From Parisi full support to deterministic QAOA circuits

A long reasoning record

19 November 2026. Mathematical sources: OpenAI's full-support manuscript of 27 September 2026 and QAOA manuscript of 25 September 2026, at the pinned public revision.

AI assistance. Will Cook directed the work. AI agents based on large language models produced and revised this manuscript.

OpenAI's full-support manuscript proves that an admissible integrable minimizer of the zero-temperature Parisi functional for the pure zero-field Sherrington–Kirkpatrick model has Stieltjes support equal to every overlap scale in [0,1).

The proof turns a hypothetical support gap into a constant interval for the order parameter. Variational contact at the ends of that interval then conflicts with a strict curvature inequality inside it. Along the auxiliary diffusion, the spatial gradient of the value function is a bounded martingale. Full support makes its second moment equal to time, and its terminal correlation with Brownian motion gives the ground-state value.

The manuscript's further argument gives a smooth Stieltjes density, strictly positive on $(0, 1)$, and no atom at zero. Sections 9.1–9.2 derive this stronger ordinary conclusion separately from the selected formal support statement.

We reconstruct these arguments and follow the finite Gaussian approximation into the separate QAOA manuscript, which claims that finite deterministic circuits approach the SK optimum when system size tends to infinity before circuit depth increases. Its circuit construction needs four further arguments: approximation, synthesis, seed removal and transfer. The two formal statements concern full support and the scalar value consequences, and neither states QAOA circuit convergence. Sections 12–13 describe what they cover. See OpenAI (2026a), §§1–2 and OpenAI (2026b), Theorem 1.1.

The support and scalar arguments assume real analysis, stochastic calculus and variational minimization. Section 8 also uses Hilbert-space tensor products, symmetric Fock space, unbounded operators and quantum control. Its weighted-domain and control estimates are further analytic inputs from the QAOA manuscript, and the text gives each one's precise use and proof destination.

The companion short paper concentrates on full support and the strictly positive interior density. The exposition introduces the support argument through examples. Throughout, OpenAI (2026a) and OpenAI (2026b) name the two manuscripts.

Sections 1–8 give the continuous proof: full support, the terminal value, finite Gaussian correlations and, for each accuracy, one ordinary QAOA word. Section 9 derives the density and finite-disorder consequences. Section 10 gives an independent scalar comparison and a heat-family lower bound. Section 11 treats alternative circuit constructions and further limits. Section 12 gives the exact formal statements and compares them with the model, and Section 13 lists the sources and the finite calculation.

1. The SK model and the Parisi variational problem

For independent standard Gaussian couplings J_{ij}, put

$$H_n(\sigma) = \frac{1}{\sqrt{n}} \sum_{i<j} J_{ij}\sigma_i\sigma_j, \qquad \sigma \in \{-1,1\}^n, \qquad P_* = \lim_{n\to\infty} \frac{1}{n} \mathbb{E} \max_{\sigma} H_n(\sigma).$$

For overlap $R_{12} = n^{-1}\sum_i \sigma_i^1\sigma_i^2$, the covariance is $nR_{12}^2/2 - 1/2$. Thus the convention is $\xi(t) = t^2/2$, with $\xi'' = 1$. A centered Gaussian of variance $1/2$, common to all configurations, removes the final constant and leaves the expected maximum unchanged. Every normalization below uses this convention.

Let $\mathcal{U}$ consist of nonnegative, nondecreasing, right-continuous functions $\gamma : [0,1) \to [0,\infty)$ with $\int_0^1 \gamma(t)\,dt < \infty$. Their Stieltjes measures satisfy $\mu_\gamma([0,t]) = \gamma(t)$, including $\mu_\gamma(\{0\}) = \gamma(0)$. They are locally finite. They need not be probability measures, and their total mass may be infinite. Support is taken relative to [0,1).

Let B be standard Brownian motion. The control value and Parisi functional are

$$\Phi_\gamma(t,x) = \sup_{|\alpha|\le 1} \mathbb{E}\left[\left|x + B_1 - B_t + \int_t^1 \gamma(s)\alpha_s\,ds\right| - \frac{1}{2}\int_t^1 \gamma(s)\alpha_s^2\,ds\right],$$

$$\mathcal{P}(\gamma) = \Phi_\gamma(0,0) - \frac{1}{2}\int_0^1 t\gamma(t)\,dt.$$

Controls are progressively measurable for Brownian increments after t. The control formula fixes the solution, and the PDE is $\Phi_t = -\frac{1}{2}(\Phi_{xx} + \gamma\Phi_x^2)$, with $\Phi(1,x) = |x|$. For a general admissible coefficient, the solution and its spatial derivatives are locally absolutely continuous in time on compact strips, and the time equation holds almost everywhere.

Auffinger and Chen proved the zero-temperature Parisi formula and its attainment, which identify P_* with $\min_{\gamma\in\mathcal{U}} \mathcal{P}(\gamma)$. The support theorem itself takes a minimizer as given and proves a property of it. See OpenAI (2026a), §§1–2 and Auffinger and Chen (2017), Theorem 1 and proof of Lemma 3.

Write $u = \Phi_x$, $a = \Phi_{xx}$, and $dX_t = \gamma(t)u(t,X_t)\,dt + dB_t$, $X_0 = 0$. Write $q(t) = \mathbb{E}[u(t,X_t)^2]$ for the second moment of this gradient along X.

The first result assumes an admissible minimizer. It constructs its Stieltjes measure and an associated diffusion, proves $\operatorname{supp}\mu_\gamma = [0,1)$ and strict increase of γ, and obtains $\mathbb{E}u(t,X_t)^2 = t$ and $\mathbb{E}a(t,X_t)^2 = 1$ for every associated diffusion. The selected formal statement has exactly this conditional scope and asserts no smooth density.

Its Lean implementation at the fixed source revision implements the separately stated formal challenge, and the Comparator configuration identifies the intended comparison. The evidence kept with this draft does not establish a successful research-target replay.

The second result concerns the scalar value. A bounded martingale ends at a spin-valued random variable whose Brownian correlation represents P_*. Finite Gaussian coefficients approximate that correlation at each fixed $T < 1$, and then $T \uparrow 1$. Its formal statement is separate and keeps the minimizer and diffusion hypotheses where they are needed.

The QAOA theorem goes further. For each requested accuracy it selects one finite deterministic circuit, independent of size and disorder, whose limiting expected energy approaches P_*. Neither

formal statement asserts that circuit conclusion. The scalar clauses are in a separate Lean implementation, with their own formal challenge and Comparator configuration. Those links identify the selected statements and give no completed replay receipt for either research target.

2. Analytic preliminaries and prior work

2.1 Outline of the proof

We begin with regularity on strips bounded away from terminal time, which makes the diffusion X and the gradient moment q available, so that a first variation can show where the minimizer is in contact. Then, on a constant interval, the shape inequalities and the rational certificate give the strict curvature needed to contradict that contact. The zero-interval case uses a separate heat argument, and the possible terminal plateau needs its own exclusion. Together these steps give full support and the two moment equalities.

The value argument then uses bounded-martingale convergence and the attained Parisi formula to express P_* as a terminal Brownian correlation and as an integral of expected curvature. Section 6 defines the terminal martingale variable and proves these equalities. The fixed-T Gaussian approximation recovers this integral through Euler coupling, orthogonal normalization and mean-preserving rounding.

The QAOA theorem needs four more steps: rooted energy identification, classical synthesis, removal of temporary seeds and transfer to fixed angles. The main route is §§3–5, then §§6–7, then §§8.1–8.6, and §12 compares the two formal interfaces. The original analytic details are in OpenAI (2026a), §§2–7, and the circuit construction is in OpenAI (2026b), §§2–6.

The density theorem in §§9.1–9.2 follows from support and the PDE, and neither the scalar argument nor the circuit construction uses it. The positive-temperature route in §11.1 independently supplies the same bounded classical input used for synthesis. The probability consequence in §11.2 requires a separate fixed-word second-moment theorem. These branches show why the two selected formal statements fall short of a formal proof of the complete circuit theorem.

The run-in paragraphs below separate the premises of the support proof from its consequences and locate the detailed arguments behind the short paper's support and density theorem.

The diffusion and moment identity. For any admissible γ, the compact-strip estimates justify defining X, q and computing $q' = \mathbb{E}a^2$; see §2.4 and §3.1.

Variational contact. Minimality, admissible mass addition and compact mass removal give $G \geq 0$, its vanishing on support and the contact slopes. The steps are in §3.2, §3.3 and §3.4.

Strict curvature. This step uses only admissibility and a positive constant c on an interval. The shape comparisons, the algebra and the limiting argument are in §4.3, §4.4 and §4.5.

Exclusion of gaps. We combine contact with the positive, zero or terminal argument, as the case requires. Each case has its own destination: terminal, bounded and initial gaps.

Positive density. Full support comes first. Differentiated moments then give smoothness in §9.1, and a local comparison gives strict positivity in §9.2.

Value and circuits. We keep the same minimizer and diffusion through the value identity in §6 and the finite Gaussian construction in §7. The circuit argument in §8 requires additional theorems.

2.2 Compact-strip regularity and the terminal bound

Proposition 2.1 (compact-strip regularity). For $\gamma \in \mathcal{U}$ and $0 \le T < 1$, the value has bounded, jointly continuous spatial derivatives of every positive finite order on $[0,T] \times \mathbb{R}$. If admissible coefficients converge in L^1 and their smooth, even, convex, one-Lipschitz terminal values converge uniformly to $|x|$, these spatial derivatives converge uniformly on each such strip. The bounds depend on the derivative order, the fixed strip and a common coefficient mass bound. Time derivatives need only exist almost everywhere across coefficient jumps. On an open constant interval, the heat formula supplies the additional time smoothness used in §4. The terminal estimate is

$$|x| \le \Phi_\gamma(t,x) \le |x| + \sqrt{2(1-t)/\pi} + \frac{1}{2}\int_t^1 \gamma(s)\,ds.$$

It controls the approach to $|x|$ without differentiating the cusp. The strip estimates justify the interior calculations and finite meshes, while the terminal bound and integrability justify the later endpoint limits. We keep the two roles separate, since each depends on the region and the limit.

On a fixed $[0,T]$ with $T < 1$, bounded spatial derivatives and locally absolutely continuous time dependence justify the Itô identities and Euler coupling. Across a jump of γ, the spatial derivative moments remain continuous and the time equations hold almost everywhere. This is enough for the contact slopes and the endpoint chord comparison. Inside a constant interval, classical time differentiation permits the strict-curvature calculation.

As $T \uparrow 1$ the controls change. The displayed value bound, integrable coefficient tails and martingale convergence justify the first variation and the terminal value, so the argument uses no differentiated estimate at the terminal cusp.

These are OpenAI (2026a), Lemma 2.1 and its diffusion consequences, and §2.4 reconstructs their uniformity.

The hypotheses also have distinct uses. Zero field supplies evenness, $u(0,0) = 0$ and reflection in the shape argument. The convention $\xi'' = 1$ fixes the curvature coefficients. Monotonicity gives bounded coefficients before time one and nonnegative coefficient jumps, and it turns a support gap into a constant interval. Integrability controls terminal tails and Fubini interchanges, while local finiteness of the Stieltjes measure is enough for compact mass removal.

Minimality first enters through the variational inequality in §3. The constant-interval curvature estimate applies to any admissible coefficient, minimiser or not. OpenAI (2026a), §§1–2, OpenAI (2026a), §2 and OpenAI (2026a), §6.

2.3 Prior results and their roles

The ground-state formula of Auffinger and Chen supplies two facts with different uses. Their Theorem 1 identifies the ground-state limit with an infimum over $\mathcal{U}$. To choose a coefficient attaining that infimum, we also need their construction in the proof of Lemma 3. In the cited arXiv v2, the theorem is on p. 3, the compactness construction on pp. 7–9 and the lower-bound proof on pp. 11–12.

The support proof is conditional on a minimizer. A reader who grants that hypothesis can continue directly with the compact-strip estimates in §2.4. The construction below shows how attainment supplies the minimizer later used in the value and circuit arguments.

Here is the part of that construction which supplies the witness. Let α_β be a positive-temperature Parisi distribution function. The source bounds $\beta\alpha_\beta(t)$ uniformly on each interval ending before one. Helly selection and a diagonal subsequence produce a nonnegative, nondecreasing right-continuous limit γ_0. In the SK normalization, the source's weighted integral bound also controls $\int_0^1 \beta\alpha_\beta(t)\,dt$. For every nonnegative nondecreasing h, splitting at $1/2$ gives $\int_0^1 h(t)\,dt \le 4\int_0^1 th(t)\,dt$: the first half is bounded by $h(1/2)/2$, and on the second half $t \ge 1/2$. Fatou's lemma then gives $\gamma_0 \in \mathcal{U}$.

Mass can concentrate at time one during this limit, so compact convergence alone does not identify the limiting control problem. Write the limiting measure of the time coefficients as

$$\nu_0(dt) = \gamma_0(t)\,dt + m\delta_1(dt), \qquad m \ge 0.$$

For an arbitrary admissible control, let S denote its terminal state with coefficient γ_0. To see why the terminal choice is admissible, let the original control be α and put $S_s = B_s + \int_0^s \gamma_0(r)\alpha_r\,dr$. For a cutoff $\tau < 1$ use α_s before τ and $g_k(S_s)$ afterwards, where g_k is a continuous function bounded by one and converging pointwise to the sign with $\operatorname{sgn}(0) = 1$. This control uses only information available at time s. At fixed τ, k, we first take the temperature limit and then let $\tau \uparrow 1$ and $k \to \infty$. The continuous terminal evaluation has thereby approached $\operatorname{sgn}(S)$ without anticipating a future state. The endpoint atom then contributes

$$|S + m\operatorname{sgn}(S)| - |S| = m$$

to the reward and

$$\tfrac{1}{2}m(\operatorname{sgn}(S)^2 + 1) = m$$

to the combined control and Parisi penalties. These contributions cancel. Taking the supremum over the remaining controls proves $P_* \ge \mathcal{P}(\gamma_0)$. The upper bound proved earlier in that paper gives $P_* \le \inf_{\gamma\in\mathcal{U}} \mathcal{P}(\gamma)$. Therefore

$$P_* \le \inf_{\gamma\in\mathcal{U}} \mathcal{P}(\gamma) \le \mathcal{P}(\gamma_0) \le P_*.$$

Thus γ_0 attains the infimum. The cancellation is a step in the imported existence proof and adds no assumption about the Stieltjes measure $d\gamma_0$. In particular, ν_0 here is the limit of coefficient-times-Lebesgue measures, distinct from the order parameter's Stieltjes measure, and the two must not be identified. Auffinger and Chen, arXiv:1606.05335v2, pp. 7–12.

The cited theorem is stated for the almost-sure normalized maximum. Its expected-value formulation used here follows from Gaussian concentration. For $L_n = n^{-1}\max_\sigma H_n(\sigma)$, the Lipschitz constant in the Gaussian couplings is $\sqrt{(n-1)/(2n^2)}$. Hence $L_n - \mathbb{E}L_n \to 0$ in probability. Since the almost-sure limit is the deterministic Parisi value, the deterministic sequence $\mathbb{E}L_n$ has the same limit. The full-support theorem below remains conditional on a minimizer, and neither of the selected formal interfaces constructs one. OpenAI full-support manuscript, Introduction, ground-state formula and its footnote.

Auffinger, Chen and Zeng proved infinite support, which still allows an interior gap. Their terminal-plateau observation anticipates one endpoint obstruction, and the argument below supplies the quantitative boundary-layer proof used here.

The full-support manuscript develops a rational coercivity route and credits Chen's 2026 theorem for the same full-support conclusion and smooth absolute continuity. Chen's Proposition 4.2 has the conditional crossing form $q''(t) = 0 \Rightarrow q'''(t) > 0$. The cited rational certificate gives $q'''(t) > 0$ throughout a positive constant interval. These are different sufficient mechanisms, and this record

makes no priority claim for full support. Auffinger, Chen and Zeng (2020), Theorem 1 and Remark 7; Chen (2026), Theorem 1.3 and Proposition 4.2; OpenAI (2026a), §§1–2.

The transformed density comparisons also have antecedents in Lopatto's positive-temperature analysis and Chen's zero-temperature analysis. The rational coercive identity is the step that distinguishes the full-support proof.

Zhou's positive-temperature interval-support result applies near the critical temperature, whereas Lopatto's cited theorem covers the low-temperature zero-field phase. Neither transfers full Stieltjes support to zero temperature on its own. For the minimizers, $\beta\mu_\beta$ converges vaguely on $[0,1)$ to the zero-temperature Stieltjes measure, while the unscaled Parisi probability measures converge weakly on $[0,1]$ to δ_1. These are different measures and topologies, and convergence without an interval-specific lower bound can lose support information. The direct zero-temperature gap proof below is therefore essential. Chen (2026), Remarks 1.4–1.5; OpenAI (2026a), §§1–2.

The coefficient classes in this comparison can be translated explicitly. In the positive-temperature convention, write

$$U_t = -\frac{\beta^2}{2}(U_{yy} + \alpha_\nu U_y^2), \qquad U(1,y) = \log\cosh y.$$

Set $\Phi_\beta(t,x) = \beta^{-1}U(t,\beta x)$ and $\gamma_\beta = \beta\alpha_\nu$. Direct differentiation gives

$$(\Phi_\beta)_t = -\tfrac{1}{2}((\Phi_\beta)_{xx} + \gamma_\beta(\Phi_\beta)_x^2), \qquad \Phi_\beta(1,x) = \beta^{-1}\log\cosh(\beta x).$$

Dividing the associated diffusion by β likewise gives unit Brownian noise and drift $\gamma_\beta(\Phi_\beta)_x$. Conversely, every bounded nonnegative, nondecreasing, right-continuous coefficient on $[0,1)$ has this form: choose $\beta > 0$ at least its supremum, take distribution function $\alpha_\nu = \gamma/\beta$ below one, and put the remaining probability mass at one. This identifies coefficient classes, not minimizers at different temperatures. The terminal values differ from $|x|$ by at most $(\log 2)/\beta$, but passing a strict inequality to the cusp and to coefficients unbounded near one still needs uniform estimates. A qualitative crossing statement alone supplies no positive limiting fourth-moment margin. Lopatto, arXiv:2607.11756v4, equation (1.2), §2 and Proposition 5.1.

Two related works sharpen the comparison without entering the pinned September proof. Lopatto, arXiv:2607.11756v4, Proposition 5.1 and (5.8)–Corollary 5.6, proves a conditional curvature crossing on a positive constant interval. Its integrating factor makes a transformed curvature derivative strictly increasing, so the associated contact derivative decreases and then increases. It does not assert an unconditional positive third derivative.

Lopatto, arXiv:2610.00127v1, Theorem 2.3 and §8, works at positive temperature and positive external field. Its comparison requires one-sided bounds on the coefficient before and after the chosen time, a moment inequality and a stationarity equality. Several of these conditions are automatic in its density application: monotonicity supplies the one-sided bounds at a continuity point, contact on a full-support interval makes the scalar inequality an equality, and the quotient formula for the coefficient supplies stationarity. Its moment bootstrap then gives smooth, strictly positive density on a nontrivial support interval, and a singleton support is a separate case.

Appendix B, Theorem B.1 and equation (B.1), also prove strict positivity in the zero-field positive-temperature setting, where the extra noncentered-law condition is unnecessary.

These results are therefore close analytic antecedents. The distinction used below is the unconditional quantitative fourth-moment margin for the comparison coefficient and its controlled passage to the zero-temperature cusp.

Here is an explicit example of the information lost in such a limit. It concerns admissible measures, not Parisi minimizers. Put $g = \mathbf{1}_{[0,1/3]\cup[2/3,1)}$ and $q_n = 1 - n^{-2}$, and for $n \geq 2$ define

$$\mu_n(dt) = \frac{g(t) + n^{-1}}{n} \mathbf{1}_{[0,q_n]}(t)\, dt + \left(1 - \frac{1}{n}\int_0^{q_n} (g(t) + n^{-1})\, dt\right) \delta_{q_n}(dt).$$

The continuous part has mass at most $(2/3 + 1/n)/n < 1$, so these are probability measures with a positive endpoint atom. Their density is positive throughout $[0, q_n]$, giving $\operatorname{supp} \mu_n = [0, q_n]$. Their continuous mass tends to zero and $q_n \to 1$, hence $\mu_n \Rightarrow \delta_1$ weakly on $[0, 1]$. For a compactly supported continuous test function on $[0, 1)$, the endpoint atom eventually lies outside its support. Consequently $n\mu_n \to g(t)\, dt$ vaguely there, although the limit vanishes on $(1/3, 2/3)$. Even the cumulative coefficients converge in the topology used for the PDE: with $\gamma_n(t) = n\mu_n([0, t])$ and $\gamma(t) = \int_0^t g(s)\, ds$, their difference is t/n below q_n, and $\gamma_n = n$ on the terminal strip of length n^{-2}. Thus

$$\|\gamma_n - \gamma\|_{L^1(0,1)} \leq \frac{1}{2n} + \frac{1}{n} \longrightarrow 0.$$

Every γ_n is admissible, but $d\gamma$ has an interior gap. Minimality and the gap argument supply information that these convergences alone do not. An atom added at time one is likewise invisible to the coefficient on $[0, 1)$ and hence to its PDE and variational functional, so the zero-temperature problem selects no endpoint mass.

Control and gradient-martingale methods have antecedents in Auffinger–Chen and Jagannath–Tobasco. A finite-temperature theorem with smooth terminal value $\log\cosh$ does not alone establish the result with terminal cusp $|x|$. The required compact-strip and terminal estimates are supplied in OpenAI (2026a), §2; Chen–Handschy–Lerman supplies earlier zero-temperature regularity, control and uniqueness results (Theorem 4 for uniqueness). This is distinct from the earlier finite-temperature uniqueness result of Auffinger and Chen. Montanari and El Alaoui–Montanari–Sellke inform normalised discretisation and optimisation. The particular shifted Gaussian coefficient identity used here is proved in OpenAI (2026a), §7.

2.4 Recovering the compact-strip estimates

The curvature argument needs estimates uniform over the approximating coefficients. We derive them from the control formula, with every bound taken on a fixed strip ending before time one.

For finite-step coefficients and a smooth, even, convex, one-Lipschitz terminal function ψ, the evolution through a step of height c is

$$\Phi(t, x) = \begin{cases} c^{-1} \log P_{b-t}(e^{c\Phi(b,\cdot)})(x), & c > 0, \\ P_{b-t}\Phi(b, \cdot)(x), & c = 0, \end{cases}$$

where P_h denotes convolution with a centred Gaussian of variance h. The heat formula preserves evenness, convexity and the spatial Lipschitz bound. Along a controlled path $dY_s = \gamma(s)\alpha_s\, ds + dB_s$, the drift of the value minus the accumulated control cost is $-\gamma(s)(\alpha_s - \Phi_x(s, Y_s))^2/2$. It is nonpositive, and is zero for the feedback control. This proves the stated control formula for the smooth step problem, including attainability by its feedback diffusion.

Using the same control in two such problems and then taking suprema gives

$$\|\Phi_{\gamma,\psi} - \Phi_{\tilde\gamma,\tilde\psi}\|_\infty \leq \|\psi - \tilde\psi\|_\infty + \frac{3}{2}\|\gamma - \tilde\gamma\|_{L^1(0,1)}.$$

The drift part costs at most the coefficient difference in L^1, by the terminal Lipschitz bound, and the quadratic cost contributes another half of that difference. The supremum norm is of the difference of the values, which is bounded even though each value grows linearly in x. With $\psi_M(x) = M^{-1}\log(2\cosh Mx)$, $0 \le \psi_M - |x| \le (\log 2)/M$. Choose simultaneously $\gamma_m \to \gamma$ in L^1 and $M_m \to \infty$. These are two distinct requirements: $M_m \ge \|\gamma_m\|_\infty$ alone does not force terminal convergence. This compatible approximation determines a unique uniform limiting value. Its terminal bound, given in §2.2, also follows directly: choose the zero control for the lower bound, and use $|\alpha| - \alpha^2/2 \le 1/2$ for the upper bound.

To obtain spatial derivatives, choose $T < b < 1$. Monotonicity gives the uniform coefficient estimate

$$\gamma(t) \le \frac{\|\gamma\|_1}{1-b} \quad (0 \le t \le b).$$

An L^1-convergent family thus has a common coefficient bound Γ on this larger strip. For its smooth step approximants, put $U = \Phi_x$. The differentiated mild equation is

$$U(t) = P_{b-t}U(b) + \int_t^b \gamma(s)P_{s-t}[U(s)U_x(s)]\,ds.$$

Since $\|U\|_\infty \le 1$ and $\|\partial_x P_h F\|_\infty \le Ch^{-1/2}\|F\|_\infty$,

$$\|U_x(t)\|_\infty \le \frac{C}{\sqrt{b-t}} + C\Gamma\int_t^b \frac{\|U_x(s)\|_\infty}{\sqrt{s-t}}\,ds.$$

Substituting this inequality once into its own integral uses $\int_t^s ((r-t)(s-r))^{-1/2}\,dr = \pi$. It reduces the estimate to an ordinary backward Gronwall inequality with integrable forcing. This bounds U_x uniformly on $[0,T]$ without any terminal derivative bound. Repeating on successively smaller strips bounds each higher spatial derivative: put one derivative on the heat kernel and the rest on UU_x, and only the term $U\partial_x^m U$ contains the new highest derivative. The remaining products have already been bounded.

Uniform convergence of the values then yields convergence of every positive spatial derivative. For a bounded twice differentiable difference F with bounded second derivative, Taylor's formula gives

$$\|F'\|_\infty \le \frac{2\|F\|_\infty}{h} + \frac{h}{2}\|F''\|_\infty \qquad (h > 0).$$

First choose h small using the common second-derivative bound, then make $\|F\|_\infty$ small. Iterate on derivative differences. Passing to the integral form of the differentiated PDE proves local absolute continuity in time and its almost-everywhere equations. No classical time derivative at a coefficient jump, and no derivative of $|x|$ at its cusp, is being assumed.

Finally, couple all feedback diffusions with the same Brownian motion. On $[0,T]$, let L be a common spatial Lipschitz bound for their drifts. The difference of the two integral equations gives, for example,

$$\sup_{s\le T}|X_s^n - X_s| \le e^{LT}\left(\|\gamma_n - \gamma\|_1 + \|\gamma_n\|_1\|u_n - u\|_{\infty,[0,T]\times\mathbb{R}}\right) \longrightarrow 0.$$

This is a deterministic bound under the coupling. The derivative bounds one order higher ensure uniform continuity when derivatives are evaluated at these moving points. Every fixed polynomial in the bounded derivatives therefore converges in expectation. This is precisely the convergence needed for the integrated curvature expression in §4, and none of the forward density scores or their quotients need converge. OpenAI (2026a), Lemmas 2.1–2.3.

2.5 The order of the approximation limits

Each limit below fixes its background objects first, so that one symbol for a small error never hides a change of object. A single admissible minimiser and its associated diffusion stay fixed throughout the scalar construction.

Compact-strip regularity. Fix a derivative order $k \geq 1$, a cutoff $T < 1$ and a common coefficient L^1 bound. Coefficients converge in L^1 within that bound, and spatial derivatives converge uniformly on the strip.

Shape inequalities on a plateau. Fix the height $c > 0$ and a closed interior neighbourhood. Choose step coefficients equal to c there and smooth terminal values with $M_m \geq \|\gamma_m\|_\infty$ and $M_m \to \infty$. The coefficient and terminal approximations may be chosen as one compatible sequence.

First variation. Fix γ, a feasible $\eta \in L^1$ and initially $\epsilon > 0$. Remove the terminal cutoff before letting $\epsilon \downarrow 0$. The terminal remainder divided by ϵ is bounded by an L^1 tail of η.

Terminal value. Keep the chosen minimiser and diffusion fixed. Bounded martingale convergence and integrability of γ justify $t \uparrow 1$ without any terminal derivative bound.

Finite Gaussian coefficients. Fix γ, X and $T < 1$. Refine N using the fixed-strip bounds, then control the remaining value tail by $1 - T$.

For the plateau transfer, tail estimates for a fixed approximant may depend on that approximant. The passage to the limit uses only the integrated polynomial in the scaled second, third and fourth spatial derivatives of the value function, together with the common strip bounds. The auxiliary density scores need no uniform bounds.

Sections 3 and 4 give the variation and plateau-transfer arguments, and §7 gives a complete scalar error budget. The further choices used to produce a circuit appear in §8.6. OpenAI (2026a), Lemma 2.1, Lemma 6.1, Proposition 5.1 and Proposition 7.2.

3. From minimality to contact

Proposition 3.1 (variational contact). For an admissible minimizer, the potential $G(s) = \int_s^1 (q(t) - t)\, dt$ is nonnegative on $[0, 1)$ and vanishes on $S = \operatorname{supp} \mu_\gamma$. At every $s \in S$, $q(s) = s$ and $q'(s) \leq 1$, with a right derivative at zero.

The minimisation takes place over nonnegative Stieltjes measures with an integrable cumulative function, and their total mass is free. We first derive the variation of the functional, then use changes of measure which preserve this admissible class. The result is contact on the *existing* support. Gaps are excluded afterwards, in §5.

3.1 The moment function before minimisation

Fix any $\gamma \in \mathcal{U}$, without yet assuming that it minimises $\mathcal{P}$. Compact-strip regularity and Itô's formula give

$$u(t, X_t) = \int_0^t a(s, X_s)\, dB_s, \qquad q(t) := \mathbb{E}u(t, X_t)^2 = \int_0^t \mathbb{E}a(s, X_s)^2\, ds.$$

The initial term vanishes because zero field makes Φ even and $u(0, 0) = 0$. Thus $q(0) = 0$, $0 \leq q(t) \leq 1$, and

$$q'(t) = \mathbb{E}a(t, X_t)^2 \qquad (0 \leq t < 1),$$

with a right derivative at zero. The integrand is continuous on each compact time strip, so $q \in C^1[0,1)$. These facts hold for every admissible coefficient, and the identity $q(t) = t$ is still to be proved. The source proves them by smooth approximation on $[0,T]$, $T < 1$, and never differentiates the terminal cusp $|x|$ at time one. OpenAI (2026a), Lemmas 2.1–2.2.

3.2 The feasible right derivative

Let $\eta \in L^1(0,1)$ and suppose that $\gamma + \varepsilon\eta \in \mathcal{U}$ for $0 \le \varepsilon \le \varepsilon_0$, where $\varepsilon_0 > 0$. This is a feasibility hypothesis: an arbitrary signed bump in γ need not preserve monotonicity. Since $\mathcal{U}$ is convex, it is enough to check that the endpoint $\gamma + \varepsilon_0\eta$ is admissible. The formula below is the zero-field, $\xi'' = 1$ specialization of Chen, Handschy and Lerman, arXiv:1609.04368v4, Proposition 4. Their chord $\gamma_\theta = (1-\theta)\gamma + \theta\gamma_1$ gives this direction by taking $\gamma_1 = \gamma + \varepsilon_0\eta$ and $\theta = \varepsilon/\varepsilon_0$. Their class $\mathcal{U}$ permits unnormalized cumulative measures with finite cumulative integral, so this change of parameter also covers the scaling direction $-\gamma$. We reconstruct the detailed PDE proof from OpenAI's Lemma 6.1.

For $0 < \varepsilon \le \varepsilon_0$, put $\beta = \gamma + \varepsilon\eta$, $\tilde{u} = \Phi_{\beta,x}$ and $D = \Phi_\beta - \Phi_\gamma$. Subtracting the two PDEs gives

$$D_t = -\frac{1}{2}D_{xx} - \frac{1}{2}\beta(\tilde{u} + u)D_x - \frac{\varepsilon}{2}\eta u^2.$$

The choice of diffusion

$$dY_t^\varepsilon = \frac{1}{2}\beta(t)(\tilde{u} + u)(t, Y_t^\varepsilon)\,dt + dB_t, \qquad Y_0^\varepsilon = 0,$$

cancels the first-order spatial term when Itô's formula is applied to $D(t, Y_t^\varepsilon)$. Its drift has absolute value at most $\beta(t)$, which is integrable. On $[0,T]$ the resulting identity is

$$D(0,0) = \mathbb{E}D(T, Y_T^\varepsilon) + \frac{\varepsilon}{2}\int_0^T \eta(t)\mathbb{E}[u(t, Y_t^\varepsilon)^2]\,dt.$$

At coefficient jumps the time derivatives need only exist almost everywhere. The source obtains this identity by approximating both coefficients by monotone step functions with the same smooth terminal datum, and then passing to the limit using the compact-strip estimates.

The control representation, applied with the same control on each side, gives the time-local difference bound

$$\sup_x |D(T,x)| \le \frac{3\varepsilon}{2}\int_T^1 |\eta(t)|\,dt.$$

Hence $T \uparrow 1$, at fixed ε, removes the terminal term even after division by ε. We obtain the exact finite-change identity

$$\frac{\Phi_{\gamma+\varepsilon\eta}(0,0) - \Phi_\gamma(0,0)}{\varepsilon} = \frac{1}{2}\int_0^1 \eta(t)\mathbb{E}[u(t, Y_t^\varepsilon)^2]\,dt.$$

As $\varepsilon \downarrow 0$, the coefficients converge in L^1 and have uniformly bounded L^1 norms. On each fixed strip, their gradients converge uniformly and have common spatial Lipschitz bounds. Coupling with the same Brownian motion and applying Gronwall's inequality therefore gives $Y^\varepsilon \to X$ uniformly on that strip. The integrand converges to $\eta(t)q(t)$ for each $t < 1$, and its absolute value is at most $|\eta(t)|$. Dominated convergence, followed by subtraction of the linear Parisi penalty, now yields

$$D_+\mathcal{P}(\gamma)[\eta] := \lim_{\varepsilon\downarrow 0}\frac{\mathcal{P}(\gamma+\varepsilon\eta) - \mathcal{P}(\gamma)}{\varepsilon} = \frac{1}{2}\int_0^1 \eta(t)(q(t) - t)\,dt.$$

If γ is a minimiser, this derivative is nonnegative for every feasible direction. It need not vanish: the minimisation has one-sided constraints. OpenAI (2026a), Lemma 6.1, equations (6.1)–(6.2).

3.3 Admissible changes of measure

Write $\mu = \mu_\gamma$ and $S = \operatorname{supp}\mu$, with support relative to $[0,1)$. Tonelli's theorem gives the useful identity

$$\int_0^1 \gamma(t)\,dt = \int_{[0,1)}\int_s^1 dt\,\mu(ds) = \int_{[0,1)} (1-s)\,\mu(ds) < \infty.$$

This is a weighted finiteness condition, and $\mu([0,1))$ may be infinite. In particular, mass added at one point need not be removed elsewhere, since no normalisation constrains the total.

Define the tail potential

$$G(s) = \int_s^1 (q(t)-t)\,dt \qquad (0 \le s < 1).$$

It is continuous, satisfies $|G(s)| \le 1-s$, and has $G'(s) = s - q(s)$ and $G''(s) = 1 - q'(s)$ on compact strips, with one-sided interpretations at zero. The following two competitors suffice.

Change to the Stieltjes measure	Direction in the cumulative function	Allowed parameter	Consequence at a minimiser
Add $\varepsilon\delta_s$, $0 \le s < 1$	$\eta(t) = \mathbf{1}_{[s,1)}(t)$	$\varepsilon \ge 0$	$G(s) \ge 0$
Remove $\varepsilon\nu$, where $\nu = \mu\vert_K$ and $K \subset [0,1)$ is compact	$\eta_K(t) = -\nu([0,t])$	$0 \le \varepsilon \le 1$	$\int_K G(s)\,\mu(ds) \le 0$

For atom addition, the new cumulative function is nonnegative, nondecreasing and right-continuous. Its integral increases by $\varepsilon(1-s)$, so it remains in $\mathcal{U}$. This includes $s = 0$: an atom there adds a constant to γ on all of $[0,1)$. The point 1 is outside the measure's domain and is not an additional competitor. The derivative formula gives $D_+\mathcal{P}(\gamma)[\mathbf{1}_{[s,1)}] = G(s)/2$, proving $G \ge 0$ everywhere below one.

For removal, choose $T < 1$ with $K \subset [0,T]$. Then $\nu(K) \le \gamma(T) < \infty$. The measure $\mu - \varepsilon\nu$ is nonnegative for $0 \le \varepsilon \le 1$, so its cumulative function is nonnegative, nondecreasing and right-continuous. It lies between zero and γ, hence is integrable, and $|\eta_K| \le \nu(K)$. Fubini is justified quantitatively by

$$\int_K\int_s^1 |q(t)-t|\,dt\,\mu(ds) \le \int_K (1-s)\,\mu(ds) \le \nu(K) < \infty.$$

It follows that

$$0 \le D_+\mathcal{P}(\gamma)[\eta_K] = -\frac{1}{2}\int_0^1 \nu([0,t])(q(t)-t)\,dt = -\frac{1}{2}\int_K G(s)\,\mu(ds).$$

Since $G \ge 0$, its integral on every compact K is zero. The compact intervals $[0,T]$ exhaust $[0,1)$, so $G = 0$ μ-almost everywhere. If a support point s had $G(s) > 0$, continuity would give a relative neighbourhood on which G stays positive. After restricting to a compact interval containing that

neighbourhood, the defining property of support would give it positive μ-mass, contradicting the zero integral. Therefore

$$G(s) = 0 \qquad (s \in S).$$

This step applies to atoms, diffuse mass and singular continuous mass. Removing individual atoms alone would not establish it. OpenAI (2026a), Lemma 6.2.

3.4 Contact values and endpoint slopes

At $s \in S \cap (0,1)$, the function G has an interior minimum of value zero. Thus $G'(s) = 0$ and $G''(s) \geq 0$, giving

$$q(s) = s, \qquad q'(s) \leq 1.$$

If $0 \in S$, the argument needs the zero-field initial condition in addition to the one-sided minimum. We have $G(0) = 0$ and $q(0) = 0$, so

$$G(h) = G(0) + \int_0^h (t - q(t))\, dt = \frac{1}{2}(1 - q'(0))h^2 + o(h^2).$$

Nonnegativity for $h > 0$ implies $q'(0) \leq 1$. This works whether or not zero is an atom. The equality $q(0) = 0$ comes from symmetry, rather than from differentiating a one-sided minimum. If $0 \notin S$, no condition $G(0) = 0$ has been obtained or used. OpenAI (2026a), Lemma 6.2.

We have now proved, without any no-gap assumption,

$$G \geq 0 \text{ on } [0,1), \qquad G = 0 \text{ on } S, \qquad q(s) = s \text{ and } q'(s) \leq 1 \text{ for } s \in S.$$

These are precisely the contact data used in §5. They make no assertion that every zero of G belongs to S, or that $q(t) = t$ at a point outside S. Terminal, interior and initial gaps remain to be excluded.

3.5 The reach of the first-order criterion

The gap proof needs only the contact data above, so a reader may continue to the curvature argument in §4. We give here an alternative contact proof and the converse optimality criterion.

There is also a global alternative to compact mass removal. Atom addition still gives $G \geq 0$. Now replace μ by $(1 - \varepsilon)\mu$, or equivalently γ by $(1 - \varepsilon)\gamma$, for $0 \leq \varepsilon \leq 1$. This preserves admissibility, and its direction $-\gamma$ lies in L^1, although it may be unbounded. The first variation in §3.2 therefore applies. The signed interchange is justified by

$$\int_{[0,1)} \int_s^1 |q(t) - t|\, dt\, \mu(ds) = \int_0^1 \gamma(t)|q(t) - t|\, dt \leq \int_0^1 \gamma(t)\, dt < \infty.$$

Consequently

$$0 \leq D_+\mathcal{P}(\gamma)[-\gamma] = -\frac{1}{2}\int_{[0,1)} G(s)\, \mu(ds).$$

Together with $G \geq 0$, this gives $\int G\, d\mu = 0$. Continuity then gives $G = 0$ on relative support, including zero, as in §3.3. Finite weighted mass is sufficient throughout, with no assumption of finite total mass or of a density for μ. This ordinary argument uses the same integrable-direction first variation as the local removal proof.

For the actual moment function associated with γ, the two potential conditions $G \geq 0$ and $G = 0$ on S are also sufficient for minimality. This ordinary deduction explains exactly how much variational information has been extracted. For another admissible coefficient β, let $\nu = \mu_\beta$. The weighted identity above and $|G(s)| \leq 1 - s$ make both measure integrals absolutely convergent, even when either measure has infinite total mass. Fubini gives

$$2D_+\mathcal{P}(\gamma)[\beta - \gamma] = \int_{[0,1)} G(s)\,\nu(ds) - \int_{[0,1)} G(s)\,\mu(ds) \geq 0.$$

Here the second integral is zero because μ is concentrated on its relative support. The control formula also shows that $\mathcal{P}$ is convex in its coefficient: for a fixed control, the terminal absolute value is convex in the coefficient and both penalties are linear, and taking the supremum over the same control class preserves convexity. Therefore

$$\mathcal{P}(\beta) - \mathcal{P}(\gamma) \geq D_+\mathcal{P}(\gamma)[\beta - \gamma] \geq 0.$$

This proves the converse. It agrees with the chord optimality criterion of Chen, Handschy and Lerman, arXiv:1609.04368v4, Proposition 4, printed p. 14. Chen's zero-temperature treatment obtains the same contact integral by testing 0 and 2γ and then using weighted Fubini Chen (2026), Proposition 3.2, equations (3.11)–(3.12). The compact-removal proof above keeps the local mass operation explicit.

The converse concerns G built from the PDE and diffusion of the same γ. It does not say that an arbitrary curve with these contact properties comes from a Parisi solution. Nor does it prove full support: the missing input is the strict curvature on an interval where the coefficient is constant, established next. The formal chord statement and the ordinary derivative bridge are distinguished in §12.1.

4. Strict curvature on a constant interval

The gap argument needs a property of the derivative q', which is distinct from convexity of the spatial function Φ. We establish the following analytic input before using minimality. Let $\gamma \in \mathcal{U}$ be constant with value c on an open interval $I \subset (0,1)$, and let X be its diffusion started at zero. Then

$$\begin{cases} q'''(t) \geq \dfrac{1}{1000c^2}\mathbb{E}[(c\Phi_{xx}(t,X_t))^4] > 0, & c > 0, \\ q''(t) = \mathbb{E}[\Phi_{xxx}(t,X_t)^2] > 0, & c = 0, \end{cases} \qquad t \in I.$$

Thus q' is strictly convex for $c > 0$ and strictly increasing for $c = 0$. The coefficient may jump outside I and may be unbounded near one, provided it is nonnegative, nondecreasing, right-continuous and integrable. The statement holds for every such coefficient, whatever its support and whether or not it minimises. The positive-time restriction matters when a diffusion density is used. Interval endpoints are handled later by continuity. OpenAI (2026a), Proposition 5.1.

The proof has two stages. In §§4.1–4.4 the coefficient is a positive monotone finite-step function with smooth terminal data. Shape comparisons and boundary estimates justify the polynomial correction and its zero expectation. In §4.5, approximants γ_n, Φ_n, X^n approach an arbitrary admissible coefficient. Only the integrated expression in spatial derivatives passes to the limit, and the auxiliary forward scores need not converge. Bounds depend on a fixed compact strip and a common coefficient mass bound. Finally, we justify the strictly positive right side for the limiting solution itself.

4.1 Why the uncorrected expression has no pointwise sign

On a positive constant interval put $r = c\Phi_x$, $v = r_x$, $w = r_{xx}$ and $z = r_{xxx}$. In forward time, $r_t = -\frac{1}{2}r_{xx} - rr_x$ and $dX_t = r(t, X_t)\,dt + dB_t$. Differentiating the PDE and applying Itô's formula gives

$$dr(t, X_t) = v(t, X_t)\,dB_t,$$
$$dv(t, X_t) = -v(t, X_t)^2\,dt + w(t, X_t)\,dB_t, \qquad dw(t, X_t) = -3v(t, X_t)w(t, X_t)\,dt + z(t, X_t)\,dB_t.$$

Since $q' = c^{-2}\mathbb{E}v^2$, the drift of v^2 yields $c^2q'' = \mathbb{E}(w^2 - 2v^3)$. The drifts of w^2 and v^3 are, respectively, $z^2 - 6vw^2$ and $3vw^2 - 3v^4$. Therefore

$$c^2q''' = \mathbb{E}\big[z^2 - 12vw^2 + 6v^4\big].$$

These are interior-time identities. On a smaller constant interval the heat formula gives classical time smoothness and bounded spatial derivatives of sufficiently high order. Hence the local martingale terms have zero mean, and the expectation differentiations are justified. The identities leave the sign of q'' undetermined. OpenAI (2026a), Lemma 2.3.

A concrete example shows that the last display cannot be read pointwise. For the smooth finite-step problem $c = M = 1$ on $[0, 1)$, with terminal value $\log(2\cosh x)$, the exact solution is

$$\Phi(t, x) = \log(2\cosh x) + \frac{1 - t}{2}, \qquad r = \tanh x.$$

Indeed, $\Phi_{xx} + \Phi_x^2 = 1$. Writing $R = r^2$ gives $v = 1 - R$, $w = -2rv$ and $z = -2v^2 + 4Rv$. At $R = 1/2$, $v = 1/2$, $w^2 = 1/2$ and $z = 1/2$, so

$$z^2 - 12vw^2 + 6v^4 = -\frac{19}{8} < 0.$$

This example lies in the smooth analytic class used below. Its diffusion density, obtained from the heat transition formula, is $f_t(x) = e^{-t/2}\cosh(x)p_t(x)$, and it integrates to one because $\mathbb{E}\cosh(B_t) = e^{t/2}$.

The example shows that the integration argument is necessary. It uses a smooth terminal value and makes no claim to be the zero-temperature minimiser. The negative value above is an exact elementary calculation, and the heat formula itself is OpenAI (2026a), Lemma 2.1.

The same pointwise obstruction occurs with the actual terminal datum $|x|$. For the admissible constant coefficient $\gamma_c \equiv c > 0$, write $\tau = 1 - t > 0$ and use the same heat formula:

$$\Phi_c(t, x) = c^{-1}\log \mathbb{E}e^{c|x + \sqrt{\tau}Z|}.$$

At $c = 0$ its continuous extension is the heat value, whose derivatives are

$$a_0 = \sqrt{\frac{2}{\pi\tau}}e^{-x^2/(2\tau)}, \qquad b_0 = -\frac{x}{\tau}a_0, \qquad d_0 = \left(\frac{x^2}{\tau^2} - \frac{1}{\tau}\right)a_0.$$

At $x = \sqrt{\tau}$, $d_0 = 0$ and $b_0^2 = a_0^2/\tau > 0$. For fixed t, x, Gaussian convolution may be differentiated in x through its kernel and expanded in c near zero, since exponential moments dominate every resulting polynomial factor. The apparent singularity of $c^{-1}\log \mathbb{E}e^{c|x+\sqrt{\tau}Z|}$ is removable, and its spatial derivatives through order four are smooth in c. Hence

$$d_c^2 - 12ca_c b_c^2 + 6c^2 a_c^4 = -\frac{12a_0^3}{\tau}c + O(c^2) < 0$$

for all sufficiently small positive c, at that fixed point. The sign follows from the constant-coefficient formula, in which all four derivatives come from one solution. The example asserts nothing about the minimality of γ_c and is consistent with the positive expectation proved below.

It shows that, even with the absolute-value terminal value, that expectation cannot be replaced by a pointwise inequality.

4.2 The smooth finite-step problem and its boundary control

Take $0 = t_0 < \cdots < t_N = 1$, $0 < c_1 \le \cdots \le c_N \le M$, and $\gamma(t) = c_i$ on $[t_{i-1}, t_i)$. The terminal value is $M^{-1}\log(2\cosh Mx)$. On a step of height c, set $l = c\Phi$ and $h = e^l$. Then $h_t = -\frac{1}{2}h_{xx}$. For $a < t$ in that step the diffusion transition density is

$$p_{t-a}(x-y)\frac{h(t,x)}{h(a,y)}, \qquad p_t(x) = (2\pi t)^{-1/2}e^{-x^2/(2t)}.$$

Its mass is one by $h(a,\cdot) = P_{t-a}h(t,\cdot)$. Factoring the positive-time density as $f = e^l g$ gives $g_t = \frac{1}{2}g_{xx}$ on each step. Put $s = -\partial_x \log g$. On the first step, $g(t,x) = p_t(x)e^{-l(0,0)}$ and $s = x/t$. At a forward jump $\delta = c_+ - c_- \ge 0$, continuity of f gives $g_+ = g_- e^{-\delta\Phi}$. The functions r and s are odd. OpenAI (2026a), §3, before Lemma 3.1.

The comparison arguments take place on unbounded spatial domains, so the boundary estimates must precede the signs. For these fixed smooth data, Gaussian convolution and the finite jump rules give, as $x \to +\infty$,

$$l(t,x) = cx + C(t) + O(e^{-\kappa x}),$$
$$g(t,x) = \exp\left(-\frac{x^2}{2t} + \beta(t)x + C_0(t)\right)(1 + O(e^{-\kappa x})).$$

The errors may be differentiated any fixed finite number of times, and the estimates are uniform on each closed step, with positive time bounded away from zero for the forward estimate. Their constants may depend on the step data.

For example, after completing the square in a Gaussian convolution, the error is an expectation at $Y_x = \lambda x + b + \sqrt{V}Z$ with λ bounded below by a positive constant. Split at $Y_x = \mathbb{E}Y_x/2$. The first part inherits the exponential tail. On the second, the Gaussian probability dominates the permitted exponential growth of the error and all its specified derivatives. This proves the differentiated estimates by finite induction through heat evolution and jumps. In particular, $s = x/t - \beta(t) + O(e^{-\kappa x})$, $s_x = 1/t + O(e^{-\kappa x})$ and $s_{xx} = O(e^{-\kappa x})$. OpenAI (2026a), Lemma 3.1.

The comparison principle used below is also explicit. If $U_\tau - \frac{1}{2}U_{xx} - bU_x - dU \ge 0$ on a compact cylinder, $d \le C$, the initial values are nonnegative and the lateral values are at least $-\varepsilon$, then

$U \geq -\varepsilon e^{(\max(C,0)+1)(\tau-\tau_0)}$. Multiply by the corresponding decreasing exponential and apply the minimum principle. Exhaustion is legitimate when the upper bound for d is independent of the cylinder and the lateral errors tend to zero. The first-order coefficient need only be bounded on each cylinder, which permits the $1/x$ coefficients that appear below. OpenAI (2026a), comparison preceding Lemma 3.2.

4.3 Backward and forward comparisons, including jumps

We work in backward time τ on a single step, where the equation is $r_\tau = \frac{1}{2}r_{xx} + rr_x$. The logarithmic heat formula preserves strict positivity of v: its second spatial derivative is the tilted mean of the old second derivative plus the tilted variance of the old first derivative. Initially $v = cM\operatorname{sech}^2(Mx) > 0$, and positive rescaling at a backward jump preserves this property. Consequently $r(x) > 0$ for $x > 0$.

Twice differentiating gives $w_\tau = \frac{1}{2}w_{xx} + rw_x + 3vw$. At the terminal endpoint $w = -2cM^2\operatorname{sech}^2(Mx)\tanh(Mx) \leq 0$ for $x > 0$. The boundary value at zero is zero, and the differentiated tails make the far boundary tend uniformly to zero. Comparison applied to $-w$ gives $w \leq 0$ on the positive half-line.

Define

$$H = \frac{r_\tau}{r} = \frac{w}{2r} + v.$$

Oddness and $r_x(0) = v(0) > 0$ make w/r, hence H, smooth and even at zero. Direct differentiation gives

$$H_\tau = \tfrac{1}{2}H_{xx} + (v/r + r)H_x + vH,$$
$$(H_x)_\tau = \tfrac{1}{2}(H_x)_{xx} + (v/r + r)(H_x)_x + ((v/r)_x + 2v)H_x + wH.$$

The terminal signs are $H = (cM - M^2)\operatorname{sech}^2(Mx) \leq 0$ and $H_x \geq 0$ on $x > 0$, using $M \geq c$. For the first comparison write $v/r + r = 1/x + \eta(x)$, where η is smooth and odd near zero. The radial extension $H(|y|)$ to $\mathbb{R}^3$ turns $\frac{1}{2}H_{xx} + x^{-1}H_x$ into $\frac{1}{2}\Delta H$. It is smooth at the origin, as is the extra drift, and H tends uniformly to zero at infinity. Comparison on expanding balls proves $H \leq 0$.

For the second comparison, $wH \geq 0$ and $(v/r)_x = w/r - v^2/r^2 < 0$, so the zeroth-order coefficient has upper bound $2\sup v$. On $[\varepsilon, A]$, the two lateral errors are $H_x(\varepsilon) = O(\varepsilon)$ and $H_x(A) = o_A(1)$, uniformly on the closed step. Taking $\varepsilon \downarrow 0$ and $A \to \infty$ gives $H_x \geq 0$. At a backward jump $r_{\text{new}} = \alpha r_{\text{old}}$ with $0 < \alpha \leq 1$,

$$H_{\text{new}} = H_{\text{old}} - (1-\alpha)v_{\text{old}}, \qquad (H_x)_{\text{new}} = (H_x)_{\text{old}} - (1-\alpha)w_{\text{old}}.$$

Thus both signs survive every jump. OpenAI (2026a), Lemma 3.2.

For the forward score use $a_s = s_x$, temporarily distinguishing it from the Hessian $a = \Phi_{xx}$. The forward heat equation gives

$$s_t = \tfrac{1}{2}s_{xx} - ss_x, \qquad (a_s)_t = \tfrac{1}{2}(a_s)_{xx} - s(a_s)_x - a_s^2,$$
$$(s_{xx})_t = \tfrac{1}{2}(s_{xx})_{xx} - s(s_{xx})_x - 3a_s s_{xx}.$$

The first step has $s = x/t$. Thereafter $a_s - 1/t$ satisfies the linear equation with drift $-s$ and zeroth-order coefficient $-(a_s + 1/t)$. For fixed smooth data this coefficient is bounded above on a closed step, independently of spatial truncation. This uses the tail bound and continuity, before its sign is known. Comparison gives $a_s \geq 1/t$. The last displayed equation then gives $s_{xx} \leq 0$ on $x > 0$. At a forward jump,

$$s_+ = s_- + \delta\Phi_x, \quad (a_s)_+ = (a_s)_- + \delta\Phi_{xx}, \quad (s_{xx})_+ = (s_{xx})_- + \delta\Phi_{xxx}.$$

The backward signs and $\delta \geq 0$ preserve the forward signs.

Concavity alone is insufficient here, so we need one further comparison. On $x > 0$ set $d_s = s_x/s$, $d_r = r_x/r$ and $\rho = d_s - d_r$. The heat equation gives $s_t = \frac{1}{2}s_{xx} - ss_x$, whereas $r_t = -rH$. Taking logarithmic derivatives yields

$$(d_s)_t = \tfrac{1}{2}(d_s)_{xx} + d_s(d_s)_x - s_{xx}, \qquad (d_r)_t = -H_x.$$

Since $\frac{1}{2}(d_r)_{xx} + d_r(d_r)_x = \frac{1}{2}(w/r)_x = H_x - w$, substitution of $d_s = d_r + \rho$ gives its equation

$$\rho_t = \tfrac{1}{2}\rho_{xx} + d_s\rho_x + (d_r)_x\rho + 2H_x - w - s_{xx}.$$

All three terms in the source are nonnegative by the comparisons already proved, and $(d_r)_x < 0$. On the first step, concavity gives $r(x) = \int_0^x v(y)\,dy \geq xv(x)$, so $\rho = 1/x - v/r \geq 0$. Odd Taylor expansions give $d_s = 1/x + O(x)$, $d_r = 1/x + O(x)$ and $\rho = O(x)$ near zero, and the tail estimates give $\rho \to 0$ at infinity. Comparison on $[\varepsilon, A]$ therefore proves $\rho \geq 0$ on each subsequent step provided its initial sign is preserved. Define

$$j = s\rho = s_x - (s/r)v.$$

At a forward jump, with $\alpha = c_+/c_-$ and $\lambda = (c_+ - c_-)/c_-$, $r_+ = \alpha r_-$, $v_+ = \alpha v_-$, $s_+ = s_- + \lambda r_-$ and $(s_x)_+ = (s_x)_- + \lambda v_-$. Substitution cancels the added terms and gives $j_+ = j_-$. Since $s_+ > 0$ on $x > 0$, the sign of ρ is preserved. The quotient s/r extends smoothly and evenly across zero, with $j(0) = 0$. Reflection then gives $j \geq 0$ everywhere. OpenAI (2026a), Lemma 3.3.

This ratio comparison also follows from the earlier lower-history inequalities in Zhou (2026), Appendix E.3, (E.37) and (E.41). To compare the finite-step problems, choose $M > \|\gamma_{\text{step}}\|_\infty$ and put $T = M^2t$, $y = Mx$, $U(T, y) = M\Phi(t, x)$ and $m = c/M$. Then U has terminal data $\log(2\cosh y)$, whose additive constant has no effect on the derivatives, and $0 < m < 1$. In Zhou's notation, $B = U_y$, $C = U_{yy}$ and $V_y = y/T + W_y$. The correspondence is $r = MmB$, $s = MV_y$, so

$$j = M^2\left(V_{yy} - \frac{CV_y}{B}\right) = M^2\left(\frac{B - yC}{TB} + B\,\partial_y\frac{W_y}{B}\right) \geq 0.$$

Here $B \geq yC$ follows from decreasing C on the positive half-line (Zhou, p. 130), while (E.37) implies (E.41), $\partial_y(W_y/B) \geq 0$. At a mass jump $m \mapsto m + d$, Zhou's update $W \mapsto W + dU$ in (E.28) adds dB to V_y and dC to V_{yy}, leaving the displayed expression for j unchanged. This identifies an earlier source for the ratio sign and its jump invariance in the smoothed finite-step problem; the approximation and coercivity arguments below are still needed for the full-support result.

4.4 The polynomial correction with zero expectation

Here the coefficient is a positive constant c on the step containing the chosen time. The variables are the rescaled spatial derivatives

$$r = c\Phi_x, \qquad v = c\Phi_{xx}, \qquad w = c\Phi_{xxx}, \qquad z = c\Phi_{xxxx}.$$

Write $h = e^{c\Phi}$ and factor the forward diffusion density as $f = hg$, with auxiliary score $s = -\partial_x \log g$. These are the backward and forward quantities compared in §4.3. On the positive half-line, where $r, s > 0$, set

$$H = \frac{w}{2r} + v, \qquad j = s_x - \frac{s}{r}v, \qquad \rho = \frac{s_x}{s} - \frac{v}{r}, \qquad j = s\rho.$$

In this certificate only, $b = -w/r$ is a quotient. In the unscaled curvature formula, b denotes instead the derivative Φ_{xxx}. At a fixed positive time inside the step define

$$R = r^2, \quad K = rs, \quad b = -w/r, \quad B = b - 2v, \quad L = -rs_{xx}, \quad S_1 = z + bv - 2Rb.$$

The subscript on S_1 distinguishes it from the support set. The previous comparisons give

$$R, K, v, B, j, L, S_1 \geq 0, \qquad b = B + 2v \geq 0, \qquad Rb^2 = w^2,$$

because $B = -2H$ and $S_1 = 2rH_x$. Here $s = -\partial_x \log g$ belongs to the forward heat factor, while $\partial_x \log f = r - s$ is the logarithmic derivative of the diffusion density. The signs have separate sources: $B \geq 0$ comes from $H \leq 0$, $S_1 \geq 0$ from $H_x \geq 0$ on the positive half-line, $j = s\rho \geq 0$ from the relative-slope comparison, and $L = -rs_{xx} \geq 0$ from forward concavity.

Reflection extends these signs. The quotients were defined for $x > 0$, where $r > 0$ and $v > 0$ by the tilted mean-and-variance argument in §4.3. At zero, r and w are odd, $r_x(0) = v(0) > 0$, and division by r is therefore division by a function with a simple zero. This gives the smooth even extension $b(0) = -z(0)/v(0)$, and the other comparison identities give $j(0) = 0$ and $S_1(0) = 0$. Thus the integration has no boundary at zero.

For $\mathcal{D} = r\partial_x$, direct differentiation gives

$$\mathcal{D}(R, K, v, b, j) = (2Rv,\ 2vK + Rj,\ -Rb,\ -z - bv,\ -jv + Kb - L).$$

Also $f_x/f = r - s$. Hence $\mathcal{W}(F) = \mathcal{D}F + (v + R - K)F$ satisfies

$$f\mathcal{W}(F) = \partial_x(rfF).$$

By itself the identity leaves the boundary terms at infinity to be controlled. The tail estimates give $r = O(1)$, exponential decay of v, w, z, b and their derivatives, $s = x/t + O(1)$, $j = 1/t + O((1 + x)e^{-\kappa x})$ and $K = O(1 + x)$. Every polynomial F in R, K, v, b, j and its differentiated expression therefore has at most polynomial growth. Meanwhile $f(t, x) \leq C \exp(-x^2/(2t) + C|x|)$. Thus every term used below is absolutely integrable and $rfF \to 0$ at both infinities. Smooth extension across zero prevents an interior boundary term. Integrating on $[-A, A]$ and taking $A \to \infty$ now gives $\mathbb{E}_f \mathcal{W}(F) = 0$. OpenAI (2026a), Lemma 3.4.

The curvature estimate rests on four separate steps. The comparisons supply the signs above, and the tail estimates make the correction integrate to zero. The rational identity below gives a positive remainder, and finite arithmetic verifies this third step. Finally, §4.5 passes the resulting expectation inequality to the original coefficient.

Set

$$(A, C, k, T_0, g_0, h_0) = \left(\frac{1}{13}, \frac{53}{24}, \frac{19}{16}, \frac{73}{48}, \frac{2331}{10000}, \frac{87}{2500}\right),$$
$$(x_0, y, m, n_0) = \left(\frac{4663}{5000}, \frac{6853}{10000}, -\frac{1051}{400}, -\frac{1073}{500}\right).$$

Define

$$\begin{aligned} Q &= Avb + Cv^2 + kKv - 2kRb + T_0Rv, \qquad T_1 = v(g_0R + h_0j), \\ F_2 &= (Av - 2kR)b^2 + b(2Cv^2 + 2kKv + (g_0 + 2T_0)Rv + h_0jv) \\ &\quad + x_0v^2j + yv^3 + mKv^2 + n_0Rv^2. \end{aligned}$$

The terms containing z motivate this choice. Indeed,

$$\begin{aligned} \partial_b F_2 &= 2(Av - 2kR)b + 2Cv^2 + 2kKv + (g_0 + 2T_0)Rv + h_0jv \\ &= 2Q + T_1, \\ [z]\mathcal{W}(F_2) &= -\partial_b F_2 = -(2Q + T_1). \end{aligned}$$

Here $[z]$ means the coefficient of z, and only $\mathcal{D}b = -z - bv$ contributes it. The square $(z+Q)^2$ and S_1T_1 contribute $2Q+T_1$, so these terms cancel exactly. Therefore subtracting $(z+Q)^2+S_1T_1+\mathcal{W}(F_2)$ from the curvature expression cancels every term involving z. The other favourable derivative is $\partial_j F_2 = x_0v^2 + h_0bv \geq 0$, so the $-L$ in $\mathcal{D}j$ leaves a nonnegative correction.

Put

$$\begin{aligned} P_0 &= Rj(b - 7v)^2, & P_1 &= RB(b - \tfrac{7}{3}v)^2, \\ P_2 &= Kv(b - \tfrac{9}{2}v)^2, & P_3 &= vB(K + \tfrac{1}{4}R - \tfrac{1}{7}v)^2, \\ P_4 &= (vb - \tfrac{14}{5}v^2 - \tfrac{13}{4}Kv + \tfrac{4}{3}Rv)^2, \end{aligned}$$
$$(p_0, p_1, p_2, p_3, p_4) = \left(\frac{337}{10000}, \frac{183}{2500}, \frac{411}{10000}, \frac{11461}{5000}, \frac{647}{10000}\right).$$

Exact expansion, with $b = B + 2v$, gives

$$z^2 - 12Rvb^2 + 6v^4 = (z+Q)^2 + S_1T_1 + \sum_{i=0}^{4} p_iP_i + \mathcal{W}(F_2) + L(x_0v^2 + h_0bv) + P_{\mathrm{rem}}.$$

The table gives the entire remainder, and every unlisted coefficient is zero.

Monomial	Coefficient	Monomial	Coefficient
R^2v^2	223/160000	R^2vB	4007/240000
R^2B^2	1097/120000	RKv^2	83/80000
$RKvB$	3343/24000	Rv^3	16259/23400000
Rv^2B	30179/5460000	Rv^2j	26/625
RvB^2	107/15000	$RvBj$	3/2500
RB^3	121/32500	RB^2j	11/10000
K^2v^2	579/20000	K^2vB	207/2500
Kv^3	911/975000	Kv^2B	349/420000
Kv^2j	5011/5000	KvB^2	133/130000
$KvBj$	87/2500	v^4	1661507/1521000000
v^3B	633799/621075000	v^2B^2	10657/1690000

All square prefactors $Rj, RB, Kv, vB, 1$ are nonnegative, as are S_1T_1, $L(x_0v^2 + h_0bv)$ and every remainder monomial. Integration removes only $\mathcal{W}(F_2)$, giving

$$\mathbb{E}_f[z^2 - 12vw^2 + 6v^4] \geq \frac{1661507}{1521000000}\mathbb{E}_f v^4 \geq \frac{1}{1000}\mathbb{E}_f v^4.$$

The surplus of the first coefficient over 1/1000 is exactly $140507/1521000000 > 0$. The rational identity and the positivity of the table reduce to finite arithmetic. Their application to this diffusion also uses the signs, quotient behaviour, integrability and vanishing flux proved above. Dividing the integrated inequality by c^2 returns

$$\mathbb{E}[\Phi_{xxxx}^2 - 12c\Phi_{xx}\Phi_{xxx}^2 + 6c^2\Phi_{xx}^4] \geq \frac{c^2}{1000}\mathbb{E}\Phi_{xx}^4.$$

This is the unscaled estimate needed by the gap argument. Section 4.5 passes it to the original coefficient and supplies strictness for the limiting solution. OpenAI (2026a), Lemma 4.1 and Proposition 4.2.

4.5 Coercivity at a fixed time and on plateaus

Proposition 4.1 (coercivity at a fixed time). Let $\gamma \in \mathcal{U}$ and let X be its associated diffusion, started at zero. Suppose that $0 < t < 1$, γ is continuous at t, and $c = \gamma(t) > 0$. With $a = \Phi_{xx}$, $b = \Phi_{xxx}$ and $d = \Phi_{xxxx}$, evaluated at (t, X_t), one has

$$\mathcal{C}_\gamma(t) := \mathbb{E}[d^2 - 12cab^2 + 6c^2a^4] \geq \frac{c^2}{1000}\mathbb{E}a^4. \tag{4.1}$$

The proposition assumes neither minimality nor a constant neighbourhood of t, and it is the common estimate behind plateau curvature and positive density.

When γ is smooth near t, Itô's formula gives $q'' = \mathbb{E}b^2 - 2\gamma\mathbb{E}a^3$ and

$$(\mathbb{E}b^2)' = \mathbb{E}d^2 - 6\gamma\mathbb{E}(ab^2), \qquad (\mathbb{E}a^3)' = 3\mathbb{E}(ab^2) - 3\gamma\mathbb{E}a^4.$$

Differentiating the first identity therefore gives

$$q'''(t) + 2\gamma'(t)\mathbb{E}a^3 = \mathcal{C}_\gamma(t). \tag{4.2}$$

On a plateau the coefficient derivative vanishes. At a smooth minimizer, $q = t$ makes the third derivative of q vanish instead.

Proof. Fix $t < T < T' < 1$ and a shrinking interval $[t - \delta_n, t + \delta_n) \subset (0, T)$. Approximate γ by

$$h_n = \max\{\epsilon_n, \min\{K_n, \eta_n \lfloor \gamma/\eta_n \rfloor\}\},$$

where $0 < \epsilon_n < c < K_n$, $\epsilon_n, \eta_n \downarrow 0$ and $K_n \uparrow \infty$. Use $\min(c, h_n)$ to the left of the interval, c on it, and $\max(c, h_n)$ to its right. Monotonicity makes these positive, nondecreasing, right-continuous finite-step coefficients γ_n. Clipping toward c outside the interval cannot increase the error, and

$$\|\gamma_n - \gamma\|_1 \le \epsilon_n + \eta_n + \int_0^1 (\gamma - K_n)_+ \, ds + 2\delta_n \sup_{|s-t| \le \delta_n} |\gamma(s) - c| \longrightarrow 0. \tag{4.3}$$

Integrability controls the capped tail and continuity at t the last term. Choose smooth terminal parameters $M_n \ge \max(n, K_n)$. The terminal values then converge uniformly to $|x|$, and the coefficient masses have a common bound. The finite-step inequality of §4.4 applies at t, where every γ_n equals the same positive c.

On a fixed strip containing t and ending before one, compact-strip regularity gives uniform spatial convergence and bounds through order four. The next derivative bounds give a common spatial Lipschitz constant for those derivatives. Coupling with the same Brownian motion and applying Gronwall gives a deterministic bound $\sup_{s \le T} |X_s^n - X_s| \to 0$. If $D_k^n = \partial_x^k \Phi_n$ and $D_k = \partial_x^k \Phi$, then, for $1 \le k \le 4$,

$$|D_k^n(t, X_t^n) - D_k(t, X_t)| \le \|D_k^n - D_k\|_{\infty, [0,T] \times \mathbb{R}} + C_{k,T} |X_t^n - X_t| \longrightarrow 0.$$

After the nonnegative terms are discarded, the two surviving integrands $z_n^2 - 12 v_n w_n^2 + 6 v_n^4$ and v_n^4 are uniformly bounded, with $v_n = cD_2^n$, $w_n = cD_3^n$ and $z_n = cD_4^n$. Dominated convergence therefore passes their expectations to the limit.

To make this quantitative, let V, W, Z bound both the limiting and approximating v, w, z, and let e_v, e_w, e_z bound their coupled evaluation errors. Uniform spatial convergence and the path bound above give these errors with $e_v, e_w, e_z \to 0$. Difference-of-powers identities give

$$\begin{aligned} |\mathbb{E}\mathcal{C}(v_n, w_n, z_n) - \mathbb{E}\mathcal{C}(v, w, z)| &\le (12W^2 + 24V^3) e_v + 24VW e_w + 2Z e_z, \\ |\mathbb{E} v_n^4 - \mathbb{E} v^4| &\le 4V^3 e_v, \end{aligned}$$

where $\mathcal{C}(v, w, z) = z^2 - 12vw^2 + 6v^4$. These constants depend only on the common strip and the coefficient mass bound, and in particular are independent of the plateau width. After division by c^2, this is (4.1).

Only spatial moments pass to the limit. The auxiliary quantities s_n and j_n have already disappeared after integration, so the transfer needs neither their convergence nor a uniform bound on their tail constants. OpenAI (2026a), Lemmas 2.1–2.3; Proposition 5.1.

The spatial inequality at a coefficient jump. Proposition 4.1 and the sampled formal interface assume continuity at t, and the spatial-moment argument permits a little more. Fix $c = \gamma(t) > 0$ and replace γ by c on $[t - \delta, t + \delta)$, leaving it unchanged elsewhere. Monotonicity preserves both joins and right continuity. For fixed $0 < \delta_0 < \min(t, 1 - t)$ and $0 < \delta \le \delta_0$,

$$\|\gamma_\delta - \gamma\|_1 \le 2\delta \, \gamma(t + \delta_0) \longrightarrow 0.$$

Thus local boundedness suffices for the same strip stability and coupled expectation limit. Each replacement has a positive plateau at t, so (4.1) holds at a jump too, with the stated right-continuous value c. This extension concerns spatial moments only. It supplies no time derivative $q'''(t)$ or $\gamma'(t)$ at a jump, and the formal declaration keeps every one of its hypotheses. Neither application below uses the extension.

Corollary 4.2 (positive plateaus). Suppose now that $\gamma = c > 0$ on an open interval $I \subset (0,1)$ and $t \in I$. Then (4.2) gives $q'''(t) = \mathcal{C}_\gamma(t)$. To establish strictness for this limiting solution, choose $b \in I$ later than t and put $\psi = \Phi_\gamma(b, \cdot)$. This is smooth, convex, even and a bounded distance from $|x|$, so ψ' cannot be constant. The Cole–Hopf formula gives

$$\Phi_{xx}(t,x) = \mathbb{E}^{\mathrm{tilt}}_{t,x}\psi'' + c\,\mathrm{Var}^{\mathrm{tilt}}_{t,x}(\psi'),$$
$$\nu_{t,x}(dy) = \frac{e^{c\psi(y)}p_{b-t}(y-x)}{\int_{\mathbb{R}} e^{c\psi(z)}p_{b-t}(z-x)\,dz}\,dy.$$

The tilted density is positive everywhere. Hence the tilted variance of the nonconstant continuous function ψ' is strictly positive, while the tilted mean of ψ'' is nonnegative. Thus $v(t,x) > 0$ for every x, and $\mathbb{E}v(t,X_t)^4 > 0$. Combining the limiting inequality with the differentiated identity proves $q'''(t) \geq (1000c^2)^{-1}\mathbb{E}v(t,X_t)^4 > 0$. In particular, q' is strictly convex on I. OpenAI (2026a), Lemma 2.3; Proposition 5.1.

4.6 The zero interval as a separate case

When $c = 0$, the heat equation gives $q''(t) = \mathbb{E}\Phi_{xxx}(t,X_t)^2$ without division by c. Start the drift-free evolution at an earlier point of I. Its Gaussian transition kernel gives the law of X_t a positive density everywhere. If the expectation were zero, continuity would force $\Phi_{xxx}(t,\cdot) \equiv 0$. Then $\Phi(t,\cdot)$ would be an even quadratic polynomial. Its one-Lipschitz bound forces that polynomial to be constant, contradicting its bounded distance from $|x|$. Therefore $q''(t) > 0$, giving strict increase of q'. OpenAI (2026a), Lemma 2.3.

The analytic certificate is thus a proved input to the gap argument in this source, under the stated model and regularity assumptions. An abstract polynomial or flux lemma assumes those analytic hypotheses, and the finite-step comparisons and the transfer argument discharge them. Section 13 separates inspection of the corresponding Lean sources, which shows that construction, from compilation, scoped dependency extraction and Comparator replay, and reports the observed outcome of each.

5. Excluding terminal, interior and initial gaps

Theorem 5.1 (full support). In the pure zero-field SK model with $\xi(t) = t^2/2$, let $\gamma \in \mathcal{U}$ minimize $\mathcal{P}$. Then $\operatorname{supp}\mu_\gamma = [0,1)$ and γ is strictly increasing. For every associated diffusion started at zero, $\mathbb{E}u(t,X_t)^2 = t$ and $\mathbb{E}a(t,X_t)^2 = 1$ for $0 \leq t < 1$.

We keep the same minimiser, diffusion and moment function q throughout. Section 3 has established $G \geq 0$, $G = 0$ on the support $S = \operatorname{supp}\mu_\gamma$, and the contact conditions $q(s) = s$, $q'(s) \leq 1$ at support points. At zero, the value identity comes from symmetry, and the slope bound requires $0 \in S$. Section 4 supplies curvature on constant intervals for every admissible coefficient. These premises are available before any support gap is excluded. We first deal with the endpoint at one, and then use the curvature estimate on the remaining intervals.

For a terminal interval $(a,1)$, including empty support $[0,1)$, the cusp estimate gives $1-q(t) \gtrsim \sqrt{1-t}$ and hence $G<0$ near one. On a positive plateau with two contact endpoints, strict convexity puts q' below its endpoint chord, although its integral must equal the interval length. On a zero plateau the heat argument makes q' strictly increasing, below the right contact slope. At an initial gap we use the one-sided contact slope if $0 \in S$. Otherwise the coefficient is zero, and symmetry supplies $q(0)=0$.

A terminal plateau violates the variational inequality. Suppose that $\gamma = c$ on $(a,1)$, where $0 \le a < 1$ and $c \ge 0$ is finite. The endpoint 1 lies outside the support domain and carries no contact condition, so an endpoint-chord argument is unavailable. Instead we prove

$$1-q(t) \ge C\sqrt{1-t} \tag{5.1}$$

for some $C>0$ and every t sufficiently close to one. The proof separates two facts: the diffusion puts positive probability density near zero, and the smoothed terminal slope has a definite deficit from ± 1 in a shrinking window around zero.

Fix $s \in (a,1)$ and let ρ_s be the law of X_s. Choose $R<\infty$ such that $p_0 = \rho_s([-R,R]) > 0$. For $c>0$, set

$$h(t,y) = \mathbb{E}e^{c|y+\sqrt{1-t}Z|}, \qquad \Phi(t,y) = c^{-1}\log h(t,y),$$

where Z is standard normal. The drift on this interval is $c\Phi_x = h_x/h$, and the transition density is

$$k_{s,t}(x,y) = p_{t-s}(y-x)\frac{h(t,y)}{h(s,x)}, \qquad p_v(z) = (2\pi v)^{-1/2}e^{-z^2/(2v)}. \tag{5.2}$$

The heat-semigroup identity normalises this kernel. Equivalently, for Brownian motion started at (s,x), $h(t,B_t)/h(s,x)$ is the positive mean-one change of measure which adds the drift h_x/h. For $c=0$, the same kernel formula applies with $h=1$, and the value function is the ordinary heat evolution of $|x|$. OpenAI (2026a), Lemma 6.3, proof of (6.6).

Put $L = 1-s$ and $\delta = L/2$. For $s+\delta \le t < 1$, $|x| \le R$ and $|y| \le 1$,

$$p_{t-s}(y-x) \ge (2\pi L)^{-1/2}e^{-(R+1)^2/(2\delta)} =: k_* > 0,$$

while $h(t,y) \ge 1$ and $h(s,x) \le e^{cR}\mathbb{E}e^{c\sqrt{L}|Z|} =: M < \infty$. Integrating (5.2) over the part of ρ_s in $[-R,R]$ therefore gives a density f_t of X_t with

$$f_t(y) \ge p_0 k_*/M =: m > 0 \quad (s+\delta \le t < 1,\ |y| \le 1). \tag{5.3}$$

At time s we used only the positive mass of a compact set, with no density assumption on ρ_s.

Now write $\varepsilon = 1-t$ and restrict to $|y| \le \sqrt{\varepsilon}$. Differentiating the terminal heat formula gives, with the unweighted interpretation for $c=0$,

$$u(t,y) = \frac{\mathbb{E}[\operatorname{sgn}(y+\sqrt{\varepsilon}Z)e^{c|y+\sqrt{\varepsilon}Z|}]}{\mathbb{E}e^{c|y+\sqrt{\varepsilon}Z|}}.$$

The denominator is at most $D = e^c\mathbb{E}e^{c|Z|} < \infty$. Each sign has unnormalised mass at least $p_* := \mathbb{P}(Z \ge 1) > 0$, since the threshold $|y/\sqrt{\varepsilon}|$ is at most one, and the exponential weight is at least one. Hence each sign has tilted probability at least $\theta = p_*/D \in (0,1/2)$. It follows that

$$|u(t,y)| \le 1-2\theta, \qquad 1-u(t,y)^2 \ge 4\theta(1-\theta) =: \kappa > 0.$$

Multiplying this uniform deficit by the density lower bound (5.3) and the window width $2\sqrt{\varepsilon}$ yields

$$1 - q(t) = \int_{\mathbb{R}} (1 - u(t,y)^2) f_t(y)\, dy \geq 2m\kappa\sqrt{1-t}.$$

This proves (5.1) with $C = 2m\kappa$. OpenAI (2026a), Lemma 6.3, density and terminal-window estimates.

We can now read off the contradiction. For s' sufficiently close to one, the estimate holds at every $t \in [s', 1)$, so

$$\begin{aligned} G(s') &= \int_{s'}^{1} (q(t) - t)\, dt \\ &\leq \frac{(1-s')^2}{2} - \frac{2C}{3}(1-s')^{3/2} < 0. \end{aligned}$$

The final inequality holds once $\sqrt{1-s'} < 4C/3$. It contradicts the nonnegative cost of adding an atom at s'. This proof covers $c = 0$ as well as $c > 0$. In particular, empty support is impossible: then $\mu_\gamma = 0$ and $\gamma = 0$ on all of $[0, 1)$. The terminal-plateau conclusion appears already in Auffinger, Chen and Zeng (2020), Remark 7, and the estimates above supply the quantitative argument needed here.

This excludes the terminal gap. The support proof resumes with the remaining finite components. The next argument gives a shorter route to the same exclusion, without its quantitative bound.

A shorter qualitative exclusion. The curvature input of §4 also excludes a terminal support gap, with no estimate of its boundary layer. One elementary calculus step suffices. Suppose $q \in C^1([r, 1))$, $q(r) = r$, $q'(r) \leq 1$, $q \leq 1$, and q' is strictly convex on $(r, 1)$. If $q(t) \geq t$ at some $r < t < 1$, strict integration below the finite endpoint chord gives

$$t - r \leq \int_r^t q'(v)\, dv < \frac{t-r}{2}(q'(r) + q'(t)),$$

so $q'(t) > 1$. For $t < s < 1$, convexity then gives

$$q'(s) \geq q'(t) + \frac{s-t}{t-r}(q'(t) - q'(r)) \geq q'(t) > 1.$$

Consequently $q(s) \geq t + (s-t)q'(t) > 1$ for s sufficiently close to one, a contradiction. Thus $q(t) < t$ throughout $(r, 1)$. All chord endpoints lie below one, so no derivative at one is used.

The same inputs close the empty-support case. Empty Stieltjes support forces $\gamma = 0$ everywhere. The zero-interval argument makes q strictly convex on $[0, 1)$, whereas the positive-plateau argument above makes q' strictly convex. Here $q(0) = 0$ and $0 \leq q \leq 1$. For $0 < t < s < 1$, convexity gives

$$q(t) \leq \frac{t}{s} q(s) \leq \frac{t}{s}.$$

Letting $s \uparrow 1$ in this bound gives $q(t) \leq t$. Now fix $t < u < 1$. Strict convexity on the finite interval $[0, u]$ gives

$$q(t) < \frac{t}{u} q(u) \leq t.$$

Thus the bounded continuous integrand $q(t) - t$ is strictly negative on $(0, 1)$, so $G(0) < 0$, contrary to the allowed addition of mass at zero. The argument assumes no value or derivative at one. The

explicit zero-coefficient calculation also gives $D_+\mathcal{P}(0)[1] = 1/4 - 1/\pi$, and the qualitative argument needs only its sign.

A nonempty support with a terminal complementary component has a last point $r < 1$. Its plateau has positive height, since a neighbourhood of r has positive measure. The contact values and slope at r, including the one-sided case $r = 0$, and positive-plateau curvature meet the calculus hypotheses. Since $0 \le q \le 1$, integration gives $G(r) < 0$, contradicting atom addition. This follows from the inputs already stated. We keep the quantitative proof above for its stronger estimate (5.1).

The remaining gaps have a finite right contact point.

The support is relatively closed in $[0, 1)$. Its complement is a union of relative open intervals, and the Stieltjes measure vanishes on each such interval. For $x < y$ in one interval,

$$\gamma(y) - \gamma(x) = \mu_\gamma((x, y]) = 0,$$

so γ is constant there. The terminal argument excludes a component ending at one. Every remaining component therefore ends at a point $b < 1$ of S.

First consider a gap (a, b) with both endpoints in S, allowing $a = 0$. Compact-strip regularity gives continuity of q' on $[a, b]$, even if γ jumps at an endpoint. Contact gives

$$q(a) = a, \qquad q(b) = b, \qquad q'(a), q'(b) \le 1. \tag{5.4}$$

If the constant coefficient is $c > 0$, Section 4 makes q' strictly convex in the gap. Continuity extends the chord comparison to its endpoints, giving

$$q'(t) < \frac{b-t}{b-a} q'(a) + \frac{t-a}{b-a} q'(b) \le 1, \qquad a < t < b. \tag{5.5}$$

Integrating contradicts $q(b) - q(a) = b - a$. Equivalently, the mean value theorem applied to (5.4) requires an interior point with $q' = 1$, whereas (5.5) excludes every such point. The endpoint derivatives need only be at most one.

With $h = q'$ and $C_{a,b}$ its endpoint chord, the contradiction is the exact area identity

$$(b-a) - \int_a^b h = (b-a)\left(1 - \frac{h(a) + h(b)}{2}\right) + \int_a^b (C_{a,b} - h) > 0.$$

The first term is nonnegative, and strict convexity makes the second positive. Only continuity of q' is needed at the ends. For example, a jump $\Delta\gamma$ at s gives the one-sided identity

$$q''(s+) - q''(s-) = -2\Delta\gamma\, \mathbb{E} a(s, X_s)^3.$$

It follows by applying Itô's formula to a^2 and using continuity of its derivative moments on a compact strip. Taking $\gamma = 0$ before $s = 1/2$ and $\gamma = 1/2$ afterward makes this jump strictly negative, because the terminal Cole–Hopf solution has positive Hessian. Thus continuity of q' gives no classical higher time derivatives across coefficient jumps. OpenAI (2026a), proof of Corollary 6.4.

If $c = 0$, the separate heat argument in Section 4 makes q' strictly increasing. Then $q'(t) < q'(b) \le 1$ for $a < t < b$, and the same contradiction follows. The left slope bound is unnecessary in this case. OpenAI (2026a), proof of Theorem 1.1 after Lemma 6.3.

There are two distinct possibilities at the left endpoint. If $(0, b)$ is a gap and $0 \in S$, the preceding proof applies using the one-sided contact slope bound at zero. Because $(0, b)$ contains no mass, the support condition at zero forces an atom there and hence $c = \gamma(0) > 0$. At this stage the atom has not yet been excluded.

If instead the complementary component is $[0, b)$ with $0 \notin S$, there is no mass anywhere in $[0, b)$. Thus $\gamma = 0$ there. The identity $q(0) = 0$ comes from zero-field symmetry, and the right endpoint gives $q(b) = b$, $q'(b) \leq 1$. Here $q''(t) > 0$ for $0 < t < b$, and at zero evenness gives $\Phi_{xxx}(0, 0) = 0$ and hence $q''(0) = 0$. Strict increase on the open interval still forces $q' < 1$ throughout $(0, b)$, an impossibility.

Full support and the moment identities.

The cases are exhaustive, so $S = [0, 1)$. Only now can contact be applied at every $0 < t < 1$, giving $q(t) = t$ throughout the domain. The value at zero was already known. Differentiating and using the gradient-martingale identity gives

$$\mathbb{E}a(t, X_t)^2 = q'(t) = 1, \qquad 0 \leq t < 1,$$

with the right derivative at zero.

Finally, every nonempty open subinterval (a, b) receives positive mass. Therefore

$$\gamma(b) - \gamma(a) = \mu_\gamma((a, b]) > 0 \qquad (0 \leq a < b < 1).$$

The proof establishes full support, strict increase and the two moment identities. Smooth absolute continuity, including the later conclusion $\gamma(0) = 0$, uses the source's further argument from saturated consistency, and neither contact nor the exclusion of an initial gap depends on it. OpenAI (2026a), conclusion of the proof of Theorem 1.1.

A weaker analytic condition suffices for the bounded-gap contradiction. If $h \in C([a, b]) \cap C^2((a, b))$ has the displayed endpoint and integral data, then $h' = 0 \Rightarrow h'' > 0$ is already impossible: either $h \leq 1$ everywhere, forcing $h \equiv 1$ by its integral, or it attains an interior maximum above one, where $h' = 0$ and $h'' \leq 0$. This is an elementary deduction from these premises. Chen's conditional crossing theorem, with its own hypotheses, has this form. Section 4 proves the stronger $q''' > 0$ throughout a positive plateau, for any admissible coefficient. See Chen, arXiv:2607.18032v1, Proposition 4.2 and the proof of Proposition 4.3. Neither statement supplies terminal contact at the excluded endpoint one.

6. Recovering the value from a martingale

Proposition 6.1 (terminal value). For the minimizer and diffusion of Theorem 5.1 there is a spin-valued terminal variable U_1 such that $u(t, X_t) = \mathbb{E}[U_1 \mid \mathcal{F}_t]$ and

$$P_* = \mathbb{E}[U_1 B_1] = \int_0^1 \mathbb{E}a(t, X_t)\, dt. \tag{6.1}$$

Its finite-cutoff tail lies in $[0, 1-T]$. Besides the support theorem, the proposition uses the attained Parisi formula.

The bounded martingale $u(t, X_t)$ converges almost surely and in L^2 to U_1. Its second moment tends to one because $q(t) = t$. Since $|U_1| \leq 1$, $\mathbb{E}U_1^2 = 1$ forces $U_1^2 = 1$ almost surely. Uniform integrability identifies $u(t, X_t) = \mathbb{E}[U_1 \mid \mathcal{F}_t]$, which is the conditional-expectation identity used below.

The drift bound $|\gamma u| \le \gamma \in L^1$ gives convergence of X_t to X_1, and uniform terminal convergence of Φ to $|x|$, together with convexity, gives $u(t, x_t) \to \operatorname{sgn}(x)$ when $x_t \to x \neq 0$. Thus $U_1 X_1 = |X_1|$, including when $X_1 = 0$, so we never need the probability of that event.

Itô's formula gives

$$\mathbb{E}\Phi(t, X_t) = \Phi(0,0) + \frac{1}{2}\int_0^t \gamma(s) s \, ds.$$

The terminal limit first gives $\Phi(0,0) = \mathbb{E}|X_1| - \frac{1}{2}\int_0^1 s\gamma(s) ds$. Subtracting the second half from the definition of the functional yields $\mathcal{P}(\gamma) = \mathbb{E}|X_1| - \int_0^1 s\gamma(s)\, ds$. The ground-state Parisi formula identifies this with P_*. The diffusion equation, conditional expectation and Fubini give

$$\mathbb{E}[U_1 X_1] = \mathbb{E}[U_1 B_1] + \int_0^1 \gamma(s)\mathbb{E}[u(s, X_s)^2]\, ds.$$

Bounded u, U_1 and integrable γ justify these terms. Combining the identities and using stochastic integral covariance gives

$$P_* = \mathbb{E}|X_1| - \int_0^1 t\gamma(t)\, dt = \mathbb{E}[U_1 B_1] = \int_0^1 \mathbb{E}a(t, X_t)\, dt.$$

The last integral is absolutely bounded by one because $\mathbb{E}|a| \le (\mathbb{E}a^2)^{1/2} = 1$. Full support supplies $\mathbb{E}a^2 = 1$, whereas the value uses $\int \mathbb{E}a$, so substituting the second moment into the value identity would change the assertion. OpenAI (2026a), §7.

The prediction error and the energy tail have different sizes:

$$\|U_1 - u(T, X_T)\|_2^2 = 1 - T,$$

$$P_* - \int_0^T \mathbb{E}a(t, X_t)\, dt = \mathbb{E}[(U_1 - u(T, X_T))(B_1 - B_T)] \in [0, 1-T].$$

The first identity is orthogonal projection onto $\mathcal{F}_T$. Martingale orthogonality removes the past Brownian component in the second. Cauchy–Schwarz bounds the product by $1-T$, and convexity makes it nonnegative, so no uniform terminal curvature bound is needed.

Sellke gave a direct antecedent of this stochastic value calculation in Sellke, arXiv:2105.03506v4, Definition 1.3 and Lemma 3.11. There, "optimizable" means the contact identity $\mathbb{E}u(t, X_t)^2 = t$ on $[q, 1)$, with q the infimum of the support of the coefficient function. Lemma 3.11 derives the finite-cutoff value identity by applying Itô's formula to Φ and xu, before taking the terminal limit. This is the same value-recovery mechanism used here, specialized to zero field and $\xi'' = 1$.

A cutoff must retain the full Parisi penalty. Set $\Psi(T, x) = \Phi(T, x) - xu(T, x)$ and $I(T) = \int_0^T \mathbb{E}a(t, X_t)\, dt$. Subtracting the complete penalty from the finite-time Itô identity gives

$$\mathcal{P}(\gamma) - I(T) = \mathbb{E}\Psi(T, X_T) + \frac{1}{2}\int_0^T \gamma(t)(q(t) - t)\, dt - \frac{1}{2}\int_T^1 t\gamma(t)\, dt.$$

Under contact the middle term vanishes. The final term is absent from the displayed equality in El Alaoui, Montanari and Sellke, arXiv:2001.00904v1, §6.2, pp. 32–33. Its sign is nonpositive, so the upper bound by $\mathbb{E}\Psi(T, X_T)$ used there remains valid, and its magnitude tends to zero by integrability. This finite-cutoff correction is consistent with the exact terminal identity above.

A coefficient budget from the first curvature moment. Full support also bounds the minimizing coefficient. Put $m(t) = \mathbb{E}a(t, X_t)$. The compact-strip identity

$$da(t, X_t) = -\gamma(t)a(t, X_t)^2\, dt + b(t, X_t)\, dB_t$$

holds before any smoothness conclusion for γ, and the bounded strip derivatives make its stochastic integral a martingale. OpenAI (2026a), proof of Corollary 6.4. At zero, $X_0 = 0$, $a(0, 0) \geq 0$ and $\mathbb{E}a(0, X_0)^2 = 1$ give $a(0, 0) = 1$. Taking expectations and using $\mathbb{E}a(t, X_t)^2 = 1$ therefore yields

$$m(t) = 1 - \int_0^t \gamma(s)\, ds, \qquad 0 \leq t < 1.$$

Convexity gives $m(t) \geq 0$. Letting $t \uparrow 1$ and using monotonicity of γ on every tail gives

$$\int_0^1 \gamma(s)\, ds \leq 1, \qquad 0 \leq \gamma(t) \leq \frac{m(t)}{1-t} \leq \frac{1}{1-t}.$$

Indeed, $(s-t)\gamma(t) \leq \int_t^s \gamma \leq m(t)$ for $t < s < 1$. Integrating the first-moment identity also gives

$$I(T) = T - \int_0^T (T-s)\gamma(s)\, ds, \qquad P_* = 1 - \int_0^1 (1-s)\gamma(s)\, ds.$$

The second equality uses (6.1) and so depends on its ground-state identification. Both follow directly in the normalization $\xi'' = 1$. The budget alone leaves the terminal first moment undetermined.

Proposition 6.2 (terminal normalization in zero field). For the minimizer and diffusion of Theorem 5.1,

$$\lim_{t\uparrow 1} m(t) = 0, \qquad \int_0^1 \gamma(s)\, ds = 1, \qquad m(t) = \int_t^1 \gamma(s)\, ds,$$

and consequently

$$P_* = \int_0^1 s\gamma(s)\, ds.$$

Thus $\gamma(s)\, ds$ has total mass one. This says nothing about the total mass of the Stieltjes measure $d\gamma$. The proposition is an analytic consequence outside the selected formal endpoints.

Proof. The backward comparison in §4.3 gives $\Phi_{xxx}(t, x) \leq 0$ for $x > 0$ in every positive finite-step approximation. Compact-strip convergence of the third spatial derivative passes this inequality to γ. Together with evenness and convexity, it says that $a(t, x)$ is nonnegative, even and nonincreasing in $|x|$. These inputs are OpenAI (2026a), Lemma 3.2 and Lemma 2.1, and neither requires differentiability of γ. Uniform terminal convergence to $|x|$ and convex difference quotients give $u(t, x) \to 1$ at every fixed $x > 0$. For each $\delta > 0$,

$$0 \leq a(t, \delta) \leq \frac{2}{\delta} \int_{\delta/2}^{\delta} a(t, x)\, dx = \frac{2}{\delta}(u(t, \delta) - u(t, \delta/2)) \longrightarrow 0.$$

The spatial monotonicity is essential here, since convergence of convex functions alone would not imply convergence of their second derivatives.

It remains to control the probability of the shrinking region where curvature can be large. Oddness and monotonicity give $u(t,x) \geq 0$ for $x \geq 0$, so the drift points away from zero. Tanaka's formula for $R_t = |X_t|$ gives

$$R_t = W_t + V_t + L_t, \qquad W_t = \int_0^t \operatorname{sgn}(X_s)\, dB_s, \qquad V_t = \int_0^t \gamma(s) u(s, R_s)\, ds.$$

Here L is the local time of X at zero. The occupation-time formula and $d\langle X\rangle_s = ds$ imply $\int_0^t \mathbf{1}_{\{X_s=0\}}\, ds = 0$. Thus $\langle W\rangle_t = t$, and the continuous martingale W is Brownian motion. The process V is nondecreasing and starts at zero, and L increases only when $R = 0$. The reflection formula therefore gives, pathwise,

$$R_t = W_t + V_t - \min\left(0, \inf_{s\leq t}(W_s + V_s)\right) \geq W_t - \min\left(0, \inf_{s\leq t} W_s\right) =: R_t^0.$$

Indeed $0 \leq V_s \leq V_t$ bounds the subtracted minimum by $\min(0, \inf_{s\leq t} W_s) + V_t$. No independence of V and W is needed. Reflected Brownian motion R_t^0 has the one-time law $|B_t|$, so for $t > 0$,

$$\Pr(|X_t| \leq \delta) \leq \Pr(|B_t| \leq \delta) \leq \sqrt{\frac{2}{\pi}} \frac{\delta}{\sqrt{t}}.$$

Splitting at $|X_t| = \delta$, using spatial monotonicity outside and Cauchy–Schwarz with $\mathbb{E}a(t, X_t)^2 = 1$ inside, yields

$$0 \leq m(t) \leq a(t,\delta) + \sqrt{\Pr(|X_t| \leq \delta)}.$$

First let $t \uparrow 1$ and then $\delta \downarrow 0$. This proves $m(1-) = 0$. The compact-strip identity for m now gives total mass one and the tail formula. Tonelli and (6.1) give the stated expression for P_*. In particular,

$$0 \leq P_* - I(T) = \int_T^1 (s - T)\gamma(s)\, ds \leq (1 - T)m(T) = o(1 - T).$$

No numerical rate for $m(T) \to 0$ or for the required spatial derivative bounds follows from this proof.

The endpoint distinction also matters when comparing editions. Sellke, arXiv:2105.03506v4, Lemma 2.9, equation (2.7) states the expected-Hessian tail equality. In version 6, Lemma 2.9, equation (2.8), and Remark 2.10, the general statement is an inequality, and the remark obtains equality indirectly for even mixtures. The specialized zero-field proof above supplies its own endpoint argument from the shape and contact identities. It never substitutes $t = 1$ into a compact-strip formula, and it avoids the older equality in its full stated generality.

7. A finite Gaussian spin and its adjacent correlations

Proposition 7.1 (finite Gaussian approximation). For a fixed minimizer and its diffusion, the construction below has positive active normalizers for all sufficiently fine meshes at every fixed $0 < T < 1$. Its adjacent correlation sum converges to $\int_0^T \mathbb{E}a(t, X_t)\, dt$, and this integral tends to P_* as $T \uparrow 1$. The limits are taken in that order.

We keep the minimizer and its diffusion from §6 and construct a finite Gaussian sign F_N with two correlation vectors whose *adjacent* pairing

$$S_{T,N} = \sum_{j=0}^{N} A_j B_{j+1} \longrightarrow \int_0^T \mathbb{E}a(t, X_t)\, dt.$$

The shift by one is essential for the rooted energy identity in §8.1, and §10.4 computes a case where the unshifted pairing has a different value.

We first fix $0 < T < 1$, then a mesh $N \geq 1$, $\Delta = T/N$, $t_j = j\Delta$. On independent standard normals $Z_1, \dots, Z_{N+1}$, set

$$Y_0 = 0, \quad Y_j = Y_{j-1} + \sqrt{\Delta} Z_j + \Delta \gamma(t_{j-1}) u(t_{j-1}, Y_{j-1}),$$
$$c_j = (\mathbb{E} a(t_{j-1}, Y_{j-1})^2)^{-1/2}, \qquad g_j = c_j a(t_{j-1}, Y_{j-1}) Z_j.$$

For sufficiently large N, all c_j, $1 \leq j \leq N$, are finite and positive. Couple the increments to Brownian motion. On the fixed strip, the gradient and curvature are bounded and Lipschitz. Monotonicity gives, for the left-grid projection π_N,

$$\int_0^T |\gamma(s) - \gamma(\pi_N(s))| \, ds \leq \Delta(\gamma(T) - \gamma(0)).$$

Drift error estimates and discrete Gronwall yield $\mathbb{E} \max_{j \leq N} |Y_j - X_{t_j}|^2 \to 0$, hence $\max_j |c_j - 1| \to 0$. The g_j are orthonormal in L^2: they have unit variance and conditional mean zero given their past. Their predictable multipliers may depend on earlier coordinates, so orthonormality is weaker than independence.

With F_{G} the standard Gaussian distribution function, define

$$F_N = \operatorname{sgn}(u(T, Y_N) + 2F_{\mathrm{G}}(Z_{N+1}) - 1).$$

$$A_j = \mathbb{E}[F_N g_j], \quad B_i = \mathbb{E}[F_N Z_i], \quad A_0 = 0.$$

The extra coordinate is independent and its distribution function is uniform. For any $v \in [-1, 1]$, the random sign of $v + 2U - 1$ has mean v. Therefore $\mathbb{E}[F_N \mid Z_1, \dots, Z_N] = u(T, Y_N)$ exactly. A deterministic sign of the gradient would not preserve this conditional mean. Rounding preserves correlations with functions of the first N coordinates, which is exactly what the two coefficient vectors need. It does not make the whole rounded spin close in L^2 to that conditional mean. Indeed, with $m_N = u(T, Y_N)$,

$$\mathbb{E}[(F_N - m_N)^2] = 1 - \mathbb{E} m_N^2 \longrightarrow 1 - T.$$

Here $m_N \to u(T, X_T)$ in L^2 at fixed T and full support gives $\mathbb{E} u(T, X_T)^2 = T$. The independent rounding variance disappears only in the later endpoint limit.

The approximation

$$M_N = \sum_{j=1}^{N} \sqrt{\Delta} \, a(t_{j-1}, Y_{j-1}) Z_j = u(T, Y_N) + o_{L^2}(1)$$

gives a common residual for both coefficient families. In $\mathbb{E}[M_N Z_i]$, a summand with $j < i$ is independent of centered Z_i, and one with $j > i$ vanishes on conditioning on the first $j-1$ coordinates. The $j = i$ term is $\sqrt{\Delta} \, \mathbb{E} a(t_{i-1}, Y_{i-1})$. Also $\mathbb{E}[M_N g_j] = \sqrt{\Delta}/c_j$ by orthonormality.

Since the union of the two families need not be orthogonal, we apply Bessel's inequality separately to the orthonormal family (g_j) and to the family (Z_i). This gives the vector estimates

$$(A_j)_{j=1}^N = (\sqrt{\Delta}/c_j)_{j=1}^N + o_{\ell^2}(1),$$

$$(B_i)_{i=1}^N = (\sqrt{\Delta} \, \mathbb{E} a(t_{i-1}, Y_{i-1}))_{i=1}^N + o_{\ell^2}(1).$$

A sum of growing length needs these vector estimates, since coordinatewise errors would not control it. Put

$$e_N = \|u(T, Y_N) - M_N\|_2, \quad \alpha_j = \sqrt{\Delta}/c_j, \quad \beta_j = \sqrt{\Delta}\,\mathbb{E}a(t_{j-1}, Y_{j-1}) \quad (1 \le j \le N).$$

The fixed-T Euler argument gives $e_N \to 0$. Bessel's inequality gives $\|A_{1:N} - \alpha\|_2, \|B_{1:N} - \beta\|_2 \le e_N$. Also $\|B_{1:N}\|_2 \le 1$ and $\|\alpha\|_2 \le C_T$ for sufficiently large N, with C_T independent of the mesh. Cauchy–Schwarz gives

$$\begin{aligned}\left|\sum_{j=1}^{N-1}(A_jB_{j+1} - \alpha_j\beta_{j+1})\right| &\le \|A - \alpha\|_2\|B_{2:N}\|_2 + \|\alpha_{1:N-1}\|_2\|B - \beta\|_2 \\ &\le (1 + C_T)e_N.\end{aligned}$$

The unindexed differences in this inequality use coordinates $1, \dots, N$, and restricting to the indicated subvector only decreases their norms. The rounding coordinate contributes at most $|A_NB_{N+1}| \le e_N + \sqrt{\Delta}/c_N$. Thus the dimension-free estimate is

$$\left|\sum_{j=0}^{N} A_jB_{j+1} - \Delta\sum_{j=1}^{N-1}\frac{\mathbb{E}a(t_j, Y_j)}{c_j}\right| \le (2 + C_T)e_N + \frac{\sqrt{\Delta}}{c_N} \longrightarrow 0.$$

The index shift selects the Hessian at t_j, and Z_{N+1} serves the rounding, with no Euler increment of its own. This estimate is a fixed-strip scalar correspondence, and the circuit energy identity comes only in §8.1. Hence

$$\sum_{j=0}^{N} A_jB_{j+1} = \Delta\sum_{j=1}^{N-1}\frac{\mathbb{E}a(t_j, Y_j)}{c_j} + o(1) \longrightarrow \int_0^T \mathbb{E}a(t, X_t)\,dt.$$

The same vector estimates also give an unshifted sum with this limit:

$$\left|\sum_{j=1}^{N} A_jB_j - \sum_{j=1}^{N}\alpha_j\beta_j\right| \le (1 + C_T)e_N \longrightarrow 0,$$

$$\sum_{j=1}^{N} A_jB_j = \Delta\sum_{j=1}^{N}\frac{\mathbb{E}a(t_{j-1}, Y_{j-1})}{c_j} + o(1) \longrightarrow \int_0^T \mathbb{E}a(t, X_t)\,dt.$$

The latter is a left Riemann sum, since $c_j \to 1$ uniformly and the Euler coupling replaces Y_{j-1} by $X_{t_{j-1}}$. Equal continuum limits therefore leave open which finite expression the circuit needs. The rooted energy calculation in §8.1 selects the adjacent pairing, and the two finite pairings can differ, as §10.4 shows.

Only afterward let $T \uparrow 1$, obtaining P_*. The mesh count N is unrelated to the spin-system size n. As in OpenAI (2026a), Proposition 7.2, these coefficients describe a finite Gaussian function. The circuit on finite SK disorder is built only in §8.

The QAOA scalar input also uses the tail estimate $0 \le P_* - \int_0^T \mathbb{E}a(t, X_t)\,dt \le 1 - T$, from non-negative curvature and $\mathbb{E}a^2 = 1$. This estimate and the fixed-strip regularity facts must be carried beside the coefficient limit, since some of them lie outside the selected finite-Gaussian statement. They enter the construction in OpenAI (2026b), §3.

Thus, for every accuracy, the same chosen minimizer supplies a cutoff T and then one finite function F_N with $|S_{T,N} - P_*|$ below that accuracy. This is the finite scalar input. Turning it into spins for a finite SK matrix uses the separate AMP theorem in §9.4, and turning it into a common QAOA word uses §§8.1–8.6. Both operations lie outside the Gaussian limit.

7.1 Error estimates at a fixed cutoff

We fix the minimizer first and then $T < 1$. The constants in the bounds below depend on that minimizer and strip. They give an $N^{-1/2}$ rate there, with no uniformity as $T \uparrow 1$. The derivative bounds are finite, but no numerical values for them are available as inputs to an algorithm that chooses a mesh.

The calculation is a deduction from the source's strip estimates and goes beyond the qualitative selected formal statement. Write $S_{T,N} = \sum_{j=0}^{N} A_j B_{j+1}$ and $f(t) = \mathbb{E}a(t, X_t)$, so $0 \le f(t) \le 1$. On the fixed strip let $K_T = \sup_{t \le T, x} |a(t,x)|$ and choose a spatial Lipschitz constant L_T for a. Define

$$\delta_N = \max_{0 \le j \le N} \|Y_j - X_{t_j}\|_2, \qquad d_N = \max_{1 \le j \le N} \left|\mathbb{E}a(t_{j-1}, Y_{j-1})^2 - 1\right|.$$

These errors have a rate at each fixed strip, even if γ has jumps. Put $\Gamma = \gamma(T)$ and $V = \gamma(T) - \gamma(0)$. Bounded spatial derivatives and the PDE give temporal Lipschitz constants

$$J = \tfrac{1}{2}\|u_{xx}\|_\infty + \Gamma K_T, \qquad H = \tfrac{1}{2}\|a_{xx}\|_\infty + \Gamma(K_T^2 + L_T)$$

for u and a, respectively, on this strip. No derivative of γ is used. Brownian coupling gives $\|X_s - X_{\pi_N(s)}\|_2 \le \sqrt{\Delta} + \Gamma\Delta$. Subtract the Euler and diffusion equations and split the drift difference on $[t_j, t_{j+1}]$ as

$$\gamma(s)(u(s, X_s) - u(t_j, Y_j)) + (\gamma(s) - \gamma(t_j))u(t_j, Y_j).$$

Put $A_T = \int_0^T \gamma(s)\, ds \le 1$ and $w_j = \int_{t_j}^{t_{j+1}} \gamma(s)\, ds$. Taking a pathwise maximum over grid times, Minkowski and discrete Gronwall give, with

$$R_N = A_T K_T \sqrt{\Delta} + [A_T(J + \Gamma K_T) + V]\Delta,$$

the bound

$$\delta_N \le \left\| \max_{j \le N} |Y_j - X_{t_j}| \right\|_2 \le e^{K_T A_T} R_N \le e^{K_T} R_N.$$

The recurrence weights the preceding maximal norm on interval j by $K_T w_j$, and these weights sum to $K_T A_T$. The term $V\Delta$ is the monotone quadrature bound for the coefficient. For the martingale residual, Itô isometry and the same increment estimate give

$$e_N \le K_T \delta_N + \sqrt{T}\{L_T \delta_N + L_T\sqrt{\Delta} + (H + \Gamma L_T)\Delta\}.$$

Thus $\delta_N, e_N = O_{\gamma,T}(\sqrt{\Delta})$. These are deductions from the compact-strip bounds and stochastic identities. The selected formal statement contains no such quantitative clause.

Since a is bounded and Lipschitz, $d_N \le 2K_T L_T \delta_N \to 0$. Whenever $d_N \le 1/2$,

$$\sqrt{2/3} \le c_j \le \sqrt{2}, \qquad |1/c_j - 1| \le d_N, \qquad \|\alpha\|_2 \le \sqrt{T(1 + d_N)}.$$

These estimates give an explicit sufficient condition for all the required normalisers to be positive, and they assume nothing about coarse meshes.

The first-moment identity in §6 makes $f = m$ nonincreasing, with $f(0) = 1$. Its right Riemann sum therefore satisfies

$$0 \le I(T) - \Delta \sum_{j=1}^{N} f(t_j) \le \Delta(f(0) - f(T)).$$

Removing the last term adds $\Delta f(T)$, so $0 \le I(T) - \Delta \sum_{j=1}^{N-1} f(t_j) \le \Delta$. The difference between the shifted interior sum and the integral is bounded by

$$\left| \Delta \sum_{j=1}^{N-1} \frac{\mathbb{E}a(t_j, Y_j)}{c_j} - \int_0^T f(t)\, dt \right| \le TK_T d_N + TL_T \delta_N + \Delta.$$

The first two terms replace the normalisers and Euler positions, and the last is the preceding monotone quadrature estimate. Combining this with the vector and rounding-endpoint estimate gives

$$\begin{aligned} |S_{T,N} - P_*| \le (1-T)m(T) + (2 + \sqrt{T(1+d_N)})e_N + \sqrt{\Delta(1+d_N)} \\ + TK_T d_N + TL_T \delta_N + \Delta. \end{aligned}$$

In particular, for each fixed minimizer and $T < 1$ there are finite $C_{\gamma,T}$ and $N_0(\gamma, T)$ such that

$$|S_{T,N} - P_*| \le (1-T)m(T) + C_{\gamma,T}\sqrt{T/N} \qquad (N \ge N_0(\gamma, T)).$$

The constants are not uniform as $T \uparrow 1$. The coefficient budget bounds $\int \gamma$ and γ on each fixed strip, but the argument gives no numerical bounds for the strip derivative constants and no effective mesh threshold. The scalar tail bound $(1-T)m(T) \le 1 - T$ uses only monotonicity of the first moment, and Proposition 6.2 further makes it $o(1-T)$.

All terms after this scalar tail tend to zero at each fixed $T < 1$. Given a scalar tolerance $\zeta > 0$, first choose T with $1 - T < \zeta/2$, then choose $N_0(\gamma, T, \zeta)$ so that their sum is below $\zeta/2$ for every $N \ge N_0$. This gives the required finite choice of a Gaussian spin. The formula exposes the errors, but it gives no numerical values of K_T, L_T, δ_N, e_N and no effective mesh threshold. OpenAI (2026a), Lemma 2.1; OpenAI (2026a), Proposition 7.2.

These thresholds also let us choose a diagonal. For prescribed $T_m \uparrow 1$, choose increasing N_m large enough to make the mesh-dependent part of the fixed-T_m bound less than $1/m$. An arbitrary increasing mesh schedule does not follow from the iterated limit alone. For example, the scalar array

$$h(T, N) = \min\{1, [N(1-T)]^{-1}\}$$

tends to zero in N at every fixed $T < 1$, but $h(1 - 1/N, N) = 1$ for every $N \ge 2$. The array only illustrates the quantifier issue and asserts no error law for the SK construction.

8. From the scalar formula to one ordinary QAOA word

On a single spin, let $Z = \mathrm{diag}(1, -1)$ and $X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The operators Z_i and X_i act on spin i and as the identity on the other spins. Also set $Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ and $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Then $X|\pm\rangle = \pm|\pm\rangle$ and $Z|+\rangle = |-\rangle$. The cost and mixer are

$$C_{n,J} = n^{-1/2} \sum_{i<j} J_{ij} Z_i Z_j, \qquad B_n = \sum_i X_i.$$

The two evolutions act differently:

$$e^{-i\theta C_{n,J}}|\sigma\rangle = e^{-i\theta H_n(\sigma)}|\sigma\rangle, \qquad e^{-i\beta B_n} = \prod_{i=1}^{n} (\cos\beta\, I - i \sin\beta\, X_i).$$

The cost evolution changes phases while preserving spin-basis probabilities, and the mixer combines amplitudes of different spin configurations. Cost evolution alone leaves the initial uniform probabilities unchanged, and a mixer applied first only multiplies $|+\rangle^{\otimes n}$ by $e^{-in\beta}$. Alternating them lets the cost-dependent phases change subsequent measurement probabilities. Farhi, Goldstone, Gutmann and Zhou (2022), §2, equations (1)–(6).

A depth-p state, with p counting cost–mixer pairs, is

$$|\psi_{n,p,J}(\gamma,\beta)\rangle = e^{-i\beta_p B_n} e^{-i\gamma_p C_{n,J}} \cdots e^{-i\beta_1 B_n} e^{-i\gamma_1 C_{n,J}} |+\rangle^{\otimes n}.$$

Here the real angle vectors γ, β are deterministic, and γ is separate from the Parisi coefficient $\gamma(t)$. A *word* means a finite ordered product of cost and mixer exponentials. Its pattern and angles are fixed independently of n, J, while its cost operator depends on n, J. Define

$$v_p(\gamma,\beta) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_J \langle \psi_{n,p,J} | C_{n,J} | \psi_{n,p,J} \rangle, \qquad Q_p = \sup_{\gamma,\beta\in\mathbb{R}^p} v_p.$$

The fixed-depth, fixed-angle infinite-size limit is the input from Farhi, Goldstone, Gutmann and Zhou, and the later depth conclusion lies beyond it; see Farhi, Goldstone, Gutmann and Zhou (2022). The source target is $\lim_{p\to\infty} Q_p = P_*$. Its needed lower bound has the literal quantifiers

$$\forall \varepsilon > 0 \quad \exists p \in \mathbb{N}, \gamma, \beta \in \mathbb{R}^p : \quad \lim_{n\to\infty} \frac{1}{n} \mathbb{E}_J \langle \psi_{n,p,J} | C_{n,J} | \psi_{n,p,J} \rangle \geq P_* - \varepsilon.$$

The same finite list of gates and angles serves every n, J, and the conclusion concerns the limiting expected energy. A cost expectation is bounded above by its maximum eigenvalue, and zero-angle padding makes Q_p nondecreasing. It therefore suffices to construct one finite word for every accuracy. Fixed-instance convergence at increasing depth is a different quantifier order. Farhi, Goldstone and Gutmann's original QAOA argument obtains that convergence by adiabatic evolution and its Trotter approximation, with depth allowed to depend on the instance. That argument gives no size-independent word of the kind required here. Farhi–Goldstone–Gutmann (2014), equation (10) and §VI, pp. 10–11; OpenAI (2026b), Theorem 1.1.

The construction first describes local operations on a large tree, where neighbouring subtrees supply Gaussian inputs. Temporary random labels let those operations implement the finite classical function of §7. The hard steps are to implement its aggregations using cost and mixer gates, and then to replace every use of the labels by ordinary gates while preserving the energy. Only the resulting ordinary word is transferred to SK.

The choices are nested: fix ε, choose $T < 1$, choose a finite mesh, fix the bounded classical formula and helper probabilities, then choose the finite gate approximations and remove the seeds. The constants in a replacement may depend on the surrounding word and on those positive helper probabilities. §8.6 collects the error allocation before the final $n \to \infty$ limit. We state the principal circuit inputs as four local lemmas. Their proofs are in the cited manuscript, and §§8.2–8.5 reconstruct the mechanisms and show where each hypothesis enters.

Smooth replacement. Fix the height, a finite expression and channel list, and the required derivatives. Uniform recursive weighted bounds, convergence of operators and adjoints, and strongly commuting signals with the stated joint exponential bounds permit substitution inside a fixed surrounding sequence in the exclusion-gate class defined in §8.2. Constants depend on that fixed family. This is the input used in §8.3 to replace classical aggregates by finite gates; see the foundations, smooth topology and gate calculus.

Helper contraction and synthesis. Fix a bounded, finite, type-excluded odd formula, zero on helpers, and positive helper masses. The contraction converges smoothly at every cut and for the specified responses. For each $\eta > 0$, finite seeded synthesis gives

$$\mathcal{E}(v) \geq \mathcal{E}(m) - 2\eta - 2\sqrt{p_H}.$$

Section 8.4 starts from this one fixed seeded word. The source inputs are the synthesis statements and helper theorem.

Seed removal. Fix that finite seeded word and an energy tolerance. Reuse the same Gaussian at every occurrence on a vertex. Packet control and the one-particle comparison then produce one finite ordinary word with the required energy error. The seed-removal theorem supplies an energy conclusion, which is stronger than convergence of seed marginals, and so makes the ordinary-word transfer in §8.5 applicable.

Fixed-word transfer. Fix one ordinary word and all its angles before either size or degree varies. Its positive tree scaled-edge limit equals its SK expected-energy limit, with the convention conversion in §8.5. This fixed-word transfer closes the thermodynamic-first theorem in §8.6.

One cost–mixer pair. A single cost–mixer pair fixes the energy normalization before we introduce the rooted representation. Use the positive cost $C_{n,J}$ of §8 and $U = e^{-i\beta B_n} e^{-i\theta C_{n,J}}$, where the symbol θ is a cost angle. Put $c = \cos(2\beta)$ and $s = \sin(2\beta)$. The Pauli identities are

$$e^{i\beta X} Z e^{-i\beta X} = cZ + sY, \qquad e^{i\alpha Z_i Z_j} Y_j e^{-i\alpha Z_i Z_j} = Y_j \cos(2\alpha) + Z_i X_j \sin(2\alpha).$$

The mixed term $Z_i Y_j$ therefore has cost-state expectation

$$\sin(2\theta J_{ij}/\sqrt{n}) \prod_{k \neq i,j} \cos(2\theta J_{jk}/\sqrt{n}).$$

The other mixed term exchanges i and j. The $Z_i Z_j$ term has zero expectation. The $Y_i Y_j$ term can be nonzero on an individual complete graph, but its expectation after the cost gate is independent of J_{ij}, because $Y_i Y_j$ commutes with that edge cost. Multiplying by J_{ij} and averaging therefore removes it, even in the presence of triangles. For a standard Gaussian J,

$$\mathbb{E}\cos(aJ) = e^{-a^2/2}, \qquad \mathbb{E}[J \sin(aJ)] = a e^{-a^2/2}.$$

Averaging the mixed terms and summing the $n(n-1)/2$ unordered edges gives

$$\frac{1}{n} \mathbb{E}_J \langle C_{n,J} \rangle = \frac{n-1}{n} \theta \sin(4\beta) e^{-2\theta^2 (n-1)/n}.$$

This agrees with Farhi et al., equation (62). With energy maximization, $\theta = 1/2$, $\beta = \pi/8$ gives the limiting value $1/(2\sqrt{e})$, and their minimization example uses $\beta = -\pi/8$ and the negative value.

For the positive cost $D^{-1/2} \sum_{\{i,j\}} Z_i Z_j$ on the $(D+1)$-regular tree, each endpoint has D other neighbours. The same calculation now has no common-neighbour contribution and yields

$$\langle Z_o Z_v \rangle = \sin(4\beta) \sin(2\theta/\sqrt{D}) \cos^D(2\theta/\sqrt{D}).$$

Multiplying by $\sqrt{D}/2$ and letting $D \to \infty$ gives $\theta e^{-2\theta^2} \sin(4\beta)$, the same limit as finite SK. This confirms the factor $1/2$ and the degree convention in §8.5.

We recover the value just obtained once more from the rooted insertion in §8.1.1, which shows which part of the limiting state carries the energy.

8.1 Rooted realisation and the observable it preserves

The fixed-depth spin–boson representation has an antecedent in Boulebnane et al., Theorem 2 (informal), equations (10)–(11), p. 3: as the degree grows with depth fixed, the symmetric tree state converges to a spin coupled to finitely many bosonic modes. The QAOA manuscript credits that representation and then proves the occupation, insertion and surrounding-circuit estimates needed below. These estimates control the energy observable and substitutions inside a circuit, which goes beyond identifying a limiting state. OpenAI (2026b), §2.

We now realise the finite Gaussian spin from §7 in a space where its shifted correlation sum becomes an energy observable. We identify that sum with a rooted quadratic form, and then relate the same form to the scaled edge energy of a fixed circuit. Section 8.5 supplies the final transfer to SK.

For an ordinary depth-p word, this tree calculation uses only finitely many spins. Conjugate Z_iZ_j backwards through the circuit. A mixer rotates the spins already in its support without enlarging that support. In a cost layer all edge factors commute, so factors disjoint from the current support commute past the other factors and cancel. The remaining factors enlarge support by at most one graph step. After p pairs, the edge expectation therefore depends only on its radius-p backward light cone. On the $(D+1)$-regular tree this contains at most $2\sum_{r=0}^{p} D^r$ vertices, and the spins outside factor out of the product-state expectation.

Thus fixed depth gives fixed tree height, although the number of children grows with D. This locality explains the finite tree calculation, and the equality with dense SK is the separate transfer theorem of §8.5. Farhi, Goldstone and Gutmann (2014), §II, equations (14)–(26).

On a rooted tree, factor a circuit as $U = (\bigotimes_j U_j)u$, where each U_j contains operations wholly in child subtree j. Conjugating a root observable P reduces to u^*Pu, and its insertion vector is $u^*Pu\,\mathrm{vac}$, centered when its vacuum mean is zero. This factorization explains how child operations enter the recursion below.

The first change of setting preserves a specified quadratic form. At height zero the pointed space is $\mathsf{k} = \mathbb{C}^2 \otimes L^2(\text{seed})$, with distinguished vector $\omega = |+\rangle \otimes 1$. Recursively,

$$\mathcal{H}_0 = \mathsf{k}, \qquad \mathcal{H}_{h+1} = \mathsf{k} \otimes \Gamma_s(\mathcal{H}_h^\circ),$$

where $\mathcal{H}_h^\circ$ is the orthogonal complement of the vacuum and Γ_s is symmetric Fock space. A particle here is one child subtree whose state differs from its vacuum. It is not one physical spin. The occupation embedding identifies a symmetric state with m nonvacuum children among D available children with its m-particle vector, using the factor $\binom{D}{m}^{-1/2}$. Adding empty descendants gives compatible embeddings between heights. OpenAI (2026b), §2, pointed spaces and occupation embeddings.

For a child direction f, write $q(f) = a^*(f) + a(f)$ and $W(f) = e^{iq(f)}$. Explicitly, $\Gamma_s(H) = \bigoplus_{m\ge 0} H^{\otimes_s m}$, with vacuum Ω in sector zero. On sector m, $a^*(f)\psi = \sqrt{m+1}\,\mathrm{Sym}(f\otimes\psi)$, and $a(f)$ is its adjoint. The number operator is multiplication by m, and more generally $d\Gamma(A)$ sums A over those m child factors. The vacuum characteristic function $\langle\Omega, e^{itq(f)}\Omega\rangle = e^{-t^2\|f\|^2/2}$ fixes the Gaussian normalization.

For any finite list of fixed centred real child vectors, their parent positions $\mathsf{G}(f) = q(f)$ are jointly centred Gaussian, and for f, g,

$$\mathbb{E}[\mathsf{G}(f)\mathsf{G}(g)] = \langle f, g\rangle, \qquad \|[\mathsf{G}(f) - \mathsf{G}(g)]\mathrm{vac}\| = \|f - g\|.$$

The directions are fixed elements of the child space, even when their values depend on the child's Gaussian coordinates. The receiving root's own seed is a separate, independent tensor factor.

The root shift

$$\mathsf{L}f = \omega \otimes a^*(f)\Omega$$

is an isometry on centred vectors, and is extended by zero on the vacuum. Its adjoint extracts the root-vacuum, one-child component, including a constant receiving-root seed. For a classical function of orthonormal Gaussian positions, this is its first Gaussian chaos. The next calculation uses the covariance, independence and projection properties in distinct roles. OpenAI (2026b), §2, root shift and classical first chaos OpenAI (2026b), §2, classical Gaussian positions.

We start with the standard root seed $g_0 = S$ and recursively set $Z_{j+1} = \mathsf{G}(g_j)$. Suppose $g_0, \dots, g_{j-1}$ are centred orthonormal vectors. The joint Gaussian law of their lifts and the covariance identity give independent standard normals $Z_1, \dots, Z_j$. In the scalar construction of §7, the multiplier of Z_j in g_j depends only on its predecessors. Conditional centring therefore makes g_j orthogonal to $g_1, \dots, g_{j-1}$, and the chosen normaliser c_j gives unit norm. Independence of the receiving-root seed gives orthogonality to g_0. This closes the induction. Compatible height embeddings ensure that the earlier variables are the same variables when the formula is embedded at greater height.

Oddness of u and evenness of a make Y_j and g_j odd under simultaneous seed reversal. The extra centred rounding variable $2F_{\mathrm{G}}(Z_{N+1}) - 1$ is odd as well, so F_N is centred. It is independent of the receiving-root seed, hence $\langle F_N, g_0\rangle = 0 = A_0$. With A_j, B_{j+1} as in §7, first-chaos projection gives

$$\mathsf{L}^* F_N = \sum_{j=0}^{N} B_{j+1} g_j, \qquad \mathcal{E}(F_N) := \mathrm{Re}\langle F_N, \mathsf{L}F_N\rangle = \sum_{j=0}^{N} A_j B_{j+1}.$$

Thus the construction identifies both the scalar law and the required rooted energy. The Gaussian covariance rule does not declare distinct calculations on shared descendant subtrees independent. OpenAI (2026b), §3, "A finite Gaussian formula and its rooted energy".

Connecting this energy to an edge observable needs a separate statement. Let f_D be centred, invariant under rooted permutations of complete child subtrees, and supported at a fixed finite radius. Suppose its occupation embedding converges strongly to v, with a uniformly bounded expected top-generation occupation $\langle f_D, \mathsf{N} f_D\rangle$. Put $M = D + 1$ and $R_M = \sqrt{1 - \mathsf{N}/M}$.

To obtain the decomposition, we delete the distinguished edge and project onto the vacuum on either resulting half-tree. There are four pairings. The vacuum–vacuum pairing is zero by centring. If only one side is excited, the other side forces the root and all remaining branches into their vacuum, and root symmetry gives the term $M^{-1/2}\langle \mathsf{L}^* f_D, R_M f_D\rangle$ or its conjugate. In the pairing where both sides are excited, each projected vector has squared norm $M^{-1}\langle f_D, \mathsf{N} f_D\rangle$ by symmetry among the M branches. Cauchy–Schwarz bounds their pairing by that same quantity. Finally, deleting one vacuum child multiplies the m-particle sector by $\sqrt{\binom{M-1}{m}/\binom{M}{m}} = \sqrt{1 - m/M}$, which identifies R_M. Thus the exact edge decomposition is

$$\langle f_{D,i}, f_{D,j}\rangle = \frac{2}{\sqrt{M}}\,\mathrm{Re}\langle \mathsf{L}^* f_D, R_M f_D\rangle + \varepsilon_D, \qquad |\varepsilon_D| \le \frac{\langle f_D, \mathsf{N} f_D\rangle}{M}.$$

Vacuum deletion also gives $\|R_M f_D - f_D\|^2 \le M^{-1}\langle f_D, \mathsf{N} f_D\rangle$. Consequently,

$$\lim_{D\to\infty} \frac{\sqrt{D}}{2}\langle f_{D,i}, f_{D,j}\rangle = \mathcal{E}(v).$$

Strong convergence alone does not control the scaled remainder, and the occupation estimate for the relevant finite circuits comes later. Norm convergence of the vectors cannot replace the occupation assumption. For orthogonal unit vectors e_m in sector m, $\psi_k = \sqrt{1-1/k}\,\Omega + k^{-1/2}e_{k^2}$ tends to Ω in norm, whereas $\langle \psi_k, \mathsf{N}\psi_k\rangle = k$.

The elementary continuity bound

$$|\mathcal{E}(v) - \mathcal{E}(w)| \le (\|v\| + \|w\|)\|v - w\|$$

controls every later approximation once the observable has been identified. OpenAI (2026b), §2, "Energy as a root shift," including its exact finite-degree decomposition.

The three identifications use different information:

First, passing from the finite Gaussian formula to the rooted classical sector preserves the joint Gaussian law and the same F_N. This uses covariance, conditional centring and independence of the receiving-root seed.

Second, the root-vacuum and first-chaos projection identify the energy itself: $\mathcal{E}(F_N) = S_{T,N}$.

Third, for a finite-degree circuit insertion tending to v, root invariance, fixed radius, strong convergence and uniform occupation control imply that the scaled edge expectation tends to $\mathcal{E}(v)$.

We gave the first two calculations above. The third uses the source's edge decomposition, and the circuit estimates it requires enter in §8.2.

8.1.1 The same energy in the rooted representation We return to the value $\theta e^{-2\theta^2}\sin(4\beta)$ obtained in the finite-SK and tree calculation. We keep its cost angle θ, mixer angle β, $c = \cos(2\beta)$ and $s = \sin(2\beta)$. Here Ω is the Fock vacuum, $\mathrm{vac} = |+\rangle \otimes \Omega$, and $q(e) = a^*(e) + a(e)$. The root shift L defined in §8.1 creates one child excitation and puts the receiving root in its vacuum. Let $e = Z\mathrm{vac} = |-\rangle \otimes \Omega$ and $Q = q(e)$ in the rooted description. Its vacuum law is standard normal. The depth-one root operation is $u = e^{-i\beta X}e^{-i\theta ZQ}$. Using $Y|+\rangle = -i|-\rangle$, its insertion is

$$v = u^* Z u\,\mathrm{vac} = |-\rangle \otimes [c - is\cos(2\theta Q)]\Omega + |+\rangle \otimes s\sin(2\theta Q)\Omega.$$

The root shift retains only the plus-root, one-particle component. The first-chaos coefficient of the sine is $\mathbb{E}[G\sin(2\theta G)] = 2\theta e^{-2\theta^2}$, so

$$\mathsf{L}^* v = 2\theta s e^{-2\theta^2} e, \qquad \mathrm{Re}\langle v, \mathsf{L}v\rangle = 2\theta s c e^{-2\theta^2} = \theta e^{-2\theta^2}\sin(4\beta).$$

Here $\mathrm{Re}\langle e, v\rangle = c$ supplies the final factor. Thus the observable itself, as well as the Gaussian law, agrees in all three representations. The example uses the ordinary root updates and root shift. It calibrates one depth, and at higher depth the occupation and replacement hypotheses remain necessary.

8.2 Finite regularisation and the calculus needed for compilation

We fix T, N first. The independent continuous rounding coordinate makes the event at the sign discontinuity null. Bounded continuous odd approximations to that sign therefore converge in L^2 by dominated convergence. If the bounded output is m^*, shrink it to $(1-\kappa)m^*$, with fixed $0 < \kappa < 1$, before taking $\theta_\kappa = \frac{1}{2}\arcsin((1-\kappa)m^*)$. This margin keeps the derivative of inverse sine bounded on its used range, and the output loss is at most κ in L^2.

We then truncate the transmitted functions and approximate the resulting bounded maps by trigonometric polynomials. The source makes these choices from output toward leaves: each outer Lipschitz function is fixed before its inputs are approximated. The Gaussian isometry above converts an error in a transmitted child vector into the same L^2 error in its receiving position. Energy continuity then bounds the accumulated loss. The bounded output has the form $m = \sin(2\theta)$. The transmitted columns are bounded, while the Gaussian fields themselves remain unbounded. OpenAI (2026b), §3, bounded trigonometric approximation.

The formulas to be synthesized. A mask is an event in the local seed space invariant under seed reversal. Fix a finite partition into masks, and write p_R for the probability of mask R. Probabilities used as denominators are positive. For disjoint masks R, S, aggregation on the finite regular tree means

$$(A_{RS}f)_i = \mathbf{1}_R(i)D^{-1/2}\sum_{j\sim i}\mathbf{1}_S(j)f_j.$$

A formula has a finite expression tree, with bounded measurable functions of the local seed as leaves. Internal operations are such aggregations or pointwise functions of previously formed quantities. Pointwise functions are smooth in their aggregate arguments, with polynomial bounds on all derivatives, uniformly in the seed. Seed dependence itself need not be smooth. Transmitted quantities are odd under simultaneous seed reversal, and sent nonlinear functions and output angles are bounded, while aggregate Gaussian fields may be unbounded.

An aggregation from S to R excludes R if its transmitted formula uses no vertex of type R below that call. All ancestor exclusions are retained, so a mask cannot recur on one active expression branch. It may recur on a different branch, and distinct branches may share descendants, since exclusion asserts no independence between them. Gate targets are likewise excluded from the cross-edge dependencies of their parameters. This makes the physical all-neighbour calculation agree with the intended descendant calculation.

For synthesis, we reserve disjoint helper masks of positive mass, one per positive aggregation depth and absent from the output dependencies. The bounded odd output angle vanishes on their union. OpenAI (2026b), Definition 3.3; Theorem 4.1.

The types can be encoded together with a Gaussian seed in one raw Gaussian variable: its sign supplies the odd symmetry, while disjoint binary digits of a uniform transform of its absolute value supply independent coordinates. The resulting finite type label is even and the Gaussian seed is odd under reversal. The associated pullback preserves vacuum, parity and L. Thus adding types preserves the energy identification and needs no additional circuit resource. OpenAI (2026b), §3, colours, exclusions and helpers.

For the already fixed formula, the exclusion estimate is

$$\|m^{\mathrm{col}} - m\|^2 \le C(\alpha + p_H),$$

where α is the largest nonhelper type mass and p_H is the total helper mass. Its constant is independent of the number of types. The proof uses orthogonality of the disjoint masks before applying the Lipschitz estimate in the original, fixed list of formula coordinates. Treating the refined types as an arbitrary growing list of independent arguments would not provide this bound. The resulting classical near-optimality proposition supplies a finite bounded odd formula, zero on helpers, with energy as close to P_* as required. OpenAI (2026b), §3, "Type exclusion at small cost" and "Classical near-optimality in synthesis form".

The topology of the later substitutions. The gate calculus uses recursive weights $\mathsf{W}_{h+1} = \mathsf{W}_0 + d\Gamma(\Lambda_h)$, where Λ_h is the previous weight restricted to the centred child space. The local weight includes the unbounded raw seed when needed. Its smooth domain is the intersection of all weighted domains. Smooth bounds mean polynomial bounds, with a finite loss of weight powers, for the maps, their adjoints and every specified parameter derivative. The finite height, fixed expression and fixed number of channels matter in these estimates.

Strong convergence together with uniform stronger weighted bounds gives convergence in each weaker weighted norm by a spectral cutoff. This is the topology in which one replaces a gate inside a fixed surrounding circuit. OpenAI (2026b), §2, weights and the topology of convergence.

More precisely, the imported smooth-convergence contract is: for each output power m and parameter multi-index α, there are an input power r, exponent k and constant C such that

$$\|\mathsf{W}_D^m \partial_x^\alpha F_D(x)\xi\| \le C(1+|x|)^k \|\mathsf{W}_D^r \xi\|, \qquad \xi \in \bigcap_{q\ge 0} \mathrm{Dom}(\mathsf{W}_D^q).$$

The bound is uniform in degree D and parameter x. The same bound and strong convergence on smooth inputs are required for the adjoints and all specified parameter derivatives, with compatible limiting adjoints. Constants may depend on the fixed expression, height, channel list and derivative order. Bounded adjoint derivatives alone do not supply their convergence. At bounded local seeds $\mathsf{W}_0 = 1$, and for polynomial Gaussian growth the source uses the vacuum-preserving weight $P_0 + Q_0(1+|s|)Q_0$, where P_0 projects to the local vacuum and $Q_0 = 1 - P_0$. These recursive weights control every generation.

They differ from the top-generation number weight in §8.4's packet argument. No uniformity in growing word length, vanishing helper mass or the later seed-removal parameter is implicit in this contract. OpenAI (2026b), §2, smooth-family bound.

One of the essential inputs retains a term beyond the Gaussian field. Let $K_D(x)$ be a smooth unitary family with $K_D(0) = 1$, uniformly controlled derivatives and adjoint derivatives that converge smoothly in the stated weighted topology to those of a limiting family K, and centred first vacuum derivatives. Write $K_\alpha = \partial_\alpha K(0) = iT_\alpha$ and $K_{\alpha\lambda} = \partial_\alpha \partial_\lambda K(0)$. For a fixed finite parameter list a, the source proves

$$K_D(a/\sqrt{D})^{\otimes D} \longrightarrow \exp\left\{ i\sum_\alpha a_\alpha q(T_\alpha \mathrm{vac}) + \frac{1}{2}\sum_{\alpha,\lambda} a_\alpha a_\lambda \left\langle K_{\alpha\lambda} + \tfrac{1}{2}\{T_\alpha, T_\lambda\}\right\rangle_0 \right\}.$$

The second term is imaginary by unitarity. It survives because a second-order change of size D^{-1} occurs in each of D children. Its scalar coefficients become a possibly nonconstant controlled operator when commuting signals replace a. For example, $K(x) = e^{ix^2}$ has zero first derivative at zero but $K(a/\sqrt{D})^D = e^{ia^2}$ exactly. Keeping only the first derivative would miss even this elementary contribution. The example explains why the term is necessary, and the source's weighted product lemma supplies the general convergence and domain bounds. OpenAI (2026b), §2, "Products of small unitaries".

The smooth-substitution lemma makes the required joint control explicit. For a fixed finite list of strongly commuting self-adjoint signals $S_D = (S_{D,1}, \dots, S_{D,b})$, their polynomials must converge smoothly, and their joint exponentials must converge strongly on smooth inputs and satisfy, for every output weight m,

$$\|\mathsf{W}_D^m e^{it\cdot S_D}\xi\| \le C_m(1+|t|)^{k_m}\|\mathsf{W}_D^{r_m}\xi\|,$$

uniformly in D and $t \in \mathbb{R}^b$. The coefficient family $F_D(x)$ and its adjoint obey the smooth bounds above and converge with all their jets on compact parameter sets. Each coefficient must commute with the signals' spectral projections. For unbounded coefficients this means commutation in the closed-operator sense. Seed-dependent substitution also requires strong commutation with the receiving seed multiplication algebra. These hypotheses permit the joint spectral and Fourier calculation of $F_D(S_D)$, including its derivatives and adjoints. A formal commutator identity on a single domain would not supply them. No nonsingular covariance matrix or inverse Gram matrix is required. OpenAI (2026b), Lemma 2.10, "Smooth commuting substitution".

For the formula fixed here, its finite expression tree supplies the finite channel list. Exclusion supplies the commuting physical control algebra, and normalized-sum limits and the preceding product estimates supply the signal limits and weighted bounds. Conjugation by the already fixed preceding circuit preserves strong commutation, including when signals share descendants. The gate-calculus theorem applies these facts inductively to the formula and its adjoints, at each cut and inside the fixed surrounding sequence. It is this theorem, with these constructed inputs, that licenses the later replacements. Its conditional root-seed bounds also give the masked-column estimates used in §8.3. OpenAI (2026b), Theorem 2.12.

The exact cut recursion records the effect of such shared lower calculations. Let y be the incoming boundary data, $V(y)$ the root matrix, and $m(y)$ its outgoing signals in the current frame. Factoring out the operations wholly inside each child gives

$$E^{[d]}(y) = V(y) \prod_{j \text{ child}} K_j^{[d-1]}(D^{-1/2} m(y)), \qquad K^{[d]}(y) = u^* E^{[d]}(0)^* E^{[d]}(y) u.$$

Every ratio $K^{[d]}$ equals one at zero. Nevertheless $m(0)$ generally does not vanish, because removing the parent input leaves the root's own signals. The actual zero-boundary root update is $u_{\text{new}} = E^{[d]}(0)u$. It is not recovered by setting the entire ratio to one and discarding the baseline operation. Applying the unitary product lemma to this recursion, and then commuting substitution, is the source's smooth gate calculus. Its replacement statement covers a fixed finite surrounding sequence and its required parameter derivatives. It asserts no uniformity for arbitrary sequences whose length or channel count grows. OpenAI (2026b), §2, smooth gate calculus and its cut recursion.

8.3 Seeded synthesis: helper contraction and nested responses

We temporarily add longitudinal seed gates $\exp(-it \sum_i s_i Z_i)$, using an independent Gaussian s_i at each vertex and reusing that same variable whenever a later seed gate acts there. The synthesis theorem turns the finite classical formula into a finite word of these gates, costs and mixers. Its ideal local Y rotations have Z insertion

$$v_{\text{cl}} = |+\rangle m + |-\rangle n, \qquad m = \sin(2\theta).$$

Here n is the remaining seed-dependent coefficient of the $|-\rangle$ component, not the number of spins. Both coefficients may depend on descendant seeds. The second term does not contribute to $\mathcal{E}$: L puts the root spin in $|+\rangle$, and shifting the $|-\rangle$ term puts a spin excitation into a child, orthogonal to the classical sector. Thus $\mathcal{E}(v_{\text{cl}}) = \mathcal{E}(m)$. OpenAI (2026b), §2, classical rotations OpenAI (2026b), §4, "Classical synthesis".

There are two different approximation operations. First, seed conjugations of X generate $\cos(2ts)X$ and $\sin(2ts)Y$. Trigonometric approximation and matrix commutators give the required bounded measurable masked spin controls. Sign averaging of masked X flips isolates an interaction between

chosen types. Since the cost terms commute, this isolation is exact. This sign averaging is the refocusing method of Jones and Knill (1999), §II, p. 2: orthogonal sign rows cancel unwanted Ising couplings, while identical rows retain a selected coupling. The source applies it to the finite partition of seed masks.

Second, finite commutator products then approximate the desired pointwise matrices with smooth bounds. They can be substituted after the surrounding circuit has been fixed. The local masks are measurable functions of the seed. Smoothness is required only in the continuous aggregate inputs, and the mask labels need none. OpenAI (2026b), §4, elementary masked controls and matrix products.

The second operation implements an aggregation. Let I be the receiver region, J the sender region, and K an unused helper region of fixed positive probability p_K. They are disjoint. The lower seed-only formula f excludes these regions along its cross-edge paths. With $A_{ij} = D^{-1/2}1_{i\sim j}$, the desired propagation has generator

$$H = \sum_{i\in I, j\in J} P_i A_{ij} f_j.$$

Partition J into parts J_l of probabilities p_l. Two levels of spin commutators generate the commuting Hamiltonian

$$\widetilde{H}_{\mathcal{P}} = \frac{1}{p_K}\sum_l \sum_{i\in I,\, j,j'\in J_l,\, k\in K} P_i A_{ij} X_j A_{jk} X_k A_{kj'} X_{j'} f_{j'}.$$

For a fixed receiver–sender edge $i \sim j$, the terms with $j = j'$ contain $X_j^2 = 1$, and f_j is seed-only and commutes with these spin factors. On helper-pinned states $X_k = 1$, so their sum is exactly

$$\frac{1}{p_K} P_i A_{ij} f_j \sum_{k\in K} A_{jk}A_{kj} = P_i A_{ij} f_j \frac{N_K(j)}{Dp_K},$$

where $N_K(j)$ counts helper neighbours of j. Conditional on the receiver edge, its endpoint $i \in I$ is not a helper and the other D neighbours have mean helper count Dp_K. Its binomial variance is $Dp_K(1-p_K)$, so the ratio tends to one in L^2 at fixed $p_K > 0$. Terms with $j \neq j'$ are simple paths and must disappear under partition refinement. The factor $1/p_K$ is harmless only after a positive helper mass has been fixed. OpenAI (2026b), §4, helper generators and contraction.

Helper pinning is a statement about every helper spin in the light cone. Its finite-degree projection is

$$\Pi_{K,V} = \prod_{v\in V}\left[1 - 1_K(s_v)\frac{1-X_v}{2}\right].$$

It retains the seed variables and helper-labelled descendant spaces while fixing the helper spins in $|+\rangle$. The ideal propagation, the exact Hamiltonian $\widetilde{H}_{\mathcal{P}}$, and the surrounding gates that avoid K preserve this compressed space. Multiplying an insertion by $1_{K^c}(s_o)$ only masks the root label and does not impose this descendant condition. Individual finite commutator factors may leave the compressed space. Accordingly, the proof first establishes the exact contraction there, then chooses finite commutator products on the full space to approximate it. OpenAI (2026b), §4, "Simultaneous helper contraction".

The estimate controlling a growing partition is

$$\rho^2 = \sum_l p_l^2 \le p_J \max_l p_l \longrightarrow 0.$$

At a helper root, each part supplies two commuting fields with columns d_l, e_l on disjoint sender masks, of weighted norm at most $C\sqrt{p_l}$. The remaining correction is a sum of centred quadratics

$$\sum_l [q(d_l)q(e_l) - \langle d_l, e_l\rangle].$$

Normal ordering and disjoint-mask orthogonality bound its quadratic kernels by $C\rho$ in Hilbert–Schmidt norm. The source separately bounds the exponentials in the weighted topology, uniformly in the number of parts. Smallness of the generator on smooth vectors alone would not justify inserting its exponential into the surrounding circuit. OpenAI (2026b), §4, "Quadratics with many orthogonal columns".

The helper estimate alone does not control all cuts. Put $\epsilon_{\mathcal{P}} = \max_l \sqrt{p_l}$. The receiver and helper corrections have weighted size $O(\rho)$, and a sender correction is $O(\sqrt{p_l})$ conditional on its label. The remaining lower-formula responses have size $O(\epsilon_{\mathcal{P}})$. Both quantities tend to zero with f, the number of lower channels and $p_K > 0$ fixed.

There remains a response inside the lower formula f. A cut can pass through a descendant used to compute f, rather than through one of the displayed three-edge paths. The source first writes the exact difference of the affected Hamiltonian terms before applying the preceding subtree circuit. That circuit is fixed with respect to the new cut input. Exclusions ensure that the downward coefficients are independent of that active boundary argument. The resulting additive cut-response rule retains both first derivatives and mixed second derivatives:

$$\sum_\alpha a_\alpha q(H_\alpha \mathrm{vac}) + \frac{1}{2}\sum_{\alpha,\lambda} a_\alpha a_\lambda \langle H_{\alpha\lambda}\rangle_0.$$

Here H_α is centred by parity, so no limit of the extensive zero-input Hamiltonian is required. The self-adjoint generator is obtained from the limiting time group of cut ratios. OpenAI (2026b), §2, "Additive cut responses".

After substituting the signals of a lower formula, a boundary difference contains, in particular,

$$\begin{aligned}&\sum_\alpha [m_\alpha(y) - m_\alpha(0)]q(r_\alpha)\\ &\quad + \frac{1}{2}\sum_{\alpha,\lambda}[m_\alpha(y)m_\lambda(y) - m_\alpha(0)m_\lambda(0)]c_{\alpha\lambda}.\end{aligned}$$

The mixed coefficients $c_{\alpha\lambda}$ must remain even when two channels share descendants. A bounded scalar example makes the issue visible: for a shared standard normal Q, take $H(y_1, y_2) = \cos(Q + y_1 + y_2) - \cos Q$. At zero both first derivatives are $-\sin Q$ and are centred, but $\mathbb{E}H_{12} = -e^{-1/2}$. The two cross terms in the half-sum therefore contribute $-e^{-1/2}a_1a_2$. Centred first responses do not remove a mixed second response. This example checks the algebraic reason for retaining the term, and the cut-response theorem supplies the operator and domain justification.

At the actual root there is no parent input to subtract: the corresponding unsubtracted $m(0)$ terms enter the root update. The source propagates small terminal responses through the fixed depth and fixed list of lower channels. It refines the calling sender mask while keeping those lower formulas fixed. This is what controls all cuts in a simultaneous replacement of several calls. A root-seed scalar phase may remain on helper labels, where it cancels in insertions and child conjugations. OpenAI (2026b), §4, lower-formula responses and completion of the contraction proof.

Compilation now proceeds from outer aggregations to inner ones. Reserve a separate unused helper type for each aggregation rank. At a given rank, establish the simultaneous exact contraction, fix the sender partitions, and only then choose finite commutator approximations. Subsequent inner calls use their own helpers, which remain unused. Each inner accuracy is chosen after its surrounding finite sequence is fixed. This order avoids assuming uniformity over all helper masses, ranks or circuit lengths. OpenAI (2026b), §4, order of compilation.

If the resulting insertion v obeys $\|1_{H^c}(v - v_{\mathrm{cl}})\| < \eta$, seed commutation and unitarity give $\|1_H v\| = \sqrt{p_H}$. First omit the helper part of v, at energy cost at most $2\sqrt{p_H}$. Compare the two insertions projected off helpers. Their norm difference is below η, so their energies differ by at most 2η. The projected classical insertion still has energy $\mathcal{E}(m)$ because $m = 0$ on helpers. Thus

$$|\mathcal{E}(v) - \mathcal{E}(m)| \le 2\eta + 2\sqrt{p_H}.$$

All projected insertion vectors have norm at most one. This completes the interface from a bounded classical formula to a finite seeded word. It uses the source's smooth gate and helper-contraction theorems, and neither the scalar theorem nor a finite-degree central limit theorem alone implies it. OpenAI (2026b), §4, synthesis statement and its proof.

8.4 Removing the seed controls without changing their joint law

Seed removal must simulate one Gaussian variable per vertex, reused at all seed gates. Replacing each occurrence with a fresh independent Gaussian would change the circuit. The source proves a theorem for any already fixed finite seeded word and any energy tolerance, producing one finite ordinary word. Its proof works in the fixed-word tree limit and never takes a degree limit simultaneously with its small parameter. OpenAI (2026b), §5, "Removal of seed controls".

A small component supplies a large response. Fix $b > 0$ and $0 < r < 1$, and put

$$R_* = e^{-i\pi X/4}, \qquad \delta_r = \arcsin r, \qquad \gamma_r = b/r, \qquad e = Z\mathrm{vac}, \qquad \ell_r = Y e^{-2i\gamma_r Z q(e)}\mathrm{vac}.$$

Apply, in chronological order, cost γ_r, mixer δ_r, cost $-\gamma_r$ and mixer $\pi/4$. Just before the inverse cost the insertion is $v_{2,r} = \cos(2\delta_r)e + \sin(2\delta_r)\ell_r$. Define

$$d_{0,r} = \gamma_r(v_{2,r} - e) = -2br\, e + 2b\sqrt{1 - r^2}\, \ell_r, \qquad W_Z(f) = e^{iZq(f)}.$$

Expanding the first mixer and using the Weyl identity gives the exact split

$$u_r = a_{0,r}U_{O,r} + a_{1,r}U_{S,r}, \qquad U_{O,r} = R_* W_Z(d_{0,r}), \quad U_{S,r} = R_* X W_Z(-2\gamma_r e - d_{0,r}),$$

where

$$a_{0,r} = \cos\delta_r\, e^{i\vartheta_r}, \quad a_{1,r} = -ire^{-i\vartheta_r}, \quad \vartheta_r = \gamma_r^2\, \mathrm{Im}\langle v_{2,r}, e\rangle.$$

Both component operators are unitary. The special component $\psi_{S,r} = a_{1,r}U_{S,r}\mathrm{vac}$ therefore has norm exactly r. Its normalized contribution to the insertion is

$$w_r = u_r^* Z\psi_{S,r}/r, \qquad \|w_r\| = 1. \tag{8.1}$$

This is the direction amplified in the echo below. The two components need not be orthogonal. The ordinary component has a bounded displacement $d_{0,r} = 2b\ell_r + o(1)$, whereas the special component contains a displacement of order $1/r$. Subsequent unitaries preserve each component's

norm. Only the undisplaced packet tails have bounded occupation moments. Unitarity carries the norm remainders, and no such bounds are assigned to them.

The source's high-direction lemma gives that ℓ_r, w_r are centred unit vectors, become mutually orthogonal, and escape every fixed occupation sector. Splitting a uniformly bounded family of occupation moments at a fixed sector cutoff shows that its pairings with these directions vanish: first let $r \downarrow 0$, then send the cutoff to infinity. The special component must nevertheless be retained because the later echo magnifies its insertion contribution by $1/r$. OpenAI (2026b), §5.2, exact initialization and high directions.

A packet has the form $W(d_r/r)\eta_r$, where the scaled centre d_r is bounded and the tail η_r has uniformly bounded occupation moments. Packets with centres separated by a fixed positive amount have overlap smaller than every fixed power of r. A short probe along a Pauli axis P sees a finite rotation determined by $-2\,\mathrm{Im}\langle v_P, d_r\rangle$, as well as a small Weyl displacement. The direction $v_P = u^*Pu\,\mathrm{vac}$ is the actual current insertion. Toggling a cost rotates both its root Pauli axis and this direction, so the two cannot be selected independently. OpenAI (2026b), §5, packet estimates.

For the finite list of nonzero special centres $h_j \in \mathbb{R}^2$, the source constructs the same rotation $e^{-i\theta X}$ at every centre, with *zero clocks*: the signed pulse durations on each of two perpendicular axes in the ZY plane sum exactly to zero. Choose the axes so that $c_{j,a} = h_j \cdot P_a \neq 0$, and the square pairs $(c_{j,1}^2, c_{j,2}^2)$ are distinct except for equal or opposite centres. Brackets have zero clock components and generate iX coefficients $c_{j,1}c_{j,2}F(c_{j,1}^2, c_{j,2}^2)$. Polynomial interpolation makes all these coefficients one. Equal or opposite centres have the same product, so they cause no obstruction to this common rotation.

This does not give arbitrary independent rotations: opposite centres obey $U_{-h} = XU_hX$. Exact finite reachability of the required common rotation follows by taking finitely many reachable conjugates spanning the generated Lie algebra. Their product map contains an identity neighbourhood, and subdivision reaches the whole one-parameter path. Closure alone would not give exact cancellation. The Lie-bracket and polynomial method follows Li and Khaneja (2005), §II, Examples 1–2, pp. 2–3; the finite-family, equal/opposite-centre and exact-clock assertions are proved in the QAOA source itself.

The distinction is exact: zero net probe clocks and the common target rotation are assertions about the finite ensemble of limiting centres. A finite pulse list realizing those assertions is fixed first. Its action on the actual finite-r packets is approximate. The separate accuracy lemma then follows the actual insertion drift through every probe prefix. Its ordinary-component error is $o(r)$, and the error on the normalised special component is $o(1)$ and hence also contributes $o(r)$ to the state. These bounds are uniform over the bounded-tail families needed for insertions, and their constants refer to the chosen finite pulse list. OpenAI (2026b), §5, "Selective rotations with zero clocks" and "Accuracy of selective probes".

The shared seed family is essential even for a single spin. Two gates e^{-itSZ} applied to $|+\rangle$, with the same standard Gaussian S, give mean X expectation $\mathbb{E}\cos(4tS) = e^{-8t^2}$. Replacing the second seed by an independent copy gives e^{-4t^2}. Agreement of marginal seed laws does not preserve a circuit's value.

The error scale needed below follows from the source's quantitative probe estimate. For a fixed finite pulse list, separated packet centres and bounded designated tail moments, write $\varepsilon_r = o(r)$ for the discarded vacuum remainder. At each probe prefix j, with current insertion $v_{P,r}^{(j)}$ and huge-cost

direction v_r, the proof gives

$$\max_j |\operatorname{Im}\langle v_{P,r}^{(j)}, v_r\rangle| \le C_M(r^2 + \varepsilon_r + r^M).$$

Choose a fixed $M > 1$. The right side is $o(r)$: within each ordinary packet the relevant Hermitian expectation is real, distinct packet centres have overlap $O(r^M)$, and terms containing only special packets have size $O(r^2)$. Unitarity carries the norm remainder, which receives no unproved occupation bound. The exact zero clocks then cancel the first-order displacement on ordinary packets, leaving $o(r)$ error there. The prescribed action on a normalized special packet has $o(1)$ error, multiplied by its $O(r)$ amplitude. These are the unequal accuracy requirements that the next echo uses. OpenAI (2026b), §5, accuracy of selective probes.

The estimates must hold on the next input. The repeated echo uses more than the initial special norm in (8.1). Let N count particles in the top child Fock space. Call a family (η_r) bounded smooth when $\sup_r \|(1+\mathsf{N})^k \eta_r\| < \infty$ for every fixed k. At each stage between simulated gates, the source propagates a linear ordinary/special decomposition on every such family, with

$$\begin{aligned} u_r\eta_r &= \Psi_{O,r}(\eta_r) + \Psi_{S,r}(\eta_r), \\ \|\Psi_{S,r}(\eta_r)\| &= r\|\eta_r\|, \\ \Psi_{O,r}(\eta_r) &= \zeta_r A_r \eta_r + o(r). \end{aligned}$$

Here $|\zeta_r| = 1$, and the low unitary A_r and its adjoint preserve all these seminorms uniformly. A separate part of the induction transports the special insertion:

$$u_r^* Z \psi_{S,r}/r = w_r + o(1),$$

with the negative sign between a selective flip and its undoing. The norm remainders remain separate from the designated smooth tails, and no occupation estimate is imposed on them.

This forward estimate gives the insertion approximation used at the next gate. For $P = Y, Z$, the vector $\hat{v}_{P,r} = A_r^* P A_r \mathrm{vac}$ is itself a bounded smooth family. Since $A_r \hat{v}_{P,r} = P A_r \mathrm{vac}$, unitarity gives

$$\begin{aligned} \|u_r^* P u_r \mathrm{vac} - \hat{v}_{P,r}\| &= \|P u_r \mathrm{vac} - u_r \hat{v}_{P,r}\| \\ &\le \|P(u_r - \zeta_r A_r)\mathrm{vac}\| \\ &\quad + \|(u_r - \zeta_r A_r)\hat{v}_{P,r}\| = O(r). \end{aligned}$$

The second term explains why an estimate on the vacuum alone would not suffice: its input changes with r. The scalar phase cancels in this comparison. This $O(r)$ estimate supplies smooth approximants for actual insertions and is never divided by r. The singular echo instead uses the more accurate difference derived next. OpenAI (2026b), §5, packet invariant and insertion-approximation lemma.

The echo amplifies a controlled difference. Consider a forward cost t/r, a selective flip of the special component, the reverse cost $-t/r$, and a selective undoing of that flip. The packet invariant transports the special component so that its contribution to the insertion changes sign. For the actual insertion directions before and after the first flip, it gives

$$v_r' - v_r = -2r\, w_r + o(r).$$

The Weyl product of the two costs is formed with these actual directions before making any approximation. Their large centres cancel, leaving a bounded displacement

$$\frac{t}{r}(v_r' - v_r) = -2t\, w_r + o(1).$$

Approximating each direction by a low component with merely $O(r)$ error before this subtraction would leave an uncontrolled order-one error. The remaining scalar Weyl phase is independent of the root spin and cancels in insertions, so it need not itself converge. The second flip restores the transported special component so that the same construction can be used again. Mixers receive analogous selective compensations. OpenAI (2026b), §5, packet invariant, insertion approximation and "One echo".

In the limit, ℓ_r and w_r become two orthogonal external real Gaussian modes at each root. The initialization uses $q(\ell)$ and every simulated seed gate uses the same $q(w)$: choosing $t = \sigma/2$ implements $e^{-i\sigma Z q(w)}$. Different vertices have independent local copies. This is the required joint law of the seeded word. A finite induction uses the Weyl identity

$$\langle W(f)\Omega, W(g)\Omega\rangle = \exp\bigl(-\tfrac{1}{2}\|g-f\|^2 + i\,\mathrm{Im}\langle f, g\rangle\bigr)$$

to match the relevant finite Gram matrices. Escaping high directions supply precisely the vanishing pairings between the new external modes and endogenous child directions. OpenAI (2026b), §5, external-mode limit.

The energy pairing after seed removal. Matching the seed law and its finite Gram matrices is not yet enough. For a centred low insertion $h = \sum_j s_j \otimes W(f_j)\Omega$, the one-particle coefficient of $W(f_j)\Omega$ is $ie^{-\|f_j\|^2/2} f_j$. Hence

$$\langle h, \mathsf{L}h\rangle = -i \sum_j \langle s_j, +\rangle e^{-\|f_j\|^2/2} \langle f_j, h\rangle.$$

In the seeded representation, L places the root-seed factor in its vacuum and creates h only in the endogenous child space. External components therefore contribute zero to the last pairing. The high-direction estimate yields the same zeros in the actual limit. The actual insertion differs from its low insertion by $O(r)$, so energy continuity completes the comparison. Matching Gram matrices without this one-particle calculation would not identify L or its energy. OpenAI (2026b), §5, "Finite Gram induction," including the energy calculation.

Why the selective rotations are available at every required stage. The limiting projection of each special centre must be nonzero. The source imposes finitely many visibility conditions on the angles of the formal seeded word. At each relevant standard stage they have form

$$-2\,\mathrm{Im}\langle v_Z^{\mathrm{form}}, -2bz\, e + yD^{\mathrm{form}}\rangle \neq 0, \qquad z, y \in \{\pm 1\},$$

where $e = Z\mathrm{vac}$ and D^{form} accumulates echoes whose reverse costs have already occurred. These functions are real analytic by the finite Weyl formulas. Setting all target angles to zero, while retaining initialization, gives $4bze^{-8b^2} \neq 0$, so none is identically zero. An arbitrarily small angle perturbation avoids their finite union of zero sets and the zero echo-time hyperplanes. The huge-cost stages need a separate transverse-projection estimate, after which their limiting projection pairs also have a nonzero coordinate.

Finite pulse lists can now be chosen successively from limits already established at earlier stages. No list is chosen from a limit that assumes its own successful implementation. OpenAI (2026b), §5, huge-stage transverse projection, visibility and construction order.

The choices are therefore nested. We make b small to control initialization, perturb the fixed target angles to ensure visibility and choose all finite probe lists. Finally we choose r small. A value of r fixes every large angle and every finite product. The theorem allows those angles to be large or negative. It makes no claim of uniform bounds as the target word length grows or as the tolerance tends to zero. OpenAI (2026b), §5, proof of seed removal.

8.5 Transfer of the resulting ordinary word

After all these choices there is one finite ordinary word W, fixed independently of degree, size and disorder. On the $(D+1)$-regular tree use $C_D = D^{-1/2}\sum_{\{i,j\}\in E} Z_i Z_j$, counting each edge once. The edge expectation is defined on the finite light cone described in §8.1. The source applies the fixed-word transfer theorem

$$v_W = \lim_{D\to\infty} \frac{\sqrt{D}}{2}\langle +|W_D^* Z_i Z_j W_D|+\rangle.$$

The full preprint of Basso et al., arXiv:2110.14206v3, Theorem 1, equates the limiting tree recursion with the infinite-size SK recursion for every fixed depth and fixed angle vector. Its proof is in §6, especially (6.6)–(6.31), with the symmetry statement proved in Appendix B and the iteration stabilisation in Appendix A.4. The shorter proceedings edition states the theorem but refers to this full preprint for the proof.

Basso et al. use a negative tree cost in (2.3). Let $V = \prod_{i\in A} X_i$ on one bipartition of the finite light cone. Every edge has exactly one endpoint in A, hence

$$VZ_iZ_jV = -Z_iZ_j, \qquad VB_DV = B_D, \qquad V|+\rangle = |+\rangle.$$

Consequently, if $C_D^- = -C_D$, then $VW(C_D^-, B_D)V = W(C_D, B_D)$ for the same fixed word and angles, while its measured edge expectation changes sign. This converts the negative cut convention in their (2.6) to the positive energy above. Their SK cost (6.1) already has the positive convention used here. The result applies after seed removal and does not directly transfer an arbitrary seeded or controlled word. It also supplies no joint limit in which depth or angles change with D or n. OpenAI (2026b), §2, "Fixed-word transfer".

For an ordinary cost or mixer the root updates are, respectively,

$$u \longmapsto e^{-itZq(v_Z)}u, \qquad u \longmapsto e^{-i\beta X}u, \qquad v_Z = u^* Z u\,\mathrm{vac}.$$

The source proves smooth convergence for these updates, their adjoints and fixed angle derivatives. The resulting occupation bounds supply the hypothesis of the root-shift energy proposition in §8.1. Thus its energy is the scaled edge limit required for transfer. Finally merge consecutive gates of the same type and insert zero-angle gates where needed to express W in alternating QAOA form. OpenAI (2026b), §2, "Ordinary root updates".

8.6 Fixing the word before the thermodynamic limit

Approximating a specified finite observable. First fix T, N and the positive finite normalizers in §7. The rooted construction in §8.1 has energy exactly $S_{T,N}$. Let z be a smoothed version of its rounded spin F_N, let $m = \sin(2\theta)$ be the bounded formula's spin output, and let m^{col} be its color-excluded approximation. Write

$$\delta_s = \|z - F_N\|_2, \qquad \delta_b = \|m - z\|_2, \qquad \delta_c = \|m^{\mathrm{col}} - m\|_2.$$

These measure errors in the spin outputs themselves. The rooted shift is a contraction, so

$$|\mathcal{E}(f) - \mathcal{E}(g)| \le (\|f\|_2 + \|g\|_2)\|f - g\|_2.$$

All these outputs have norm at most one. Apply synthesis with tolerance η_s and total helper probability p_H, then remove the seeds with energy tolerance η_r. The source's absolute synthesis bound and seed-removal theorem, followed by fixed-word transfer, give one ordinary word W with

$$|v(W) - S_{T,N}| \le 2(\delta_s + \delta_b + \delta_c) + 2\eta_s + 2\sqrt{p_H} + \eta_r. \tag{8.2}$$

The bounded formula is chosen before the colors and helper law. The excluded receiver and ancestor colors prevent a dependency from returning through a traversed branch. Reserved helper colors are absent from every dependency, and the output is zero on helpers. Color refinement and small positive helper probabilities jointly control δ_c and p_H. Only then is synthesis performed with that law fixed, using the same sitewise Gaussian seed at every seed gate. Each term in (8.2) can thus be made smaller than its allotted tolerance in this order. OpenAI (2026b), Lemmas 3.1–3.4; synthesis energy bound; seed removal.

This gives an absolute approximation of each specified finite $S_{T,N}$ in the construction, whether or not it is already near P_*. It preserves the energy and makes no claim about the entire classical output law. The cutoff and mesh choices then give near-optimality.

We allocate the errors after choosing the desired $\varepsilon > 0$, and apply the classical near-optimality proposition with the two tolerances $\varepsilon/4$ and $\eta_H = (\varepsilon/8)^2$. It supplies one formula and its helper law together, with classical loss below $\varepsilon/4$ and $0 < p_H < \eta_H$. In its proof the uncoloured formula is fixed first, then the small positive helper probabilities and the excluded-colour approximation are selected. Thus changing a helper law does not silently preserve an already fixed formula. With $\eta = \varepsilon/8$, synthesis gives

$$\mathcal{E}(v) \ge \mathcal{E}(m) - 2\eta - 2\sqrt{p_H} > P_* - 3\varepsilon/4.$$

Seed removal uses less than $\varepsilon/4$. The resulting fixed ordinary word therefore satisfies $v_W > P_* - \varepsilon$ after the fixed-word transfer. Its thermodynamic limit is taken only after all approximation choices have been made. Since $Q_p \le P_*$ and zero padding makes Q_p nondecreasing, this proves the source's thermodynamic-first conclusion from its stated analytic inputs. OpenAI (2026b), §6.

The choices come in a fixed order. We choose the accuracy, then the minimizer and a fixed $T < 1$, then the finite scalar mesh. Next come the smoothing and bounded approximations, the finite types and positive helper masses, and the exact helper contractions and finite commutator products, after which one seeded word is fixed. We then choose the initialization size, a generic angle perturbation, finite selective pulse lists and a sufficiently small r, which fixes one ordinary word. Only then do we take its $n \to \infty$ limit.

The degree limits identify the rooted calculus and the transferred value for each already fixed word, and they never allow that word to change with the finite instance.

The reconstruction makes the source's circuit mechanisms explicit and keeps their evidence classes separate. The scalar construction and its formal interfaces in §§6–7 and 12 do not by themselves establish the weighted unitary calculus, simultaneous helper contraction or seed removal. Those analytic theorems are additional inputs from the QAOA manuscript, with the domain, commutation and uniformity hypotheses specified above. No quantitative depth bound, finite-size rate, efficient angle-selection procedure, or independent formal verification of this entire circuit construction follows from the selected formal statements.

This completes the main support-to-circuit proof.

The later comparisons have separate purposes. §11.1 gives another classical input, and §11.2 adds a second-moment sampling theorem. §§11.3–11.5 compare quantifiers, other models and later finite-size estimates. None of them is a premise of the proof just completed.

9. Density and finite-disorder consequences

9.1 Smoothness of the minimizing coefficient

A density for the Stieltjes measure needs an argument beyond full support. In the source manuscript it uses the PDE again, after the gap contradiction has established $\mathbb{E}a(t, X_t)^2 = 1$. We reconstruct this argument because it gives a further conclusion of the original paper, outside the selected full-support formal statement.

Put $f_k = \partial_x^k \Phi_\gamma$, so $f_1 = u$ and $f_2 = a$. All formulas in this subsection are on a fixed compact time strip, where every finite collection of spatial derivatives is bounded. Differentiating the PDE and applying Itô's formula along the associated diffusion gives

$$df_k(t, X_t) = -\frac{\gamma(t)}{2} \sum_{j=1}^{k-1} \binom{k}{j} f_{j+1}(t, X_t) f_{k-j+1}(t, X_t)\, dt + f_{k+1}(t, X_t)\, dB_t.$$

The sum is empty when $k = 1$. This identity needs no derivative of γ.

For an initially nonsmooth coefficient, we apply it to finite-step approximations and pass to the integrated identity. There the coefficient converges in L^1, the spatial factors converge uniformly, and the stochastic integrals converge by the Itô isometry. OpenAI full-support manuscript, smooth-density proof, lines 35–50.

Set $b = f_3$. The case $k = 2$ gives $da = -\gamma a^2 dt + b\, dB_t$, and consequently

$$0 = \frac{d}{dt}\mathbb{E}a^2 = \mathbb{E}b^2 - 2\gamma(t)\mathbb{E}a^3 \quad \text{for almost every } t < 1.$$

Convexity gives $a \geq 0$, and the already established second moment gives $\mathbb{E}a^3 \geq (\mathbb{E}a^2)^{3/2} = 1$. Thus the quotient

$$\tilde{\gamma}(t) = \frac{\mathbb{E}b(t, X_t)^2}{2\mathbb{E}a(t, X_t)^3}$$

is well defined and continuous, before any continuity of γ is known. It equals γ almost everywhere.

Approaching an arbitrary $t < 1$ from the right through the full-measure equality set, continuity of $\tilde{\gamma}$ and right continuity of γ give equality at that t as well. This proves continuity of γ. At zero, $X_0 = 0$, and evenness of $\Phi(0, \cdot)$ gives $b(0, 0) = 0$; hence $\gamma(0) = 0$. OpenAI full-support manuscript, lines 52–66; Chen, Proposition 4.6, pp. 25–26.

The bootstrap must treat all polynomial moments simultaneously. For a polynomial P in finitely many variables, write

$$M_P(t) = \mathbb{E}P(f_1(t, X_t), \dots, f_m(t, X_t)).$$

The preceding stochastic identities and Itô's product rule provide polynomials Q_P, R_P, involving only finitely many additional derivatives, such that

$$M_P(t) - M_P(0) = \int_0^t \left(M_{Q_P}(s) + \gamma(s) M_{R_P}(s)\right) ds.$$

Explicitly, if

$$D_k(y) = -\tfrac{1}{2}\sum_{j=1}^{k-1}\binom{k}{j} y_{j+1}y_{k-j+1},$$

then

$$Q_P = \tfrac{1}{2}\sum_{k,l}\partial_{kl}P\, y_{k+1}y_{l+1}, \qquad R_P = \sum_k \partial_k P\, D_k.$$

All these moments are continuous to begin with. Once γ is continuous, the integral formula makes every moment C^1, and the quotient formula makes γ itself C^1. If all moments and γ are C^r, the same argument makes them C^{r+1}. Induction proves $\gamma \in C^\infty([0,1))$, with derivatives at zero taken from the right. Since $\gamma(0) = 0$, there is no atom at zero. Its Stieltjes measure is $\gamma'(t)\,dt$, and monotonicity gives $\gamma' \geq 0$. The source credits this polynomial-moment method to Auffinger–Chen and to Chen's zero-temperature treatment. OpenAI full-support manuscript, lines 68–101; Auffinger and Chen, proof of Theorem 2(ii), p. 16; Chen, Appendix B.6, p. 38.

The exact arXiv edition of Auffinger–Chen's polynomial-moment proof has a local display discrepancy: p. 16 of arXiv:1303.3573v1 prints a square in F_3 below (40), whereas Proposition 3, p. 9, has the cube. The identity $d(a^2) = (-2\gamma a^3 + b^2)dt + 2ab\,dB$ requires the cube used above, and the moment induction itself is unchanged.

9.2 Strict positivity of the density away from zero

Strict positivity is a further consequence of full support, and neither full support nor §6 uses it.

A smooth increasing function can have zero derivative at an interior point. For example, $g(t) = (t - \frac{1}{2})^3 + \frac{1}{8}$ on $[0,1)$ is nonnegative and strictly increasing, and dg has full support, although $g'(\frac{1}{2}) = 0$. The example is no minimizing coefficient, and it refutes only the inference from strict increase to a positive derivative. For the minimizing density we need a further estimate.

The smoothness just proved permits (4.2), now with $q(t) = t$. Writing $b = \Phi_{xxx}$ and $d = \Phi_{xxxx}$, it becomes

$$2\gamma'(t)\mathbb{E}a^3 = \mathcal{C}_\gamma(t).$$

Fix $t \in (0,1)$ and set $c = \gamma(t) > 0$, using full support and $\gamma(0) = 0$. The coefficient is now smooth, so Proposition 4.1 applies at this time to the same diffusion. Its proof uses nonminimizing local constant replacements, with compact-strip bounds independent of the widths of the inserted plateaus. Combining (4.1) with the preceding identity gives

$$2\gamma'(t)\mathbb{E}a^3 \geq \frac{\gamma(t)^2}{1000}\mathbb{E}a^4 > 0, \qquad 0 < t < 1.$$

The final strict inequality uses $\mathbb{E}a^4 \geq (\mathbb{E}a^2)^2 = 1$. Cauchy–Schwarz also gives $\mathbb{E}a^3 \leq \sqrt{\mathbb{E}a^2\mathbb{E}a^4} = \sqrt{\mathbb{E}a^4} \leq \mathbb{E}a^4$. Therefore the same estimate proves the quantitative ordinary consequence

$$\gamma'(t) \geq \frac{\gamma(t)^2}{2000} > 0 \quad (0 < t < 1), \qquad \mu_\gamma((r,s]) \geq \frac{s-r}{2000}\gamma(r)^2 \quad (0 < r < s < 1).$$

The second inequality follows by integration and monotonicity. Its lower bound depends on this minimizer and the left endpoint, so it gives no uniform mass estimate for a family of finite-temperature measures.

At the left endpoint the argument proves only $\gamma'(0) \geq 0$. It makes no claim that the minimizing coefficient itself is constant on a neighbourhood, and it passes none of the approximants' third time

derivatives to the limit. Only a spatial-derivative expectation is transferred. OpenAI full-support manuscript, lines 102–152.

The density result is a separate addition to the scalar approximation in §7. The next consequence concerns the geometry of near-ground spin configurations.

9.3 Near-ground configurations at a prescribed overlap

We choose one minimizer supplied by the attained Parisi formula and apply the full-support theorem to that same coefficient. The external landscape theorem below then applies at every point of its Stieltjes support. For every fixed $r \in [0,1)$ and $\varepsilon, \eta > 0$, there is a constant $K = K(r, \varepsilon, \eta) > 0$, independent of n, such that

$$\mathbb{P}\left(\exists \sigma, \tau \in \{-1,1\}^n : \left|\frac{1}{n}\sum_{i=1}^{n} \sigma_i \tau_i - r\right| < \varepsilon, \quad \frac{H_n(\sigma)}{n}, \frac{H_n(\tau)}{n} \geq P_* - \eta\right) \geq 1 - Ke^{-n/K}, \qquad n \geq 1.$$

The probability is over the SK couplings, and the configurations may depend on them. The display specializes Auffinger–Chen–Zeng, Proposition 1, equation (5), p. 5 to the full-support minimizer above, since their theorem applies at each point of its Stieltjes support.

Their proof, on pp. 15–16, uses the coupled free energy constrained to a neighbourhood of that overlap, passes to the coupled maximum and applies Gaussian concentration. The coupled maximum approaches $2P_*$, while the unrestricted maximum approaches P_*. These two facts force both members of a maximizing pair to be near ground energy.

The covariance normalization needs one small adjustment. Writing $R = n^{-1}\sum_i \sigma_i \tau_i$, our Hamiltonian has covariance $nR^2/2 - 1/2$. Adjoin a common independent Gaussian C of variance $1/2$. Then $\widetilde{H}_n(\sigma) = H_n(\sigma) + C$ has covariance $n\xi(R)$ with $\xi(t) = t^2/2$, as in the cited mixed-spin theorem. Its expected maximum equals that of H_n, so its ground-state limit is the same P_*. Apply the theorem with energy tolerance $\eta/2$. On $|C| \leq \eta n/2$, the resulting pair also satisfies the displayed energy inequalities for H_n. The omitted event has probability at most $2e^{-\eta^2 n^2/4}$, which can be absorbed into $Ke^{-n/K}$ by enlarging K. Since the conclusion for H_n depends only on its couplings, this also removes the auxiliary Gaussian from the probability statement.

The pointwise result also gives a simultaneous statement at any fixed positive resolution. Given $\varepsilon, \eta > 0$, choose finitely many centres $r_j \in [0,1)$ whose open $\varepsilon/2$-neighbourhoods cover $[0,1]$. Apply the preceding result at each centre with overlap tolerance $\varepsilon/2$ and energy tolerance η. A finite union bound gives constants $C, c > 0$, depending only on these tolerances, such that

$$\mathbb{P}(\forall r \in [-1,1]\ \exists \sigma, \tau \in \{-1,1\}^n : |R(\sigma,\tau) - r| < \varepsilon, \quad H_n(\sigma)/n, H_n(\tau)/n \geq P_* - \eta) \geq 1 - Ce^{-cn}. \tag{9.1}$$

Indeed, on the event where every centre has a suitable pair, choose a nearest centre and use the triangle inequality. For negative targets, reverse one spin configuration: the pure two-spin Hamiltonian preserves its energy and changes the overlap's sign. The finite net covers the endpoints as well, with centres strictly below one.

At a fixed finite size the picture is different. If σ, τ differ at k sites, precisely the $k(n-k)$ edges crossing that set change their spin product. Their independent Gaussian contributions give

$$\operatorname{Var}(H_n(\sigma) - H_n(\tau)) = \frac{4k(n-k)}{n}.$$

For $0 < k < n$ this is positive, so the two energies tie with probability zero. A finite union over pairs, together with $H_n(-\sigma) = H_n(\sigma)$, shows that almost surely the exact maximum has one maximizing pair $\{\sigma_*, -\sigma_*\}$. The overlap conclusion above concerns nearly optimal pairs at intermediate separations. Its threshold $P_* - \eta$ is deterministic, and setting $\eta = 0$ would not turn it into exact maximization of a finite sample. This elementary deduction uses only the displayed Gaussian Hamiltonian and is separate from the external asymptotic landscape theorem.

The overlap deduction uses fixed tolerances and target-dependent pairs. It gives neither exact arbitrary real overlaps at finite n, a typical-pair law, a search procedure, nor a guarantee for tolerances shrinking with n. It uses the external landscape theorem, full support and attainment, and lies outside the two selected formal statements in §12. The separate algorithmic consequence is in §9.4.

9.4 A classical optimization algorithm

The scalar Gaussian construction is not itself an algorithm on the matrix of SK couplings. We obtain a finite-disorder algorithm from the theorem of El Alaoui, Montanari and Sellke. In their cited preprint, Assumption 2 asks for a strictly increasing minimizer of the zero-temperature Parisi functional on $\mathcal{U}$. The support theorem gives this property, and the imported attainment argument supplies the minimizer.

Their Corollary 2.2 then returns, for each fixed $\varepsilon > 0$, a randomized spin configuration within εn of the finite-instance maximum with probability tending to one. Its operation count is the one in the displayed Theorem 3 on p. 7, which the preprint's running references call "Theorem 3.2." El Alaoui, Montanari and Sellke, arXiv:2001.00904v1, pp. 7–8.

To match that theorem's Gaussian input exactly, form a symmetric matrix with

$$A_{ij} = J_{ij}/\sqrt{n} \quad (i < j), \qquad A_{ii} = \sqrt{2/n}\, G_i,$$

where the G_i are additional independent standard Gaussians. For $\widehat{H}_n(x) = \frac{1}{2} x^{\mathsf{T}} A x$,

$$\mathbb{E}\widehat{H}_n(\sigma)\widehat{H}_n(\tau) = \frac{(\sigma^{\mathsf{T}}\tau)^2}{2n}, \qquad \widehat{H}_n(\sigma) = H_n(\sigma) + \frac{1}{\sqrt{2n}} \sum_i G_i.$$

The final term is the same for all spin configurations. Therefore the maximizing spin and every optimization gap are unchanged. The extra diagonal randomness is included in the algorithm's randomness. The input convention thus matches, and the target optimization problem is unchanged. OpenAI full-support manuscript, algorithm application, lines 30–47.

The imported operation count is $C(\varepsilon)(\chi + n) + n\chi_1$, where χ is the cost of evaluating the full gradient and χ_1 is the cost of one coordinate of the distinct-index gradient. Here $\nabla\widehat{H}_n(x) = Ax$ costs $O(n^2)$, and one coordinate of $\nabla H_n(x)$ costs $O(n)$. The resulting count is $C(\varepsilon)n^2$, with $C(\varepsilon)$ independent of n.

Real sums and multiplications have unit cost in the cited model, and the scalar functions and normalizers prescribed for the chosen accuracy are part of that construction. This statement supplies neither a bound for their numerical preprocessing nor a finite-precision bit-complexity bound. El Alaoui, Montanari and Sellke, Remark 2.1 and Theorem 3, p. 7; OpenAI full-support manuscript, lines 1–11 and 49–59.

A distinct classical construction is Jekel, Sandhu and Shi, *Potential Hessian Ascent*, version 3, Assumption 1.4 and Theorem 1.5. Its hypothesis is finite-temperature interval support $\operatorname{supp}\mu_\beta = [0, q^*_\beta]$, with $q^*_\beta \to 1$. Its inputs include oracles for the Parisi PDE, its derivatives and the minimizer's distribution function. For each fixed accuracy the stated runtime is polynomial in n, with an exponential dependence on inverse accuracy, and it excludes the cost of implementing those oracles.

Its Ginibre Hamiltonian has covariance nR^2: on spins it has the law $\sqrt{2}H_n$ plus a Gaussian term independent of the configuration. Thus its energy and accuracy must be divided by $\sqrt{2}$ to compare with this paper's normalization. This is a separate algorithm theorem with its own analytic inputs, and it gives no new reading of the finite Gaussian spin in §7.

The consequence is classical. It supplies no efficient construction of QAOA angles, no quantum depth bound and no clause of either selected formal interface. The proof of the separate QAOA statement still needs its own rooted energy, synthesis, seed-removal and transfer arguments.

10. An independent scalar comparison

10.1 Contact from convexity on constant intervals

A second route obtains $q = t$ from the variational inequalities. It uses convexity on intervals where the tail potential is positive, without first proving full Stieltjes support. The scalar implementation uses this argument, which is independent of the formal full-support theorem.

In the scalar model put $q(t) = \mathbb{E}[u(t, X_t)^2]$, $f(t) = q(t) - t$ and $G(t) = \int_t^1 f(s)\,ds$. The source first obtains $G \geq 0$, $G(1) = 0$, $q(0) = 0$ and $G' = -f$. Compact-mass variations also show that on any interval where $G > 0$ the coefficient is constant: if its cumulative mass increased between two points there, integration of the strictly positive G against that positive increment would contradict the mass-removal variational inequality. This is the content of the scalar contact-integral argument.

Suppose now that $G(x) > 0$ for some $0 < x < 1$. Choose the component (a, b) of this positivity set containing x, allowing $a = 0$ or $b = 1$. Its coefficient is constant, say $c \geq 0$, and $G(b) = 0$. At the left endpoint, $f(a) = 0$: this is the contact derivative when $a > 0$, and $q(0) = 0$ when $a = 0$.

If $c > 0$, one also has $G(a) = 0$. The only extra case is $a = 0$: right continuity gives $\gamma(0) = c > 0$, so the mass-removal condition at that initial atom forces $G(0) = 0$. The mean-value theorem on $[a, x]$ and $[x, b]$ now provides points $r < s$ with $f(r) < 0 < f(s)$. There is a later point $z \leq b$, with $z < 1$, for which $f(z) \leq f(s)$. For $b < 1$ take $z = b$, since contact gives $f(b) = 0$. For $b = 1$, choose z sufficiently close to one and use $f(z) = q(z) - z \leq 1 - z < f(s)$.

The three intervals $[a, r]$, $[r, s]$ and $[s, z]$ therefore have secant slopes for f of signs respectively negative, positive and nonpositive. Applying the mean-value theorem once more gives three ordered points where f' has those signs. This is impossible when $f' = q' - 1$ is convex: its middle value cannot exceed the larger of its two exterior values. The scalar source supplies precisely this nonstrict convexity on a constant interval. See the positive-hump argument and its endpoint condition and the curvature-moment convexity lemma.

If $c = 0$, the source instead gives convexity of $f = q - t$. A mean-value argument between x and b gives $f(s) > 0$, and the same right-endpoint argument produces $z > s$ with $f(z) \leq f(s)$. Since $f(a) = 0$,

$$\frac{f(s) - f(a)}{s - a} > 0 \geq \frac{f(z) - f(s)}{z - s},$$

contradicting monotonicity of adjacent secant slopes of a convex function. This case does not need $G(a) = 0$. See the zero-coefficient hump lemma.

Thus G cannot have a positive component, and $G \equiv 0$ on $(0, 1)$. Differentiation gives $q(t) = t$ there, with the origin supplied separately. The scalar assembly then obtains the unit curvature moment, finite Gaussian approximation and terminal value identities from this contact equality. Nonstrict convexity suffices for this scalar route. The strict curvature estimate in Sections 4–5 has the further job of excluding gaps in Stieltjes support. These are source-level proof reconstructions, not a fresh Lean acceptance or a claim that the two formal models have been composed.

10.2 A universal bound for predictable correlations

The scalar implementation obtains the value equality by two inequalities. Its finite-spin limit follows from superadditivity and a uniform upper bound. A finite-profile Gaussian comparison supplies $P_* \le \mathcal{P}(\gamma)$ for every admissible coefficient. The reverse inequality at a minimizer comes from a more general finite-dimensional fact: predictable orthonormal Gaussian functions paired with any spin have a shifted correlation sum at most P_*. We reconstruct that fact below. It supplies the inequality that §7's mesh limit consumes, without invoking the literature's Parisi formula used in §6.

Write $e_n = n^{-1}\mathbb{E}\max_\sigma H_n(\sigma)$, with the Hamiltonian of §1. The finite-spin convergence and upper comparison are separate inputs to this subsection, and their precise suppliers are the superadditive limit and the upper comparison. The proof below concerns only the lower comparison, and it builds neither a finite-disorder algorithm nor a circuit.

Proposition 10.1 (predictable Gaussian comparison). Fix $K < \infty$, independent standard normals $z_0, \dots, z_K$, and real functions $f_j = f_j(z_0, \dots, z_j)$ for $0 \le j < K$. Assume measurability, square integrability and $\mathbb{E}[f_i f_j] = \delta_{ij}$. For any measurable $S = S(z_0, \dots, z_K) \in \{-1, 1\}$, the conclusion is

$$\sum_{j=0}^{K-1} \mathbb{E}[f_j S]\,\mathbb{E}[S z_{j+1}] \le P_*.$$

These assumptions allow nonconstant, dependent f_j, and even $\mathbb{E}f_j = 0$ is unnecessary. Their order of dependence ensures $\mathbb{E}[f_i z_{j+1}] = 0$ whenever $i \le j$. The next fresh coordinate, rather than centring of f_i, is what will remove the projection terms. Scalar implementation, predictable-coefficient theorem.

10.2.1 Revealing one Gaussian column For $w \in \mathbb{R}^n$, put $Q(w) = w \otimes w/\sqrt{2}$. An independent standard Gaussian matrix G gives a centered field $\langle Q(w), G\rangle$ with covariance $\langle w, w'\rangle^2/2$. Given a unit vector v, define

$$R_v w = w - \langle v, w\rangle v, \qquad C_v(w) = \langle v, w\rangle R_v w + \frac{\langle v, w\rangle^2}{\sqrt{2}} v.$$

Write $\alpha = \langle v, w\rangle$, $\beta = \langle v, w'\rangle$ and $r = \langle R_v w, R_v w'\rangle$. For an independent standard Gaussian column Z and matrix G', the covariance of $\langle C_v(w), Z\rangle + \langle Q(R_v w), G'\rangle$ is

$$\alpha\beta r + \frac{1}{2}\alpha^2\beta^2 + \frac{1}{2}r^2 = \frac{1}{2}(r + \alpha\beta)^2 = \frac{1}{2}\langle w, w'\rangle^2.$$

Thus the new field has the old joint Gaussian law. The same statement holds for $v = 0$: then R_v is the identity and $C_v = 0$. Allowing that case will avoid assuming empirical linear independence.

For a finite nonempty family of candidates w_s and offsets d_s, define

$$\mathcal{V}(w,d) = \mathbb{E}_G \max_s \{\langle Q(w_s), G\rangle + d_s\}.$$

The covariance identity and Fubini give the exact recursion

$$\mathcal{V}(w,d) = \mathbb{E}_Z \mathcal{V}((R_v w_s)_s, (d_s + \langle C_v(w_s), Z\rangle)_s).$$

Also $d_s \le \mathcal{V}(w,d)$ for each s, since the residual Gaussian term has mean zero. These are the two facts needed for adaptive choices. The source derives the equality by Gaussian comparison in both directions and proves the required integrability in the one-column decomposition.

10.2.2 Adaptive directions and a finite-spin inequality We choose each zero-or-unit direction measurably from columns already revealed. Conditional on that history, the current residual candidate vectors and offsets are fixed, so the preceding recursion applies. More precisely, conditional on the revealed history, the expected next continuation value, including its accumulated offsets, equals the current one. Iteration therefore preserves the initial expected maximum. For any measurable choice of a candidate from the final history, its accumulated offset is bounded by the remaining conditional expected maximum. Hence

$$\mathbb{E}\, d^{\mathrm{final}}_{s(\mathrm{history})} \le \mathcal{V}(w,0).$$

Contraction of each residual, the finite candidate set and Gaussian column moments justify the expectations. Adaptivity does not turn a future column into part of the conditioning history. Scalar implementation, adaptive recursion.

Take all spins $w_s \in \{-1,1\}^n$ as candidates. The full quadratic matrix has off-diagonal coefficients $(G_{ij} + G_{ji})/\sqrt{2}$, which are independent standard normals, and a centered diagonal contribution common to every spin. Thus

$$\mathcal{V}(\mathrm{spins}, 0) = \sqrt{n}\, \mathbb{E} \max_\sigma H_n(\sigma).$$

Dividing the offset inequality by $n\sqrt{n}$ gives a bound by e_n. The implementation proves this normalization through equality of Gaussian increment distances in the finite adaptive SK comparison.

To obtain the desired coefficients, take independent rows $z^{(1)}, z^{(2)}, ...,$ each with the law of $(z_0, \dots, z_K)$, and put

$$\hat{g}_n = n^{-1/2}(g(z^{(1)}), \dots, g(z^{(n)})).$$

Let $v_{n,j}$ be the normalized Gram–Schmidt vector obtained from $\hat{f}_{0,n}, \dots, \hat{f}_{j,n}$, where a zero remainder is normalized to zero. Start the adaptive procedure with a zero direction to reveal column z_0. After seeing columns through z_j, $v_{n,j}$ is available, and the next column is z_{j+1}. This explains both the extra column and the shift.

Let $R_{n,j}$ project away $v_{n,0}, \dots, v_{n,j-1}$ and set

$$E_{n,j} = \langle C_{v_{n,j}}(R_{n,j}\widehat{S}_n), \hat{z}_{j+1,n}\rangle.$$

The raw spin and fresh column are each $\sqrt{n}$ times their empirical versions, and C_v is quadratic in its input. Consequently their revealed energy is exactly $n\sqrt{n}E_{n,j}$. Transposing the array between independent rows and independent columns preserves its Gaussian law. The finite comparison therefore gives, for every $n > 0$,

$$\mathbb{E} \sum_{j<K} E_{n,j} \le e_n.$$

This is an actual choice of the adaptive directions and final rowwise spin, not an uninstantiated conditional inequality. Empirical directions and normalization; row–column transposition and finite bound.

10.2.3 Orthogonalization preserves the limiting correlations We keep K fixed while $n \to \infty$. The strong law applies to the finitely many needed products, which are integrable by Cauchy–Schwarz. The empirical Gram matrix tends to the identity. Inductively in Gram–Schmidt, every projection coefficient tends to zero and every remaining norm tends to one; hence $\|v_{n,j} - \hat{f}_{j,n}\| \to 0$ almost surely. In particular,

$$\langle v_{n,j}, \widehat{S}_n \rangle \to a_j := \mathbb{E}[f_j S], \qquad \langle \widehat{S}_n, \hat{z}_{j+1,n} \rangle \to b_{j+1} := \mathbb{E}[S z_{j+1}],$$

while $\langle v_{n,i}, \hat{z}_{j+1,n} \rangle \to 0$ for $i \le j$. The last limit uses precisely the fresh-coordinate cancellation stated at the start of §10.2. Gram–Schmidt limits.

Expand the revealed energy before taking its limit:

$$\begin{aligned} E_{n,j} = \langle v_{n,j}, \widehat{S}_n \rangle \langle R_{n,j+1} \widehat{S}_n, \hat{z}_{j+1,n} \rangle \\ + \frac{\langle v_{n,j}, \widehat{S}_n \rangle^2}{\sqrt{2}} \langle v_{n,j}, \hat{z}_{j+1,n} \rangle. \end{aligned}$$

All terms introduced by projecting away earlier directions vanish, as does the second line. The remaining product tends to $a_j b_{j+1}$. Thus $\sum_{j<K} E_{n,j} \to \sum_{j<K} a_j b_{j+1}$ almost surely. This establishes the random-variable limit, but an expectation bound still requires the next step. Scalar implementation, empirical energy limit.

10.2.4 Passing the expectation bound and returning to the value Since $S^2 = 1$, $\|\widehat{S}_n\| = 1$ for $n > 0$, and every residual projection contracts. The column formula gives $|E_{n,j}| \le 1 + \|\hat{z}_{j+1,n}\|^2$. Define

$$c_n = \sum_{j<K} (1 + \|\hat{z}_{j+1,n}\|^2).$$

Then $\sum_j E_{n,j} + c_n \ge 0$, $\mathbb{E} c_n = 2K$ exactly, and $c_n \to 2K$ almost surely. Fatou's inequality, the finite bound and $e_n \to P_*$ imply

$$\sum_{j<K} a_j b_{j+1} + 2K \le \liminf_n \mathbb{E}[\sum_{j<K} E_{n,j} + c_n] \le P_* + 2K.$$

Cancelling $2K$ proves the claimed predictable-coefficient inequality. The source's constant-limit helper establishes the same nonnegative comparison by clipping at the limiting constant and applying dominated convergence. It shifts to dimensions $n+1$ to avoid dimension zero. It does not apply Fatou directly to the signed energies or presume uniform integrability. Column bound and correction; expectation limit and final theorem.

For the Euler construction in §7, take $K = N$, relabel its Z_{j+1} as z_j, and set $f_j = g_{j+1}$. Each normalized increment depends only on $z_0, \dots, z_j$ and the increments are orthonormal when their raw second moments are positive. The rounded sign is spin-valued almost surely: its tie event is null by the independent last coordinate. The scalar implementation uses a measurable Bool spin equal to that sign almost surely. It follows that

$$S_{T,N} = \sum_{j=1}^{N} A_j B_{j+1} \le P_*.$$

This finite inequality assumes admissibility, $0 \le T < 1$, $N > 0$ and positive active second moments, but no minimizer or diffusion. Literal scalar specialization.

At a minimizer, the separate contact and diffusion arguments establish the fixed-strip inputs and eventual positivity. First take the mesh limit to obtain $I(T) \le P_*$, then let $T \uparrow 1$. The scalar terminal assembly identifies that last limit with $\mathcal{P}(\gamma)$. Together with $P_* \le \mathcal{P}(\gamma)$ this proves equality. The third scalar endpoint constructs its own diffusion before making this application. These sources therefore supply an independent route to the equality, alongside the imported Parisi formula in §6. Contact assembly. The inspection here does not change the research replay outcome in §13.

10.3 The zero coefficient

The admissible coefficient $\gamma_0 = 0$ makes the use of minimality in the normalization step explicit. Its diffusion is $X_t = B_t$, and the zero-coefficient heat formula gives

$$u_0(t,x) = 2F_{\mathrm{G}}\left(\frac{x}{\sqrt{1-t}}\right) - 1, \qquad a_0(t,x) = \sqrt{\frac{2}{\pi}}\frac{e^{-x^2/(2(1-t))}}{\sqrt{1-t}}.$$

Since $\mathbb{E}e^{-\lambda B_t^2} = (1+2\lambda t)^{-1/2}$,

$$\mathbb{E}a_0(t,B_t) = \sqrt{\frac{2}{\pi}}, \qquad \mathbb{E}a_0(t,B_t)^2 = \frac{2}{\pi\sqrt{1-t^2}}, \qquad q_0(t) = \frac{2}{\pi}\arcsin t.$$

The last identity follows by integrating $q_0' = \mathbb{E}a_0^2$ from $q_0(0) = 0$. This heat and late-step variation argument is the zero-field specialization of Chen, Handschy and Lerman, Theorem 6 and its proof, arXiv:1609.04368v4, pp. 13–16. Strict convexity of arcsin gives $q_0(t) < t$ for $0 < t < 1$. Thus $G_0(s) = \int_s^1 (q_0(t) - t)\,dt < 0$ for every $s < 1$. Inserting an atom at s is admissible, and its first variation is $G_0(s)/2 < 0$. In particular, the derivative for inserting mass at zero is $1/4 - 1/\pi < 0$. Hence this coefficient does not minimize.

Here the Euler scheme is exact: $Y_j = B_{t_j}$ under the Brownian coupling. Nevertheless its normalizers are

$$c_j = \sqrt{\frac{\pi}{2}}\,(1 - t_{j-1}^2)^{1/4},$$

so they approximate this nonconstant function of time, rather than converging uniformly to one. Already $c_1 = \sqrt{\pi/2}$ for every mesh. The directions g_j remain orthonormal by construction and the rounding still preserves conditional means. The missing input to the value limit is the minimizing identity $\mathbb{E}a(t,X_t)^2 = 1$.

The terminal spin also survives: $U_1 = \operatorname{sgn}(B_1)$ almost surely. But its correlation is

$$\mathbb{E}[U_1 B_1] = \sqrt{\frac{2}{\pi}} = \mathcal{P}(\gamma_0),$$

the value of this nonminimizing candidate. The negative first variation proves that this value is strictly above P_*. Admissibility, exact Euler sampling, and a spin-valued terminal martingale therefore do not supply the minimizing normalization or identify the ground-state value.

10.4 An exact one-step calculation

To test the definitions on a finite example, we use the admissible heat coefficient $\gamma_0 \equiv 0$, whose solution has

$$u_0(t,x) = 2F_{\mathrm{G}}(x/\sqrt{1-t}) - 1, \qquad a_0(0,0) = \sqrt{2/\pi}.$$

This coefficient is not a minimiser. The calculation therefore tests the construction and its indices, and leaves the minimiser-dependent convergence theorem aside. Take $T = 4/5$ and $N = 1$. Then

$$Y_1 = \frac{2}{\sqrt{5}} Z_1, \qquad c_1 = \sqrt{\pi/2}, \qquad g_1 = Z_1, \qquad m_1 = u_0(T, Y_1) = 2F_{\mathrm{G}}(2Z_1) - 1.$$

Strict monotonicity and symmetry of the Gaussian distribution function give

$$F_1 = \operatorname{sgn}(F_{\mathrm{G}}(2Z_1) + F_{\mathrm{G}}(Z_2) - 1) = \operatorname{sgn}(2Z_1 + Z_2)$$

almost surely, so the additional Gaussian is visibly the rounding coordinate. Set $V = (2Z_1 + Z_2)/\sqrt{5}$, a standard normal. Gaussian regression gives $\mathbb{E}[Z_1 \mid V] = 2V/\sqrt{5}$ and $\mathbb{E}[Z_2 \mid V] = V/\sqrt{5}$. Since $\mathbb{E}|V| = \sqrt{2/\pi}$,

$$A_0 = 0, \qquad A_1 = B_1 = 2\sqrt{\frac{2}{5\pi}}, \qquad B_2 = \sqrt{\frac{2}{5\pi}}.$$

Consequently

$$\sum_{j=0}^{1} A_j B_{j+1} = \frac{4}{5\pi} \quad \text{whereas} \quad A_1 B_1 = \frac{8}{5\pi}.$$

Dropping the index shift changes the answer by a factor of two.

At $N = 1$ the interior Riemann sum is empty, and the entire product comes from the rounding endpoint. That endpoint term disappears only asymptotically, and at a finite mesh it need not vanish. The displayed coefficients follow from exact Gaussian regression, so the example needs no numerical quadrature or finite computation.

10.4.1 The same test at every mesh size For the nonminimizing coefficient $\gamma = 0$, Euler sampling is exact: $Y_j = B_{t_j}$. The curvature moment in §10.3 gives

$$c_j = \sqrt{\pi/2}(1 - t_{j-1}^2)^{1/4}.$$

Monotonicity and symmetry of the Gaussian CDF turn the rounded spin into

$$F_N = \operatorname{sgn}(B_T + \sqrt{1-T}\, Z_{N+1}) \quad \text{almost surely.}$$

The argument of this sign is standard normal. Gaussian regression gives $B_i = \sqrt{2/\pi}\sqrt{\Delta}$ for $1 \le i \le N$ and $B_{N+1} = \sqrt{2/\pi}\sqrt{1-T}$. To calculate A_j, we condition on Brownian motion through t_j and use the heat martingale. Gaussian integration by parts then gives

$$\mathbb{E}[Z_j u_0(t_j, x + \sqrt{\Delta} Z_j)] = \sqrt{\Delta}\, a_0(t_{j-1}, x).$$

Multiplication by $c_j a_0(t_{j-1}, B_{t_{j-1}})$ and expectation yield $A_j = c_j\sqrt{\Delta}\, \mathbb{E} a_0(t_{j-1}, B_{t_{j-1}})^2 = \sqrt{\Delta}/c_j$. Consequently, for every $N \ge 1$,

$$S_{T,N} = \sqrt{\frac{2}{\pi}} \left(\Delta \sum_{j=1}^{N-1} \frac{1}{c_j} + \frac{\sqrt{\Delta(1-T)}}{c_N} \right).$$

The one-step value above is its $N = 1$ case. At fixed $T < 1$, its limit is

$$\frac{2}{\pi}\int_0^T (1-t^2)^{-1/4}\,dt,$$

whereas the unnormalized candidate covariance is $I_0(T) = T\sqrt{2/\pi}$. The subsequent $T \uparrow 1$ limit is $4\Gamma(3/4)/(\sqrt{\pi}\,\Gamma(1/4)) \approx 0.76275976$, distinct from $\mathcal{P}(0) = \sqrt{2/\pi} \approx 0.79788456$. The general predictable comparison of §10.2 turns this limit into a lower bound for P_*, as shown next. It determines neither P_* nor the value of the zero coefficient.

Exact Euler simulation, orthonormal predictable increments and correct rounding all hold here. The missing identity is $\mathbb{E}a_0(t, B_t)^2 = 1$, and consequently the normalizers do not approach one. The calculation extends the heat and rounding derivations of §§10.3–10.4 and isolates the role of minimality in the convergence theorem.

Corollary 10.2 (the heat test). In this normalization,

$$P_* \geq \frac{4\Gamma(3/4)}{\sqrt{\pi}\,\Gamma(1/4)} \approx 0.7627597635. \tag{10.2}$$

For each fixed $T < 1$ and $N \geq 1$, the heat normalizers c_j are positive. The increments g_j have unit second moments and are orthogonal, since their multipliers depend only on preceding Gaussians. The rounded F_N is a measurable spin after an arbitrary choice on its null tie event. Thus Proposition 10.1 applies with $K = N$, $z_j = Z_{j+1}$ and $f_j = g_{j+1}$, giving $S_{T,N} \leq P_*$. First let $N \to \infty$ at fixed T, using the displayed exact sum, and then let $T \uparrow 1$. The integrand is nonnegative and its endpoint singularity has exponent $1/4 < 1$. Substitution $s = t^2$ yields

$$P_* \geq \frac{2}{\pi}\int_0^1 (1-t^2)^{-1/4}\,dt = \frac{1}{\pi}B\Big(\frac{1}{2},\frac{3}{4}\Big) = \frac{4\Gamma(3/4)}{\sqrt{\pi}\,\Gamma(1/4)}.$$

The bound follows from the heat calculation and the general Gaussian comparison. Minimality is unnecessary for the inequality. It is what made the different, optimizing construction converge to P_* itself.

11. Alternative constructions and circuit limits

11.1 The separate positive-temperature route

The positive-temperature appendix supplies a separate classical input to the same synthesis and seed-removal theorems. Fix $\beta_0 > 1$, let μ_{β_0} be its Parisi measure, and write $q = q_{\beta_0} = \max\operatorname{supp}\mu_{\beta_0}$. Here $\gamma(t) = \beta_0\mu_{\beta_0}([0,t])$, $m = \phi_{\beta_0,x}$, $a = \phi_{\beta_0,xx}$ and X belong to the positive-temperature PDE, whose terminal value is $\beta_0^{-1}\log(2\cosh(\beta_0 x))$. They are distinct from the zero-temperature objects in §§1–8. The input credited to Lopatto, arXiv:2607.11756v3, Theorem 1.1 gives $\operatorname{supp}\mu_{\beta_0} = [0,q]$ for $\beta_0 > 1$, and its smooth density below q gives the regularity used on a fixed $[0,T]$ with $0 < T < q$.

The normalization can be checked directly against Lopatto's PDE. Writing $\widehat{\Phi}_{\beta_0}(t,y)$ for his solution with terminal value $\log\cosh y$ and covariance factor β_0^2, set

$$\phi_{\beta_0}(t,x) = \frac{\widehat{\Phi}_{\beta_0}(t,\beta_0 x) + \log 2}{\beta_0}, \quad m = \widehat{\Phi}_y(t,\beta_0 x), \quad a = \beta_0\widehat{\Phi}_{yy}(t,\beta_0 x).$$

His diffusion $\widehat{X}$ has noise coefficient β_0, hence $X = \widehat{X}/\beta_0$ has unit noise and drift $\beta_0\mu_{\beta_0}([0,t])m$. His identity $\beta_0^2\mathbb{E}\widehat{\Phi}_{yy}(t,\widehat{X}_t)^2 = 1$ is exactly $\mathbb{E}a(t,X_t)^2 = 1$. This is a fixed-temperature change of variables, not convergence of either Parisi measure as $\beta_0 \to \infty$. Lopatto, arXiv:2607.11756v3, equations (1.3), (2.1)–(2.2) and Lemma 2.2.

The source derives $\mathbb{E}a(t,X_t)^2 = 1$ on $[0,q]$ and $1-q \le \beta_0^{-1}$. The value estimate is

$$\int_0^T \mathbb{E}a(t,X_t)\,dt \ge P_* - \frac{1/4+\log 2}{\beta_0} - (q-T).$$

To recover its constants, martingale contact first gives $\int_0^t \mathbb{E}a(s,X_s)^2 ds = t$ on $[0,q]$, and continuity yields unit second moment including the endpoints. On $[q,1]$, $\gamma = \beta_0$, so Cole–Hopf gives

$$\phi(q,x) = \beta_0^{-1}\log(2\cosh\beta_0 x) + \beta_0(1-q)/2, \quad m(q,x) = \tanh(\beta_0 x), \quad a(q,x) = \beta_0\operatorname{sech}^2(\beta_0 x).$$

The endpoint identity $\mathbb{E}a(q,X_q)^2 = 1$ is also needed. Cauchy–Schwarz gives

$$1-q = \mathbb{E}\operatorname{sech}^2(\beta_0 X_q) \le \left(\mathbb{E}\operatorname{sech}^4(\beta_0 X_q)\right)^{1/2} = \beta_0^{-1}.$$

This is the endpoint argument in OpenAI (2026b), Appendix A, Lemma "A positive-temperature value bound". Itô's product rule and the value evolution give

$$\int_0^q \mathbb{E}a = \mathbb{E}[X_q m(q,X_q)] - \int_0^q t\gamma(t)dt, \quad \mathbb{E}\phi(q,X_q) = \phi(0,0) + \tfrac{1}{2}\int_0^q t\gamma(t)dt.$$

To compare this integral with the spin optimum, define the expected free energy in energy units by

$$F_{n,\beta_0} = (n\beta_0)^{-1}\mathbb{E}\log\sum_\sigma e^{\beta_0 H_n(\sigma)}, \qquad F_{\beta_0} = \lim_{n\to\infty} F_{n,\beta_0}.$$

The fixed-temperature limit is supplied by the positive-temperature Parisi formula cited below. The finite-size soft-maximum inequality gives $F_{\beta_0} \ge P_*$. Using $x\tanh(\beta_0 x) \ge \beta_0^{-1}\log(2\cosh\beta_0 x) - \beta_0^{-1}\log 2$ and retaining the penalty on $[q,1]$ yields

$$\int_0^q \mathbb{E}a \ge F_{\beta_0} - \frac{\beta_0(1-q)^2}{4} - \frac{\log 2}{\beta_0}.$$

The coefficient $1/4$ comes from the two tail contributions, since $\gamma = \beta_0$ above q:

$$\frac{\beta_0}{2}\int_q^1 t\,dt - \frac{\beta_0}{2}(1-q) = -\frac{\beta_0}{4}(1-q)^2.$$

The cutoff removes at most $q-T$, because $0 \le \mathbb{E}a \le 1$. This proves the displayed estimate. Choose first a sufficiently large finite β_0, then $T < q$ close enough to q. This route obtains the existence and value of the ground-state limit from the positive-temperature Parisi formula of Talagrand (2006), Theorem 1.1, pp. 222–223. At each fixed β_0, his covariance hypotheses (1.3)–(1.5) hold with $\xi(t) = \beta_0^2 t^2/2$ and error $c(n) = \beta_0^2/(2n)$. His partition sum (1.14) is unnormalized, and dividing its logarithmic free energy by β_0 gives the convention here, including the $\log 2/\beta_0$ entropy term. The finite-size inequality used above is

$$F_{n,\beta_0} - (\log 2)/\beta_0 \le n^{-1}\mathbb{E}\max H_n \le F_{n,\beta_0}.$$

It does not invoke the zero-temperature companion to justify that step. Write $M_n = n^{-1}\mathbb{E}\max H_n$. The same uniform entropy bound gives

$$|F_{n,\beta_0} - P_*| \le |M_n - P_*| + \frac{\log 2}{\beta_0}.$$

Thus arbitrary joint $n, \beta_0 \to \infty$ are permitted for these scalar free-energy values. This says nothing about joint convergence of measures, PDE derivatives, message calculations or circuit words. OpenAI (2026b), Appendix A.

The remaining limits have distinct tasks. At fixed β_0, T, δ, L, with $\delta > 0$, integer $L \ge 1$ and $L\delta \le T$, the degree limit $D \to \infty$ identifies the finite message law and its edge value. Let $U_0, U_1, \dots$ be independent standard normals. Put

$$X_0^\delta = \sqrt{\delta}\, U_0, \qquad X_{r+1}^\delta = X_r^\delta + \delta\gamma(r\delta)m(r\delta, X_r^\delta) + \sqrt{\delta}\, U_{r+1}, \qquad c_{\delta,r} = (\mathbb{E}a(r\delta, X_r^\delta)^2)^{-1/2}.$$

The finite-degree output is a message calculation on the $(D+1)$-regular tree, where ∂i is the set of neighbours of vertex i. Write $t_r = r\delta$ and $b(t,x) = \gamma(t)m(t,x)$. Initialise independent standard normal labels g_i, with $u_{i\to j}^0 = u_i^0 = g_i$, $A_{i\to j}^{-1} = 1$ and $x_{i\to j}^0 = x_i^0 = \sqrt{\delta}\, g_i$. For $0 \le r < L$ set

$$u_{i\to j}^{r+1} = D^{-1/2} \sum_{\substack{v \in \partial i \\ v \ne j}} A_{v\to i}^{r-1} u_{v\to i}^r, \qquad u_i^{r+1} = (D+1)^{-1/2} \sum_{v \in \partial i} A_{v\to i}^{r-1} u_{v\to i}^r.$$

Each x then takes the Euler step $x^{r+1} = x^r + \delta b(t_r, x^r) + \sqrt{\delta}\, u^{r+1}$ with its corresponding directed-edge or vertex message. The directed-edge law fixes the common normaliser:

$$c_{D,r} = (\mathbb{E}a(t_r, x_{i\to j}^r)^2)^{-1/2}, \quad A_{i\to j}^r = c_{D,r}a(t_r, x_{i\to j}^r), \quad B_i^r = c_{D,r}a(t_r, x_i^r), \quad z_i^L = \sqrt{\delta}\sum_{r=1}^{L} B_i^{r-1} u_i^r.$$

The one-step lag in $A^{r-1}u^r$ gives conditional centring from the new outer Gaussian layer, although successive finite-degree messages need not be independent. At fixed δ, L, the joint vertex messages converge to the independent U_r above, and z_i^L converges to $\sqrt{\delta}\sum_{r=1}^{L} c_{\delta,r-1}a(t_{r-1}, X_{r-1}^\delta)U_r$. The source's separate edge calculation gives

$$\lim_{D\to\infty} \frac{D+1}{2\sqrt{D}} \mathbb{E}[z_o^L z_v^L] = \delta \sum_{r=1}^{L-1} c_{\delta,r}\mathbb{E}a(r\delta, X_r^\delta), \qquad o \sim v.$$

Only afterwards take $\delta \downarrow 0$ with $L = \lfloor T/\delta \rfloor$. The source bounds the discrepancy from $\int_0^T \mathbb{E}a(t, X_t)\, dt$ by $C(\beta_0, T)\sqrt{\delta}$. A separate norm-convergence argument with a uniform first occupation moment identifies this edge limit with the rooted energy. A finite-dimensional central limit theorem alone does not identify that observable. OpenAI (2026b), Appendix A.

Source comparison for the finite weighted-tree formula. This auxiliary comparison concerns the positive-temperature route, and a reader following the circuit argument can go directly to §11.1.1 or §11.2.

The comparison is with El Alaoui–Montanari–Sellke, arXiv:2111.06813v2, 3 February 2023. Proposition 2.1 (p. 7) sums the adjacent-time term over $\ell = 2, \dots, L$. Applying Lemma 4.3, equation (4.6) (p. 14), replaces the expectation in that summand by the coefficient at $r = \ell - 1$, so the resulting range is $r = 1, \dots, L-1$. The displayed ranges in (4.5) (p. 13) and (4.18) (p. 16) are instead $2, \dots, L$.

Separately, the same-time display (2.21) in Lemma 2.8 (p. 9) omits a factor $1/k$ that is present in its proof and in Proposition 2.1. The QAOA manuscript's Appendix A already uses the latter proposition, the reindexed sum and spatial Lipschitz bounds for both a and b. The two following calculations establish the finite discrepancies and their precise scope. They do not compare formulae in the journal edition.

The test weights expose the missing factor. Work on the k-regular tree, $k \geq 2$, and let all initial labels g_i be independent standard Gaussians. In particular, the adjacent vertices o, v have distinct labels g_o, g_v. Set $u_i^0 = u_{i\to j}^0 = g_i$, $A_{i\to j}^{-1} = 1$, and use constant algorithmic update functions $F_r = H_r = 1$. These satisfy EAMS Assumption 2 and its later boundedness Assumption 3. For the auxiliary tests in Lemma 2.8 choose $Y_o^r = g_o$ and $\widetilde{Y}_v^r = g_v$ for every $r \geq 0$. Each test is measurable with respect to its required initial-label sigma-field, and all moments are finite. The tests are separate from the constant update functions.

At $t = 1$, the recursion gives

$$u_o^1 = k^{-1/2} \sum_{x\sim o} g_x, \qquad u_v^1 = k^{-1/2} \sum_{y\sim v} g_y.$$

Hence

$$\mathbb{E}[u_o^1 u_v^1 g_o g_v] = \frac{1}{k} \sum_{x\sim o,\, y\sim v} \mathbb{E}[g_x g_y g_o g_v] = \frac{1}{k}.$$

Only $x = v, y = o$ survives. The printed right side of (2.21) is $\mathbb{E}[g_v^2 g_o^2] = 1$, and multiplying it by $1/k$ agrees with the calculation. Both initial labels matter. With both test weights equal to one, both sides are zero. With just the neighbouring test omitted, the left side is an odd Gaussian moment and is again zero.

A bounded version uses $Y_o^r = \operatorname{sgn}(g_o)$ and $\widetilde{Y}_v^r = \operatorname{sgn}(g_v)$. Only the same edge pair survives, so

$$\mathbb{E}[u_o^1 u_v^1 \operatorname{sgn}(g_o) \operatorname{sgn}(g_v)] = \frac{1}{k}(\mathbb{E}|g|)^2 = \frac{2}{\pi k},$$

whereas the printed right side is $2/\pi$. These tests are bounded and initial-label measurable. Both witnesses concern the auxiliary weighted identity, and neither substitutes arbitrary test weights into a Parisi output theorem.

The shifted endpoint violates the printed uniform finite bound. For a separate test retain the initial labels and set

$$\delta = \varepsilon = \frac{1}{4}, \qquad L = 2, \qquad b(t,x) = 0, \qquad a(t,x) = \begin{cases} 1, & 0 \leq t < 3/8, \\ -1, & 3/8 \leq t < 1. \end{cases}$$

Use the source's actual initial states $x_i^0 = x_{i\to j}^0 = \sqrt{\delta}\, g_i$ and its definitions $A_{i\to j}^r = a(r\delta, x_{i\to j}^r)$, $B_i^r = a(r\delta, x_i^r)$. The state initialization has not been set to zero. It drops out of these coefficients because a is independent of x. This fixed function is bounded by $K = 1$, has spatial Lipschitz constant zero, and satisfies $a^2 = 1$, so all directed-edge normalizers are exactly one. The drift has Lipschitz constant $M = 0$, and $L\delta = 1/2 \leq 1 - \varepsilon$. It therefore meets the printed hypotheses of Theorem 4.1 and Assumptions 2–3. That finite theorem does not require temporal continuity or positivity of a.

At the mesh points, $a_0 = a_1 = 1$ and $a_2 = -1$. Thus

$$u_i^2 = \frac{1}{\sqrt{k(k-1)}} \sum_{w\sim i} \sum_{y\sim w,\, y\neq i} g_y, \qquad z_i^2 = \sqrt{\delta}\,(u_i^1 + u_i^2).$$

The equal-time covariances across the edge are zero. Each adjacent-time covariance is $\sqrt{k-1}/k$: exactly $k-1$ labels occur in both sums, and their product normalization is $k\sqrt{k-1}$. Therefore

$$k\mathbb{E}[z_o^2 z_v^2] = 2\delta\sqrt{k-1}, \qquad \frac{k}{2\sqrt{k-1}}\mathbb{E}[z_o^2 z_v^2] = \delta.$$

The target printed in (4.5) uses $a(2\delta, x_o^2) = -1$, so it is $-2\delta\sqrt{k-1}$. Its absolute error is $4\delta\sqrt{k-1}$, which is unbounded as $k \to \infty$ while all arguments of the claimed constant $C(M, K, \delta, \varepsilon)$ remain fixed. This refutes auxiliary Theorem 4.1 as printed. The example is not a Parisi curvature: it changes sign. Consequently it is not by itself a counterexample to the specialized assertion (4.18) in Proposition 4.7, and the endpoint correction there comes from the exact proposition and reindexing explained above. Neither test refutes the main EAMS Theorem 1.7.

A finite estimate with the required hypotheses. The source argument does provide a usable replacement. Fix $\delta > 0$, $L \geq 1$ and $L\delta < 1$. Allow the coefficients a_k, b_k to depend on degree, but require bounds independent of k at the finitely many mesh times: $|a_k| \leq K$ with $K \geq 1$, and spatial Lipschitz constants L_a, L_b for a_k, b_k. Use the original initialization and the same normalized function for node and directed-edge coefficients, with $\mathbb{E}(A_{i\to j}^r)^2 = 1$. Then

$$\left| k\mathbb{E}[z_o^L z_v^L] - 2\sqrt{k-1}\,\delta \sum_{r=1}^{L-1} \mathbb{E}B_o^r \right| \leq C(K, L_a, L_b, \delta, L).$$

The constant is independent of k, and the sum is empty when $L = 1$. The coefficient expectation on the right uses the vertex law, while its normalization was defined using the directed-edge law. The following proof keeps the two laws distinct. It reconstructs the finite estimate from EAMS Proposition 2.1 and Lemmas 3.2–3.4 and 4.2–4.3. It does not claim that changing an index alone repairs every hypothesis of the printed theorem.

Put $D = k-1$. Conditional on the earlier cavity labels, the new message is centered Gaussian with variance $(\tau^r)^2$. Its recursion is

$$(\tau_{i\to j}^{r+1})^2 = \frac{1}{D} \sum_{\substack{w \in \partial i \\ w \neq j}} (A_{w\to i}^{r-1})^2 (\tau_{w\to i}^r)^2, \qquad \tau^0 = 1.$$

This conditional Gaussianity gives $\Delta_r = \mathbb{E}[((\tau^r)^2 - 1)^2] \leq V_r/D$, where

$$V_0 = 0, \qquad V_{r+1} = 2K^4 V_r + 2(K^4 + 1).$$

Indeed, split the conditional-variance recursion into the average of $(A^{r-1})^2((\tau^r)^2 - 1)$ and the average of $(A^{r-1})^2 - 1$. Jensen bounds the first square, while independence of child cavities and exact coefficient normalization bound the second by a constant over D. The moment recursion also gives $\mathbb{E}(u_{i\to j}^r)^2 \leq K^{2r}$.

Splitting off the missing neighbour as in the source gives $\mathbb{E}|u_i^r - u_{i\to j}^r|^2 \leq 4K^{2r}/k$. Subtract the node and directed-edge Euler recursions. Their equal initial states and the drift Lipschitz bound give

$$\mathbb{E}|x_i^r - x_{i\to j}^r|^2 \leq d_r/k, \quad d_0 = 0, \qquad d_{r+1} = 2(1 + \delta L_b)^2 d_r + 8\delta K^{2(r+1)}.$$

The bound $\|B_i^r - A_{i\to j}^r\|_2 \leq L_a\sqrt{d_r/k}$ now uses spatial Lipschitzness of a. This property is invoked in the proof of EAMS Lemma 4.2 but omitted from the printed hypotheses of Theorem 4.1. It is available in the QAOA manuscript's compact-strip application.

For $s = \ell - 1$, the adjacent-time summand of Proposition 2.1 is

$$Q_s = \mathbb{E}[A^{s-1}_{v\to o}(u^s_{v\to o})^2 B^{s-1}_v B^s_o].$$

Let $\mathcal{G}^{r,+}_{i\to j}$ be the labels within distance r of i in its cavity after removing edge ij. Condition on $\mathcal{G}^{s-1,+}_{v\to o} \vee \mathcal{G}^{s,+}_{o\to v}$. The three coefficient factors are measurable there: their vertex balls split into these two disjoint cavity balls. The newest layer driving $u^s_{v\to o}$ is outside this field. Adding the initial label at v to the source's minus field does not change the conditional law, since that message and its variance exclude that label. Conditioning therefore replaces $(u^s_{v\to o})^2 - 1$ by $(\tau^s_{v\to o})^2 - 1$. Its contribution is bounded by $K^3\sqrt{V_s/D}$. Replacing the two node coefficients by directed-edge coefficients then costs at most $K^2 L_a(\sqrt{d_{s-1}} + \sqrt{d_s})/\sqrt{k}$. The two directed-edge trees are independent, and exact normalization reduces their product expectation to $\mathbb{E}A^s_{o\to v}$. One final comparison with $\mathbb{E}B^s_o$ proves

$$|Q_s - \mathbb{E}B^s_o| \le \frac{C_s}{\sqrt{D}}, \quad C_s = K^3\sqrt{V_s} + K^2 L_a\sqrt{d_{s-1}} + (K^2+1)L_a\sqrt{d_s}.$$

For the same-time remainder in that proposition, $|R_\ell| \le K^{2\ell+2}$ by boundedness and Cauchy–Schwarz. Hence the asserted finite bound holds, for example, with

$$C = 2\delta \sum_{s=1}^{L-1} C_s + \delta \sum_{\ell=1}^{L} K^{2\ell+2}.$$

No constant uniform in increasing horizon, temperature or terminal cutoff has been asserted.

Why the continuum value survives the correction. The QAOA manuscript already checks the convergence and uniform boundedness of its finitely many degree-dependent normalizers, and supplies both spatial Lipschitz bounds, before taking the degree limit. Its limiting coefficient $m_r = c_{\delta,r}\mathbb{E}a(t_r, X^\delta_r)$ satisfies $|m_r| \le 1$ by Cauchy–Schwarz and the definition of $c_{\delta,r}$. The shifted and corrected sums differ exactly by

$$\delta \sum_{r=2}^{L} m_r - \delta \sum_{r=1}^{L-1} m_r = \delta(m_L - m_1), \qquad |\delta(m_L - m_1)| \le 2\delta.$$

Thus the error in the printed unnormalized finite theorem can grow like $\sqrt{k}$ at fixed mesh even though the correction disappears in the subsequent normalized continuum limit. The fixed-temperature coupling and separate strong realization in the QAOA manuscript justify that limiting application. These checks neither reject nor independently certify the whole EAMS or QAOA theorem. OpenAI (2026b), Appendix A.

After clipping and bounded approximation, the result enters the same synthesis and seed-removal route as the primary construction. Its source dependencies and fixed-temperature choice of parameters remain separate from the retained zero-temperature formal interfaces. OpenAI (2026b), Appendix A.

11.1.1 Where the temperature routes meet

Both routes produce a bounded finite classical calculation whose rooted energy is close to P_*. They reach that common input by different theorems and different finite approximations.

Variational input and support. The zero-temperature route fixes an attained minimizer $\gamma \in \mathcal{U}$ with terminal value $|x|$ and uses full relative Stieltjes support on $[0, 1)$. The positive-temperature

route fixes a finite $\beta_0 > 1$ and its Parisi probability measure, with terminal value $\beta_0^{-1}\log(2\cosh\beta_0 x)$. It uses Lopatto's interval-support result on $[0, q_{\beta_0}]$.

Scalar error. Before discretization, the zero-temperature loss is at most $1 - T$ for $T < 1$. The positive-temperature loss is at most

$$(1/4 + \log 2)/\beta_0 + (q_{\beta_0} - T), \qquad T < q_{\beta_0}.$$

Finite energy. In the zero-temperature route, a first-chaos projection identifies the explicit Gaussian coefficient sum with $\mathrm{Re}\langle F_N, \mathsf{L}F_N\rangle$. The alternative route first calculates finite-degree messages and edge energy, then uses a rooted norm and occupation argument.

Order of approximation. At zero temperature, fix T before refining the scalar mesh, and choose a finite mesh before regularization and synthesis. At positive temperature, fix β_0, T, δ, L for the message degree limit, then refine δ at fixed β_0, T.

Both routes then use the same bounded approximation, exclusions, seeded synthesis, seed removal and fixed-word transfer. The quantum construction is shared, and its classical input has been supplied independently.

Here the displayed positive-temperature estimate is exactly the one derived above, rather than an assertion of convergence of zero-temperature supports. One can choose β_0 and T finitely for the required error and then stop. The later circuit argument needs that finite choice and never asks for a joint limit of the temperature, message mesh, graph degree and circuit depth. OpenAI (2026b), Appendix A, Proposition "The independent positive-temperature construction"; §6.

11.1.2 Finite-temperature values and exact Gibbs samples

This comparison samples the exact Gibbs distribution. The QAOA Born law in §11.2 is a separate distribution, and none of the following bounds identifies the two. For one fixed instance, let $L_n = \max_\sigma H_{n,J}(\sigma)/n$, $\mathcal{F}_{n,\beta} = (n\beta)^{-1}\log\sum_\sigma e^{\beta H_{n,J}(\sigma)}$, and let $\mathcal{U}_{n,\beta}$ be the mean energy per spin under the Gibbs law proportional to $e^{\beta H_{n,J}(\sigma)}$. Then

$$0 \le \mathcal{F}_{n,\beta} - L_n \le \frac{\log 2}{\beta}, \qquad 0 \le L_n - \mathcal{U}_{n,\beta} \le \frac{\log 2}{\beta}.$$

The first is the log-sum bound. The second uses $\mathcal{F} = \mathcal{U} + S/(n\beta)$, where the Gibbs entropy satisfies $0 \le S \le n\log 2$, together with $\mathcal{U} \le L_n \le \mathcal{F}$. Thus both disorder averages approach P_* along every joint $n \to \infty$, $\beta \to \infty$, by the already established convergence of $\mathbb{E}L_n$.

There is also a finite probability estimate. Every configuration has Gaussian coefficient-vector norm $\sqrt{(n-1)/(2n^2)}$ after division by n, and their maximum has the same Lipschitz bound. Gaussian concentration gives, for $n \ge 2$,

$$\Pr_J(|L_n - \mathbb{E}L_n| > s) \le 2\exp\{-n^2 s^2/(n-1)\}.$$

For an exact Gibbs draw Σ^{G}, count the at most 2^n configurations with deficit at least s and bound the partition sum below by its maximal term. This gives, for every instance,

$$\Pr\{H_{n,J}(\Sigma^{\mathrm{G}})/n \le L_n - s \mid J\} \le \min\{1, e^{n(\log 2 - \beta s)}\}.$$

The lower tail below $P_* - s$ is bounded by $\Pr_J(L_n < P_* - s/2) + \min\{1, e^{n(\log 2 - \beta s/2)}\}$; the upper tail is at most the upper tail of L_n. Hence an exact Gibbs sample also approaches P_* in joint probability. These are elementary scalar consequences of the Gibbs law, entropy and Gaussian concentration. They supply neither an efficient sampler nor QAOA angles.

11.2 What is averaged, and a fixed-word probability consequence

The probability conclusion here uses a fixed-word second-moment theorem in addition to the expected-value construction of §§8.1–8.6.

The theorem in §8 is an existence statement about limiting expected energy, and two random choices lie behind that expectation. First the couplings J are sampled. Then a computational-basis measurement of the state produces a spin string Σ. For the single word W already chosen, put

$$E_n(J,\Sigma) = H_{n,J}(\Sigma)/n, \qquad \overline{E}_n(J) = \mathbb{E}[E_n(J,\Sigma) \mid J].$$

The quantity defining v_W is $\mathbb{E}_J\overline{E}_n(J)$. It is neither the energy of every measured string nor a worst-disorder guarantee. The temporary Gaussian labels used during synthesis add no randomness to the final angle vectors.

Farhi, Goldstone, Gutmann and Zhou supply the additional input. They prove, with the same Hamiltonian and gate-sign convention, that for every fixed word

$$\mathbb{E}_{J,\Sigma}E_n \longrightarrow v_W, \qquad \mathbb{E}_{J,\Sigma}E_n^2 \longrightarrow v_W^2.$$

Their second expression is the quantum expectation of $(C_{n,J}/n)^2$ averaged over disorder, not merely the square of the conditional mean. Consequently

$$\mathbb{E}_{J,\Sigma}(E_n - v_W)^2 \longrightarrow 0, \qquad \Pr_{J,\Sigma}(|E_n - v_W| > s) \longrightarrow 0 \quad (s > 0).$$

The law of total variance separates what vanishes:

$$\mathbb{E}(E_n - v_W)^2 = \mathbb{E}_J \operatorname{Var}(E_n \mid J) + \mathbb{E}_J(\overline{E}_n(J) - v_W)^2.$$

Both terms are nonnegative and hence tend to zero. Concentration of the conditional mean alone would not control a measured string. This is the elementary second-moment deduction from Farhi et al., arXiv:1910.08187v4, §4, equations (13)–(22).

Now run the circuit construction with accuracy $\varepsilon/2$ and freeze its resulting word. Since $v_W \geq P_* - \varepsilon/2$,

$$\Pr_{J,\Sigma}\{E_n < P_* - \varepsilon\} \leq \frac{4}{\varepsilon^2}\mathbb{E}_{J,\Sigma}(E_n - v_W)^2 \longrightarrow 0.$$

This follows from the expected-value construction **together with** the external second-moment theorem, and it lies outside both selected formal interfaces.

If $b_n(J) = \Pr(E_n < P_* - \varepsilon \mid J)$, then for every fixed $\tau > 0$,

$$\Pr_J\{b_n(J) > \tau\} \leq \mathbb{E}_J b_n(J)/\tau \longrightarrow 0.$$

Thus the same word also has small conditional measurement-failure probability for most disorder instances, in this precise sense. We claim nothing for every instance, for arbitrarily varying depths, or with a quantitative size threshold.

For K fresh state preparations at the same disorder instance, the conditional failures are independent. Their joint failure probability is therefore $\mathbb{E}_J b_n(J)^K$, which in general differs from $(\mathbb{E}_J b_n(J))^K$. A fraction α of instances that always fail gives α for the former at every K. A bound $b_n(J) \leq \tau$ on a particular instance does give τ^K there. Repeatedly measuring an already collapsed state is a different experiment.

Comparison with the same finite instance. The concentration of the finite maximum in §11.1.2 strengthens this to a near-optimal sampled configuration. For every $\varepsilon > 0$, there is one finite deterministic word W, independent of n and J, such that

$$\Pr_{J,\Sigma}\Big\{H_{n,J}(\Sigma) \geq \max_{\sigma} H_{n,J}(\sigma) - \varepsilon n\Big\} \longrightarrow 1.$$

To see this, choose the word with $v_W \geq P_* - \varepsilon/4$ and write $r_n = \mathbb{E}(E_n - v_W)^2 \to 0$ and $M_n = \max_\sigma H_{n,J}(\sigma)/n$. The convergence $\mathbb{E}M_n \to P_*$ makes its bias at most $\varepsilon/4$ for all sufficiently large n. The Gaussian coefficient-vector bound from §11.1.2 and Chebyshev's inequality for E_n then give, for $n \geq 2$,

$$\Pr_J\{M_n > P_* + \varepsilon/2\} \leq 2\exp\left\{-\frac{n^2\varepsilon^2}{16(n-1)}\right\}, \qquad \Pr_{J,\Sigma}\{E_n < v_W - \varepsilon/4\} \leq \frac{16 r_n}{\varepsilon^2}.$$

Outside these two events, $M_n - E_n \leq \varepsilon$. Their union therefore has probability tending to zero. The maximum and the measured energy use the same disorder, and the union bound needs no independence.

One fresh preparation and measurement suffices for this asymptotic statement. Its inputs are the ordinary-word theorem, the external fixed-word second-moment theorem and Gaussian concentration, and it lies outside the selected formal endpoints. The unspecified sequence r_n and mean-maximum bias still give no effective size threshold or efficient angle-selection procedure.

11.3 Depth, angle choice and the order of quantifiers

The complete witness is a finite ordered word, including every real angle. Its existence for each error tolerance asserts uniformity in system size and disorder that fixed-instance convergence does not provide. We read the quantifiers literally.

The source theorem says that for every $\varepsilon > 0$, there is one finite word W_ε such that

$$\lim_{n\to\infty} \mathbb{E}E_n(W_\varepsilon) \geq P_* - \varepsilon.$$

The cost operator uses J, n, but this word and its angles do not.

In fixed-instance QAOA convergence, by contrast, the chosen depth and angles may depend on the fixed finite cost matrix and the requested accuracy. The performance of a prescribed family of nonzero angles with $p = p(n) \to \infty$ needs further estimates. A depth allowance alone can be met by the padded existence construction below.

The distinction already occurs for the elementary array $a_{n,p} = p/(n+p)$. For each fixed n, its limit as $p \to \infty$ is one, and for each fixed p, its limit as $n \to \infty$ is zero. This is a quantifier example and makes no claim to model QAOA. The finite-dimensional convergence theorem of Binkowski, Koßmann, Ziegler and Schwonnek, arXiv:2302.04968v1, Theorem 12 fixes the optimization problem before selecting a sufficiently long circuit. It does not replace the construction of a size-independent word here.

An existential schedule under a prescribed depth allowance.

There is a narrower diagonal consequence for sampled output. Choose W_k with $P_* - 1/k \leq v_{W_k} \leq P_*$. Fixed-word second-moment convergence supplies N_k so that, for every $n \geq N_k$, $\mathbb{E}(E_n(W_k) - v_{W_k})^2 \leq k^{-2}$. Given any prescribed integer depth allowance $d(n) \to \infty$, enlarge these strictly increasing

thresholds until $d(n) \geq \mathrm{depth}(W_k)$ for every $n \geq N_k$. On $N_k \leq n < N_{k+1}$, use W_k padded with zero-angle layers to depth $d(n)$. Padding changes no state, and

$$\mathbb{E}(E_n(W_k) - P_*)^2 \leq 2\mathbb{E}(E_n(W_k) - v_{W_k})^2 + 2(v_{W_k} - P_*)^2 \leq 4/k^2.$$

Thus a suitable schedule approaches the optimum in joint L^2, even under an arbitrarily slowly diverging allowance. The word family and thresholds are selected nonconstructively. This says nothing about an independently prescribed family of nonzero angles, and is not an interchange of limits or an effective depth bound.

Computational meaning of a finite word.

Nor does a supremum over angles assert that an optimizing angle vector exists at any particular depth. The proof constructs a sufficiently good word directly, and monotonicity from identity padding then gives the limit of Q_p. Each QAOA round contains a full cost evolution and a mixer evolution. In the dense SK model a cost evolution involves all $\binom{n}{2}$ pair couplings, so the round count is not itself a bound on two-qubit gate depth, connectivity cost, precision or physical runtime.

Three computational questions remain distinct: evaluating the limiting energy at supplied angles, finding suitable angles, and implementing and sampling the resulting quantum circuit. The Basso et al. iteration evaluates the fixed-angle limit in $O(p^2 4^p)$ arithmetic time, which bounds the cost of evaluation. The search needed to find the existence theorem's angles is a separate problem.

Spin–boson representations provide another route to evaluation. The earlier Boulebnane et al. mapping, arXiv:2505.07929v2, and the later mixed-spin decoupling paper, arXiv:2609.40309v1, §§II–III and Supplement S2.1, must be distinguished. The latter is a separate work, distinct from the reference called BKL2026 in the original QAOA manuscript. Its fixed-depth mapping and numerical tensor-network performance do not supply an efficient angle-selection theorem for the present construction. Its discussion also gives fixed-word finite-size estimates. The absence of a finite-size rate **in the construction here** is not a claim that the wider literature contains none.

Why the two-spin hypothesis matters.

The statement is specific to the quadratic, zero-field SK model. The pure even higher-spin model has a different thermodynamic-first conclusion. For every even interaction order $q \geq 4$, Basso, Gamarnik, Mei and Zhou, arXiv:2204.10306v2, Theorem 4, (3.15), and its proof in §4.2 show

$$v_p^{(q)}(\theta, \beta) \leq \eta_{\mathrm{OGP}}(G_q) < \eta_{\mathrm{OPT}}(G_q)$$

for every finite p and every fixed angle vector, in that source's normalization. The same ceiling bounds their supremum over all finite p and angles. The proof combines its dense–sparse value identity with the coupled-overlap-gap obstruction of Chou, Love, Sandhu and Shi, arXiv:2108.06049v3, Parameter 4.1 and Corollary 4.4. That local-algorithm obstruction applies to sparse random signed Max-q-XOR for even $q \geq 4$, at sufficiently large fixed degree and under the source's logarithmic-depth bound. The Basso–Gamarnik–Mei–Zhou transfer above only needs the resulting obstruction at each fixed depth. It requires no interchange of depth and system size. The exclusion $q \geq 4$ is essential, so this is not an obstruction for the SK two-spin model treated here.

For comparison, Goh, arXiv:2404.06087v6, Theorem 4.2 and Corollary 4.2.1, treats even interaction order at least four. Its logarithmic-depth obstruction concerns sparse random-regular Max-q-XORSAT, and the dense spin-glass corollary is explicitly conditional on swapping limits. The source states on p. 12 that its sparse/dense equivalence is only established at fixed depth, so it does

not prove a dense logarithmic-depth obstruction. These hypotheses do not include quadratic SK. Neither a conclusion for all mixed spin glasses nor a general quantum advantage follows from §8.

11.4 Limits of the regular-MaxCut consequence

The original manuscript also gives a consequence for MaxCut. For each $\eta > 0$ it selects one finite depth and normalized angles, and then a degree threshold, such that every finite simple unweighted $(D+1)$-regular graph of girth greater than $2p+1$ has expected cut fraction at least

$$\frac{1}{2} + \frac{P_* - \eta}{\sqrt{D}}$$

for all sufficiently large D. The usual MaxCut cost is $H_{\mathrm{MC}} = \frac{1}{2}\sum_{\{i,j\}}(I - Z_i Z_j)$, so its angles are $2\gamma/\sqrt{D}$: the normalized angles are fixed, while these physical cost angles depend on degree. The sign conversion uses a bipartition of the relevant tree neighbourhood of each edge, so the whole graph need not be bipartite.

The further optimality comparison is for a uniform simple labelled $(D+1)$-regular graph $G_{n,D+1}$, along sizes with $n(D+1)$ even. Write $\mathrm{OPT}(G)$ for the maximum cut size divided by $|E(G)|$, $q_{p,G}$ for the circuit's expected cut fraction, and $\Delta_D(G) = \mathrm{OPT}(G) - q_{p,G}(2\gamma/\sqrt{D}, \beta)$. The same word can be chosen so that

$$\lim_{D\to\infty} \limsup_{\substack{n\to\infty \\ n(D+1)\ \text{even}}} \Pr_G\{\sqrt{D}\,\Delta_D(G_{n,D+1}) > \eta\} = 0.$$

Size goes to infinity first at fixed degree and word. The probability is over the random graph, and the cut fraction has already averaged the quantum measurement. For fixed D, p, McKay, Wormald and Wysocka (2004), Theorem 1, gives tight counts of cycles of lengths at most $2p+3$. Every edge with a nontree neighbourhood is within bounded distance of one of these cycles, so their fraction tends to zero.

The optimum is supplied separately by Dembo, Montanari and Sen, arXiv:1503.03923v2, Theorem 1.6(b): dividing their cut size per vertex by $(D+1)/2$ gives the leading cut fraction $1/2 + P_*/\sqrt{D+1}$, with an error $o(D^{-1/2})$ in the stated iterated regime. A high-girth lower bound alone would not prove optimality on every high-girth graph. See OpenAI (2026b), Corollary 1.2 and its complete proof. We include this separate consequence to keep the source's scope in view. It adds no selected formal endpoint and no joint size–degree limit.

11.5 A later finite-size estimate and its limited depth consequence

Boulebnane, Khan, Subrahmanya, Herman, Farhi, Villalonga and Shaydulin later gave a quantitative comparison with the fixed-angle limit arXiv:2609.40309v1, Corollary S3.1, pp. 46–47. They use the ordered-tuple Gaussian mixed-spin cost with covariance polynomial $\xi(t) = \sum_q c_q^2 t^q$ of finite degree. Let $\Gamma > 0$ bound every cost angle in a depth-p word. Choose $x > 0$ bounding $\xi(2), \xi'(1), \xi''(33)$, with $\Gamma^2 p^2 x \geq 1$. The mixer angles need no bound. Set

$$B_p = 256p^4(2e\Gamma^2 p^2 x)^p, \qquad E_p = 70\Gamma^2 p^2 x + p\log 2, \qquad R_p = (2E_p/\log 2)^{2p}.$$

For

$$0 < \varepsilon \leq 4\Gamma p x \min\{e^{-1/2}, 2^{-(8p^2+3)/100}\},$$

the source asserts $|\nu_n - \nu_\infty| \le \varepsilon$ whenever

$$n \ge T_p := \max\left\{ 2^{8p^2+3} B_p \left(\frac{2^{\deg\xi+4}\Gamma^2 p^2 x^2}{\varepsilon^2} \right)^{R_p} \left[2\log_2 \frac{4\Gamma p x}{\varepsilon} + \deg\xi \right]^2, \frac{16\Gamma p x}{\varepsilon} \right\}.$$

Here ν_n is the disorder-averaged energy per spin.

The estimate applies to each near-optimal word from §8.6 separately. Given $\zeta > 0$, choose a finite word W_ζ with $\nu_\infty(W_\zeta) \ge P_* - \zeta/2$. Its depth p_ζ and cost angles are finite. We pad by a zero layer if needed and take $\Gamma_\zeta = \max\{1, \max_j |\theta_j(W_\zeta)|\}$. Choose a positive $\varepsilon_\zeta \le \zeta/2$ satisfying the displayed small-error condition. With these fixed parameters, the threshold gives

$$\nu_n(W_\zeta) \ge P_* - \zeta \qquad (n \ge T_{p_\zeta}).$$

The threshold is explicit in the chosen word and error. The existential construction supplies no quantitative bounds for p_ζ or Γ_ζ as $\zeta \downarrow 0$, so it gives no bound in terms of the accuracy alone. The fixed-word application needs no common angle bound, while the uniform growing-depth consequence below does.

For SK take $c_2 = 1/\sqrt{2}$. Combining the two ordered off-diagonal couplings recovers our standard Gaussian J_{ij}, and the diagonal adds an independent scalar Gaussian of variance $1/2$. That scalar changes the circuit by a global phase and has zero mean, so the expected-energy comparison uses the same normalization. A fixed $x \ge \max\{2, \Gamma^{-2}\}$ satisfies the above bounds for every $p \ge 1$.

For fixed mixture, Γ and x, put $\varepsilon_p = e^{-p^2}$. The small-error condition holds eventually, and direct logarithms give

$$\log R_p = 4p\log p + O(p), \qquad \log\log T_p = 4p\log p + O(p + \log p).$$

Consequently, for any deterministic sequence of words W_n with depths $p(n) \to \infty$, one common cost-angle bound Γ, and $p(n)\log p(n) = o(\log\log n)$, the displayed hypotheses hold and $|\nu_n(W_n) - \nu_\infty(W_n)| \le e^{-p(n)^2}$ eventually. This is a uniform comparison along that specified bounded-angle class. It gives neither an efficient way to find near-optimal angles nor concentration of a quantum measurement in this growing-depth regime. The fixed-word second-moment theorem in §11.2 has separate quantifiers.

12. The formal statements and their scope

12.1 The support statement and the derivative bridge

The first formal statement assumes a Brownian system and an admissible minimizer. It supplies a Stieltjes measure, an associated diffusion, full relative support and strict increase, and the two moment identities for every associated diffusion. Its proof assembles contact, exclusion of a terminal plateau, curvature for positive and zero constant intervals, and uniqueness to transfer the moments between associated diffusions. This describes the inspected proof's mathematical route, and we claim no successful research-target replay.

Its quantifiers are $\forall W\, \forall\gamma$, followed by the minimizer hypothesis, and it supplies both witnesses on the given Brownian system. The moment conclusion is universal over every diffusion satisfying the specified equation, and covers exactly $0 \le t < 1$. The diffusion equation itself holds through time one, simultaneously for all those times outside one null set. Controls are progressive for the natural filtration of Brownian increments and bounded by one pointwise. These are not additional

regularity or minimizer-existence assumptions. See the exact support statement and the companion source-interface review.

The manuscript's first variation and its formal counterpart differ in one useful respect. The formal result initially gives a derivative along chord lengths $1/(n+1)$. For this particular control functional, that sampled limit determines the ordinary feasible right derivative. With a control fixed, the terminal absolute value is convex in the coefficient while the penalty is linear. The supremum over the same control class, and then subtraction of the linear Parisi penalty, preserve convexity.

The secant slopes therefore vary monotonically as the chord shortens, and each intervening chord lies between two successive sampled chords and has the same limiting slope. If η is an integrable feasible direction, choose $\varepsilon_0 > 0$ so that $\gamma + \varepsilon_0\eta$ is admissible and rescale the chord. This bridge is an ordinary deduction from the definitions and the sampled limit, and no separately named formal theorem states it.

At the origin, the manuscript smoothly extends auxiliary quotient variables. The formal coercivity proof for the spatial diffusion density instead works on the positive half-line, controls the zero flux at the origin and reflects. The two must agree in the resulting integrated inequality, while auxiliary values at zero may differ. In particular, the declared generic flux lemma has a global derivative hypothesis and is not the supplier used here. The finite-density proof uses the Gaussian half-line coercivity lemma and reflection. A totalized quotient value at zero cannot silently satisfy the generic global hypothesis. The precise source spans are in the correspondence record.

The fixed-time supplier `actual_coercivity` assumes positive time and continuity of the coefficient at that time. Its scaled polynomial inequality is (4.1) multiplied by c^2, and division requires $c > 0$. This gives a direct source destination for Proposition 4.1 without adding a minimizing or plateau premise.

12.2 The separate scalar statement

The second formal statement has three clauses. Given a minimizer and an associated diffusion, it establishes convergence of the finite-spin ground-state values, a bounded martingale with a spin-valued terminal limit, convergence of the gradient and the value identities. Under the same hypotheses and fixed $0 < T < 1$, it gives eventual positivity of the Gaussian normalizers and convergence of the finite coefficient sum.

These are the declarations `scalar_value_target` and `finite_gaussian_target`. Their caller supplies the same minimizer and associated diffusion. The implementation derives compact-strip bounds from `sourceStripRegularity` and the unit curvature moment from contact for those objects before applying its fixed-strip helper. Those internal inputs are not additional hypotheses of the exported clauses. The convergence clause establishes existence of the finite-spin limit.

The index match in the Gaussian statement is exact. For $\Delta = T/N$ and $t_j = j\Delta$, the Euler path uses $Z_1, \dots, Z_N$ and produces $Y_0, \dots, Y_N$. The source normalizer indexed from zero corresponds to manuscript $c_1, \dots, c_N$, and its eventual positivity covers precisely those normalizers. A separate Z_{N+1} performs mean-preserving rounding. It enters B_{N+1} and is no Euler increment. Thus the finite sum is $\sum_{j=0}^{N} A_j B_{j+1}$ with $A_0 = 0$, using exactly the displayed index ranges. Total definitions at $N = 0$ make no extra mathematical claim about the construction for $N \geq 1$.

More explicitly, the zero-based formal quantities `coordinate N j`, `rawCoefficient ... j` and `normalizer ... j` correspond to Z_{j+1}, $a(t_j, Y_j)$ and c_{j+1} for $0 \leq j < N$. The positivity quantifiers

are $\exists N_0\, \forall N \geq N_0\, \forall j < N$, simultaneously over the growing mesh, for each already fixed T. The extra coordinate with formal index N is Z_{N+1}.

The formal martingale stops at its terminal random variable U for all later times. Then $U^2 = 1$ almost surely, and the gradient converges to U almost surely and in L^2 from below one. The internal terminal-gradient lemma proves $UX_1 = |X_1|$ without requiring X_1 to have no atom at zero. That identity is used inside the implementation and is no extra conjunct of the endpoint. See the scalar statement and its complete dictionary.

The third declaration, `ground_state_parisi_target`, assumes a minimizer alone and concludes $P_* = \mathcal{P}(\gamma)$. Its proof chooses $X =$ `diffusionProcess` $W\ \gamma$ and proves the diffusion predicate for that process before applying the equality with a supplied diffusion. No diffusion is required from the caller, and none of these clauses constructs a minimizer. The exact declarations are MainConclusions, lines 13–27; their analytic suppliers and applications are in ContactAssembly. The independent contact and lower-bound arguments are reconstructed in §10.1 and §10.2.

12.3 Relating the two statements

The support and scalar models encode closely related control problems. A direct formal application still needs a specified map of their objects. The following contract lists the obligations, and it asserts no equivalence of their complete interfaces.

Coefficients. Restrict the scalar real-time coefficient to $[0, 1)$. Conversely, choose an extension compatible with the scalar right-continuity predicate. Prove equality almost everywhere of the coefficients used by each payoff and penalty. The scalar endpoint and exterior values are unconstrained by the shape conditions, so two extensions need not be literally equal.

Brownian space. Bundle or unbundle the same sample space, sigma algebra, probability law and Brownian process. Identify the natural filtration and every elapsed-increment filtration. Equality in law alone does not identify admissible controls on a chosen space.

Controls and payoffs. Bundle an admissible scalar control as a support control and conversely. With $s = t + r$, prove equality between the absolute integral over $[t, 1]$ using $\alpha(s - t)$ and the elapsed integral over $[0, 1 - t]$ using $\alpha(r)$. Preserve progressive measurability and the pointwise bound $|\alpha| \leq 1$.

Value and minimizer. Equality of the payoff sets gives equality of their suprema for every x and $0 \leq t < 1$. Only then identify the first and second spatial derivatives and the Parisi penalties. Transport minimality against every admissible competitor, including restriction or extension of that competitor.

Diffusion and moments. Reindex the same process between nonnegative and real time, and stop it at one where required. Transport the integral equation on one full-measure event for all $t \in [0, 1]$, path continuity, initial value, gradient and moment integrals.

The diffusion comparison has a version issue. The support predicate asks for a globally progressive process, while the scalar predicate asks for adaptedness of the stopped process and almost-sure continuity on $[0, 1]$. An ordinary construction gives the needed progressive version. For $t \geq 0$ and $k = 0, 1, ...,$ put

$$X_t^{(k)} = X_{\min(2^{-k}\lfloor 2^k t \rfloor, 1)}.$$

Each left-step process is progressive: on every bounded time interval, its value on a step is measurable at that step's left endpoint. Define $\bar{X}_t$ as the pointwise limit where the sequence converges,

and zero elsewhere. The convergence set and this limit are progressively measurable. On the single event where X is continuous on $[0,1]$, the equality $\bar{X}_t = X_{\min(t,1)}$ holds for every $t \geq 0$. Thus stopping and reindexing preserve paths, time integrals and moments on that event, and constructing this version needs no completion of the natural filtration. The initial value follows from the scalar equation at zero and $B_0 = 0$ on their common full-measure event.

This settles the version construction at the ordinary mathematical level. Its implementation in the two formal predicates, and the coefficient, filtration and value identifications listed above, remain separate obligations. In particular, transporting the derivatives needs equality of values for every x, and agreement at $(0,0)$ is insufficient.

Once those obligations are proved, an application must use the same transported minimizer and associated diffusion in all premises. No such cross-model transport theorem is supplied here. The manuscript proof in §§3–7 and the scalar implementation's independent argument remain available without it. These distinctions come from the defining support model and scalar model themselves.

12.4 Logical limits of the argument

The full-support theorem and scalar-value theorem use related but different formal representations. Their mathematical conclusions can be discussed together, but a direct formal implication would require the coefficient, Brownian law, control value and diffusion translations described in §12, with common witnesses. The finite Gaussian result also needs its rooted energy identification before it says anything about a QAOA circuit. The boundedness, regularization, synthesis and transfer estimates in §8 are additional steps, with temperature-specific hypotheses where the positive-temperature route is used.

Infinite support allows gaps. An indefinite curvature identity gives strict convexity only once the expectation is controlled, and an algebraic certificate leaves its analytic hypotheses to be established. Coordinatewise convergence does not control a growing coefficient sum. A finite-dimensional Gaussian law identifies an energy observable only together with its operator argument. Separate witnesses cannot be merged silently. Finally, the order $\forall \varepsilon\ \exists$finite word $\lim_{n\to\infty}$ cannot be replaced by a fixed-instance depth limit.

The local comparison in §11.1 concerns two printed finite-step displays. It neither replaces the cited theorem's hypotheses nor supplies the remaining circuit estimates.

13. Source checks and finite calculation

The source collection retains the two manuscripts, formal statements, their implementations and build metadata. The derived project specifies Lean 4.34.1 and a pinned mathlib revision. The proof-assistant and library infrastructure is described by de Moura and Ullrich (2021), *The Lean 4 Theorem Prover and Programming Language* and The mathlib Community (2020), *The Lean Mathematical Library*. These references credit the software, and the retained project metadata identifies the versions used for this source.

The retained polynomial computation in §4 uses exact rational arithmetic. Its expansion matches all 22 positive remainder coefficients. The surplus of the v^4 coefficient above $1/1000$ is $140507/1521000000$. This finite check does not establish the PDE signs, integrability, boundary flux or limiting argument.

The source-byte audit verified the inspected edition and proved none of its claims. Paper assembly

checked identifiers, declared scope and document bindings. It did not check semantic correspondence or reader understanding. Tool-compatibility checks and synthetic fixtures exercised only their stated cases, which include neither selected SK theorem. Reading of the mathematical and formal sources supports the bounded comparisons reported here, without a complete independent proof acceptance or transitive axiom audit.

We keep the retained observations separate. The original graph records unrun plans and an earlier full-support attempt that could not start because the Comparator executable was unavailable. A later retained attempt, observed on 23 October 2026, built and exported the challenge, then stopped while building the implementing solution. The retained partial-observation receipt and log preserve that event: the imported object for `Diffusion/AbsIncrement` was missing. The challenge is an admitted specification, so its successful build and export are not a proof of the theorem. There was no successful solution export or research-target comparison in that observation.

The scalar-value invocation in the same original two-target run reached a terminal failure observed on 24 October 2026. Its challenge built and exported, but its solution build stopped because the imported object for `Gaussian/IntegralPositiveMulGaussian` was missing. The retained scalar receipt and log establish no solution export, kernel acceptance or comparison acceptance. The companion source review locates both records.

Both failures concern the original build configuration. A subsequent repair to the derived configuration was recorded. In the subsequent corrected replay, the full-support challenge built and its exporter was observed running. Reconciliation on 25 October found that the controller had stopped without a recovered terminal research result. At that date, Comparator and research extraction outcomes were unknown, neither rejected nor accepted. The interruption receipt retains this observation separately from the original build failures.

A subsequent source-bound observation of 26 October 2026 records a successful local Lake build of the Support implementation and v2 extraction of the single conditional full-support root. Its selected traversal contains 1,674 declarations. The root reports no sorry, unsafe or partial flag, and its recorded transitive axiom names are `propext`, `Classical.choice` and `Quot.sound`. The selected in-scope traversal completed, but dependencies outside it remain unresolved boundaries, so the extracted graph is partial. This does not establish minimiser existence, positive density, the scalar target or QAOA convergence.

The portable observation identifies the original receipts by their hashes and retains the source inventory, root and extraction limits. Its internal-consistency checks are not authenticated execution provenance. The curated paper graph records the build and documentary extraction observation separately, and it is distinct from the extracted declaration graph. We have not completed a Comparator run or a standalone kernel replay ourselves, and our scalar-target build did not finish. Two independent reruns report that Comparator accepts OpenAI's proofs of both formal statements: MathVet, on 8 October 2026, in a sandboxed run, and C. Toledo, on 10 October 2026, with a second kernel as well. Both reports note that the statements take a minimiser as hypothesis and do not mention QAOA. The earlier failed and interrupted invocations of our own remain dated history, rather than evidence of a later rejection or acceptance. Permitted axioms and checker settings describe the requested comparison and cannot supply its missing result.

This editorial revision does not establish a new formal result. The source review through 26 October 2026 records the preceding paper revision's reading coverage, finite checks, unsuccessful observations and remaining obligations, and viewing it requires repository access. This remains a draft after source-based agent review. Independent mathematical acceptance and a human reconstruction

study have not been performed.

References

- Auffinger and Chen (2015), "On properties of Parisi measures," *Probability Theory and Related Fields* **161**, 817–850. Inspected preprint: arXiv:1303.3573v1.
- Farhi, Goldstone and Gutmann (2014), *A Quantum Approximate Optimization Algorithm*, arXiv:1411.4028v1.
- Binkowski, Koßmann, Ziegler and Schwonnek (2023), *Elementary Proof of QAOA Convergence*, arXiv:2302.04968v1.
- Basso, Gamarnik, Mei and Zhou (2022), *Performance and limitations of the QAOA at constant levels on large sparse hypergraphs and spin glass models*, arXiv:2204.10306v2.
- Boulebnane, Khan, Liu, Larson, Herman, Shaydulin and Pistoia (2026), *Spin–Boson Mapping of the Quantum Approximate Optimization Algorithm*, arXiv:2505.07929v2.
- Boulebnane, Khan, Subrahmanya, Herman, Farhi, Villalonga and Shaydulin (2026), *Decoupling of the QAOA into independent spin-boson systems and high-depth performance on pure and mixed spin glasses*, arXiv 2609.40309, version 1.
- Goh (2025), *The Overlap Gap Property limits limit swapping in the QAOA*, preprint, version 6.
- OpenAI (2026a), *Full support of the zero-temperature Sherrington–Kirkpatrick order parameter*, manuscript, 27 September 2026.
- OpenAI (2026b), *QAOA attains the SK ground-state energy in the thermodynamic-first limit*, manuscript, 25 September 2026.
- Auffinger and Chen (2017), "Parisi formula for the ground state energy in the mixed p-spin model," *Annals of Probability* **45**, 4617–4631.
- Auffinger and Chen (2015), "The Parisi formula has a unique minimizer," *Communications in Mathematical Physics* **335**, 1429–1444.
- Auffinger, Chen and Zeng (2020), "The SK model is infinite step replica symmetry breaking at zero temperature," *Communications on Pure and Applied Mathematics* **73**, 921–943.
- Jagannath and Tobasco (2016), "A dynamic programming approach to the Parisi functional," *Proceedings of the American Mathematical Society* **144**, 3135–3150.
- Chen, Handschy and Lerman (2018), "On the energy landscape of the mixed even p-spin model," *Probability Theory and Related Fields* **171**, 53–95.
- Montanari (2025), "Optimization of the Sherrington–Kirkpatrick Hamiltonian," *SIAM Journal on Computing* **54**, FOCS19-1–FOCS19-38.
- El Alaoui, Montanari and Sellke (2021), "Optimization of mean-field spin glasses," *Annals of Probability* **49**, 2922–2960.
- Farhi, Goldstone, Gutmann and Zhou (2022), "The Quantum Approximate Optimization Algorithm and the Sherrington–Kirkpatrick Model at Infinite Size," *Quantum* **6**, 759.

- Basso, Farhi, Marwaha, Villalonga and Zhou (2022), "The Quantum Approximate Optimization Algorithm at High Depth for MaxCut on Large-Girth Regular Graphs and the Sherrington–Kirkpatrick Model," *LIPIcs* **232**, 7:1–7:21.
- El Alaoui, Montanari and Sellke (2023), "Local algorithms for Maximum Cut and Minimum Bisection on locally treelike regular graphs of large degree," *Random Structures and Algorithms* **63**, 689–715.
- Chen (2026), *FRSB in the SK spin glass: convergence to full-interval support at zero temperature*, arXiv:2607.18032v1.
- Lopatto (2026), *Full replica symmetry breaking in the Sherrington–Kirkpatrick model*, arXiv:2607.11756v3.
- Zhou (2026), "Existence of full replica symmetry breaking for the Sherrington–Kirkpatrick model at low temperature," *Communications on Pure and Applied Mathematics* **79**, 1746–1770.
- Sellke (2021), *Optimizing Mean Field Spin Glasses with External Field*, arXiv:2105.03506v4.
- Sellke (2024), *Optimizing Mean Field Spin Glasses with External Field*, *Electronic Journal of Probability* **29**, paper 4. Inspected revision: arXiv:2105.03506v6, 17 December 2023. Lemma 2.9 and Remark 2.10 are distinguished from the preceding v4 entry.
- Jekel, Sandhu and Shi (2026), *Potential Hessian Ascent: The Sherrington–Kirkpatrick Model*, arXiv:2408.02360v3.
- Chou, Love, Sandhu and Shi (2022), *Limitations of Local Quantum Algorithms on Random Max-k-XOR and Beyond*, arXiv:2108.06049v3.
- Lopatto (2026), *Full replica symmetry breaking in the Sherrington–Kirkpatrick model*, arXiv:2607.11756v4. This later comparison is distinct from the v3 input cited above.
- Lopatto (2026), *Parisi measure of the Sherrington–Kirkpatrick model with external field*, arXiv:2610.00127v1.
- Jones and Knill (1999), "Efficient Refocussing of One Spin and Two Spin Interactions for NMR Quantum Computation," *Journal of Magnetic Resonance* **141**, 322–325. Inspected preprint: arXiv:quant-ph/9905008v1.
- Li and Khaneja (2005), *Noncommuting vector fields, polynomial approximations and control of inhomogeneous quantum ensembles*, arXiv:quant-ph/0510012v1.
- Talagrand (2006), "The Parisi formula," *Annals of Mathematics* **163**, 221–263. Published article.
- Dembo, Montanari and Sen (2017), "Extremal Cuts of Sparse Random Graphs," *Annals of Probability* **45**, 1190–1217. Inspected preprint: arXiv:1503.03923v2.
- McKay, Wormald and Wysocka (2004), "Short Cycles in Random Regular Graphs," *Electronic Journal of Combinatorics* **11**, R66. Published article.
- de Moura and Ullrich (2021), "The Lean 4 Theorem Prover and Programming Language," *CADE 28*, 625–635. Published article.

- The mathlib Community (2020), "The Lean Mathematical Library," *CPP 2020*, 367–381. Published article.
- Zhou (2026), "On the Gardner Transition in the Ising Pure p-Spin Glass II". arXiv:2608.06523v1, 6 August 2026. Appendix E.3, (E.28), (E.37), (E.41), pp. 127–128, and p. 130.