Conclusion:

I can see that the performance is terrible when the ratio (cutoff/array size) is less

and reaches some optimal value around 0.1 to 0.3

As threads increase, performance gets better too for array sizes just over a million