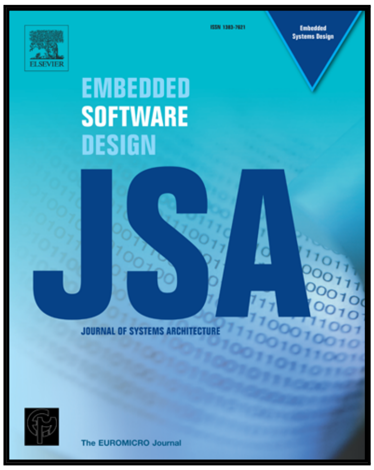
Accepted Manuscript



Exploring the Processing-in-Memory Design Space

Marko Scrbak , Mahzabeen Islam , Krishna M. Kavi ,
Mike Ignatowski , Nuwan Jayasena

PII: S1383-7621(16)30086-8
DOI: 10.1016/j.sysarc.2016.08.001
Reference: SYSARC 1378

To appear in: *Journal of Systems Architecture*

Received date: 3 May 2016
Accepted date: 6 August 2016

Please cite this article as: Marko Scrbak , Mahzabeen Islam , Krishna M. Kavi , Mike Ignatowski , Nuwan Jayasena , Exploring the Processing-in-Memory Design Space, *Journal of Systems Architecture* (2016), doi: 10.1016/j.sysarc.2016.08.001

Exploring the Processing-in-Memory Design Space

Marko Scrbak[1], Mahzabeen Islam[1], Krishna M. Kavi[1], Mike Ignatowski[2], and Nuwan Jayasena[2]

[1]University of North Texas, USA
{markoscrbak,mahzabeenislam}@my.unt.edu,krishna.kavi@unt.edu
[2]AMD Research - Advanced Micro Devices, Inc., USA
{mike.ignatowski,nuwan.jayasena}@amd.com

1 ABSTRACT

With the emergence of 3D-DRAM, Processing-in-Memory has once more become of great interest to the research community and industry. Here we present our observations on a subset of the PIM design space. We show how the architectural choices for PIM core frequency and cache sizes will affect the overall power consumption and energy efficiency. We include a detailed power consumption breakdown for an ARM-like core as a PIM core. We show the maximum possible number of PIM cores we can place in the logic layer with respect to a predefined power budget. Additionally, we catalog additional sources of power consumption in a system with PIM such as 3D-DRAM link power and discuss the possible power reduction techniques. We describe the shortcomings of using ARM-like cores for PIM and discuss other alternatives for the PIM cores. Finally, we explore the optimal design choices for the number of cores as a function of performance, utilization, and energy efficiency.

Keywords: Processing in memory; Heterogeneous Computing; High Performance Computing; 3D-DRAM; Energy Efficient Computing

2 Introduction

Over the last decade, we have witnessed the Big Data processing evolution. Existing commodity systems, which are widely used in the Big Data processing community, are becoming less energy efficient and fail to scale in terms of power consumption and area [21] clearly shows that this is also true for any Scale-Out workloads in general. Therefore using hardware accelerators to aid the Big Data processing is becoming more and more prominent. With the evolution of new emerging DRAM technologies, in particular 3D-DRAM, Processing-in-Memory (PIM) has again become of great