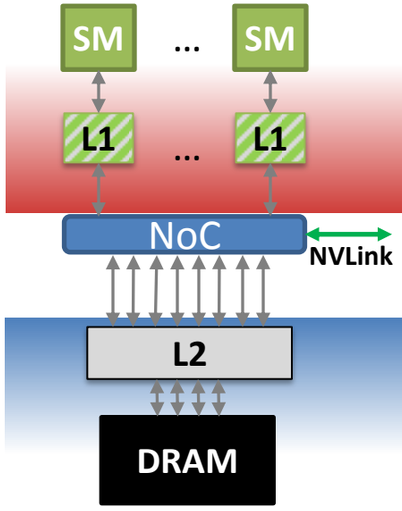
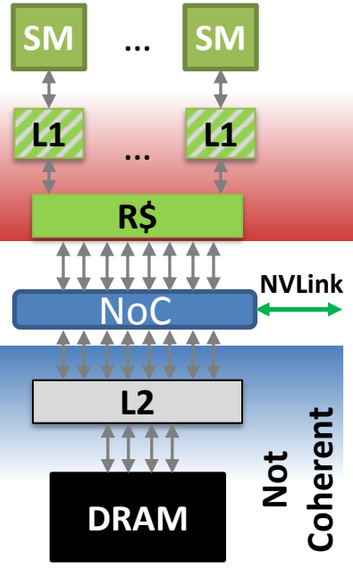
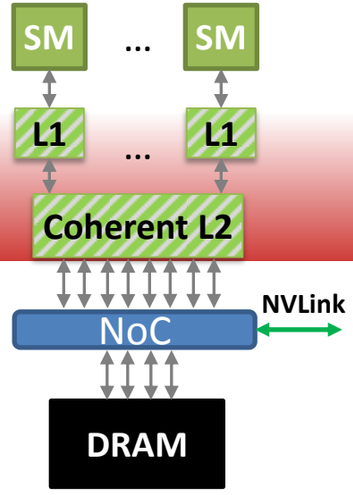


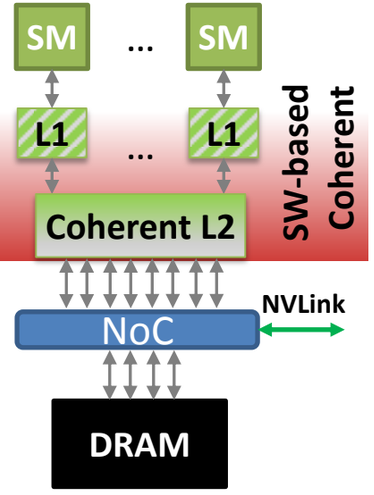
(a) Mem-Side Local Only L2



(b) Static R$



(c) Coherent L2



(d) NUMA-Aware Coherent L2

NUMA-Aware Cache Partitioning Algorithm

0) Allocate half the ways for local and another half for remote data
1) Monitor NVLink and local DRAM outgoing BW
2) If NVLink is saturated and local DRAM BW not
   - *RemoteWays++* and *LocalWays--*
3) If local DRAM BW is saturated and NVLink not
   - *RemoteWays--* and *LocalWays++*
4) If both are saturated
   - *Equalize allocated ways (++ and --)*
5) None of them is saturated
   - *Do nothing*
6) Go back to 1) after *SampleTime* cycles