1. Power Wall is the problem of...

2. Frequency Wall is the problem of...

3. Memory Wall is the problem of...

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

5. 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

6. The SPE can load and store to and from (circle all that apply):
   a) its own local memory
   b) local memories of all other SPEs
   c) local memories of SPEs located next to it on the EIB
   d) L1 cache
   e) L2 cache
   f) the main memory

7. The SPE can read and write using DMA (circle all that apply):
   a) its own local memory
   b) local memories of all other SPEs
   c) local memories of SPEs located next to it on the EIB
   d) L1 cache
   e) L2 cache
   f) the main memory

8. The SPE's Local Store is (circle all that apply):
   a) completely incoherent with other Local Stores and the main memory
   b) coherent with other Local Stores, but incoherent with the main memory
   c) coherent with the main memory, but incoherent with other Local Stores
   d) coherent with L1 cache
   e) coherent with L2 cache
9. **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) L1 cache
   e) L2 cache

10. **The GPU SM can read and write using DMA (circle all that apply):**
    a) its own shared memory
    b) shared memories of all other SMs
    c) L1 cache
    d) L2 cache
    e) the GPU's device memory (on-board, off-chip)

11. **The GPU SM's shared memory is (circle all that apply):**
    a) not coherent with shared memories of other SM
    b) coherent with other SMs, but not device memory (off-chip)
    c) coherent with device memory, but not other SMs
    d) coherent with L1 cache
    e) coherent with L2 cache

12. **The SPE has (circle one):**
    a) equal number of scalar and vector registers
    b) small number of scalar registers and huge number of vector registers
    c) small number of vector registers and huge number of scalar registers
    d) only vector registers
    e) only scalar registers

13. **The GPU SM has (circle one):**
    a) equal number of scalar and vector registers
    b) small number of scalar registers and huge number of vector registers
    c) small number of vector registers and huge number of scalar registers
    d) only vector registers
    e) only scalar registers

14. **The dual-issue capability means that the SPU can issue in one cycle (circle one):**
    a) two load instructions as long as they use different registers
    b) two store instructions as long as they use different registers
    c) two arithmetic instructions as long as they use different registers
    d) two permutation instructions as long as they use different registers
    e) one arithmetic instruction and one permutation instruction as long as they use different registers
    f) all of the above
    g) none of the above

15. **The SPU registers are statically partitioned among all threads:**
    a) TRUE
    b) FALSE

16. **The SPU registers are dynamically shared among all threads:**
    a) TRUE
    b) FALSE
17. The GPU SM's registers are statically partitioned among all threads:
   a) TRUE
   b) FALSE

18. The GPU SM's registers are dynamically shared among all threads:
   a) TRUE
   b) FALSE

19. **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

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

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

22. **Coalescing means (circle one):**
   a) hardware ability to combine accesses to discontinuous memory locations
   b) software necessity to organize accesses such that they reference continuous memory locations