GEOM theoretical performance and scalability limits

Based on ZERO GEOM class.

```
# kldload geom_zero
# sysctl kern.geom.zero.clear=0
# diskinfo -v /dev/gzero
/dev/gzero
        512                     # sectorsize
        1152921504606846976     # mediasize in bytes (1.0E)
        2251799813685248        # mediasize in sectors
```

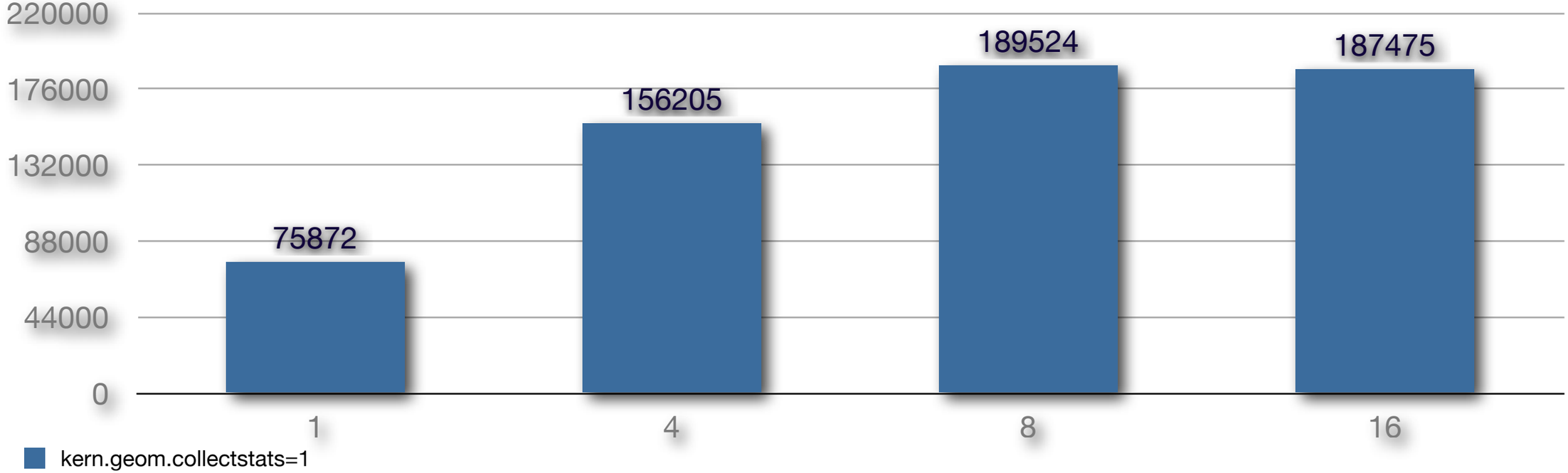
```
CPU Intel(R) Xeon(R) CPU X5272 @ 3.40GHz
FreeBSD/SMP: Multiprocessor System Detected: 4 CPUs
FreeBSD/SMP: 2 package(s) x 2 core(s)
real memory = 8589934592 (8192 MB)
```

```
static void
g_zero_start(struct bio *bp)
{
        int error;

        switch (bp->bio_cmd) {
        case BIO_READ:
                if (g_zero_clear)
                        memset(bp->bio_data, g_zero_byte, bp->bio_length);
                /* FALLTHROUGH */
        case BIO_WRITE:
        case BIO_DELETE:
                bp->bio_completed = bp->bio_length;
                error = 0;
                break;
        default:
                error = EOPNOTSUPP;
                break;
        }
        g_io_deliver(bp, error);
}
```

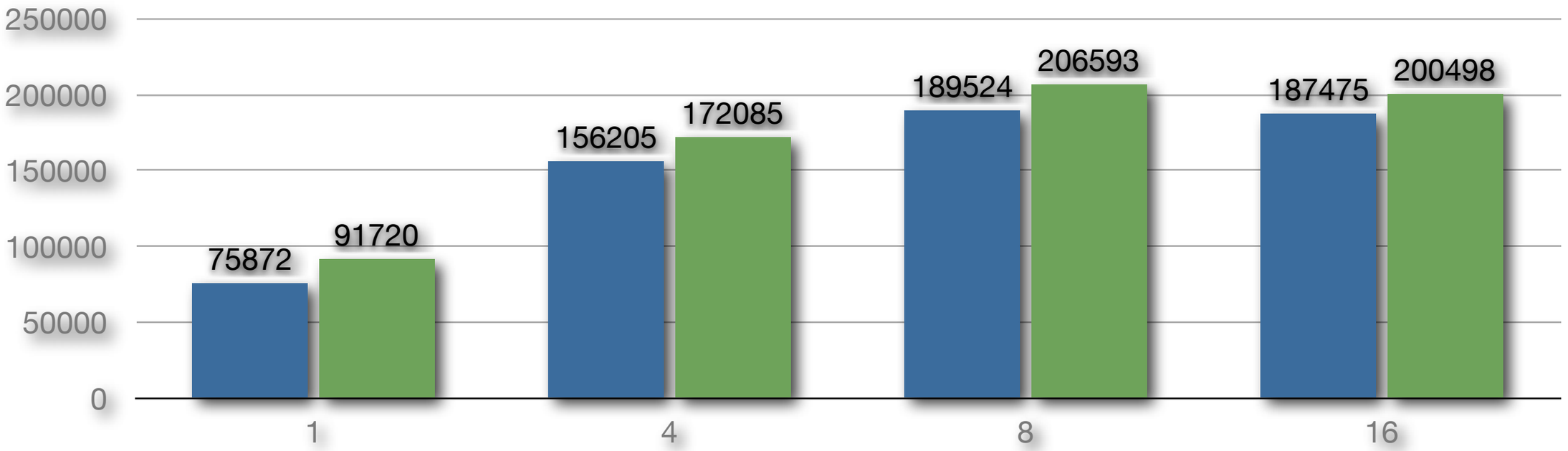
kern.geom.collectstats=1

raidtest test -r -d /dev/gzero -n <NPROC>



kern.geom.collectstats=0

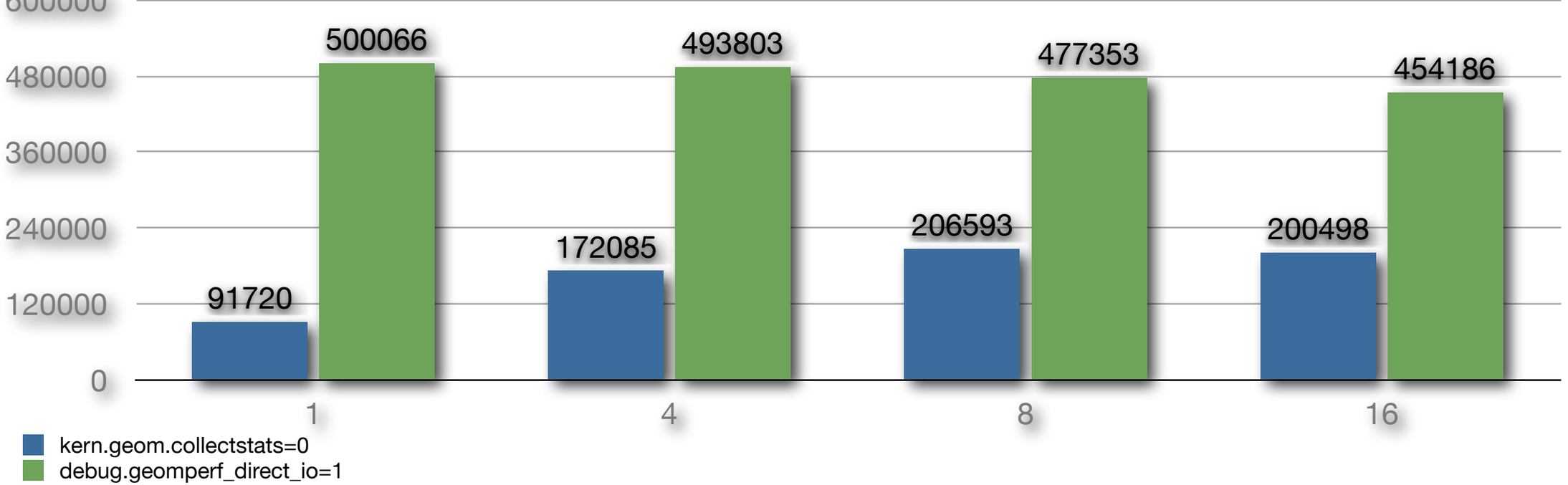
Difference at 95% confidence
15880 +/- 1862.84
10.1661% +/- 1.19256%
(Student's t, pooled s = 821.868)



debug.geomperf_direct_io=1

Eliminates the g_up/g_down threads and passes the I/O requests directly in both directions.

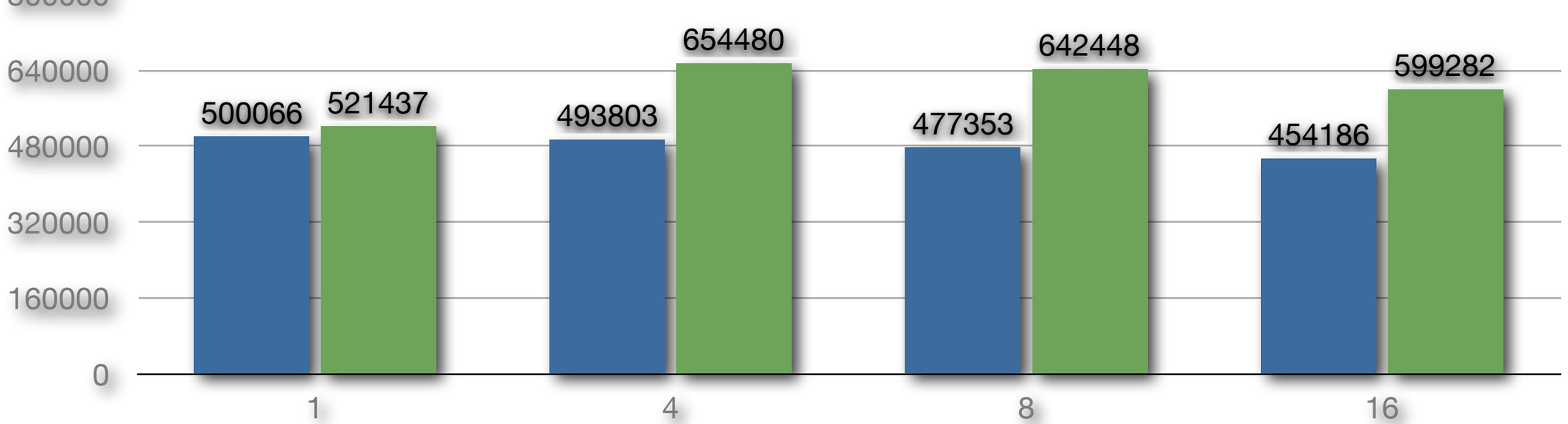
Difference at 95% confidence
321718 +/- 4812
186.953% +/- 2.79629%
(Student's t, pooled s = 2123.01)



debug.geomperf_dev_eternal=1

Adds the MAKEDEV_ETERNAL flag to make_dev_p() call in GEOM DEV class.
This avoids contentions on the global devmtx mutex.

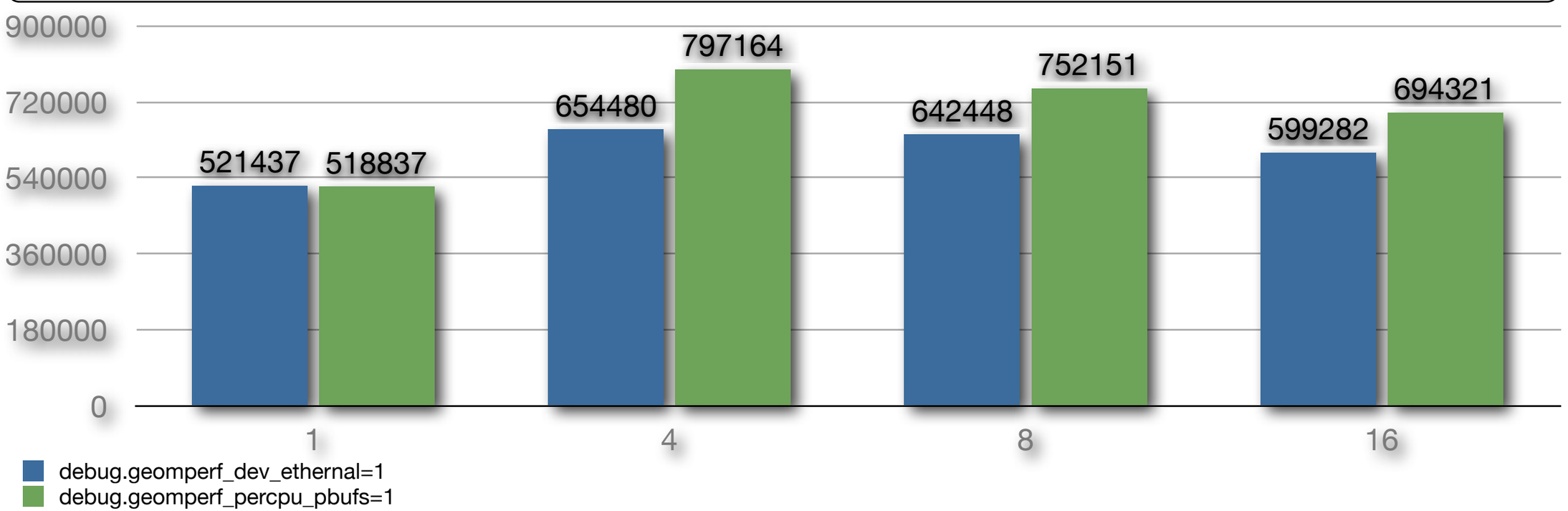
Difference at 95% confidence
160677 +/- 10450.8
32.5387% +/- 2.1164%
(Student's t, pooled s = 4610.81)



debug.geomperf_percpu_pbufs=1

Uses per-CPU pbufs.
This avoids contentions on the global pbuf_mtx mutex.

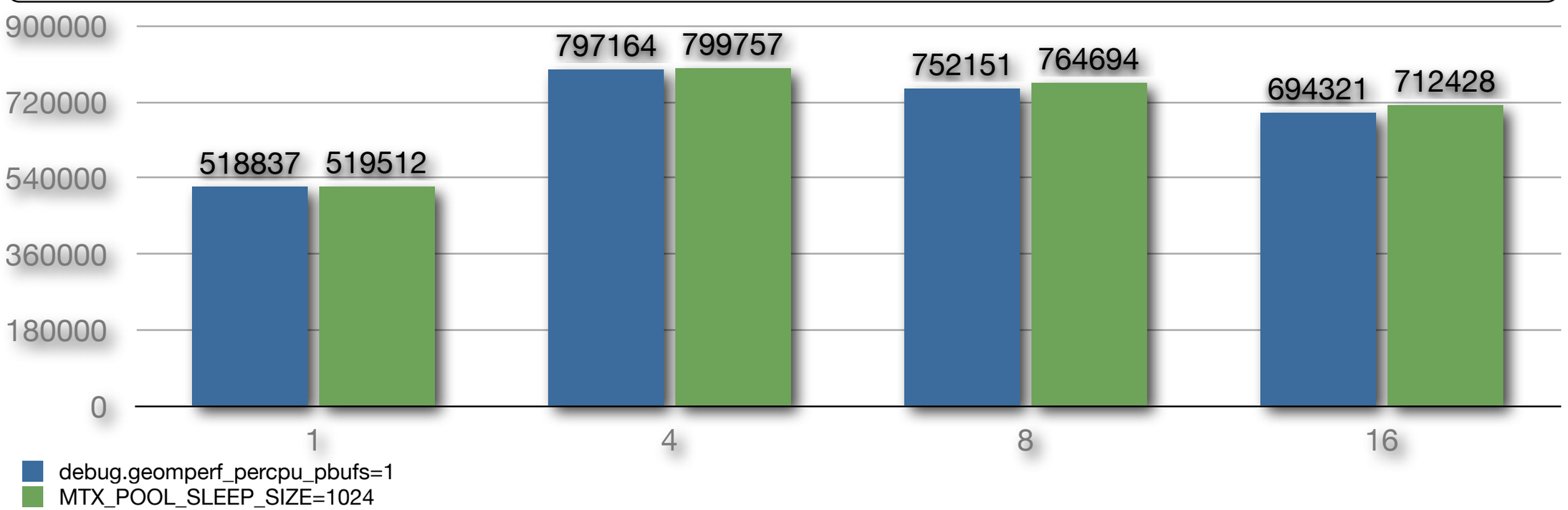
Difference at 95% confidence
142685 +/- 20086.5
21.8012% +/- 3.06908%
(Student's t, pooled s = 8861.98)



-#define MAX_POOL_SLEEP_SIZE 128
+#define MAX_POOL_SLEEP_SIZE 1024

This reduces contention on mtxpool.

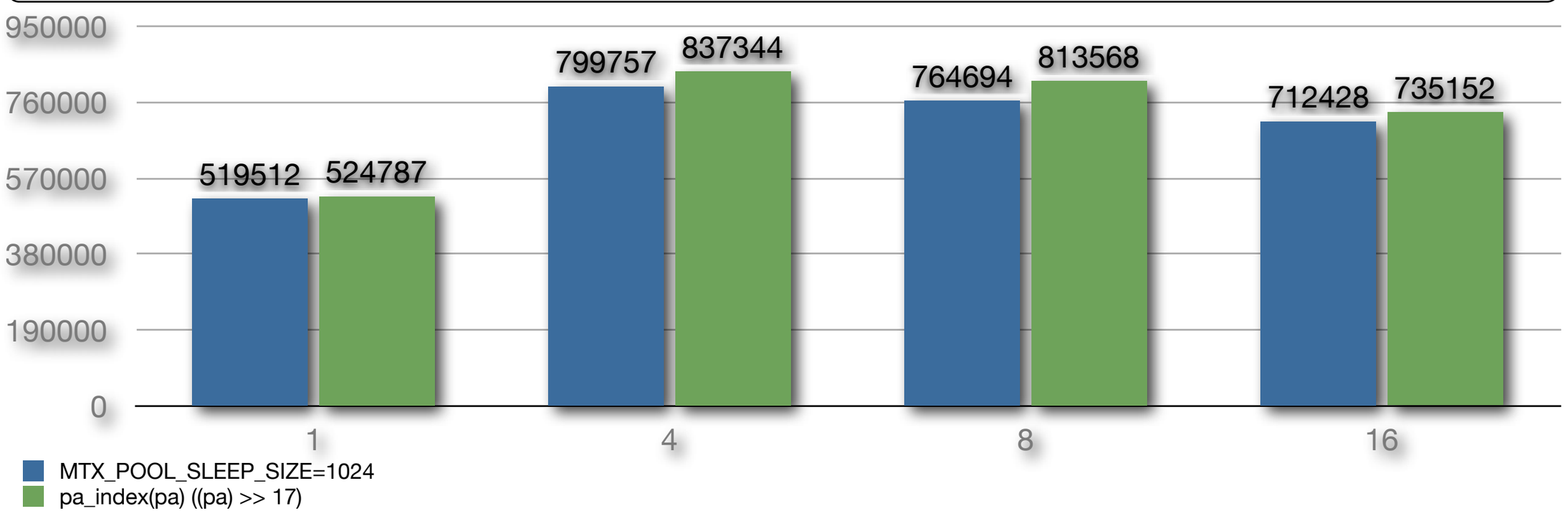
No difference proven at 95% confidence

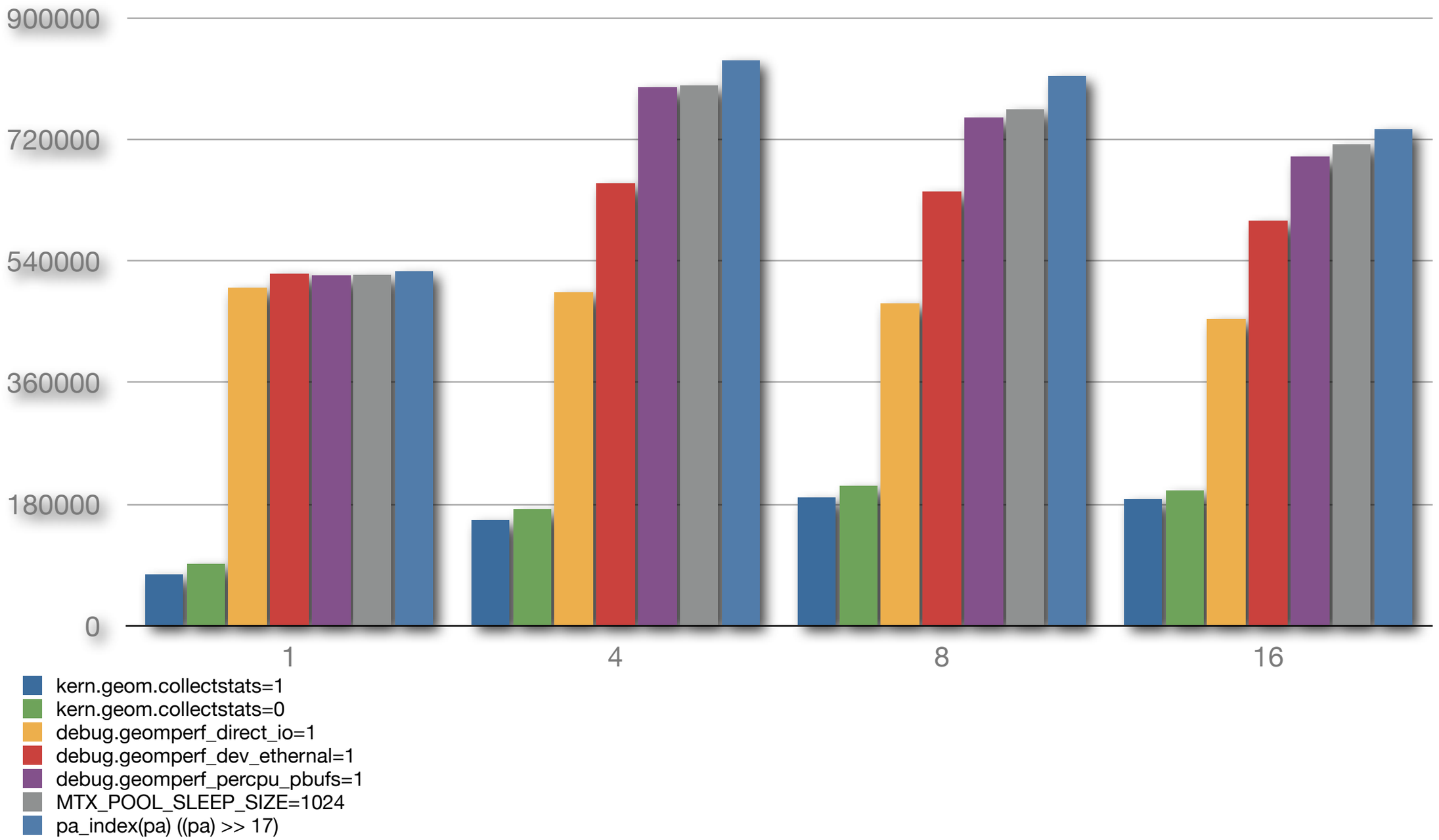


-#define pa_index(pa) ((pa) >> PDRSHIFT)
+#define pa_index(pa) ((pa) >> 17)

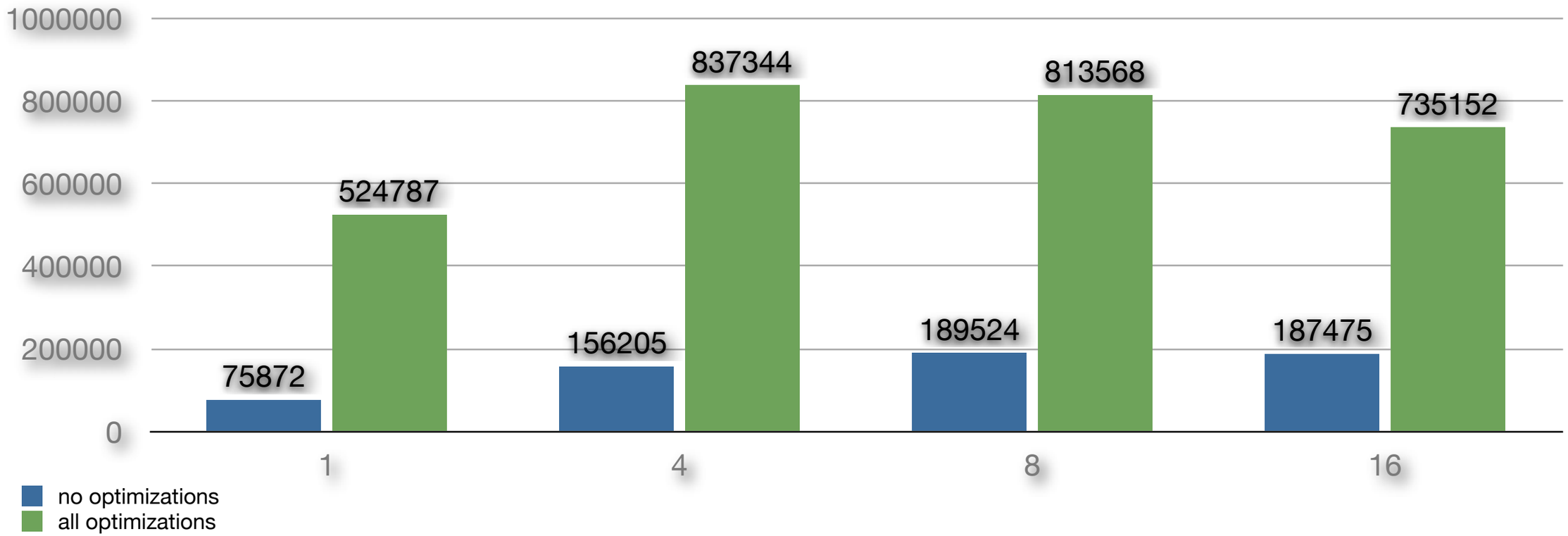
This reduces contention on vm_page lock.

Difference at 95% confidence
37586.3 +/- 27027.2
4.69972% +/- 3.37942%
(Student's t, pooled s = 11924.1)




900000
720000
540000
360000
180000
0
1
4
8
16
kern.geom.collectstats=1
kern.geom.collectstats=0
debug.geomperf_direct_io=1
debug.geomperf_dev_ethernal=1
debug.geomperf_percpu_pbufs=1
MTX_POOL_SLEEP_SIZE=1024
pa_index(pa) ((pa) >> 17)

Difference at 95.0% confidence
681139 +/- 13138.2
436.055% +/- 8.41088%
(Student's t, pooled s = 5796.45)



	1	4	8	16
kern.geom.collectstats=1	75872	156205	189524	187475
kern.geom.collectstats=0	91720	172085	206593	200498
debug.geomperf_direct_io=1	500066	493803	477353	454186
debug.geomperf_dev_ethernal=1	521437	654480	642448	599282
debug.geomperf_percpu_pbufs=1	518837	797164	752151	694321
MTX_POOL_SLEEP_SIZE=1024	519512	799757	764694	712428
pa_index(pa) ((pa) >> 17)	524787	837344	813568	735152
kern.geom.collectstats=1	75872	156205	189524	187475
pa_index(pa) ((pa) >> 17)	524787	837344	813568	735152