Cerebras SDK technical overview

- Programming model
    - CS-2 system already supports frameworks like PyTorch and Tensorflow
    - Cerebras SDK allows programmers to write lower-level code to target WSE's microarchs
    - Host CPU (python) targets groups of cores on the WSE, specified by the programmer
    - programmer to specify the communication paths through the on-wafer network for data communication within each group and between the groups
    - **Data movement**: in 32-bit packets called wavelets
        - Each wavelet is tagged with a 5bit virtual channel (color)
    - **Syntax**: similar to C/C++
        - Types, Functions and control structures
        - Modules: allows to import other CSL files
        - Tasks: Implements dataflow programs (mapped to colors)
            - Two ways to trigger a task
                - On incoming wavelet color
                - Programmatic activation and unblock
    - Instructions work on tensors whose operands are specified in data structure descriptors (DSDs)
        - Data structure registers (DSRs) are used to hold DSDs
- Simulator
    - Apparently cycle-accurate
    - Allows for debugging