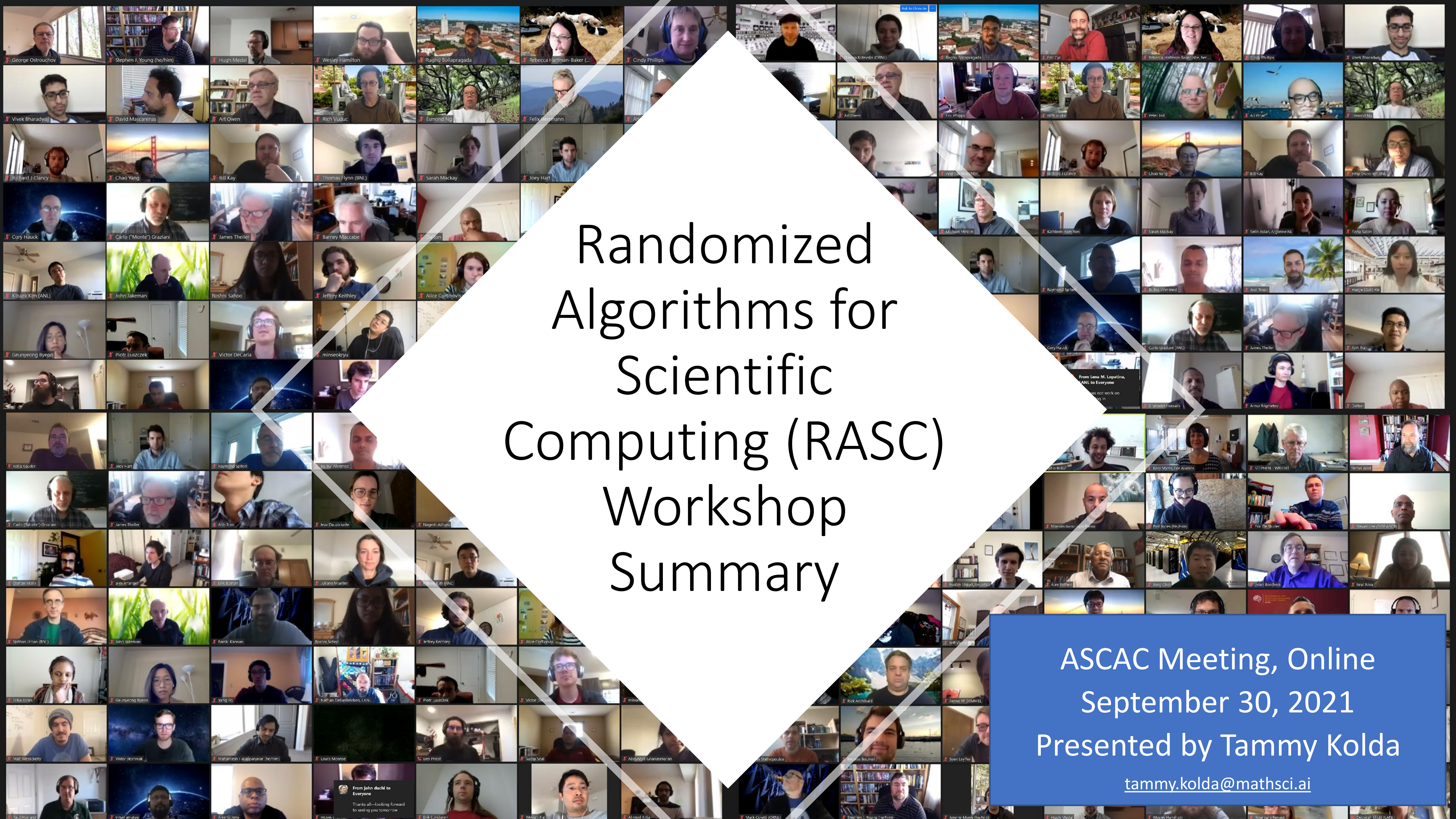


Randomized Algorithms for Scientific Computing (RASC) Workshop Summary

ASCAC Meeting, Online

September 30, 2021

Presented by Tammy Kolda

tammy.kolda@mathsci.ai

Acknowledgements

Program Manager

Steven Lee
US Department of Energy, ASCR

Workshop Co-Chairs (plus me)

Aydin Buluc
Lawrence Berkeley National Laboratory

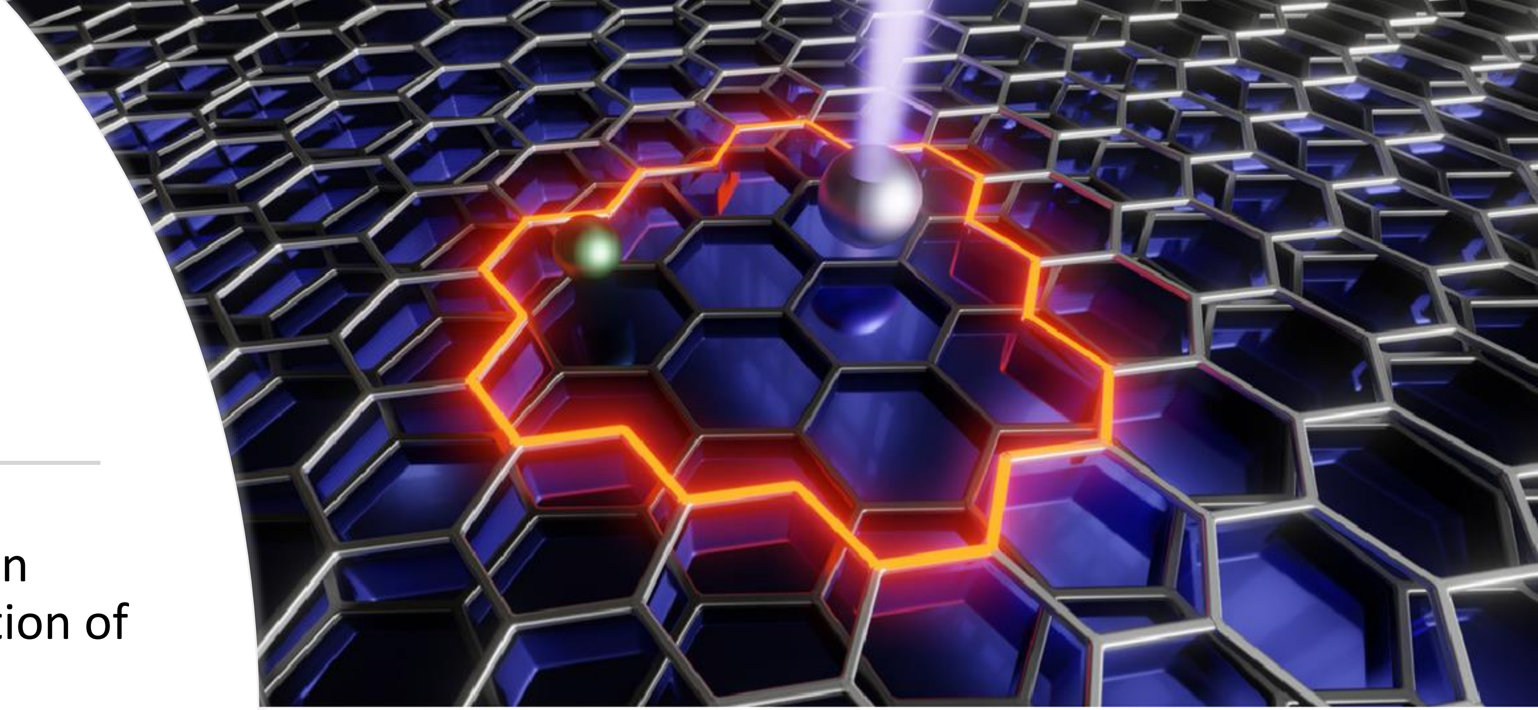
Stefan M. Wild
Argonne National Laboratory

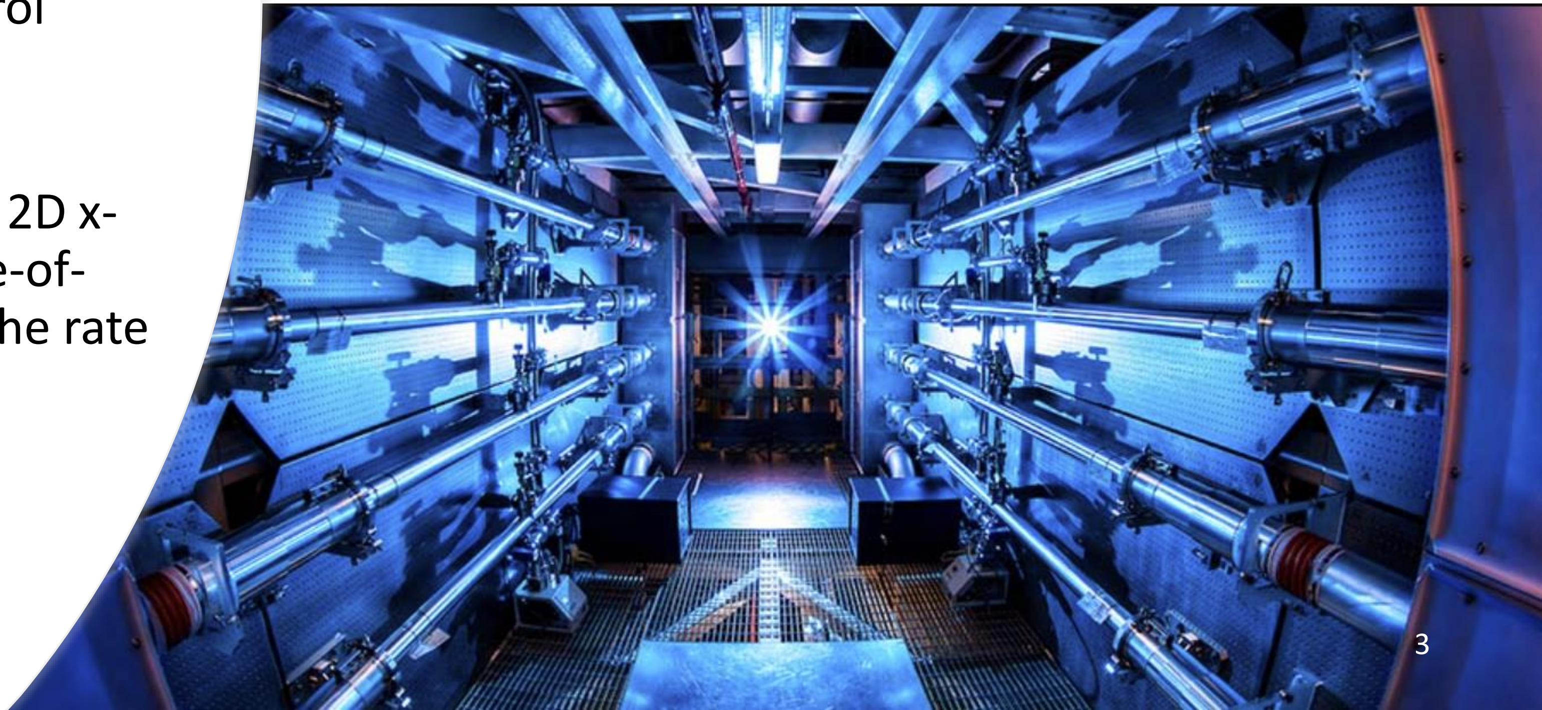
Huge thanks as well to organizing committee, invited speakers and panelists, writing committee, additional report authors, breakout leaders, workshop participants, and ORISE (esp. Jody Crisp, Andrew Fowler, and Paul Hudson).

Need for Randomization Motivated by Data Tsunami

Top: ORNL’s Scanning transmission electron microscopes (STEMs) can modify the location of a silicon atom in a graphene hole with randomized algorithms for real-time control

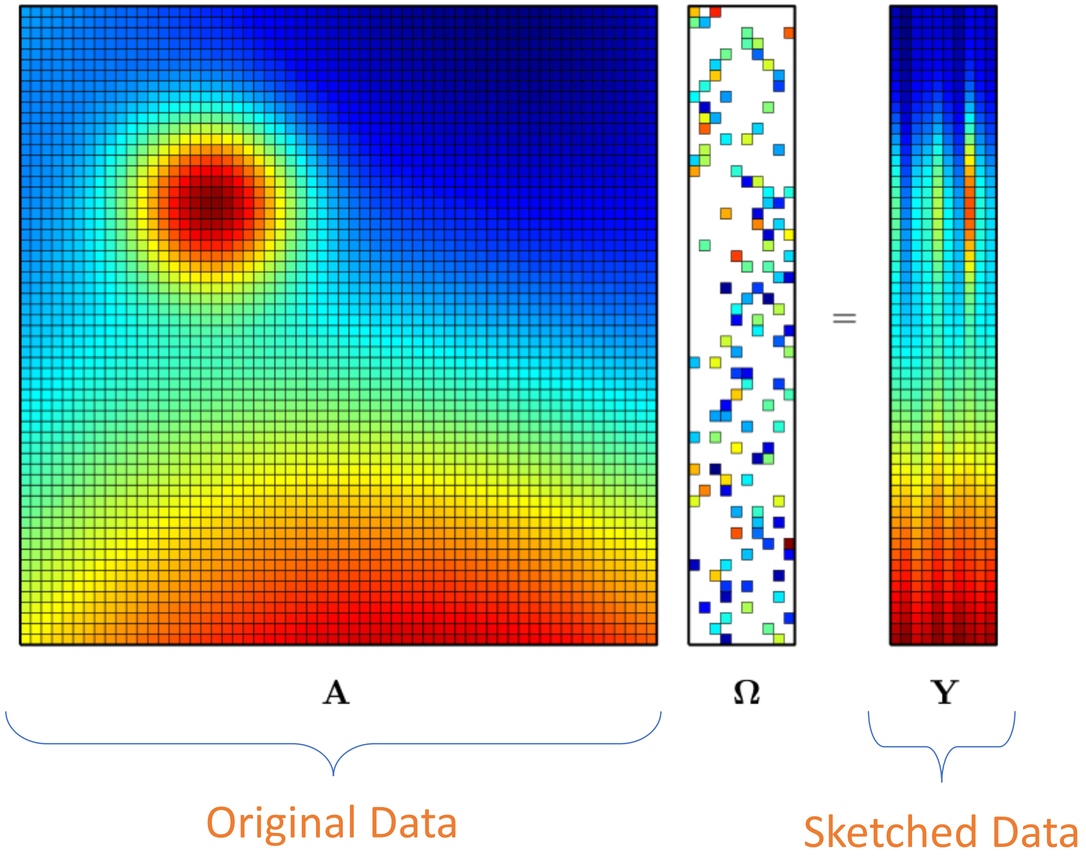
Bottom: LLNL’s ITER has a suite of unique diagnostic tools including 3D neutron and 2D x-ray imaging instruments and neutron time-of-flight detectors. These generates data at the rate of 2TB/day and will require randomized algorithms for efficiency.





Randomized Algorithms Employ Randomness in Internal Algorithmic Decisions

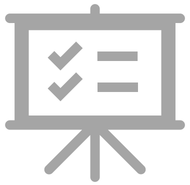
- De rigueur in Machine Learning (ML) and Artificial Intelligence (AI)
    - Ex: Stochastic gradient descent (SGD)
- Employ randomization *within* algorithm
    - Even for deterministic inputs
    - Complementary to but distinct from methods for stochastic inputs
- Probabilistic guarantees for *finite* samples
- Building block for advancing Scientific Computing and AI for Science, Engineering, & Technology
    - Accelerate time to solution
    - Increase scalability
    - Improve reliability
    - Expand impact across DOE



Workshop in Two Parts

Bootcamp (Dec 2-3, 2020) – 453 active participants

- Interactive tutorial presentations
- Panel sessions with application specialists
- Extensive Q&A with experts
- One limited *random* breakout, preview for second part



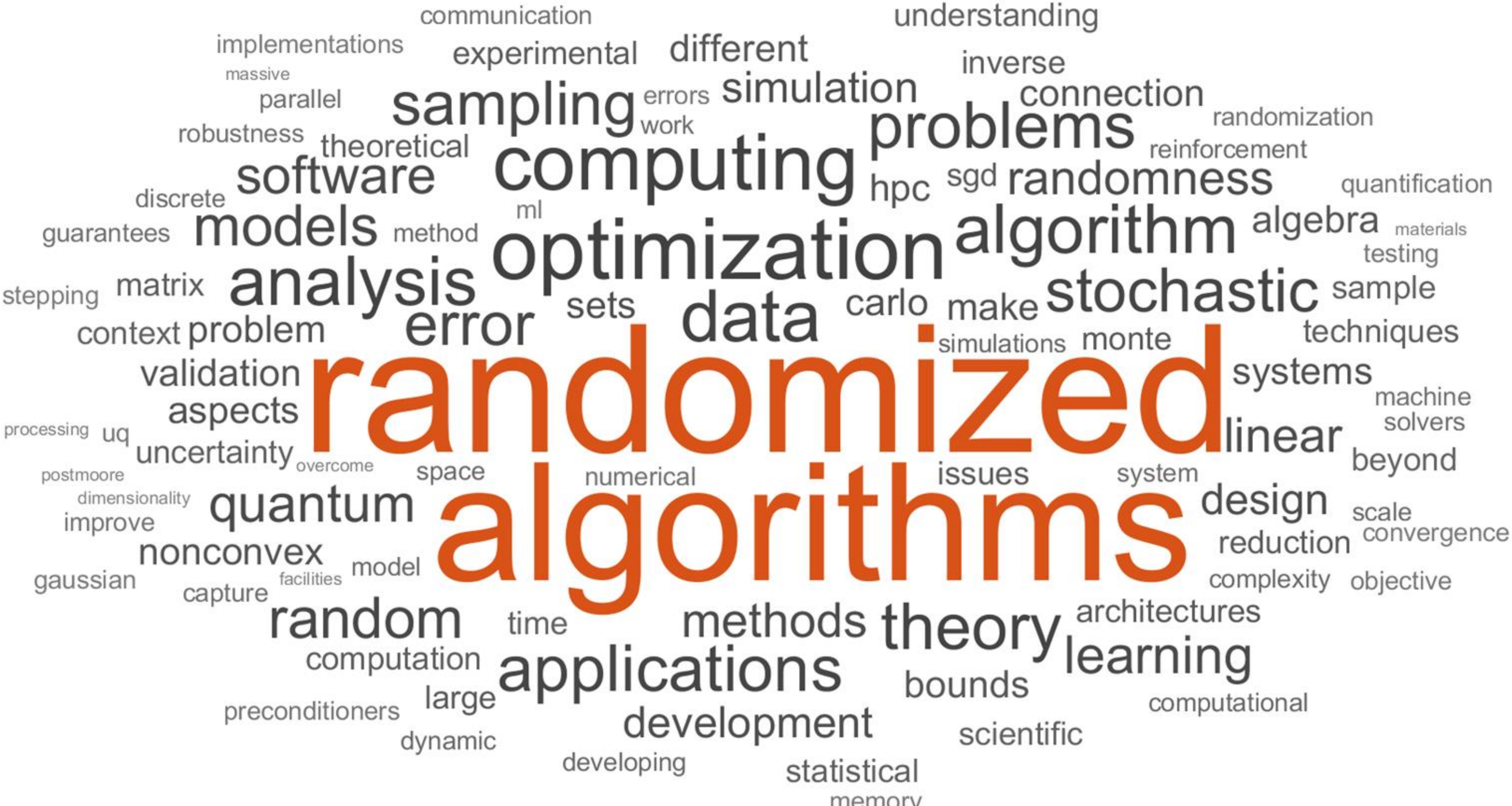
Brainstorming and Writing (Jan 6-7, 2021) – 204 active participants

- Open only to Bootcamp attendees, required 200-word idea for registration
- Panel session with funding program managers
- Breakouts into themes identified from idea submissions
- Identification of main themes to be put forth in final report

A Few Words About the Organization of First Such Online Workshop

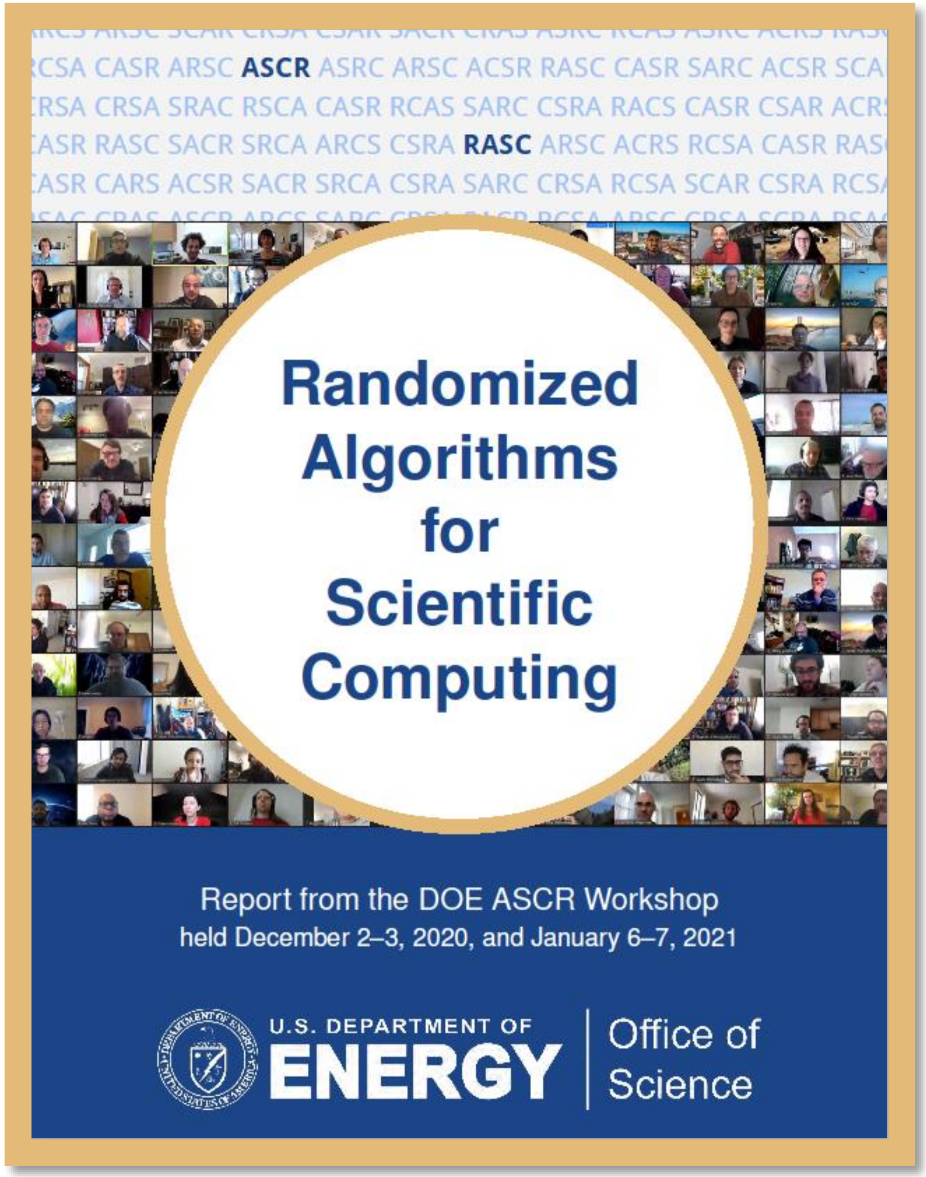
- Online worked surprisingly well
  - Higher and more diverse engagement
  - Allowed non-traditional scheduling, with time for thinking in between sessions
- Organization split into two parts was a big success
  - Huge engagement with bootcamp
  - Healthy degree of continued engagement for writing sessions
- Huge online *meeting*, not a *webinar* (thanks ORISE)
  - Every attendee had the ability to speak, display video, and participate in chat
- Lots of effort put forth to make the meeting engaging and inclusive
  - Non-intense schedule with ample breaks
  - “Seeded” Q&A sessions
  - Polls, breakout sessions, online chatting
- Online meetings for writing committee worked well, report submitted in 2 months

Part 1 Breakout Summaries



Submitted Workshop Report

A. Buluc, T. G. Kolda, S. M. Wild, and others. **Randomized Algorithms for Scientific Computing (RASC)**, 2021, http://arxiv.org/abs/2104.11079



Six Challenges/Opportunities

1. Randomized algorithms are essential to future computational capacity

- Rate of growth in the computational capacity of integrated circuits is expected to slow while data collection is expected to grow exponentially
- Makes randomized algorithms— which depend on sketching, sampling, and streaming computations—essential to the future of computational science and AI for Science

2. Novel approaches by reframing long-standing challenges

- Potential for randomized algorithms goes beyond keeping up with the onslaught of data
- Opens door to novel approaches to long-standing challenges
- Includes scenarios where uncertainty is unavoidable, such as real-time control and experimental steering, design under uncertainty, and mitigating stochastic failures in novel materials

3. Randomness intrinsic to next-generation hardware

- Computing efficiencies can be realized by *purposely* allowing random imprecision
- Imprecision inherent in emerging architectures such as quantum and neuromorphic computers
- Future computing systems will benefit from co-design of randomized algorithm alongside hardware that favors certain instantiations of randomness

Six Challenges/Opportunities

4. Technical hurdles requiring theoretical and practical advances

- Need to crafting sophisticated approaches that break the “curse of dimensionality” via sublinear sampling, sketching, and online algorithms requires sophisticated analysis
- Tackled thus far only in a small subset of scientific computing problems
- Foundational research in theory and algorithms needs to be multiplied many times over in order to cover the breadth of DOE applications

5. Reconciliation of randomness with user expectations

- Users are conditioned to certain expectations, such as viewing machine precision as sacrosanct, even when fundamental uncertainties make such precision ludicrous
- New metrics for success can expand opportunities for scientific breakthroughs by accounting for tradeoffs among speed, energy consumption, accuracy, reliability, and communication

6. Need for expanded expertise in statistics and other areas

- Need to increase emphasis on statistics, theoretical computer science, data science signal processing, and emerging hardware expertise
- Continued emphasis along in traditional domains of applied mathematics, computer science, and engineering and science domain expertise

Recommended Research Directions

1. Theoretical foundations
2. Algorithmic foundations
3. Application integration
4. Next-gen architectures
5. Interdisciplinary research/outreach
6. Workflow standardization

EXPRESS Randomized Algorithms for Extreme-Scale Science

- Posted 3/26
- Closed 5/19
- Awards 8/24
- $2.8M in funding

Laura Balzano (Michigan)	• Get Non-Real: Randomized Sketching for High-Dimensional Non-Real-Valued Data
Petros Drineas (Purdue)	• A Contextually-Aware Sensitivity Analysis to Guide the Design of Randomized Least Squares Solvers in Applications
Per-Gunnar Martinsson (Texas)	• Randomized Algorithms for Accelerating Linear Algebraic Computations
Christopher Musco (NYU)	• Leverage Score Sampling for Parametric PDEs
Katarzyna Swirydowicz (PNNL)	• RATFISH: Randomized Techniques for Iterative Solvers in Heterogeneous environments
Stephen Young (PNNL)	• SHARWK: Scalable Hypergraph Analysis via Random Walk Kernels

Contributions from the Office of Science Research Community & Beyond

Substantial volunteer investments made this workshop successful. We provide explicit acknowledgments in the next few slides but want to stress this still does not do justice to the generosity of many individuals as well as the community as a whole.

Breakout Leads for Part 1

Ali Pinar
Anthony DeGennaro
Aric Hagberg
Chandrika Kamath
Chao Yang
Cindy Phillips
David Womble
Draguna Vrabie
Eric Cyr
Eric Phipps
Esmond Ng

Hemanth Kolla
Jed Duersch
Jeff Hittinger
Jelani Nelson
Jim Ahrens
Joey Hart
John Jakeman
Juan Restrepo
Kary Myers
Osni Marques
Paul Hovland

Ramki Kannan
Ray Tuminaro
Rebecca Hartman-Baker
Rick Archibald
Rich Lehoucq
Sherry Li
Stephen J Wright
Stephen Young
Sven Leyffer
Vishwas Rao

Report Authors

- Aydin Buluc (co-chair), LBNL
- Tamara G. Kolda (co-chair), Sandia
- Stefan M. Wild (co-chair), Argonne

- Mihai Anitescu, Argonne
- Anthony DeGennaro, Brookhaven
- John Jakeman, Sandia
- Chandrika Kamath, LLNL
- Ramakrishnan (Ramki) Kannan, ORNL
- Miles E. Lopes, U. California, Davis
- Per-Gunnar Martinsson, U. Texas, Austin
- Kary Myers, LANL
- Jelani Nelson, U. California, Berkeley
- Juan M. Restrepo, ORNL
- C. Seshadhri, U. California, Santa Cruz
- Draguna Vrabie, PNNL
- Brendt Wohlberg, LANL
- Stephen J. Wright, U. Wisconsin, Madison
- Chao Yang, LBNL
- Peter Zwart, LBNL

Additional Report Authors

- Rick Archibald, Oak Ridge National Laboratory
- David Barajas-Solano, Pacific Northwest National Laboratory
- Andrew Barker, Lawrence Livermore National Laboratory
- Charles Bouman, Purdue University
- Moses Charikar, Stanford University
- Jong Choi, Oak Ridge National Laboratory
- Aurora Clark, Washington State University
- Victor DeCaria, Oak Ridge National Laboratory
- Zichao Wendy Di, Argonne National Laboratory
- Jack Dongarra, University of Tennessee, Knoxville
- Jed Duersch, Sandia National Laboratories
- Ethan Epperly, California Institute of Technology
- Benjamin Erichson, University of California, Berkeley
- Maryam Fazel, University of Washington, Seattle
- Andrew Glaws, National Renewable Energy Laboratory
- Carlo Graziani, Argonne National Laboratory
- Cory Hauck, Oak Ridge National Laboratory
- Paul Hovland, Argonne National Laboratory
- William Kay, Oak Ridge National Laboratory
- Nathan Lemons, Los Alamos National Laboratory
- Ying Wai Li, Los Alamos National Laboratory
- Dmitriy Morozov, Lawrence Berkeley National Laboratory
- Cynthia Phillips, Sandia National Laboratories
- Eric Phipps, Sandia National Laboratories
- Benjamin Priest, Lawrence Livermore National Laboratory
- Ruslan Shaydulin, Argonne National Laboratory
- Katarzyna Swirydowicz, Pacific Northwest National Laboratory
- Ray Tuminaro, Sandia National Laboratories

Organizing Committee and Support

- Co-chairs
  - Aydin Buluc (LBNL)
  - Tammy Kolda (SNL)
  - Stefan Wild (ANL)
- Committee Members
  - Anthony DeGennaro (BNL)
  - Mihai Anitescu (ANL)
  - Brendt Wohlberg (LANL)
  - Draguna Vrabie (PNNL)
  - Maryam Fazel (U. Washington)
  - Chandrika Kamath (LLNL)
  - Ramki Kannan (ORNL)
- Committee Members (cont'd)
  - Kary Myers (LANL)
  - Jelani Nelson (UCB)
  - Joel Tropp (Caltech)
  - Juan Restrepo (ORNL)
  - Steve Wright (U. Wisconsin)
  - Chao Yang (LBNL)
- DOE Sponsor
  - Steve Lee (ASCR)
- Logistics and Support
  - Jody Crisp (ORISE)
  - Paul Hudson (ORISE)