ASTRA-sim and Chakra Tutorial:
Closing Remarks

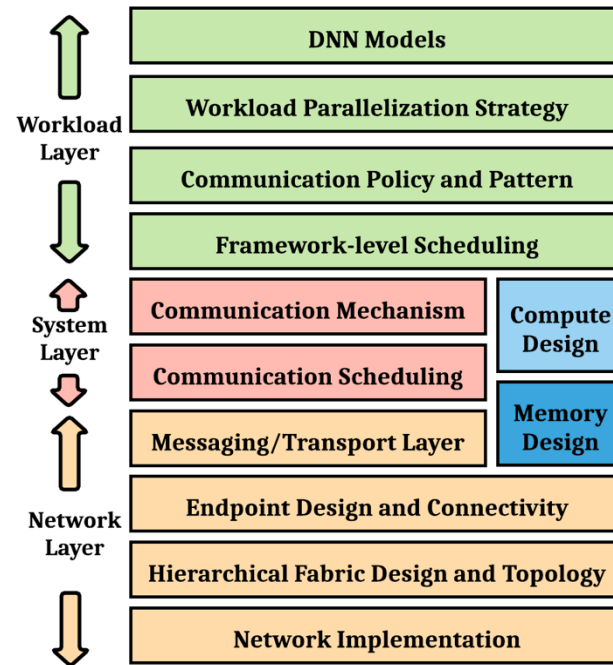Tushar Krishna

Associate Professor

School of ECE, Georgia Institute of Technology

tushar@ece.gatech.edu

Recap: Motivation of this Tutorial

- Large-scale distributed ML is an ongoing open-research area
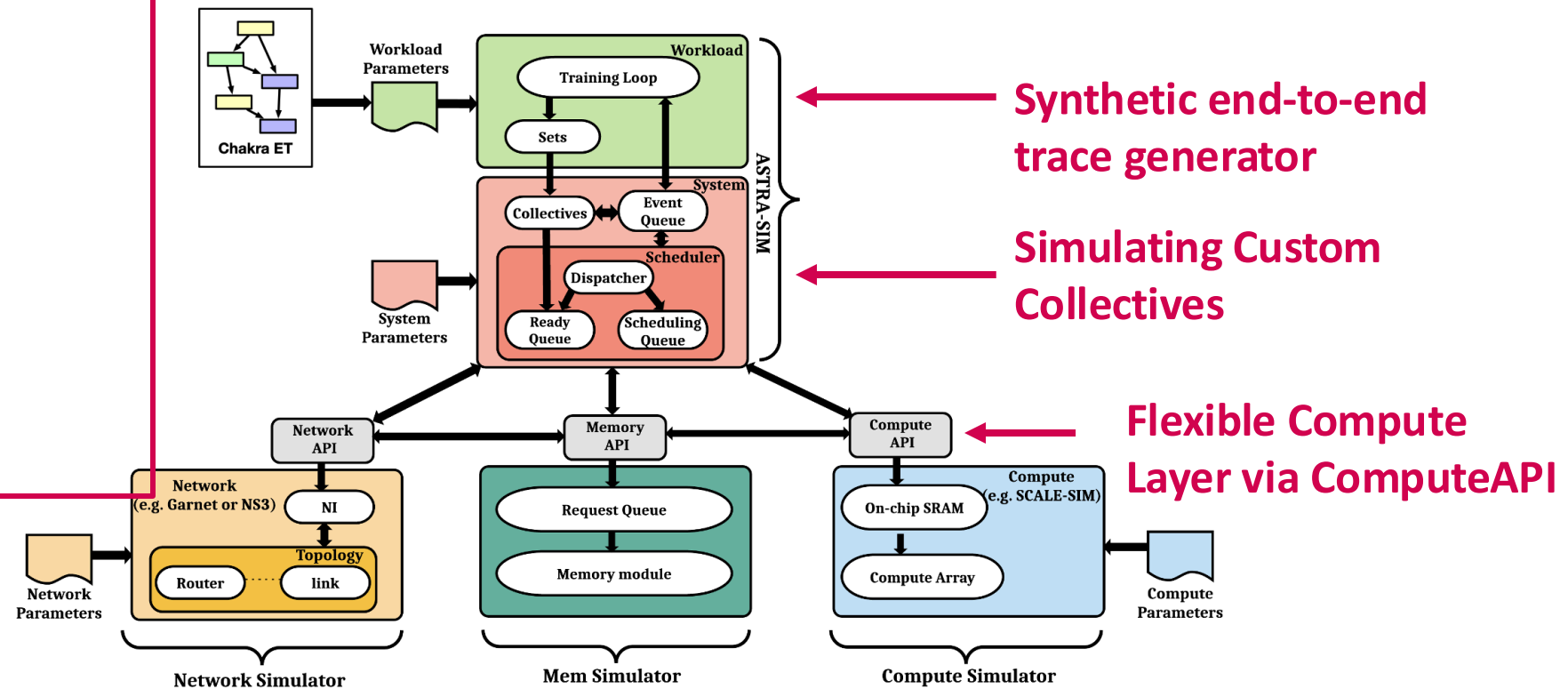- Various co-design aspects for performance, efficiency, sustainability

- Many emerging supercomputing systems being designed specifically for this problem!

ASTRA-sim is evolving

- Highlighted works in progress + many more collaborative efforts

Wish List (Call for Collaboration)
- More detailed statistics
- Hierarchical Multi-fidelity Network simulation
- SST as Network Backend
- Help with validation
- ...

Synthetic end-to-end trace generator

Simulating Custom Collectives

Flexible Compute Layer via ComputeAPI

Contribution and Participation

- **Chakra and ASTRA-sim are open-source!**
  - Feel free to raise GitHub issues
  - Contribute via pull-requests

- Reach out to use via Chakra WG members mailing list (chakra@mlcommons.org) for Chakra-specific and ASTRA-sim User mailing list (astrasim-users@googlegroups.com) for ASTRA-sim specific discussions

Summary of Resources

- **[Chakra]**
  - **[Working Group]** http://mlcommons.org/working-groups/research/
  - **[Github]** https://github.com/mlcommons/chakra
  - **[Wiki]** https://github.com/mlcommons/chakra/wiki
  - **[Chakra Paper]** https://arxiv.org/abs/2305.14516

- **[ASTRA-sim]**
  - **[Website]** https://astra-sim.github.io
    - Including links to tutorials
  - **[GitHub]** https://github.com/astra-sim/astra-sim
  - **[Wiki]** https://astra-sim.github.io/astra-sim-docs/index.html
  - **[Validation]** https://astra-sim.github.io/astra-sim-docs/validation/validation.html
  - **[ASTRA-sim2.0 paper]** https://arxiv.org/abs/2303.14006
  - **[Mailing List Sign-up]** https://forms.gle/18KVS99SG3k9CGXm6

Presenters and Contributors

Presenters

Tushar Krishna
Associate Professor, ECE
Georgia Tech
tushar@ece.gatech.edu

William Won
Ph.D. Candidate, CS
Georgia Tech
william.won@gatech.edu

Joongun Park
Post Doctoral Researcher
Georgia Tech
jpark3234@gatech.edu

Taekyung Heo
Senior HPC Middleware
Developer, NVIDIA
theo@nvidia.com

Vinay Ramakrishnaiah
Senior Member of Technical
Staff. AMD
vinay.ramakrishnaiah@amd.com

Contributors & Collaborators

Georgia Tech
Jinsun Yoo
Changhai Man
Ziwei Li
Divya Kiran Kadiyala

Meta
Saeed Rashidi
Louis Feng
Sheng Fu
Brian Coutinho
Darshan Sanghani
Adi Gangidi

NVIDIA
Srinivas Sridharan

Intel
Sudarshan Srinivasan

AMD
Ruchi Shah
Brad Beckmann
Furkan Eris
+more

ML Commons

Chakra: Advancing Benchmarking and Co-design for Future AI Systems

Announcing Chakra, execution traces and benchmarks working group

+ many more industry/academic researchers & engineers

Discussions, Q/A, and Conclusion

Thank you!