

Supplementary material for: Running orbital AI data centers hot: chip leakage matters as much as temperature tolerance

Equations, tables and figures numbered with the prefix S belong to this supplement; numbers without the prefix refer to the main text. The three leakage shapes are flat ($b = 2584$ K), median ($b = 3315$ K) and steep ($b = 4515$ K); the three reference radiators are advanced, nominal and heavy (Table 2). A fourth, pessimistic radiator is used only as a sensitivity case (Section S9.10). Unless stated otherwise, results refer to $f_L = 0.15$, the median shape, the nominal radiator, a fixed nominal clock, $\Delta T_{jc} = \Delta T_{hbm} = 40$ K, a present chip limit $T_{cap} = 85$ °C with an HBM limit of 95 °C, and, when the limits are raised, a chip limit T_{cap} with an HBM limit of $\max(T_{cap}, 95$ °C). The leakage-driven power growth s of Eq. (4) is given in % K⁻¹ and evaluated at the hottest chip of the present design; s_0 is its break-even value, at which raising the limits to 150 °C stops saving mass, and s_{10} , s_5 and s_2 are the values at which the saving falls to 10%, 5% and 2%. The leakage fractions at these points are $f_{L,0}$, $f_{L,10}$, $f_{L,5}$ and $f_{L,2}$.

The following short forms are used: “hot chip” is the hottest chip, at the loop outlet; “loop rise” is the loop temperature rise ΔT_{loop} ; s_{85} is s evaluated at 85 °C, equal to s of the main text for the present design; the radiator share is ϕ_{eff} and the radiator mass share ϕ_{rad} ; “raising the limits to 150 °C” means raising the chip limit from 85 °C and the HBM limit from 95 °C to 150 °C; and “a stack of X K” means $\Delta T_{jc} = \Delta T_{hbm} = X$ K. Temperatures in the equations and in the solution grids are absolute (K); chip, HBM and coolant temperatures in the results are given in °C and radiating temperatures in K. # S1. Parameters, power closure and one-at-a-time sensitivity

The sources of all model parameters, the closed power and mass balance, the pump model and the one-at-a-time sensitivities around the nominal design are given below.

S1.1. Parameters and their basis

The values are listed in Table S1.

Table S1. Model parameters. “OAT range” is the low and high value used in the one-at-a-time sensitivity (Section S1.6); other ranges are scan ranges used in the main text.

Symbol	Description	Nominal	Range	Basis
ϵ	radiator emissivity	0.90	0.80–0.92 (OAT)	[3]; upper value [1]
F	view factor to deep space	0.85	0.75–0.95 (OAT)	[3]
η_{fin}	fin efficiency	0.85; heavy radiator 0.819	0.65–0.95 (OAT); 0.61–0.95 (design-rule sample, Section S9.3)	assumed; heavy value calibrated to the ISS photovoltaic radiator (see below); ground test implies 0.61 (Section S4) [37]
σ_{rad}	radiator areal mass per m ² of planform	10 kg m ⁻²	5–17.5 kg m ⁻² (OAT)	5 kg per m ² of emitting face [3]; 17.46 kg m ⁻² for the ISS photovoltaic radiator [20]
q_{env}	absorbed environmental load, both faces	127.3 W m ⁻²	×0.5 to ×2 (OAT)	Eq. (S5); end-of-life variant ×1.3395 (Section S9.12)
ΔT_{jc}	junction-to-local-coolant difference at the power drawn at 87.5 °C	40 K	15–50 K (OAT)	[5, 21]
ΔT_{hbm}	HBM-to-coolant-outlet difference	40 K	30–50 K (OAT); 15–40 K (Section S9.6)	assumed
T_{cap}	chip limit on the hottest junction	85 °C	83–92 °C (OAT); 85–150 °C (scan)	83 °C V100 [22]; 85 °C A100 [23]
—	HBM limit	95 °C	85–105 °C (OAT); max(T_{cap} , 95 °C) in the scan; variants fixed at 95 °C and capped at 105 °C	H100 HBM specification [24]; [41]
f_{L}	leakage fraction at 87.5 °C	0.15	0.10–0.22 (OAT); 0–0.35 (scan)	Section S3
b	leakage activation temperature	3315 K	2584–4515 K	Section S2
T_{ref}	anchor of Eqs. (2) and (3)	87.5 °C	—	Section S1.3
—	kW _{ref} normalization junction temperature	80 °C	—	Section S1.3
n	chip positions along the loop	41	—	Section S1.2
c_p, ρ	coolant specific heat; density	4000 J kg ⁻¹ K ⁻¹ ; 1000 kg m ⁻³	3600; 1030 (glycol-like variant)	water-like coolant
Δp_{ref}	loop pressure drop at $\Delta T_{\text{ref}} = 10$ K	200 kPa	100–400 kPa (OAT)	assumed (cold plates and radiator manifold)
η_{pump}	pump efficiency	0.5	0.35–0.70 (OAT)	assumed
ΔT_{loop}	loop temperature rise	optimized	{2, 3, 4, 6, 8, 10, 12, 15, 20, 25, 30, 40, 50} K	design variable
T_{out}	coolant outlet (radiator inlet) temperature	optimized	290–440 K in 0.25 K steps	design variable
ΔT_{hx}	intermediate heat-exchanger difference	0 K	0–8 K (OAT), with 0.5 kg kW ⁻¹ when present	assumed
SP_{pv}	solar-array specific power (array level)	100 W kg ⁻¹	50–300 W kg ⁻¹ (OAT)	system sizing estimate
η_{PMAD}	bus-side PMAD efficiency (MPPT, bus conversion, harness)	0.92	0.88–0.96 (OAT)	assumed
m_{bat}	seasonal-eclipse storage mass	2 kg kW ⁻¹	0–6 kg kW ⁻¹ (OAT)	estimate for a dawn–dusk orbit
$m_{\text{IT}}, m_{\text{fix}}$	compute-hardware; fixed mass	7; 6 kg kW _{ref} ⁻¹	4–10.5; 3–12 (OAT)	system sizing estimate
f_{other}	non-chip bus load per chip power	0.05	0.02–0.10 (OAT)	assumed

Symbol	Description	Nominal	Range	Basis
m_{comp}	compressor specific mass	4.6 kg kW _e ⁻¹	0.92–9.2 (OAT)	calibrated at the prototype rated COP of 2.3 [13] (Section S5)
m_{evap}	evaporator specific mass	1.5 kg kW _{th} ⁻¹	0.5–2.72 (OAT)	upper value [11]
—	heat-pump component efficiencies	nominal set	conservative and optimistic sets (OAT)	Section S5

Areal-mass basis. The model rejects heat from both faces of a panel, and $q(T)$ in Eq. (5) is the net flux per m² of planform. With Turyshev’s inputs [3] the model gives an emitted flux of 650.9 W m⁻² per face at 350 K, 2495 m² of emitting face for 1 MW and 12.48 kg kW⁻¹ at 5 kg per m² of face; the same 1 MW needs 1248 m² of planform, which at 10 kg m⁻² gives the same 12.48 kg kW⁻¹. The nominal 10 kg m⁻² planform therefore equals Turyshev’s 5 kg per m² of emitting face. The ISS photovoltaic radiator rejects 14 kW with 740.7 kg on a 3.12 m × 13.6 m panel [20], i.e. 17.46 kg m⁻² of planform.

Fin efficiency of the heavy radiator. The heavy radiator combines the 17.5 kg m⁻² areal mass of the ISS photovoltaic radiator with the fin efficiency at which the model reproduces that radiator’s rated flux. The rating of 14 kW on 3.12 m × 13.6 m of planform corresponds to 329.9 W m⁻². For an ammonia loop cooling linearly from 291 to 276 K (assumed), with $\varepsilon = 0.9$, $F = 0.85$ and the nominal environmental load, the fin efficiency is found by root finding on the model’s own area integral, Eq. (5), in which Q/A is the harmonic mean of the local net flux over the fluid temperature drop. The root is 0.81895, and 0.819 is used; with this value the model gives 330.0 W m⁻², within 0.01% of the rating. Averaging the local flux arithmetically instead gives 0.81466. With the end-of-life environmental load of Section S9.12 the two methods give 0.89678 and 0.89165; a higher environmental load would thus require a higher fin efficiency to match the rating, and the calibrated 0.819 lies on the conservative side.

Power basis. Chip power is module power (die and on-package HBM, as measured by board power telemetry), which includes the losses of the on-board voltage regulators. η_{PMAD} therefore covers only the bus side (maximum-power-point tracking, bus conversion and harness), so that point-of-load conversion is not counted twice. The leakage fraction f_L is defined on the same module basis (Section S3.1).

S1.2. Power closure

Each loop carries chips at 41 equally spaced positions between the radiator outlet and the radiator inlet; chip i sees the local coolant temperature $T_{c,i} = T_{\text{out}} - \Delta T_{\text{loop}}(1 - \xi_i)$ with $\xi_i = (i - 1)/40$, so that the hottest chip sits at the loop outlet. Eq. (2) is solved for each position, and the chip power per kW_{ref} is the loop average

$$P_{\text{chip}} = \frac{1}{41} \sum_{i=1}^{41} \frac{P(T_{j,i})}{P(T_{80})}, \quad (\text{S1})$$

where P is Eq. (3), equal to 1 at 87.5 °C, and $T_{80} = 80$ °C. Non-chip loads and pumping power are added as

$$P_{\text{bus}} = (1 + f_{\text{other}}) P_{\text{chip}}, \quad (\text{S2})$$

$$\kappa \equiv \frac{P_{\text{pump}}}{P_{\text{tot}}} = \frac{\dot{m} \Delta p}{\rho \eta_{\text{pump}} P_{\text{tot}}} = \frac{\Delta p_{\text{ref}} (\Delta T_{\text{ref}} / \Delta T_{\text{loop}})^2}{\rho c_p \eta_{\text{pump}} \Delta T_{\text{loop}}}, \quad \dot{m} = \frac{P_{\text{tot}}}{c_p \Delta T_{\text{loop}}}, \quad (\text{S3})$$

$$P_{\text{tot}} = \frac{P_{\text{bus}}}{1 - \kappa}. \quad (\text{S4})$$

All bus power ends up as heat, so the radiator rejects $Q = P_{\text{tot}}$. The radiator area follows from Eq. (5) with the radiator inlet at $T_{\text{out}} - \Delta T_{\text{hx}}$ and its outlet at $T_{\text{in}} - \Delta T_{\text{hx}}$, where $T_{\text{in}} = T_{\text{out}} - \Delta T_{\text{loop}}$; the integral is evaluated with the trapezoidal rule on 41 points of a linear fluid temperature drop, and a design is infeasible wherever $q(T) \leq 0$. The solar

array is sized for P_{tot} through $SP_{\text{pv}}\eta_{\text{PMAD}}$, the storage mass is $m_{\text{bat}}P_{\text{tot}}$, and an intermediate heat exchanger, when present, adds 0.5 kg per kW of P_{tot} (Eq. 1).

The environmental load absorbed by a vertical two-sided panel in a dawn–dusk Sun-synchronous orbit is

$$q_{\text{env}} = 2 \left[\varepsilon F_E \sigma T_E^4 + \alpha_s F_E a S_0 f_a \right] + \alpha_s S_0 f_p, \quad (\text{S5})$$

with Earth view factor $F_E = 0.25$ per face, Earth effective temperature $T_E = 255$ K, solar absorptance $\alpha_s = 0.15$, albedo $a = 0.30$ with a terminator factor $f_a = 0.5$, solar constant $S_0 = 1361$ W m⁻² and a pointing-error fraction $f_p = 0.02$ of direct sunlight on one face. With $\varepsilon = 0.9$ the three terms are 107.9 (Earth infrared), 15.3 (albedo) and 4.1 W m⁻² (direct sunlight), 127.3 W m⁻² in total; the end-of-life variant and the pessimistic radiator use 170.5 W m⁻² (Section S9.12). Because the infrared term scales with ε , the emissivity cases of the one-at-a-time sensitivity (Section S1.6) also change q_{env} .

The effective radiating temperature of a design is

$$T_{\text{eff}} = \left[\frac{P_{\text{tot}}/A + q_{\text{env}}}{2\varepsilon F \sigma} \right]^{1/4}, \quad (\text{S6})$$

i.e. the temperature of an isothermal surface with $\eta_{\text{fin}} = 1$ and the same ε , F and q_{env} that achieves the same net flux. The area elasticity in Eq. (6) is $|\partial \ln a / \partial T_c|$, where $a = A/Q$ is the area per unit heat at the design loop rise, evaluated by a central difference of ± 0.05 K in the outlet temperature; its value at the nominal optimum, 0.01554 K⁻¹, is the normalizing constant of ϕ_{eff} .

S1.3. Normalization points

Two normalizations are used and are kept separate.

- 87.5 °C (*leakage anchor*). f_L and $g(T)/g(T_{\text{ref}})$ in Eq. (3) are anchored at $T_{\text{ref}} = 87.5$ °C, so $P(87.5$ °C) = 1. The fixed point of Eq. (2), the chip-level loop gain $\Delta T_{\text{jc}} dP/dT_j$ and the chip terms of the system-level gain (Section S9.9) use this normalization; ΔT_{jc} is thus the junction-to-local-coolant difference at the power drawn at 87.5 °C. Expressed at the power drawn at 80 °C, the nominal 40 K corresponds to 38.7 K.
- 80 °C (kW_{ref}). Only the mass model divides chip power by $P(T_{80})$ (Eq. S1), so that 1 kW_{ref} is the compute of chips drawing 1 kW at 80 °C and nominal clock. For $f_L = 0.15$ and the median shape $P(T_{80}) = 0.9683$. The heat-load coupling term of the system-level gain carries the factor $1/P(T_{80})$ through the mass model.
- In the clock-scaled chip model (Section S6) chip power is normalized to 1 at clock ratio 1 and 87.5 °C inside the fixed point, and the mass model divides by its value at clock ratio 1 and 80 °C.

S1.4. Solution procedure

1. The coolant outlet temperature T_{out} is scanned from 290 to 440 K in 0.25 K steps, and for each value all 13 loop rises ΔT_{loop} of Table S1 are evaluated.
2. Eq. (2) is solved at each chip position by direct iteration from $T_j = T_c + \Delta T_{\text{jc}}$, with convergence when successive iterates differ by less than 10^{-5} K within 80 iterations. An iterate above 600 K or a non-converged iteration means no operating point, and the design is infeasible. The per-chip solutions are tabulated on a coolant grid of 1/16 K (224–460 K) and interpolated linearly between nodes; the interpolation error is of order 10^{-8} .
3. A design is feasible if the hottest junction does not exceed T_{cap} and $T_{\text{out}} + \Delta T_{\text{hbm}}$ does not exceed the HBM limit.
4. Eqs. (S1)–(S4) give the bus power in closed form, Eq. (5) the radiator area and Eq. (1) the mass.
5. For each T_{out} the lightest loop rise is kept, and the optimum is the lightest T_{out} . The binding constraint is identified by testing the next grid point (chip limit, HBM limit, runaway or interior optimum). When the optimum

lies on a limit, the limit is located within the 0.25 K grid step by bisection to 10^{-3} K, so that savings vary smoothly with f_L ; the cold-plate arrangements of Section S9.4 keep the grid resolution.

6. Literature mappings, runaway limits and lifetime use a stricter solver (5000 iterations, 10^{-9} K, divergence above 1500 K) and accept only roots with chip-level loop gain below 1.
7. Leakage-fraction thresholds (saving from raising the limits to 150 °C falling to 10%, 5% or 2%) are found by bisection on f_L to 10^{-4} ; the break-even $f_{L,0}$ is the value at which the saving falls below 0.01%. The corresponding s values are evaluated with Eq. (4) at the hottest chip of the present optimum, with f_L set to the root.

S1.5. Pump model

With Eq. (S3) and the nominal parameters, the pumping power is $10^4/\Delta T_{\text{loop}}^3$ W per kW of rejected heat (ΔT_{loop} in K): 10 W kW⁻¹ at 10 K, 5.8 W kW⁻¹ at 12 K and 1250 W kW⁻¹ at 2 K. The loop pressure drop is taken per kW, independent of system size. Pumping power is drawn from the bus and rejected by the radiator. At the nominal optimum for $T_{\text{cap}} = 85$ °C the loop rise is 10 K and the pump draws 10.6 W kW⁻¹_{ref}; at $T_{\text{cap}} = 150$ °C the optimum loop rise is 15 K and the pump draws 3.4 W kW⁻¹_{ref}. In the two-phase variant the loop is isothermal at T_{out} , all chips see T_{out} , the radiator is isothermal and the pumping power is fixed at 0.3% of the rejected heat.

The saving from raising the limits from 85 to 150 °C is split into three parts (Table S2, main-text Fig. 5c). At each of the two optima, the pumping-power penalty is the mass with pumping power minus the mass without it at the same outlet temperature and loop rise; the loop-rise penalty is the pump-free mass at that loop rise minus the pump-free mass at the same outlet with the loop rise reoptimized; and the radiator part is the difference between the two pump-free reoptimized masses. The three parts add up to the saving. The loop-rise penalty arises because a single-phase loop with a large temperature rise delivers the heat of its outlet chip, which sets the coolant temperature, to a radiator whose mean temperature is lower.

Table S2. Decomposition of the saving from raising chip and HBM limits from 85 to 150 °C, in % of the mass at 85 °C (flat / median / steep shape; loop rise and pumping power for the median shape).

Case	f_L	Saving	Radiator part	Pumping-power part	Loop-rise part	$\Delta T_{\text{loop}}, 85 \rightarrow 150$ °C (K)	Pumping power, 85 $\rightarrow$ 150 °C (W kW ⁻¹ _{ref})
Advanced	0.10	3.45 / 2.18 / 1.09	2.02 / 0.99 / 0.14	0.56 / 0.19 / 0.18	0.87 / 1.00 / 0.76	12 $\rightarrow$ 15	6.1 $\rightarrow$ 3.3
Advanced	0.15	1.24 / 0.55 / 0.14	0.35 / -0.08 / -0.14	0.18 / 0.18 / 0.18	0.71 / 0.45 / 0.10	12 $\rightarrow$ 15	6.1 $\rightarrow$ 3.2
Advanced	0.22	0.08 / 0.00 / 0.00	-0.09 / 0.00 / 0.00	0.00 / 0.00 / 0.00	0.17 / 0.00 / 0.00	15 $\rightarrow$ 15	3.0 $\rightarrow$ 3.0
Nominal	0.10	8.83 / 6.07 / 3.45	6.58 / 3.99 / 1.62	0.53 / 0.52 / 0.51	1.73 / 1.56 / 1.32	10 $\rightarrow$ 15	10.6 $\rightarrow$ 3.5
Nominal	0.15	4.55 / 2.59 / 1.05	2.73 / 1.03 / -0.10	0.51 / 0.51 / 0.50	1.31 / 1.05 / 0.65	10 $\rightarrow$ 15	10.6 $\rightarrow$ 3.4
Nominal	0.22	1.30 / 0.38 / 0.01	0.18 / -0.24 / -0.14	0.20 / 0.20 / 0.20	0.92 / 0.42 / -0.05	12 $\rightarrow$ 15	6.0 $\rightarrow$ 3.2
Heavy	0.10	15.68 / 11.09 / 6.58	12.50 / 8.12 / 3.92	0.60 / 0.58 / 0.57	2.58 / 2.39 / 2.09	10 $\rightarrow$ 15	10.6 $\rightarrow$ 3.7
Heavy	0.15	9.18 / 5.58 / 2.57	6.48 / 3.19 / 0.79	0.58 / 0.57 / 0.34	2.12 / 1.82 / 1.44	10 $\rightarrow$ 15	10.6 $\rightarrow$ 3.5
Heavy	0.22	3.62 / 1.53 / 0.31	1.68 / 0.06 / -0.25	0.56 / 0.55 / 0.00	1.38 / 0.92 / 0.57	10 $\rightarrow$ 15	10.5 $\rightarrow$ 3.3

At 85 °C the chip limit holds the coolant at 45.45 °C, the radiator is large and cold, and the pumping penalty forces a moderate loop rise; once the limits are raised the coolant warms, the loop rise widens to 15 K and the pumping power falls from 10.6 to 3.4 W kW⁻¹_{ref} with the nominal radiator. For $f_L = 0.15$ and the nominal radiator, the 2.59% saving consists of 1.03 percentage points from radiator area, 0.51 from pumping power and 1.05 from the loop-rise penalty; in

kilograms, pumping and loop-rise penalties fall from 0.317 and 0.518 to 0.090 and 0.048 kg kW_{ref}⁻¹. With the advanced radiator the radiator part is slightly negative and the whole saving comes from the two loop penalties (Table S3).

Table S3. Sensitivity of the saving (85 → 150 °C), the leakage fraction for a 5% saving and the break-even values to the pump and loop model (flat / median / steep shape). The two-phase loop is isothermal with pumping power fixed at 0.3% of the rejected heat.

Case	Pump model	M at 85 °C, kg kW _{ref} ⁻¹ (median)	Saving, $f_L =$ 0.10 (%)	Saving, $f_L =$ 0.15 (%)	Saving, $f_L =$ 0.22 (%)	$f_{L,5}$	$f_{L,0}$	s_0 (% K ⁻¹)
Nominal	nominal (200 kPa, η_{pump} 0.5)	44.680	8.83 / 6.07 / 3.45	4.55 / 2.59 / 1.05	1.30 / 0.38 / 0.01	0.143 / 0.112 / 0.081	0.310 / 0.270 / 0.219	0.763 / 0.800 / 0.825
Nominal	Δp_{ref} 100 kPa	44.521	8.47 / 5.73 / 3.17	4.27 / 2.35 / 0.87	1.10 / 0.28 / 0.00	0.139 / 0.108 / 0.078	0.297 / 0.256 / 0.214	0.729 / 0.758 / 0.806
Nominal	Δp_{ref} 400 kPa	44.894	9.30 / 6.48 / 3.77	4.92 / 2.85 / 1.26	1.52 / 0.53 / 0.04	0.149 / 0.117 / 0.085	0.310 / 0.270 / 0.229	0.763 / 0.800 / 0.863
Nominal	η_{pump} 0.35	44.787	9.07 / 6.28 / 3.66	4.74 / 2.73 / 1.14	1.39 / 0.47 / 0.04	0.146 / 0.115 / 0.083	0.310 / 0.270 / 0.229	0.763 / 0.800 / 0.863
Nominal	η_{pump} 0.70	44.589	8.70 / 5.93 / 3.31	4.41 / 2.45 / 0.93	1.20 / 0.32 / 0.00	0.142 / 0.111 / 0.080	0.306 / 0.256 / 0.214	0.751 / 0.758 / 0.806
Nominal	no pumping power	43.846	6.84 / 4.23 / 1.89	2.93 / 1.24 / 0.18	0.37 / 0.00 / 0.00	0.120 / 0.092 / 0.063	0.256 / 0.213 / 0.169	0.627 / 0.629 / 0.634
Nominal	glycol-like (c_p 3600, ρ 1030, 400 kPa)	44.923	9.35 / 6.52 / 3.81	4.96 / 2.89 / 1.29	1.57 / 0.53 / 0.04	0.149 / 0.118 / 0.085	0.310 / 0.270 / 0.229	0.763 / 0.800 / 0.863
Nominal	two-phase loop	43.840	6.47 / 3.89 / 1.61	2.64 / 1.03 / 0.09	0.26 / 0.00 / 0.00	0.116 / 0.088 / 0.060	0.248 / 0.205 / 0.161	0.608 / 0.606 / 0.603
Advanced	nominal	35.601	3.45 / 2.18 / 1.09	1.24 / 0.55 / 0.14	0.08 / 0.00 / 0.00	0.078 / 0.059 / 0.041	0.235 / 0.204 / 0.173	0.577 / 0.603 / 0.647
Advanced	no pumping power	35.286	2.19 / 1.11 / 0.30	0.45 / 0.05 / 0.00	0.00 / 0.00 / 0.00	0.061 / 0.045 / 0.030	0.190 / 0.158 / 0.126	0.465 / 0.466 / 0.469
Advanced	glycol-like	35.709	3.78 / 2.50 / 1.31	1.49 / 0.71 / 0.16	0.16 / 0.02 / 0.00	0.082 / 0.063 / 0.044	0.255 / 0.225 / 0.173	0.626 / 0.664 / 0.647
Advanced	two-phase loop	35.340	1.99 / 0.96 / 0.20	0.36 / 0.02 / 0.00	0.00 / 0.00 / 0.00	0.058 / 0.043 / 0.028	0.184 / 0.152 / 0.119	0.450 / 0.448 / 0.446
Heavy	nominal	59.704	15.68 / 11.09 / 6.58	9.18 / 5.58 / 2.57	3.62 / 1.53 / 0.31	0.198 / 0.157 / 0.116	0.364 / 0.302 / 0.254	0.899 / 0.898 / 0.958
Heavy	no pumping power	58.062	13.06 / 8.59 / 4.37	6.86 / 3.55 / 1.06	1.96 / 0.41 / 0.00	0.171 / 0.132 / 0.094	0.304 / 0.253 / 0.202	0.748 / 0.751 / 0.758
Heavy	glycol-like	60.172	16.32 / 11.67 / 7.14	9.74 / 6.07 / 2.99	4.01 / 1.85 / 0.52	0.204 / 0.163 / 0.122	0.364 / 0.317 / 0.271	0.899 / 0.945 / 1.024
Heavy	two-phase loop	57.954	12.54 / 8.08 / 3.91	6.41 / 3.16 / 0.79	1.67 / 0.26 / 0.00	0.166 / 0.127 / 0.089	0.295 / 0.244 / 0.192	0.726 / 0.724 / 0.721

The pressure-drop and pump-efficiency settings move the saving with the nominal radiator by −0.24 to +0.30 percentage points. Removing pumping power or using a two-phase loop lowers it to about half, because both remove the pumping and loop-rise penalties that a higher coolant temperature relieves. The break-even values s_0 of these two variants are 0.63 and 0.61% K⁻¹ for the nominal radiator and nearly independent of the leakage shape, because all chips then sit at the same temperature.

S1.6. One-at-a-time sensitivity

Each parameter of Table S1 is set in turn to its low and high OAT value around the nominal design (Table S4, Fig. S1). At the center the on-orbit mass is $44.680 \text{ kg kW}_{\text{ref}}^{-1}$, the optimum coolant outlet temperature $45.45 \text{ }^{\circ}\text{C}$ with the chip limit binding, the saving from raising the limits to $150 \text{ }^{\circ}\text{C}$ 2.59% (2.12% with the HBM held at $95 \text{ }^{\circ}\text{C}$), the break-even values $f_{L,0} = 0.270$ and $s_0 = 0.800\% \text{ K}^{-1}$, the penalty of a heat pump forced to $T_{\text{eff}} = 350 \text{ K}$ 4.7% and the best heat-pump design 3.4% heavier than the passive one.

All 46 optima lie on a constraint corner: 43 on the chip limit and three on the HBM limit ($\Delta T_{\text{jc}} = 15 \text{ K}$, $\Delta T_{\text{hbm}} = 50 \text{ K}$ and an HBM limit of $85 \text{ }^{\circ}\text{C}$; Section S9.2). The six heat-pump rows do not change the passive design, so 37 of the 40 rows that affect it are pinned by the chip limit. The optimum coolant temperature moves only with ΔT_{jc} , T_{cap} , ΔT_{hbm} and the HBM limit and stays at $45.45 \text{ }^{\circ}\text{C}$ in all other cases. The mass is most sensitive to σ_{rad} , SP_{pv} , η_{fin} , m_{fix} and ΔT_{jc} . The saving from higher ratings is most sensitive to f_L (6.07 / 0.38%), q_{env} (1.60 / 6.08%), σ_{rad} (0.55 / 5.24%), SP_{pv} (0.84 / 5.38%), ΔT_{jc} (5.12 / 3.24%), η_{fin} (4.78 / 1.96%) and b (4.55 / 1.05%). Across all rows the break-even leakage-driven power growth s_0 spans 0.603–0.980% K^{-1} , lowest with the 5 kg m^{-2} radiator and highest with twice the environmental load, and it follows the shifts in the radiator share ϕ_{eff} ; the leakage itself, f_L and b , changes s_0 by at most 5%. The break-even fraction $f_{L,0}$ spans 0.204–0.390 and also moves with b (0.310 / 0.219); with $\Delta T_{\text{jc}} = 15 \text{ K}$ it rises to 0.390 while s_0 moves only to 0.853% K^{-1} . The compressor and evaporator masses and the heat-pump efficiency set affect only heat-pump designs.

Table S4. One-at-a-time sensitivity around the nominal design (low / high value). The saving is that of raising the chip limit to $150 \text{ }^{\circ}\text{C}$ with the HBM limit rising with it. Heat-pump columns give the mass change relative to the passive optimum with the same parameters; positive values mean the heat-pump design is heavier. An asterisk marks an optimum pinned by the HBM limit; all others are pinned by the chip limit.

Parameter	Low / high value	M (kg kW $_{\text{ref}}^{-1}$)	Optimum coolant outlet (°C)	Saving 85 → 150 °C (%)	$f_{L,0}$	s_0 (% K $^{-1}$)	Heat pump to $T_{\text{eff}} = 350$ K (%)	Best heat-pump design (%)
Center	—	44.680	45.45	2.59	0.270	0.800	4.7	3.4
ϵ	0.80 / 0.92	47.024 / 44.275	45.45 / 45.45	3.13 / 2.49	0.281 / 0.267	0.834 / 0.793	2.7 / 5.0	2.0 / 3.6
F	0.75 / 0.95	47.703 / 42.415	45.45 / 45.45	3.48 / 1.96	0.289 / 0.253	0.859 / 0.750	1.8 / 6.9	1.2 / 4.9
η_{fin}	0.65 / 0.95	51.940 / 42.415	45.45 / 45.45	4.78 / 1.96	0.297 / 0.253	0.882 / 0.750	6.0 / 5.6	−1.9 / 4.9
q_{env}	×0.5 / ×2	42.891 / 49.721	45.45 / 45.45	1.60 / 6.08	0.240 / 0.329	0.709 / 0.980	7.6 / −3.0	5.4 / −3.2
σ_{rad} (kg m $^{-2}$)	5 / 17.5	35.601 / 58.234	45.45 / 45.45	0.55 / 5.24	0.204 / 0.297	0.603 / 0.884	14.3 / −4.1	7.9 / −4.1
ΔT_{jc} (K)	15 / 50	40.841* / 47.748	55.00 / 35.56	5.12 / 3.24	0.390 / 0.267	0.853 / 0.792	6.0 / 2.7	5.1 / 1.0
T_{cap} (°C)	83 / 92	44.920 / 44.031	43.79 / 51.11	3.11 / 1.15	0.285 / 0.222	0.815 / 0.748	4.7 / 5.4	3.2 / 4.4
HBM limit (°C)	85 / 105	44.743* / 44.680	45.00 / 45.45	2.72 / 2.59	0.279 / 0.270	0.809 / 0.800	4.5 / 4.7	3.2 / 3.4
ΔT_{hbm} (K)	30 / 50	44.680 / 44.743*	45.45 / 45.00	2.59 / 2.72	0.270 / 0.279	0.800 / 0.809	4.7 / 4.5	3.4 / 3.2
f_L	0.10 / 0.22	44.782 / 44.432	45.30 / 45.66	6.07 / 0.38	0.270 / 0.270	0.800 / 0.800	4.7 / 4.7	3.4 / 3.2
b (K)	2584 / 4515	44.729 / 44.598	45.37 / 45.58	4.55 / 1.05	0.310 / 0.219	0.763 / 0.825	4.7 / 4.6	3.4 / 3.3
Δp_{ref} (kPa)	100 / 400	44.521 / 44.894	45.45 / 45.45	2.35 / 2.85	0.256 / 0.270	0.758 / 0.800	4.3 / 5.0	3.0 / 3.5
η_{pump}	0.35 / 0.70	44.787 / 44.589	45.45 / 45.45	2.73 / 2.45	0.270 / 0.256	0.800 / 0.758	5.0 / 4.4	3.5 / 3.2
ΔT_{hx} (K)	0 / 8	44.680 / 47.666	45.45 / 45.45	2.59 / 3.39	0.270 / 0.288	0.800 / 0.856	4.7 / −1.9	3.4 / −3.1
SP_{pv} (W kg $^{-1}$)	50 / 300	56.106 / 37.018	45.45 / 45.45	0.84 / 5.38	0.213 / 0.312	0.629 / 0.930	11.0 / −1.6	6.0 / −1.6
η_{PMAD}	0.88 / 0.96	45.203 / 44.201	45.45 / 45.45	2.47 / 2.70	0.266 / 0.273	0.790 / 0.809	5.0 / 4.3	3.6 / 3.2
m_{bat} (kg kW $^{-1}$)	0 / 6	42.566 / 48.902	45.45 / 45.45	3.15 / 1.75	0.283 / 0.246	0.842 / 0.727	3.2 / 7.3	2.4 / 4.8
m_{IT} (kg kW $_{\text{ref}}^{-1}$)	4 / 10.5	41.680 / 48.180	45.45 / 45.45	2.77 / 2.40	0.270 / 0.269	0.801 / 0.799	5.0 / 4.3	3.6 / 3.1
m_{fix} (kg kW $_{\text{ref}}^{-1}$)	3 / 12	41.680 / 50.680	45.45 / 45.45	2.77 / 2.28	0.270 / 0.269	0.801 / 0.798	5.0 / 4.1	3.6 / 3.0
f_{other}	0.02 / 0.10	43.775 / 46.189	45.45 / 45.45	2.57 / 2.62	0.270 / 0.270	0.800 / 0.800	4.6 / 4.7	3.3 / 3.4
m_{comp} (kg kW $_{\text{e}}^{-1}$)	0.92 / 9.2	44.680 / 44.680	45.45 / 45.45	2.59 / 2.59	0.270 / 0.270	0.800 / 0.800	1.7 / 8.4	1.3 / 5.3
m_{evap} (kg kW $_{\text{th}}^{-1}$)	0.5 / 2.72	44.680 / 44.680	45.45 / 45.45	2.59 / 2.59	0.270 / 0.270	0.800 / 0.800	2.3 / 7.6	1.0 / 6.3
Heat-pump efficiency set	conservative / optimistic	44.680 / 44.680	45.45 / 45.45	2.59 / 2.59	0.270 / 0.270	0.800 / 0.800	9.7 / 2.3	6.7 / 1.6

S2. Evidence on the temperature shape of leakage

The leakage activation temperature b of Eq. (3) is constrained by leakage–temperature curves from vendor power-estimation tools and by transistor measurements converted to off-state current; the sources, fitting rules, percentiles and sensitivities are given below (main-text Fig. 2a).

S2.1. Sources

The values are listed in Table S5.

Table S5. Evidence on the temperature shape of leakage. b from fits of Eq. (3) over 50–150 °C (50–125 °C for the AMD tool).

Group	Object	Temperature coverage	Curves	b range (median), K
TI AM62x power tool [26]	TSMC N16FFC SoC, core-supply leakage formula	0–125 °C (tool input range)	1	3225
TI AM62D power tool [27]	TSMC N16FFC, 6 device libraries $\times$ 3 process corners $\times$ 2 voltages	–55–150 °C (22 points)	36	2473–4799 (3588)
AMD UltraScale+ XPE [28]	16 nm FinFET FPGAs, core-supply (VCCINT) static current, typical and maximum corners, 28 die types	up to 125 °C	58	3042–3677 (3329)
Kim et al., FinFET [29]	imec bulk FinFET	25–125 °C	5	2046–4081 (3063)
Kim et al., GAA nanowire [29]	imec two-level stacked gate-all-around nanowire	25–125 °C	5	2275–4515 (3395)
Simoen et al. [30]	imec gate-all-around double nanosheet	25–200 °C	6	2846–5004 (3655)
de Souza et al. [31]	CEA-Leti SOI tri-gate nanowire and nanosheet	27–307 °C	10	2753–5165 (3959)
All (group-weighted)			121	10th / 50th / 90th percentile 2584 / 3315 / 4515

Vendor tools. The AM62x tool gives core-supply leakage as a pure exponential in junction temperature (doubling every 22.6 K). The AM62D tool contains a leakage scaling table for six device libraries (three threshold-voltage classes and two channel-length classes), weak, nominal and strong process corners and supply voltages of 0.75 and 0.85 V. In the AMD tool the core-supply static current is a fifth-order polynomial in junction temperature up to 125 °C; speed grade, temperature grade and supply voltage enter only as constant multipliers, so that the temperature shape of a given die and corner does not depend on them within the tool. Coefficient sets shared by several devices are counted once, giving 29 typical-corner and 29 maximum-corner curves for 190 devices. These tools give leakage scaling for the respective processes and are not measurements of individual dies.

Device data. The literature devices report the temperature coefficient of the threshold voltage and the subthreshold swing $SS(T)$ rather than leakage curves. Off-state current is obtained as $I_{\text{off}} \propto 10^{-V_{\text{th,sat}}(T)/SS(T)}$ with a temperature-independent prefactor, and the saturation threshold voltage at 25 °C, V_{th0} , is set to 0.15, 0.20, 0.25, 0.30 and 0.35 V because the threshold of a product is a design choice; the double-nanosheet device is also evaluated at its own V_{th0} of 0.183 V. The inputs are $dV_{\text{th}}/dT = -0.43$ (FinFET) and -0.44 mV K^{–1} (nanowire), $SS = 69$ and 65 mV dec^{–1} at 25 °C and $dSS/dT = 0.24$ mV dec^{–1} K^{–1} [29]; $V_{\text{th,sat}}$ falling from 182 to 79 mV between 25 and 200 °C (-0.600 mV K^{–1}) and SS rising from 65.3 to 99.8 mV dec^{–1} [30]; and -0.539 and -0.588 mV K^{–1} with $SS = 61.5$ mV dec^{–1} at 300 K and $SS \propto T$ [31]. The coefficients of [29] are linear-region values; steeper saturation-region coefficients would raise b . The gate-induced drain leakage and junction leakage curves of [31] ($b = 708$ and 1481 K) are not included in the evidence range.

S2.2. Fitting rule

Each curve is fitted as $\ln(y/T^2) = \ln a - b/T$ by linear least squares over 50–150 °C, or 50–125 °C for the AMD tool, which ends at 125 °C. The percentiles are weighted so that each of the seven source groups carries equal weight and the curves within a group share it equally. This rule was fixed before fitting.

S2.3. Results and sensitivities

The group-weighted 10th, 50th and 90th percentiles are 2584, 3315 and 4515 K, and leakage then doubles every 27.3, 22.3 and 17.2 K near 87.5 °C. All 121 curves lie between 2046 and 5165 K; the unweighted percentiles are 2835, 3368 and 4376 K. The vendor tools alone (95 curves, three groups) give 3042, 3225 and 3963 K, and the device data alone (26 curves, four groups) give 2275, 3395 and 4603 K. Variants of the AMD-tool treatment and of the fitting window change the median by −90 to +71 K and the 90th percentile by at most +46 K (Table S6).

Table S6. Sensitivity of the percentiles of b (K) to the choices made for the evidence set.

Variant	10th	50th	90th
Main rule (7 groups, 121 curves)	2584	3315	4515
AMD tool, typical corner only	2584	3376	4515
AMD tool, maximum corner only	2584	3225	4515
AMD tool with additional low-voltage supply rails	2584	3251	4515
AMD tool fitted over 25–125 °C	2584	3283	4515
AMD tool fitted with an additional constant term	2584	3386	4515
All curves fitted over 50–125 °C	2582	3337	4561
Unweighted (121 curves)	2835	3368	4376
Vendor tools only (95 curves)	3042	3225	3963
Device data only (26 curves)	2275	3395	4603

What sets b . Threshold voltage dominates. In the AM62D table the high-threshold libraries give 3804–4799 K and the low-threshold libraries 2473–3962 K, while the two supply voltages differ by only 20–40 K for the same library and corner. In the AMD tool the maximum (fast) corner is flatter than the typical corner, with medians of 3084 and 3376 K. For the device data, b grows by 1018–1124 K, on average 1085 K, per 0.1 V of V_{th0} , which corresponds to subthreshold ideality factors of 1.03–1.14; the spread of the device group therefore reflects the assumed range of V_{th0} rather than measured scatter. A prefactor proportional to T^p would raise b by 185 K on average for $p = 0.5$ and by 369 K for $p = 1$.

Fit quality and extrapolation. Eq. (3) fits the low-threshold AM62D curves to within 0.6–2.2% and the high-threshold ones to within 2.7–21.7% (largest for the weak corner at 0.85 V), and the AMD core-supply curves to within 0.9–7.9%. Fits over 50–100 °C extrapolated to the end of each window underpredict the TI curves by 1–36% and the AMD curves by 0–15% and overpredict the device curves by 1.6–3.2%. Within the 85–150 °C scan the shape is covered directly by the TI tool up to 150 °C and by the device data up to 200 and 307 °C.

Temperature-independent static power. In the AMD tool a constant term accounts for a median of under 3% of core-supply static current at 87.5 °C (0.8–13.9%), but for about 70% in the auxiliary and I/O supplies; total programmable-logic static power therefore has an effective b of 2097–2873 K when fitted without a constant. Temperature-independent static power belongs to $1 - f_L$ in Eq. (3), so only the core-supply shape enters the evidence.

Data-center GPU. Free fits to the V100 idle cooling curves of Section S3 give 2625 K (bootstrap 2300–3025 K) and 2975 K (1900–3525 K) for the two identifiable GPUs, and a common fit to the three nominal-clock GPUs gives 2850 K (profile interval 1925–3750 K; stratified bootstrap 2475–3100 K). These values lie in the flatter half of the evidence range, below the 10th percentile of the vendor tools.

S3. Evidence on the leakage fraction

The definition of f_L , the H100 and V100 estimates, their summary, the direction of their biases, the coupling between f_L and b , and the data license are given below (main-text Fig. 2b).

S3.1. Definition

f_L is defined at module level (GPU die and on-package HBM, i.e. board power as reported by nvidia-smi), at full load (the 250 W power limit for V100 PCIe; 700 W for H100 SXM) and nominal clock (1380 MHz for V100; H100 clocks were not locked): it is the temperature-dependent leakage power at a junction temperature of 87.5 °C divided by the full-load module power. Static power that does not vary with temperature belongs to $1 - f_L$. HBM leakage and refresh power measured in board power are part of the definition, because the model's kW_{ref} has no separate HBM power. Voltage-regulator losses included in board power leave the ratio unchanged to first order when the regulator efficiency is constant.

All estimates are converted with Eq. (3) and the three leakage shapes.

S3.2. H100: liquid- versus air-cooled nodes

Latif et al. [21] ran the same eight training workloads on a liquid-cooled and an air-cooled HGX H100 node and report the average power of the eight GPUs (Tables I–II of the preprint version). With module power $P = D + a g(T)$ and the same D on both nodes,

$$a = \frac{P_{\text{air}} - P_{\text{liq}}}{\bar{g}_{\text{air}} - \bar{g}_{\text{liq}}}, \quad f_L = \frac{a g(T_{\text{ref}})}{700 \text{ W}}, \quad (S7)$$

where $\bar{g}$ is the mean of g over the GPUs of a node. The leakage power is thus obtained from the power difference at the measured, partial load and divided by the 700 W full-load module power.

Temperatures. Per-GPU temperatures are read from the published traces of Qwen2.5-32B (Fig. 6 of [21]) and Vita-CLIP (Fig. 7) over the steady part of the runs (17–85% of the plotted time axis). The liquid-cooled node shows two clusters of four GPUs; node means are 39.76 °C (liquid) and 52.94 °C (air) for Qwen2.5-32B, a difference of 13.19 K, and 43.03 and 55.66 °C for Vita-CLIP, a difference of 12.62 K. The digitized air-node power agrees with the tabulated per-GPU means to within 1% (485.1 vs 480.5 W) and 0% (566.0 vs 566.1 W). The other language-model workloads use the Qwen2.5-32B pair and the other vision-language workloads the Vita-CLIP pair. As a sensitivity, the midpoints of the temperature ranges quoted for the GPU Burn stress test (41–50 °C and 54–72 °C, i.e. 45.5 and 63.0 °C) are used for all workloads (Table S7).

Table S7. H100 leakage-fraction estimates for the eight workloads of [21]. Power is the per-GPU average of the node.

Workload	Utilization, liquid / air (%)	Power, liquid → air (W)	Difference (%)	Temperature pair (ΔT , K)	f_L , flat / median / steep	f_L , stress-test pair (median shape)
Mistral-7B-v0.3	95.8 / 95.3	437.9 → 457.3	4.24	Qwen2.5-32B (13.19)	0.205 / 0.214 / 0.245	0.131
Llama-3.1-8B	94.9 / 95.2	438.1 → 456.6	4.05	Qwen2.5-32B (13.19)	0.196 / 0.204 / 0.234	0.125
Mistral-Nemo- Base-2407	95.8 / 95.6	444.5 → 464.6	4.33	Qwen2.5-32B (13.19)	0.213 / 0.222 / 0.254	0.136
Gemma-2-27B	92.7 / 93.1	467.8 → 489.0	4.35	Qwen2.5-32B (13.19)	0.225 / 0.234 / 0.269	0.143
Qwen2.5-32B	93.4 / 92.9	458.9 → 480.5	4.50	own (13.19)	0.229 / 0.239 / 0.273	0.146
X-CLIP	58.6 / 58.1	408.5 → 422.9	3.40	Vita-CLIP (12.62)	0.146 / 0.148 / 0.164	0.097
EVL	60.0 / 59.3	290.0 → 293.9	1.32	Vita-CLIP (12.62)	0.039 / 0.040 / 0.044	0.026
Vita-CLIP	97.4 / 97.8	536.9 → 566.1	5.17	own (12.62)	0.297 / 0.302 / 0.333	0.197

The figure-based temperature pairs give estimates 1.53–1.64 times those of the stress-test pair. A regression of per-GPU power on per-GPU temperature within the air-cooled node, where the eight GPUs fall into two temperature clusters (about 46–52 and 58–63 °C), gives a slope of 1.21 W K^{-1} ($r = 0.76$) and $f_L = 0.150 / 0.152 / 0.166$ for Qwen2.5-32B, and 1.08 W K^{-1} ($r = 0.89$) and $0.124 / 0.123 / 0.129$ for Vita-CLIP. This within-node estimate is free of node-to-node differences but is confounded by chip-to-chip variation of leakage and has few degrees of freedom. The EVL row has a per-GPU power difference of only 3.9 W at about 60% utilization, and the two nodes differ in their CPUs (EPYC 9474F with 48 cores and EPYC 9534 with 64 cores); the row is retained.

S3.3. V100: idle cooling at a fixed clock

Data and sampling. The MIT Supercloud Datacenter Challenge dataset [32, 33] contains nvidia-smi telemetry of V100 PCIe 32 GB GPUs at 100 ms intervals, including GPU and memory temperature, board power, GPU and memory utilization and, in the more recent files, the SM clock. Thirteen files (about 190 MB) of the 2022-01 GPU release were analyzed: six chosen at random in directories 0000–0002 and seven chosen from 80 candidates because their header contains the clock column.

Selection rule. The following criteria were fixed before the analysis. Idle samples have both utilizations at zero and the clock equal to the level considered, belong to an idle run of at least 30 samples (3 s) and draw less than 100 W. A unit (file $\times$ GPU $\times$ clock level) needs at least 200 idle samples, integer-temperature bins with at least 5 samples, at least 4 such bins and a temperature span of at least 8 K. Files without a clock column cannot satisfy the clock criterion and are excluded. All qualifying units at the nominal 1380 MHz enter the headline result with equal weight and without exclusion.

Fit. The median power of each integer-temperature bin is fitted, with weights $\sqrt{n}$ for a bin of n samples, by

$$P_{\text{idle}}(T) = c + a g(T), \quad f_L = \frac{a g(T_{\text{ref}})}{250 \text{ W}}, \quad (\text{S8})$$

for each of the three shapes and, separately, with b free on a 500–12 000 K grid in 25 K steps. Uncertainty is estimated with a block bootstrap of 2000 replicates in which the blocks are the cooling segments (contiguous qualifying idle runs), or 60 s chunks within segments for units with fewer than 8 segments; each replicate re-applies the bin and span criteria, and failed replicates are counted. A common b is also fitted to the three 1380 MHz units with unit-specific c and a (Table S8).

Table S8. V100 units from the MIT Supercloud telemetry [33]. f_L uses the 250 W denominator; for the control units, which are not at full-load clock, it serves only for comparison.

Job ID	GPU	Clock (MHz)	Temperature, °C (span, K)	Idle samples / segments	Slope (W K ⁻¹)	f_L , flat / median / steep	Bootstrap 95% of f_L (median shape)	Free b , K (bootstrap 95%)	Use
10713835868076	0	1380	33–51 (18)	14 371 / 48	0.303	0.139 / 0.152 / 0.187	0.150–0.153	2625 (2300–3025)	estimate
60837379255107	0	1380	27–52 (25)	631 / 16	0.316	0.156 / 0.173 / 0.219	0.167–0.174	2975 (1900–3525)	estimate
65390563485290	0	1380	44–53 (9)	1422 / 54	0.211	0.081 / 0.085 / 0.097	0.075–0.102	not identifiable (2125 to grid bound)	estimate
65390563485290	0	135	34–50 (16)	943 / 8	0.165	0.077 / 0.085 / 0.105	0.083–0.086	2825 (500–3578)	same-GPU low-clock control
15474831599409	0	135	30–39 (9)	208 406 / 1 ^a	0.170	0.095 / 0.110 / 0.149	0.107–0.112	3000 (1623–4150)	low-clock control
67199342511781	0	1230	32–60 (28)	12 942 / 7 ^a	0.314	0.129 / 0.136 / 0.158	0.135–0.136	4075 (3825–4225)	non-nominal clock
67199342511781	1	1230	31–56 (25)	12 911 / 8	0.337	0.150 / 0.162 / 0.197	0.161–0.163	3625 (2683–3950)	non-nominal clock
78482278827304	0	1035	38–53 (15)	1353 / 6 ^a	0.271	0.113 / 0.120 / 0.142	0.119–0.123	3250 (500–10 025)	non-nominal clock
78482278827304	1	1035	35–48 (13)	1369 / 6 ^a	0.228	0.105 / 0.115 / 0.143	0.077–0.151	4950 (500–12 000)	non-nominal clock
41888968329786	0	135	—	200 634	—	—	—	—	excluded: 7 bins, span 6 K
60837379255107	0	135	—	518	—	—	—	—	excluded: 1 bin
39270258089; 60505412919178	—	—	—	—	—	—	—	—	excluded: fewer than 50 qualifying idle samples at any fixed clock
10028747453883; 1388292231549; 2779360139651; 73584466053757	—	—	—	—	—	—	—	—	excluded: no clock column

^a Bootstrap blocks of 60 s within segments.

The three 1380 MHz units give 0.085, 0.152 and 0.173 with the median shape (median 0.152) and 0.081–0.219 across the shapes; the bootstrap envelope over units and shapes is 0.071–0.221. For unit 65390563485290 the free b is not identifiable: 64% of the bootstrap replicates reach the upper grid bound and 409 of 2000 replicates fail the criteria (65 of 2000 for unit 60837379255107, none for 10713835868076); its idle periods are mostly short pauses of 3–8 s between loads, and the median full-load power of this job is only 136 W. Inverse-variance weighting gives 0.154 at the median shape, but the weights (0.853, 0.142, 0.005) reflect only segment noise, and the spread between GPUs (0.085–0.173) is

much wider than any single-GPU bootstrap interval. The common- b fit gives $b = 2850$ K and $f_L = 0.143, 0.161$ and 0.082 (mean 0.129 , bootstrap 0.124 – 0.135 , 761 of 1000 stratified replicates successful). Free b and f_L are strongly correlated across bootstrap replicates (Spearman 0.98 and 0.93 for the two identifiable units): a flatter shape goes with a smaller fraction.

Raising the minimum idle-run length from 3 s to 10 s and 30 s leaves f_L of unit 10713835868076 at 0.152 but moves its free b from 2625 to 2450 and 1400 K as the span narrows to 10 K; the post-load decay thus affects the shape estimate rather than the fraction. Memory and GPU temperatures are correlated with coefficients of 0.971 – 0.999 for the 1380 MHz units, so HBM and die contributions cannot be separated, which the module-level definition does not require.

S3.4. Summary of the estimates

The estimates come from two GPU types, each measured with its own method, and the eight H100 workloads share one pair of nodes. They are therefore summarized by method and group (Table S9) rather than as 11 independent points. With equal weight for the H100 and V100 groups, the median is 0.173 with the figure-based temperature pairs and 0.143 with the stress-test pair; treating the H100 node pair as one point alongside the three V100 GPUs gives 0.162 and 0.142 . The center of the evidence is therefore 0.142 – 0.173 , and the nominal 0.15 lies within it. The within-node H100 regressions, which are free of node-to-node differences, give 0.152 and 0.123 , and the common V100 fit gives 0.129 on average at $b = 2850$ K. Pooling all 11 points instead gives medians of 0.204 and 0.136 for the two pairings. These values come from three GPUs and one pair of nodes; they describe a range, not a confidence interval.

Table S9. Leakage-fraction estimates by method and group (median shape unless stated).

Method and group	Estimates	Median
H100, node-to-node, figure-based temperature pairs (8 workloads)	Mistral-7B-v0.3 0.214, Llama-3.1-8B 0.204, Mistral-Nemo-Base-2407 0.222, Gemma-2-27B 0.234, Qwen2.5-32B 0.239, X-CLIP 0.148, EVL 0.040, Vita-CLIP 0.302	0.218
H100, node-to-node, stress-test pair (8 workloads)	0.131, 0.125, 0.136, 0.143, 0.146, 0.097, 0.026, 0.197 (same order)	0.133
H100, within the air-cooled node (8 GPUs)	Qwen2.5-32B 0.152 (95% CI 0.027–0.277); Vita-CLIP 0.123 (0.059–0.186)	—
V100, three GPUs at 1380 MHz	0.152, 0.173, 0.085	0.152
V100, own free b	0.140 (b 2625 K), 0.164 (b 2975 K); third GPU not identifiable	—
V100, common $b = 2850$ K (profile 95% 1925–3750 K)	0.143, 0.161, 0.082	mean 0.129
<i>Summary</i>		
Group-equal-weight median (lower and upper weighted quartiles), figure-based / stress-test pairs	0.173 (0.148–0.222) / 0.143 (0.097–0.152)	
H100 as one point plus three V100, figure-based / stress-test pairs	0.162 / 0.142	
All 11 points pooled, median (quartiles), figure-based / stress-test pairs	0.204 (0.150–0.228) / 0.136 (0.111–0.149)	

Leakage-driven power growth of present GPUs. Each estimate is converted with Eq. (4) into the rate s_{85} at which leakage raises module power at 85 °C (Table S10). The measurements are made at 40–60 °C, so the conversion to 85 °C depends on the leakage shape: the same estimate gives values that differ by a factor of about 1.8 between the flat and steep shapes. With the median shape, the group-equal-weight medians are 0.510% K^{−1} with the figure-based pairs and 0.420% K^{−1} with the stress-test pair, the 0.42 – 0.51% K^{−1} quoted in the main text; the pooled medians are 0.603 (quartiles 0.441 – 0.675) and 0.398% K^{−1} (0.325 – 0.437). The within-node regressions give 0.446 (CI 0.078 – 0.822) and 0.360% K^{−1} (0.172 – 0.549), and the common V100 fit 0.214 – 0.423% K^{−1} (0.337 at its mean fraction). The nominal $f_L =$

0.15 corresponds to 0.365, 0.441 and 0.561% K^{-1} for the three shapes. Of the 21 estimates, only the figure-based Vita-CLIP value (0.899% K^{-1}) exceeds the nominal break-even value of 0.800% K^{-1} ; the median of all 21 is 0.436% K^{-1} .

Table S10. Leakage-driven power growth s_{85} at 85 °C implied by each estimate (% K^{-1} ; flat / median / steep shape; f_L for the median shape).

Estimate	Method	f_L	s_{85}
Mistral-7B-v0.3	H100 node-to-node, figure pairs	0.214	0.501 / 0.631 / 0.924
Llama-3.1-8B	H100 node-to-node, figure pairs	0.204	0.478 / 0.603 / 0.881
Mistral-Nemo-Base-2407	H100 node-to-node, figure pairs	0.222	0.520 / 0.656 / 0.961
Gemma-2-27B	H100 node-to-node, figure pairs	0.234	0.550 / 0.694 / 1.016
Qwen2.5-32B	H100 node-to-node, figure pairs	0.239	0.560 / 0.706 / 1.034
X-CLIP	H100 node-to-node, figure pairs	0.148	0.355 / 0.436 / 0.613
EVL	H100 node-to-node, figure pairs	0.040	0.095 / 0.117 / 0.163
Vita-CLIP	H100 node-to-node, figure pairs	0.302	0.729 / 0.899 / 1.267
Mistral-7B-v0.3	H100 node-to-node, stress-test pair	0.131	0.318 / 0.383 / 0.520
Llama-3.1-8B	H100 node-to-node, stress-test pair	0.125	0.304 / 0.366 / 0.496
Mistral-Nemo-Base-2407	H100 node-to-node, stress-test pair	0.136	0.331 / 0.398 / 0.540
Gemma-2-27B	H100 node-to-node, stress-test pair	0.143	0.349 / 0.420 / 0.571
Qwen2.5-32B	H100 node-to-node, stress-test pair	0.146	0.355 / 0.428 / 0.581
X-CLIP	H100 node-to-node, stress-test pair	0.097	0.236 / 0.283 / 0.384
EVL	H100 node-to-node, stress-test pair	0.026	0.063 / 0.076 / 0.103
Vita-CLIP	H100 node-to-node, stress-test pair	0.197	0.482 / 0.581 / 0.790
Qwen2.5-32B	H100 within node	0.152	0.365 / 0.446 / 0.621
Vita-CLIP	H100 within node	0.123	0.301 / 0.360 / 0.482
10713835868076, GPU 0	V100	0.152	0.339 / 0.446 / 0.702
60837379255107, GPU 0	V100	0.173	0.380 / 0.510 / 0.823
65390563485290, GPU 0	V100	0.085	0.197 / 0.248 / 0.361

Values reported for other devices, converted with Eq. (3) and the three shapes, are listed in Table S11.

Table S11. Other leakage-fraction values converted with Eq. (3) and the three shapes, for comparison (Fig. 2b). Their methods and denominators differ from the definition of Section S3.1.

Source	Device	Method	f_L
Price et al. [17]	Tesla K20 (28 nm planar)	fixed clocks of 705, 805 and 905 MHz, controlled 30–90 °C, fitted static power; load power as denominator	0.066–0.137 (0.061–0.097 with the 225 W board power as denominator)
Kandiah et al. [42]	Quadro GV100 (12 nm)	static component from a power-model fit at 65 °C, no temperature scan; 250 W denominator	0.163–0.266
Zapater et al. [16]	SPARC T3 server CPU (40 nm)	constant load, fan speed varied, CPU rail power, 50–90 °C	0.124–0.163
Crop et al. [39]	Xeon Platinum 8480	static component of a model fit at a locked core clock, 40–100 °C	0.204–0.280 (slope of total power: 0.081–0.087)

S3.5. Direction of biases

The bias mechanisms of the estimates and their directions are listed in Table S12.

Table S12. Bias mechanisms of the leakage-fraction estimates.

Source	Mechanism	Direction for module-level f_L	Magnitude
H100, V100	Board power includes HBM	within the definition; a die-only fraction would be lower	135 MHz to 1380 MHz slope ratio 0.78 on the same GPU and 0.54–0.56 across GPUs, a loose upper bound on the core-voltage-independent temperature-sensitive share
H100, V100	Board power includes regulator losses	unchanged to first order; slightly high if regulator efficiency falls with temperature	not quantified
H100	Clocks not locked; slightly lower throughput of the air-cooled node implies less dynamic power there, which compresses the power difference	low	throughput differences of 0.1–1.0% understate the power difference of the language-model workloads by 1.7–19.6% (frequency only) to 5.1–58.9% (frequency $\times$ voltage ²); Vita-CLIP 1.4–4.1%, X-CLIP 2.5–7.5%, EVL 29–87%
H100	Temperature pairing: stress-test midpoints instead of workload traces	low when the stress-test pair is used	figure-based pairs give 1.53–1.64 times the stress-test values; the proxy temperatures of the two low-utilization workloads are unknown
H100	Node-to-node differences in CPUs, boards and chip lots	undetermined	within-node regression 0.12–0.17 vs 0.23–0.33 between nodes for the same two workloads
H100	Node-to-node power difference proportional to workload power, suggesting a load-proportional component unrelated to leakage	high	per GPU, $\Delta P = -26.6 (\pm 1.9) \text{ W} + 0.1036 (\pm 0.0043) P_{\text{liq}}$, $R^2 = 0.990$, zero at 256 W; correlation of P_{liq} with the figure-based estimates 0.988; $\Delta P/P = 1.34\text{--}5.45\%$
All	Extrapolation from 40–60 °C to 85 °C through the leakage shape	undetermined; affects s_{85} more than f_L	s_{85} differs by a factor of about 1.8 between flat and steep shapes (Table S10)
H100	Leakage measured at partial load (290–566 W per GPU) and divided by 700 W; voltage at full load may be lower under the power cap	high	measured power is 0.60–0.81 of 700 W (EVL 0.42)
V100	250 W limit vs measured full-load power as denominator	small	0.173 (250 W) vs 0.164 (236.7 W measured for job 60837379255107)
All	Supply voltage lowered with temperature at a locked clock	low	on a Xeon Platinum 8480 the total-power slope gives 0.08 against 0.20–0.28 from the model split [39]
All	Supply voltage raised with temperature at a locked clock	high	unknown; the voltages of the two V100 clock levels are not recorded, so the sign cannot be fixed
V100	Post-load decay taken as temperature curvature	mainly affects b ; f_L stable	f_L of unit 10713835868076 0.152 for minimum idle runs of 3, 10 and 30 s; free b 2625 $\rightarrow$ 1400 K
All	Sample size and chip-to-chip variation	undetermined	three V100 GPUs (one job each) and one pair of H100 nodes; the V100 spread of 0.085–0.173 dominates the per-GPU bootstrap widths

S3.6. Coupling between leakage fraction and shape

For subthreshold leakage $I \propto T^2 \exp\{-qV_{\text{th}}(T)/(nkT)\}$ with $V_{\text{th}}(T) = V_{\text{th}0} - \kappa_V(T - T_0)$, the coefficient of Eq. (3) is $b = q(V_{\text{th}0} + \kappa_V T_0)/(nk)$. Raising the threshold voltage by ΔV multiplies the leakage at T_{ref} by $r = \exp\{-q\Delta V/(nkT_{\text{ref}})\}$ and raises b by

$$\Delta b = \frac{q\Delta V}{nk} = T_{\text{ref}} \ln \frac{1}{r}, \quad r = \frac{f_1/(1-f_1)}{f_0/(1-f_0)}, \quad (\text{S9})$$

with $T_{\text{ref}} = 360.65$ K, where the second expression for r holds when only leakage changes and dynamic power stays constant. A numerical check with $n = 1.3$, $\kappa_V = 0.5$ mV K⁻¹ and $\Delta V = 50$ mV, fitted over 50–150 °C, gives $\Delta b = 446.33$ K against 446.32 K from Eq. (S9). For the device data, 0.1 V of threshold corresponds to $\Delta b = 1085$ K and a leakage factor of 0.049 at T_{ref} (Table S13).

Table S13. Leakage reduction factor r , shape change Δb and threshold increase ΔV (for $n = 1.07$) when the leakage fraction is lowered from f_0 to f_1 by raising threshold voltages.

f_0	$f_1 = \mathbf{0.084}$	$f_1 = \mathbf{0.065}$	$f_1 = \mathbf{0.045}$
0.12	r 0.673; Δb 143 K; ΔV 13 mV	0.510; 243 K; 22 mV	0.346; 383 K; 35 mV
0.13	0.614; 176 K; 16 mV	0.465; 276 K; 25 mV	0.315; 416 K; 38 mV
0.15	0.520; 236 K; 22 mV	0.394; 336 K; 31 mV	0.267; 476 K; 44 mV
0.20	0.367; 362 K; 33 mV	0.278; 462 K; 43 mV	0.189; 602 K; 56 mV
0.23	0.307; 426 K; 39 mV	0.233; 526 K; 49 mV	0.158; 666 K; 61 mV

Because every leakage-fraction threshold falls as b grows (Section S9.1; for example the 5% threshold of the nominal radiator is 0.143, 0.112 and 0.081 for the flat, median and steep shapes), a leakage fraction lowered through threshold voltage must meet a tighter threshold than the same fraction at fixed shape. Expressed as leakage-driven power growth, the same thresholds change much less with the shape ($s_5 = 0.349, 0.329$ and 0.301% K⁻¹; $s_0 = 0.763, 0.800$ and 0.825% K⁻¹), because s already contains the steepness. Leakage lowered by raising threshold voltages is therefore assessed by its s at 85 °C, which includes the steeper shape. Applying Eq. (S9) with $n = 1.07$ to the median shape at $f_L = 0.15$, a threshold increase of 12 mV lowers the leakage fraction to 0.109 and raises b to 3447 K, which brings s at 85 °C down to the nominal 5% threshold of 0.329% K⁻¹; the advanced 5% threshold of 0.173% K⁻¹ requires 37 mV ($f_L = 0.054$, $b = 3720$ K). At a fixed shape the corresponding leakage-fraction thresholds (0.112 and 0.059, Section S9.1) would require 11 and 34 mV. Power gating and other measures that leave device thresholds unchanged lower f_L , and with it s , without changing b .

S3.7. Data license

The MIT Supercloud Datacenter Challenge dataset [33] is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 license (CC BY-NC-ND 4.0). The raw telemetry and any per-sample derived data are therefore not redistributed; the repository contains only the scripts that retrieve the files listed in Table S8 and reproduce the fits, and Table S8 reports fitted summary values. The H100 inputs are the published tables and figures of [21]; the digitized temperatures are provided with the scripts.

S4. Plausibility anchors: Starcloud environmental load, NASA GRC radiator and ISS photovoltaic radiator

The consistency checks against published calculations and the physical anchors of Table 3 are documented below.

Turyshev. With $\varepsilon = 0.9$ and $F = 0.85$, the model gives an emitted flux of 650.9 W m⁻² at 350 K (651 in [3]), 2495 m² of emitting face for 1 MW (2500), 12.48 kg kW⁻¹ (12.5) and an area–temperature elasticity of -5.20 (-5.2).

Starcloud environmental load. The white paper [1] evaluates the net flux of a radiator at 293.15 K with $\varepsilon = 0.92$, solar absorptance $\alpha = 0.09$, solar constant 1366 W m⁻², albedo 0.3, Earth view factor 0.25 and an Earth temperature of 253.15 K. Its two-sided blackbody emission of 838 W m⁻² and emission of 770.48 W m⁻² are reproduced (837.5 and 770.53 W m⁻²). Its Earth term is written as

$$q_E^{\text{wp}} = \alpha F_E (aS + \sigma T_E^4), \quad (\text{S10})$$

which gives the paper's 14.46 W m^{-2} . By Kirchhoff's law the absorptance for Earth infrared equals the infrared emissivity, and the solar absorptance applies only to the albedo part:

$$q_E = \alpha F_E aS + \varepsilon F_E \sigma T_E^4. \quad (\text{S11})$$

This gives 62.78 W m^{-2} per face. Substituting Eq. (S11) for Eq. (S10) in the white paper's balance lowers the net flux of a 20°C radiator from 633.08 to 584.8 W m^{-2} . For a linear coolant temperature drop from 35 to 5°C , the mean of T^4 exceeds the fourth power of the mean temperature by 0.53% , so a mean radiating temperature of 20°C is a close description of that radiator.

NASA GRC radiator test. The full-scale radiator test [37] rejected 6.28 kW from a $2.0 \text{ m} \times 2.7 \text{ m}$ panel with inlet water at 400 K and 0.5 kg s^{-1} in a vacuum chamber with a 250 K cold wall; the water leaves at 397.0 K . Writing the rejection as $Q = 2\varepsilon F \eta_{\text{fin}} \sigma (T_f^4 - T_s^4) A$ with the fluid temperature T_f as the fin-root temperature, the cold-wall temperature T_s and a chamber view factor of 0.88 gives $\varepsilon F \eta_{\text{fin}} = 0.481$ and, with $\varepsilon = 0.9$, $\eta_{\text{fin}} = 0.607$ at about 398.5 K . The same expression predicts 8.79 kW with $\eta_{\text{fin}} = 0.85$ ($+40\%$) and 6.31 kW with 0.61 . Because the fin parameter grows with temperature (Section S9.7), the same fin geometry would have $\eta_{\text{fin}} = 0.771, 0.750, 0.721, 0.687$ and 0.637 at $300, 313, 330, 350$ and 380 K ; at the coolant temperatures of the present designs it corresponds to about 0.75 .

ISS photovoltaic radiator. The panel rejects 14 kW from $3.12 \text{ m} \times 13.6 \text{ m}$ of planform [20], i.e. 329.9 W m^{-2} of planform at 17.46 kg m^{-2} . For an ammonia loop cooling from 291 to 276 K (assumed), the two-sided net flux of the model, Q/A from Eq. (5), is 347.3 W m^{-2} with the nominal and advanced parameters, 330.0 W m^{-2} with the heavy parameters, whose fin efficiency is calibrated to this rating (Section S1.1), and 191.7 W m^{-2} with the pessimistic parameters ($\eta_{\text{fin}} = 0.65$, $q_{\text{env}} = 170.5 \text{ W m}^{-2}$). The nominal fin efficiency is thus on the optimistic side of flight hardware, the heavy radiator matches it, and the pessimistic radiator lies 42% below it (Table S14).

Table S14. Consistency checks and plausibility anchors.

Anchor	Quantity	This model	Source
Turyshchev [3]	emitted flux at 350 K , W m^{-2} per face	650.9	651
	emitting-face area for 1 MW , m^2	2495	2500
	radiator specific mass, kg kW^{-1}	12.48	12.5
	area–temperature elasticity	-5.20	-5.2
Starcloud [1]	two-sided blackbody emission at 20°C , W m^{-2}	837.5	838
	emission with $\varepsilon = 0.92$, W m^{-2}	770.53	770.48
	Earth term as written / corrected, W m^{-2}	14.46 / 62.78	14.46 / —
	net flux at 20°C , W m^{-2}	584.8 (corrected)	633.08
NASA GRC [37]	rejected heat, kW ($\eta_{\text{fin}} 0.85 / 0.61$)	8.79 / 6.31	6.28
ISS PVR [20]	planform flux, W m^{-2} (nominal / heavy / pessimistic)	347.3 / 330.0 / 191.7	329.9
Space heat-pump prototype [13]	COP on the model's basis (measured overall efficiency, 0.2 bar evaporator pressure drop)	2.53	2.60

S5. Heat pumps

The heat-pump cycle model, the working-fluid selection, the mass model, the comparison with the prototype and the full results are given below (Fig. S2).

Configuration. The chip loop feeds an evaporator; the condenser is an isothermal radiator at the condensing temperature T_{cond} , which is at least 5 K above the chip-loop outlet. The search covers $T_{\text{cond}} = 320\text{--}460$ K in 4 K steps, chip-loop outlet temperatures of 290–377 K in 3 K steps and loop rises of 4, 8 and 12 K. The evaporator absorbs $Q_c = P_{\text{bus}}/(1 - \kappa)$, the compressor draws $W = Q_c/\text{COP}$, the radiator rejects $Q_c + W$ at T_{cond} , and the solar array and storage are sized for $Q_c + W$.

Cycle. A single-stage vapor-compression cycle is computed with CoolProp 8.0.0 properties [34]. The evaporating temperature is the chip-loop supply temperature minus an evaporator difference; the vapor leaves the evaporator 5 K superheated; for dry fluids an internal heat exchanger (pinch 3 K) superheats the suction gas until compression ends outside the two-phase region; the condenser outlet is saturated liquid, subcooled in the internal heat exchanger before throttling; motor losses are rejected on the condenser side; line pressure drops are neglected. Component efficiencies form three sets: conservative (isentropic 0.65, motor 0.90, evaporator difference 5 K), nominal (0.75, 0.92, 3 K) and optimistic (0.80, 0.95, 2 K). COP values are tabulated over supply temperatures of 274–378 K and condensing temperatures of 316–464 K and interpolated bilinearly.

Working fluid. At each operating point the fluid with the highest COP is selected from R1233zd(E), R245fa, R1234ze(Z), R1336mzz(Z), n-pentane, isopentane, cyclopentane and water, subject to a condensing temperature at least 10 K below the critical temperature, an overall pressure ratio of at most 10 (8 and 15 as sensitivities; 4.7 in the prototype demonstration), a discharge temperature of at most 473 K and a volumetric cooling capacity of at least 300 kJ m⁻³. Toluene is excluded.

Mass. The heat-pump mass is

$$M_{\text{hp}} = m_{\text{comp}}B + m_{\text{evap}}Q_c, \quad B \in \{W, Q_c, Q_c + W\}, \quad (\text{S12})$$

with compressor specific masses of 4.6 kg kW⁻¹ (scaling with electric work, nominal), 2.0 kg per kW of cooling and 1.39 kg per kW of rejected heat; the three bases give the same mass at the prototype rated COP of 2.3 [13]. The value of 2.0 kg per kW of cooling lies between the 0.4 of the prototype centrifugal compressor, the 0.45 aviation convention and about 4 for commercial and space hardware. The prototype compressor itself weighs 2 kg for 1.98 kW of electric power and 5.141 kW of cooling, i.e. 1.01 kg kW⁻¹ (0.39 kg per kW of cooling), or 3.03 kg kW⁻¹ including 4 kg of electronics; the nominal value is 4.6 times the compressor alone.

Prototype comparison. The prototype [13] runs on isopentane between saturation temperatures of 45 and 100 °C with a measured cooling of 5141 W, fluid work of 1445 W, 535 W rejected to cooling water and an overall compressor efficiency of 0.68. Its COP is 2.33 on the paper's basis and 2.60 on the basis of this model, in which motor losses are rejected on the condenser side. The model cycle at the same saturation temperatures with the measured overall efficiency and the measured evaporator pressure drop of 0.2 bar gives 2.53 (3% below), at a pressure ratio of 4.08 (4.7 measured). Without the pressure drop it gives 2.79; with an internal heat exchanger 2.99 (nominal set) and 2.53 (conservative set).

Literature radiating temperatures. A heat pump “reaching” a literature radiating temperature T_{lit} condenses at $T_{\text{lit}}/\eta_{\text{fin}}^{1/4}$, i.e. at 91.4 °C for 350 K and 127.8 °C for 385 K with $\eta_{\text{fin}} = 0.85$, and at 94.8 and 131.6 °C with the 0.819 of the heavy radiator.

Fin efficiency. Results are given with a constant fin efficiency and with one that declines with temperature at fixed fin geometry (Section S9.7). In the second case the condenser, which is isothermal and hotter than the chip loop, loses part of its fin efficiency: for the nominal radiator η_{fin} falls from 0.866 at 300 K to 0.792 at 360 K, and at the condensing temperature of the best heat-pump design (340 K) it is 0.818. Both the heat-pump design and the passive design it is compared with use the same fin-efficiency model (Table S15).

Table S15. Heat-pump designs relative to the passive optimum with the same parameters ($f_L = 0.15$, median shape; values for the flat and steep shapes differ by at most 0.2 percentage points). Positive values mean the heat-pump design is heavier; “–” means unreachable within the pressure-ratio limit, or with the declining fin efficiency at any pressure ratio up to 10; n.c., not computed. Rows with $\eta_{\text{fin}}(T)$ use the fin efficiency that declines with temperature.

Radiator	Efficiency set	Compressor scaling	η_{fin}	$T_{\text{eff}} = 350 \text{ K}$ (%)	$T_{\text{eff}} = 385 \text{ K}$, pressure ratio $\leq 8 / 10 / 15$ (%)	Best design (%)	Best design $T_{\text{cond}} \text{ (K)}$; COP	COP at 350 / 385 K
Nominal	nominal	electric work	constant	4.7	– / 35.7 / 14.8	3.4	344; 4.95	2.94 / 1.27
Nominal	nominal	electric work	$\eta_{\text{fin}}(T)$	8.9	n.c. / – / n.c.	5.0	340; 5.61	2.45 / –
Nominal	nominal	cooling	constant	5.7	– / 31.2 / 12.0	5.4	356; 3.23	2.94 / 1.27
Nominal	nominal	rejected heat	constant	5.4	– / 32.5 / 12.8	4.9	352; 3.96	2.94 / 1.27
Nominal	conservative	electric work	constant	9.7	– / 47.7 / 23.4	6.7	336; 5.06	2.38 / 1.03
Nominal	conservative	cooling	constant	9.8	– / 40.8 / 18.9	8.9	348; 3.52	2.38 / 1.03
Nominal	conservative	rejected heat	constant	9.8	– / 42.9 / 20.2	8.3	344; 3.94	2.38 / 1.03
Nominal	optimistic	electric work	constant	2.3	n.c. / 27.7 / n.c.	1.6	352; 4.49	3.32 / 1.51
Nominal	Carnot $\times 0.409$ [43]	electric work	constant	10.8	16.8 (no pressure-ratio limit)	7.2	324; 8.42	2.28 / 1.37
Advanced	nominal	electric work	constant	14.3	n.c. / 52.7 / n.c.	7.9	324; 11.16	2.94 / 1.27
Advanced	nominal	electric work	$\eta_{\text{fin}}(T)$	18.8	n.c. / – / n.c.	8.4	324; 11.16	2.44 / –
Heavy	nominal	electric work	constant	–4.9	n.c. / 21.7 / n.c.	–4.9	364; 2.98	2.73 / 1.16
Heavy	nominal	electric work	$\eta_{\text{fin}}(T)$	0.6	n.c. / – / n.c.	–1.0	356; 3.58	2.17 / –
Pessimistic	nominal	electric work	constant	–11.6	n.c. / – / n.c.	–15.5	380; 2.14	1.57 / –
Pessimistic	nominal	electric work	$\eta_{\text{fin}}(T)$	–	n.c. / – / n.c.	–6.1	368; 2.73	– / –

The COP columns give the cycle COP; with the declining fin efficiency the condenser must run hotter to reach the same effective radiating temperature, which lowers the COP at 350 K (2.45 instead of 2.94 for the nominal radiator) and makes 385 K unreachable for every radiator. For the pessimistic radiator, 385 K would require condensing beyond the envelope of the qualifying fluids at a pressure ratio of 10. The Carnot row uses a constant second-law efficiency of 0.409, the mean of a screw-chiller performance curve over its valid range [43]. With a declining fin efficiency, the heavy radiator’s advantage from a heat pump shrinks from 4.9% to 1.0%, and the pessimistic radiator’s from 15.5% to 6.1% (Table S16).

Table S16. Mass change of the best heat-pump design relative to the passive optimum versus radiator areal mass and solar-array specific power ($\eta_{\text{fin}} = 0.85$ at the calibration temperature, nominal compressor scaling with electric work; $f_L = 0.15$, median shape). Negative values mean the heat-pump design is lighter; the optimum condensing temperature (K) is given in brackets.

Efficiency set; η_{fin}	SP_{pv} (W kg ⁻¹)	$\sigma_{\text{rad}} = 2$	5	10	17.5 kg m ⁻²
Nominal; constant	100	+10.5 (320)	+7.9 (324)	+3.4 (344)	-4.1 (364)
Nominal; constant	200	+10.5 (320)	+7.2 (332)	+0.1 (360)	-8.8 (376)
Nominal; constant	300	+10.6 (320)	+6.6 (340)	-1.6 (364)	-10.9 (380)
Nominal; constant	500	+10.7 (324)	+5.8 (344)	-3.2 (368)	-12.8 (384)
Nominal; $\eta_{\text{fin}}(T)$	100	+10.8 (320)	+8.4 (324)	+5.0 (340)	-0.9 (356)
Nominal; $\eta_{\text{fin}}(T)$	200	+10.8 (320)	+8.0 (324)	+2.5 (352)	-4.7 (368)
Nominal; $\eta_{\text{fin}}(T)$	300	+10.8 (320)	+7.6 (332)	+1.2 (356)	-6.4 (372)
Nominal; $\eta_{\text{fin}}(T)$	500	+11.0 (320)	+7.2 (340)	-0.1 (360)	-7.9 (376)
Conservative; constant	100	+12.5 (320)	+10.1 (324)	+6.7 (336)	+0.4 (356)
Conservative; constant	200	+12.4 (320)	+9.5 (324)	+3.9 (348)	-4.0 (368)
Conservative; constant	300	+12.4 (320)	+9.1 (328)	+2.4 (356)	-5.9 (372)
Conservative; constant	500	+12.4 (324)	+8.6 (336)	+0.9 (360)	-7.8 (376)
Conservative; $\eta_{\text{fin}}(T)$	100	+12.9 (320)	+10.6 (320)	+7.8 (328)	+3.3 (348)
Conservative; $\eta_{\text{fin}}(T)$	200	+12.6 (320)	+10.0 (324)	+6.1 (344)	-0.1 (360)
Conservative; $\eta_{\text{fin}}(T)$	300	+12.6 (320)	+9.8 (324)	+5.0 (348)	-1.6 (364)
Conservative; $\eta_{\text{fin}}(T)$	500	+12.7 (320)	+9.6 (328)	+3.8 (352)	-3.1 (368)

A heat pump pays off only with a heavy radiator and a light power system. With the declining fin efficiency the break-even moves to roughly one step higher in solar-array specific power. With optimistic efficiencies and light heat-pump components (0.92 kg kW_e⁻¹ compressor, 0.5 kg kW_{th}⁻¹ evaporator) and constant fin efficiency, the heat pump saves up to 21.5% ($\sigma_{\text{rad}} = 17.5 \text{ kg m}^{-2}$, $SP_{\text{pv}} = 500 \text{ W kg}^{-1}$), with optimum condensing temperatures of at most 388 K.

S6. Clock scaling

The clock-scaled chip model, its parameters and the results with the clock ratio as a design variable are given below (Fig. S3).

Chip model. With clock ratio φ and supply-voltage ratio v , the power of one chip relative to its value at $\varphi = 1$ and 87.5 °C is

$$P(\varphi, T_j) = (1 - f_L - f_u) \varphi v^2 + f_u \varphi + f_L v \exp\{c_I V_{\text{nom}}(v - 1)\} \frac{g(T_j)}{g(T_{\text{ref}})}, \quad (\text{S13})$$

$$v(\varphi) = \begin{cases} 1 - (1 - v_{\min}) \frac{1 - \varphi}{1 - \varphi_v}, & \varphi \geq \varphi_v, \\ v_{\min}, & \varphi < \varphi_v, \end{cases} \quad (\text{S14})$$

where f_u is the share of power that scales with work but not with core voltage (nominally 0). The voltage–frequency curve uses $v_{\min} = 0.78$, $\varphi_v = 0.72$ and $V_{\text{nom}} = 0.91 \text{ V}$ from the measured A100 curve of [36], linearized above the knee. The voltage coefficient of leakage, $c_I = 1.0 \text{ V}^{-1}$, follows from the voltage-scaling multiplier of the core-supply static current in the AMD tool [28]. Delivering 1 kW_{ref} requires $1/\varphi$ chips, so m_{IT} scales with $1/\varphi$ while m_{fix} does not, and the chip power per kW_{ref} is the loop average of Eq. (S13) divided by φ and by its value at $\varphi = 1$ and 80 °C. The junction fixed point uses Eq. (S13), so a lower clock also lowers the junction-to-coolant difference; ΔT_{hbm} is kept at 40 K independent of clock in the main case. Three model variants are examined: ΔT_{hbm} scaled with the power of one chip, as if the memory heat flux fell with the clock; a share $f_u = 0.15$ of power that scales with work but not with voltage; and a stronger voltage dependence of leakage, $c_I = 3 \text{ V}^{-1}$. The clock ratio is optimized over 0.50–1.00 in steps of 0.05 and then 0.01 within ± 0.04 of the best value, jointly with the coolant outlet and loop rise (Table S17).

Table S17. Clock-scaled optimum (chip limit 85 °C, HBM limit 95 °C) and the saving from raising the limits to 150 °C (flat / median / steep shape; masses for the median shape). Except for the advanced radiator at $f_L = 0.22$, whose

optimum is interior, the optimum clock ratio is 0.72, at $v = 0.780$, with the coolant outlet at 55.00 °C and the HBM limit binding.

Case	f_L	M , fixed clock (kg kW _{ref} ⁻¹)	M , clock- scaled	Saving from clock scaling (%)	Hot-chip junction (°C)	Saving 85 → 150 °C, clock- scaled (%)	Saving 85 → 150 °C, fixed clock (%)	Clock-scaled at 150 °C vs fixed clock at 85 °C (%)
Advanced	0.10	35.70	28.73	19.3 / 19.5 / 19.8	72.3	1.60 / 1.13 / 0.78	3.45 / 2.18 / 1.09	20.6 / 20.4 / 20.4
Advanced	0.15	35.60	28.68	19.2 / 19.5 / 19.8	72.2	0.46 / 0.26 / 0.15	1.24 / 0.55 / 0.14	19.5 / 19.7 / 19.9
Advanced	0.22	35.37	28.54	18.9 / 19.3 / 19.6	69.7	0.00 / 0.00 / 0.00	0.08 / 0.00 / 0.00	18.9 / 19.3 / 19.6
Nominal	0.10	44.78	33.50	25.0 / 25.2 / 25.5	72.3	5.13 / 3.87 / 2.78	8.83 / 6.07 / 3.45	28.8 / 28.1 / 27.5
Nominal	0.15	44.68	33.45	24.8 / 25.1 / 25.6	72.2	2.52 / 1.71 / 1.11	4.55 / 2.59 / 1.05	26.7 / 26.4 / 26.4
Nominal	0.22	44.43	33.32	24.6 / 25.0 / 25.6	72.1	0.62 / 0.29 / 0.13	1.30 / 0.38 / 0.01	25.0 / 25.2 / 25.7
Heavy	0.10	59.87	41.39	30.6 / 30.9 / 31.2	72.3	10.50 / 8.08 / 5.87	15.68 / 11.09 / 6.58	37.9 / 36.5 / 35.2
Heavy	0.15	59.70	41.31	30.4 / 30.8 / 31.3	72.2	6.12 / 4.32 / 2.89	9.18 / 5.58 / 2.57	34.7 / 33.8 / 33.2
Heavy	0.22	59.39	41.18	30.1 / 30.7 / 31.3	72.1	2.44 / 1.44 / 0.81	3.62 / 1.53 / 0.31	31.8 / 31.7 / 31.9

At the clock-scaled optimum the chips run at about 72 °C and only the HBM limit binds, so raising the chip limit alone saves nothing. Raising the HBM limit alone saves 1.49% with the nominal radiator at $f_L = 0.15$ (1.71% when both limits rise), 0.26% (0.26%) with the advanced and 3.09% (4.32%) with the heavy radiator (Section S9.8). At the 150 °C limits all clock-scaled optima are interior. Clock scaling requires 1/0.72, about 1.39 times, as many chips (Table S18).

Table S18. Clock-scaled optimum under model variants (median shape, $f_L = 0.15$). “ ΔT_{hbm} scaled” lets the HBM-to-coolant difference fall with the power of one chip.

Case	ΔT_{hbm}	f_u	c_f (V ⁻¹)	M (kg kW _{ref} ⁻¹)	Saving vs fixed clock (%)	Coolant (°C)	Hot chip / HBM (°C)	Binding	Saving 85 → 150 °C (%)	Chip limit alone (%)
Nominal	fixed	0	1	33.451	25.1	55.00	72.2 / 95.0	HBM	1.71	0.00
Nominal	fixed	0	3	32.739	26.7	55.00	71.4 / 95.0	HBM	3.62	0.00
Nominal	fixed	0.15	1	35.363	20.9	55.00	74.1 / 95.0	HBM	1.99	0.00
Nominal	fixed	0.15	3	34.603	22.6	55.00	73.2 / 95.0	HBM	4.07	0.00
Nominal	scaled	0	1	32.951	26.3	66.56	85.0 / 85.6	chip	0.22	0.22
Nominal	scaled	0	3	31.911	28.6	67.73	85.0 / 85.6	chip	1.12	0.87
Nominal	scaled	0.15	1	34.815	22.1	64.87	85.0 / 85.7	chip	0.45	0.44
Nominal	scaled	0.15	3	33.747	24.5	66.04	85.0 / 85.6	chip	1.64	1.15
Heavy	fixed	0	1	41.313	30.8	55.00	72.2 / 95.0	HBM	4.32	0.00
Heavy	fixed	0.15	3	42.977	28.0	55.00	73.2 / 95.0	HBM	8.42	0.00
Heavy	scaled	0	1	40.038	32.9	66.56	85.0 / 85.6	chip	1.27	1.07
Heavy	scaled	0.15	1	42.760	28.4	64.87	85.0 / 85.7	chip	1.92	1.45

Across the eight variants the saving from clock scaling is 20.9–28.6% with the nominal radiator and 26.2–35.5% with the heavy one, and the saving from raising the limits under clock scaling is 0.22–4.07% and 1.27–8.42%. When the

HBM difference scales with chip power, the chip limit binds again at a coolant outlet temperature of 64.9–67.7 °C and the saving from higher ratings becomes smaller (Table S19).

Table S19. Leakage-fraction thresholds at the clock-scaled optimum (saving from raising the limits to 150 °C; flat / median / steep shape); “–” means the saving stays below the threshold for all f_L .

Case	10% saving	5% saving	2% saving
Advanced	– / – / –	0.039 / 0.030 / 0.021	0.089 / 0.075 / 0.061
Nominal	0.046 / 0.035 / 0.024	0.102 / 0.083 / 0.064	0.164 / 0.141 / 0.119
Heavy	0.105 / 0.082 / 0.060	0.167 / 0.139 / 0.112	0.233 / 0.201 / 0.173

Cost basis. Clock scaling trades power-system mass for chips. With the objective $M + m_{eq}/\varphi$, where m_{eq} is the chip purchase price per kW_{ref} divided by the launch price per kilogram (kg kW_{ref}^{−1}), the saving from clock scaling decreases with m_{eq} and vanishes at the break-even values of Table S20.

Table S20. Saving from clock scaling on the cost basis (median shape; optimum clock ratio in brackets) and break-even m_{eq} .

Case	f_L	$m_{eq} = 0$	5	10	20	50 kg kW _{ref} ^{−1}	Break-even m_{eq} (kg kW _{ref} ^{−1})
Advanced	0.15	19.5% (0.72)	12.3% (0.72)	7.0% (0.77)	2.1% (0.88)	0.0% (1.00)	37.4 (39.0–35.8 for f_L 0.10–0.22)
Nominal	0.15	25.1% (0.72)	18.7% (0.72)	13.4% (0.72)	6.4% (0.81)	0.4% (0.95)	62.6 (64.9–59.5)
Heavy	0.15	30.8% (0.72)	25.4% (0.72)	20.8% (0.72)	13.3% (0.72)	3.5% (0.89)	105.2 (108.9– 99.6)

S7. Lifetime

The Arrhenius lifetime sensitivity is defined below with two replacement models and quantified for the nominal radiator, with the HBM limit rising with the chip limit; the advanced and heavy radiators are summarized at the end.

The acceleration factor of a chip at junction temperature T_j relative to 80 °C is

$$AF = \exp \left[\frac{E_a}{k_B} \left(\frac{1}{T_{80}} - \frac{1}{T_j} \right) \right], \quad (S15)$$

averaged over the 41 chip positions of the loop at their stable junction temperatures, with activation energies E_a of 0.5, 0.7 and 0.9 eV [44]. A design life of 5 years at 80 °C gives a chip life of 5/AF years. Two replacement models bound the mass consequence.

- *Model A, deterministic wear-out.* Chips that wear out before the 5-year mission are replaced in batches, which adds $m_{IT} \max(AF - 1, 0)$ to the mass.
- *Model B, constant failure rate with spares.* Failures occur at a base rate λ_0 at 80 °C, accelerated by AF, and are covered by spares of mass $m_{IT} \lambda_0 t_{life} AF$ with $t_{life} = 5$ years; $\lambda_0 = 2\%$ per year is assumed, with 5% per year as a sensitivity.

The net saving is computed relative to the 85 °C design with its own replacement or spares mass, and the failure-rate reduction needed to keep the 85 °C life is the ratio of the mean acceleration factors (Table S21).

Table S21. Chip life, net saving with the two replacement models and required reduction of the thermally activated failure rate (nominal radiator, $f_L = 0.15$; values for $E_a = 0.5 / 0.7 / 0.9$ eV; model B with $\lambda_0 = 2\%$ per year).

Shape	T_{cap} (°C)	Life (years)	Net saving, model A (%)	Net saving, model B (%)	Required failure-rate reduction (×)
Flat	85	5.13 / 5.14 / 5.14	0 / 0 / 0	0 / 0 / 0	1 / 1 / 1
Flat	100	2.78 / 2.18 / 1.70	−9.1 / −16.9 / −27.0	2.0 / 1.3 / 0.3	1.84 / 2.36 / 3.02
Flat	110	1.98 / 1.35 / 0.91	−19.5 / −38.0 / −65.8	1.9 / 0.0 / −2.7	2.58 / 3.81 / 5.63
Flat	150	1.50 / 0.91 / 0.55	−32.0 / −65.6 / −121.8	0.8 / −2.5 / −8.0	3.42 / 5.64 / 9.33
Median	85	5.16 / 5.19 / 5.20	0 / 0 / 0	0 / 0 / 0	1 / 1 / 1
Median	100	2.83 / 2.23 / 1.75	−9.6 / −17.0 / −26.7	1.1 / 0.4 / −0.6	1.82 / 2.32 / 2.97
Median	≥110	2.33 / 1.69 / 1.21	−15.4 / −28.2 / −46.4	0.7 / −0.5 / −2.3	2.22 / 3.08 / 4.29
Steep	85	5.22 / 5.27 / 5.29	0 / 0 / 0	0 / 0 / 0	1 / 1 / 1
Steep	≥100	3.50 / 2.98 / 2.51	−5.7 / −9.6 / −14.5	0.3 / −0.1 / −0.6	1.49 / 1.77 / 2.11

The optimum stops changing once the limit no longer binds (from 110 °C for the median and 100 °C for the steep shape; the flat-shape optimum leaves the limit at a hot-chip temperature of about 117 °C), so the entries marked $\geq$ apply up to 150 °C. Without a lifetime term the median-shape saving at these limits is 2.59%. Model A turns it into a mass increase of 15.4–46.4%; model B gives −2.3 to +0.7% with $\lambda_0 = 2\%$ per year and −9.5 to +0.7% with 2–5% per year. With $f_L = 0.10$ and the median shape, a limit of 110 °C gives lives of 1.84, 1.22 and 0.81 years and net savings of −21.3 to −75.6% (model A) and −2.5 to +2.8% (model B) against 5.60% without a lifetime term. For the median shape at 150 °C, the heavy radiator gives −67.6 to −17.1% (model A) and −1.8 to +3.2% (model B) against 5.58%, and the advanced radiator −16.4 to −6.9% and −1.3 to −0.3% against 0.55%.

S8. External anchors

Published radiating temperatures of two concepts are mapped to coolant, junction and HBM temperatures with the closed mapping of Section 2.6 ($\Delta T_{\text{jc}} = \Delta T_{\text{hbm}} = 40$ K).

Starcloud. The white paper [1] gives a mean radiating temperature of 20 °C (293.15 K). In an interview the company’s chief executive described keeping the radiator at about 50 °C (323.15 K) [45].

SpaceX A11. The reported figures are an average compute payload of 120 kW, a peak of 150 kW and a deployable radiator of up to 110 m² [38]. If the full heat load is rejected from both faces of 110 m² of planform with $\varepsilon = 0.9$, $F = 0.85$ and the nominal environmental load, the radiating temperature is

$$T = \left[\frac{Q/A + q_{\text{env}}}{2\varepsilon F\sigma} \right]^{1/4}, \quad (\text{S16})$$

i.e. 344.2 K for 120 kW and 362.1 K for 150 kW. Additional rejection surfaces or a lower heat load would lower these values (Table S22).

Table S22. External anchors mapped to coolant, junction and HBM temperatures (°C). Junction temperatures with leakage are given for the flat / median / steep shape; “run” means no stable operating point at a fixed clock.

Radiator	Anchor	T (K)	Coolant	Junction and HBM, no leakage	Junction, $f_L = 0.10$	Junction, $f_L = 0.15$	Junction, $f_L = 0.22$
Nominal	Starcloud white paper, 20 °C	293.15	36.2	76.2	75.1 / 74.9 / 74.5	74.5 / 74.1 / 73.6	73.5 / 72.9 / 72.0
Nominal	Starcloud interview, 50 °C	323.15	68.5	108.5	111.6 / 112.9 / 116.2	113.7 / 116.3 / run	117.6 / 125.5 / run
Nominal	AII, 120 kW	344.2	90.4	130.4	139.3 / 145.8 / run	146.8 / run / run	run / run / run
Nominal	AII, 150 kW	362.1	109.0	149.0	166.6 / run / run	197.7 / run / run	run / run / run
Heavy	Starcloud white paper, 20 °C	293.15	39.1	79.1	78.2 / 78.0 / 77.7	77.7 / 77.4 / 76.9	76.9 / 76.4 / 75.6
Heavy	Starcloud interview, 50 °C	323.15	71.6	111.6	115.4 / 117.0 / 121.9	117.9 / 121.5 / run	123.0 / 139.5 / run
Heavy	AII, 120 kW	344.2	93.8	133.8	143.9 / 152.4 / run	152.9 / run / run	run / run / run
Heavy	AII, 150 kW	362.1	112.5	152.5	172.4 / run / run	run / run / run	run / run / run

Below the reference junction temperature of 87.5 °C, leakage is smaller than at the reference, so the junction with leakage can lie below the value without leakage feedback.

S9. Supplementary results

Supplementary results are given below on the present design point and the savings over the full leakage-fraction scan (S9.1), the condition that decides which limit binds (S9.2), the sampling and validity of the break-even relation (S9.3), cold-plate arrangements (S9.4), the loop-rise penalty and the two-phase loop (S9.5), the temperature-stack lever (S9.6), design margin, fin efficiency and end-of-life load (S9.7), the attribution of the saving to chip and HBM limits (S9.8), system-level stability (S9.9), the pessimistic radiator (S9.10), the convention in which the HBM limit equals the chip limit (S9.11) and the environmental load versus solar beta angle (S9.12).

S9.1. Present design point and savings over the leakage-fraction scan

Table S23 gives the mass-optimal designs for the present limits.

Table S23. Mass-optimal designs for present limits (chip 85 °C, HBM 95 °C; $f_L = 0.15$, median shape, range across the three shapes in brackets; masses in kg kW_{ref}⁻¹). All nine combinations of radiator and shape are pinned by the chip limit.

Quantity	Advanced	Nominal	Heavy
On-orbit mass M	35.60 (35.53–35.65)	44.68 (44.60–44.73)	59.70 (59.57–59.79)
Compute / fixed / radiator / solar array / storage	7 / 6 / 9.11 / 11.40 / 2.10	7 / 6 / 18.07 / 11.49 / 2.11	7 / 6 / 33.10 / 11.49 / 2.11
Coolant outlet (°C); loop rise (K)	45.45; 12	45.45; 10	45.45; 10
Hot-chip junction; HBM (°C); HBM margin (K)	85.00; 85.45; 9.55	85.00; 85.45; 9.55	85.00; 85.45; 9.55
T_{eff} (K)	300.0	301.0	298.2
ϕ_{rad} ; ϕ_{eff}	0.403; 0.406	0.570; 0.570	0.709; 0.715
Radiator planform per kW_{ref} (m^2)	1.821	1.807	1.891
Leakage share of the hot chip	14.0%	14.0%	14.0%
Leakage-driven power growth s of the hot chip ($\% \text{ K}^{-1}$)	0.441 (0.365–0.561)	0.441 (0.365–0.561)	0.441 (0.365–0.561)
Mass increase with coolant 5 / 10 / 15 / 25 K below the optimum (%)	0.81 / 1.98 / 3.49 / 7.79	1.82 / 4.12 / 6.97 / 14.39	2.94 / 6.47 / 10.60 / 21.26
Coolant range within 1% of the minimum mass (°C)	39.60–45.35	42.60–45.35	43.60–45.35
Optimum without chip and HBM limits: coolant (°C); hot chip (°C); T_{eff} (K); saving (%)	54.10; 95.8; 306.8; 0.55	62.10; 106.7; 314.5; 2.59	67.60; 114.9; 316.9; 5.58

With $f_L = 0.10$ all combinations are also pinned by the chip limit; at $f_L = 0.22$ the advanced radiator with the median and steep shapes has an interior optimum (coolant 44.85 and 42.35 °C). The pessimistic radiator (Section S9.10) is also pinned by the chip limit (Table S24).

Table S24. Saving from raising the chip limit from 85 to 150 °C, with the HBM limit rising with it, over the leakage-fraction scan (%; flat / median / steep shape). The last column holds the HBM limit at 95 °C. The row $f_L = 0$ describes a chip without leakage.

f_L	Advanced	Nominal	Heavy	Nominal, HBM held at 95 °C
0	14.85	23.53	32.23	5.71
0.02	11.91 / 10.70 / 8.43	20.43 / 19.09 / 16.11	28.98 / 27.55 / 24.26	5.36 / 5.29 / 5.17
0.05	7.74 / 5.94 / 4.02	15.74 / 12.69 / 8.91	24.02 / 20.51 / 14.69	4.81 / 4.61 / 4.29
0.08	4.81 / 3.33 / 1.93	11.25 / 8.20 / 5.11	18.99 / 14.25 / 9.13	4.23 / 3.89 / 3.33
0.10	3.45 / 2.18 / 1.09	8.83 / 6.07 / 3.45	15.68 / 11.09 / 6.58	3.82 / 3.39 / 2.70
0.12	2.38 / 1.35 / 0.54	6.86 / 4.41 / 2.24	12.76 / 8.54 / 4.64	3.40 / 2.88 / 2.04
0.14	1.57 / 0.77 / 0.23	5.25 / 3.12 / 1.38	10.28 / 6.47 / 3.16	2.97 / 2.38 / 1.36
0.15	1.24 / 0.55 / 0.14	4.55 / 2.59 / 1.05	9.18 / 5.58 / 2.57	2.77 / 2.12 / 1.05
0.16	0.97 / 0.38 / 0.07	3.93 / 2.12 / 0.76	8.17 / 4.79 / 2.06	2.56 / 1.85 / 0.76
0.18	0.54 / 0.16 / 0.00	2.85 / 1.37 / 0.34	6.38 / 3.43 / 1.26	2.14 / 1.31 / 0.34
0.20	0.27 / 0.02 / 0.00	2.00 / 0.78 / 0.10	4.88 / 2.36 / 0.70	1.69 / 0.78 / 0.10
0.22	0.08 / 0.00 / 0.00	1.30 / 0.38 / 0.01	3.62 / 1.53 / 0.31	1.22 / 0.38 / 0.01
0.25	0.00 / 0.00 / 0.00	0.57 / 0.08 / 0.00	2.16 / 0.67 / 0.02	0.57 / 0.08 / 0.00
0.30	0.00 / 0.00 / 0.00	0.05 / 0.00 / 0.00	0.62 / 0.02 / 0.00	0.05 / 0.00 / 0.00
0.35	0.00 / 0.00 / 0.00	0.00 / 0.00 / 0.00	0.03 / 0.00 / 0.00	0.00 / 0.00 / 0.00

What leakage takes away. Between $f_L = 0$ and 0.15 (median shape) the saving falls from 14.85% to 0.55% with the advanced radiator, from 23.53% to 2.59% with the nominal and from 32.23% to 5.58% with the heavy one, so present leakage removes 96%, 89% and 83% of the saving available to a chip without leakage. At $f_L = 0.15$ the saving with the nominal radiator is 1.156 kg $\text{kW}_{\text{ref}}^{-1}$, or 3.65% of the power-proportional mass (0.196 kg and 0.87% advanced; 3.334 kg and 7.14% heavy). Along the scan the leakage share of the hot chip grows from 14.0% to 23.7% (nominal; 18.5%

advanced, 28.1% heavy), 90% of the saving is reached at chip limits of 93.6, 99.7 and 104.4 °C, and the optimum at 150 °C is interior, with effective radiating temperatures of 306.8, 314.5 and 316.9 K and designs within 1% of the minimum mass spanning 297.3–312.5, 306.4–319.7 and 309.8–319.7 K. Without leakage the optimum reaches 362.0, 363.0 and 359.7 K. Holding the HBM limit at 95 °C matters most when leakage is low: at $f_L = 0$ it leaves 5.71% of the 23.53%, whereas at 0.15 it leaves 2.12% of 2.59% (82%; 3.79% of 5.58% with the heavy radiator) because leakage stops the optimum near an HBM temperature of 102 °C. Capping the HBM limit at 105 °C gives 2.59% (nominal), 0.55% (advanced) and 5.50% (heavy) at $f_L = 0.15$, and 10.38% (nominal) at $f_L = 0$ (Table S25).

Table S25. Leakage at which the saving from raising the limits to 150 °C falls to 10%, 5%, 2% and zero, as leakage fraction and as leakage-driven power growth s of the present hot chip (flat / median / steep shape). The last two columns give the relative spread of the break-even values across the three shapes.

Radiator	10%	5%	2%	Zero	Spread of $s_0 / f_{L,0}$
Advanced, f_L	0.033 / 0.023 / 0.014	0.078 / 0.059 / 0.041	0.129 / 0.104 / 0.079	0.235 / 0.204 / 0.173	11.5% / 30.8%
Advanced, s (% K ⁻¹)	0.080 / 0.068 / 0.052	0.188 / 0.173 / 0.151	0.313 / 0.304 / 0.292	0.577 / 0.603 / 0.647	
Nominal, f_L	0.090 / 0.066 / 0.044	0.143 / 0.112 / 0.081	0.200 / 0.163 / 0.125	0.310 / 0.270 / 0.219	7.8% / 34.1%
Nominal, s (% K ⁻¹)	0.218 / 0.194 / 0.162	0.349 / 0.329 / 0.301	0.489 / 0.479 / 0.466	0.763 / 0.800 / 0.825	
Heavy, f_L	0.142 / 0.108 / 0.074	0.198 / 0.157 / 0.116	0.254 / 0.208 / 0.161	0.364 / 0.302 / 0.254	6.6% / 36.0%
Heavy, s (% K ⁻¹)	0.347 / 0.317 / 0.276	0.484 / 0.462 / 0.432	0.623 / 0.614 / 0.604	0.899 / 0.898 / 0.958	

Present GPUs ($s = 0.365$, 0.441 and 0.561% K⁻¹ at $f_L = 0.15$ for the three shapes; 0.42–0.51% K⁻¹ for the group medians of Section S3.4) lie between s_5 and s_2 with the nominal radiator, near or below s_2 with the advanced radiator and near s_5 with the heavy one. With the HBM limit held at 95 °C the break-even values are unchanged, because near break-even the 150 °C optimum stays close to the present corner, but the 5% threshold of the nominal radiator falls to $f_L = 0.033$ ($s_5 = 0.096\%$ K⁻¹); with the HBM limit capped at 105 °C it is 0.108 (0.316% K⁻¹).

S9.2. Which limit binds

With both temperature differences referred to the coolant outlet, the hot chip reaches its limit at a coolant temperature of $T_{cap} - \Delta T_{jc}P(T_{cap})$ and the HBM at $T_{hbm, cap} - \Delta T_{hbm}$. The chip limit therefore binds first if and only if

$$D_{rating} - D_{stack} > \delta, \quad D_{rating} = T_{hbm, cap} - T_{cap}, \quad D_{stack} = \Delta T_{hbm} - \Delta T_{jc}, \quad \delta = \Delta T_{jc} [1 - P(T_{cap})], \quad (S17)$$

with P normalized at 87.5 °C as in Eq. (2). Because a chip at 85 °C draws slightly less than its power at T_{ref} , δ is a fraction of a kelvin (Table S26), and it becomes negative once the chip limit exceeds 87.5 °C. For present limits $D_{rating} = 10$ K and $D_{stack} = 0$, so the chip binds with an HBM margin of 9.55 K. The binding switches to the HBM when D_{stack} exceeds about $D_{rating} - 0.45$ K: with an HBM limit of 95 °C, ΔT_{hbm} must exceed ΔT_{jc} by about 9.5 K. This explains the three HBM-bound rows of Table S4 ($\Delta T_{hbm} = 50$ K, an HBM limit of 85 °C, or $\Delta T_{jc} = 15$ K).

Table S26. Margin δ of Eq. (S17) (K) at a chip limit of 85 °C (flat / median / steep shape), and its dependence on the chip limit (nominal case).

f_L	δ at 85 °C	T_{cap} (°C)	δ ($f_L = 0.15$, median)
0.10	0.248 / 0.300 / 0.385	75	1.980
0.15	0.372 / 0.451 / 0.578	85	0.451
0.22	0.545 / 0.661 / 0.848	95	-1.540
		105	-4.094
		150	-26.106

The condition was checked against the optimizer for 56 combinations of $D_{rating} = 0, 5, 10$ and 20 K and D_{stack} from -10 to +20 K ($\Delta T_{jc} = 40$ K, nominal case, $f_L = 0.15$, median shape); the predicted binding limit agreed in all 56. When the

HBM binds, the coolant drops 1 K for every kelvin by which D_{stack} exceeds D_{rating} ; with equal limits of 85 °C and equal differences, for example, the optimum moves to a coolant outlet temperature of 45.00 °C with the hot chip at 84.45 °C (Section S9.11).

S9.3. Break-even relation: sampling, fits and validity

Sampling. The 126 settings behind Eq. (7) comprise the nominal radiator, 25 one-at-a-time settings around it (σ_{rad} 2–17.5 kg m⁻², η_{fin} 0.61–0.95, SP_{pv} 50–300 W kg⁻¹, q_{env} $\times 0.5$ –2, m_{bat} 0–6 kg kW⁻¹) and a Latin hypercube of 100 settings over the same ranges (SP_{pv} and q_{env} log-uniform). The flat and steep shapes use the first 30 hypercube settings. For each setting the four thresholds are found by root finding on f_L and converted to s at the hot chip of that setting's present optimum. The sample spans $\phi_{\text{eff}} = 0.239$ –1.103 for the median shape; the advanced, nominal, heavy and pessimistic radiators lie at 0.406, 0.570, 0.715 and 0.925 (Table S27).

Table S27. Fits of the thresholds to ϕ_{eff} (s in % K⁻¹; coefficients from the highest order down).

Shape (settings)	Threshold	Fit in s	R^2	Residual SD / max	Same fit in f_L	R^2
Median (126)	zero, linear	$0.964 \phi_{\text{eff}} + 0.220$	0.976	0.027 / 0.072	$0.319 \phi_{\text{eff}} + 0.078$	0.974
Median (126)	zero, quadratic	$-0.530 \phi_{\text{eff}}^2 + 1.622 \phi_{\text{eff}} + 0.032$	0.991	0.016 / 0.046	$-0.185 \phi_{\text{eff}}^2 + 0.548 \phi_{\text{eff}} + 0.012$	0.991
Median (126)	2%, linear	$0.896 \phi_{\text{eff}} - 0.048$	0.980	0.022 / 0.060	$0.301 \phi_{\text{eff}} - 0.014$	0.979
Median (126)	5%, linear	$0.803 \phi_{\text{eff}} - 0.140$	0.972	0.024 / 0.077	$0.272 \phi_{\text{eff}} - 0.046$	0.971
Median (122)	10%, linear	$0.655 \phi_{\text{eff}} - 0.184$	0.948	0.026 / 0.091	$0.223 \phi_{\text{eff}} - 0.062$	0.947
Flat (30)	zero, linear	$0.932 \phi_{\text{eff}} + 0.205$	0.979	0.029 / 0.067	$0.371 \phi_{\text{eff}} + 0.088$	0.977
Steep (30)	zero, linear	$1.035 \phi_{\text{eff}} + 0.216$	0.984	0.027 / 0.070	$0.270 \phi_{\text{eff}} + 0.060$	0.984
Pooled (186)	zero, linear	$0.972 \phi_{\text{eff}} + 0.216$	0.965	0.035 / 0.122	$0.319 \phi_{\text{eff}} + 0.076$	0.821

Within one shape the two forms fit equally well, but only the form in s survives pooling the three shapes. On the same 30 hypercube settings, the break-even s_0 differs between the three shapes by 9.5% on average (at most 12.3%), and $f_{L,0}$ by 32.7% (at most 36.5%). The quadratic fit shows that the relation is slightly concave, so the linear form is high at both ends of the sampled range. For four of the 126 median-shape settings the 10% threshold has no root, so the 10% fit uses 122 settings.

Validity. Table S28 compares the break-even value found by root finding with the linear relation of Eq. (7) for settings outside the sample and for structural variants of the model. Radiator and power-system settings within the sampled range deviate by at most 8%, mostly by less than 3%. Outside it, at $\phi_{\text{eff}} = 0.17$ and 3.0, the relation overestimates s_0 by about 40%. Structural changes matter more: without pumping power the break-even value is 15% lower, with a two-phase loop 21% lower and with other cold-plate arrangements 17–19% lower, and with a smaller temperature stack 8–18% higher. The relation therefore holds for the single-phase, series-loop design with 40 K temperature differences that it was fitted to; within that design it is insensitive to the leakage shape ($b = 2850$ K: +0.8%), the HBM convention and the design margin (+3.9 to +5.4%). Of the 21 GPU estimates of Table S10, only one exceeds the nominal break-even value.

Table S28. Break-even values from root finding (“direct”) and from the linear relation of Eq. (7) (“rule”) for settings outside the sample and for structural variants (median shape). Deviation = direct/rule – 1.

Variant	ϕ_{eff}	In sampled range	s_0 direct (% K ⁻¹)	s_0 rule (% K ⁻¹)	Deviation (%)	$f_{L,0}$ direct
$\sigma_{\text{rad}} = 1 \text{ kg m}^{-2}$	0.169	no	0.223	0.383	-41.7	0.076
$\sigma_{\text{rad}} = 30 \text{ kg m}^{-2}$, $\eta_{\text{fin}} = 0.5$, $q_{\text{env}} \times 2.5$	3.019	no	1.966	3.131	-37.2	0.644
$q_{\text{env}} \times 3$	1.213	no	1.253	1.389	-9.8	0.418
$\sigma_{\text{rad}} = 25 \text{ kg m}^{-2}$, $\eta_{\text{fin}} = 0.55$	0.958	yes	1.113	1.143	-2.6	0.372
$SP_{\text{pv}} = 500 \text{ W kg}^{-1}$, $\sigma_{\text{rad}} = 17.5 \text{ kg m}^{-2}$	0.878	yes	1.038	1.066	-2.6	0.348
$SP_{\text{pv}} = 30 \text{ W kg}^{-1}$	0.323	yes	0.489	0.531	-7.9	0.166
$m_{\text{bat}} = 12 \text{ kg kW}^{-1}$	0.434	yes	0.640	0.639	+0.2	0.217
Heavy radiator	0.715	yes	0.898	0.909	-1.2	0.302
Pessimistic radiator	0.925	yes	1.089	1.112	-2.1	0.364
Nominal, $\eta_{\text{fin}} = 0.65$	0.697	yes	0.882	0.892	-1.1	0.297
Nominal, $q_{\text{env}} \times 1.34$ (end of life)	0.636	yes	0.871	0.833	+4.6	0.293
Nominal, $\Delta p_{\text{ref}} = 400 \text{ kPa}$	0.578	yes	0.800	0.777	+2.9	0.270
Nominal, $b = 2850 \text{ K}$	0.571	yes	0.776	0.770	+0.8	0.293
Nominal, HBM held at 95 °C or capped at 105 °C	0.570	yes	0.800	0.770	+3.9	0.270
Nominal, HBM limit equal to chip limit (Section S9.11)	0.574	yes	0.809	0.773	+4.7	0.279
Nominal, 10 K design margin	0.636	yes	0.878	0.833	+5.4	0.350
Nominal, no pumping power	0.542	yes	0.629	0.742	-15.2	0.213
Heavy, no pumping power	0.684	yes	0.751	0.879	-14.6	0.253
Nominal, two-phase loop (ϕ_{eff} of the single-phase design)	0.570	yes	0.606	0.770	-21.3	0.205
Heavy, two-phase loop	0.715	yes	0.724	0.909	-20.3	0.244
Nominal, all chips at the outlet	0.570	yes	0.636	0.770	-17.3	0.215
Nominal, parallel cold plates (mid-loop)	0.539	yes	0.609	0.740	-17.7	0.206
Nominal, all chips at the supply, $\Delta T_{\text{loop}} \leq 20 \text{ K}$	0.476	yes	0.553	0.679	-18.6	0.187
Nominal, $\Delta T_{\text{jc}} = \Delta T_{\text{hbm}} = 25 \text{ K}$	0.481	yes	0.740	0.683	+8.3	0.250
Nominal, $\Delta T_{\text{jc}} = \Delta T_{\text{hbm}} = 15 \text{ K}$	0.427	yes	0.699	0.632	+10.7	0.236
Heavy, $\Delta T_{\text{jc}} = \Delta T_{\text{hbm}} = 15 \text{ K}$	0.554	yes	0.890	0.754	+17.9	0.299

S9.4. Cold-plate arrangements

In the main model the chips are distributed along a series loop and the hottest chip and the HBM sit at the loop outlet. Three other arrangements are examined: every chip and the HBM at the outlet temperature; every chip and the HBM at the loop mid-point, as for parallel cold plates whose ΔT_{jc} is referred to the mean coolant temperature; and every chip and the HBM at the supply temperature, with the loop rise limited to 20 K. These variants keep the grid resolution of the outlet temperature (Section S1.4) (Table S29).

Table S29. Cold-plate arrangements ($f_L = 0.15$; saving from higher ratings for the flat / median / steep shape, other values for the median shape).

Radiator	Arrangement	M at 85 °C (kg kW _{ref} ⁻¹)	Coolant outlet (°C); ΔT_{loop} (K)	Binding	Saving 85 → 150 °C (%)	HBM held at 95 °C (%)	$f_{L,0}$	s_0 (% K ⁻¹)	$\Delta T_{40} \rightarrow$ 25 K, both (%)	ΔT_{jc} only (%)
Nominal	series loop (main)	44.680	45.45; 10	chip	4.55 / 2.59 / 1.05	2.12	0.270	0.800	7.97	6.71
Nominal	all at outlet	45.445	45.45; 10	chip	3.32 / 1.41 / 0.19	1.39	0.215	0.636	8.08	7.06
Nominal	all at mid- loop (parallel)	43.885	55.35; 20	chip	2.71 / 1.08 / 0.11	1.08	0.206	0.609	7.54	6.66
Nominal	all at supply, $\Delta T_{loop} \leq 20$ K	41.538	65.45; 20	chip	1.66 / 0.51 / 0.00	0.51	0.187	0.553	6.57	5.98
Advanced	series loop (main)	35.601	45.45; 12	chip	1.24 / 0.55 / 0.14	0.55	0.204	0.603	5.04	4.59
Advanced	all at outlet	36.191	45.45; 10	chip	0.60 / 0.09 / 0.00	0.09	0.162	0.478	5.07	4.89
Advanced	all at mid- loop (parallel)	35.346	55.35; 20	chip	0.38 / 0.03 / 0.00	0.03	0.154	0.452	4.70	4.61
Advanced	all at supply, $\Delta T_{loop} \leq 20$ K	34.171	64.10; 20	interior	0.09 / 0.00 / 0.00	0.00	0.135	0.397	4.10	4.10
Heavy	series loop (main)	59.704	45.45; 10	chip	9.18 / 5.58 / 2.57	3.79	0.302	0.898	10.97	8.89
Heavy	all at outlet	60.749	45.45; 8	chip	7.33 / 3.69 / 1.01	3.05	0.252	0.747	11.01	9.17
Heavy	all at mid- loop (parallel)	58.065	55.35; 20	chip	6.51 / 3.26 / 0.87	2.76	0.247	0.728	10.49	8.79
Heavy	all at supply, $\Delta T_{loop} \leq 20$ K	53.735	65.45; 20	chip	4.78 / 2.16 / 0.40	1.99	0.228	0.674	9.34	7.97

With the nominal radiator the saving from higher ratings is 0.51–2.59%, the break-even value 0.553–0.800% K⁻¹ and the stack lever 6.57–8.08% across the four arrangements; the stack lever is larger than higher ratings in all twelve combinations of arrangement and radiator. The main arrangement gives the largest saving from higher ratings because it pins the hottest chip at the radiator inlet, so that a large loop rise makes the radiator colder (Section S9.5). In the three other arrangements all chips share one temperature and s_0 hardly depends on the leakage shape (spread 0.8–0.9% against 7.8% for the main arrangement; nominal $s_0 = 0.639 / 0.636 / 0.634, 0.611 / 0.609 / 0.606$ and $0.555 / 0.553 / 0.550\%$ K⁻¹ for the flat / median / steep shapes).

S9.5. Loop-rise penalty and two-phase loop

The decomposition of Table S2 shows that, with the nominal radiator, 1.05 of the 2.59 percentage points saved by higher ratings come from a smaller loop-rise penalty and 0.51 from lower pumping power, leaving 1.03 from radiator area. A two-phase loop, which is isothermal at the outlet temperature and has a fixed pumping power of 0.3% of the rejected heat, has neither penalty: its mass at present limits is 43.840 kg kW_{ref}⁻¹ (nominal), 35.340 (advanced) and 57.954 (heavy), and raising the limits to 150 °C saves 2.64 / 1.03 / 0.09%, 0.36 / 0.02 / 0.00% and 6.41 / 3.16 / 0.79% (Table S3). Its break-even value is 0.606% K⁻¹ for the nominal radiator, 21% below the single-phase value, and the stack lever

from 40 to 25 K still saves 7.50% (Table S30). Removing pumping power from the single-phase model has a similar effect (1.24% saving, $s_0 = 0.629\% \text{ K}^{-1}$).

S9.6. Temperature-stack lever

The values are listed in Table S30.

Table S30. Saving from reducing the temperature differences at present limits (chip 85 °C, HBM 95 °C) relative to 40 K ($f_L = 0.15$; flat / median / steep shape), the saving from subsequently raising the limits to 150 °C, and the total relative to (40 K, 85 °C); coolant and hot-chip temperatures for the median shape. The savings are plotted in Fig. S4.

Variant	Radiator	$\Delta T_{jc} / \Delta T_{hbm}$ (K)	Saving vs 40 K (%)	Coolant outlet (°C)	Hot-chip junction (°C); binding	Then 85 → 150 °C (%)	Total (%)
Both lowered	Advanced	25 / 25	5.05 / 5.04 / 5.02	60.28	85.0; chip	0.61 / 0.20 / 0.01	5.63 / 5.23 / 5.03
	Advanced	15 / 15	7.72 / 7.69 / 7.64	70.17	85.0; chip	0.33 / 0.07 / 0.00	8.02 / 7.76 / 7.64
	Nominal	30 / 30	5.59 / 5.56 / 5.55	55.34	85.0; chip	3.61 / 2.02 / 0.77	8.99 / 7.47 / 6.28
	Nominal	25 / 25	7.98 / 7.97 / 7.93	60.28	85.0; chip	3.19 / 1.75 / 0.65	10.92 / 9.58 / 8.53
	Nominal	20 / 20	10.18 / 10.15 / 10.09	65.23	85.0; chip	2.80 / 1.50 / 0.54	12.70 / 11.51 / 10.58
	Nominal	15 / 15	12.19 / 12.15 / 12.06	70.17	85.0; chip	2.44 / 1.28 / 0.45	14.33 / 13.27 / 12.46
	Heavy	25 / 25	11.03 / 10.97 / 10.87	60.28	85.0; chip	7.42 / 4.53 / 2.12	17.63 / 15.01 / 12.76
	Heavy	15 / 15	16.77 / 16.68 / 16.54	70.17	85.0; chip	6.35 / 3.89 / 1.83	22.06 / 19.92 / 18.06
ΔT_{jc} only	Advanced	25 / 40	4.47 / 4.59 / 4.73	55.00	79.1; HBM	1.22 / 0.68 / 0.31	5.63 / 5.23 / 5.03
	Nominal	25 / 40	6.58 / 6.71 / 6.88	55.00	79.1; HBM	4.65 / 3.07 / 1.77	10.92 / 9.58 / 8.53
	Nominal	15 / 40	8.30 / 8.59 / 8.91	55.00	69.0; HBM	6.58 / 5.12 / 3.89	14.33 / 13.27 / 12.46
	Heavy	25 / 40	8.76 / 8.89 / 9.05	55.00	79.1; HBM	9.73 / 6.72 / 4.08	17.63 / 15.01 / 12.76
HBM not below 30 K	Nominal	25 / 30	7.98 / 7.97 / 7.93	60.28	85.0; chip	3.19 / 1.75 / 0.65	10.92 / 9.58 / 8.53
	Nominal	15 / 30	11.10 / 11.21 / 11.34	65.00	79.5; HBM	3.64 / 2.32 / 1.26	14.33 / 13.27 / 12.46
HBM not below 35 K	Nominal	25 / 35	7.93 / 7.91 / 7.88	60.00	84.7; HBM	3.25 / 1.81 / 0.71	10.92 / 9.58 / 8.53
	Nominal	15 / 35	9.83 / 10.04 / 10.28	60.00	74.2; HBM	4.99 / 3.59 / 2.42	14.33 / 13.27 / 12.46
Two-phase loop, both lowered	Advanced	25 / 25	4.66 / 4.67 / 4.82	58.60	83.1; –	0.07 / 0.00 / 0.00	4.73 / 4.67 / 4.82
	Nominal	25 / 25	7.51 / 7.50 / 7.48	60.28	85.0; –	1.70 / 0.58 / 0.02	9.08 / 8.04 / 7.49
	Heavy	25 / 25	10.47 / 10.45 / 10.42	60.28	85.0; –	5.10 / 2.51 / 0.62	15.03 / 12.69 / 10.97

The two-phase rows are relative to the two-phase loop at 40 K.

Independence of leakage. With equal differences the coolant rises by about 1 K for every kelvin removed from the stack while the hot chip stays at its 85 °C limit, so leakage does not grow. Across $f_L = 0.10$ – 0.22 and the three shapes, a reduction to 25 K saves 5.01–5.12% (advanced), 7.70–8.06% (nominal) and 10.61–11.12% (heavy). For a chip without leakage, reducing both differences from 40 to 25 K and raising both limits by 15 K give identical savings for every

radiator (5.12%, 8.16% and 11.13%), so within the model the stack lever is equivalent to a 15 K higher rating for a chip whose leakage does not grow.

Chip side and memory side. Because the chip binds at present and the HBM has 9.55 K of margin, most of the lever is on the chip side. Lowering ΔT_{jc} alone lets the coolant rise until the HBM binds at 55 °C and saves 6.71% at 25 K; holding ΔT_{hbm} at 30 or 35 K while ΔT_{jc} falls to 25 K gives 7.97% and 7.91%, against 7.97% when both fall. Below 25 K the memory side starts to matter: at $\Delta T_{jc} = 15$ K the three variants give 11.21%, 10.04% and 8.59% against 12.15% (Table S31).

Table S31. Value of the temperature stack and the leakage fraction above which it exceeds the saving from higher ratings (fixed clock; flat / median / steep shape).

Radiator	Value per kelvin at 40 K (kg kW _{ref} ⁻¹ K ⁻¹)	Mean value per kelvin, 40 → 25 K	Crossover f_L	Stack lever at the crossover (%, median)
Advanced	0.140 / 0.138 / 0.136	0.120 / 0.120 / 0.119	0.076 / 0.058 / 0.040	5.10
Nominal	0.277 / 0.275 / 0.272	0.238 / 0.237 / 0.236	0.107 / 0.081 / 0.055	8.06
Heavy	0.512 / 0.508 / 0.502	0.440 / 0.437 / 0.432	0.133 / 0.100 / 0.068	11.09

The crossover f_L is the leakage fraction at which raising the limits to 150 °C saves as much as reducing both differences from 40 to 25 K. The center of the present evidence, 0.142–0.173, lies above the crossover of every reference radiator and shape, so at fixed clock the stack lever is the larger one. After the stack has been reduced to 25 K, raising the limits to 150 °C adds 1.75% (nominal), 0.20% (advanced) and 4.53% (heavy). The value per kelvin varies little with f_L : 0.231–0.240 kg kW_{ref}⁻¹ K⁻¹ for the mean from 40 to 25 K with the nominal radiator over $f_L = 0.10$ –0.22. The cold-plate mass and pressure drop needed to achieve a smaller stack are not modeled (Table S32).

Table S32. Levers at the clock-scaled optimum (median shape; savings relative to the clock-scaled design with 40 K differences and present limits, %).

Radiator	f_L	M (kg kW _{ref} ⁻¹)	Both 40 → 25 K	ΔT_{jc} only	ΔT_{jc} 25 K, ΔT_{hbm} 30 K	ΔT_{hbm} only	Limits to 150 °C	Chip limit alone
Advanced	0.10	28.733	2.13	0.70	1.80	0.98	1.13	0.00
Advanced	0.15	28.677	1.64	1.05	1.62	0.26	0.26	0.00
Nominal	0.10	33.503	4.18	0.84	3.28	2.51	3.87	0.00
Nominal	0.15	33.451	3.55	1.27	3.05	1.49	1.71	0.00
Nominal	0.22	33.319	2.67	1.82	2.63	0.29	0.29	0.00
Heavy	0.10	41.388	6.64	0.98	5.06	4.37	8.08	0.00
Heavy	0.15	41.313	5.84	1.49	4.72	3.09	4.32	0.00
Heavy	0.22	41.176	4.74	2.19	4.26	1.39	1.44	0.00

Under clock scaling the chips run at about 72 °C and the HBM binds, so the memory side of the stack gains weight. The stack lever remains the larger one at $f_L = 0.15$ for all three radiators, whereas at $f_L = 0.10$ higher ratings are the larger lever with the heavy radiator (8.08% against 6.64%) and nearly equal with the nominal one (3.87% against 4.18%).

S9.7. Design margin, fin efficiency and end-of-life load

Four choices in the nominal design are varied: a design margin that lowers both present limits by 10 K (chip 75 °C, HBM 85 °C, raised to 140 °C); a constant fin efficiency between 0.61, the value implied by the GRC test, and the nominal 0.85; an end-of-life environmental load 1.3395 times the nominal one (Section S9.12); and a fin efficiency that declines with temperature at fixed fin geometry. In the last variant the fin parameter grows as $(mL)^2 \propto \varepsilon \sigma T^3$, and $\eta_{fin} = \tanh(mL)/(mL)$ is calibrated to each radiator's fin efficiency at the mean coolant temperature of its present optimum (39.5, 40.5 and 40.5 °C for the advanced, nominal and heavy radiators). The fin efficiency then falls from

0.865 at 300 K to 0.791 at 360 K for the advanced radiator, from 0.866 to 0.792 for the nominal and from 0.837 to 0.753 for the heavy one (Table S33).

Table S33. Effect of design margin, fin efficiency and environmental load on the saving from higher ratings and on the thresholds (flat / median / steep shape where three values are given).

Variant	Radiator	M at present limits (kg kW _{ref} ⁻¹)	ϕ_{eff}	Saving, $f_L = 0.10 / 0.15 / 0.22$ (% median)	Saving, $f_L = 0.15$ (%)	$f_{L,5}$	$f_{L,0}$	s_0 (% K ⁻¹)
Reference	Nominal	44.680	0.570	6.07 / 2.59 / 0.38	4.55 / 2.59 / 1.05	0.143 / 0.112 / 0.081	0.310 / 0.270 / 0.219	0.763 / 0.800 / 0.825
10 K margin	Advanced	36.160	0.455	4.53 / 2.09 / 0.40	3.08 / 2.09 / 1.27	0.114 / 0.093 / 0.072	0.303 / 0.277 / 0.255	0.646 / 0.676 / 0.725
10 K margin	Nominal	46.161	0.636	10.16 / 5.71 / 2.09	8.03 / 5.71 / 3.67	0.193 / 0.161 / 0.128	0.383 / 0.350 / 0.304	0.837 / 0.878 / 0.885
10 K margin	Heavy	62.835	0.783	16.64 / 10.29 / 4.61	14.17 / 10.29 / 6.78	0.254 / 0.214 / 0.174	0.425 / 0.384 / 0.350	0.941 / 0.975 / 1.042
$\eta_{\text{fin}} = 0.61$	Nominal	54.153	0.732	10.62 / 5.47 / 1.58	8.80 / 5.47 / 2.60	0.197 / 0.156 / 0.115	0.361 / 0.307 / 0.258	0.892 / 0.914 / 0.976
$\eta_{\text{fin}} = 0.65$	Nominal	51.940	0.697	9.59 / 4.78 / 1.27	7.81 / 4.78 / 2.20	0.186 / 0.147 / 0.108	0.359 / 0.297 / 0.249	0.886 / 0.882 / 0.941
$\eta_{\text{fin}} = 0.75$	Nominal	47.703	0.626	7.55 / 3.48 / 0.72	5.90 / 3.48 / 1.50	0.162 / 0.128 / 0.093	0.332 / 0.289 / 0.230	0.818 / 0.859 / 0.866
End-of-life load	Advanced	36.343	0.459	2.93 / 0.95 / 0.03	1.89 / 0.95 / 0.28	0.092 / 0.070 / 0.049	0.260 / 0.226 / 0.191	0.639 / 0.668 / 0.718
End-of-life load	Nominal	46.125	0.636	7.45 / 3.49 / 0.76	5.84 / 3.49 / 1.53	0.162 / 0.128 / 0.093	0.337 / 0.293 / 0.233	0.830 / 0.871 / 0.877
End-of-life load	Heavy	62.483	0.793	13.09 / 7.06 / 2.29	11.17 / 7.06 / 3.48	0.219 / 0.174 / 0.130	0.376 / 0.325 / 0.273	0.928 / 0.968 / 1.035
$\eta_{\text{fin}}(T)$	Advanced	—	—	1.57 / 0.31 / 0.00	0.78 / 0.31 / 0.04	0.065 / 0.049 / 0.034	0.214 / 0.186 / 0.157	0.524 / 0.547 / 0.587
$\eta_{\text{fin}}(T)$	Nominal	—	—	4.73 / 1.79 / 0.16	3.34 / 1.79 / 0.62	0.124 / 0.097 / 0.069	0.284 / 0.247 / 0.209	0.698 / 0.731 / 0.787
$\eta_{\text{fin}}(T)$	Heavy	—	—	8.43 / 3.83 / 0.76	6.61 / 3.83 / 1.62	0.170 / 0.134 / 0.098	0.330 / 0.287 / 0.229	0.812 / 0.852 / 0.862

The first three variants raise the radiator share ϕ_{eff} and with it the saving from higher ratings: a 10 K margin raises the saving with the nominal radiator from 2.59% to 5.71% and raises the 5% threshold by about 0.05 in f_L , while s_0 rises by only about 10%; a fin efficiency of 0.61 raises the saving to 5.47% and the end-of-life load to 3.49%. In this respect the nominal design lies on the side that disfavors running hot. A fin efficiency that declines with temperature works the other way, because a hotter radiator loses part of its emitting effectiveness: the saving with the nominal radiator falls to 1.79% and s_0 by 9%.

S9.8. Attribution of the saving to chip and HBM limits

Table S34 attributes the saving to the chip and HBM limits.

Table S34. Saving from raising the chip limit alone (HBM held at 95 °C), the HBM limit alone (chip held at 85 °C) or both to 150 °C, at fixed clock and at the clock-scaled optimum (flat / median / steep shape).

Radiator	f_L	Chip only, fixed clock (%)	Both, fixed clock (%)	HBM only, clock-scaled (%)	Both, clock-scaled (%)
Advanced	0.10	1.98 / 1.61 / 1.04	3.45 / 2.18 / 1.09	1.16 / 0.98 / 0.77	1.60 / 1.13 / 0.78
Advanced	0.15	1.06 / 0.55 / 0.14	1.24 / 0.55 / 0.14	0.46 / 0.26 / 0.15	0.46 / 0.26 / 0.15
Nominal	0.10	3.82 / 3.39 / 2.70	8.83 / 6.07 / 3.45	2.74 / 2.51 / 2.23	5.13 / 3.87 / 2.78
Nominal	0.15	2.77 / 2.12 / 1.05	4.55 / 2.59 / 1.05	1.85 / 1.49 / 1.10	2.52 / 1.71 / 1.11
Nominal	0.22	1.22 / 0.38 / 0.01	1.30 / 0.38 / 0.01	0.62 / 0.29 / 0.13	0.62 / 0.29 / 0.13
Heavy	0.10	5.80 / 5.31 / 4.50	15.68 / 11.09 / 6.58	4.65 / 4.37 / 4.01	10.50 / 8.08 / 5.87
Heavy	0.15	4.60 / 3.79 / 2.48	9.18 / 5.58 / 2.57	3.51 / 3.09 / 2.58	6.12 / 4.32 / 2.89
Heavy	0.22	2.77 / 1.51 / 0.31	3.62 / 1.53 / 0.31	2.00 / 1.39 / 0.81	2.44 / 1.44 / 0.81

At fixed clock the chip binds at present, so raising the HBM limit alone saves nothing in every case, and raising the chip limit alone captures most of the saving at present leakage (82% nominal, 68% heavy at $f_L = 0.15$) until the HBM binds at a coolant outlet temperature of 55 °C. The lower the leakage, the more the HBM rating matters (Table S24). At the clock-scaled optimum the HBM binds instead, and raising the chip limit alone saves nothing.

S9.9. System-level stability

Loop gains. At chip level the fixed point of Eq. (2) exists while

$$g_i = \Delta T_{jc} \frac{dP}{dT_j} \Big|_{T_{j,i}} < 1. \quad (S18)$$

For a built design (radiator area, coolant flow and pumping power fixed, junctions quasi-static), a heat-load change dQ raises the radiator outlet temperature by $R dQ$, where $R = \partial T_{out} / \partial Q$ at fixed A and $\dot{m}c_p$, and widens the loop rise by $dQ / (\dot{m}c_p)$. The local coolant temperature of a chip at relative distance x_i upstream of the outlet ($x_i = 0$ at the outlet, 1 at the inlet) therefore changes by $(R - x_i / \dot{m}c_p) dQ$, and the system loop gain is

$$L = \frac{1}{41} \sum_{i=1}^{41} \frac{1 + f_{other}}{P(T_{80})} \frac{1}{1 - g_i} \frac{dP}{dT_j} \Big|_{T_{j,i}} \left(R - \frac{x_i}{\dot{m}c_p} \right), \quad (S19)$$

with stability for $L < 1$. This is the loop gain of the same loop integration as the mass model. A conservative bound places all chips at the outlet, $G_{hot} = g_{hot} + \{(1 + f_{other}) / P(T_{80})\} (dP/dT_j)_{hot} R$.

At the 54 mass-optimal designs (three radiators, chip limits of 85 and 150 °C, three leakage fractions and three shapes) $L = 0.153$ – 0.706 , 1.49 – 1.93 times the chip-level gain of the hot chip, so all are stable. The bound G_{hot} , which also describes cold plates fed in parallel at the outlet temperature, stays at or below 0.94 at present limits but reaches 1.03 for the nominal radiator and 1.10 for the heavy radiator at 150 °C (1.008 for the heavy radiator with $f_L = 0.15$ and the median shape) and 1.21 for the pessimistic radiator. In such an arrangement the heavy-radiator optimum at 150 °C would have no stable operating point at nominal clock (Table S35).

Table S35. Chip-level tangent points T_j^* / T_c^* (°C) for $\Delta T_{jc} = 40$ K, and the maximum coolant outlet temperature with a stable operating point along the mass-optimal designs without chip and HBM limits (°C; chip level / system level, $L = 1$ / all chips at the outlet, $G_{hot} = 1$).

f_L	Shape	T_j^* / T_c^*	Advanced	Nominal	Heavy
0.10	flat	246.7 / 136.1	136.1 / 125.9 / 97.9	136.1 / 125.9 / 102.4	136.1 / 126.4 / 104.4
0.10	median	190.2 / 103.6	103.6 / 99.9 / 80.9	103.6 / 99.9 / 82.4	103.6 / 96.9 / 83.4
0.10	steep	148.2 / 79.1	79.1 / 77.9 / 64.9	79.1 / 76.9 / 65.9	79.1 / 74.9 / 66.4
0.15	flat	209.4 / 109.8	109.8 / 102.4 / 78.4	109.8 / 102.4 / 81.4	109.8 / 100.9 / 82.9
0.15	median	166.0 / 86.0	86.0 / 82.9 / 65.9	86.0 / 81.9 / 66.9	86.0 / 79.4 / 67.9
0.15	steep	133.0 / 68.1	68.1 / 66.9 / 54.9	68.1 / 65.4 / 55.4	68.1 / 64.4 / 55.9
0.22	flat	178.5 / 88.8	88.8 / 82.9 / 61.9	88.8 / 81.4 / 63.9	88.8 / 79.9 / 65.4
0.22	median	145.3 / 71.9	71.9 / 69.4 / 52.9	71.9 / 66.9 / 54.4	71.9 / 65.9 / 55.4
0.22	steep	119.7 / 59.4	59.4 / 57.4 / 46.4	59.4 / 56.9 / 46.9	59.4 / 55.9 / 47.4

The system-level limits lie 1.2–10.3 K below the chip-level ones, and the all-at-outlet bound a further 8.5–28.0 K lower (Table S36).

Table S36. Closed-mapping designs for the literature radiating temperatures ($\Delta T_{jc} = 40$ K): stability and the largest ΔT_{jc} with a stable operating point (K; chip level / system level; flat / median / steep shape).

Design	Coolant (°C)	$f_L = 0.10$	$f_L = 0.15$	$f_L = 0.22$
Nominal, 350 K	96.4	56.2 / 43.0 / 30.5; 55.4 / 41.7 / 27.7	45.2 / 34.7 / 24.6; 43.1 / 31.9 / 21.1	36.4 / 28.0 / 19.8; 32.1 / 22.8 / 13.8
Nominal, 385 K	133.9	40.4 / 27.5 / 15.8; 36.0 / 21.2 / 6.7	31.0 / 20.9 / 11.7; 23.9 / 11.4 / 0.0	23.6 / 15.8 / 8.7; 13.0 / 2.1 / 0.0
Heavy, 350 K	99.8	54.6 / 41.4 / 28.8; 52.9 / 39.0 / 25.6	43.8 / 33.2 / 23.1; 40.4 / 29.0 / 17.9	35.0 / 26.6 / 18.4; 29.2 / 19.6 / 10.2
Heavy, 385 K	136.6	39.4 / 26.6 / 15.0; 34.8 / 19.9 / 5.3	30.1 / 20.1 / 11.1; 22.7 / 10.1 / 0.0	22.9 / 15.1 / 8.2; 11.8 / 0.9 / 0.0

With $\Delta T_{jc} = 40$ K the nominal 350 K design is stable at $f_L = 0.10$ for the flat and median shapes ($L = 0.517$ and 0.907) and at $f_L = 0.15$ only for the flat shape (junction 158.2 °C, $L = 0.794$). The nominal 385 K design has a chip-level operating point only at $f_L = 0.10$ with the flat shape (junction 225.2 °C), where the system gain is 1.52. The heavy 350 K design is stable at system level only for the flat shape at $f_L = 0.10$ and 0.15 (junction 152.5 and 165.7 °C); with $f_L = 0.10$ and the median shape it has a chip-level operating point at 167.5 °C but a system gain of 1.12. The advanced radiator gives nearly the same results as the nominal one (closed-mapping coolant 97.5 °C for 350 K). In the perfect-radiator mapping, which gives the lowest coolant temperatures (76.9 °C for 350 K and 111.9 °C for 385 K), the junction at $f_L = 0.15$ is 125.3 and 131.2 °C for the flat and median shapes at 350 K, with no stable operating point for the steep shape, and no shape has a stable operating point at 385 K. For comparison, the chip-level gain of the hot chip at the present optimum is 0.174 (median shape).

Critical leakage shape. For a given leakage fraction, a closed design is stable only if leakage grows more slowly with temperature than a critical b (Table S37). With $f_L = 0.15$ the nominal 350 K design requires b below 2905 K at chip level and 2753 K at system level. The common fit to the V100 telemetry, $b = 2850$ K, lies above the system-level limit at $f_L = 0.15$ and below both limits at 0.13. For 385 K the system-level limit is 1978 K, flatter than every fitted leakage curve (Table S5).

Table S37. Critical leakage activation temperature b (K) below which the closed designs for the literature radiating temperatures remain stable ($\Delta T_{jc} = 40$ K; chip level / system level).

Design	Coolant (°C)	$f_L = 0.10$	0.13	0.15	0.20
Nominal, 350 K	96.4	3544 / 3433	3126 / 2987	2905 / 2753	2485 / 2302
Nominal, 385 K	133.9	2603 / 2437	2310 / 2136	2155 / 1978	1858 / 1675
Heavy, 350 K	99.8	3419 / 3251	3017 / 2827	2808 / 2604	2403 / 2181
Heavy, 385 K	136.6	2558 / 2394	2272 / 2099	2120 / 1944	1827 / 1647

S9.10. Pessimistic radiator

The pessimistic radiator combines the 17.5 kg m^{-2} areal mass of the ISS photovoltaic radiator with a fin efficiency of 0.65 and the end-of-life environmental load of 170.5 W m^{-2} . With these parameters the model gives 191.7 W m^{-2} for the ISS radiator against its rated 329.9 W m^{-2} (Section S4), so the case lies well below flight hardware and is used only to show how the results scale with a still heavier effective radiator (Table S38).

Table S38. Results for the pessimistic radiator (flat / median / steep shape where three values are given; $f_L = 0.15$ unless stated).

Quantity	Value
M at present limits ($\text{kg kW}_{\text{ref}}^{-1}$); binding; coolant (°C); loop rise (K)	76.006 (75.859–76.088); chip; 45.45; 8
ϕ_{rad} ; ϕ_{eff} ; T_{eff} (K)	0.781; 0.925; 282.4
Saving 85 → 150 °C, $f_L = 0 / 0.10 / 0.15 / 0.22$ (%; median)	40.32 / 17.76 / 10.48 / 4.25
Saving 85 → 150 °C, $f_L = 0.15$ (%)	15.82 / 10.48 / 5.64
HBM held at 95 °C; HBM capped at 105 °C (%)	7.09 / 6.29 / 4.85; 12.07 / 9.82 / 5.64
Decomposition, radiator / pumping power / loop rise (%; median)	6.98 / 1.40 / 2.10
$f_{L,10}$; $f_{L,5}$; $f_{L,2}$; $f_{L,0}$	0.200 / 0.154 / 0.109; 0.260 / 0.209 / 0.158; 0.318 / 0.263 / 0.206; 0.421 / 0.364 / 0.307
s_{10} ; s_5 ; s_2 ; s_0 (% K^{-1})	0.488 / 0.453 / 0.408; 0.638 / 0.617 / 0.590; 0.783 / 0.780 / 0.776; 1.042 / 1.089 / 1.168
Stack lever, ΔT 40 → 25 K; 40 → 15 K (%)	14.70 / 14.68 / 14.63; 22.00 / 21.95 / 21.85
Value per kelvin at 40 K ($\text{kg kW}_{\text{ref}}^{-1} \text{K}^{-1}$, median); crossover f_L	0.891; 0.158 / 0.119 / 0.080
Saving with $\eta_{\text{fin}}(T)$ (%)	9.41 / 5.75 / 2.67
Clock scaling: M ($\text{kg kW}_{\text{ref}}^{-1}$); saving (%); saving 85 → 150 °C (%)	49.25; 35.2; 11.35 / 8.45 / 5.92
Clock scaling: stack lever 40 → 25 K vs higher ratings, $f_L = 0.15$; 0.10 (%; median)	8.72 vs 8.45; 9.52 vs 13.55
Best heat-pump design, constant η_{fin} ; $\eta_{\text{fin}}(T)$ (%; condensing K)	−15.5 (380); −6.1 (368)
Lifetime at 150 °C, model A; model B with 2% per year (%; median)	−87.0 to −15.2; +0.7 to +7.8

The pessimistic radiator follows the same pattern as the heavy one, scaled up: it is still pinned by the chip limit, its break-even value lies close to Eq. (7) (−2.1%, Table S28), and the stack lever exceeds higher ratings at fixed clock for f_L above 0.080–0.158. Only for the flat shape does its crossover reach the evidence range.

S9.11. HBM limit equal to the chip limit

If the HBM limit is set equal to the chip limit (85 °C at present, rising with the chip limit), the HBM binds at present because $D_{\text{rating}} = 0$ is below δ (Section S9.2). The nominal optimum then lies at a coolant outlet temperature of 45.00 °C with the hot chip at 84.5 °C and a mass of $44.743 \text{ kg kW}_{\text{ref}}^{-1}$, raising the chip limit alone saves nothing, and every saving comes from the HBM rating. The savings from raising both limits are slightly larger than with the 95 °C HBM limit because the present design is slightly colder (Table S39); the break-even values change by about 3% in f_L and 1% in s .

Table S39. Saving from raising chip and HBM limits from 85 to 150 °C when the HBM limit equals the chip limit (%; flat / median / steep shape), and the thresholds of the nominal radiator.

Radiator	$f_L = 0$	$f_L = 0.10$	$f_L = 0.15$	$f_L = 0.22$	Chip limit alone, $f_L = 0.15$
Advanced	14.85	3.51 / 2.25 / 1.16	1.31 / 0.61 / 0.17	0.11 / 0.00 / 0.00	0
Nominal	23.53	8.93 / 6.19 / 3.59	4.68 / 2.72 / 1.19	1.42 / 0.48 / 0.03	0
Heavy	32.23	15.82 / 11.27 / 6.80	9.38 / 5.80 / 2.81	3.83 / 1.73 / 0.49	0
Nominal thresholds	$f_{L,5}$ 0.145 / 0.114 / 0.083	$f_{L,0}$ 0.318 / 0.279 / 0.227	s_5 0.351 / 0.331 / 0.303% K ⁻¹	s_0 0.771 / 0.809 / 0.827% K ⁻¹	

S9.12. Environmental load and solar beta angle

A dawn–dusk Sun-synchronous orbit at 650 km has an inclination of 97.99°, and the solar beta angle varies over the year between 58.6° and 89.9°. For a radiator whose normal stays along the velocity vector (no yaw), the orbit-average environmental load at the least favorable beta angle is 170.5 W m⁻², 1.3395 times the default of Eq. (S5), with a peak of 1.67 times the default; between 600 and 800 km the orbit-average factor is 1.28–1.35. Yaw steering that keeps the Sun at grazing incidence limits the orbit average to at most 108.7 W m⁻². The end-of-life variant (Section S9.7) and the pessimistic radiator (Section S9.10) use the orbit average at the least favorable beta angle; sizing on the orbit average neglects the thermal capacitance of the radiator, and the peak value is not used (Table S40).

Table S40. Absorbed environmental load (W m⁻², both faces) at 650 km versus solar beta angle.

Beta angle (°)	Fixed attitude, orbit average	Fixed attitude, peak	Yaw steering, orbit average	Yaw steering, peak
58.6	170.5	211.9	108.7	119.0
60	167.8	207.5	108.6	118.4
65	162.8	191.6	107.8	116.2
70	151.7	175.0	107.1	113.8
75	140.1	158.0	106.4	111.4
80	128.4	140.5	105.7	109.0
90	105.0	105.0	105.0	105.0

At a beta angle of 90° the geometric orbit average (105.0 W m⁻²) lies below the default of 127.3 W m⁻², which includes the albedo factor of 0.5 and the 2% pointing-error term.

Supplementary figures

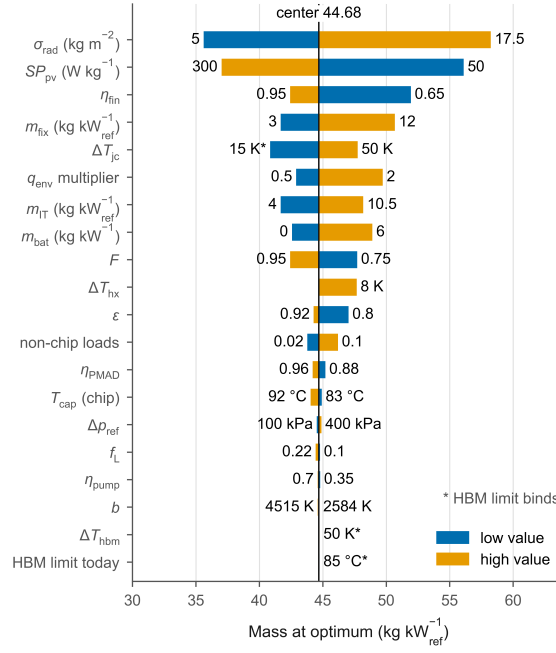


Fig. S1. One-at-a-time sensitivity around the nominal design (chip limit 85 °C, HBM limit 95 °C, $f_L = 0.15$, median shape). Bars show the on-orbit mass at the optimum when each parameter is set to the low (blue) and high (orange) values of Table S1, with the parameter value printed at the bar end; the vertical line marks the center value of 44.68 kg kW⁻¹. Asterisks mark the three settings in which the HBM limit binds instead of the chip limit ($\Delta T_{\text{jc}} = 15$ K, $\Delta T_{\text{hbm}} = 50$ K and an HBM limit of 85 °C). The heat-pump parameters, which do not affect the passive design, are omitted. Numerical values, including the saving from higher ratings and the break-even values $f_{L,0}$ and s_0 , are listed in Table S4.

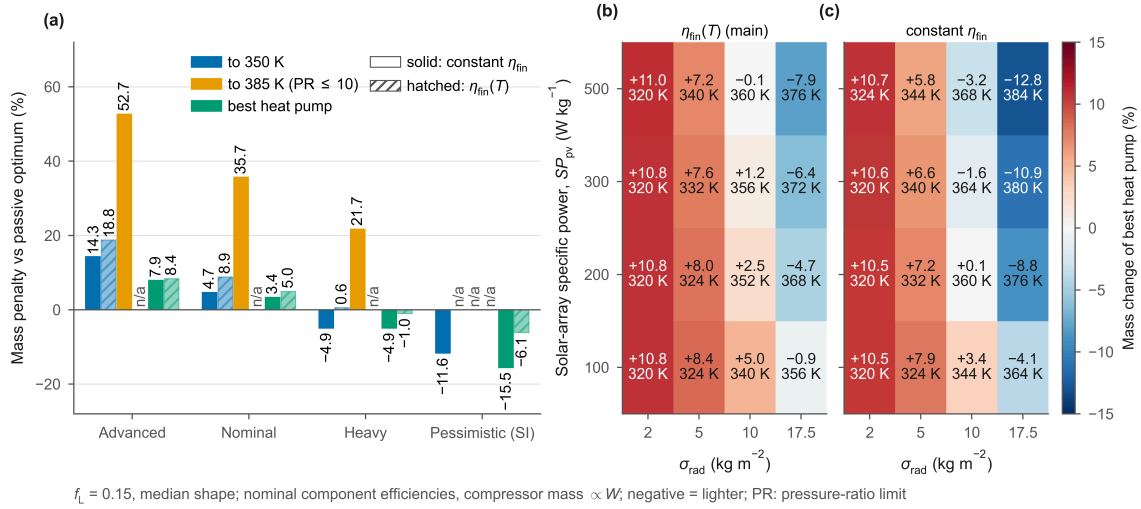


Fig. S2. Heat pumps ($f_L = 0.15$, median shape; nominal component efficiencies, compressor mass scaling with electric work). (a) Mass change relative to the passive optimum with the same parameters for heat pumps forced to an effective radiating temperature of 350 K and of 385 K (pressure ratio at most 10) and for the best heat-pump design, for the advanced, nominal, heavy and pessimistic radiators; solid bars use a constant fin efficiency and hatched bars a fin efficiency that declines with temperature; positive values mean the heat-pump design is heavier, and “n/a” marks temperatures that cannot be reached. (b, c) Mass change of the best heat-pump design relative to the passive optimum versus radiator areal mass and solar-array specific power, (b) with the declining and (c) with the constant fin efficiency ($\eta_{\text{fin}} = 0.85$ at the calibration temperature); negative values mean the heat-pump design is lighter, as in (a), and each cell gives the optimum condensing temperature. Numerical values are listed in Tables S15 and S16.

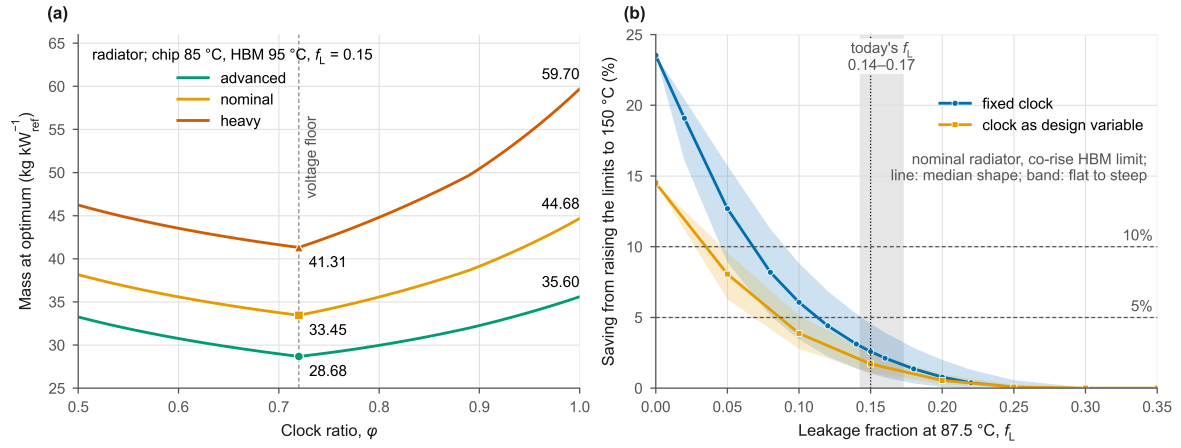


Fig. S3. Clock scaling (chip limit 85 °C, HBM limit 95 °C, $f_L = 0.15$). (a) On-orbit mass at the optimum versus clock ratio for the three reference radiators (median shape); markers show the optima at the voltage floor (clock ratio 0.72), 28.68, 33.45 and 41.31 kg kW⁻¹_{ref}, against 35.60, 44.68 and 59.70 kg kW⁻¹_{ref} at nominal clock. (b) Saving from raising the chip limit from 85 to 150 °C, with the HBM limit rising with it, versus leakage fraction at fixed clock and with the clock as a design variable (nominal radiator); lines show the median shape and bands the range from the flat to the steep shape, the gray band marks the center of the present evidence (0.14–0.17), and dashed lines mark savings of 5% and 10%.

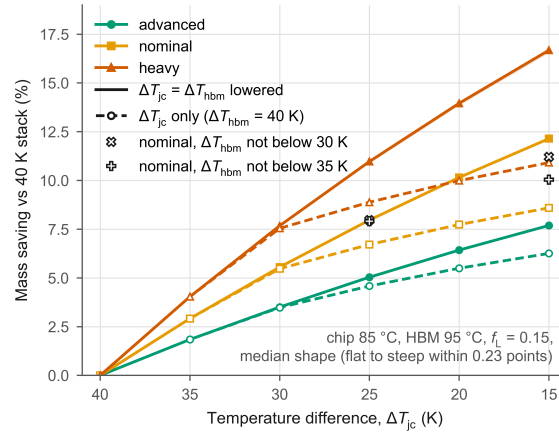


Fig. S4. Temperature-stack lever (chip limit 85 °C, HBM limit 95 °C, $f_L = 0.15$, median shape; with both differences lowered the flat and steep shapes differ by at most 0.23 percentage points). Mass saving relative to a 40 K stack for the three reference radiators versus ΔT_{jc} : solid lines lower ΔT_{jc} and ΔT_{hbm} together, dashed lines lower ΔT_{jc} only with $\Delta T_{hbm} = 40$ K, and black markers show the nominal radiator with ΔT_{hbm} not reduced below 30 K or 35 K. Numerical values are listed in Table S30.

References

- [1] E. Feilden, A. Oltean, P. Johnston, Why we should train AI in space, White Paper v1.03, Lumen Orbit (now Starcloud), September 2024. <https://starcloudinc.github.io/wp.pdf> (accessed 23 October 2026)
- [3] S.G. Turyshev, Orbital data centers: spacecraft constraints and economic viability, arXiv:2604.27197 (2026). <https://arxiv.org/abs/2604.27197>
- [5] S. Gaalema, S. Indyk, C. Staley, Reduced-mass orbital AI inference via integrated solar, compute, and radiator panels, arXiv:2604.07760 (2026). <https://arxiv.org/abs/2604.07760>
- [11] K.R. Sridhar, M. Gottmann, Thermal control systems for low-temperature heat rejection on a lunar base, NASA Contractor Report NASA-CR-190063, University of Arizona, February 1992 (grant NAG5-1572). <https://ntrs.nasa.gov/citations/19920011027>
- [13] H.J. van Gerner, G. van Donk, A. Pauw, J. van Es, S. Lapensée, A heat pump for space applications, ICES-2015-35, in: 45th International Conference on Environmental Systems, Bellevue, WA, USA, 12–16 July 2015. <http://hdl.handle.net/2346/64305>

- [16] M. Zapater, O. Tuncer, J.L. Ayala, J.M. Moya, K. Vaidyanathan, K. Gross, A.K. Coskun, Leakage-aware cooling management for improving server energy efficiency, *IEEE Trans. Parallel Distrib. Syst.* 26 (10) (2015) 2764–2777. <https://doi.org/10.1109/TPDS.2014.2361519>
- [17] D.C. Price, M.A. Clark, B.R. Barsdell, R. Babich, L.J. Greenhill, Optimizing performance-per-watt on GPUs in high performance computing: temperature, frequency and voltage effects, *Comput. Sci. Res. Dev.* 31 (4) (2016) 185–193. <https://doi.org/10.1007/s00450-015-0300-5>
- [20] The Boeing Company, Active Thermal Control System (ATCS) overview, ISS fact sheet, Boeing IDS Business Support, St. Louis, MO, USA, 2006 (undated; PDF created 22 November 2006), 19 pp. https://www.nasa.gov/wp-content/uploads/2021/02/473486main_iss_atcs_overview.pdf (accessed 23 October 2026)
- [21] I. Latif, M.A. Shafique, H. Ullah, A.C. Newkirk, X. Yu, A. Munir, Cooling matters: benchmarking large language models and vision-language models on liquid-cooled versus air-cooled H100 GPU systems, *IEEE Trans. Cloud Comput.* 14 (3) (2026) 1645–1656. <https://doi.org/10.1109/TCC.2026.3700971>
- [22] P. Sinha, A. Guliani, R. Jain, B. Tran, M.D. Sinclair, S. Venkataraman, Not all GPUs are created equal: characterizing variability in large-scale, accelerator-rich systems, in: SC22: International Conference for High Performance Computing, Networking, Storage and Analysis, Dallas, TX, USA, 13–18 November 2022, pp. 1–15. <https://doi.org/10.1109/SC41404.2022.00070>
- [23] (a) slow-zhang, Question about MIG config persistent, NVIDIA/k8s-device-plugin, GitHub issue #343, 3 November 2022. <https://github.com/NVIDIA/k8s-device-plugin/issues/343> (accessed 23 October 2026). (b) csibbitt, DCGM-Exporter crashes on startup when using MIG w/ A100, NVIDIA/dcg-exporter, GitHub issue #443, 16 January 2025. <https://github.com/NVIDIA/dcg-exporter/issues/443> (accessed 23 October 2026).
- [24] NVIDIA, NVIDIA H100 PCIe GPU product brief, PB-11133-001_v02, November 2022. https://www.nvidia.com/content/dam/en-zz/Solutions/gtcs22/data-center/h100/PB-11133-001_v01.pdf (accessed 23 October 2026)
- [26] Texas Instruments, AM62x Power Estimation Tool, AM62X-PET-CALC, version 01.00.00.00, released 30 May 2022; application note SPRAD31, May 2022. <https://www.ti.com/tool/download/AM62X-PET-CALC/01.00.00.00> (accessed 23 October 2026)
- [27] Texas Instruments, AM62D Power Estimation Tool, AM62D-PET-CALC, version 1.1, released 21 February 2025; application note SPRADO7. <https://www.ti.com/tool/download/AM62D-PET-CALC/1.1> (accessed 23 October 2026)
- [28] AMD (Xilinx), Xilinx Power Estimator (XPE) for UltraScale+, version 2023.1.2, released 16 August 2023. <https://www.amd.com/en/products/adaptive-socs-and-fpgas/technologies/power-efficiency/power-estimator.html> (accessed 23 October 2026)
- [29] S. Kim, J. Kim, D. Jang, R. Ritzenthaler, B. Parvais, J. Mitard, H. Mertens, T. Chiarella, N. Horiguchi, J.W. Lee, Comparison of temperature dependent carrier transport in FinFET and gate-all-around nanowire FET, *Appl. Sci.* 10 (8) (2020) 2979. <https://doi.org/10.3390/app10082979>
- [30] E. Simoen, C.H.S. Coelho, V.C.P. da Silva, J.A. Martino, P.G.D. Agopian, A. Oliveira, B. Cretu, A. Veloso, Performance perspective of gate-all-around double nanosheet CMOS beyond high-speed logic applications, *J. Integr. Circuits Syst.* 17 (2) (2022) 1–9. <https://doi.org/10.29292/jics.v17i2.617>
- [31] M. de Souza, A. Cerdeira, M. Estrada, S. Barraud, M. Cassé, M. Vinet, O. Faynot, M.A. Pavanello, High temperature and width influence on the GIDL of nanowire and nanosheet SOI nMOSFETs, *IEEE J. Electron Devices Soc.* 11 (2023) 672–680. <https://doi.org/10.1109/JEDS.2023.3264876>
- [32] S. Samsi, M.L. Weiss, D. Bestor, B. Li, M. Jones, A. Reuther, D. Edelman, W. Arcand, C. Byun, J. Holodnick, M. Hubbell, J. Kepner, A. Klein, J. McDonald, A. Michaleas, P. Michaleas, L. Milechin, J. Mullen, C. Yee, B. Price, A. Prout, A. Rosa, A. Vanterpool, L. McEvoy, A. Cheng, D. Tiwari, V. Gadepally, The MIT Supercloud dataset, in: 2021 IEEE High Performance Extreme Computing Conference (HPEC), Waltham, MA, USA, 20–24 September 2021, pp. 1–8. <https://doi.org/10.1109/HPEC49654.2021.9622850>
- [33] [dataset] S. Samsi, M. Weiss, D. Bestor, et al., MIT Supercloud dataset, Datacenter Challenge labelled data, release 202201 (Version 1), Registry of Open Data on AWS, 2022. <https://registry.opendata.aws/dcc/> (accessed 23 October 2026). License CC BY-NC-ND 4.0. s3://mit-supercloud-dataset/datacenter-challenge/202201/
- [34] I.H. Bell, J. Wronski, S. Quoilin, V. Lemort, Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library CoolProp, *Ind. Eng. Chem. Res.* 53 (6) (2014) 2498–2508. <https://doi.org/10.1021/ie4033999>
- [36] R. Schoonhoven, B. Veenboer, B. van Werkhoven, K.J. Batenburg, Going green: optimizing GPUs for energy efficiency through model-steered auto-tuning, in: 2022 IEEE/ACM International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS), Dallas, TX, USA, November 2022, pp. 48–59. <https://doi.org/10.1109/PMBS56514.2022.00010>

- [37] D.L. Ellis, J. Calder, J. Siamidis, Summary of the manufacture, testing and model validation of a full-scale radiator for fission surface power applications, in: Proc. Nuclear and Emerging Technologies for Space 2011 (NETS 2011), Albuquerque, NM, USA, 7–10 February 2011, paper 3181 (NASA report E-17664). <https://ntrs.nasa.gov/citations/20110012006>
- [38] L. James, Elon Musk's first-gen orbital data center craft spans wider than a Boeing 747 and runs an interchangeable chip payload — AI1 satellite compute payload is 120 kW, peaks at 150 kW, Tom's Hardware, 9 June 2026. <https://www.tomshardware.com/tech-industry/spacex-details-its-ai1-compute-satellite> (accessed 23 October 2026). Primary source: SpaceX (@SpaceX), technical update video on SpaceX's capability to manufacture, launch, and operate AI satellites at scale, X post, 8 June 2026. <https://x.com/SpaceX/status/2064099405758906727> (accessed 23 October 2026)
- [39] J. Crop, H. Moore, S. Pasricha, Revisiting “Cooler is Better”: ITD-aware per-CPU thermal optimization for sustainable data center operation, in: Proc. 16th ACM International Green and Sustainable Computing Conference (IGSC '26), Canandaigua, USA, 2026, pp. 1–7. <https://doi.org/10.1145/3797248.3815399>; arXiv:2606.11163
- [41] Vertiv, Keep it at 30 °C: a roadmap for the future chip coolant temperature, Vertiv blog, 9 April 2026. <https://www.vertiv.com/en-emea/insights/articles/blog-posts/a-roadmap-for-the-future-chip-coolant-temperature/> (accessed 23 October 2026)
- [42] V. Kandiah, S. Peverelle, M. Khairy, J. Pan, A. Manjunath, T.G. Rogers, T.M. Aamodt, N. Hardavellas, AccelWattch: a power modeling framework for modern GPUs, in: Proc. 54th Annu. IEEE/ACM Int. Symp. Microarchitecture (MICRO-54), 2021, pp. 738–753. <https://doi.org/10.1145/3466752.3480063>
- [43] U.S. Department of Energy, National Renewable Energy Laboratory, EnergyPlus reference data set Chillers.idf (ElectricEIRChiller, Trane RTWA 383 kW, COP 4.17) [dataset], EnergyPlus GitHub repository. <https://github.com/NatLabRockies/EnergyPlus/blob/develop/datasets/Chillers.idf> (accessed 23 October 2026).
- [44] JEDEC Solid State Technology Association, Failure mechanisms and models for semiconductor devices, JEDEC Publication JEP122H, September 2016.
- [45] P. Johnston, interviewed by C. Simms, AI hits a power wall. Starcloud launches data centers into orbit, Inevitable podcast (MCJ), 13 January 2026. <https://mcj.vc/inevitable-podcast/starcloud> (accessed 23 October 2026).