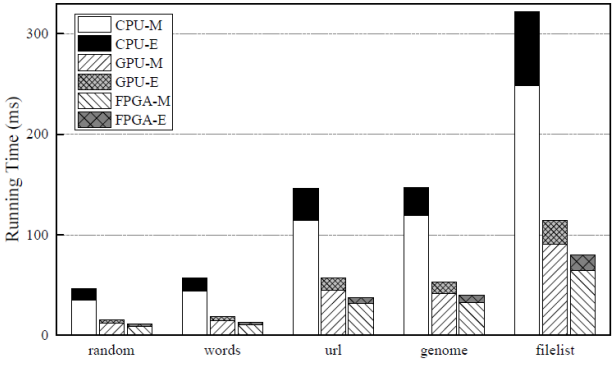
ReCSA: A Dedicated Sort Accelerator Using ReRAM-based Content Addressable Memory

Huize LI, Hai JIN, Long ZHENG, Yu HUANG, Xiaofei LIAO

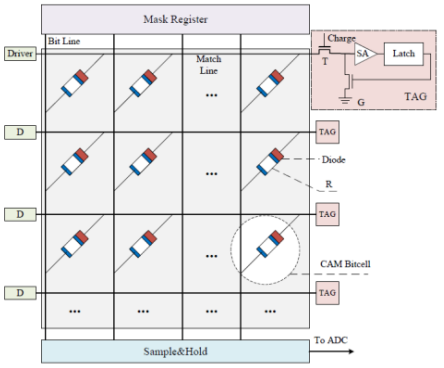
Frontiers of Computer Science, DOI: 10.1007/s11704-022-1329-9

Problems & Ideas

- Problems of current hardware sorting accelerators:
  - Traditional Von-Neumann-based sorting accelerators has memory wall since the multi-rounds features of sorting, leading to poor sorting performance.
  - Existing PIM sorting algorithm cannot process sorting efficiently for lacking feasible data mapping strategy and can not support a various of data types. 2) PIM-based sorting algorithm is designed for other domains accelerators such as ReSQM for database accelerator, which has a poor performance to perform sorting applications
- Ideas: Emerging ReCAM has in-situ processing capability and can process parallel comparison, which is the basic computational primitive in sorting. Therefore, we are motivated to design a ReCAM-based specific sorting accelerator to eliminate the memory wall in sorting. Further, we want this specific accelerator can solve the problems in current PIM-based accelerators and boosting the sorting performance.



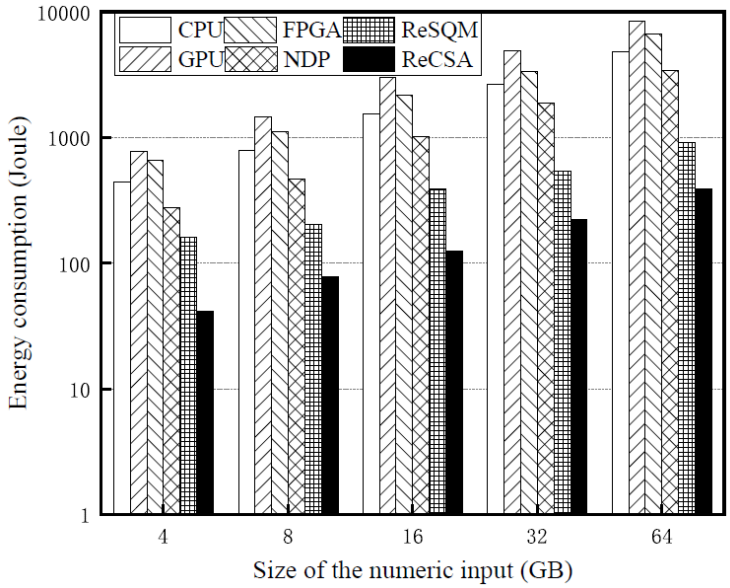
Time breakdown of 3 accelerators running five real-world datasets (CPU-M means CPU's memory access time and CPU-E means CPU's execution time).



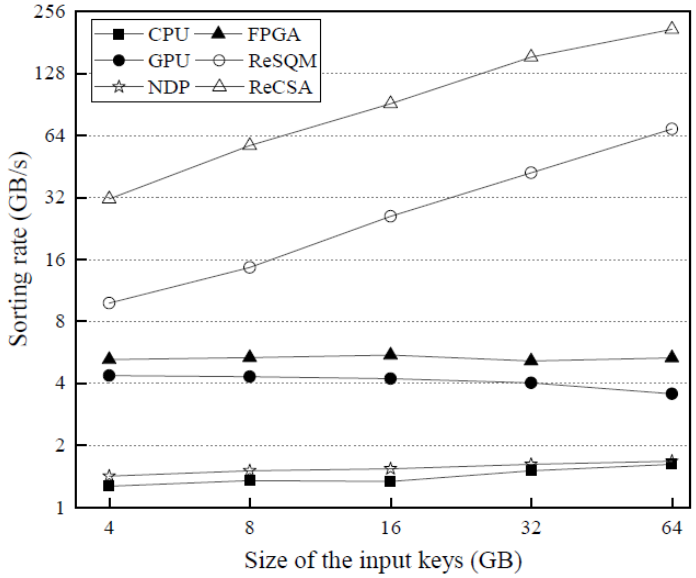
Designed 1D1R ReCAM array

Main Contributions

- Contributions:
  - We design a new type of ReCAM array that can support matrix-vector multiplication and the vector-scalar comparison operations simultaneously;
  - Using this designed ReCAM array, we propose our data mapping strategies for sorting to improve memory utilization. We also present detailed architecture and algorithm designs for various data types;
  - The experimental results show ReCSA can achieve the best time and energy performance among all state-of-the-art hardware sort accelerators.



Energy consumption when perform numeric key-only workloads at Zipf is 0.75.



Response time when perform numeric key-only workloads when Zipf is 0.75.