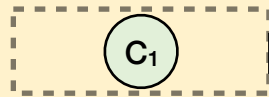
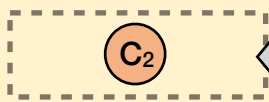


Differentiated Execution Engine (§5.1)

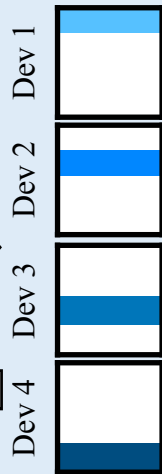
e.g.,
Fine-tuning → *Generation*



Execution:

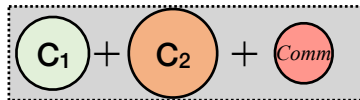


Workload-aware Transformation (§5.2)



Goal (§5.3)

Min:



$+$ $Comm$

