# PGAS in-memory data processing for the Processing Unit of the Upgraded Electronics of the Tile Calorimeter of the ATLAS Detector

Daniel Ohene-Kwofie

University of the Witwatersrand
Johannesburg, South Africa

High Energy Particle Physics Workshop 2015
February 2015

# Outline

# Introduction

- The continuously growing gap between CPU and I/O speed presents a great challenge to the performance in high performance and high-throughput computing.

# Introduction

- The continuously growing gap between CPU and I/O speed presents a great challenge to the performance in high performance and high-throughput computing.
- The compute power of CPUs has been growing in line with Moore’s law ever since it was predicted.
- The performance of storage subsystems has not kept up with the performance requirement of applications.

# Introduction

- The continuously growing gap between CPU and I/O speed presents a great challenge to the performance in high performance and high-throughput computing.
- The compute power of CPUs has been growing in line with Moore's law ever since it was predicted.
- The performance of storage subsystems has not kept up with the performance requirement of applications.
- The rapid growth in data volumes from scientific projects such as the Large Hadron Collider (LHC) for instance, further increases the computing challenge.

# Introduction

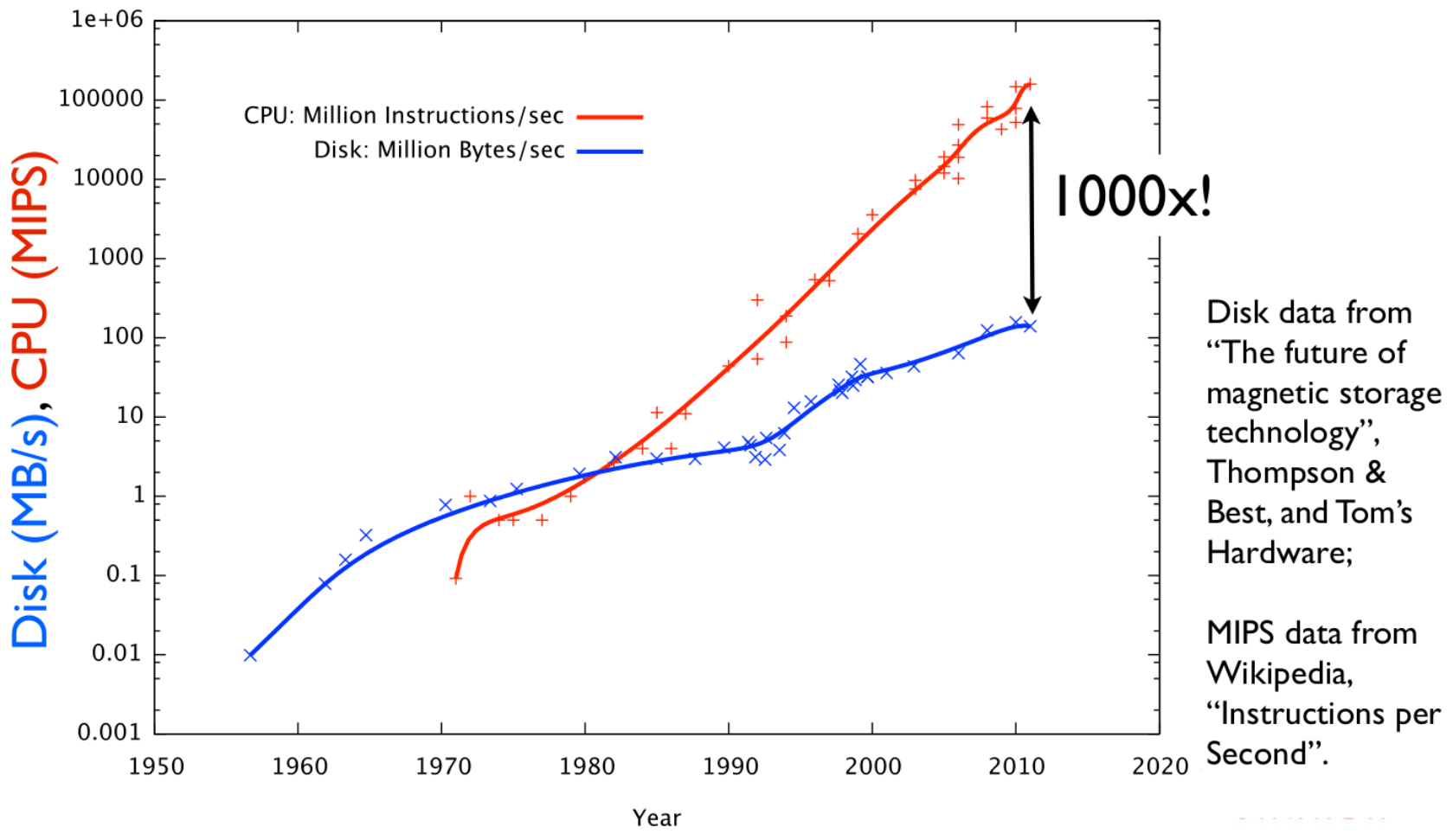


Figure: Increasing speed gap between disks and CPUs

# Introduction: ATLAS LHC Tile Calorimeter

- TileCal is the central hadronic calorimeter of the ATLAS experiment at the Large Hadron Collider (LHC) at CERN.

# Introduction: ATLAS LHC Tile Calorimeter

- TileCal is the central hadronic calorimeter of the ATLAS experiment at the Large Hadron Collider (LHC) at CERN.
- Scheduled upgrades in 2022 will result in an increase in data output from the TileCal to about *41 Tb/s*.

# Introduction: ATLAS LHC Tile Calorimeter

- TileCal is the central hadronic calorimeter of the ATLAS experiment at the Large Hadron Collider (LHC) at CERN.
- Scheduled upgrades in 2022 will result in an increase in data output from the TileCal to about *41 Tb/s*.
- Storing such massive data for offline processing is infeasible and presents a great challenge.

# Introduction: ATLAS LHC Tile Calorimeter

- TileCal is the central hadronic calorimeter of the ATLAS experiment at the Large Hadron Collider (LHC) at CERN.
- Scheduled upgrades in 2022 will result in an increase in data output from the TileCal to about *41 Tb/s*.
- Storing such massive data for offline processing is infeasible and presents a great challenge.
- The Super Read-Out Driver (sROD) is a core element of the back-end electronics after the upgrades.
  - The sROD will perform some processing before sending data to the rest of the triggering and data acquisition system.

# Introduction: ATLAS TileCal Read Out Architecture

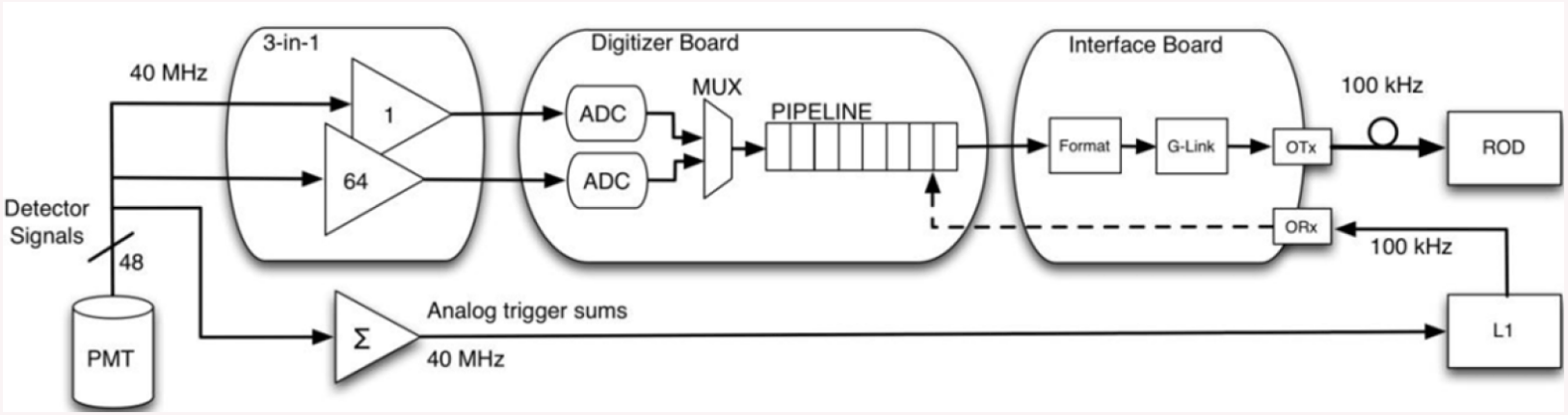


Figure: ATLAS TileCal current read out architecture

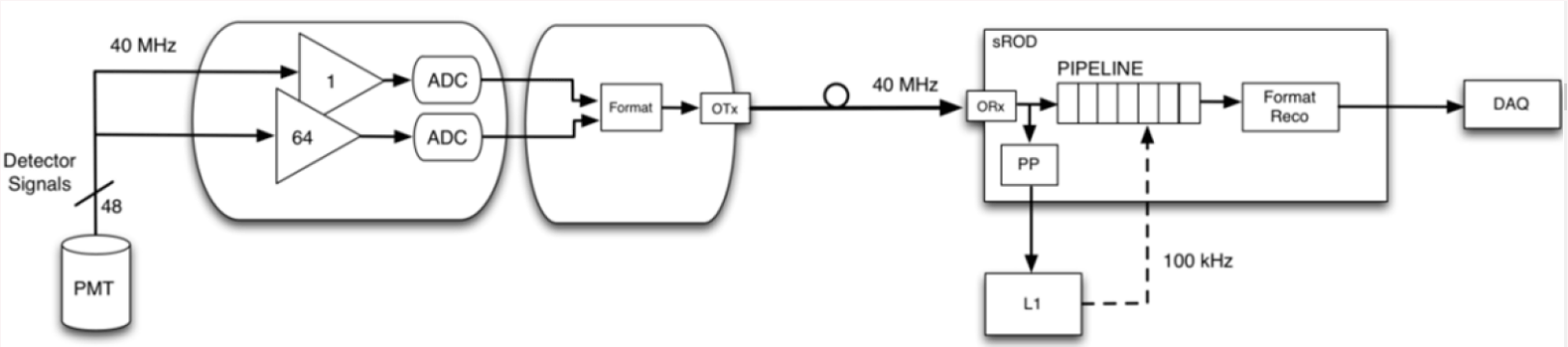


Figure: ATLAS TileCal upgraded read out architecture with sROD

# The Motivation

- The need for general purpose processing to enhance high-data throughput at the sROD of the upgraded TileCal.

# The Motivation

- The need for general purpose processing to enhance high-data throughput at the sROD of the upgraded TileCal.
- There is a general growing need for fast storage systems that match the computational speed and efficiency required for high data throughput.

# The Motivation

- The need for general purpose processing to enhance high-data throughput at the sROD of the upgraded TileCal.
- There is a general growing need for fast storage systems that match the computational speed and efficiency required for high data throughput.

# The Motivation

- The need for general purpose processing to enhance high-data throughput at the sROD of the upgraded TileCal.
- There is a general growing need for fast storage systems that match the computational speed and efficiency required for high data throughput.

MAC project at Wits is developing a cost-effective, and high data throughput Processing Unit (PU) using ARM processors.

# The Motivation

- The need for general purpose processing to enhance high-data throughput at the sROD of the upgraded TileCal.
- There is a general growing need for fast storage systems that match the computational speed and efficiency required for high data throughput.

MAC project at Wits is developing a cost-effective, and high data throughput Processing Unit (PU) using ARM processors.

The PU will serve as a general purpose co-processor to the sROD to enhance and provide new functionality that is difficult to implement on FPGA.

# General Purpose PU for the sROD

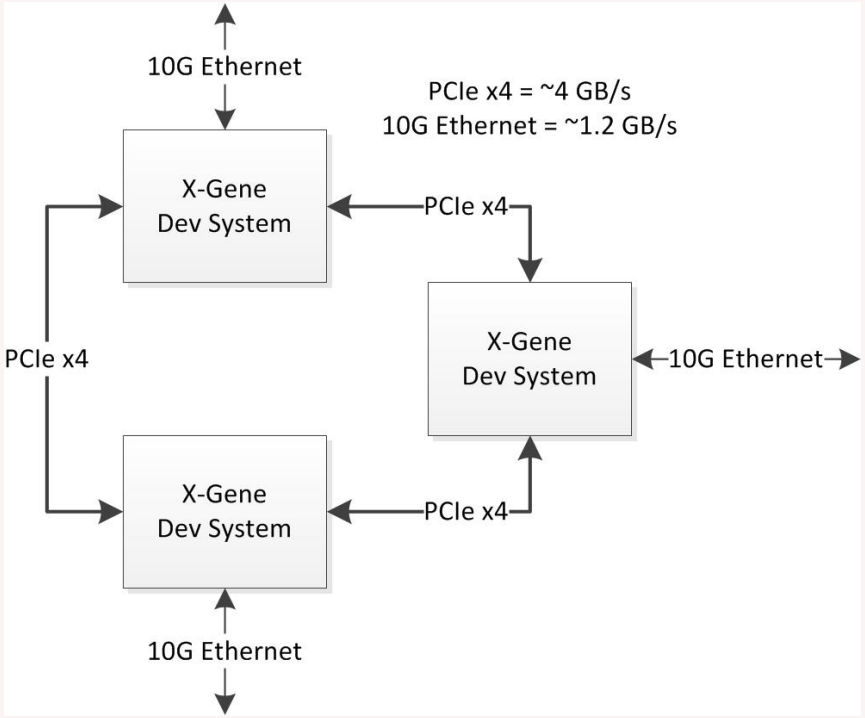


Figure: Schematic: PU architecture

# General Purpose PU for the sROD

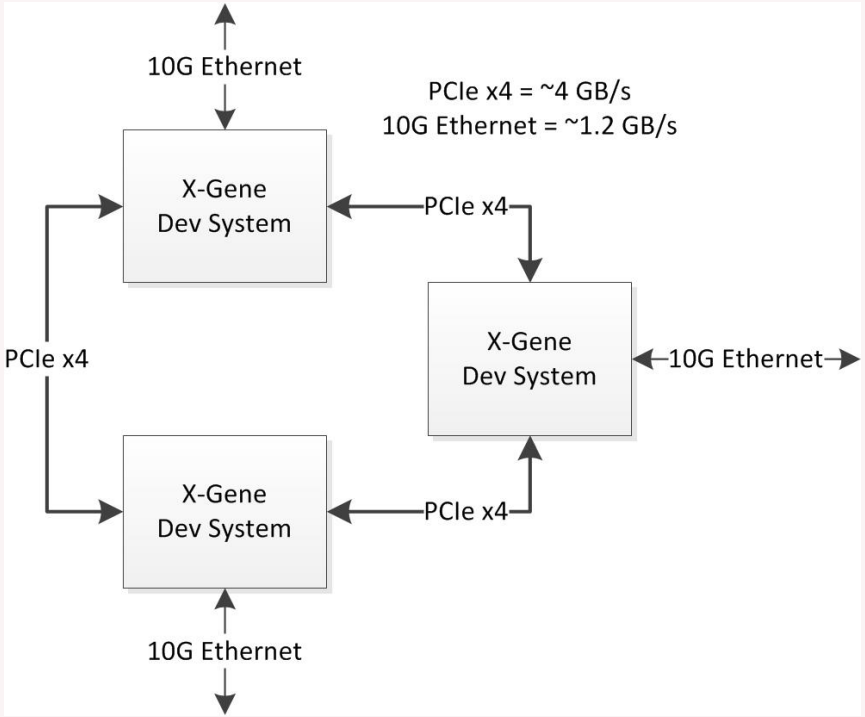


Figure: Schematic: PU architecture

How do we leverage this design to further enhance on-line data processing for high throughput for the sROD?

# The PGAS Strategy

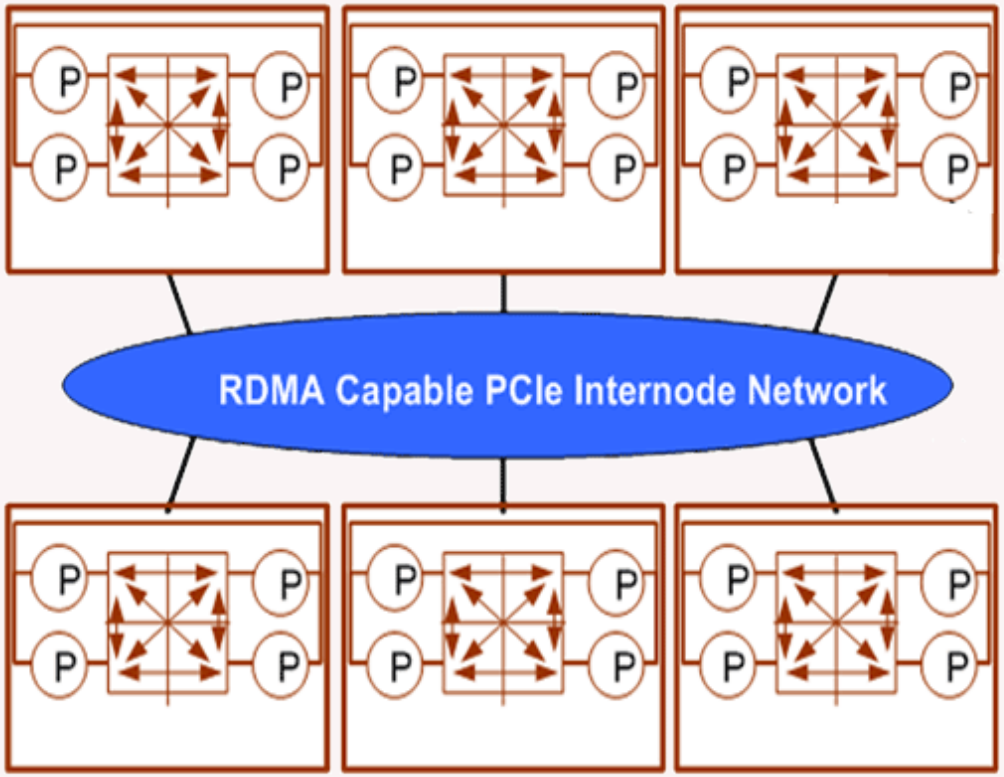


Figure: Schematic: PGAS layout with PU

# The The PGAS Strategy

Memory aggregation to provide a global logical memory address space for data processing.

# The The PGAS Strategy

Memory aggregation to provide a global logical memory address space for data processing.

## Provides..

- ease of general purpose processing using semantics similar to that of shared memory systems.
- better overlap of communication with computation

# The The PGAS Strategy

Memory aggregation to provide a global logical memory address space for data processing.

## Provides..

- ease of general purpose processing using semantics similar to that of shared memory systems.
- better overlap of communication with computation

Leverage RDMA over PCIe to enhance memory-to-memory data copy.

- *Provide low latencies*
- *High throughput*

# The PGAS Strategy

A complete RAM-based storage could yield a very high-throughput ($100 - 1000x$) at very low-latency ($100 - 1000x$).

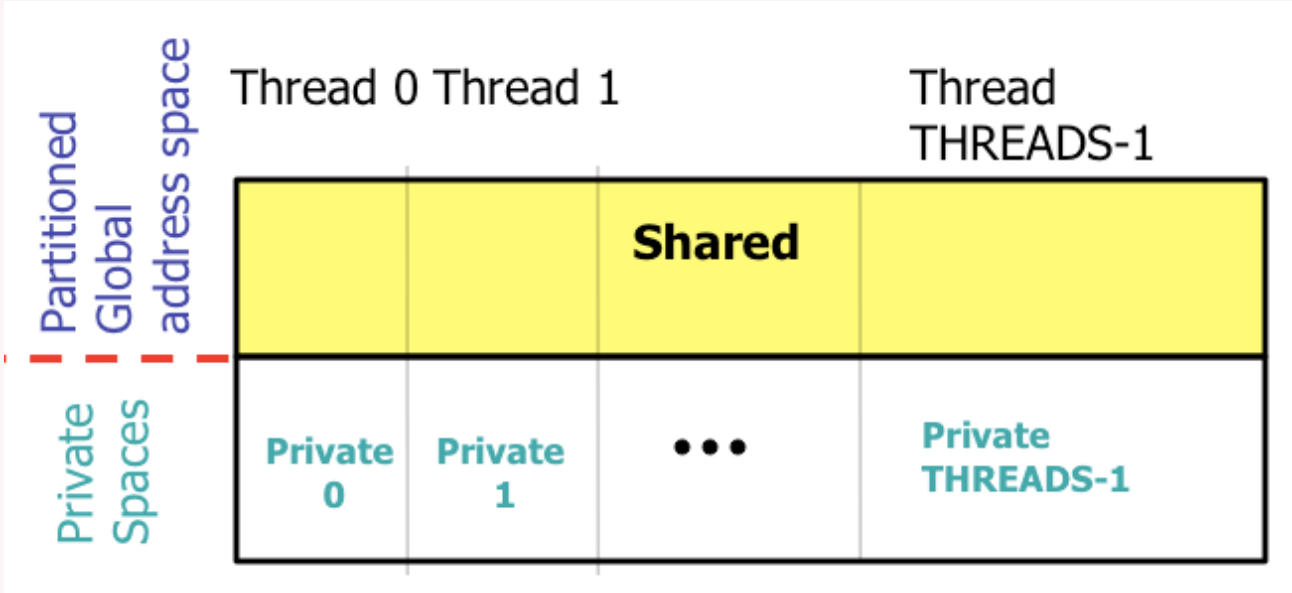


Figure: Schematic: PGAS layout with PU

# Experimental Setup

Figure: Wits Tegra K1

- 4 nodes of the Wits Tegra K1 cluster.
- 2GB of memory each and 1Gbp Ethernet interconnect between nodes.
- NASA Advanced Supercomputing (NAS) Parallel Benchmarks is used.

# Experimental Setup

## Selected Kernel

Fast Fourier Transform: This benchmark solves a 3D partial differential equation using an FFT-based spectral method, also requiring long range communication. FT performs three one-dimensional (1-D) FFT's, one for each dimension.

# Preliminary Results

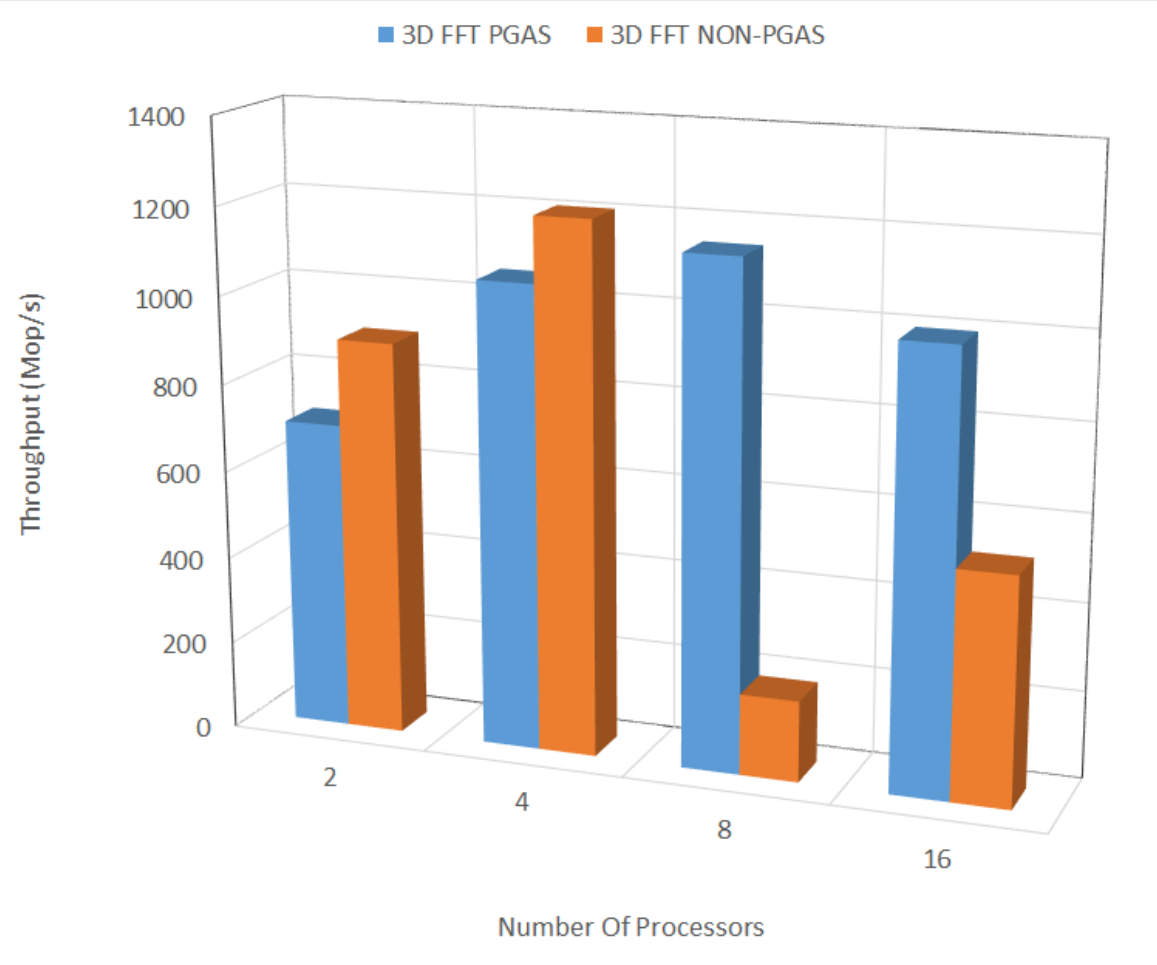


Figure: Throughput : $64 \times 64 \times 64$ 3D-grid

# Preliminary Results

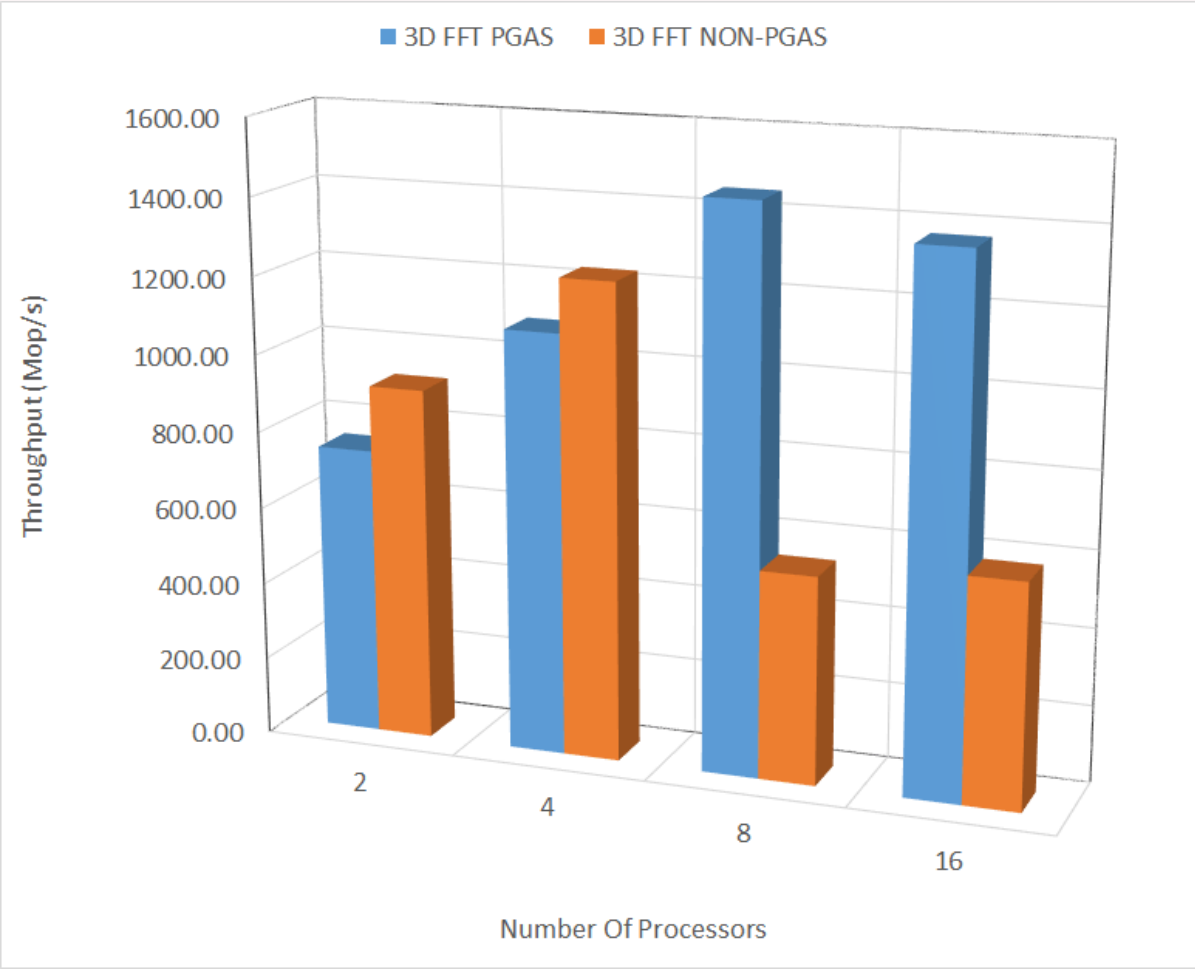


Figure: Throughput : $256 \times 256 \times 128$ 3D-grid

# Preliminary Results

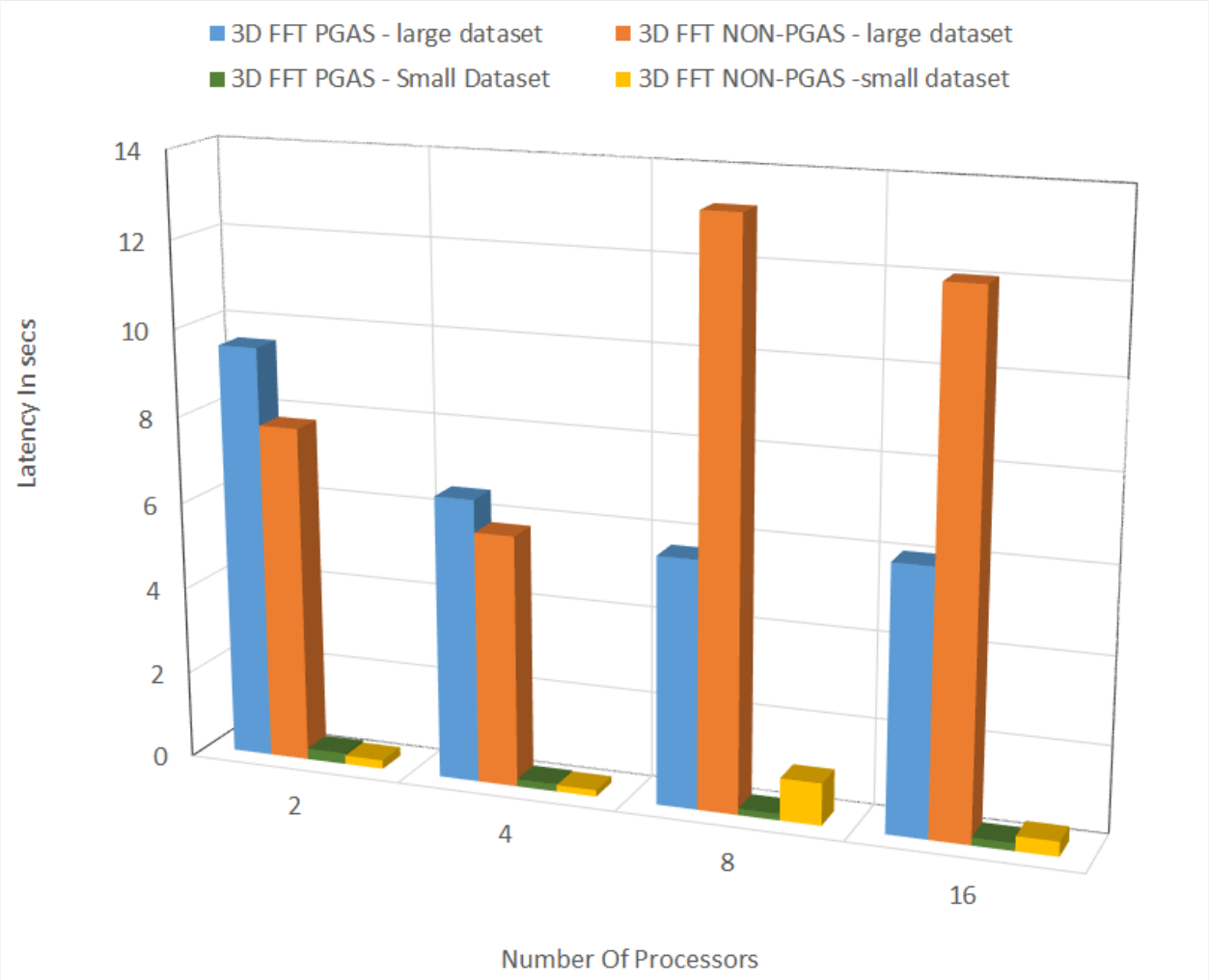


Figure: Average Latency for varying workloads

# Preliminary Results

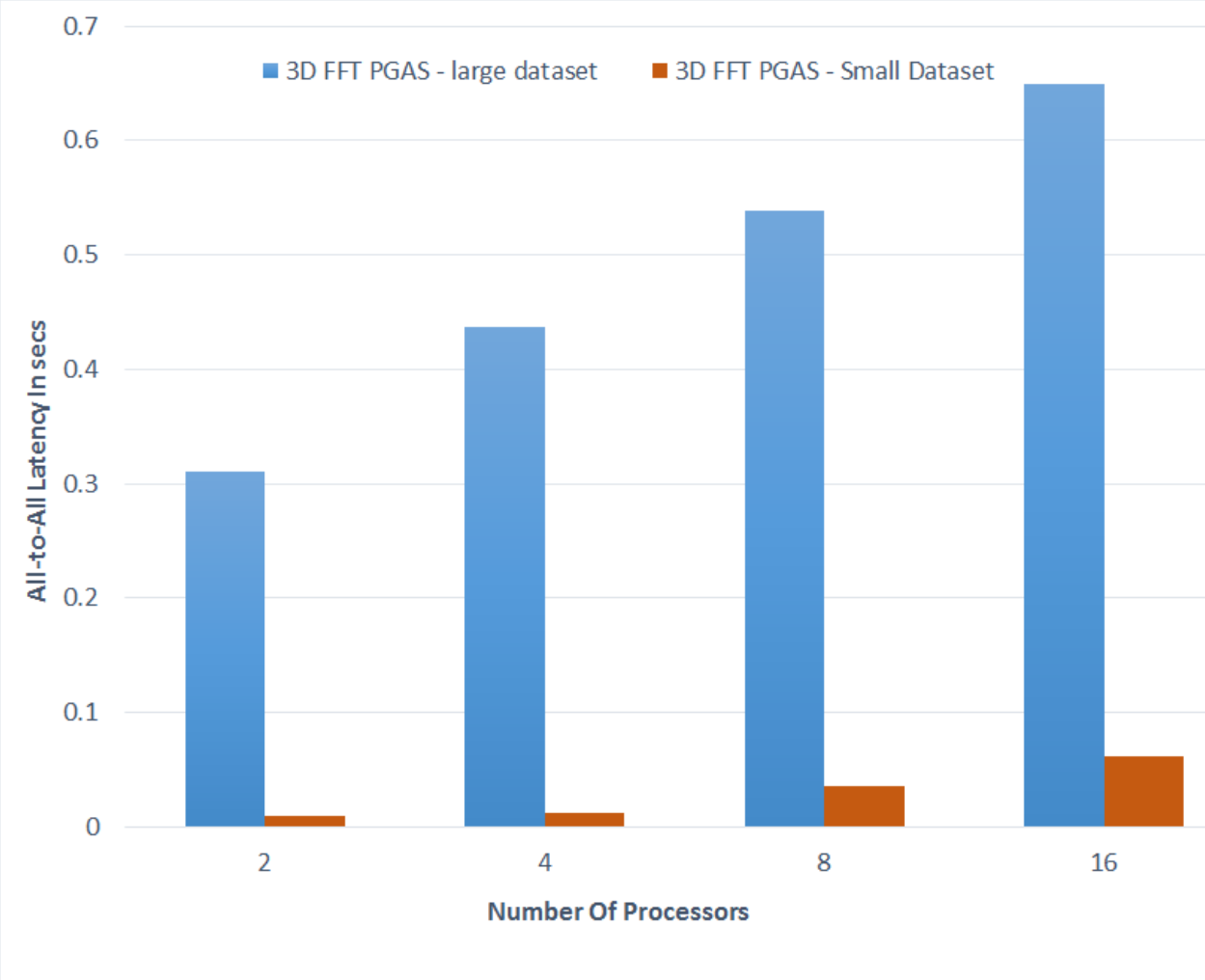


Figure: Average Latency due to All-to-All collective operation

# Conclusions

- Preliminary investigation demonstrate horizontal scalability of the architecture and thus present a low cost alternative to large data science projects such as ATLAS data processing.

- Low cost and power efficient ARM with PGAS could therefore be used as general purpose co-processors to the sROD for high throughput data processing.

# Conclusions

- Preliminary investigation demonstrate horizontal scalability of the architecture and thus present a low cost alternative to large data science projects such as ATLAS data processing.
- Low cost and power efficient ARM with PGAS could therefore be used as general purpose co-processors to the sROD for high throughput data processing.

Future investigations with Remote Direct Memory Access (RDMA) over PCIe is expected to provide much lower latency and higher throughput.

# Thank you...