

Towards an End-to-End Processing-in-DRAM Acceleration of Spectral Library Search

Tianyun Zhang*, Eric Tang*, Farzana Ahmed Siddiquet†, Kevin Skadron†, Franz Franchetti*

*Department of Electrical and Computer Engineering, College of Engineering, Carnegie Mellon University

†Department of Computer Science, School of Engineering & Applied Science, University of Virginia

Introduction

Mass-spectrometry (MS)-based proteomics generates vast amounts of data, with modern instruments producing millions of spectra per experiment. Spectral library search is a crucial workload in protein identification. Traditional compute-centric architectures struggle with data movement between memory and processors, a challenge that grows with increasing dataset sizes. Processing-in-memory (PIM) architectures offer a potential solution by performing computations directly within or near memory, reducing data movement. This work is a first look at exploring PIM's potential to accelerate spectral library searches by integrating hardware and algorithm co-design across the MS data analysis pipeline.

We make the following contributions:

- Adaptation of an approximate sum of absolute differences algorithm for PIM-based spectral matching
- A preliminary approach to integrate PIM into mass spectrometry workflows
- Early performance assessments via PIM simulation

DRAM + Processing-in-memory (PIM)

Processing-in-memory (PIM)

- Modern workloads are data-intensive and limited by memory bandwidth
- Data movement across long, narrow off-chip memory channel between CPU and DRAM incur high energy and latency costs
- Processing-in-memory (PIM), a.k.a. near-data-processing (NDP)
- Reduces data movement by adding compute capability to memory

DRAM for PIM

- DRAM-based PIM can benefit from simultaneous access to multiple DRAM arrays for parallel data processing

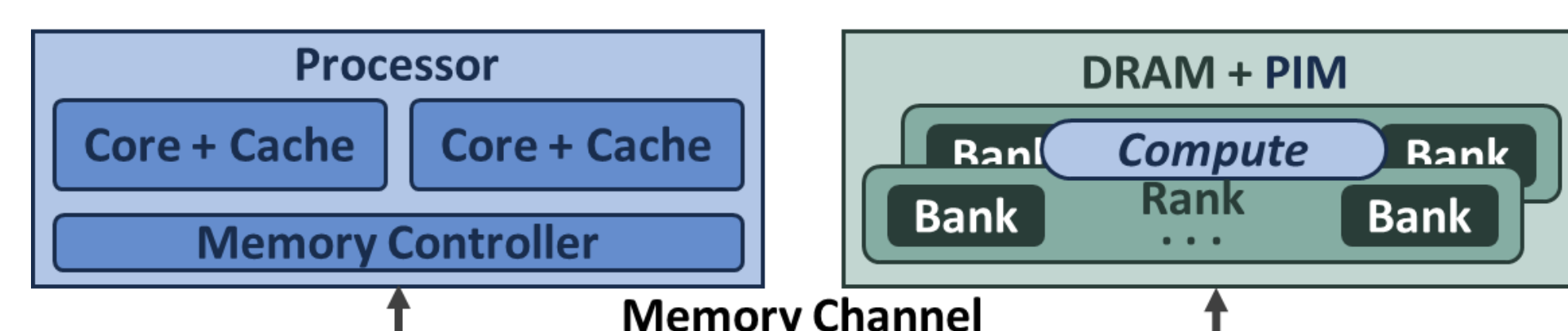
PIM types

1. Processing-using-memory (PUM)

- Use inherent circuit-level properties of memory cells to perform logic within memory arrays

2. Processing-near-memory (PNM)

- 3D-stacked memory with a logic layer
- Take advantage of the high bandwidth communication over vertical interconnects between layers



Processor and DRAM+PIM architecture alleviates data movement bottleneck by adding compute units to DRAM

Mass Spectrometry (MS) and Spectral Library Search

Mass spectrometry (MS) is a method to analyze the proteome of protein and peptide samples

- **Tandem MS (MS/MS)**: multiple staged MS for structural studies of complex molecules

Mass spectrum plot

- Mass-to-charge ratio (m/z) on the y-axis
- Ion signal intensity (percentage) on the x-axis

Spectral library search

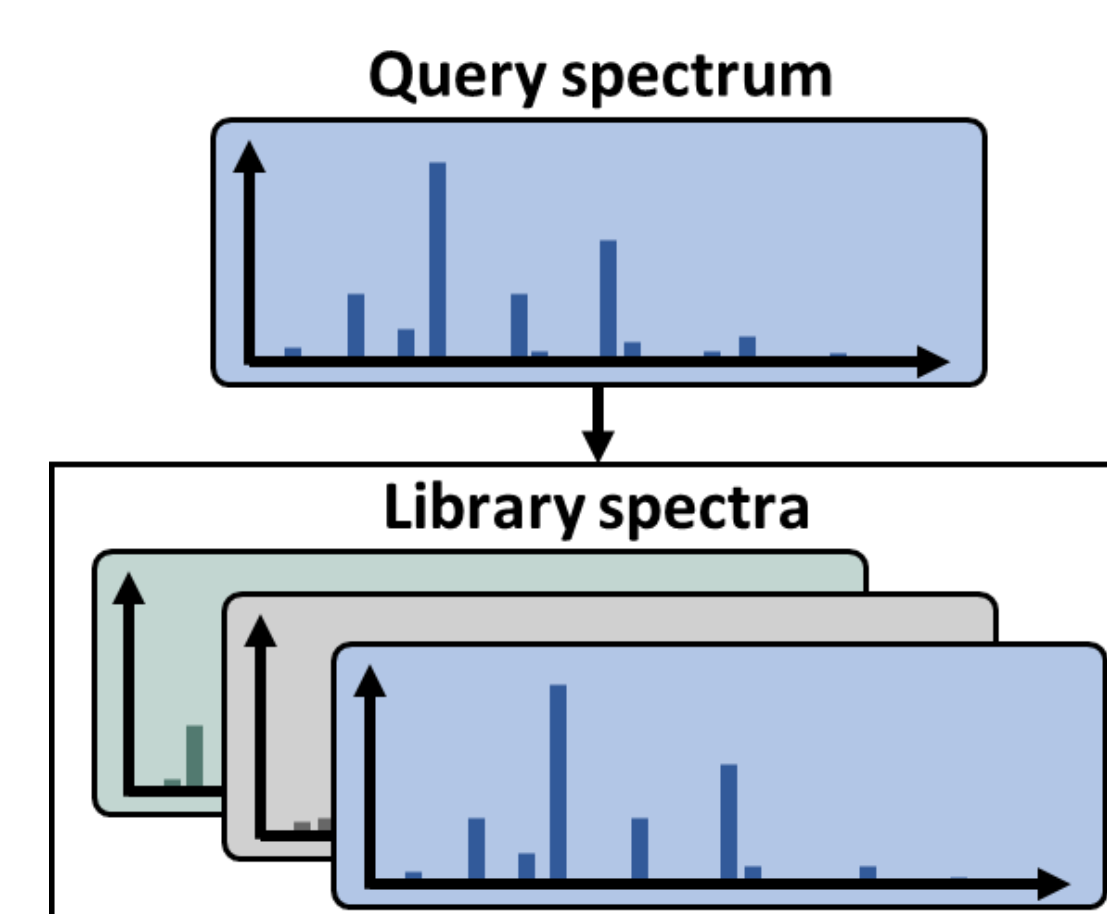
- Used to interpret MS/MS data
- The search process involves calculating similarity scores between an experimental spectrum and each library spectrum

PIM for spectral library search

- Modern MS experiments generate massive data: millions of mass spectra must be matched against libraries of known spectra to identify molecules
- Can do parallel comparisons of query spectra with many library spectra
- Suitable candidate workload for PIM

```
BEGIN IONS
TITLE=17398
RTINSECONDS=4052.711
PEPMASS=744.086138662863
CHARGE=8+
102.0587082 13
103.0541916 16
...
1417.616333 6.806824207
1443.578491 6.806824207
END IONS
```

Standard format for MS/MS data: mascot generic format (MGF).



Spectral library search identifies experimental spectrum by querying it against library spectra.

Minimum Sum of Absolute Differences

We use sum of absolute differences to calculate a similarity score between the experimental spectrum $S = (s_1, s_2, \dots, s_n)$ with each library spectrum $L_p = (\ell_{p1}, \ell_{p2}, \dots, \ell_{pn})$ for $p=1, 2, \dots, m$.

Assuming m library spectra and n elements in each spectrum, the basic algorithm seeks L_k such that:

$$k = \arg \min_p \sum_{j=1}^n |s_j - \ell_{pj}|$$

Pre-processing and Pre-filtering

Discretization

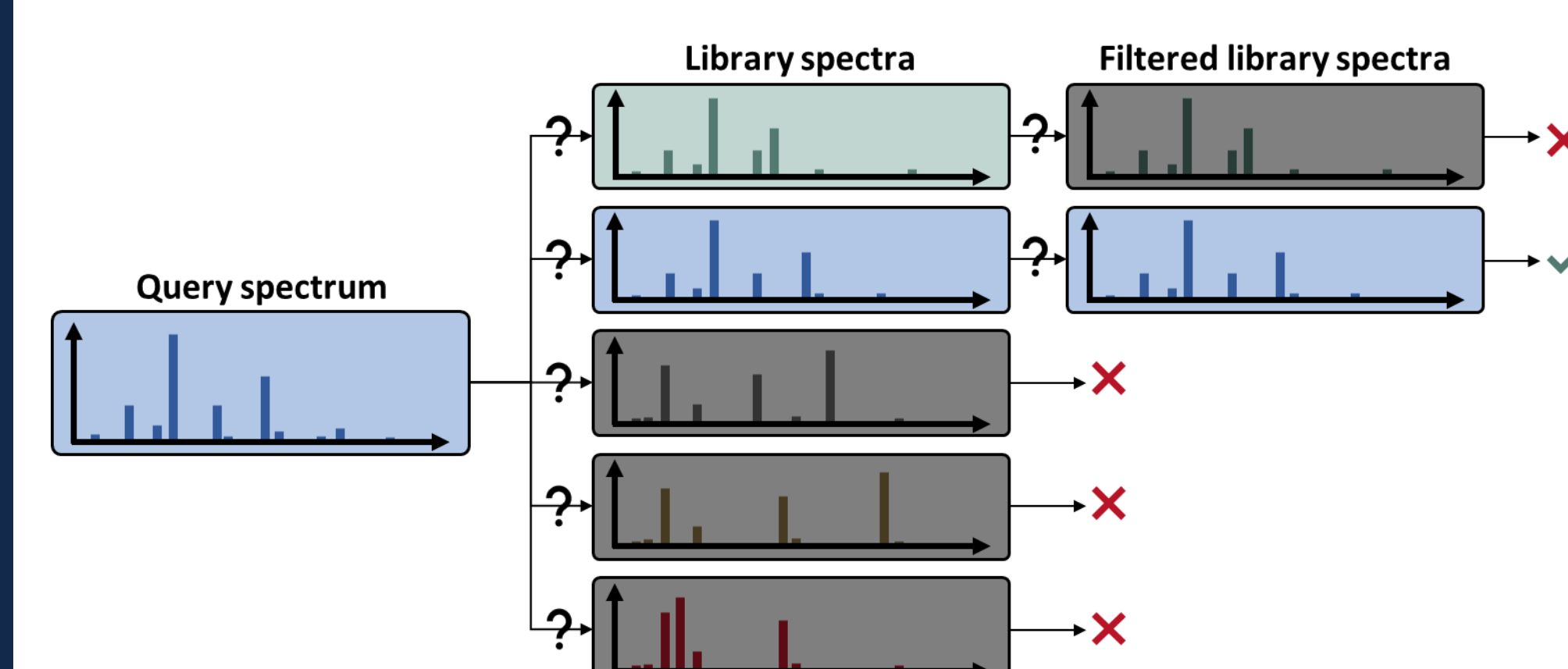
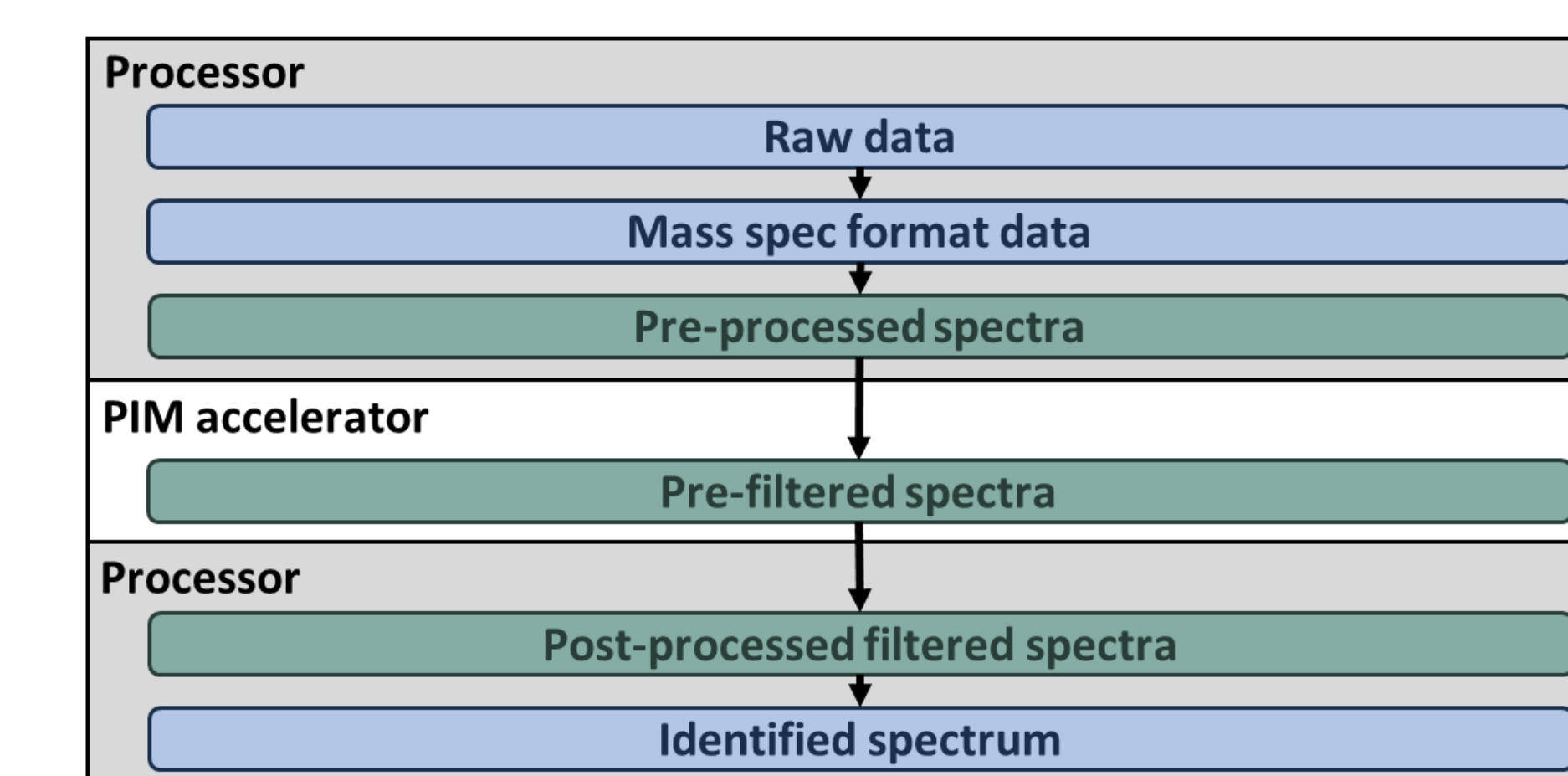
- Divide the m/z range of spectra into bins and summing intensity values in the same bin

Approximation

- Fixed-point for floating-point values with reduced bit-precision enables program execution on PIM hardware

Key insights:

- Co-design application with hardware
- Low-precision approach aligns with MS error tolerance



Pre-processed and pre-filtered spectral matching reduces noise & decreases search space.

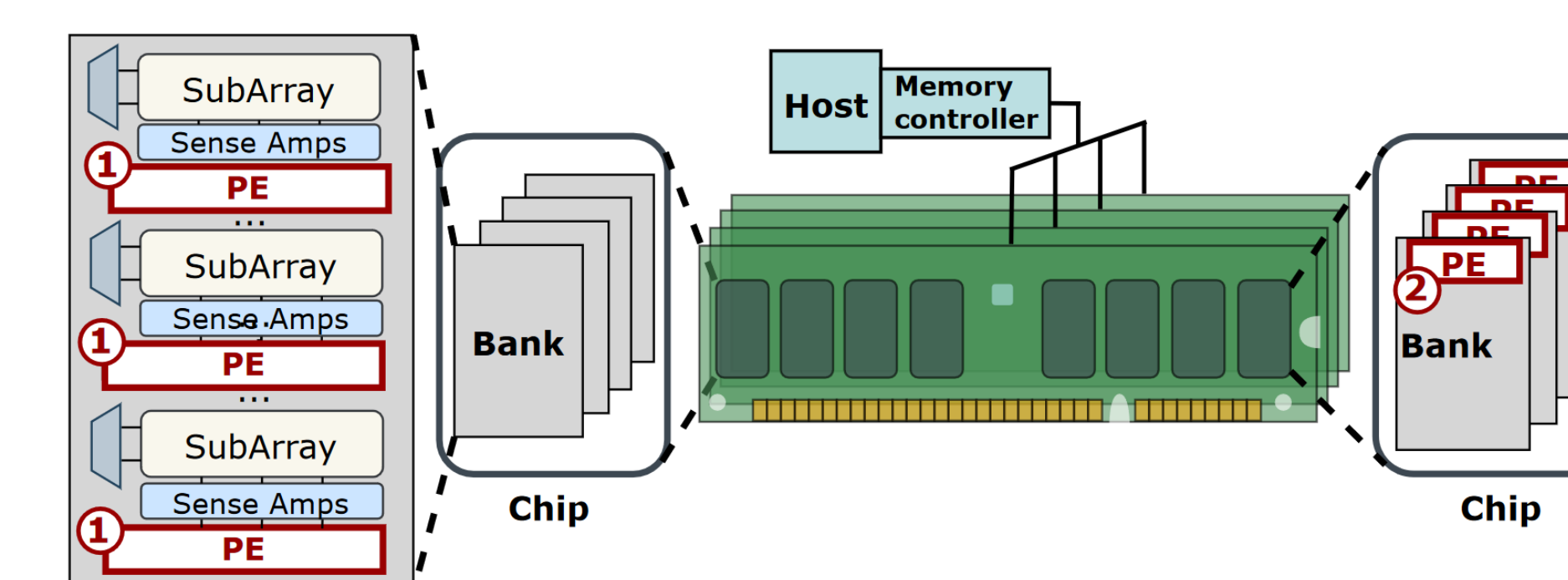
PIM Simulator

PIMEval

- Performance and energy simulator for diverse DRAM-based PIM architectures
- We model subarray-level bit serial PIM

PIM device parameters

Single rank DIMM with 8 chips, 16 banks per rank, and 32 subarrays per rank of 8192×8192 cells.



(1) Subarray-level bit-serial and bit-parallel, (2) bank-level.

Evaluation

```
for (int idx=0; idx < subvecLen; idx++) {
    pimSub(obj1, obj2, obj3);
    pimAbs(obj3, obj3);
    for (int i=idx; i+subvecLen-1 < vecLen; i+=subvecLen) {
        pimRedSumRangedInt(obj3, i, i+subvecLen-1, &sumAbsDiff);
        ...
    }
    pimRotateElementsRight(obj2);
}
```

Current 11 MB problem on 2-rank DIMM

- CPU implementation completed in 175 ms
- PIM implementation took 1368 ms

Key insights:

- Workload underutilized available hardware parallelism
- PIM performance benefits scale with increasing problem size
- Small datasets favor CPU due to lower overhead

Conclusions

We have demonstrated a proof of concept for an end-to-end, hardware-algorithm co-designed pipeline for DRAM-based PIM architecture acceleration of spectral library search, a key workload for mass spectrometry. The next step is to observe the performance from taking better advantage of the massive parallelism in DRAM for more reasonable comparisons and to connect larger volumes of data from real-world workloads to the hardware for a full-system view of PIM accelerated bioinformatics applications.

References

- [1] O. Mutlu, S. Ghose, J. Gómez-Luna, and R. Ausavarungnirun, "A modern primer on processing in memory," in Emerging Computing: From Devices to Systems. Springer, 2023, pp. 171–243.
- [2] J. Kang, W. Xu, W. Bittremieux, N. Moshiri, and T. Rosing, "Accelerating open modification spectral library searching on tensor core in high-dimensional space," Bioinformatics, vol. 39, no. 7, p. btad404, 2023.
- [3] V. R. Pagala, A. A. High, X. Wang, H. Tan, K. Kodali, A. Mishra, K. Kavdia, Y. Xu, Z. Wu, and J. Peng, "Quantitative protein analysis by mass spectrometry," Protein-protein interactions: Methods and applications, pp. 281–305, 2015.
- [4] J. M. Chick, D. Kolippakkam, D. P. Nusinow, B. Zhai, R. Rad, E. L. Huttlin, and S. P. Gygi, "A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides," Nature biotechnology, vol. 33, no. 7, pp. 743–749, 2015.
- [5] F. A. Siddique, D. Guo, Z. Fan, M. Gholamrezaei, M. Baradaran, A. Ahmed, H. Abbot, K. Durrer, K. Nandagopal, E. Ermovick et al., "Architectural modeling and benchmarking for digital dram pim.

Acknowledgements

This work was supported in part by the Jump 2.0 PRISM center, a Semiconductor Research Corporation (SRC) program sponsored by DARPA.