Problem 1. (15 points) For memory consistency models, in the order of from the strictest to the relaxed, there are strict consistency model, Sequential consistency model, Causal consistency model, and PRAM consistency model. For each model, give an example involving only two processors whose memory reference operations satisfy this model but violate the stricter ones. That is, give an example which is sequential consistent but not strict consistent, give another example which is causal consistent but not sequential consistent, etc. The example must not be isomorphic to any example in the notes, nor may it be isomorphic to an example in the notes with some extra references added.

Problem 2. (20 points) Consider the following code fragment; executed on 3 different processors. All variables are initialized to 0.

<table>
<thead>
<tr>
<th></th>
<th>P_1</th>
<th>P_2</th>
<th>P_3</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>A = 1</td>
<td>while (flag == 0);</td>
<td>C = 1</td>
</tr>
<tr>
<td>C</td>
<td>B = 1</td>
<td>x = A</td>
<td>y = B</td>
</tr>
<tr>
<td>D</td>
<td>flag = 0</td>
<td>C = 0</td>
<td>z = C</td>
</tr>
</tbody>
</table>

(a) Which operations are conflicting?
(b) Which operations are competing?
(c) Which operations are synchronization operations?
(d) How should these operations be labeled in weak ordering?
   i. How should these operations be labeled in release consistency?
   ii. Under sequential consistency, give a set of values for x, y, and z that would not be possible.

Problem 3. (15 points) Suppose that you are trying to decide which kind of interconnection network should be used on a new multiprocessor computer system. Your design requirements specify that the system should use message passing for interprocessor communication. You have also learned that your system will frequently be used to transpose $4 \times 4$ matrices.

(a) What is the sum of the lengths of all the messages required to transpose a $4 \times 4$ matrix using the following interconnection networks: cyclic shift, Illiac IV mesh, a mesh where rows and columns are arranged as cycles of four nodes (similar to DAP without wraparound connections), barrel shifter, and cube networks? Assume that the matrix is stored as a vector of length 16 and that the $i$th element of the vector is held in the $i$th processor's local memory. Distance is measured by counting the number of links that a message must traverse to reach its destination.

(b) Why is total distance traveled by all messages not a good measure of how long it will take to execute a parallel algorithm?
(c) Show that a $4 \times 4$ matrix can be transposed on a 16-node Illiac IV Mesh in less than five message-passing cycles. A message-passing cycle is the time for messages to be passed from one node to an adjacent node plus the time for each node to process all of the messages it received during that cycle. Assume that each node can process (receive, determine routing, and send) four messages at the end of every message cycle; however, each link can only be used by one message per cycle. Also assume that not all messages have to travel in the same direction during a message passing cycle.

**Problem 4.** (10 points) When a variable exhibits migratory sharing, a processor that reads the variable will be the next one to write it. What kinds of protocol optimizations could you use to reduce traffic and latency in this case, and how would you detect the situation dynamically? Describe a scheme or two in some detail. *(Hint: Think about different ways to give a requesting processor exclusive access to a block, even when only shared access is requested when it is likely that the data exhibits this migratory sharing pattern)*

**Problem 5.** (20 points) In the Origin implementation, incoming request messages to the memory/directory interface are given priority over the incoming responses unless there is a danger of responses being starved. Give reasons for this choice of giving priorities to requests was made? Describe some methods for how you might detect when to invert the priority and what would be the danger with responses being starved?