

## EDUCATION

- Carnegie Mellon University** Pittsburgh, PA, USA 2017 – Present
- Ph.D. student, Computer Science Dept. Member of Parallel Data Lab. Advisor: **Greg Ganger**.
  - Area: Software Systems. Topics: Systems for ML, ML for Systems, Distributed Storage Systems.
- University of Cambridge** Cambridge, United Kingdom 2013 – 2016
- B.A. (2.i) in **Computer Science** Tripos, at Churchill College. 1<sup>st</sup> year NatSci option: Physiology of Organisms.
  - **'Highly commended dissertation' award** for final year dissertation "*Compartmentalization of cryptographic components (OpenSSL)*". Advisor: Robert Watson.

## PHD RESEARCH

- ML for caching Spring 2020 – **Present**
- *Goal*: ML-based caching admission policy. Utilizing traces from PDL-Facebook collaboration.
  - *Role*: Lead student. Advisors: Daniel Berger, Nathan Beckmann, Greg Ganger.
- Cheap replication for distributed NVM storage. Spring 2019 – **Present**
- *Goal*: Performant and highly-available distributed replicated storage with smarter recovery.
  - *Role*: Building out the system w/ Thomas Kim. Advisors: Michael Kaminsky, Greg Ganger, Dave Andersen.
- Selective Backpropagation Aug 2018 – *Present*
- *Goal*: Accelerate DNN training by prioritizing examples with high loss at each iteration. *Role*: 2<sup>nd</sup> student.
  - *Status*: preprint on arXiv. *Under submission*. Paper: <https://arxiv.org/abs/1910.00762>, Code: <https://github.com/angelajiang/SelectiveBackprop>
- Evolutionary Weight Sharing for Neural Architecture Search. Jul 2018 – Dec 2018
- *Goal*: Make Neural Architecture Search faster, w/ evolutionary NAS controller guiding shared weights training.
  - Worked with Liam Li, Angela Jiang, Ameet Talwalkar, Greg Ganger.
- Mainstream: Dynamic Stem-Sharing for Multi-Tenant Video Processing. Sep 2017 – Jul 2018
- *Goal*: Optimize serving of multiple real-time video processing DNNs by varying sharing & sampling rate.
  - *Role*: Formulated & built a dynamic programming scheduler to replace greedy scheduler, reducing scheduling time from exponential to polynomial and allowing optimization of other objectives, e.g., max-min fairness.
  - *Published*: short paper at SysML '18, full paper in USENIX ATC '18. Open-sourced: <https://viscloud.github.io/mainstream-web/>

## WORK EXPERIENCE

### Google

- Software Engineer Intern – Platforms Performance Team. May 2019 – Aug 2019
- Student Researcher – Brain ML for Systems Team Sep 2019 – Dec 2019
- **Machine learning systems**. Towards automatic model parallelism: co-optimization of operator scheduling and device placement in TensorFlow. Collaborators: Peter Ma (host), Sudip Roy (co-host), Yanqi Zhou.
- Institute of High Performance Computing, A\*STAR, Singapore** Oct 2016 – Jul 2017

### AI Group Research Engineer.

- **Deep learning**. Explored deep RL and LSTMs for neural architecture and hyperparameter search. Contributed a workflow scheduler to ModStore, a data science experiment platform.
- Dropbox**, Software Engineer Intern – Vacuuming team. Jun 2015 – Sep 2015
- **Distributed systems**. Rewrote vacuuming system in Go to implement a robust and efficient streaming pipelined design (for petabyte-scale garbage collection of deleted Dropbox files). Implemented a distributed jobs service running across hundreds of dedicated machines. **Received full-time return offer**.
  - Removed tight coupling and reduced operational overhead by ensuring automatic failover during MySQL promotions. Decreased storage overhead by increasing system reliability, allowing it to run continuously without intervention, enabling significant cost savings and a lower storage margin (less data awaiting vacuuming) essential for Project BASE Jump (migration from AWS to Dropbox data centers).
- Facebook**, Software Engineer Intern – Core Data Cache Client team. Jul 2014 – Sep 2014
- **Distributed storage**. Improved cache consistency in widely distributed storage systems (Memcache backed by MySQL DBs, thousands of machines spread across multiple continents). **Received return offer**.
  - Resolved race condition in cache invalidation mechanism that caused read-after-write inconsistencies in geographical regions with multiple MySQL database replicas.
  - Worked across stack to diagnose & implement fix, ranging from: cache invalidation, Memcache PHP client, Wormhole (pub-sub), TAO (distributed graph data store), MySQL, Ads Infra. Coded in PHP/Hack, C++.

DANIEL WONG

## RESEARCH ATTACHMENTS

**School of Computing**, National University of Singapore. Jul 2016 – Jul 2017

- *Advisor*: Ben Leong. **Sensor systems** research on small slot sizes. Built a high-fidelity simulator, devised a theoretical model and new stochastic broadcasting patterns for faster discovery.
- *Published at IEEE MASS 2018*.

**Data Analytics Dept.**, Institute for Infocomm Research (I<sup>2</sup>R), A\*STAR, Singapore Jan 2013 – Sep 2013

- **Research Intern**. *Advisors*: Xiao-Li Li, See-Kiong Ng. Devised stochastic graph **clustering** algorithm beating state-of-the-art methods on protein interaction data. **Travel award** for oral ppt at Int'l Conf on Bioinformatics.
- *Published and presented at InCoB 2013*.

## CODING COMPETITIONS

- **Bronze** medal at **International Olympiad in Informatics** (2010).
- **Silver** in Asia-Pacific Informatics Olympiad and **Gold** in National Olympiad in Informatics.

## SELECTED DATA MINING COMPETITIONS

- *Recommendation Systems*. The Hut Challenge – 4<sup>th</sup> (2015), 6<sup>th</sup> (2014). Built Jenkins-based ML pipelines with ensembling of random forests, random walks (PLW) and collaborative filtering.
- *Data Mining*. ACM SIGKDD Cup – Top 25% (2013). Did text-mining & relational mining of publication metadata to generate confidence scores, then did hierarchical clustering to de-duplicate authors.
- *Data Mining*. Quora ML Codesprint. Top 25% (Jul 2013). Did feature engineering, utilized ensemble methods (e.g., random forests and gradient boosting machines) and sparse linear models in scikit-learn.

## SELECTED SECURITY COMPETITIONS

- Cybersecurity Challenge S'pore – Winner (2017). Organized by S'pore Cyber Security Agency and BAE Systems.
- Represented Singapore at Cybersecurity Challenge UK Masterclass in London.
- MIT CSAIL-Cambridge C2C Capture The Flag (CTF) – Team 2<sup>nd</sup> (2016). Challenges on binary exploitation (buffer overflows, return-oriented programming), remote code execution, fuzzing, password-cracking, lock-picking.
- Facebook-Cambridge Ethical Hacking Capture the Flag (CTF) – Team 1<sup>st</sup> (2015).

## SERVICE AND TEACHING

- Co-reviewer for ASE/IEEE BigData and IEEE ICDM.
- Coach for Informatics Olympiad trainings. Delivered lectures on algorithms & data structures, conducting lab practical sessions and organizing mock competitions and training camps.

## PUBLICATIONS

- Angela Jiang, Daniel Wong, Christopher Canel, Ishan Misra, Michael Kaminsky, Michael Kozuch, Padmanabhan Pillai, David Andersen, Greg Ganger, "Mainstream: Dynamic Stem-Sharing for Multi-Tenant Video Processing", USENIX Annual Technical Conference 2018 (USENIX ATC'18).
- Angela Jiang, Christopher Canel, Daniel Wong, Michael Kaminsky, Michael Kozuch, Padmanabhan Pillai, David Andersen, Greg Ganger. "Dynamic Stem-Sharing for Multi-Tenant Video Processing" Conference on Systems and Machine Learning 2018 (SysML'18).
- Shuaizhao Jin, Xiangyun Meng, Daniel Lin-Kit Wong, Zixiao Wang, Ben Leong, Yabo Dong, Dongming Lu. "Improving Neighbor Discovery by Operating at the Quantum Scale" IEEE International Conference on Mobile Ad-hoc and Sensor Systems 2018 (IEEE MASS '18).
- Daniel Lin-Kit Wong, Xiao-Li Li, Min Wu, Jie Zheng and See-Kiong Ng, "PLW: Probabilistic Local Walks for detecting protein complexes from protein interaction networks", BMC Genomics 2013 (IF: 4.4) / International Conference on Bioinformatics (InCoB) 2013.

## LANGUAGES & TECHNOLOGIES

- Languages: **Python, C, C++, Go, PHP/Hack**, SQL, Standard ML, Java, Prolog, HTML/CSS/JavaScript
- Frameworks: **TensorFlow, Keras**, scikit-learn, Protocol Buffers, Memcache, Redis, MongoDB, D-Bus (libdbus)
- Tools: GDB, DTrace, Git/Mercurial, Jenkins, Phabricator, SAP CUA/R3/BI
- Written & spoken languages: English (as 1<sup>st</sup> language), Chinese (Mandarin)

## MILITARY EXPERIENCE

Combat Service Support Command, Singapore Armed Forces Apr 2011 – Dec 2012

- *SAP Authorization Administrator*. Built a license tracking & CRM application to process 1,500 yearly user requests, enabling existing assets to support 40% more users (\$XX,000 in cost avoidance). Bridged gap b/w business process owners & SAP consultants to tailor solutions for improved productivity. Streamlined processes, cut turnaround from 2 weeks to 2 days. Promoted to Corporal First Class (<15% of cohort.)

Testimonial: <https://wonglkd.fi-de.net/military/testimonial.pdf>