



CPU read + decode



GPU dispatch
(Rank 0)



Inter-GPU collectives



GPU compute
(Triton kernels)



Log / aggregator
write (Rank 0)



Sub-step
(hairline)