HPC with the Groq MLIR compiler

Sanjif Shanmugavelu, Groq

HPC with the Groq MLIR compiler

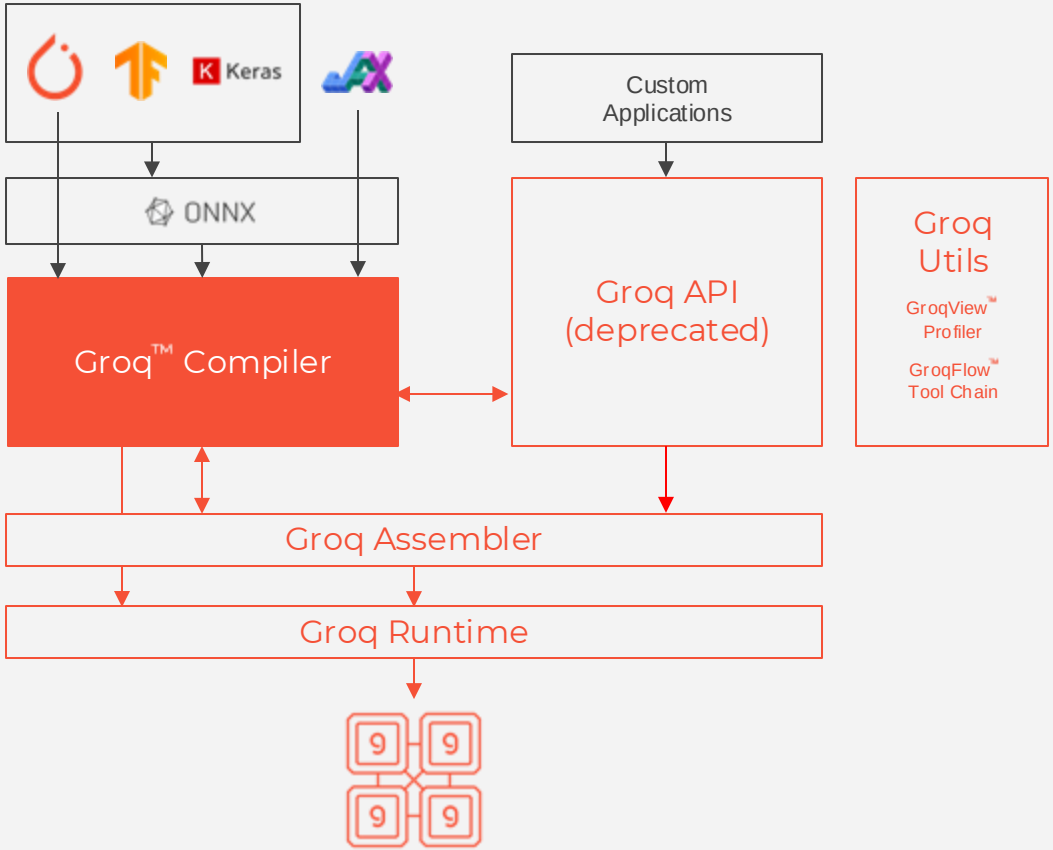
Sanjif Shanmugavelu, Machine Learning Engineer, Groq
sshanmugavelu@groq.com

A birds eye view

GroqWare

Groq's Software Stack

- Modular + Monolithic



GroqWare

Groq's Software Stack

- Modular + Monolithic
- Compiler at the core

Model Import

Layout & Vectorization

Multichip Partition

↗ Tensor Parallelism
↘ Pipeline Parallelism

Mapping to ISA

MLIR

Live State Plan

Instruction Schedule

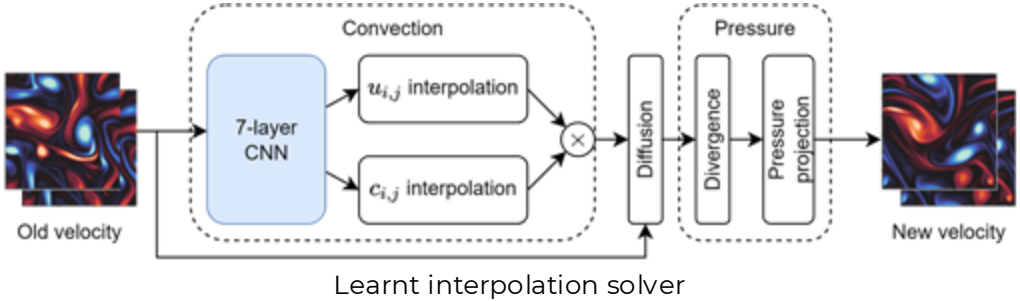
Assemble

Converged HPC and AI: CFD Super-Resolution

Conventional and AI based solvers for structured grid methods

Solver summary:

- 2D structured grid
- Incompressible airflow
- Explicit time integration
- Direct numerical simulation (DNS) with ML augmentation



Learnt interpolation solver

4 different solvers:

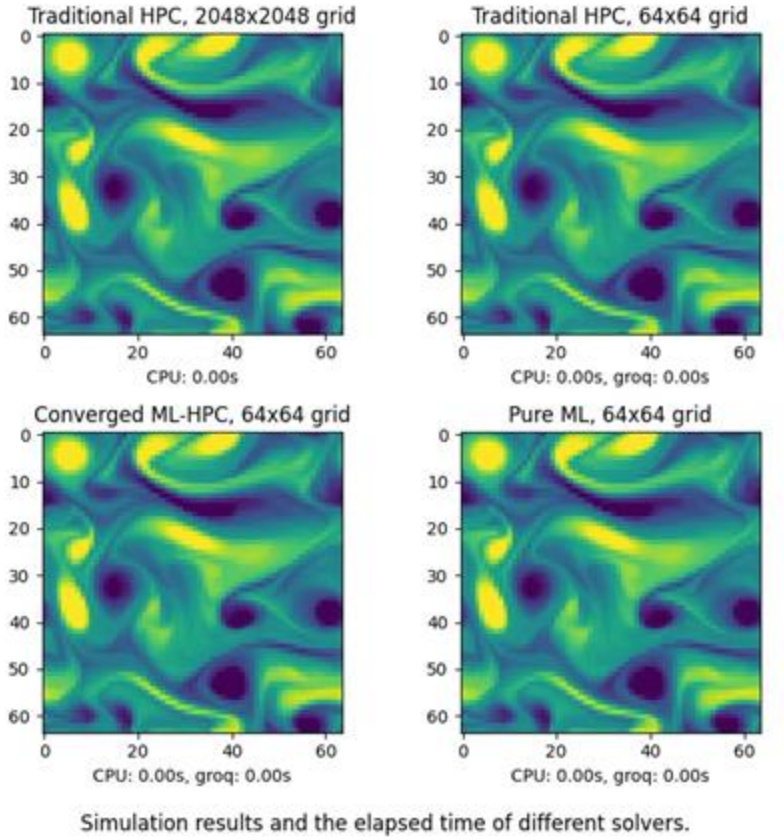
- Pure DNS: standard finite volume solver based on pressure projection algorithm
- Learned interpolation: DNS with CNN to predict the cell boundary flux
- Learned correction: DNS with CNN to supersample the simulation result
- Pure ML: encoder-process-decoder (process stage can be LSTM)

Hybrid CFD with AI Augmentation

Augment traditional HPC algorithm with AI

Four Approaches:

- Traditional DNS: standard solver based on pressure projection (high and low res)
- Learned correction: Small grid DNS with CNN-based correction
- Pure ML: LSTM-based encoder-process-decoder
- Converged ML-HPC combines high throughput and high accuracy



Simulation results and the elapsed time of different solvers.

POTENTIAL APPLICATIONS

Aerospace **Automotive** **Industrial** **Energy** **Medical**