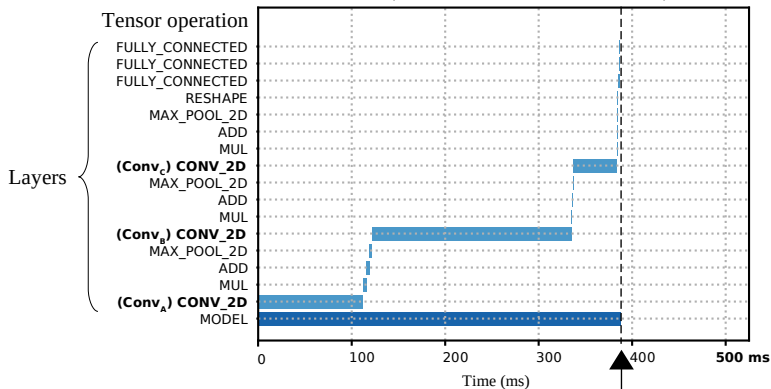


# Runtime Profiling: Latencies and Scheduling on Floating-Point and Hybrid Float 6 Accelerators

| Tensor Op         | CPU (ms) | TP FP32 (ms) | Gain | TP HF6 (ms) | Gain |
|-------------------|----------|--------------|------|-------------|------|
| Conv <sub>A</sub> |          |              |      |             |      |
| Conv <sub>B</sub> |          |              |      |             |      |
| Conv <sub>C</sub> |          |              |      |             |      |

CPU (ARM Cortex-A9 @ 666 MHz)



Total = 387 ms (2.6 inference per second)