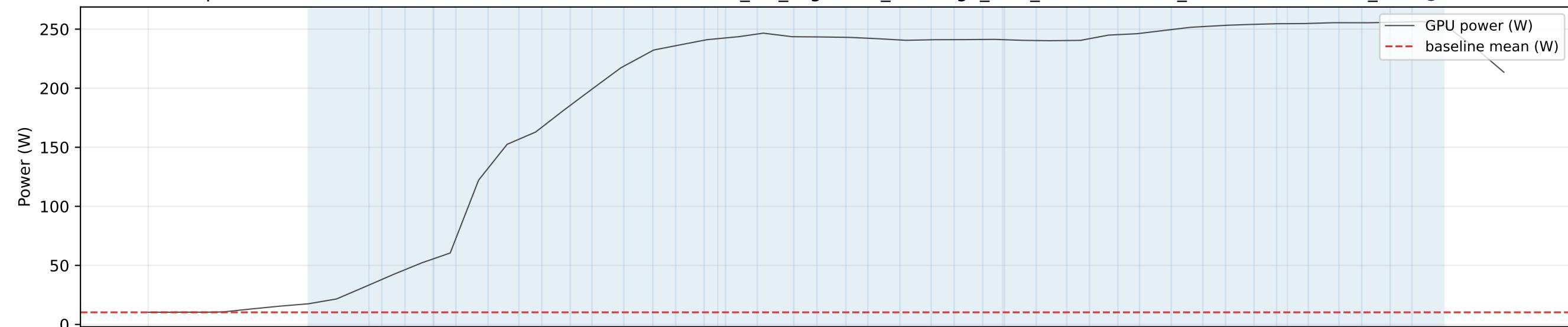


GPU power with LLM decode windows (n=41), baseline=from_llm_segments_csv, align_shift_s=-0.050, trim_baseline=workload_start@30.0s



Delta power vs baseline (decode windows shaded)

