High-Precision Computing on Low-Precision Hardware

Reet Barik[*], Aditya Kashi[ϕ], Hao Lu[χ], Piotr Luszczek[τ], Paul Lin[θ], and Balavignesh Vemparala[κ]

[*]Argonne National Laboratory (ANL)
[ϕ,χ]Oak Ridge National Laboratory (ORNL)
[τ]MIT Lincoln Laboratory
[θ]Lawrence Berkeley National Laboratory (LBNL)
[κ]Synopsys Inc

January 2026

Potential Steering/Program Committee

TBD

Abstract

This workshop aims to bring together researchers, application scientists, system developers, and industry partners to explore approaches that enable high-precision scientific computing to effectively leverage low-precision hardware. As modern accelerators increasingly favor reduced-precision arithmetic for performance and energy efficiency, many scientific applications continue to require numerical accuracy, stability, and reproducibility. Addressing this widening gap demands community-driven advances across algorithms and scientific software towards domain-agnostic precision-aware libraries and infrastructure.

The workshop will feature solicited works, invited and contributed talks highlighting mixed-precision methods, error-aware and adaptive numerical algorithms, floating-point emulation libraries, and practical experiences deploying such techniques on large-scale systems. A key emphasis will be on building shared, community-driven libraries that can be broadly adopted across scientific domains, reducing duplication of effort and accelerating impact. By fostering discussion among application developers, numerical experts, and hardware vendors, the workshop seeks to establish best practices, share lessons learned, and strengthen collaboration between academia and industry to make low-precision hardware a reliable and productive component of future scientific computing workflows.

1 Workshop Scope

Current and emerging HPC platforms increasingly deliver peak performance and energy efficiency through reduced-precision arithmetic (e.g., FP16/BF16/TF32/FP8 and integer formats) and specialized units such as tensor and matrix engines, while improvements in throughput for FP64 and often FP32 are no longer keeping pace. For instance, FP64/FP32 throughput gains from NVIDIA A100 to H100 were substantial [4], whereas the corresponding gains from H100 to B200 are comparatively modest [5]. This reflects a broader architectural shift toward performance growth concentrated in reduced precision. At the same time, a wide range of scientific applications continue to require high-precision results, numerical stability under ill-conditioning, and reproducibility. Bridging this widening gap presents an opportunity for coordinated hardware-aware innovation across algorithms, software, tools and applications.

The workshop will convene researchers, application scientists, system and library developers, and industry partners to share established and emerging approaches for delivering accurate results on low-precision hardware. Topics include mixed-precision methods with reliability safeguards, precision-adaptive and selective refinement strategies, and precision-extension techniques such as FP64 emulation for general-purpose scalar arithmetic [1]. The workshop will also explore alternative number formats, including posits and unum-inspired representations [2], in the context of HPC applications and workflows.

Equally important are methodologies and tools that help users understand and manage precision sensitivity. These include techniques for identifying numerically fragile kernels and operation chains, diagnosing sources of instability or non-reproducibility, and quantifying the impact of precision choices on end-to-end scientific conclusions. A central theme of the workshop is community building around shared artifacts. In particular, the workshop will promote the development of a community-curated inventory and eventual suite of applications and kernels with documented numerical sensitivity, along with evaluation practices and metrics that jointly capture performance, energy efficiency, and numerical fidelity. By grounding discussions in representative benchmarks such as OpenMxP [3] and real deployment experiences on leadership-class systems, the workshop aims to facilitate reproducible comparisons, disseminate best practices, and accelerate the maturation of precision-aware libraries and tools.

Within this scope, which emphasizes early-stage ideas, methods, tools, and community-driven efforts to advance precision-aware algorithm and software design for large-scale scientific applications, the workshop will solicit contributions including, but not limited to:

- Techniques, analyses, and tooling for characterizing numerical sensitivity in scientific workloads, including application kernels, solver pipelines, and full-scale simulations.
- Frameworks and methodologies for reasoning about numerical stability, conditioning, and error propagation across sequences of floating-point operations.
- Approaches for executing high-precision computations on low-precision hardware, including FP64 emulation strategies, mixed-precision algorithms, alternative number systems (e.g., posits and unum-based representations), and precision-adaptive methodologies beyond dense linear algebra.
- Compiler, runtime, and library support for precision-aware execution, including hardware-aware optimizations and automated precision selection.
- Application-level experiences and case studies demonstrating trade-offs between numerical fidelity, performance, and energy efficiency across scientific domains.

2 Workshop Goals

This workshop aims to bring together HPC researchers attempting to explore ways in which high precision scientific computing can continue on low precision hardware. The discussions and solicited works will align with the following goal:

- **Ideation** Survey existing methods, tools, and workflows that allow scientists to characterize and reason about the numerical (in-)stability and precision sensitivity of their scientific workloads.
- **Ideation:** Develop and evaluate formal frameworks for identifying well-conditioned sequences of numerical operations and for quantifying error propagation and sensitivity in ill-conditioned operation chains within scientific applications.
- **Ideation and Socialization of Early Works:** Expose emerging techniques for high-precision computation on low-precision hardware, including FP64 emulation for scalar arithmetic, alternative number formats such as posits and unum-based representations, and precision-adaptive software libraries that extend beyond dense linear algebra.
- **Ideation and Building Bridges Between Interdisciplinary Communities:** Discuss and establish a community-curated benchmark suite of scientific applications and kernels with known or documented numerical sensitivity, for evaluating precision-aware methods, emulation techniques, and hardware support.

- **Building Bridges Between Interdisciplinary Communities and Offering a Forum for HPC Practitioners:** Facilitate dialogue between practitioners from industry and the scientific computing community to share practical experiences and best practices for enabling high-precision computation on low-precision hardware, including hardware-aware software support, compiler and library capabilities, and application-level refactoring strategies that balance numerical fidelity with performance.

3 Relevance and Impact to SC Attendees

4 Expected Outcome from the Workshop

The workshop will foster interdisciplinary collaboration between researchers and practitioners and distill practical guidance for mixed-precision strategies, application refactoring, and tool support, while producing concrete, community-facing outcomes that advance high-precision scientific computing on low-precision hardware.

- **Archival Publications:** Accepted papers will be peer-reviewed and published in the SC workshop proceedings, providing a citable and archival record of the presented research.
- **Curated Technical Artifacts:** Publicly archive workshop invited talks, paper presentations, and panel discussion summaries capturing key technical insights and emerging ideas.
- **Numerical Sensitivity Benchmarks and Use Cases:** Establish an initial, community-curated inventory of scientific applications and kernels with documented numerical sensitivity for evaluating precision-aware methods and hardware–software approaches.
- **Open Challenges and Research Directions:** Synthesize discussion outcomes into a concise summary of gaps, challenges, and promising research directions in numerical stability, error propagation, and precision-adaptive computing.

5 Workshop Format

The workshop will feature invited talks, a series of presentations based on peer-reviewed research submissions, and a panel discussion with experts from academia, government research, and industry. A tentative schedule for a half-day workshop that aligns with the SC conference timeline can be found below:

09:00 – 09:10 Welcome and Introductory Remarks (Organizing Committee)

09:10 – 09:45 Invited Talk I

09:45 – 10:15 Accepted Paper Session I

- Paper 1: 7 min presentation + 3 min Q/A
- Paper 2: 7 min presentation + 3 min Q/A
- Paper 3: 7 min presentation + 3 min Q/A

10:15 – 10:30 Break

10:30 – 11:05 Invited Talk II

11:05 – 11:45 Accepted Paper Session II

- Paper 4: 7 min presentation + 3 min Q/A
- Paper 5: 7 min presentation + 3 min Q/A
- Paper 6: 7 min presentation + 3 min Q/A
- Paper 7: 7 min presentation + 3 min Q/A

11:45 – 12:15 Panel Discussion

12:15 – 12:30 Concluding Remarks and Open Discussion

5.1 Expected Timeline

March xx, 2026: Call for Papers
August xx, 2026: Paper submission deadline (soft)
August xx, 2026: Paper submission deadline (hard)
September xx, 2026: Notifications sent to authors
September xx, 2026: Camera-ready versions due
November xx, 2026: Post workshop content online

6 Inclusivity Plan

The organizing committee is committed to fostering an inclusive environment that brings together participants from industry, national laboratories, and academia, spanning a range of career stages and technical backgrounds. We will make a deliberate effort to ensure balanced representation among invited speakers, panelists, and contributors, in alignment with the SC diversity and inclusivity guidelines, with the goal of supporting broad participation and the continued growth of the HPC community as a whole.

7 Advertising Plan

References

[1] Yozo Hida, Xiaoye S Li, and David H Bailey. "Library for double-double and quad-double arithmetic". In: *NERSC Division, Lawrence Berkeley National Laboratory* (2007), p. 19.

[2] John L Gustafson and Isaac T Yonemoto. "Beating floating point at its own game: Posit arithmetic". In: *Supercomputing frontiers and innovations* 4.2 (2017), pp. 71–86.

[3] Hao Lu et al. "Climbing the Summit and Pushing the Frontier of Mixed Precision Benchmarks at Extreme Scale". In: *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. SC '22. Dallas, Texas: IEEE Press, 2022. ISBN: 9784665454445. DOI: 10.1109/SC41404.2022.00083.

[4] NVIDIA Corporation. *NVIDIA Hopper Architecture In-Depth*. Whitepaper. Accessed: 2026-02-02. 2022. URL: https://resources.nvidia.com/en-us-hopper-architecture/nvidia-h100-tensor-c.

[5] NVIDIA Corporation. *NVIDIA Blackwell Architecture Technical Brief*. Technical Brief. Accessed: 2026-02-02. 2024. URL: https://resources.nvidia.com/en-us-blackwell-architecture.