CS3411 Test #4, 27 November 2007. **Must be turned in in class on Thursday, 29 November.** Answer all questions. The number of points (out of 100) each problem is worth is shown in (parentheses). Be as concise as you can while still answering the question. Open book, open Google, open notes, open ITV videos, open past test keys, but it must be an individual effort. No talking about the test, not even at a high level.

1. (30) Write "true" or "false".

* On a hard drive, a track can be made up of many sectors. **True**

* On a hard drive, a sector can be made up of many tracks. **False**

* An AS/400 traces its roots back to the IBM System/360, more-so than to the System/38. **False**

* The MIPS ISA traces its roots back to the IBM System/360, more-so than to the System/38. **True**

* RAID 0 has no striping of data, and therefore no performance benefit. **False**

* RAID 4 improves on RAID 3 by adding a parity check, where RAID 3 simply mirrors the data. **False**

* RAID 5 has a bottleneck on back-to-back writes for the parity disk, something that is not a problem for RAID 4. **False**

* A device would typically connect to the south bridge if there was not
enough memory bandwidth for its DMA needs on the north bridge.

False

* Graphics cards exploit SIMD concepts, that is, Single Instruction Multiple Data, for increased performance.

True

* DMA allows the CPU to do useful work instead of spending time doing data transfers to/from I/O devices.

True

* The AS/400 and Series i line of computers dedicates an I/O processor to each I/O device because the applications developed for this series of computers are typically CPU-bound, large scientific computations and the CPU needs to be free to maximize the number of GFLOPS.

False

* The RPG language is specialized for I/O-intensive report generation from large sets of data such as customer records.

True

* A typical GPU has a two-dimensional memory that allows for interpolation.

True

* One reason that stacking multiple metal layers in a CPU does not truly make it a 3D chip is that the vias are more expensive in terms of delay, implementation, etc. than a normal wire.

True

* The power wall is the only reason why we are trending towards multicore chips.

False
2. (50) Multiple Choice (circle all that apply):

*Which of the following types of issues would a GPU designer need to worry about?

a. Bus bandwidth to/from the GPU's memory
b. TLB flushes during context switches
c. Choosing a good virtual memory page size
d. Delay from the GPU's memory
e. Cache miss and hit rates for caches in the GPU

*Which of the following will you typically **not** find on a graphics GPU?

a. Virtual memory support for process memory protection
b. Precise interrupts
c. A rasterizer
d. The ability to do floating point operations

*Which of these are ways of ameliorating data hazards in a scalar pipeline?:

a. A multi-ported memory
b. Data forwarding
c. A bubble/stall
d. Delay slot

*Which of the following data hazards can be completely resolved in the MIPS pipeline described in the book using data forwarding (no bubble required)?

a. `add $t0 $t1 $t2` followed by `sw $t0 0($t5)`

b. `add $t0 $t1 15` followed by `sw $t0 0($t5)`

c. `lw $t0 0($t5)` followed by `add $t0 $t1 15`

d. `lw $t0 0($t5)` followed by `add $t0 $t1 $t2`

*This question has an error, we'll talk about it in class.*
*Melanie mentioned that the Itanium had a slightly better power/performance ratio than their model predicted. Which of these are plausible explanations for this?
  a. The Itanium has a single-cycle datapath
  b. Itanium has lower resistance than other metals
  c. The Itanium supports static multiple issue
  d. The Itanium doesn't support virtual memory

* Differences between scaling of biological organisms vs. scaling of CPUs include:
  a. The number of dimensions, i.e., 3 vs. 2ish
  b. Transistors can continually be made smaller, cells typically cannot
  c. The exponent they scale with is different
  d. CPUs always stayed the same size, only transistor density has changed

3. (5) There is a rule that empirically relates by a power law the number of pins or wires coming out of an integrated circuit with the number of internal components (i.e. logic gates). What is this rule called?

   Rent's Rule

4. (5) In the discussion of die yield on a wafer in Chapter 1 of the book, what are the two reasons that die cost increases as die size increases?

   1. Die yield decreases due to a higher probability of defects
   2. You can't fit as many larger dies on a wafer
5. (10) Write down, for each transformation, what compiler optimization has occurred (note that these optimizations are performed by the compiler, typically on an intermediate form, not on the C code or assembly code itself):

<table>
<thead>
<tr>
<th>Before optimization</th>
<th>After optimization</th>
<th>What?</th>
</tr>
</thead>
<tbody>
<tr>
<td>for (i=0;i&lt;4;i++) {</td>
<td>z = 4;</td>
<td>Code Motion</td>
</tr>
<tr>
<td>z = 4;</td>
<td>for (i=0;i&lt;4;i+=2) {</td>
<td></td>
</tr>
<tr>
<td>x[i] = y[z-i];</td>
<td>x[i] = y[z - i];</td>
<td></td>
</tr>
<tr>
<td>}</td>
<td>}</td>
<td></td>
</tr>
<tr>
<td>addi $s3 $zero 32</td>
<td>addi $s3 $zero 172</td>
<td></td>
</tr>
<tr>
<td>addi $s3 $s3 40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>addi $s3 $s3 100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i = i * 8 + 100;</td>
<td>i = i &lt;&lt; 3 + 100;</td>
<td>Strength Reduction</td>
</tr>
<tr>
<td>for (i=0;i&lt;100;i++) {</td>
<td>for (i=0;i&lt;100;i+=2) {</td>
<td>Loop Unrolling</td>
</tr>
<tr>
<td>x[i] = y[i] * z;</td>
<td>x[i] = y[i] * z;</td>
<td></td>
</tr>
<tr>
<td>}</td>
<td>x[i+1] = y[i+1] * z;</td>
<td></td>
</tr>
<tr>
<td>add eax 5</td>
<td>add eax 5</td>
<td>Dead Code Elimination</td>
</tr>
<tr>
<td>cmp eax eax</td>
<td>shr eax 2</td>
<td></td>
</tr>
<tr>
<td>jne LL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sub $t5 $t6 $t7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LL:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>shr eax 2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Some of these could be construed as multiple optimizations, but write whatever the best fit is.)