

2. Consider the following matrix transpose code

```
int i, j ,k;
double *A, *B, *C;
A = (double *)malloc(sizeof(double)*N*N);
B = (double *)malloc(sizeof(double)*N*N);
init_data(A, N*N);
for(i = 0; i < N; i++)
    for(j = 0; j < N; j++)
        B[i*N+j] = A[j*N+i];
// assume load A[j*N+i] and then store B[i*N+j]
output_data(B, N*N);
```

Assume that the starting address of array A is 0x20000 and array B is 0x40000.

Assume $N = 128$. Please answer the following questions:

- (1) Assuming that you have an AMD Bulldozer microarchitecture that has a 16K, 4-way, 64-byte blocked L1 data blocked L1 data cache, please estimate the cache miss rate.

(2) Continued from the previous question, how many of the misses are compulsory misses? How many of them are conflict misses?

3. You are building a system around a single-issue in-order processor running at 2 GHz and the processor has a base CPI of 1 if all memory accesses are hits. The only instructions that read or write data from memory are loads (20% of all instructions) and stores (5% of all instructions). The memory system for this computer is composed of a split L1 cache that imposes no penalty on hits. Both the I-cache and D-cache are direct mapped and hold 32KB each. You may assume the caches use write-allocate and write-back policies. The L1 I-cache has a 2% miss rate and the L1 D-cache has a 5% miss rate. Also, 50% of all blocks replaced from L1 D-cache are dirty. The 512KB write-back, unified L2 cache has an access time of 12ns. Of all memory references sent to the L2 cache in this system, 80% are satisfied without going to main memory. Also 25% of all blocks replaced are dirty. The main memory has an access latency of 60ns.

(1) What is the overall CPI, including memory accesses?

- (2) You are considering replacing the 2 GHz CPU with one that runs at 3 GHz, but is otherwise identical. How much faster does the system run with a faster processor? Assume the L1 cache still has no hit penalty, and that the speeds of the L2 cache, and main memory remain the same in absolute terms (e.g. the L2 cache still has a 12 ns access time).