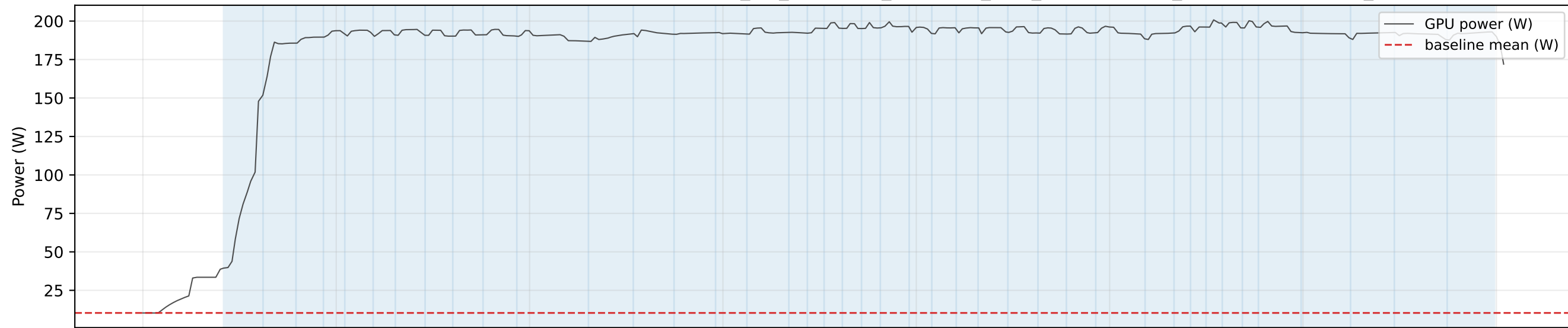


GPU power with LLM decode windows (n=41), baseline=from_llm_segments_csv, align_shift_s=0.031, trim_baseline=workload_start@29.9s



Delta power vs baseline (decode windows shaded)

