3. Threads within a block can share data using “shared” variables declared within the kernel with the prefix “\_\_shared\_\_”
4. No. All threads in a warp are either running the same instruction or sleeping. They all start and finish at the same time regardless of the number of instruction cycles any particular one needs.
5. An if-then-else block is executed in the following manner:
   1. All threads calculate the value of the conditional
   2. The “then” block is executed
      1. If the conditional result was true for a thread, the thread executes these instructions
      2. If the conditional result was false, the thread sleeps
   3. The “else” block is executed
      1. If the conditional result was true for a thread, the thread sleeps
      2. If the conditional result was false, the thread executes these instructions
   4. All threads continue executing the remaining instructions
6. Unified virtual addressing keeps variables stored on the host (computer/RAM) consistent with variables of the same name and address in the device (GPU). This is done automatically as needed without the programmer needing to explicitly copy data back and forth. It allows the same variables/pointers/data to be used in both the host and device code without having to worry about data consistency or valid host/device addresses.