3f.

To do matrix multiplication, we are actually accessing matrix A in row-major order and matrix B in column-major order. The time for mmul3 is shorter than mmul4 because mmul3 maintains the spatial locality for both matrix A and B by storing A and B in the order that matrix multiplication operates on the the two matrixes. However, for mmul4 matrix A is stored in column-major order and B is stored in row-major order so spatial locality is violated. Accordingly, cache hits will be more often for mmul3 than mmul4 so the time for mmul3 will be shorter than mmul4.