

Problem 1. Integrating a processor with a data cache

In this problem, we will consider a program executing on a processor connected to a data cache. The processor is a 5-stage pipelined TinyRV1 processor with **full bypassing** to resolve data hazards. The data cache is a **two-line direct-mapped cache with 8B cache lines**.

The program is calculating accumulated sum of an array.

```

1 int accu_sum( int src[], int size ) {
2   int sum = 0;
3   for ( int i = 0; i < size; i++ ) {
4     sum += src[i]
5     src[i] = sum;
6   }
7   return sum;
8 }

```

The simplified assembly instructions which correspond to the loop above are as shown below. Study the assembly instructions and make sure that you understand how it corresponds to the C++ program above. **For simplicity, assume the following:**

- x1 : holds the accumulated sum of the src[] array
- x2 : holds the base address of the src[] array
- x3 : holds the size or loop count

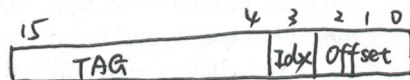
```

1 loop:
2 lw   x4, 0(x2)
3 add  x1, x1, x4
4 sw   x1, 0(x2)
5 addi x2, x2, 4
6 addi x3, x3, -1
7 bne  x3, x0, loop
8 opA
9 opB
10 ...

```

Part 1.A Cache Behavior

Assume the data cache is empty initially. We also assume the base address of the src array stored in **x2 is 0x2000**. The table below shows the first few dynamic memory transactions. Fill in the table. What's the cache miss rate of this program? Can you think of any way to reduce cache miss rate? Will larger cache help?



	tag	idx	hit/miss	Set	
				0	1
rd 0x2000	0x200	0	m		
wr 0x2000	0x200	0	h	0x200	
rd 0x2004	0x200	0	h	0x200	
wr 0x2004	0x200	0	h	0x200	
rd 0x2008	0x200	1	m	0x200	
wr 0x2008	0x200	1	h	0x200	0x200
rd 0x200c	0x200	1	h	0x200	0x200
wr 0x200c	0x200	1	h	0x200	0x200
rd 0x2010	0x200	0	m	0x200	0x200
wr 0x2010	0x200	0	h	0x201	0x200
rd 0x2014	0x200	0	h	0x201	0x200
wr 0x2014	0x200	0	h	0x201	0x200
rd 0x2018	0x201	1	m	0x201	0x200

miss rate is 25%.

larger cache cannot help reduce cache miss rate of this program.

Solutions: ① large cache lines.

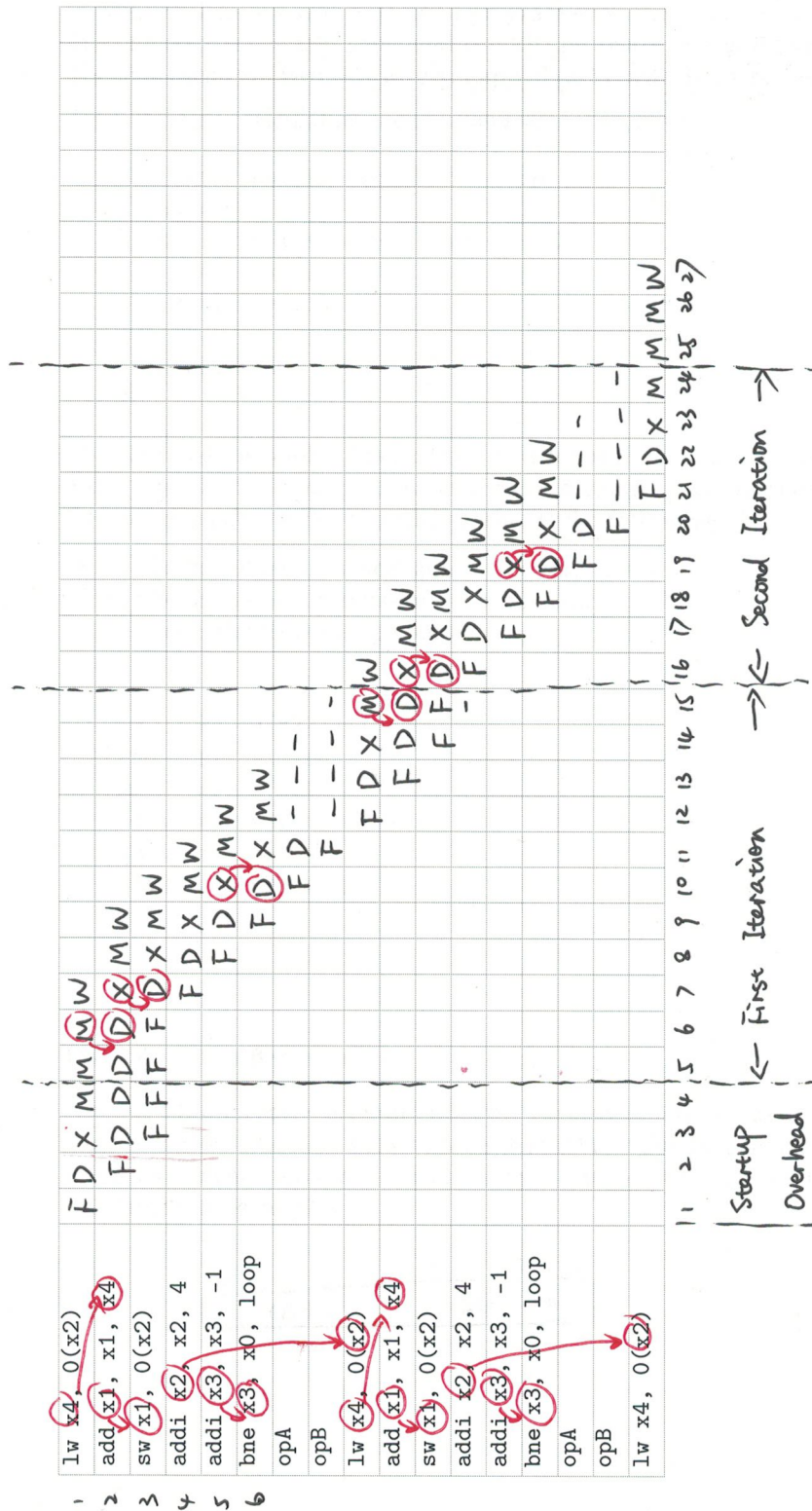
② software/hardware prefetching.

Part 1.B Pipeline Diagram Illustrating Cache Behavior

Now let's try to draw the pipeline diagram of the first two iterations of the loop. We also include the first instruction of the third iteration as shown in the instruction sequence.

Assume the miss penalty is two cycles. So on a cache miss, the processor stalls for **two extra cycles** in the M stage. Note that the **cache is not pipelined** so please be aware of potential structural hazards in the M stage.

Draw arrows in your pipeline diagram to indicate any data or control hazards. Ignoring the startup overhead, what's the CPI of the first iteration? What's the CPI of the second iteration? What's the overall CPI if loop size is 64? If cache lines are 16B instead of 8B, what's the overall CPI?



16B Cache Lines

CPI of the first iteration: $\frac{11}{6}$

CPI of the second, third, fourth iteration: $\frac{9}{6}$

overall CPI: $\frac{11 \times 16 + 9 \times 48}{6 \times 64} = \frac{19}{12} (< \frac{5}{3})$

CPI of the first iteration: $\frac{11}{6}$

CPI of the second iteration: $\frac{9}{6} = \frac{3}{2}$

overall CPI: $\frac{11 \times 32 + 9 \times 32}{6 \times 64} = \frac{10}{6} = \frac{5}{3}$