CS 330 - Winter 2020
Assignment W5

Due: Wednesday, March 11, 2020 (~~start of class~~ **5pm**)

You should submit a physical copy of your written homework. Be sure to include a collaboration statement with your assignment, even if you worked alone. This is worth 2 points.

[40 points] Problem 1 - GPU Scheduling

a) By looking at the abstract, introduction, and conclusion, make a bullet point list of the main contributions of Amert et al. [2017].

b) List two open problems stated as future work.

c) The presentation and the discussion in class focused on thread resources. What are the other two resources constraining execution on NVIDIA GPUs, as described in the paper?

d) The full experiment of the basic GPU scheduling rules is described in detail in Figures 3 and 4, and in Table 1. Consider the time labeled (b) in the plot (corresponding to inset (b) in Figure 4). (i) Why does K2 not execute concurrently with K1? (ii) Why does K5 not execute concurrently with K4, even though K5 is at the head of the EE queue?

e) Figure 5 in the paper depicts a NULL stream scheduling experiment. Given the rules described in the paper and the discussion from class, state in your own words (i) why K3 and K4 do not execute concurrently, and (ii) why K4 and K6 do not execute concurrently.

[24 points] Problem 2 - Uniform Heterogeneous Platforms

a) By looking at the abstract, introduction, and conclusion, make a bullet point list of the main contributions of Yang and Anderson [2017].

b) List two open problems stated as future work.

c) How does the uniform multiprocessor model differ from the identical and unrelated multiprocessor models?

d) The ideal schedule defined on page 3 is highly reminiscent of something else we have seen related to multiprocessor scheduling. What is it similar to, and how is it used in this paper?

[34 points] Problem 3 - The CGLP

a) By looking at the abstract, introduction, and conclusion, make a bullet point list of the main contributions of Nemitz et al. [2019].

b) List two open problems stated as future work.

c) Clearly state the two fundamental problems introduced by nested resource requests, and give a brief sketch illustrating each.

d) Consider the request graph depicted in Figure 4. Assume the resources required by request R_4 are changed from $\{\ell_a, \ell_b\}$ to $\{\ell_b, \ell_d\}$. Draw the new graph. Has the minimum number of required colors changed? If so, give a new coloring of the graph.

e) Consider again the request graph depicted in Figure 4. This time, assume that request R_1 is not in the system. Draw the new graph. Has the minimum number of required colors changed? If so, give a new coloring of the graph.