTC) thanks to being easily divided into digestible chunks for concurrent processing
- A **roadmap** defines future hardware architecture and software systems (e.g., tools, interfaces, libraries, data and model commons); programming models and compilers; algorithms and theory; principles and practices; workforce development and diversity

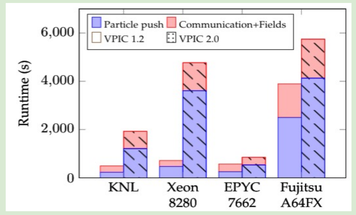
[1] https://research.cs.wisc.edu/htcondor/htc.html

Taxonomy of High-throughput Applications

High-throughput applications include **data analysis + computing**



Deelman et al. Computing in Science Engineering, vol. 21, no. 4, pp.22–36, 2019



Code portability across platforms may come with a **performance portability costs**

← Comparison of performance for CPU-based platforms, including chips from AMD, Intel, and ARM/Fujitsu. Results are shown for both VPIC 1.2 (a version of the code was optimized for the specific platform) vs. VPIC 2.0 (a single version of the code was made portable across platforms with Kokkos)

Robert Bird, Nigel Tan, Scott V Luedtke, Stephen Harrell, Michela Taufer, and Brian Albright. VPIC 2.0: Next Generation Particle-in-Cell Simulations. IEEE Transactions on Parallel and Distributed Systems, page 2734 – 2748, 2021.

Reasoning on differences in **reproduced results** is not trivial

Results from binary coalescence searches presented in the LIGO's GW150914 discovery paper from *Abbott et al.* and in the reproducibility work of *Brown et al*. Differences are likely due to the gravitational-wave stain datasets originally used vs. those made open-access to the public. →

Duncan et al.. Sci. Eng., 23(2):73–82, 2021
Abbott et al. Phys.Rev. Lett.,vol.116,p.061102, Feb2016,

Figure 4 of *Abbott et al.*

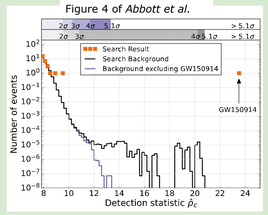
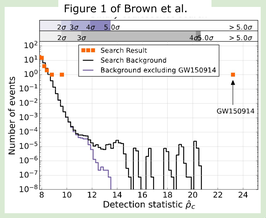


Figure 1 of Brown et al.



A Virtual World Café (VWC) …

A **Virtual World Café (VWC)** is an online event that fosters knowledge sharing and creates opportunities for actions. Café participants are distributed across several breakout sessions, with participants switching sessions periodically and getting introduced to the previous discussion at their new session by a session lead.

… for Scientists Working with High-throughput Applications

Three key requirements to achieve robust science:

- **Performance scalability:** high-throughput applications must meet both hardware and software performance expectations when executed despite heterogeneous resources and large scale systems.
- **Trustworthiness:** individuals must trust technology (i.e., hardware and software), people (e.g., collaborators across scientific domains), and organizations hosting the applications' execution and data (e.g., a cloud provider such as IBM, AWS, or Google hosting scientific data) to behave as specified or expected.
- **Reproducibility:** individuals must be able to draw the same scientific conclusions using the knowledge encapsulated in the original computational experiment.

Acknowledgement:
The authors thank the participants in the February 2021 VWC (https://robustscience.org/) for the vibrant discussions. Findings and recommendations in this poster are the results of those discussions.
The work in this poster is funded by the National Science Foundation (NSF) under grants #2028881, #2028923, #2028930, #2028955, and #2028956.

https://robustscience.org

Outcomes of the Virtual World Café (VWC): Findings and Recommendations

Active Engagement of Scientists

Findings:
Data-driven and HPC scientific simulations are amenable to high-throughput computing thanks to being easily divided into digestible chunks for concurrent processing. Examples of such applications include: the Transitory Exoplanet Sky Survey (TESS), reproducing GW150914 (i.e., the first observation of gravitational waves from a binary black hole merger), molecular dynamics simulations, neural network architecture search, and connectomics. These applications have defined methods for dividing data and processing chunks in parallel with little to no interaction between chunks.

Recommendations:
- Work with the communities to create a taxonomy of data-driven applications and develop standards for data manipulation.
- Consider developer time and costs when prioritizing the applications to target.

Standards for Scalability, Trustworthiness, and Reproducibility

Findings:
Definitions for these three metrics may change across domains; the lack of an explicit definition in interdisciplinary projects may slow down collaborations. Scalability is limited by hardware (e.g., memory bandwidth restrictions, I/O bottlenecks, network bandwidth) and software (e.g., lack of parallelism caused by complex or inefficient communication, or algorithms that were not well designed). Reproducibility is difficult if not impossible at large scale. Resources, funding, and workforce support for reproducibility are often limited. Models and executions are trustworthy if explainable. Annotated executions are vital for determining trustworthiness in disciplines with rapidly evolving models and data. Stochasticity and "messy" data are major challenges in trustworthiness of high-throughput applications.

Recommendations:
- Work with the community to identify and develop standards; curate standards to make sure they remain relevant across projects and time.
- Create processes and mechanisms for deciding what data should be kept or thrown out; store applications' recurring patterns.
- Establish trust and reproducibility in published work; make reproducibility studies a standard submission of journals and conferences.
- Share intermediate results for validation; design and disseminate tools and APIs that support workflow traceability adoptable across communities.

Machine Learning for HTC Workflows

Findings:
Scientific data from different sources (e.g., biological, astrophysics, materials science) is messy, unsteady, varying, and lacks a general annotation format. It is hard to extract the data/information AI developers need from what is given by the scientists, leading to trust issues in the data and model. Moreover, in most AI applications, there is insufficient validation data. The inner stochasticity in AI (e.g., drop-out layers, random seeds for NN weights) hinders the trustworthiness and reproducibility of the applications.

Recommendations:
- Create common annotation standards across applications and scientific domains. Automate the end-to-end pipeline of data generation and analysis (traceability) and executions (explainability).
- Provide validation datasets and benchmarks for every high-throughput application.
- Explain AI models in order to trust them; it is better to have simple and understandable models with comprehensive records more than complex architectures that the end-user cannot trust or replicate the results.

Workforce Development

Findings:
The frantic pace of the academic community exposes the challenging trade-off between performance and scalability vs. trustworthiness and reproducibility. The human participation in raw data processing, analysis of results, or other stages in high-throughput applications impose a bottleneck especially in terms of performance scalability and trustworthiness. Students in particular tend to focus on performance and scalability to the detriment of trustworthiness and reproducibility. Students are often pressured to produce impactful results under tight time constraints for their degree.

Recommendations:
- Work with communities to curate standards. Invest resources to train students and scientists in best practices and standardization.
- Provide necessary infrastructure, such as repositories and tools, to make applications reproducible, scalable, and trustworthy.
- Foster collaboration from cloud alternatives, GitHub (repositories), ACM Badges, and XSEDE resources to promote scalability, trustworthiness, and reproducibility.