Chi Zhang

Resume

Personal Inforamtion

Name Chi Zhang

Phone 757-338-7196

Email chz54@pitt.edu

ShortBio Active researcher and software engineer. Experienced and interested in the areas of **computer architecture, compiler, deep learning and AI.**

Education

2014–2020 **Ph.D. in Computer Science**, *University of Pittsburgh*.
Arts and Science Fellowship

2010–2014 **B.E. in Computer Science**, *Xidian University*, China.
Graduate with Honor(3%)

Work Experience

2019 **Facebook**, *Software Engineer PhD Intern*, Menlo Park, CA.
PyTorch Glow Runtime team
(Glow: A machine learning compiler and execution engine for hardware accelerators. The compiler is designed to allow state of the art compiler optimizations and code generation of neural network graphs.)
Support for debugging in Glow

- Add functionality to track and dump all changes that happened in the graph compilation and optimization phases of Glow.
- Implemented log-based debugging tools to
  - reconstruct the node graph at any certain fixed compilation phase of Glow.
  - filter and infer all nodes transformations related to one given node.
  - collect basic statistics of nodes at any phase or bewtween any pair of compilation phases.

2018 **Facebook**, *Software Engineer PhD Intern*, Menlo Park, CA.
PyTorch Glow Runtime team
Quantization Support for GPU backend of Glow

- Add quantization support for more than 20 GPU operators of Glow.
- Enable the weights and data of the neural network to be stored in quantized format (INT8) other than 32bits (INT32). **Reduced the entire memory usage by 75%.**

2017 **Google**, *Software Engineer PhD Intern*, Mountain View, CA.
Google Payment Infrastructure
Backend support for ads billing and payments data

- Help migrate the BigTable/MapReduce based backend system to a new Google F1 based system.
- Implement the API to fetch/render/process the billing documents of Google customers.

2016 **Bosch Research & Technology Center**, *Research Engineer Intern*, Pittsburgh, PA.
PRIVACY AND SECURITY TEAM
Cloud-based encrypted search engine based on SSE (Searchable Symmetric Encryption)
- Designed and implemented an encrypted search engine infrastructure that is based on SSE.
- Achieve scalability for this infrastructure by utilizing Apacha Lucene/Solr and deployment on AWS.

Significant Projects

Transactional Row-Column Store DBMS.
- Designed and implemented a transactional DBMS that supports aggregated query and concurrent execution of transactions.
- Employed Strict 2-Phase Locking to ensure serializability. Adopted an undo recovery/no-redo strategy to achieve atomicity. Utilized wait-for graph to meet dead lock detection.

Mini-Google: Document Indexing and Querying.
- Effectuated a simple Map-Reduce framework based on RPC to index and search large documents.
- Delivered a replicated and reliable client/server model, consisting of: the client, the server (with indexing and querying masters), the helpers (for mapping and reducing), and the name-server (for name resolution).

Live Code Update Strategies for Energy Harvesting Devices.
- Explore efficient live update strategies for code images of IoT devices in energy harvesting environment.
- Proposed a novel strategy based on in-place code updating and code trampoline. It effectively eliminates system down time and minimizes resource demands for updates.

Exploring potential of Escape Analysis on state-of-art Java Virtual Machine.
- Conducted research on Escape Analysis on HotSpot Java VM.
- Applied Merlin Algorithm by Hertz et al to acquire the most accurate lifetime of objects. Compared this profiling information with the state-of-art escape analysis on HotSpot Java VM.

Awards

2012 **National Lizhi Scholarship**, Ministry of Education, China

2011 **National Scholarship**, Ministry of Education, China

Skills

Languages C++/C, Python, Java, SQL

Tools LaTeX, CUDA, OPENCL, PyTorch, Caffe, Linux/Unix, GDB debugging

CS Knowledge Solid algorithms skills, Computer Systems concepts (e.g. I/O system, compiler, organization)

Publications

2018 **Locality-Aware Software Throttling for Sparse Matrix Operation on GPUs**, Y. Chen, A. Hayes, C. Zhang, T. Salmon, E.Z. Zhang. Proceedings of the USENIX Annual Technical Conference *(USENIX ATC 2018)*, Boston, MA, July 2018.

2016 **Live Code Update for IoT Devices in Energy Harvesting Environments**, C.Zhang, W.Ahn, Y.Zhang, B.Childers. Non-Volatile Memory Systems and Applications Symposium *(NVMSA)*, 2016 5th. IEEE, 2016.

Services

2018 **Artifact Evaluation Committee Member for PPoPP 2019**

2016 **Artifact Evaluation Committee Member for CGO-PPoPP 2017**

(757) 338 7196 • raymond.chizhang@gmail.com
linkedin.com/in/raymondchizhang