GENESIS: A COMPOSABLE HARDWARE ACCELERATION FRAMEWORK FOR GENOME ANALYSIS

LISA WU WILLS, DUKE UNIVERSITY
My Research Goal #1: **Design and deploy highly efficient domain-specific hardware accelerators to advance state-of-the-art emerging applications.**

My Research Goal #2: **Democratize end-to-end accelerated system development and deployment while leveraging hardware acceleration to advance emerging application domains in the interdisciplinary area of computer architecture and healthcare.**
FPGA ACCELERATED INDEL REALIGNMENT IN THE CLOUD [HPCA 2019]

LISA WU, DAVID BRUNS-SMITH, FRANK NOTHAFT, QIJING HUANG, SAGAR KARANDIKAR, HOWARD MAO, BRENDA SWEENEY, KRSTÉ ASANOVIC, DAVID PATTERTON, AND ANTHONY JOSEPH

GENESIS: A HARDWARE ACCELERATION FRAMEWORK FOR GENOMIC DATA ANALYSIS [ISCA 2020, IEEE MICRO TOP PICK 2021]

TAE JUN HAM, DAVID BRUNS-SMITH, BRENDA SWEENEY, YEJIN LEE, SEONG HOON SEO, U GYEONG SONG, YOUNG H OH, KRSTÉ ASANOVIC, JAE W LEE, AND LISA WU WILLS
GOAL OF GENOMIC ANALYSIS:
Identify the nucleotide differences (or variants) between an individual genome and the reference genome at a given position, with acceptable accuracy.
ALGORITHM REFEMENT IS THE SLOWEST PIPELINE

Read Mapping: 14%
Alignment Refinement: 58%
Variant Calling: 28%

Genomic Analysis Execution Time Breakdown:

- Read Mapping: ~17 Hours
- Alignment Refinement: ~72 Hours
- Variant Calling: ~36 Hours

Genomic Analysis Execution Time Breakdown
WHAT DOES DATABASE ANALYTICS AND GENOMIC ANALYSIS HAVE IN COMMON?

Generic Data Manipulations (i.e. aggregation, sorting)!

- Domain-Specific Language: SQL+
- Genomic Hardware Library
- Accelerator Composer

ADEPT-EOP-Wills
Accelerator Composer Design Templates

Accelerator System
(contains $n$ Abstract Cores of Type $t$ and $m$ Abstract Cores of Type $s$)

Abstract Collection
(contains $n$ Cores of Type $t$)

Abstract Core

Read Channels

Write Channels

RoCC cmds

RoCC resps

n:1 channel arbiter

r:1 channel arbiter

w:1 channel arbiter

Abstract Collection
(contains $m$ Cores of Type $s$)

Abstract Core

Read Channels

Write Channels

RoCC cmds

RoCC resps

r+w:1 channel arbiter

Abstract Core

Read Channels

Write Channels

n+m:1 channel arbiter

To/From Memory
OUR ACCELERATED IR SYSTEM PERFORMS 81X BETTER THAN SOFTWARE RUNNING 8 THREADS

42 hours —> roughly 30 minutes
IMPACTS/SIGNIFICANCE

ACCELERATING COMMON PRIMITIVES ACROSS DOMAINS ALLOWS THE SHARING, REUSING, AND COMPOSITION OF ACCELERATED SYSTEMS ACROSS DOMAINS, LOWERING DEVELOPMENT EFFORT.
LEVERAGE AN ALREADY-STANDARDIZED LANGUAGE AS THE DSL AND CONSTRUCT PRIMITIVE OPERATORS THAT DIRECTLY MAP SOFTWARE PRIMITIVES TO HARDWARE BLOCKS PRODUCES EFFICIENT ACCELERATED SYSTEMS.
IMPACTS/SIGNIFICANCE

This development methodology can be adopted for various domains beyond database and genomic analytics.