A Case for Richer Cross-layer Abstractions: Bridging the Semantic Gap with Expressive Memory

Nandita Vijaykumar

Abhilasha Jain, Diptesh Majumdar, Kevin Hsieh, Gennady Pekhimenko

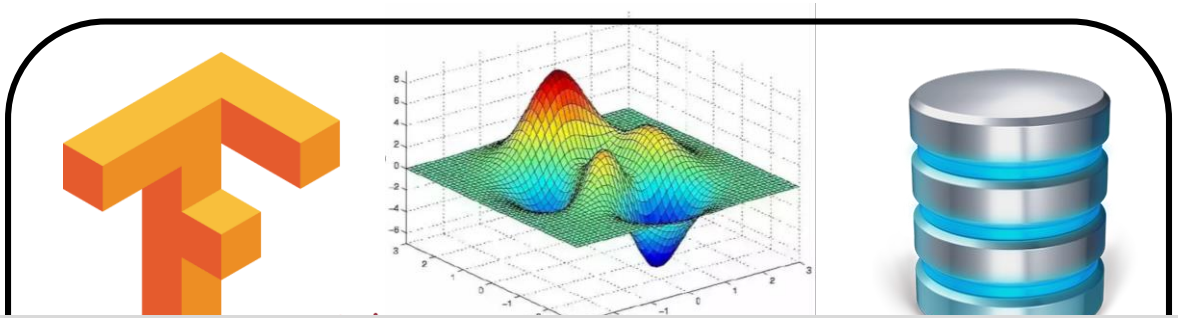
Eiman Ebrahimi, Nastaran Hajinazar, Phillip B. Gibbons, Onur Mutlu

Carnegie Mellon University

UNIVERSITY OF TORONTO

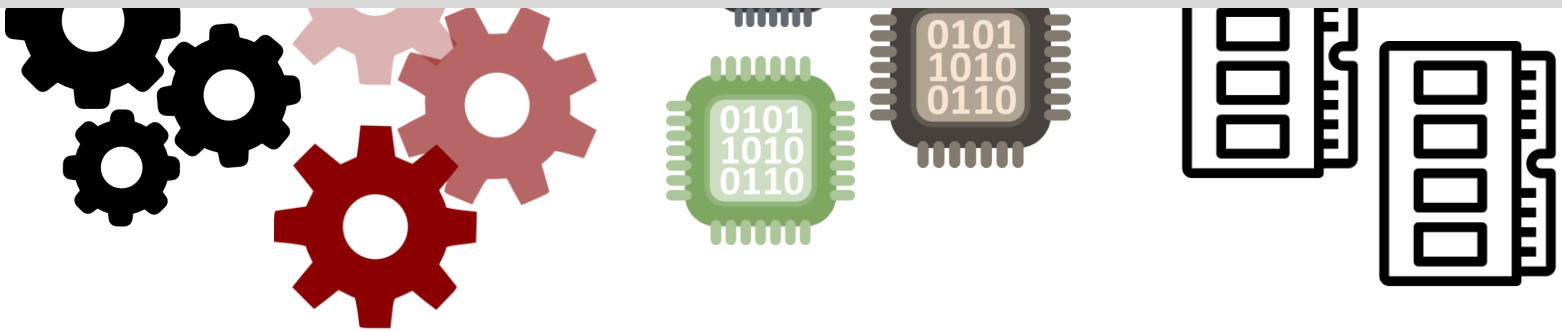
SFU

ETH zürich

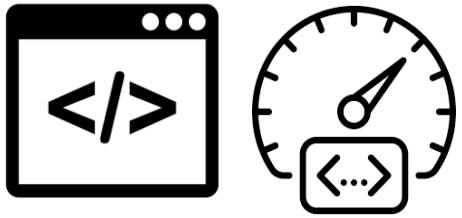


Today's cross-layer abstractions are designed to primarily convey functionality,

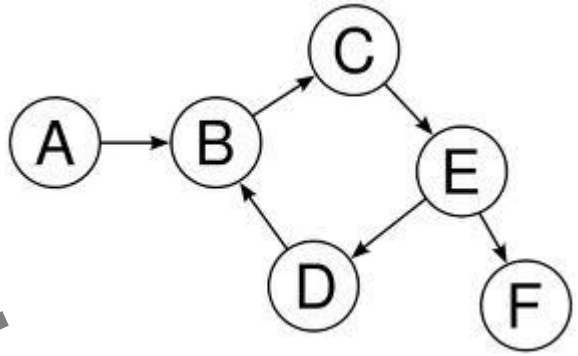
not to aid performance optimization



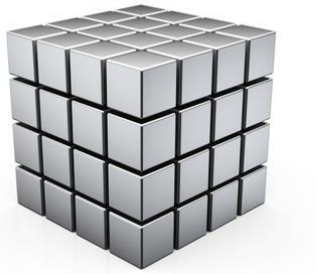
The narrow abstractions lose key information



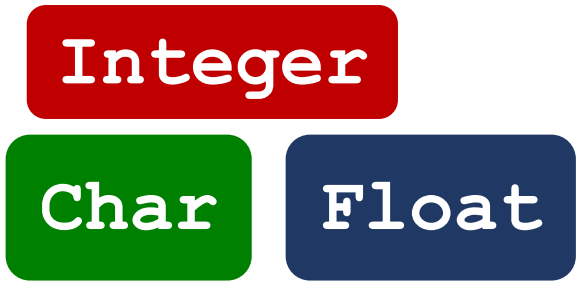
Code Optimizations



Access Patterns



Data Structures



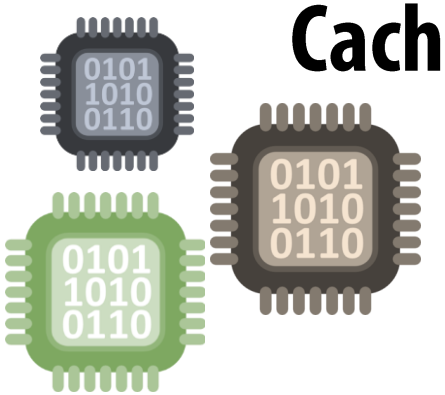
Data Type

Instructions 100011111…

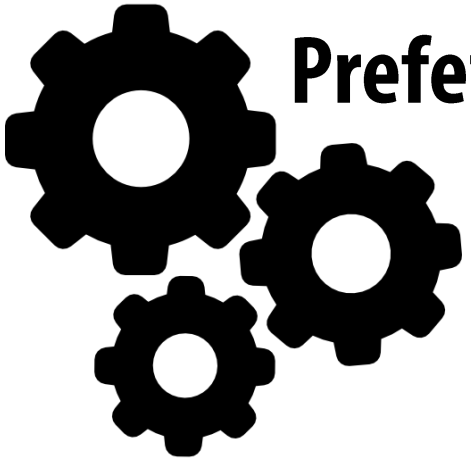
Memory Addresses 101010011…

Consequence 1: The Hardware Approach

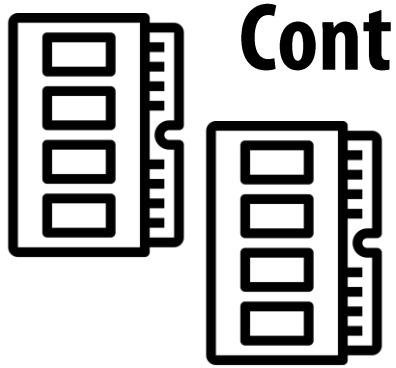
We design hardware to infer and predict program behavior to optimize for performance



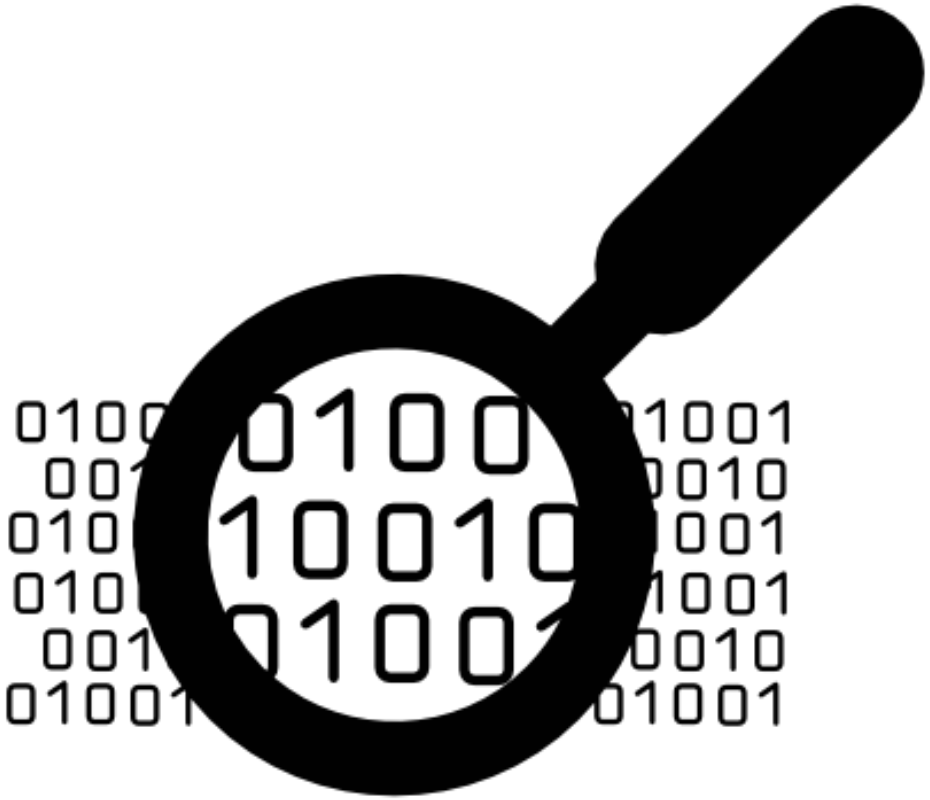
Caches



Prefetcher



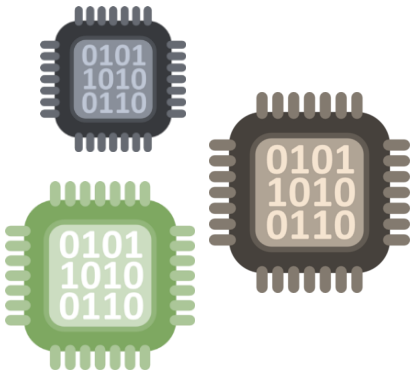
Memory Controller



✘ Performance on the table

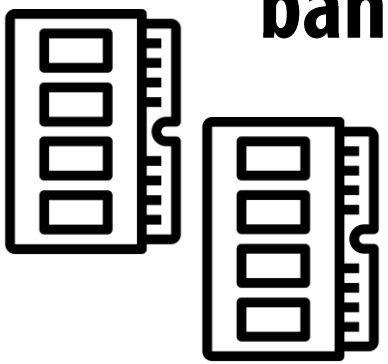
Consequence 2: The Software Approach

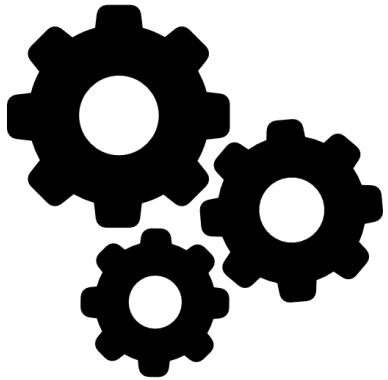
Software is tuned to the specifics of hardware architecture when optimizing for performance



Cache space available?

How many DRAM banks?



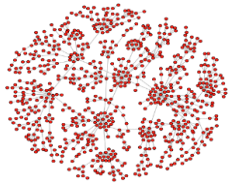


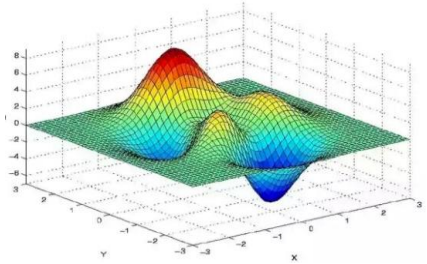
Hardware optimizations?

- ✘ Programmability
- ✘ Portability

With growing HW/SW sophistication, traditional interfaces limit optimization effectiveness

Time for a richer interface between hardware and software?

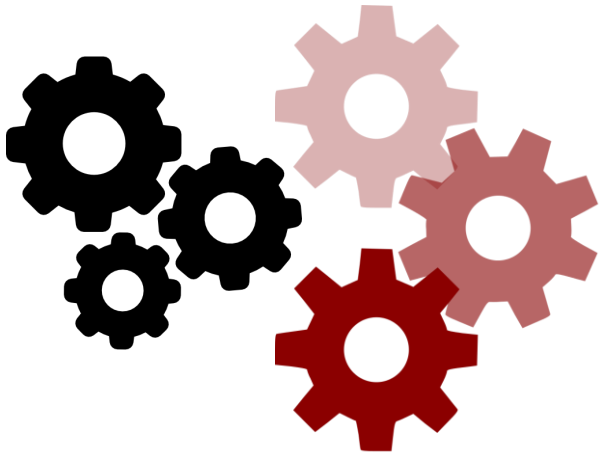
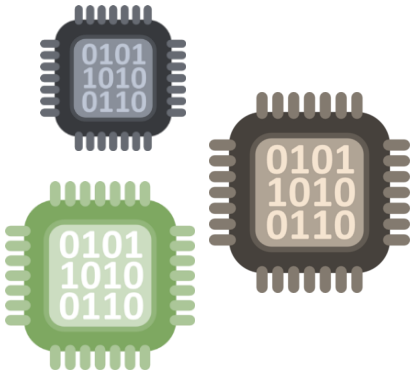


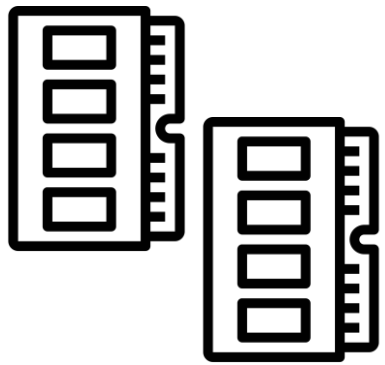


ISA
Virtual Memory

Higher-level
Program
Semantics

Expressive
Memory
"XMem"



A fresh approach to traditional optimizations

Cache Management

Data Placement in DRAM

Data Compression

Approximation

DRAM Cache Management

NVM Management

NUCA/NUMA Optimizations

….

✓Programmability

✓Portability

✓Resource Efficiency

A Case for Richer Cross-layer Abstractions: Bridging the Semantic Gap with Expressive Memory

Nandita Vijaykumar

Abhilasha Jain, Diptesh Majumdar, Kevin Hsieh, Gennady Pekhimenko

Eiman Ebrahimi, Nastaran Hajinazar, Phillip B. Gibbons, Onur Mutlu

Carnegie Mellon University

UNIVERSITY OF TORONTO

SFU

ETH zürich