

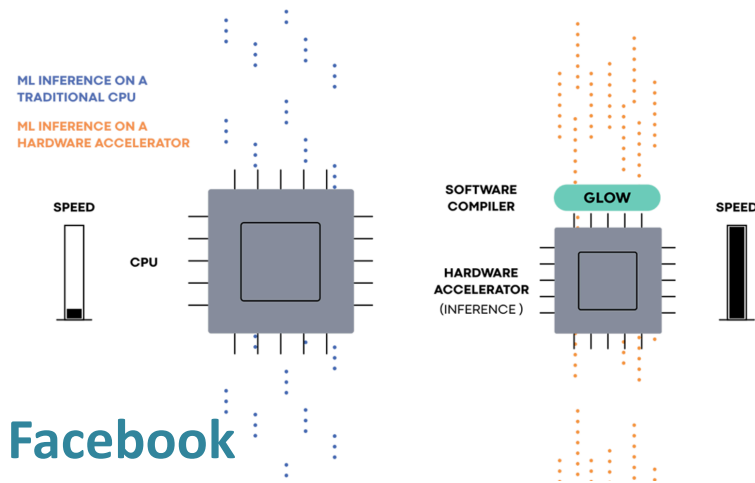
**How do we leverage deep learning  
investments to pay for HPC  
compiler technology?**

Mary Hall, University of Utah

# Deep Learning Compiler Technology

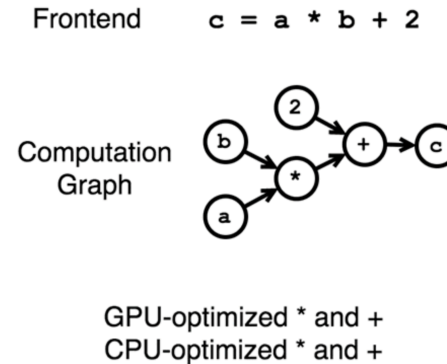
POSTED ON SEP 13, 2018 TO AI RESEARCH, ML APPLICATIONS

Glow: A community-driven approach to AI infrastructure



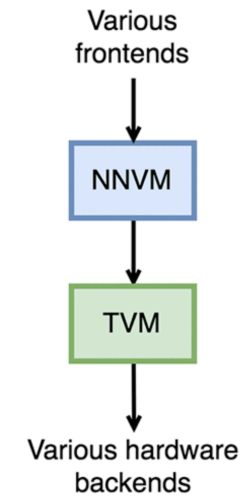
Facebook

A typical framework



Amazon

NNVM Compiler



Google

MLIR: Multi-Level Intermediate Representation  
Compiler Infrastructure

2019 European LLVM Developers Meeting

Challenges and opportunities:

- Domain-specific
- Many frontends
- Many target architectures
- Abundant parallelism and data reuse
- Must scale to large problems

<http://code.fb.com/ml-applications/glow-a-community-driven-approach-to-ai-infrastructure/>

<http://aws.amazon.com/blogs/machine-learning/introducing-nnvm-compiler-a-new-open-end-to-end-compiler-for-ai-frameworks/>

# Convolutional Neural Network Forward Layer Code (in C)

```
for (n=0; n<N; n++) { // minibatch size
  for (k=0; k<K; k++) { // output feature map
    for (c=0; c<C; c++) { // input feature map
      for (p=0; p<P; p++) { // output height
        ij = p * u; // input height
        for (q =0; q<Q; q++) { // output width
          ii = q * v; // input width
          for (r=0; r<R; r++) { // filter height
            for (s =0; s< S; s++) { // filter width
              output_seq[n][k][p][q] +=
                input [n][c][ij+r][ii+s] * weight[k][c][r][s];
            }
          }
        }
      }
    }
  }
}
```

# Current State of HPC Compilers

## Proprietary

- Robust
- High-quality implementations for supported architectures
- Support HPC community
- Code not performance portable across systems
- Often conservative

## Open Source

- Research compilers
  - State-of-the-art
  - Experimental, untrusted
  - Difficult to track language changes
  - Gaps, such as Fortran frontend
- LLVM and gcc
  - Gaps in HPC support
  - Conservative

# Current State of HPC Compilers, cont.

## Challenges:

- HPC market not large enough to drive significant change to open source or even proprietary compilers
- Meanwhile, research systems not sufficiently robust for production codes

## Impact:

- Productivity improvements for HPC not being exploited
- Heterogeneity will make this a bigger concern

# Goal: HPC Support in Open Source Compilers

## Short-term

(ECP time frame)

### Extend LLVM

- Parallel IR
- Loop transformations
- OpenMP/OpenACC
- Autotuning
- Fortran frontend

## Longer term

(But need to start now)

### Work with Google on MLIR?

- Higher level of abstraction
- Composability of different views (parallelism?)
- Built-in polyhedral transformations and code generation
- Multiple backends via LLVM
- Missing frontends