PageRank Acceleration for Large Graphs with Scalable Hardware and Two-Step SpMV

Fazle Sadi, Joe Sweeney, Scott McMillan, Tze Meng Low, James C. Hoe, Larry Pileggi and Franz Franchetti

Contact: fsadi@cmu.edu

Carnegie Mellon University

Electrical & Computer ENGINEERING

Research Problem

PageRank is a widely used iterative algorithm that ranks the vertices of a graph according to their relative importances. However, it suffers from poor performance and efficiency due to notorious memory access behavior. More importantly, when graphs become bigger and sparser, PageRank applications are inhibited as most solutions profoundly rely on large random access fast memory, which is not easily scalable.

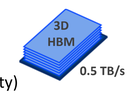
PageRank vector: $\boldsymbol{x}_i = \underbrace{\alpha\, \boldsymbol{x}_i^T \boldsymbol{A}}_{\text{SpMV}} + (1-\alpha)\, \boldsymbol{x}_i^T \frac{\boldsymbol{e}\boldsymbol{e}^T}{N}$

Target Graphs

- Very large (~billion nodes)
- Highly sparse (average degree < 10)
- No exploitable non-zero pattern

Key Goals

- Streaming DRAM access
- Full utilization of streaming bandwidth
- Off-chip DRAM traffic reduction
- Low requirement of fast on-chip memory (scalability)



Proposed Solution

We designed a 16nm ASIC (currently under fabrication) based shared memory platform for PageRank that fundamentally accelerates **Sparse Matrix dense Vector multiplication (SpMV),** the core kernel of PageRank.

Custom Hardware with 3D High Bandwidth Memory

- Efficient SpMV implementation
- Scalable – less dependence on SRAM/eDRAM

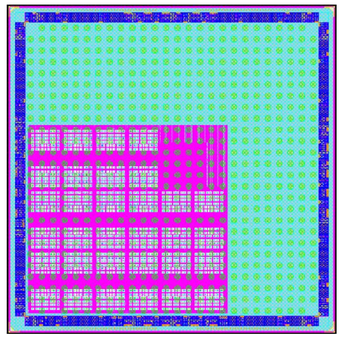
Two-Step SpMV

- Guarantees DRAM streaming
- No dependence on non-zero pattern or structure

Iteration Overlapped PageRank

- Saturates HBM bandwidth
- Reduces off-chip DRAM traffic

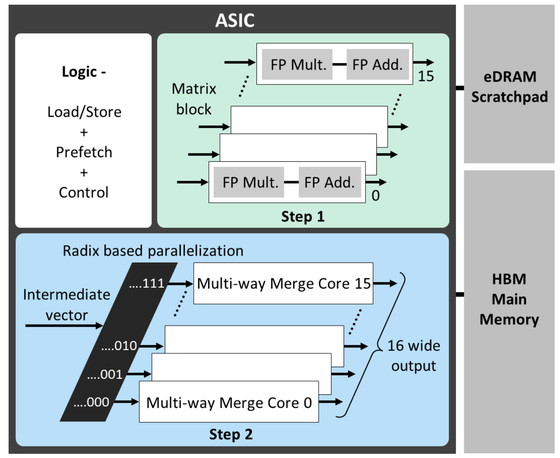
16nm FinFET ASIC for PageRank Acceleration (currently being fabricated)



Freq.: 1.4 GHz Leakage power: 0.10 W

Area: 7.5 mm² Dynamic power: 3.01 W

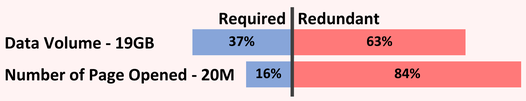
Block Diagram of the Accelerator



Background on SpMV

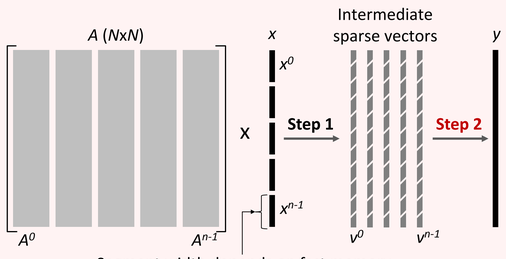
SpMV, the core operation of PageRank, requires random access to dense vector which is too large for the last level cache. This results in random access to DRAM, affecting performance and efficiency in many ways. Below is a comparison of required vs redundant data transfer and dram page openings of baseline SpMV on a example problem.

Example: 80M x 80M matrix, 3 NNZ per row, 1KB DRAM page, 64B cache block, double precision data



Two-Step SpMV Algorithm

This algorithm conducts SpMV in two separate steps. It requires column blocking of the matrix and segmentation of the source vector.



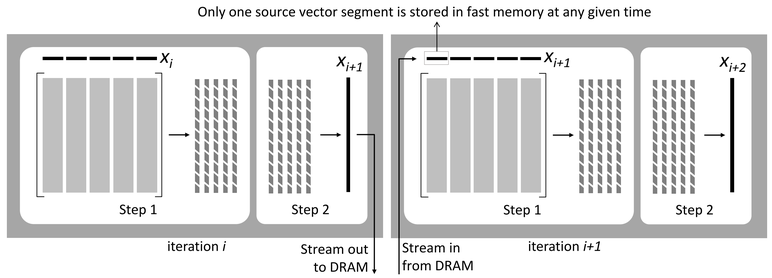
Segment width depends on fast memory

- ❑ **Guarantees full DRAM streaming access**
- ❑ **Reduces off-chip traffic and Enables high bandwidth utilization**
- ❑ **Requires custom hardware for efficient multi-way merge**

Less Dependence on SRAM - More Scalable

Solution	Fast memory (MB)	Max. nodes reported
FPGA [3]	8.4	2.3M
ASIC [4]	32	8M
CPU (single socket) [5]	20	95M
CPU (single socket) [6]	50	118M
PR_TS_Opt (proposed)	11	2B
PR_TS (proposed)	11	4B

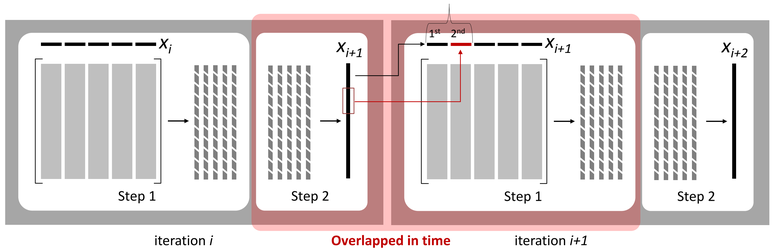
PageRank using Two-Step SpMV (PR_TS)



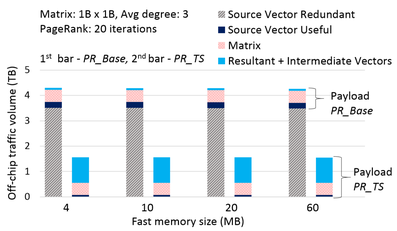
- ❑ **Two-Step SpMV is conducted independently in each iteration of PageRank**
- ❑ **Resultant vector of one iteration is the source vector of the next, communicated via DRAM**
- ❑ **Either Step1 or Step2 is conducted at any given time, keeping other part of the circuit inactive**

Optimized PageRank by Iteration Overlap (PR_TS_Opt)

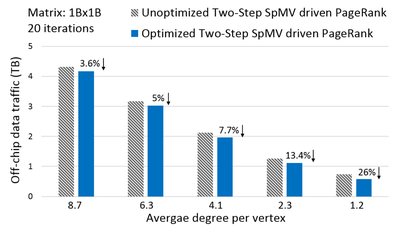
Two source vector segment storages in fast memory are required:
1) for computation of Step1 in iteration *i+1* and 2) for storing output of Step 2 in iteration *i*.



- ❑ **Step 2 of an iteration is conducted simultaneously with Step 1 of the next iteration**
- ❑ **Reduces off-chip traffic by eliminating DRAM round trip of source and resultant vectors**
- ❑ **Simultaneous execution of Step 1 and 2 almost doubles the throughput and saturates HBM bandwidth**

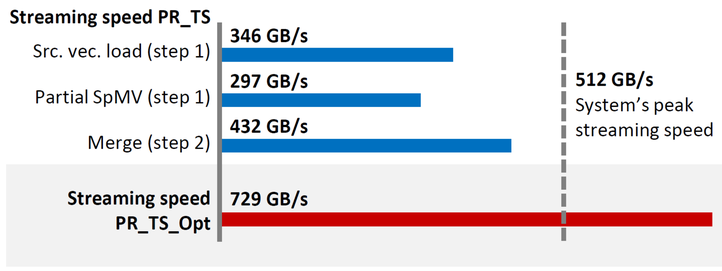


PageRank off-chip traffic comparison: PR_TS vs Baseline

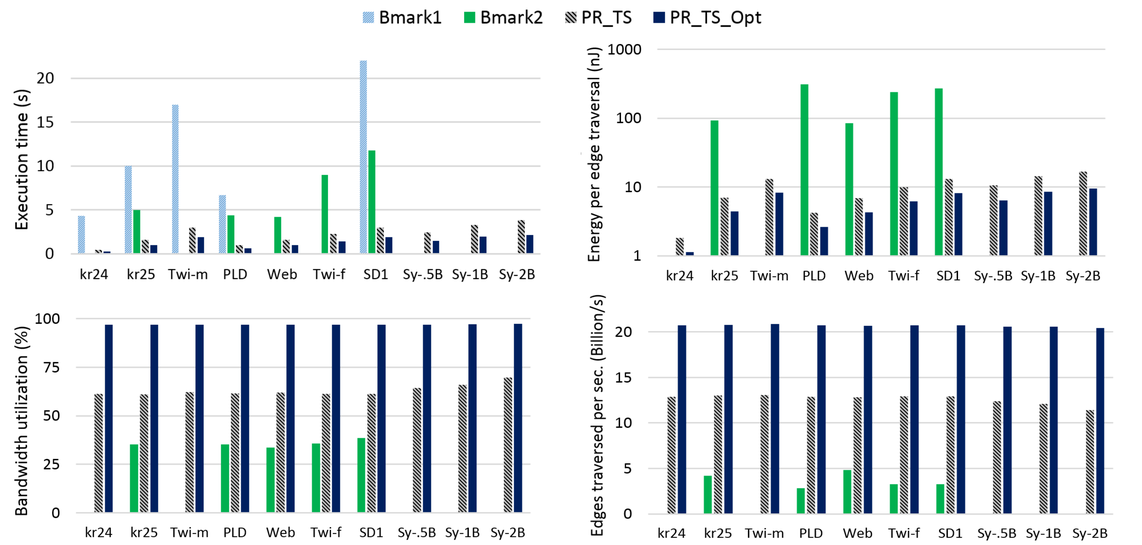


PageRank off-chip traffic comparison: PR_TS vs PR_TS_Opt

Experimental Results



Optimized Two-Step driven PageRank (PR_TS_Opt) attains much higher throughput with the same silicon area and saturates HBM's extreme off-chip bandwidth .



Bmark1 – Single socket CPU implementation [5], Bmark2 – Dual socket CPU implementation [6]

Graph	kr24	kr25	Twi-m	PLD	Web	Twi-f	SD1	Sy-.5B	Sy-1B	Sy-2B
# Nodes (M)	16.7	33.5	52.5	42.9	118	61.6	94.9	500	1000	2000
Avg. degree	16.1	31.3	37.4	14.5	8.6	23.8	20.4	3	2	1.1
# Edges (M)	268	1047	1963	623	1014	1468	1936	1500	2000	2200

Acknowledgements

This work was supported in part by Defense Advanced Research Projects Agency (DARPA) contract HR0011-16-C-0038, "Circuit Realization At Faster Timescales (CRAFT)". This material is also based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center [DM18-0845]. Additionally, this work was sponsored by DARPA contract HR0011-13-2-0007, "Power Efficiency Revolution for Embedded Computing Technologies (PERFECT)". The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.