

EDUCATION

- **Rice University** Houston, TX
Ph.D. of Science in Computer Science; GPA: 4.0 *Aug. 2018 – Aug. 2023 (expected)*
 - **Course:** Programming & Formal Methods, Compiler, IOT Programming, Multicore Computing, Computer Security, Computer Networks, Advanced Logics, Reasoning of Software, Artificial Intelligence, Computer Architecture
- **Shanghai Jiao Tong University** Shanghai, China
Bachelor of Computer Science with Honors in Zhiyuan College; GPA: 3.9 *Jul. 2014 – Jul. 2018*
 - **Selected Course:** Operating System, Mobile Networks, FPGA Programming, Electronic Circuit Design
- **Cornell University** Ithaca, NY
Computer Science Department, Visiting Student *Jun. 2017 – Jul. 2017*

PROFESSIONAL EXPERIENCE

- **Meta** Remote
Part-time Researcher *Aug. 2022 - Oct. 2022*
 - Developed a privacy-enhancing technology (**PET**) that protects users' **differential privacy** by rewriting SQL queries constructed as semantic trees in Meta's Unified Programming Model (**UPM**).
 - Integrated this PET into Meta's infrastructure with a multi-level cache using **ZippyDB** and **Laser**, which handles SQL queries (millions per day) from data analytics pipelines used for advertising applications.
 - Wrote **Python** programs to import logs to **Scuba** and visualize results on **Unidash** to enlighten engineers on why their queries may cause unsafe dataflow.
- **Meta** Boston, MA
Software Engineer Intern *May. 2022 - Aug. 2022*
 - Developed algorithms to detect SQL queries from **Presto** and **Scuba** that involve unsafe data aggregation.
 - Built a **Python** library that rewrites unsafe queries by adding noise to the computation of SQL aggregation in place.
 - Deployed the tool using **Conveyor** and **Tupperware** in the infrastructure, which handles ~10,000 pipelines per day.
- **Facebook** Remote
Software Engineer Intern *May. 2021 - Aug. 2021*
 - Developed algorithms for analyzing the patterns of SQL queries from **Apache Hive** that may cause unsafe data flow.
 - Wrote a **Python** library that automatically groups SQL queries with similar data-flow patterns into templates.
 - Developed a template cache using **Memcache** with an 80% hit rate that increases the speed of PES by 30x.

RESEARCH PROJECTS

- **Software Engine for High-performance Regular Expression Matching** 2022-2023
 - Developed efficient algorithms to match regular expressions with bounded repetitions.
 - Implemented a **Rust** engine for regular expression matching that integrates our efficient algorithms and other optimizations based on the use of Intel Streaming SIMD Extensions (**SSE**).
 - Experimentally validated that our engine is on average 10-100 times faster than state-of-the-art engines for >99% cases over regular expressions used in many real-world applications.
- **Software-Hardware Codesign for Efficient Regular Pattern Matching** 2021-2022
 - Developed a **Java** library that performs static analysis over regular expressions, which allows us to exponentially reduce the size of the automata representation of regular patterns.
 - Built a Java compiler that translates POSIX regular expressions into hardware-readable code.
 - Developed a specialized in-memory **ASIC** architecture and evaluated the hardware implementation using a simulator based on circuit parameters collected by SPICE simulation in TSMC 28nm CMOS process; Experiments concerning real-world workloads show our design achieves substantial energy/area reduction compared to SOTA designs such as Micron's AP and CAMA.
- **Query Language for Safe Parallel Stream Processing** 2020-2023

- Developed a **Rust** engine that supports a rich set of operators for programming streaming computations over time series targeting healthcare applications.
- Proposed a programming framework and algorithms to achieve parallelism with the preservation of sequential semantics, i.e., safe parallelism.
- Evaluated this engine against state-of-the-art engines and results show it is 1.2-50 times faster in single-threaded executions and provides superior scalability in parallel executions.
- **Query Language for Complex Analysis over Streaming Data** 2018-2020
 - Proposed a language that provides high-level programming abstractions for stream processing with formal semantics.
 - Implemented the language as a **Java** library with a rich set of operators, which is 5x faster than prior advanced tools.
 - Used the proposed language to prototype algorithms for healthcare monitoring and high-frequency trading.
- **Blockchain-based crowdsensing system** 2018
 - Built a crowd-sensing system based on blockchain to make up for the paucity of traditional crowd-sensing systems with security and low service fee.
 - Implemented BCS based on **Ethereum** with **Smart Contracts**.
 - Demonstrated the security and efficiency of BCS with both theoretical analysis and experimental evaluation.
- **Simulator for Market Analysis under Bancor Protocol** 2018
 - Built a **Python** simulator to simulate the trading of virtual currencies in the market under **Bancor** protocol.
 - Revealed that the Bancor protocol is flawed by experimental results in the following aspects: (1) The problem about “Double Coincidence of Wants” Bancor wants to solve is unsubstantiated in the real world. (2) The price of the smart token may fluctuate significantly, as Bancor neglects potential human behavior. (3) Severe cancellation of concurrent transactions occurs to Bancor under limited order.
- **Acemap: Academic Map System** 2015-2018
 - Developed tools in **Python** to visualize the relation between papers and authors for Acemap, a website like Google Scholar with a similar scale of data volumes.
 - Built a **Redis** in-memory database on more than 200M papers to decrease average response time by 93% using **PHP**.
 - Proposed two statistic models, EBM and MSM, for analyzing and predicting the evolution of academic networks.

PROGRAMMING SKILLS

- **Languages:** Java, Rust, Python, R, SQL, C/C++, C#, JavaScript, Coq, PHP, CUDA
- **Tools:** Smart Contract, Block Chain, Flink, Storm, Beam, Spark, Redis, Kafka Streams, React, Angular

PATENTS

- J. He, Y. Huang, **L. Kong**, J. Shen, C. Liu, Y. Jia, H. Xiao, W. Tang, T. Hu, L. Fu, X. Wang, “An Method to Construct & Visualize the Heterogeneous Topic Network Based on Text Information”, CHN No. 106372147A.

SELECTED PUBLICATIONS

- K. Mamouras, A. Le Glaunec, **L. Kong**. *Regular Expression Matching Using Bit Vector Automata*, OOPSLA’23 (to appear).
- **L. Kong**, Q. Yu, A. Chattopadhyay, A. Le Glaunec, Y. Huang, K. Mamouras, and K. Yang. *Software-Hardware Codesign for Efficient In-Memory Regular Pattern Matching*, PLDI’22.
- **L. Kong**, K. Mamouras. *StreamQL: A Query Language for Processing Streaming Time Series*, OOPSLA’20.
- J. Huang, **L. Kong**, L. Kong, Z. Liu, Z. Liu, and G. Chen. *Blockchain-based Crowd-sensing System*, HotICN’18.
- L. Fu, S. Ma, **L. Kong**, S. Shi, X. Wang, *FINE: A Framework for Distributed Learning on Incomplete Observations for Heterogeneous Crowdsensing Networks*, ToN’18.

SELECTED SCHOLARSHIP & HONORS

- **China National Scholarship:** Highest honor for undergraduates in China, top 0.2% nationwide 2015 & 2017
- **Zhiyuan Honor Scholarship:** Award for academic performance 2014 & 2015 & 2016 & 2018