Processor - Features

- Branch Prediction → BHT (1 Bit), Global Predictor → 11 bit
- Decode crack and group formation
  - groups are dispatched in order
  - Results are committed when group finished

(MUCH MORE - NOT COVERED)
L1
- 2 state coherence: Valid, Invalid
- 2 way associative
- By definition, all data stored in L1 is available in L2.

L2
- L2 is implemented as 3 identical controllers (~5mb each).
- Cache lines are hashed across 3 controllers
- 8 way associative cache
- 4 *coherency processors* in each controller
- 4 *snoop processors* in each controller

L2 View

Coherency processor
- Controls the return of data from L2 (hit) or from the fabric (miss) to the core
- Updates L2 Directory
- Issues fabric commands on L2 misses
- Initiates invalidates to a processor which has some data in its L1 via the CIU resulting from a store in one core

Snoop Processor
- Sends back invalidate request to the cores to invalidate a cache line in its L1 data cache
- Read data from L2 Cache
- Update the cache lines directory state
- Issue a push operation to the fabric to write modified data back to memory
- Source data to another L2 from this L2

L2 Cache Coherence
- Enhanced Version of MESI with 7 states
- I (invalid state): data is invalid. Initial state
- SL (shared, can be sourced to L2 caches on the same module): entered as read request from L1 which resulted in a miss
- S: data cannot be sourced to another L2. Entered when prev state was SL and a read occurred
L2 Cache Coherence

- **M**: modified. Data can be sent to other L2 caches
- **Me**: exclusive. Data is owned, but not modified
- **Mu**: unsolicited modified. Entered when data comes from another L2 where it was in M state. Data is modified.
- **T**: tagged. Entered when a snoop read occurs in M state. Data will not be sent to other L2s until messages are received that no other L2 is sourcing the data

Sanity Check

<table>
<thead>
<tr>
<th>L2 State</th>
<th>L1 Data Cache</th>
<th>Send to Other L2s</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>L, V</td>
<td>L, S, S, T</td>
</tr>
<tr>
<td>S</td>
<td>L, V</td>
<td>L, S, T</td>
</tr>
<tr>
<td>M, Me or Mu</td>
<td>L, V</td>
<td>L, S, S, T</td>
</tr>
<tr>
<td>T</td>
<td>L, V</td>
<td>L, S, S, T</td>
</tr>
</tbody>
</table>

L3 cache (8 way associative)

L3 Coherence

- **I**: data is invalid
- **S**: shared
- **T**: tagged, data is modified wrt the memory copy, but may be in shared state in other L2 or L3s
- **Trem**: same as tagged, but data was sourced from memory attached to another chip
- **O**: data in L3 identical to memory. Data was sourced from memory attached to this L3. Status in other L2 or L3 caches is unknown.

Interconnect- 1 module

Interconnect-Multiple modules
Comparisons with sun fire

• Sun doesn’t have multiple levels of cache

• One core per processor

• 5 state protocol, extension of MESI

• ZhangHao’s presentation