Zhezhi (Elliot) He

Ph.D. Candidate
School of Electrical, Computer and Energy Engineering
Arizona State University
⚲ 650 E Tylor Mall
Tempe, Arizona. 85281 U.S.A.

📞 +1-541-602-9448
✈ zhezhihe@asu.edu
🏠 elliothe.github.io
in linkedin.com/in/zhezhi-he
G+ Scholar.google.com
○ github.com/elliothe

Areas of specialization

- Deep Learning
- Low-power VLSI
- Neuromorphic Computing
- Post-CMOS Device

Education

2020   **Ph.D., Electrical Engineering**
🏛 Arizona State University, Tempe, AZ, US
ADVISOR: Deliang Fan

2015   **M.eng., Electrical and Computer Engineering**
🏛 Oregon State University, Corvallis, OR, US

2012   **B.S., Information Engineering**
🏛 Southeast University, Nanjing, Jiangsu, CHN

Professional experience

2019-2020   Research Assistant
Arizona State University

2019   Research Scientist Intern
Machine Learning System Group, ByteDance AI Lab

2016-2019   Research Assistant
University of Central Florida

2016-2017   Teaching Assistant
University of Central Florida

Grants, honors & awards

2019   TECHCON travel grant, Semiconductor Research Corporation
2019   ACM SIGDA scholarship for Ph.D. forum, Design Automation Conference
2018   Best Paper Award, IEEE Computer Society Annual Symposium on VLSI [C20] (1/104)
2017   Best Paper Award, IEEE Computer Society Annual Symposium on VLSI [C37] (1/94)
2014   ISAS scholarship, Oregon State University

Publications & Presentations

= indicates equal contribution among authors.

RESEARCH DIRECTION HIGHLIGHT

- Security and Efficiency in Deep Learning:
  CVPR-2020 [C1, C3], CVPR-2019 [C9, C13], ICCV-2019 [C12], AAAI-2020 [C4], DAC-2020 [C5, C2]

- Brain-inspired In-memory Computing:
  DAC-2019 [C11], DAC-2018 [C21, C17], ICCAD-2018 [C16], DATE-2017 [C33], ICCD-2017 [C30], TCAD [J7], TETC [J11]

- Neural Network Accelerator on FPGA:
  GLSVLSI-2019 [C14], ISLPED-2018 [C24]

- Post-CMOS Device modeling:
  DRC-2019 [C8], MAGL [J10], GLSVLSI-2017 [C32], ISLPED-2016 [C38]

JOURNAL (TOTAL: 11)

[J1] Shaahin Angizi, <u>Zhezhi</u> He, An Chen, and Deliang Fan. "Hybrid Spin-CMOS Polymorphic Logic Gate With Application in In-Memory Computing". In: *IEEE Transactions on Magnetics* 56.2 (2020), pp. 1–15. 📄.

[J2] Durjoy Dev, Adithi Krishnaprasad, Mashiyat Shawkat, <u>Zhezhi</u> He, Sonali Das, Hee-Suk Chung, Deliang Fan, Yeonwoong Jung, and Tania Roy. "2D MoS2 Based Threshold Switching Memristor For Artificial Neuron". In: *IEEE Electron Device Letters* (2020). (Accepted).

[J3] <u>Zhezhi</u> He, Li Yang, Shaahin Angizi, Adnan Siraj Rakin, and Deliang Fan. "Sparse BD-Net: A Multiplication-less DNN with Sparse Binarized Depth-wise Separable Convolution". In: *ACM Journal on Emerging Technologies in Computing Systems (JETC)* 16.2 (2020), pp. 1–24. 📄.

[J4] Xiaolong Ma, Sheng Lin, Shaokai Ye, <u>Zhezhi</u> He, Linfeng Zhang, Geng Yuan, Sia Huat Tan, Zhengang Li, Deliang Fan, Xuehai Qian, et al. "Non-Structured DNN Weight Pruning–Is It Beneficial in Any Platform?" In: *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)* (2020). (Accepted).

[J5] Shaahin Angizi, <u>Zhezhi</u> He, Amro Awad, and Deliang Fan. "MRIMA: An MRAM-based In-Memory Accelerator". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* (2019). 📄.

[J6] Zhibo Wang, <u>Zhezhi</u> He, Milan Shah, Teng Zhang, Deliang Fan, and Wei Zhang. "Network-based multi-task learning models for biomarker selection and cancer outcome prediction". In: *Oxford academic Bioinformatics* (2019). 📄.

[J7] Shaahin Angizi, <u>Zhezhi</u> He, Nader Bagherzadeh, and Deliang Fan. "Design and Evaluation of a Spintronic In-Memory Processing Platform for Nonvolatile Data Encryption". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 37.9 (2018), pp. 1788–1801. 📄.

[J8] <u>Zhezhi</u> He, Yang Zhang, Shaahin Angizi, Boqing Gong, and Deliang Fan. "Exploring A SOT-MRAM based In-Memory Computing for Data Processing". In: *IEEE Transactions on Multi-Scale Computing Systems* (2018). 📄.

[J9] Farhana Parveen, Shaahin Angizi, <u>Zhezhi He</u>, and Deliang Fan. "IMCS2: Novel Device-to-Architecture Co-Design for Low-Power In-Memory Computing Platform Using Coterminous Spin Switch". In: *IEEE Transactions on Magnetics* 54.7 (2018), pp. 1–14. 📄.

[J10] <u>Zhezhi He</u>, Shaahin Angizi, and Deliang Fan. "Current -induced dynamics of multiple skyrmions with domain-wall pair and skyrmion-based majority gate design". In: *IEEE Magnetics Letters* 8 (2017), pp. 1–5. 📄.

[J11] <u>Zhezhi He</u> and Deliang Fan. "Energy efficient reconfigurable threshold logic circuit with spintronic devices". In: *IEEE Transactions on Emerging Topics in Computing* 5.2 (2017), pp. 223–237. 📄.

CONFERENCE (TOTAL: 38)

[C1] <u>Zhezhi He</u>, Adnan Siraj Rakin, Jingtao Li, Chaitali Chakrabarti, and Deliang Fan. "Defending and Harnessing the Bit-Flip based Adversarial Weight Attack". In: *Conference on Computer Vision and Pattern Recognition (CVPR)* (2020). (Accepted).

[C2] Jingtao Li, Adnan Siraj Rakin, Yan Xiong, Liangliang Chang, <u>Zhezhi He</u>, Deliang Fan, and Chaitali Chakrabarti. "Defending Bit-Flip Attack through DNN Weight Reconstruction". In: *57th Design Automation Conference (DAC)* (2020). (Accepted).

[C3] Adnan Siraj Rakin, <u>Zhezhi He</u>, and Deliang Fan. "TBT: Targeted Neural Network Attack with Bit Trojan". In: *Conference on Computer Vision and Pattern Recognition (CVPR)* (2020). (Accepted).

[C4] Li Yang, <u>Zhezhi He</u>, and Deliang Fan. "Harmonious Coexistence of Structured Weight Pruning and Ternarization for Deep Neural Networks". In: *Thirty-third AAAI Conference on Artificial Intelligence (AAAI)* (2020). (**Spotlight**).

[C5] Li Yang, <u>Zhezhi He</u>, and Deliang Fan. "Non-uniform DNN Structured Subnets Sampling for Dynamic Inference". In: *57th Design Automation Conference (DAC)* (2020). (Accepted).

[C6] Shaahin Angizi, <u>Zhezhi He</u>, and Deliang Fan. "ParaPIM: a parallel processing-in-memory accelerator for binary-weight deep neural networks". In: *Proceedings of the 24th Asia and South Pacific Design Automation Conference (ASP-DAC)*. ACM. 2019, pp. 127–132. 📄.

[C7] Shaahin Angizi, <u>Zhezhi He</u>, Dayane Reis, Sharon Xiaobo Hu, Wilman Tsai, Shy Jay Lin, and Deliang Fan. "Accelerating Deep Neural Networks in Processing-in-Memory Platforms: Analog or Digital Approach?" In: *2019 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. 2019, pp. 197–202. 📄.

[C8] Durjoy Dev, Adithi Krishnaprasad, <u>Zhezhi He</u>, Sonali Das, Mashiyat Sumaiya Shawkat, Madison Manley, Olaleye Aina, Deliang Fan, Yeonwoong Jung, and Tania Roy. "Artificial Neuron using Ag/2D-MoS 2/Au Threshold Switching Memristor". In: *2019 Device Research Conference (DRC)*. IEEE. 2019, pp. 193–194.

[C9] <u>Zhezhi He</u> and Deliang Fan. "Simultaneously Optimizing Weight and Quantizer of Ternary Neural Network using Truncated Gaussian Approximation". In: *Conference on Computer Vision and Pattern Recognition (CVPR)* (2019). 📄.

[C10] <u>Zhezhi He</u>, Boqing Gong, and Deliang Fan. "Optimize Deep Convolutional Neural Network with Ternarized Weights and High Accuracy". In: *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2019). 🎧 📄.

[C11] <u>Zhezhi He</u>, Jie Lin, Rickard Ewetz, Jiann-Shiun Yuan, and Deliang Fan. "Noise Injection Adaption: End-to-End ReRAM Crossbar Non-ideal Effect Adaption for Neural Network Mapping". In: *56-th Design Automation Conference (DAC)* (2019). 🎧 📄.

[C12] <u>Zhezhi He</u>=, Adnan Siraj Rakin=, and Deliang Fan. "Bit-Flip Attack: Crushing Neural Network with Progressive Bit Search". In: *International Conference on Computer Vision (ICCV)*. 2019. 🎧 📄.

[C13] <u>Zhezhi He</u>=, Adnan Siraj Rakin=, and Deliang Fan. "Parametric Noise Injection: Trainable Randomness to Improve Deep Neural Network Robustness against Adversarial Attack". In: *Conference on Computer Vision and Pattern (CVPR)* (2019). 🎧 📄.

[C14] Li Yang, <u>Zhezhi He</u>, and Deliang Fan. "Binarized Depthwise Separable Neural Network for Object Tracking in FPGA". In: *Great Lakes Symposium on VLSI (GLVLSI)* (2019). 📄.

[C15] Shaahin Angizi, <u>Zhezhi He</u>, Yu Bai, Jie Han, Mingjie Lin, Ronald F DeMara, and Deliang Fan. "Leveraging Spintronic Devices for Efficient Approximate Logic and Stochastic Neural Networks". In: *Proceedings of the 2018 on Great Lakes Symposium on VLSI*. ACM. 2018, pp. 397–402. 📄.

[C16] Shaahin Angizi, <u>Zhezhi He</u>, and Deliang Fan. "DIMA: a depthwise CNN in-memory accelerator". In: *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*. IEEE. 2018, pp. 1–8. 📄.

[C17] Shaahin Angizi, <u>Zhezhi He</u>, and Deliang Fan. "PIMA-logic: a novel processing-in-memory architecture for highly flexible and energy-efficient logic computation". In: *Proceedings of the 55th Annual Design Automation Conference (DAC)*. ACM. 2018, p. 162. 📄.

[C18] Shaahin Angizi, <u>Zhezhi He</u>, Farhana Parveen, and Deliang Fan. "Imce: energy-efficient bit-wise in-memory convolution engine for deep neural network". In: *Proceedings of the 23rd Asia and South Pacific Design Automation Conference*. IEEE Press. 2018, pp. 111–116. 📄.

[C19] <u>Zhezhi He</u>, Shaahin Angizi, and Deliang Fan. "Accelerating Low Bit-Width Deep Convolution Neural Network in MRAM". In: *2018 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE. 2018, pp. 533–538. 📄.

[C20] <u>Zhezhi He</u>, Shaahin Angizi, Adnan Siraj Rakin, and Deliang Fan. "BD-NET: A Multiplication-Less DNN with Binarized Depthwise Separable Convolution". In: *2018 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE. 2018, pp. 130–135. **Best Paper Award** 📄.

[C21] <u>Zhezhi He</u>=, Shaahin Angizi=, Adnan Siraj Rakin, and Deliang Fan. "CMP-PIM: an energy-efficient comparator-based processing-in-memory neural network accelerator". In: *Proceedings of the 55th Annual Design Automation Conference (DAC)*. ACM. 2018, p. 105. 📄.

[C22] Farhana Parveen, <u>Zhezhi He</u>, Shaahin Angizi, and Deliang Fan. "HielM: Highly flexible in-memory computing using STT MRAM". In: *2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC)*. IEEE. 2018, pp. 361–366. 📄.

[C23] Adnan Siraj Rakin, Shaahin Angizi, <u>Zhezhi He</u>, and Deliang Fan. "PIM-TGAN: A Processing-in-Memory Accelerator for Ternary Generative Adversarial Networks". In: *2018 IEEE 36th International Conference on Computer Design (ICCD)*. IEEE. 2018, pp. 266–273. 📄.

[C24] Li Yang, <u>Zhezhi He</u>, and Deliang Fan. "A Fully Onchip Binarized Convolutional Neural Network FPGA Impelmentation with Accurate Inference". In: *Proceedings of the International Symposium on Low Power Electronics and Design (ISLPED)*. ACM. 2018, p. 50. 📄.

[C25] Shaahin Angizi, <u>Zhezhi He</u>, Ronald F DeMara, and Deliang Fan. "Composite spintronic accuracy-configurable adder for low power digital signal processing". In: *2017 18th International Symposium on Quality Electronic Design (ISQED)*. IEEE. 2017, pp. 391–396. 📄.

[C26] Shaahin Angizi, <u>Zhezhi He</u>, and Deliang Fan. "Energy efficient in-memory computing platform based on 4-terminal spin hall effect-driven domain wall motion devices". In: *Proceedings of the on Great Lakes Symposium on VLSI 2017*. ACM. 2017, pp. 77–82. 📄.

[C27] Shaahin Angizi, <u>Zhezhi He</u>, Farhana Parveen, and Deliang Fan. "Rimpa: A new reconfigurable dual-mode in-memory processing architecture with spin hall effect-driven domain wall motion device". In: *2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE. 2017, pp. 45–50. 📄.

[C28] Deliang Fan, Shaahin Angizi, and <u>Zhezhi He</u>. "In-memory computing with spintronic devices". In: *2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE. 2017, pp. 683–688. 📄.

[C29]   Deliang Fan, <u>Zhezhi He</u>, and Shaahin Angizi. "Leveraging spintronic devices for ultra-low power in-memory computing: Logic and neural network". In: *2017 IEEE 6oth International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE. 2017, pp. 1109–1112. ⧉.

[C30]   <u>Zhezhi He</u>, Shaahin Angizi, and Deliang Fan. "Exploring stt-mram based in-memory computing paradigm with application of image edge extraction". In: *2017 IEEE International Conference on Computer Design (ICCD)*. IEEE. 2017, pp. 439–446. ⧉.

[C31]   <u>Zhezhi He</u>, Shaahin Angizi, Farhana Parveen, and Deliang Fan. "High performance and energy-efficient in-memory computing architecture based on sot-mram". In: *2017 IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH)*. IEEE. 2017, pp. 97–102. ⧉.

[C32]   <u>Zhezhi He</u>, Shaahin Angizi, Farhana Parveen, and Deliang Fan. "Leveraging dual-mode magnetic crossbar for ultra-low energy in-memory data encryption". In: *Proceedings of the on Great Lakes Symposium on VLSI 2017*. ACM. 2017, pp. 83–88. ⧉.

[C33]   <u>Zhezhi He</u> and Deliang Fan. "A tunable magnetic skyrmion neuron cluster for energy efficient artificial neural network". In: *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017*. IEEE. 2017, pp. 350–355. ⧉.

[C34]   <u>Zhezhi He</u> and Deliang Fan. "Developing All-Skyrmion Spiking Neural Network". In: *Neuromorphic Computing Symposium (NCS)* (2017). ⧉.

[C35]   Farhana Parveen, Shaahin Angizi, <u>Zhezhi He</u>, and Deliang Fan. "Hybrid polymorphic logic gate using 6 terminal magnetic domain wall motion device". In: *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE. 2017, pp. 1–4. ⧉.

[C36]   Farhana Parveen, Shaahin Angizi, <u>Zhezhi He</u>, and Deliang Fan. "Low power in-memory computing based on dual-mode SOT-MRAM". In: *2017 IEEE/ACM International Symposium on Low Power Electronics and Design (ISLPED)*. IEEE. 2017, pp. 1–6. ⧉.

[C37]   Farhana Parveen, <u>Zhezhi He</u>, Shaahin Angizi, and Deliang Fan. "Hybrid polymorphic logic gate with 5-terminal magnetic domain wall motion device". In: *2017 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*. IEEE. 2017, pp. 152–157. **Best Paper Award** ⧉.

[C38]   <u>Zhezhi He</u> and Deliang Fan. "A Low Power Current-mode Flash ADC with Spin Hall Effect Based Multi-threshold Comparator". In: *Proceedings of the 2016 International Symposium on Low Power Electronics and Design*. ACM. 2016, pp. 314–319. ⧉.

Under-review or Archived

[UA1]   Wuyang Zhang, <u>Zhezhi He</u>, Luyang Liu, Zhenhua Jia, Yunxin Liu, Marco Gruterser, Dipankar Raychaudhuri, and yanyong zhang. "Elf: Accelerate High-resolution Mobile Deep-Vision with Content-aware Distributed Offloading". In: *18th USENIX Symposium on Networked Systems Design and Implementation (NSDI)* (2021).

[UA2]   Adnan Siraj Rakin, <u>Zhezhi He</u>, Yanzhi Wang, Liqiang Wang, and Deliang Fan. "Robust Sparse Regularization: Adversarial Defense with Sparsity and Compactness". In: *arXiv* (2020).

[UA3]   Adnan Siraj Rakin, <u>Zhezhi He</u>, Boqing Gong, and Deliang Fan. "Robust Pre-Processing: A Robust Defense Method Against Adversary Attack". In: *arXiv preprint arXiv:1802.01549* (2018).

Talk

[T1]   *Secure and Efficient AI Computing System: A Software and Hardware Co-design Perspective.* University of Illinois at Chicago. Chicago, Illinois, USA, Mar. 2020.

[T2]   *Secure and Efficient AI Computing System: A Software and Hardware Co-design Perspective.* Binghamton University. Binghamton, New York, USA, Mar. 2020.

[T3] *Secure and Efficient AI Computing System: A Software and Hardware Co-design Perspective.* University of Central Florida. Orlando, Florida, USA, Mar. 2020.

[T4] *Secure and Efficient AI Computing System: A Software and Hardware Co-design Perspective.* University of Nevada, Reno. Reno, Nevada, USA, Mar. 2020.

[T5] *Noise Injection Adaption: End-to-End ReRAM Crossbar Non-ideal Effect Adaption for Neural Network Mapping.* 56-th Design Automation Conference (DAC). Las Vegas, NV, USA, June 2019.

[T6] *High Performance and Energy-efficient In-memory Computing Architecture based on SOT-MRAM.* 2017 IEEEACM International Symposium on Nanoscale Architectures. Newport, Rhode Island, USA, July 2017.

Poster

[P1] *Design and Evaluation of in-Memory Processing Unit (MPU) for Data-Centric Computations.* Semiconductor Research Corporation TECHCON. Austin, TX, USA, Sept. 2019.

[P2] *Parametric Noise Injection: Trainable Randomness to Improve Deep Neural Network Robustness against Adversarial Attack.* Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA, June 2019.

[P3] *Simultaneously Optimizing Weight and Quantizer of Ternary Neural Network using Truncated Gaussian Approximation.* Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA, June 2019.

Projects

2017-2020 "Non-Volatile In-Memory Processing Unit: Memory, In-Memory Logic and Deep Neural Network" Jointly Funded by National Science Foundation (NSF) and Semiconductor Research Corporation (SRC) nCore ($185K).

- Assist the principal investigator with proposal writing.

- Develop in-memory computing system for data-centric applications, and open-source tools.

Teaching

2016-2017 EEL3801 Computer Organization and Design
EGN3373 Introduction to Electrical System I
EEE3342 Digital System: Lab instructor
EEE4309 Electronics II: Lab instructor
EEL4742 Embedded System: Lab instructor

Service to the profession

Technical Program Committee Member

2020 30th edition of the ACM Great Lakes Symposium on VLSI (GLSVLSI)

Journal review

2020 Reviewer of IEEE Transactions on Parallel and Distributed Systems (TPDS)
2020 Reviewer of IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
2020 Reviewer of IEEE Transactions on Cloud Computing (TCC)

2020	Reviewer of IEEE Design & Test
2020	Reviewer of ElSEVIER Microprocessors and Microsystems
2020	Reviewer of CCF Transactions on High Performance Computing
2019	Reviewer of ELSEVIER Journal of Artificial Intelligence (AI)
2019-2020	Reviewer of IEEE Transactions on Circuits and Systems I: Regular Papers (TCAS-I)
2019	Reviewer of IOPscience Applied Physics Express
2019	Reviewer of IEEE Transactions on Electron Devices (TED)
2019	Reviewer of IEEE Access
2018	Reviewer of Microelectronics Journal
2017-2018	Reviewer of IEEE Transactions on Nanotechnology (TNANO)
2017-2020	Reviewer of Integration, the VLSI Journal
2016-2017	Reviewer of IEEE Transactions on Computers (TC)

Conference review

2020	Reviewer of Design, Automation and Test in Europe Conference (DATE)
2020	Reviewer of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
2020	Reviewer of IEEE International Symposium on High-Performance Computer Architecture (HPCA)
2020	Reviewer of IEEE Winter Conference on Application of Computer Vision (WACV)
2018-2020	Expert reviewer of Design Automation Conference (DAC)
2018-2019	Reviewer of International Conference On Computer Aided Design (ICCAD)
2018-2019	Reviewer of ACM Great Lakes Symposium on VLSI (GLSVLSI)
2019	Reviewer of International Symposium on Quality Electronic Design (ISQED)
2018-2019	Reviewer of IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)
2018	Reviewer of IEEE Non-Volatile Memory Systems and Applications Symposium
2017	Reviewer of IEEE Computer Society Annual Symposium on VLSI (ISVLSI)
2017	Reviewer of Asia and South Pacific Design Automation Conference (ASP-DAC)

Courses

- Machine learning
- VLSI system design
- Analog CMOS integrated circuit
- Information theory
- Electric and magnetic fields
- Signals and systems
- Probability and random signals
- Computer architecture
- Linear system
- RF microwave circuit design

Skills

Languages

- English (professional-proficiency)
- Mandarin (native)

Scripting

Machine Learning Framework (Pytorch, Torch, TensorFlow)
Scientific programming (Python, MATLAB, C++/C)
System language (Bash shell, Tcl/Tk)
Hardware language (Verilog/VHDL, SPICE)
Digital typesetting (Latex, Microsoft Word)

EDA tools

- Cadence (IC)
- ADS (RF)
- Mentor Pads (PCB)
- Design-Compiler (VLSI)
- ISE design suit (FPGA)
- OOMMF (Magnetic)

References

Available upon request