What Use is Hardware Without Software?

Deborah Guillon, MASc
Machine Learning Lead, YetiWare Inc.

YetiWare INC.

Communities in Isolation

Software without hardware is math	Hardware without software is a heater

?

Need to bring together to accelerate **heterogeneous** computing … missing piece

YetiWare
INC.

2
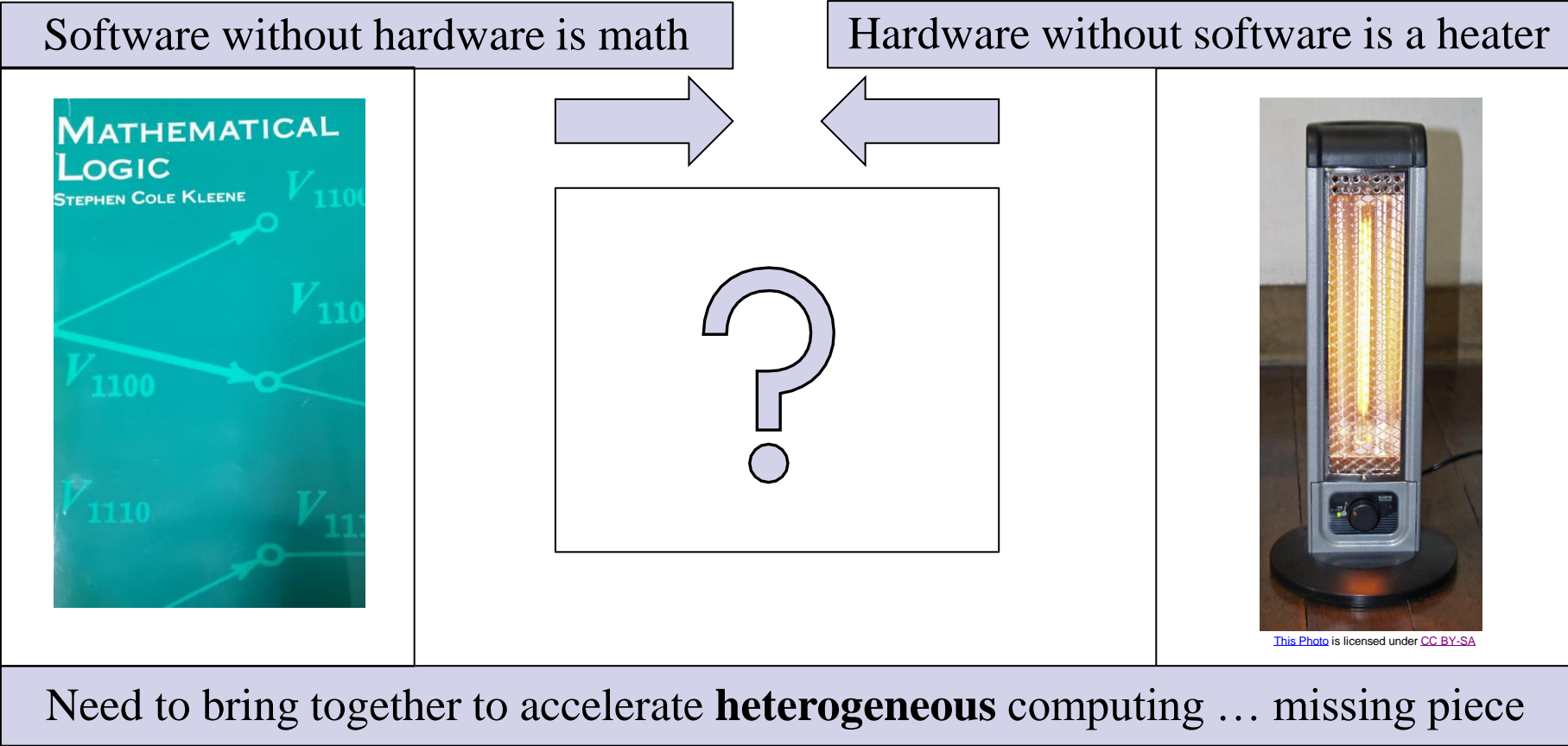
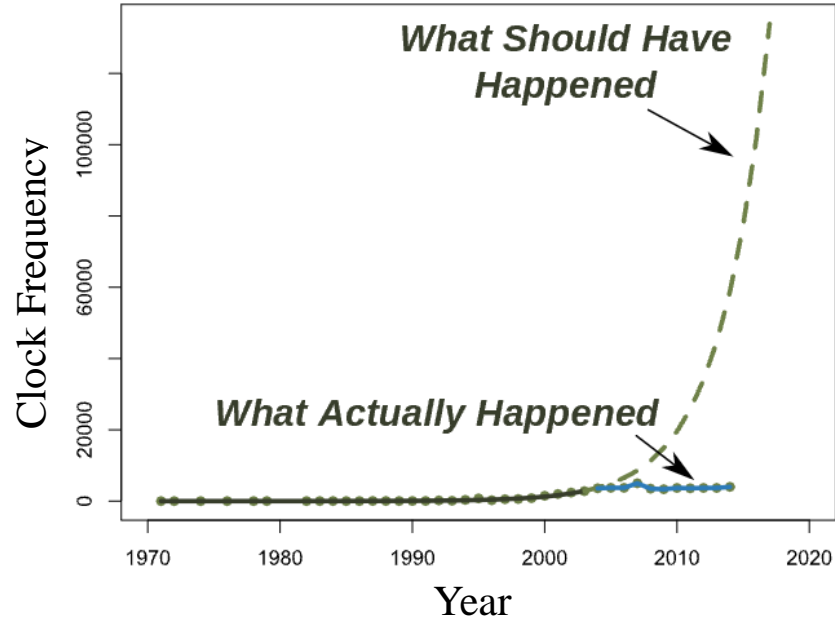# How Did We Get Here: The Free Lunch

Software

Golden era
- Applications
- New markets
- Grow customers

Processor Clock Frequency Over Time

What Should Have Happened

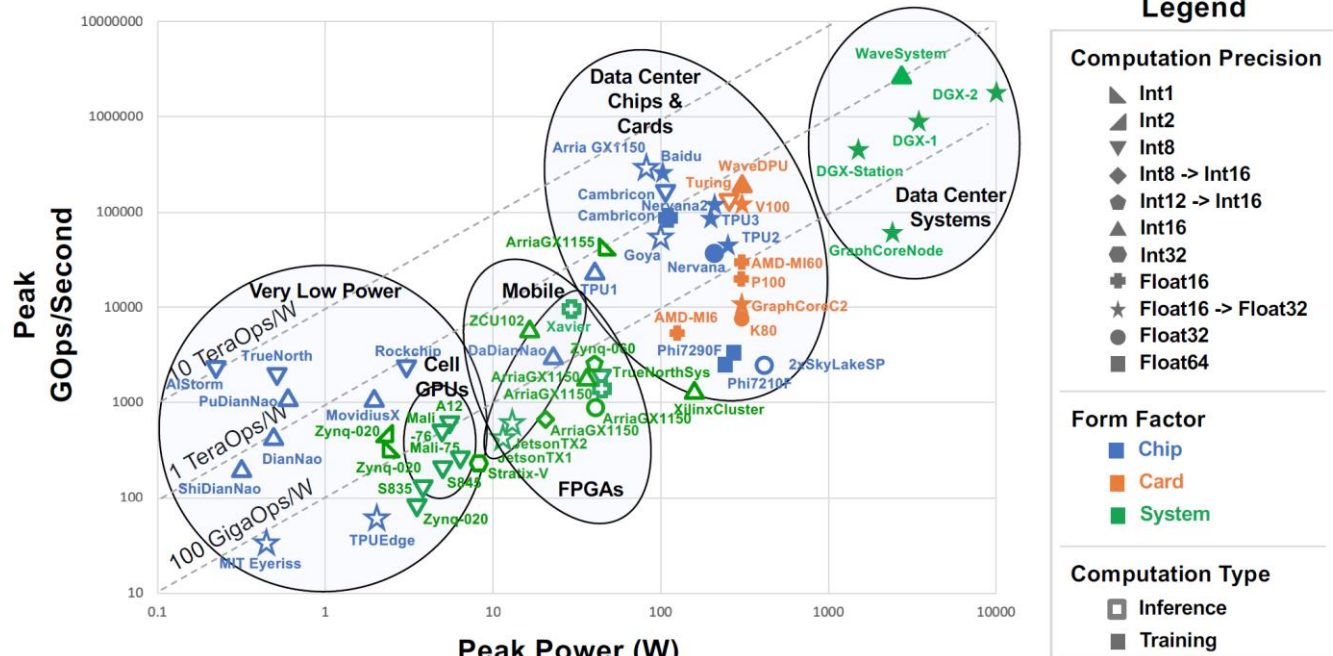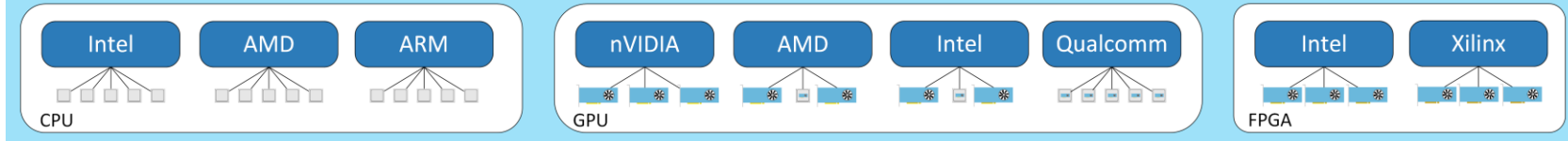What Actually Happened
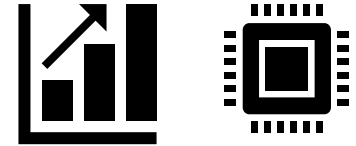
Hardware

Processor performance increase

End of general-purpose compute

What Happened When the Free Lunch Ended?

Hardware: More or application-specific processors

Legend — Computation Precision: Int1, Int2, Int8, Int8 -> Int16, Int12 -> Int16, Int16, Int32, Float16, Float16 -> Float32, Float32, Float64. Form Factor: Chip, Card, System. Computation Type: Inference, Training.

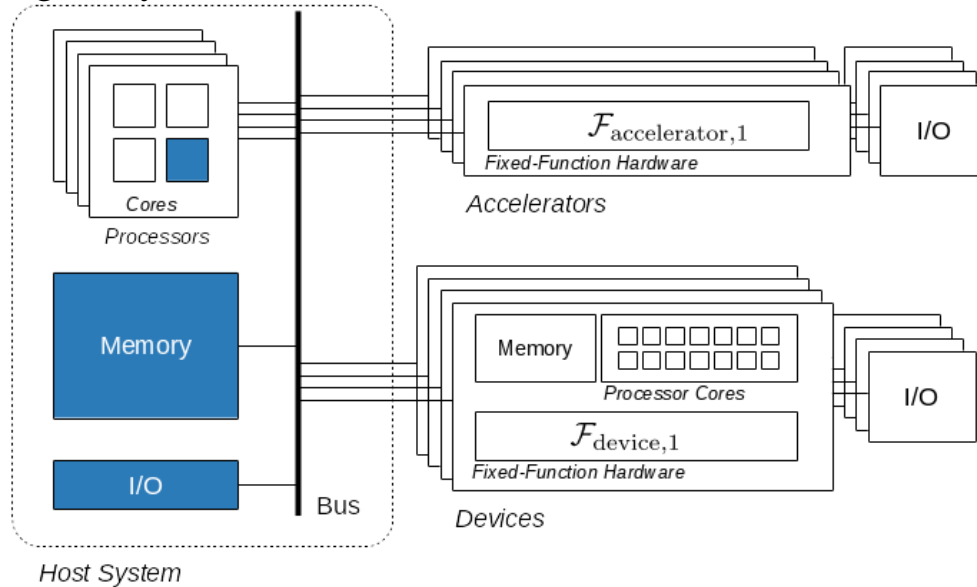From: Reuther, Michaleas, Jones, et al. Survey and Benchmarking of Machine Learning Accelerators. 2019 Sept. Link

YetiWare INC.

Software Didn't Change

Software is broken

No one is using your hardware efficiently

Stuck in a 1950s model
No parallelism, no heterogeneity

Use ML?

We need to think differently to solve this problem

YetiWare INC.

Heterogeneous Compute Space

Deployment Environment

PC   Mobile   Cluster   Cloud   IoT

Specifications

OpenCL   CUDA   Vulkan   HSA   Metal

Processors

Intel CPU   AMD CPU   ARM CPU   Qualcomm GPU   AMD GPU   Intel GPU   nVIDIA GPU   Intel FPGA   Xilinx FPGA

Complex compute space requires deployment, specification, and processor specific code

YetiWare INC.

6

Compute Space Chosen

Reduce to homogeneous problem using NVIDIA GPUs with 1 version of CUDA

YetiWare INC.

Frameworks Layered On Top

ML Frameworks

TensorFlow PyTorch

Deployment Environment

PC Mobile Cluster Cloud IoT

Specifications

OpenCL CUDA Vulkan HSA Metal

Processors

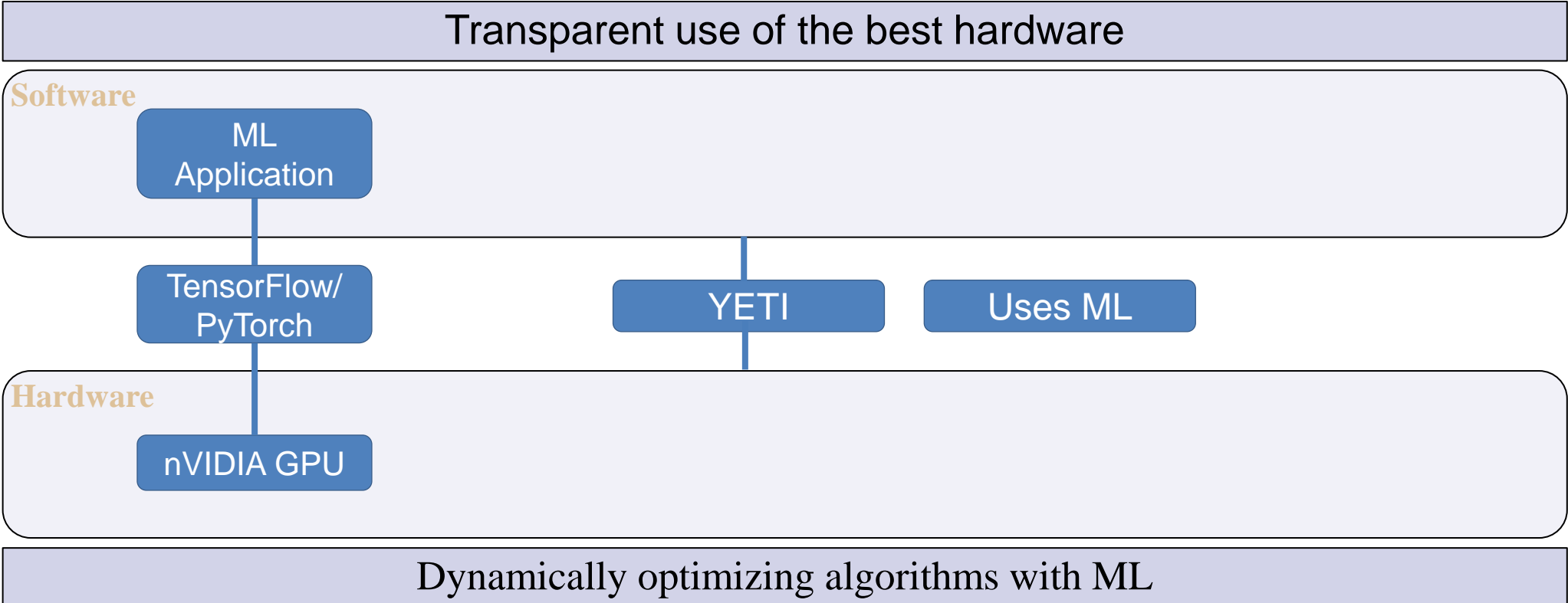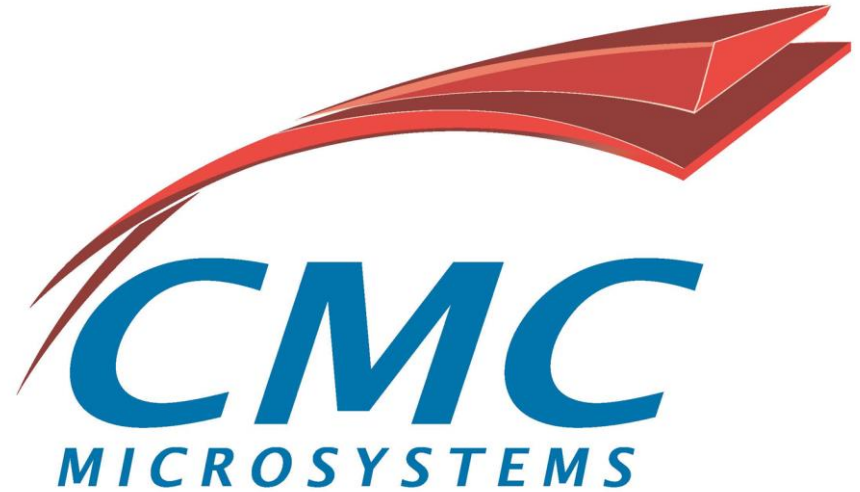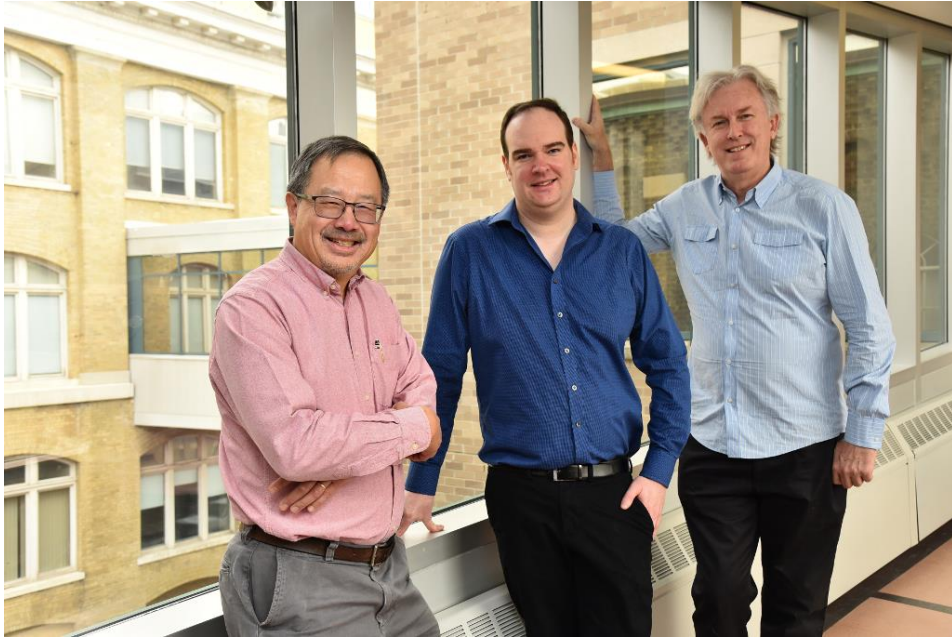Intel CPU AMD CPU ARM CPU Qualcomm GPU AMD GPU Intel GPU nVIDIA GPU Intel FPGA Xilinx FPGA

Shift problem to HW, which must integrate with ML frameworks

YetiWare INC.

Our Vision

Transparent use of the best hardware

Software

ML Application

TensorFlow/ PyTorch

YETI

Uses ML

Hardware

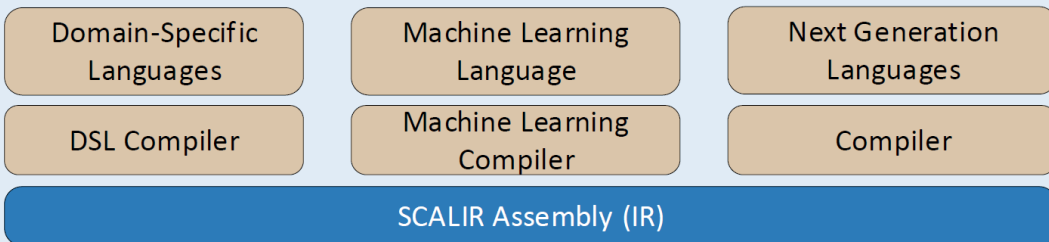nVIDIA GPU

Dynamically optimizing algorithms with ML

YetiWare INC.

YetiWare and CMC Prototyped First Component
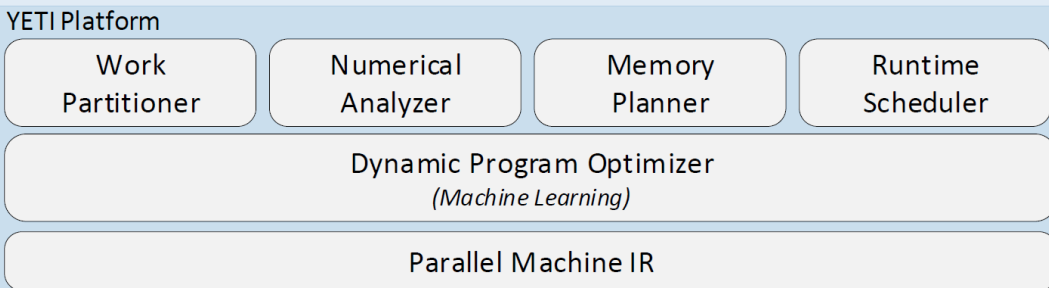

https://account.cmc.ca/WhatWeOffer/Products/CMC-00200-05319.aspx

YetiWare

Our Vision: The YETI Platform

Productivity Layer 📈

Write code only once.

Domain-Specific Languages	Machine Learning Language	Next Generation Languages
DSL Compiler	Machine Learning Compiler	Compiler

SCALIR Assembly (IR)

Performance Layer ⛰️

Continuous performance improvement through updated runtimes.

YETI Platform

Work Partitioner	Numerical Analyzer	Memory Planner	Runtime Scheduler

Dynamic Program Optimizer
(Machine Learning)

Parallel Machine IR
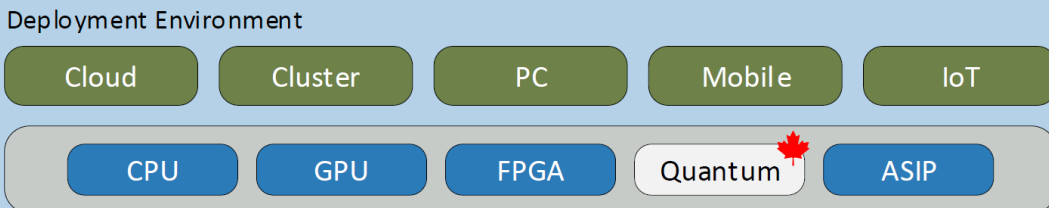
Portability Layer

Any processor. Any vendor. Any deployment environment.

Deployment Environment

Cloud	Cluster	PC	Mobile	IoT

CPU	GPU	FPGA	Quantum	ASIP

Heterogeneous Processor Targets

YetiWare INC.

Communities in Isolation

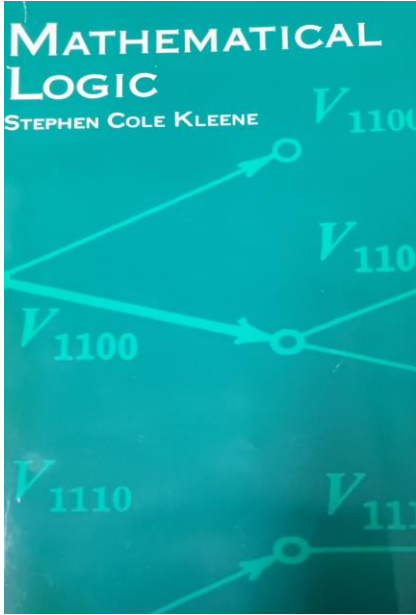Software without hardware is math

Hardware without software is a heater

Need to bring together to accelerate **heterogeneous** computing … missing piece

Communities Reconnected

YETI will connect SW and HW in a heterogeneous environment

Productivity

Performance

Portability
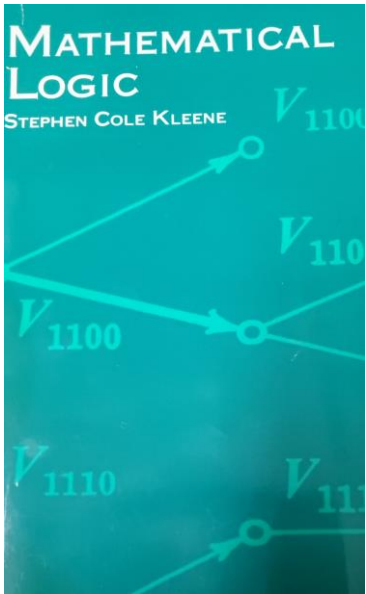
This Photo is licensed under CC BY-SA

Bringing about a fundamental paradigm shift