CS/ECE 757: ADVANCED COMPUTER ARCHITECTURE II
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN—MADISON

Prof. Mark D. Hill
TA David Mulvihill

Midterm Examination I
In-Class
Wednesday, February 18, 2004
Weight: 25%

1:15 minutes.

CLOSED BOOK, etc., but one cheat sheet allowed (two-sided 8.5x11 page).

The exam is two-sided and has EIGHT pages, including two blank pages at the end.
Plan your time carefully, since some problems are longer than others.

NAME: __________________________________________________________

ID# ______________________________________________________________

<table>
<thead>
<tr>
<th>Problem Number</th>
<th>Maximum Points</th>
<th>Actual Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>
Problem 1: Dynamic Load Balancing (12 points)

(a) What is dynamic load balancing?

(b) How might the author of an application or library implement dynamic load balancing?

(c) What factors make dynamic load balancing preferred to static load balancing?
Problem 2: Theory and Scaling (12 points)

Consider parallel multiplication of two N by N matrices.

(a) In theory, how long does this matrix multiply take on an unlimited number of ideal processors (e.g., a PRAM). Your answer should be a function of N, e.g., O(n). Justify your answer.

(b) Assume that executing this matrix multiply on near-ideal P-processor multiprocessor takes time T. About how long will it take to execute this matrix multiply with 2P processors? Justify your answer. (If needed, assume that both P and 2P are much less than N).
Problem 3: Coherence (12 points)

(a) Consider invalidation-based coherence protocols using either two states (*VI:* Valid & Invalid) or three states (*MSI:* Modified, Shared, & Invalid). Why might one protocol be preferred to the other for (i) write-back or (ii) write-through?

(b) In most snooping coherence protocols, memory must determine whether it should respond to a coherence request with data. Discuss alternative methods of implementing this determination.
Problem 4: Miscellaneous (12 points)

(a) Write pseudo-code for a library function that *atomically swaps* the values of two memory locations (given as arguments R and S) on shared memory hardware that supports *load*, *store*, and *test-and-set*.

(b) *Translation Lookaside Buffers* (TLBs) cache *Page Table Entries* (PTEs) from memory. Since a cache coherence protocol keeps memory and caches coherent, does it also keep TLBs coherent? Justify your answer.
Problem 5: Communication (12 points)

Consider a parallel program where a processor P2’s thread determines that to continue computing it needs a B-byte array of data recently created by a processor P1’s thread. Assume that P1 does not know this in advance.

(a) What actions must P2 do to obtain the data using *shared memory*?

(b) What actions must P2 do to obtain the data using *message passing*?

(c) What factors determine whether using shared-memory or message-passing might be more effective?
Scratch Sheet 2 of 2 (in case you need additional space for some of your answers)