# SecureLoop:
# Design Space Exploration of Secure DNN Accelerators

**Kyungmi Lee**, Mengjia Yan, Joel S. Emer*, Anantha P. Chandrakasan

MIT, *MIT/NVIDIA

MICRO 2023

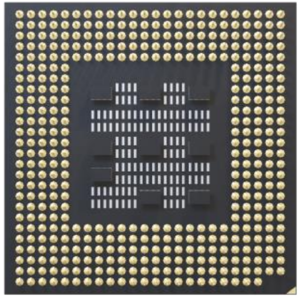
# ML Needs Both Security and Performance

Autonomous Driving

Biometric Authentication

Healthcare

**Applications require security**

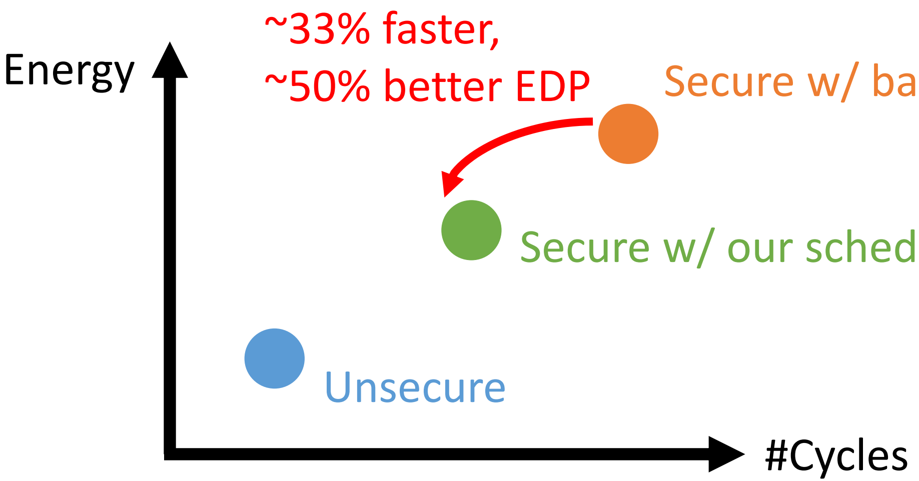


**Accelerator design requires performance, area, energy**

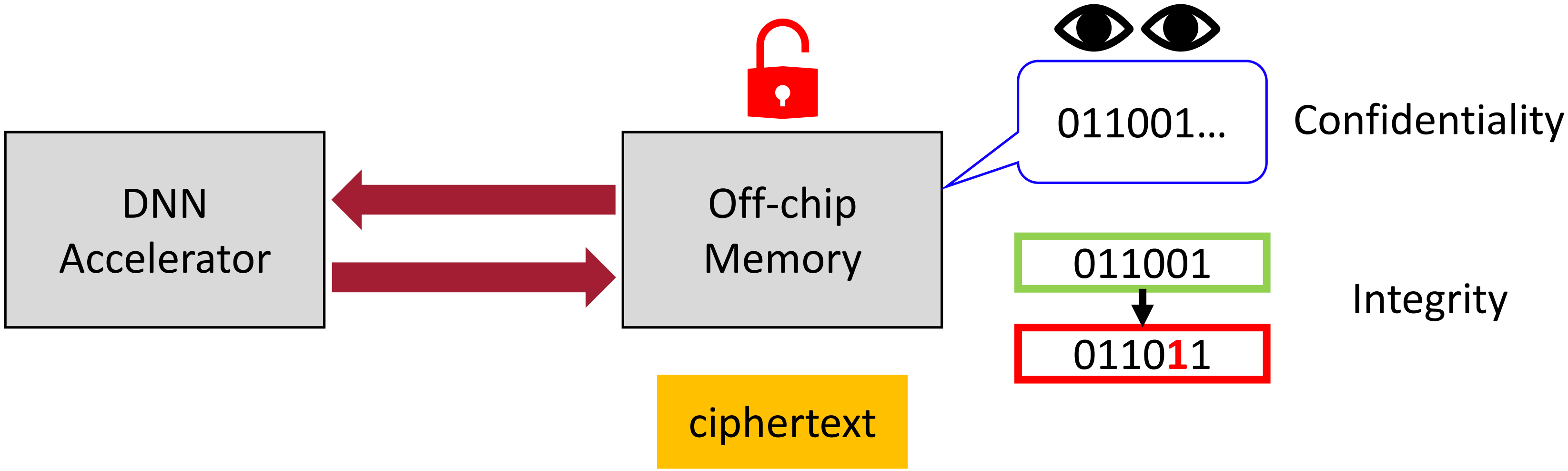
Design space exploration for “secure” DNN accelerators

# SecureLoop

A tool for **design space exploration of secure DNN accelerators** equipped with cryptographic engines

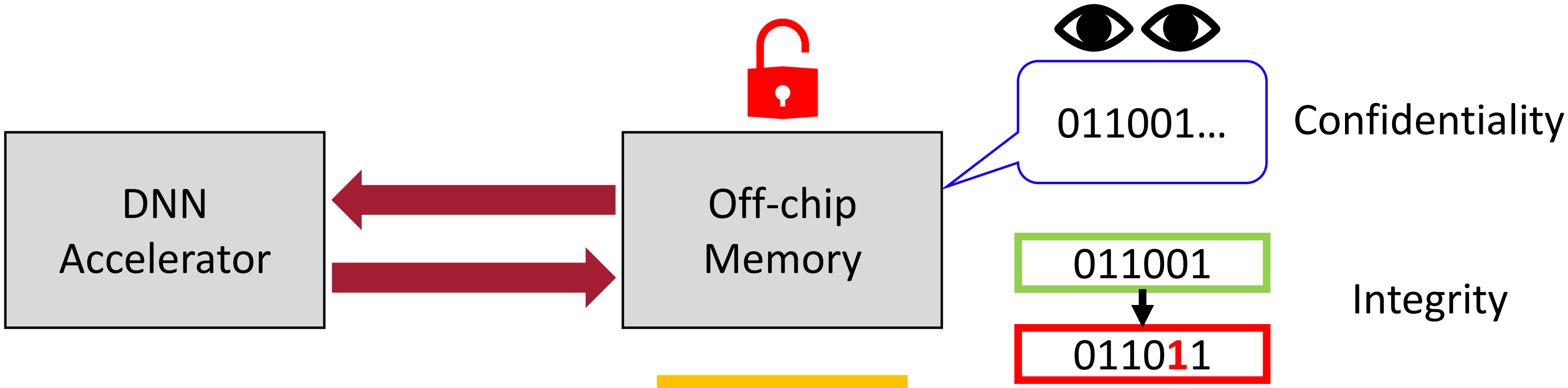


# Background: TEE and DNN Accelerators



Support confidentiality with cryptographic encryption

# Background: TEE and DNN Accelerators



DNN Accelerator

Off-chip Memory

011001…

Confidentiality

011001

011011

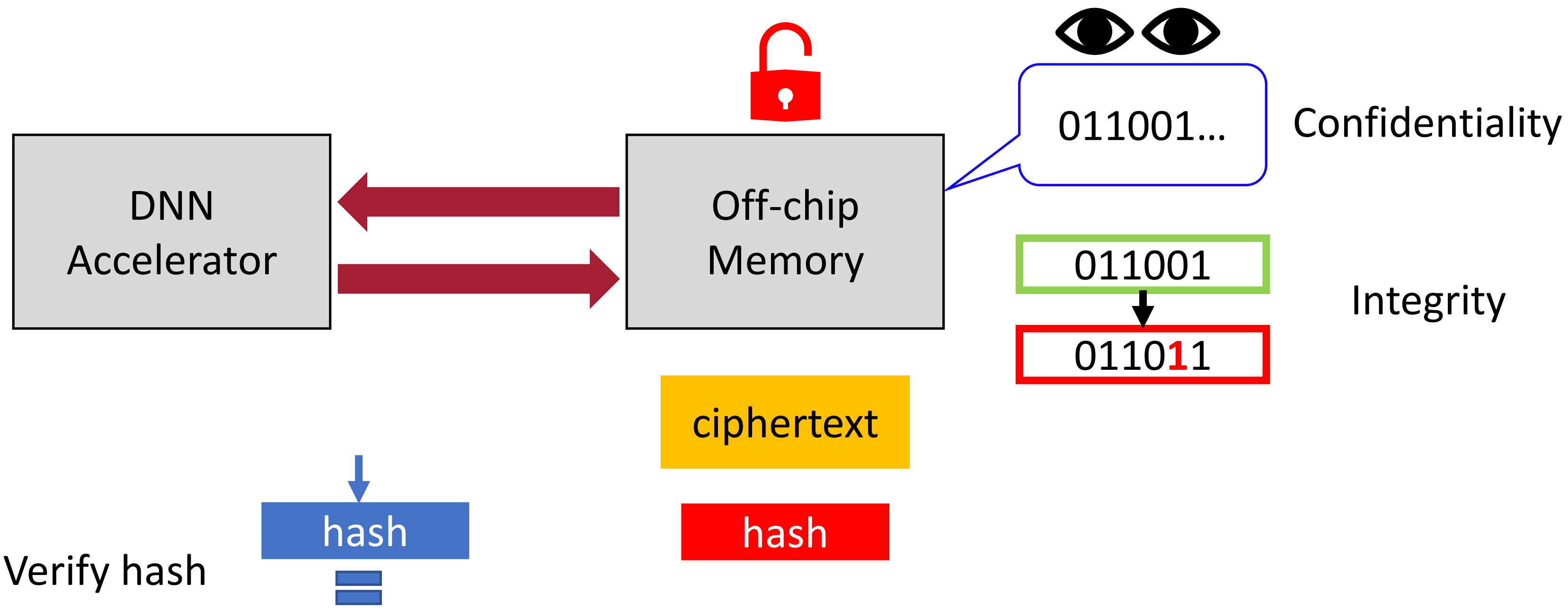
Integrity

Support integrity with cryptographic hashes

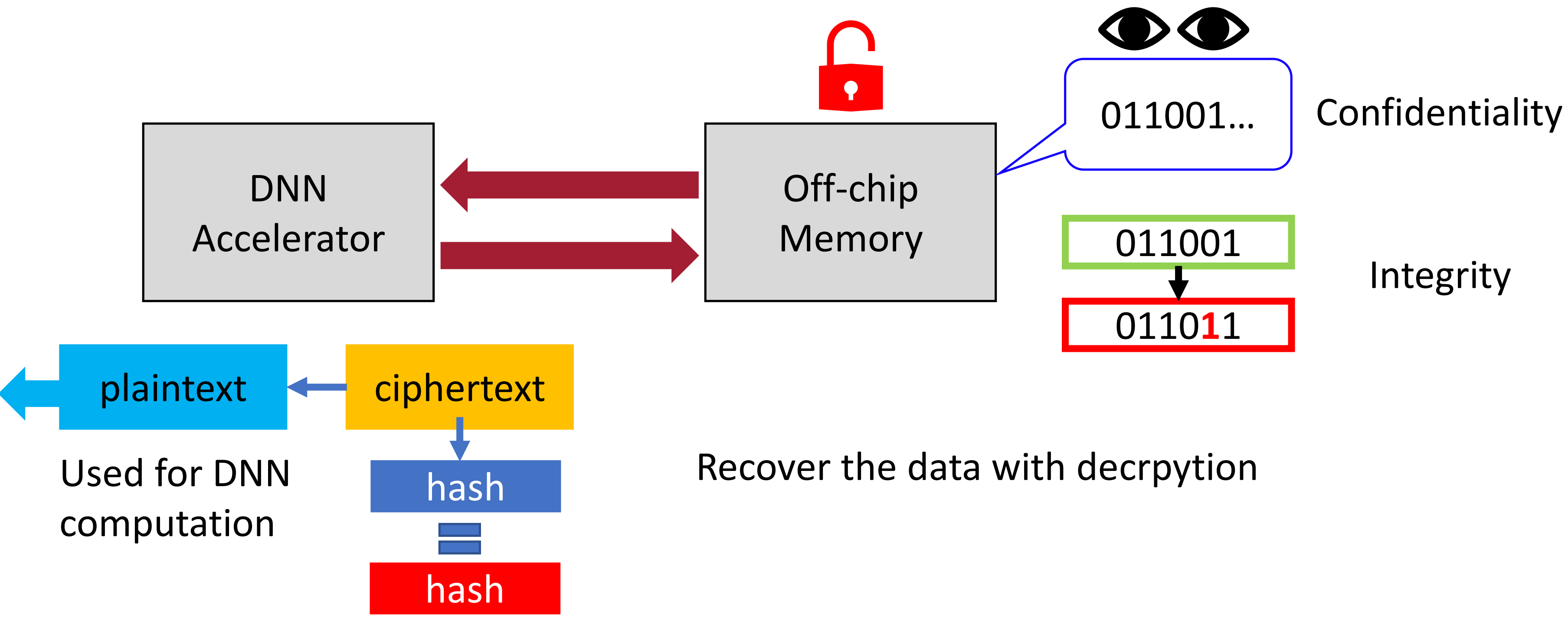
ciphertext

hash

# Background: TEE and DNN Accelerators



# Background: TEE and DNN Accelerators



DNN Accelerator

Off-chip Memory

011001…

Confidentiality

011001

011011

Integrity

plaintext

ciphertext

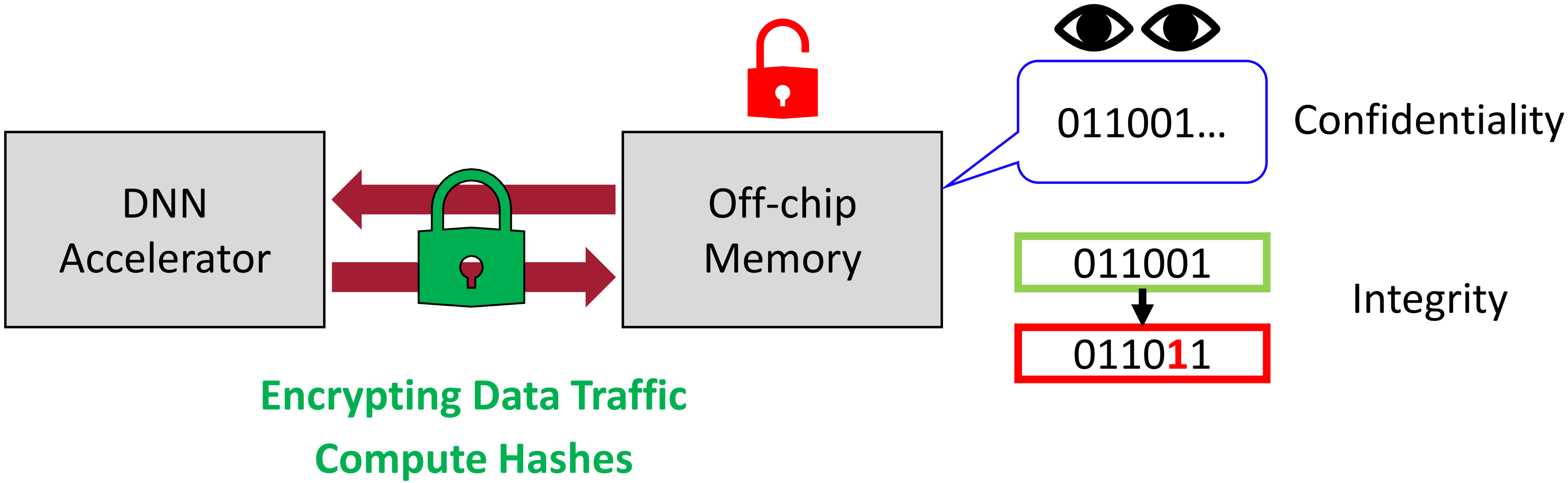
hash

hash

Used for DNN computation

Recover the data with decrpytion

# Background: TEE and DNN Accelerators

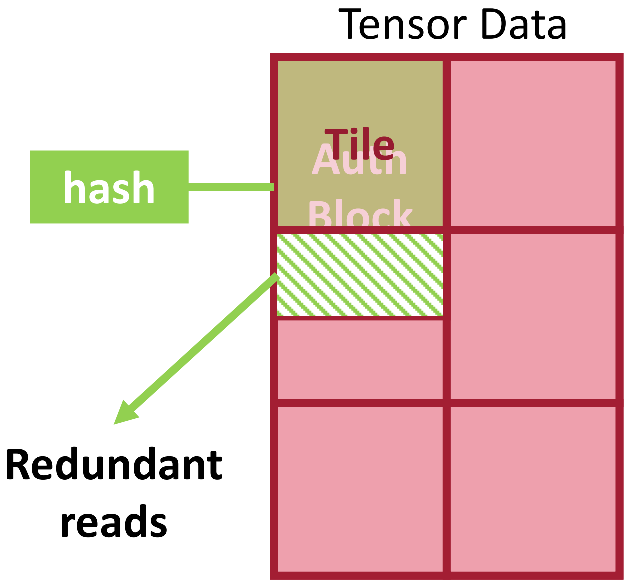


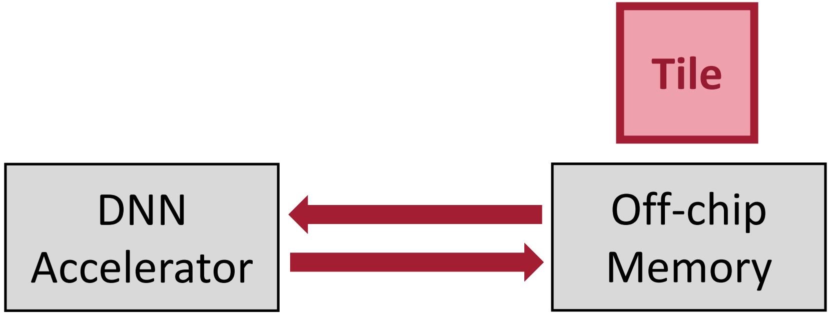
Encrypting Data Traffic

Compute Hashes

- Accelerator scheduling has to be coordinated with cryptographic operations

# What if Tile != Authentication Block

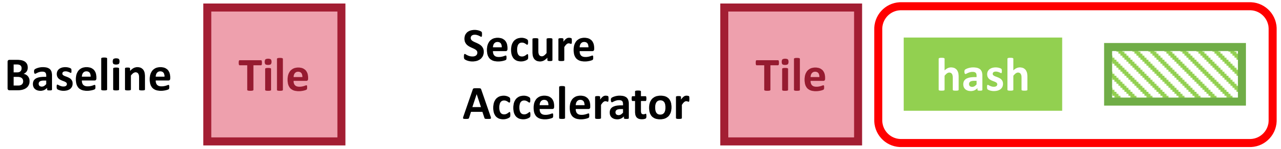




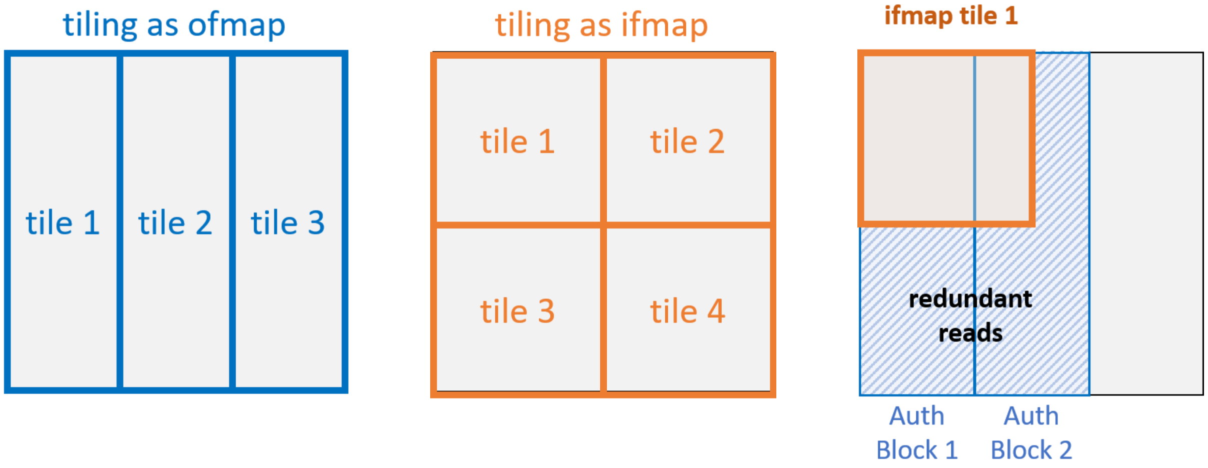
Can’t get a hash only with this ▨ta

Fetch extra “redundant” data for hash

additional off-chip traffic



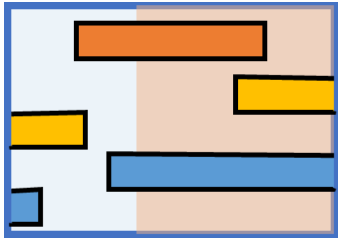
# Tile-as-an-AuthBlock is not optimal



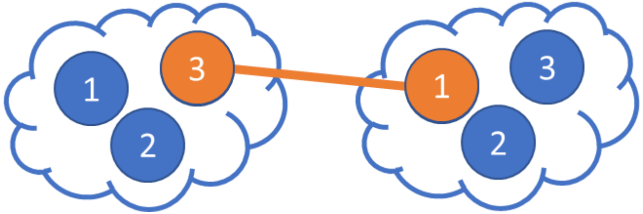
- Option 1: Tile-as-an-AuthBlock based on the output tiling ☹
- Option 2: Rehash between layers ☹

# Summary of our techniques

Analytical approach to identify the optimal AuthBlock assignment

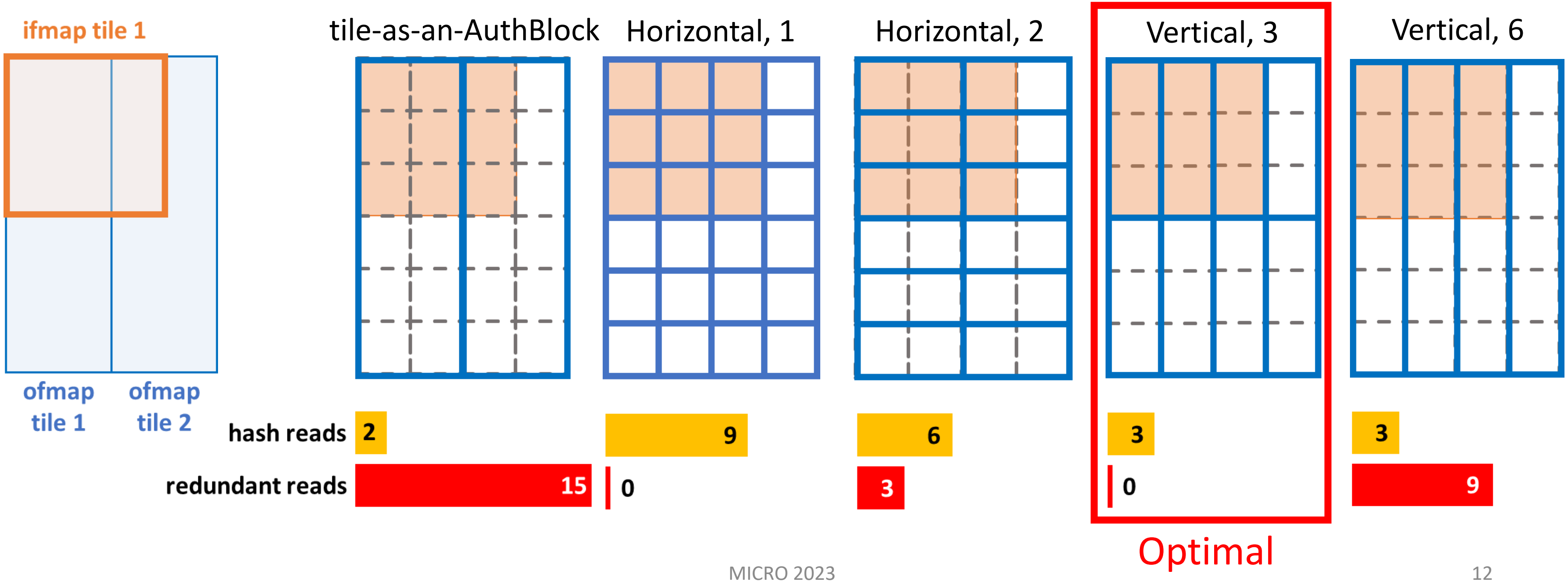


Cross-layer fine tuning from the loopnest schedule



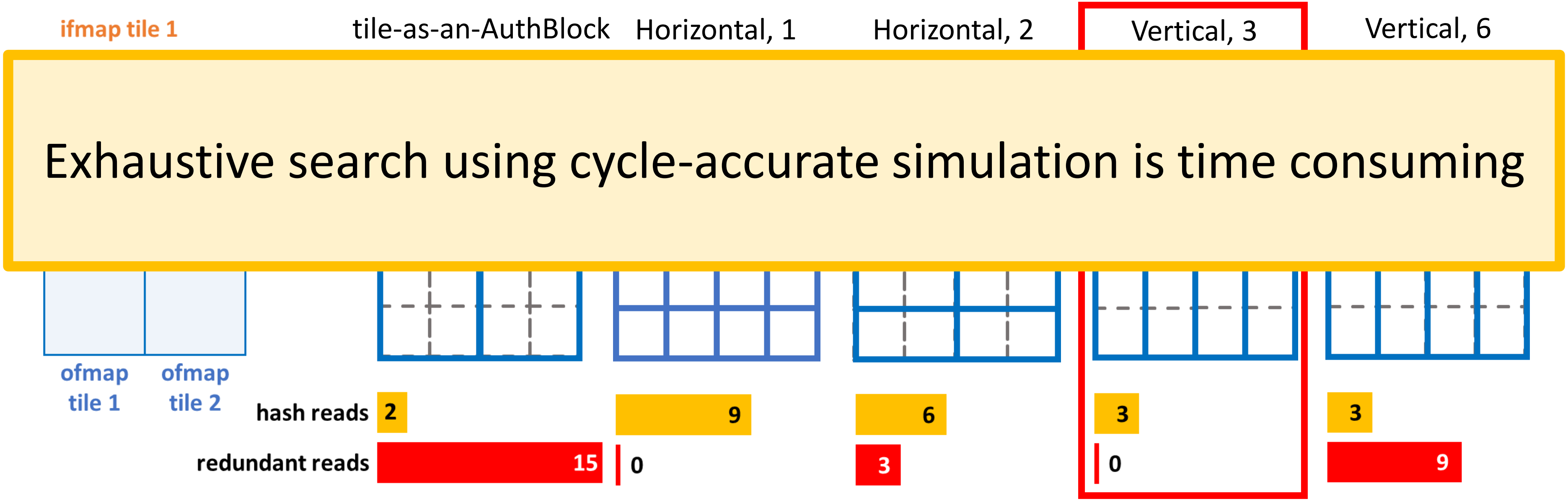
# Search space of AuthBlocks is complex

- Find the AuthBlock assignment that minimizes the additional off-chip traffic
- Both size and orientation of AuthBlocks matter



# Search space of AuthBlocks is complex

- Find the AuthBlock assignment that minimizes the additional off-chip traffic
- Both size and orientation of AuthBlocks matter

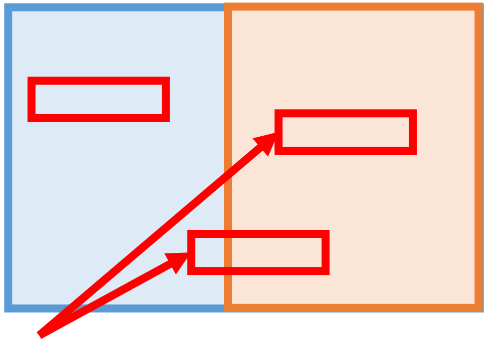


Exhaustive search using cycle-accurate simulation is time consuming

Optimal

# Analytical approach to AuthBlock assignment

- Counting how many AuthBlocks intersect a tile



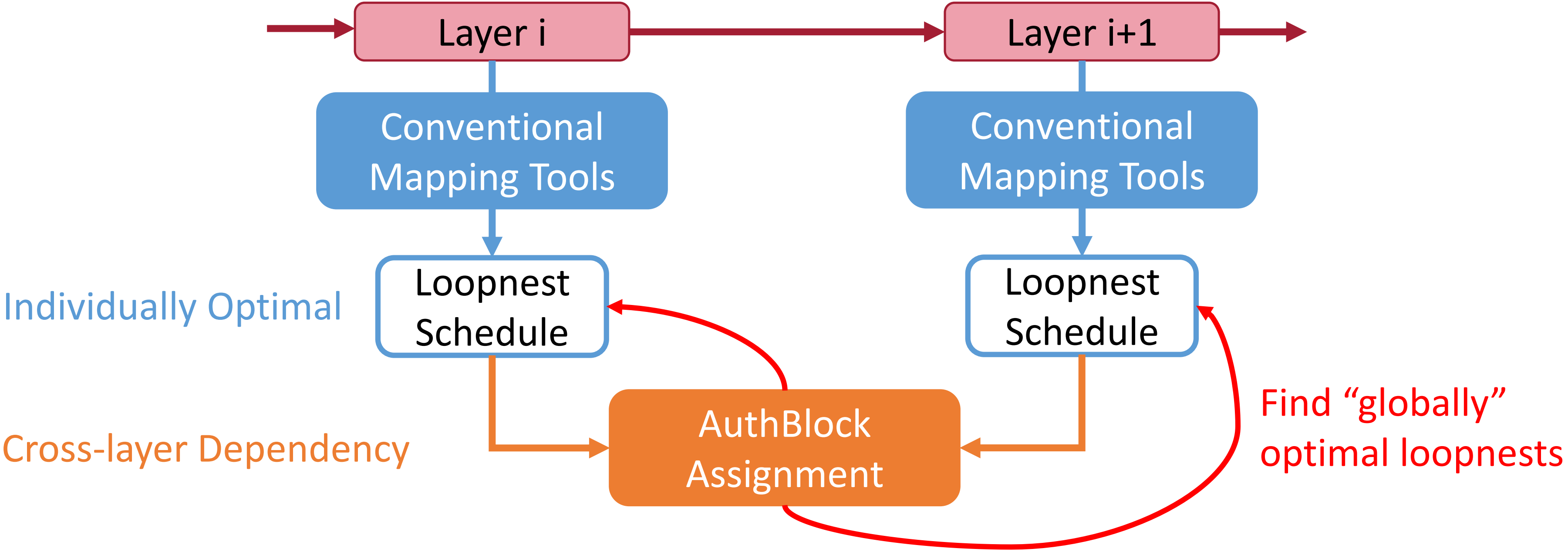
Count these AuthBlocks

Please refer to the paper for details

**Linear congruence problem**

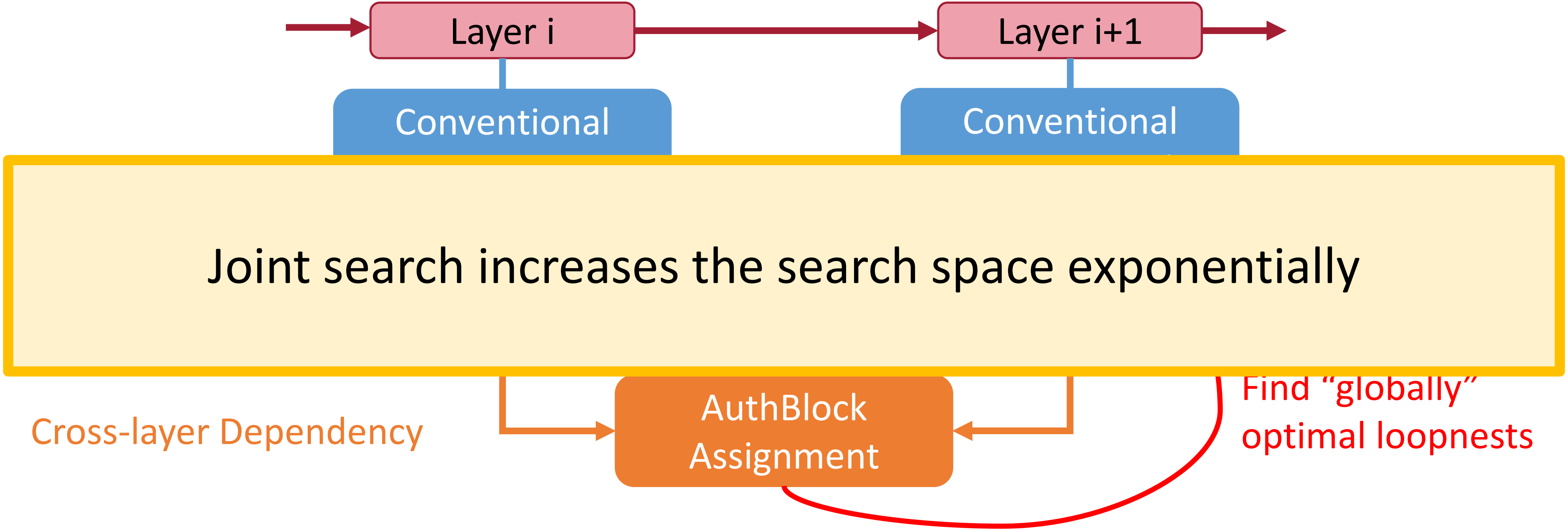
$$u \times k \equiv \min\left(w_i - w_j - u + 1, 0\right)$$
$$, \ldots, w_i - 1 \bmod w_i$$

# Cross-layer dependency from the loopnest level



**Loopnest schedules optimal for individual layer ↛ globally optimal**

# Cross-layer dependency from the loopnest level

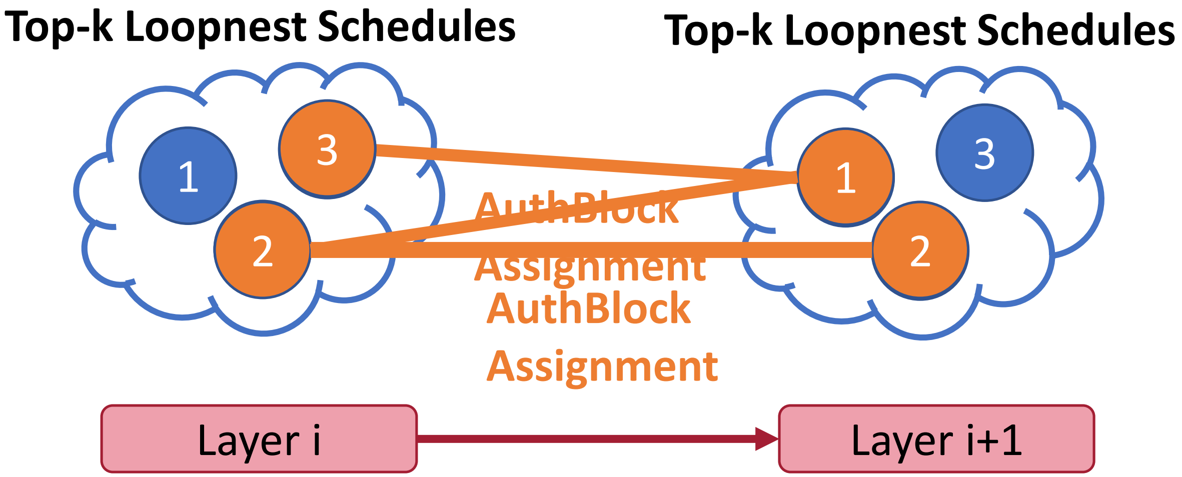


Joint search increases the search space exponentially

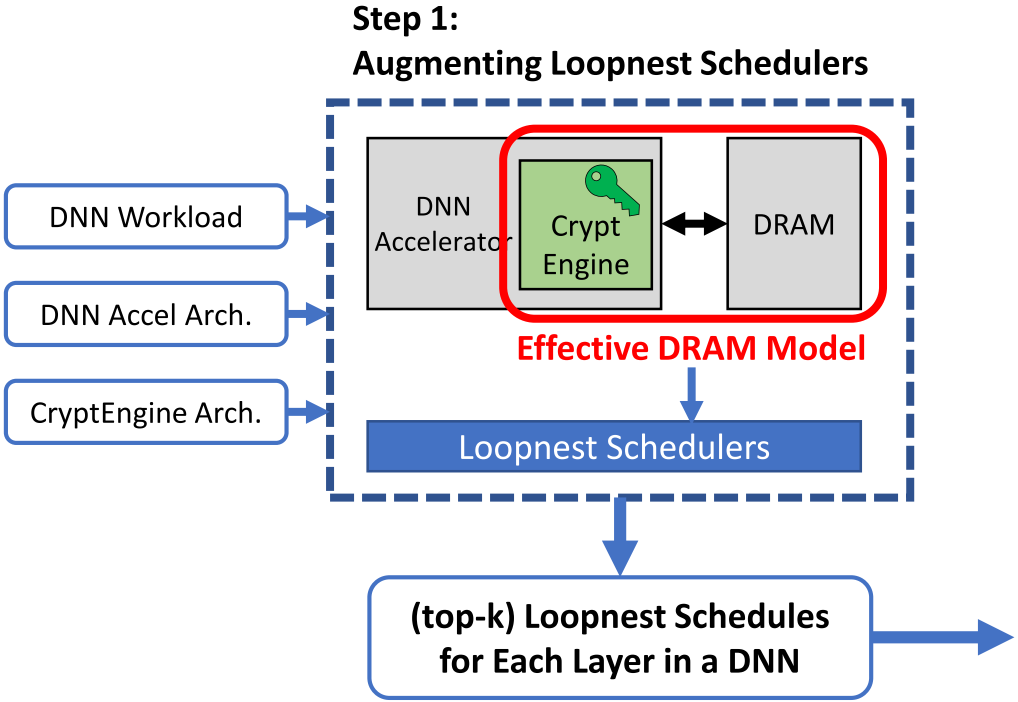
**Loopnest schedules optimal for individual layer ↛ globally optimal**

# Heuristic approach to joint optimization

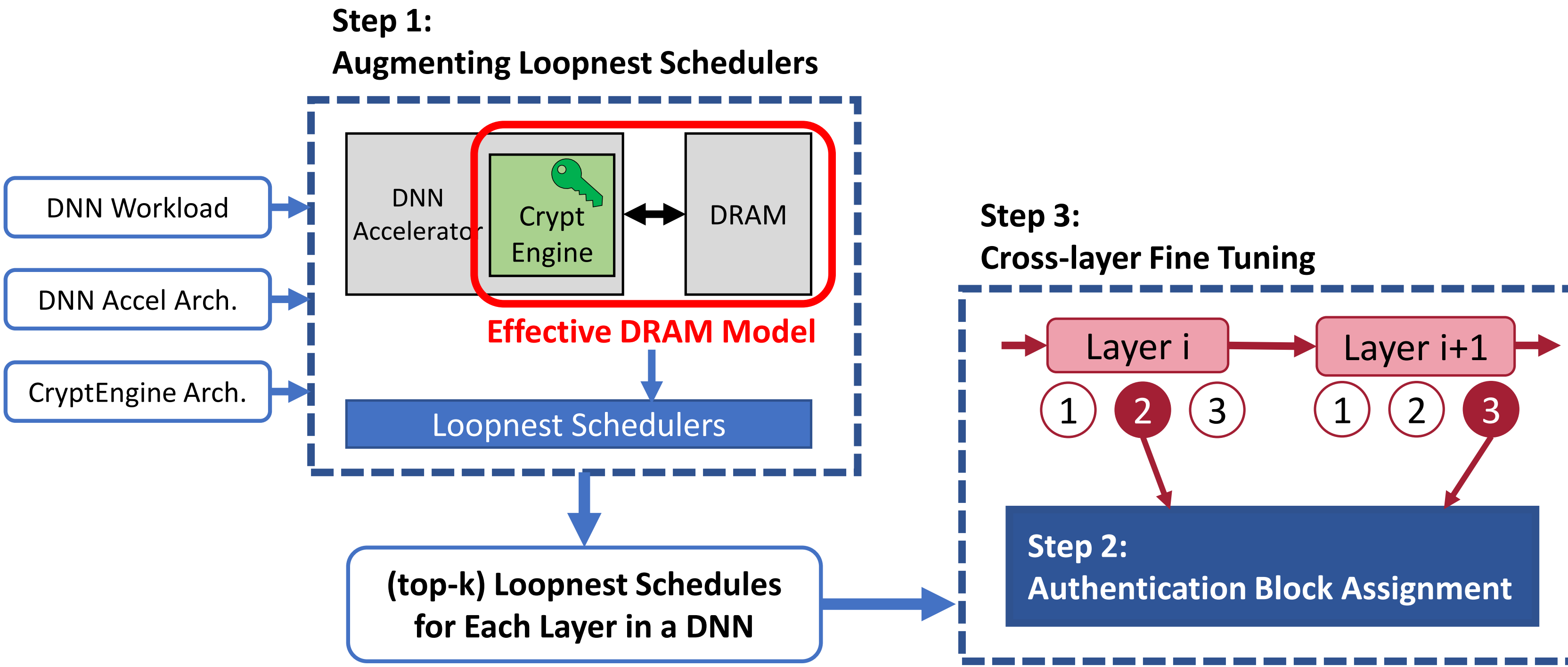
- Simulated annealing to find the approximate solution



# SecureLoop: Scheduling Search Engine



# SecureLoop: Scheduling Search Engine



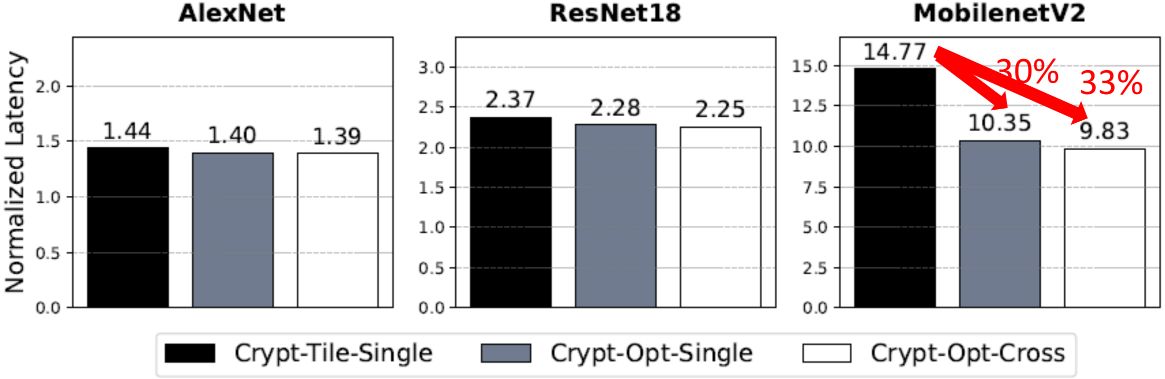
# Comparing scheduling algorithms

## Setup

- Eyeriss-like + AES-GCM
- Mostly conv workloads
- Different scheduling algo.

*shallow,*
*unavoidable rehashing*

*deeper, ↑ opportunity*



Step 1 only
+ tile-as-an-AuthBlock

Step 1 + Step 2

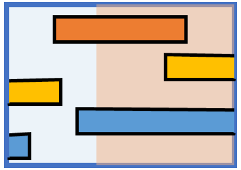
All three steps

**Summary:** ~33% faster, ~50% better in EDP compared to the “tile-as-an-AuthBlock”

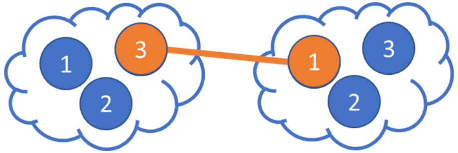
# SecureLoop

- Scheduling algorithm considering cryptographic operations

Analytical approach to identify the optimal AuthBlock assignment



Cross-layer fine tuning from the loopnest schedule



- Design space exploration for secure accelerators

more in paper, including sweeps for different design choices and trade-off curves..