CACHE CONTROLLER

Two level Cache Controller with Direct Mapped L1 Cache and Four way Set-Associative L2 Cache

RAGHAV GOYAL
18110135

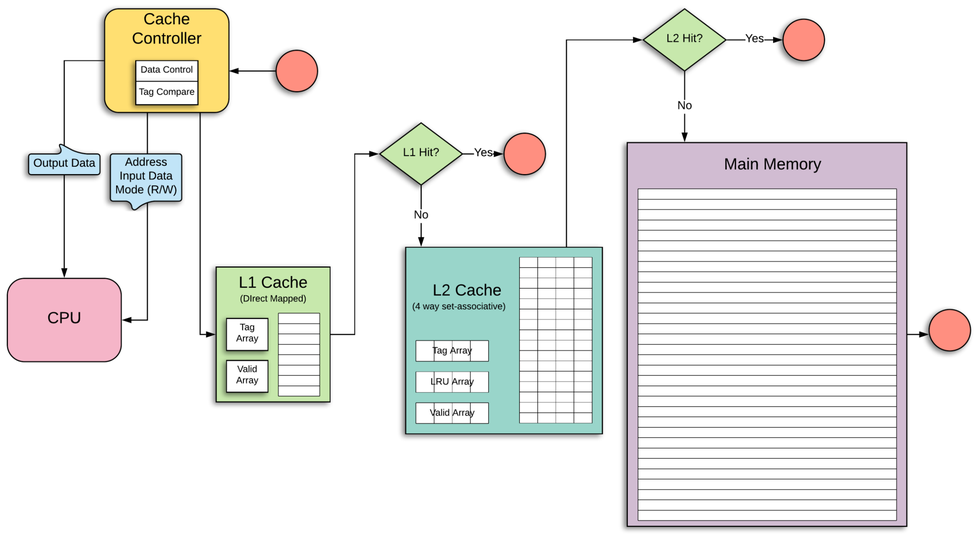
SACHIN YADAV
18110148

JAY RAHUL SHAH
18110154

VIVEK MODI
18110190

Indian Institute of Technology Gandhinagar

BLOCK DIAGRAM



Cache Controller is a hardware which acts as an intermediate between the processor and the cache memory. It executes the read and write requests from the processor and copies or replaces data within different levels of cache memory and main memory to reduce the average time taken by the processor to retrieve data from an address.

Here in this project, we have implemented a Cache Controller for two layers of Cache Memory. The L1 Cache Memory is Direct Mapped and the mapping used for L2 Cache is four-way set associative. In direct mapping, each line in the next memory level (L2 Cache here) is mapped to a specific line in the Cache (L1 Cache here). Direct Mapping being the simplest mapping policy is easiest to implement but has some disadvantages such as if two memory blocks which map to the same line in Cache are continuously referred, then the two blocks are continuously swapped in the Cache Memory.

Set-Associative Mapping is an enhanced form of Direct Mapping where the disadvantages of direct mapping are removed. Instead of mapping a block on a single line in Cache, it is mapped to a set (a group of lines). In case the set is full, then the least recently used block in Cache is removed to make space for the next block to be stored in Cache.

METHODOLOGY (ALGORITHM AND POLICIES USED)

WRITE-BACK POLICY

Optimizes system speed

Risk of data loss

In write-back policy, the data is updated in a cache level every time a write instruction is issued by the processor, but it is written into higher levels of cache or main memory only when the memory block evicts from the lower cache level.

NO-WRITE ALLOCATE POLICY

Fewer spurious evictions

Slower data read from write location

In no-write allocate policy, when a write miss occurs in a lower level of cache memory, the data is updated in the higher level of cache memory or in main memory (wherever found), but is not loaded into the lower level cache memory.

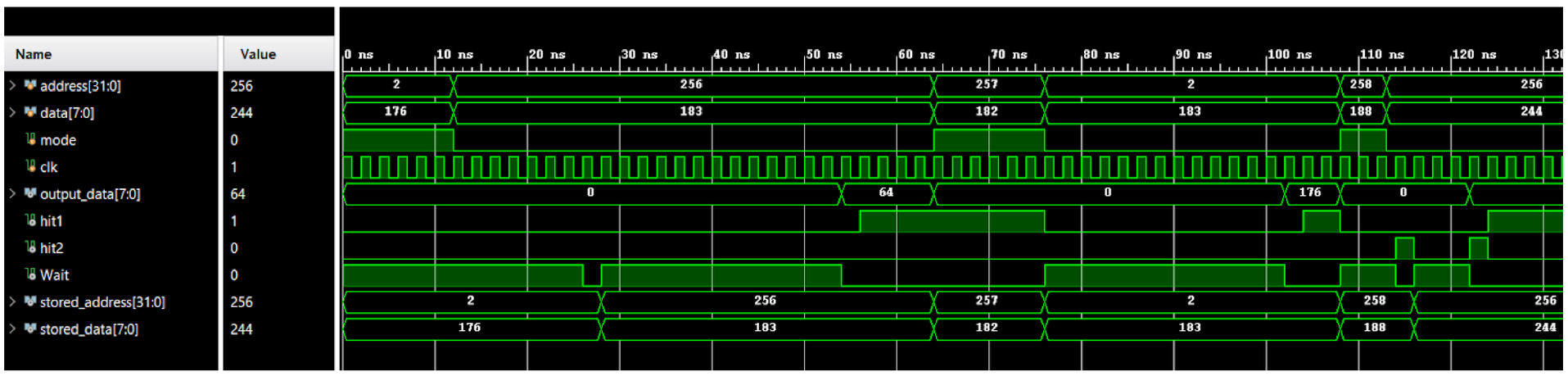
LEAST RECENTLY USED REPLACEMENT ALGORITHM

Minimizes unnecessary evictions

Time and space consuming

LRU Replacement Algorithm works on the idea that a block of memory that has been heavily used in the last few instructions is likely to be used again in the next few instructions. Thus, the memory block lying unused for the longest time is thrown out whenever required.

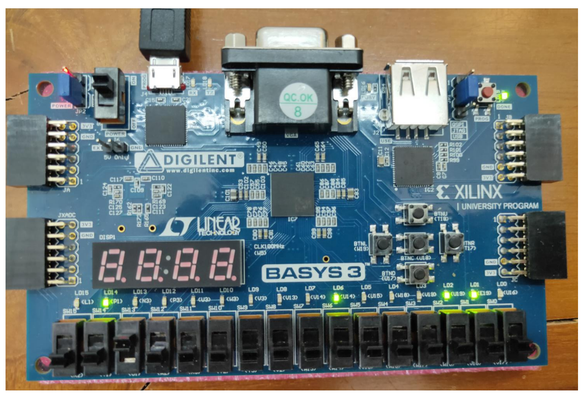
SIMULATION WAVEFORMS



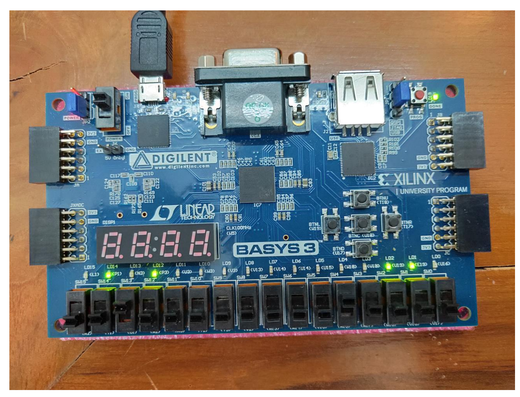
The sequence of the operations executed on the Cache Controller are:

1. Write at Block 0 (in main memory)
2. Read at Block 64 (in main memory, copied to L1 Cache thereafter)
3. Read at Block 64 (in L1 Cache)
4. Write at Block 64 (in L1 Cache)
5. Read at Block 0 (in Main memory, copied to L1 Cache replacing Block 64 thereafter)
6. Write at Block 64 (in L2 Cache)
7. Read at Block 64 (in L2 Cache, copied to L1 Cache thereafter)

IMPLEMENTATION ON BASYS 3 FPGA



Signal	Value
Address	00000000000 (11 bit)
Mode	0 (Read)
Output Data	0110 (4 bit)
Wait	0
L1 Hit	1 (Found in L1 Cache)
L2 Hit	0



Signal	Value
Address	00000010000 (11 bit)
Mode	0 (Read)
Output Data	0110 (4 bit)
Wait	1 (Processor in wait state)
L1 Hit	0 (Not Found in L1 Cache)
L2 Hit	0 (Not Found in L2 Cache)

SYNTHESIS REPORT

Site Type	Used	Fixed	Available	Util%
Slice LUTs*	6953	0	20800	33.43
LUT as Logic	6889	0	20800	33.12
LUT as Memory	64	0	9600	0.67
LUT as Distributed RAM	64	0		
LUT as Shift Register	0	0		
Slice Registers	1763	0	41600	4.24
Register as Flip Flop	1763	0	41600	4.24
Register as Latch	0	0	41600	0.00
F7 Muxes	220	0	16300	1.35
F8 Muxes	32	0	8150	0.39

- Number of Look Up Tables utilised = 6953
- Percentage of Look Up Tables utilised = 33.43%
- Number of Flip Flops Used = 1763
- Number of F7 Multiplexers used= 220
- Number of F8 Multiplexers used= 32

Design Timing Summary

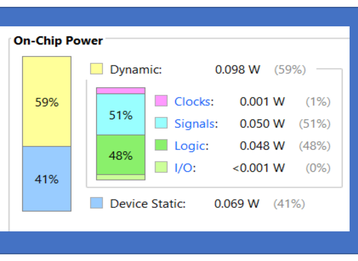
Setup		Hold		Pulse Width	
Worst Negative Slack (WNS):	4.852 ns	Worst Hold Slack (WHS):	0.265 ns	Worst Pulse Width Slack (WPWS):	4.500 ns
Total Negative Slack (TNS):	0.000 ns	Total Hold Slack (THS):	0.000 ns	Total Pulse Width Negative Slack (TPWS):	0.000 ns
Number of Failing Endpoints:	0	Number of Failing Endpoints:	0	Number of Failing Endpoints:	0
Total Number of Endpoints:	64	Total Number of Endpoints:	64	Total Number of Endpoints:	34

All user specified timing constraints are met.

- Worst Negative Slack = 4.852 ns
- Worst Hold Slack = 0.265 ns
- Worst Pulse Width Slack =4.500 ns

Power analysis from Implemented netlist. Activity derived from constraints files, simulation files or vectorless analysis.

Total On-Chip Power:	**0.167 W**
Design Power Budget:	**Not Specified**
Power Budget Margin:	**N/A**
Junction Temperature:	**25.8°C**
Thermal Margin:	59.2°C (11.8 W)
Effective θJA:	5.0°C/W
Power supplied to off-chip devices:	0 W



- Total On-Chip Power = 0.167 W
- Junction Temperature = 25.8^{0} C