Recent Advances in Flash Solutions

Sangyeun Cho

Vice President at Samsung (Memory Solutions Lab)

ABSTRACT

In this keynote talk, I will deep dive into two topics related with now common flash memory solutions: Multi-streaming and In-Storage Computing (ISC). Multi-streaming allows upper-level software to see “streams” — internal parallelism within a solid-state drive (SSD) — in a disciplined way. Proper use of multiple streams segregates data having different properties. Our study using popular NoSQL DB engines shows that significant gains can be achieved in terms of steady-state SSD performance and device lifetime. ISC refers to new capabilities of running user-written codes within the SSD, helped by a software framework (and hardware IPs). By exploiting ample bandwidth of flash chips and processing cycles within the SSD, one can significantly improve the system throughput by balancing work between compute resources and/or shortening data trips. I will discuss results of several case studies using database and data analytics workloads.

SPEAKER BIOGRAPHY

Sangyeun Cho received the BS degree in computer engineering from Seoul National University in 1994 and the PhD degree in computer science from the University of Minnesota in 2002. In 1999, he joined the System LSI Division of Samsung Electronics Co. and contributed to the development of Samsung’s flagship embedded processor core family CalmRISC™. He was a lead architect of CalmRISC-32, a 32-bit microprocessor core, and designed its memory hierarchy including caches, DMA, and stream buffers. In 2004, he joined the faculty of the Computer Science Department at the University of Pittsburgh and was promoted to the rank of tenured associate professor in 2010. He has recently joined Samsung’s Memory Division to lead systems-related research efforts, where he is a VP for advanced solutions research and development. His research interests are in the area of computer architecture and systems with particular focus on performance, power and reliability of memory and storage hierarchy design for next-generation data centers.