

- 1) (13/40) Let us consider a 2-level-Pap branch-predictor: the Branch History Shift Register (BSHR) has 2 elements ( $m=1$ ) of 12 bits each ( $k=12$ ) and we have 2 Pattern History Tables (PHT), i.e.,  $p=1$ . The predictor uses a finite-state machine represented by a 2-bit saturation-counter ( $j=2$ ). The Branch Target Buffer (BTB) has 4 elements. Draw the complete scheme of the predictor and highlight the connections with the rest of a standard 5-stages pipeline (e.g., MIPS). Represent the content of BHSR, PHT, BTB for the following branch sequence (hexadecimal addresses) BIA $\rightarrow$ BTA (Branch Instruction Address $\rightarrow$ Branch Target Address): 1040 $\rightarrow$ 1000, 1080 $\rightarrow$ 1000, 1040 $\rightarrow$ 1000, 1080 $\rightarrow$ 1000.
  
- 2) (14/40) Write a CUDA program (explain at best the possibilities of parallelization offered by such programming model) to parallelize the dot-product of two vectors of size 10'000 integers.
  
- 3) (13/40) In a quad-core (4 cores) shared-memory multiprocessor the coherence protocol is MESI. Each core has a private direct-access cache ( $A=$ Associativity $=1$ ), the block size is 64 bytes and the total capacity of the cache is 256 bytes. Write the final state of each copy of a given block in each of the 4 caches for the following list of accesses to the cache (notation: core $\rightarrow$ hexadecimal\_address(R for read or W for write)):  
  
0 $\rightarrow$ 0x1000 (R), 1 $\rightarrow$ 0x1000(R), 1 $\rightarrow$  0x1000 (W), 2 $\rightarrow$ 0x1000(R), 3 $\rightarrow$ 0x1000(R), 0 $\rightarrow$ 0x1000(W).