

## Profile

I am an ML compiler compiler engineer at NVIDIA. During my time at Google, I was working on MLIR and the IREE compiler. I have been working with and contributing to the open-source MLIR compiler project since 2021. My research at university was concerned with code generation, domain-specific languages (DSLs), programming models and memory access optimizations for SIMD/GPU architectures. I am passionate about compilers, high-performance computing (especially GPUs), programming language design and parallel computing.

## Work Experience

- Since May 2024 **NVIDIA Switzerland AG, Zürich, Switzerland**, Deep Learning Compiler Engineer.  
Working on an MLIR-based compiler infrastructure for deep learning, targeting NVIDIA Blackwell GPUs.
- Mar. 2021 **Google Switzerland GmbH, Zürich, Switzerland**, Research Engineer, Google Research.
- Apr. 2024 Core  MLIR (Multi-Level IR Compiler Framework) contributor with >1000 commits. See personal website for RFCs. Contributions to the  IREE (Intermediate Representation Execution Environment) Compiler.
- Contributions to core APIs: builders/rewriters, listener infrastructure, greedy pattern rewriter, dialect conversion, operation definition specification (ODS), etc.
  - One-Shot Bufferize: Design, implementation and maintenance of an MLIR pass for buffer assignment/management in tensor IR. Supported its adoption in IREE and XLA/TensorFlow (see MLIR Open Design Meeting presentation). Improvements and maintenance of the ownership-based buffer deallocation pass.
  - Contributions to the Transform dialect, a scheduling language for compiler transformations: driving rewrite pattern application from transform scripts, core transform ops, integration of existing MLIR transformations, etc.
  - Loop Peeling: Implemented a for loop peeling and loop body simplification transformation in the SCF dialect.
  - Value Bounds Inference: Designed and implemented an infrastructure for computing bounds of integer SSA values.
  - Various contributions to the tiling, fusion and vectorization infrastructure in the Linalg and Tensor dialects.
  - Various debugging and usability improvements such as runtime op verification, rewrite API verification checks, Graphviz export of IR graphs, pattern application fuzzing, transform dialect interpreter hardening.
  - Various improvements and bug fixes in the core MLIR infrastructure (interfaces, vector warp distribution on GPUs, MemRef/SCF/SparseTensor/Tensor/Transform/Vector dialect improvements, etc.)
- Oct. 2019 **Google Japan G.K. 【グーグル合同会社】**, Tokyo, Japan, Software Engineer, ChromeOS Platform.
- Feb. 2021 Worked on ARCVM (Android Runtime for Chrome), a virtual machine that runs Android. Identified and fixed bugs in the ChromeOS/Android camera stacks and worked on overall performance optimizations. Gained experience in virtualization, kernel/OS development and large-scale software development (mainly C++, some Java and Golang).
- 2017, 2018 **Google LLC, United States of America**, Software Engineering Intern (Summer Internship).
- 2016, 2014 (4 × 3 months)  [2018; Mountain View, CA; Host: DeLesley Hutchins, Ph.D.] Worked on an auto-batching system for tree-structured RNNs. Implemented linear algebra operations and TPU-specific optimizations with TensorFlow XLA.
- [2017; Mountain View, CA; Host: DeLesley Hutchins, Ph.D.] Worked on  LLGTM (Low-level Library for Gradients, Tensors, and Matrices), a deep learning C++ framework for dynamic computation graphs. Implemented Eigen/CUDA kernels and TensorFlow kernel adapters.
  - [2016; Seattle, WA; Host: Vijay Menon, Ph.D.] Part of the Dart programming language team. Worked on a  Dart-to-Java compiler, focusing on performance optimizations and language interoperability.
  - [2014; Boulder, CO; Host: Craig Wright] Worked on a business event process engine for an internal Google payments system, using Megastore, F1/Spanner, Java and Guice.
- April 2012 **Tokyo Institute of Technology / Hasso Plattner Institute**, Teaching Assistant.
- June 2017 Teaching assistant for exchange students and courses: Information Literacy (Prof. Morozov, Ph.D.), Mathematics II (Dr. habil. Börner), Software Architecture (Prof. Dr. Hirschfeld), Software Engineering I (Prof. Dr. Hirschfeld)
- Aug. 2012 **Senacor Technologies AG, Munich, Germany**, Software Engineering Intern.
- Oct. 2012 Developed SOA components and tests (with Java EE, Spring Framework), supporting a merger of two banks.
- Aug. 2011 **TNG Technology Consulting GmbH, Munich, Germany**, Software Engineering Intern.
- Oct. 2011 Developed plugins for the team's continuous integration software stack (Atlassian JIRA/Confluence, Hudson/Jenkins).

---








## Education

- Oct. 2015 **Tokyo Institute of Technology** 【東京工業大学】 , *Meguro-ku, Tokyo, Japan*,  
- Sept. 2019 Programming Research Group, Department of Mathematical and Computing Sciences  
Doctor of Philosophy (Ph.D.). Academic advisor: Prof. Dr. Hidehiko Masuhara.  
○ *Doctoral thesis*: Memory-Efficient Object-Oriented Programming on GPUs  
○ *Research areas*: Compilers, Program Optimization, GPGPU, Modularity, Context-oriented Programming  
○ *Relevant coursework*: Programming Language Design, Practical Parallel Computing, Distributed Computing
- Sept. 2014 **Hasso Plattner Institute, University of Potsdam**, *Potsdam, Brandenburg, Germany*,  
- Sept. 2015 Master of Science, IT Systems Engineering, overall grade: 1.0 (A+).  
○ *Master's project*: Spur to go faster: Low-level Functionality in a High-level Language  
○ *Master's thesis*: Nested Class Modularity in Squeak/Smalltalk  
Project and thesis supervised by Prof. Dr. Robert Hirschfeld, Tim Felgentreff, Tobias Pape  
○ *Relevant coursework*: VMs and Execution Environments, Context-oriented Programming
- Sept. 2013 **University of California San Diego**, *La Jolla, CA, United States of America*,  
- June 2014 Visiting student, Department of Computer Science and Engineering, GPA: 4.0.  
○ Full tuition and living expenses covered by *UC Education Abroad Program* and *DAAD Scholarship*  
○ *Relevant coursework*: Advanced Compilers (CSE 131/231), Programming Languages (CSE 130/230),  
Adv. Algorithms (CSE 190/202/203A), Parallel Computation (CSE 260), Database Analytics (CSE 190)
- Aug. 2010 **Hasso Plattner Institute, University of Potsdam**, *Potsdam, Brandenburg, Germany*,  
- July 2013 Bachelor of Science, IT Systems Engineering, overall grade: 1.0 (A+), rank 1/74.  
○ *Bachelor's project*: Evolving Applications: Object-migration with Ruby and GemStone  
○ *Bachelor's thesis*: Inter-language Collaboration in an Object-oriented Virtual Machine  
Project and thesis supervised by Prof. Dr. Robert Hirschfeld, Tim Felgentreff, Tobias Pape  
○ *Relevant coursework*: Software Architecture, Software Engineering I, Advanced Modularity, Database Systems I/II,  
Internet and WWW Technologies, Designing Interactive Systems (HCI)

Publications on Google Scholar: <https://scholar.google.com/citations?user=EvHvYtMAAAAJ>

---

## Research Experience

- Oct. 2014 **Programming Languages/Compilers**, *Tokyo Institute of Technology / Hasso Plattner Institute*,  
- Sept. 2019 Research with Prof. Hidehiko Masuhara and Prof. Robert Hirschfeld.  
○  **DYNASOAR**: A lock-free CUDA memory allocator with SOA layout, based on hierarchical bitmaps.  
○  **COMPACTGPU**: Fully parallel GPU memory defragmentation for better vector/cache performance.  
○  **IKRA-CPP**: A C++/CUDA DSL for object-oriented programming with Structure-of-Arrays data layout.  
○  **IKRA-RUBY**: A GPGPU library for Ruby (Ruby-to-CUDA compiler) with parallel array operations.  
○  **MATRIONA**: A module system for Squeak/Smalltalk based on class nesting/parameterization.  
○  **CONTEXTAMBER**: A COP (Context-oriented Programming) library for Amber Smalltalk.  
○ *Other Projects*: Minor contributions to the  **RSQUEAK** VM and to the Truffle-based JRUBY implementation.
- March 2014 **Database Research**, *University of California San Diego (with Prof. Yannis Papakonstantinou)*.  
- Nov. 2014 Evaluated algorithms and data structures for queries on compressed data in relational/graph database systems.
- May 2011 **Internet Technologies**, *Hasso Plattner Institute, Internet Technologies and Systems Group*.  
- Apr. 2013 ○ **SOA SECURITY LAB**: A browser-based simulation system for modelling and executing web service scenarios.  
○ **TELE-LAB**: A platform for teaching and simulating network security scenarios.

---

## Achievements and Prizes

- Feb. 2021 **Seiichi Tejima Doctoral Dissertation Award** 【手島精一記念研究賞（博士論文賞）】 .  
Honoring authors of outstanding dissertations at Tokyo Institute of Technology. Category: *Information Science*.
- July 2019, **ACM Student Research Competition**, *Association for Computing Machinery*.
- Nov. 2018, ○ [PLDI 2019 (Phoenix, AZ)] 1<sup>st</sup> place, graduate category for **COMPACTGPU**  
Feb. 2018 ○ [SPLASH 2018 (Boston, MA)] 1<sup>st</sup> place, graduate category for **SOAALLOC** (now called **DYNASOAR**)  
○ [CGO 2018 (Vienna, Austria)] 3<sup>rd</sup> place, graduate category for **IKRA-CPP**
- April 2018 **Research Fellowship for Young Scientists (JSPS DC2)** 【日本学術振興会特別研究員DC2】 .  
- Sept. 2019 Japanese government fellowship, covering living expenses and research expenses (18J14726).

- Oct. 2015 **Monbukagakusho (MEXT) Scholarship** 【文部科学省奨学金】 .
- Mar. 2018 Japanese government scholarship for doctoral students, covering tuition and living expenses.
- Oct. 2014 **Hasso Plattner Scholarship**.
- Sept. 2015 One-year scholarship awarded to the best Bachelor graduates of each year, covering living expenses.
- Sept. 2013 **German Academic Exchange Service Scholarship (DAAD Jahresstipendium)**.
- May 2014 German government scholarship, covering tuition and living expenses at a North American university.
- 2007 **German Federal Competition in Computer Science (Bundeswettbewerb Informatik)**.
- 2010 Participated three years in a row. 30/around 1100 participants are invited to the final round.
  - o [2009/2010] 1<sup>st</sup> prize in first two rounds, invited to final round (*University of Freiburg*)
  - o [2008/2009] 2<sup>nd</sup> prize in first two rounds
  - o [2007/2008] 1<sup>st</sup> prize in first two rounds, invited to final round (*Max Planck Institute for CS*)
- 2010/2011 **informatiCup 2011**, organized by the *Gesellschaft für Informatik, Bonn, Germany*.  
Participated in the first round and in the final round (6/38 teams invited). Solved an optimization problem using Simulated Annealing, Tabu Search and greedy algorithms.
- (various) **Academic Honors**, Graduation with distinction (Bachelor's, Master's), 2x Provost's Honors at UCSD.
- (various) **Travel Grants**, ECOOP Summer School 2016, SIGPLAN-PLMW (POPL 2017), SIGPLAN-PAC (PLDI 2017), Google MUC Compiler and PL Summit (2017, 2018), ACM SRC (CGO 2018, SPLASH 2018, PLDI 2019).

---

## Academic Service

**Program Committee**, CROW 2016, COP 2017, COP 2018.

**Reviewer**, APLAS 2016, ARRAY 2017, ECOOP 2020, LASSY 2017, MPLR 2019, SIMPAT.

**Student Volunteer**, ECOOP (2015, 2016, 2017, 2019), PLDI (2016, 2017, 2018), SPLASH 2018.

**Organizing Committee**, Student volunteer co-chair at PLDI 2019.

---

## Other Projects

- Sept. 2014 **ME310 Global Team-based Product Innovation & Engineering**,
- July 2015 *Collaboration between Hasso Plattner Institute and Stanford University, including several exchange visits.*  
Worked on a design challenge by Audi USA with Stanford mechanical engineering students. Prototyped concepts for communication between pedestrians, passengers and autonomous cars, using design thinking.
- June 2013 **⦿ Athens for Amber Smalltalk**, *Google Summer of Code 2013 Project*,
- Sept. 2013 *European Smalltalk User Group (ESUG), Mentors: Nicolas Petton, Igor Stasenko.*  
Implemented the ATHENS vector graphics library in Amber Smalltalk, a Smalltalk system running in the web browser, using HTML5 Canvas. Developed a Morpich-like framework for building GUIs with Athens.
- Dec. 2012 **⦿ MagLev Database Explorer**, *Part of Bachelor's project at HPI, Software Architecture Group.*
- June 2013 Developed an IDE running in a web browser for exploring Ruby/Smalltalk objects persisted in a GemStone/S 64 image, writing Ruby/Smalltalk code and debugging Rails/Sinatra applications interactively.

---

## Skills

- Programming C, C++, CUDA, Dart, Go, HTML, Java, LLVM, MLIR, NumPy, Python, Ruby, Smalltalk (Amber, GemStone, Pharo, Squeak), SQL, Visual Basic.
- Software Eng. Continuous Integration, Design Patterns, Git, Mercurial, Scrum, Subversion, TDD, UML.

---

## Personal Information

- Languages German (native speaker), English (CEFR C2, TOEFL iBT score 118/120), Japanese (JLPT N3 147/180).

Date: August 31, 2024