1. Power Wall is the problem of...

2. Frequency Wall is the problem of...

3. Memory Wall is the problem of...

4. An accelerator is... (extra credit for a cool definition)

5. The PetaFlop barrier was breached by a system of (circle one):
   a) Cell processors
   b) GPUs
   c) a homogeneous system
   d) a hybrid design including Cell processor
   e) a hybrid design including GPUs
   f) a hybrid design including Cell processors and GPUs

6. The fastest unclassified supercomputer in the world is built of (circle one):
   a) Cell processors
   b) GPUs
   c) a homogeneous system
   d) a hybrid design including Cell processor
   e) a hybrid design including GPUs
   f) a hybrid design including Cell processors and GPUs

7. We can expect the next generation Cell processor (circle one):
   a) To have 32 SPEs
   b) To have longer vectors
   c) To have higher memory bandwidth
   d) none of the above
   e) all of the above

8. We can expect the next generation Nvidia GPUs (circle one):
   a) To double the number of transistors
   b) To double the performance
   c) To be more power efficient
   d) To double memory bandwidth
   e) none of the above
   f) all of the above

9. The performance price of not vectorizing Cell code was (circle one):
   a) minimal
   b) half the performance of vectorized code
   c) one fourth the performance of vectorized code
   d) more than an order of magnitude less performance than vectorized code

10. In order to get good performance all threads in a ……. have to execute the same instructions (branch together).

11. The advantage of the FMA instruction ovet the MADD instruction is (circle one):
    a) twice the performance
    b) twice the accuracy
    c) better performance
    d) better accuracy

12. A GPU SM can load and store to and from (circle all that apply):
    a) the GPU's device memory (on-board, off-chip)
    b) the SM's local shared memory (on-chip)
    c) the shared memories of other SM's
    d) the L1 caches of other SMs

13. Threads can communicate through registers:
    a) TRUE
    b) FALSE

14. Threads can communicate through shared memory:
    a) TRUE
    b) FALSE

15. Thread blocks can communicate through shared memory:
    a) TRUE
    b) FALSE

16. SMs can communicate through shared memory:
    a) TRUE
    b) FALSE

17. GPU thread divergence happens if (circle one):
    a) different SMs execute different code
    b) different thread blocks in one SM execute different code
    c) different warps in one thread block execute different code
    d) none of the above
    e) all of the above

18. What happens when two GPU threads access the same address in shared memory? (circle one):
    a) bank conflict
    b) broadcast
    c) coalescing

19. What happens when two GPU threads access the same bank in shared memory? (circle one):
    a) bank conflict
    b) broadcast
    c) coalescing

20. Fermi will perform one memory transaction if (circle all that apply):
    a) consecutive threads in a warp access consecutive 128 bytes
    b) threads in a warp access consecutive 128 bytes in a random pattern
    c) all threads in a warp access the same address
    d) threads in a warp access cache-line aligned 128 bytes in a random pattern