Ana Gainaru

Computer Scientist, Oak Ridge National Laboratory
http://ana-gainaru.com
gainarua@ornl.gov

RESEARCH INTEREST	High Performance Computing, Big Data, Performance modeling, Data Analytics, Storage Systems, Performance optimizations, Sparse Computations, Irregular Applications, Data movement

EDUCATION

University of Illinois at Urbana-Champaign, IL, USA   **Aug 2011 - Aug 2015**
PhD in Computer Science
Advisor: Marc Snir

University Politehnica of Bucharest, Romania   **Oct 2008 - Jul 2010**
Master of Science in Computer Science
Ranked 1st in my year.

University Politehnica of Bucharest, Romania   **Oct 2003 - Jul 2008**
Bachelor of Science in Computer Science
Performance Scholarship 6 semesters out of 10

PROFESSIONAL EXPERIENCE

Computer Scientist   **Sep 2020 - present**
Oak Ridge National Laboratory, CSMD
Develop techniques for profiling, managing, and analyzing large-scale data from simulations and experiments. Focus on data streaming and campaign management designed for heterogeneous I/O architectures.

Research Assistant Professor   **Jul 2017 - Aug 2020**
Vanderbilt University, Electrical Engineering and Computer Science Department
Performance analysis and optimization for large sparse structures and algorithms driven by big-data simulations. Developing resource scheduling, fault tolerance and performance scaling techniques for large scale HPC applications. Working with the neuroscience department on understanding their needs and designing methods and tools to facilitate their deployment on HPC.

HPC Architect   **Oct 2015 - Jun 2017**
Mellanox Technologies Inc
Designed and optimized Mellanox' internal collective communication library for extreme-scale systems. Implemented a small data all-to-all algorithm using different data layout patterns at intermediate steps by selectively replacing the CPU based buffer packing and unpacking with Mellanoxs InfiniBand support for Host Channel Adapter (HCA) hardware scatter/gather.

Research Assistanship at NCSA   **Aug 2011 - May 2015**
National Center for Supercomputing Applications, Integrated Systems Console (ISC)
Worked done in the context of the NCSA/UIUC/INRIA Joint Lab for Petascale Computing and the Blue Waters project. Designed and implemented parts of the fault tolerance framework used by the Blue Waters system. Developed methods for studying "big data" system logs from different production HPC systems and investigated the best way to model failure distribution and characteristics for several hardware failures.

Collaboration with ENS Lyon and INRIA Bordeaux   **Oct 2013 - present**
Developed methods for analyzing the congestion experienced by multiple IO intensive applications running concurrently and sharing resources in a cluster. Modeled the application I/O behavior and designed I/O scheduling policies that optimize burst buffer usage and decrease the overall system congestion. Analyzing prefetching and I/O dependencies schemes based on the needs of neuroscience applications at Vanderbilt.

PUBLICATIONS   For a complete list of publications (19 conference papers and 5 journal papers) visit my dblp page (`https://dblp.org/pers/hd/g/Gainaru:Ana`). The following list reflects the most recent and/or most relevant publications:

1. Hongyang Sun, <u>Ana Gainaru</u>, Manu Shantharam and Padma Raghavan. *Selective Protection for Sparse Iterative Solvers to Reduce the Resilience Overhead* [IEEE 32nd International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), 2020]

2. <u>Ana Gainaru</u>, Brice Goglin, Valentin Honor, Guillaume Pallez, Padma Raghavan, Yves Robert, Hongyang Sun. *Reservation and Checkpointing Strategies for Stochastic Jobs* [IPDPS 2020]

3. <u>Ana Gainaru</u>, Guillaume Pallez, Hongyang Sun, Padma Raghavan. *Speculative Scheduling for Stochastic HPC Applications* [ICPP 2019]

4. <u>Ana Gainaru</u>, Guillaume Pallez. *Making Speculative Scheduling Robust to Incomplete Data* [SCALA@SC 2019]

5. Guillaume Aupy, <u>Ana Gainaru</u>, Valentin Le Fevrez. *I/O scheduling strategy for periodic applications* [ACM Transactions on Parallel Computing 2019]

6. <u>Ana Gainaru</u>, Hongyang Sun, Guillaume Aupy, Yuankai Huo, Bennett A. Landman, Padma Raghavan *On-the-fly scheduling vs. reservation-based scheduling for unpredictable workflows* [Special Issue of the IJHPCA 2019]

7. Guillaume Aupy, <u>Ana Gainaru</u>, Valentin Honor, Padma Raghavan, Yves Robert, Hongyang Sun *Reservation Strategies for Stochastic Jobs* [IPDPS 2019]

8. Richard Graham, <u>Ana Gainaru</u>, Artem Polyaiov and Gilad Shainer *Using InfiniBand Hardware Gather-Scatter Capabilities to Optimize MPI All-to-All* [EuroMPI 2016]

9. <u>Ana Gainaru</u>, Guillaume Aupy, Anne Benoit, Franck Cappello, Yves Robert, Marc Snir *Scheduling the I/O of HPC applications under congestion* [IPDPS 2015]

10. <u>Ana Gainaru</u> , Franck Cappello, Marc Snir, William Kramer *Fault prediction under the microscope: A closer look into HPC systems* [SC 2012]

11. Eric Heien, Derrick Kondo, <u>Ana Gainaru</u> , Dan LaPine, Bill Kramer, Franck Cappello *Modeling and Tolerating Heterogeneous Failures in Large Parallel Systems* [SC 2011]

AWARDS,
HONORS AND
COMPETITIONS

- 2015 AfterCollege Engineering Student Scholarship recipient, April 2015
- Special Prize for Excellence, Gala of Romanian Students Abroad (LSRS), judged by the Romanian Academy (the most important scientific and cultural forum in Romania), awarded for my entire PhD work, 2015
- 2014 Student of the Year Award, NCSA Employee Recognition Award, 2014
- First place at the Intel Parallel Universe Computing Challenge, 2014
- 2013 Technical Achievement Award, NCSA Employee Recognition Award, 2013
- CS Excellence Fellowship (Saburo Muroga Endowed Fellowship) granted by the University of Illinois at Urbana-Champaign, 2011-2012
- Grad Cohort for Women in Computing Research Association (CRAW) Travel Grant, 2011
- SC Travel Support Grant Scholarship, 2010
- Excellence Title for the Master Program, granted by Automatic Control and Computers Faculty, University Politehnica of Bucharest, graduated first of my class, 2010
- Third prize at the Students Scientific Conference from the University Politehnica of Bucharest for my undergrad thesis: "Mobility Model based on Social Networks for VANETs", 2008.
- Oracle diploma received for my results on Database and Database Design courses studied in Automatic Control and Computer Science Faculty, top 10%, 2007

PANEL	*Fault Tolerance/Resilience at Petascale/Exascale: Is it Really Critical? Are Solutions Necessarily Disruptive?* [SC 2013] Moderator: Franck Cappello, Panelists: Marc Snir, Bronis De Supinski, Al Geist, John Daly, <u>Ana Gainaru</u> , Satoshi Matsuoka
BOOK CHAPTER	*Fault-Tolerance Techniques for High-Performance Computing* [Springer Book, Computer Communications and Networks series, 2015], Editors: Thomas Herault and Yves Robert Chapter 1: Fault and failures (including: source detection, root cause analysis, silent errors, predictors), Authors: <u>Ana Gainaru</u> and Franck Cappello

PROFESSIONAL AND COMMUNITY

- NSF Panel review member, 2019/2020
- Organizer for the Convergence Computing Infrastructure Workshop (CCIW), 2019/2020
- Journal editorial board:
  - IEEE Transactions on Parallel and Distributed Systems (TPDS), part of the Review Board for an 18 month term since February 2020
  - Journal of Parallel and Distributed Computing (JPDC) as as a subject area editor (AE) since August 2019
  - International Journal for High Performance Computing Applications (IJHPCA), since January 2019
- Vice chair or Co-chair:
  - Vice chair for the Tutorials SC 2020 Committee
  - Vice chair for the Poster SC 2018 Committee
  - Co-chair for the FTS 2017 Workshop on Fault Tolerant Systems in conjunction with CCGRID 2017
- Member of the Program Committee for:
  - HPDC 2020, the International Symposium on High-Performance Parallel and Distributed Computing
  - SC 2019/2018/2017/2016, the International Conference for High Performance Computing, Networking, Storage and Analysis, technical papers program committee for the Algorithms area
  - ICPP 2019, the International Conference on Parallel Processing
  - IPDPS 2019/2018/2014, the IEEE International Parallel and Distributed Processing Symposium, technical program committee for the System software track
  - ICS 2017, External Review Committee for the ACM International Conference on Supercomputing
  - CCGRID 2016, the 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing
  - EuroMPI 2017, 2016, the MPI users and developers conference
- Reviewer for the following international journals: IJHPCA since June 2014, TPDS since October 2015, FGCS since November 2015, JPDC since January 2016, PARCO since March 2017, SUSCOM since June 2017

TEACHING EXPERIENCE	**Course instructor** *Course: SC 3260 / SC 5260 High Performance Computing*	**Spring 2019, Vanderbilt**
	Invited lecture on MPI *Course: CS 6320 Algorithms for parallel computing*	**Fall 2019, Vanderbilt**
	Teaching assistant *Course: CS 425 Distributed systems*	**Fall 2014, UIUC**

Invited instructor   **Spring 2011, UIUC**
Course: CS 498 Hot Topic in High Performance Computing:
Networks and Fault Tolerance

Teaching assistant   **Spring 2008 - Fall 2009, UPB**
Courses: Architecture of Computing Systems (seniors level),
Parallel Processing Architecture (seniors and grad level)
and The Utilization of Operating Systems (freshman level)

PATENTS

Richard Graham, <u>Ana Gainaru</u>
Using Hardware Gather-Scatter Capabilities to Optimize MPI All-to-All
U.S. Patent No. 15/681, 390, 20 August 2017

STUDENT
MENTORING

Vanderbilt University, 2019 (summer internship), Co-supervision of Loc Jouans from ENS Lyon. The work focused on analyzing neural networks resiliency to silent failures and designing redundancy computations within different neural network layers to increase the intrinsic tolerance.

Vanderbilt University, 2018 (summer internship), Co-supervision of Valentin Honor from INRIA Bordeaux. The main focus of the work was on designing optimal solutions for scheduling stochastic applications. The work during Valetins internship was published in IPDPS 2019 and continued online (the new results published at IPDPS 2020).

OPEN SOURCE
SOFTWARE

All my software is available on my github account (`https://github.com/anagainaru`).

iSBatch
Batch Scheduler Interface
A tool for generating the sequence of walltime requests an application needs to use for submissions on HPC systems (optimal sequence based on a historic log of execution times for an application on a given system configuration). In the testing phase to be used by the neuroscience department at Vanderbilt.

ScheduleFlow
Scheduler Simulator
A python package consisting of a series of scripts and classes that offer an API allowing users to create simulation scenarios for online and reservation-based batch schedulers for large-scale computational systems. Also used as an educational tool to teach advanced topics for the SC 3260 HPC class.

HELO
Hierarchical Event Log Organizer
A tool for extracting event templates from large datasets and updating them as new events get generated. HELO presents an intuitive output to system administrators. It has been integrated in the Blue Water software stack at NCSA.