

Exercise 2: Memory Locality Optimizations

July 8, 2014

1. What are the possible ways of improving the cache performance of an application?
2. What are the two different ways of reusing data in cache?
3. What aspects of the architectural design of caches may impact application performance?
4. What is the safety constraint of each optimization below? How can it enhance the memory performance of a program?
 - (a) Loop interchange
 - (b) Loop blocking
 - (c) Loop unrolling
 - (d) Loop unroll&jam
 - (e) array blocking
 - (f) scalar replacement

Try optimize each of the following code fragments.

```
1. void dgemm(double *a, double *b,
             double *c, double beta, int n)
{
    int i, j, k;
    for (j = 0; j < n; j++)
        for (i = 0; i < n; i++)
        {
            c[j*n+i] = beta*c[j*n+i];
            for (k = 0; k < n; k++)
                c[j*n+i] += a[k*n+i] * b[j*n+k];
        }
}

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