18-645: How to Write Fast Code

Project 3:- Memory Design and Kernel Implementation

This phase of the semester project is to implement the kernels previously designed, and to examine the data layout required to ensure that the kernels are running efficiently. At the end of this phase, a combined report is required.

Requirements. The final design document should contain the following:

1. *Updated Kernel Design and Implementation.* You are expected to have implemented your kernels and verified that they are correct. This report should report the design of the kernels that was implemented.

   Please highlight any differences and the reasons for those differences between the kernel implementation and the previous design. We are looking for reasons why this new kernel design is more performant than the previous design. For example, you might have increased the size of the kernel. You might have also fused some of your previous kernels into a single kernel. If the code is a faithful implementation of the kernel design previously submitted, a single statement stating that there is no changes is sufficient.

   Please restrict this section to no more than 2 pages per implemented kernel. Please also indicate which team member(s) worked on each kernel. If multiple team members worked on a single kernel, please assign an approximate percentage of the work for that particular kernel to each team members.

2. *Memory Hierarchy Design.* Based on the implemented kernel design, you should examine if an intermediate data layout is required. If it is determined that an intermediate layout is required, a description of the layout is expected. Similarly, if no intermediate layout is required, explain why.

   Note that designing the memory layout for individual kernels may result in multiple intermediate layouts. As such, it may be useful to examine if a single memory layout could work across multiple kernels.

3. *Performance Plots.* You should include performance plots for each kernel that shows the following features across a number of data sizes (when appropriate):

   - *Theoretical Peak.* You should have identified what is the bottleneck for your kernel. Compute the theoretical peak, and plot that as a horizontal line to indicate peak performance.
   - *Kernel Performance.* The performance of your implementation should be included on the same plot as a measurement against theoretical peak.

Submission. The report (in pdf format) is to be submitted by the due date. *Submissions through Canvas must be submitted individually.*