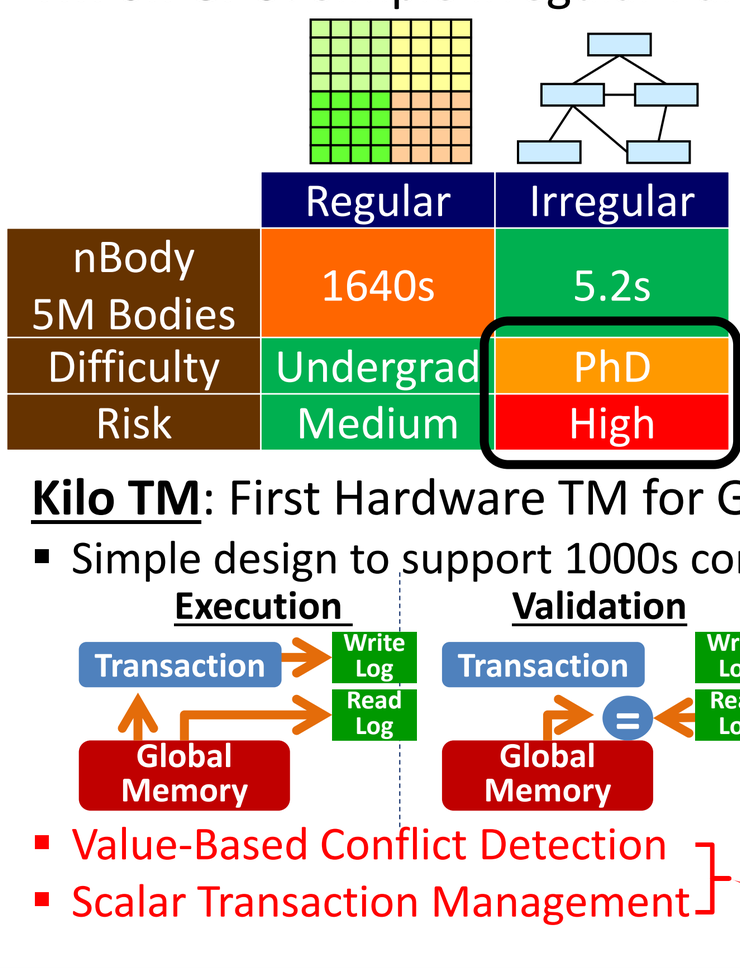
# Energy Efficient GPU Transactional Memory via Space-Time Optimizations

**Wilson W. L. Fung** wwlfung@ece.ubc.ca
**Tor M. Aamodt** aamodt@ece.ubc.ca
University of British Columbia

a place of mind

## TM on GPU: Energy Concern

**TM on GPU**: Simple Irregular Parallelism on GPUs




TM Reduces Risk in Software Development
- Predictable Dev Time
- No Deadlocks
- More Maintainable Code

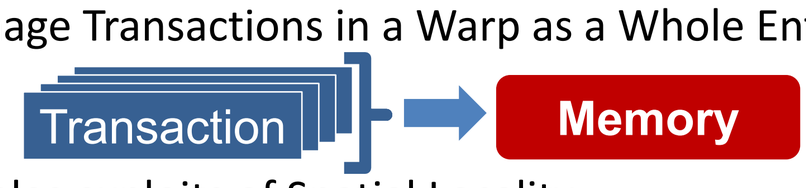
**Kilo TM**: First Hardware TM for GPU
- Simple design to support 1000s concurrent transactions




- Value-Based Conflict Detection
- Scalar Transaction Management

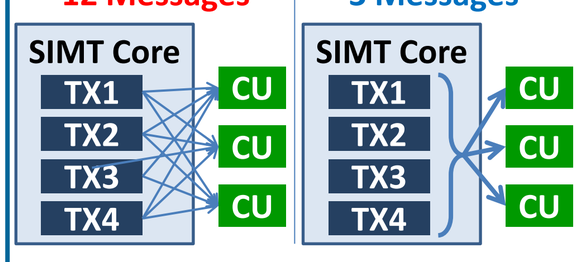
## Warp Level Transaction Management

Key Idea
- Manage Transactions in a Warp as a Whole Entity

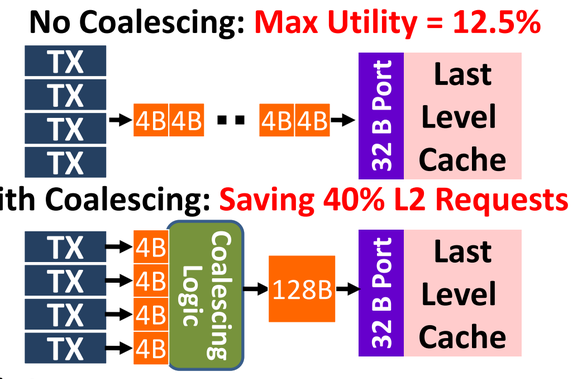



- Enables exploits of Spatial Locality:

**Aggregate Control Messages**



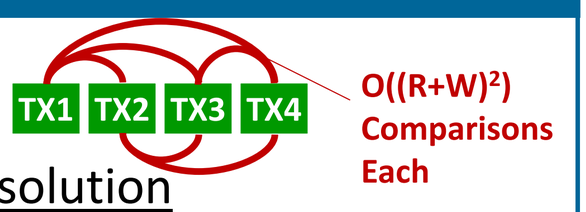

**Validation and Commit Coalescing**




- Challenge: Intra-Warp Conflict

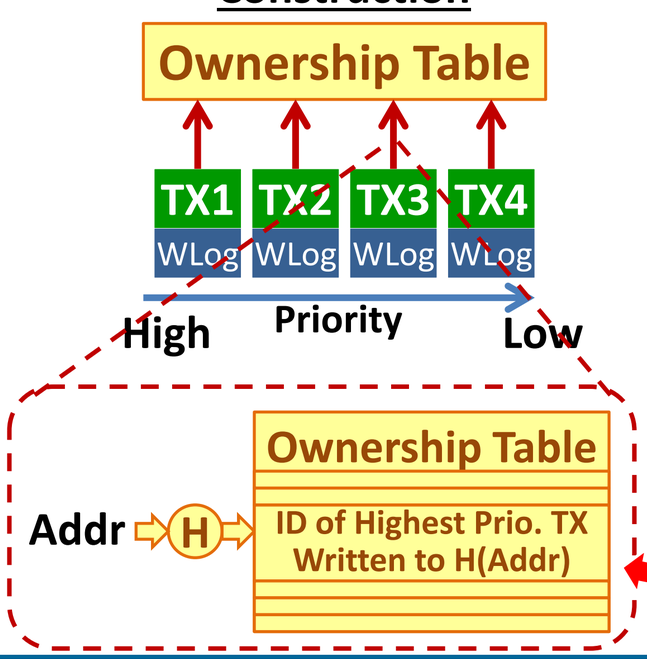
## Intra-Warp Conflict Resolution

Naïve Conflict Resolution
- $O(T^2 \times (R+W)^2)$



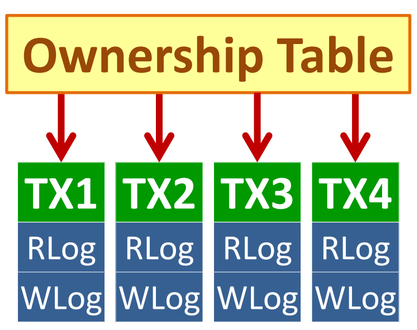

2-Phase Parallel Conflict Resolution
- Insight: Fixed priority for conflict resolution enables parallel resolution

**1: Ownership Table Construction**




Stored in Shared Memory (On-Chip Per-Core Scratchpad)

**2: Parallel Match**




**Owner ID < My ID → Abort**
**Owner ID = My ID → Pass**
**Owner ID = NULL → Pass**

## Temporal Conflict Detection

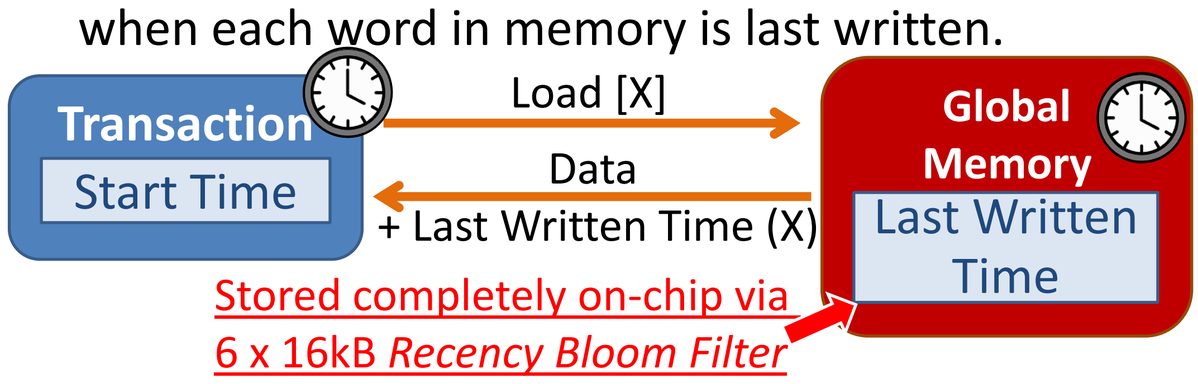
Motivation: Skip Value-Based Conflict Detection for Conflict-Free Read-Only Transactions
- Transaction may skip write to memory due to data dependent control flow.
- Programmer may introduce them for memory consistency.

Examples:

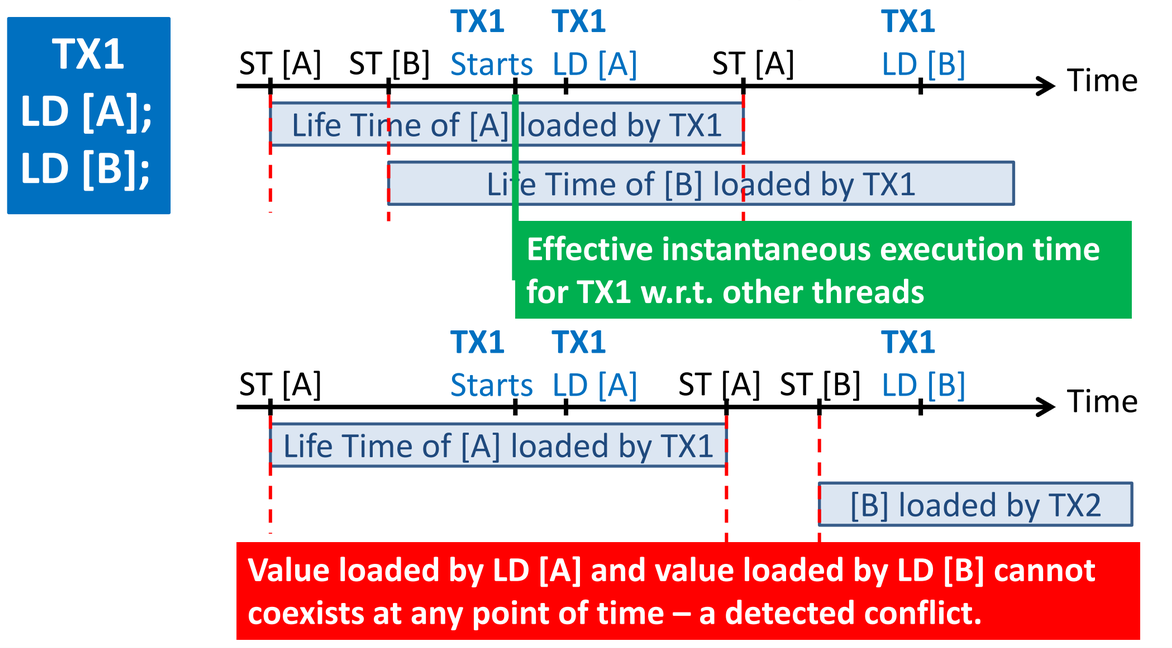
```
TX1
if (C == 0)
  B = B + 1;
```

```
TX2
int K;
K = X + Y;
```

Key Idea
- Use globally synchronous on-chip timers to record when each word in memory is last written.




- If every value loaded by a read-only transaction has not been written since the start of the transaction, the transaction can commit silently.




## Results

GPGPU-Sim 3.2.1 + GPUWattch

| | |
|---|---|
| HT-[H/M/L] – Hash Table Construction | ATM – Bank Transactions |
| BH-[H/L] – Barnes Huts (N-Body) | CL/CLto – Cloth Simulation |
| CC – Maxflow/Mincut Graph | AP – Data Mining |