Nikola Samardzic

US Citizen | phone: +1-650-660-3463 | mail: nsamar@mit.edu | web: n-samar.github.io

EDUCATION

Massachusetts Institute of Technology Boston, MA
Ph.D. Computer Science Sept. 2020 – Present

- Advisors: Daniel Sanchez, Arvind
- Areas: Computer Architecture, Computer Systems
- Awarded Jae S. and Kyuho Lim Graduate Fellowship for 2020-21 academic year.

University of California, Los Angeles Los Angeles, CA
B.Sc. Computer Science Sept. 2016 – June 2020

- Summa Cum Laude & Phi Beta Kappa

Awards & Grants

- *National Science Foundation CAPA Research Grant*: $16,000 award; the only undergraduate to be funded by this grant; 2-3 labs awarded annually across the US; The NSF/Intel Partnership on Computer Assisted Programming for Heterogeneous Architectures (CAPA) "aims to address the problem of effective software development for diverse hardware architectures through groundbreaking university research" (award #1723773); awarded based on my work with Prof. Jason Cong's lab for the 2018-2019 academic year; award subsequently extended to Summer 2019.
- *Norton Rodman Endowed Engineering Scholarship*: $4,500 award; merit-based academic UCLA engineering scholarship; one recipient annually out of 4,000 engineering students; received three years in a row (2017-2019).
- *UCLA Women's Faculty Club Scholarship*: $3,000 award; twelve awardees annually out of 39,000 undergraduate students; merit-based, campus-wide scholarship selected by committee of current and former UCLA professors; preference given to students pursuing research.
- *UCLA Achievement Scholarship*: $10,000 annually; 1% of UCLA students receive the award; awarded in recognition of students' academic excellence; received for 4 years in a row (2016-19).

RESEARCH EXPERIENCE

UCLA VAST Lab (Prof. Jason Cong) Los Angeles, CA
Undergraduate Researcher April 2017 – Present

- **Bonsai sorter** (see Publications, bullet 1):
  - Developed state-of-the-art sorter on an FPGA with higher throughput than all prior work on FPGA, CPU, and GPU for 2-16GB problem size; results submitted for publication at the International Symposium of Computer Architecture (h5-index: 56).
  - Modeled how merge tree unrolling and pipelining should be used to optimize performance of sorters on any hardware setup and problem type. This model

enables automatic optimal reconfiguration of the sorter to any data center or cloud workloads.

- **PRAM to AT^2 translation** (see Publications, bullet 2):
  - Derived an algorithm for translating PRAM-optimal algorithms to AT^2 Interconnect optimal circuits.
  - Specified programming model that maps PRAM-optimal algorithms in C language to AT^2 Interconnect optimal circuits.
- Other Contributions:
  - Helped review FPGA '20 submissions; Drafted funding proposals.
  - Received two $8,000 National Science Foundation Research Grants for my work.

Goldman Sachs New York City, NY
Research Intern Summer 2018

- *Motivation*: Automating text-based accounting tasks using latest Natural Language Processing (NLP) models.
- *My contribution*: Created custom NLP model for performing a specific accounting classification task. This task was previously performed full time by two employees. The model now performs the task at 95% accuracy without any manual input.
- Interned with team of six PhD's in areas of Machine Learning and NLP.
- Successfully reproduced results from a family of LSTM with Attention papers in NLP using keras and tensorflow. Automated hyperparameter search for our models. Used AWS GPUs to reduce NLP model training time by 5x.
- Used tensorflow, keras, CUDA, jupyter, python, conda, and AWS. Got return offer.

UCLA Institute for the Risk Sciences Los Angeles, CA
Undergraduate Researcher 2016 – 2017

- *Motivation*: As custom-built hardware is 100x more expensive to manufacture than commodity electronics, using commodity components allows for big savings on satellites projects. Due to low reliability of commodity electronics, the amount of redundancy required to meet specifications must be modeled.
- *My contribution*: Created a tool which uses statistical methods to automatically decide how much redundancy is required for a certain low-reliability component to perform to satellite's specification. Tool showed that redundancy can be used to get many commodity components to perform at levels of custom-built parts, while costing 100x less to manufacture. Tool uses probability, statistics, expert opinion, and manufacturer specifications to derive failure probability over time.
- My code was eventually integrated into a software package used by engineers at JPL.

PUBLICATIONS

- **Nikola Samardzic**, Weikang Qiao, Vaibhav Aggarwal, M.C. Frank Chan, Jason Cong. *Bonsai: High-Performance Adaptive Merge Tree Sorting.* International Symposium on Computer Architecture **(ISCA 2020)**
  - *Motivation*: As sorting is often a bottleneck in many data processing pipelines (e.g. MapReduce, Hive, SparkSQL), improving sorters has a direct impact on the performance of these systems. Further, the ability to automatically adapt the sorter to

any type of hardware and problem size is important for hard-to-predict data center workloads.
    - *Results*: Developed state-of-the-art FPGA sorting kernel with higher throughput than all prior work on FPGA, CPU, and GPU for 2-16GB problem size. Simulation results on other hardware and problem size suggest that our model can tune our architecture to achieve 16x latency improvement over previous work in terabyte-scale sorting with NVMe Flash.
    - *Author Contributions*: I developed the merge tree architecture, the Bonsai optimization model, and wrote the paper; others developed the data loader which is able to feed data to the merge tree at peak memory bandwidth, ran the experiments, and worked with me on the implementation of prior work as part of a homework project that started this research.
- **Nikola Samardzic**, Jason Cong. *A Method for Translating PRAM-Optimal Algorithms to AT^2 Interconnect-Optimal Circuits.* **(Preprint)**
    - *Motivation*: There is currently no widely agreed-upon metric for analyzing the asymptotic complexity of circuit designs, although many have been proposed. The *Parallel Random Access Machines (PRAM)* model has been extensively studied, but is considered not sufficiently realistic for measuring scalability of circuits. Conversely, the *AT^2 (Area x Time squared) Interconnect model* is considered a good circuit model as it considers both wire length and resource cost. However, it has not been widely studied. The goal is to show that many optimality statements proven in context of PRAM also apply under the AT^2 model.
    - *Results*: I proved that PRAM-optimal algorithms can be translated intoAT^2 Interconnect-optimal circuit designs under very limited assumptions. This result is significant because: (a) it shows the systolic array scales optimally for a wide range of problems; (b) it defines a simple algorithm for translating PRAM-optimal designs to AT^2 Interconnect optimal ones.
    - *Author Contributions*: The result is entirely my work, but the research topic was suggested and guided by Prof. Cong.

WORK EXPERIENCE

NAND Capital San Francisco, CA
Software Engineering Intern Winter 2020

- Will work on developing an analytics engine for processing terabytes of trading data in real-time in the cloud.
- NAND Capital is an information theory hedge fund backed by Founders Fund and Paradigm.

SpaceX Los Angeles, CA
Software Engineering Intern Summer 2017

- *Motivation*: Develop tool to assist Test Engineers in finding propulsion issues.
- *My Contribution*: Created software to automate defining and testing of all propulsion joints on SpaceX rockets. Created tool for automatically generating repair and inspection plans for propulsion joints. Software saves 7.5 weeks of engineering time per rocket design.

- Was offered to work full time immediately and leave school.

TEACHING EXPERIENCE

- **UCLA Math Circle (2016-2018):** Tutored small groups of gifted high school students to help them prepare for the International Mathematical Olympiad; Created and graded homework sets.
- **Lectures in Mathematical Grammar School, Serbia:** Was invited to give lectures on the separating hyperplane theorem to students at my high school.
- **Computing Theory, UCLA:** Graded homework for the undergraduate computing theory class taught by Prof. Alexander Sherstov.
- **Stochastic Processes (graduate), UCLA**: Prepared lecture slides based on Prof. Suhas Diggavi's notes for his graduate course on stochastic processes.

MISCELLANEOUS

- GRE scores: 163 verbal; 166 math; 4.5 writing
- Won Bronze Medal at International Junior Physics Olympiad in Tehran, Iran (2012).
- Lived in Serbia before coming to college in the US.