# Hybrid Cost Model for Input Row Activation

PIM Optimizer Team

December 30, 2025

## 1 Introduction

Accurately estimating DRAM Row Activations for Input Feature Maps is critical for performance optimization. Simple analytical models (e.g., assuming contiguous access) fail to capture complex interactions between:

- **Tile Alignment**: The relative offset of a tile within a DRAM Row.
- **Loop Order**: The specific traversal order (e.g., Row-Major vs. Block-Major) within a tile.
- **Row Thrashing**: The penalty incurred when a single tile repeatedly crosses DRAM Row boundaries.

This document describes a **Hybrid Cost Model** that combines **Number Theory (GCD)** for efficient quantity estimation and **Micro-Trace Simulation** for accurate unit cost calculation.

## 2 The Hybrid Approach: Categorization & Sampling

The total Input Row Activation cost is derived by categorizing tiles into two states: **Safe** (aligned within a row) and **Crossing** (straddling a row boundary).

$$\text{TotalCost} = N_{safe} \cdot C_{safe} + N_{crossing} \cdot C_{crossing} \tag{1}$$

### 2.1 1. Cost Determination ($C$)

We use **Micro-Trace Simulation** to determine the unit costs.

- **Safe Cost ($C_{safe}$)**: The cost when the tile is perfectly aligned (Offset = 0).

$$C_{safe} = \text{MicroTrace}(0) \tag{2}$$

- **Crossing Cost ($C_{crossing}$)**: The expected cost when the tile is in the "Danger Zone". We sample offsets specifically from the crossing region.

$$C_{crossing} = \frac{1}{|\mathcal{O}_{cross}|} \sum_{o \in \mathcal{O}_{cross}} \text{MicroTrace}(o) \tag{3}$$

### 2.2 2. Quantity Determination ($N$)

Instead of simulating every tile to check if it crosses, we use **Number Theory** to calculate the exact counts.

Let $P$ be the Period (Row Buffer Size) and $S$ be the Step (Stride in linear memory). The "Danger Zone" $Z$ is the set of offsets that cause a tile of width $W_{tile}$ to cross a boundary:

$$Z = [P - W_{tile} + 1, P - 1] \tag{4}$$

The sequence of offsets is $o_i = (i \cdot S) \pmod P$. This sequence visits the set of values $\{0, g, 2g, \dots\}$ where $g = \gcd(S, P)$. The probability of landing in the Danger Zone is:

$$Prob_{crossing} = \frac{\text{Count of multiples of } g \text{ in } Z}{P/g} \tag{5}$$

Thus, the exact quantities are:

$$N_{crossing} = N_{total} \times Prob_{crossing} \quad (6)$$
$$N_{safe} = N_{total} - N_{crossing} \quad (7)$$

# 3 Micro-Trace Simulation

To determine the cost of a single tile at a specific offset $o$, we use a lightweight simulation.

## 3.1 Strict Ordering (Realistic Hardware)

We simulate the exact memory access sequence generated by the hardware loop nest (e.g., Row-Major $H, W$).

**Algorithm 1** Micro-Trace Simulation (Strict Order)

```
ActiveRow ← −1
Activations ← 0
for h ∈ [0, TileH − 1] do
    for w ∈ [0, TileW − 1] do
        Addr ← BaseAddr + o + h · W_total + w
        RowID ← Addr//RowBufferSize
        if RowID ≠ ActiveRow then
            Activations ← Activations + 1
            ActiveRow ← RowID
        end if
    end for
end for
return Activations
```

This captures **Row Thrashing** where a tile straddles a row boundary and the access pattern causes repeated switching (e.g., $A \to B \to A \to B$).

# 4 Validation Results

We compared the Hybrid Model against a full exhaustive simulation of all tiles.

| Scenario | Params (H, W, Tile, Stride) | Ground Truth | Hybrid Model | Error |
|---|---|---|---|---|
| Aligned | 100, 1024, 3x3, 1 | 3.0000 | 3.0000 | 0.00% |
| Misaligned | 224, 224, 3x3, 1 | 1.4324 | 1.4375 | 0.35% |
| Large Stride | 224, 224, 3x3, 2 | 1.4324 | 1.4375 | 0.35% |
| Large Tile | 224, 224, 7x7, 2 | 2.3119 | 2.3125 | 0.02% |

Table 1: Validation of Hybrid Cost Model vs. Exhaustive Simulation

# 5 Implementation in ILP

The ILP optimizer will use this model to precompute a lookup table or cost function:

1. For a candidate mapping $(TileH, TileW)$, calculate $E[\text{Cost}]$.
2. Use $E[\text{Cost}]$ as the coefficient for the Input Row Activation term in the objective function.
3. This ensures the optimizer penalizes mappings that result in high thrashing probabilities.