Austin Physical Design and Integration Intern Project

Summer 2025 | Alec Bender | IBM z18 | Samsung 3nm

Life of an L3 Integrator

© 2023 IBM Corporation

What is the L3 Cache?

Summary

- CPU Cache memory
- Stores frequently accessed data
- Reduces data fetch time
- Comprised of SRAM

L3 Characteristics

- Speed: 2.8 GHz (half speed of core)
- Memory: 36MB Cache, 12 instances (432 MB total)
- Size: 7000 µm x 3000 µm, 12 instances (252 mm^2 total, 40% total chip area)

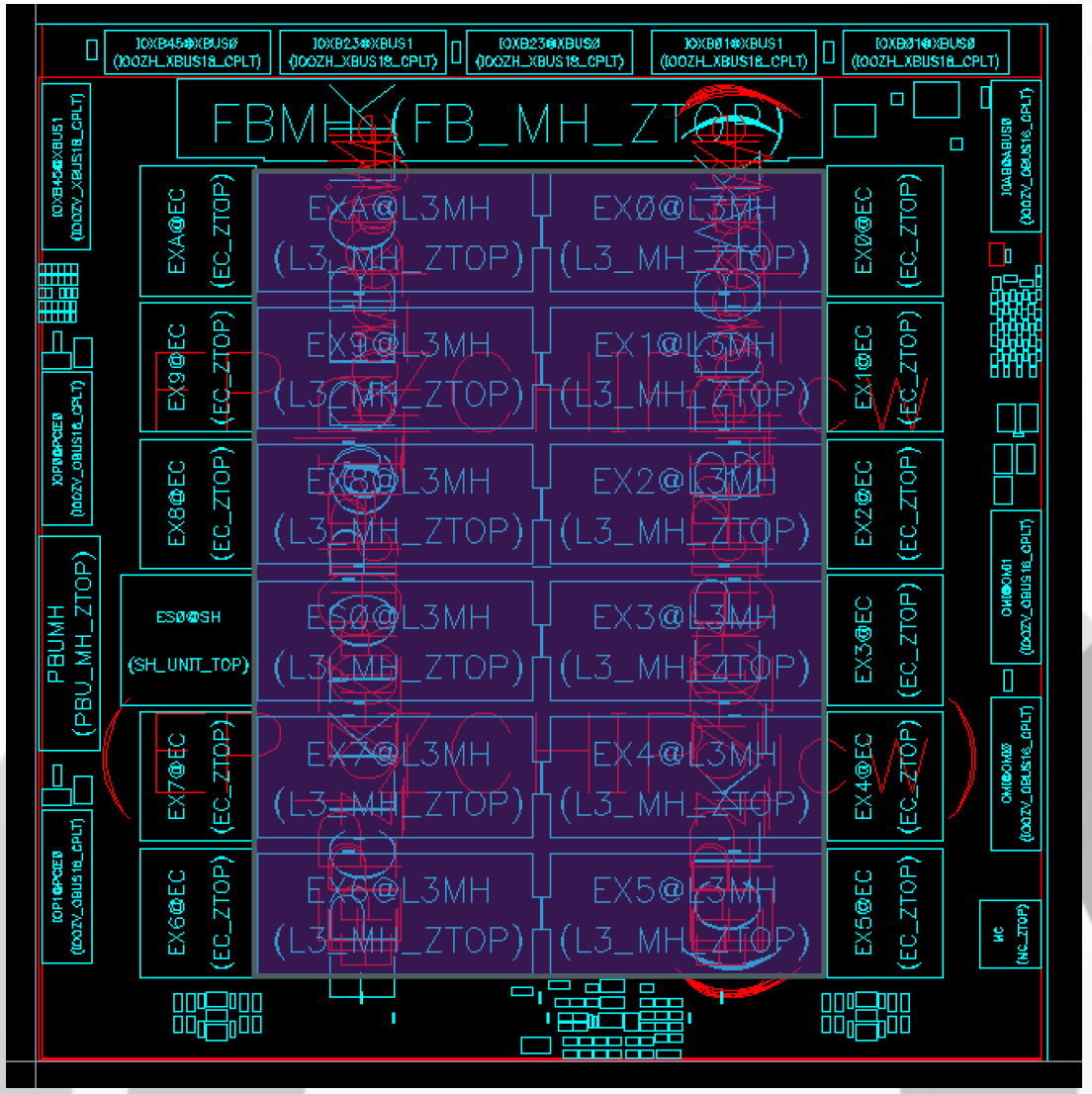


Figure 1: z18 chip top level (L3 cache highlighted), zrh10_ep_257021_0_T1_fs

© 2023 IBM Corporation

Goal of an L3 Integrator

Goal

- Continually reduce FOM until a timing wall is met
  - FOM (Figure of Merit) evaluates performance, related to timing and slack analysis
  - Timing wall is where PHYSICAL design improvements are negligible (LOGIC potentially can improve)

How

- Improve the generic build methodology for L3

© 2023 IBM Corporation

The L3 Integration Flow

© 2023 IBM Corporation

The "Flow"

MSC

- Methodology Script Cutting
  - Guide flow for dividing up small blocks (macros) and generating large blocks (continents)
  - Part of the S2L (Small to Large) methodology

S2L Methodology

- Uses Pangea
  - Fully abutted hierarchy
  - Separate continents grouped together by functionality or spatial relationships

MSF

- Methodology Script Flow
  - Produces populated Continents

Large Block MSI

- Methodology Script Integration
  - Guide flow for integration construction
  - Connects the constructed continents together

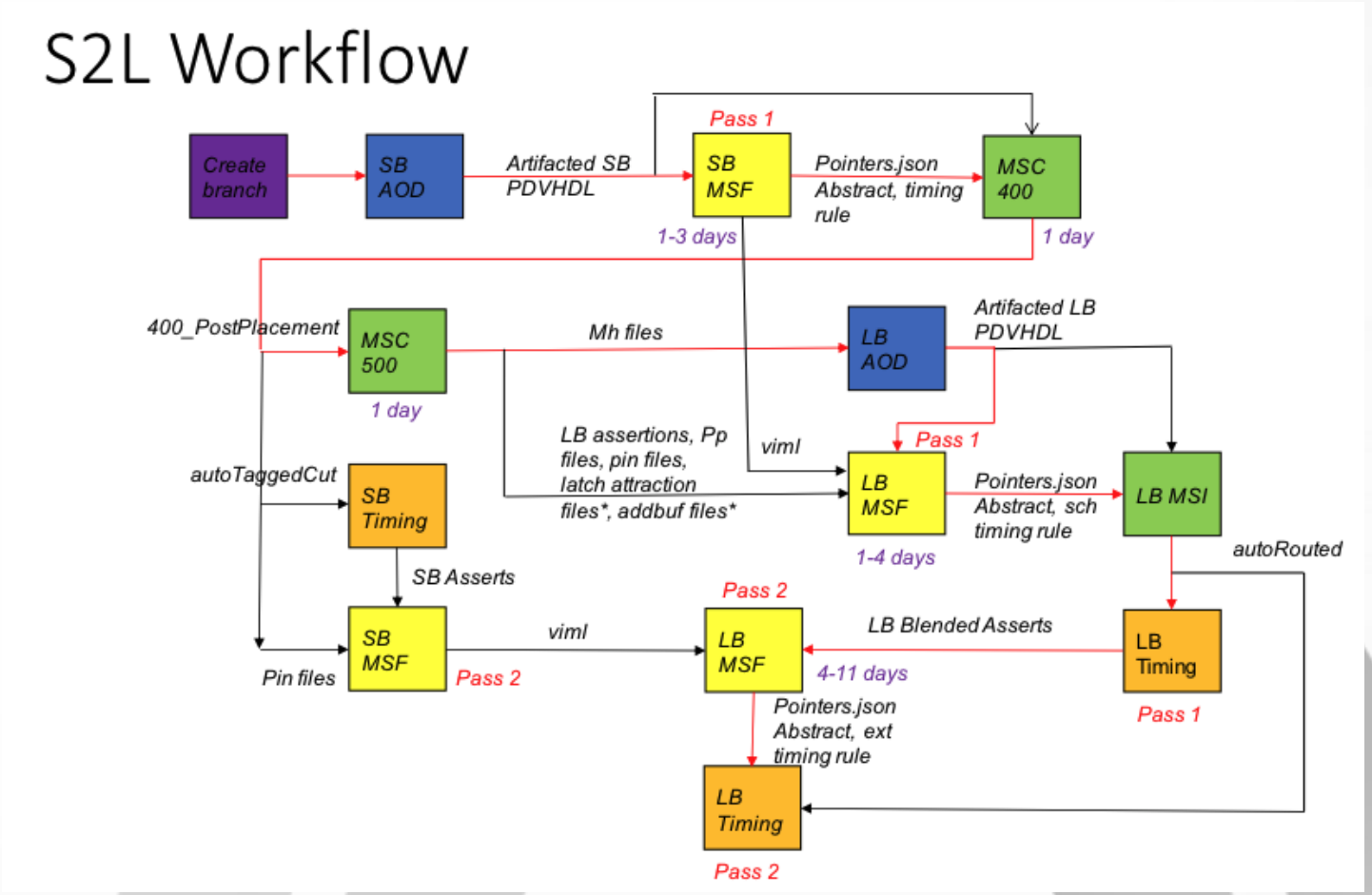


Figure 2: S2L Workflow, s2l_education.pptx by Jesse Surprise

© 2023 IBM Corporation

Knobs

- Collection of run IDs that control the behavior of a particular step or process within the flow
- Run IDs can be combined and are separated by a "."

Alec	L3D1	deleted	hlbs_latchopt.hlbs_parent.iclless.flat_dsa
Alec	L3D1	deleted	hlbs_parent.iclless.flat_dsa
Alec	L3D1	-	flat_dsa.flat_pfcc.hlbs_parent.iclless.tvis.250526v2
Alec	L3D1	-	flat_dsa.hlbs_parent.iclless.tvis.250526v2
Alec	L3D1	-	rrbucket.flat_dsa.flat_pfcc.hlbs_parent.iclless.tvis.250526v3
Alec	L3D1	-	alec.hideBx.flat_dsa.hlbs_parent.iclless.tvis.250526v4
Alec	L3D1	-	flat_dsa.flat_pfcc.hlbs_parent.iclless.tvis.250526v4
Alec	L3D1	COMP	flat_dsa.hlbs_parent.iclless.tvis.250526v4
Alec	L3D1	-	hideBx.flat_dsa.hlbs_parent.iclless.tvis.250526v4
Alec	L3D1	-	rrbucket.flat_dsa.flat_pfcc.hlbs_parent.iclless.tvis.250526v4

Figure 3: Example Runs with Various Run IDs

© 2023 IBM Corporation

Flattening

- Eliminates hierarchical boundaries pre-synthesis, giving the flow control in placing latches

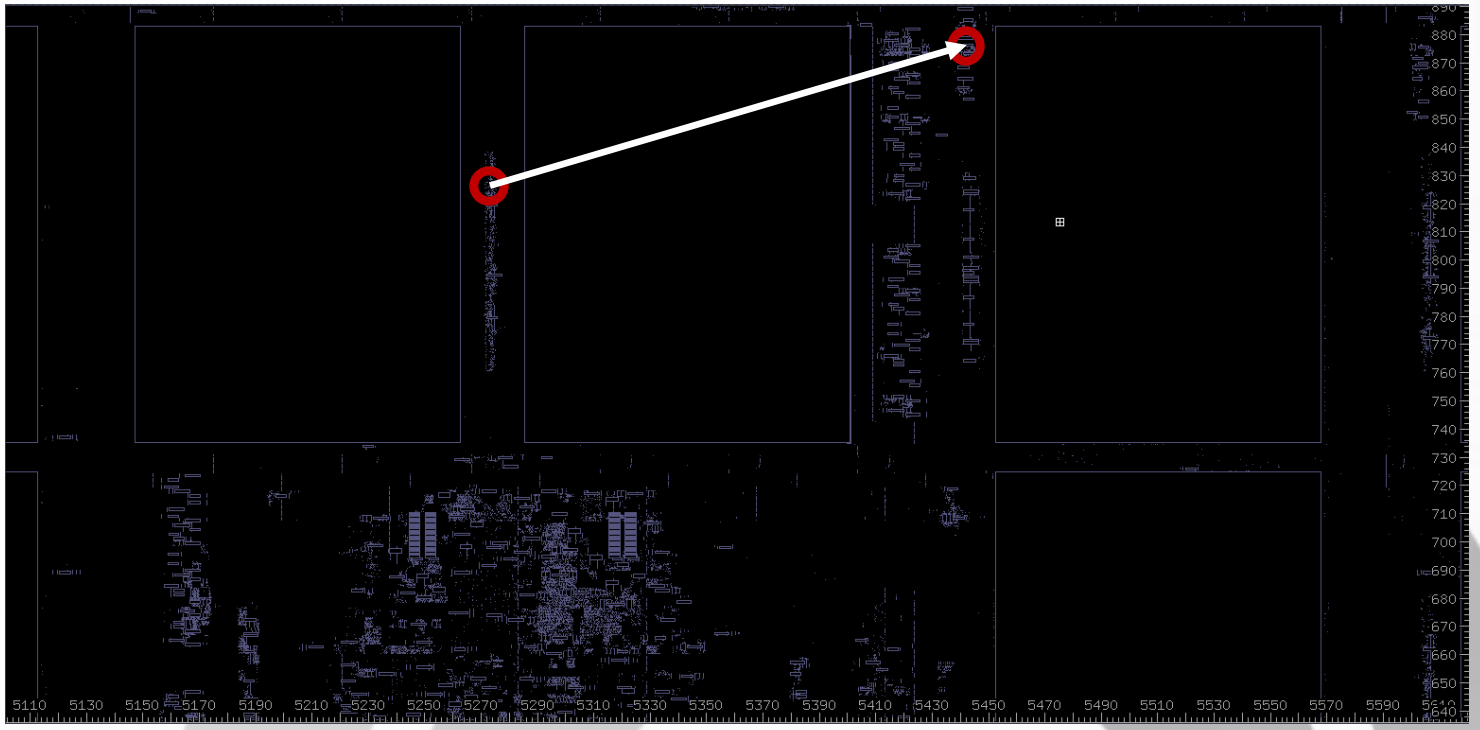


Figure 4: L3D1 Latch Flattened

© 2023 IBM Corporation

The "Flow"

MSC

- Serves as the input to cutting the large block model
- Cuts the continent and places macros in their respective continents
- Small block model allows for fast timing feedback
- Completes in a few days

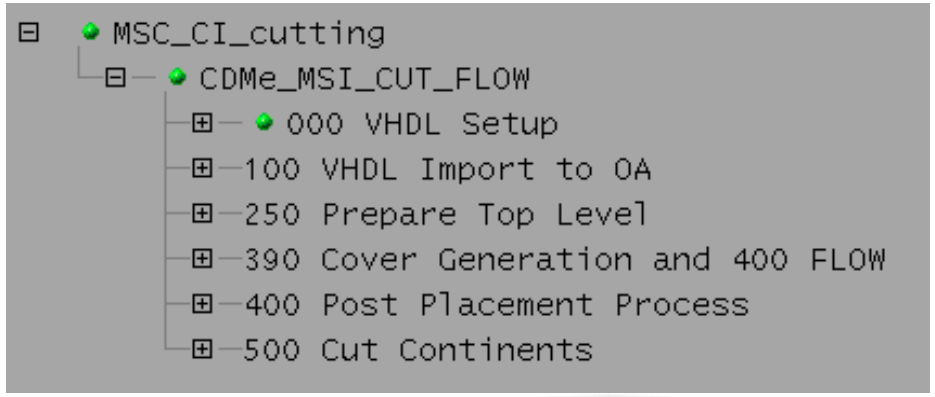


Figure 5: MSC Flow

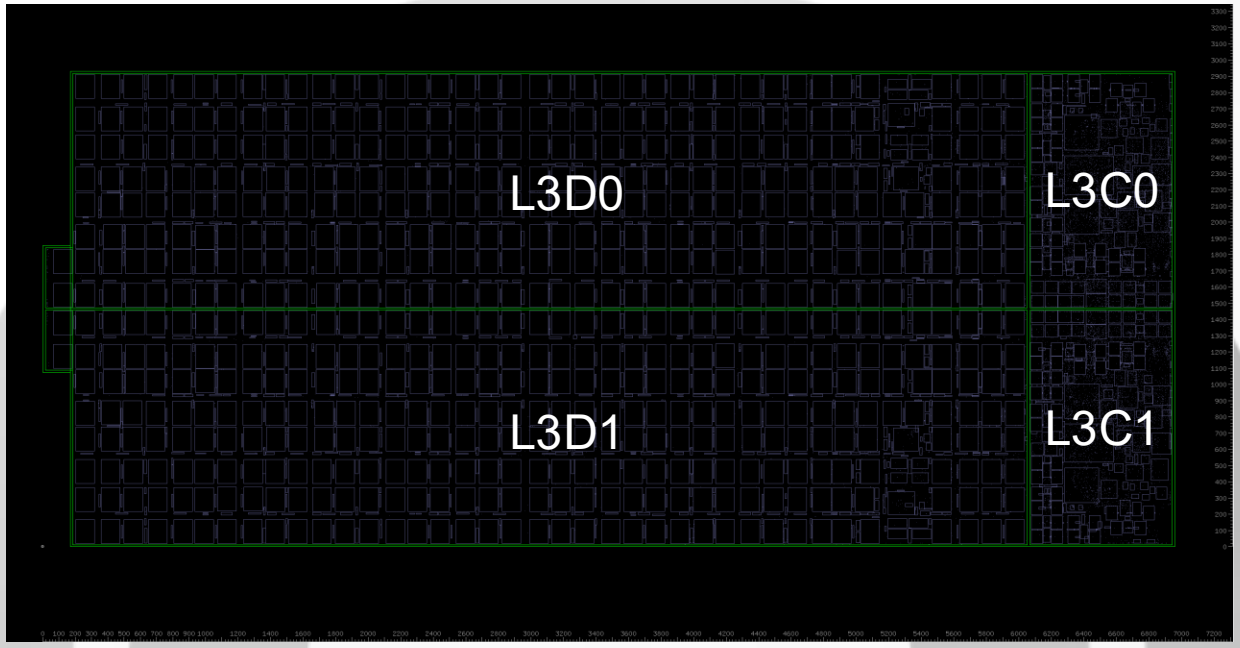


Figure 6: Cut Continents with Small Blocks, MSC Output

© 2023 IBM Corporation

The "Flow"

MSF

- HLBS (Hierarchical Large Block Synthesis)
  - Constructs continents and connects logic
  - Can utilize "flattening"
  - Completes in more than a week

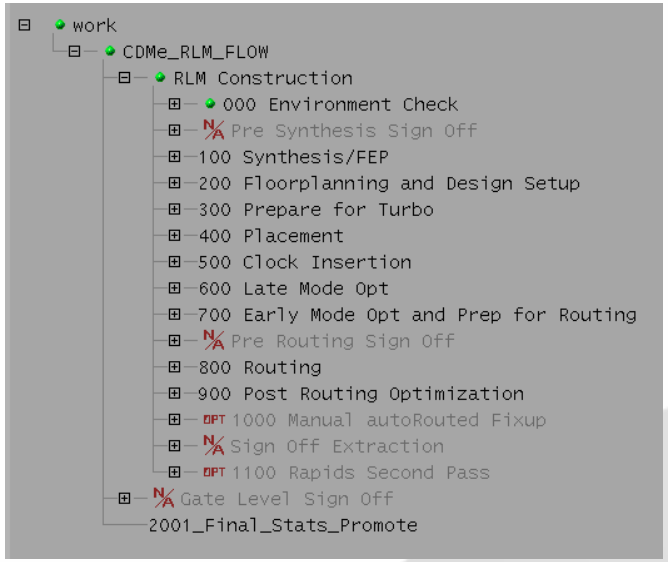


Figure 7: MSF Flow

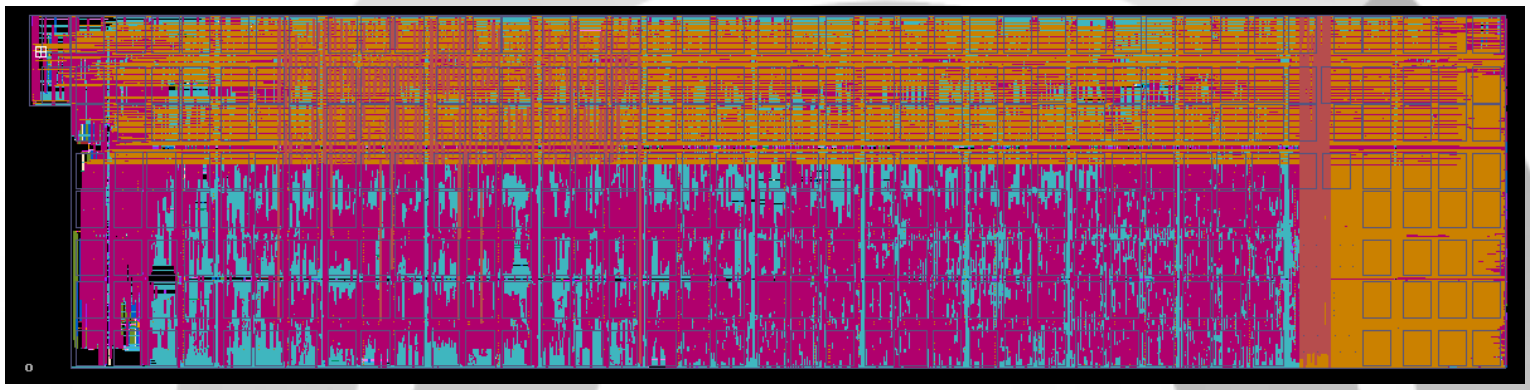


Figure 8: L3D1 Populated with Nets/Logic, MSF Output

© 2023 IBM Corporation

LBMSI

- Integrates continents back to a simplified top level
- Completes in about 6 hours

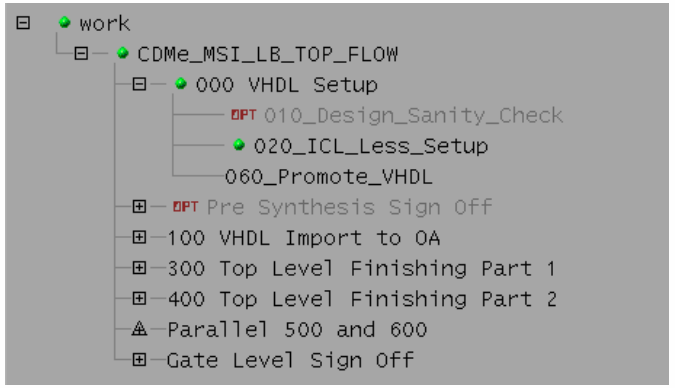


Figure 9: MSI Flow

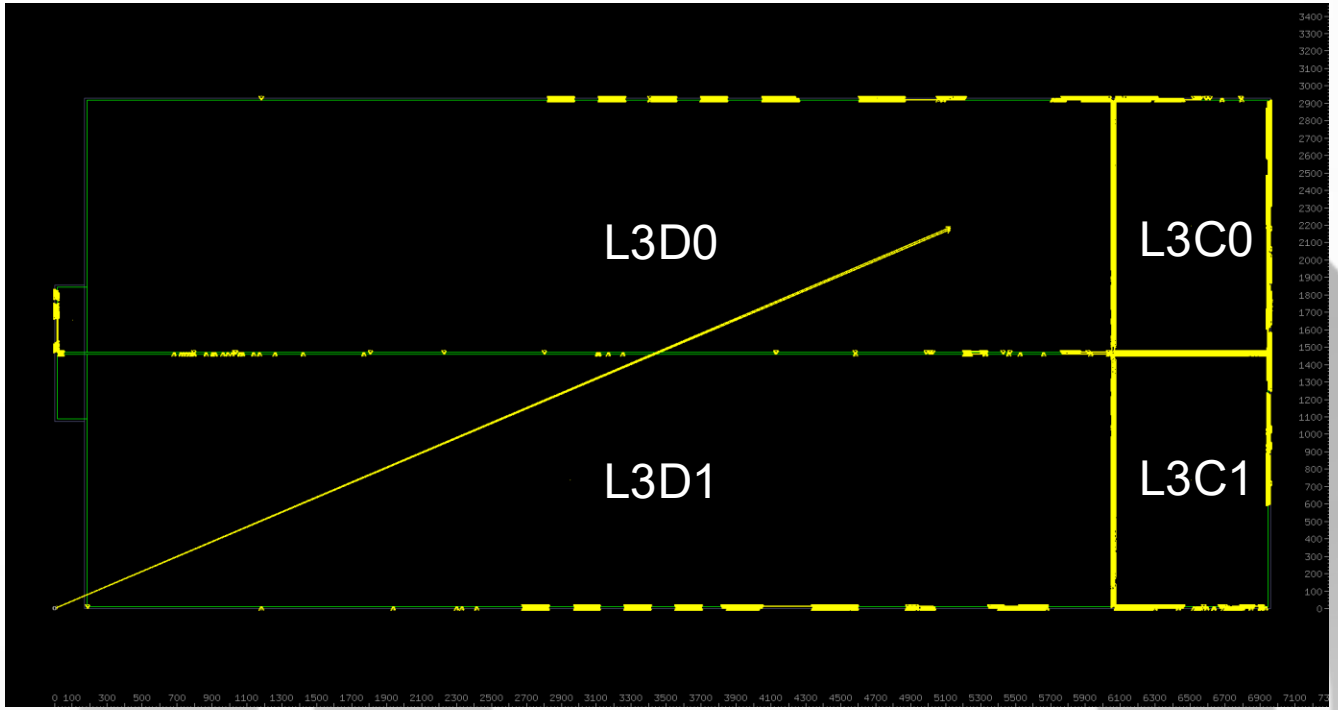


Figure 10: L3D1 Populated with Pins, MSI Output

© 2023 IBM Corporation

Contributions in Advancing the Flow

© 2023 IBM Corporation

Fixing Cross-Continent Net Failures

Problem

- Continent boundaries had multiple pins for the same net

Investigation

- Nets capable of reroute without double boundary crossing using iBuf

Solution

- Debugging from Pangea and parms updated

Figure 11: Net Double Crossing Continents (L3C1 → L3D1 → L3D0 → L3D1)

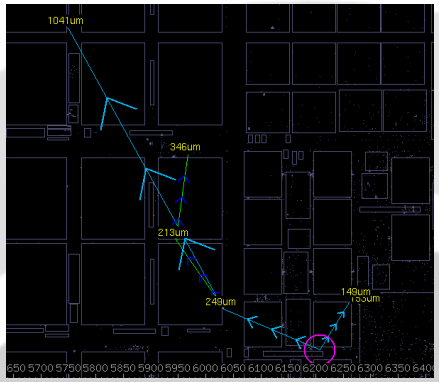


Figure 12: Net Fixed with iBuf (L3C1 → L3D1 → L3D0)

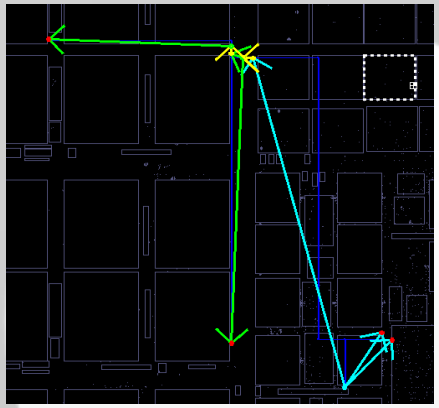


Figure 13: Net with Correct Behavior (L3C1 → L3D1 → L3D0)

© 2023 IBM Corporation

Investigating Routing Detours

Problem

- First continent builds
  - wSlack: -3.5k
  - FOM: -1.9M

Investigation

- Worst 40 nets had major routing detours
- Synthesis flattened buffers and routed limited fan outs in a point to point fashion
- Manual reroute in iBuf demonstrated significant improvement during the flow

Solution

- Did not allow flattening of these buffers
- After Fix
  - wSlack: -500 (down 86%)
  - FOM: -350k (down 81%)

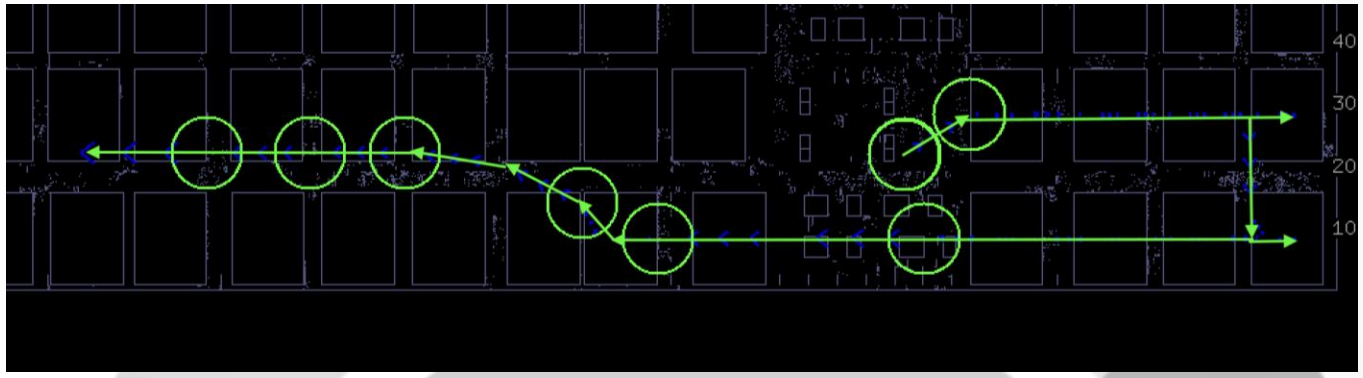


Figure 14: Net with Routing Detour

© 2023 IBM Corporation

Reducing B-Metal Congestion

Problem

- Massive increase in congestion (wACE4 95 to 140-270) during early optimization (step 620)

Solution

- EDA (pds-hlbs) suggested to hide B wires
  - Eliminating usage of highest available wire until necessary
  - New hideBx run ID

Results

- Significant decrease in final wACE4 and FOM
  - flat_dsa.hlbs_parent.icllless.tvis.250526v4
    - wACE4: 121 → 95
    - FOM: -1,710,142 → -952,964
  - hide45.flat_dsa.flat_pfncc.hlbs_parent.icllless.tvis.250707v2:
    - wACE4: 184 → 100
    - FOM: -729,851 → -541,098

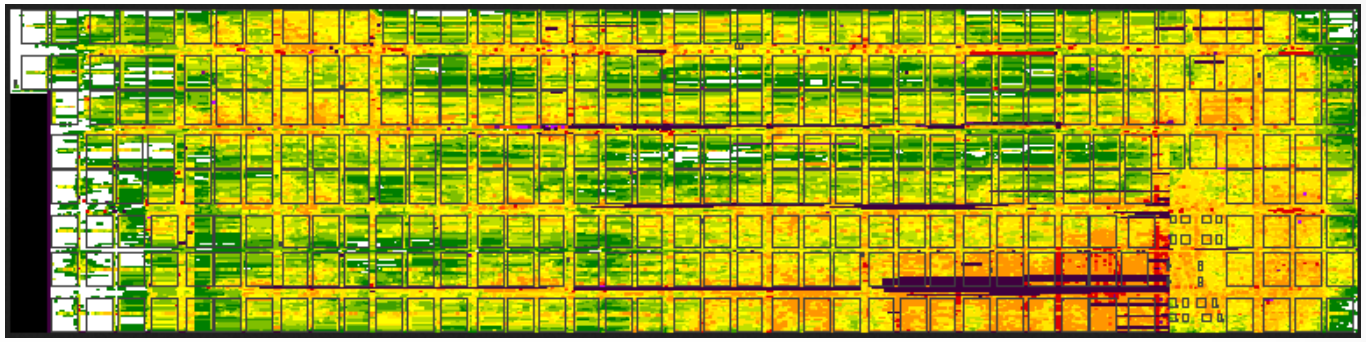


Figure 15: L3D1 Congestion Map (All Metal Layers)

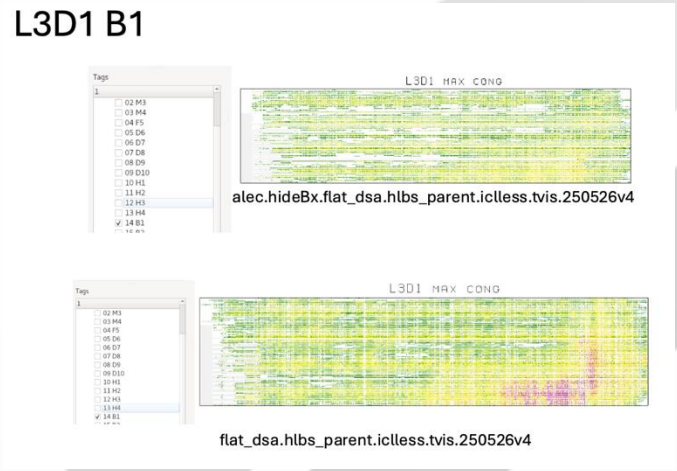


Figure 16: L3D1 Congestion Map (B1 Metal), w/ HideBx Top w/o HideBx Bottom

Figure 17: L3D1 Congestion Map (H3 Metal), w/ HideBx Top w/o HideBx Bottom

© 2023 IBM Corporation

Visualizing Latch Displacement after Flattening

Problem

- We do not necessarily know the scale that latches move during flattening

Solution

- Curated a Python script to summarize latch displacement
  - Parses .gz files, matches latches based on name and instance hierarchy, and computes coordinate differences

Results

- Output summary, csv, and histograms
- Showed significant movement much greater than hypothesized
  - Some latches moved as much as 8 SRAMs away

```
L3DSA_L21_P21_02_MH_ZTOP:
  Total Instances      : 18
------------------------------------------------------------

  L3@@L3D1@@L3DRW_DSA_L21_P21_02_1:
    Total Latches        : 166
      ├── Not Found      : 0
      └── Found          : 166
          ├── Flat       : 166
          ├── Static     : 0
          ├── ΔX Total   : 17916.672 µm
          └── ΔX/Latch   : 107.932 µm

              ├── → X Right      : 9096.864 µm
              ├── ← X Left       : 8819.808 µm
              ├── SRAM x →       : 30
              └── SRAM x ←       : 47
          ├── ΔY Total   : 16800.8 µm
          └── ΔY/Latch   : 101.2 µm

              ├── ↑ Y Up         : 0.0 µm
              ├── ↓ Y Down       : 16800.8 µm
              ├── SRAM y ↑       : 0
              └── SRAM y ↓       : 0
```

Figure 18: Summary Output Excerpt

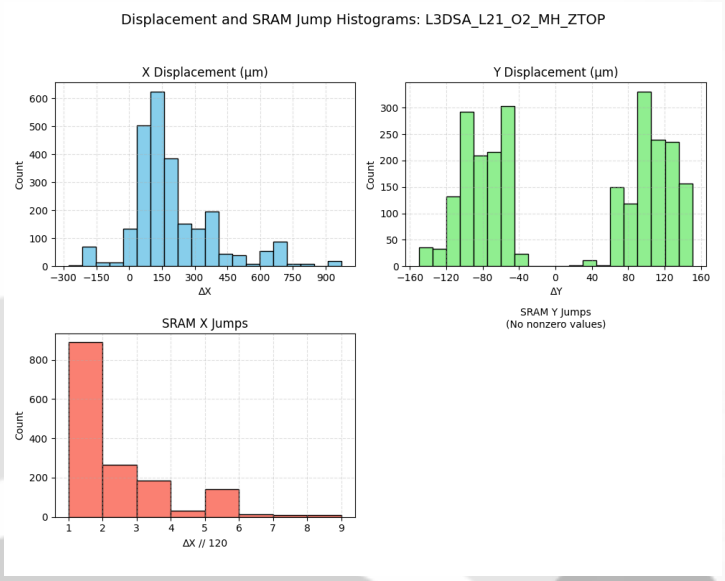


Figure 19: Latch Name Plot

Instance	default_x	default_y	flat_x	flat_y	dx	dy	dX_SRAM	dY_SRAM	Displacemer	Status	X_Dir	Y_Dir
L3@@L3D1@	467.52	638	1431.744	561.2	964.224	-76.8	8	0	967.278	Mismatch	Right	Down
L3@@L3D1@	467.52	638.4	1438.56	561.2	971.04	-77.2	8	0	974.104	Mismatch	Right	Down
L3@@L3D1@	467.52	638.8	1438.56	561.2	971.04	-77.6	8	0	974.136	Mismatch	Right	Down
L3@@L3D1@	467.52	638.8	1431.744	561.2	964.224	-77.6	8	0	967.342	Mismatch	Right	Down
L3@@L3D1@	467.52	638.4	1431.744	561.6	964.224	-76.8	8	0	967.278	Mismatch	Right	Down
L3@@L3D1@	467.52	637.2	1438.56	561.6	971.04	-75.6	8	0	973.978	Mismatch	Right	Down
L3@@L3D1@	467.52	636.8	1438.56	561.6	971.04	-75.2	8	0	973.947	Mismatch	Right	Down
L3@@L3D1@	467.52	636.4	1438.56	560.8	971.04	-75.6	8	0	973.978	Mismatch	Right	Down
L3@@L3D1@	467.52	636.8	1431.744	560.8	964.224	-76	8	0	967.215	Mismatch	Right	Down

Figure 20: CSV Output Excerpt

© 2023 IBM Corporation

Visualizing Latch Displacement after Flattening

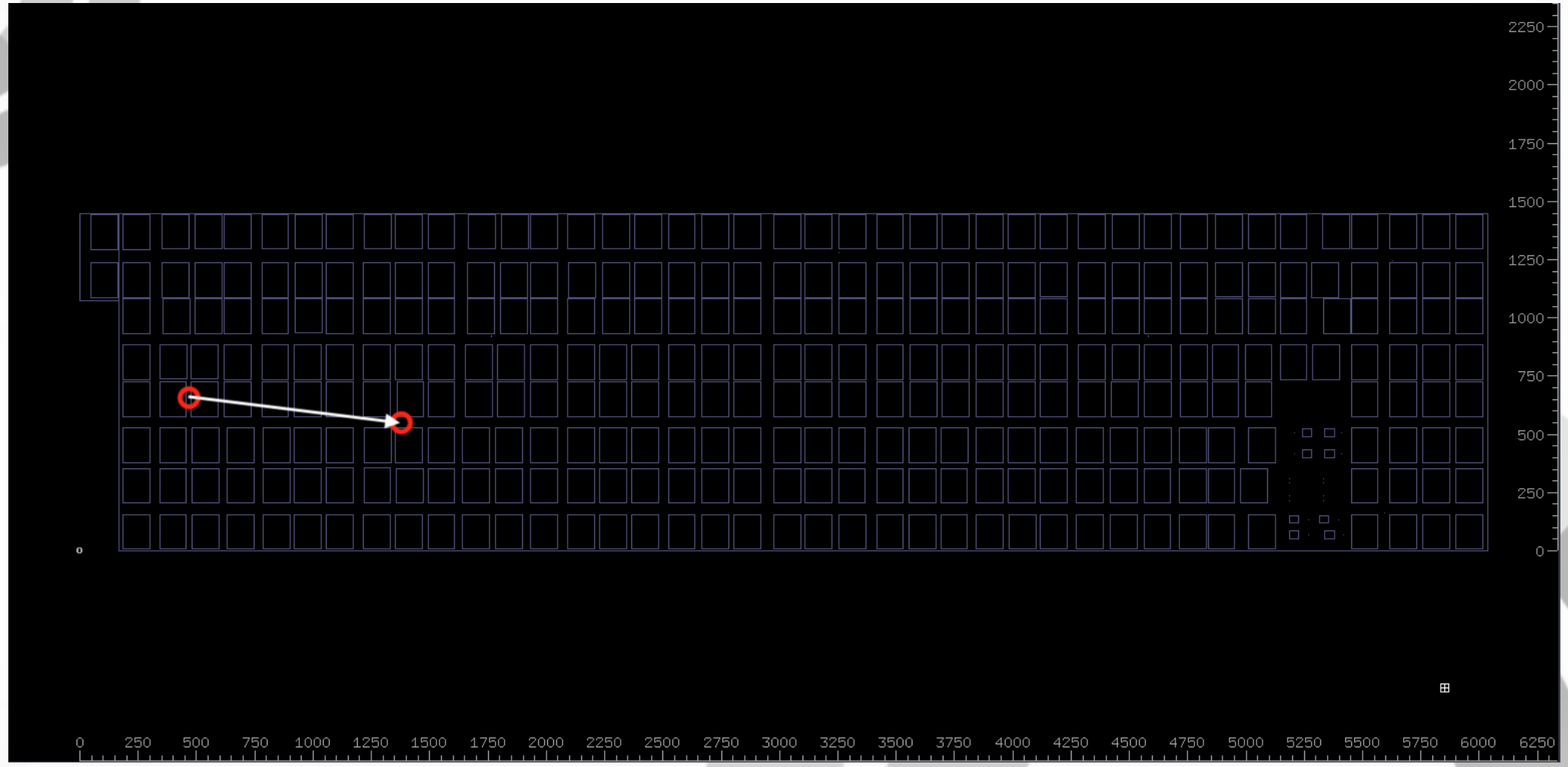


Figure 21: Net with Latch Displacement of 8 SRAM Widths

© 2023 IBM Corporation

Visualizing Latch Displacement after Flattening

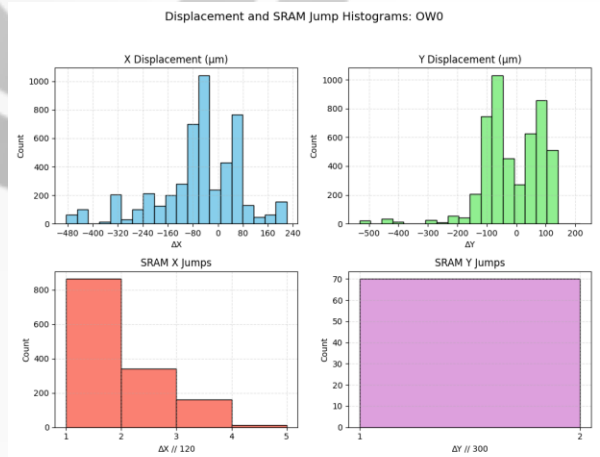


Figure 22: OW0 Plot

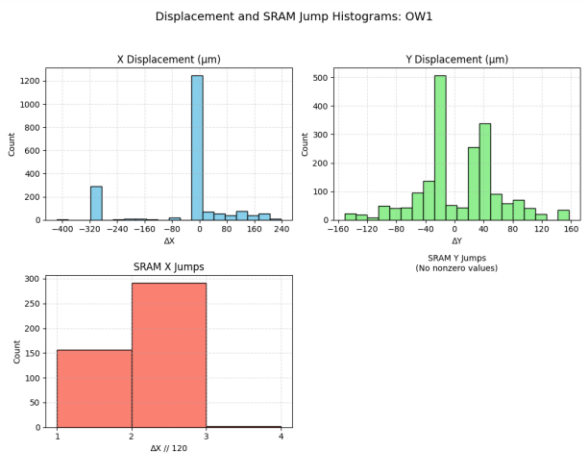


Figure 23: OW1 Plot

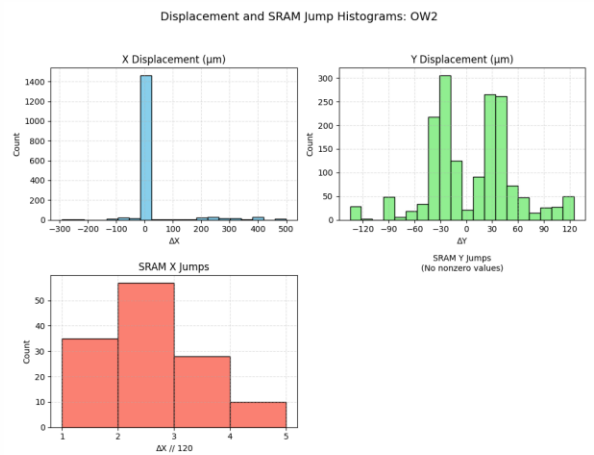


Figure 24: OW2 Plot

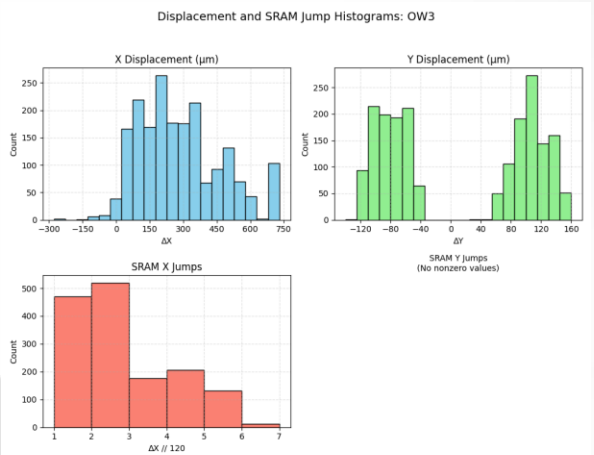


Figure 25: OW3 Plot

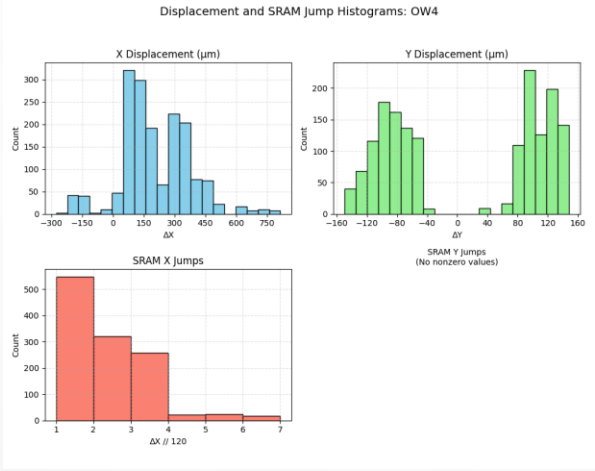


Figure 26: OW4 Plot

Figure 27: OW5 Plot

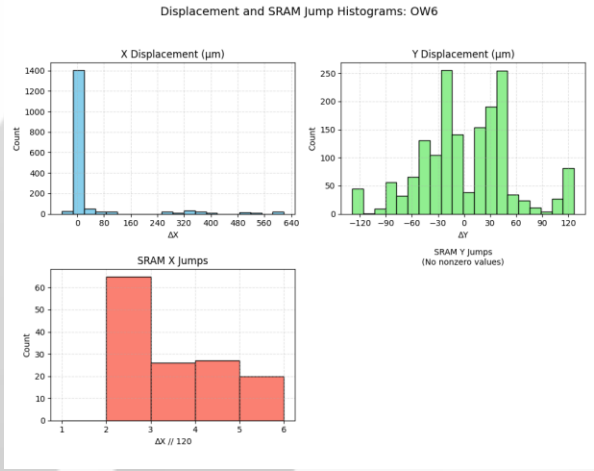


Figure 28: OW6 Plot

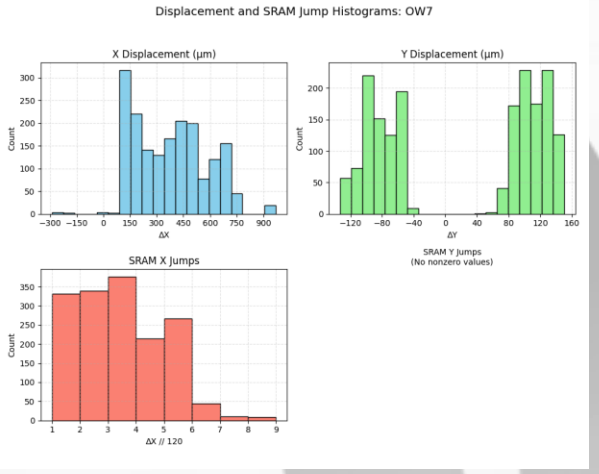


Figure 29: OW7 Plot

© 2023 IBM Corporation

Project Overview

© 2023 IBM Corporation

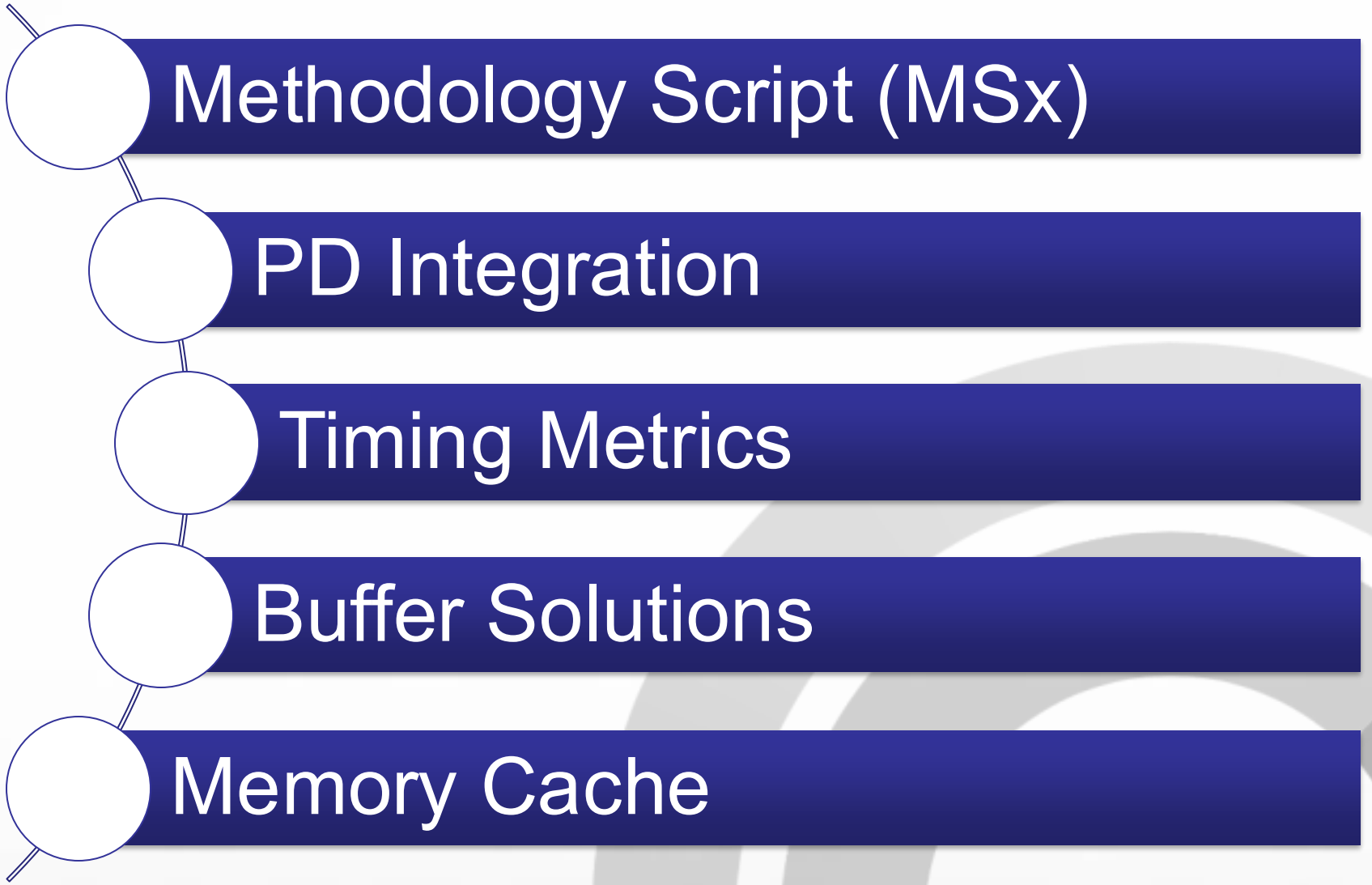
L3D1 Start (June 2025)

- FOM ≈ -1.9M
- wSlack ≈ -3.5k

L3D1 Current (August 2025)

- FOM ≈ -300k to -700k
- wSlack ≈ -200 to -400

© 2023 IBM Corporation



© 2023 IBM Corporation

Acknowledgements

Calist Friedman

Asher Lazarus

Jesse Surprise

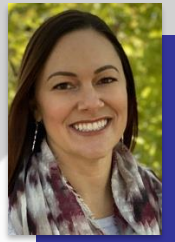
Chandler Brown

Jonathan Chor

Ben Stolt

David Yan

Malkam Hawkins



Nicole Strevig

© 2023 IBM Corporation

Questions

© 2023 IBM Corporation