## Avani Wildani née Gadani, Ph.D.

| CONTACT<br>INFORMATION     | 580 Hayes St. #201<br>San Francisco, CA, 94102<br><i>c.v.</i> date: March 31, 2022  | <i>Voice:</i> +1 909.437.8626<br><i>E-mail:</i> agadani@gmail.com<br><i>Citizenship:</i> United States  |  |  |
|----------------------------|---|---|--|--|
| Summary                    | As a scientist and an engineer, I am fascinated with information: what it is, how it is stored, how it is accessed, and how it leads to ethical decisions. My doctoral research focused on how patterns in disk accesses allow a system to predict what data is commonly accessed together along with how this knowledge of momentary grouping can be used to make systems more available and efficient.  |   |  |  |
|                            | My current interest is in using quantitative models derived from machine learning to characterize<br>the space of information usage and leverage these rigorous, reproducible models to system tuning<br>as well as understanding high dimensional interconnected systems such as airborne pollutants<br>My lab focuses on exascale and cloud storage systems, though applications exist in archives, deep<br>learning backends, and general distributed systems. |   |  |  |
|                            | We are also pursuing the dual<br>how to arrange storage to mee<br>science the metrics are observa<br>the brain as a storage problem   | of the storage problem. Whereas comp<br>t specific metrics such as fault tolerance<br>ble but the system unknown. I am word<br>to better learn how we collect and inter | puter scientists have defined<br>e and access speed, in neuro-<br>king to model information in<br>pret signals from our world. |  |
| Education                  | <b>The Salk Institute of Biologic</b><br>Postdoctoral Fellow (2013-<br>Advisor: Tatyana Sharpee   | <b>al Studies</b><br>2015)  |  |  |
|                            | <b>University of California</b> , Santa Cruz, CA USA<br>Ph.D. Computer Science (September 2013)<br>Committee: Jiri Schindler, Ahmed Amer, Darrell D.E. Long, Ethan L. Miller (Advisor).   |   |  |  |
|                            | <b>University of New Mexico</b> , A<br>M.S. Computer Science : M<br>Advisor: Terran Lane  | buquerque, NM USA<br>Iachine Learning (May 2007)  |  |  |
|                            | <b>Harvey Mudd College</b> , Clarer<br>B.S. Joint Math / Compute<br>Advisor: Mike Erlinger  | nont, CA USA<br>r Science (May 2003)  |  |  |
| Professional<br>Experience | <b>Emory University:</b> Assistant I  | Professor, Math and Computer Science  | July 2015 - present  |  |
|                            | The Salk Institute: Postdoctor  | al Research Associate   | August 2013 - July 2015  |  |
|                            | Postdoctoral training in the CN   | NL-T lab under Dr. Tatyana Sharpee.   |  |  |
|                            | IBM Almaden Research Cent   | er: Research Intern   | Summer 2011  |  |
|                            | Worked with a small team add<br>mary storage by using my the<br>time systems as well as writing<br>ICDE 2013.   | fressing the disk bottleneck problem in<br>sis work in data grouping. Involved lo<br>g Python libraries to simulate de-duplica  | data de-duplication for pri-<br>w-overhead analysis of real-<br>ation. This work appeared in                                   |  |

## Sandia Labs - Computer Science Research Institute: Research Intern

Analyzed block I/O data from a multi-use production system in order to define and identify similarity between particular data blocks. Looked into creating access groupings based on the similarity data. Research on similarity was published in SYSTOR 2011 and TOS 2016.

#### Google: Intern

Implemented a PID-based control system in Python to automatically balance customers for Bigtable (a key/value store). Extensively researched reward functions and SLA/SLO management techniques and developed a detailed experiment plan to investigate the tradeoffs between economic and standard scheduling techniques. Worked with the Census team to expand their SLO framework to include SLAs.

#### **IBM Almaden Research Center:** Research Intern Summer 2008 (continued through fall)

Worked with a team of 4 researchers to develop a highly accurate method to automatically classify log excerpts from commercially deployed storage systems to speed up problem isolation. This involved coding in Java, particularly the Lucene search engine, and Python for the backend and data-manipulation respectively. I also did a significant portion of the system design and analysis. This work was published in SRDS 2009.

#### University of California, Santa Cruz, CA USA:

Research Assistant

Storage Systems research under Dr. Ethan Miller. (see Projects)

#### Teaching Assistant

Developed lecture slides, developed and graded homework, and ran recitation sections and occasional lectures for Introduction to Programming in Python, for Dr. Darrell Long.

#### Google: Intern

Helped componentize monolithic codebase including visualizing dependency structures in Python, UNIX Shell, and C++ scripts.

#### **Google:** Intern

Documented Internal Build System Components using HTML/CSS, Javascript, and Wiki. Used Python to implement hidden Markov models to analyze web data.

#### University of New Mexico, Albuquerque, NM USA:

#### Research Assistant

Machine Learning research under Dr. Terran Lane. (see Projects)

#### Teaching Assistant

Lead teaching assistant for Introduction to Java Programming. Wrote and graded assignments, Taught Labs, Managed and coordinated with 5 TAs, and maintained the class web site

#### **CENIC:** Systems Administrator

Configured and supported a variety of internal solutions for CENIC's network management. Managed **Cisco optical equipment** (15808 and 15540), and led deployment of a complete Cisco VoIP solution for the office. Also administered 5-7 OS X and Solaris servers including mail, news, Request Tracker trouble ticketing system, and Ciscoworks.

#### Summer 2007

#### Summer 2006

# Aug 2004 - Dec 2004

Jan 2005 - May 2007

# Oct 2003 - May 2004

## **Summer 2010**

**Summer 2009** 

March 2008 - August 2013

January 2012 - May 2012

## Harvey Mudd College, Claremont, CA USA

#### Research Assistant

May 2002 - Jan 2008 IDXP/BEEP Intrusion Detection project at Harvey Mudd College, with Prof. Mike Erlinger, Head of the IDWG. Over Summer 2002, we formed a joint team with engineers from the Aerospace Corporation and Silicon Defense to **implement a functional IDS** using Perl and Java. We also monitored traffic using a modification of the snort and NFR tools and sent these alerts into a MySQL database.

#### Copy Editor and Textbook Assistant

**ETFX ed** and **Edited** the 2<sup>nd</sup> edition of Ordinary Differential Equations: A Modeling Approach, by Profs. Robert Borrelli and Courtney Coleman of Harvey Mudd.

#### Computer Science Student Staff

#### October 1999 - May 2003

Sep 2001 - Aug 2003

Maintain and Secure several Solaris, Linux, and Windows machines. Used shell scripts and limited experience with SQL databases. Also assisted users in using a lab of NCD terminals. Responsible for software installation, OS patching, and user troubleshooting.

- Summer 01 : **Re-structured the entire file-system** using RAID and NetApp.
- Summer 02 : Supervised the upgrading of the entire cluster including gcc and Solaris.

#### Grader/Tutor

#### January 2000 - May 2003

Jun 1997 - May 1999

Graded and Tutored classes including Algorithms, Intermediate Computer Science, Architecture and Operating Systems, Differential Equations, Linear Algebra, and Discrete Mathematics.

#### Museum of Scientific Discovery: Instructor and Exhibit Technician

Taught courses on animal anatomy and optics to groups of 30 middle school students. Maintained and assisted with development of interactive educational museum exhibits.

- 2021 Sandia Labs Research Grant: Cosine Similarity for High Performance Computing \$56,529 (per 6mo; to be renewed in 2022)
  - 2021 NSF Medium Grant: Collaborative Research: CNS Core: Medium: Optimizing Storage Caches via Adaptive and Reconfigurable Tiering \$800,000. Lead PI - Emory
  - 2020 Intel Loihi Research Grant: Real Time MRI Reconstruction via Spiking Deep Denoising Autoencoders \$6000
  - 2018 Google Research Grant: Smart Agents on the Capacity Scaling of Wireless Mesh Networks \$10,000. Single PI.
  - 2018 Intel Labs Equipment Grant: Machine Learning in Storage \$5,000. Single PI.
  - 2018 Emory University Research Council Halle Institute Grant: Developing Satellite-based PM2.5 Air Quality Models in Urban East Asia \$39,816. Co-PI with Dr. Yang Liu.
  - 2018 NSF CRII Grant: CSR: Skeletor: Building a Platform for Quantitative Workload Characterization \$174,860. Single PI.
  - 2018 CMBC Pilot Grant: A theoretical framework for feature–based parameter space investigation in computational neuroscience

\$3,650. Supporting Faculty for PI Dr. David B. Stockton.

- 2018 Center for Faculty Development and Excellence Seed Grant: Understanding Physical Security \$300. Single PI.
- 2017 Google Cloud Grant. \$1900. Single PI
- 2016 NVIDIA Equipment Grant: Deep Network Architectures for Analyzing Ultrasonic Mouse Vocalizations \$1,200. Single PI.
- 2016 Microsoft Azure Educator Grant Award. \$6,600. Single PI.
- 2015 Emory PERS Grant: Data Collection for Characterization of Big Data Workloads \$22,000 awarded. Single PI.
- 2014 Pioneer Postdoctoral Fellowship: Computational Learning Approaches to Fetal and Adult Breast Cancer

#### FUNDING

*Gene Expression Analysis.* \$31,250. Single PI.

2013 NIH T32 Training Grant. *Vision Research* \$50,000 Trainee; PI: Dr. Tom Albright.

AWARDS

PROJECTS

- 2019 Emory Innovation of 2019 Award for Autonomic Capacity Scaling of Wireless Mesh Networks.
  - 2019 Co-Mentor (with Dr. Lars Ruthotto) of Emory Team that won top honors at the International MCM competition.
  - 2014 Pioneer Postdoctoral Fellowship, Pioneer Foundation.
  - 2014 Winner of SciChat-Scientist Idol Competition for best science communication lecture
  - 2014 Google FAST Grant
  - 2014 Grace Hopper Scholar
  - 2013 Best Short Paper Award. Proceedings of the 6th Annual International Systems and Storage Conference (SYSTOR 2013), June 2013.
  - 2013 CRA-W Workshop Fellowship
  - 2012 File and Storage Technologies (FAST) Student Grant Recipient
  - 2010 File and Storage Technologies (FAST) Student Grant Recipient
  - 2009 Grace Hopper Scholar
  - 2007 UC Regents Scholarship
  - 2005 NSF CSEMS Initiative Computer Science and Mathematics (CSAM) Scholarship
  - 2003 International Computation in Modeling, Meritorious research)

CURRENT Machine-Learning for Storage Tuning

To better apply our prediction models, we are also working to characterize and distill features from common workload types with a goal of building a characteristic database of workloads using learned features. Research from these projects has been published at ICDE, SYSTOR, PDSW, MASCOTS, HotDep, and SRDS.

#### Imputation and Analysis for PM2.5 Concentration in Third-World Environments

Combining multiple data sources with high proportions (50+%) of missing values, and building random forests, CNNs, and MCMC models to predict particulate content in the air from AOD with methods that are both interpretable and can leverage the domain expertise of our collaborators. We are also pioneering work in environmental transfer learning. Research from this project has been published in ISES and the Journal for Environmental Science & Technology.

#### Archival Economics

We are studying the economics of storing data of unknown future value for an indeterminate amount of time. We simulate the costs, taking into account factors such as hardware, power, maintenance, and data migration, of different configurations of hardware and performance constraints for data stored and accessed over several human lifetimes. Our current focus is on building heterogeneous storage systems to capture the cost reductions in solid state devices and project future device configurations that would be ideally fit for a permanent archive. Research from this project has been published at MASCOTS, HotStorage, and MSST.

#### Wireless Capacity Scaling

We are exploring super-linearity in the scaling of fixed wireless mesh networks through the implementation of overlays also known as software-defined networks (SDNs). We've built an emulation environment through ns3 and plan to do a hardware deployment in the near future. Results from this work have been published in WiMob, BDA, and NSDI.

#### Computational Neuroscience

Designing models that explore familiar tradeoffs, such as power vs. reliability, in the context of neural power efficiency and dynamics.

#### SELECTED PAST Neural Receptive Field Classification

PROJECTS

Using topological classification to better understand the information network involved in biological vision. Current work involves characterizing populations of neurons in the primary visual cortex by deriving the persistent homology of the electrical spikes the neurons release for various stimuli. We use Python and Matlab for data analytics and simulations. Our preliminary results indicate the presence of distinct cell populations in the presence of natural stimuli. This project was funded by a T32 NIH research grant as well as partially published in ICML-Topology.

#### Biophotonic Data Analysis

Using simple machine learning such as ICA along with careful image alignment and pre-processing, we are working to help researchers from the Waitt Biophotonics Center extract information about neural and glial activation in the rodent spinal cord during motor tasks and eventually correlate this data to activation in the motor cortex. This analysis is novel because the sampling methodology, one-photon sampling affixed to the subject, was developed at Waitt, and so all analysis procedures are being developed in house from scratch.

#### Breast Cancer Subtyping

We apply techniques from machine learning and computational neurobiology to gene network analysis for breast cancer. Our goal is to improve on current hierarchical subtyping methods, which conflate tumor types with divergent prognoses. We have promising early results using shrunken centroid analysis to determine interesting genes to analyze using functional bases techniques, and we have computationally corroborated experimental results showing the importance of an immune pathway in LumB tumors.

#### MIND Institute

This research involved using fMRI images to detect interconnections within the brains of both healthy and schizophrenic patients. From this, we worked on deriving an efficient technique to perform clustering and hidden variable Bayesian Structure Search analyses. I have also used Support Vector Machines to obtain a baseline for our other classification methods, and worked extensively with **Spectral Graph Clustering** methods to handle the patients' categorical data.

#### Sandia National Laboratories Clinic

(Students advised or co-advised are underlined)

Program Manager of a team working with Dr. Kevin Boyack at Sandia to cluster discrete data points and produce a visualization tool to help with large data-set analysis. I implemented density clustering and also researched validity metrics to integrate with the tool we wrote in MATLAB.

## Refereed

PUBLICATIONS

- J6 Vishwanath Seshagiri, Darby Huye, Lan Liu, Raja R Sambasivan, Avani Wildani, Identifying mismatches between microservice testbeds and industrial perceptions of microservices, Accepted; To Appear in Journal of Systems Research, 2022
- J5 Yafeng Wu, Yulai Xie, Xuelong Liao, Pan Zhou, Dan Feng, Lin Wu, Xuan Li, Avani Wildani, Darrell Long, Paradise: Real-time, Generalized, and Distributed Provenance-Based Intrusion Detection, IEEE Transactions on Dependable and Secure Computing, Mar. 2022.
- C23 Si Chen, Jianqiao Liu, Avani Wildani. CENSUS: Counting Interleaved Workloads on Shared Storage, 36th International Conference on Massive Storage Systems and Technology (MSST20), Oct. 2020.
- C22 Tyler Estro, Pranav Bhandari, Avani Wildani, Erez Zadok. Position: Desperately Seeking ... Optimal Multi-Tier Cache Configurations, 12th USENIX Workshop on Hot Topics in Storage and File Systems (HotStor'20), Jul. 2020.
- C21 Sergio Gramacho, Felipe Grmacho, Avani Wildani, Autonomic Formation of Large-Scale Wireless Mesh Networks, IEEE Conference on Network Softwarization (NetSoft 2020), Jun. 2020.

- J4 Jianzhao Bi, Avani Wildani, Howard H Chang, Yang Liu. Incorporating Low-Cost Sensor Measurements into High-Resolution PM2.5 Modeling at A Large Spatial Scale, Environmental Science & Technology, Jan. 2020.
- C20 <u>Sergio Gramacho</u>, Avani Wildani. Autonomic Partitioning for the Smart Control of Wireless Mesh Networks, The 6<sup>th</sup> Intl. Workshop on Cooperative Wireless Networks (WiMOB-CWN2019), Oct. 2019.
  - J3 Jianzhao Bi, Jessica H Belle, Yujie Wang, Alexei I Lyapustin, Avani Wildani, and Yang Liu.*Impacts* of snow and cloud covers on satellite-derived PM2. 5 levels, Remote Sensing of Environment, Feb. 2019.
- C19 <u>Si Chen</u>, Avani Wildani, Chasing the Signal: Statistically Separating Multi-Tenant I/O Workloads, ML for Systems (co-located with NeurIPS 2018), Dec. 2018.
- C18 Juncheng Yang, Reza Karimi, Trausti Sæmundsson, Avani Wildani, Ymir Vigfusson, MITHRIL: Mining Sporadic Associations for Cache Prefetching, ACM Symposium on Cloud Computing 2017 (SoCC '17), Oct. 2017.
  - J2 <u>Xuefei Hu</u>, Jessica Hartmann Belle, Xia Meng, Avani Wildani, Lance Waller, Matthew Strickland, Yang Liu, Estimating PM2. 5 Concentrations in the Coterminous United States Using the Random Forest Approach, Environmental Science & Technology, Jun. 2017.
- C17 Avani Wildani, Fighting for a Niche: An Evolutionary Model of Storage, 9th USENIX Workshop on Hot Topics in Storage and File Systems (HotStor'17), Jul. 2017.
- C16 <u>Wenxuan Wang</u>, Avani Wildani, Snapshot Judgments: Obtaining Data Insights without Tracing, 9th USENIX Workshop on Hot Topics in Storage and File Systems (HotStor'17), Jul. 2017.
- C15 <u>Preeti Gupta</u>, Darrell D. E. Long, Ethan L. Miller, David S.H. Rosenthal, Avani Wildani, Effects of Prolonged Media Usage and Long-term Planning on Archival Systems, 32nd International Conference on Massive Storage Systems and Technologies (MSST2016), May 2016.
- C14 Avani Wildani, Ian Adams, A Case for Rigorous Workload Classification, 23rd IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2015), Oct. 2015.
- J1 Avani Wildani, Ethan L. Miller, Can We Group Storage?: Statistical Techniques to Identify Predictive Groupings in Storage System Accesses. ACM Transactions on Storage, Sep. 2015.
- C13 Avani Wildani, Ethan L. Miller, Ian Adams, Darrell D. E. Long, *PERSES: Data Layout for Low Impact Failures*, 22th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2014), Sep. 2014.
- C12 <u>Preeti Gupta</u>, Avani Wildani, Daniel Rosenthal, Ethan L. Miller, Ian Adams, Christina Strong, Andy Hospodor, An Economic Perspective of Disk vs. Flash Media in Archival Storage, Proceedings of the 22th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2014), Sep. 2014.
- C11 Avani Wildani, Tatyana Sharpee, Persistent Homology for Characterizing Stimuli Response in the Primary Visual Cortex, Proceedings of the 31st International Conference on Machine Learning – Workshop on Topological Methods for Machine Learning. (ICML-Topology 2014), Jun. 2014.
- C10 Avani Wildani, Ian Adams, Ethan L. Miller, Single-Snapshot File System Analysis, Proceedings of

the 21st IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2013), Aug. 2013.

**C9** Ian Adams, Mark W. Storer, Avani Wildani, Ethan L. Miller, Brian Madden, Validating Storage System Instrumentation, Proceedings of the 21st IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2013), Aug. 2013.

- C8 Aleatha Parker-Wood, Brian Madden, Michael McThrow, Darrell D. E. Long, Ian Adams, Avani Wildani, Examining Extended and Scientific Metadata for Scalable Index Designs, Proceedings of the 6th Annual International Systems and Storage Conference (SYSTOR 2013), Jun. 2013.
- **C7** Avani Wildani, Ethan L. Miller, Ohad Rodeh, *HANDS: A Heuristically Arranged Non-Backup In-line Deduplication System*, Proceedings of the 29th IEEE International Conference on Data Engineering (ICDE 2013), Apr. 2013.
- C6 Avani Wildani, Lee Ward, Ethan L. Miller, Efficiently Identifying Working Sets in Block I/O Streams, Proceedings of the 4th Annual International Systems and Storage Conference (SYSTOR 2011), May 2011.
- C5 Avani Wildani, Ethan L. Miller, Semantic Data Placement for Power Management in Archival Storage, Proceedings of the 5th International Workshop on Petascale Data Storage (PDSW 2010), held in conjunction with SC2010, Nov. 2010.
- **C4** Wendy Belluomini, Binny Gill, Avani Wildani, Pin Zhou, GAUL: Gestalt Analysis of Unstructured Logs for Diagnosing Recurring Problem in Large Enterprise Storage Systems, 29th IEEE International Symposium on Reliable Distributed Systems (SRDS 2010), Nov. 2010.
- C3 Ari Rabkin, Wei Xu, Avani Wildani, Armando Fox, Dave Patterson, Randy Katz, A Graphical Representation for Identifier Structure in Logs, Workshop on Managing Systems via Log Analysis and Machine Learning Techniques (SLAML 2010), Oct. 2010.
- **C2** Avani Wildani, Thomas Schwarz, Ethan L. Miller, Darrell D. E. Long, *Protecting Against Rare Event Failures in Archival Systems*, Proceedings of the 17th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2009), Sep. 2009.
- C1 Kevin Greenan, Darrell D. E. Long, Ethan L. Miller, Thomas Schwarz, Avani Wildani, Building Flexible, Fault-Tolerant Flash-based Storage Systems, Proceedings of the Fifth Workshop on Hot Topics in System Dependability (HotDep 2009), Jun. 2009.
- OTHER TR5 **Avani Wildani**, *The Promise of Data Grouping in Large Scale Systems*, Ph.D. Thesis. Sep. 2013. PUBLICATIONS
  - TR4 Avani Wildani, Ethan L. Miller, Ian Adams, Darrell D. E. Long, PERSES: Data Layout for Low Impact Failures, Technical Report UCSC-SSRC-12-06, Sep. 2012.
  - TR3 Avani Wildani, Ethan L. Miller, Ohad Rodeh, HANDS: A Heuristically Arranged Non-Backup Inline Deduplication System, Technical Report UCSC-SSRC-12-03, Mar. 2012.

Avani Wildani, FAST 2011 Work-In-Progress Summaries, USENIX ;login: Magazine, 2011.

Avani Wildani, FAST 2010 Work-In-Progress Summaries, USENIX ;login: Magazine, 2010.

- TR2 Avani Wildani, Thomas Schwarz, Ethan L. Miller, Darrell D. E. Long, Protecting Against Rare Event Failures in Archival Systems, Technical Report UCSC-SSRC-09-03, April 2009. Preliminary version of a paper that appeared in MASCOTS 2009. [Archival Storage] [Reliable Storage]
- TR1 Avani Gadani, Eric Wu, Daniel Lowd, Brian Roney, Belinda Thom, *Implementation of and Experimentation with a Clustering Tool*, Harvey Mudd Clinic Final Report, June 2003.

ACADEMIC SERVICE Journal Reviewer

ACM Transactions on Storage ACM Transactions on Computing ACM Transactions on Parallel and Distributed Systems IEEE Transactions on Services Computing PLoS Computational Biology PLoS One Environmental Science & Technology Applied Sciences IEEE Micro Atmosphere Nature Sustainability

#### **Program Committee Chair**

2020 The 2nd International Workshop for System Characterization (SAC2020)

- 2020 The 18th USENIX Conference on File and Storage Technologies, Work-In-Progress Session (FAST WIPS '20)
- 2018 The 23rd Annual Conference on Modeling, Analysis, and Simulation of Computer and Telecommunication System (MASCOTS 2018). Program committee co-chair
- 2018 The 19th Annual Grace Hopper Celebration of Women in Computer Science. Program committee co-chair: Computer Systems and Engineering.
- 2017 The 18th Annual Grace Hopper Celebration of Women in Computer Science. Program committee co-chair: Computer Systems and Engineering.
- 2016 1st International Workshop in Systems Analytics and Characterization (SAC). Conference co-chair
- 2016 The 17th Annual Grace Hopper Celebration of Women in Computer Science. Program committee co-chair: Computer Systems and Engineering (Inaugural year).

#### Panels

Oct 2018 NSF CSR:CRII Panel

Feb 2017 NSF CSR:Small Panel

Oct 2016 NSF Campus Cyberinfrastructure Panel

### **Program Committee Member**

- 2022 2022 USENIX Annual Technical Conference (USENIX ATC '22)
- 2022 Supercomputing Tutorials (SC-Tut 2022)
- 2021 2021 USENIX Annual Technical Conference (USENIX ATC '21)
- 2021 International Symposium on Computer Architecture (ISCA 2021)
- 2021 Supercomputing Tutorials (SC-Tut 2021)
- 2021 USENIX HotStorage Workshop (HotStor 2021)
- 2021 The 37th IEEE Symposium on Massive Storage Systems and Technologies (MSST2021), Research Track.
- 2021 The 14th ACM International Systems and Storage Conference (SYSTOR '21)
- 2020 IEEE International Conference on Workload Characterization (IISWC 2020)
- 2020 The Workshop on Performance Debugging in Modern Computer Systems (PDMCS '20)
- 2020 The 13th ACM International Systems and Storage Conference (SYSTOR '20)
- 2020 USENIX HotStorage Workshop (HotStor 2020)
- 2020 2020 USENIX Annual Technical Conference (USENIX ATC '20)
- 2020 The 36th IEEE Symposium on Massive Storage Systems and Technologies (MSST2020), Research Track.
- 2020 18th USENIX Conference on File and Storage Technologies (FAST '20)

- 2019 USENIX HotStorage Workshop (HotStor 2019)
- 2019 International Conference on Distributed Computing Systems (ICDCS 2019), Distributed Algorithms and Theory Track
- 2019 The 35th IEEE Symposium on Massive Storage Systems and Technologies (MSST2019), Research Track.
- 2019 2019 USENIX Security and AI Networking Conference (ScAINet 19)
- 2018 Eurosys (External)
- 2017 The 33rd IEEE Symposium on Massive Storage Systems and Technologies (MSST2017), Research Track.
- 2016 The 5th Non-Volatile Memory Systems and Applications Symposium (NVMSA16)
- 2015 The 10th IEEE International Conference on Networking, Architecture, and Storage (NAS 2015).
- 2015 The 4th Non-Volatile Memory Systems and Applications Symposium (NMVSA15)
- 2015 The 16th Annual Grace Hopper Celebration of Women in Computer Science. Data Science Track.
- 2013 The 29th IEEE Symposium on Massive Storage Systems and Technologies (MSST2013), Research Track. Subreviewer

#### **Faculty Committee Service**

- 2019 Emory Computer Science and Informatics graduate admission committee
- 2019 Emory Computer Science faculty search committee
- 2019 Emory faculty planning committee
- 2018 Emory Woodruff Scholars selection committee
- 2017 Emory Computer Science and Informatics graduate admission committee

#### Ph.D. Dissertation Committee

- 2021 Lei Zhang
- 2020 Sergio Gramacho
- 2020 Jianzhao Bi (Environmental Health)
- 2020 Joe Natale (Physics)

#### Ph.D. Dissertation Proposal Committee

- 2022 Vishwanath Seshagiri
- 2022 Reza Karimi
- 2020 Pranav Bhandari
- 2020 Devashish Purandure
- 2019 Sinjoni Mukhopadhyay
- 2019 Sergio Gramacho
- 2019 Joe Natale (Physics)
- 2018 Jianzhao Bi (Environmental Health)
- 2017 Safoora Yousefi
- 2015 Preeti Gupta

#### **Undergraduate Honors Thesis Committee**

2019 Evan Craft 2017 Chi Le

#### **Community Outreach**

500WomeninSTEM:
Member. Panel Moderator for Women/Harassment in STEM panel (200 attendees). Various associated outreach activities.
Odd Salon Fellow:
Public outreach for general science topics including robotics, neuroscience, and physics. (2017-

|               | present)<br>Women who Code:   |
|---------------|---|
|               | Mentor for Emory coding group aimed at improving diversity (2015-present)   |
|               | Harvey Mudd College Alumni Association Board of Governors:  |
|               | Salk Science Day:   |
|               | Taught the public about the work in the computational neuroscience group at the Salk Institute (2014,2015)  |
|               | SciChat Lecturer:   |
|               | WitsOn Mentor:  |
|               | Mentoring women in science (Piazza group)   |
|               | Taught Introduction to Soldering, 2011  |
|               | CS Graduate Student Association Treasurer, 2006-2007  |
|               |   |
| INVITED TALKS | Myths and Mysteries of Microservices, Rivian. January 2022.   |
|               | Data Skeletons: I/O Workload Characterization for the Modern Age, NERSC. January 2020.<br>Privacy in AI Systems (Panelist + Moderator), 2019 USENIX Security and AI Networking Conference |
|               | (ScAINet '19).  |
|               | Ethics in Computer Science (Panel Discussion), TiGer at Emory Law, March 2019.  |
|               | NeuroComputing 101, Invited Talk: Lesbians Who Tech and Allies, February 2019.  |
|               | Your Cerebellum as an Attack Surface: How Does the Brain Stay Secure?, Invited Talk: SchmooCon XIV, January 2018  |
|               | January 2010.   |
|               | Skeletons in Your Data: Why and How Should We Characterize Storage Workloads?, Invited Talk: IBM Almaden Research Center, October 2017.   |
|               | <i>New Interfaces in Neural Computing</i> , Keynote presentation, 2017 ACM Richard Tapia Celebration of Diversity in Computing, September 2017.   |
|               | <i>Skeletons in Your Data: Why and How Should We Characterize Storage Workloads?,</i> Invited Talk: Intel Labs, August 2017.  |
|               | Towards music process understanding through deep neural network song feature identification in mice and song birds, Invited Talk: Pandora December 2016.                                  |
|               | Feedback Mechanisms in Neural Network Validation, Invited Talk: University of Reykjavik. June 2016.   |
|               | Can We Group Storage?, Invited Talk: Seagate. February 2015.  |
|               | Topological Characterization of Visual Invariants, Salk Science Retreat: Lightning Talk. October 2014.  |
|               | <i>Emerging Topics in Computational Biology: Beyond Gene Sequencing,</i> Technical Talk (Data Science Track): Grace Hopper Celebration of Women in Computing. October 2014.               |
|               | Data Grouping and Vision, Invited Talk: Vicarious. September 2014.  |
|               | The Promise of Data Grouping, Invited Talk: IBM Research. August 2013.  |
|               | PERSES, Invited Talk: NetApp. November 2012.  |
|               |   |
|               |   |

| Other<br>Presentations           |       | Pranav Bhandari, Avani Wildani, Dimitris Skourtis, Vasily Tarasov, Deepavali Bhagwat, Lukas Rupprecht, Ali Anwar, Position: Can Microservices Drive a Renaissance in Workload-Aware Storage Management?, HotStorage Short Presentation.                          |  |  |
|----------------------------------|-------|--|--|--|
|                                  |       | <u>Vishwanath Seshagiri</u> , Abigail Julian, <b>Avani Wildani</b> , <i>Clustering Workloads for Storage Optimisation</i> .<br>FAST 2020 Work-in-Progress. (Talk + Poster).  |  |  |
|                                  |       | Si Chen, Jianqiao Liu Avani Wildani, CENSUS: Counting Interleaved Functional Tenants on Shared Stor-<br>age. FAST 2020 Work-in-Progress. (Talk + Poster).  |  |  |
|                                  |       | Si Chen, Avani Wildani, Chasing the Signal: Statistically Separating Multi-Tenant I/O Workloads. FAST 2019 Work-in-Progress (Talk + Poster).   |  |  |
|                                  |       | Pranav Bhandari, Avani Wildani, Automated Metric Extractor for Filesystem Traces, FAST 2018 Work-<br>in-Progress. (Talk + Poster)  |  |  |
|                                  |       | Juncheng Yang, Reza Karimi, Avani Wildani, Ymir Vigfusson, A Simple Cache Prefetching Layer<br>Based on Block Correlation, FAST 2017 Work-in-Progress. (Talk + Poster)   |  |  |
|                                  |       | Zhike Zhang, <u>Preeti Gupta</u> , <b>Avani Wildani</b> , Ignacio Corderi, and Darrell D.E. Long, <i>Reverse Deduplication: Optimizing for Fast Restore</i> , FAST 2013 Work-in-Progress. (Talk + Poster)  |  |  |
|                                  |       | Avani Wildani, Ethan L. Miller, Grouping Data for Faster Rebuilds: The Art of Failing Silently, FAST 2012 Work-in-Progress. (Talk+Poster)  |  |  |
|                                  |       | Avani Wildani, Ethan L. Miller, <i>Probabilistic Reputation for Personal Trust Networks</i> , FAST 2009 Work-<br>in-Progress. (Talk+Poster)  |  |  |
| Press A                          | Aug   | 2015 New faculty members showcase interdisciplinary strengths of Emory College by Laura Douglas-Brown. Emory News Center.  |  |  |
| Dec                              |       | http://news.emory.edu/stories/2015/08/er_bts_new_faculty/campus.html<br>2016 Mind over matter: Her interest in the brain led her to computer science<br>by Carol Clark eScienceCommons   |  |  |
| De                               | Dec 2 | http://news.emory.edu/stories/2016/12/esc_mind_over_matter_wildani/campus.html<br>2014 Next Generation: Avani Wildani. Inside Salk.  |  |  |
|                                  |       | https://inside.salk.edu/wp-content/uploads/2016/05/inside_salk_12-14.pdf   |  |  |
| Refereed Poster<br>Presentations |       | Pranav Bhandari, Dimitri Skourtis, Vasily Tarasov, Deepavali Bhagwat, Lukas Rupprecht, Ali An-<br>war, <b>Avani Wildani</b> <i>Position: Can Microservices Drive a Renaissance in Workload-Aware Storage Man-</i><br><i>agement?</i> . HotStorage'20. July 2020. |  |  |
|                                  |       | <u>Vishwanath Seshagiri</u> , Abigail Julian, <b>Avani Wildani</b> , <i>Clustering Workloads for Storage Optimisation</i> .<br>FAST 2020 Work-in-Progress. (Talk + Poster).  |  |  |
|                                  |       | <u>Si Chen, Jianqiao Liu</u> <b>Avani Wildani</b> , <i>CENSUS: Counting Interleaved Functional Tenants on Shared Stor-age</i> . FAST 2020 Work-in-Progress. (Talk + Poster).   |  |  |
|                                  |       | Si Chen, Avani Wildani, Chasing the Signal: Statistically Separating Multi-Tenant I/O Workloads. FAST 2019 Work-in-Progress (Talk + Poster).   |  |  |
|                                  |       | <u>Sérgio Gramacho</u> , <b>Avani Wildani</b> , <i>Capacity Scaling on Self-Organizing Wireless Mesh Networks</i> . NSDI 2019.   |  |  |
|                                  |       |  |  |  |

<u>Jianzhao Bi</u>, **Avani Wildani**, Yujie Wang, Alexei Lyapustin, Yang Liu, *Incorporating Snow and Cloud Fractions in Random Forest to Estimate High Resolution PM2.5 Exposures in New York State*, ISEE Conference 2018.

<u>Sérgio Gramacho</u>, **Avani Wildani**, *Smart Agents on the Capacity Scaling of Wireless Mesh Networks*, BRASCON 2018.

<u>Pranav Bhandari</u>, **Avani Wildani**, *Automated Metric Extractor for Filesystem Traces*, FAST 2018 Workin-Progress. (Talk + Poster)

Juncheng Yang, Reza Karimi, Avani Wildani, Ymir Vigfusson, A Simple Cache Prefetching Layer Based on Block Correlation, FAST 2017 Work-in-Progress. (Talk + Poster)

Jianzhao Bi, B. Vu, A. Wildani, Yujie Wang, A. Lyapustin, Yang Liu ,*Citywide Validation and Improvement of the MAIAC Aerosol Product in Lima, Peru*, 27th meeting of the International Society of Exposure Science (ISES 2017), Oct 2017.

<u>Clarissa Tuxen</u>, <u>Sérgio Gramacho</u>, **Avani Wildani**, *Scaling in Socially-Driven Computer Networks*, Grace Hopper Celebration (GHC 2016), Oct 2016. **Best Undergraduate Poster** 

<u>Clarissa Tuxen</u>, Sérgio Gramacho, **Avani Wildani**, *Socially Driven Computer Networks*, Biological Distributed Algorithms (BDA 2016), Aug 2016.

Avani Wildani, Ian F. Adams, A Case for Rigorous Workload Classification, FAST 2015 Work-in-Progress.

<u>Preeti Gupta</u>, **Avani Wildani**, Ethan L. Miller, Darrell D.E. Long, David S.H. Rosenthal, *Which Media Should Be Used In Long-term*, FAST 2015.

**Avani Wildani**, Tatyana Sharpee, *Characterizing Vision with Persistent Homology*, Joint Symposium on Neural Computation, May 2014.

Zhike Zhang, <u>Preeti Gupta</u>, **Avani Wildani**, Ignacio Corderi, and Darrell D.E. Long, *Reverse Deduplication: Optimizing for Fast Restore*, FAST 2013 Work-in-Progress. (Talk + Poster)

Ian F. Adams, Ethan L. Miller, Mark W. Storer, **Avani Wildani**, and Yangwook Kang, *Improved Analysis and Trace Validation Using Metadata Snapshots*, FAST 2013.

**Avani Wildani**, and Ethan L. Miller, *Fault Isolation: Dynamically Re-arranging Data to Increase Availability*, FAST 2013.

Yan Li, Christina Strong, Ignacio Corderi, **Avani Wildani**, Aleatha Parker-Wood, Andy Hospodor, Thomas M. Kroeger, and Darrell D. E. Long, *Energy-Aware Storage*. FAST 2013.

**Avani Wildani**, Ethan L. Miller, *Grouping Data for Faster Rebuilds: The Art of Failing Silently*, FAST 2012 Work-in-Progress. (Talk+Poster)

**Avani Wildani**, Ethan L. Miller, *Probabilistic Reputation for Personal Trust Networks*, FAST 2009 Workin-Progress. (Talk+Poster)

#### TEACHING Courses Designed

**Analysis of Algorithms**. Theory of algorithms, dynamic programming, graph algorithms, and basic theory of computation for junior and senior computer science majors.

Advanced Computer Systems (Graduate) Developed course to cover major topics in distributed systems including virtualization, concurrency, security, and reliability, along with some current

research in field.

**Neural Computing** (Graduate) Developed course to cover current research in neural computing and relationships between biological, computational, and thermodynamic neural models.

Instructor of Record

- Fall 2021 Analysis of Algorithms (CS 326), Computer Science Department, Emory University
- Spr 2021 Graduate Computer Systems (CS 562), Computer Science Department, Emory University
- Spr 2021 Analysis of Algorithms (CS 326), Computer Science Department, Emory University
- Fall 2020 Sabbatical (See Leaves)
- Spr 2020 Analysis of Algorithms (CS 326), Computer Science Department, Emory University
- Spr 2019 Data Structures and Algorithms (CS 323), Computer Science Department, Emory University
- Spr 2019 Graduate Computer Systems (CS 562), Computer Science Department, Emory University
- Fall 2018 Data Structures and Algorithms (CS 323), Computer Science Department, Emory University
- Spr 2018 *Leave* (See Leaves)
- Fall 2017 Leave (See Leaves)
- Fall 2017 Topics in Advanced Computer Systems (CS 584), Math and Computer Science Department, Emory University (Class taught partially online)
- Spr 2017 Introduction to Computer Science II (CS 171), Math and Computer Science Department, Emory University
- Spr 2017 Data Structures and Algorithms (CS 323), Math and Computer Science Department, Emory University
- Fall 2016 *Leave* (See Leaves)
- Fall 2016 Neural Computing (CS 584-0), Joint Neuroscience Program, Emory University (Class taught online)
- Spr 2016 Introduction to Computer Science II (CS 171), Math and Computer Science Department, Emory University
- Spr 2016 Topics in Advanced Computer Systems (CS 584), Math and Computer Science Department, Emory University
- Fall 2015 Introduction to Computer Science II (CS 171), Math and Computer Science Department, Emory University

Every semester, I also taught Graduate Directed Study (CS 597R), Dissertation Research (CS 799), and frequently Rotation Project (CS 598R) and Undergraduate Directed Study (CS 497R).

Teaching Assistant

University of California, Santa Cruz: Lead TA for **Introduction to Programming in Python**. Wrote most course material including lecture slides and assignments, taught labs, and coordinated other TAs and graders. 72% of reviews "above average."

University of New Mexico: Lead TA for **Introduction to Java Programming**. Wrote and graded assignments, Taught Labs, Managed 5 TAs, and maintained the class web site.

Harvey Mudd College: Grader and Tutor for - Algorithms and Data Structure, Discrete Mathematics, Linear Algebra 1, Differential Equations, Architecture and Operating Systems, CS for Smart People

#### STUDENTS Ph.D. Students

- 2020 present Shrey Gupta
- 2020 present Vishwanath Seshagiri
- 2017 present Pranav Bhandari
- 2017 present Si Chen
- 2017 2020 Jianzhao Bi: "Assessment of High-Resolution PM2.5 Exposures and Changes in PM2.5 Cardiorespiratory Disease Associations Over Time" (co-advised with Prof. Yang Liu, Dept. of Environmental Health). Postdoc at University of Washington.
- 2016 2020 Sergio Gramacho: "Autonomic Formation of Large Scale Wireless Mesh Networks." Software Engineer at Emory University.

2013 - 2016 Preeti Gupta (ABD). Independent educational consultant.

#### M.S. Students

- 2019 2020 Ana Noriega: "The Importance of Allowing Cycles / Some Qualities of Biologically Structured Random Block Matrix Networks."
- 2016 2018 Benjamin Bolte: "Generative Adversarial Networks for Birdsong Classification." Software Engineer at Facebook

#### **Undergraduate Research Students**

- 2016 2019 Safa Tinaztepe. Amazon Web Services.
- 2017 2019 Eli Hanover. Blockchain Startup.
- 2017 2019 Justus Schmidt.
- 2016 2018 Matt Ulmer. Local IT.
- 2015 2017 Wenxuan Wang. Finance Industry.
- 2017 2017 Chi Le (Senior Honors Project): "Modeling chloride dynamics to examine homeostatic plasticity in central pattern generators." Medical School.

#### **Rotation Students**

|                      | Shrey Gupta<br>Vishwanath Seshagiri<br>Ana Noriega<br>Si Chen<br>Pranav Bhandari<br>Jianzhao Bi<br>Mani Sotoodeh<br>Hale Soloff<br>Sergio Gramacho  |
|----------------------|---|
| Software<br>Released | Perses Disk Fault Simulator: https://bitbucket.org/agadani/perses-code<br>pmv automatic PDF renamer: https://github.com/agadani/pdf_namer<br>Vim bindings for ciao  |
| Tools                | <ul> <li>Programming and Markup Languages I've used for large projects:<br/>Python (including NumPy/Numeric, SciPy, and SimPy), C++, Java, Prolog, MATLAB, LATEX, Shell Scripting, XML, HTML</li> <li>Programming Languages I have used for smaller projects:</li> </ul>          |
|                      | Ruby, Lua, R, C, SML, Lisp, Ciao, Rex, ns2 simulator, Tcl   |
| LEAVES               | Sep 2016 - Jan 2017: Recovery from pedestrian-automobile collision (bilateral tibia-fibia fractures).<br>Nov 2017 - Feb 2018: Recovery from on-campus fall (ulna, radius, femur, and tibia fractures)<br>Feb 2018 - May 2018: Maternity Leave<br>Fall 2020: Pre-Tenure Sabbatical |
| AFFILIATIONS         | Affiliate Faculty: Pioneer Institute.<br>Core Faculty: Emory Center for Mind, Brain, and Culture<br>Associate Faculty: Emory Neuroscience Program<br>Member: ACM, IEEE, USENIX, SfN<br>Affiliate: Storage Systems Research Center (SSRC), UC Santa Cruz                           |