

Kroemer’s Bose–Fermi partition-function bound beyond the continuum classical regime: exact discrete-spectrum bounds, prefactor corrections, and an orbital-diagonal interacting extension

scientific-author-59*

September 7, 2026

Abstract

Counting the states of an ideal gas as though its particles were distinguishable gives a canonical partition function that lies between the exact bosonic and the exact fermionic one, so the weight carried by configurations with a doubly occupied one-particle state is precisely the largest error that counting can make. Kroemer bounded that weight by a fixed multiple of either exact partition function, for a spinless non-interacting monatomic gas in a continuum box and in a system large enough that removing two particles does not move the chemical potential. We show that a bound of the same shape holds exactly, with no large-system hypothesis, for every finite discrete one-particle spectrum, every temperature and every particle number of at least two, with degeneracies entering as repeated levels. The prefactor is Kroemer’s with the square of the particle number replaced by that number times one less, and that replacement is the whole of the small-system correction: without it the lower side of the bound misses by a relative amount of order the reciprocal of the particle number, however dilute the gas. Since the classical value is trapped between the two exact ones, every bound of this kind transfers to the logarithm of the partition function, and so to the free energy. Both slacks of the prefactor interval are then decomposed into explicitly non-negative series, producing a monotone hierarchy of upper endpoints and a matching lower endpoint whose relative error in the dilute three-dimensional limit falls as the inverse square of the diluteness instead of its inverse, with closed-form leading coefficients. Pair energies that are diagonal in a fixed single-particle basis are included at arbitrary coupling strength. Interactions that mix orbitals, and interacting fermions whose pair term does not vanish by exclusion, remain outside the result.

1 Introduction

A canonical gas of N identical, non-interacting particles has three natural partition functions. Counting the many-body states as though the particles were distinguishable and then dividing by $N!$ gives the classical, or Maxwell–Boltzmann, value $Z_1(\beta)^N/N!$, where Z_1 is the one-particle partition function and β the inverse temperature. Counting them in the symmetric and in the antisymmetric sector gives the exact bosonic value B_N and the exact fermionic value F_N . The three are not comparable by accident: as Lemma 2.2 shows, the classical value always sits between the two exact ones. The difference

$$M_N = B_N - F_N \tag{1}$$

*Corresponding author: `scientific-author-59`. Solution repository: `scientific-author-59/kroemer-bose-fermi-discrete-bound`.

is therefore the full width of the window inside which classical counting has to land, and it has a direct meaning: it is the total thermal weight of those occupation configurations in which some one-particle state carries two or more particles. Asking how incorrect the classical partition function is amounts to asking how large M_N can be relative to the exact values that bracket it.

Kroemer answered that question for the dilute ideal gas [3]. For a spinless, non-interacting, monatomic gas in a continuum box, and in a system large enough that removing two particles does not change the chemical potential, he obtained the two-sided bound

$$f B_N > M_N > f F_N, \quad f = \frac{1}{2^{3/2}} \frac{N^2}{n_Q V}, \quad (2)$$

where n_Q is the quantum concentration and V the volume, so that $n_Q V = Z_1(\beta)$ and the dimensionless product $n_Q V$ measures how dilute the gas is. The bound is a statement about a prefactor: the same number f multiplies the bosonic partition function from above and the fermionic one from below. Four restrictions are built into it — no internal degeneracy, no interactions, a continuum of one-particle levels in a macroscopic box, and a particle number large enough for the chemical potential to be insensitive to the loss of two particles — and the published problem that this paper addresses asks how much of the bound survives when they are lifted [2].

Each of the four restrictions excludes a situation in which the question is worth asking. Trapped and mesoscopic systems — atoms in an optical or magnetic trap, electrons in a quantum dot, vibrational ladders of a molecule — have genuinely discrete one-particle spectra, and for them the parameter that controls quantum statistics is not the volume but the number of orbitals within a few $k_B T$ of the ground state, which can be small. Internal spin or orbital degeneracy multiplies every thermal sum by the degeneracy and so changes the prefactor itself. Interactions are what make double occupancy physically interesting in the first place: an on-site repulsion of Hubbard type suppresses it, and a bound on the weight of doubly occupied configurations is a bound on that suppression. And the fourth restriction is the sharpest one, because at fixed particle number the replacement of N^2 by $N(N-1)$ is exactly the difference between an inequality that holds and one that does not (Proposition 5.5).

This paper proves that a bound of Kroemer's shape holds with no large-system hypothesis at all. For every finite discrete one-particle spectrum, degeneracies included as repeated levels, every inverse temperature and every $N \geq 2$, the multiple-occupancy part is trapped between two spectrum-independent expressions (Theorem 4.1), and the prefactor form holds with

$$f_N = \frac{Z_1(2\beta)}{Z_1(\beta)^2} N(N-1) \quad (3)$$

in place of f (Theorem 5.3). In the spinless three-dimensional continuum limit, where $Z_1(2\beta) = 2^{-3/2} Z_1(\beta)$, f_N is Kroemer's prefactor multiplied by $(N-1)/N$; the small-system correction is thus isolated exactly rather than estimated, and it is the only correction needed. The one-particle quantity $Z_1(2\beta)/Z_1(\beta)^2$ is what carries the spectral information: for a discrete spectrum it replaces the continuum ratio $2^{-3/2}/(n_Q V)$ and remains meaningful when the level spacing is comparable to $k_B T$, when the spectrum is a trap, or when a degeneracy multiplies it.

The two sides of the prefactor interval are then sharpened, and this is where most of the technical content lies. The upper endpoint $f_N B_N$ differs from M_N for two separate reasons, and both have exact, explicitly non-negative decompositions: the injection that produces the bound $M_N \leq Z_1(2\beta) B_{N-2}$ overcounts by a series whose coefficients are determined by the even thermal sums $Z_1(2k\beta)$ alone (Theorem 6.1), and the passage from that bound to the prefactor form loses a further non-negative series (Theorem 6.4). Subtracting as much of the two series as one likes

therefore produces a monotone hierarchy of upper endpoints (Corollary 6.6). Retaining the first two terms of each gives an endpoint $U_N^{(2)}$ whose relative error in the dilute three-dimensional limit falls as the inverse square of the diluteness rather than as its inverse, with a closed-form coefficient (Theorem 7.3). The lower side has a matching endpoint $L_N^{(2)}$ obtained from a single coefficientwise inequality for the exponential, with the same inverse-square relative error (Theorem 7.2), and its own prefactor slack is certified non-negative coefficientwise in the monomial basis (Proposition 7.4).

Interactions enter through the simplest class that still changes the physics: pair energies u_i attached to the double occupancy of a fixed orbital, so that a configuration with occupation ν_i in orbital i carries the additional energy $\sum_i u_i \binom{\nu_i}{2}$. Because the exclusion principle makes every fermionic occupation zero or one, the fermionic partition function is untouched by such a term, which is what allows the difference M_N to remain a well-posed object. The same injection that gives the free upper bound survives at arbitrary coupling strength, with a prefactor that becomes $\sum_i e^{-\beta(2\varepsilon_i + u_i)}$ for repulsive couplings (Theorem 8.1), and restricting the injection to fermionic images gives a lower bound that holds for attractive couplings as well (Theorem 8.2). Both reduce to the free bounds when the couplings vanish, and both are exact at $N = 2$ and $N = 3$. What does not survive is the coefficientwise factorisation on which the free lower bound rests: its cubic coefficient turns strictly negative under repulsion, so the free route has no direct interacting analogue and the restricted injection is the substitute (Proposition 8.3).

The mechanism behind all of this is small enough to state in one paragraph. The generating functions of B_N and F_N differ by an exponential $B(x) = F(x)e^{\Delta(x)}$ whose exponent Δ contains only even powers of x with non-negative coefficients (Section 2). Every lower bound in this paper comes from truncating $e^\Delta - 1$ coefficientwise, and every upper bound comes from a single injection that sends a bosonic configuration with a double occupancy to the pair formed by its first doubly occupied orbital and the configuration left after those two particles are removed. The rest of the paper is the bookkeeping of what those two moves leave over, and of what the leftovers do in the dilute limit. Since the classical value always lies inside $[F_N, B_N]$, the same bookkeeping bounds what reaches a thermodynamic observable: the absolute error of $\log Z$, and hence of the free energy, is at most $\log(1 + M_N/F_N)$ (Corollary 2.3).

The theorems are the evidence; the computation reported in Section 9 is corroboration. Every identity and inequality was re-evaluated in exact deterministic arithmetic at 200-bit precision over a grid of six one-particle spectra — including a spin-degenerate spectrum and a spectrum with no structure at all — six inverse temperatures and particle numbers up to eight, together with the model continuum spectrum at degeneracies one, two and three and dilutenesses up to 10^8 , and an interacting grid with attractive, repulsive and mixed couplings. No quantity in this paper is sampled, fitted or extrapolated from a stochastic computation, and no error bar appears anywhere; where an asymptotic coefficient is quoted, it is quoted in closed form and then measured against the exact recurrences. The derivation notes and the deterministic implementation from which this consolidation was made are archived as reference [1].

Three boundaries should be visible from the start. First, the interacting results require the interaction to be diagonal in a fixed single-particle basis and to act on the occupancy of each orbital separately; two-body matrix elements that mix orbitals, and interactions on fermions that do not vanish by exclusion, are open. Second, the continuum statements are asymptotics of the model spectrum $Z_1(k\beta) = s k^{-3/2}$ at fixed N and large diluteness $s = n_Q V$; they are not derived from a finite box, and the identity $Z_1(2\beta) = 2^{-3/2} Z_1(\beta)$ that they use is itself a large-volume approximation. Third, the lower bounds are dilute-regime statements: as the temperature drops towards a unique ground state the upper bound becomes exact while the lower one retains only the fraction $2/N$ of M_N for even N and $2/(N - 1)$ for odd N (Proposition 4.2).

Section 2 fixes the model and the notation. Section 3 states precisely what is being extended. Sections 4–7 carry the ideal-gas results: the two universal bounds, the exact prefactor, the decomposition of both slacks and the second-order endpoints. Section 8 treats on-site interactions. Section 9 reports the deterministic verification, Section 10 collects the scope and the physical reading, and Section 11 summarises. Appendices A–D hold the derivation of the canonical recurrences and the symmetric-function background, the continuum expansions to three subleading orders, the interacting weight calculus, and the full verification protocol.

2 The model: discrete spectra and canonical counting

2.1 Spectrum, configurations and the three partition functions

Fix a finite list of one-particle energies

$$\varepsilon_1 \leq \varepsilon_2 \leq \cdots \leq \varepsilon_K, \quad K \geq 1, \quad (4)$$

and an inverse temperature $\beta > 0$. Repeated entries are degeneracies, so a spin or an orbital multiplicity is not a separate structure: a particle with g internal states living on a spatial level ϵ contributes g copies of ϵ to (4). This is the sense in which everything below applies to degenerate particles without further hypotheses. Write

$$x_i = e^{-\beta \varepsilon_i}, \quad p_k = Z_1(k\beta) = \sum_{i=1}^K x_i^k = \sum_{i=1}^K e^{-k\beta \varepsilon_i}, \quad k = 1, 2, \dots, \quad (5)$$

so that $p_1 = Z_1(\beta)$ is the ordinary one-particle partition function and p_k is the same sum evaluated at k times the inverse temperature. Only energy differences enter any relative statement below, and p_2/p_1^2 is invariant under a shift of the zero of energy; a shift rescales B_N , F_N and M_N by the common factor $e^{-\beta N \varepsilon_1}$.

An occupation configuration is a vector $\nu = (\nu_1, \dots, \nu_K)$ of non-negative integers with $|\nu| = \sum_i \nu_i = N$, and its one-body energy is $E(\nu) = \sum_i \varepsilon_i \nu_i$. The three partition functions are

$$B_N = \sum_{|\nu|=N} e^{-\beta E(\nu)}, \quad F_N = \sum_{\substack{|\nu|=N \\ \nu_i \in \{0,1\}}} e^{-\beta E(\nu)}, \quad Z_N^{\text{MB}} = \frac{p_1^N}{N!}, \quad (6)$$

for bosons, for fermions and for classical Maxwell–Boltzmann counting respectively. Both exact values are symmetric polynomials in the weights x_i : $B_N = h_N(x)$ is the complete homogeneous one and $F_N = e_N(x)$ the elementary one. The object of this paper is

$$M_N = B_N - F_N = \sum_{\substack{|\nu|=N \\ \max_i \nu_i \geq 2}} e^{-\beta E(\nu)} \geq 0, \quad (7)$$

the total thermal weight of the configurations that place at least two particles in the same one-particle state. Throughout,

$$B_r = F_r = e_r = h_r = 0 \quad (r < 0 \text{ or } r > K), \quad B_0 = F_0 = e_0 = h_0 = 1, \quad (8)$$

so that $F_N = 0$ automatically once N exceeds the number of orbitals and $M_N = B_N$ there.

Two standard facts are used constantly, and both are proved in Appendix A from the generating functions.

Lemma 2.1 (canonical recurrences). *For every $n \geq 1$,*

$$nB_n = \sum_{k=1}^n p_k B_{n-k}, \quad nF_n = \sum_{k=1}^n (-1)^{k+1} p_k F_{n-k}. \quad (9)$$

Proof. Appendix A: both follow from comparing coefficients in the logarithmic derivatives of the two products in (13), and neither uses anything beyond the finiteness of the spectrum. $\square$

The two recurrences differ only in the sign of the odd- k terms, which is the combinatorial shadow of symmetrisation. They are the reason a discrete spectrum is computationally no harder than a continuum one: the entire canonical thermodynamics of N particles follows from the N numbers $Z_1(\beta), \dots, Z_1(N\beta)$.

Lemma 2.2 (the classical value is bracketed). *For every $N \geq 0$ and every spectrum,*

$$F_N \leq Z_N^{\text{MB}} = \frac{p_1^N}{N!} \leq B_N. \quad (10)$$

Consequently M_N bounds the error of classical counting against either exact value: $|Z_N^{\text{MB}} - B_N| \leq M_N$ and $|Z_N^{\text{MB}} - F_N| \leq M_N$.

Proof. Compare the three generating functions coefficientwise. For $t \geq 0$, $1 + t \leq e^t \leq (1 - t)^{-1}$, and multiplying over $t = x_i z$ gives

$$F(z) = \prod_{i=1}^K (1 + x_i z) \leq \exp(p_1 z) \leq \prod_{i=1}^K (1 - x_i z)^{-1} = B(z) \quad (11)$$

as an inequality between series with non-negative coefficients. The middle series has coefficient $p_1^N / N!$ at z^N . $\square$

Lemma 2.2 is what makes M_N the right object to bound. It is not an estimate of the classical error in one direction; it is the sharp common error bar for both directions at once, and any bound on M_N in terms of B_N or F_N converts immediately into a bound on the classical partition function.

The same window also says how much of that error reaches a thermodynamic observable, and this is the point on which the physical reading of the bound turns.

Corollary 2.3 (error that reaches the logarithm). *Let $F_N > 0$. For every finite spectrum, every $\beta > 0$ and every $N \geq 2$,*

$$0 \leq \log B_N - \log Z_N^{\text{MB}} \leq \log\left(1 + \frac{M_N}{F_N}\right), \quad 0 \leq \log Z_N^{\text{MB}} - \log F_N \leq \log\left(1 + \frac{M_N}{F_N}\right). \quad (12)$$

Consequently every upper bound on M_N in terms of F_N bounds the absolute error that classical counting makes in $\log Z$, and hence in the free energy $-\beta^{-1} \log Z$. When $F_N = 0$, which by (8) happens exactly for $N > K$, the left inequalities still hold and the right ones are vacuous.

Proof. $\log$ is increasing, so (10) gives both left inequalities, and each difference is bounded by their sum, $\log B_N - \log F_N = \log(B_N / F_N) = \log(1 + M_N / F_N)$. $\square$

The cap is close to sharp rather than a crude consequence of monotonicity: on the verification grid of Section 9 the larger of the two errors in (12) reaches 0.992 of it, for a two-level spectrum at $\beta = 0.25$ and $N = 3$.

In the dilute model spectrum at fixed N , Theorem 7.3 below gives $M_N/F_N = f_N(1 + o(1))$, so the absolute error of $\log Z$ is $N(N-1)(2^{3/2}s)^{-1} + o(s^{-1})$, a quantity that vanishes as the gas dilutes, while $\log Z_N^{\text{MB}} = N \log p_1 - \log N!$ grows like $N \log s$. The two requirements are therefore of different strength: the interval $[f_N F_N, f_N B_N]$ carries information only when $f_N \ll 1$, whereas the bound on the logarithm is small compared with the logarithm itself as soon as $f_N \ll N \log s$. This is the precise sense in which an estimate of the partition function can be loose and the corresponding statement about the free energy still be tight.

2.2 Generating functions and the even exponent

Let

$$B(z) = \sum_{N \geq 0} B_N z^N = \prod_{i=1}^K \frac{1}{1 - x_i z}, \quad F(z) = \sum_{N \geq 0} F_N z^N = \prod_{i=1}^K (1 + x_i z). \quad (13)$$

Taking logarithms and expanding,

$$\log B(z) - \log F(z) = \sum_{i=1}^K \left[-\log(1 - x_i z) - \log(1 + x_i z) \right] = \sum_{i=1}^K \sum_{k \geq 1} \frac{x_i^{2k}}{k} z^{2k} = \sum_{k \geq 1} \frac{p_{2k}}{k} z^{2k} =: \Delta(z). \quad (14)$$

The odd powers cancel exactly. Hence

$$B(z) = F(z) e^{\Delta(z)}, \quad \sum_{N \geq 0} M_N z^N = B(z) - F(z) = F(z)(e^{\Delta(z)} - 1), \quad (15)$$

and the structural fact that drives every lower bound in this paper is that Δ has non-negative coefficients and only even powers. Truncating $e^\Delta - 1$ coefficientwise can only lose weight, never gain it, and the loss is explicit.

2.3 The continuum model spectrum and the diluteness parameter

For a spinless monatomic gas of mass m in a three-dimensional box of volume V , the thermal wavelength is $\lambda = h/\sqrt{2\pi m k_B T}$, the quantum concentration is $n_Q = \lambda^{-3} = (m k_B T / 2\pi \hbar^2)^{3/2}$, and in the continuum (large-volume) approximation

$$Z_1(k\beta) = \frac{V}{\lambda(k\beta)^3} = k^{-3/2} n_Q V, \quad \text{that is} \quad p_k = s k^{-3/2}, \quad s := n_Q V. \quad (16)$$

With an internal degeneracy g the right-hand side is multiplied by g . The dimensionless number s is the diluteness: $s \gg 1$ means many thermal wavelengths fit in the box per particle, which is the classical regime, and $s \ll 1$ means the gas is quantum degenerate. Equation (16) is an asymptotic statement about a large box, not an identity for a finite one; the paper uses it only inside limits that are labelled as such, and every exact statement is made for a finite list (4).

Two dimensionless combinations recur. The first is

$$q = \frac{p_2}{p_1^2} = \frac{Z_1(2\beta)}{Z_1(\beta)^2}, \quad \frac{1}{K} \leq q \leq 1, \quad (17)$$

where the upper bound is the observation that $(\sum_i x_i)^2 \geq \sum_i x_i^2$ and the lower one is Cauchy–Schwarz. In the continuum model (16) with degeneracy g ,

$$q = \frac{1}{2^{3/2} g s}, \quad (18)$$

so q is the reciprocal of the effective number of thermally accessible one-particle states, up to the factor $2^{3/2}$ that comes from evaluating the same sum at twice the inverse temperature. That reading survives verbatim for a discrete spectrum: at high temperature every orbital becomes equally occupied, $q \rightarrow 1/K$, and at low temperature a single level dominates, $q \rightarrow 1$. The bounds below are informative precisely when $qN(N-1)$ is small, which is the discrete-spectrum form of Kroemer’s diluteness condition $N^2 \ll n_Q V$.

The second combination is the *relative slack* of an endpoint. For an upper endpoint $U \geq M_N$ it is $(U - M_N)/M_N$, and for a lower endpoint $L \leq M_N$ it is $(M_N - L)/M_N$. Both are dimensionless, both are zero exactly when the endpoint is sharp, and they are what the tables and figures report. Because M_N can be exponentially large or small, absolute gaps are not comparable across spectra or temperatures and are reported only where they are the worst case of a check.

3 What is being extended

Kroemer’s note [3] asks how incorrect the classical partition function of the ideal gas is, and answers with the two-sided bound (2). Its hypotheses are worth listing separately, because each of them is a restriction on the object rather than on the technique: the gas is spinless and monatomic, so the one-particle spectrum has no internal multiplicity; the particles do not interact; the box is a continuum, so that $Z_1(k\beta) = k^{-3/2} Z_1(\beta)$ is available; and the system is large enough that the chemical potential is unchanged by the removal of two particles, which is what allows a grand-canonical argument to be read as a statement about fixed N . The conclusion is a strict inequality with a prefactor proportional to N^2/s .

At $N = 2$ the same content is the textbook exchange correction. From (6),

$$B_2 = \frac{p_1^2 + p_2}{2}, \quad F_2 = \frac{p_1^2 - p_2}{2}, \quad Z_2^{\text{MB}} = \frac{p_1^2}{2}, \quad M_2 = p_2, \quad (19)$$

so the classical value sits exactly halfway between the two exact ones and the relative error of classical counting is $p_2/p_1^2 = q$: one part in $2^{3/2} g s$ for a three-dimensional gas with degeneracy g . What the prefactor form adds is the statement that the same relative size, multiplied by the number of ways two particles can be chosen out of N , controls the error at every N . That reading is exactly right, and it explains the finite- N correction that this paper isolates. A doubly occupied orbital is produced by a pair of particles, and there are $N(N-1)$ ordered pairs of *distinct* particles against the N^2 ordered pairs that allow a particle to be paired with itself. The prefactor $f_N = qN(N-1)$ of (3) counts the pairs that can actually produce a double occupancy; the literal $f = qN^2$ counts the rest as well, and at fixed N the excess is a relative amount $1/N$ that no amount of dilution removes.

The published problem [2] asks for the bound and its prefactor outside that setting, in four directions: (a) spin and orbital degeneracy; (b) small systems and stronger quantum degeneracy; (c) interparticle interactions; and (d) discrete level structures or confining geometries. What follows reaches (d) completely, (a) completely, (b) completely for the ideal gas, and (c) for the subclass of interactions that are diagonal in a fixed single-particle basis.

Direction (d) is where the exact statement is strongest. Nothing in Theorems 4.1 and 5.3 refers to a box, to a volume, to a dispersion relation or to a continuum: the input is a finite list of energies, and the output holds at every temperature and every $N \geq 2$. A harmonic trap, a hard-wall box on a finite mode grid, a two-level system, a spectrum with no structure at all and a spectrum whose levels are degenerate are all instances of (4), and the verification of Section 9 runs over six of them. Direction (a) is then free: a degeneracy is a repeated level, and g identical copies of a spatial spectrum change p_k by the factor g , hence q by $1/g$ (18) — the prefactor is suppressed by the degeneracy, as it must be, because a particle has more orbitals available in which not to meet another one. Direction (b) is the content of Proposition 5.5, which shows that the literal prefactor fails on its lower side at fixed N and that the replacement of N^2 by $N(N-1)$ is precisely the repair.

Direction (c) is reached only in part, and the part that is reached is worth stating exactly. Section 8 treats pair energies of the form $\sum_i u_i \binom{\nu_i}{2}$ with arbitrary real u_i , at any coupling strength. The upper bound survives with the prefactor $\sum_i e^{-\beta(2\varepsilon_i + u_i)}$ for repulsion, the lower bound survives in a restricted form for attraction as well, and both collapse to the free bounds at $u = 0$. What is not reached — and this is the main open question left by this work — is any interaction with matrix elements that mix orbitals, and any interaction on the fermionic side that does not vanish by exclusion. The injection used throughout pairs a doubly occupied orbital with an arbitrary remaining configuration, and that pairing stops being available when the interaction is not diagonal in the occupation basis.

Two further differences from [3] are matters of form rather than of scope, but they change how the result can be used. Every inequality here is non-strict and exact — an identity or an inequality between explicitly defined finite sums — so it can be applied at any temperature, including the degenerate one, without checking that an asymptotic parameter is small enough. And both slacks of the prefactor interval are decomposed into series with explicitly non-negative coefficients, so the distance between the bound and the truth is itself computable to any desired order rather than merely bounded.

4 Two spectrum-independent bounds

4.1 Statement and proof

Theorem 4.1 (universal two-sided bound). *Let $\varepsilon_1 \leq \dots \leq \varepsilon_K$ be any finite one-particle spectrum, $\beta > 0$, and $p_k = Z_1(k\beta)$. Then for every $N \geq 2$,*

$$\sum_{k=1}^{\lfloor N/2 \rfloor} \frac{p_{2k}}{k} F_{N-2k} \leq M_N \leq p_2 B_{N-2}. \quad (20)$$

Both inequalities are equalities for $N = 2$ and $N = 3$. Neither side refers to the geometry, the volume, the level spacing or the degeneracy of the spectrum, and neither requires N to be large.

Proof. Lower bound. By (15), $M_N = [z^N] F(z)(e^{\Delta(z)} - 1)$. The series Δ has non-negative coefficients and vanishes at $z = 0$, so the scalar inequality $e^w - 1 \geq w$, valid for $w \geq 0$, applies coefficientwise: $e^\Delta - 1 = \Delta + \frac{1}{2}\Delta^2 + \dots \geq \Delta$. Multiplication by F , whose coefficients are also non-negative, preserves the coefficientwise order, and extracting $[z^N]$ from $F\Delta$ gives

$$M_N \geq [z^N] F(z)\Delta(z) = \sum_{k \geq 1} \frac{p_{2k}}{k} F_{N-2k} = \sum_{k=1}^{\lfloor N/2 \rfloor} \frac{p_{2k}}{k} F_{N-2k}, \quad (21)$$

the last equality because $F_r = 0$ for $r < 0$ by (8).

Upper bound. Order the orbitals as in (4) and give every configuration ν with some $\nu_i \geq 2$ its *first doubly occupied orbital* $i_*(\nu) = \min\{i : \nu_i \geq 2\}$. The map

$$\nu \mapsto (i_*(\nu), \nu - 2e_{i_*(\nu)}) \quad (22)$$

is a bijection from those configurations onto the set of pairs (i, ν') with $|\nu'| = N - 2$ and $\nu'_j \leq 1$ for every $j < i$; its inverse is $(i, \nu') \mapsto \nu' + 2e_i$, and the condition $\nu'_j \leq 1$ for $j < i$ is exactly what makes i the *first* doubly occupied orbital of the image. Removing two particles from orbital i multiplies the Boltzmann weight by $e^{-2\beta\varepsilon_i} = x_i^2$, so

$$M_N = \sum_{i=1}^K x_i^2 \sum_{\nu' \in D_i} e^{-\beta E(\nu')}, \quad D_i = \{\nu' : |\nu'| = N - 2, \nu'_j \leq 1 \text{ for } j < i\}. \quad (23)$$

Each D_i is a subset of all $(N - 2)$ -particle bosonic configurations, whose total weight is B_{N-2} , and every summand is non-negative. Extending the inner sum therefore only increases it:

$$M_N \leq \sum_{i=1}^K x_i^2 B_{N-2} = p_2 B_{N-2}. \quad (24) \quad \square$$

The identity (23) is stronger than the bound and is what makes the refinement of Section 6 possible: the difference between the two sides is exactly the weight of those ν' that still contain an earlier double occupation, and that set can be enumerated.

The two endpoints have a clean reading. The leading term of the lower bound is $p_2 F_{N-2}$: choose an orbital, put two particles in it, and distribute the remaining $N - 2$ *fermionically*. The upper bound is $p_2 B_{N-2}$: choose an orbital, put two particles in it, and distribute the remaining $N - 2$ *bosonically*. The whole gap between them is the effect of exchange symmetrisation on the $N - 2$ particles that are left over, and it is the only place where the bounds can be loose. The higher terms $p_{2k} F_{N-2k}/k$ of the lower bound add configurations with k separated double occupancies, with the factor $1/k$ removing their ordering.

4.2 Exactness, and a worked instance

For $N = 2$ the lower bound is $p_2 F_0 = p_2$ and the upper one is $p_2 B_0 = p_2$, while $M_2 = p_2$ by (19): both are equalities. For $N = 3$ they are $p_2 F_1 = p_2 p_1$ and $p_2 B_1 = p_2 p_1$, and $M_3 = p_1 p_2$: equalities again. The first non-trivial case is $N = 4$, where

$$M_4 = \frac{p_1^2 p_2 + p_4}{2}, \quad p_2 B_2 - M_4 = \frac{p_2^2 - p_4}{2} = \frac{1}{2} \sum_{i \neq j} x_i^2 x_j^2, \quad M_4 - \left(p_2 F_2 + \frac{p_4}{2} \right) = \frac{p_2^2}{2}. \quad (25)$$

The upper excess $\frac{1}{2} \sum_{i \neq j} x_i^2 x_j^2$ is precisely the weight of the configurations with two doubly occupied orbitals — those are the elements of D_i that the injection overcounts, since a configuration with double occupancies in orbitals $j < i$ is counted once for each of them. The lower deficit $p_2^2/2$ is what the truncation $e^\Delta - 1 \geq \Delta$ discards at order z^4 , namely $\frac{1}{2}\Delta^2 \supset \frac{1}{2}p_2^2 z^4$. Both quantities are non-negative for every spectrum, and both vanish only in the trivial case of a single orbital.

4.3 Two limits

The bounds behave in opposite ways at the two ends of the temperature range, and it is worth being explicit about that, because it determines where each of them is useful.

Proposition 4.2 (low-temperature limit). *Let the lowest level be non-degenerate and placed at zero, $\varepsilon_1 = 0 < \varepsilon_2$, and let $\Delta_\varepsilon = \varepsilon_2 - \varepsilon_1$. As $\beta \rightarrow \infty$,*

$$B_N = 1 + O(e^{-\beta\Delta_\varepsilon}), \quad F_N = O(e^{-\beta\Delta_\varepsilon}), \quad M_N = 1 + O(e^{-\beta\Delta_\varepsilon}), \quad (26)$$

the upper endpoint of (20) is asymptotically exact, $p_2 B_{N-2}/M_N = 1 + O(e^{-\beta\Delta_\varepsilon})$, the lower endpoint retains the fraction

$$\frac{1}{M_N} \sum_{k=1}^{\lfloor N/2 \rfloor} \frac{p_{2k}}{k} F_{N-2k} \longrightarrow \begin{cases} 2/N, & N \text{ even}, \\ 2/(N-1), & N \text{ odd}, \end{cases} \quad (27)$$

and $q \rightarrow 1$, so the prefactor endpoints of Theorem 5.3 tend to $N(N-1)$ and to 0.

Proof. Here $x_1 = 1$ and $x_i \leq e^{-\beta\Delta_\varepsilon}$ for $i \geq 2$. Every $p_k = 1 + O(e^{-\beta\Delta_\varepsilon})$, and $B_N = h_N(x)$ has the single monomial $x_1^N = 1$ together with terms carrying at least one excited index, so $B_N = 1 + O(e^{-\beta\Delta_\varepsilon})$; likewise $F_N = e_N(x)$ has every monomial carrying at least $N-1$ excited indices, so $F_N = O(e^{-\beta\Delta_\varepsilon})$ for $N \geq 2$, and $p_2 B_{N-2}$ has the same leading value 1 as M_N . In the lower bound, F_{N-2k} survives the limit only when $N-2k = 0$, which contributes $p_N/(N/2) \rightarrow 2/N$ and requires N even, or when $N-2k = 1$, which contributes $p_{N-1}F_1/((N-1)/2) \rightarrow 2/(N-1)$ and requires N odd. Finally $q = p_2/p_1^2 \rightarrow 1$. $\square$

Physically, at low temperature almost all of the bosonic weight sits in the single configuration that puts every particle in the ground state. That configuration has a double occupancy as soon as $N \geq 2$, its first doubly occupied orbital is the ground state, and the leftover configuration is empty: the injection (22) is onto up to exponentially small terms, which is why the upper bound is exact there. The lower bound has no such luck, because it insists on distributing the remaining $N-2$ particles fermionically, and at low temperature that costs an excitation energy. Its limiting fraction $2/N$ is the residue of the single term in which all N particles form $N/2$ separated pairs. The prefactor form degenerates in the same limit: $q \rightarrow 1$ means $f_N \rightarrow N(N-1)$, an upper bound that grows with N while M_N stays at 1. The prefactor interval is a dilute-regime instrument, and Section 7 quantifies how it sharpens as s grows.

At the opposite end the bounds turn into pure combinatorics.

Corollary 4.3 (high-temperature, or infinite-volume, combinatorial form). *Let $\beta \rightarrow 0$ at fixed K , so that every $x_i \rightarrow 1$. Then $B_N \rightarrow \binom{N+K-1}{N}$, $F_N \rightarrow \binom{K}{N}$ and $p_k \rightarrow K$, and Theorem 4.1 becomes the pair of inequalities*

$$\sum_{k=1}^{\lfloor N/2 \rfloor} \frac{K}{k} \binom{K}{N-2k} \leq \binom{N+K-1}{N} - \binom{K}{N} \leq K \binom{N+K-3}{N-2}, \quad (28)$$

while Theorem 5.3 becomes, with $f_N \rightarrow N(N-1)/K$,

$$\frac{N(N-1)}{K} \binom{K}{N} \leq \binom{N+K-1}{N} - \binom{K}{N} \leq \frac{N(N-1)}{K} \binom{N+K-1}{N}. \quad (29)$$

All of these are exact rational inequalities between binomial coefficients, holding for all integers $K \geq 1$ and $N \geq 2$. For $K > 1$ the pair (29) is weaker than (28) on both sides, which is what one expects from a bound whose endpoints no longer refer to B_{N-2} and F_{N-2} but only to B_N and F_N themselves.

Proof. Every quantity involved is a polynomial with non-negative integer coefficients in the x_i , so the limit $x_i \rightarrow 1$ is a substitution; the counts of bosonic configurations and of fermionic ones are the stated binomial coefficients. The bounds are preserved because they hold for every $\beta > 0$. In the same limit $q = p_2/p_1^2 \rightarrow K/K^2 = 1/K$, hence $f_N \rightarrow N(N-1)/K$ and (29) is the limit of Theorem 5.3. $\square$

Corollary 4.3 was checked in exact rational arithmetic for every $K = 1, \dots, 13$ and $N = 2, \dots, 15$, all 182 pairs, with no failures for either (28) or (29) (Table 5). It is the cleanest statement of the fact that nothing in (20) is analytic: the bounds are properties of the counting, and the temperature only supplies the weights.

5 The exact discrete prefactor

The bounds of Theorem 4.1 are exact, but they are not in Kroemer's form: their endpoints are p_2 times a partition function of $N-2$ particles, not a number times a partition function of N particles. Two elementary inequalities convert one into the other, and the price paid in the conversion turns out to be exactly the finite-size correction.

5.1 Two coefficient inequalities

Lemma 5.1 (boson side). *For every $n \geq 2$ and every finite spectrum,*

$$n(n-1) B_n \geq p_1^2 B_{n-2}. \quad (30)$$

Proof. The recurrence (9) expresses nB_n as a sum of non-negative terms, one of which is $p_1 B_{n-1}$, so $nB_n \geq p_1 B_{n-1}$; the same recurrence at $n-1$ gives $(n-1)B_{n-1} \geq p_1 B_{n-2}$. Multiplying the two inequalities and dividing by the positive number B_{n-1} yields (30). $\square$

The estimate loses information twice, once at each use of the recurrence, and both losses are the discarded terms $p_k B_{n-k}$ with $k \geq 2$. Section 6 computes exactly what they amount to.

Lemma 5.2 (fermion side). *For every $n \geq 2$ and every finite spectrum,*

$$p_1^2 F_{n-2} \geq n(n-1) F_n, \quad (31)$$

with strict inequality whenever $F_n > 0$.

Proof. Write $F_n = e_n(x)$, the elementary symmetric polynomial of degree n in the weights $x_1, \dots, x_K$, and $p_1 = e_1(x)$. Expand the product $e_1^2 e_{n-2}$ and separate it by how the two factors of e_1 meet the subset that e_{n-2} selects. For an ordered pair of distinct indices $i \neq j$ let

$$e_{n-2}^{(i,j)} = \sum_{\substack{|T|=n-2 \\ i,j \notin T}} \prod_{t \in T} x_t \quad (32)$$

be the part of e_{n-2} whose subsets avoid both i and j . Then $x_i x_j e_{n-2}^{(i,j)}$ is exactly the sum of x^S over those n -element subsets S that contain both i and j , so summing over ordered pairs counts every n -subset once for each ordered pair inside it:

$$\sum_{i \neq j} x_i x_j e_{n-2}^{(i,j)} = n(n-1) e_n. \quad (33)$$

On the other hand

$$p_1^2 e_{n-2} = \left(\sum_i x_i^2 + \sum_{i \neq j} x_i x_j \right) e_{n-2} = \sum_i x_i^2 e_{n-2} + \sum_{i \neq j} x_i x_j e_{n-2} \geq \sum_{i \neq j} x_i x_j e_{n-2}^{(i,j)}, \quad (34)$$

because $e_{n-2} - e_{n-2}^{(i,j)}$ is a sum of monomials with non-negative coefficients — those subsets that meet $\{i, j\}$ — and the diagonal terms $\sum_i x_i^2 e_{n-2}$ are non-negative as well. Combining this with (33) gives (31). The inequality is strict whenever $e_n > 0$, because then either some diagonal term or some subset meeting $\{i, j\}$ is present and positive. $\square$

Both lemmas are inequalities between symmetric functions with non-negative coefficients; neither uses the spectrum, the temperature or N beyond the stated range, and both are saturated by a single-orbital spectrum.

5.2 The prefactor theorem

Theorem 5.3 (exact discrete Kroemer prefactor). *Let $\varepsilon_1 \leq \dots \leq \varepsilon_K$ be any finite one-particle spectrum with $p_1 = Z_1(\beta) > 0$, let $\beta > 0$, and let $N \geq 2$. With*

$$q = \frac{p_2}{p_1^2} = \frac{Z_1(2\beta)}{Z_1(\beta)^2}, \quad f_N = q N(N-1), \quad (35)$$

the multiple-occupancy part satisfies

$$f_N F_N \leq M_N \leq f_N B_N. \quad (36)$$

Proof. Theorem 4.1 gives $p_2 F_{N-2} \leq M_N \leq p_2 B_{N-2}$. Lemma 5.1 multiplied by $q = p_2/p_1^2$ reads $qN(N-1)B_N \geq p_2 B_{N-2}$, which is the upper half of (36); Lemma 5.2 multiplied by the same factor reads $qN(N-1)F_N = p_2 N(N-1)F_N/p_1^2 \leq p_2 F_{N-2}$, which is the lower half. $\square$

Three features of (36) deserve emphasis. It is an exact statement at every temperature, including the degenerate one, and every $N \geq 2$; it makes no reference to a volume, a chemical potential or a continuum; and its prefactor is a single measurable spectral ratio $q = Z_1(2\beta)/Z_1(\beta)^2$ times a purely combinatorial factor. By (17) the prefactor obeys $N(N-1)/K \leq f_N \leq N(N-1)$, so the bound is informative — f_N small — exactly when the number of thermally accessible orbitals is large compared with N^2 , which is the discrete-spectrum version of the diluteness condition.

At $N = 2$ the interval is exactly symmetric about M_2 .

Corollary 5.4 ($N = 2$ gap). *For every spectrum and every $\beta > 0$,*

$$f_2 B_2 - M_2 = M_2 - f_2 F_2 = q p_2 > 0, \quad \frac{f_2 B_2 - M_2}{M_2} = \frac{M_2 - f_2 F_2}{M_2} = q. \quad (37)$$

Proof. Insert $B_2 = (p_1^2 + p_2)/2$, $F_2 = (p_1^2 - p_2)/2$, $M_2 = p_2$ from (19) and $f_2 = 2p_2/p_1^2$ into both differences; each reduces to $p_2^2/p_1^2 = qp_2$. $\square$

So at two particles the prefactor interval has relative width $2q$, and q is the same number that measures the relative error of classical counting at $N = 2$ (Section 3). Over the verification grid of Section 9 the tightest relative slack attained anywhere in (36) is 3.78×10^{-2} , at $N = 2$ for a three-dimensional harmonic trap at $\beta = 0.1$, and Corollary 5.4 identifies it as the value of q there; for $N \geq 4$ the tightest slack on the grid is 0.21.

5.3 Recovery of the continuum prefactor

In the continuum model (16) with degeneracy g , $p_k = g s k^{-3/2}$, so

$$q = \frac{gs \cdot 2^{-3/2}}{(gs)^2} = \frac{1}{2^{3/2} g s}, \quad f_N = \frac{N(N-1)}{2^{3/2} g s}, \quad f = \frac{N^2}{2^{3/2} g s} \quad (\text{Kroemer's, for } g = 1), \quad (38)$$

and $f_N/f = (N-1)/N$. Two of the four extension directions are settled by this line: the degeneracy g enters the prefactor exactly as a suppression by g , and the small-system content of the bound is the single factor $(N-1)/N$, with no other correction of that order. The identification is verified numerically in Table 5 by computing q from the model spectrum at $g = 1, 2, 3$ and comparing it with $1/(2^{3/2}gs)$.

5.4 What the literal prefactor misses

The factor $(N-1)/N$ looks harmless, and for large N it is. At fixed N it is the difference between a bound and a statement that is false on one side.

Proposition 5.5 (the literal N^2 prefactor at fixed N). *In the continuum model spectrum $p_k = s k^{-3/2}$, at fixed $N \geq 2$ and as $s \rightarrow \infty$,*

$$\frac{f B_N}{M_N} \rightarrow \frac{N}{N-1}, \quad \frac{f F_N}{M_N} \rightarrow \frac{N}{N-1}, \quad f = \frac{N^2}{2^{3/2} s}. \quad (39)$$

Consequently the upper advertisement $M_N \leq f B_N$ holds with a relative slack tending to $1/(N-1)$, which does not vanish as the gas dilutes, while the lower advertisement $M_N \geq f F_N$ fails for all sufficiently dilute gases, with a relative deficit tending to $-1/(N-1)$. For the exact prefactor f_N of (35) both relative slacks tend to zero.

Proof. Appendix B establishes the expansions

$$B_N(s) = \frac{s^N}{N!} + \frac{2^{-5/2} s^{N-1}}{(N-2)!} + O(s^{N-2}), \quad F_N(s) = \frac{s^N}{N!} - \frac{2^{-5/2} s^{N-1}}{(N-2)!} + O(s^{N-2}), \quad (40)$$

in which the coefficients of s^{N-2} are common to B_N and F_N , so that

$$M_N(s) = \frac{2^{-3/2} s^{N-1}}{(N-2)!} + O(s^{N-3}). \quad (41)$$

Multiplying (40) by $f = N^2/(2^{3/2}s)$ gives, for both B_N and F_N , the leading value

$$f \frac{s^N}{2^{3/2} s N!} = \frac{N}{N-1} \cdot \frac{2^{-3/2} s^{N-1}}{(N-2)!} = \frac{N}{N-1} M_N(s) (1 + O(s^{-2})), \quad (42)$$

which is (39). The statement about f_N is Theorem 7.3, whose leading relative slack is $N(N-1)/(4\sqrt{2}s)$. $\square$

Table 1 measures both sides of (39) against the exact recurrences. The literal relative slacks converge to $\pm 1/(N-1)$ from either side, agreeing to within 1.2×10^{-3} already at $s = 10^4$ and to within 10^{-6} at $s = 10^8$, while the exact prefactor's slacks fall as s^{-1} .

Proposition 5.5 should be read together with the hypothesis that Kroemer imposed. His derivation assumes a system large enough that removing two particles leaves the chemical potential unchanged, which is a statement that N is large; a relative deficit of order $1/N$ is then beyond the

Table 1: Relative slacks of the literal continuum prefactor $f = N^2/(2^{3/2}s)$ and of the exact discrete prefactor $f_N = N(N-1)/(2^{3/2}s)$ in the model spectrum $p_k = s k^{-3/2}$, computed from the exact canonical recurrences at 200-bit precision. Upper slacks are $(fB_N - M_N)/M_N$ and $(f_N B_N - M_N)/M_N$; lower slacks are $(M_N - fF_N)/M_N$ and $(M_N - f_N F_N)/M_N$, so a negative lower entry means the advertised inequality fails. The literal lower side fails at every N and every s shown, and its deficit tends to $-1/(N-1)$, whereas the exact slacks fall as s^{-1} .

N	s	literal f		exact f_N	
		upper	lower	upper	lower
2	10^4	+1.00007	-0.99993	3.54×10^{-5}	3.54×10^{-5}
2	10^8	+1.00000	-1.00000	3.54×10^{-9}	3.54×10^{-9}
4	10^4	+0.33362	-0.33305	2.12×10^{-4}	2.12×10^{-4}
4	10^8	+0.33333	-0.33333	2.12×10^{-8}	2.12×10^{-8}
6	10^4	+0.20064	-0.19936	5.30×10^{-4}	5.30×10^{-4}
6	10^8	+0.20000	-0.20000	5.30×10^{-8}	5.30×10^{-8}
8	10^4	+0.14399	-0.14173	9.90×10^{-4}	9.90×10^{-4}
8	10^8	+0.14286	-0.14286	9.90×10^{-8}	9.90×10^{-8}

accuracy claimed. What the proposition says is that the N^2 prefactor is an asymptotically large- N object, that its defect is exactly of order $1/N$ on both sides, and that replacing N^2 by $N(N-1)$ repairs both sides at every $N \geq 2$ without weakening anything else. At $N = 2$ the failure is total rather than asymptotic, and it can be exhibited in closed form.

Proposition 5.6 ($N = 2$ witness). *Impose the continuum relation $p_2 = 2^{-3/2}p_1$ and write $s = p_1$. Then $B_2 = (s^2 + 2^{-3/2}s)/2$, $F_2 = (s^2 - 2^{-3/2}s)/2$, $M_2 = 2^{-3/2}s$ and $f = \sqrt{2}/s$, so that*

$$fB_2 - M_2 = \frac{s}{2\sqrt{2}} + \frac{1}{4} > 0 \quad \text{for every } s > 0, \quad (43)$$

while

$$M_2 - fF_2 = \frac{1}{4} - \frac{s}{2\sqrt{2}} \quad (44)$$

changes sign at $s = 2^{-1/2}$. The lower advertisement of (2) therefore fails at $N = 2$ for every diluteness above $2^{-1/2}$, while $F_2 > 0$ requires only $s > 2^{-3/2}$, so the failure occurs throughout the classical regime $s \gg 1$. The exact prefactor has no such failure: by Corollary 5.4 its two gaps are equal, and in this model both are $qp_2 = 1/8$, independently of s .

Proof. With $p_2 = 2^{-3/2}p_1$ and $s = p_1$, (19) gives the three stated values of B_2 , F_2 and M_2 , and $f = N^2/(2^{3/2}s) = \sqrt{2}/s$ at $N = 2$. Then $fB_2 - M_2 = (\sqrt{2}/s)(s^2 + 2^{-3/2}s)/2 - 2^{-3/2}s = s/(2\sqrt{2}) + 1/4$ and $M_2 - fF_2 = 2^{-3/2}s - (\sqrt{2}/s)(s^2 - 2^{-3/2}s)/2 = 1/4 - s/(2\sqrt{2})$, which is positive, zero and negative according as $s < 2^{-1/2}$, $s = 2^{-1/2}$ or $s > 2^{-1/2}$. For the exact prefactor, $f_2 = 2q = 2 \cdot 2^{-3/2}/s$ and Corollary 5.4 makes both gaps $qp_2 = (2^{-3/2}/s)(2^{-3/2}s) = 1/8$. $\square$

Table 2 tabulates (43) and (44). The sign change is the whole content: the literal prefactor is not too small on the upper side, it is too large on the lower one, and the exact discrete prefactor is symmetric about M_2 instead.

One caution belongs with the witness. The relation $p_2 = 2^{-3/2}p_1$ is the large-volume limit of a three-dimensional box, not an identity that a finite box obeys, so Proposition 5.6 is a statement

Table 2: The $N = 2$ witness of Proposition 5.6 in the model continuum spectrum $p_2 = 2^{-3/2}p_1$, $s = p_1 = n_Q V$. Both gaps are computed in exact arithmetic from the closed forms (43) and (44). The fermionic partition function is positive only for $s > 2^{-3/2} = 0.3536$, so the first row carries no lower advertisement to test; from $s = 0.8$ onwards the literal lower gap is negative, while the exact prefactor's gap is the constant $qp_2 = 1/8$ on both sides.

$s = n_Q V$	$F_2 > 0$	$fB_2 - M_2$	$M_2 - fF_2$	exact gap, both sides
0.3	no	+0.35607	+0.14393	0.125
0.4	yes	+0.39142	+0.10858	0.125
0.5	yes	+0.42678	+0.07322	0.125
0.6	yes	+0.46213	+0.03787	0.125
0.7	yes	+0.49749	+0.00251	0.125
$2^{-1/2}$	yes	+0.50000	0	0.125
0.8	yes	+0.53284	-0.03284	0.125
1.0	yes	+0.60355	-0.10355	0.125
2.0	yes	+0.95711	-0.45711	0.125
10.0	yes	+3.78553	-3.28553	0.125

about the model spectrum (16) and not about an arbitrary confining geometry. For a genuine finite spectrum the exact prefactor of Theorem 5.3 is the object that holds without qualification, and the witness only identifies which step of the continuum argument does not transplant: it is the replacement of $N(N - 1)$ by N^2 , and nothing else.

6 The exact structure of both slacks

The upper endpoint of Theorem 4.1 is exact at $N = 2, 3$ and too large from $N = 4$ on, and the upper endpoint of Theorem 5.3 is larger still. Both excesses are computable exactly, and both turn out to be sums of non-negative terms, which is what allows them to be subtracted one layer at a time. Write

$$E_N = p_2 B_{N-2} - M_N \geq 0, \quad S_N = f_N B_N - p_2 B_{N-2} \geq 0, \quad (45)$$

for the *overcount* of the injection and the *prefactor slack* of the conversion; the second is non-negative by Lemma 5.1. The two are independent sources of looseness, and $f_N B_N - M_N = S_N + E_N$.

6.1 The injection overcount

Set $b_i = x_i^2 = e^{-2\beta\varepsilon_i}$, so that $p_{2k} = \sum_i b_i^k$, and introduce the complete homogeneous symmetric polynomials of the b_i ,

$$H(s) = \prod_{i=1}^K \frac{1}{1 - b_i s} = \sum_{n \geq 0} H_n(b) s^n, \quad H_n(b) = \sum_{m_1 + \dots + m_K = n} b_1^{m_1} \dots b_K^{m_K}. \quad (46)$$

Theorem 6.1 (first-double-occupation overcount). *For every finite spectrum, every $\beta > 0$ and every $N \geq 2$,*

$$p_2 B_{N-2} - M_N = \sum_{m=2}^{\lfloor N/2 \rfloor} c_m F_{N-2m}, \quad c_m = \sum_{\substack{r \in \mathbb{Z}_{\geq 0}^K \\ |r|=m}} (\ell(r) - 1) b^r \geq 0, \quad (47)$$

where $\ell(r) = \#\{i : r_i > 0\}$ is the number of orbitals that r occupies and $b^r = \prod_i b_i^{r_i}$. Equivalently $c_m = p_2 H_{m-1}(b) - H_m(b)$, and the first two coefficients are

$$c_2 = \frac{p_2^2 - p_4}{2} = e_2(b), \quad c_3 = \frac{p_2^3 - p_6}{3}. \quad (48)$$

Proof. The generating function of E_N is $E(z) = p_2 z^2 B(z) - \sum_N M_N z^N$. Using $B = F e^\Delta$ from (15) and $\sum_N M_N z^N = F(e^\Delta - 1)$,

$$E(z) = F(z) \left[p_2 z^2 e^\Delta - e^\Delta + 1 \right] = F(z) \left[1 - (1 - p_2 z^2) e^{\Delta(z)} \right]. \quad (49)$$

Each factor of e^Δ is $(1 - x_i z)^{-1} (1 + x_i z)^{-1} = (1 - b_i z^2)^{-1}$, so $e^{\Delta(z)} = H(z^2)$ with H as in (46), and the bracket is a series in $s = z^2$:

$$1 - (1 - p_2 s) H(s) = \sum_{m \geq 2} (p_2 H_{m-1}(b) - H_m(b)) s^m = \sum_{m \geq 2} c_m s^m, \quad (50)$$

the $m = 1$ coefficient being $p_2 H_0 - H_1 = p_2 - p_2 = 0$. Multiplying by F and extracting $[z^N]$ gives (47) with F_{N-2m} in place of $[z^{N-2m}]F$.

For the sign and the weight form, expand $p_2 H_{m-1}(b) = \sum_i b_i \sum_{|r'|=m-1} b^{r'}$. A monomial b^r of total degree $m \geq 2$ is produced once for each orbital i that carries a positive power in r , that is $\ell(r)$ times, whereas $H_m(b)$ contains it once. The coefficient of b^r in c_m is therefore $\ell(r) - 1$, which is non-negative because $|r| = m \geq 1$ forces $\ell(r) \geq 1$. It vanishes exactly when the whole degree m sits in a single orbital.

The two displayed values follow from $H_1 = p_2$, $H_2 = (p_2^2 + p_4)/2$ and $H_3 = (p_2^3 + 3p_2 p_4 + 2p_6)/6$; the identity $c_2 = e_2(b)$ is the observation that $p_2^2 = p_4 + 2 \sum_{i < j} b_i b_j$. $\square$

The reading of (47) is the one anticipated in Section 4.2: the injection (22) counts a configuration once for each of its doubly occupied orbitals, and $c_m F_{N-2m}$ is the weight of the configurations with m separated double occupancies that are therefore counted m times too often. The correction is not an estimate of that overcount but its exact value.

Corollary 6.2 (refined injection endpoint). *For $N \geq 4$,*

$$M_N \leq p_2 B_{N-2} - \frac{p_2^2 - p_4}{2} F_{N-4}, \quad (51)$$

and the inequality is an equality for $N = 4$ and $N = 5$.

Proof. Drop the terms $m \geq 3$ from the non-negative sum (47). For $N = 4, 5$ the sum has the single term $m = 2$, so nothing is dropped. $\square$

6.2 A closed form in the even thermal sums alone

The coefficients c_m do not require the individual weights b_i ; they are determined by the numbers $Z_1(2k\beta)$, which is what makes them usable when the spectrum is given through its density of states rather than level by level. Put

$$P_k = \sum_{i=1}^K b_i^k = Z_1(2k\beta), \quad k = 1, 2, \dots, \quad (52)$$

so that $P_1 = p_2$. Taking logarithms in (46),

$$H(s) = \exp\left(\sum_i \sum_{k \geq 1} \frac{b_i^k s^k}{k}\right) = \exp\left(\sum_{k \geq 1} \frac{P_k s^k}{k}\right), \quad (53)$$

and differentiating once with respect to s gives the Newton recurrence

$$mH_m = \sum_{k=1}^m P_k H_{m-k}, \quad H_0 = 1. \quad (54)$$

Corollary 6.3 (power-sum form of the correction). *Every $c_m = P_1 H_{m-1} - H_m$ with $m \geq 2$ is a polynomial with rational coefficients in $Z_1(2\beta), \dots, Z_1(2m\beta)$, obtained from (54) in finitely many steps, and the whole overcount is*

$$p_2 B_{N-2} - M_N = \sum_{m=2}^{\lfloor N/2 \rfloor} (P_1 H_{m-1} - H_m) F_{N-2m}. \quad (55)$$

Proof. (54) determines H_m from $H_0, \dots, H_{m-1}$ and $P_1, \dots, P_m$, so induction on m gives each H_m , and with it each $c_m = P_1 H_{m-1} - H_m$, as a polynomial with rational coefficients in $P_1, \dots, P_m = Z_1(2\beta), \dots, Z_1(2m\beta)$. Substituting that expression for c_m into (47) gives (55). $\square$

Corollary 6.3 was checked against the literal weight enumeration of c_m over the whole verification grid: the two forms agree to 2.0×10^{-56} at 200-bit precision, and the enumeration form agrees with the closed form to 1.0×10^{-58} on the spectra small enough to enumerate (Table 5).

6.3 The prefactor slack

The second source of looseness is the conversion of Lemma 5.1 itself. Write

$$A(z) = \sum_{k \geq 1} p_k z^k, \quad D(z) = A(z)^2 - p_1^2 z^2 + zA'(z) - A(z) = \sum_{j \geq 2} d_j z^j. \quad (56)$$

Theorem 6.4 (boson prefactor slack). *For every finite spectrum, every $\beta > 0$ and every $N \geq 2$,*

$$N(N-1)B_N - p_1^2 B_{N-2} = \sum_{j=2}^N d_j B_{N-j}, \quad d_2 = p_2, \quad d_j = \sum_{a=1}^{j-1} p_a p_{j-a} + (j-1)p_j \geq 0 \quad (j \geq 3), \quad (57)$$

and therefore

$$S_N = f_N B_N - p_2 B_{N-2} = \frac{p_2}{p_1^2} \sum_{j=2}^N d_j B_{N-j} \geq 0. \quad (58)$$

Proof. Since $\log B(z) = \sum_{k \geq 1} p_k z^k / k$, one has $z \partial_z \log B = A(z)$, hence $z \partial_z B = BA$ and

$$((z \partial_z)^2 - z \partial_z) B = z \partial_z (BA) - BA = B(A^2 + zA' - A). \quad (59)$$

The left-hand side has coefficient $N^2 B_N - N B_N = N(N-1)B_N$ at z^N . On the right, subtracting and adding $p_1^2 z^2$ inside the bracket splits it into $B(z)(D(z) + p_1^2 z^2)$, whose z^N coefficient is $\sum_{j \geq 2} d_j B_{N-j} + p_1^2 B_{N-2}$. Equating the two gives the first displayed identity, and $B_r = 0$ for $r < 0$ truncates the sum at $j = N$. The coefficients of D follow from $A^2 = \sum_{j \geq 2} (\sum_{a=1}^{j-1} p_a p_{j-a}) z^j$ and $zA' - A = \sum_{k \geq 1} (k-1)p_k z^k$: at $j = 2$ the term $p_1^2 z^2$ is removed by construction and only $(2-1)p_2$ survives, while for $j \geq 3$ both contributions are present and non-negative. Multiplying the identity by p_2/p_1^2 and using $f_N = (p_2/p_1^2)N(N-1)$ turns it into (58). $\square$

The coefficients d_j are the recurrence terms that Lemma 5.1 discards, arranged by order: d_j multiplies the products $p_a p_{j-a}$ of two thermal sums together with the single sum $(j-1)p_j$ that the second use of the recurrence loses. Like the c_m , they are non-negative for every spectrum, so they can be subtracted rather than merely bounded.

6.4 An exact identity and a hierarchy of upper endpoints

Putting the two decompositions together gives M_N itself as the prefactor endpoint minus two explicit non-negative series.

Theorem 6.5 (exact decomposition of the multiple-occupancy part). *For every finite spectrum, every $\beta > 0$ and every $N \geq 2$,*

$$M_N = f_N B_N - \frac{p_2}{p_1^2} \sum_{j=2}^N d_j B_{N-j} - \sum_{m=2}^{\lfloor N/2 \rfloor} c_m F_{N-2m}, \quad (60)$$

with $c_m \geq 0$ from (47) and $d_j \geq 0$ from (57).

Proof. By (45), $f_N B_N - M_N = S_N + E_N$; (58) expresses S_N as $(p_2/p_1^2) \sum_{j=2}^N d_j B_{N-j}$ and (47) expresses E_N as $\sum_{m=2}^{\lfloor N/2 \rfloor} c_m F_{N-2m}$. Moving both to the right-hand side gives (60), and the sums truncate because $B_r = F_r = 0$ for $r < 0$ and $r > K$. $\square$

Corollary 6.6 (monotone hierarchy of upper bounds). *For every $2 \leq k \leq N$ and every $2 \leq \ell \leq \lfloor N/2 \rfloor$, with empty sums read as zero,*

$$M_N \leq f_N B_N - \frac{p_2}{p_1^2} \sum_{j=2}^k d_j B_{N-j} - \sum_{m=2}^{\ell} c_m F_{N-2m}, \quad (61)$$

and the right-hand side decreases as either k or ℓ increases, reaching M_N at $k = N$, $\ell = \lfloor N/2 \rfloor$. In particular the two one-term refinements are

$$M_N \leq f_N B_N - \frac{p_2^2}{p_1^2} B_{N-2}, \quad M_N \leq f_N B_N - \frac{p_2^2 - p_4}{2} F_{N-4} \quad (N \geq 4). \quad (62)$$

Proof. Subtract any initial segment of either non-negative series from (60). Monotonicity is immediate, and at full retention both series are exhausted, so (60) is an equality. The first displayed refinement uses $d_2 = p_2$ and the second uses $c_2 = (p_2^2 - p_4)/2$. $\square$

Corollary 6.6 is the reason the decompositions are worth having: they convert an inequality whose slack was unknown into a sequence of inequalities whose slack is a series with computed non-negative coefficients, and any truncation of that series is still a bound. Section 7 uses the first two terms of each and measures what they buy in the dilute limit. The identity (60) itself was verified over the whole grid to 7.8×10^{-54} , and each of its 1800 truncated instances was checked to be an upper bound, with no violations (Table 5).

7 Second-order endpoints and the dilute limit

Corollary 6.6 supplies upper bounds of any desired order. The question this section answers is what the first few terms are worth in the regime where the prefactor form is used, namely a dilute three-dimensional gas at fixed N , and whether the lower side can be sharpened to match.

7.1 The second-order upper endpoint

Retaining d_2 , d_3 and c_2 in (61) gives the endpoint that the rest of this section analyses. Recall from (57) that $d_2 = p_2$ and $d_3 = 2p_1p_2 + 2p_3$, and from (48) that $c_2 = (p_2^2 - p_4)/2$.

Theorem 7.1 (second-order upper endpoint). *For every finite spectrum, every $\beta > 0$ and every $N \geq 4$,*

$$U_N^{(2)} := f_N B_N - \frac{p_2}{p_1^2} \left(p_2 B_{N-2} + (2p_1p_2 + 2p_3) B_{N-3} \right) - \frac{p_2^2 - p_4}{2} F_{N-4} \geq M_N. \quad (63)$$

The difference $U_N^{(2)} - M_N$ equals

$$\frac{p_2}{p_1^2} \sum_{j=4}^N d_j B_{N-j} + \sum_{m=3}^{\lfloor N/2 \rfloor} c_m F_{N-2m}, \quad (64)$$

a sum of explicitly non-negative terms.

Proof. This is (61) at $k = 3$, $\ell = 2$, and (64) is what (60) leaves after the three retained terms are moved to the other side. $\square$

7.2 The second-order lower endpoint

The lower side has no decomposition of the same kind, but the exponential itself supplies the missing non-negative terms.

Theorem 7.2 (second-order lower endpoint). *For every finite spectrum, every $\beta > 0$ and every $N \geq 4$,*

$$f_N F_N \leq L_N^{(2)} := p_2 F_{N-2} + \frac{1}{2} p_2^2 F_{N-4} \leq M_N. \quad (65)$$

For $N = 4$ and $N = 5$ the endpoint $L_N^{(2)}$ also dominates the lower bound of Theorem 4.1; for larger N the two are not ordered, and their maximum is again a valid lower bound.

Proof. The scalar inequality $e^w - 1 \geq w + w^2/2$ holds for every $w \geq 0$. The series $\Delta(z) = \sum_{k \geq 1} p_{2k} z^{2k}/k$ has non-negative coefficients, so the inequality applies coefficientwise, and restricting w^2 to its leading term $p_2^2 z^4$ preserves it:

$$e^{\Delta(z)} - 1 \geq p_2 z^2 + \frac{1}{2} p_2^2 z^4 \quad (66)$$

coefficientwise. Multiplying by the non-negative series $F(z)$ and extracting $[z^N]$ gives the right half of (65). The left half is Lemma 5.2 multiplied by p_2/p_1^2 , namely $p_2 F_{N-2} \geq f_N F_N$, together with the fact that $L_N^{(2)}$ exceeds $p_2 F_{N-2}$ by a non-negative amount. At $N = 4, 5$ the lower bound of Theorem 4.1 is $p_2 F_{N-2} + \frac{1}{2} p_4 F_{N-4}$ and $p_2^2 \geq p_4$ gives the domination; for $N \geq 6$ that bound carries further terms $p_{2k} F_{N-2k}/k$ with $k \geq 3$, which $L_N^{(2)}$ does not, while $L_N^{(2)}$ carries the larger coefficient $\frac{1}{2} p_2^2$ at $k = 2$. $\square$

7.3 Dilute-limit asymptotics

Everything so far is exact for a finite spectrum. The next statement is the only one in the paper that is asymptotic, and it is about the model continuum spectrum (16); its derivation, including the expansions of B_n and F_n to three subleading orders, is in Appendix B.

Theorem 7.3 (continuum sharpness). *Let $p_k = s k^{-3/2}$ with $s = n_Q V$, and fix $N \geq 4$. As $s \rightarrow \infty$,*

$$\frac{f_N B_N - M_N}{M_N} = \frac{M_N - f_N F_N}{M_N} = \frac{N(N-1)}{4\sqrt{2}s} + o(s^{-1}), \quad (67)$$

$$\frac{f_N B_N - c_2 F_{N-4} - M_N}{M_N} = \frac{4N-6}{4\sqrt{2}s} + o(s^{-1}), \quad (68)$$

$$\frac{U_N^{(2)} - M_N}{M_N} = \frac{(N-2)(N-3)}{s^2} \left[\frac{N^2 - 9N + 23}{24} + \frac{2\sqrt{3}}{9} \right] + o(s^{-2}), \quad (69)$$

$$\frac{M_N - L_N^{(2)}}{M_N} = \frac{(N-2)(N-3)}{s^2} \left[\frac{\sqrt{2}}{8} + \frac{(N-4)(N-5)}{48} \right] + o(s^{-2}). \quad (70)$$

With an internal degeneracy g , so that $p_k = g s k^{-3/2}$, the same formulas hold with gs in place of s .

All four coefficients are derived in Appendix B, from the expansions of B_n and F_n to three subleading orders in s^{-1} , and each is measured against the exact recurrences in Table 3. The remainder is in fact $O(s^{-3})$ there; the weaker $o(s^{-2})$ is what is stated because it is what the derivation needs.

Both sides of the prefactor interval are therefore asymptotically sharp at the rate s^{-1} , and the second-order endpoints improve that rate to s^{-2} . The improvement is not marginal. Dividing (69) by (67),

$$\begin{aligned} \frac{U_N^{(2)} - M_N}{f_N B_N - M_N} &= \frac{4\sqrt{2}(N-2)(N-3)}{N(N-1)s} \left[\frac{N^2 - 9N + 23}{24} + \frac{2\sqrt{3}}{9} \right] \\ &= \frac{16(N-2)(N-3)A_N}{N^2(N-1)^2} f_N (1 + o(1)), \end{aligned} \quad (71)$$

with A_N the bracket in (69); the coefficient of f_N tends to $2/3$ as N grows. The second-order endpoint is better than the first whenever the prefactor bound is itself informative, that is whenever f_N is small, and the same conclusion holds on the lower side with A_N replaced by the bracket of (70). The intermediate endpoint (68), which subtracts only the first overcount term, keeps the s^{-1} rate but reduces its coefficient by the factor $(4N-6)/(N(N-1))$, which behaves as $4/N$ for large N : the first-double-occupation overcount is the smaller of the two sources of looseness at large N , and the prefactor slack is the one that has to be removed to change the order.

Table 3 compares the four closed-form coefficients with values measured from the exact recurrences, and Figure 1 shows the whole range of diluteness. The measured coefficients are the relative slacks multiplied by s for the first-order endpoints and by s^2 for the second-order ones; agreement at $s = 10^6$ is already at the level of 10^{-5} and improves by two further orders of magnitude at $s = 10^8$, which is the numerical content of the $o(s^{-1})$ and $o(s^{-2})$ remainders.

Table 3: Leading coefficients of the relative slacks in the model continuum spectrum $p_k = s k^{-3/2}$, at fixed N . Measured values are the relative slack times s for the first-order endpoints and times s^2 for the second-order ones, evaluated from the exact canonical recurrences at 200-bit precision; predicted values are the closed forms (67)–(70). The columns are the upper prefactor endpoint $f_N B_N$, the endpoint refined by $c_2 F_{N-4}$, the second-order upper endpoint $U_N^{(2)}$ and the second-order lower endpoint $L_N^{(2)}$. The exact lower prefactor endpoint $f_N F_N$ was measured as well and agrees with the $f_N B_N$ column in every digit shown, as (67) states. The largest deviation between measured and predicted entries over the whole grid $N = 4, \dots, 8$, $s = 10^2, \dots, 10^8$ is 8.7×10^{-5} and occurs at $s = 10^6$; at $s = 10^2$ it is 0.87, which is the size of the remainder terms where the asymptotic regime has not yet been reached.

N	s	$f_N B_N$		$f_N B_N - c_2 F_{N-4}$		$U_N^{(2)}$		$L_N^{(2)}$	
		meas.	pred.	meas.	pred.	meas.	pred.	meas.	pred.
4	10^6	2.121322	2.121320	1.767769	1.767767	1.019801	1.019800	0.353553	0.353553
4	10^8	2.121320	2.121320	1.767767	1.767767	1.019800	1.019800	0.353553	0.353553
5	10^6	3.535538	3.535534	2.474879	2.474874	3.059406	3.059401	1.060660	1.060660
5	10^8	3.535534	3.535534	2.474874	2.474874	3.059401	3.059401	1.060660	1.060660
6	10^6	5.303310	5.303301	3.181993	3.181981	7.118818	7.118802	2.621321	2.621320
6	10^8	5.303301	5.303301	3.181981	3.181981	7.118802	7.118802	2.621320	2.621320
7	10^6	7.424640	7.424621	3.889113	3.889087	15.198045	15.198004	6.035538	6.035534
7	10^8	7.424621	7.424621	3.889088	3.889087	15.198004	15.198004	6.035534	6.035534
8	10^6	9.899528	9.899495	4.596244	4.596194	30.297094	30.297005	12.803311	12.803301
8	10^8	9.899495	9.899495	4.596195	4.596194	30.297006	30.297005	12.803301	12.803301

7.4 A coefficientwise certificate for the lower prefactor slack

The lower prefactor slack $p_2 F_{N-2} - f_N F_N$ is non-negative by Lemma 5.2, but it does not decompose in the basis $\{F_{N-j}\}$ the way its bosonic counterpart does in Theorem 6.4. In the monomial basis it does, and the certificate is short enough to be worth stating because it shows that the missing decomposition is a property of the basis and not an obstruction.

Proposition 7.4 (monomial certificate). *Let $x_i = e^{-\beta \varepsilon_i}$ and let e_k be the elementary symmetric polynomial of degree k in $x_1, \dots, x_K$, so that $F_k = e_k$ and $p_1 = e_1$. For $k \geq 2$ put*

$$A_k = e_1 e_{k-1} - k e_k, \quad A_1 = 0. \quad (72)$$

Then each A_k is a sum of unit-coefficient monomials,

$$A_k = \sum_{|T|=k-1} \sum_{i \in T} x_i^2 \prod_{j \in T \setminus \{i\}} x_j \geq 0, \quad (73)$$

the identity

$$p_1^2 F_{N-2} - N(N-1) F_N = e_1 A_{N-1} + (N-1) A_N \quad (74)$$

holds for every $N \geq 2$, and consequently the cleared lower prefactor slack is coefficientwise non-negative:

$$p_1^2 (p_2 F_{N-2} - f_N F_N) = p_2 [e_1 A_{N-1} + (N-1) A_N]. \quad (75)$$

Proof. Expanding $e_1 e_{k-1}$ produces each squarefree monomial x^S with $|S| = k$ exactly k times, once for each index of S taken from the factor e_1 , and each monomial $x_i^2 \prod_{j \in T''} x_j$ with $|T''| = k-2$

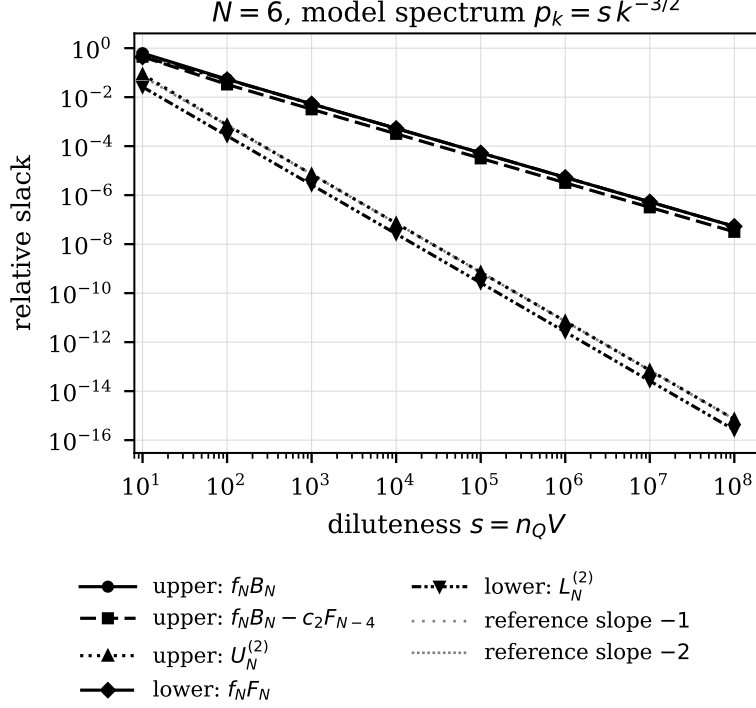


Figure 1: Relative slack of the five continuum endpoints at $N = 6$ in the model spectrum $p_k = s k^{-3/2}$, as a function of the diluteness $s = n_Q V$; both axes are logarithmic and every point is computed from the exact canonical recurrences at 200-bit precision. Upper endpoints are plotted as $(\text{endpoint} - M_N)/M_N$ and lower endpoints as $(M_N - \text{endpoint})/M_N$, so every curve is non-negative and a lower curve is a sharper bound. The two light reference lines have slopes -1 and -2 : the three prefactor-type endpoints $f_N B_N$, $f_N B_N - c_2 F_{N-4}$ and $f_N F_N$ follow the first at large s , with coefficients 5.30, 3.18 and 5.30 respectively, while the second-order endpoints $U_N^{(2)}$ and $L_N^{(2)}$ follow the second, with coefficients 7.12 and 2.62. At $s = 10$ the asymptotes are not yet reached and the curves lie above them, which is the visible size of the remainders.

exactly once. The first kind is what ke_k subtracts, and the second is enumerated by (73) with $T = \{i\} \cup T''$; monomials carrying a cube or a higher power occur on neither side. For (74), substituting the definition (72) gives

$$\begin{aligned} e_1 A_{N-1} + (N-1) A_N &= e_1^2 e_{N-2} - (N-1) e_1 e_{N-1} + (N-1) e_1 e_{N-1} - N(N-1) e_N \\ &= p_1^2 F_{N-2} - N(N-1) F_N, \end{aligned} \quad (76)$$

the two mixed terms cancelling. Multiplying by p_2 and using $f_N = (p_2/p_1^2)N(N-1)$ gives (75). $\square$

Identity (74) was checked over the whole verification grid to 2.1×10^{-53} , and (75) to 3.3×10^{-52} ; the smallest value of any A_k encountered is -5.8×10^{-59} , which is a round-off representation of an exact zero, namely A_8 for a six-orbital spectrum, where e_8 and e_7 both vanish identically (Table 5).

8 Interactions diagonal in the orbital basis

8.1 The model

Keep the spectrum (4) and attach to each orbital a pair energy $u_i \in \mathbb{R}$ that is paid once for every pair of particles sharing that orbital. A configuration ν then has the energy

$$E_u(\nu) = \sum_{i=1}^K \left[\varepsilon_i \nu_i + u_i \binom{\nu_i}{2} \right], \quad x_i = e^{-\beta \varepsilon_i}, \quad y_i = e^{-\beta u_i}, \quad (77)$$

and its Boltzmann weight factorises over orbitals, $w^{(u)}(\nu) = \prod_i x_i^{\nu_i} y_i^{\binom{\nu_i}{2}}$. This is the density-density, or on-site, part of a two-body interaction in a fixed single-particle basis: it is the atomic limit of a Hubbard model, and for a contact interaction in a trap it is what survives when the matrix elements between different orbitals are dropped. The interacting partition functions are

$$B_N^{(u)} = \sum_{|\nu|=N} w^{(u)}(\nu), \quad F_N^{(u)} = \sum_{\substack{|\nu|=N \\ \nu_i \in \{0,1\}}} w^{(u)}(\nu) = F_N, \quad M_N^{(u)} = B_N^{(u)} - F_N \geq 0. \quad (78)$$

The fermionic value is unchanged, because $\binom{\nu_i}{2} = 0$ whenever $\nu_i \in \{0,1\}$: exclusion makes the interaction invisible on the antisymmetric subspace. That is the entire reason a difference of the form $B - F$ remains a meaningful object here, and it is also the reason the extension stops at this class of interactions.

8.2 The injection survives

The map (22) needs no modification: $\nu \mapsto (i_*(\nu), \nu - 2e_{i_*})$ is still a bijection onto the pairs (i, ν') with $|\nu'| = N - 2$ and $\nu'_j \leq 1$ for $j < i$. What changes is the weight bookkeeping. Since

$$\binom{\nu'_i + 2}{2} = \binom{\nu'_i}{2} + 2\nu'_i + 1, \quad (79)$$

the two particles placed in orbital i contribute the factor $x_i^2 y_i^{2\nu'_i+1}$, where the exponent $2\nu'_i + 1$ counts the pairs they form with each other and with the ν'_i particles already present. Hence, with $D_i = \{\nu' : |\nu'| = N - 2, \nu'_j \leq 1 \text{ for } j < i\}$,

$$M_N^{(u)} = \sum_{i=1}^K x_i^2 \sum_{\nu' \in D_i} y_i^{2\nu'_i+1} w^{(u)}(\nu'), \quad (80)$$

an exact representation from which both bounds follow by estimating the middle factor from above and by restricting the inner sum from below. On D_i the occupation ν'_i ranges over $0, \dots, N - 2$, so $2\nu'_i + 1$ ranges over the odd integers from 1 to $2N - 3$.

Theorem 8.1 (interacting upper bound). *For every finite spectrum, every real on-site couplings u_i , every $\beta > 0$ and every $N \geq 2$,*

$$M_N^{(u)} \leq \left(\sum_{i=1}^K x_i^2 c_i \right) B_{N-2}^{(u)}, \quad c_i = \max(y_i, y_i^{2N-3}) = \begin{cases} y_i, & u_i \geq 0, \\ y_i^{2N-3}, & u_i \leq 0. \end{cases} \quad (81)$$

For repulsive couplings the prefactor is

$$\tilde{p}_2 = \sum_{i=1}^K x_i^2 y_i = \sum_{i=1}^K e^{-\beta(2\varepsilon_i + u_i)}, \quad (82)$$

the inequality is an equality at $N = 2$, and at $u = 0$ it reduces to $M_N \leq p_2 B_{N-2}$.

Proof. On D_i one has $1 \leq 2\nu'_i + 1 \leq 2N - 3$, so $y_i^{2\nu'_i + 1} \leq \max(y_i, y_i^{2N-3}) = c_i$: if $y_i \leq 1$ the largest power is the smallest exponent and if $y_i \geq 1$ the reverse. Substituting into (80) and extending the inner sum from D_i to all $(N - 2)$ -particle configurations, whose total weight is $B_{N-2}^{(u)}$, gives (81). At $N = 2$ the set D_i contains only the empty configuration, $c_i = y_i$ because $2N - 3 = 1$, and $B_0^{(u)} = 1$, so both sides are $\sum_i x_i^2 y_i$. Setting every $u_i = 0$ gives $c_i = 1$ and $B^{(0)} = B$, which is Theorem 4.1. $\square$

Theorem 8.2 (interacting lower bound). *For every finite spectrum, every real u_i , every $\beta > 0$ and every $N \geq 2$,*

$$M_N^{(u)} \geq L_N^{(u)} := \sum_{i=1}^K x_i^2 \left[y_i e_{N-2}^{(i)} + y_i^3 x_i e_{N-3}^{(i)} \right], \quad (83)$$

where $e_m^{(i)}$ is the elementary symmetric polynomial of degree m in the weights $\{x_j : j \neq i\}$, with $e_m^{(i)} = 0$ for $m < 0$ or $m > K - 1$. Equality holds at $N = 2$ and $N = 3$, and at $u = 0$ the bound reduces to $M_N \geq p_2 F_{N-2}$, the leading term of the free lower bound.

Proof. Restrict the inner sum in (80) to the fermionic images $E_i = \{\nu' \in D_i : \nu'_j \in \{0, 1\} \forall j\}$, a subset of D_i . On E_i the occupation ν'_i is 0 or 1, so $y_i^{2\nu'_i + 1}$ is y_i or y_i^3 . Splitting the sum over E_i accordingly, the configurations with $\nu'_i = 0$ are the subsets of $\{1, \dots, K\} \setminus \{i\}$ of size $N - 2$ and contribute $y_i e_{N-2}^{(i)}$, while those with $\nu'_i = 1$ contribute $y_i^3 x_i e_{N-3}^{(i)}$. The set E_i is empty when $N - 2 > K$, and (83) vanishes identically when $N - 3 > K - 1$, in which case it reads $M_N^{(u)} \geq 0$. At $N = 2$ only the first term survives and $e_0^{(i)} = 1$; at $N = 3$ both do, and together they enumerate every three-particle configuration with a double occupancy, including the triple occupancy $\nu_i = 3$, which corresponds to $\nu' = e_i$. At $u = 0$, $e_{N-2}^{(i)} + x_i e_{N-3}^{(i)} = e_{N-2}$ by the subset recursion, so $L_N^{(0)} = \sum_i x_i^2 e_{N-2} = p_2 F_{N-2}$. $\square$

The two theorems are the controlled analogue of Theorem 4.1 for this class of interactions: they hold at arbitrary coupling strength, they are exact at the smallest particle numbers, and they reduce to the free bounds when the interaction is switched off. For a uniform repulsion $u_i = U > 0$ the upper prefactor becomes $\tilde{p}_2 = e^{-\beta U} p_2$ whenever the u_i are all equal, so the bound on the weight of doubly occupied configurations is suppressed by the Boltzmann factor of the interaction energy — the qualitative statement that repulsion suppresses double occupancy, made quantitative without any assumption that U is small compared with $k_B T$ or with the level spacing.

8.3 Why the free lower bound does not transplant

The free lower bound came from a coefficientwise inequality for a product of series. The natural interacting factorisation is

$$B^{(u)}(t) - F(t) = F(t) \left[\prod_{i=1}^K R_i(t) - 1 \right], \quad R_i(t) = \frac{S_i(t)}{1 + x_i t}, \quad S_i(t) = \sum_{n \geq 0} x_i^n y_i^{\binom{n}{2}} t^n, \quad (84)$$

and it fails at the first non-trivial order.

Proposition 8.3 (cubic obstruction). *The coefficient of t^m in R_i is $r_m^{(i)} = x_i^m \sum_{k=0}^m (-1)^{m-k} y_i^{(k)}_i$. Then $r_1^{(i)} = 0$, $r_2^{(i)} = x_i^2 y_i \geq 0$ and*

$$r_3^{(i)} = x_i^3 (y_i^3 - y_i) = x_i^3 y_i (y_i^2 - 1), \quad (85)$$

which is strictly negative for every repulsive coupling $u_i > 0$. Hence $\prod_i R_i - 1$ has negative coefficients, and no lower bound obtained by truncating it coefficientwise can hold.

Proof. Expand $(1 + x_i t)^{-1} = \sum_m (-1)^m x_i^m t^m$ and collect the coefficient of t^m . Since $\binom{0}{2} = \binom{1}{2} = 0$, $r_1^{(i)} = x_i(1 - 1) = 0$; $r_2^{(i)} = x_i^2(1 - 1 + y_i) = x_i^2 y_i$; and $r_3^{(i)} = x_i^3(-1 + 1 - y_i + y_i^3)$, which is (85). For $u_i > 0$, $y_i < 1$ and $y_i^3 - y_i < 0$. At $y_i = 0.7$ the coefficient is $-0.357 x_i^3$. $\square$

The mechanism is physical rather than technical: with repulsion, the sequence of weights $y^{(n)}_i$ for $n = 0, 1, 2, 3$ is $1, 1, y, y^3$, and y^3 is much smaller than the value that an alternating reconstruction from the first two terms would require. The restricted injection of Theorem 8.2 avoids the issue by never forming the product.

8.4 Weak-coupling response

Although the bounds hold at any coupling, the first-order effect of the interaction has an exact closed form.

Proposition 8.4 (first-order response). *For every $N \geq 2$ and every orbital i ,*

$$\left. \frac{\partial M_N^{(u)}}{\partial u_i} \right|_{u=0} = -\beta x_i^2 T_{N-2}^{(i)}, \quad T_m^{(i)} = \sum_{|\nu|=m} \binom{\nu_i + 2}{2} \prod_{j=1}^K x_j^{\nu_j}, \quad (86)$$

so that $M_N^{(u)} = M_N - \beta \sum_i u_i x_i^2 T_{N-2}^{(i)} + R_N(\beta u)$ with the remainder controlled by

$$|R_N(\beta u)| \leq \frac{\beta^2}{2} e^{\beta \binom{N}{2} \sum_i |u_i|} \binom{N}{2}^2 \left(\sum_i |u_i| \right)^2 M_N. \quad (87)$$

In particular a repulsive coupling lowers M_N to first order, and the whole interacting interval stays within a neighbourhood of order $\beta^2 \|u\|^2$ of the free one.

The derivation of (86) and (87) is in Appendix C; the derivative was checked against a central difference of the exact enumeration to 1.2×10^{-23} at $N = 4$, $\beta = 1$ and four orbitals, the coefficient $T_{N-2}^{(i)}$ was checked against the generating-function form $T_m^{(i)} = [t^m] B(t)/(1 - x_i t)^2$ derived there to 1.0×10^{-59} , and the sign is fixed by $T_{N-2}^{(i)} > 0$: repulsion lowers the weight of multiply occupied configurations.

8.5 What the interacting bounds are worth

The two theorems are exact statements, and they are the only ones available here without further hypotheses, but their tightness depends strongly on the coupling sign and on the number of orbitals. Over the interacting grid of Section 9 — three spectra with $K = 2, 5, 8$ orbitals, four inverse temperatures, three deterministic draws of couplings in $[-0.8, 0.8]$ per spectrum and temperature,

Table 4: Interacting on-site extension: coverage and outcome of the deterministic verification of Theorems 8.1 and 8.2, Proposition 8.3 and Proposition 8.4. Couplings are drawn from a fixed deterministic generator in the interval $[-0.8, 0.8]$ per orbital; the signed sub-grids use the absolute values (repulsive) and their negatives (attractive). Relative slacks are worst cases over the grid and are large by construction where the number of orbitals is small.

quantity	value	coverage
trials of (81), (83)	468	3 spectra, 4 β , 3 draws, $N = 2..6$
violations	0	—
equality at $N = 2$, both bounds	residual 1.0×10^{-59}	36 cases
equality at $N = 3$, lower bound	residual 1.0×10^{-59}	36 cases
reduction to $p_2 F_{N-2}$, $p_2 B_{N-2}$ at $u = 0$	exact	60 cases
trivial lower bound ($N - 3 > K - 1$)	48 of 468	two orbitals, $N = 5, 6$
worst upper slack, mixed couplings	5.8	$N = 6$, $\beta = 1.5$, 5 levels
worst upper slack, repulsive	40	$N = 5$, $\beta = 1.5$, 8 levels
worst upper slack, attractive	1.3	$N = 5$, $\beta = 0.5$, 5 levels
worst lower slack (non-trivial)	1.000, 0.9995, 1.000	mixed, repulsive, attractive
first-order response (86)	residual 1.2×10^{-23}	$N = 4$, $\beta = 1$, 4 orbitals
cubic coefficient (85) at $y = 0.7$	$-0.357 x_i^3$	exact

particle numbers $N = 2, \dots, 6$, together with all-repulsive and all-attractive sub-grids — there were no violations of (81) or (83) in 468 trials, the bounds were equalities at $N = 2$ to 1.0×10^{-59} and at $N = 3$ to the same accuracy, and both reduced to the free bounds at $u = 0$.

The slack, however, is not small in general. The worst relative slack of the upper bound over that grid is 5.8 for mixed-sign couplings and 40 for purely repulsive ones; both occur at $N = 5$ or $N = 6$ at the coldest temperature of the interacting grid, $\beta = 1.5$, where the bosonic weight of the remaining $N - 2$ particles concentrates on a few configurations that the restriction $\nu'_j \leq 1$ for $j < i$ excludes. The worst relative slack of the lower bound is essentially 1 for all three coupling signs, meaning that the bound can be negligible compared with $M_N^{(u)}$: the fermionic images that (83) retains are a small part of D_i once the coupling or the temperature makes several particles share one orbital favourable. In 48 of the 468 trials, all of them on the two-orbital spectrum at $N \geq 5$, where $N - 3 > K - 1$ and both elementary polynomials vanish, the lower bound is the trivial statement $M_N^{(u)} \geq 0$. Both weaknesses are properties of the two estimates that produce the bounds — pointwise control of $y_i^{2\nu'_i+1}$ from above, and restriction to fermionic images from below — and neither is a defect in the inequalities themselves. Improving them is part of the open question in Section 10.

9 Deterministic verification

9.1 Protocol

Nothing in Sections 4–8 depends on a numerical result: each theorem is proved, and each asymptotic coefficient is derived in closed form. The computation reported here has two purposes, to corroborate the algebra against independent evaluation and to measure the asymptotic coefficients and the tightness of the bounds where the closed forms do not by themselves say how large the remainders are.

All values were computed in exact deterministic arithmetic at 200-bit floating point precision, about sixty decimal digits, with no sampling and no random input: where a coupling or a spectrum

is described as drawn, it comes from a fixed integer recurrence with a stated seed and is reproducible bit for bit. Each canonical partition function was obtained twice, once from the recurrences (9) and once by literal enumeration of the occupation configurations, and the two agree to 8.3×10^{-55} in the worst of 54 compared cases, which is round-off at the working precision. Every identity was then checked as an equality of two independently evaluated numbers, and every inequality as a signed test with a tolerance. The tolerances are 10^{-25} absolute for identities and 10^{-18} relative for inequalities: both are thirty orders of magnitude above the round-off of the arithmetic and many orders of magnitude below any quantity of physical interest. A case was counted as a violation only when the signed residual exceeded the tolerance after scaling by the magnitude of the quantities involved.

The grid is listed in full in Appendix D. It comprises six one-particle spectra with $K = 2, 4, 6, 8$ and two with $K = 27$ — a two-level system, the same with a two-fold degeneracy, a one-dimensional harmonic trap truncated at eight levels, a three-dimensional isotropic harmonic trap truncated at $n_x, n_y, n_z \leq 2$, a three-dimensional hard-wall box on the mode grid $n_x, n_y, n_z \in \{1, 2, 3\}$, and six unstructured levels drawn from a fixed deterministic generator; six inverse temperatures $\beta \in \{0.1, 0.25, 0.5, 1, 2, 4\}$ spanning the dilute and the degenerate regime; and particle numbers $N = 2, \dots, 8$, which for the two-level spectra extends beyond K into the range where $F_N = 0$ exactly. The continuum model spectrum was evaluated at degeneracies $g = 1, 2, 3$, dilutenesses $s = 10^2, \dots, 10^8$ and $N = 4, \dots, 8$. The interacting grid uses three spectra with $K = 2, 5, 8$, four inverse temperatures, three deterministic coupling draws per case in $[-0.8, 0.8]$ together with all-repulsive and all-attractive sub-grids, and $N = 2, \dots, 6$. Two limits were checked separately: the cold limit at $\beta = 30$ for a spectrum with a unique ground state, and the $\beta \rightarrow 0$ limit in exact rational arithmetic, where no floating point enters at all.

9.2 Results

Table 5 lists every check, its coverage and its worst residual. There were no violations anywhere on the grid: 252 evaluations of each of the ideal-gas inequalities, 1800 truncated instances of the hierarchy (61), 375 continuum evaluations, 468 interacting trials, 182 combinatorial pairs in exact rational arithmetic and 54 recurrence-versus-enumeration comparisons.

Three entries need a word of interpretation. Quantities that vanish identically for structural reasons — F_N when $N > K$, e_k when $k > K$, c_m for a single-orbital spectrum — evaluate to numbers of magnitude 10^{-59} whose sign is decided by round-off; the smallest value of any A_k encountered in Proposition 7.4 is -5.8×10^{-59} , and it is an exact zero, not a counterexample. The continuum row reports the deviation of a measured leading coefficient from its closed form, which is large at the smallest diluteness simply because the remainder terms have not yet died away: at $s = 10^2$ the deviation is 0.87 for the second-order upper endpoint, at $s = 10^6$ it is 8.7×10^{-5} and at $s = 10^8$ it is below 10^{-6} , which is the numerical content of the $o(s^{-1})$ and $o(s^{-2})$ claims. The first-order response row compares the closed form (86) with a central difference of the exact enumeration at step $h = 10^{-12}$, where the truncation error of the difference is of order h^2 and the round-off is of order $10^{-60}/h$.

A second, independent implementation of the same recurrences and bounds, written before the proofs were assembled and run at a different working precision, reproduces every entry of Table 5 within the ratio of the two precisions [1]. The tightness figures differ between the two runs only because the grids differ, and both are reported here: on the grid of Appendix D the tightest relative slack attained anywhere in the prefactor interval (36) is 3.78×10^{-2} , at $N = 2$ for the three-dimensional harmonic trap at $\beta = 0.1$, where Corollary 5.4 identifies it as the value of q ; restricted to $N \geq 4$ the tightest slack is 0.21 on the lower side and 0.24 on the upper side, again

Table 5: Deterministic verification at 200-bit precision over the grid of Appendix D. “Residual” is the worst absolute deviation of an exact identity, or the number of signed-test violations for an inequality, over the stated coverage. Round-off at this precision is of order 10^{-60} relative, so residuals at the level of 10^{-52} are zeros.

statement checked	coverage	worst residual
recurrences (9) against direct enumeration	54 cases, $N \leq 4$	8.3×10^{-55}
classical window, Lemma 2.2	252 cases	0 violations
logarithmic error, Corollary 2.3	201 cases with $F_N > 0$	0 violations
two-sided bound, Theorem 4.1	252 cases	0 violations
equality at $N = 2, 3$	72 cases	9.6×10^{-58}
small- N values (19), (25)	108 cases	1.0×10^{-56}
prefactor interval, Theorem 5.3	252 cases	0 violations
$N = 2$ symmetric gap, Corollary 5.4	36 cases	1.7×10^{-58}
$1/K \leq q \leq 1$	36 spectra and temperatures	$q \in [0.0378, 0.99996]$
overcount, Theorem 6.1	252 cases	2.0×10^{-54}
two forms of c_m , Corollary 6.3	108 coefficients	2.0×10^{-56}
c_m against literal weight enumeration	54 coefficients	1.0×10^{-58}
refined injection endpoint, Corollary 6.2	180 cases, $N \geq 4$	0 violations
equality at $N = 4, 5$	72 cases	6.1×10^{-56}
prefactor slack, Theorem 6.4	252 cases	1.3×10^{-52}
exact decomposition, Theorem 6.5	252 cases	7.8×10^{-54}
hierarchy, Corollary 6.6	1800 truncated instances	0 violations
second-order endpoints, Theorems 7.1, 7.2	180 cases, $N \geq 4$	0 violations
monomial certificate, Proposition 7.4	252 cases	3.3×10^{-52}
continuum coefficients, Theorem 7.3	375 evaluations	8.7×10^{-5} for $s \geq 10^6$
literal prefactor, Proposition 5.5	21 evaluations	1.1×10^{-3} at $s = 10^4$
$N = 2$ witness, Proposition 5.6	10 values of s	sign change at $2^{-1/2}$
cold limit, Proposition 4.2	$N = 2, \dots, 8$ at $\beta = 30$	1.4×10^{-13}
combinatorial limit, Corollary 4.3	182 pairs, exact rationals	0 failures
interacting bounds, Theorems 8.1, 8.2	468 trials	0 violations
interacting reduction at $u = 0$	60 cases	exact
interacting generating function $T_m^{(i)}$, Section 8	4 cases	1.0×10^{-59}
first-order response, Proposition 8.4	central difference, $h = 10^{-12}$	1.2×10^{-23}

at the most dilute point of the grid. The tightest value of the injection bound (20) on the upper side is 1.1×10^{-10} , at $\beta = 4$ for the three-dimensional box at $N = 4$, which is the low-temperature exactness of Proposition 4.2 appearing inside the main grid.

9.3 What the bounds are worth away from the dilute limit

Figure 2 shows the relative slacks of all six ideal-gas endpoints for a single concrete discrete spectrum, the three-dimensional isotropic harmonic trap truncated at $n_x, n_y, n_z \leq 2$, so $K = 27$ levels $\varepsilon = n_x + n_y + n_z$ with multiplicities 1, 3, 6, 7, 6, 3, 1. At $\beta = 0.25$ the trap is dilute: $p_1 = Z_1(\beta) = 13.57$, meaning that the thermal population is spread over roughly fourteen of the twenty-seven orbitals, and $q = 4.2 \times 10^{-2}$. The second-order interval $[L_N^{(2)}, U_N^{(2)}]$ has relative width 1.4×10^{-2} at $N = 4$, 4.5×10^{-2} at $N = 5$ and 5.7×10^{-1} at $N = 8$, and at $N = 4$ the six endpoints span $0.77 M_N$ to $1.27 M_N$, the whole of that range being carried by the two prefactor ones: the injection and second-order endpoints already lie between $0.956 M_N$ and $1.039 M_N$. At $\beta = 1$ the same spectrum has $p_1 = 3.40$ and $q = 0.133$: the thermal population covers only a

few orbitals, the gas is degenerate at every $N \geq 4$, and the relative width of the same interval grows from 0.25 at $N = 4$ to 5.4 at $N = 8$. The prefactor endpoints deteriorate fastest, because $f_N = qN(N - 1)$ grows quadratically in N while q has already risen by a factor of three; the injection endpoints deteriorate more slowly, and on the upper side they remain the sharper of the two families throughout, at 0.23 against 0.75 at $\beta = 0.25$, $N = 6$ and at 0.29 against 3.1 at $\beta = 1$, $N = 6$.

One ordering in the figure is the reverse of the asymptotic one and deserves to be stated. In the dilute continuum limit the second-order endpoints are the sharpest available, by Theorem 7.3, but on a discrete spectrum at fixed N with q not small they are built on the prefactor family and inherit its factor f_N : at $\beta = 1$, $N = 8$ the second-order upper endpoint has a relative slack of 4.7, ten times worse than the injection upper endpoint at 0.45. Which family to use is therefore decided by the value of $qN(N - 1)$, and the two are complementary rather than ordered. Nothing in the paper claims otherwise: Theorem 7.3 is an asymptotic statement about the model continuum spectrum, while Theorems 7.1 and 7.2 are exact for every spectrum and are simply not sharp when f_N is large.

The figure is therefore a picture of the boundary of the result rather than of a failure. Every curve in it is a valid bound at every point, and none of them crosses M_N ; what changes is how much information the bound carries. The control parameter is the product $qN(N - 1) = f_N$: the interval is informative when $f_N \ll 1$, which for a discrete spectrum means that the number of thermally accessible orbitals is large compared with N^2 , and it degenerates when that stops being true. At $\beta = 0.25$ this product is 0.50 at $N = 4$ and 2.34 at $N = 8$; at $\beta = 1$ it is 1.60 and 7.45 at the same two particle numbers, and the ordering of the six curves in the two panels follows it. This is the same condition that makes the classical counting usable in the first place, so the bounds are sharp exactly where they are needed, and the second-order endpoints extend the useful range by the factor computed in (71).

10 Discussion

10.1 What has been removed from the hypotheses

Kroemer's bound was derived for a spinless, non-interacting monatomic gas in a three-dimensional continuum box, in the classical limit, and in a system large enough that removing two particles does not change the chemical potential [3]. Every one of those five conditions is a hypothesis about the spectrum or about the size of the system, and none of them is needed for Theorem 4.1. What replaces them is a single object, the one-particle spectrum, which enters the bounds only through the numbers $Z_1(k\beta)$, $k = 1, \dots, N$. Degeneracies — including spin and orbital degeneracy — are repeated levels of that spectrum, so the extension to degenerate systems is not an extra argument but the absence of a restriction; the same is true of any confining geometry, which fixes the ε_i and nothing else. Small systems are covered at every $N \geq 2$ with no reference to a reservoir or a chemical potential, since all partition functions here are canonical at fixed N . The price of this generality is visible in the prefactor: $f_N = qN(N - 1)$ depends on the spectrum through $q = Z_1(2\beta)/Z_1(\beta)^2$, whereas Kroemer's f was the number $2^{-3/2}/n_Q V$ because the continuum fixes q to $2^{-3/2}$. The continuum value is recovered exactly in the limit (39), and Proposition 5.5 shows that the finite- N factor $(N - 1)/N$ cannot be dropped: it is not a refinement of the statement but part of it, and a prefactor proportional to N^2 satisfies the lower advertisement for no dilute gas whatsoever once N is fixed.

The witness of Proposition 5.6 sharpens this. In the continuum at $N = 2$ the literal prefactor

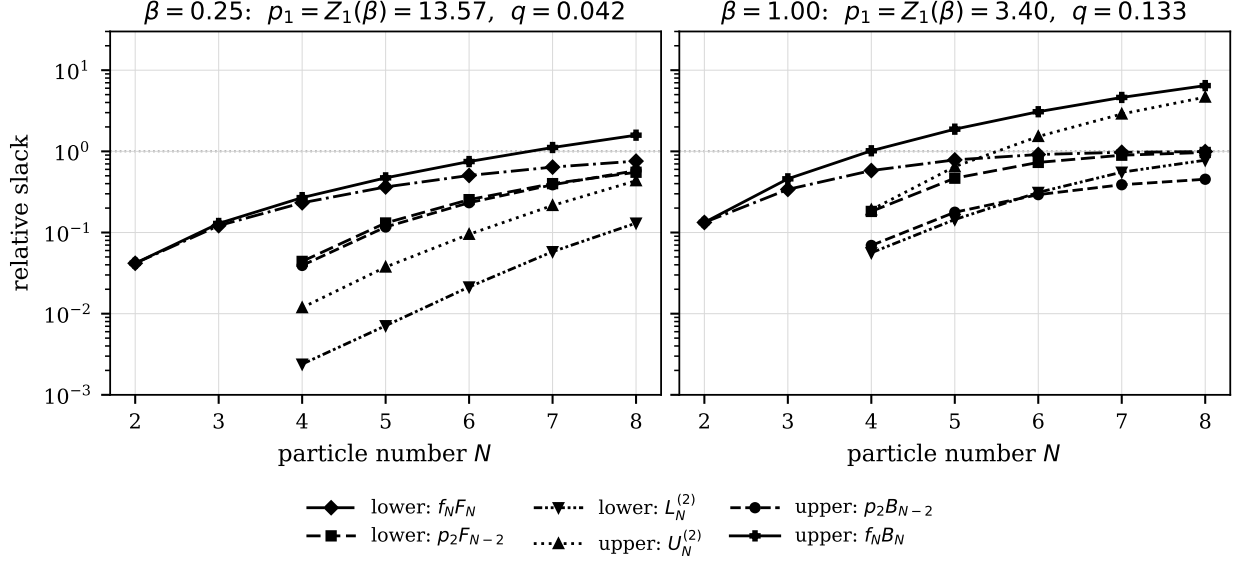


Figure 2: Relative slack of the six ideal-gas endpoints against M_N for a three-dimensional isotropic harmonic trap with 27 levels $\varepsilon = n_x + n_y + n_z \leq 2$, as a function of the particle number N at two temperatures; the vertical axis is logarithmic, upper endpoints are plotted as $(\text{endpoint} - M_N)/M_N$ and lower endpoints as $(M_N - \text{endpoint})/M_N$, and every value is computed from the exact canonical recurrences at 200-bit precision. The two injection endpoints $p_2 F_{N-2}$ and $p_2 B_{N-2}$ coincide with M_N at $N = 2, 3$ by Theorem 4.1 and are omitted there rather than plotted at zero. The dotted horizontal line marks a relative slack of one, that is, an endpoint that carries no information. At $\beta = 0.25$ (left), where $q = 4.2 \times 10^{-2}$ and f_N reaches 2.34 at $N = 8$, the six endpoints span $0.77 M_N$ to $1.27 M_N$ at $N = 4$, the four non-prefactor ones lying between $0.956 M_N$ and $1.039 M_N$; at $N = 8$ the worst of them is the lower prefactor endpoint, at $0.24 M_N$, and the upper prefactor endpoint stands at $2.58 M_N$. At $\beta = 1$ (right), where $q = 0.133$ and f_N reaches 7.45, the two prefactor endpoints already span $0.42 M_N$ to $2.02 M_N$ at $N = 4$ and $0.006 M_N$ to $7.46 M_N$ at $N = 8$; there the upper injection endpoint, at $1.45 M_N$, is the closest of the six to M_N , the second-order endpoints span $0.22 M_N$ to $5.66 M_N$, and the lower injection endpoint, at $0.031 M_N$, degenerates together with the fermionic weight F_{N-2} that carries it.

obeys the upper advertisement for every value of $n_Q V$ but violates the lower one as soon as $n_Q V > 2^{-1/2}$, which is to say in the entire regime in which one would ask whether the classical counting is legitimate. The failure is therefore not confined to the deep quantum regime: it is the regime where the bound is meant to be informative. The exact discrete prefactor removes it and restores the symmetric gap of Corollary 5.4, in which both slacks are equal to q by an identity rather than by an estimate.

10.2 How much of the slack is left

The interval $[f_N F_N, f_N B_N]$ has a relative width of order f_N , so it is informative exactly when $qN(N-1) \ll 1$, that is, when the number of thermally accessible one-particle states is large compared with the square of the particle number. Two independent mechanisms narrow it. The injection mechanism — counting configurations by their first doubly occupied orbital — gives an upper endpoint $p_2 B_{N-2}$ and a lower endpoint $p_2 F_{N-2}$ that are exact at $N = 2, 3$ and, at low

temperature, asymptotically exact on the upper side (Proposition 4.2). The prefactor mechanism gives the second-order endpoints $U_N^{(2)}$ and $L_N^{(2)}$, which in the dilute three-dimensional limit reduce the leading relative error from order s^{-1} to order s^{-2} with the closed-form coefficients of Theorem 7.3: the improvement factor is $(4N - 6)/(N(N - 1))$ on the upper side, approaching $4/N$ for large N , and on the lower side the second-order endpoint beats $f_N F_N$ by a factor that grows linearly in the diluteness, the numerical ratio of the two leading coefficients being 6.0 at $N = 4$ and 0.77 at $N = 8$. Both mechanisms are exact statements about the same quantity, and neither subsumes the other: the injection endpoints are sharper at low temperature, where the bosonic weight is concentrated on the ground orbital, and the second-order endpoints are sharper in the dilute limit at fixed N .

What remains unexplained is the full asymptotic expansion. The measured coefficients of Table 3 agree with the closed forms to five or six significant digits at $s = 10^6$ and to all displayed digits at $s = 10^8$, which is consistent with the existence of an expansion in powers of s^{-1} with rational coefficients built from the even power sums, but nothing here proves that such an expansion exists beyond the two orders displayed. The monomial certificate of Proposition 7.4 identifies the natural objects for a further step: the lower prefactor slack is $p_2[e_1 A_{N-1} + (N - 1)A_N]/p_1^2$ with $A_k = e_1 e_{k-1} - k e_k \geq 0$, so the question of how much smaller the lower endpoint can be made is a question about the size of the Newton-type coefficients A_k , which are zero exactly when the spectrum is flat.

10.3 Boundaries of the result

Four limitations bound the reach of every theorem above, and each is worth stating precisely rather than leaving to the reader.

The discrete statements assume $K < \infty$, or more generally $Z_1(\beta) < \infty$. A true continuum has no occupation-number basis in the sense used here, and the model spectrum $p_k = s k^{-3/2}$ of (16) is a discrete proxy for it, chosen so that the power sums reproduce the three-dimensional classical values $Z_1(k\beta) = k^{-3/2} Z_1(\beta)$ exactly. The asymptotic statements are therefore statements about that proxy in a limit in which it becomes dense; they are not a derivation from a continuum Hamiltonian, and nothing here addresses the thermodynamic limit $N, V \rightarrow \infty$ at fixed density, in which s stays finite and the dilute expansions do not apply.

The bounds concern the difference $B_N - F_N$ of two exactly solvable canonical partition functions, and nothing is claimed about the approach to Bose–Einstein condensation, about the grand-canonical ensemble, or about fluctuations of the occupation numbers. The quantity M_N is the total weight of configurations that are not single-valued in the orbitals; it is the exact measure of how much the distinguishable counting overcounts, but it is not itself a physical observable.

Theorem 8.1 and Theorem 8.2 cover pair energies $u_i \binom{\nu_i}{2}$ that are diagonal in the same basis as the one-particle spectrum. This class includes the atomic limit of a Hubbard model and the on-site part of a contact interaction, at arbitrary coupling strength and sign. It excludes every interaction with off-diagonal matrix elements, and it excludes interacting fermions: the whole construction rests on $F_N^{(u)} = F_N$, which holds because exclusion annihilates $\binom{\nu_i}{2}$, and a fermion–fermion interaction changes the reference value against which M_N is measured. For orbital-mixing couplings the injection survives as a map but not as a weight estimate, since the two particles moved out of orbital i_* no longer carry a factorisable weight; whether a controlled two-sided bound exists there is open.

The interacting bounds are valid everywhere on the grid of Section 9 and exact at $N = 2, 3$, but they are loose in the mean: the worst observed relative slack of the upper bound is 40 for

repulsive couplings and the lower bound can be negligible compared with $M_N^{(u)}$. Neither weakness is a failure of the theorems; both come from the two estimates in their proofs, pointwise control of $y_i^{2\nu'_i+1}$ and restriction to fermionic images, and both would be reduced by any sharper handle on the distribution of ν'_i inside D_i .

10.4 Why the discrete statement is the right one

A bound whose prefactor is a number is more portable than one whose prefactor is a function of the spectrum, and Kroemer's form has that advantage. The results here suggest the opposite ranking. The discrete statement is proved once, for all spectra, and each physical situation is then read off by evaluating $Z_1(\beta)$ and $Z_1(2\beta)$: a harmonic trap, a box, a two-level system and a degenerate multiplet are all special cases of the same theorem, and the correction to the continuum prefactor is not an asymptotic refinement to be estimated but the exact factor $(N-1)/N$ times the exactly computable q . For the small systems in which counting arguments are usually deployed — cold atoms in a trap, few-electron quantum dots, nucleons in a light nucleus — that factor is not close to one, and the accompanying slack decompositions say how much of the remaining error is structural overcount and how much is prefactor approximation. The cost is a prefactor that must be evaluated rather than recited, which for any finite spectrum is a sum of exponentials.

One further point of contact with the original note deserves to be made explicit, because it explains what a bound of this shape is for. Kroemer's own summary of his result is that the classical value differs from the exact one by what can be a large factor even well inside the classical regime, and that this nevertheless leaves the observables alone, because they are derivatives of $\log Z$ rather than of Z , and the fractional error of the logarithm is negligible [3]. Corollary 2.3 is the exact discrete form of that second half: the absolute error of $\log Z$ is at most $\log(1 + M_N/F_N)$, so any upper bound on the ratio M_N/F_N turns into a bound on the free energy with no further work, and in the dilute model spectrum Theorem 7.3 supplies one: the bound is $f_N + o(f_N)$ against a $\log Z$ of size $N \log s$. The two halves of the paper therefore answer different questions: the prefactor and its sharpenings measure how wrong the counting of Z can be, while Corollary 2.3 measures how much of that wrongness survives into a quantity one can observe. Nothing here bounds the temperature derivatives of M_N , so the transfer stops at $\log Z$ itself.

11 Conclusion

The net weight of multiply occupied configurations, $M_N = B_N - F_N$, is bracketed for every finite one-particle spectrum, every temperature and every $N \geq 2$ by

$$p_2 F_{N-2} \leq M_N \leq p_2 B_{N-2}, \quad f_N F_N \leq M_N \leq f_N B_N, \quad f_N = \frac{p_2}{p_1^2} N(N-1), \quad (88)$$

with $p_k = Z_1(k\beta)$. Both pairs are exact at $N = 2, 3$; the second pair has the shape of Kroemer's bound and holds without the continuum, spinlessness, large-system or thermodynamic hypotheses, with degeneracy carried by repeated levels. The prefactor is Kroemer's with N^2 replaced by $N(N-1)$, and that replacement is necessary: in the three-dimensional continuum a prefactor proportional to N^2 violates the lower side for every diluteness above $2^{-1/2}$ at $N = 2$, whereas the exact prefactor leaves a symmetric gap equal to $q p_2$ on both sides.

The remaining slack is decomposed exactly rather than estimated. The first-double-occupation overcount is a series $\sum_{m \geq 2} c_m F_{N-2m}$ with explicitly non-negative coefficients given by a finite recurrence in the even power sums alone; the prefactor slack is a series $(p_2/p_1^2) \sum_{j \geq 2} d_j B_{N-j}$ with

$d_j \geq 0$; and the two together yield the identity $M_N = f_N B_N - (p_2/p_1^2) \sum_j d_j B_{N-j} - \sum_m c_m F_{N-2m}$, so that every partial retention of the two sums is a valid upper bound and the endpoints form a monotone hierarchy. Keeping the first term of each series gives the second-order endpoints, whose relative error in the dilute three-dimensional limit falls as s^{-2} instead of s^{-1} , with closed-form leading coefficients verified numerically to five and six digits. On the lower side the prefactor slack admits the monomial certificate $p_1^2(p_2 F_{N-2} - f_N F_N) = p_2[e_1 A_{N-1} + (N-1)A_N]$ with $A_k = e_1 e_{k-1} - k e_k \geq 0$.

Pair energies diagonal in the one-particle basis are included at arbitrary coupling strength and sign, with the fermionic reference value unchanged because exclusion annihilates the on-site term; the interacting bounds are exact at $N = 2, 3$, reduce to the free ones at zero coupling, and have an exact first-order weak-coupling response. A factorised lower bound of the free type is ruled out at cubic order for repulsive couplings, where the corresponding series coefficient is strictly negative.

Every statement was checked in exact arithmetic at 200-bit precision against independent evaluation, with no violations anywhere on the grid of Appendix D, and the asymptotic coefficients were confirmed against the exact recurrences. Three questions remain open: whether an asymptotic expansion of $M_N/f_N B_N$ exists beyond the two orders established here, whether the interacting bounds can be tightened by a better handle on the occupation of the injected orbital, and whether controlled two-sided bounds exist for interactions with off-diagonal matrix elements, or for interacting fermions, where the reference value F_N is no longer interaction-independent.

A Recurrences, symmetric polynomials and the enumeration

A.1 Proof of the canonical recurrences

Write $B(z) = \sum_{N \geq 0} B_N z^N$ and $F(z) = \sum_{N \geq 0} F_N z^N$ as in (13). For the bosonic series,

$$\frac{zB'(z)}{B(z)} = z \frac{d}{dz} \sum_{i=1}^K [-\log(1 - x_i z)] = \sum_{i=1}^K \frac{x_i z}{1 - x_i z} = \sum_{i=1}^K \sum_{k \geq 1} x_i^k z^k = \sum_{k \geq 1} p_k z^k, \quad (89)$$

where the exchange of the two sums is legitimate because every term is non-negative for $z \geq 0$ and the i -sum is finite. Since $zB'(z) = \sum_{n \geq 1} n B_n z^n$, multiplying (89) by $B(z)$ and comparing the coefficient of z^n gives

$$n B_n = \sum_{k=1}^n p_k B_{n-k}, \quad (90)$$

the first half of (9); the terms with $k > n$ cannot contribute because B has no negative powers. For the fermionic series the same computation runs with $1 + x_i z$ in place of $(1 - x_i z)^{-1}$:

$$\frac{zF'(z)}{F(z)} = \sum_{i=1}^K \frac{x_i z}{1 + x_i z} = \sum_{i=1}^K \sum_{k \geq 1} (-1)^{k+1} x_i^k z^k = \sum_{k \geq 1} (-1)^{k+1} p_k z^k, \quad (91)$$

and comparing coefficients gives $n F_n = \sum_{k=1}^n (-1)^{k+1} p_k F_{n-k}$. Both recurrences are therefore consequences of the product form of the generating functions alone, and both are exact: no approximation, no large- N limit and no continuum enters.

The recurrences determine B_n and F_n for every n from $B_0 = F_0 = 1$ and the n numbers $p_1, \dots, p_n$, at a cost of $O(n^2)$ arithmetic operations per particle number. This is the computational reason a discrete spectrum is no harder to handle than a continuum one: the canonical thermodynamics of N particles requires only the one-particle partition function evaluated at the N inverse temperatures $\beta, 2\beta, \dots, N\beta$.

A.2 The symmetric-polynomial dictionary

Expanding the products in (13) term by term identifies the coefficients with the two classical families of symmetric polynomials in $x_1, \dots, x_K$:

$$\prod_{i=1}^K \frac{1}{1 - x_i z} = \prod_{i=1}^K \sum_{\nu_i \geq 0} x_i^{\nu_i} z^{\nu_i} = \sum_{N \geq 0} \left(\sum_{|\nu|=N} x^\nu \right) z^N = \sum_{N \geq 0} h_N(x) z^N, \quad (92)$$

$$\prod_{i=1}^K (1 + x_i z) = \sum_{N \geq 0} \left(\sum_{|T|=N} \prod_{i \in T} x_i \right) z^N = \sum_{N \geq 0} e_N(x) z^N, \quad (93)$$

so that $B_N = h_N$ and $F_N = e_N$. The two sums in brackets are precisely the occupation-configuration sums of (6): the first runs over all vectors ν of non-negative integers with $|\nu| = N$, the second over subsets T of size N , which is the same as restricting to $\nu \in \{0, 1\}^K$. Two consequences are used throughout and are worth isolating. First, $e_N = F_N = 0$ for $N > K$, because there is no subset of size larger than K ; this is an exact zero, not a small number, and it is what makes $M_N = B_N$ beyond the Pauli limit. Second, h_N and e_N are related by the involution that negates the odd power sums, which is the algebraic content of the sign difference between (89) and (91) and of the even-only exponent Δ in (14).

The recurrences (9) are Newton's identities for h_n and e_n written in physical notation. The same identities in their standard form are

$$n e_n = \sum_{j=1}^n (-1)^{j-1} p_j e_{n-j}, \quad n h_n = \sum_{j=1}^n p_j h_{n-j}, \quad p_n = \sum_{j=1}^{n-1} (-1)^{j-1} e_j p_{n-j} + (-1)^{n-1} n e_n, \quad (94)$$

and the third of these, which expresses a power sum through the elementary polynomials, is the one that underlies the power-sum form (54) of the overcount coefficients.

A.3 Independent enumeration

Every exact statement in this paper was evaluated twice, once from (9) and once from the literal configuration sums,

$$B_N = \sum_{\substack{\nu \in \mathbb{Z}_{\geq 0}^K \\ |\nu|=N}} \exp\left(-\beta \sum_i \varepsilon_i \nu_i\right), \quad F_N = \sum_{\substack{\nu \in \{0,1\}^K \\ |\nu|=N}} \exp\left(-\beta \sum_i \varepsilon_i \nu_i\right), \quad (95)$$

generated by recursive composition of N out of K in the second case and of all weak compositions in the first, with each Boltzmann factor accumulated in full working precision rather than in machine double precision. The two routes share no algebra: (95) never uses a generating function, and (9) never enumerates a configuration. Their agreement to 8.3×10^{-55} in the worst of 54 comparisons over the grid of Appendix D is therefore a check of the algebra rather than of the arithmetic. The same enumeration supplies the literal weight form of the overcount coefficients c_m in (47), which was compared with the power-sum form (54) in Table 5.

B Continuum expansions of the model spectrum

This appendix derives every asymptotic statement about the model continuum spectrum (16): the expansions (40) and (41) used in Proposition 5.5, and the four leading coefficients of Theorem 7.3.

Throughout, $s = n_Q V \rightarrow \infty$ at fixed particle number, and $p_k = s k^{-3/2}$; an internal degeneracy g replaces s by gs everywhere and changes nothing else. Put

$$a_k = k^{-5/2}, \quad a_1 = 1, \quad a_2 = 2^{-5/2}, \quad a_3 = 3^{-5/2} = \frac{\sqrt{3}}{9}, \quad a_4 = 2^{-5}, \quad (96)$$

and

$$T = \frac{2^{-3/2} s^{N-1}}{(N-2)!}, \quad (97)$$

which is the common scale of every gap computed below.

B.1 Three subleading orders of B_n and F_n

From (13) and (5),

$$\log B(z) = \sum_{k \geq 1} \frac{p_k}{k} z^k = sA(z), \quad \log F(z) = \sum_{k \geq 1} \frac{(-1)^{k+1} p_k}{k} z^k = sA_-(z), \quad A(x) = \sum_{k \geq 1} a_k x^k, \quad (98)$$

and A_- is A with a_k replaced by $(-1)^{k+1} a_k$, so that a_2 and a_4 change sign while a_3 does not. Expanding the exponential,

$$B_n(s) = [x^n] e^{sA(x)} = \sum_{m=0}^n \frac{s^m}{m!} [x^{n-m}] (1 + a_2 x + a_3 x^2 + a_4 x^3 + \dots)^m, \quad (99)$$

because $A = x(1 + a_2 x + a_3 x^2 + \dots)$. Only $m = n, n-1, n-2, n-3$ contribute to the four leading orders, and the coefficient in brackets counts the ways of writing $n-m$ as an ordered sum of parts taken from $\{1, 2, 3\}$ with weights a_2, a_3, a_4 :

$$\begin{aligned} [x^0](\dots)^n &= 1, & [x^1](\dots)^{n-1} &= (n-1)a_2, \\ [x^2](\dots)^{n-2} &= \binom{n-2}{2} a_2^2 + (n-2)a_3, \\ [x^3](\dots)^{n-3} &= \binom{n-3}{3} a_2^3 + (n-3)(n-4)a_2 a_3 + (n-3)a_4. \end{aligned} \quad (100)$$

Multiplying by $s^m/m!$ and dividing by $s^n/n!$ converts each line into a polynomial in n times a power of s^{-1} , giving

$$B_n(s) = \frac{s^n}{n!} \left[1 + \frac{b_1(n)}{s} + \frac{b_2(n)}{s^2} + \frac{b_3(n)}{s^3} + O(s^{-4}) \right], \quad (101)$$

with

$$\begin{aligned} b_1(n) &= a_2 n(n-1), \\ b_2(n) &= \left[\frac{1}{2}(n-3)a_2^2 + a_3 \right] n(n-1)(n-2), \\ b_3(n) &= \left[\frac{1}{6}(n-4)(n-5)a_2^3 + (n-4)a_2 a_3 + a_4 \right] n(n-1)(n-2)(n-3), \end{aligned} \quad (102)$$

each vanishing when its factorial factor does. The fermionic expansion follows from the same computation with $a_2, a_4 \mapsto -a_2, -a_4$:

$$F_n(s) = \frac{s^n}{n!} \left[1 - \frac{b_1(n)}{s} + \frac{b_2(n)}{s^2} - \frac{b_3(n)}{s^3} + O(s^{-4}) \right]. \quad (103)$$

The sign pattern is the substance of the parity argument: b_1 and b_3 are odd under the replacement and b_2 is even, because b_2 receives a_2 only through a_2^2 and a_3 only linearly. Truncating (101) and (103) after b_1 gives (40), since $b_1(n)/n! = a_2/(n-2)!$. Subtracting them gives

$$M_N(s) = \frac{s^N}{N!} \left[\frac{2b_1(N)}{s} + \frac{2b_3(N)}{s^3} + O(s^{-5}) \right] = T \left[1 + \frac{m(N)}{s^2} + O(s^{-4}) \right], \quad (104)$$

where the first equality uses $2b_1(N)s^{N-1}/N! = 2a_2s^{N-1}/(N-2)! = T$ because $2a_2 = 2^{-3/2}$, and

$$m(N) = \frac{b_3(N)}{b_1(N)} = (N-2)(N-3) \left[\frac{1}{6}(N-4)(N-5)a_2^2 + (N-4)a_3 + \frac{a_4}{a_2} \right] \quad (105)$$

with the numerical values $a_2^2/6 = 1/192$, $a_3 = \sqrt{3}/9$ and $a_4/a_2 = 2^{-5/2} = \sqrt{2}/4$. Equation (104) is (41) together with its first correction; the absence of an s^{N-2} term is what makes $M_N/T = 1 + O(s^{-2})$ and is used repeatedly below.

B.2 The literal prefactor

Kroemer's literal prefactor and the exact one of (35) are, in this model,

$$f = \frac{N^2}{2^{3/2}s}, \quad f_N = qN(N-1) = \frac{2^{-3/2}N(N-1)}{s}, \quad \frac{f}{f_N} = \frac{N}{N-1}, \quad (106)$$

a ratio that is a pure number, independent of the spectrum. From (101) and (104), using $f_N B_N = T[1 + b_1(N)s^{-1} + \dots]$ and $f_N F_N = T[1 - b_1(N)s^{-1} + \dots]$,

$$\frac{fB_N - M_N}{M_N} = \frac{\frac{N}{N-1} \left[1 + \frac{b_1(N)}{s} + O(s^{-2}) \right] - \left[1 + O(s^{-2}) \right]}{1 + O(s^{-2})} = \frac{1}{N-1} + \frac{N b_1(N)}{(N-1)s} + O(s^{-2}), \quad (107)$$

$$\frac{M_N - fF_N}{M_N} = \frac{\left[1 + O(s^{-2}) \right] - \frac{N}{N-1} \left[1 - \frac{b_1(N)}{s} + O(s^{-2}) \right]}{1 + O(s^{-2})} = -\frac{1}{N-1} + \frac{N b_1(N)}{(N-1)s} + O(s^{-2}). \quad (108)$$

These are the two limits of Proposition 5.5 with their first corrections: the literal prefactor over-advertises on the upper side and under-advertises on the lower side by the same relative amount $1/(N-1)$, which does not vanish as the gas dilutes, while the corrections $Nb_1(N)/[(N-1)s] = 2^{-5/2}N^2/s$ do. The entries of Table 1 are exact recurrences rather than these truncated series, which is why they approach $\pm 1/(N-1)$ only as s grows.

B.3 First-order gaps

The two ingredients beyond (101) are

$$q = \frac{p_2}{p_1^2} = \frac{2^{-3/2}}{s}, \quad c_2 = \frac{p_2^2 - p_4}{2} = \frac{s^2}{16} - \frac{s}{16}, \quad \frac{F_{N-4}}{T} = \frac{2^{3/2}(N-2)(N-3)}{s^3} \left[1 + O(s^{-1}) \right], \quad (109)$$

the last one being (112) below at $k = 4$. Since $f_N B_N/T = 1 + b_1(N)s^{-1} + O(s^{-2})$ and $M_N/T = 1 + O(s^{-2})$,

$$\frac{f_N B_N - M_N}{T} = \frac{b_1(N)}{s} + O(s^{-2}) = \frac{a_2 N(N-1)}{s} + O(s^{-2}), \quad (110)$$

and, using $c_2 F_{N-4}/T = (s^2 - s)2^{3/2}(N-2)(N-3)/(16s^3)[1 + O(s^{-1})] = a_2(N-2)(N-3)s^{-1} + O(s^{-2})$ because $2^{3/2}/16 = a_2$,

$$\frac{f_N B_N - c_2 F_{N-4} - M_N}{T} = \frac{a_2 [N(N-1) - (N-2)(N-3)]}{s} + O(s^{-2}) = \frac{a_2(4N-6)}{s} + O(s^{-2}). \quad (111)$$

Dividing (110) and (111) by $M_N/T = 1 + O(s^{-2})$ and using $a_2 = 2^{-5/2} = 1/(4\sqrt{2})$ gives (67) and (68). The same two lines in unscaled form are $f_N B_N - M_N = N(N-1)s^{N-2}/[16(N-2)!] + O(s^{N-3})$ and $f_N B_N - c_2 F_{N-4} - M_N = (4N-6)s^{N-2}/[16(N-2)!] + O(s^{N-3})$.

B.4 Second-order gaps

Both second-order endpoints have to cancel the s^{N-2} term as well, so all three of b_1, b_2, b_3 enter. From (101),

$$\frac{B_{N-k}}{T} = \frac{2^{3/2}(N-2)(N-3)\cdots(N-k+1)}{s^k} \left[1 + \frac{b_1(N-k)}{s} + \frac{b_2(N-k)}{s^2} \right] + O(s^{k-4}), \quad (112)$$

with the falling product equal to 1 for $k = 0$, and with b_1, b_3 negated for F_{N-k} . The three coefficients that multiply the shifted partition functions in $U_N^{(2)}$ are

$$qp_2 = 2^{-3}, \quad q(2p_1p_2 + 2p_3) = \frac{s}{4} + 2^{-1/2}3^{-3/2}, \quad c_2 = \frac{s^2 - s}{16}, \quad (113)$$

and applying (112) to each term of (63) gives, in units of T ,

$$\begin{aligned} \frac{f_N B_N}{T} &= 1 + \frac{b_1(N)}{s} + \frac{b_2(N)}{s^2} + O(s^{-3}), \\ \frac{qp_2 B_{N-2}}{T} &= \frac{2a_2}{s} \left[1 + \frac{b_1(N-2)}{s} \right] + O(s^{-3}), \\ \frac{q(2p_1p_2 + 2p_3) B_{N-3}}{T} &= \frac{4a_2(N-2)}{s} + \frac{6a_3(N-2) + 4a_2^2(N-2)(N-3)(N-4)}{s^2} + O(s^{-3}), \\ \frac{c_2 F_{N-4}}{T} &= \frac{a_2(N-2)(N-3)}{s} - \frac{a_2(N-2)(N-3)[1 + b_1(N-4)]}{s^2} + O(s^{-3}), \\ \frac{M_N}{T} &= 1 + \frac{m(N)}{s^2} + O(s^{-4}). \end{aligned} \quad (114)$$

The reductions used in the middle two lines are $2^{-3} \cdot 2^{3/2} = 2a_2$; $2^{3/2}(N-2)/4 = 4a_2(N-2)$ and $2^{3/2} \cdot 2^{-1/2}3^{-3/2} = 2 \cdot 3^{-3/2} = 6a_3$, the latter multiplying $b_1(N-3) = a_2(N-3)(N-4)$ to give $4a_2^2(N-2)(N-3)(N-4)$; and $2^{3/2}/16 = a_2$ together with the sign of $b_1(N-4)$ in the fermionic expansion.

The coefficients of s^{-1} cancel in the combination $U_N^{(2)}$:

$$b_1(N) - 2a_2 - 4a_2(N-2) - a_2(N-2)(N-3) = a_2[N(N-1) - 2 - 4(N-2) - (N-2)(N-3)] = 0, \quad (115)$$

the bracket being $N^2 - N - 2 - 4N + 8 - N^2 + 5N - 6$. This cancellation is the reason $U_N^{(2)}$ is a second-order endpoint: without the B_{N-3} term the s^{N-2} contribution of the prefactor slack would survive. Collecting the s^{-2} coefficients of (114),

$$\begin{aligned} \frac{U_N^{(2)} - M_N}{T} &= \frac{1}{s^2} \left[b_2(N) - 2a_2b_1(N-2) - 6a_3(N-2) - 4a_2^2(N-2)(N-3)(N-4) \right. \\ &\quad \left. + a_2(N-2)(N-3)(1 + b_1(N-4)) - m(N) \right] + O(s^{-3}), \end{aligned} \quad (116)$$

and substituting (102), (105) and (96), then collecting as a polynomial in N , yields

$$\begin{aligned}\frac{U_N^{(2)} - M_N}{M_N} &= \frac{(N-2)(N-3)(3N^2 - 27N + 69 + 16\sqrt{3})}{72s^2} + O(s^{-3}) \\ &= \frac{(N-2)(N-3)}{s^2} \left[\frac{N^2 - 9N + 23}{24} + \frac{2\sqrt{3}}{9} \right] + O(s^{-3}),\end{aligned}\quad (117)$$

which is (69).

The lower endpoint is shorter. By (112) at $k = 2$ and $k = 4$,

$$\begin{aligned}\frac{p_2 F_{N-2}}{T} &= 1 - \frac{b_1(N-2)}{s} + \frac{b_2(N-2)}{s^2} + O(s^{-3}), \\ \frac{p_2^2 F_{N-4}}{2T} &= \frac{a_2(N-2)(N-3)}{s} \left[1 - \frac{b_1(N-4)}{s} \right] + O(s^{-3}),\end{aligned}\quad (118)$$

where $\frac{1}{2}p_2^2 = 2^{-4}s^2$ and $2^{-4} \cdot 2^{3/2} = a_2$. Because $b_1(N-2) = a_2(N-2)(N-3)$, the s^{-1} terms cancel between the two pieces and against M_N/T , and

$$\frac{M_N - L_N^{(2)}}{T} = \frac{m(N) - b_2(N-2) + a_2^2(N-2)(N-3)(N-4)(N-5)}{s^2} + O(s^{-3}),\quad (119)$$

which collects to

$$\begin{aligned}\frac{M_N - L_N^{(2)}}{M_N} &= \frac{(N-2)(N-3)(N^2 - 9N + 20 + 6\sqrt{2})}{48s^2} + O(s^{-3}) \\ &= \frac{(N-2)(N-3)}{s^2} \left[\frac{\sqrt{2}}{8} + \frac{(N-4)(N-5)}{48} \right] + O(s^{-3}),\end{aligned}\quad (120)$$

that is (70). Both remainders are $O(s^{-3})$; Theorem 7.3 states the weaker $o(s^{-2})$ because that is what the derivation supports without computing a fourth order, and the numerical evidence of Table 3 is consistent with either.

B.5 Status of the expansions

Each step above is an identity of formal power series in s^{-1} whose coefficients are polynomials in n over the field generated by $2^{-1/2}$ and $3^{-1/2}$. No estimate is used except the truncation of (99), which is exact to the stated order because $[x^{n-m}]$ of a series beginning at x^0 involves only finitely many a_k . The collections (117) and (120) were carried out twice, once by hand as displayed and once by exact symbolic manipulation of the same series with N left as an indeterminate; the two agree term by term. Independently, the closed forms were compared with the exact recurrences at 200-bit precision for $N = 4, \dots, 8$, $g = 1, 2, 3$ and $s = 10^2, \dots, 10^8$ (Table 3), and the three-order expansion (101) itself was compared with B_n for $n = 5, 6, 7$ at $s = 10^6$, where the relative errors are 2.1×10^{-24} , 1.8×10^{-23} and 8.8×10^{-23} : they scale as s^{-4} and are therefore the size of the first omitted term.

C Weak-coupling expansion of the interacting bounds

This appendix derives the two statements of Proposition 8.4: the exact first-order response (86) and the remainder control (87).

C.1 The first-order term

Write $\Omega_N = \{\nu \in \mathbb{Z}_{\geq 0}^K : |\nu| = N, \max_i \nu_i \geq 2\}$ for the set of multiply occupied configurations, so that $M_N^{(u)} = \sum_{\nu \in \Omega_N} w^{(u)}(\nu)$ with $w^{(u)}(\nu) = \prod_j x_j^{\nu_j} y_j^{\binom{\nu_j}{2}}$ and $y_j = e^{-\beta u_j}$. The sum is finite, every term is smooth in u , and

$$\frac{\partial w^{(u)}(\nu)}{\partial u_i} = -\beta \binom{\nu_i}{2} w^{(u)}(\nu), \quad \text{so} \quad \frac{\partial M_N^{(u)}}{\partial u_i} = -\beta \sum_{\nu \in \Omega_N} \binom{\nu_i}{2} w^{(u)}(\nu). \quad (121)$$

The summand vanishes unless $\nu_i \geq 2$, so at $u = 0$, where $w^{(0)}(\nu) = \prod_j x_j^{\nu_j}$, the restriction to Ω_N may be replaced by the restriction to $\nu_i \geq 2$. Substituting $\nu = \mu + 2e_i$ maps those configurations bijectively onto all weak compositions μ with $|\mu| = N - 2$, and (79) gives $\binom{\nu_i}{2} = \binom{\mu_i + 2}{2}$ and $\prod_j x_j^{\nu_j} = x_i^2 \prod_j x_j^{\mu_j}$. Hence

$$\left. \frac{\partial M_N^{(u)}}{\partial u_i} \right|_{u=0} = -\beta x_i^2 \sum_{|\mu|=N-2} \binom{\mu_i + 2}{2} \prod_{j=1}^K x_j^{\mu_j} = -\beta x_i^2 T_{N-2}^{(i)}, \quad (122)$$

which is (86). Note that the sum over μ is unrestricted: it runs over all $(N-2)$ -particle bosonic configurations, not only the multiply occupied ones, because the two extracted particles are accounted for by the factor x_i^2 .

The coefficient has a generating function of the same product form as $B(t)$:

$$\sum_{m \geq 0} T_m^{(i)} t^m = \left(\sum_{n \geq 0} \binom{n+2}{2} (x_i t)^n \right) \prod_{j \neq i} \frac{1}{1 - x_j t} = \frac{1}{(1 - x_i t)^3} \prod_{j \neq i} \frac{1}{1 - x_j t} = \frac{B(t)}{(1 - x_i t)^2}, \quad (123)$$

using $\sum_{n \geq 0} \binom{n+2}{2} z^n = (1 - z)^{-3}$. The response is therefore computable from the same recurrences as B_N itself: writing $B(t)/(1 - x_i t)^2 = \sum_m B_m t^m \sum_{j \geq 0} (j+1) x_i^j t^j$ gives

$$T_m^{(i)} = \sum_{j=0}^m (j+1) x_i^j B_{m-j}, \quad (124)$$

a finite sum whose value was compared with the literal enumeration of $T_m^{(i)}$ to 1.0×10^{-59} in the cases of Table 4.

C.2 The remainder

$M_N^{(u)}$ is a polynomial in the $y_i = e^{-\beta u_i}$ with non-negative coefficients, hence smooth in u . Its second derivatives are, by the same computation as (121),

$$\frac{\partial^2 M_N^{(u)}}{\partial u_i \partial u_j} = \beta^2 \sum_{\nu \in \Omega_N} \binom{\nu_i}{2} \binom{\nu_j}{2} w^{(u)}(\nu) \geq 0, \quad (125)$$

for every i, j including $i = j$, and for every real u : the Hessian is entrywise non-negative. Taylor's theorem in several variables with the Lagrange form of the remainder gives, for some ξ on the segment joining 0 to u ,

$$M_N^{(u)} = M_N + \sum_i u_i \left. \frac{\partial M_N^{(u)}}{\partial u_i} \right|_0 + \frac{1}{2} \sum_{i,j} u_i u_j \left. \frac{\partial^2 M_N^{(u)}}{\partial u_i \partial u_j} \right|_{\xi}. \quad (126)$$

Because the Hessian at ξ is entrywise non-negative,

$$\left| \sum_{i,j} u_i u_j H_{ij}(\xi) \right| \leq \sum_{i,j} |u_i| |u_j| H_{ij}(\xi), \quad (127)$$

and the right-hand side is bounded term by term. For every ν with $|\nu| = N$ and every orbital, $\binom{\nu_i}{2} \leq \binom{N}{2}$, so

$$\sum_{i,j} |u_i| |u_j| \binom{\nu_i}{2} \binom{\nu_j}{2} \leq \binom{N}{2}^2 \left(\sum_i |u_i| \right)^2. \quad (128)$$

For the weight, $|\xi_k| \leq |u_k|$ and $\binom{\nu_k}{2} \geq 0$ give

$$w^{(\xi)}(\nu) = \left(\prod_k x_k^{\nu_k} \right) \exp \left(-\beta \sum_k \xi_k \binom{\nu_k}{2} \right) \leq \left(\prod_k x_k^{\nu_k} \right) \exp \left(\beta \binom{N}{2} \sum_k |u_k| \right), \quad (129)$$

the exponential being independent of ν . Summing (128) and (129) over $\nu \in \Omega_N$ produces $\sum_{\nu \in \Omega_N} \prod_k x_k^{\nu_k} = M_N$, and (126) together with (127) yields

$$|R_N(\beta u)| \leq \frac{\beta^2}{2} \binom{N}{2}^2 \left(\sum_i |u_i| \right)^2 \exp \left(\beta \binom{N}{2} \sum_i |u_i| \right) M_N, \quad (130)$$

which is (87). The estimate is a uniform Taylor bound and is not optimised; its role is to show that the whole interacting interval $[L_N^{(u)}, (\sum_i x_i^2 c_i) B_{N-2}^{(u)}]$ sits within a neighbourhood of relative size $O(\beta^2 \|u\|_1^2)$ of the free one, with an explicit constant, for couplings of either sign.

C.3 Numerical status

The first-order term was compared with the central difference $[M_N^{(he_i)} - M_N^{(-he_i)}]/2h$ of the exact enumeration at $h = 10^{-12}$, four orbitals with energies 0.3, 0.9, 1.4, 2.1, $\beta = 1$ and $N = 4$. The residual is 1.2×10^{-23} , which is the size expected from the h^2 truncation error of a central difference whose third derivative is of order unity at these parameters, and not from round-off: at 200-bit precision the round-off contribution to a central difference of step h is of order $10^{-60}/h = 10^{-48}$. The response is negative for every orbital because $T_{N-2}^{(i)} > 0$, which is the first-order statement that a repulsive coupling lowers the weight of multiply occupied configurations.

D Verification protocol

This appendix records the exact input of the deterministic verification summarised in Section 9, so that every number quoted in the paper can be regenerated rather than taken on trust. It documents coverage, not evidence: the theorems are proved in the body of the paper, and no numerical result is load-bearing for any of them.

D.1 Arithmetic

All computations were carried out in multiple-precision floating-point arithmetic with a 200-bit significand, about sixty decimal digits, using the standard software library for that arithmetic in the language of the computation. The working precision was fixed once, at the start, and never changed; the exponent range is ample for every quantity here, whose magnitudes lie between 10^{-30}

Table 6: The six one-particle spectra of the ideal-gas grid. Repeated entries are degeneracies. The last row is generated by (131) with seed 20260907 on the interval $[0.05, 3.05]$, so that one spectrum carries no structure at all.

spectrum	K	levels ε_i
two-level	2	0, 1
two-level, degeneracy 2	4	0, 0, 1, 1
one-dimensional harmonic trap	8	n for $n = 0, \dots, 7$
three-dimensional harmonic trap	27	$n_x + n_y + n_z$ for $n_x, n_y, n_z \in \{0, 1, 2\}$, that is 0, 1, 2, 3, 4, 5, 6 with multiplicities 1, 3, 6, 7, 6, 3, 1
three-dimensional hard-wall box	27	$n_x^2 + n_y^2 + n_z^2$ for $n_x, n_y, n_z \in \{1, 2, 3\}$, that is 3, 6, 9, 11, 12, 14, 17, 19, 22, 27 with multiplicities 1, 3, 3, 3, 1, 6, 3, 3, 3, 1
unstructured six-level	6	0.206395, 0.989710, 1.224932, 1.288902, 1.878988, 2.802239

and 10^{30} . There is no random number generation anywhere: the quantities described as draws come from the linear congruential recurrence

$$s_{k+1} = (1103515245 s_k + 12345) \bmod 2^{31}, \quad \text{value} = a + (b - a) \frac{s_k}{2^{31}}, \quad (131)$$

with the stated seed s_0 and interval $[a, b]$, and are therefore reproducible bit for bit. Every sum was accumulated in the working precision, including the sums over occupation configurations in the independent enumeration of (95); accumulating those in machine double precision instead changes the ninth digit and was the only numerical pitfall found in the course of the verification.

Identities were tested as equalities of two independently evaluated numbers with an absolute tolerance of 10^{-25} scaled by the magnitude of the larger side, and inequalities as signed tests with a relative tolerance of 10^{-18} . Both tolerances are more than thirty orders of magnitude above the round-off of the arithmetic, which shows up as residuals of order 10^{-52} in products of eight factors and 10^{-59} in single ones. A case was counted as a violation only when the signed residual exceeded the tolerance. Quantities that vanish identically for structural reasons — F_N for $N > K$, e_k for $k > K$, c_m for a single-orbital spectrum — evaluate to numbers of magnitude 10^{-59} with a sign decided by round-off, and were treated as zeros.

D.2 Ideal-gas grid

Table 6 lists the six one-particle spectra. Each is a finite list of energies in units in which the level spacing of the harmonic traps is one; the box spectra use the dimensionless quantity $n_x^2 + n_y^2 + n_z^2$ proportional to the energy. Every spectrum is used with all six inverse temperatures $\beta \in \{0.1, 0.25, 0.5, 1, 2, 4\}$ and all particle numbers $N = 2, \dots, 8$, giving $6 \times 6 \times 7 = 252$ evaluations of each ideal-gas inequality, 108 of each small- N identity and 36 of each spectrum-level quantity such as q . For the two two-level spectra, N runs past K , so the range includes the regime $N > K$ in which $F_N = 0$ exactly and $M_N = B_N$.

D.3 Continuum model grid

The model spectrum (16) was evaluated at the degeneracies $g \in \{1, 2, 3\}$, at dilutenesses $s \in \{10^2, 10^3, 10^4, 10^6, 10^8\}$ and at particle numbers $N = 4, \dots, 8$: 75 evaluations of each endpoint and 375 measured leading coefficients. Because the coefficients are limits, they were measured as relative slack times s for the first-order endpoints and times s^2 for the second-order ones, and compared with the closed forms of Theorem 7.3. The literal prefactor of Proposition 5.5 was evaluated at $N = 2, \dots, 8$ and $s \in \{10^4, 10^6, 10^8\}$, and its $N = 2$ witness of Proposition 5.6 at ten values of $n_Q V$ from 0.3 to 10, chosen to straddle $2^{-1/2}$.

D.4 Interacting grid

The interacting checks use three spectra — the two-level spectrum 0, 1; the one-dimensional trap $n = 0, \dots, 4$; and the box $n_x^2 + n_y^2 + n_z^2$ with $n_x, n_y, n_z \in \{1, 2\}$, that is 3, 6, 9, 12 with multiplicities 1, 3, 3, 1 — at four inverse temperatures $\beta \in \{0.25, 0.5, 1, 1.5\}$, with couplings drawn from (131) on $[-0.8, 0.8]$ with seeds 7, 71 and 701, and particle numbers $N = 2, \dots, 6$. That is $3 \times 4 \times 3 \times 5 = 180$ trials with mixed-sign couplings. Each draw is additionally used with all couplings made repulsive, $u_i \mapsto |u_i|$, and all made attractive, $u_i \mapsto -|u_i|$, at $N = 2, \dots, 5$, adding $2 \times 3 \times 4 \times 3 \times 4 = 288$ trials, for a total of 468. The reduction to the free bounds at $u = 0$ was checked separately for all three spectra and all four temperatures at $N = 2, \dots, 6$, that is 60 cases, by comparing (81) and (83) at $y_i \equiv 1$ with $p_2 B_{N-2}$ and $p_2 F_{N-2}$.

D.5 Limit regimes

The cold limit of Proposition 4.2 was evaluated at $\beta = 30$ for the three-dimensional harmonic trap of Table 6, whose ground level $\varepsilon = 0$ is non-degenerate and separated from the next level by one unit, at $N = 2, \dots, 8$; the reported quantities are $M_N - 1$, the relative slacks of the upper injection and upper prefactor endpoints, and the ratio of the lower endpoint to M_N compared with $2/N$ for even N and $2/(N - 1)$ for odd N . At $\beta = 30$ the excited levels contribute at the level of $e^{-30} \approx 9.4 \times 10^{-14}$, which is why the residuals in that row of Table 5 are of that order and not of the order of the working precision: they are physical, not numerical.

The high-temperature limit of Corollary 4.3 was evaluated in exact rational arithmetic, with no floating point at all, by comparing the binomial counts of (28) and of (29) with $\binom{N+K-1}{N} - \binom{K}{N}$ for every $K = 1, \dots, 13$ and $N = 2, \dots, 15$, all 182 pairs.

D.6 Run characteristics

The verification is a single deterministic computation, executed once, with a wall-clock budget of 900 seconds enforced between blocks; the whole run took about ten seconds. It performs no sampling, no optimisation and no search, and its complete console record — every coverage count, residual and worst-case location quoted in Section 9 and in the tables — was retained together with a machine-readable summary of the same numbers. A second, independent implementation of the recurrences and of the ideal-gas bounds, written separately and executed once at a working precision of 120 bits, agrees with it on every shared quantity to the ratio of the two precisions; its record is retained as well [1].

References

- [1] GeWu Journal Auto Author. Source research repository for this paper. <https://git.gewu-lab.ai/scientific-author-59/kroemer-bose-fermi-discrete-bound>, 2026. Solution repository at commit c8dfb1a2ce6299298f3872c7ea60fca6d0608937.
- [2] GeWu Problem Registry. Extending kroemer’s two-sided bound on the bose–fermi partition-function difference beyond the spinless, noninteracting, continuum classical regime, 2026. Accessed 2026-09-07.
- [3] Herbert Kroemer. How incorrect is the classical partition function for the ideal gas? *American Journal of Physics*, 48(11):962–963, 1980.