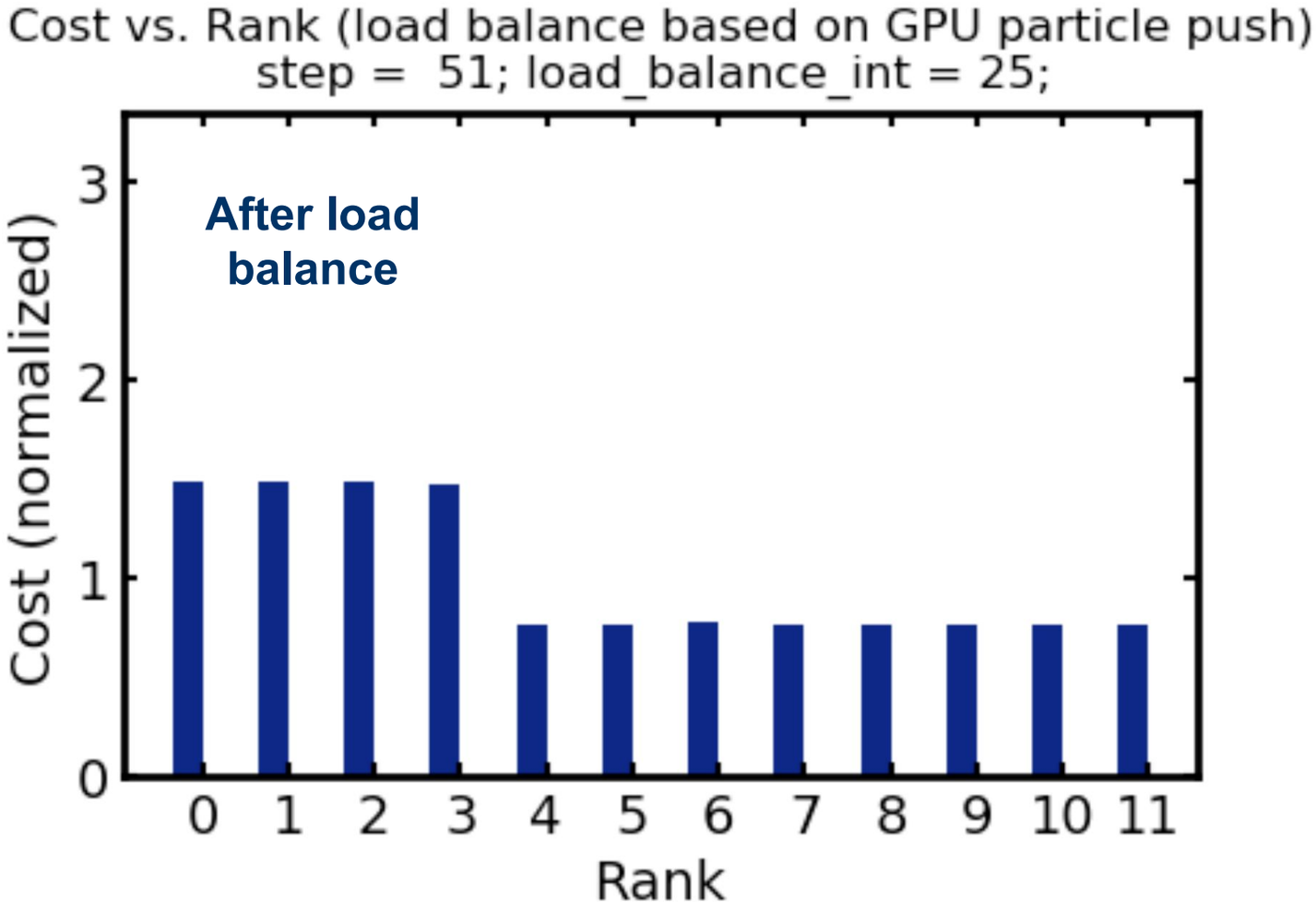
Used CUPTI Callback timing as input to load balancing modules for a WarpX test problem

- Plot to the right shows time evolution of ‘cost’ per GPU
- Load balance every 25 steps (for this case)
- Cost initially imbalanced (ranks, 4 – 7 do most of the work)
- Work is more evenly distributed after step 25



In this test case, get a speedup of ~2x

- Red curve shows case with no load balancing
- Blue curve shows case with load balance every 125 steps
- Prior to the first load balance (step=125), work is unevenly distributed over GPUs
- After load balance, work is distributed more evenly → performance improvement