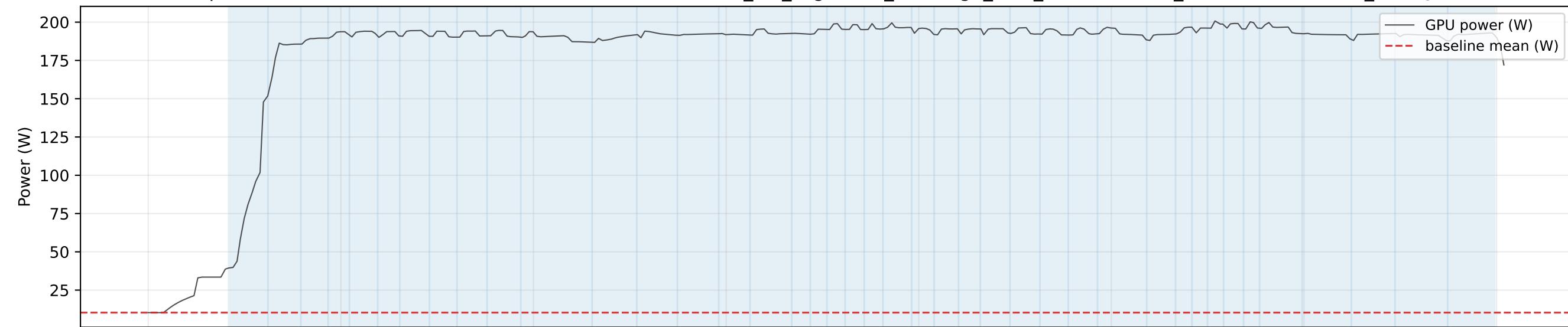


GPU power with LLM decode windows (n=41), baseline=from\_llm\_segments\_csv, align\_shift\_s=0.031, trim\_baseline=workload\_start@29.9s



Delta power vs baseline (decode windows shaded)

