

# Theodore M. Wong, PhD 黃明道

tmw@tmwong.org

<http://www.tmwong.org>

## Current position

**23andMe, Inc.**, Sunnyvale, CA

*Engineering Manager, Health Machine Learning (September 2022 - Present)*

*Team Lead (July 2021 - September 2022)*

*Technical Lead (April 2021 - July 2021)*

*Senior Software Engineer (January 2018 - April 2021)*

Manager for the Health Machine Learning Engineering team, which designs and builds the pipelines for training machine learning models underpinning many of the direct-to-consumer reports in the 23andMe Health product. My responsibilities include managing project timelines and implementation quality, mentoring team members on their career development, and collaborating with other cross-departmental teams. Additionally, while at 23andMe I have worked on projects to:

- Optimize and maintain the pipeline for training polygenic risk score models for the Health product.
- Design and implement a production pipeline to perform genome-wide association studies over 23andMe customer data.
- Modularize the prediction of customer genetic traits from the 23andMe Product infrastructure.

## Previous positions

**Human Longevity, Inc.**, San Diego, CA / Mountain View, CA

*Software Engineer (September 2014 - November 2017)*

I worked on several projects that span basic scientific research, infrastructure development, and customer-facing product deployment.

- I co-developed a software framework for data analysis pipelines to track the version dependencies between output data, code used to generate the output, and input data.
- I designed and developed the authorization/authentication server to secure and manage client connections to our cancer genome database.
- I studied the relationship between genetics, voice, and demographic data as part of a research study to predict faces and other phenotypes from the genome.

**Illumina, Inc.**, San Diego, CA / San Francisco, CA

*Staff Software Engineer (July 2013 - September 2014)*

I re-engineered a human mutation mapping pipeline SaaS system to support the identification of the genetic variations that drive the onset and growth of cancers.

**IBM Research - Almaden**, San Jose, CA

*Research Staff Member (December 2003 - July 2013)*

I led the development of the IBM Neuro Synaptic Core Simulator (NSCS) as part of the IBM SyNAPSE project. NSCS models a reconfigurable cortical hardware circuit capable of capturing the various cognitive abilities of the brain, and is intended to evaluate the expected behavior of cognitive applications, such as image processing, when deployed on hardware implementations. Evaluations performed with NSCS demonstrated the potential and power of neuronal algorithms in advance of hardware implementations, thus enabling efficient research and development within this new problem-solving domain.

Prior to NSCS development, I was the technical leader for the IBM Virtual Mission Bus (VMB) project. The VMB was a middleware system for supporting distributed, adaptive, hard real-time applications for a dynamic cluster of satellites, under the aegis of the DARPA System F6 program. I led the engineering team that designed and implemented the VMB, and that produced a successful technology demonstration of the VMB.

My projects at IBM include:

- The IBM SyNAPSE Cognitive Computing project for DARPA SyNAPSE (*October 2010 - July 2013*)
- The Virtual Mission Bus middleware system, as part of the Pleiades architecture for DARPA System F6 (*February 2008 - October 2010*)
- Kybos: Self-management for distributed brick-based storage (*December 2003 - February 2008*)

### **Baskin School of Engineering, UC Santa Cruz, Santa Cruz, CA**

*Research Associate (April 2006 - April 2010)*

I collaborated with faculty and students of the Institute for Scalable Scientific Data Management on research in storage systems for high-end computing, with a focus on resource management and end-to-end I/O performance guarantees.

### **Storage Systems Program, Hewlett-Packard Laboratories, Palo Alto, CA**

*Summer Research Intern (June 1999 - August 1999)*

I conducted research into cooperative caching methods for high-end storage systems, and demonstrated that a simple algorithm could yield useful (sometimes substantial) speedups. See *My cache or yours? Making storage more exclusive* under "Refereed publications" for details.

### **Information Technology Section, Mann Library, Cornell University, Ithaca, NY**

*Senior Programmer/Analyst (January 1997 - August 1997)*

I designed and implemented a storage and retrieval system for delivering digitally scanned monographs over the Internet, as part of the Core Historical Literature of Agriculture project. I took responsibility for system development from its design through to its deployment.

### **Isis Distributed Systems, Ithaca, NY**

*Software Engineer, Message Distribution Service (MDS) (May 1996 - January 1997)*

*Software Engineer, Isis Software Developer's Kit (SDK) (August 1995 - May 1996)*

I collaborated in the development of a new MDS release that increased performance. I also provided maintenance engineering and quality assurance testing for the SDK. My experience included porting MDS to Windows NT, adding performance and stability enhancements to the code base, and writing extensive new MDS user documentation.

### **Cornell Information Technologies, Cornell University, Ithaca, NY**

*Technical Consultant (October 1993 - April 1995)*

I provided front-line computer support for Windows- and MSDOS-based platforms.

### **JP Morgan & Co., Incorporated, New York, NY**

*Summer Intern, Global Technology and Operations (June 1994 - August 1994)*

I designed and developed software for computing credit exposure on interest-rate derivatives.

### **Refereed publications**

- Christoph Lippert *et al.* Identification of individuals by trait prediction using whole-genome sequencing data. In *Proceedings of the National Academy of Sciences (PNAS)*, September 2017 (early edition)
- Arnon Amir *et al.* Cognitive Computing Programming Paradigm: A Corelet Language for Composing Networks of Neurosynaptic Cores. In *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN 2013)*, August 2013

- Andrew S. Cassidy et al. Cognitive Computing Building Block: A Versatile and Efficient Digital Neuron Model for Neurosynaptic Cores. In *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN 2013)*, August 2013
- Steve K. Esser et al. Cognitive Computing Systems: Algorithms and Applications for Networks of Neurosynaptic Cores. In *Proceedings of the 2013 International Joint Conference on Neural Networks (IJCNN 2013)*, August 2013
- Emmett McQuinn, Pallab Datta, Myron D. Flickner, William P. Risk, Dharmendra S. Modha, Theodore M. Wong, Raghavendra Singh, Steven K. Esser, and Rathinakumar Appuswamy. Connectivity of a Cognitive Computer Based on the Macaque Brain. *Science*, 339(6119), pp. 512–513, 1 February 2013
- Robert Preissl, Theodore M. Wong, Pallab Data, Raghavendra Singh, Steven Esser, William Risk, Horst Simon, Myron Flickner, and Dharmendra Modha. Compass: A Scalable Simulator for an Architecture for Cognitive Computing. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis (SC 2012)*, November 2012
- David M. LoBosco, Glen E. Cameron, Richard A. Golding, and Theodore M. Wong. The Pleiades fractionated space system architecture and the future of national security space. In *Proceedings of the AIAA SPACE 2008 Conference*, September 2008
- Tim Kaldewey, Theodore M. Wong, Richard Golding, Anna Povzner, Scott Brandt, and Carlos Maltzahn. Virtualizing disk performance. In *Proceedings of the 14th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS 2008)*, April 2008 (Best student paper)
- Anna Povzner, Tim Kaldewey, Scott Brandt, Richard Golding, Theodore M. Wong, and Carlos Maltzahn. Efficient guaranteed disk request scheduling with Fahrrad. In *Proceedings of the ACM SIGOPS/EuroSys European Conference on Computer Systems 2008 (EuroSys 2008)*, April 2008
- Richard A. Golding and Theodore M. Wong. Walking toward moving goalposts: agile management for evolving systems. In *Proceedings of the First Workshop on Hot Topics in Autonomic Computing*, June 2006
- Theodore M. Wong, Richard A. Golding, Caixue Lin, and Ralph A. Becker-Szendy. Zygaria: Storage performance as a managed resource. In *Proceedings of the 12th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS 2006)*, April 2006
- Winfried W. Wilcke et al. IBM Intelligent Bricks project—Petabytes and beyond. *IBM Journal of Research and Development*, 50(2/3), pp. 181–198, March–May 2006
- Theodore M. Wong, Chenxi Wang, and Jeannette M. Wing. Verifiable secret redistribution for archive systems. In *Proceedings of the First International IEEE Security in Storage Workshop*, December 2002
- Theodore M. Wong and John Wilkes. My cache or yours? Making storage more exclusive. In *Proceedings of the USENIX Annual Technical Conference*, June 2002, pp. 161-175

## **Notable technical reports**

- James R. Ashenurst, Sarah B. Laskey, Jianan Zhan Aditya Ambati, Rafaela Bagur Quetglas, Anna Guan, Jingran Wen, Peter Wilton, Iris Yang, Bo Yoo, Sahil Hansalia, Yashashree Kokje, Ishana Raghuram, Akele Reed, Marcos Torres, 23andMe Research Team, Subarna Sinha, Theodore M. Wong, Bertram L. Koelsch. A Generalized Method for the Creation and Evaluation of Polygenic Scores: Accounting for Genetic Ancestry as a Continuum. 23andMe White Paper 23-25, December 2023
- Theodore M. Wong, Robert Preissl, Pallab Datta, Myron Flickner, Raghavendra Singh, Steven K. Esser, Emmett McQuinn, Rathinakumar Appuswamy, William P. Risk, Horst D. Simon, Dharmendra S. Modha. 10<sup>14</sup>. IBM Technical Paper RJ10502, November 2012
- Theodore M. Wong, Richard A. Golding, Harvey M. Ruback, Wilfred Plouffe, and Scott A. Brandt. The Virtual Mission Bus. IBM Technical Paper RJ10472, September 2010
- Theodore M. Wong, Richard A. Golding, Joseph S. Glider, Elizabeth Borowsky, Ralph A. Becker-Szendy, Claudio Fleiner, Deepak R. Kenchammana-Hosekote, and Omer A. Zaki. Kybos: Self-management for

distributed brick-based storage. IBM Technical Paper RJ10356, August 2005

- Theodore M. Wong. *Decentralized recovery for survivable storage systems*. PhD dissertation (Technical Report CMU-CS-04-119), School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, May 2004
- Theodore M. Wong and Jeannette M. Wing. Verifiable secret redistribution. Technical Report CMU-CS-01-155, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, October 2001

## Patents

- Nelly Fazio, Richard A. Golding, Theodore M. Wong. Method for authenticated communication in dynamic federated environments. US Patent 9,130,757 (8 September 2015)
- Steven K. Esser, Dharmendra S. Modha, Theodore M. Wong. Cortical simulator for object-oriented simulation of a neural network. US Patent 9,130,757 (28 April 2015)
- Scott A. Brandt, Richard A. Golding, Theodore M. Wong. Integration of dissimilar job types into an earliest deadline first (EDF) schedule. US Patent 8,607,240 (10 December 2013)
- Ralph A. Becker-Szendy, Richard A. Golding, Caixue Lin, Theodore M. Wong, and Omer A. Zaki. System and method for managing storage system performance as a resource. US Patent 7,962,563 (14 June 2011)
- Richard A. Golding, Theodore M. Wong, and Omer A. Zaki. Computer program and method for managing resources in a distributed storage system. US Patent 7,694,082 (6 April 2010)
- John Wilkes and Theodore M. Wong. Exclusive caching in computer systems. US Patent 6,851,024 (1 February 2005)
- John Wilkes and Theodore M. Wong. Adaptive data insertion for caching. US Patent 6,728,837 (27 April 2004)

## Professional activities

- 15th USENIX Conf. on File & Storage Technologies (FAST 2017): Program committee
- 14th USENIX Conf. on File & Storage Technologies (FAST 2016): Program committee
- 29th IEEE Conf. on Massive Data Storage (MSST 2013): Program co-chair
- 10th USENIX Conf. on File & Storage Technologies (FAST 2012): Program committee
- 2nd Wksp. on Hot Topics in Storage & File Systems (HotStorage 2010): Program committee
- 6th USENIX Conf. on File & Storage Technologies (FAST 2008): Program committee
- 24th IEEE Conf. on Mass Storage Systems & Technologies (MSST 2007): Program committee
- 13th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS 2007): Track A committee

## Educational background

**Carnegie Mellon University** (August 1997 - May 2004)

Doctor of Philosophy in Computer Science

**Cornell University** (January 1994 - August 1995)

Master of Engineering in Computer Science

**Oxford University** (October 1989 - June 1993)

Bachelor of Arts in Engineering Science