

# RAINBOW FALLS

## Sun's Next Generation CMT Processor

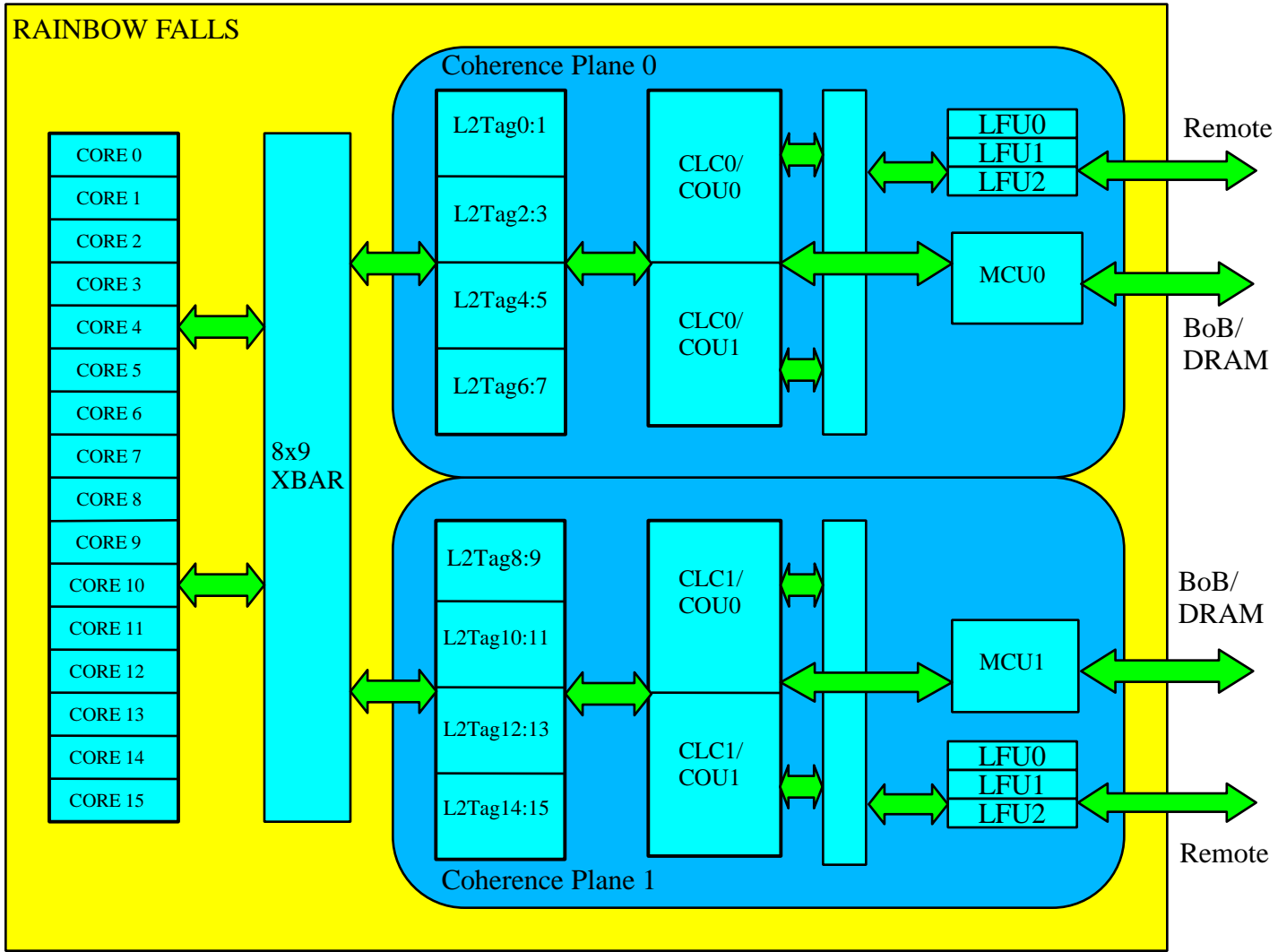
Presenter: Sanjay Patel

August 24, 2009

## Rainbow Falls (RF) at a High-Level

- Provides continuity to UltraSparc T2 and T2+.
- 16 cores in a chip.
- Glueless node-to-node coherent interconnect extended.
- Other generational improvements made.
- *Focus of this presentation is the challenge faced in integrating double the # of threads/cores and minimizing impact of supporting related bandwidth both on-chip and off-chip.*

# Rainbow Falls – Massive Data Flow across the Hierarchy



## Challenges facing Rainbow Falls

- On-Chip Connectivity
  - > Massive buses throughout.
  - > 16 cores interfacing to 16 L2 Banks.
  - > 4 Coherence Units, a set of 2 each interfacing independently to 3 high-speed links for remote node connectivity.
- Reducing Interface Count
  - > Performance requirements must be carefully balanced with need to reduce area.
  - > Core to CCX port reduction.
    - > *CCX is a crossbar with queuing that interfaces Cores to L2 Cache.*
  - > CCX to L2 Bank port reduction.

## Challenges facing Rainbow Falls (continued)

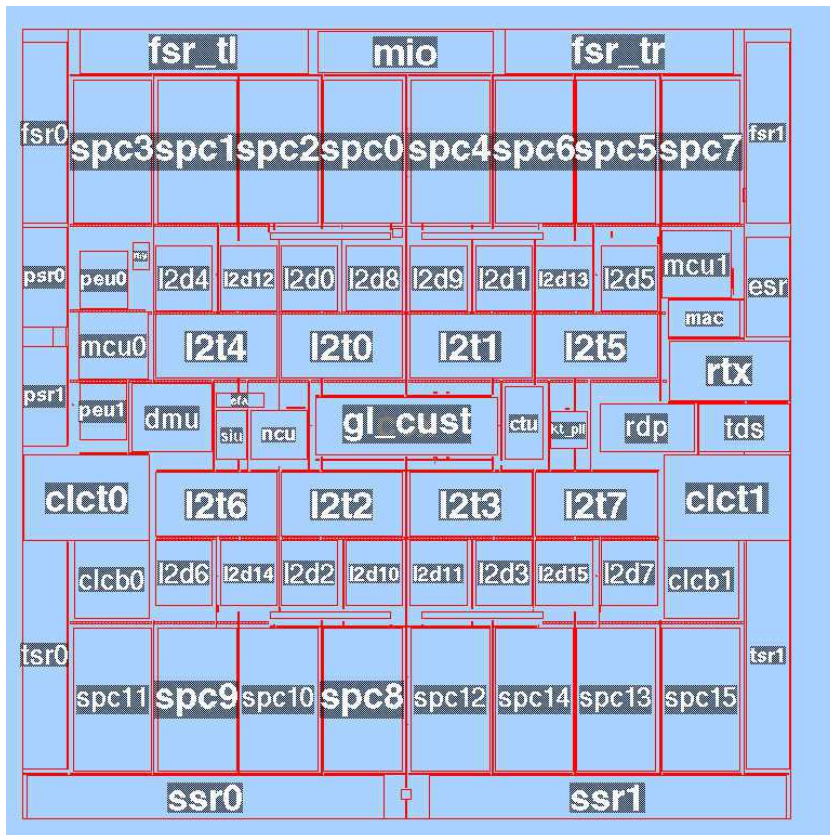
- Reducing Area of Larger Components.
  - > L2 Tags specifically.
    - > *L2 Cache is logically banked into 16 L2Banks, each with own tags (L2Tags) and data array (L2Data).*
  - > Merge multiple L2 buffers across a pair of L2Tag pipelines that were otherwise independent.
  - > Convert Directory from CAM to SRAM.
- Demands of Four Coherency Units
  - > *A Coherency Unit (COU) supports requests to a memory address whether homed local or remote, and maintains coherence across SMP cluster.*
  - > Greater link bandwidth requirements.
  - > Support for more outstanding transactions.
  - > Area cost becomes prohibitive unless dealt with upfront.

## Challenges facing Rainbow Falls (continued)

- Chip-to-Chip Connectivity
  - > RF Supports two coherence planes.
  - > Coherence Planes can be mirrored in adjacent chips to optimize board-level routing.
    - > *A Coherency Plane is a partition of the physical address along with associated resources – L2Banks, Coherency Units and Links.*
  - > Links can be flexibly configured, adding to possible optimizations at board-level.

*These specific challenges and how they are addressed will be discussed in upcoming sections.*

# Rainbow Falls – A Bird's Eye View



- Crossbar centric floorplan
- Optimize to reduce L2 Miss Latencies
- Large Pin Count to support greater Memory and Coherence Links.
- Balanced distribution of “hot spots” creates uniform thermal profile.

## Core to Cache Connectivity.

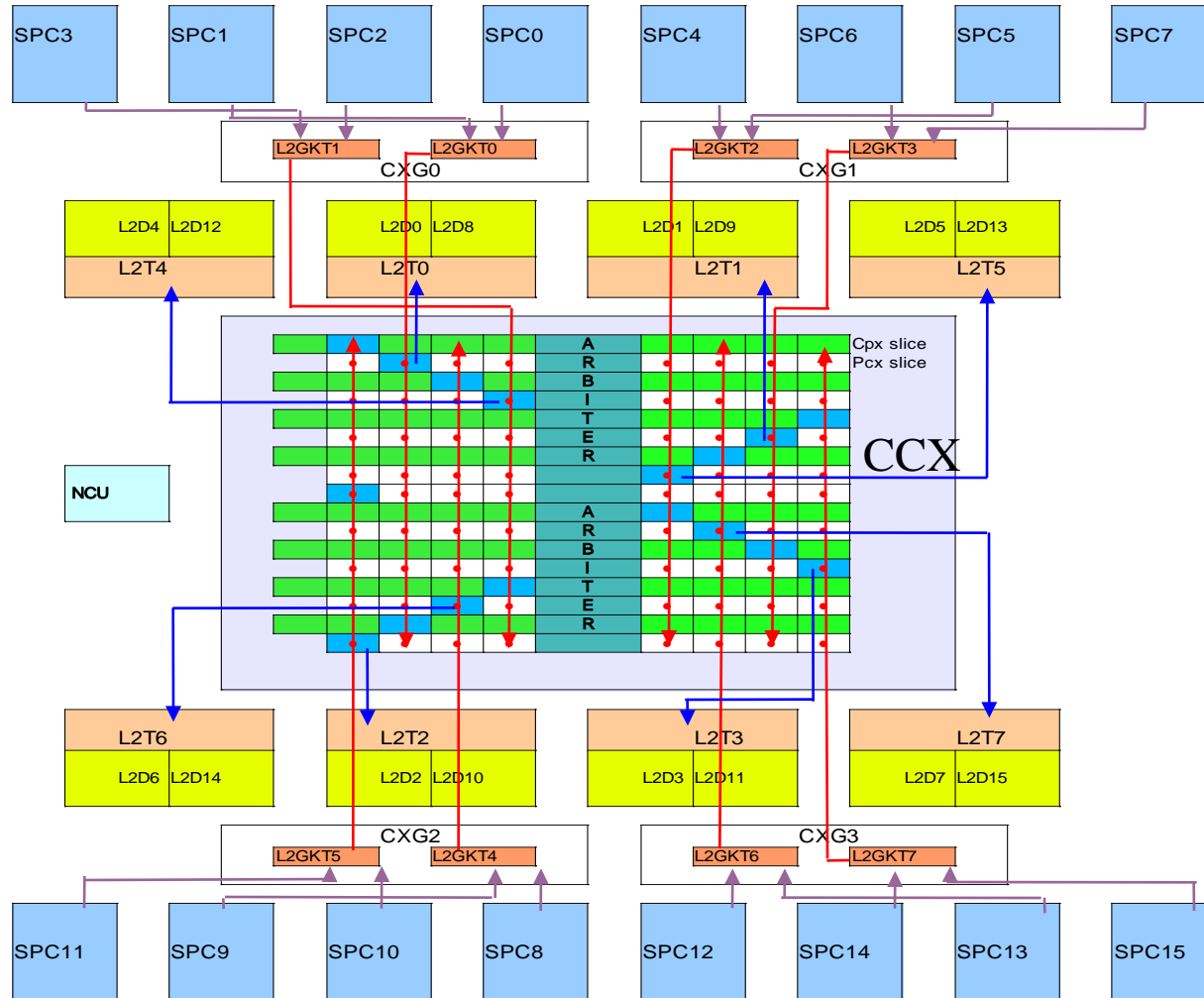
- Core to L2 Bank Crossbar(CCX)
  - > Buses are wide to support L1 cache-line requests – 32B bytes for L1I and 16B for L1D.
  - > Core requests are 135b wide, L2 Bank responses are 146b wide.
  - > Fully connected CCX means 16 request ports and 16 response ports.
  - > Functional tracks contest with Power and Clock grid.
  - > *Fully sizing makes CCX unroutable within allocated area or leads to blowout in area and poor latencies on L1 miss !*



## Core to Cache Connectivity (continued).

- Core to L2 Bank Crossbar(CCX) : Solution
  - > Two cores share common port to CCX through intermediate gasket.
  - > Gasket placement and queuing balanced to maintain high throughput on ingress and egress interfaces.
  - > L2Banks are paired up through a common L2Tag with dual-pipes to access two L2Data arrays but common input port.
  - > CCX differentiates between near and far ports to allow for asymmetric latencies internal to CCX.
  - > Result : 8x9 CCX with compact dimensions operating at highest chip frequency.

# Rainbow Falls – Centralized Crossbar Implementation (Request Side)



MUX ouput cell

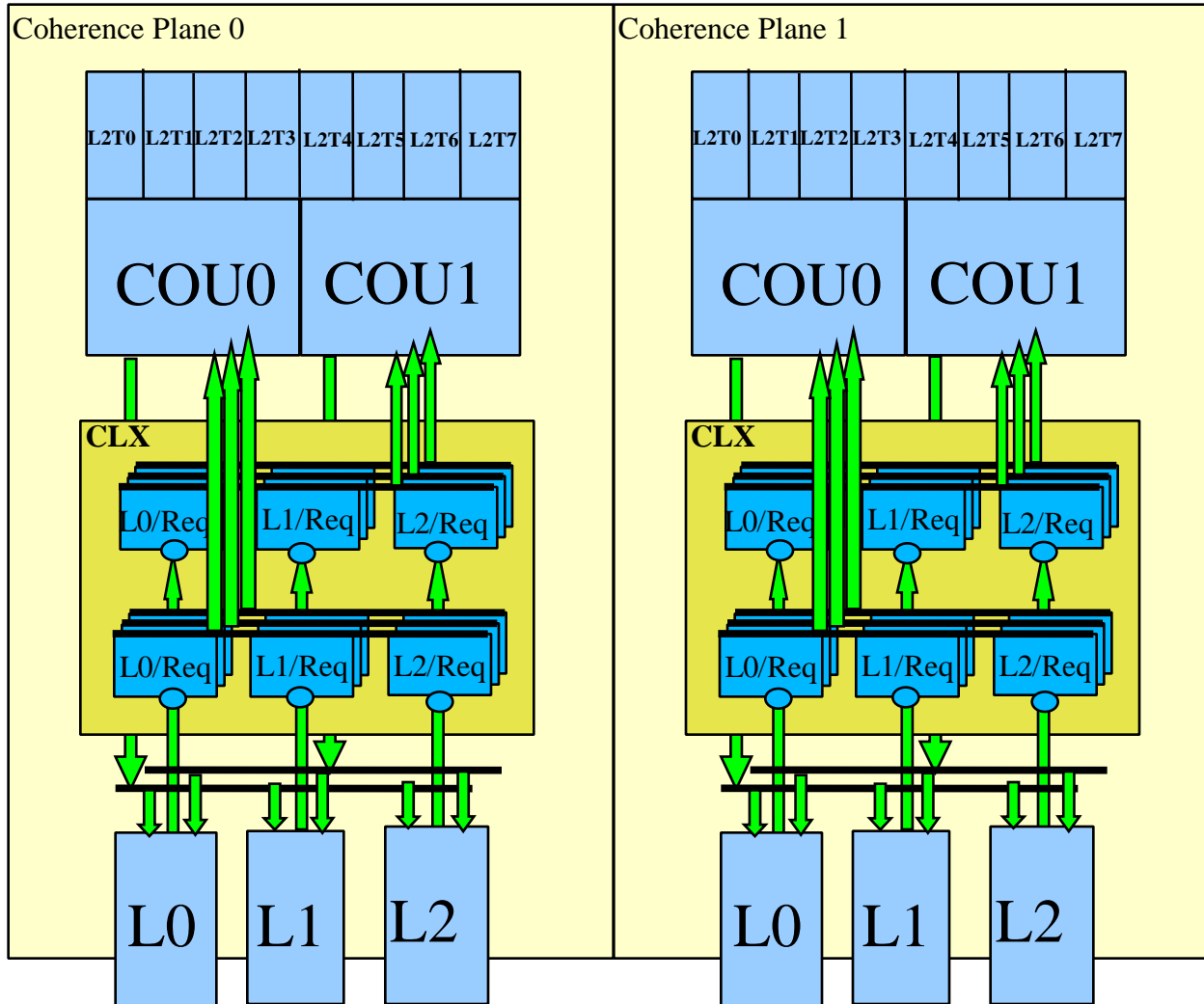
## Coherence Unit (COU) to Link Connectivity.

- Connectivity requirements vary based on whether configured for 2-way or 4-way cluster.
- Chip divided into two coherence planes each servicing 8 L2Banks.
- Remote connectivity in a plane provided through 3 high-speed links.
- COU to Link Crossbar (CLX) provides ingress queuing.
- Total # of Queues per Coherence plane – 3(links) x 2 (COUs) x 3(message types) = 18 total queues each up to 128 entries deep.
- Total # of outstanding transactions are up to 1024.
- *How does one limit impact of fully-sized queuing and guarantee functional correctness with undersized queuing ?*

## Coherence Unit (COU) to Link Connectivity (continued).

- Introduce Flow-Control across links : Solution
  - > Flow-Control is credit-based and bi-directional.
  - > Credits are updated as a transaction layer link message.
  - > Avoids the need to fully size queues, reducing area requirements.
  - > In-flight queue residence time reduced per transaction thus positively impacting FIT rate for transactions.
  - > Guarantees functional correctness – in-flight transactions are numerous and all possible combinations are impossible to verify. Instead, verify flow-control, a much simpler problem.

# Coherence Unit(COU) to Links Crossbar(CLX) : Two Coherence Planes



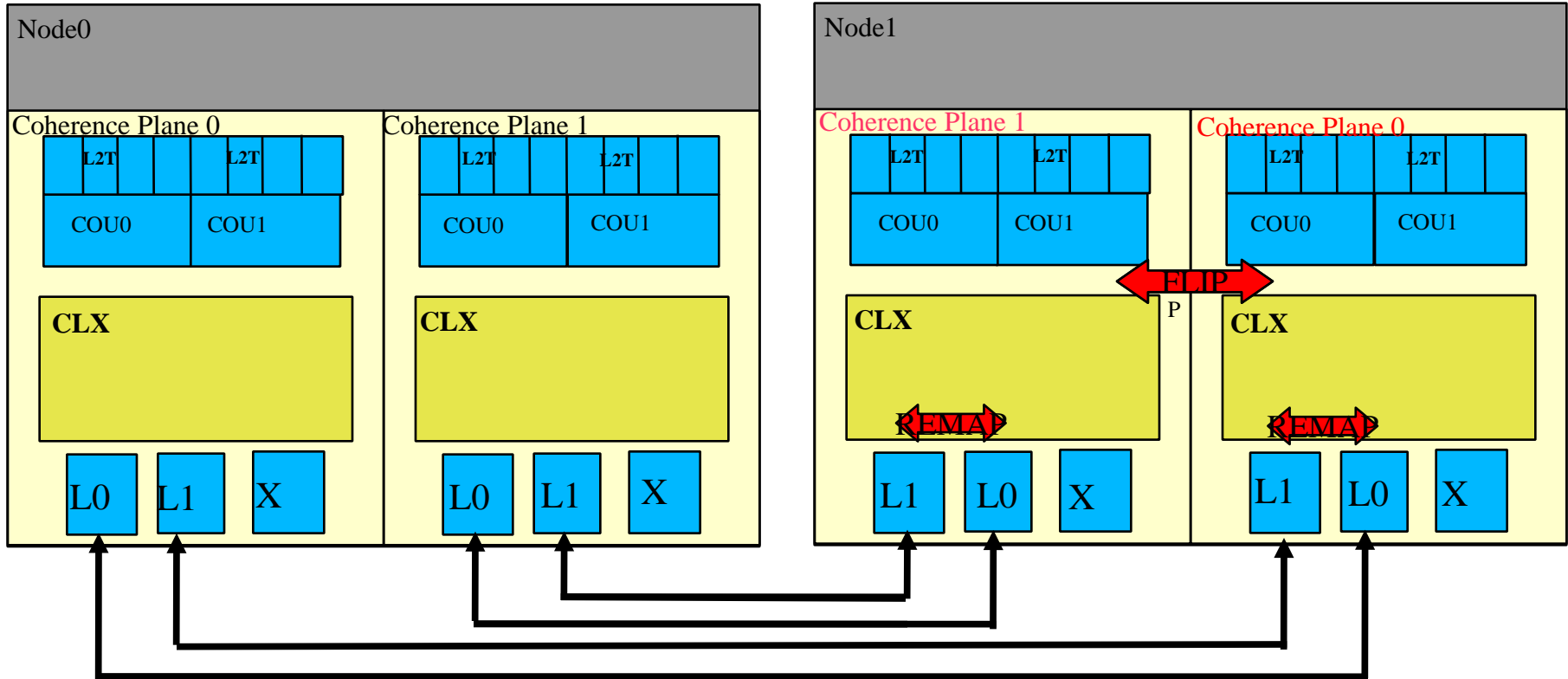
## Reducing Area of Larger Components.

- Focus L2Cache – 16 Banks and Larger Size to support double the # of threads. *How do we compensate ?*
- L2Bank consists of merged L2Tag with two L2Data Arrays.
- Common input port to L2Tag but two independent pipelines to access two L2Data arrays.
- Fill Buffer shared between two pipes in a single L2Tag.
- Copyback and Writeback buffers merged.
- L2 Directory, used to track L1 cache lines for inclusive nature, changed from CAM to RAM for power and area savings.
- *All steps together result in significant area savings.*

## Chip-to-Chip Connectivity

- In multi-node system, board-level routing needs to be factored into chip design.
- Otherwise routing is complicated. Further, flight-time impacts replay-on-CRC-error requirements for links.
- Configuration differs from system to system – Need complete configurability.
- Solution :
  - > Two Coherency Planes, in two vertical halves of chip, can be mirrored in left or right adjacent chip to minimize routing.
  - > Flexible mapping of links to COUs.

# Configuring Chip to Chip Connectivity



*This represents a hypothetical arrangement of complete non-overlap. With multiple layers of routing on board and other constraints, systems may be designed completely differently.*



## Summary

- Challenges exist in integrating more cores and more threads on chip. Coherent node-to-node interconnect adds another dimension.
- Rainbow Falls addresses various technical challenges in a unique manner that can be leveraged to future generations.
- With Rainbow Falls, Sun once again delivers next-generation CMT scaling on schedule.

## Acknowledgements

- Primary Authors :
  - > Sanjay Patel, Steve Phillips, Allan Strong.
- Contributors to design in presentation :
  - > Link Framing Unit : Bruce Chang, Suresh Thirumalai
  - > Coherence Unit : Yoganand Chillarige
  - > L2Cache : Prashant Jain, Srinivasan Iyengar.
  - > Crossbar : Sandip Das.
- Physical : Tim Johnson, Penny Li.
- Team Managers/Leads :
  - > Kenneth K Chan, Kalon Holdbrook, Tom Karabinas, Sanjay Patel, John Saba, Mike Schmidt, Gregory Smith.

Thank you ...

[sanjay.1.patel@sun.com](mailto:sanjay.1.patel@sun.com)