

KAMIL M ROCKI



ARTIFICIAL INTELLIGENCE
SUPERCOMPUTING
COMPUTER ARCHITECTURE

EXPERIENCE

ZYPHRA TECHNOLOGIES

VP - ENGINEERING
November 2023 -

STABILITY AI

HEAD OF PERFORMANCE
January 2023 - November 2023

NVIDIA

DENSE LINEAR ALGEBRA TEAM LEAD
May 2022 - January 2023
SENIOR MATH LIBRARIES ENGINEER
November 2021 - May 2022

NEURALINK CORP

Brain-Machine Interfaces
SOFTWARE ENGINEER, BRAIN-MACHINE
INTERFACES
October 2020 - November 2021

CEREBRAS SYSTEMS

Advanced Technology Group
MEMBER OF TECHNICAL STAFF
Mar 2019 - October 2020

IBM RESEARCH

Machine Intelligence Group
RESEARCH SCIENTIST
Jul 2014 - Feb 2019
Record promotion time at IBM
POSTDOCTORAL RESEARCHER
Oct 2013 - Jul 2014
1st employee of the MI Group

UNIVERSITY OF TOKYO

Department of Computer Science
POSTDOCTORAL RESEARCHER
Sep 2011 - Sep 2013

UNIVERSITY OF TOKYO

PHD IN COMPUTER SCIENCE
September 2011
UT Fellowship Award

LOCATION : LOS GATOS, CA

Website: <https://www.krocki.ai>

Email: kamil.rocki@gmail.com

Twitter: [@rockikamil](https://twitter.com/rockikamil)

Github: [krocki](https://github.com/krocki)

LinkedIn: [krocki](https://www.linkedin.com/in/krocki)

SELECTED PROJECTS

MINDPONG | [LINK](#) | 2021

- Development of low-level code for embedded, low-power sensing device
- Lowered the latency by on-chip real-time adaptive data compression
- Bare-metal training/inference inside on-device

WAFERSCALE FLUID DYNAMICS | [LINK](#) | 2019-2020

- Conjugate-gradient solver achieving 0.86 PF/s within 15kW power budget
- First-ever public evaluation of a waferscale architecture, published at SC20

NINTENDO LEARNING ENVIRONMENT | [LINK](#) | 2018-

- Faster than OpenAI Gym by a factor of 10, CPU + CUDA versions.

GAMEBOY SUPERCOMPUTER | [GITHUB](#) | [VICE ARTICLE](#) | [MEDIUM POST](#) | 2018-

- A \$2M FPGA-based supercomputer, 1296 nodes connected in a 3D mesh
- Hardware implementation of Nintendo Gameboy for Reinforcement Learning.
- 1B+ frames per s aggregate, over 200x speedup vs Xeon CPU

ADAPTIVE INSTRUCTION SET COMPUTER [AISC] | 2017-

- RISC-V CPU augmented with an ability to shape its own architecture given a task

CONTENT ADDRESSABLE MEMORY | [GITHUB A](#), [GITHUB B](#) | 2016-2018

- Memory Augmented NN with approximate key-value map
- Allows rapid constant time lookups in a large distributed system

LOW-LEVEL PARALLEL PRIMITIVES [L2P2] | [GITHUB](#) | 2016

- Provide portable low-level templates for parallel programming and enable auto-tuning of compute kernels using ML

RECURRENT MEMORY ARRAY | [GITHUB](#) | 2015-2016

- Extended SOTA on wikipedia character-level prediction task (twice)
- Fastest LSTM implementation at the time (sustained 1TF/s on a Titan GPU)
- Improvements to existing models (Array LSTM, Surprisal-Feedback)

DARPA SACCADIC VISION PROJECT | [GITHUB](#) | 2015-2016

- Image recognition using sequences of 2D patches as inputs
- Unsupervised training using RBM and LSTM builds predictive model
- Supervised task-based learning using reinforcement learning

HIERARCHICAL TEMPORAL MEMORY | [VIDEO](#) | 2013-2014

- Joint IBM-Numenta project
- Analyzed parallel implementation of the HTM algorithm and hardware architectures
- Classifying EMG sensor data for sub-vocal communication

STOCHASTIC LOCAL OPTIMIZATION | [GITHUB](#) | [HPC WIRE ARTICLE](#) | 2012-2013

- Fastest GPU-based Travelling Salesman Solver (sustained 2TF/s)
- Algorithmic innovations applicable to wide range of irregular problems

PARALLEL MONTE CARLO TREE SEARCH | [GITHUB](#) | [PHD THESIS](#) | 2008-2011

- Full implementation of distributed MCTS on TSUBAME2 supercomputer (2048 CPUs + 256 GPUs, 3.5M threads). 10000x faster than the baseline.

DATA PROCESSING USING GPU | MSc THESIS | [GALLERY](#) | 2006-2008

- Used OpenGL graphics pipeline for general-purpose computation (pre-CUDA). Reduced full-frame (1024x768x24b) image denoising time from 1s to 10ms.

ROBOTIC ARM MANIPULATION USING NEURAL NET | [GALLERY](#) | 2005-2007

- Used ARTMAP fuzzy neural net to guide a robotic arm (pixels -> torque)

PUBLICATIONS, PAGE 1

2020

FAST STENCIL-CODE COMPUTATION ON A WAFER-SCALE PROCESSOR Kamil Rocki, Dirk Van Essendelft, Ilya Sharapov, Robert Schreiber, Michael Morrison, Vladimir Kibardin, Andrey Portnoy, Jean Francois Dietiker, Madhava Syamlal, Michael James

<https://arxiv.org/abs/2010.03660> + SC 20: International Conference for High Performance Computing, Networking, Storage and Analysis

2019

INFERENCE OF LONG-SHORT TERM MEMORY NETWORKS AT SOFTWARE-EQUIVALENT ACCURACY USING 2.5 M ANALOG PHASE CHANGE MEMORY DEVICES H Tsai, S Ambrogio, C Mackin, P Narayanan, RM Shelby, K Rocki, A Chen, GW Burr

<http://olab.is.s.u-tokyo.ac.jp/~kamil.rocki/vlsi19.pdf> + 2019 Symposium on VLSI Technology

2018

IBM NEURAL COMPUTER KM Rocki, A Asseman, C Cox, P Narayanan, AS Ozcan, WW Wilcke

<https://towardsdatascience.com/a-gameboy-supercomputer-33a6955a79a4>

LEARNING TO REMEMBER, FORGET AND IGNORE USING ATTENTION CONTROL IN MEMORY TS Jayram, Y Bouhadjar, RL McAvoy, T Kornuta, A Asseman, KM Rocki, AS Ozcan

<https://arxiv.org/pdf/1809.11087.pdf>

2016

UTILIZATION OF DEEP REINFORCEMENT LEARNING FOR SACCADIC-BASED OBJECT VISUAL SEARCH T Kornuta, KM Rocki

<https://arxiv.org/pdf/1610.06492.pdf>

SURPRISAL-DRIVEN ZONEOUT KM Rocki, T Kornuta, T Maharaj

<https://arxiv.org/pdf/1610.07675.pdf> + NIPS 2016 workshop proceedings

SURPRISAL-DRIVEN FEEDBACK IN RECURRENT NETWORKS KM Rocki

<https://arxiv.org/pdf/1608.06027.pdf>

RECURRENT MEMORY ARRAY STRUCTURES KM Rocki

<https://arxiv.org/pdf/1607.03085.pdf>

TOWARDS MACHINE INTELLIGENCE KM Rocki

<https://arxiv.org/pdf/1603.08262.pdf>

2015

A PROBABILISTIC VIEW OF THE SPATIAL POOLER IN HIERARCHICAL TEMPORAL MEMORY M Leake, L Xia, KM Rocki, W Imaino

International Journal of Computer, Electrical, Automation, Control and Information Engineering

EFFECT OF SPATIAL POOLER INITIALIZATION ON COLUMN ACTIVITY IN HIERARCHICAL TEMPORAL MEMORY M Leake, L Xia, KM Rocki, W Imaino

29th AAAI Conference Proceedings

2014

THE FUTURE OF ACCELERATOR PROGRAMMING: ABSTRACTION, PERFORMANCE OR CAN WE HAVE BOTH? KM Rocki, M Burtcher, R Suda

29th ACM Symposium On Applied Computing (SAC)

<https://www.hpcwire.com/2014/01/09/future-accelerator-programming/>

PUBLICATIONS, PAGE 2

2013

REGISTER LEVEL SORT ALGORITHM ON MULTI-CORE SIMD PROCESSORS T Xiaochen, KM Rocki, R Suda
2013 Supercomputing Conference Proceedings

巡回セールスマン問題に対する反復局所探索の大規模並列アルゴリズム KM Rocki, R Suda
TSUBAME Supercomputer e-Science Journal Vol.10
https://www.gsic.titech.ac.jp/sites/default/files/TSUBAME_ESJ_10jp_0.pdf

OPENCL-BASED APPROACH TO HETEROGENEOUS PARALLEL TSP OPTIMIZATION KM Rocki, R Suda
2013 International Workshop for OpenCL

HIGH PERFORMANCE GPU ACCELERATED LOCAL OPTIMIZATION IN TSP KM Rocki, R Suda
2013 IEEE International Symposium on Parallel and Distributed Processing (IPDPS) Proceedings

2012

AN EFFICIENT GPU IMPLEMENTATION OF A MULTI-START TSP SOLVER FOR LARGE PROBLEM INSTANCES KM Rocki, R Suda
GECCO '12 Proceedings

HIGH PERFORMANCE GPU ACCELERATED TSP SOLVER KM Rocki, R Suda
2012 Supercomputing Conference Proceedings

ACCELERATING 2-OPT AND 3-OPT LOCAL SEARCH USING GPU IN THE TRAVELLING SALESMAN PROBLEM KM Rocki, R Suda
2012 High Performance Computing and Simulation Conference (HPCS) Proceedings
http://olab.is.s.u-tokyo.ac.jp/~kamil.rocki/rocki_hpcs2012.pdf

2011

GPU による大規模モンテカルロ木探索 / LARGE SCALE MONTE CARLO TREE SEARCH ON GPU KM Rocki
PhD Thesis, University of Tokyo
http://olab.is.s.u-tokyo.ac.jp/~kamil.rocki/phd_thesis.pdf

PARALLEL MONTE CARLO TREE SEARCH SCALABILITY KM Rocki, R Suda
AI 2011: Advances in Artificial Intelligence: 24th Australasian Joint Conference

LARGE-SCALE PARALLEL MONTE CARLO TREE SEARCH ON GPU
KM Rocki, R Suda
2011 IEEE International Symposium on Parallel and Distributed Processing (IPDPS) Proceedings

2010

MASSIVELY PARALLEL MONTE CARLO TREE SEARCH
KM Rocki, R Suda
2010 VECPAR Conference

2009

PARALLEL MINIMAX TREE SEARCH ON GPU
KM Rocki, R Suda
2009 International Conference on Parallel Processing and Applied Mathematics (PPAM)
