A Scalable Circular Pipeline Design for Multi-Way Stream Joins in Hardware

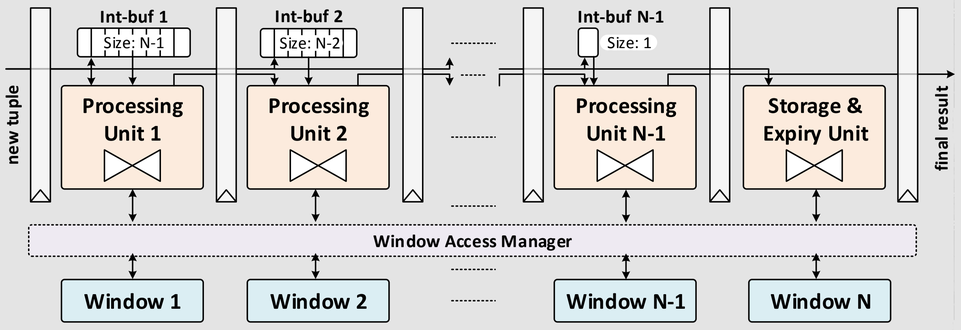
Mohammadreza Najafi[1], Mohammad Sadoghi[2], Hans-Arno Jacobsen[1]

[1] Technical University of Munich
[2] Exploratory Systems Lab
[2] University of California, Davis

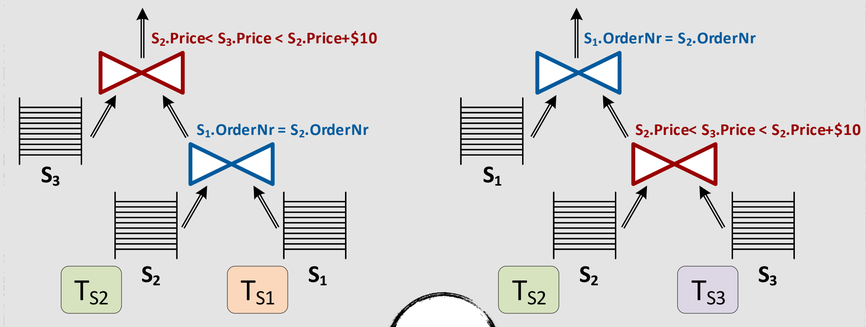
1 Linear Pipeline Multi-Way Stream Join Design

It is **non-trivial** to build multi-way join operators by cascading operators designed for two streams

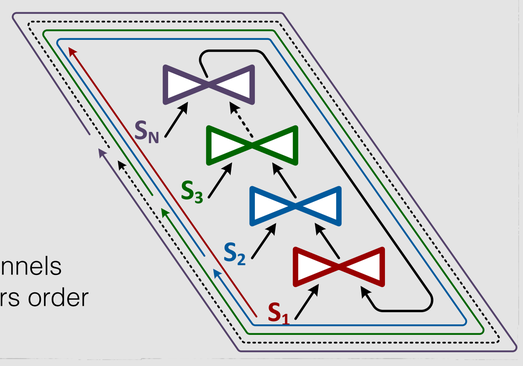
Arbitrary join operators order
Not scalable due to the N-to-N connection in the "Window Access Manager" crossbar



2 Operators Reordering Issue in Hardware



3 Circular Design



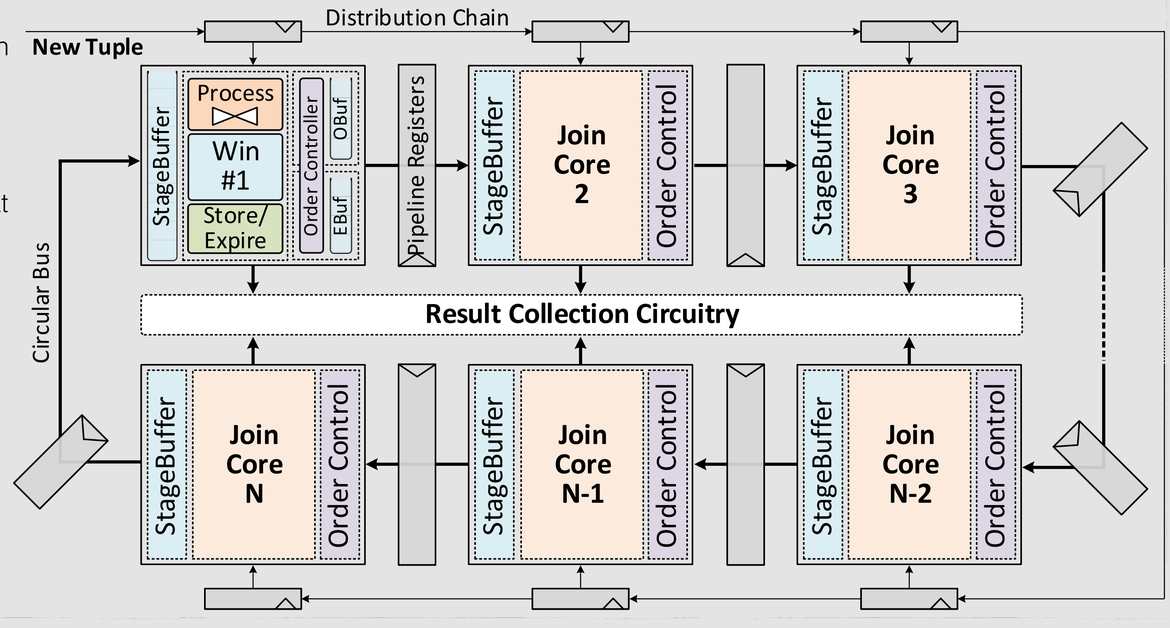
Circular data-path
Multiple input channels
Fixed join operators order

4 Circular-MJ Architecture

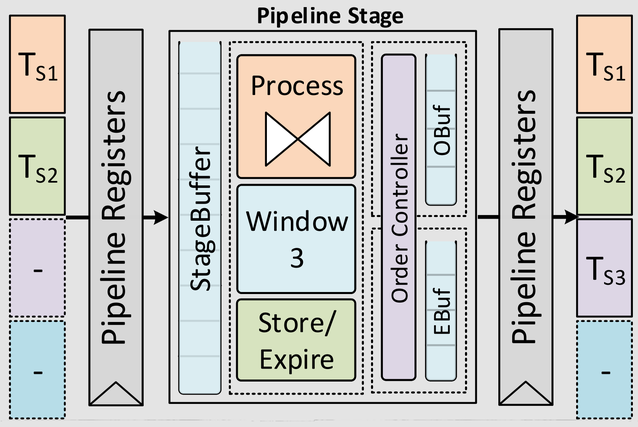
A **dedicated stage** for each stream sliding window **limits data dependency** between stages

A **scalable** architecture that is centered around direct neighbor-to-neighbor communication

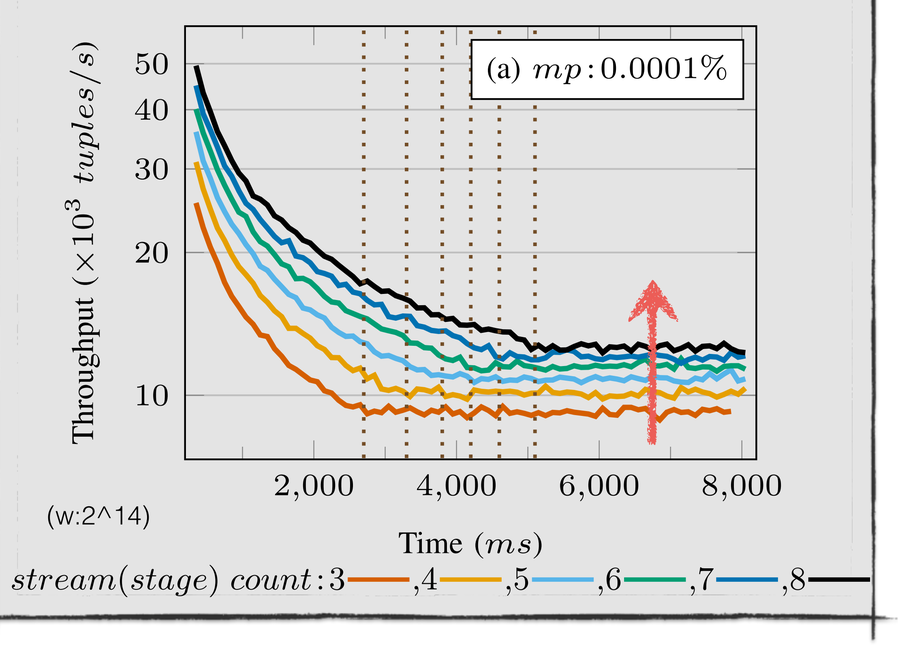
Eliminate the join operator reordering problem by moving the reordering task to tuple insertion circuitry using a **pipelined distribution chain**



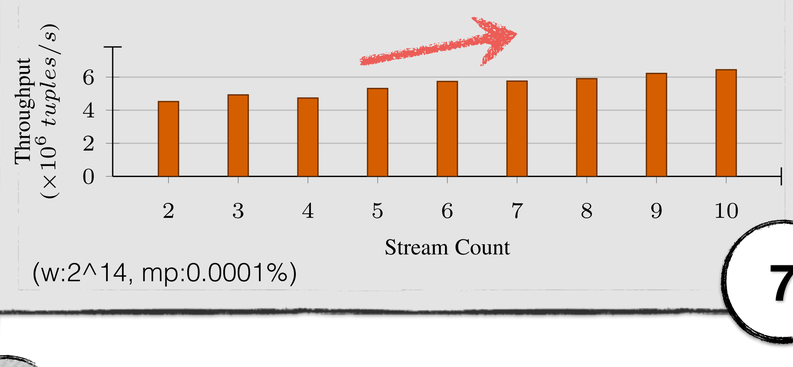
5 Intermediate Result Generation in a Stage



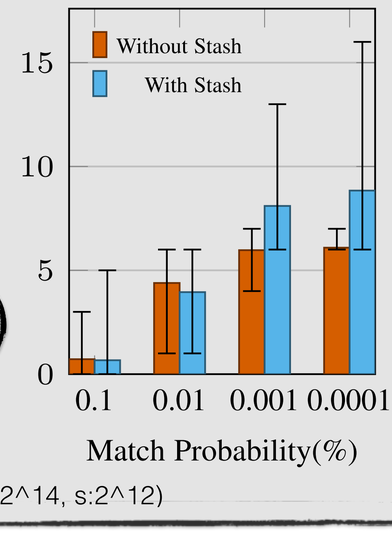
6 Throughput with Nested-Loop Approach



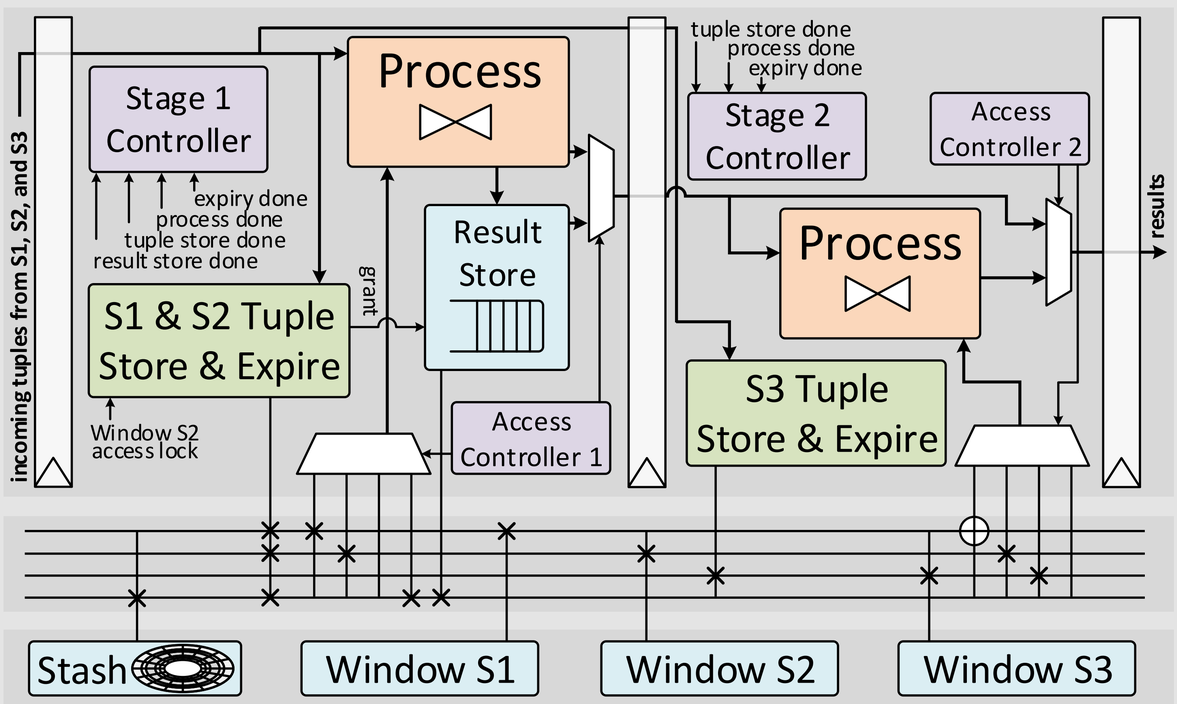
7 Throughput with Hash-Based Approach



3 Stash Effect on Throughput



1 Stashed-MJ Architecture



A design to **avoid the re-computation** of already processed data

2 Stash Design and its Internal Components