Perf Tool New Features (v5.15..v6.1-rc1)

Earlier : 《Deep dive into Linux perf tool》、perf-doc、perf-wiki、 tracing-summit

v5.15

- Uncore PMU **aliases**
- ARM Coresight ETE decoder (HW Assisted Tracing, Armv9)
- More **fused instructions** in 'perf annotate' 7efbcc8c075c & 3149733584c8

v5.16

- Armv8 (Neoverse-v1, AWS), **riscv64**
- **Armv8.2 Statistical Profiling Extension (SPE)**
- Multithreaded perf.data readers
- Arm Coresight **snapshot** mode
- PID tracing、task context switch; (ARM hw-based)

v5.17

- $ perf **ftrace trace** -G check_preempt_curr sleep 0.00001
- $ perf **ftrace latency** (-b/--use-bpf) -T __handle_mm_fault -a sleep 1
- Arm Neoverse N2

v5.18

- $ perf lock report -c/--**combine-locks** // merge lock instances per class
- $ **perf record –threads** // per data files for each cpu
- intel-pt, CFE (**Control Flow** Event) and EVD (Event Data)
  - Each Event Trace event is represented by a CFE packet that is preceded by zero or more EVD packets.

INTR、IRET、SMI、RSM、SIPI、INIT、VMENTRY、VMEXIT、VMEXIT_INTR、SHUTDOWN

v5.19

- **Intel hybrid PMU、AMD Zen4 Instruction Based Sampling (IBS)、AMD PerfMonV2、AMD Fam19h Branch Sampling**
- **Enable off-cpu analysis with BPF** edc41a1099c2
  - $ perf record --off-cpu // off-cpu time in read and write
  - tracking child processes (in v6.0)

v6.0

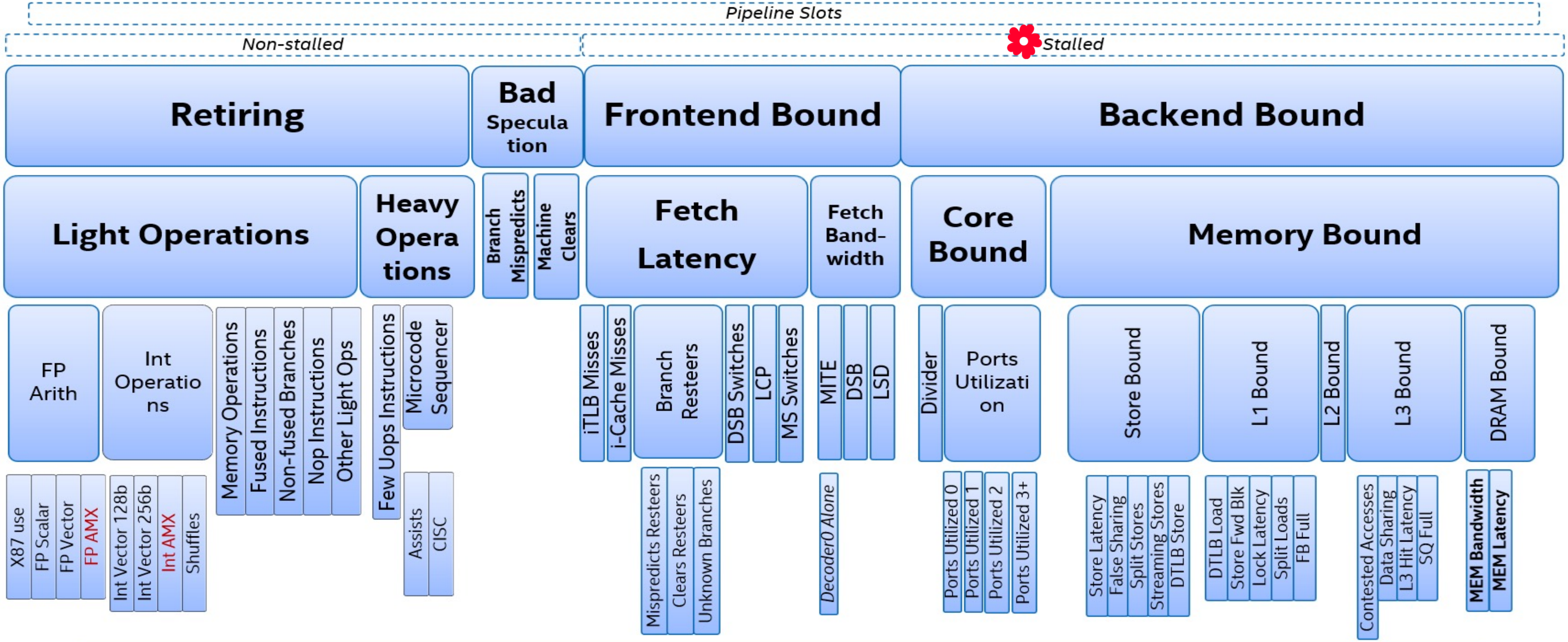
- $ **perf lock contention -b** // based on the lock contention events
- **perf c2c on ARM64** (at least Neoverse)
  - trace physical address for Arm SPE events (unchanging VMX TSC Offset & no VMX TSC Scaling)
- **intel_pt, tracing guest user-space on the host**
- **perf kwork** 0f70d8e9db4f // bpf counters
  - trace time properties of kernel work (such as irq, softirq, and workqueue), including runtime, latency, and timehist

v6.1

- AMD Zen4 **Last Branch Record** Extension Version 2 (LbrExtV2)
- **perf mem/c2c on AMD** (load、store、load latency)
- HiSilicon **PCIe PMU**
  - monitors bandwidth, latency
  - bus utilization and buffer occupancy

Yes, it may under/over count.

Topdown v4.4 : tune single-thread payload



Pipeline Slots

Non-stalled

Stalled

Retiring

Bad Speculation

Frontend Bound

Backend Bound

Light Operations

Heavy Operations

Branch Mispredicts

Machine Clears

Fetch Latency

Fetch Band-width

Core Bound

Memory Bound

FP Arith

Int Operations

Memory Operations

Fused Instructions

Non-fused Branches

Nop Instructions

Other Light Ops

Few Uops Instructions

Microcode Sequencer

iTLB Misses

i-Cache Misses

Branch Resteers

DSB Switches

LCP

MS Switches

MITE

DSB

LSD

Divider

Ports Utilization

Store Bound

L1 Bound

L2 Bound

L3 Bound

DRAM Bound

X87 use

FP Scalar

FP Vector

FP AMX

Int Vector 128b

Int Vector 256b

Int AMX

Shuffles

Assists

CISC

Mispredicts Resteers

Clears Resteers

Unknown Branches

Decoder0 Alone

Ports Utilized 0

Ports Utilized 1

Ports Utilized 2

Ports Utilized 3+

Store Latency

False Sharing

Split Stores

Streaming Stores

DTLB Store

DTLB Load

Store Fwd Blk

Lock Latency

Split Loads

FB Full

Contested Accesses

Data Sharing

L3 Hit Latency

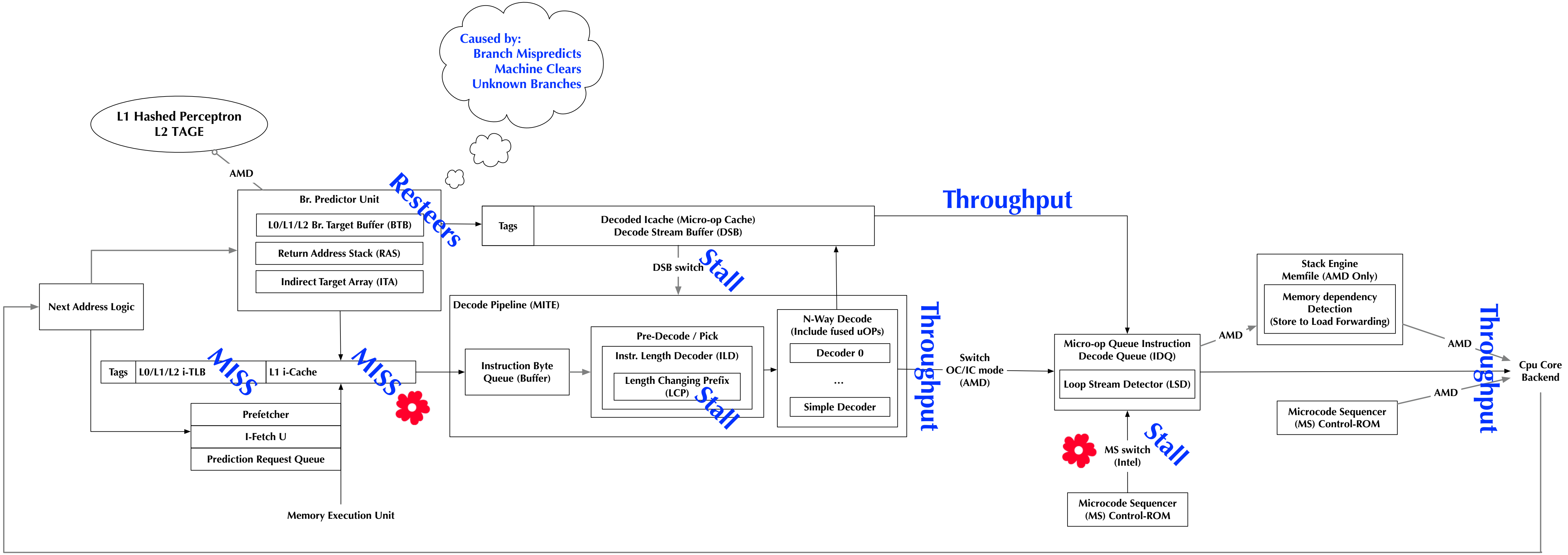
SQ Full

MEM Bandwidth

MEM Latency

* Reference paper: A. Yasin, “A Top-Down Method for Performance Analysis and Counters Architecture”, ISPASS 2014

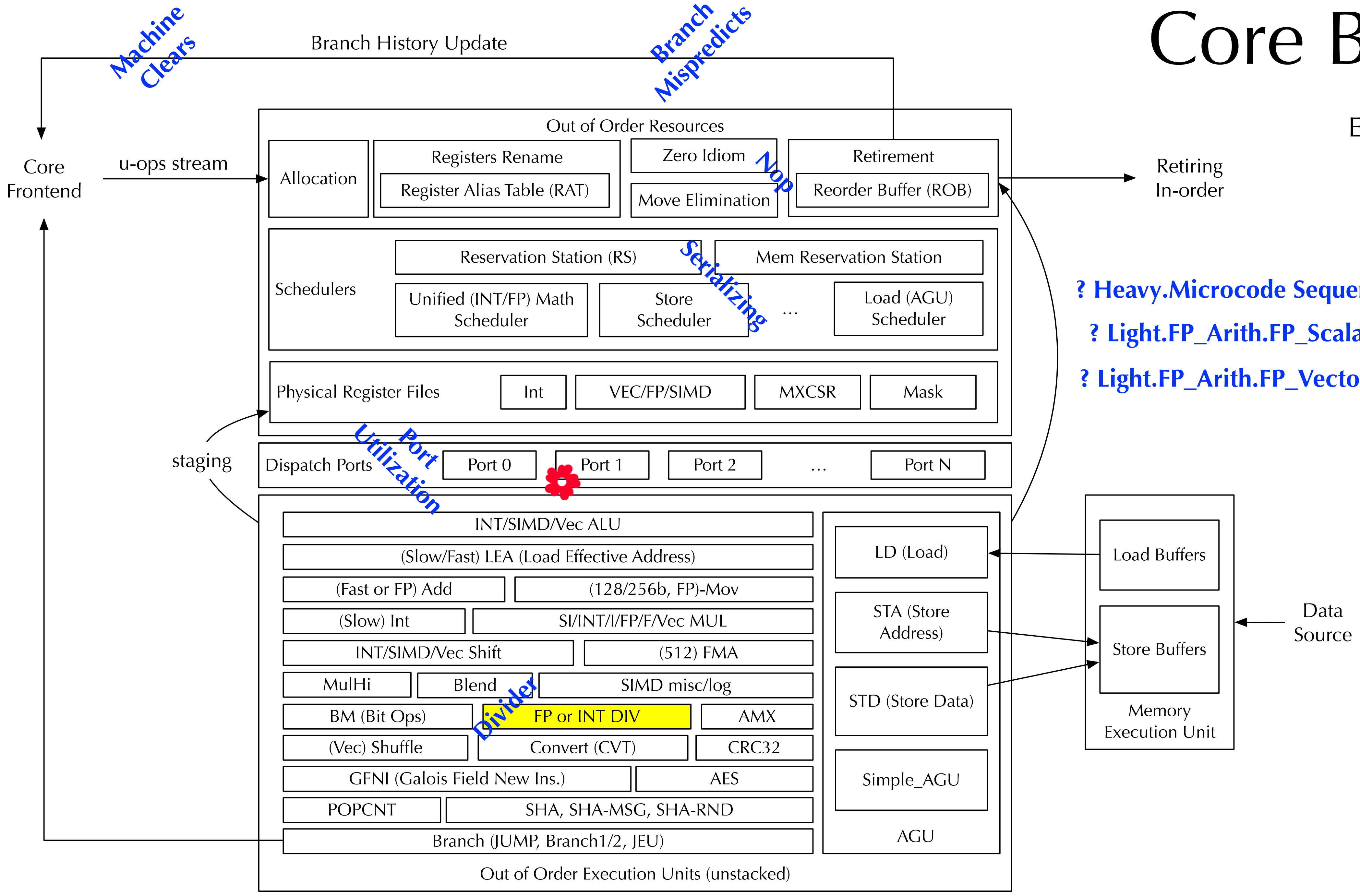
Core Frontend (Predictability of Code)



Caused by:
Branch Mispredicts
Machine Clears
Unknown Branches
L1 Hashed Perceptron
L2 TAGE
AMD
Br. Predictor Unit
L0/L1/L2 Br. Target Buffer (BTB)
Return Address Stack (RAS)
Indirect Target Array (ITA)
Resteers
Throughput
Tags
Decoded Icache (Micro-op Cache)
Decode Stream Buffer (DSB)
DSB switch
Stall
Next Address Logic
Decode Pipeline (MITE)
Tags
L0/L1/L2 i-TLB
MISS
L1 i-Cache
MISS
Instruction Byte Queue (Buffer)
Pre-Decode / Pick
Instr. Length Decoder (ILD)
Length Changing Prefix (LCP)
Stall
N-Way Decode (Include fused uOPs)
Decoder 0
...
Simple Decoder
Throughput
Switch OC/IC mode (AMD)
Prefetcher
I-Fetch U
Prediction Request Queue
Memory Execution Unit
Micro-op Queue Instruction Decode Queue (IDQ)
Loop Stream Detector (LSD)
MS switch (Intel)
Stall
Microcode Sequencer (MS) Control-ROM
AMD
Stack Engine
Memfile (AMD Only)
Memory dependency Detection
(Store to Load Forwarding)
AMD
Throughput
AMD
Microcode Sequencer (MS) Control-ROM
Cpu Core Backend

Core Backend

Execution Latency
Execution Throughput



Machine Clears

Branch History Update

Branch Mispredicts

Core Frontend

u-ops stream

Out of Order Resources

Allocation

Registers Rename

Register Alias Table (RAT)

Zero Idiom

Move Elimination

Nop

Retirement

Reorder Buffer (ROB)

Retiring In-order

Schedulers

Reservation Station (RS)

Mem Reservation Station

Serializing

Unified (INT/FP) Math Scheduler

Store Scheduler

…

Load (AGU) Scheduler

Physical Register Files

Int

VEC/FP/SIMD

MXCSR

Mask

staging

Dispatch Ports

Port Utilization

Port 0

Port 1

Port 2

…

Port N

INT/SIMD/Vec ALU

(Slow/Fast) LEA (Load Effective Address)

(Fast or FP) Add

(128/256b, FP)-Mov

(Slow) Int

SI/INT/I/FP/F/Vec MUL

INT/SIMD/Vec Shift

(512) FMA

MulHi

Blend

SIMD misc/log

BM (Bit Ops)

Divider

FP or INT DIV

AMX

(Vec) Shuffle

Convert (CVT)

CRC32

GFNI (Galois Field New Ins.)

AES

POPCNT

SHA, SHA-MSG, SHA-RND

Branch (JUMP, Branch1/2, JEU)

Out of Order Execution Units (unstacked)

LD (Load)

STA (Store Address)

STD (Store Data)

Simple_AGU

AGU

Load Buffers

Store Buffers

Memory Execution Unit

Data Source

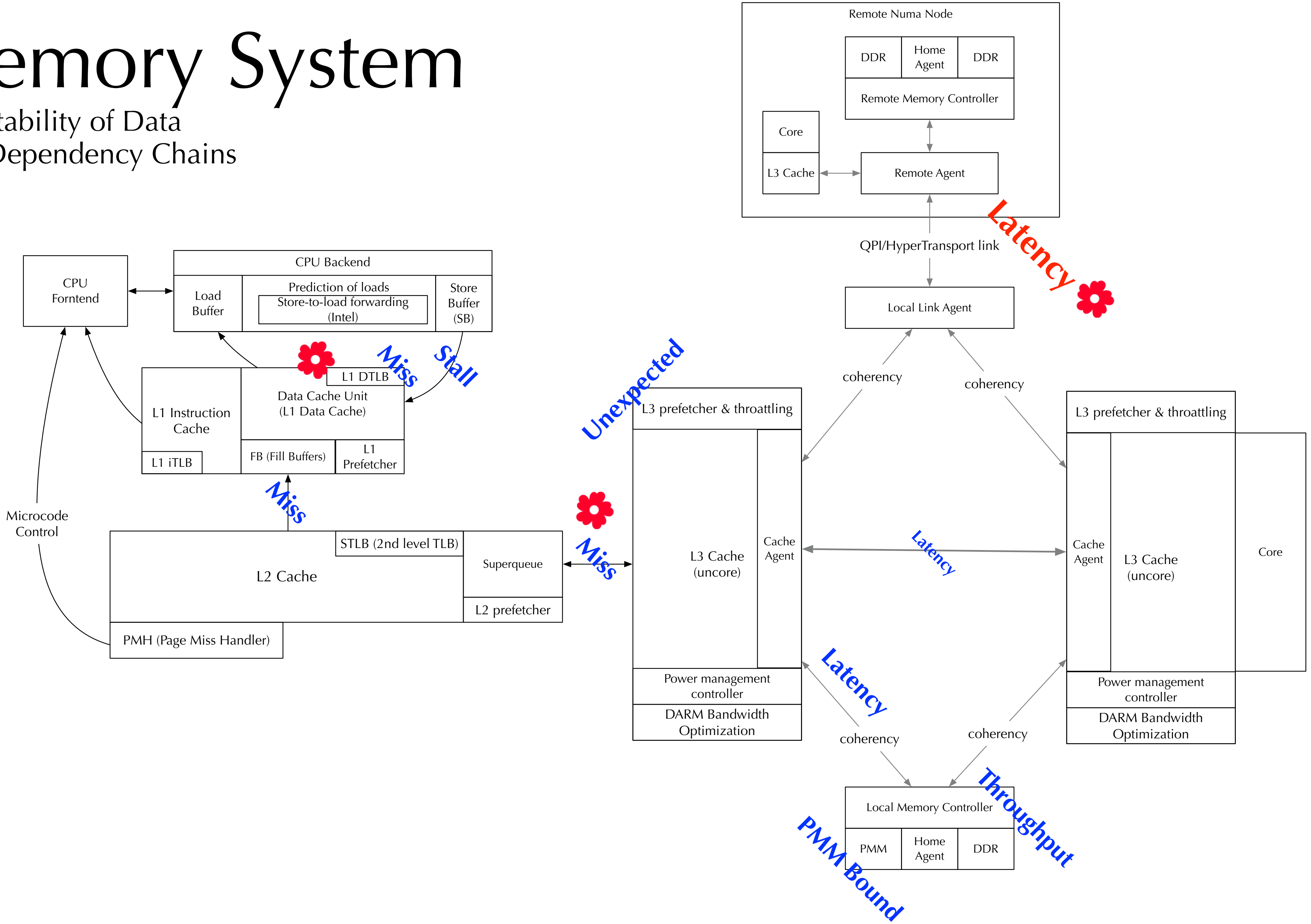
? Heavy.Microcode Sequencer

? Light.FP_Arith.FP_Scalar

? Light.FP_Arith.FP_Vector

Memory System

Predictability of Data
Data Dependency Chains


Remote Numa Node
DDR
Home Agent
DDR
Remote Memory Controller
Core
L3 Cache
Remote Agent
QPI/HyperTransport link
Latency
Local Link Agent
CPU Forntend
CPU Backend
Load Buffer
Prediction of loads
Store-to-load forwarding (Intel)
Store Buffer (SB)
Miss
Stall
L1 DTLB
Data Cache Unit (L1 Data Cache)
L1 Instruction Cache
L1 iTLB
FB (Fill Buffers)
L1 Prefetcher
Miss
Microcode Control
STLB (2nd level TLB)
L2 Cache
Superqueue
L2 prefetcher
PMH (Page Miss Handler)
Miss
Unexpected
coherency
coherency
L3 prefetcher & throattling
L3 Cache (uncore)
Cache Agent
Latency
Cache Agent
L3 Cache (uncore)
Core
L3 prefetcher & throattling
Power management controller
DARM Bandwidth Optimization
Power management controller
DARM Bandwidth Optimization
Latency
coherency
coherency
Throughput
Local Memory Controller
PMM
Home Agent
DDR
PMM Bound