

OCCUPIER FLOW

Inputs

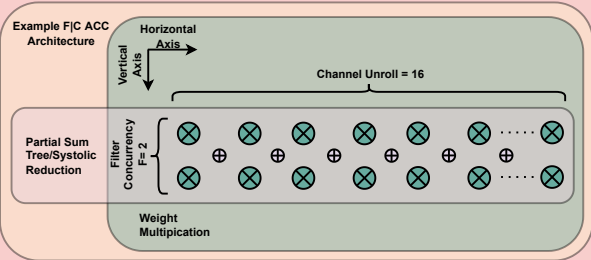
Arch Config

PE Budget Range
Filter Unroll Range
Channel Unroll Range
Kernel Sizes supported Directly
GEMM Lowering/Lifting Mode

Model Library



Select Arch Config



Select Layer

Get Layer Utilization Statistics



Reshape Weight Tensor to Matrix

Tile Based On Architecture then pad tiles

Weigh low utilization by cycles spent evaluating OFMAP

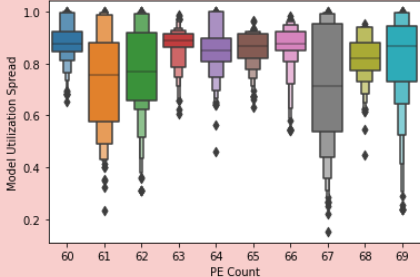
Get avg utilization of model

Log config's average model performance

Get Next Config

Report Top-1 avg model utilization for each Arch config at each PE budget Value

Utilization optimal configuration for HERO Architecture under different PE Budgets



Get Next Model

Config search Complete