CSE 6230, Fall 2014

Lab 6

Ankit Srivastava, gtID: 902838136

October 17, 2014

1

Question
Report the speed of data transfer using explicit copies vs. implicit copies on pinned memory blocks.
Solution
I built and ran the code for reversing arrays and got the following running times for, 121 trials:

- Reversal with explicit copies took 0.0375731 seconds
  $\implies$ Observed effective bandwidth was $\boxed{3.57217 \text{ GB/s}}$
- Reversal without explicit copies took 0.0165643 seconds
  $\implies$ Observed effective bandwidth was $\boxed{8.10281 \text{ GB/s}}$

2

Question
Report the results for GPU-to-GPU communication with and without using an explicit intermediate buffer on the CPU.
solution
I built and ran the code for GPU-to-GPU communication and got the following results:

- Time taken for GPU-to-GPU copy, buffered through CPU memory, for 15 trials was 0.0668967 seconds
  $\implies$ Observed effective bandwidth was $\boxed{2.00634 \text{ GB/s}}$

- Time taken for direct GPU-to-GPU copy for 26 trials was: 0.0399213 seconds
  $\implies$ Observed effective bandwidth was $\boxed{3.36206 \text{ GB/s}}$

3

Question
After you run the codes, inspect the .o* job script output. Report the nodes used, problem size, effective GFLOP/s. Do this for both the CPU (MKL BLAS) and GPU (CUDA) programs.
Solution

- For the CPU (MKL BLAS) run, following values were obtained:
  Nodes used: $\boxed{\text{jinx5, jinx7}}$
  Problem dimension: $\boxed{n = 4096}$
  Effective performance: $\boxed{210.6 \text{ GFLOP/s}}$

- For the initial CUDA run, following values were obtained:
  Nodes used: $\boxed{\text{jinx9, jinx10}}$
  Problem dimension: $\boxed{n = 4096}$
  Effective performance: $\boxed{647.7 \text{ GFLOP/s}}$

4

Question
Submit your (hopefully working) code by transferring your repository to us, as you've done in previous labs. Report the nodes used and effective GFLOP/s for your implementation.
Solution
I have transferred the repository back after making necessary changes for using pinned memory.
Observed values for the final implementation were:
Nodes used: $\boxed{\text{jinx9, jinx10}}$
Problem dimension: $\boxed{n = 4096}$
Effective performance: $\boxed{690.2 \text{ GFLOP/s}}$