Comparative Architectures Supervision Problem (Set 3)

Wei Song 25/02/2016

The following problems are mostly open-ended. Diagrams or charts are not needed in the answers. Therefore it is highly recommended to type the answers. Please keep your answers organised and tidy. If you feel difficult to answer some questions or would like a discussion during the supervision session, please label it in your answer.

Lecture 9 & 10. Cache

Q1. What does it mean if a cache memory hierarchy adopts a multi-level inclusion policy? What might influence a decision on whether to adopt a multi-level inclusion or exclusion policy?

Q2. A naive programmer writes the following code for performing the matrix multiply-add function C=AB+C on square matrices:

```
for (i=0; i<N; ++i) {
    for (j=0; j<N; ++j) {
        for (k=0; k<N; ++k) {
            C[k][i] = C[k][i] + ( A[k][j] * B[j][i] );
        }
    }
}
```

(Where X[v][u] refers to the element in row v, column u. Arrays are stored in memory row by row, i.e. X[0][0], X[0][1], X[0][2], ...X[0][N], X[1][0], ... etc.)

1. When used to multiply very large matrices, performance of the programmer's algorithm is very poor. Explain what is happening.
2. The algorithm can be improved simply by changing the order of the loops. Demonstrate how and why.
3. Show how further improvement can be obtained through a technique known as cache blocking
4. Could the algorithm be successfully parallelised to run a on a microprocessor supporting Simultaneous Multithreading (SMT)? Briefly justify your answer.

Q3. Describe at least 3 different ways to improve the performance of a directly mapped cache.

Q4. Describe the three difference reasons of cache misses and the potential software/hardware methods to reduce the cache miss overhead caused by each reason.

Q5. A multicore processor has 8 scalar cores, each of which has an 8KB private write-through L1. A shared L2 of 2MB keeps all L1s in coherent state. Instead of maintaining a directory for each L2 cache line, this L2 chooses to maintain a directory by shadowing L1 tags. Can you figure out the reason? What if each L1 is 128KB, or if the L1 is write-back, would you design differently?

Lecture 11. Vector Machine

Q6. Both VLIW and Vector (SIMD) improve ILP. Describe the pros and cons of VLIW and Vector machines.

Q7. What does support for vector chaining and tailgating allow in a vector processor?

Q8. Why does a vector processor offers a particularly energy efficient solution to execute some type of program?

Q9. In which situations might a vector processor perform worse than a simple pipelined processor?