Title: Intelligent Memory for Efficient Hardware Accelerator

Memory architecture and design have been critical for digital systems to achieve ample storage, low latency, fast access time, and energy efficiency, especially for battery-operated devices. The increase of data generated by many devices such as mobile, sensors, communications, and security not only increased the requirements on memory capacity but also increased the challenges on memory access and energy. The memory interface has limited throughput and high latency, which has not been scaling at the same rate as data size or processing speed; this limits the performance of accessing the data, which refer to as the memory wall. In addition to the negative impact on latency and performance, large data movement results in high energy consumption. Research has been focusing on elevating the memory wall issue by engineering more memory hierarchy and increasing local on-chip memory. This has partially reduced the timing issue but did not address the high leakage and active energy consumption. It is estimated that more than 60% of energy spent on most computing platforms is spent on data movements and memory access. The new era of big data and artificial intelligence-based applications has increased the urgency to solve memory capacity, data movement energy, and memory wall issues. Some solutions have brought processing into centralized cloud computing, with high performance and large memory hardware capacity available. However, this brought a new challenge to communications, privacy, security, and latency, especially for real-time applications.

The goal of this lecture is to highlight the after mentioned challenges and to present a new paradigm of computing beyond von Neuman's architecture to enable processing as close to the data source as possible. This includes in-memory computing, near memory computing architecture. Both existing and emerging memory technologies will be explored. Since the new computing paradigm is more data-centric than traditional processing-centric, the traditional single architecture for all applications is not feasible, but rather a domain-specific architecture and hardware solutions need to be adopted. Popular high computing functions such as Query, MAC, hamming distance, and image compression will be presented as an example of in-memory hardware accelerators.

Dr. Baker Mohammad is the director of the System on Chip center and professor of EECS at Khalifa University. Before joining Khalifa University, he was a Senior Staff Engineer/Manager at Qualcomm, Austin, Tx, USA, for 6-years, where he was engaged in designing high-performance and low-power DSP processors used for communication and multi-media application. Before joining Qualcomm, he worked for ten years at Intel Corporation on a wide range of microprocessors design from high-performance server chips > 100Watt (IA-64) to mobile embedded processors low power sub 1 watt (xscale). He has over 16 years of industrial experience in microprocessor design, emphasizing memory, low power circuit, and physical design. Baker earned his PhD from the University of Texas at Austin in 2008, his M.S. degree from Arizona State University, Tempe, and his BS degree from the University of New Mexico, Albuquerque, all in ECE. His research interests include VLSI, power-efficient computing, embedded memory and in-memory computing, neuromorphic computing, emerging technology such as Memristor, STTRAM, hardware accelerators for Cyber-Physical Systems and AI. He is also engaged in a microwatt range computing platform for wearable electronics and WSN, focusing on energy harvesting, power management, and power conversion, including efficient dc/dc, ac/dc converters.

Baker authored/co-authored over 200 referred journals and conference proceedings, >5 books, >20 US patents, multiple invited seminars/panellists, and the presenter of >3 conference tutorials, including one tutorial on Energy Harvesting and Power Management for WSN at the 2015 (ISCAS). Baker is on the advisory board for the secure systems research center part of the Technology Innovation Institute. Baker is an associate editor for IEEE Transaction on VLSI (TVLSI), IEEE Access, and Scientific Reports journals. Dr Mohammad participates in technical committees at IEEE conferences and reviews for TVLSI, IEEE Circuits and Systems journals. He has received several awards, including the KUSTAR staff excellence award in intellectual property creation, IEEE TVLSI best paper award, 2016 IEEE MWSCAS Myrill B. Reed best paper award, and Qualcomm Qstar award for excellence in performance and leadership. SRC Techon's best session papers for 2016 and 2017. 2009 Best paper award for Qualcomm Qtech conference and Intel Involve in the community award for volunteer and impact on the community