

Our superscalar reservation station has 2 RS banks, each handles 1-way dispatch and 1-way issue. The size of RS bank is 8 entries so in total we have 16 entries in our superscalar RS. Our ROB has 32 entries and each entry will record PC, exception, halt, Tag_old, Tag_new, inst, rd/wr_mem, branch/predict_taken and other necessary information.

In our EX stage, we have 2 ALU, 2 8-stage MULT and 2 Branch resolution units. We can handle 2 branches at the same time and take the right result. Also the EX stage can send 2 load or store instructions to LSQ and the LSQ can solve 2 instructions at the same time. However due to the memory's bandwidth requirement, we can only send one load or store request to memory so we added two buffers between LSQ and Dcache.

- **branch prediction enhancement**

Our branch prediction includes three components: direction predictor (DIRP), branch target buffer (BTB) and return address stack (RAS).

1. DIRP: we designed a gshare direction predictor with global branch history table and per-address pattern history table. The size of PHT is 32 vertically and 32 horizontally, which means we use 5-bit of branch PC address to determine which pattern history table we use for each branch and 5-bit index to determine which pattern history state we use for predicting direction.
2. BTB: The size of our branch target buffer is 32 as well. The structure of BTB is essentially a direct mapped cache, in which the tag is the branch PC address and the data is the corresponding target PC address. We use lower [6:2] bits of branch PC address to index BTB.
3. RAS: In addition to the above features, we also added a return address stack(size is 16) and a predecoder. Predecoder is used to determine whether an instruction is a call PC instruction (JAL) or a return PC instruction (JALR). Every time when an instruction is pre-decoded as a JAL, it will be pushed into RAS or when an instruction is pre-decoded as a JALR, RAS will pop out the target return PC address.

- **load-store processing enhancement**

To resolve load-store dependency, we added four components: store queue, load queue, load buffer, post-retirement store buffer.

1. Store queue: we use store queue to orderly record store instructions and add age logic to detect if store queue can forward value to a load instruction.
2. Load queue: load instructions are orderly recorded in the load queue. Also we add age logic in the load queue to detect if a load instruction executes early before store and cause memory violation. If it is , we set a memory violation bit for corresponding load instruction in ROB. When it's ready to retire, we roll back the whole system.
3. Load buffer: if load needs to get value from Dcache or memory, we push it into the load buffer and Dcache and memory can handle these read requests sequentially. Once Dcache returns the value, we store it in the load buffer therefore the common data bus(CDB) can read the load instruction from the load buffer for broadcast.
4. Post-retirement store buffer: Since dcache and memory need multiple cycles to write value. So after store instruction retires, we put it into the store buffer. So the cache and memory can handle those write requests sequentially.

- **memory system enhancement**

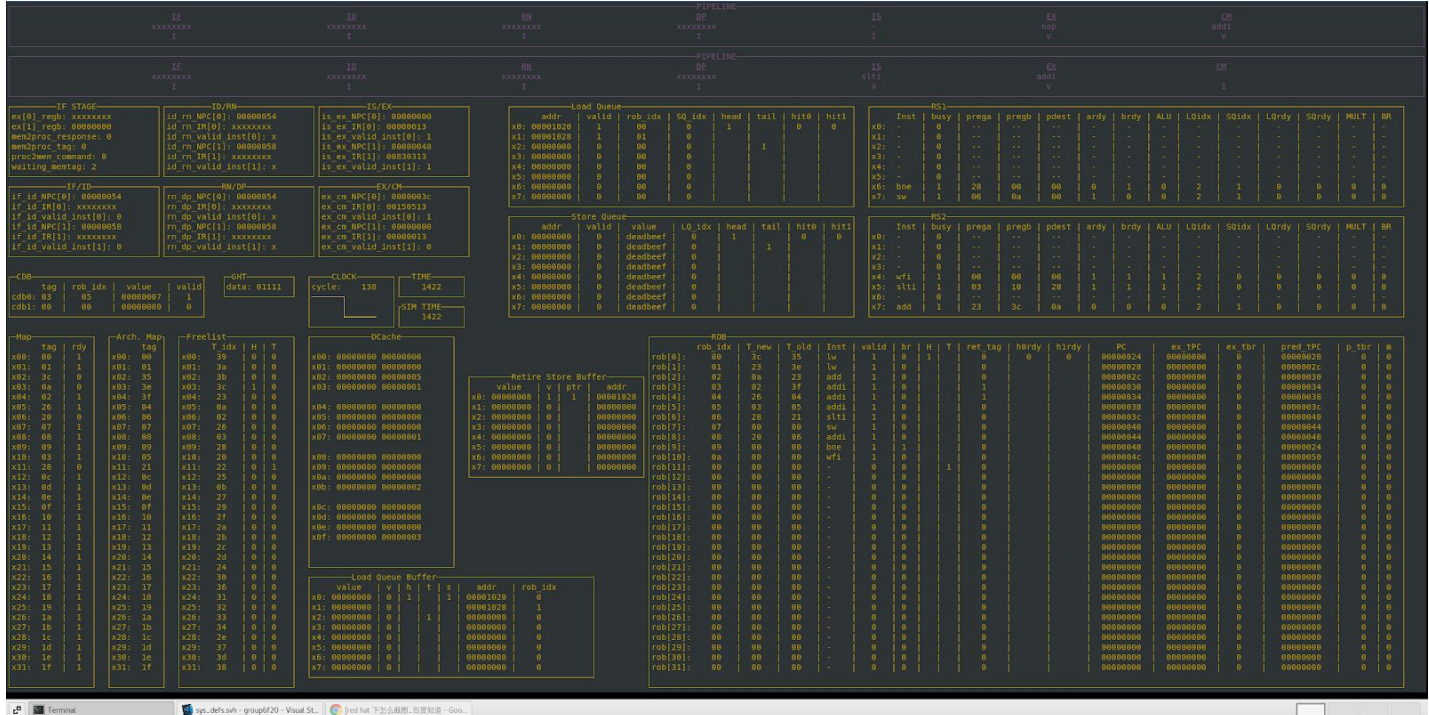
In our system we designed Icache and Dcache separately.

1. The Icache is a direct mapped cache with 16 cache lines. Each cache line has 8 Bytes. The prefetch policy in the Icache is that it will continuously prefetch instruction from the main memory until it fetches all zero data. There is a circular buffer that can record the response from memory and can send the correct instruction back to IF stage. If a branch or memory exception happens, the buffer will be reset and start again until a cache miss happens.
2. For our Dcache, we designed a parameterizable N-way associative cache. Currently we are using a 4-way cache with 4 sets. This cache is a write-back cache and the eviction policy is LRU. We only set the recent bit when a write operation happens, which means if there is a write request from memory or there is a store hit from the processor, the recent bit of this cache line will be set. If a load instruction hits the cache, it will not change the recent bit. Also if there are load requests and store requests both happening from memory, the Dcache will handle the store first and start working on load requests until it resolves all store requests.

Debug and Performance Test

- **Visual Debugger**

We implemented a visual debugger that can show the run-time information in each module.

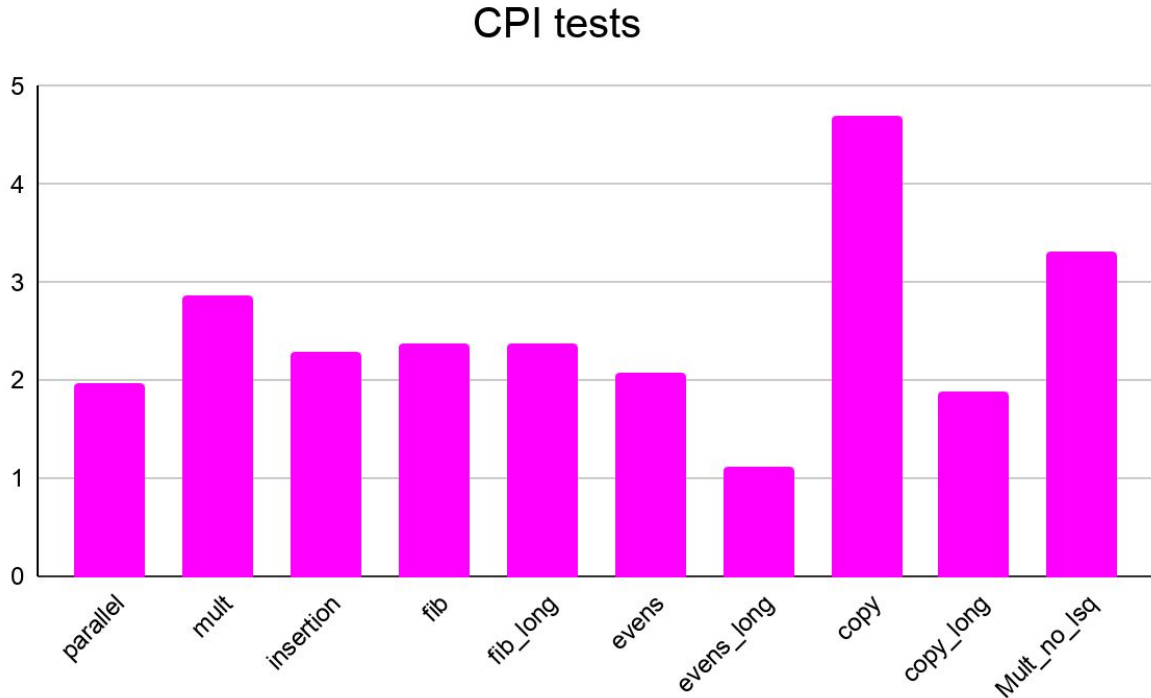


- **Test on superscalar**

We tested our superscalar using a simple assembly code without any dependency and the CPI we achieved is 0.61 and the IPC is 1.64. This result is close to what we expected. However when we did tests on provided code, the average CPI was 2.49. The reason for this result might be the number of instructions of these assembly code is small and the penalty of cache miss, miss predict and 8-stage mult is high.

The assembly code we used to test superscalar is shown below:

```
li x1, 0x0
li x2, 0x0
li x3, 0x0
li x4, 0x0
li x5, 2000
loop: addi x1, x1, 0x1
      addi x2, x2, 0x1
      addi x3, x3, 0x1
      addi x4, x4, 0x1
      bne x1, x5, loop
wfi
```



- **Test on branch predictor**

We tested our branch predictor with a nested loop assembly code. The CPI we achieved is 0.89.

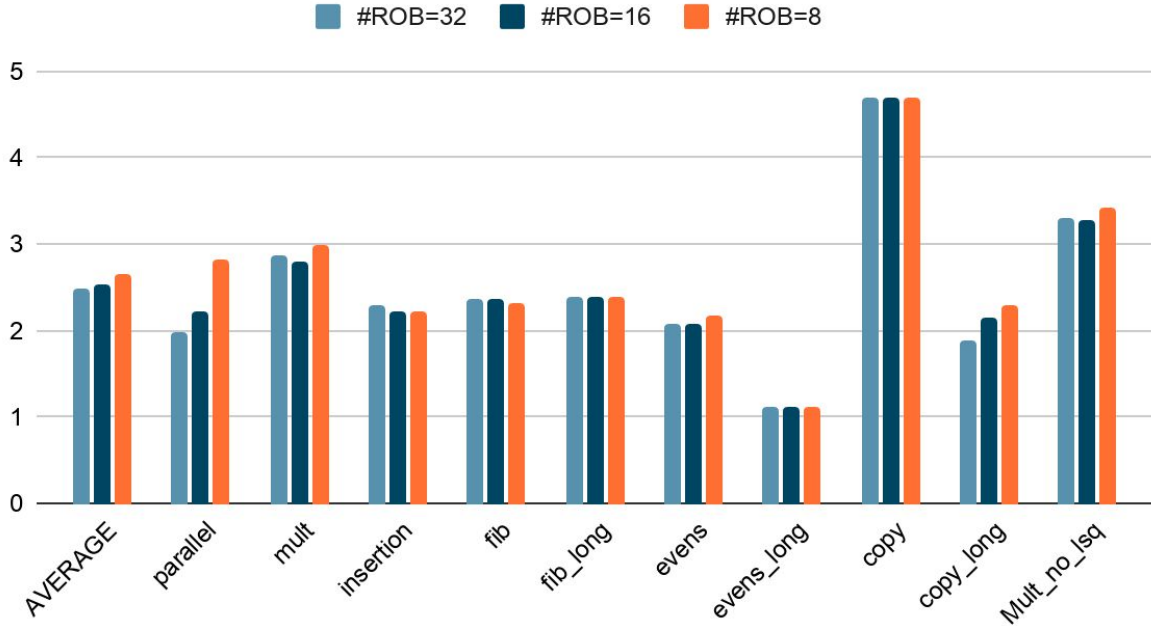
```
li x1, 0xef
li x2, 0xff
loop1: addi x1, x1, -1
loop2: addi x2, x2, -1
       bne x1, x0, loop1
       bne x2, x0, loop2
wfi
```

Performance Analysis

- **ROB size**

The number of entries in our ROB is 32. We did tests when ROB size was 16 or 8. For all test programs, the ROB will not be full and will not cause structure hazard if the size is 32. For rv32_parallel, rv32_copy_long, the ROB size has a significant impact on performance. The average CPI is 2.49, 2.53, 2.64 when the number of ROB entries is 32, 16, 8, separately.

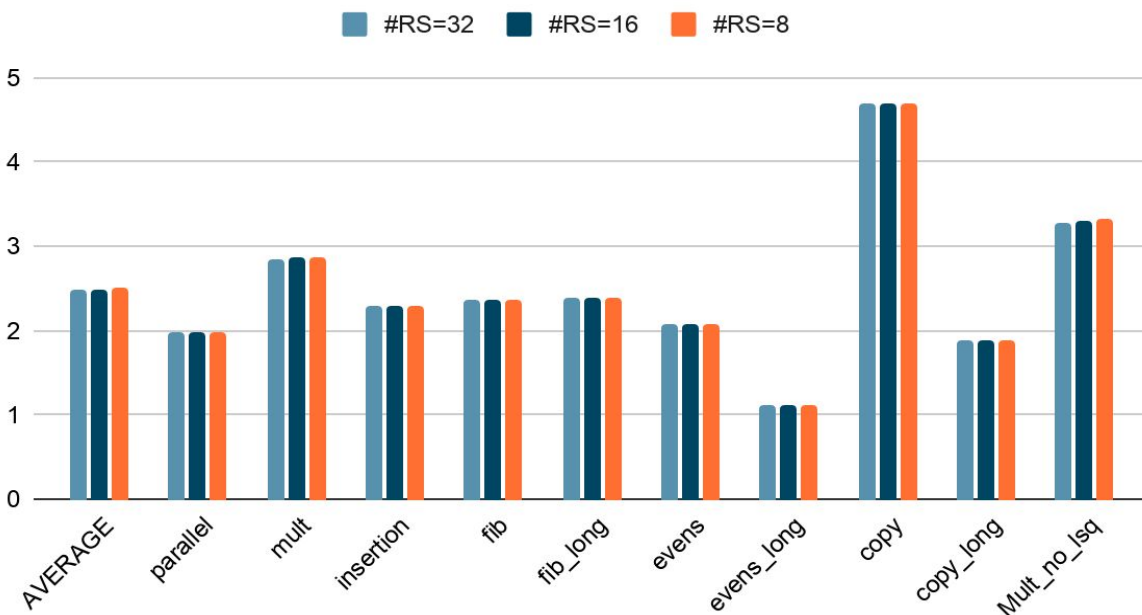
ROB size tests



- RS size**

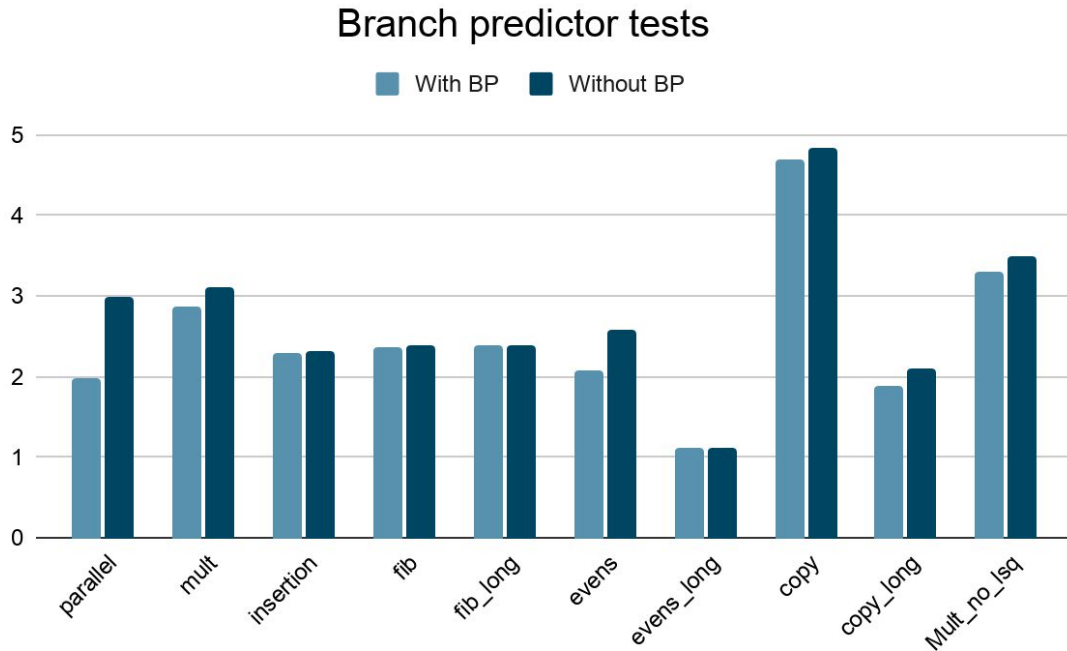
We did tests on different RS sizes. We found that for most programs the RS sizes did not affect the CPI. Only for rv32_mult.s and mult_no_lsq.s, the RS will be full and cause structure hazard so the size will slightly influence the performance. The average CPI is 2.49, 2.496, 2.501 when the number of RS entries is 32, 16, 8, separately.

RS size tests



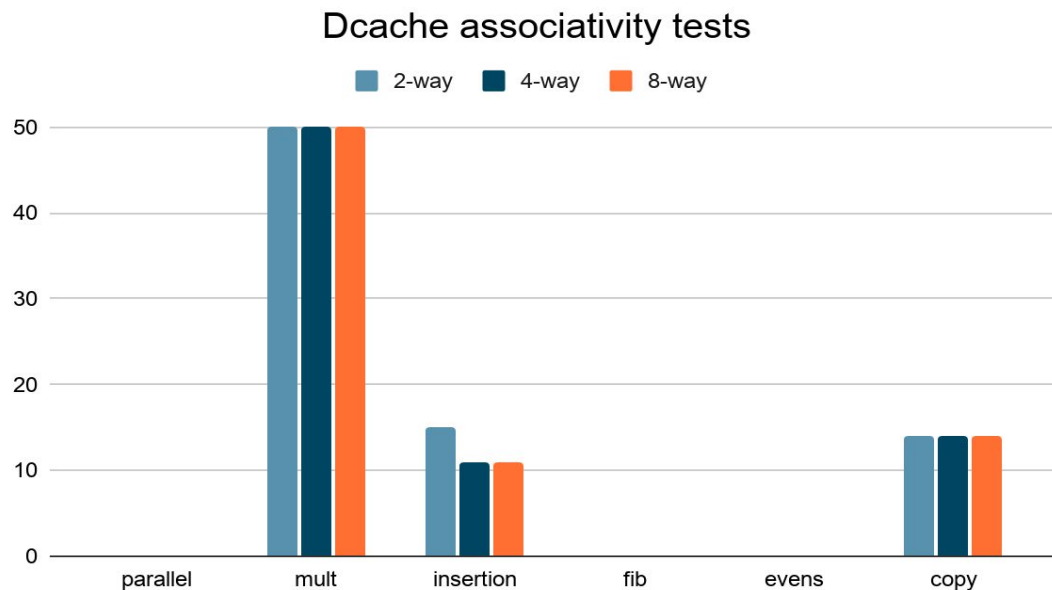
- **Branch prediction**

We tried testing our system without our branch predictor (predict always not taken). The average CPI without branch predictor is 2.735, which is larger than CPI with branch predictor(2.49). This result shows that our branch predictor effectively improved the system performance.



- **Dcache associativity**

Our Dcache is parameterized so we tried 2-way, 4-way and 8-way. The chart below shows the eviction times for each program that has load or store instructions. For parallel, fib and evens programs, 16 cache lines is enough so there is no eviction happening. For mult and copy programs different dcache has the same result. The only difference is rv32_insertion.s and 2-way cache will evict 15 times but 4-way and 8-way only evict 11 times.



Summary

In summary, our design achieved 2-way superscalar and other performance improvement as a R10K style out-of-order processor. The advanced features we implemented are listed below. However in our tests, the performance did not achieve what we expected. The reason for that is the penalty of cache miss, branch predict miss and multiple stage of multiplication is high when the code size is relatively small. Also our system still has some bugs which will cause unnecessary Icache miss, so the overall performance is reduced.

- Advanced feature:
 - Superscalar (width = 2)
 - Store-to-load forwarding in LSQ
 - Loads issue out-of-order past pending stores(non-speculative)
 - Post-retirement store buffer
 - Multiple outstanding load misses
 - Next-line prefetching for instructions
 - Write-back data cache
 - Data cache associativity > 1
 - Gshare branch predictor with BTB
 - Return address stack

Contribution

- **Jiajun Cao:** ROB, RS, LSQ, Pipeline Integration, System verification
- **Rui Sun:** ROB, RS, I/D cache, Pipeline Integration, System verification
- **Wenbo Duan:** Maptable, Pipeline Integration, Visual Debugger, System verification
- **Guoqin Yin:** Freelist, LSQ