

Time per FMA (log-scale)

Right matrices per thread: 1

Right matrices per thread: 2

Right matrices per thread: 4

16x16

32x32

64x64

128x128

256x256

16x16

32x32

64x64

128x128

256x256

16x16

32x32

64x64

128x128

256x256

Matrices size (log-scale)

Grouped overlaps per thread

1

2

4

1 ps

500 fs

300 fs

