

Entropy, Correlations, and Observable Reversal in a Dilute Hard-Sphere Gas

(Dated: Working manuscript, revised September 5, 2026)

How much microscopic information distinguishes a gas that retraces its past from a freshly prepared gas with the same one-particle state? For deterministic hard spheres in the Boltzmann–Grad limit, we connect the entropy increase, multiparticle correlations, preparation distinguishability, and an observable reversal probability. We work with the native grand-canonical preparation of Deng, Hani, and Ma on a fixed interval where a regular Boltzmann solution satisfies the specified bounds. Entropic chaos holds through a squared-logarithmic factorial order, while the mean within-sector total correlation per particle tends to the Boltzmann entropy increase. Under explicit positivity and moment assumptions, the log likelihood of the exactly reversed preparation relative to a fresh hard-core preparation concentrates at this same entropy density. This identifies the matching exponential probability of specified reversal tests, the optimal fixed-power discrimination exponent, and an exact relative-entropy budget for changing the test success probability. For continuous positive endpoint profiles, finite phase-space bins and global moment measurements retain any prescribed fraction arbitrarily close to the full exponent. The readout is fixed before the kinetic limit and extends to sufficiently accurate certified bounded readout errors. The results quantify a preparation-dependent arrow within reversible mechanics; they do not require a kinetic restart of the fresh ensemble, nor do they establish a full path large-deviation principle or a physical minimum-work protocol.

I. FROM ENTROPY ACCOUNTING TO A REVERSAL EXPERIMENT

An isolated hard-sphere gas evolves by an invertible, volume-preserving billiard flow. Its fine-grained entropy is constant. The Boltzmann equation describing its dilute limit instead has a nondecreasing one-particle entropy. The information identity relating these statements is elementary: the difference between a many-particle law and the product of its marginals measures correlations omitted by the one-particle description. The difficulty is to pass this identity through a singular kinetic limit, because convergence in total variation alone does not control entropy.

The derivation of Deng, Hani, and Ma (DHM) [1] supplies the hard-sphere convergence estimates on any prescribed regular solution interval satisfying their hypotheses. We use the whole-space, grand-canonical setting of version 3 of that work. The time interval is fixed before the sphere diameter tends to zero. No assertion of global existence for arbitrary Boltzmann data is needed or made.

The physical question pursued here goes beyond the entropy balance. Evolve a freshly prepared gas up to time T , then reverse every velocity while preserving its correlations. Compare this law, denoted A_ε , with an independently prepared hard-core gas B_ε whose limiting one-particle density is the same. Under A_ε , the microscopic dynamics retraces the previous history. How unlikely is it for B_ε to pass a specified observation of that reversal?

Let $a = h(f_T) - h(f_0) > 0$, with natural logarithms and $k_B = 1$. We prove that the full preparation likelihood concentrates at a per mean particle. A family of reversal tests therefore satisfies

$$A_\varepsilon(E_\varepsilon) \longrightarrow 1, \quad B_\varepsilon(E_\varepsilon) = \exp[-\bar{N}a + o(\bar{N})]. \quad (1)$$

The tests use two logarithmic endpoint observables and may include finitely many bounded observations at prescribed intermediate times. They are not an unrestricted event of entropy decreasing, and Eq. (1) is not a path

large-deviation principle.

An exact log likelihood is an ideal observable. We therefore also construct a finite-output readout with a physically interpretable tail control: finitely many phase-space bin counts, together with a global measurement of $1 + |x| + |v|^2$ at each endpoint. For every fixed $0 < \eta < a$, a readout chosen before $\varepsilon \rightarrow 0$ has asymptotic false-positive exponent between $a - \eta$ and a , while accepting A_ε with probability tending to one. The global moment is essential in this construction: it includes particles outside the histogram’s field of view.

There are two analytic routes in the argument. Entropic chaos at growing orders uses positive collision-history bounds and an explicit cutoff replay of the DHM source estimates. The reversal likelihood needs only the first two factorial estimates from DHM, supplemented by weighted endpoint control. The latter does not inherit the stronger collision-history entropy argument as an additional hypothesis. The subsequent probability and readout bounds are exact consequences of likelihood concentration and pointwise measurement certificates.

II. MODEL, PREPARATION, AND ANALYTIC ASSUMPTIONS

Consider hard spheres of diameter ε in $\mathbb{R}^d$, $d \geq 2$, and set $\mu = \varepsilon^{-(d-1)}$. In the sector of N labeled coordinates, $\mathcal{D}_{N,\varepsilon} = \{Z_N : |x_i - x_j| > \varepsilon, i \neq j\}$. The probability law is symmetric, and for a normalized density u ,

$$Q_{\mu,u}(N, dZ_N) = \frac{\mu^N}{N! \Xi_u} \mathbf{1}_{\mathcal{D}_{N,\varepsilon}} \prod_{i=1}^N u(z_i) dZ_N, \quad (2)$$

$$\Xi_u = \sum_{N \geq 0} \frac{\mu^N}{N!} \int_{\mathcal{D}_{N,\varepsilon}} \prod_i u(z_i) dZ_N. \quad (3)$$

The $1/N!$ convention is part of the ensemble definition. We keep the actual random count, its conserved law π_N ,

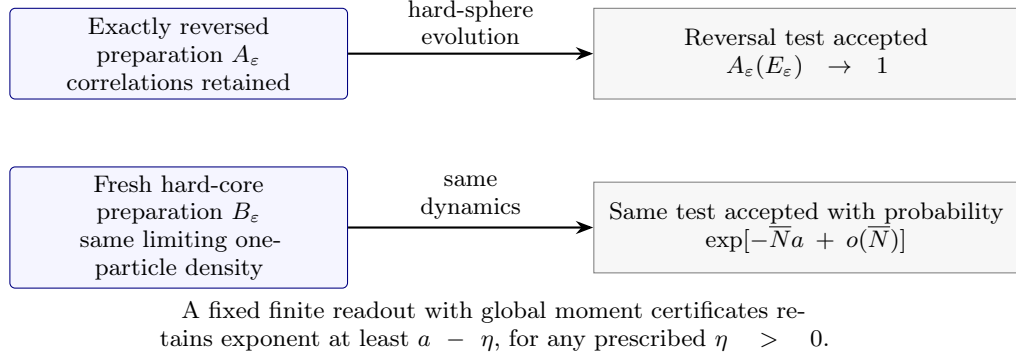


FIG. 1. The comparison is between preparations, with a common microscopic evolution and a common observation. The matching exponent shown at right concerns the shrinking-tolerance tests of Theorem 3. A single fixed finite readout has the controlled-loss bounds of Sec. V. This is a schematic of the theorem, not simulation data.

and its conserved mean $\bar{N}$. The factorial density $c_{\varepsilon,q}(t)$ is normalized by μ^q , whereas the probability density $\rho_q(t)$ is normalized by $\mathbb{E}(N)_q$; these normalizations must not be interchanged.

We adopt the following common sufficient class so that all statements can be read on one set of hypotheses. The entropy-only argument also allows weaker data, as recorded in its source proof. For some $\beta, c, a_0, b_0 > 0$, the normalized C^1 density $f_0 > 0$ satisfies

$$\sum_{m \in \mathbb{Z}^d} \sup_{x \in m + [0,1]^d} e^{2\beta|v|^2} (f_0 + |\nabla_x f_0|) < \infty, \quad (4)$$

$$f_0(x, v) \geq ce^{-a_0\sqrt{1+|x|^2}-b_0|v|^2}, \quad \int |x|^2 f_0 \, dx \, dv < \infty. \quad (5)$$

Let f_t be a regular solution of $\partial_t f + v \cdot \nabla_x f = Q(f, f)$ on a fixed interval $[0, T]$, with

$$\sup_{t \in [0, T], x, v} e^{2\beta|v|^2} f_t(x, v) < \infty. \quad (6)$$

For the finite-bin construction assume in addition that the endpoint profiles are continuous. This is satisfied by the smooth local examples discussed in Appendix B. The positivity and moment conditions give finite one-particle entropies; for the whole-space Shannon conversion one may take a spatial reference proportional to $\exp(-a_0\sqrt{1+|x|^2})$.

The solution bound (6) is a supremum in space. It is not, by itself, a global velocity-moment bound. The spatial summability in (4), its propagated version, and energy matching are used explicitly in Appendix B. All phase-space coordinates and logarithms refer to one fixed nondimensional reference measure.

Let Φ_t be the almost-everywhere defined hard-sphere flow and let $R(x, v) = (x, -v)$, acting particlewise. Then

$$P_t = (\Phi_t)_\# Q_{\mu, f_0}, \quad R\Phi_t R = \Phi_{-t}. \quad (7)$$

The kinetic input from DHM Theorem 1, Eq. (1.18), is

$$\sup_{0 \leq t \leq T} \|c_{\varepsilon,q}(t) - \mathbf{1}_{\mathcal{D}_{q,\varepsilon}} f_t^{\otimes q}\|_1 \leq \varepsilon^\theta, \quad q \leq |\log \varepsilon|, \quad (8)$$

for sufficiently small ε and a positive dimension-dependent θ , with the harmless adjustment of constants allowed in the source statement. Only $q = 1, 2$ are needed for the reversal theorem. No numerical value of the exponent is assigned here. The growing-order entropy estimate uses the additional source construction specified in Appendix A, rather than extending the displayed range of (8) by assumption.

III. ENTROPY PRODUCTION AND COLLECTIVE INFORMATION

The information stored by the full gas must be separated from information created by mixing different particle numbers. This distinction is essential in a grand-canonical kinetic limit. We use the preparation of Eq. (2) throughout, without replacing it by a prescribed canonical sequence. Define

$$Z_N^f = \int \mathbf{1}_{\mathcal{D}_{N,\varepsilon}} f_0^{\otimes N} \, dZ_N, \quad \pi_N = \frac{\mu^N Z_N^f}{N! \Xi_f}, \quad (9)$$

$$P_N^t = (\Phi_t)_\# \left(\frac{\mathbf{1}_{\mathcal{D}_{N,\varepsilon}} f_0^{\otimes N}}{Z_N^f} \right), \quad u_N = P_{N:1}^t.$$

Here $\Xi_f = \Xi_{f_0}$ is the partition function in Eq. (2). Only sectors of positive weight are used. The normalized conditional density P_N^t is with respect to ordered Lebesgue measure dZ_N ; the factor $1/N!$ appears in the count law. Let

$$F_q = \mathbb{E}_\pi(N)_q, \quad \bar{N} = F_1, \quad w_N^{(q)} = \frac{\pi_N(N)_q}{F_q}, \quad (10)$$

$$\rho_q = \sum_{N \geq q} w_N^{(q)} P_{N:q}^t, \quad \sigma_q = \sum_{N \geq q} w_N^{(q)} u_N.$$

Here $(N)_q = N(N-1) \cdots (N-q+1)$ and $F_0 = 1$. The true one-coordinate marginal of ρ_q is σ_q . In general $\sigma_q \neq \rho_1$: separately normalized factorial densities do not form a consistent marginal hierarchy.

We distinguish three information quantities:

$$K_\varepsilon(t) = \sum_{N \geq 1} \pi_N D(P_N^t \| u_N^{\otimes N}), \quad (11)$$

$$C_\varepsilon(t) = \sum_{N \geq 1} w_N^{(1)} D(u_N \| \rho_1), \quad (12)$$

$$\mathcal{I}_{\text{gc}}^\varepsilon(t) = D(P_t \| \Pi_{\bar{N}, \rho_1}). \quad (13)$$

The first is the mean total correlation *within* a sector. The second is the mutual information between particle number and one tagged particle, when sectors are sampled with weight $w^{(1)}$. The comparison law

$$\Pi_{n,u}(N, dZ_N) = e^{-n} \frac{n^N}{N!} u^{\otimes N}(Z_N) dZ_N \quad (14)$$

has no hard-core constraint. It is used to measure information, rather than as a hard-sphere preparation. The exact decomposition is

$$\mathcal{I}_{\text{gc}}^\varepsilon = D(\pi \| \text{Pois}_{\bar{N}}) + K_\varepsilon + \bar{N} C_\varepsilon. \quad (15)$$

Consequently, a result for $\mathcal{I}_{\text{gc}}^\varepsilon$ alone cannot identify its entire extensive contribution with within-sector correlation.

Theorem 1 (Entropy and conditional information). *Assume the whole-space initial-data and fixed-time conditions of Sec. II. Write $\lambda = |\log \varepsilon|$ and $s = \lceil \lambda^2 \rceil$. The DHM source estimates and the cutoff replay specified in Appendix A give, at every fixed $t \in [0, T]$, probability references $m_\varepsilon(x, v) = w_\varepsilon(x) M_{\beta_m}(v)$ and constants $b > 0$, $p < \infty$, and $C_t < \infty$ such that*

$$\sup_{1 \leq q \leq s} |D(\rho_q \| m_\varepsilon^{\otimes q}) - qD(f_t \| m_\varepsilon)| \leq C_t \varepsilon^b \lambda^p, \quad (16)$$

$$\sup_{1 \leq q \leq s} D(\rho_q \| \rho_1^{\otimes q}) \leq C_t \varepsilon^b \lambda^p. \quad (17)$$

These are absolute block errors. There is a nonnegative upper bound

$$B_{\text{info}}(\varepsilon) = \frac{d-1}{\lambda} + O_t(\lambda^{-2}) + O_t(\varepsilon^{b'} \lambda^{p'}), \quad b' > 0, \quad (18)$$

for which $C_\varepsilon(t) \leq B_{\text{info}}(\varepsilon)$ and, for $1 \leq q \leq s$,

$$\begin{aligned} & \sum_{N \geq q} w_N^{(1)} D(P_{N;q}^t \| u_N^{\otimes q}) \\ & \leq \sum_{N \geq q} w_N^{(1)} D(P_{N;q}^t \| \rho_1^{\otimes q}) \leq q B_{\text{info}}(\varepsilon). \end{aligned} \quad (19)$$

The sums retain their physical weights and are not renormalized after omitting $N < q$. In particular, these two averaged conditional KL divergences vanish whenever $q = o(|\log \varepsilon|)$.

The Shannon entropies in the present initial-data class are finite. For each fixed q ,

$$h(\rho_q(t)) \longrightarrow qh(f_t), \quad h(u) = - \int u \log u. \quad (20)$$

Moreover,

$$\boxed{\frac{K_\varepsilon(t)}{\bar{N}}, \quad \frac{\mathcal{I}_{\text{gc}}^\varepsilon(t)}{\bar{N}} \longrightarrow \Delta h(t) := h(f_t) - h(f_0)}. \quad (21)$$

Both convergences also hold in $L^1(0, T)$. In distributions on $(0, T)$,

$$\partial_t \frac{K_\varepsilon}{\bar{N}}, \quad \partial_t \frac{\mathcal{I}_{\text{gc}}^\varepsilon}{\bar{N}} \longrightarrow \mathcal{D}_B[f_t], \quad (22)$$

where the collision normalization is $\mu \varepsilon^{d-1} = 1$ and

$$\begin{aligned} \mathcal{D}_B[f] &= \frac{1}{4} \int B(v - v_*, \omega) (f' f'_* - f f_*) \\ &\quad \times \log \frac{f' f'_*}{f f_*} dx dv dv_* d\omega \geq 0. \end{aligned} \quad (23)$$

The references m_ε are technical probability envelopes with uniform spatial log-Lipschitz constants and a fixed Gaussian reserve. They are not equilibrium states and are not used to define the physical entropy. The squared-logarithmic range in Eq. (17) is stronger than the range for the conditional averages in Eq. (19); the two assertions concern different probability laws. Neither range identifies a sharp correlation threshold. No growing-order Shannon rate is inferred from the reference-KL rate.

The limiting identity has a simple finite-system origin. Volume preservation and particle-number conservation imply

$$\mathcal{I}_{\text{gc}}^\varepsilon(t) - \mathcal{I}_{\text{gc}}^\varepsilon(0) = \bar{N} [h(\rho_1(t)) - h(\rho_1(0))]. \quad (24)$$

Initial insertion estimates make both the initial matched-Poisson information and the count KL subextensive. The growing-block argument then proves $C_\varepsilon \rightarrow 0$, which, through Eq. (15), attributes the limiting entropy increase to K_ε itself. This makes precise the correlation interpretation of kinetic irreversibility discussed by Spohn [2]; it does not change the reversibility of the underlying flow or explain the choice of the initial low-information preparation.

There is a second way to express the collective nature of this information. Put

$$\begin{aligned} I_{N,r} &= D(P_{N;r}^t \| u_N^{\otimes r}), & I_N &= I_{N,N}, \\ J_N &= D(P_N^t \| u_N \otimes P_{N-1}^t), & \mathcal{J}_\varepsilon &= \sum_{N \geq 1} w_N^{(1)} J_N. \end{aligned} \quad (25)$$

Thus $J_N = I(Z_1; Z_{2:N} | N)$, evaluated in a fixed sector; we set $J_1 = 0$.

Corollary 1 (A particle, a partner, and the remaining gas). *For each fixed ε , $\mathcal{J}_\varepsilon(t)$ is finite and*

$$\mathcal{J}_\varepsilon(t) \geq \frac{K_\varepsilon(t)}{\bar{N}}, \quad \liminf_{\varepsilon \rightarrow 0} \mathcal{J}_\varepsilon(t) \geq \Delta h(t). \quad (26)$$

At the same time,

$$\sum_{N \geq 2} w_N^{(1)} I(Z_1; Z_2 | N) \leq 2B_{\text{info}}(\varepsilon) \longrightarrow 0. \quad (27)$$

The sampling rule in both statements first chooses a sector with particle-biased weight $w^{(1)}$ and then tags particles in that sector. If $\Delta h(t) > 0$, a tagged particle has vanishing mean information about a second tagged particle, while its information about the joint state of the rest has a positive asymptotic lower bound. This does not give an equality or a uniform upper bound for $\mathcal{J}_\varepsilon$, a statement in every predetermined canonical sector, or a spatial location for the information. An invertible velocity reflection preserves all these within-sector KL divergences, so the conclusions also hold for the exactly reversed preparation. This observation is consistent with the distinction between a reversed ensemble and a freshly prepared chaotic ensemble [3].

IV. REVERSED PREPARATIONS AND THEIR STATISTICAL COST

We now ask how the correlations in an evolved gas affect an operational test of its preparation. Let $R(x, v) = (x, -v)$ act on densities and configurations, and retain the positive initial-data class and fixed solution interval of Sec. II. Set

$$\begin{aligned} P_t^\varepsilon &= (\Phi_t)_\# Q_{\mu, f_0}, & A_\varepsilon &= R_\# P_T^\varepsilon, \\ g &= R f_T, & k &= R f_0, & B_\varepsilon &= Q_{\mu, g}. \end{aligned} \quad (28)$$

The notation P_t^ε makes explicit the diameter dependence of the microscopic law P_t . Here A_ε is the exact velocity reversal of the evolved preparation, whereas B_ε is a fresh hard-core preparation with input density g . Their limiting one-particle densities agree at the preparation time. Their actual finite- ε intensities and particle-count laws need not agree. The conserved mean count under A_ε is $\bar{N}$, and

$$a := \Delta h = h(f_T) - h(f_0) > 0. \quad (29)$$

The future is observed using the same flow Φ_t under both laws. Reversibility gives $(\Phi_t)_\# A_\varepsilon = R_\# P_{T-t}^\varepsilon$ for $0 \leq t \leq T$.

Theorem 2 (Endpoint likelihood concentration). *The two preparations are equivalent. Their log likelihood ratio is exactly*

$$\begin{aligned} L_\varepsilon(Z) &:= \log \frac{dA_\varepsilon}{dB_\varepsilon}(Z) = \log \Xi_g - \log \Xi_{f_0} \\ &+ \sum_{i=1}^N \log k(z_i(T)) - \sum_{i=1}^N \log g(z_i(0)). \end{aligned} \quad (30)$$

Moreover,

$$\frac{L_\varepsilon}{\mu} \longrightarrow a \quad \text{in } L^1(A_\varepsilon), \quad \frac{D(A_\varepsilon \| B_\varepsilon)}{\bar{N}} \longrightarrow a. \quad (31)$$

The identity follows by a sectorwise change of variables. The factors $\mu^N/N!$ cancel, so it already includes the fluctuating particle number; an additional count correction

would be incorrect. The partition term divided by μ is $O(\varepsilon)$. The substantive input is concentration of the two unbounded logarithmic observations. Appendix B proves

$$|\log f_t(x, v)| \leq C_T(1 + |x| + |v|^2) \quad (32)$$

and a weighted empirical L^1 law of large numbers. Its ingredients are the $q = 1, 2$ conclusions of DHM, spatial localization through DHM Proposition A.1, matching of microscopic and Boltzmann energy, and the initial spatial second moment. A velocity Gaussian bound uniform in x alone is insufficient for the whole-space integrals. No growing particle-order estimate or kinetic restart from g is used here.

Define the empirical observable

$$M_\varphi(t, Z) = \mu^{-1} \sum_{i=1}^N \varphi(z_i(t)). \quad (33)$$

Choose any fixed finite list of bounded measurable one-particle tests φ_j at times $0 \leq t_j \leq T$, and put $b_j = \int \varphi_j R f_{T-t_j}$. For fixed $\delta > 0$, let $E_{\varepsilon, \delta}$ impose

$$\begin{aligned} |M_{\log g}(0) + h(g)| &\leq \delta, \\ |M_{\log k}(T) + h(k)| &\leq \delta, \\ |M_{\varphi_j}(t_j) - b_j| &\leq \delta \quad \text{for every } j. \end{aligned} \quad (34)$$

The bounded tests may sample a reversed macroscopic history at finitely many chosen times. The logarithmic endpoint tests are part of the event, rather than conclusions inferred from its bounded tests.

Theorem 3 (Probability of the specified reversal test). *For every fixed $\delta > 0$, $A_\varepsilon(E_{\varepsilon, \delta}) \rightarrow 1$, and*

$$\begin{aligned} a - 2\delta &\leq \liminf_{\varepsilon \rightarrow 0} -\mu^{-1} \log B_\varepsilon(E_{\varepsilon, \delta}) \\ &\leq \limsup_{\varepsilon \rightarrow 0} -\mu^{-1} \log B_\varepsilon(E_{\varepsilon, \delta}) \leq a + 2\delta. \end{aligned} \quad (35)$$

The lower and upper limiting exponents therefore both tend to a when $\delta \downarrow 0$ after $\varepsilon \rightarrow 0$. Equivalently, there is a deterministic sufficiently slow choice $\delta_\varepsilon \downarrow 0$ such that $E_\varepsilon = E_{\varepsilon, \delta_\varepsilon}$ satisfies

$$A_\varepsilon(E_\varepsilon) \rightarrow 1, \quad B_\varepsilon(E_\varepsilon) = \exp[-\bar{N}a + o(\bar{N})]. \quad (36)$$

The proof uses the exact change of measure

$$B_\varepsilon(E) = \mathbb{E}_{A_\varepsilon}[e^{-L_\varepsilon} \mathbf{1}_E]. \quad (37)$$

On Eq. (34), L_ε/μ lies within $2\delta + O(\varepsilon)$ of a . Since the event is A_ε -typical, bounding the integrand above and below proves Eq. (35). This gives a matching exponent for the specified finite-observation event. It does not give a full path large-deviation principle, an exponent for every entropy-decreasing event, an inner limit at every fixed tolerance, or a prescribed rate for shrinking δ_ε . Finite precision of the observation requires further control because the endpoint logarithms are unbounded.

Corollary 2 (Optimal one-sided discrimination). *For fixed $0 < \alpha < 1$, define*

$$\beta_\varepsilon(\alpha) = \inf_{\substack{0 \leq \varphi \leq 1 \\ \mathbb{E}_{A_\varepsilon} \varphi \geq 1-\alpha}} \mathbb{E}_{B_\varepsilon} \varphi. \quad (38)$$

Then $-\mu^{-1} \log \beta_\varepsilon(\alpha) \rightarrow a$. The events in Eq. (36) attain this exponent with acceptance tending to one.

Thus the entropy increase determines the best false-positive exponent for distinguishing these preparations at fixed allowed false-rejection probability. This is an information-spectrum conclusion with no independent-particle or independent-time sampling assumption. It is not a symmetric error exponent or a result for every prescribed vanishing false-rejection tolerance.

Corollary 3 (Statistical cost of successful preparation). *For the events in Eq. (36), set $q_\varepsilon = B_\varepsilon(E_\varepsilon)$. For any desired $p \in [0, 1]$,*

$$\inf_{C: C(E_\varepsilon)=p} D(C \| B_\varepsilon) = d(p \| q_\varepsilon), \quad (39)$$

where d denotes Bernoulli relative entropy. The minimum is attained by

$$C_{\varepsilon,p}^* = pB_\varepsilon(\cdot | E_\varepsilon) + (1-p)B_\varepsilon(\cdot | E_\varepsilon^c), \quad (40)$$

and $\mu^{-1} d(p \| q_\varepsilon) \rightarrow pa$. If instead arbitrary competing preparations obey

$$\limsup_{\varepsilon \rightarrow 0} \mu^{-1} D(C_\varepsilon \| B_\varepsilon) \leq r, \quad (41)$$

then $\limsup C_\varepsilon(E_\varepsilon) \leq \min\{1, r/a\}$.

The laws C_ε need not be product preparations or satisfy DHM initial-data hypotheses: binary data processing suffices once this event's probability is known. The cost is a statistical variational quantity relative to B_ε , not a minimum physical work or an implemented control protocol. Sampling the conditional mixture may itself be difficult.

A subextensive budget forces $C_\varepsilon(E_\varepsilon) \rightarrow 0$, but does not preserve the exponent a . Taking $p_\varepsilon = 1/\mu$ in Eq. (40), for example, gives

$$D(C_{\varepsilon,p_\varepsilon}^* \| B_\varepsilon) \rightarrow a, \quad C_{\varepsilon,p_\varepsilon}^*(E_\varepsilon) = \mu^{-1}. \quad (42)$$

Even this bounded cost changes an exponentially small probability into a polynomially small one. It still cannot make success have a nonzero limiting probability.

The positive-production class is nonempty. One example is $f_0(x, v) = Z^{-1} e^{-c\sqrt{1+|x|^2}} e^{-v^\top \Sigma^{-1} v/2}$, with $c > 0$ and positive definite Σ having two distinct eigenvalues. A positive local construction preserves the required Gaussian envelope and spatial summability, and anisotropy gives strictly positive initial collision entropy production. Appendix B6 establishes a nonzero short-time interval. For a larger fixed T , existence and the stated DHM solution bound on $[0, T]$ remain explicit hypotheses.

V. FINITE-RESOLUTION READOUT OF THE REVERSAL EXPONENT

The logarithmic observations in Theorem 3 use an unbounded weight on every particle. We now replace them by finitely many phase-space bin counts and a global moment measurement, all reported at fixed finite precision. The additional global measurement controls the contribution of particles outside the histogram box. This is essential to the construction: an overflow count alone does not bound the energy carried by those particles.

Retain the hypotheses and notation of Theorem 2, with $a = \Delta h > 0$, and additionally assume that the positive endpoint profiles g and k are continuous. Write

$$M_\psi(t) = \frac{1}{\mu} \sum_{i=1}^N \psi(z_i(t)), \quad W(x, v) = 1 + |x| + |v|^2. \quad (43)$$

Coordinates are expressed in the fixed reference units of the model. At time zero the limiting reversed density is g , and at time T it is k . The device observes M_{1_C} for finitely many bounded rectangular cells C , together with M_W , at these two times. Every channel has a fixed range, a fixed quantization step, and an overflow symbol. The moment includes all particles, including those outside the histogram box. It can be supplied by count, radial-position, and total-energy channels; the conserved energy channel can be shared between the endpoints.

Theorem 4 (A fixed finite-alphabet readout). *For every prescribed loss $0 < \eta < a$, there exist fixed cells, readout ranges and positive quantization steps, a finite alphabet $\mathcal{Y}_\eta$, and an acceptance set $\mathcal{F}_\eta \subseteq \mathcal{Y}_\eta$, such that the measurement $Y_{\varepsilon,\eta}$ described above satisfies, with $F_{\varepsilon,\eta} = \{Y_{\varepsilon,\eta} \in \mathcal{F}_\eta\}$,*

$$A_\varepsilon(F_{\varepsilon,\eta}) \rightarrow 1, \quad (44)$$

$$a - \eta \leq \liminf_{\varepsilon \rightarrow 0} \frac{-\log B_\varepsilon(F_{\varepsilon,\eta})}{\mu}, \quad (45)$$

$$\limsup_{\varepsilon \rightarrow 0} \frac{-\log B_\varepsilon(F_{\varepsilon,\eta})}{\mu} \leq a. \quad (46)$$

The alphabet, acceptance set, cells and precisions are chosen before taking $\varepsilon \rightarrow 0$ and are independent of ε .

The map $Y_{\varepsilon,\eta}$ still uses the known activity μ to normalize collective readings. Thus the precision is fixed for normalized counts and moments, not for each particle coordinate. For a different desired loss η , a different finite device may be needed. A single fixed device is not asserted to attain the exact exponent a , or even to have an existing limiting exponent. Because $\bar{N}/\mu \rightarrow 1$, all displayed exponent and information bounds remain valid with μ replaced by $\bar{N}$.

The construction provides a record-by-record certificate. For one endpoint profile $u \in \{g, k\}$, partition a compact box K into cells C_j , and choose

$$w_j^- = \inf_{C_j} W, \quad \ell_j = \log u(\zeta_j), \quad \zeta_j \in C_j. \quad (47)$$

For floor-quantized readings $q_j = s \lfloor M_{1_{C_j}}/s \rfloor$ and $q_W = s \lfloor M_W/s \rfloor$, reject overflow and compute

$$r_u = q_W + s - \sum_j w_j^- q_j, \quad \hat{m}_u = \sum_j \ell_j q_j. \quad (48)$$

The observable condition $r_u \leq \tau_u$ bounds the entire unobserved weighted tail:

$$M_{W \mathbf{1}_{K^c}} \leq r_u \leq \tau_u. \quad (49)$$

If $|\log u| \leq C_u W$, the grid oscillation of $\log u$ is at most ω_u , and the readout range is H , then on this certificate

$$|M_{\log u} - \hat{m}_u| \leq \omega_u(H + s) + C_u \tau_u + s \sum_j |\ell_j|. \quad (50)$$

Integrability of Wu and compact continuity allow all three errors to be made small, while the certificate remains typical under A_ε . Comparing the two endpoint scores then proves Theorem 4; Appendix C gives the full parameter selection and probability bounds. No dynamical exponential tail estimate under B_ε is used.

Corollary 4 (Information retained by the readout). *Let A_ε^Y and B_ε^Y be the output laws of the device in Theorem 4. Then*

$$a - \eta \leq \liminf_{\varepsilon \rightarrow 0} \frac{D(A_\varepsilon^Y \| B_\varepsilon^Y)}{\mu}, \quad (51)$$

$$\limsup_{\varepsilon \rightarrow 0} \frac{D(A_\varepsilon^Y \| B_\varepsilon^Y)}{\mu} \leq a. \quad (52)$$

For each fixed $0 < \alpha < 1$, define

$$\beta_Y(\alpha) = \inf_{\substack{0 \leq \varphi \leq 1 \\ \mathbb{E}_{A_\varepsilon} \varphi(Y) \geq 1 - \alpha}} \mathbb{E}_{B_\varepsilon} \varphi(Y). \quad (53)$$

The exponent $-\mu^{-1} \log \beta_Y(\alpha)$ obeys the same limit-inferior and limit-superior bounds as Eqs. (45)–(46). The conclusion also holds if the infimum is restricted to deterministic acceptance sets.

Consequently, any prescribed fraction $1 - \gamma$ of the full information density, with $0 < \gamma < 1$, can be retained by choosing $\eta = \gamma a$. A finite output alphabet has bounded Shannon entropy, but its relative entropy against another output law can grow linearly in μ : that law assigns exponentially small probability to the accepted output.

Corollary 5 (Calibrated bounded readout errors). *The same construction applies to finite output codes that report guaranteed intervals*

$$M_{1_{C_j}} \in [l_j, u_j], \quad M_W \in [l_W, u_W], \quad u_W \leq H_*. \quad (54)$$

Require nonsaturation probability tending to one under A_ε and sufficiently small interval widths. Set $c_j = (l_j + u_j)/2$, $r_j = (u_j - l_j)/2$, and accept only if

$$r_u = u_W - \sum_j w_j^- l_j \leq \tau_u. \quad (55)$$

On every accepted record,

$$\left| M_{\log u} - \sum_j \ell_j c_j \right| \leq \omega_u H_* + C_u \tau_u + \sum_j |\ell_j| r_j. \quad (56)$$

Theorems and bounds above hold with these endpoint error budgets. Randomized readout is included if both preparations use the same conditional readout kernel and the interval guarantees hold almost surely on every nonsaturated joint state-readout record.

The intervals describe calibrated bounded errors. A fixed nonzero analog error budget cannot be removed by refining digital quantization alone. Unflagged interval failures with probability merely tending to zero are not covered: their false positives can dominate an exponentially rare acceptance event. Nor does a finite output alphabet itself provide the global-moment error guarantee.

Proposition 1 (Sufficiency and one-bit information). *The exact endpoint score*

$$S = \sum_i \log k(z_i(T)) - \sum_i \log g(z_i(0)) \quad (57)$$

is sufficient for the binary experiment $(A_\varepsilon, B_\varepsilon)$:

$$D(S_\# A_\varepsilon \| S_\# B_\varepsilon) = D(A_\varepsilon \| B_\varepsilon). \quad (58)$$

For the shrinking-band events E_ε of Theorem 3, the single bit $\mathbf{1}_{E_\varepsilon}$ retains the full leading information density:

$$\frac{d(A_\varepsilon(E_\varepsilon) \| B_\varepsilon(E_\varepsilon))}{\mu} \rightarrow a, \quad (59)$$

where d denotes binary relative entropy.

Sufficiency concerns discrimination between these two preparations; it does not reconstruct the microscopic state. Computing S and then quantizing it is also different from the histogram measurement of Theorem 4, which supplies an observable error certificate before any logarithmic score is accepted.

Corollary 6 (Conditioning on the narrow reversal band). *For the shrinking-band events E_ε of Theorem 3, let $A_{\varepsilon,E} = A_\varepsilon(\cdot | E_\varepsilon)$ and $B_{\varepsilon,E} = B_\varepsilon(\cdot | E_\varepsilon)$. Then*

$$\frac{D(A_{\varepsilon,E} \| B_{\varepsilon,E})}{\mu} \rightarrow 0, \quad \frac{D(B_{\varepsilon,E} \| A_{\varepsilon,E})}{\mu} \rightarrow 0. \quad (60)$$

More precisely, there is a deterministic $\omega_\varepsilon \rightarrow 0$ such that for every measurable event G ,

$$e^{-\mu \omega_\varepsilon} B_{\varepsilon,E}(G) \leq A_{\varepsilon,E}(G) \leq e^{\mu \omega_\varepsilon} B_{\varepsilon,E}(G). \quad (61)$$

These conclusions concern the leading exponential scale. They do not imply convergence in total variation or equality of the conditioned microscopic ensembles. They use the two-sided narrow likelihood band of E_ε and are not asserted for the one-sided finite-device acceptance event $F_{\varepsilon,\eta}$.

VI. PHYSICAL INTERPRETATION AND RELATION TO EARLIER WORK

The preparation comparison answers a concrete question about the Loschmidt reversal. A one-particle density does not specify the full microscopic preparation. Preserving the correlations of an evolved gas and freshly sampling its one-particle profile yield different laws, despite the same limiting marginal. The exactly reversed law has the specified reverse macroscopic history by reversibility of the already controlled forward evolution. No additional Boltzmann derivation starting from $g = Rf_T$ is used. A statement about the *typical forward kinetic evolution* of the fresh law would separately require that restart to satisfy the relevant hypotheses.

This mechanism belongs to the classical discussion of reversibility, initial data, and correlations. Spohn's treatments [2, 4] and the one-sided convergence analysis of Bodineau, Gallagher, Saint-Raymond, and Simonella [3] are relevant antecedents. We do not claim the information identity, the qualitative correlation mechanism, or the distinction between reversed and fresh preparations as discoveries. The quantitative claims evaluated here are the entropy convergence, the precise grand-canonical attribution, the likelihood concentration, and the specified readout exponents under the stated dynamical hypotheses.

The relation between dissipation and preparation relative entropy also has established formulations, including the driven canonical Hamiltonian setting of Kawai, Parrondo, and Van den Broeck [5] and coarse-grained bounds in driven Brownian systems [6]. Here the reference is a fresh hard-core law at a nonequilibrium kinetic endpoint, and no work source or thermal protocol is specified. Accordingly the relative-entropy preparation budget is a statistical optimization, not a theorem about minimum work or dissipated heat. Likewise, the likelihood method gives an exponent for a restricted family of tests without requiring the full fluctuation or path large-deviation results studied for hard-sphere gases [7, 8].

The coexistence of small fixed-order mutual information and an extensive total correlation does not preclude useful collective measurements. A histogram of all particles is not the observation of a fixed number of tagged particles. Moreover, the readout compares two times. One binary decision can have relative entropy of order $\bar{N}$ when its false-positive probability is exponentially small, even though the Shannon entropy of that decision is at most $\log 2$. These facts reconcile the order-escape result with the finite-output discrimination theorem.

A. Connection to Gibbs mixing: a prospective extension

The conservation of fine-grained Gibbs entropy addressed here must be distinguished from the Gibbs mixing paradox, which concerns particle identity, distinguishabil-

ity, and thermodynamic operations [9–11]. Our model has one component and a prescribed $1/N!$ convention. The theorem neither derives that convention nor compares the thermodynamic work of mixing identical and different substances.

One consistency check is immediate. Let q erase the purely mathematical ordering of particles while retaining their unordered configuration and count. For the two symmetric preparations, the conditional distribution over the orderings is the same. The chain rule therefore gives the exact identity

$$D(A_\varepsilon \| B_\varepsilon) = D(q_\# A_\varepsilon \| q_\# B_\varepsilon). \quad (62)$$

Thus the preparation information is not an artifact of counting particle permutations.

A possible extension introduces finitely many passive, physically readable labels C , with constant fractions x_c , normalized phase-space densities f_c , and finite displayed entropies. Put $F(c, z) = x_c f_c(z)$ and $f = \sum_c x_c f_c$. Direct expansion yields

$$I(C; Z) = \sum_c x_c D(f_c \| f), \quad (63)$$

$$h(C, Z) = h(Z) + H(C) - I(C; Z), \quad (64)$$

$$\Delta h(C, Z) = \Delta h(Z) + I_0(C; Z) - I_t(C; Z). \quad (65)$$

These are information identities. For two equally populated, initially separated labels, complete mixing would reduce $I(C; Z)$ from $\log 2$ to zero. If the total profile were unchanged, this would give an additional $\log 2$ of labeled one-particle entropy per particle. This endpoint calculation is not a proof that such complete mixing occurs in the present whole-space model or within its fixed time interval.

The missing dynamical input is a marked hard-sphere limit tracking initial labels jointly with later particle states. Unmarked single-time convergence does not automatically provide it. A work-extraction statement would additionally need an explicit class of label-sensitive controls or separation operations. Equation (65) is included as a direction for further study, not as a new mixing theorem.

B. Limits of the present arrow

The finite readout theorem assumes a moment channel covering all particles. An overflow count alone cannot replace it: one particle with energy of order μ can change a normalized logarithmic score by order one. Unflagged sensor errors of probability merely $o(1)$ can likewise dominate an intended $\exp(-\bar{N}a)$ false-positive probability. Such errors need a pointwise certificate or appropriately strong probability control under the fresh preparation.

The remaining quantitative problems include usable finite-particle acceptance bounds, calibration of the endpoint profiles, and the resources needed to implement the collective measurements. The asymptotic readout theorem fixes normalized precision; it does not fix the total

cost of measuring a growing number of particles. Other extensions include a canonical entropy limit at prescribed particle number, the scale at which the escaping correlations become visible, collision-flux trace convergence, and a full nonequilibrium path large-deviation principle. Finally, the low-entropy preparation is an input. The results quantify its dynamical consequences and the cost of changing preparation statistics; they do not explain the origin of a low-entropy past.

VII. REPRODUCIBILITY AND ANALYTIC DEPENDENCIES

The appendices give the analytic derivations and distinguish the DHM source estimates from the probability arguments added here. The accompanying source package retains the longer entropy, endpoint, and sensor proofs, with immutable input manifests and a theorem-to-source map. Finite-state verification scripts check information identities, likelihood algebra, and measurement certificates. They do not verify the continuum hard-sphere limit, replace the DHM argument, or establish literature priority. No numerical simulation or experimental data are used to prove the results of this manuscript.

Appendix A: Proof of the entropy and information results

This appendix separates the collision-history input from the entropy argument. The input is the proof-level positive representation and estimates of DHM, with an explicit change of cutoffs. The published L^1 theorem alone does not give the entropy conclusions, and its stated logarithmic observation range is not silently evaluated at a larger order. All DHM equation and page numbers below refer to version 3 of Ref. [1]. The terminal time and the number of time layers remain fixed as $\varepsilon \rightarrow 0$.

1. Static insertion and the physical count law

Let $R_0 = \|\int f_0(x, v) dv\|_\infty$ and $c = v_d R_0$. If X is a configuration sampled from the initial grand-canonical law, let $\mathcal{A}_q(X)$ be the probability that q independent points sampled from f_0 avoid both X and one another. Expanding the nonnegative grand partition series gives the exact insertion identity

$$F_q = \mu^q \mathbb{E} \mathcal{A}_q(X). \quad (\text{A1})$$

The case $q = 1$ implies $\bar{N} \leq \mu$. A union bound for the inserted-old and inserted-inserted pairs yields

$$\mu^q (1 - a_q) \leq F_q \leq \mu^q, \quad a_q = c \varepsilon^d \left(q\mu + \frac{q(q-1)}{2} \right). \quad (\text{A2})$$

The lower bound is useful when $a_q < 1$, and the upper bound holds at every order. In particular, $\bar{N}/\mu = 1 + O(\varepsilon)$. Writing $\Xi(z)$ for the partition function at activity z , the same insertion argument gives $1 - c z \varepsilon^d \leq \partial_z \log \Xi(z) \leq 1$. Integration gives

$$0 \leq \mu - \log \Xi_f \leq \frac{c}{2} \mu^2 \varepsilon^d = \frac{c}{2} \mu \varepsilon. \quad (\text{A3})$$

Sequential insertion also implies, whenever $cN\varepsilon^d \leq 1/2$,

$$Z_N^f \geq \prod_{j=0}^{N-1} (1 - c j \varepsilon^d), \quad -\log Z_N^f \leq cN(N-1)\varepsilon^d. \quad (\text{A4})$$

We will need quantitative comparison of the different count weights. The exact overlap polynomial is

$$N\{(N-1)_{q-1}\}^2 = \sum_{j=0}^{q-1} \binom{q-1}{j}^2 j! (N)_{2q-1-j}. \quad (\text{A5})$$

It counts two ordered collections sharing their distinguished root, according to their number of further shared labels. Since $w_N^{(q)}/w_N^{(1)} = F_1(N-1)_{q-1}/F_q$, insertion gives

$$1 + \chi^2(w^{(q)} \| w^{(1)}) \leq \frac{\exp((q-1)^2/\mu)}{(1 - a_q)^2}. \quad (\text{A6})$$

Indeed the numerator sum is bounded by $\mu^{2q-1} \sum_{j \geq 0} (q-1)^{2j} / (j! \mu^j)$ before normalization. Let $v_1 = \kappa_1 = 0$ and, for $q \geq 2$, set

$$v_q = \min \left\{ 2, \left[\frac{e^{(q-1)^2/\mu}}{(1 - a_q)^2} - 1 \right]^{1/2} \right\},$$

$$\kappa_q = (q-1)^2/\mu - 2 \log(1 - a_q). \quad (\text{A7})$$

Then $\|w^{(q)} - w^{(1)}\|_1 \leq v_q$ and $D(w^{(q)} \| w^{(1)}) \leq \kappa_q$. For $q \leq s = \lceil \lambda^2 \rceil$, these bounds give

$$v_q = O(\lambda^2 \sqrt{\varepsilon}), \quad \kappa_q = O(\lambda^4 \varepsilon). \quad (\text{A8})$$

The identity $(N-q)(N)_q = (N)_{q+1}$ and comparison with a geometric distribution on the nonnegative integers imply

$$H(w^{(q)}) \leq h_q := \log \left[e \left(1 + \frac{\mu}{1 - a_q} \right) \right]. \quad (\text{A9})$$

One may first apply this comparison to truncated distributions, so entropy finiteness is a conclusion of the bound. All count weights are unchanged by the hard-sphere flow.

2. The collision-history input and the cutoff replay

For clarity we state the precise source-dependent analytic input used below. There are symmetric events B_N of the complete physical history, common to all selections of

observed particles, and normalized good and bad factorial laws such that

$$\rho_q = (1-p_q)g_q + p_q b_q, \quad p_q = \sum_{N \geq q} w_N^{(q)} P_N^0(B_N). \quad (\text{A10})$$

Terms of zero weight are omitted. For each fixed positive time, the DHM positive representation, replayed as specified below, gives constants $a > 0$ and $k < \infty$, independent of $1 \leq q \leq s$, such that

$$\|\rho_q - f_t^{\otimes q}\|_1 + p_q \leq C_t \varepsilon^a \lambda^k, \quad (\text{A11})$$

$$f_t/m_\varepsilon \leq M_t,$$

$$\log \max\{2, \|g_q/m_\varepsilon^{\otimes q}\|_\infty\} \leq \lambda^{k_t}. \quad (\text{A12})$$

The geometric definitions, restricted-flow injectivity, graph estimates, and leading-tree cancellation used to obtain these statements are external analytic inputs from DHM. We give their cutoff dependence and the cap argument here; we do not reprove the underlying collision-history construction. In particular, Eq. (A11) for a positive physical failure probability is not inferred from a small signed correlation error.

Choose L time layers of length $\tau = t/L < 1$ as in DHM. Replace the terminal cutoff in their (1.32) by

$$A_L = \lceil \lambda^2 \rceil, \quad \Lambda_j = A_j^{10d}, \quad A_{j-1} = \Lambda_j^{10d}. \quad (\text{A13})$$

All cutoffs are fixed powers of λ . With $R = (10d)^2$, one has $A_j = A_L^{R^{L-j}}$, and hence

$$A_j \Lambda_j^2 < A_{j-1}, \quad \Lambda_j/(\lambda A_j) \longrightarrow \infty, \quad \frac{\Lambda_j}{\lambda \sum_{i>j} A_i \Lambda_i^2} \longrightarrow \infty. \quad (\text{A14})$$

The source constant hierarchy and the smallness threshold for ε may depend on the fixed terminal power 2.

The finite identities of DHM Definition 1.6 and Proposition 5.16 request preceding orders at most $A_j \Lambda_j^2$, covered by Eq. (A14). Their initial cumulant estimate (5.81)–(5.82) uses $C|H_0|\varepsilon^d \leq \varepsilon^{d-1/2}$, still valid at these cutoffs. Proposition 7.2 gives graph counts $C^m \lambda^{K\rho}$ for m atoms and source complexity ρ ; this includes root permutations and initial-link choices at $|H_0| \leq \lambda^{K_0}$. It is essential to retain the joint Gaussian and time-volume estimate (9.65)–(9.81), rather than charge a power of λ for every normal collision component. In source notation its relevant form is

$$I_{M'}(Q') \leq (C_\tau)^m \lambda^{K(\rho+g)} \varepsilon^{\nu g - (d-1)b_0}, \quad \nu = 3^{-d-1}. \quad (\text{A15})$$

Here g and b_0 count good and bad final components. The geometric charge estimate (9.51) gives

$$\begin{aligned} \frac{\nu}{2}g + (d-1)(|H| - b_0) &\geq c_*\rho, \\ \frac{\nu}{2}g + (d-1)(|H| - b_0) &\geq \nu/2. \end{aligned} \quad (\text{A16})$$

A convex combination, the root factor $\varepsilon^{(d-1)|H|}$, and the Gaussian dyadic summation then give $\tau^{m/9} \varepsilon^{2\nu/5 + c_*\rho/5} \lambda^{K\rho}$. The graph sum therefore retains the positive cumulant gain $\varepsilon^{\nu/3 + b_*|H|}$ for some fixed $b_* > 0$. For error molecules, $m \geq \Lambda_j$ absorbs the extra factor $\varepsilon^{-C_*A_j}$ and leaves the positive bound $\tau^{\Lambda_j/10}$. These are bounds for the absolute molecule integrals entering (3.17)–(3.18), not just their signed sums.

The leading-tree cancellation in Section 14 is retained. Its remaining geometric series has ratio $C\tau^{1/2} < 1$, and its discarded tail is $O((C\tau^{1/2})^{\Lambda_j - O(1)})$. The mean-value and exceptional-input losses are polynomial in λ and are absorbed in the fixed positive source exponent gaps. If R_j is the positive propagated error mass, the recurrence (3.19), (5.46) becomes

$$R_0 = 0, \quad R_j \leq \tau^{\Lambda_j/10} + \varepsilon^{-2(d-1)\Lambda_j^2 A_j} R_{j-1}. \quad (\text{A17})$$

Its expansion has logarithms bounded by $-c_\tau \Lambda_j + 2(d-1)\lambda \sum_{i>j} \Lambda_i^2 A_i$. They tend to $-\infty$ faster than $-a\lambda$ for every fixed a by Eq. (A14).

To control physical failure, use the original singleton positive majorant in (5.98)–(5.100), (3.20), and Section 15. First-failure coupling to the original truncated flow bounds the global physical failure probability by $O(\varepsilon^{\nu/3})$ under $w^{(1)}$. Enlarging all cutoffs enlarges the physical good event. Transferring from $w^{(1)}$ to $w^{(q)}$ costs at most v_q in Eq. (A8); since $\nu/3 < 1/2$, this preserves a positive power uniformly for $q \leq s$. The event is global and permutation invariant, even when its witness lies outside the selected roots. Coupling gives the corresponding physical truncation error. Finally the tensor-product error is bounded by $q\varepsilon^\delta(1+\varepsilon^\delta)^{q-1}$, the nonempty cumulants by $\varepsilon^{\nu/3} q\varepsilon^{b_*} \exp(q(\varepsilon^\delta + \varepsilon^{b_*}))$, and the forbidden-pair product mass by $Cq^2\varepsilon^d$. The normalization μ^q/F_q is controlled by Eq. (A2). This proves the stated form of Eq. (A11) from those source estimates.

3. A probability reference and a direct good-density cap

Cube envelopes avoid a covering problem with unit balls on $\mathbb{Z}^d$ in high dimensions. Finitely many translates of the unit-ball lattice cover each unit cube; applying the translation-covariant DHM estimates with common constants transfers the source norms to cubes. At a fixed endpoint, let a_k^0 dominate the cube Gaussian suprema of the finitely many target profiles, and let $r_k^\varepsilon \geq 0$ collect the positive profile remainders of Proposition 6.1. Their sums satisfy $\sum a_k^0 < \infty$ and $\sum r_k^\varepsilon \leq C_t \varepsilon^\theta$. Choose a fixed positive summable sequence b_k comparable to $e^{-|k|}$, and set

$$\begin{aligned} W_\varepsilon(x) &= \sum_k (a_k^0 + r_k^\varepsilon + b_k) e^{-\gamma|x-k|}, \\ w_\varepsilon &= W_\varepsilon / \int W_\varepsilon, \quad m_\varepsilon = w_\varepsilon M_{\beta_m}. \end{aligned} \quad (\text{A18})$$

Here M_{β_m} is a normalized Maxwellian. The normalizers are bounded above and below. Termwise comparison and exponential convolution give uniform bounds

$$\begin{aligned} w_\varepsilon(x+h) &\leq e^{\gamma|h|} w_\varepsilon(x), & \|w_\varepsilon\|_\infty &\leq C_t, \\ w_\varepsilon(x) &\geq c_t e^{-\gamma|x|}, & |\log w_\varepsilon(x)| &\leq C_t + \gamma|x|. \end{aligned} \quad (\text{A19})$$

The approximate input profiles obey $|f^A(j\tau, x, v)| \leq C_t w_\varepsilon(x) e^{-\beta_*|v|^2}$ with fixed $\beta_* > 0$. Choose β_m below the surviving Gaussian exponent.

For a prescribed source molecule with p particle lines and q final roots, write (y_i, u_i) for its input points and (x_i, v_i) for the common final points, continuing lines freely outside their active times. Elasticity and path length give

$$E = \sum_i |u_i|^2 = \sum_i |v_i|^2, \quad \sum_i |y_i - x_i| \leq T\sqrt{pE}. \quad (\text{A20})$$

For $0 < b < \beta$, a product input envelope is therefore bounded on the molecule support by

$$\begin{aligned} C_0^p \prod_i w_\varepsilon(y_i) e^{-\beta|u_i|^2} \\ \leq C_0^p e^{\gamma^2 T^2 p/[4(\beta-b)]} \prod_i w_\varepsilon(x_i) e^{-b|v_i|^2}. \end{aligned} \quad (\text{A21})$$

The prescribed flow is injective on its molecule indicator, as used before DHM (5.94). The non-root variables in Definition 3.23 and (3.12) are already final coordinates. Only after the preceding inequality may one drop the indicators and integrate those variables; there is no change of variables on a fixed-root fiber. Initial-link factors also lie in $[0, 1]$. Thus

$$\begin{aligned} |\text{IN}_M|(z_q) &\leq \mu^{p-q} C_0^p e^{\gamma^2 T^2 p/[4(\beta-b)]} (\pi/b)^{d(p-q)/2} \\ &\times \prod_{i=1}^q w_\varepsilon(x_i) e^{-b|v_i|^2}. \end{aligned} \quad (\text{A22})$$

The activity factor is retained: its logarithm is $(d-1)(p-q)\lambda$.

All molecule sizes are bounded by a fixed power of λ . Their finite graph encodings cost at most $\exp(CP_{\max} \log(2 + P_{\max}))$, with $P_{\max} = \sum_j A_j \Lambda_j^2$. Consequently every single-layer positive error operator has envelope norm with logarithm at most λ^{k_t} . Assign Gaussian exponents

$$b_j = \beta_*/2 - j\beta_*/(4L), \quad \beta_m = \beta_*/8. \quad (\text{A23})$$

If $H_{j,q}$ bounds the entire error family at level j , the pointwise recurrence (5.52)–(5.54), including the ordinary/error union following (5.57), gives

$$H_{j,q} \leq e^{\lambda^{k_t}} + e^{\lambda^{k_t}} \max_{q \leq p \leq q\Lambda_j^2} H_{j-1,p}. \quad (\text{A24})$$

The preceding error enters once. The preceding orders are admissible by Eq. (A14), and there are only L iterations.

Source cumulants obey the same cap by Eq. (A22). Their subset sum in (3.15) costs at most 2^q , which is harmless on the logarithmic scale of this cap. The full positive truncated density is therefore bounded by $e^{\lambda^{k_t}} m_\varepsilon^{\otimes q}$. The physical good sublaw is dominated by this positive truncated law: on each survival domain the two flows agree, and dropping the survival indicator increases the positive recursion. Division by F_q/μ^q and by $1 - p_q$ proves Eq. (A12).

4. The full-sector likelihood and bad-event entropy

The initial envelope has a strict Gaussian reserve: $f_0 \leq C_0 m_\varepsilon e^{-\zeta|v|^2}$ for some fixed $\zeta > 0$. For a physical trajectory let $E_N = \sum_i |v_i|^2$ be its conserved energy. Volume preservation and the actual path lengths imply

$$\begin{aligned} \log \frac{dP_N^t}{dm_\varepsilon^{\otimes N}} &\leq -\log Z_N^f + N \log C_0 - \zeta E_N + \gamma T \sqrt{N E_N} \\ &\leq -\log Z_N^f + C_T N. \end{aligned} \quad (\text{A25})$$

This also proves finiteness of the sector KL: its positive logarithmic part is bounded above, and the negative part of a likelihood ratio integrated against its probability law is at most $1/e$. For any symmetric event B_N , put $p_N = P_N^0(B_N)$. Conditioning and transport give a symmetric endpoint $R_{N,B}^t$ satisfying

$$\begin{aligned} p_N D(R_{N,B}^t \| m_\varepsilon^{\otimes N}) \\ \leq p_N (C_T N - \log Z_N^f) + p_N \log(1/p_N). \end{aligned} \quad (\text{A26})$$

For a symmetric N -particle law with finite product-reference KL, write $D_j = D(P_{:,j} \| m^{\otimes j})$. The increments satisfy

$$D_j - D_{j-1} = D(P_{:,j} \| m) + I(Z_j; Z_{1:j-1}). \quad (\text{A27})$$

They are nondecreasing by symmetry and data processing. It follows that

$$\frac{D(P_{:,q} \| m^{\otimes q})}{q} \leq \frac{D(P_{:,r} \| m^{\otimes r})}{r}, \quad q \leq r \leq N. \quad (\text{A28})$$

This fractional tensorization is applied inside one sector, before mixing any count weights.

Apply Eq. (A28) to Eq. (A26) and mix with $w^{(q)}$. Convexity gives

$$\begin{aligned} \frac{p_q}{q} D(b_q \| m_\varepsilon^{\otimes q}) &\leq C_T p_q + \sum_N w_N^{(q)} p_N \frac{-\log Z_N^f}{N} \\ &+ \sum_N \frac{w_N^{(q)}}{N} p_N \log(1/p_N). \end{aligned} \quad (\text{A29})$$

The last sum is bounded by both $p_q \log(1/p_q)/q$ and $[e\mu(1-a_q)]^{-1}$, since $\sum_N w_N^{(q)}/N \leq F_{q-1}/F_q$. On $N \leq 4\mu$, Eq. (A4) bounds the partition term by $4c\varepsilon p_q$. On

$N > 4\mu$ multiply by the physical weight before using $z \log(1/z) \leq 1/e$. If $Y \sim \text{Pois}_\mu$, factorial shifting gives the bound

$$\mathcal{U}_q = \frac{e^\mu \mu^q}{4e\mu \Xi_f F_q} \Pr\{Y + q > 4\mu\} \leq \mu^{-1} e^{-c_1 \mu + o(\mu)}. \quad (\text{A30})$$

The last bound is uniform for $q \leq s$, by Eqs. (A2)–(A3) and Poisson Chernoff bounds. We have proved the finite-parameter estimate

$$\begin{aligned} \frac{p_q}{q} D(b_q \| m_\varepsilon^{\otimes q}) &\leq (C_T + 4c\varepsilon) p_q + \mathcal{U}_q \\ &\quad + \min \left\{ \frac{p_q \log(1/p_q)}{q}, \frac{1}{e\mu(1-a_q)} \right\}. \end{aligned} \quad (\text{A31})$$

All series are finite by these same bounds. The extensive likelihood in Eq. (A25) has been divided by N before Palm mixing. No requirement $p_q \bar{N} \rightarrow 0$ is used.

5. Entropy continuity and factorial information

If $p/m, g/m \leq B$ on a probability reference m , $B \geq 2$, and $\delta = \|p - g\|_1$, then

$$|D(p\|m) - D(g\|m)| \leq \Phi(\delta, B) := \delta \log(eB^2/\delta), \quad (\text{A32})$$

with value zero at $\delta = 0$. To see this, set $F(x) = x \log x$. For $a, b \in [0, B]$ and $u = |a - b|$, rescaling to $[0, 1]$ gives $|F(a) - F(b)| \leq u(1 + 2 \log B) - u \log u$. Integrate and apply Jensen to $-u \log u$. The bound is independent of the dimension and remains valid for a changing reference.

For Eq. (A10), set

$$\begin{aligned} \eta_q &= \min \left\{ 2, \frac{\|\rho_q - f_t^{\otimes q}\|_1 + 2p_q}{1 - p_q} \right\}, \\ B_q &= \max\{2, \|g_q/m_\varepsilon^{\otimes q}\|_\infty, M_t^q\}. \end{aligned} \quad (\text{A33})$$

The good density is within η_q of $f_t^{\otimes q}$. The binary mixture label J gives the exact finite chain rule

$$\begin{aligned} D(\rho_q \| m_\varepsilon^{\otimes q}) &= (1 - p_q) D(g_q \| m_\varepsilon^{\otimes q}) \\ &\quad + p_q D(b_q \| m_\varepsilon^{\otimes q}) - I(J; Z_q). \end{aligned} \quad (\text{A34})$$

Here $0 \leq I(J; Z_q) \leq h_2(p_q)$. Consequently the absolute reference error Δ_q is bounded by

$$\begin{aligned} \Delta_q &:= |D(\rho_q \| m_\varepsilon^{\otimes q}) - qD(f_t \| m_\varepsilon)| \\ &\leq \Phi(\eta_q, B_q) + p_q D(b_q \| m_\varepsilon^{\otimes q}) \\ &\quad + p_q q \log M_t + h_2(p_q). \end{aligned} \quad (\text{A35})$$

Equations (A11), (A12), and (A31) give $\sup_{q \leq s} \Delta_q \leq C_t \varepsilon^b \lambda^p$. At $t = 0$, insertion gives $\rho_q(0) \leq (1 - a_q)^{-1} f_0^{\otimes q}$ and a positive-power L^1 error, so Eq. (A32) gives the same conclusion directly. This proves Eq. (16).

We next remove the technical reference without pretending that ρ_1 is the marginal of ρ_q . If $f/m \leq M$ and $\delta = \|p - f\|_1$, the bounded KL variational test $\phi = \log \max(f/m, \eta)$ gives

$$D(p\|m) \geq D(f\|m) - \Omega_M(\delta), \quad (\text{A36})$$

where $\Omega_M(0) = 0$ and, for $\delta > 0$,

$$\Omega_M(\delta) = \delta \max\{|\log \eta|, \log M\} + \eta, \quad \eta = \min(\delta, 1/2). \quad (\text{A37})$$

Indeed $\int e^\phi dm \leq 1 + \eta$ and $\int f \phi \geq D(f\|m)$. This proof handles zeros of f and moving references. Marginalization gives $\|\sigma_q - f_t\|_1 \leq \delta_q := \|\rho_q - f_t^{\otimes q}\|_1$. The product-reference identity and the count-kernel data-processing inequality yield

$$\begin{aligned} D(\rho_q \| \sigma_q^{\otimes q}) &\leq \Delta_q + q\Omega_{M_t}(\delta_q), \\ D(\sigma_q \| \rho_1) &\leq \kappa_q, \\ D(\rho_q \| \rho_1^{\otimes q}) &\leq \Delta_q + q\Omega_{M_t}(\delta_q) + q\kappa_q. \end{aligned} \quad (\text{A38})$$

This proves Eq. (17), including its absolute error bound throughout $q \leq s$.

6. Conditional information from a growing block

We prove the finite-parameter implication used to remove count mixing. Let P_N be symmetric, $u_N = P_{N:1}$, and suppose

$$D(P_N \| m^{\otimes N}) \leq Nb_N < \infty. \quad (\text{A39})$$

Let ν be a probability on $N \geq 1$, ω a probability on $N \geq s$, and assume $\sum_N (\nu_N + \omega_N) b_N < \infty$ and $H(\omega) < \infty$. Set $r = \sum \nu_N u_N$ and $R_s = \sum \omega_N P_{N:s}$. Suppose $f/m \leq M$ and

$$s^{-1} D(R_s \| m^{\otimes s}) \leq D(f\|m) + e, \quad e \geq 0. \quad (\text{A40})$$

Put

$$\begin{aligned} \mathcal{T} &= \sum_N |\nu_N - \omega_N| b_N, \quad L_q = \sum_{N < q} \nu_N b_N, \\ E &= e + H(\omega)/s + \mathcal{T} + \Omega_M(\|r - f\|_1). \end{aligned} \quad (\text{A41})$$

Then

$$\begin{aligned} \sum_N \nu_N D(u_N \| r) &\leq E, \\ \sum_{N \geq q} \nu_N D(P_{N:q} \| r^{\otimes q}) &\leq q(E + L_q), \quad 1 \leq q \leq s. \end{aligned} \quad (\text{A42})$$

The same upper bound holds with the smaller own-sector-product KL.

Here is a proof that keeps the omitted low-count sectors explicit. Write $K_{N,q} = D(P_{N:q} \| m^{\otimes q})$ when $N \geq q$ and zero otherwise. The joint count/block chain rule gives

$$\begin{aligned} \sum_N \omega_N K_{N,s} &= D(R_s \| m^{\otimes s}) + I(N; Z_s) \\ &\leq s\{D(f\|m) + e\} + H(\omega). \end{aligned} \quad (\text{A43})$$

Fractional tensorization and $K_{N,q} \leq qb_N$ imply

$$q^{-1} \sum_N \nu_N K_{N,q} \leq D(f\|m) + e + H(\omega)/s + \mathcal{T}. \quad (\text{A44})$$

For $q = 1$, subtracting Eq. (A36) from the identity $\sum \nu_N D(u_N\|m) = D(r\|m) + \sum \nu_N D(u_N\|r)$ proves the first assertion. To prove the second put $C_N = D(u_N\|r)$; it is finite because $u_N \leq r/\nu_N$ in every positively weighted sector. The exact finite identity is

$$\begin{aligned} & \sum_{N \geq q} \nu_N D(P_{N;q}\|r^{\otimes q}) \\ &= \sum_{N \geq q} \nu_N K_{N,q} - qD(r\|m) \\ &+ q \sum_{N < q} \nu_N D(u_N\|m) - q \sum_{N < q} \nu_N C_N. \end{aligned} \quad (\text{A45})$$

Dropping only the last nonnegative tail and using Eqs. (A36) and (A44) proves Eq. (A42). Finiteness follows from the assumed summable envelope and the count/block chain rule, so no subtraction of infinite entropies has occurred.

For the hard-sphere application, Eq. (A25) supplies $b_N = C_T - \log Z_N^f/N$. On $N \leq 2\mu$, $b_N \leq B_\varepsilon := C_T + 2c\varepsilon$. For $Y \sim \text{Pois}_\mu$, define

$$\begin{aligned} \mathcal{R}_j &= \left(C_T + \frac{1}{2e\mu}\right) \frac{e^{c\mu\varepsilon/2}}{1 - a_j} \Pr\{Y + j > 2\mu\}, \\ \mathcal{L}_q &= B_\varepsilon \frac{e^{c\mu\varepsilon/2}}{1 - a_1} \Pr\{Y \leq q - 2\} \quad (q \geq 2), \quad \mathcal{L}_1 = 0. \end{aligned} \quad (\text{A46})$$

Multiplication by the physical partition weight, the inequality $Z_N^f(-\log Z_N^f) \leq 1/e$, and Poisson factorial shifting give

$$\begin{aligned} & \sum_{N > 2\mu} w_N^{(j)} b_N \leq \mathcal{R}_j, \\ & \sum_N |w_N^{(1)} - w_N^{(s)}| b_N \leq B_\varepsilon v_s + \mathcal{R}_1 + \mathcal{R}_s, \\ & \sum_{N < q} w_N^{(1)} b_N \leq \mathcal{L}_q. \end{aligned} \quad (\text{A47})$$

These tails are exponentially small in μ , uniformly for $j, q \leq s$; the factor $e^{c\mu\varepsilon/2}$ is only $e^{o(\mu)}$. Apply Eq. (A42) with $\nu = w^{(1)}$, $\omega = w^{(s)}$, $r = \rho_1$, $f = f_t$, and $e = \Delta_s/s$. An admissible bound is

$$\begin{aligned} B_{\text{info}} &= \Delta_s/s + h_s/s + B_\varepsilon v_s + \mathcal{R}_1 + \mathcal{R}_s \\ &+ \Omega_{M_t}(\|\rho_1 - f_t\|_1) + \max_{q \leq s} \mathcal{L}_q. \end{aligned} \quad (\text{A48})$$

Since $h_s/s = (d-1)/\lambda + O(\lambda^{-2})$, all other terms have positive-power errors. This proves Eqs. (18)–(19). Count/singleton convergence alone would not imply the whole-block bound; the transfer in Eq. (A44) is required.

7. Shannon convergence and strict within-sector attribution

The assumed finite initial second spatial moment supplies the needed whole-space entropy domain. Indeed Eq. (A19) implies a uniform bound $|\log m_\varepsilon(x, v)| \leq C_t(1 + |x| + |v|^2)$. For a fixed q , sample the ordered labels initially with weight $w^{(q)}$ and follow these same physical particles. Initial insertion gives $\rho_q(0) \leq (1 - a_q)^{-1} f_0^{\otimes q}$. Symmetry and conservation of total energy make the expected kinetic energy of each selected particle equal to its initial expectation. The path inequality

$$|X_i(t) - X_i(0)|^2 \leq t \int_0^t |V_i(r)|^2 dr \quad (\text{A49})$$

therefore bounds its spatial second moment uniformly in ε and $t \in [0, T]$. The limiting target has the same bound by L^1 convergence and Fatou's lemma. It follows that the spatial cross weights $|\log w_\varepsilon|$ are uniformly integrable.

The initial selected kinetic energy converges to that of f_0 , by insertion domination and the initial L^1 limit. The regular Boltzmann solution conserves this energy. Hence the two endpoint kinetic-energy integrals agree in the limit. Combined with L^1 convergence this gives

$$\int |v_i|^2 |\rho_q - f_t^{\otimes q}| dZ_q \longrightarrow 0. \quad (\text{A50})$$

For example, extract an almost-everywhere convergent subsequence; the minimum of the two densities, multiplied by $|v_i|^2$, is dominated by the target weighted density. Dominated convergence and convergence of the two total moments prove the claim, and the subsequence criterion gives the full limit. This step requires matching energies, not merely finite energy bounds. Spatial truncation using the uniform second moment and Eq. (A50) now imply

$$\int (\rho_q - f_t^{\otimes q}) \log m_\varepsilon^{\otimes q} dZ_q \longrightarrow 0. \quad (\text{A51})$$

Expanding the reference KL in Eq. (16) proves Eq. (20). This argument is for fixed q ; it does not supply a quantitative growing-order Shannon tail estimate.

All exact entropy identities can also be made finite before taking a limit. Fix a positive reference m_* constructed from the initial cube envelope and an exponential spatial floor, with the same Gaussian reserve. The full grand-canonical likelihood estimate gives

$$\begin{aligned} D(P_t\|\Pi_{\bar{N}, m_*}) &\leq A_\varepsilon + C_T \bar{N}, \\ A_\varepsilon &= \bar{N} - \log \Xi_f + \bar{N} \log(\mu/\bar{N}). \end{aligned} \quad (\text{A52})$$

Direct computation identifies $A_\varepsilon = D(P_0\|\Pi_{\bar{N}, f_0})$ and yields

$$0 \leq A_\varepsilon \leq \mu - \log \Xi_f = O(\mu\varepsilon). \quad (\text{A53})$$

The product-reference projection is

$$D(P_t\|\Pi_{\bar{N}, m_*}) = \mathcal{I}_{\text{gc}}^\varepsilon(t) + \bar{N} D(\rho_1(t)\|m_*). \quad (\text{A54})$$

One way to justify finiteness first is to use the Poisson exponential formula with a bounded positive one-particle test u :

$$D(P_t \| \Pi_{\bar{N}, m_*}) \geq \bar{N} \left[\int \rho_1 \log u - \int u m_* + 1 \right]. \quad (\text{A55})$$

Taking the supremum gives the finite singleton KL. Its logarithm is then integrable, allowing the likelihood decomposition and Eq. (A54). The initial moment bounds, Eq. (A49), and the sector likelihood bound also imply $\sum_N \pi_N \int P_N^t |\log P_N^t| < \infty$. Physical partition weighting controls this sum even at arbitrarily large N . Poisson domination at each fixed ε makes the count entropy and all needed count cross terms finite.

Expansion of these finite quantities gives

$$\mathcal{I}_{\text{gc}}^\varepsilon(t) = D(\pi \| \text{Pois}_{\bar{N}}) - \sum_N \pi_N h(P_N^t) + \bar{N} h(\rho_1(t)). \quad (\text{A56})$$

The first two terms are conserved by the flow, proving Eq. (24). Expanding each sector KL around its own marginal gives Eq. (15). Data processing at $t = 0$ and Eq. (A54) imply

$$0 \leq D(\pi \| \text{Pois}_{\bar{N}}) \leq A_\varepsilon, \quad 0 \leq \mathcal{I}_{\text{gc}}^\varepsilon(0) \leq A_\varepsilon. \quad (\text{A57})$$

Combining the Shannon limit, exact balance, initial bound, and $C_\varepsilon \rightarrow 0$ proves Eq. (21) for K_ε as well as $\mathcal{I}_{\text{gc}}^\varepsilon$.

For time integration, the fixed reference in Eq. (A52) supplies the bound

$$0 \leq K_\varepsilon(t)/\bar{N} \leq \mathcal{I}_{\text{gc}}^\varepsilon(t)/\bar{N} \leq C_T + A_\varepsilon/\bar{N}, \quad 0 \leq t \leq T. \quad (\text{A58})$$

The measurable fixed-time limits therefore converge in $L^1(0, T)$ by dominated convergence. Testing their distributional derivatives against $\varphi \in C_c^\infty(0, T)$ gives $-\int \Delta h \varphi' = \int \mathcal{D}_B[f_t] \varphi$, using Eq. (B28), which proves Eq. (22). This is not convergence of pointwise microscopic rates or of incoming collision traces.

8. The particle–rest lower bound

For a fixed symmetric sector with finite product-reference KL, the projection identity gives

$$D(P_{N:r} \| m^{\otimes r}) = I_{N,r} + r D(u_N \| m) < \infty. \quad (\text{A59})$$

The KL chain rule gives $J_N = I_N - I_{N,N-1}$. The increments $a_r = I_{N,r} - I_{N,r-1} = I(Z_r; Z_{1:r-1})$, with $a_1 = 0$, are nondecreasing by symmetry and data processing. Therefore

$$0 \leq J_N \leq I_N, \quad J_N = a_N \geq I_N/N. \quad (\text{A60})$$

There are no differences of infinite Shannon entropies in this argument.

For completeness, the physical average of J_N is finite at each ε . Equation (A25) bounds $J_N \leq I_N \leq C_T N - \log Z_N^f$. Since $0 \leq Z_N^f \leq 1$,

$$\begin{aligned} \mathbb{E}_\pi N^2 &\leq (\mu^2 + \mu) e^\mu / \Xi_f, \\ \sum_{N \geq 1} \pi_N N (-\log Z_N^f) &\leq \mu e^\mu / (e \Xi_f). \end{aligned} \quad (\text{A61})$$

These establish finiteness, without a uniform asymptotic upper bound. Multiplication of Eq. (A60) by $N \pi_N / \bar{N}$ and summation proves $\mathcal{J}_\varepsilon \geq K_\varepsilon / \bar{N}$; the entropy limit gives its stated liminf. Equation (27) is the $q = 2$ case of Eq. (19). This completes the proof of Corollary 1.

Appendix B: Endpoint control and the reversal theorems

This appendix proves Sec. IV under the whole-space hypotheses of Sec. II. Only the one- and two-particle conclusions of DHM Theorem 1 and the spatial propagation estimate in its Proposition A.1 are used [1]. No kinetic theorem restarted from g , or entropy theorem for higher particle orders, is an additional input.

1. Positivity and global spatial integrability

Write $Q(f, f) = Q^+(f, f) - \nu_f f$. Angular integration gives $\nu_f(x, v) = c_d \int |v - v_*| f(x, v_*) dv_*$, with $c_d < \infty$. The solution's uniform Gaussian bound therefore implies

$$0 \leq \nu_{f_t}(x, v) \leq C_T(1 + |v|). \quad (\text{B1})$$

Only velocity is integrated here; no spatial integrability has yet been assumed. The characteristic formula and $Q^+ \geq 0$ imply

$$\begin{aligned} f_t(x, v) &\geq f_0(x - tv, v) e^{-C_T t(1 + |v|)} \\ &\geq c_T e^{-a - \sqrt{1 + |x|^2} - b_T |v|^2} > 0. \end{aligned} \quad (\text{B2})$$

Use $\sqrt{1 + |x - tv|^2} \leq \sqrt{1 + |x|^2} + t|v|$ and absorb linear velocity terms into the Gaussian. Together with $f_t \leq A$, this proves Eq. (32).

The Gaussian supremum cannot itself be integrated over $\mathbb{R}_x^d$. Apply the localization estimate of DHM Proposition A.1 with

$$\beta_*(t) = 2\beta - \frac{\beta t}{T + 1} > \beta. \quad (\text{B3})$$

Its decreasing-weight supremum bound follows from the assumed $e^{2\beta|v|^2} f_t$ bound, and its initial spatial and derivative norms follow from the initial cube sum. Here is the localization argument with cubes, which avoids assuming that integer-centered unit balls cover space in every dimension.

Set $Q_j = j + [0, 1]^d$ and

$$U_j(t) = \sup_{x \in Q_j, v} e^{\beta_*(t)|v|^2} f_t(x, v),$$

$$M_\kappa(t) = \sum_j e^{-\kappa|j|} U_j(t), \quad 0 < \kappa \leq 1. \quad (\text{B4})$$

The cutoff sum is finite even before propagation is proved, since $U_j(t) \leq A$. In the collision integral from s to t , pair energy conservation leaves

$$e^{-\beta_*(t)|v_*|^2} e^{-[\beta_*(s) - \beta_*(t)](|v|^2 + |v_*|^2)}. \quad (\text{B5})$$

Half of the second Gaussian absorbs the hard-sphere collision factor at cost $C(t-s)^{-1/2}$. Split the remaining integral into dyadic displacement shells $(t-s) \max(|v|, |v_*|) \sim K$, $K = 1, 2, 4, \dots$, including displacement at most one in the first shell. The remaining Gaussian is at most $C_T e^{-c_T K^2}$. For terminal $x \in Q_i$, the common source position belongs to Q_j with $|i-j| \leq C_d(1+K)$. Bound one weighted density by A and the other by $U_j(s)$. Pair energy conservation makes this valid for the gain term, including its postcollision velocities.

The weight transfer in the spatial sum costs at most $e^{C_d(1+K)}$, and at most $C_d(1+K)^d$ cubes contribute. The shell sum $\sum_K (1+K)^d e^{C_d(1+K)} e^{-c_T K^2}$ is finite uniformly in κ . The free term has the same displacement bound without the collision factor. Dropping the nonpositive loss term gives

$$M_\kappa(t) \leq C_T M_\kappa(0) + C_T \int_0^t (t-s)^{-1/2} M_\kappa(s) ds. \quad (\text{B6})$$

The iterated kernels have summable bounds proportional to $C^n t^{n/2} / \Gamma(1+n/2)$. Hence $\sup_t M_\kappa(t) \leq C_T$, uniformly in κ , by the unweighted initial cube norm. Monotone convergence as $\kappa \downarrow 0$ proves spatial summability. In particular,

$$F_t(x) := \sup_v e^{\beta|v|^2} f_t(x, v) \leq A,$$

$$\sup_{t \leq T} \int F_t(x) dx < \infty. \quad (\text{B7})$$

This is the spatially localized content of Proposition A.1 needed here. It bounds the full (x, v) integral of every fixed velocity polynomial times f_t , uniformly in time.

Pair energy conservation further gives

$$Q^+(f_t, f_t) + \nu_{f_t} f_t \leq C F_t(x)^2 (1 + |v|) e^{-\beta|v|^2}$$

$$\leq C' F_t(x) (1 + |v|) e^{-\beta|v|^2}. \quad (\text{B8})$$

The collision integrals against fixed velocity polynomials are therefore absolutely integrable. Collision changes of variables give $\int Q(f_t, f_t) dv = \int |v|^2 Q(f_t, f_t) dv = 0$ at each x . Test the equation with $\chi(x/R)$ and $\chi(x/R)|v|^2$, where χ is a compact smooth cutoff equal to one near the origin. Velocity cutoffs can first be removed using

Eq. (B8). The time-integrated transport boundary terms are at most $CR^{-1} \int_0^T \int |v| f_t$ and $CR^{-1} \int_0^T \int |v|^3 f_t$, respectively, and vanish as $R \rightarrow \infty$. Thus

$$\int f_t = 1, \quad \int |v|^2 f_t = E_0 := \int |v|^2 f_0, \quad 0 \leq t \leq T. \quad (\text{B9})$$

2. The actual grand-canonical normalization

For a normalized profile u with bounded spatial density $R_u = \|\int u(x, v) dv\|_\infty$, fix ε and vary the activity λ :

$$\Xi_u(\lambda) = \sum_{N \geq 0} \frac{\lambda^N}{N!} \int \mathbf{1}_{\mathcal{D}_{N, \varepsilon}} \prod_i u(z_i) dZ_N. \quad (\text{B10})$$

The configuration integrals lie in $[0, 1]$, which justifies differentiating the exponential series. Particle insertion gives

$$\frac{d}{d\lambda} \log \Xi_u(\lambda) = \frac{n_u(\lambda)}{\lambda}$$

$$= \mathbb{E}_{Q_{\lambda, u}} \int u(z) \prod_i \mathbf{1}_{\{|x-x_i| > \varepsilon\}} dz. \quad (\text{B11})$$

At $\lambda = 0$, use the continuous value one. Acceptance is at most one, so $n_u(\lambda) \leq \lambda$. A union bound gives

$$1 - v_d R_u \varepsilon^d \lambda \leq \frac{d}{d\lambda} \log \Xi_u(\lambda) \leq 1. \quad (\text{B12})$$

Independence of the present particles is not assumed. Integration to, and evaluation at, $\mu = \varepsilon^{-(d-1)}$ yield

$$0 \leq \mu - \log \Xi_u(\mu) \leq \frac{1}{2} v_d R_u \mu \varepsilon, \quad (\text{B13})$$

$$1 - v_d R_u \varepsilon \leq n_u(\mu) / \mu \leq 1. \quad (\text{B14})$$

These apply to $u = f_0, g$, and prove the partition correction in Theorem 2.

The activity-rescaled ordered factorial densities of P_t^ε are characterized by

$$\mathbb{E}_{P_t^\varepsilon} \sum_{i_1, \dots, i_q \text{ distinct}} \psi(z_{i_1}, \dots, z_{i_q}) = \mu^q \int \psi c_{\varepsilon, q}(t). \quad (\text{B15})$$

They are not probability-normalized marginals: $\int c_{\varepsilon, 1} = \bar{N} / \mu$. Initial insertion gives

$$c_{\varepsilon, 1}(0, z) = f_0(z) \mathbb{E}_{Q_{\mu, f_0}} \prod_i \mathbf{1}_{\{|x-x_i| > \varepsilon\}}. \quad (\text{B16})$$

The upper bound $c_{\varepsilon, 1}(0) \leq f_0$ first bounds the actual neighbor intensity. A further union bound then proves

$$(1 - c_0 \varepsilon) f_0 \leq c_{\varepsilon, 1}(0) \leq f_0, \quad c_0 = v_d R_{f_0}. \quad (\text{B17})$$

Consequently every nonnegative f_0 -integrable observable has its actual initial intensity moment converging to its f_0 moment. In particular

$$E_\varepsilon := \int |v|^2 c_{\varepsilon,1}(0) \rightarrow E_0, \\ \int |x|^2 c_{\varepsilon,1}(0) \leq X_0 := \int |x|^2 f_0. \quad (\text{B18})$$

Two-particle insertion gives $c_{\varepsilon,2}(0) \leq f_0^{\otimes 2}$; hence $\mathbb{E}(N)_2 \leq \mu^2$ and $\mathbb{E}(N/\mu)^2 \leq 1 + \mu^{-1}$. The microscopic flow conserves the count. These statements do not assert a Poisson count law.

3. Weighted endpoint concentration from orders one and two

The DHM conclusions used here, uniformly on the fixed interval, are

$$\|c_{\varepsilon,1}(t) - f_t\|_1 \leq \varepsilon^\theta, \\ \|c_{\varepsilon,2}(t) - f_t^{\otimes 2} \mathbf{1}_{\mathcal{D}_{2,\varepsilon}}\|_1 \leq \varepsilon^\theta. \quad (\text{B19})$$

The two-particle excluded volume under $f_t^{\otimes 2}$ is at most $v_d R_T \varepsilon^d$, where $R_T = \sup_t \|\int f_t dv\|_\infty < \infty$. For bounded real φ , the exact factorial identities are

$$\mathbb{E} M_\varphi(t) = \int \varphi c_{\varepsilon,1}(t), \\ \mathbb{E} M_\varphi(t)^2 = \int \varphi(z_1) \varphi(z_2) c_{\varepsilon,2}(t) + \mu^{-1} \int \varphi^2 c_{\varepsilon,1}(t). \quad (\text{B20})$$

The diagonal contribution in the final term cannot be omitted. Using $\int c_{\varepsilon,1} \leq 1$ gives

$$\sup_{t \leq T} \mathbb{E} \left| M_\varphi(t) - \int \varphi f_t \right|^2 \leq \|\varphi\|_\infty^2 (3\varepsilon^\theta + v_d R_T \varepsilon^d + \mu^{-1}). \quad (\text{B21})$$

The same bound applies to a measurable family φ_t with a common supremum bound. Taking $\varphi = 1$ proves $N/\mu \rightarrow 1$ in L^2 .

A bounded mean energy alone would not justify logarithmic tails. The microscopic flow conserves total energy, rather than each particle's energy, and thus

$$\int |v|^2 c_{\varepsilon,1}(t) = E_\varepsilon. \quad (\text{B22})$$

For $V(v) = |v|^2$ and $M > 0$, compare this with the conserved Boltzmann energy:

$$\int (V - M)_+ c_{\varepsilon,1}(t) \\ \leq |E_\varepsilon - E_0| + M\varepsilon^\theta + \int (V - M)_+ f_t. \quad (\text{B23})$$

The last term tends to zero uniformly in t as $M \rightarrow \infty$ by Eq. (B7). The inequality $V \mathbf{1}_{\{V > 2M\}} \leq 2(V - M)_+$ proves uniform integrability of the microscopic energy tail after taking the large-particle limit first. No uniform fourth microscopic velocity moment is invoked.

For positions, pathwise continuity and Cauchy–Schwarz give

$$\sum_i |x_i(t)|^2 \leq 2 \sum_i |x_i(0)|^2 + 2t^2 \sum_i |v_i(0)|^2. \quad (\text{B24})$$

Taking expectations yields $\int |x|^2 c_{\varepsilon,1}(t) \leq X_T := 2X_0 + 2T^2 E_0$, whence $\int |x| \mathbf{1}_{\{|x| > R\}} c_{\varepsilon,1}(t) \leq X_T/R$. Applying Eq. (B19) to $|x|^2 \wedge K$ and then taking $K \rightarrow \infty$ gives $\int |x|^2 f_t \leq X_T$. With $W(x, v) = 1 + |x| + |v|^2$, these estimates imply

$$\lim_{K \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \sup_{t \leq T} \int W \mathbf{1}_{\{W > K\}} c_{\varepsilon,1}(t) = 0, \\ \lim_{K \rightarrow \infty} \sup_{t \leq T} \int W \mathbf{1}_{\{W > K\}} f_t = 0. \quad (\text{B25})$$

Indeed $W \mathbf{1}_{\{W > K\}} \leq 3 \sum_{\xi \in \{1, |x|, |v|^2\}} \xi \mathbf{1}_{\{\xi > K/3\}}$ reduces this to the separate tails.

The same bounds justify the Boltzmann entropy identity throughout the common data class, not only for the example below. Indeed the transport entropy flux is uniformly integrable, since

$$\int |v| f_t |\log f_t| \leq C_T \left[1 + \int |v|^3 f_t \right. \\ \left. + \left(\int |x|^2 f_t \int |v|^2 f_t \right)^{1/2} \right] < \infty. \quad (\text{B26})$$

For the collision term write $\mathcal{L} = |\log f| + |\log f_*| + |\log f'| + |\log f'_*|$. The logarithmic envelope, pair energy conservation, and integration of the partner against the pointwise Gaussian cap give

$$\int B f f_* \mathcal{L} dx dv dv_* d\omega \\ \leq C_T \int f [1 + |v|^3 + |x|(1 + |v|)] dx dv < \infty. \quad (\text{B27})$$

By Cauchy–Schwarz, the mixed spatial–velocity moment is finite. The same estimate controls the gain term by the pre/postcollision change of variables. Both bounds are uniform on the fixed interval. For the regular solution, first use compact spatial and velocity cutoffs and a regularized logarithm. Remove the logarithmic regularization and velocity cutoff by these absolute-integrability bounds. The spatial transport boundary is bounded by C/R times Eq. (B26), and tends to zero. Collision symmetrization therefore gives

$$h(f_t) - h(f_0) = \int_0^t \mathcal{D}_B[f_s] ds, \quad \mathcal{D}_B[f] \in L^1(0, T). \quad (\text{B28})$$

This supplies the entropy-identity hypothesis needed for the distributional derivative in Eq. (22).

Lemma 1 (Weighted empirical law of large numbers). *If a real measurable family ψ_t , independent of ε , satisfies $|\psi_t| \leq CW$ on $[0, T]$, then*

$$\sup_{t \leq T} \mathbb{E}_{P_t^\varepsilon} \left| \mu^{-1} \sum_i \psi_t(z_i) - \int f_t \psi_t \right| \rightarrow 0. \quad (\text{B29})$$

Proof. Clip ψ_t to $[-K, K]$, writing $\psi_t^{[K]}$. The expectation in Eq. (B29) is bounded by

$$\begin{aligned} & \left(\mathbb{E} \left| M_{\psi_t^{[K]}}(t) - \int f_t \psi_t^{[K]} \right|^2 \right)^{1/2} \\ & + \int |\psi_t| \mathbf{1}_{\{|\psi_t| > K\}} (c_{\varepsilon,1}(t) + f_t). \end{aligned} \quad (\text{B30})$$

For fixed K , the first term vanishes uniformly in t by Eq. (B21). The second vanishes after $\limsup_{\varepsilon \rightarrow 0}$ and then $K \rightarrow \infty$, by Eq. (B25). $\square$

The logarithmic envelope gives $\int f_t |\log f_t| < \infty$. Apply Lemma 1 to $\log f_t$, and use the reversal identity for A_ε , to obtain

$$M_{\log g}(0) \rightarrow -h(g), \quad M_{\log k}(T) \rightarrow -h(k) \quad \text{in } L^1(A_\varepsilon). \quad (\text{B31})$$

Every additional bounded test in Eq. (34) converges in probability to b_j . A finite union bound suffices; tests at different times need not be independent.

4. Likelihood and probability exponent

Work on the common invariant conull set where the hard-sphere flow is reversible and volume preserving. The exceptional set is null for all absolutely continuous sector laws considered here. Reversal preserves admissibility and partition functions. The identity $\Phi_t R = R \Phi_{-t}$ gives $A_\varepsilon = (\Phi_{-T})_\# Q_{\mu,k}$; its density at Z is the $Q_{\mu,k}$ density at $\Phi_T Z$. Dividing by the density of $Q_{\mu,g}$ at Z , positivity gives equivalence and cancellation of $\mu^N/N!$ gives Eq. (30), including the empty sector. By Eq. (B13),

$$c_\varepsilon = (\log \Xi_g - \log \Xi_{f_0})/\mu = O(\varepsilon). \quad (\text{B32})$$

Equation (B31) now proves $L_\varepsilon/\mu \rightarrow h(g) - h(k) = a$ in $L^1(A_\varepsilon)$. All log expectations are finite by the same moment bounds. Taking expectations and using $\bar{N}/\mu \rightarrow 1$ proves the KL limit and Theorem 2.

The event in Eq. (34) has $p_{\varepsilon,\delta} := A_\varepsilon(E_{\varepsilon,\delta}) \rightarrow 1$. On that event $|L_\varepsilon/\mu - a| \leq 2\delta + |c_\varepsilon|$, so change of measure gives

$$\begin{aligned} e^{-\mu(a+2\delta+|c_\varepsilon|)} p_{\varepsilon,\delta} & \leq B_\varepsilon(E_{\varepsilon,\delta}) \\ & \leq e^{-\mu(a-2\delta-|c_\varepsilon|)} p_{\varepsilon,\delta}. \end{aligned} \quad (\text{B33})$$

Since $\mu^{-1} \log p_{\varepsilon,\delta} \rightarrow 0$, this proves Eq. (35) and its iterated tolerance limit. To obtain a slow diagonal, choose $\varepsilon_j \downarrow 0$ such that for every $\varepsilon \leq \varepsilon_j$ all failures in Eq. (34) at tolerance $1/j$ have total A_ε -probability at most $1/j$. Make the thresholds strictly decreasing, and set $\delta_\varepsilon = 1/j$ on $(\varepsilon_{j+1}, \varepsilon_j]$. Then the acceptance tends to one and Eq. (B33) gives Eq. (36). This existential construction supplies no prescribed rate, completing Theorem 3.

5. Testing and preparation budgets

For $0 \leq \varphi \leq 1$ and real c , change of measure gives

$$\mathbb{E}_{B_\varepsilon} \varphi \geq e^{-c} [\mathbb{E}_{A_\varepsilon} \varphi - A_\varepsilon(L_\varepsilon > c)]_+. \quad (\text{B34})$$

Restrict the expectation to $\{L_\varepsilon \leq c\}$ and bound $\varphi \leq 1$ on the complement. At $c = \mu(a + \delta)$, every feasible test in Eq. (38) therefore has false acceptance at least $e^{-\mu(a+\delta)}(1 - \alpha - o(1))$. Conversely $\varphi = \mathbf{1}_{\{L_\varepsilon \geq \mu(a-\delta)\}}$ is eventually feasible and has false acceptance at most $e^{-\mu(a-\delta)}$. Taking limits and then $\delta \downarrow 0$ proves Corollary 2.

For an event with $q = B(E) \in (0, 1)$, put $p = C(E)$. If $C \ll B$, its KL is infinite. Otherwise the event-indicator chain rule gives the exact extended-real identity

$$\begin{aligned} D(C\|B) & = d(p\|q) + pD(C(\cdot | E)\|B(\cdot | E)) \\ & \quad + (1-p)D(C(\cdot | E^c)\|B(\cdot | E^c)), \end{aligned} \quad (\text{B35})$$

omitting zero-weight conditional terms. Here

$$d(p\|q) = p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q}. \quad (\text{B36})$$

The lower bound in Eq. (39) follows by nonnegativity. The mixture in Eq. (40) sets both conditional KL terms to zero and attains the bound, including $p = 0, 1$.

The exact largest event probability under a finite budget K is

$$\sup_{D(C\|B) \leq K} C(E) = \begin{cases} 1, & K \geq \log(1/q), \\ p_K, & 0 \leq K < \log(1/q), \end{cases} \quad (\text{B37})$$

where $p_K \in [q, 1]$ uniquely solves $d(p_K\|q) = K$. For $p > q$, $\partial_p d(p\|q) = \log[p(1-q)/(q(1-p))] > 0$; at $K = 0$, $p_K = q$.

For the events of Theorem 3, $-\mu^{-1} \log q_\varepsilon \rightarrow a$ and $0 < q_\varepsilon < 1$ eventually. Writing h_2 for binary Shannon entropy gives

$$\begin{aligned} d(p\|q) & = p \log(1/q) - h_2(p) + (1-p) \log(1/(1-q)) \\ & \geq p \log(1/q) - \log 2. \end{aligned} \quad (\text{B38})$$

For fixed p , division by μ gives $d(p\|q_\varepsilon)/\mu \rightarrow pa$. For arbitrary $p_\varepsilon = C_\varepsilon(E_\varepsilon)$, Eqs. (B35) and (B38) prove the budget bound of Corollary 3. It is sharp: the conditional mixture with $p = \min\{1, r/a\}$ attains its limiting coefficient. If an exact finite- ε cap is required instead, use Eq. (B37) at each parameter. The actual law A_ε has

budget $\mu a + o(\mu)$ and acceptance tending to one, and thus reaches the unit-success threshold in this statistical sense.

For Eq. (42), choose $p_\varepsilon = 1/\mu$ in Eq. (40) once $\mu > 1$. Its exact cost equals

$$\frac{\log(1/q_\varepsilon)}{\mu} - h_2(1/\mu) + (1 - 1/\mu) \log \frac{1}{1 - q_\varepsilon} \rightarrow a. \quad (\text{B39})$$

Its success probability $1/\mu$ has zero exponential rate. This is a law on the same preparation space, not an implemented experimental intervention.

6. A nonempty positive-production class

Take

$$\begin{aligned} f_0(x, v) &= p_c(x) q_\Sigma(v), \\ p_c(x) &= Z_c^{-1} e^{-c\sqrt{1+|x|^2}}, \\ q_\Sigma(v) &= \frac{e^{-v^\top \Sigma^{-1} v/2}}{(2\pi)^{d/2} \sqrt{\det \Sigma}}, \end{aligned} \quad (\text{B40})$$

with $c > 0$ and positive definite Σ having two distinct eigenvalues. This smooth density has the required lower envelope, all spatial moments, and the initial cube-summable Gaussian norm of f_0 and $\nabla_x f_0$ for sufficiently small $\beta > 0$.

A positive short-time construction is explicit. Choose $b, K > 0$ with $f_0 \leq K w(x) e^{-b|v|^2}$ and $w(x) = e^{-c\sqrt{1+|x|^2}}$. Choose C_b such that the collision integral against $e^{-b|v_*|^2/2}$ is bounded by $C_b(1 + |v|)$. Set

$$H = 2KC_b + (c + 2KC_b)^2/4, \quad t_0 = \min\{b/2, \log 2/H\}. \quad (\text{B41})$$

Then $\mathcal{W}(t, x, v) = K e^{Ht} w(x) e^{-(b-t)|v|^2}$ satisfies

$$\begin{aligned} \frac{(\partial_t + v \cdot \nabla_x) \mathcal{W}}{\mathcal{W}} &\geq H + |v|^2 - c|v| \\ &\geq 2KC_b(1 + |v|) \geq \frac{Q^+(\mathcal{W}, \mathcal{W})}{\mathcal{W}}. \end{aligned} \quad (\text{B42})$$

Use $w \leq 1$, $K e^{Ht} \leq 2K$, and pair energy conservation for the last step. Start $F^{(0)} = 0$, and define the positive characteristic iteration

$$\begin{aligned} (\partial_t + v \cdot \nabla_x) F^{(n+1)} + \nu_{F^{(n)}} F^{(n+1)} &= Q^+(F^{(n)}, F^{(n)}), \\ F^{(n+1)}(0) &= f_0. \end{aligned} \quad (\text{B43})$$

Equation (B42) gives $0 \leq F^{(n)} \leq \mathcal{W}$. For the norm $d_n(t) = \sup_{x,v} e^{(b-t)|v|^2} |F^{(n+1)} - F^{(n)}|$, differences obey

$$d_{n+1}(t) \leq C \int_0^t [1 + (t-s)^{-1/2}] d_n(s) ds. \quad (\text{B44})$$

Retain nonnegative damping on the left of the difference equation, bound the gain and loss-coefficient differences bilinearly, and use $(1 + |v|) e^{-(t-s)|v|^2} \leq C[1 + (t-s)^{-1/2}]$.

The iterated kernels have summable Gamma-function bounds, proving uniform convergence and uniqueness in this Gaussian class. Dominated convergence in the collision integrals makes every iterate continuous, and the uniform limit is continuous. Spatial difference quotients give the corresponding regularity for the smooth datum; the spatial and derivative norms also propagate by DHM Proposition A.1. Taking $2\beta < b/2$ supplies the required solution bound on this nonempty local interval. Positivity follows from Eq. (B2).

For this example the stronger common envelope $f_t \leq C w(x) e^{-b|v|^2/2}$, together with the logarithmic bound, provides a common integrable majorant for the collision entropy-production integrand:

$$C w(x)^2 e^{-c'(|v|^2 + |v_*|^2)} |v - v_*| (1 + |x| + |v|^2 + |v_*|^2). \quad (\text{B45})$$

Hence

$$\begin{aligned} \mathcal{D}_B[f] &= \frac{1}{4} \int ((v - v_*) \cdot \omega)_+ (f' f'_* - f f_*) \log \frac{f' f'_*}{f f_*} \\ &\quad \times dx dv dv_* d\omega \end{aligned} \quad (\text{B46})$$

is finite and continuous for small t . Spatial and velocity cutoffs followed by a regularized logarithm, with dominated convergence under Eq. (B45), justify

$$h(f_t) - h(f_0) = \int_0^t \mathcal{D}_B[f_s] ds. \quad (\text{B47})$$

The transport boundary flux vanishes under the common spatial envelope.

Choose orthonormal eigenvectors e_1, e_2 of Σ with unequal eigenvalues, and take $v = r e_1$, $v_* = -r e_1$, $\omega = (e_1 - e_2)/\sqrt{2}$, $r > 0$. The outgoing velocities are $r e_2, -r e_2$; the cross section is positive, and the Gaussian products before and after collision are unequal. By continuity the nonnegative integrand is positive on a set of positive measure. Thus $\mathcal{D}_B[f_0] > 0$ and Eq. (B47) gives $a > 0$ on a nonzero short-time interval. This establishes nonemptiness, not existence on an arbitrary prescribed large interval.

Appendix C: Proof of the finite-resolution readout results

1. Inputs and the endpoint moment concentration

We use the hypotheses of Theorem 2 and the additional continuity of g, k stated in Sec. V. For $u = g$ at time zero and $u = k$ at time T , its endpoint argument gives

$$|\log u| \leq C_u W, \quad w_u := \int W u dz < \infty, \quad (\text{C1})$$

$$M_W \rightarrow w_u \quad \text{in } L^1(A_\varepsilon), \quad (\text{C2})$$

$$M_{1_C} \rightarrow \int_C u dz \quad \text{in } A_\varepsilon\text{-probability} \quad (\text{C3})$$

for every fixed measurable cell C . In particular, a spatially uniform Gaussian velocity cap is not being integrated over all space. The spatial integrability and weighted-tail bounds are those established in the endpoint proof.

For clarity, the extension from bounded observations to Eq. (C2) uses uniform integrability, not a two-particle estimate for an unbounded weight. With $W_M = W \wedge M$, the first two correlation functions imply concentration of M_{W_M} for fixed M . Moreover,

$$\begin{aligned} & \mathbb{E}_{A_\varepsilon} |M_W - w_u| \\ & \leq \mathbb{E}_{A_\varepsilon} \left| M_{W_M} - \int W_M u \, dz \right| \\ & \quad + \mathbb{E}_{A_\varepsilon} M_{W-W_M} + \int (W - W_M) u \, dz. \end{aligned} \quad (C4)$$

The first term vanishes at fixed M , and the other two vanish after taking $\limsup_{\varepsilon \rightarrow 0}$ and then $M \rightarrow \infty$. This is the weighted-tail input already proved for the actual reversed ensemble. Joint concentration of finitely many endpoint observations follows by a finite union bound.

The exact likelihood and its concentration are

$$\frac{L_\varepsilon}{\mu} = c_\varepsilon + M_{\log k}(T) - M_{\log g}(0), \quad (C5)$$

$$c_\varepsilon = O(\varepsilon),$$

$$L_\varepsilon / \mu \longrightarrow a \quad \text{in } L^1(A_\varepsilon), \quad (C6)$$

$$D(A_\varepsilon \| B_\varepsilon) / \mu \longrightarrow a. \quad (C7)$$

The two laws are equivalent, so for every measurable event F ,

$$B_\varepsilon(F) = \mathbb{E}_{A_\varepsilon} [e^{-L_\varepsilon} \mathbf{1}_F]. \quad (C8)$$

These are the only dynamical inputs to the readout construction.

2. A pointwise certificate for the omitted particles

Fix one endpoint profile u . Choose a compact rectangle K and finitely many rectangular cells $C_1, \dots, C_J$ partitioning K ; assign boundaries to a single cell. With $\zeta_j \in C_j$, let

$$w_j^- = \inf_{C_j} W \geq 1, \quad \ell_j = \log u(\zeta_j). \quad (C9)$$

Suppose that $\sup_{C_j} |\log u - \ell_j| \leq \omega_u$ for every cell. Select a fixed range

$$H > \max(w_g, w_k) + 1 \quad (C10)$$

and a step $0 < s < 1$. The quantizer reports $Q_s(y) = s \lfloor y/s \rfloor$ for $0 \leq y \leq H$ and an overflow symbol for $y > H$. Set

$$\begin{aligned} y_j &= M_{\mathbf{1}_{C_j}}, & y_W &= M_W, \\ q_j &= Q_s(y_j), & q_W &= Q_s(y_W). \end{aligned} \quad (C11)$$

Each channel has at most $\lfloor H/s \rfloor + 2$ symbols, including overflow, independently of ε . Reject any overflow record. For all other records, $q_j \leq y_j < q_j + s$ and $q_W \leq y_W < q_W + s$. Consequently

$$\begin{aligned} 0 &\leq M_{W \mathbf{1}_{K^c}} \leq M_W - \sum_j w_j^- y_j \\ &\leq q_W + s - \sum_j w_j^- q_j =: r_u. \end{aligned} \quad (C12)$$

This is a statement for each configuration, including configurations with a highly energetic outlier. It does not assert that outliers are exponentially unlikely under B_ε .

Let $\hat{m}_u = \sum_j \ell_j q_j$ and require $r_u \leq \tau_u$. Split the logarithmic sum into the compact part, its complement, and the rounding error. Equations (C1) and (C12) give

$$\begin{aligned} |M_{\log u} - \hat{m}_u| &\leq \omega_u M_{\mathbf{1}_K} + C_u M_{W \mathbf{1}_{K^c}} + s \sum_j |\ell_j| \\ &\leq e_u := \omega_u(H + s) + C_u \tau_u + s \sum_j |\ell_j|. \end{aligned} \quad (C13)$$

Here $M_{\mathbf{1}_K} \leq M_W < H + s$. Thus the count cap needed for rounding is supplied by the global moment channel itself.

3. Fixed parameter choices and typical acceptance

We show that e_u can be made smaller than any prescribed $b > 0$, while the certificate has A_ε -probability tending to one. First choose $\tau_u > 0$ with $C_u \tau_u < b/4$. By integrability, choose K so that

$$\int_{K^c} W u \, dz < \tau_u/4. \quad (C14)$$

Continuity and positivity of u imply uniform continuity of $\log u$ on K . Refine its rectangular grid until

$$\begin{aligned} \max_j \sup_{C_j} (W - w_j^-) &< \tau_u/4, \\ \omega_u &< \frac{b}{4(H+1)}. \end{aligned} \quad (C15)$$

Finally, with this finite grid fixed, choose $0 < s < 1$ satisfying

$$s \left(1 + \sum_j w_j^- \right) < \tau_u/4, \quad s \sum_j |\ell_j| < b/4. \quad (C16)$$

These choices give $e_u < 3b/4 < b$.

Let $p_j = \int_{C_j} u \, dz$. By Eqs. (C2)–(C3), the unrounded residual converges in probability to

$$\begin{aligned} r_u^0 &:= w_u - \sum_j w_j^- p_j \\ &= \int_{K^c} W u \, dz + \sum_j \int_{C_j} (W - w_j^-) u \, dz < \tau_u/2. \end{aligned} \quad (C17)$$

On any nonoverflow record,

$$r_u \leq M_W - \sum_j w_j^- y_j + s \left(1 + \sum_j w_j^- \right). \quad (\text{C18})$$

The strict margins in Eqs. (C16)–(C17) show that $r_u \leq \tau_u$ with probability tending to one. Overflow also has vanishing probability, since $M_W \rightarrow w_u < H$ and $y_j \leq M_W$. The floor outputs themselves need not converge to a single code; the margins make quantization boundaries harmless.

This construction applies to both endpoints. All parameters are fixed before taking $\varepsilon \rightarrow 0$. It yields no quantitative estimate of the number of cells required as a function of b , or of the particle number at which the limiting performance becomes accurate.

4. Testing exponent and observable information

Let $\mathcal{H}_\varepsilon$ denote passage of both endpoint certificates. Write $e = e_g + e_k$ and choose $\zeta > 0$ with $2e + \zeta < \eta$. The fixed acceptance rule is

$$F_{\varepsilon,\eta} = \mathcal{H}_\varepsilon \cap \{\hat{m}_k(T) - \hat{m}_g(0) \geq a - e - \zeta\}. \quad (\text{C19})$$

On $\mathcal{H}_\varepsilon$,

$$\left| \frac{L_\varepsilon}{\mu} - c_\varepsilon - \hat{m}_k(T) + \hat{m}_g(0) \right| \leq e. \quad (\text{C20})$$

Since $A_\varepsilon(\mathcal{H}_\varepsilon) \rightarrow 1$, likelihood concentration and the positive ζ margin give $A_\varepsilon(F_{\varepsilon,\eta}) \rightarrow 1$. On every accepted record,

$$L_\varepsilon/\mu \geq a - 2e - \zeta - |c_\varepsilon|. \quad (\text{C21})$$

Changing measure,

$$B_\varepsilon(F_{\varepsilon,\eta}) \leq e^{-\mu(a-2e-\zeta-|c_\varepsilon|)} A_\varepsilon(F_{\varepsilon,\eta}). \quad (\text{C22})$$

This proves the lower bound on the exponent.

For the sharper converse, fix $d > 0$ and define $G_\varepsilon = \{L_\varepsilon/\mu \leq a + d\}$. Then

$$B_\varepsilon(F_{\varepsilon,\eta}) \geq e^{-\mu(a+d)} A_\varepsilon(F_{\varepsilon,\eta} \cap G_\varepsilon). \quad (\text{C23})$$

The final factor tends to one. Taking logarithms, then $\limsup_{\varepsilon \rightarrow 0}$, and finally $d \downarrow 0$, proves Theorem 4. The upper exponent is a , rather than $a + \eta$, because any event typical under A_ε obeys this converse.

For Corollary 4, put $p = A_\varepsilon(F_{\varepsilon,\eta})$ and $q = B_\varepsilon(F_{\varepsilon,\eta})$. Data processing and the binary relative-entropy bound imply

$$\begin{aligned} D(A_\varepsilon \| B_\varepsilon) &\geq D(A_\varepsilon^Y \| B_\varepsilon^Y) \geq d(p \| q), \\ d(p \| q) &\geq p \log(1/q) - \log 2. \end{aligned} \quad (\text{C24})$$

Combine $p \rightarrow 1$, Theorem 4, and Eq. (C7).

The constructed deterministic test is eventually feasible for every fixed acceptance requirement $1 - \alpha$. Conversely, every test $0 \leq \varphi(Y) \leq 1$ satisfying that requirement obeys

$$\begin{aligned} \mathbb{E}_{B_\varepsilon} \varphi(Y) &\geq e^{-\mu(a+d)} \mathbb{E}_{A_\varepsilon} [\varphi(Y) \mathbf{1}_{G_\varepsilon}] \\ &\geq e^{-\mu(a+d)} [1 - \alpha - A_\varepsilon(G_\varepsilon^c)]. \end{aligned} \quad (\text{C25})$$

This lower bound is uniform over tests. Taking the infimum proves the claimed optimal exponent bounds, for randomized tests and also for the restricted deterministic class.

5. Guaranteed intervals and shared random readout

Suppose that each nonoverflow output reports intervals

$$0 \leq l_j \leq y_j \leq u_j, \quad 0 \leq l_W \leq M_W \leq u_W \leq H_*. \quad (\text{C26})$$

Let d_j and d_W bound their widths, and set $c_j = (l_j + u_j)/2$, $r_j = (u_j - l_j)/2$. Then, record by record,

$$M_{W \mathbf{1}_{K^c}} \leq M_W - \sum_j w_j^- y_j \leq u_W - \sum_j w_j^- l_j =: \tau_u. \quad (\text{C27})$$

If $\tau_u \leq \tau_u$, the same compact/tail decomposition as before gives

$$\begin{aligned} \left| M_{\log u} - \sum_j \ell_j c_j \right| &\leq \omega_u u_W + C_u \tau_u + \sum_j |\ell_j| r_j \\ &\leq \omega_u H_* + C_u \tau_u + \frac{1}{2} \sum_j |\ell_j| d_j. \end{aligned} \quad (\text{C28})$$

For typicality, the observable residual satisfies

$$\tau_u \leq M_W - \sum_j w_j^- y_j + d_W + \sum_j w_j^- d_j. \quad (\text{C29})$$

Choose the box and grid so that the unrounded residual has limit below $\tau_u/2$, then require $d_W + \sum_j w_j^- d_j < \tau_u/4$. Provided nonoverflow probability tends to one, the certificate is typical. Smaller widths reduce the last error in Eq. (C28). The proof from Eq. (C19) onward now applies with these deterministic endpoint error bounds. Arbitrarily small target loss requires correspondingly adequate calibrated error budgets; it is not a conclusion for an arbitrary fixed noisy device.

Nonsaturation is a separate condition: a device that always reports overflow satisfies every conditional interval guarantee but has no acceptance power. One sufficient implementation contract is a fixed readout range with the limiting moment and bin values strictly in its interior, bounded pre-quantization errors smaller than that margin, and finite quantization. Enlarge each quantization interval by the calibrated error bound and clip negative lower endpoints to zero. Overflow then implies that a true

input has left a fixed neighborhood of its limit, so its probability vanishes by Eqs. (C2)–(C3).

For randomized readout, let the same Markov kernel $K_\varepsilon(dy | Z)$ be applied under both preparations. On the joint physical-state and readout space, define

$$\begin{aligned}\tilde{A}_\varepsilon(dZ, dy) &= A_\varepsilon(dZ)K_\varepsilon(dy | Z), \\ \tilde{B}_\varepsilon(dZ, dy) &= B_\varepsilon(dZ)K_\varepsilon(dy | Z).\end{aligned}\quad (\text{C30})$$

For every bounded measurable f ,

$$\int f d\tilde{A}_\varepsilon = \int f(Z, y) e^{L_\varepsilon(Z)} d\tilde{B}_\varepsilon, \quad (\text{C31})$$

so $d\tilde{A}_\varepsilon/d\tilde{B}_\varepsilon = e^{L_\varepsilon(Z)}$. The physical-state marginal under $\tilde{A}_\varepsilon$ is unchanged; therefore likelihood concentration and its mean limit also remain unchanged. If the interval guarantee holds almost surely on every nonoverflow joint record, all pointwise bounds apply there and the preceding change-of-measure proof is valid. The kernel may depend on Z , and its channel errors need not be independent.

The interval guarantee cannot be replaced by an unflagged failure probability merely $o(1)$. For instance, a common kernel that occasionally reports an accepted code regardless of the input can produce subexponentially rare false positives. Such a code violates the required error certificate on the corresponding joint records. A separate exponentially adequate bound would be needed to include these failures.

6. Exact sufficiency and one-bit relative entropy

To prove Proposition 1, write $L_\varepsilon = \gamma_\varepsilon + S$, where $\gamma_\varepsilon = \mu c_\varepsilon$ is deterministic. For a bounded measurable function f ,

$$\begin{aligned}\int f(s) d(S_{\#}A_\varepsilon)(s) &= \int f(S) e^{\gamma_\varepsilon + S} dB_\varepsilon \\ &= \int f(s) e^{\gamma_\varepsilon + s} d(S_{\#}B_\varepsilon)(s).\end{aligned}\quad (\text{C32})$$

Thus the pushforward likelihood is exactly $e^{\gamma_\varepsilon + s}$. Integrating its logarithm proves Eq. (58); the reverse KL identity holds as an extended-real identity as well. On this standard Borel configuration space, the conditional laws also satisfy

$$A_\varepsilon(\cdot | S) = B_\varepsilon(\cdot | S) \quad (\text{C33})$$

for almost every score under either law, because the density factor is constant after conditioning on S .

For the events of Theorem 3, let $p = A_\varepsilon(E_\varepsilon) \rightarrow 1$ and $q = B_\varepsilon(E_\varepsilon)$ with $-\mu^{-1} \log q \rightarrow a > 0$. The exact binary expression is

$$d(p||q) = p \log(1/q) - h_2(p) + (1-p) \log \frac{1}{1-q}. \quad (\text{C34})$$

The first term divided by μ tends to a ; $0 \leq h_2(p) \leq \log 2$; and the last term is bounded by $-\log(1-q) \rightarrow 0$. This proves Eq. (59). Combining it with Eq. (C7) shows that the information discarded by this particular bit is $o(\mu)$. The assertion concerns leading relative entropy, not microscopic state reconstruction or vanishing total-variation error between conditional microstate ensembles.

7. Conditioning on a narrow likelihood band

For completeness, the conditioning statement follows from a finite parameter estimate, without any conditional uniform-integrability assumption. Let E have positive probability under both equivalent laws, write $p = A(E)$ and $q = B(E)$, and suppose pointwise on E that

$$\mu l \leq L \leq \mu u, \quad l \leq u. \quad (\text{C35})$$

Changing measure on E gives $e^{-\mu u} \leq q/p \leq e^{-\mu l}$. The conditional density ratio is therefore bounded pointwise:

$$\begin{aligned}\log \frac{dA_E}{dB_E} &= L + \log(q/p), \\ \left| \log \frac{dA_E}{dB_E} \right| &\leq \mu(u-l).\end{aligned}\quad (\text{C36})$$

Integrating the signed logarithm under either conditional law yields

$$D(A_E||B_E) \leq \mu(u-l), \quad D(B_E||A_E) \leq \mu(u-l). \quad (\text{C37})$$

Exponentiating Eq. (C36) and integrating over G proves the two event inequalities as well. For the shrinking-band events of Theorem 3, the pointwise band width $\omega_\varepsilon = u_\varepsilon - l_\varepsilon$ tends to zero. This proves Corollary 6. A one-sided likelihood lower bound alone supplies no such conclusion.

Vanishing conditional KL density and a subexponential density ratio do not imply convergence in total variation. For example, the two Bernoulli laws with parameters $1 - e^{-\sqrt{\mu}}$ and $e^{-\sqrt{\mu}}$ have logarithmic density ratios of order $\sqrt{\mu} = o(\mu)$ and both KL densities vanish, while their total variation distance tends to one. This example identifies the statistical strength of the conclusion; it is not an additional hard-sphere construction.

- arXiv:2408.07818 [math.AP].
- [2] H. Spohn, Loschmidt's reversibility argument and the H-theorem, in *Pioneering Ideas for the Physical and Chemical Sciences*, edited by W. Fleischhacker and T. Schönfeld (Springer, 1997) pp. 153–157.
 - [3] T. Bodineau, I. Gallagher, L. Saint-Raymond, and S. Simonella, One-sided convergence in the Boltzmann-Grad limit (2016), arXiv:1612.03722.
 - [4] H. Spohn, Microscopic time reversibility and the Boltzmann equation, in *Chance in Physics*, Lecture Notes in Physics, Vol. 574 (Springer, 2001) pp. 55–59.
 - [5] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck, Dissipation: The phase-space perspective, *Physical Review Letters* **98**, 080602 (2007).
 - [6] A. Gomez-Marín, J. M. R. Parrondo, and C. Van den Broeck, Lower bounds on dissipation upon coarse graining, *Physical Review E* **78**, 011107 (2008).
 - [7] T. Bodineau, I. Gallagher, L. Saint-Raymond, and S. Simonella, Statistical dynamics of a hard sphere gas: fluctuating Boltzmann equation and large deviations (2020), arXiv:2008.10403.
 - [8] F. Bouchet, Is the Boltzmann equation reversible? a large deviation perspective on the irreversibility paradox (2020), arXiv:2002.10398.
 - [9] S. Saunders, The Gibbs paradox, *Entropy* **20**, 552 (2018).
 - [10] A. E. Allahverdyan and T. M. Nieuwenhuizen, Explanation of the Gibbs paradox within the framework of quantum thermodynamics, *Physical Review E* **73**, 066119 (2006).
 - [11] F. Paillusson, The Real Gibbs paradox and a composition-based resolution, *Entropy* **25**, 833 (2023).