

PROJECT 2 REPORT

For this project I have implemented 4 kernels, each with various optimizations in the kernel. I have run a number of experiments with the kernels, and when the results are accurate, a trend has shown that the block size of 32 often result in the fastest performance, with block size 64 being second place. For this reason, for the next few pages are figures showing the timings of each algorithm (as annotated by their heading). I've shown a sampling of runs with block sizes 32 and 64, but it should be noted that the trend is consistent, regardless of input size, that block size 32 is the fastest.

Tiled Algorithm without Privatization (PDH Kernel3) run times:

```
[tanmay@c4cuda17 cuda-proj-1]$ make pull && make && ./SDH 10000 500.0 32
git pull origin
The authenticity of host 'github.com (192.30.253.112)' can't be established.
RSA key fingerprint is 16:27:ac:a5:76:28:2d:36:63:1b:56:4d:eb:df:a6:48.
Are you sure you want to continue connecting (yes/no)? yes
Failed to add the host to the list of known hosts (/home/t/tanmay/.ssh/known_hosts).
Enter passphrase for key '/home/t/tanmay/.ssh/id_rsa':
remote: Enumerating objects: 5, done.
remote: Counting objects: 100% (5/5), done.
remote: Compressing objects: 100% (1/1), done.
remote: Total 3 (delta 2), reused 3 (delta 2), pack-reused 0
Unpacking objects: 100% (3/3), done.
From ssh://github.com/tkotha/cuda-proj-1
 823bbfb..ee68a2a  master -> origin/master
Updating 823bbfb..ee68a2a
Fast-forward
 SDH.cu | 2 +-
 1 files changed, 1 insertions(+), 1 deletions(-)
nvcc SDH.cu -o SDH
Starting GPU...
blockcount: 313
numbuckets: 80
shmemsiz3: 768
shmemsiz4: 1088
CUDA EVENT: Running time for GPU version: 32.53706 ms

00:      2076 |      14212 |      37870 |      70863 |      113190 |
05:     162616 |     219792 |     281373 |     348231 |     418908 |
10:     492542 |     568919 |     643293 |     723787 |     799933 |
15:     878146 |     953657 |     1025635 |     1097551 |     1161963 |
20:     1225131 |     1285543 |     1340874 |     1389657 |     1435607 |
25:     1471150 |     1505326 |     1534652 |     1554499 |     1566213 |
30:     1574002 |     1572648 |     1562633 |     1550305 |     1527753 |
35:     1500146 |     1463004 |     1419940 |     1372128 |     1313777 |
40:     1249134 |     1177114 |     1097835 |     1012343 |     920271 |
45:     823626 |     723835 |     632337 |     548947 |     472904 |
50:     405296 |     344273 |     289347 |     240911 |     197652 |
55:     160654 |     129093 |     101564 |     78499 |     60069 |
60:     44360 |     32554 |     23054 |     16448 |     11533 |
65:     7835 |     5295 |     3371 |     2289 |     1372 |
70:     804 |     450 |     222 |     104 |     34 |
75:     16 |     8 |     2 |     0 |     0 |
T:49995000
***** Total Running Time of Kernel = 32.53706 ms *****
```

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 512000 500.0 32
Starting GPU...
blockcount: 16000
numbuckets: 80
shmemsiz3: 768
shmemsiz4: 1088
CUDA EVENT: Running time for GPU version: 66465.44531 ms

00:      5501396 |      37444290 |      98427880 |      185423339 |      295440468 |
05:     425708103 |     573511517 |     736172188 |     911273945 |     1096423184 |
10:     1289231615 |     1487641295 |     1689463811 |     1892879819 |     2095890086 |
15:     2296613707 |     2493587809 |     2685331702 |     2870075682 |     3046469709 |
20:     3213269222 |     3369188027 |     3513248105 |     3644273302 |     3761060343 |
25:     3863275796 |     3949722026 |     4019824382 |     4073176567 |     4109097355 |
30:     4127299575 |     4127323151 |     4108782509 |     4071576785 |     4015434179 |
35:     3940543050 |     3846464030 |     3733138914 |     3601079592 |     3450144521 |
40:     3280989595 |     3093916526 |     2888724142 |     2665988785 |     2426328214 |
45:     2169925121 |     1904887813 |     1659304152 |     1437414727 |     1236913696 |
50:     1056268982 |     894354063 |     750132845 |     622459515 |     510525916 |
55:     413240592 |     329618096 |     258646002 |     199380204 |     150732486 |
60:     111724435 |     81064669 |     57684232 |     40370308 |     27891449 |
65:     19060657 |     12808595 |     8396577 |     5343754 |     3276307 |
70:     1916846 |     1060050 |     546518 |     257605 |     106930 |
75:     38185 |     10437 |     1866 |     132 |     0 |
T:131071744000
***** Total Running Time of Kernel = 66465.44531 ms *****
```

Tanmay J. Kotha (U31037110)
Programming Massively Parallel Systems
CIS 6930

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 512000 500.0 64
Starting GPU...
blockcount: 8000
numbuckets: 80
shmeme3: 1536
shmeme4: 1856
CUDA EVENT: Running time for GPU version: 66764.96875 ms

00:      5501396 |      37444290 |      98427880 |      185423339 |      295440468 |
05:      425708103 |      573511517 |      736172188 |      911273945 |      1096423184 |
10:      1289231615 |      1487641295 |      1689463811 |      1892879819 |      2095890086 |
15:      2296613707 |      2493587809 |      2685331702 |      2870075682 |      3046469709 |
20:      3213269222 |      3369188027 |      3513248105 |      3644273302 |      3761060343 |
25:      3863275796 |      3949722026 |      4019824382 |      4073176567 |      4109097355 |
30:      4127299575 |      4127323151 |      4108782509 |      4071576785 |      4015434179 |
35:      3940543050 |      3846464030 |      3733138914 |      3601079592 |      3450144521 |
40:      3280989595 |      3093916526 |      2888724142 |      2665988785 |      2426328214 |
45:      2169925121 |      1904887813 |      1659304152 |      1437414727 |      1236913696 |
50:      1056268982 |      894354063 |      750132845 |      622459515 |      510525916 |
55:      413240592 |      329618096 |      258646002 |      199380204 |      150732486 |
60:      111724435 |      81064669 |      57684232 |      40370308 |      27891449 |
65:      19060657 |      12808595 |      8396577 |      5343754 |      3276307 |
70:      1916846 |      1060050 |      546518 |      257605 |      106930 |
75:      38185 |      10437 |      1866 |      132 |      0 |
T:131071744000
***** Total Running Time of Kernel = 66764.96875 ms *****
```

Tiled Kernel with Privatization (PDH Kernel4), run times (block size = 32):

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 10000 500.0 32
Starting GPU...
blockcount: 313
numbuckets: 80
shmeme3: 768
shmeme4: 1088
CUDA EVENT: Running time for GPU version: 33.06493 ms

00:      2076 |      14212 |      37870 |      70863 |      113190 |
05:      162616 |      219792 |      281373 |      348231 |      418908 |
10:      492542 |      568919 |      643293 |      723787 |      799933 |
15:      878146 |      953657 |      1025635 |      1097551 |      1161963 |
20:      1225131 |      1285543 |      1340874 |      1389657 |      1435607 |
25:      1471150 |      1505326 |      1534652 |      1554499 |      1566213 |
30:      1574002 |      1572648 |      1562633 |      1550305 |      1527753 |
35:      1500146 |      1463004 |      1419940 |      1372128 |      1313777 |
40:      1249134 |      1177114 |      1097835 |      1012343 |      920271 |
45:      823626 |      723835 |      632337 |      548947 |      472904 |
50:      405296 |      344273 |      289347 |      240911 |      197652 |
55:      160654 |      129093 |      101564 |      78499 |      60069 |
60:      44360 |      32554 |      23054 |      16448 |      11533 |
65:      7835 |      5295 |      3371 |      2289 |      1372 |
70:      804 |      450 |      222 |      104 |      34 |
75:      16 |      8 |      2 |      0 |      0 |
T:49995000
***** Total Running Time of Kernel = 33.06493 ms *****
```

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 512000 500.0 32
Starting GPU...
blockcount: 16000
numbuckets: 80
shmeme3: 768
shmeme4: 1088
CUDA EVENT: Running time for GPU version: 67873.29688 ms

00:      5501396 |      37444290 |      98427880 |      185423339 |      295440468 |
05:      425708103 |      573511517 |      736172188 |      911273945 |      1096423184 |
10:      1289231615 |      1487641295 |      1689463811 |      1892879819 |      2095890086 |
15:      2296613707 |      2493587809 |      2685331702 |      2870075682 |      3046469709 |
20:      3213269222 |      3369188027 |      3513248105 |      3644273302 |      3761060343 |
25:      3863275796 |      3949722026 |      4019824382 |      4073176567 |      4109097355 |
30:      4127299575 |      4127323151 |      4108782509 |      4071576785 |      4015434179 |
35:      3940543050 |      3846464030 |      3733138914 |      3601079592 |      3450144521 |
40:      3280989595 |      3093916526 |      2888724142 |      2665988785 |      2426328214 |
45:      2169925121 |      1904887813 |      1659304152 |      1437414727 |      1236913696 |
50:      1056268982 |      894354063 |      750132845 |      622459515 |      510525916 |
55:      413240592 |      329618096 |      258646002 |      199380204 |      150732486 |
60:      111724435 |      81064669 |      57684232 |      40370308 |      27891449 |
65:      19060657 |      12808595 |      8396577 |      5343754 |      3276307 |
70:      1916846 |      1060050 |      546518 |      257605 |      106930 |
75:      38185 |      10437 |      1866 |      132 |      0 |
T:131071744000
***** Total Running Time of Kernel = 67873.29688 ms *****
```

Tanmay J. Kotha (U31037110)
Programming Massively Parallel Systems
CIS 6930

Naïve Kernel without Privatization (PDH kernel1):

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 10000 500.0 32
Starting GPU...
blockcount: 313
numbuckets: 80
shmeme3: 768
shmeme4: 1088
CUDA EVENT: Running time for GPU version: 33.71155 ms

00:      2076 |      14212 |      37870 |      70863 |      113190 |
05:      162616 |      219792 |      281373 |      348231 |      418908 |
10:      492542 |      568919 |      643293 |      723787 |      799933 |
15:      878146 |      953657 |      1025635 |      1097551 |      1161963 |
20:      1225131 |      1285543 |      1340874 |      1389657 |      1435607 |
25:      1471150 |      1505326 |      1534652 |      1554499 |      1566213 |
30:      1574002 |      1572648 |      1562633 |      1550305 |      1527753 |
35:      1500146 |      1463004 |      1419940 |      1372128 |      1313777 |
40:      1249134 |      1177114 |      1097835 |      1012343 |      920271 |
45:      823626 |      723835 |      632337 |      548947 |      472904 |
50:      405296 |      344273 |      289347 |      240911 |      197652 |
55:      160654 |      129093 |      101564 |      78499 |      60069 |
60:      44360 |      32554 |      23054 |      16448 |      11533 |
65:      7835 |      5295 |      3371 |      2289 |      1372 |
70:      804 |      450 |      222 |      104 |      34 |
75:      16 |      8 |      2 |      0 |      0 |
T:49995000
***** Total Running Time of Kernel = 33.71155 ms *****
```

```

Total Running Time of Kernel = 33.71155 ms
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 10000 500.0 64
Starting GPU...
blockcount: 157
numbuckets: 80
shmeme3: 1536
shmeme4: 1856
CUDA EVENT: Running time for GPU version: 34.22051 ms

00:      2076 |      14212 |      37870 |      70863 |      113190 |
05:      162616 |      219792 |      281373 |      348231 |      418908 |
10:      492542 |      568919 |      643293 |      723787 |      799933 |
15:      878146 |      953657 |      1025635 |      1097551 |      1161963 |
20:      1225131 |      1285543 |      1340874 |      1389657 |      1435607 |
25:      1471150 |      1505326 |      1534652 |      1554499 |      1566213 |
30:      1574002 |      1572648 |      1562633 |      1550305 |      1527753 |
35:      1500146 |      1463004 |      1419940 |      1372128 |      1313777 |
40:      1249134 |      1177114 |      1097835 |      1012343 |      920271 |
45:      823626 |      723835 |      632337 |      548947 |      472904 |
50:      405296 |      344273 |      289347 |      240911 |      197652 |
55:      160654 |      129093 |      101564 |      78499 |      60069 |
60:      44360 |      32554 |      23054 |      16448 |      11533 |
65:      7835 |      5295 |      3371 |      2289 |      1372 |
70:      804 |      450 |      222 |      104 |      34 |
75:      16 |      8 |      2 |      0 |      0 |
T:49995000
***** Total Running Time of Kernel = 34.22051 ms *****
```

```

Total Running Time of Kernel = 34.22051 ms
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 10000 500.0 128
Starting GPU...
blockcount: 79
numbuckets: 80
shmeme3: 3072
shmeme4: 3392
CUDA EVENT: Running time for GPU version: 34.36486 ms

00:      2076 |      14212 |      37870 |      70863 |      113190 |
05:      162616 |      219792 |      281373 |      348231 |      418908 |
10:      492542 |      568919 |      643293 |      723787 |      799933 |
15:      878146 |      953657 |      1025635 |      1097551 |      1161963 |
20:      1225131 |      1285543 |      1340874 |      1389657 |      1435607 |
25:      1471150 |      1505326 |      1534652 |      1554499 |      1566213 |
30:      1574002 |      1572648 |      1562633 |      1550305 |      1527753 |
35:      1500146 |      1463004 |      1419940 |      1372128 |      1313777 |
40:      1249134 |      1177114 |      1097835 |      1012343 |      920271 |
45:      823626 |      723835 |      632337 |      548947 |      472904 |
50:      405296 |      344273 |      289347 |      240911 |      197652 |
55:      160654 |      129093 |      101564 |      78499 |      60069 |
60:      44360 |      32554 |      23054 |      16448 |      11533 |
65:      7835 |      5295 |      3371 |      2289 |      1372 |
70:      804 |      450 |      222 |      104 |      34 |
75:      16 |      8 |      2 |      0 |      0 |
T:49995000
***** Total Running Time of Kernel = 34.36486 ms *****
```

Tanmay J. Kotha (U31037110)
Programming Massively Parallel Systems
CIS 6930

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 512000 500.0 32
Starting GPU...
blockcount: 16000
numbuckets: 80
shmeme3: 768
shmeme4: 1088
CUDA EVENT: Running time for GPU version: 69625.33594 ms

00: 5501396 | 37444290 | 98427880 | 185423339 | 295440468 |
05: 425708103 | 573511517 | 736172188 | 911273945 | 1096423184 |
10: 1289231615 | 1487641295 | 1689463811 | 1892879819 | 2095890086 |
15: 2296613707 | 2493587809 | 2685331702 | 2870075682 | 3046469709 |
20: 3213269222 | 3369188027 | 3513248105 | 3644273302 | 3761060343 |
25: 3863275796 | 3949722026 | 4019824382 | 4073176567 | 4109097355 |
30: 4127299575 | 4127323151 | 4108782509 | 4071576785 | 4015434179 |
35: 3940543050 | 3846464030 | 3733138914 | 3601079592 | 3450144521 |
40: 3280989595 | 3093916526 | 2888724142 | 2665988785 | 2426328214 |
45: 2169925121 | 1904887813 | 1659304152 | 1437414727 | 1236913696 |
50: 1056268982 | 894354063 | 750132845 | 622459515 | 510525916 |
55: 413240592 | 329618096 | 258646002 | 199380204 | 150732486 |
60: 111724435 | 81064669 | 57684232 | 40370308 | 27891449 |
65: 19060657 | 12808595 | 8396577 | 5343754 | 3276307 |
70: 1916846 | 1060050 | 546518 | 257605 | 106930 |
75: 38185 | 10437 | 1866 | 132 | 0
T:131071744000
***** Total Running Time of Kernel = 69625.33594 ms *****
```

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 512000 500.0 64
Starting GPU...
blockcount: 8000
numbuckets: 80
shmeme3: 1536
shmeme4: 1856
CUDA EVENT: Running time for GPU version: 70086.48438 ms

00: 5501396 | 37444290 | 98427880 | 185423339 | 295440468 |
05: 425708103 | 573511517 | 736172188 | 911273945 | 1096423184 |
10: 1289231615 | 1487641295 | 1689463811 | 1892879819 | 2095890086 |
15: 2296613707 | 2493587809 | 2685331702 | 2870075682 | 3046469709 |
20: 3213269222 | 3369188027 | 3513248105 | 3644273302 | 3761060343 |
25: 3863275796 | 3949722026 | 4019824382 | 4073176567 | 4109097355 |
30: 4127299575 | 4127323151 | 4108782509 | 4071576785 | 4015434179 |
35: 3940543050 | 3846464030 | 3733138914 | 3601079592 | 3450144521 |
40: 3280989595 | 3093916526 | 2888724142 | 2665988785 | 2426328214 |
45: 2169925121 | 1904887813 | 1659304152 | 1437414727 | 1236913696 |
50: 1056268982 | 894354063 | 750132845 | 622459515 | 510525916 |
55: 413240592 | 329618096 | 258646002 | 199380204 | 150732486 |
60: 111724435 | 81064669 | 57684232 | 40370308 | 27891449 |
65: 19060657 | 12808595 | 8396577 | 5343754 | 3276307 |
70: 1916846 | 1060050 | 546518 | 257605 | 106930 |
75: 38185 | 10437 | 1866 | 132 | 0
T:131071744000
***** Total Running Time of Kernel = 70086.48438 ms *****
```

Naïve Kernel with Privatization (PDH kernel2):

```
***** Total Running Time of Kernel = 0.03255 sec *****
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 10000 500.0 32
Starting GPU...
CUDA EVENT: Running time for GPU version: 32.57367 ms

00: 2076 | 14212 | 37870 | 70863 | 113190 |
05: 162616 | 219792 | 281373 | 348231 | 418908 |
10: 492542 | 568919 | 643293 | 723787 | 799933 |
15: 878146 | 953657 | 1025635 | 1097551 | 1161963 |
20: 1225131 | 1285543 | 1340874 | 1389657 | 1435607 |
25: 1471150 | 1505326 | 1534652 | 1554499 | 1566213 |
30: 1574002 | 1572648 | 1562633 | 1550305 | 1527753 |
35: 1500146 | 1463004 | 1419940 | 1372128 | 1313777 |
40: 1249134 | 1177114 | 1097835 | 1012343 | 920271 |
45: 823626 | 723835 | 632337 | 548947 | 472904 |
50: 405296 | 344273 | 289347 | 240911 | 197652 |
55: 160654 | 129093 | 101564 | 78499 | 60069 |
60: 44360 | 32554 | 23054 | 16448 | 11533 |
65: 7835 | 5295 | 3371 | 2289 | 1372 |
70: 804 | 450 | 222 | 104 | 34 |
75: 16 | 8 | 2 | 0 | 0
T:49995000
***** Total Running Time of Kernel = 0.03257 sec *****
```

Tanmay J. Kotha (U31037110)
Programming Massively Parallel Systems
CIS 6930

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 10000 500.0 64
Starting GPU...
CUDA EVENT: Running time for GPU version: 33.29366 ms

00:      2076 |      14212 |      37870 |      70863 |      113190 |
05:     162616 |     219792 |     281373 |     348231 |     418908 |
10:     492542 |     568919 |     643293 |     723787 |     799933 |
15:     878146 |     953657 |    1025635 |    1097551 |    1161963 |
20:    1225131 |    1285543 |    1340874 |    1389657 |    1435607 |
25:    1471150 |    1505326 |    1534652 |    1554499 |    1566213 |
30:    1574002 |    1572648 |    1562633 |    1550305 |    1527753 |
35:    1500146 |    1463004 |    1419940 |    1372128 |    1313777 |
40:    1249134 |    1177114 |    1097835 |    1012343 |     920271 |
45:     823626 |     723835 |     632337 |     548947 |     472904 |
50:     405296 |     344273 |     289347 |     240911 |     197652 |
55:     160654 |     129093 |     101564 |     78499 |     60069 |
60:     44360 |     32554 |     23054 |     16448 |     11533 |
65:     7835 |     5295 |     3371 |     2289 |     1372 |
70:     804 |     450 |     222 |     104 |     34 |
75:     16 |     8 |     2 |     0 |     0 |
T:49995000
***** Total Running Time of Kernel = 0.03329 sec *****
```

```
[tanmay@c4cuda17 cuda-proj-1]$ ./SDH 512000 500.0 32
Starting GPU...
CUDA EVENT: Running time for GPU version: 66491.96094 ms

00:      5501396 |      37444290 |      98427880 |      185423339 |      295440468 |
05:      425708103 |      573511517 |      736172188 |      911273945 |     1096423184 |
10:     1289231615 |     1487641295 |     1689463811 |     1892879819 |     2095890086 |
15:     2296613707 |     2493587809 |     2685331702 |     2870075682 |     3046469709 |
20:     3213269222 |     3369188027 |     3513248105 |     3644273302 |     3761060343 |
25:     3863275796 |     3949722026 |     4019824382 |     4073176567 |     4109097355 |
30:     4127299575 |     4127323151 |     4108782509 |     4071576785 |     4015434179 |
35:     3940543050 |     3846464030 |     3733138914 |     3601079592 |     3450144521 |
40:     3280989595 |     3093916526 |     2888724142 |     2665988785 |     2426328214 |
45:     2169925121 |     1904887813 |     1659304152 |     1437414727 |     1236913696 |
50:     1056268982 |     894354063 |     750132845 |     622459515 |     510525916 |
55:     413240592 |     329618096 |     258646002 |     199380204 |     150732486 |
60:     111724435 |     81064669 |     57684232 |     40370308 |     27891449 |
65:     19060657 |     12808595 |     8396577 |     5343754 |     3276307 |
70:     1916846 |     1060050 |     546518 |     257605 |     106930 |
75:     38185 |     10437 |     1866 |     132 |     0 |
T:131071744000
***** Total Running Time of Kernel = 66.49196 sec *****
```

Experiment Findings:

Of the four kernels, PDH_kernel1, PDH_kernel2, PDH_kernel3, and PDH_kernel4, presently, PDH_kernels 2 and 3 run the fastest, with 3 being slightly faster.

PDH kernel3 Algorithm:

Listed below is the pseudocode used to implement the PDH_kernel3 algorithm:

```
_global__ void PDH_kernel3(unsigned Long Long* d_histogram,
                           double* d_atom_x_list,
                           double* d_atom_y_list,
                           double* d_atom_z_list,
                           Long Long acnt, double res)
{
    //where shared memory size is 3*BLOCKSIZE*sizeof(double)
    extern __shared__ double R[];

    int cur_id = blockIdx.x * blockDim.x + threadIdx.x;
    int i, j, h_pos;
    double Lx, Ly, Lz, Rx, Ry, Rz;
    double dist;
    if(cur_id < acnt)
    {
        Lx = d_atom_x_list[cur_id];
        Ly = d_atom_y_list[cur_id];
        Lz = d_atom_z_list[cur_id];
        for(i = blockIdx.x + 1; i < gridDim.x; i++)
        {
            cur_id = i * blockDim.x + threadIdx.x; //only valid threads may load into shared memory for block i
            if(cur_id < acnt)
            {
                R[threadIdx.x] = d_atom_x_list[cur_id];
                R[threadIdx.x + blockDim.x] = d_atom_y_list[cur_id];
                R[threadIdx.x + blockDim.x*2] = d_atom_z_list[cur_id];
            }
            __syncthreads();
            for(j = 0; j < blockDim.x; j++)
            {
                cur_id = i * blockDim.x + j; //now this prevents us from writing junk data for thread j
                if(cur_id < acnt)
                {
                    Rx = R[j];
                    Ry = R[j + blockDim.x];
                    Rz = R[j + blockDim.x*2];
                    dist = sqrt((Lx - Rx)*(Lx - Rx) + (Ly - Ry)*(Ly - Ry) + (Lz - Rz)*(Lz - Rz));

                    h_pos = (int)(dist/res);
                    atomicAdd(&d_histogram[h_pos], 1);
                }
            }
            __syncthreads();
        }
    }

    //now load the L values into R
    R[threadIdx.x] = Lx;
    R[threadIdx.x + blockDim.x] = Ly;
    R[threadIdx.x + blockDim.x*2] = Lz;
    __syncthreads();
    for(i = threadIdx.x + 1; i < blockDim.x; i++)
    {
        cur_id = blockIdx.x * blockDim.x + i; //we only proceed with valid threads for each thread i
        if(cur_id < acnt)
        {
            Rx = R[i];
            Ry = R[i + blockDim.x];
            Rz = R[i + blockDim.x*2];
            dist = sqrt((Lx - Rx)*(Lx - Rx) + (Ly - Ry)*(Ly - Ry) + (Lz - Rz)*(Lz - Rz));

            h_pos = (int)(dist/res);
            atomicAdd(&d_histogram[h_pos], 1);
        }
    }
}
```

Techniques used for the GPU kernels:

- 1) I converted the data structures from an Array of Vectors into 3 separate arrays of doubles, one for each axis. This was to enable Structure of Array accesses in the memory, which would help with memory coalescing. Also I converted the histogram buckets into an array of unsigned long long int instead, since that was the only piece of data in that struct, it made sense to use an array directly.
- 2) For the fastest kernel (PDH_kernel3), I implemented a tiling algorithm based on Algorithm 3 of the SDH paper, without histogram privatization, where L is stored in 3 double registers, and R in a dedicated shared memory, of size $\text{blockDim.x} * 3 * \text{sizeof}(\text{double})$. Bounds checking is used to ensure correctness regardless of block size or array size.
- 3) I also made sure to write all kernel code inline (that is, no external function calls except for sqrt, as sadly the intrinsic functions introduced errors). This way, we could avoid any penalty incurred by calling functions as much as possible. Loops in the code are aligned as much as possible (with syncthreads to ensure correctness), so that for almost all blocks, the warps ran in lockstep, so that no divergence arises until the very last block (which cannot be avoided due to edge cases).
- 4) Correctness was ensured by the following steps:
 - a. Ensuring that the original naïve algorithm matched the project 1 code for the CPU in smaller test cases using the difference histogram ($T = 0$), so as to experimentally determine correctness with CPU code.
 - b. For each output histogram that was very large (512k data points) for each subsequent algorithm kernel, I compared the output to the naïve algorithm (which was previously established correct) by using an online diffchecker on the histogram buffer contents. When the diffchecker confirmed the two buffers identical, that established the subsequent kernels as correct as well
- 5) I used the CUDA event timing code as recommended by the project 2 specification.
- 6) After much experimentation, it has been determined that for virtually any input size, the algorithms shown perform the best when block size is 32

References:

- Algorithms and Framework for Computing 2-body Statistics on GPUs, by Tu et al.

Issues:

- I have gone through the paper and added the Tiling algorithm, the histogram privatization, and even the lane id optimization (up to and including Algorithm 4). However, despite these changes, only the Tiling algorithm showed any speedup, and it was marginal compared to the naïve algorithm. Interestingly, if I change the datatype for the atoms from double to float, although the results are incorrect (because of precision loss), the speedup follows the trend exactly described in the paper. Thus my kernels seem to have a slowdown issue regarding double precision that does not exist with floating point precision. It should be noted though that the kernels submitted are accurate despite this. Just unusually slower than I expected.
- When I attempted the Tiled kernel with histogram privatization, I came across a strange issue. As I increased the block size, the accuracy of the kernel changed. At 32 block size, and at 512-

Tanmay J. Kotha (U31037110)

Programming Massively Parallel Systems

CIS 6930

1024 block size, the results are correct. But from 64-256 block size (up to 512 that is, if you're not using powers of 2), the results slightly differed (very slightly, by roughly .0000001%). I am presently not sure what is causing this issue, as the algorithm seems correct (and even is correct at certain block sizes).

- Even when the histogram privatization was fully correct, I found it to be slightly slower than the tiled algorithm.