

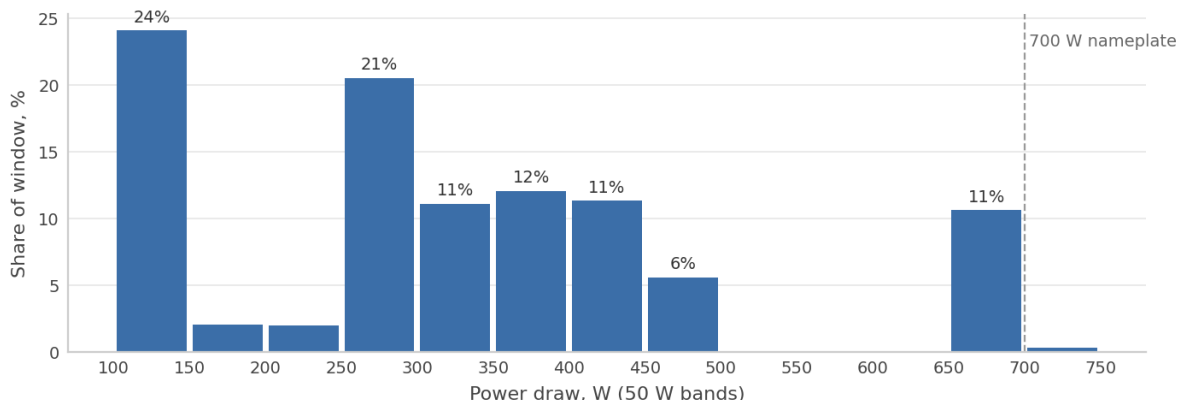
Sample headroom output: one H100 SXM, 18.1 minutes

Dynamic Margin Systems · dynamicmarginsystems.com · what the daemon's headroom file says about a real run

What was measured

One NVIDIA H100 SXM, 700 W factory limit, read-only, on July 14, 2026. The daemon sampled the card ten times a second for 1,086 seconds: 10,847 ticks, zero unobserved seconds. The card ran a mixed load: a stretch pinned at its limit, a long stretch around 300 to 450 W, and a stretch idle with a CUDA context attached. Every number below is computed by the daemon's replay mode from that log; nothing is modeled.

The shape of the draw



Mean draw 332 W. Peak 720.5 W, above the 700 W factory limit, for a tenth of a second. The card spent 24 percent of the window under 150 W, the idle-with-context floor; 61 percent between 250 and 500 W; and 11 percent at or above 650 W. A feed sized at 700 W for this card was fully used for 11 percent of the window.

What a cap would have cost

Cap	Share of window the cap would have bound	Energy above the cap
630 W (90%)	11.0%	2.3 Wh
560 W (80%)	11.0%	4.6 Wh
490 W (70%)	11.1%	6.9 Wh
420 W (60%)	23.4%	10.2 Wh
350 W (50%)	40.1%	16.9 Wh

Read the table from the top. Down to 490 W the cap bites only during the 11 percent of the window the card spent at its limit; every cap between 630 and 490 W costs the same seconds, because nothing ran between those levels. At 420 W it also catches the 420 to 470 W phase, and at 350 W it is binding 40 percent of the time. The honest ceiling for this workload sits near 490 W: 30 percent of nameplate given back to the feed, bound 11 percent of the time.

What this becomes on a full node

One card cannot show how many more cards fit, because that depends on cards not peaking together. On an eight-card node the same file carries the node's coincident peak against its 5,600 W nameplate, the same cap table per card and for the node, and the count of extra cards the same feed would carry at each cap, with the share of the window the cap would bind as the cost. If every card runs pinned at 700 W all day, the file says so, and the answer is zero.

Questions about the method, or the raw log behind this page: liam@dynamicmarginsystems.com