CHEN Han

Last update: March 6, 2025

CONTACT INFORMATION

Tel: +86-13305014345
Tel: +65-89420214
Wechat: concyclics
✉ E-mail:chenhan@u.nus.edu
Github:www.github.com/Concyclics
Address: 18-07, Blue Horizon, 23 West Coast Crescent, Singapore 128046

EDUCATION

National University of Singapore, Singapore. 2023–2025
- Master of Computing, Computer Science Specialization. GPA: 4.4/5.0

South China University of Technology (SCUT), Guangzhou, China. 2019–2023
- B.Eng., Software Engineering. GPA: 3.6/4.0

PRIZES AND AWARDS

- **Excellent Degree Dissertation of South China University of Technology** 2023
- **Honorable Mention** in Mathematical Contest in Modeling 2023
- **National Scholarship** 2022
- ***Bronze Medal (46th)*** in ICPC Asia-East Continent Final(Xi An) 2022
- ***101/1608*** in CCF-DBCI Competition of "Small Sample Data Classification" 2022
- ***Silver Medal (46th)*** in ICPC Asia Regional Contest(Ji Nan) 2021
- ***44/3567*** in CCF-DBCI Competition of "Recognition of figure skaters' skeleton points based on Paddle" 2021
- ***First Prize*** in National Olympiad in Informatics in Province(NOIP) 2017

PREPRINTS AND PUBLICATIONS

- ***CHEN Han***, *Tao Huang, Phuong Pham, Shuang Wu.* **HiAE: A High-Throughput Authenticated Encryption Algorithm for Cross-Platform Efficiency** 2025
- ***CHEN Han***, *Zicong Jiang, Zining Zhang, Bingsheng He.* **LogQuant: Log-Distributed 2-Bit Quantization of KV Cache with Superior Accuracy Preservation**. Accepted at ICLR 2025 Workshop on Sparsity in LLMs 2025
- *Wenqi Pei, Hailing Xu, Henry Hengyuan Zhao,* ***CHEN Han****, Zining Zhang, Shizheng Hou, Luo Pingyi, Bingsheng He.* **Optimizing Small Language Models for NL2SQL**. Accepted at ICLR 2025 Third Workshop on Deep Learning for Code 2025

PROJECT EXPERIENCE

- ***Research Assistant: optimization for large language Model inference*** in **National University of Singapore**
  Mentor: Prof. Bingsheng HE May-Sept. 2024
  - Design a new 2-bit KV Cache quantization for LLMs base on attention patterns. achieve over 200% accuracy improvement at same compression rate.
  - Implement a adaptive API for popular inference frameworks like Python's `transformers`, boosts batch size by 60% without increasing memory consumption.
  - Accepted at ICLR 2025 Workshop on Sparsity in LLMs. [Paper] [Code]
- ***Internship: Cryptography Engineer*** in **SG Digital Trust Lab, Singapore Research Center, 2012 Laboratory**
  Mentor: Dr. Tao HUANG Jan.-Dec. 2024
  - Design a *'XAXX'* structure to efficiently utilize the distinct pipelines of both ARM and x86 (with AES-NI) architectures, achieving high IPC.

- Build a new AEAD (Authenticated Encryption with Associated Data) cipher named *'HiAE'* based on the *'XAXX'* structure, which is 5× faster than AES-256-GCM across different platforms and outperforms all existing AEAD ciphers on latest ARM and x86 processors.
- Create a new record of 340Gbps throughput for single-threaded and single-stream AEAD encryption. Applied to the various products.
- Open-source on Cryptology ePrint Archive [Paper] [Code]

- ***Symmetric Matrix Solving Algorithm Parallel Optimization for ARM Architecture***
  Mentor: Prof. Deyou TANG May-Dec. 2022
  - Optimize and parallel Bounded Bunch-Kaufman Algorithm(*sysv_rk subroutine of LAPACK) for solving symmetric matrix on ARM server processor with NEON instruction set and openMP.
  - Implement a parallel column reordering method in row swap of solving symmetric matrix to enhance memory access locality for column major matrix for better cache hit rate and parallelism, achieving a performance improvement from 320Gflops to 580Gflops.
  - Implement the same optimization on Skylake Intel processor and achieve 2-5x multi-core speedup than MKL library for *sytrs_3 subroutine of LAPACK.
  - Awarded as the Excellent Degree Dissertation of South China University of Technology.

Technical Skills

- *English*: IELTS(6.5), CET-4, CET-6.
- *Programming Languages*: C/C++, Fortran, p4-16, Python, SQL, LaTeX.
- *Technical Skills*: openMP, SIMDs(NEON, AVX512), MPI, PyTorch, CUDA.
- *TestDemo Certificate*: C++, TOP 10%, LINUX, TOP 10%, PYTHON, TOP 10%.
- *Kaggle Certificate*: Data Visualization, Intro to Machine Learning, Intro to Deep Learning, Intro to Game AI and Reinforcement Learning.

Exchange Experience

- **Online Academic Program on Machine Learning, McGill University** Jan.-Feb. 2022

陈涵

最近更新: **2025 年 3 月 6 日**

联系方式

Tel: +86-13305014345
Tel: +65-89420214
微信: concyclics
✉ E-mail:chenhan@u.nus.edu
Github:www.github.com/Concyclics
Address: 18-07, Blue Horizon, 23 West Coast Crescent, Singapore 128046

教育经历

新加坡国立大学, 新加坡 2023–2025
- 计算机科学硕士, 计算机科学方向. GPA: 4.4/5.0

华南理工大学, 广东省广州市 2019–2023
- 工学学士, 软件工程专业. GPA: 3.6/4.0

获奖荣誉

- **华南理工大学本科优秀毕业设计(论文)** 2023
- **二等奖** 美国大学生数学建模竞赛 (MCM/ICM) 2023
- **铜牌** 第46届ICPC国际大学生程序设计竞赛亚洲区决赛 2022
- ***101/1608*** CCF-DBCI "小样本数据分类算法" 竞赛 2022
- **国家奖学金** 2022
- **银牌** 第46届ICPC国际大学生程序设计竞赛(济南站) 2021
- ***44/3567*** CCF-DBCI "基于飞浆实现花样滑冰选手骨骼点识别" 竞赛 2021
- **一等奖** 全国青少年信息学奥林匹克联赛(NOIP) 2017

预印本和发表论文

- ***CHEN Han***, *Tao Huang, Phuong Pham, Shuang Wu.* **HiAE: A High-Throughput Authenticated Encryption Algorithm for Cross-Platform Efficiency** 2025
- ***CHEN Han***, *Zicong Jiang, Zining Zhang, Bingsheng He.* **LogQuant: Log-Distributed 2-Bit Quantization of KV Cache with Superior Accuracy Preservation**. Accepted at ICLR 2025 Workshop on Sparsity in LLMs 2025
- *Wenqi Pei, Hailing Xu, Henry Hengyuan Zhao,* ***CHEN Han***, *Zining Zhang, Shizheng Hou, Luo Pingyi, Bingsheng He.* **Optimizing Small Language Models for NL2SQL**. Accepted at ICLR 2025 Third Workshop on Deep Learning for Code 2025

项目经历

- **科研助理: 大语言模型推理优化**: 新加坡国立大学 2024/05–2024/09
  导师: 何丙胜教授
  - 设计了一种基于注意力模式的2位KV Cache量化方法, 在相同压缩率下, 提高了200%的准确率。
  - 实现了一个适应性API, 用于流行的推理框架, 如Python的transformers, 在不增加内存消耗的情况下, 将批处理大小提高了60%。
  - 该项目已被ICLR 2025 Sparsity in LLMs Workshop接受。 [Paper] [Code]
- **实习生: 密码算法工程师**: 华为2012实验室新加坡研究所谢尔德实验室 2024/01–2024/12
  导师: 黄涛博士
  - 设计了一种新的*'XAXX'*算法结构, 可以高效利用ARM和x86(带AES-NI)架构的流水线, 实现了高IPC。
  - 基于*'XAXX'*结构构建了一种新的AEAD(带关联数据的认证加密)密码算法, *'HiAE'*, 在多种平台上相较AES-256-GCM提升5倍以上性能, 在最新的ARM和x86处理器上优于所有现有的AEAD密码算法。

- 创造了单线程单流AEAD加密的340Gbps新纪录, 并应用于多种华为产品。
- 该项目已在Cryptology ePrint Archive上开源。 [Paper] [Code]

- **对称矩阵函数求解*BBK*算法的并行优化** 2022/04-2022/12

  导师: 汤德佑教授

  - 在ARM处理器上利用NEON指令集和openMP对Bounded Bunch-Kaufman算法(LAPACK库*sysv_rk 函数)进行并行优化。
  - 实现了一种并行列重排方法, 在列优先矩阵的行交换中改进访存局部性, 使得缓存命中率和并行性能得到提高, 在鲲鹏920-6426处理器上的单精度性能从320Gflops提升到580Gflops。
  - 将该方法移植到Intel Skylake处理器上, 对比MKL库的*sytrs_3函数, 实现了2-5倍的并行性能提升。
  - 该项目获评华南理工大学本科优秀毕业设计。

专业技能

- 英语认证水平: CET-4, CET-6, IELTS(6.5).
- 编程语言: C/C++, Fortran, p4-16, Python, SQL, LaTeX.
- 编程技能: openMP, SIMDs(NEON, AVX512), MPI, PyTorch, CUDA.
- *TestDemo* 编程技能认证: C++, TOP 10%, LINUX, TOP 10%, PYTHON, TOP 10%
- *Kaggle* 课程认证: 数据可视化, 机器学习, 深度学习, 强化学习

交换经历

- **机器学习线上访学项目, 麦吉尔大学** 2022/01–2022/02