

Certified global heat optimality of two full swaps for a qutrit cooling task

Yuxuan Zhang

Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland

Department of Physics, Princeton University
Princeton, New Jersey 08544, USA

Version 1
22 October 2026

Based on candidate M08-v1. Proof and primary-source comparisons received separate-context internal AI review; computational checks were replayed. No external expert review or formal verification is claimed.

Abstract

We determine the minimum discarded heat for a specified two-collision qutrit cooling problem. A target with energies $(0, 1, 2)$ starts at inverse temperature one and must end at inverse temperature two. Each collision uses a fresh thermal qutrit with a scaled copy of the target Hamiltonian, with gaps in $[1, 3]$ and total gap at most four. Arbitrary joint unitaries and classical randomization are allowed. We prove that two optimized full swaps attain the global optimum, including the original bound on gross supplied work. The optimal gaps are $(a_*, 2)$, where a_* is the unique root of an explicit scalar equation, and the minimum heat is approximately 0.464789080273467. An affine terminal witness covers randomized protocols with an exact unconditioned target. Concavity reduces the quantum support problem to 280 population partitions. Separate semiconcavity bounds, exact integer arithmetic, and a complete interval certificate then establish a lower bound on the entire continuous gap domain. A local Hessian certificate treats the equality point. The result concerns this finite qutrit task, rather than arbitrary higher-dimensional cooling schedules.

1 The cooling task and main theorem

Finite-machine cooling requires simultaneous optimization of the machine spectra and the interactions. For qubit collision models, Taranto et al. establish optimality results for swap-based coherent cooling [1]. The higher-dimensional problem has additional population and coherence degrees of freedom. Here we resolve a fixed qutrit instance by proving a lower bound for every allowed quantum protocol and matching it with a feasible swap protocol.

Set $\hbar = k_B = \beta = 1$, and write

$$\begin{aligned} H_S &= \text{diag}(0, 1, 2), & p_j(g) &= \frac{e^{-gj}}{1 + e^{-g} + e^{-2g}}, \\ \tau_g &= \text{diag}(p_0(g), p_1(g), p_2(g)), & e(g) &= p_1(g) + 2p_2(g). \end{aligned} \tag{1.1}$$

The target starts in τ_1 . Machine i has Hamiltonian $H_i = g_i H_S$ and starts in τ_{g_i} , with

$$1 \leq g_1, g_2 \leq 3, \quad g_1 + g_2 \leq 4. \tag{1.2}$$

The target interacts once with each machine, in order, through arbitrary $U_i \in U(9)$; the machine is then discarded. There are no ancillary memories, measurements, or postselection. Each interaction has available time π and control norm bound one in the interaction picture. Every $U(9)$ is allowed by this budget: a Hermitian logarithm with eigenvalues in $[-\pi, \pi]$, divided by π , is an admissible constant generator.

Let ρ_{i-1} be the target input, and define

$$\rho'_{Si} = U_i(\rho_{i-1} \otimes \tau_{g_i})U_i^\dagger, \quad \rho_i = \text{Tr}_{M_i} \rho'_{Si}, \quad (1.3)$$

$$Q_i = \text{Tr}[H_i(\text{Tr}_S \rho'_{Si} - \tau_{g_i})], \quad (1.4)$$

$$w_i = \text{Tr}[(H_S + H_i)(\rho'_{Si} - \rho_{i-1} \otimes \tau_{g_i})]. \quad (1.5)$$

The final target must equal τ_2 , and gross supplied mean work is charged as $[w_1]_+ + [w_2]_+ \leq 2$. Classical randomization is allowed, with work charged on branches and the unconditioned final state required to be τ_2 . Denote the infimum of total heat by Q_* .

Full swaps with gaps $(a, 2)$, $1 \leq a \leq 2$, have heat

$$q(a) = a[e(1) - e(a)] + 2[e(a) - e(2)], \quad Q_{\text{swap}} = \min_{1 \leq a \leq 2} q(a). \quad (1.6)$$

Theorem 1.1. *For the task (1.1)–(1.5),*

$$Q_* = Q_{\text{swap}} = q(a_*), \quad (1.7)$$

where $a_* \in (1, 2)$ is the unique solution of

$$e(1) - e(a_*) + (2 - a_*)e'(a_*) = 0. \quad (1.8)$$

Two full swaps with gaps $(a_*, 2)$ attain the optimum. Their heat and gross supplied work are

$$q(a_*) = 0.4647890802734670691 \dots, \quad (1.9)$$

$$q(a_*) + e(2) - e(1) = 0.1890623706570403958 \dots < 2, \quad (1.10)$$

with $a_* = 1.4459206660687878334 \dots$. The lower bound remains valid after dropping the work constraint and allowing branch-dependent admissible gaps.

The decimals display quantities defined by exact equations. The proof combines analytic reductions with the finite certificate described in Sections 5 and 6; the supplied checker recomputes every decisive inequality.

2 The optimized swap benchmark and terminal witness

With $t = e^{-a}$, direct differentiation gives

$$e'(a) = -\frac{t(t^2 + 4t + 1)}{(1 + t + t^2)^2}, \quad (2.1)$$

$$e''(a) = \frac{t(1 - t)(1 + t)(t^2 + 7t + 1)}{(1 + t + t^2)^3} > 0 \quad (a > 0). \quad (2.2)$$

Hence

$$q''(a) = -2e'(a) + (2 - a)e''(a) > 0 \quad (1 \leq a \leq 2).$$

Moreover, $q'(1) = e'(1) < 0$ and $q'(2) = e(1) - e(2) > 0$. This proves existence and uniqueness of the root in (1.8) and global optimality within the swap class.

Choose the exact vector and diagonal operator

$$\lambda = (\lambda_0, 3, 0), \quad \Lambda = \text{diag}(\lambda), \quad \lambda_0 = \frac{e(a_*) - e(2) - 2e'(2) - 3p'_1(2)}{p'_0(2)}. \quad (2.3)$$

The denominator is positive. The certificate encloses $\lambda_0 = 5.6959953529968454488\dots$, in particular $3 < \lambda_0 < 6$. Define the constant

$$z_* = q(a_*) - \lambda \cdot p(2). \quad (2.4)$$

We prove the support inequality

$$Q_1 + Q_2 - \text{Tr}(\Lambda \rho_2) \geq z_* \quad (2.5)$$

for every allowed gap pair and every pair of quantum unitaries, without imposing final-state or work constraints. Taking an average of (2.5) over classical branches preserves its affine left side. The unconditioned constraint $\mathbb{E}\rho_2 = \tau_2$ then implies $Q_1 + Q_2 \geq q(a_*)$. Individual branches need not end at τ_2 and may have final coherences. This is the reason for certifying an affine terminal witness rather than optimizing only protocols that reach the target branch by branch.

3 Reducing arbitrary quantum unitaries to partitions

The use of squared unitary entries as doubly stochastic maps is a standard population reduction, also used in Ref. [1], Appendix B.1. Here the additional step is a concave continuation objective: it allows the two-collision terminal witness to be bounded at first-stage permutation vertices.

3.1 Concavity of the second-collision support

Fix $g_2 = b$ and define, for a target density operator ρ ,

$$f_b(\rho) = \min_{U \in U(9)} \text{Tr}[(I \otimes bH_S - \Lambda \otimes I)U(\rho \otimes \tau_b)U^\dagger] - be(b). \quad (3.1)$$

As an infimum of affine functions, f_b is concave. It is invariant under target input conjugations, since such a conjugation can be absorbed into the unrestricted minimizing unitary.

For the first collision, put $a = g_1$ and $x = p(1) \otimes p(a)$. Diagonalize its target output by a target-only conjugation of the joint output. This leaves the first-machine heat unchanged and leaves f_b unchanged. In this basis the joint diagonal y satisfies $y = Dx$, where D is the matrix of squared moduli of a unitary and is doubly stochastic. The first-machine heat is linear in y , and the target marginal is given by its row sums. Thus

$$G(y) = a \sum_{i,j} j y_{ij} - ae(a) + f_b \left(\text{diag} \left(\sum_j y_{0j}, \sum_j y_{1j}, \sum_j y_{2j} \right) \right) \quad (3.2)$$

is concave in y . By the Birkhoff decomposition, y is a convex combination of permutations of x . Concavity therefore gives

$$G(y) \geq \min_{\sigma \in S_9} G(\sigma x).$$

This is a lower bound for the original quantum protocol. Conversely, every permutation vertex is implemented by a unitary, so it remains to optimize a finite population problem.

At a permutation vertex the second input is diagonal. If its target marginal is m , its nine joint eigenvalues are $m_i p_j(b)$, and the destination costs are

$$c_{ij} = bj - \lambda_i. \quad (3.3)$$

A second doubly stochastic decomposition and the rearrangement inequality show that the minimum linear cost pairs decreasing input masses with increasing costs. One can also see rearrangement directly: if two masses and their costs have the same strict ordering, exchanging their assigned costs lowers the objective. Repeated exchanges reach the stated pairing. Permutation unitaries attain it, so no relaxation remains at this step.

3.2 The 280 representatives

Partition the nine labeled initial masses into three output rows. The optimized second-stage value depends only on the unordered row sums. Within each row, placing masses in decreasing order at machine occupations 0, 1, 2 minimizes the first heat, without altering those sums. Row labels are immaterial by input-unitary invariance of f_b . The number of unordered partitions into triples is therefore

$$|\mathcal{P}| = \frac{9!}{(3!)^3 3!} = 280. \quad (3.4)$$

A unique enumeration places the smallest remaining label in the next triple and chooses its two companions recursively. Labels are retained even when masses coincide; coincidences only repeat attainable objective values.

For $P \in \mathcal{P}$, let y^P be its table with each row decreasing and let $m_i = \sum_j y_{ij}^P$. Define

$$\begin{aligned} F_P(a, b) &= a \sum_{i,j} j y_{ij}^P - ae(a) + \sum_{r=1}^9 (m \otimes p(b))_r^\downarrow c_r^\uparrow - be(b), \\ F(a, b) &= \min_{P \in \mathcal{P}} F_P(a, b). \end{aligned} \quad (3.5)$$

The preceding arguments establish that $F(a, b)$ is the exact minimum of the left side of (2.5) over all quantum unitaries at fixed gaps, after the work constraint is dropped. The support problem is now to prove $F(a, b) \geq z_*$ throughout the triangle (1.2).

4 Lower bounds between evaluated points

The minimizing partition and the rearrangement ordering can change with the gaps. A smooth fit to sampled minima would not control those changes. We instead establish upper bounds on each coordinate second derivative of every fixed permutation branch.

For the thermal law on $\{0, 1, 2\}$,

$$p'_j(g) = p_j(g)(e(g) - j), \quad p''_j(g) = p_j(g)[(e(g) - j)^2 - \text{Var}_g(j)]. \quad (4.1)$$

Since a random variable in $[0, 2]$ has variance at most one,

$$\|p'(g)\|_1 \leq \sqrt{\text{Var}_g(j)} \leq 1, \quad \|p''(g)\|_1 \leq 2 \text{Var}_g(j) \leq 2. \quad (4.2)$$

Also $\sum_j p'_j = \sum_j p''_j = 0$.

Lemma 4.1. *On the whole square $[1, 3]^2$, every fixed pair of population permutations has support function v satisfying $v_{aa} \leq 28$ and $v_{bb} \leq 22$. The finite minimum F is separately semiconcave with those constants. On a rectangle of widths h_a, h_b ,*

$$F(a, b) - z_* \geq \min_{\text{four corners}} (F - z_*) - \frac{28h_a^2 + 22h_b^2}{8}. \quad (4.3)$$

The same assertion holds for the minimum over any subcollection of the first-stage partitions.

Proof. For a fixed first permutation, the first heat can be written $a \sum_j (A_j - j)p_j(a)$, where $0 \leq A_j \leq 2$. Its second derivative is at most $2 \cdot 2 \cdot 1 + 3 \cdot 2 \cdot 2 = 16$, by (4.2). The terminal cost $bj - \lambda_i$ lies in an interval of length $2b + \lambda_0 < 12$. After averaging over the other fixed input factors, its coefficients as a function of $p(a)$ still have range less than 12. Center those coefficients at their midrange; the centering changes no second derivative because $\sum_j p_j'' = 0$. Their contribution is at most $(12/2) \|p''\|_1 \leq 12$. This proves $v_{aa} \leq 28$.

For the second coordinate, the second-stage expression has the form

$$b \sum_j (A_j - j)p_j(b) - \sum_j B_j p_j(b), \quad 0 \leq A_j \leq 2, \quad 0 \leq B_j \leq \lambda_0.$$

The first term contributes at most 16. Centering B as above bounds the second by $\lambda_0 < 6$, proving $v_{bb} \leq 22$.

For each fixed b , subtracting $14a^2$ makes every permutation branch concave in a ; the pointwise minimum of concave functions is concave. The same argument with $11b^2$ holds in the other coordinate. The sorting choices in (3.5) are themselves minima over finitely many fixed branches, so their boundaries cause no difficulty. A semiconcave function on an interval lies above its endpoint chord minus $C(x-l)(u-x)/2$. Apply this successively in the two coordinates and use $(x-l)(u-x) \leq (u-l)^2/4$ to obtain (4.3). Restricting the branch collection leaves every step valid. $\square$

5 The complete global certificate

A finite dyadic subdivision of $[1, 3]^2$ separates a small box containing the equality point from the rest of the domain. The distinguished box is

$$B_0 = [7/5, 3/2] \times [39/20, 41/20]. \quad (5.1)$$

The stored cover has 1006 remote leaves and 18 local leaves. The checker reconstructs binary splits from the root square, always splitting the longest side and choosing the first coordinate in a tie. It stops at a listed leaf or a cell whose lower-left coordinate sum is at least four. It verifies that every listed leaf is visited, and that every local leaf is contained in B_0 . Thus every feasible interior point is covered. The feasible edge $a + b = 4$ follows by continuity from the interior.

At each remote leaf, the checker evaluates lower bounds for all four corners and applies (4.3). Every resulting lower bound is strictly positive; a conservatively rounded lower margin is $2.70 \cdot 10^{-6}$. The next subsection explains how the corner values are certified.

5.1 Integer arithmetic and the uniform input error

The exact root a_* is enclosed by sign-certified bisection starting at $[1.44, 1.46]$. The thermal probabilities, multiplier, and z_* are enclosed with 200-bit Arb ball arithmetic [2]. Each primitive thermal component and λ_0 is rounded down to an integer multiple of $1/S$, where

$S = 2^{60}$. The checker certifies error at most $1/S$ in each component and preserves the full interval uncertainty in z_* .

After this initialization, all products, sorting, costs, and minima in (3.5) use Python integers and a common rational denominator. For rational gaps a, b , put $D = \text{lcm}(\text{den } a, \text{den } b)$. The common objective denominator is DS^4 . To verify that input rounding is harmless uniformly across ordering changes, consider any fixed permutation branch.

Each thermal vector has L^1 error at most $3/S$, and its rounded-down mass is at most one. The first joint law has error at most $6/S$, its row sums have no larger error, and the second joint law has error at most $9/S$. The three objective pieces have the following error bounds:

Objective piece	Absolute error bound
First heat: output and input machine energy	$54/S$
Second terminal expectation, including λ_0	$55/S$
Second input machine energy	$18/S$
Total	$127/S$

For the first and third lines, the largest machine energy is six. In the second line, $|bj - \lambda_i| \leq 6$, the joint-law error is at most $9/S$, and the multiplier error adds at most $1/S$. These bounds apply before minimization. If every branch changes by at most a fixed error, so does its minimum, irrespective of ties or a change in the minimizing order. Subtracting $1000/S$ from the integer objective is therefore a safe allowance. Conversion of its rational value back to an Arb ball encloses the exact integer ratio rather than treating a floating value as exact.

The remote check uses only these lower corner values, the exact dyadic widths, and (4.3). Search values stored during discovery are not used to decide the certified sign.

6 The equality neighborhood

The local box contains the point $(a_*, 2)$, at which the support equals z_* . A strictly positive remote-style bound cannot hold there. We instead isolate the swap branch and prove convexity on the entire box.

With original label $3i + j$ for initial mass $p_i(1)p_j(a)$, the relevant partition is

$$P_{\text{sw}} = \{(0, 3, 6), (1, 4, 7), (2, 5, 8)\}. \quad (6.1)$$

It is index 155 in the enumerator, counted from zero. Its row sums are $p(a)$ and its first heat is $a[e(1) - e(a)]$. Throughout B_0 , the decreasing order of the second input masses has indices

$$(0, 3, 1, 6, 4, 2, 7, 5, 8),$$

while the increasing destination-cost order is

$$(0, 1, 3, 2, 4, 6, 5, 7, 8).$$

All adjacent exponent and cost differences are affine expressions with strictly positive interval lower bounds on B_0 . Their pairing is the full swap. The corresponding support is therefore the smooth function

$$s(a, b) = ae(1) + (b - a)e(a) - be(b) - \lambda \cdot p(b). \quad (6.2)$$

Equations (1.8) and (2.3) give $\nabla s(a_*, 2) = 0$ and $s(a_*, 2) = z_*$.

Its Hessian entries are

$$s_{aa} = -2e'(a) + (b-a)e''(a), \quad s_{ab} = e'(a), \quad (6.3)$$

$$s_{bb} = -2e'(b) - be''(b) - \lambda_0 p_0''(b) - 3p_1''(b). \quad (6.4)$$

The checker divides B_0 into 20×20 rational rectangles and encloses each expression directly using 200-bit Arb arithmetic. Every rectangle has

$$s_{aa} > 0.6523, \quad \det \nabla^2 s > 0.2056. \quad (6.5)$$

These are conservatively rounded display margins; the checker asserts positivity of the enclosed quantities themselves. The coordinate balls enclose the exact rational endpoints, a condition also checked in the separate review controls. The Hessian is positive definite everywhere on the convex box, so

$$s(a, b) \geq s(a_*, 2) = z_* \quad ((a, b) \in B_0). \quad (6.6)$$

It remains to exclude a different partition falling below the fixed support floor z_* . Apply Lemma 4.1 to the other 279 partitions. A separate adaptive rational subdivision of B_0 closes with 31 leaves, each having a strictly positive lower bound for

$$\min_{P \neq P_{\text{sw}}} F_P(a, b) - z_*. \quad (6.7)$$

A conservatively rounded minimum margin is $5.71 \cdot 10^{-6}$. This comparison is with the constant z_* ; no pointwise ordering between all competing partitions and $s(a, b)$ is needed or asserted. Combining the swap branch and (6.7) proves the support bound on B_0 , completing the global certificate.

Certified component	Size or lower margin
Unordered first-stage partitions	280
Remote leaves / local cover leaves	1006 / 18
Remote support margin	$> 2.70 \cdot 10^{-6}$
Local Hessian rectangles	400
Local s_{aa} / determinant	$> 0.6523 / 0.2056$
Competing-partition leaves	31
Competing support-floor margin	$> 5.71 \cdot 10^{-6}$
Input rounding scale / allowance	$S = 2^{60}; 1000/S$

Table 1: Finite certificate for the continuum support inequality. All signs are recomputed with enclosed or exact arithmetic; the displayed margins are rounded conservatively.

7 Feasibility and attainment

The support inequality (2.5) now holds for all quantum protocols at every gap pair satisfying (1.2). Its affine averaging argument proves the desired heat lower bound even when gaps depend on the classical branch and gross work is unrestricted.

For attainment, choose two full swaps with gaps $(a_*, 2)$. Since $1 < a_* < 2$, both gaps lie in $[1, 3]$ and their sum is below four. The target becomes τ_{a_*} and then exactly τ_2 . Each swap is an admissible $U(9)$ with the prescribed time and norm resources. For a swap with machine gap g ,

$$Q = g(e_{\text{in}} - e_{\text{out}}), \quad w = (g - 1)(e_{\text{in}} - e_{\text{out}}). \quad (7.1)$$

Both stages have positive work because e decreases and $1 < a_* < 2$. Hence gross work equals net work and is

$$q(a_*) + e(2) - e(1) < 2, \quad (7.2)$$

as certified by the same root enclosure. The protocol is feasible under the original work constraint and attains the lower bound. This proves Theorem 1.1.

8 Reproduction and scope

The accompanying M08-v1 evidence contains the original statement, proof, `cover.json`, and three checker modules. The top-level checker reconstructs the entire dyadic cover and all 280 representatives; it then reruns the remote integer certificate, the local order and Hessian certificate, and the competing-partition subdivision. It also compares an integer-oracle value with a direct Arb calculation at a selected point, and verifies that deliberately raising the support floor by 0.1 is rejected. The checker uses `python-flint` 0.9.0 with FLINT 3.6.0.

Separate review controls establish exact root-bracket and local-coordinate endpoint containment. They also compare the reduced oracle with all 9! first-stage permutations at four gap pairs. Those binary64 comparisons are implementation diagnostics; the continuum theorem depends on the analytic reduction, semiconcavity, and the enclosed finite certificate. The package’s reproduction driver verifies the frozen input hashes and runs the checks in clean copies.

The comparison with Ref. [1], Appendix B.1, is a comparison with its qubit swap-optimality result. The theorem here addresses the stated two-qutrit task, including its particular target, gap domain, randomization convention, and work budget. It does not prove general swap optimality for arbitrary qutrit schedules, arbitrary dimensions, or other targets. Its lower bound also gives no claim of a coherence advantage: all allowed coherent protocols are bounded, and full swaps attain the bound. The proof and code have undergone separate-agent internal review within one model family. External expert review and comprehensive novelty assessment remain pending.

AI-assisted research

This work was developed in an author-directed, AI-assisted research workflow. AI systems assisted with mathematical derivation, literature checks, computational verification, and manuscript drafting. The internal reviews and separate implementations described above were produced within this workflow and do not constitute external expert review.

References

- [1] P. Taranto, P. Lipka-Bartosik, N. A. Rodríguez-Briones, M. Perarnau-Llobet, N. Friis, M. Huber, and P. Bakhshinezhad, *Efficiently Cooling Quantum Systems with Finite Resources: Insights from Thermodynamic Geometry*, Physical Review Letters **134**, 070401 (2025). [doi:10.1103/PhysRevLett.134.070401](https://doi.org/10.1103/PhysRevLett.134.070401); [arXiv:2404.06649v2](https://arxiv.org/abs/2404.06649v2).
- [2] F. Johansson, *Arb: Efficient Arbitrary-Precision Midpoint-Radius Interval Arithmetic*, IEEE Transactions on Computers **66**(8), 1281–1292 (2017). [doi:10.1109/TC.2017.2690633](https://doi.org/10.1109/TC.2017.2690633); [arXiv:1611.02831](https://arxiv.org/abs/1611.02831).