https://wonglkd.fi-de.net
https://linkedin.com/in/wonglkd
https://github.com/wonglkd

DANIEL WONG

wonglkd@cmu.edu

EDUCATION

Carnegie Mellon University *Pittsburgh, PA, USA* 2017 – 2024

- Ph.D. student, Computer Science Dept. Member of Parallel Data Lab. Advisor: **Greg Ganger**.
- *Committee:* Greg Ganger (chair), Daniel S Berger, Nathan Beckmann, Dave Andersen.
- *Thesis:* Machine learning for flash caching in bulk storage systems.
- *Area:* Software Systems. *Topics:* ML for Systems, Systems for ML, Distributed Storage Systems.
- *Teaching Assistant,* 15-719 Adv Cloud Computing (Spring 2020 & 2024).
- Coursework (systems, theory, PL, ML, neuroscience): *Adv Cloud Computing*, *Neural Signal Processing*, *Analytical Performance Modeling & Design of Computer Systems*, *Probabilistic Graphical Models*, *Logical Foundations of Cyber-Physical Systems*, *Adv Operating Systems & Distributed Systems*, *Computer Architecture.*

University of Cambridge *Cambridge, United Kingdom* 2013 – 2016

- B.A. (2.i) in **Computer Science** (*with Physiology*) at Churchill College. Dissertation advisor: Robert Watson.
- '**Highly commended dissertation'** award for **"***Compartmentalization of cryptographic components (OpenSSL)"*.

PHD RESEARCH

Automatic mitigation of workload drift in caching Jan 2024 – **Present**

- *Problem:* ML for systems models can age poorly, leading to accuracy drops despite retraining.
- *Goal:* Automatic detection & mitigation of drift in caching.

Machine Learning for Flash Caching. Jan 2020 – Jan 2024

- Lead student. *Goal:* ML for smart cache admission & prefetching. Collaboration with Meta's CacheLib team.
- *Advisors:* Daniel Berger, Nathan Beckmann, Greg Ganger. *Published* at FAST; awarded USENIX Student Grant.

High availability in cheap distributed key value storage with NVM. Feb 2019 – Aug 2020

- *Goal:* Use a new fast crash recovery protocol, enabling low-latency storage at lower cost & higher availability.
- Work with Thomas Kim. *Advisors*: Michael Kaminsky, Greg Ganger, Dave Andersen. *Published* at ACM SoCC.

Selective Backpropagation. Aug 2018 – May 2020

- Sped up DNN training by prioritizing examples with high loss. Work with A. Jiang. *Published* on arXiv.

Evolutionary Weight Sharing for Neural Architecture Search (NAS). Jul 2018 – Dec 2018

- *Goal:* Faster NAS. Worked with Liam Li, Ameet Talwalkar, Angela Jiang, 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:* Devised dynamic programming scheduler, making scheduling tractable (from exponential to polynomial time) & enabling other objectives, e.g., max-min fairness. Work with A. Jiang. *Published* at SysML, USENIX ATC.

WORK EXPERIENCE

Google

Software Engineer Intern – Platforms Performance Team. May 2019 – Aug 2019

Student Researcher – Google Brain, ML for Systems Team Sep 2019 – Dec 2019

- **Machine learning systems**. Co-optimization of operator scheduling & device placement in TensorFlow for automatic model parallelism. Optimized RNNLM and devised a multi-segment version of Transformer-XL.
- Collaborators: Peter Ma, Sudip Roy, Yanqi Zhou. *Published* in IEEE Micro and NeurIPS. 2 patents filed.

Institute of High Performance Computing, A*STAR, Singapore Oct 2016 – Jul 2017

AI Group Research Engineer.

- **Deep learning**. Explored deep RL for neural architecture search 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 storage**. Helped automate vacuuming (petabyte-scale garbage collection of deleted files). Did design review, rewrote system in Go, eliminated biggest failure source (MySQL promotions). Reduced storage margin and helped team hit milestones for migration from AWS to on-prem. **Received full-time return offer**.

Facebook, Software Engineer Intern – Core Data Cache Client team. Jul 2014 – Sep 2014

- **Distributed storage**. Fixed cache inconsistency in DataTypes (Memcache backed by MySQL). Made cache invalidation (Mars) replica-aware, eliminating read-after-write inconsistencies due to race between MySQL replicas. Worked in PHP/Hack & C++ across: Memcache, Wormhole (pub-sub) / glostic, Ads Infra.

Continued on next page

DANIEL WONG

PUBLICATIONS

1. Daniel Lin-Kit Wong, Hao Wu, Carson Molder, Sathya Gunasekar, Jimmy Lu, Snehal Khandkar, Abhinav Sharma, Daniel S. Berger, Nathan Beckmann, Gregory R. Ganger. "Baleen: ML Admission & Prefetching for Flash Caches". *USENIX FAST* 2024.
2. Yanqi Zhou, Sudip Roy, Amirali Abdolrashidi, Daniel Wong, Peter Ma, Qiumin Xu, Hanxiao Liu, Phitchaya Phothilimtha, Shen Wang, Anna Goldie, Azalia Mirhoseini, James Laudon. "Transferable Graph Optimizers for ML Compilers.". NeurIPS 2020.
3. Thomas Kim, Daniel L.-K. Wong, Gregory R. Ganger, Michael Kaminsky, David G. Andersen. "High availability in cheap distributed key value storage." *ACM Symposium on Cloud Computing* 2020 (SOCC'20).
4. Yanqi Zhou, Sudip Roy, Amirali Abdolrashidi, Daniel Wong, Peter C. Ma, Qiumin Xu, Ming Zhong, Hanxiao Liu, Anna Goldie, Azalia Mirhoseini, James Laudon. "GDP: Generalized Device Placement for Dataflow Graphs". arXiv (Sep 2019).
5. Yanqi Zhou, Sudip Roy, Amirali Abdolrashidi, Daniel Wong, Peter Ma, Qiumin Xu, Azalia Mirhoseini, James Laudon. "A Single-shot Generalized Device Placement for Large Dataflow Graphs". IEEE Micro (Aug 2020).
6. Angela H. Jiang, Daniel L.-K. Wong, Giulio Zhou, David G. Andersen, Jeffrey Dean, Gregory R. Ganger, Gauri Joshi, Michael Kaminksy, Michael Kozuch, Zachary C. Lipton, Padmanabhan Pillai. "Accelerating Deep Learning by Focusing on the Biggest Losers". arXiv. Oct '19.
7. Angela Jiang, Daniel Lin-Kit Wong, Christopher Canel, Ishan Misra, Michael Kaminsky, Michael A. Kozuch, Padmanabhan Pillai, David G. Andersen, Gregory R. Ganger. "Mainstream: Dynamic Stem-Sharing for Multi-Tenant Video Processing" *USENIX ATC* 2018.
8. 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).
9. 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).
10. 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.

OTHER RESEARCH EXPERIENCE

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 '18.*

Data Analytics Dept., Institute for Infocomm Research (I²R), A*STAR, Singapore Jan 2013 – Sep 2013

- *Advisors*: Xiao-Li Li, See-Kiong Ng. **First author** paper on stochastic **graph clustering** algorithm for protein interactions. **Travel award** for talk at *Int'l Conf on Bioinformatics*. *Published in BMC Genomics.*

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 2nd (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 1st (2015).

SELECTED DATA MINING COMPETITIONS

- *Recommendation Systems.* The Hut Challenge – 4th (2015), 6th (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), Quora ML Codesprint – Top 25% (2013). Text mining, relational mining, hierarchical clustering for author deduplication, feature engineering.

SELECTED CODING COMPETITIONS

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

SERVICE

- Co-reviewer for ASE/IEEE BigData, IEEE ICDM.
- Coach for Informatics Olympiad. Taught algorithms & data structures to high school students.

LANGUAGES & TECHNOLOGIES

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

MILITARY EXPERIENCE

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

- *SAP Authorization Administrator*. Built a license tracking & CRM app enabling ($XX,000 in cost avoidance from supporting 40% more users with existing assets). Bridged gap b/w business process owners & SAP consultants to customize solutions. Streamlined process, cutting turnaround from 2 weeks to 2 days. Promoted to Corporal First Class (<15% of cohort.) Testimonial: https://wonglkd.fi-de.net/military/testimonial.pdf