Ex 1: CUDA Kernel Statistics 1x1:

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ------------ ---------- ---------- -------------------------------------------

100.0 2324192252 1 2324192252.0 2324192252 2324192252 addVectorsInto(float\*, float\*, float\*, int)

Ex 2: CUDA Kernel Statistics 1024x1024:

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ----------- --------- --------- -------------------------------------------

100.0 117072495 1 117072495.0 117072495 117072495 addVectorsInto(float\*, float\*, float\*, int)

Ex 3: CUDA Kernel Statistics 256 x 256:

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ----------- --------- --------- -------------------------------------------

100.0 137745480 1 137745480.0 137745480 137745480 addVectorsInto(float\*, float\*, float\*, int)

Ex 3: CUDA Kernel Statistics 1024 x 256:

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ----------- --------- --------- -------------------------------------------

100.0 156402367 1 156402367.0 156402367 156402367 addVectorsInto(float\*, float\*, float\*, int)

Ex 4: CUDA Kernel Statistics 64 x 64:

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ----------- --------- --------- -------------------------------------------

100.0 149091484 1 149091484.0 149091484 149091484 addVectorsInto(float\*, float\*, float\*, int)

Ex 5: CUDA Kernel Statistics 512 x 512:

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ----------- --------- --------- -------------------------------------------

100.0 141202799 1 141202799.0 141202799 141202799 addVectorsInto(float\*, float\*, float\*, int)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CUDA Kernel Statistics (1,1):

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ------------ ---------- ---------- -------------------------------------------

100.0 2311179867 1 2311179867.0 2311179867 2311179867 addVectorsInto(float\*, float\*, float\*, int)

CUDA Kernel Statistics (32, (threadsPerBlock+N/threadsPerBlock)+1):

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ----------- --------- --------- -------------------------------------------

100.0 132060904 1 132060904.0 132060904 132060904 addVectorsInto(float\*, float\*, float\*, int)

CUDA Memory Operation Statistics (by time):

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- ------- ------- ------- ---------------------------------

100.0 29881575 4886 6115.8 2174 160095 [CUDA Unified Memory memcpy HtoD]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

---------- ---------- ------- ------- ------- ---------------------------------

131072.000 4886 26.826 4.000 960.000 [CUDA Unified Memory memcpy HtoD]

CUDA Memory Operation Statistics (by time):

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- ------- ------- ------- ---------------------------------

100.0 21200323 768 27604.6 1599 160064 [CUDA Unified Memory memcpy DtoH]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

---------- ---------- ------- ------- -------- ---------------------------------

131072.000 768 170.667 4.000 1020.000 [CUDA Unified Memory memcpy DtoH]

## **Asynchronous Memory Prefetching**

**CUDA Memory Operation Statistics (by time):**

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- ------- ------- ------- ---------------------------------

60.9 41 631 267 832 50037.6 1599 320343 [CUDA Unified Memory memcpy DtoH]

39.1 26 702 163 3027 8821.3 2143 128444 [CUDA Unified Memory memcpy HtoD]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

---------- ---------- ------- ------- -------- ---------------------------------

262144.000 832 315.077 4.000 2048.000 [CUDA Unified Memory memcpy DtoH]

131072.000 3027 43.301 4.000 764.000 [CUDA Unified Memory memcpy HtoD]

**CUDA Memory Operation Statistics (by time):**

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- ------- ------- ------- ---------------------------------

100.0 21 198 570 768 27602.3 1599 160637 [CUDA Unified Memory memcpy DtoH]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

---------- ---------- ------- ------- -------- ---------------------------------

131072.000 768 170.667 4.000 1020.000 [CUDA Unified Memory memcpy DtoH]

**CUDA Memory Operation Statistics (by time):**

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- ------- ------- ------- ---------------------------------

55.8 26 758 415 3062 8738.9 2175 128318 [CUDA Unified Memory memcpy HtoD]

44.2 21 199 639 768 27603.7 1599 159997 [CUDA Unified Memory memcpy DtoH]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

---------- ---------- ------- ------- -------- ---------------------------------

131072.000 768 170.667 4.000 1020.000 [CUDA Unified Memory memcpy DtoH]

131072.000 3062 42.806 4.000 764.000 [CUDA Unified Memory memcpy HtoD]

**CUDA Kernel Statistics (without cudaMemPrefetchAsync):**

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ---------- -------- -------- -------------------------------------------

96.6 48 818 082 3 16272694.0 16028696 16466357 initWith(float, float\*, int)

3.4 1 709 365 1 1709365.0 1709365 1709365 addVectorsInto(float\*, float\*, float\*, int)

CUDA Memory Operation Statistics (by time):

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- ------- ------- ------- ---------------------------------

100.0 21 199 034 768 27602.9 1631 160671 [CUDA Unified Memory memcpy DtoH]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

---------- ---------- ------- ------- -------- ---------------------------------

131072.000 768 170.667 4.000 1020.000 [CUDA Unified Memory memcpy DtoH]

**CUDA Kernel Statistics (with cudaMemPrefetchAsync for GPU):**

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- --------- ------- ------- -------------------------------------------

52.1 1 861 073 3 620357.7 614971 625307 initWith(float, float\*, int)

47.9 1 708 370 1 1708370.0 1708370 1708370 addVectorsInto(float\*, float\*, float\*, int)

CUDA Memory Operation Statistics (by time):

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- ------- ------- ------- ---------------------------------

100.0 21 201 113 768 27605.6 1630 160639 [CUDA Unified Memory memcpy DtoH]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

---------- ---------- ------- ------- -------- ---------------------------------

131072.000 768 170.667 4.000 1020.000 [CUDA Unified Memory memcpy DtoH]

**CUDA Kernel Statistics (with cudaMemPrefetchAsync for CPU):**

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- --------- ------- ------- -------------------------------------------

52.0 1 860 793 3 620264.3 614387 625587 initWith(float, float\*, int)

48.0 1 716 956 1 1716956.0 1716956 1716956 addVectorsInto(float\*, float\*, float\*, int)

CUDA Memory Operation Statistics (by time):

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- -------- ------- ------- ---------------------------------

100.0 20 434 873 64 319294.9 319065 320281 [CUDA Unified Memory memcpy DtoH]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

---------- ---------- -------- -------- -------- ---------------------------------

131072.000 64 2048.000 2048.000 2048.000 [CUDA Unified Memory memcpy DtoH]

# SAXPY (without pre fetch)

CUDA Kernel Statistics:

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- ---------- -------- -------- -----------------------

100.0 24356523 1 24356523.0 24356523 24356523 saxpy(int\*, int\*, int\*)

CUDA Memory Operation Statistics (by time):

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- ------- ------- ------- ---------------------------------

99.8 9836431 1044 9421.9 2142 128220 [CUDA Unified Memory memcpy HtoD]

0.2 24381 4 6095.3 1663 10591 [CUDA Unified Memory memcpy DtoH]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

--------- ---------- ------- ------- ------- ---------------------------------

49152.000 1044 47.080 4.000 764.000 [CUDA Unified Memory memcpy HtoD]

128.000 4 32.000 4.000 60.000 [CUDA Unified Memory memcpy DtoH]

# SAXPY (with pre fetch)

CUDA Kernel Statistics:

Time(%) Total Time (ns) Instances Average Minimum Maximum Name

------- --------------- --------- -------- ------- ------- -----------------------

100.0 198268 1 198268.0 198268 198268 saxpy(int\*, int\*, int\*)

CUDA Memory Operation Statistics (by time):

Time(%) Total Time (ns) Operations Average Minimum Maximum Operation

------- --------------- ---------- -------- ------- ------- ---------------------------------

99.7 8196460 24 341519.2 339801 343672 [CUDA Unified Memory memcpy HtoD]

0.3 24509 4 6127.3 1663 10560 [CUDA Unified Memory memcpy DtoH]

CUDA Memory Operation Statistics (by size in KiB):

Total Operations Average Minimum Maximum Operation

--------- ---------- -------- -------- -------- ---------------------------------

49152.000 24 2048.000 2048.000 2048.000 [CUDA Unified Memory memcpy HtoD]

128.000 4 32.000 4.000 60.000 [CUDA Unified Memory memcpy DtoH]