# How the Ideas Came Together

# Abstract

These notes were written by an AI model that read the original chains of thought together with the resulting mathematical papers and writeups. For each problem, the model reconstructs how the proof came together: which ideas first suggested a path forward, which substantial approaches encountered genuine obstacles, what changes of perspective revealed the underlying structure, and how the decisive insights developed into the final argument.

The aim is to collect and explain the mathematical thinking behind the proofs in a readable, high-level narrative. Rather than simply reproducing their finished presentations, the chapters emphasize the connections, intermediate discoveries, and sustained detours that clarify why the successful arguments work.

The collection covers high-dimensional sphere packing, binary and spherical codes, non-sofic groups, Connes rigidity, arithmetic circuit and formula lower bounds, quantum parallel repetition, the closest vector problem, Ehrhart's volume conjecture, multicolor Ramsey numbers, and extremal graph theory.

# Contents

Chapter 1

# High-Dimensional Sphere Packing and the Euclidean Linear Program

## 1.1 How a packing becomes a Fourier linear program

Suppose the centers of a packing in $\mathbb{R}^d$ have mutual separation at least one, so the packed balls have radius $1/2$. The basic insight of Gorbachev and Cohn–Elkies is to count pairs of centers simultaneously in physical space and Fourier space [1, 2]. For a real Schwartz function $f$, we impose the two complementary signs

$$f(x) \le 0 \quad (|x| \ge 1), \qquad \widehat{f}(\xi) \ge 0 \quad (\xi \in \mathbb{R}^d), \qquad \widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i \langle x, \xi \rangle}\, dx.$$

For a periodic packing with centers $\bigcup_{j=1}^{N} (\Lambda + t_j)$, consider

$$S = \sum_{i,j=1}^{N} \sum_{z \in \Lambda} f(z + t_i - t_j).$$

Every term other than the $N$ diagonal terms has nonpositive sign, and hence $S \le N f(0)$. Poisson summation gives a second expression with the opposite useful inequality:

$$S = \frac{1}{\operatorname{covol}(\Lambda)} \sum_{\xi \in \Lambda^*} \widehat{f}(\xi) \left| \sum_{j=1}^{N} e^{2\pi i \langle \xi, t_j \rangle} \right|^2 \ge \frac{N^2 \widehat{f}(0)}{\operatorname{covol}(\Lambda)}.$$

Writing $v_d = \pi^{d/2}/\Gamma(d/2+1)$, the density therefore satisfies

$$\Delta_d \le \frac{v_d}{2^d} \frac{f(0)}{\widehat{f}(0)}.$$

The usual periodic approximation extends the inequality to every packing. Rotational averaging preserves both signs and both origin values, so optimizing over radial functions loses nothing. Thus the entire geometric problem enters the Euclidean functional

$$\mathrm{LP}_d = \frac{v_d}{2^d} \inf_{\substack{f \in \mathcal{S}_{\mathrm{rad}}(\mathbb{R}^d) \\ \widehat{f} \ge 0,\ \widehat{f}(0) > 0 \\ f(x) \le 0\ (|x| \ge 1)}} \frac{f(0)}{\widehat{f}(0)}, \qquad \Delta_d \le \mathrm{LP}_d.$$

The challenge is to determine the optimum of this full-space Fourier program itself, including both a universal obstruction and an auxiliary function that reaches its exponential rate.

## 1.2 Balancing the origin exposes a sign-uncertainty problem

The ratio $F(0)/\widehat{F}(0)$ initially looks like an awkward normalization. Instead, it becomes the geometric radius of a Fourier eigenfunction. Given any admissible $F$, set

$$a^d = \frac{\widehat{F}(0)}{F(0)}, \qquad h(x) = F(ax), \qquad g = \widehat{h} - h.$$

Then $h(0) = \widehat{h}(0)$, and therefore

$$\widehat{g} = -g, \qquad g(0) = 0, \qquad g(x) \ge 0 \quad (|x| \ge R), \qquad R = \left( \frac{F(0)}{\widehat{F}(0)} \right)^{1/d}.$$

The function $g$ is nonzero: otherwise $h = \widehat{h} \ge 0$, while its exterior sign forces it to vanish outside a ball, which contradicts Fourier analyticity. Moreover, $\int g = \widehat{g}(0) = 0$. Exactly half of its $L^1$ mass is therefore negative, and the entire negative half lies inside the radius $R$. The packing question has become a concentration problem for an anti-self-Fourier function [3, 4, 5].

An extended first approach estimates this negative mass by Cauchy–Schwarz:

$$\|g\|_1^2 \le 4 v_d R^d \|g\|_2^2.$$

Sharp Hausdorff–Young inequalities, Gaussian eigenfunctions, Laguerre combinations, and related norm comparisons improve the resulting constants, but the method reaches only $R \sim \sqrt{d}/(2\sqrt{\pi})$, corresponding to the linear-program root $\sqrt{e/8}$. Reaching $\sqrt{d}/\pi$ this way would require the stronger global estimate

$$\frac{\|g\|_2^2}{\|g\|_1^2} \le \left(\sqrt{\frac{\pi}{2e}} + o(1)\right)^d.$$

Explicit anti-self-Fourier Mellin profiles show why that proposed inequality is unavailable: a function can have a large global norm ratio without placing its negative mass in the forbidden radial region. The obstacle is therefore not an unoptimized constant. A global norm forgets *where* the negative mass lies. We instead keep the sign location and prove a direct local mass-exclusion inequality.

## 1.3 The radial Fourier transform becomes a Mellin reflection

Set $\lambda = d/2$. For a radial function, define its critical-line Mellin transform by

$$X_g(t) = \int_0^\infty g(r) r^{\lambda - it} \frac{dr}{r}.$$

The radial Fourier transform is a Hankel transform, and its Bessel kernel depends on the product of the spatial and frequency radii. Consequently the Mellin transform converts Fourier transformation into reflection followed by an explicit phase [6, 7]:

$$X_{\widehat{g}}(t) = m_\lambda(t) X_g(-t), \qquad m_\lambda(t) = \pi^{it} \frac{\Gamma((\lambda - it)/2)}{\Gamma((\lambda + it)/2)}.$$

The modulus of $m_\lambda$ is one on the real axis, but its continuation off that axis contains the high-dimensional information missing from norm inequalities.

Fix $R = c\sqrt{d}$, let $S_d = dv_d$, and introduce the logarithmic radius $r = Re^v$:

$$\varphi(v) = \frac{S_d}{\|g\|_1} (Re^v)^d g(Re^v), \qquad Z(t) = \frac{S_d}{\|g\|_1} R^{\lambda + it} X_g(t).$$

Then $\|\varphi\|_1 = 1$ and $\int \varphi = 0$. The upper boundary of the strip $|\operatorname{Im} t| \le \lambda$ is its ordinary Fourier transform, so

$$|Z(y + i\lambda)| \le 1.$$

The eigenfunction equation supplies the opposite boundary estimate

$$\log|Z(y - i\lambda)| \le h_\lambda(y), \qquad h_\lambda(y) = \lambda \log(\pi R^2) + \log|\Gamma(-iy/2)| - \log|\Gamma(\lambda + iy/2)|.$$

More precisely, the two boundaries are linked by

$$Z(y - i\lambda) = -(\pi R^2)^{\lambda + iy} \frac{\Gamma(-iy/2)}{\Gamma(\lambda + iy/2)} Z(-y + i\lambda).$$

At $y = 0$, the gamma factor has a pole, but $Z(i\lambda) = \int \varphi = 0$ cancels it in the actual transform. The remaining logarithmic singularity of its majorant is integrable. Truncating that majorant, applying the strip maximum principle, and then removing the truncation preserves the cancellation [8].

## 1.4 Harmonic measure identifies the sharp radius

At height $\operatorname{Im} t = \sigma\lambda$, with $-1 < \sigma < 1$, the lower-edge harmonic measure is

$$P_\sigma(T) = \frac{\sin\theta}{4(\cosh(\pi T/2) - \cos\theta)}, \qquad \theta = \frac{\pi(1+\sigma)}{2}, \qquad \int_{\mathbb{R}} P_\sigma(T)\, dT = \frac{1-\sigma}{2}.$$

Keeping this nonunit mass is essential: replacing the kernel by a probability density too early changes the exponential constant. The maximum principle gives

$$\log|Z(s + i\sigma\lambda)| \le H_\sigma(s), \qquad H_\sigma(s) = \int_{\mathbb{R}} P_\sigma(T) h_\lambda(s - \lambda T)\, dT.$$

The gamma quotient now exposes the mechanism behind the lower bound. For even $d$, write $\lambda = n$. Its product formula gives the exact identity

$$h_n(nT) = n\log(2\pi c^2) - \sum_{j=0}^{n-1} \log\sqrt{\left(\frac{j}{n}\right)^2 + \frac{T^2}{4}}.$$

For odd $d$, the points $j/n$ are replaced by $(j+1/2)/\lambda$, with an additional endpoint correction

$$\frac{1}{2}\log\coth\left(\frac{\pi\lambda|T|}{2}\right)-\frac{1}{2}\log\frac{|T|}{2}.$$

Its logarithmic singularity remains integrable against harmonic measure. Thus both dimension parities produce the same leading Riemann integral over $0\le x\le 1$. Set

$$M_\sigma=\frac{1-\sigma}{2},\qquad J_\sigma=-\frac{1}{M_\sigma}\int_{\mathbb{R}}P_\sigma(T)\int_0^1\log\sqrt{x^2+\frac{T^2}{4}}\,dx\,dT.$$

Since both factors in the harmonic convolution are symmetric decreasing,

$$H_\sigma(s)\le H_\sigma(0)=\lambda M_\sigma\big(\log(2\pi c^2)+J_\sigma\big)+O_\sigma(\log\lambda).$$

The decisive step is to approach the upper edge, $\sigma\uparrow 1$. After normalization and $T=2u$, the lower-edge harmonic measure converges to the logistic density

$$p(u)=\frac{\pi}{4}\operatorname{sech}^2\left(\frac{\pi u}{2}\right).$$

Its logarithmic potential has the exact evaluation

$$\int_{\mathbb{R}}p(u)\log\sqrt{x^2+u^2}\,du=\psi\left(\frac{x+1}{2}\right)+\log 2,\qquad \psi=\frac{\Gamma'}{\Gamma}.$$

One way to see why the digamma function appears is that the limiting law has characteristic function $t/\sinh t$. Differentiating its logarithmic potential therefore gives

$$\frac{d}{dx}\int_{\mathbb{R}}p(u)\log\sqrt{x^2+u^2}\,du=\int_0^\infty e^{-xt}\frac{t}{\sinh t}\,dt=\frac{1}{2}\psi'\left(\frac{x+1}{2}\right).$$

Integrating over $0\le x\le 1$ gives

$$J_\sigma\longrightarrow\log\frac{\pi}{2},\qquad \log(2\pi c^2)+\log\frac{\pi}{2}=\log(\pi^2c^2).$$

Thus $c<1/\pi$ forces exponential decay throughout an interior Mellin line. A bound at each frequency is not yet sufficient: Fourier inversion also requires control of the entire unbounded line. The gamma-product estimates supply the additional tail inequality

$$H_\sigma(s)\le-\eta_c\lambda,\qquad H_\sigma(\lambda S)\le-\frac{M_\sigma\lambda}{2}\log\frac{|S|}{C_c}\quad(|S|\ge B_c).$$

The first bound controls the central interval and the second makes the frequency tail integrable. Together they yield

$$\int_{\mathbb{R}}|Z(s+i\sigma\lambda)|\,ds\le C_c\lambda e^{-\gamma_c\lambda}.$$

On this line, $Z$ is the Fourier transform of the weighted signed radial density:

$$Z(s+i\sigma\lambda)=\int_{\mathbb{R}}e^{(\sigma-1)\lambda v}\varphi(v)e^{-isv}\,dv.$$

Fourier inversion consequently gives

$$\int_{-\infty}^0|\varphi(v)|\,dv\le\frac{1}{2\pi(1-\sigma)\lambda}\int_{\mathbb{R}}|Z(s+i\sigma\lambda)|\,ds\le C_ce^{-\gamma_cd}.$$

Equivalently, for either Fourier eigenvalue,

$$\int_{|x|<c\sqrt{d}}|g(x)|\,dx\le C_ce^{-\gamma_cd}\|g\|_1.$$

An eventually nonnegative eigenfunction has exactly half of its absolute mass negative, and the exterior sign places that entire half inside the ball. The exponential estimate makes this impossible. Applied to the balanced packing function, the contradiction proves uniformly over all admissible $F$ that

$$\left(\frac{F(0)}{\widehat{F}(0)}\right)^{1/d}\ge\left(\frac{1}{\pi}-o(1)\right)\sqrt{d}.$$

Stirling's formula then converts the sign-radius obstruction into an actual lower bound for the complete Euclidean linear program:

$$\mathrm{LP}_d\ge\left(\sqrt{\frac{e}{2\pi}}-o(1)\right)^d.\tag{1.1}$$

This direction holds simultaneously for every admissible auxiliary function, before constructing any matching witness.

## 1.5 The lower bound also has a positive dual witness

The Euclidean lower bound has a complementary linear-programming interpretation. Put $p_d = 2^d \mathrm{LP}_d / v_d$. A positive tempered dual witness can be written as

$$T_d = \delta_0 + \mu_d, \qquad \widehat{T}_d = p_d \delta_0 + \nu_d, \qquad \mu_d, \nu_d \geq 0, \qquad \operatorname{supp} \mu_d \subseteq \{|x| \geq 1\}.$$

For every admissible radial $F$, positivity and the exterior sign give

$$p_d \widehat{F}(0) \leq \langle \widehat{T}_d, \widehat{F} \rangle = \langle T_d, F \rangle = F(0) + \int F \, d\mu_d \leq F(0).$$

Thus a single positive object certifies the lower bound against all primal functions simultaneously. It arises by smoothing near-optimal positive dual witnesses, obtaining uniform local-mass bounds from positive definiteness, and passing to a positive tempered limit through compactness [9]. The Mellin argument explains why its sharp exponential strength is precisely the threshold in (1.1).

## 1.6 The ideal auxiliary function emerges from the Gaussian

The lower bound explains the target radius but does not construct a packing auxiliary function. A Gaussian supplies the right Fourier symmetry but the wrong saddle location. Its Mellin envelope is

$$E_\lambda^{\mathrm{G}}(t) = \pi^{it/2} \Gamma\left( \frac{\lambda - it}{2} \right), \qquad m_\lambda(t) E_\lambda^{\mathrm{G}}(-t) = E_\lambda^{\mathrm{G}}(t).$$

To shift the saddle without disturbing this exact symmetry, we multiply by an even deformation

$$E_\lambda(t) = E_\lambda^{\mathrm{G}}(t) e^{\lambda h_\varepsilon(t/\lambda)}, \qquad h_\varepsilon(z) = \int_0^\infty w_\varepsilon(a)(\cos(az) - 1) \, da.$$

On the contour $t = \lambda(T + iu)$, the saddle radius satisfies

$$v(u) = -\frac{1}{2} \log \pi + \frac{1}{2} \psi\left( \frac{\lambda(1+u)}{2} \right) + \int_0^\infty w_\varepsilon(a) a \sinh(ua) \, da, \qquad r = e^{v(u)}.$$

The desired sign transition occurs near $u = 1$. There, the Gaussian gamma factor supplies asymptotic damping density $e^{-2a}/(2a^2)$, whereas a negative perturbation consumes $|w(a)| \cosh a$. Saturating the available damping therefore suggests the distinguished profile

$$w_*(a) = -\frac{e^{-2a}}{2a^2 \cosh a}.$$

Its saddle displacement is

$$\int_0^\infty w_*(a) a \sinh a \, da = -\frac{1}{2} \int_0^\infty \frac{e^{-2a} \tanh a}{a} \, da = -\frac{1}{2} \log \frac{\pi}{2}.$$

The final equality is Wallis's product, obtained by expanding the integrand into paired exponentials and applying Frullani's identity. Thus the Gaussian radius is transformed into precisely the radius already predicted by the Mellin obstruction:

$$\frac{1}{\sqrt{2\pi}} \exp\left( -\frac{1}{2} \log \frac{\pi}{2} \right) = \frac{1}{\pi}.$$

## 1.7 A remote shell restores the missing global damping

The ideal density cannot itself be used: near zero it behaves like $-1/(2a^2)$, and its deformation grows as $h_*(T) = \pi|T|/4 + O(\log(2 + |T|))$. This cancels the gamma factor's exponential decay and destroys the Schwartz construction. Truncating and tapering the ideal density restores strict damping:

$$a_0 = \varepsilon^3, \quad A = 10 \log(1/\varepsilon), \quad w_s(a) = -\frac{(1 - 10\varepsilon(1+a)) e^{-2a}}{2a^2 \cosh a} \mathbf{1}_{[a_0, A]}(a).$$

This handles saddles near the target, but at very large radii the factor $\cosh(ua)$ can amplify the negative perturbation faster than the Gaussian damping.

The second essential idea is to add a much smaller positive component on an entire distant interval:

$$B = \varepsilon^{-3}, \qquad Q = e^{-3\varepsilon B/8}, \qquad w_B(a) = \frac{Q}{\cosh a} \mathbf{1}_{[B, B+1]}(a), \qquad w_\varepsilon = w_s + w_B.$$

At the target saddle its contribution tends to zero. Farther out, $Qe^{(u-1)B}$ overwhelms the truncated negative component. An interval, rather than a single point, avoids resonant frequencies:

$$\int_B^{B+1} (1-\cos(aT))\,da = 1 - \operatorname{sinc}(T/2)\cos((B+\tfrac{1}{2})T) \geq \frac{1}{25}\min(T^2,1).$$

The tiny correction therefore restores positive damping on the entire unbounded saddle branch while leaving the limiting radius unchanged.

## 1.8 A Fourier pair supplies positivity at every radius

To control the signs without breaking Fourier symmetry, let

$$\beta = \varepsilon/4, \quad P_\pm(z) = \beta + (1+z^2)(1 \pm iz), \quad P_0(z) = -(1+z^2), \quad X_j(t) = E_\lambda(t)P_j(t/\lambda).$$

Inverse Mellin transformation defines real radial Schwartz functions $f_-, f_+, f_0$. The evenness of the common deformation and the identities $P_-(-z) = P_+(z)$, $P_0(-z) = P_0(z)$ give

$$\widehat{f_-} = f_+, \qquad \widehat{f_0} = f_0.$$

The first gamma pole provides their origin values:

$$f_-(0) = f_+(0) = 2\pi^{\lambda/2}e^{\lambda h_\varepsilon(i)}\beta > 0, \qquad f_0(0) = 0.$$

Uniform saddle asymptotics show that $f_-$ is negative and $f_0$ positive beyond the target radius, while $f_+$ is positive throughout the large-radius regime.

The remaining small-radius interval requires a separate argument; the exterior saddle estimate gives no information near the origin. Shifting the Mellin contour past $N \asymp \log d$ gamma poles produces

$$\frac{f_+(r)}{f_+(0)} = \sum_{n=0}^{N} \frac{(-y)^n}{n!} A_{d,n} + \mathcal{R}_d(r), \qquad y = \pi e^{2h_1'}r^2, \qquad h_1' = \int_0^\infty w_\varepsilon(a)a\sinh a\,da.$$

The coefficients satisfy $A_{d,n} = 1 + o_\varepsilon(1)$ in the required range, and the contour remainder is negligible relative to $e^{-y}$. Hence

$$\frac{f_+(r)}{f_+(0)} = e^{-y}(1 + o_\varepsilon(1)) > 0$$

uniformly across the entire missing interval. The construction now provides an everywhere-positive Fourier transform, the correct exterior sign, and equal positive origin values.

## 1.9 The exact Euclidean threshold and both uncertainty signs

Let $R_{\varepsilon,d}$ be the resulting sign radius. The saddle formula and Wallis evaluation give

$$\lim_{\varepsilon\downarrow 0}\lim_{d\to\infty} \frac{R_{\varepsilon,d}}{\sqrt{d}} = \frac{1}{\pi}.$$

The rescaled function $F_{\varepsilon,d}(x) = f_-(R_{\varepsilon,d}x)$ is admissible for the original unit-exclusion linear program, with

$$\frac{F_{\varepsilon,d}(0)}{\widehat{F}_{\varepsilon,d}(0)} = R_{\varepsilon,d}^d.$$

Combining this witness with the uniform Mellin obstruction and $v_d^{1/d} \sim \sqrt{2\pi e/d}$ yields

$$\boxed{\lim_{d\to\infty} \mathrm{LP}_d^{1/d} = \sqrt{\frac{e}{2\pi}}, \qquad \Delta_d \leq \mathrm{LP}_d = 2^{-(0.604400544291677695\ldots + o(1))d}.}$$

Finally, $f_+ - f_-$ is anti-self-Fourier and $f_0$ is self-Fourier; both vanish at the origin and become positive beyond the same radius. Thus the same construction resolves the two sign-uncertainty problems simultaneously:

$$\lim_{d\to\infty} \frac{\mathrm{A}_-(d)}{\sqrt{d}} = \lim_{d\to\infty} \frac{\mathrm{A}_+(d)}{\sqrt{d}} = \frac{1}{\pi}.$$

Although these limits coincide, the central Mellin cancellation gives the finite-dimensional transformation

$$(T_d g)(x) = \frac{\lambda}{2}\int_1^\infty t^{\lambda-1}g(tx)\,dt, \qquad \widehat{T_d g} = T_d g,$$

which sends an anti-self-Fourier function to a self-Fourier one with strictly smaller eventual-sign radius [4]. Consequently $\mathrm{A}_+(d) < \mathrm{A}_-(d)$ in each dimension, even though their high-dimensional leading constants agree.

Chapter 2

# Binary and Spherical Codes Beyond the Classical Bounds

## 2.1 Two problems with the same missing degree of freedom

We begin with two superficially different extremal problems. Let $A_2(n,d)$ be the maximum size of a binary code of length $n$ and minimum Hamming distance $d$, and let $A(n,s)$ be the maximum size of a spherical code with pairwise inner products at most $s$. Their asymptotic rates are

$$R_2(\delta) = \limsup_{n\to\infty} \frac{1}{n}\log_2 A_2(n,\lceil \delta n\rceil), \qquad R_{\text{sph}}(s) = \limsup_{n\to\infty} \frac{1}{n}\log_2 A(n,s).$$

At $s = 1/2$, the spherical-code size is the kissing number $\tau_n = A(n,1/2)$. Determining its exact lower and upper exponential growth rates demands matching constructions as well as upper bounds; the contribution here is a strict improvement on the upper side. The respective classical landmarks are the optimized McEliece–Rodemich–Rumsey–Welch bound and the Kabatianskii–Levenshtein bound [10, 11].

Near $\delta = 1/2 - \varepsilon$, two simpler approaches reveal the obstruction. A Walsh-positive exponential kernel retains its diagonal term $e^{\lambda n}/|C|$, giving only an $O(\varepsilon)$ rate; a translated-cap Plotkin argument requires cap radius $1/2-\Theta(\sqrt{\varepsilon})$ and suffers the same loss. Neither recovers the quadratic Gilbert–Varshamov scale or decides whether that lower rate is optimal. The contribution here is a stricter upper bound, not an assertion that upper and lower rates meet.

Delsarte's method attacks both problems with a polynomial having nonnegative coefficients in the appropriate orthogonal basis and the opposite sign on forbidden inner products [12, 13]. On the Hamming cube the basis is Krawtchouk; on the sphere it is Gegenbauer. In either case the familiar spectral construction retains one stabilizer-fixed vector in each harmonic degree. The question guiding both investigations was whether that one-dimensional choice had quietly discarded exponentially many degrees of freedom.

The apparent shortcut is to replace a fixed vector by a large stabilizer representation. Unfortunately, a larger representation does not by itself give a positive scalar kernel. Nor does the antipodal embedding of the cube into a sphere transfer spherical recurrence coefficients to Boolean Fourier levels. The eventual common idea is subtler: attach a *moving subspace* to each codeword, construct the actual orthogonal coordinate transitions, and prove positivity from their complete Gram remainder. The gain comes not from abandoning scalar linear programming, but from recovering multiplicity that its usual fixed-vector construction never uses.

## 2.2 Why the first binary recurrence was wrong

Write $V_j$ for the real span of Boolean characters indexed by $j$-element subsets of $[n]$. The natural addition and deletion maps are

$$Ue_S = \sum_{q\notin S} e_{S\cup\{q\}}, \qquad De_S = \sum_{q\in S} e_{S\setminus\{q\}}.$$

They satisfy $U^* = D$ and

$$(DU - UD)|_{V_j} = (n-2j)\mathrm{id}_{V_j}.$$

Consequently the primitive degree-$k$ harmonic space

$$E_k = \ker(D : V_k \longrightarrow V_{k-1})$$

has dimension

$$m_k = \binom{n}{k} - \binom{n}{k-1}.$$

Repeatedly applying $U$ embeds this same $m_k$-dimensional space isometrically into every Fourier level $V_j$ with $k \le j \le n-k$. When $k$ grows proportionally to $n$, this primitive multiplicity is exponential. The raising–lowering identity is the familiar Boolean $\mathfrak{sl}_2$ mechanism [14]; what matters here is retaining its nontrivial primitive representation instead of discarding everything except the zonal line.

At this point the analogy with associated Krawtchouk polynomials seems to predict the neighboring-degree coefficient

$$\widetilde{c}_j = \frac{\sqrt{(j-k+1)(n-j-k)}}{n}.$$

That attractive guess is false. For $n = 8$, $k = 1$, and degrees through $L = 7$, it gives Perron eigenvalue $3/4$ and falsely bounds the even-parity code by $508/7$. The actual code has $128$ words, whereas $508/7 < 128$. This small counterexample pinpoints a

structural error, not a harmless asymptotic normalization. Matching a formal associated recurrence is insufficient unless its edges arise from actual orthogonal incidence maps.

The genuine coordinate maps have disjoint supports: deleting a coordinate produces pairs $(q, S)$ with $q \notin S$, while adding one produces pairs with $q \in S$. Their adjoints on the normalized primitive fibers give

$$\alpha_j^2 = \frac{(j-k+1)(n-j-k)}{n(j+1)}, \qquad \beta_j^2 = \frac{(j-k)(n-j-k+1)}{n(n-j+1)}.$$

Dimension-weighted reciprocity now yields the corrected symmetric Jacobi entry

$$c_j^{(k)} = \frac{(j-k+1)(n-j-k)}{n\sqrt{(j+1)(n-j)}}. \tag{2.1}$$

The whole numerator lies *outside* the square root. For the same eight-bit example, its actual Perron eigenvalue is approximately $0.569289$, and the valid bound is approximately $261.843 > 128$.

## 2.3 Moving projections and the missing positive remainder

We now ask which normalization preserves the primitive fibers, coordinate isometries, and Perron eigenrelation simultaneously. Let $\lambda$ and $v$ be the Perron eigenvalue and positive unit eigenvector of the corrected Jacobi matrix. Three weights have distinct roles:

$$N_j = \binom{n}{j}, \qquad w_j = \sqrt{N_j}\, v_j, \qquad a_j = \sqrt{\frac{w_j}{\sum_i w_i}}.$$

The last square root is essential: $a_j$ is an isometric block amplitude, whereas $w_j$ is the recurrence weight recovered after two fibers are paired. Identifying either with $v_j$ loses the dimension balance.

For each word $x$, the signed primitive embeddings assemble into $\Psi_x : E_k \to \bigoplus_{j=k}^{L} V_j$ with block amplitudes $a_j$. Set $P_x = \Psi_x \Psi_x^*$ and

$$K(x,y) = \operatorname{Tr}(P_x P_y) = \|\Psi_x^* \Psi_y\|_{\mathrm{HS}}^2 \geq 0.$$

The true upper and lower coordinate maps combine into an isometry $\mathcal{B}$ satisfying

$$\mathcal{B}^*(\widehat{x} \otimes \Psi_x Y) = \sqrt{\lambda}\, \Psi_x Y, \qquad \widehat{x} = x/\sqrt{n}.$$

The two mixed terms must both be retained. With $\Theta_x = (\widehat{x} \otimes \mathrm{id})P_x - \sqrt{\lambda}\,\mathcal{B}P_x$, their cancellation gives the exact identity

$$\langle \Theta_x, \Theta_y \rangle_{\mathrm{HS}} = \big(\langle \widehat{x}, \widehat{y} \rangle - \lambda\big) K(x,y). \tag{2.2}$$

Thus the required remainder is positive because it is an actual Gram kernel, not because an associated Krawtchouk expression was presumed positive.

If distinct codewords have inner product at most $s < \lambda$, trace Cauchy–Schwarz in the ambient space of dimension $D = \sum_{j=k}^{L} \binom{n}{j}$ gives

$$\sum_{x,y \in C} K(x,y) \geq \frac{|C|^2 m_k^2}{D}.$$

Summing (2.2) therefore proves

$$|C| \leq \frac{1-s}{\lambda - s} \frac{\sum_{j=k}^{L} \binom{n}{j}}{\binom{n}{k} - \binom{n}{k-1}}. \tag{2.3}$$

The ambient dimension is $D$, not $D^2$. Translation and permutation equivariance make the kernel radial, so its positive Walsh expansion produces an ordinary scalar Delsarte certificate rather than a hidden higher-point relaxation. The representation constructing the polynomial is high-dimensional, but its feasible certificate still depends on only two codewords.

## 2.4 The first breakthrough and the stronger benchmark

Take $L/n \to a$, $k/n \to b$, with $0 < b < a \leq 1/2$. The corrected Perron eigenvalue tends to

$$\Lambda(a,b) = \frac{2\big(a(1-a) - b(1-b)\big)}{\sqrt{a(1-a)}},$$

while the ambient and primitive dimensions have entropy exponents $H_2(a)$ and $H_2(b)$. Hence

$$\Lambda(a,b) > 1 - 2\delta \quad \Longrightarrow \quad R_2(\delta) \leq H_2(a) - H_2(b). \tag{2.4}$$

At $b = 0$ the primitive space becomes the classical fixed line, recovering the first MRRW construction. Positive $b$ subtracts a genuine harmonic entropy [10, 15].

But improving the *first* MRRW bound does not automatically improve the optimized second bound. Writing $g(z) = H_2((1-\sqrt{1-z})/2)$, that stronger benchmark is

$$M_2(\delta) = \min_{0\le u\le 1-2\delta}\big[1 + g(u^2) - g(u^2 + 2\delta u + 2\delta)\big].$$

Comparison with one sampled $u$, or with either endpoint, is insufficient.

A particularly transparent strict separation occurs at

$$\delta = \frac{4}{13}, \qquad a = \frac{1}{25}, \qquad b = \frac{1}{1500}.$$

The spectral margin reduces to the exact positive rational number

$$4\big(a(1-a) - b(1-b)\big)^2 - \left(\frac{5}{13}\right)^2 a(1-a) = \frac{3182386369}{213890625000000}.$$

Rigorous bounds over the *entire* MRRW minimization interval give

$$R_2(4/13) \le H_2(1/25) - H_2(1/1500) < \frac{469}{2000} < M_2(4/13). \tag{2.5}$$

The small spectral margin concerns an eventual asymptotic theorem; it does not claim an improvement for short block lengths.

## 2.5 Why constant-weight shells are needed for every distance

Our whole-cube bound alone does not beat $M_2(\delta)$ everywhere. For example, around $\delta = 0.1$ its optimized value is about $0.700$, whereas the second MRRW bound is about $0.693$. This failure explains why a much longer second construction cannot be replaced by optimizing (2.4) more carefully.

Restrict the code to a weight-$\alpha n$ shell. Bassalygo–Elias averaging costs $1 - H_2(\alpha)$ when returning to the whole cube [16]. A point stabilizer now acts separately on the support and its complement. Attaching primitive harmonics of degrees $\beta n$ and $\gamma n$ to these two coordinate blocks gives the resulting exponent

$$1 - H_2(\alpha) + H_2(u) - \alpha H_2\left(\frac{\beta}{\alpha}\right) - (1-\alpha)H_2\left(\frac{\gamma}{1-\alpha}\right), \tag{2.6}$$

subject to the correct associated-Hahn spectral threshold.

At $\beta = \gamma = 0$, the Johnson recurrence recovers the full classical second-MRRW objective, including interior minimizers. Starting at such an interior minimizer, choose $\beta = \gamma = \varepsilon > 0$ and increase the retained degree by $O(\varepsilon)$ to maintain spectral feasibility. The ambient entropy rises by $O(\varepsilon)$, but the two primitive spaces contribute

$$2\varepsilon \log_2(1/\varepsilon) + O(\varepsilon).$$

Their logarithmic advantage dominates: activating a tiny stabilizer representation beats the spectral cost of restoring feasibility. When the classical minimizer is the endpoint corresponding to the first MRRW bound, the established whole-cube argument supplies the improvement instead.

The endpoint and interior arguments together give the full result

$$R_2(\delta) \le \min\{\kappa_{\mathrm{H}}(\delta), \kappa_{\mathrm{CW}}(\delta)\} < M_2(\delta), \qquad 0 < \delta < \tfrac{1}{2}.$$

The separate rational witness (2.5) makes the strict comparison particularly tangible. Higher-order Delsarte relaxations provide useful context, but are not the mechanism behind this two-point improvement [17].

## 2.6 From zonal lines to moving spherical harmonics

For $x \in \mathbb{S}^{n-1}$, the classical spherical construction retains one $SO(n-1)$-fixed line in every ambient harmonic space $\mathcal{H}_i(\mathbb{R}^n)$. Its zonal generator is a normalized Gegenbauer polynomial [13, 11]. Guided by the binary construction, we replace this fixed line by the moving stabilizer representation

$$E_x = \mathcal{H}_k(x^\perp), \qquad d_k = \binom{n+k-2}{k} - \binom{n+k-4}{k-2}.$$

Rotating $x$ to $y$ simultaneously transports $E_x$ and all its embeddings into $\mathcal{H}_i(\mathbb{R}^n)$.

This differs from the fixed-base-point matrix-valued semidefinite programs of Bachoc and Vallentin [18]: the stabilizer representation here moves with the codeword, while the final overlap $\mathrm{Tr}(P_x P_y)$ remains a scalar two-point kernel. The

distinction is crucial, because the improvement belongs to ordinary spherical linear programming rather than to a stronger three-point method.

The adjacent harmonic degrees are governed by

$$c_i^{(k)} = \frac{(i-k+1)(i+k+n-2)}{\sqrt{(i+1)(i+n-2)(2i+n-2)(2i+n)}}. \tag{2.7}$$

At $k=0$, these are the classical Gegenbauer coefficients. For $k>0$, the same moving-projection identity (2.2) gives the rank saving

$$A(n,s) \le \frac{1-s}{\lambda-s}\frac{\sum_{i=k}^{L}\dim \mathcal{H}_i(\mathbb{R}^n)}{\dim \mathcal{H}_k(x^\perp)}.$$

An individual associated Gegenbauer channel need not be positive: already in dimension four, $2t^2-1=\frac{3}{2}P_2(t)-\frac{1}{2}$ has a negative constant coefficient. Positivity belongs to the full weighted Gram remainder, not to its individual summands.

## 2.7 A hierarchy rather than one improved polynomial

Define the spherical entropy

$$h_{\text{sph}}(u) = (1+u)\log_2(1+u) - u\log_2 u.$$

For $L/n \to a$ and $k/n \to b$, the one-row construction has exponent $h_{\text{sph}}(a) - h_{\text{sph}}(b)$ and threshold

$$\frac{2(a-b)(1+a+b)}{(1+2a)\sqrt{a(1+a)}} > s.$$

Again the new term is exactly the entropy of the transported stabilizer fiber.

The next step is to allow general $SO(n-1)$ representations. The multiplicity-free branching of $SO(n)$ to $SO(n-1)$ gives interlacing row parameters

$$a_1 > b_1 > a_2 > \cdots > b_r > a_{r+1} \ge 0.$$

With $x_i = a_i(1+a_i)$, $y_j = b_j(1+b_j)$, and $q(a) = \sqrt{a(1+a)}/(1+2a)$, the residues

$$R_i = \frac{\prod_{j=1}^{r}(x_i - y_j)}{\prod_{\ell\neq i}(x_i - x_\ell)}$$

are positive and sum to one. The resulting spectral and dimension quantities become

$$\Gamma_r = \sum_i R_i q(a_i), \qquad \Phi_r = \sum_i h_{\text{sph}}(a_i) - \sum_j h_{\text{sph}}(b_j).$$

The finite representation-graph estimate gives the spherical rate $\Phi_r$ whenever $2\Gamma_r > s$.

A subtle detour is that improving every fixed parameter tuple does not automatically improve an optimized infimum: paired interlacing rows can escape together to infinity. Compactifying these escaping pairs produces a lower-level configuration with a scale factor $c \le 1$; its entropy shift is $-\log_2 c$. This missing boundary analysis is what allows a genuine comparison between consecutive hierarchy levels.

## 2.8 Comparing with the right spherical benchmark

We must compare with the correct classical spherical benchmark. The direct Kabatianskii–Levenshtein exponent is not the full classical benchmark. A spherical-cap reduction first changes $s$ to $t \le s$ at entropy cost

$$\frac{1}{2}\log_2\frac{1-t}{1-s}.$$

The optimized classical exponent is therefore

$$B_{\text{KL}}(s) = \inf_{0\le t\le s}\left[h_{\text{sph}}\left(\frac{(1-t^2)^{-1/2}-1}{2}\right) + \frac{1}{2}\log_2\frac{1-t}{1-s}\right],$$

not the whole-sphere value at $t=s$ [19, 11]. The same cap optimization accompanies every higher-level bound. If $\overline{\kappa}_r$ denotes its cap-optimized rate, the strict hierarchy is

$$R_{\text{sph}}(s) \le \inf_{r\ge 0}\overline{\kappa}_r(s) < \overline{\kappa}_1(s) < \overline{\kappa}_{\text{row}}(s) < B_{\text{KL}}(s), \qquad 0<s<1.$$

The whole-cube rational witness, Johnson-layer refinement, tangent harmonics, and multi-row representation graph therefore fit into one mathematical progression. In both geometries the classical positive scalar kernel survives, while replacing its discarded fixed line by an exponentially large moving harmonic space yields strict improvement for every fixed binary distance and spherical angle.

Chapter 3

# Constructing a Non-Sofic Group

## 3.1 Finite permutation models and the missing obstruction

We begin by asking whether every finitely presented group is sofic. A negative answer requires one finite multiplication table and a fixed $\varepsilon > 0$ for which no permutations simultaneously satisfy

$$d_Y(\sigma(gh), \sigma(g)\sigma(h)) < \varepsilon, \qquad d_Y(\sigma(g), 1) > 1 - \varepsilon.$$

The degree $|Y|$ is unrestricted, so defects may disappear into larger and larger approximate actions [20].

The eventual counterexample combines a binary Leavitt algebra, property (T), two self-similar compressions, and Thompson's group $V$. Its central difficulty is extracting one usable expander from many expanding components.

The early shortcuts fail for revealingly different reasons. A nonembeddable tracial algebra need not be a group factor with its canonical trace; projecting a linear-system group onto a central eigenspace does not preserve approximate multiplication; and undecidability provides no uniform bound on the size of permutation witnesses. Even stability and ascending HNN constructions leave the defect free to escape between sheets. What we need is an explicit combinatorial amplifier, not an indirect pathology.

## 3.2 Expanders with a forbidden centralizer

If a sofic approximation of $K \times J$ has a single uniformly expanding $K$-generator graph, Kun and Thom prove that $J$ is locally embeddable into finite groups, or LEF [21]. Expansion separates approximately central permutations into Hamming clusters, and repairing products turns those clusters into a finite multiplication table.

If approximately central permutations have commuting defect at most $\delta|Y|$, their agreement set has small boundary. Expansion forces their Hamming distance to be either at most $4\delta/\lambda$ or at least $1 - 4\delta/\lambda$; repairing these separated clusters produces a finite multiplication table.

We therefore want $K$ to commute with an infinite finitely presented simple group $J$, which cannot be LEF. The difficulty is obtaining *one* expander carrying the $J$ action rather than a disjoint union of unrelated expanders [22, 21].

## 3.3 Why a self-similar ring became the right playground

Central extensions and one-sided inverses did not prevent permutation defects from escaping among many sheets. Elementary matrices suggested something stronger: their commutators encode ring multiplication, their groups can have property (T), and a self-similar coefficient ring can compress the entire configuration into a proper corner.

Choose the binary Leavitt algebra over $\mathbb{F}_2$,

$$R = \mathbb{F}_2\langle s_0, s_1, t_0, t_1\rangle / \left(t_i s_j = \delta_{ij}, \quad s_0 t_0 + s_1 t_1 = 1\right).$$

For a binary prefix $a$, put $e_a = s_a t_a$. Incomparable prefixes satisfy $t_a s_b = 0$, while $t_a s_a = 1$, so $s_a r t_b$ behave as matrix units in the corresponding cylinder corners. A complete prefix family makes that corner the whole ring; in particular,

$$R \cong M_2(R), \qquad r \longmapsto (t_i r s_j)_{i,j\in\{0,1\}}.$$

Leavitt introduced rings with exactly this kind of nonclassical module type [23]. Its decisive feature is that an entire elementary group compresses into a proper corner without losing any coefficients.

## 3.4 Putting property (T) and Thompson's group into the same ring

For a prefix family $E$, write

$$\mathrm{EL}_E(R)$$

for the elementary group in its corresponding matrix corner. An elementary root has the concrete form

$$1 + s_a r t_b, \qquad a, b \in E, \quad a \neq b, \quad r \in R.$$

When the corner is proper, extend its units by the identity on the complement. Since $R$ is generated by four elements as a unital ring, the theorem of Ershov and Jaikin-Zapirain applies to the elementary groups of ranks at least three and supplies property (T) [24].

The same ring also contains the prefix-table model of Thompson's group $V$. Given a bijection between two complete binary prefix codes,

$$a_i \longmapsto b_i,$$

the corresponding ring unit is

$$w = \sum_i s_{b_i} t_{a_i}.$$

Composition of prefix tables agrees with multiplication of these units, and refining a table does not change the represented element. Acting on infinite binary strings shows that this copy of $V$ is faithful.

Characteristic two gives an elegant bridge. For incomparable prefixes, $P = s_a t_b$ and $Q = s_b t_a$ satisfy $P^2 = Q^2 = 0$, $PQ = e_a$, and $QP = e_b$; hence their cylinder transposition factors as

$$\tau_{a,b} = (1 + P)(1 + Q)(1 + P).$$

An auxiliary cylinder handles transpositions within one block. Thus the property-(T) elementary group and local Thompson group interact inside the same explicit algebra.

The relevance of $V$ is that it is infinite, simple, and finitely presented [25, 26, 27]; a finitely presented LEF group is residually finite, so

$$V \text{ is not LEF.}$$

We now have a concrete candidate for the forbidden centralizing group $J$.

## 3.5 The nine cylinders and the two compressions

To combine property (T), proper compression, a commuting local copy of $V$, and ambient generation, use the complete prefix code

$$\begin{aligned}
\alpha &= (000, 001, 01),\\
\beta &= (1000, 1001, 101),\\
\nu &= (1100, 1101, 111),\\
D &= (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \nu_1, \nu_2, \nu_3).
\end{aligned}$$

Its nine cylinders partition Cantor space. Set

$$G = \mathrm{EL}_D(R) \cong \mathrm{EL}_9(R), \qquad \Gamma = \mathrm{EL}_{(\alpha_1,\alpha_2,\alpha_3)}(R).$$

Both are finitely generated and have property (T).

To build the compressors, introduce the auxiliary code

$$\eta = (100, 101, 11)$$

and defined two prefix-table units by

| | $\alpha_i$ | $\beta_i$ | $\nu_i$ |
|---|---|---|---|
| $u$ | $\alpha_i 0$ | $\alpha_i 1$ | $\eta_i$ |
| $v$ | $\alpha_i 0$ | $\eta_i$ | $\alpha_i 1$. |

Both maps send the $\alpha_i$ cylinder into its left child. Therefore

$$u\Gamma u^{-1} = v\Gamma v^{-1} = \mathrm{EL}_{(\alpha_1 0, \alpha_2 0, \alpha_3 0)}(R) =: K \subseteq \Gamma.$$

The difference $v^{-1}u$ centralizes $\Gamma$. This shared compression was the structural reason for introducing *two* units rather than one: they realize the same copy of $\Gamma$ while exposing different complementary cylinders.

The ambient generation statement $G = \langle \Gamma, u, v \rangle$ requires a real coefficient calculation. Conjugating an $\alpha$-root by $u^{-1}$ produces the full matrix

$$\begin{pmatrix} t_0 r s_0 & t_0 r s_1 \\ t_1 r s_0 & t_1 r s_1 \end{pmatrix}$$

as $r$ ranges through $R \cong M_2(R)$. Thus $u$ generates the combined $\alpha, \beta$ blocks; $v$ generates the $\alpha, \nu$ blocks; and a commutator through an $\alpha$ cylinder supplies the missing $\beta, \nu$ roots.

Inside the right child of the first $\alpha$ cylinder, place

$$l = 0001, \qquad J = V_l \cong V.$$

The group $K$ lives entirely on the left-child cylinders $\alpha_i 0$, whereas $J$ lives on the disjoint cylinder $\alpha_1 1$. Consequently

$$[K, J] = 1, \qquad K \cap J = \{1\}, \qquad K \times J \hookrightarrow G.$$

At this point every required algebraic feature had a transparent geometric meaning: compression acts on left children, the forbidden Thompson group lives on a right child, and the two compressors together recover the full ambient elementary group.

## 3.6 The many-expander obstruction

Suppose, toward a contradiction, that $G$ has sofic models $Y_n$, with $N_n = |Y_n| \to \infty$. Kun's theorem edits $o(N_n)$ generator edges to decompose a property-(T) model into uniform expanders [22]. Applying it to $\Gamma$ and $G$ gives component partitions $\mathcal{Q}_n$ and $\mathcal{A}_n$; transporting $\mathcal{Q}_n$ through $t_1 = \sigma_n(u)$ and $t_2 = \sigma_n(v)$ gives the $K$-component partitions $\mathcal{P}_{i,n}$.

One sustained detour tried to convert property (T) directly into mixing. That requires a lazy or anti-bipartite averaging set: a bipartite graph can have spectrum near $-1$ despite a Kazhdan gap at 1. Correcting the averaging cures this spectral problem without choosing a component. The crucial mismatch is that Kun gives *many* expanders, whereas Kun–Thom requires *one*. On $Q \sqcup Q$, for example, the swap commutes exactly with identical expanding $K$ actions but preserves neither component. The self-similar configuration must overcome this specific obstruction.

## 3.7 The first comparison: component mass cannot drift freely

For a $\Gamma$ component $Q \in \mathcal{Q}_n$, define the size function

$$M(x) = |Q| \qquad (x \in Q).$$

Because $u\Gamma u^{-1}$ and $v\Gamma v^{-1}$ both lie in $\Gamma$, the transported generators correspond to fixed $\Gamma$-words. Their edges therefore cross the original $\Gamma$ partition on few vertices, provided all relevant cleaning and transport errors are controlled in the same finite model.

For a transported component $P \in \mathcal{P}_{i,n}$, let $D(P)$ be an original component maximizing $|P \cap D(P)|$, and put

$$L(P) = |P| - |P \cap D(P)|.$$

Expansion bounds the unmatched mass by crossing edges:

$$\sum_{P \in \mathcal{P}_{i,n}} L(P) = o(N_n).$$

Because a transported component has the same size as its source, a large intersection implies the one-sided inequality

$$M(t_i x) \geq (1 - \eta_n) M(x)$$

for most $x$, with $\eta_n \to 0$. The $\Gamma$ generators preserve $M$, but iteration does not suffice: $M$ is unbounded, and small relative losses may accumulate.

## 3.8 From logarithmic midranks to a bounded median

A substantial intermediate approach binned $\log M$ on a randomly shifted grid. Random shifting makes crossings of bin boundaries rare even when component sizes vary widely; replacing the resulting classes by bounded midranks then converts almost-monotonicity into concentration. The identity $\mathrm{Var}(\mathrm{midrank}) = (1 - \sum_j p_j^3)/12$ revealed the essential principle: average a bounded monotone function of component size, never the unbounded size itself.

The final argument implements that principle more cleanly. In each ambient expanding component $A$, choose a vertex-weighted median $m_A$ of the original sizes $M(x)$ and define

$$f(x) = \frac{M(x)}{M(x) + m_A}, \qquad x \in A.$$

Then $0 < f < 1$, and $1/2$ is a median on every $A$. The elementary inequality

$$\frac{(1-\eta)x}{(1-\eta)x + a} \geq \frac{x}{x+a} - \eta \qquad (x, a > 0)$$

converts the one-sided transport estimate into

$$f(sx) \geq f(x) - \eta_n$$

away from $o(N_n)$ vertices for every positive ambient generator $s$. Because $s$ is a permutation,

$$\sum_x (f(sx) - f(x)) = 0.$$

The boundedness of $f$ therefore makes both its positive and negative variation $o(N_n)$. On the edited ambient expander,

$$\sum_{\{x,y\} \in E(L_G)} |f(x) - f(y)| = o(N_n).$$

The finite coarea identity gives

$$\int_0^1 |\partial_A\{x : f(x) > t\}|\, dt = \sum_{\{x,y\}\in E(L_G[A])} |f(x) - f(y)|.$$

Splitting at the median and applying expansion to the smaller side of each cut yields

$$\gamma_G \sum_{x\in A} \left| f(x) - \frac{1}{2} \right| \le \sum_{\{x,y\}\in E(L_G[A])} |f(x) - f(y)|.$$

Thus, outside a set $E_n$ of $o(N_n)$ vertices,

$$\left| f(x) - \frac{1}{2} \right| \le \delta_n, \qquad \delta_n \longrightarrow 0.$$

For nonexceptional $x, y$ in the same ambient component,

$$\rho_n^{-1} \le \frac{M(y)}{M(x)} \le \rho_n, \qquad \rho_n = \left( \frac{1 + 2\delta_n}{1 - 2\delta_n} \right)^2 \longrightarrow 1.$$

This deterministic median argument is the final substitute for the earlier randomized grid: it compares component sizes precisely where the transport endpoints meet inside one ambient expander.

## 3.9 From median concentration to component matching

Consider a transported component

$$P = t_i Q$$

that meets an original component $D(P)$ on at least

$$(1 - \eta_n)|P|$$

vertices. Suppose further that there is a witness $x \in Q$ such that

$$t_i x \in P \cap D(P), \qquad x, t_i x \notin E_n,$$

and $x, t_i x$ belong to the same ambient component. The first condition compares the source component $Q$ with its target $D(P)$; median concentration then gives

$$\rho_n^{-1}|P| \le |D(P)| \le \rho_n |P|,$$

and hence

$$|P \triangle D(P)| \le \big(\rho_n - 1 + 2\eta_n\big)|P| = o(|P|).$$

For sufficiently large $n$, the intersection contains more than half of $D(P)$. Since distinct transported components are disjoint, they cannot both have majority intersection with the same original component. The assignment

$$P \longmapsto D(P)$$

is therefore injective on the good components. This is stronger than merely showing that the two partitions have approximately the same block sizes: it identifies actual blocks inside the same finite model.

Transport crossings, exceptional sets, and nondominant components occupy only $o(N_n)$ vertices, so the matched intersections cover $N_n - o(N_n)$ vertices.

Every fixed $\Gamma$-word preserves the original partition outside $o(N_n)$ vertices. On matched intersections, injectivity forces it to preserve the transported partition as well:

$$\#\{x : P(x) \ne P(wx)\} = o(N_n).$$

This resolves the factor-crossing difficulty without asserting that a compressor preserves the original components just because its central ratio does. The preserved objects are the matched transported components, obtained through the common ambient median and majority intersection.

## 3.10 Selecting one component carrying every relevant word test

For a fixed radius, collect into one exceptional set all failed word equalities, failed distinctness tests, generator exits, edited edges, and nonretained components. Each such set has size $o(N_n)$. Choose the word radius

$$r_n \longrightarrow \infty$$

slowly enough that the combined global exceptional proportion $e_n$ still tends to zero. Averaging over retained components with weights $|P|$ supplies one transported block satisfying

$$|P_n \cap B_n| \le 2e_n|P_n|.$$

Therefore every $K \times J$ relation, commutation test, and freeness test of radius $r_n$ holds simultaneously on this same component. A good vertex distinguishes all relevant $K$ words, so infinitude of $K$ implies

$$|P_n| \longrightarrow \infty.$$

The generators of $K$ and $J$ act on most of $P_n$, but their restrictions are only partial permutations. Their missing domains and ranges have the same size, so complete each partial bijection by matching the missing points arbitrarily. This changes only the controlled boundary vertices. The resulting full permutations continue to satisfy each fixed product, commutation, and freeness test asymptotically.

However, completing permutations can destroy expansion on a tiny set. The correct intermediate conclusion is an *additive* Cheeger inequality,

$$|\partial X| \ge \gamma \min\big(|X|, |P_n \setminus X|\big) - a_n|P_n|, \qquad a_n \longrightarrow 0.$$

We cannot discard the additive error: for a small set $X$, it may be larger than the desired lower bound. Instead, fix

$$0 < \lambda < \gamma$$

and choose a maximal set

$$B_n \subseteq P_n, \qquad |B_n| \le \frac{1}{2}|P_n|,$$

for which

$$|\partial B_n| < \lambda|B_n|.$$

The additive inequality forces

$$|B_n| \le \frac{a_n}{\gamma - \lambda}|P_n| = o(|P_n|).$$

Maximality then shows that the induced graph on

$$Z_n = P_n \setminus B_n$$

has a genuine uniform expansion constant, with the near-half-size case handled separately using the same additive inequality. Completing the remaining partial permutations on $Z_n$ preserves this expansion and changes only $o(|P_n|)$ vertices.

The outcome is precisely the object needed by the Kun–Thom theorem: an approximation of $K \times J$ on one expanding $K$ component. That theorem therefore implies that $J$ is LEF, contradicting

$$J \cong V.$$

The maximal-cut repair is essential: it turns many loosely compatible expanding components into the single expander required by the centralizer obstruction. Since $V$ is infinite, simple, and finitely presented, it cannot be LEF: finitely presented LEF groups are residually finite [28]. Thus $G$, and consequently the full unit group $L_{\mathbb{F}_2}(1,2)^\times$, are not sofic.

## 3.11 Extracting an infinite finitely presented counterexample

The finite-presentation step applies to the established counterexample. If $G$ fails the finite test $(F, \varepsilon)$, put $T = \{1\} \cup F \cup F^2$ and form

$$H_F = \langle X_a \ (a \in T) \mid X_1 = 1, \quad X_{gh} = X_g X_h \ (g, h \in F) \rangle .$$

Evaluation $X_a \mapsto a$ preserves the nontrivial tested generators. If $H_F$ were sofic, its defining relations would give exactly the forbidden approximation of $F$ with the same tolerance. Enlarging $F$ to contain generators of $G$ also makes $H_F \twoheadrightarrow G$, hence $H_F$ infinite. Thus one obtains an infinite finitely presented non-sofic group.

For a free presentation $F_d \twoheadrightarrow G$ with kernel $N$, soficity of $G$ is equivalent to co-soficity of the invariant random subgroup $\delta_N$. Thus the example also gives a non-co-sofic normal point mass [29], without asserting non-hyperlinearity.

The decisive progression is from $R \cong M_2(R)$ to two compatible compressions, then from logarithmic midranks to bounded-median matching. Simultaneous word tests and maximal-cut repair finally produce the single expander that forces $V$ to be LEF. Its contradiction establishes both non-soficity and the infinite finitely presented counterexample.

Chapter 4

# A Counterexample to Connes Rigidity

## 4.1 The obstacle was not finding equal factors

Connes's rigidity conjecture asks whether an ICC group with Kazhdan's property (T) can be reconstructed from its group von Neumann algebra. We seek two groups

$$\Gamma \ncong \Lambda, \qquad L(\Gamma) \cong L(\Lambda),$$

with both groups ICC and both having property (T). Equal group algebras without these extra conditions are abundant: Connes's classification makes the factors of amenable ICC groups exceptionally nonrigid [30]. The difficulty is to preserve both rigidity hypotheses while hiding an honest difference between the underlying groups.

The first sustained attempt began with nonisomorphic finite groups having the same complex group algebra. A finite normal subgroup destroys ICC, while distributing the finite difference over lamp coordinates in $F^{(\mathcal{J})} \rtimes K$ destroys property (T): Bernoulli states with sparser and sparser nontrivial lamps yield almost invariant vectors. Naive cohomological pushouts and root-adjoining merely relocate this obstruction: they retain finite central conjugacy classes or identify fibers without intertwining the acting group. A separate construction instead couples an infinite symplectic $\mathbb{F}_2$-module to a root-graded property-(T) group. Its finite-matrix ideal contains arbitrarily large finite symplectic groups supporting a nontrivial pseudosymplectic cocycle, while its Laurent-polynomial quotient excludes locally finite normal subgroups. The binary-carry route below makes the same measured-versus-algebraic distinction more transparent.

The decisive change is to ask what a crossed product actually remembers. For an abelian group $A$ acted on by $K$,

$$L(A \rtimes K) \cong L^\infty(\widehat{A}, m_{\widehat{A}}) \rtimes K.$$

The algebra on the right depends on the probability space $\widehat{A}$, its Haar measure, and the measurable $K$ action. It need not remember the compact group law on $\widehat{A}$. Thus the correct target was not two isomorphic $K$-modules. It was two *different compact group laws on the same $K$ probability space*. This distinction between measurable conjugacy and algebraic conjugacy organizes the whole construction [31].

## 4.2 Choosing an acting group that cannot hide order-four torsion

We need an ICC, torsion-free, property-(T) acting group with a large characteristic-two quotient. The useful choice was

$$K = \ker\left(\mathrm{SL}_4(\mathbb{Z}[t]) \xrightarrow{g(t)\mapsto g(0) \bmod 3} \mathrm{SL}_4(\mathbb{F}_3)\right).$$

This is a finite-index subgroup of a universal lattice. Suslin's elementary-generation theorem identifies the relevant elementary and special linear groups, while the theorem of Ershov and Jaikin-Zapirain gives property (T) for the universal lattice and therefore for $K$ [32, 24].

Minkowski's congruence lemma makes the level-three constant-term subgroup torsion-free [33]. If $g(t) \in K$ had finite order, then $g(0) = \mathbf{1}$; writing

$$g(t) = \mathbf{1} + t^r A(t), \qquad r \geq 1, \qquad A(0) \neq 0,$$

gives

$$g(t)^m \equiv \mathbf{1} + mt^r A(0) \pmod{t^{r+1}},$$

contradicting $g(t)^m = \mathbf{1}$. Thus $K$ has no torsion.

At the same time, reduction modulo two is surjective:

$$\pi_2 : K \twoheadrightarrow Q := \mathrm{SL}_4(\mathbb{F}_2[t]).$$

Indeed,

$$\mathbf{1} + 3f(t)E_{ij} \in K$$

reduces modulo two to every elementary matrix $\mathbf{1} + \overline{f}(t)E_{ij}$, and those matrices generate $Q$. Hence the module calculations can take place in characteristic two even though the actual acting group is torsion-free. This separation is essential: $Q$ itself contains elements of order four and could not distinguish the two final groups by torsion.

The same elementary conjugations show that $K$ is ICC. A nonidentity matrix cannot commute with every $E_{ij}$ unless it is scalar; a scalar element would have finite order and hence must be trivial. Conjugating by

$$\mathbf{1} + 3f(t)E_{ij}, \qquad f(t) \in \mathbb{Z}[t],$$

then produces infinitely many distinct conjugates.

## 4.3 The quadratic Boolean module that makes the carry equivariant

Set

$$R = \mathbb{F}_2[t], \qquad V = R^4, \qquad b(v) = v \otimes v.$$

The tensor product here is over $\mathbb{F}_2$, not over $R$. Set

$$B = \operatorname{span}_{\mathbb{F}_2}\{v \otimes v : v \in V\} \subseteq V \otimes_{\mathbb{F}_2} V.$$

This choice took some work. An ordinary symmetric-square quotient would obscure the characteristic-two polarization and the Boolean coordinates needed for the later spectral estimate. The actual divided square keeps both the diagonal vectors and the polarized terms:

$$b(u+v) - b(u) - b(v) = u \otimes v + v \otimes u.$$

The split abelian module is

$$D = V \oplus B.$$

It has exponent two and sits in the split exact sequence

$$0 \longrightarrow V \longrightarrow D \longrightarrow B \longrightarrow 0.$$

Write

$$X = \operatorname{Hom}_{\mathbb{F}_2}(V, \mathbb{F}_2), \qquad Y = \operatorname{Hom}_{\mathbb{F}_2}(B, \mathbb{F}_2).$$

Then

$$\widehat{D} = X \times Y$$

with coordinatewise addition and product Haar measure. The action of $K$ on these compact coordinates factors through $Q$.

We now seek a second group law on exactly the same coordinates. For $a \in \mathbb{F}_2$, let $\widetilde{a} \in \{0, 1\}$ be its integer lift, and define

$$F_{\ell,q}(v) = \widetilde{\ell(v)} + 2\widetilde{q(b(v))} \pmod 4, \qquad (\ell, q) \in X \times Y.$$

The elementary binary-carry identity

$$\widetilde{a} + \widetilde{a'} = \widetilde{a + a'} + 2\widetilde{a}\,\widetilde{a'} \pmod 4$$

produces the decisive cocycle. If

$$r_{\ell,\ell'} = (\ell \otimes \ell')|_B,$$

then

$$F_{\ell,q} + F_{\ell',q'} = F_{\ell+\ell',\, q+q'+r_{\ell,\ell'}}.$$

Thus the functions $F_{\ell,q}$ form a compact subgroup

$$C \subseteq (\mathbb{Z}/4\mathbb{Z})^V,$$

and the same coordinate space $X \times Y$ acquires the alternative law

$$(\ell, q) \star (\ell', q') = \big(\ell + \ell', q + q' + r_{\ell,\ell'}\big).$$

The map

$$\Phi : X \times Y \longrightarrow C, \qquad \Phi(\ell, q) = F_{\ell,q},$$

is a $K$-equivariant homeomorphism. Crucially, it also preserves Haar measure. Translation in $C$ becomes the triangular transformation

$$(\ell, q) \longmapsto \big(\ell + \ell', q + q' + r_{\ell,\ell'}\big).$$

Successive translations in $X$ and $Y$ preserve product measure, which is therefore Haar for the $\star$ law.

This was the conceptual breakthrough: the identity on the underlying coordinates is an equivariant probability-space isomorphism, but it is not an isomorphism between the two compact group laws. The carry is invisible to the measured action and visible to the discrete dual.

## 4.4 The same factor, but different torsion

Let

$$E = \widehat{C}.$$

Pontryagin duality supplies a nonsplit exact sequence

$$0 \longrightarrow V \longrightarrow E \longrightarrow B \longrightarrow 0,$$

parallel to the split sequence defining $D$. The evaluation character

$$\varepsilon_v(F) = i^{F(v)}, \qquad F \in C,$$

satisfies

$$\varepsilon_v^2 = \iota(v), \qquad \sigma(\varepsilon_v) = b(v).$$

Taking $e = (1,0,0,0)$ and a functional $\ell$ with $\ell(e) = 1$ gives

$$\varepsilon_e(F_{\ell,0}) = i,$$

so $\varepsilon_e$ has order four. Hence $E$ cannot be isomorphic to the exponent-two group $D$.

Form the two semidirect products

$$\Gamma = E \rtimes K, \qquad \Lambda = D \rtimes K.$$

Fourier transform and the equivariant Haar-preserving map $\Phi$ give

$$\begin{aligned} L(\Gamma) &\cong L^\infty(\widehat{E}, m_{\widehat{E}}) \rtimes K \\ &\cong L^\infty(X \times Y, m_X \otimes m_Y) \rtimes K \\ &\cong L(\Lambda). \end{aligned}$$

There is no requirement that $\Phi$ respect the compact group law: only measure preservation and intertwining of the $K$ actions enter the crossed product.

On the other hand, $\Gamma$ contains the order-four element $\varepsilon_e$. Any finite-order element $(d,k)$ of $\Lambda$ must project to a finite-order element of $K$. Torsion-freeness forces $k = 1$, and then $d \in D$ has order at most two. Therefore

$$\Gamma \ncong \Lambda.$$

The characteristic-two quotient builds the carry; its torsion-free lift keeps the order-four signature visible.

## 4.5 Why the groups should be ICC

For an abelian-by-$K$ group, a convenient ICC criterion is that $K$ is ICC and every nonzero element of the abelian normal subgroup has an infinite $K$ orbit. The remaining issue is therefore the common modules $V$ and $B$.

For nonzero $v \in V$, choose a nonzero coordinate $v_j$ and $i \neq j$. The elementary matrices

$$g_n = \mathbf{1} + t^n E_{ij}$$

give distinct vectors

$$g_n v = v + t^n v_j e_i.$$

For the tensor module, the useful correction is to keep the two tensor variables independent:

$$V \otimes_{\mathbb{F}_2} V \cong \mathbb{F}_2[t_1, t_2]^{16}.$$

After removing the maximal common power of $t_2$ and specializing at $t_2 = 0$, equality of two orbit points would force

$$(t_1^n + t_1^m)z = 0$$

for a nonzero vector $z$ over $\mathbb{F}_2[t_1]$. Since that ring is an integral domain, $n = m$. Every nonzero element of $B$ therefore has an infinite orbit.

For $D = V \oplus B$, projection onto a nonzero coordinate proves the criterion directly. For $E$, its exact sequence reduces the argument to the same two cases: an element with nonzero image in $B$ has infinite orbit, and an element in the kernel is a nonzero element of $V$. This explains why the split and carry extensions have parallel ICC arguments despite being different as groups.

## 4.6 The critical property-(T) argument is relative

The most delicate remaining issue is property (T). Knowing that $K$ has property (T) is *not* enough: the groups $D$ and $E$ are infinite abelian, and an arbitrary extension of a Kazhdan group by an infinite abelian group need not be Kazhdan. We must instead establish relative property (T) of

$$(D \rtimes K, D) \quad \text{and} \quad (E \rtimes K, E).$$

This is exactly the point at which the unsuccessful wreath-product attempt had failed, so it could not be treated as a formality.

The spectral criterion is conveniently expressed on the compact dual. Given an abelian $K$-module $A$, suppose there are a finite set $J \subset A$ and $c > 0$ such that every $K$-invariant probability measure $\mu$ on $\widehat{A}$ satisfies

$$\sum_{a \in J} \int_{\widehat{A}} |\chi(a) - 1|^2 \, d\mu(\chi) \geq c\big(1 - \mu(\{1\})\big).$$

Property (T) of $K$ converts almost invariant vectors into $K$-invariant ones. Their spectral measures satisfy this inequality, which forces an atom at the trivial character and hence full invariance. This relative property-(T) step does not follow from universal-lattice property (T) alone [34].

To establish the inequality, let

$$V_N = \{v \in \mathbb{F}_2[t]^4 : \deg v_i < N \text{ for every } i\}.$$

A vector is primitive when its four coordinates generate the unit ideal. Elementary row reduction makes all primitive vectors a single $Q$-orbit. Counting by their monic greatest common divisor gives the exact formula

$$|V_N| = 2^{4N}, \qquad |\operatorname{Prim}(V_N)| = 7 \cdot 2^{4N-3} + 1.$$

Thus primitive vectors occupy asymptotic density $7/8$.

For $(\ell, q) \in X \times Y$, the two detection functions are

$$v \longmapsto \ell(v), \qquad v \longmapsto q(v \otimes v).$$

The first is Boolean linear and the second is Boolean quadratic in the $4N$ coefficients of $v$. Every nonzero Boolean polynomial of degree at most two is nonzero on at least one quarter of the cube. Since only one eighth of the cube is nonprimitive, at least one eighth consists of primitive detecting vectors. Relative to the $7/8$ primitive density, this yields the uniform lower bound

$$\mu\left\{(\ell, q) : \big(\ell(e), q(e \otimes e)\big) \neq (0,0)\right\} \geq \frac{1}{7}\big(1 - \mu(\{(0,0)\})\big)$$

for every $K$-invariant probability measure $\mu$.

For $D$, the two detecting bits correspond to the two elements

$$(e, 0), \qquad (0, e \otimes e).$$

For $E$, they are encoded together in the low and high bits of

$$F_{\ell,q}(e) \in \mathbb{Z}/4\mathbb{Z},$$

so the single character $\varepsilon_e$ detects both. The same $1/7$ estimate therefore supplies the spectral criterion for both semidirect products. Rank four now has a clear explanation: in rank $j$, the corresponding lower bound is

$$\frac{1/4 - 2^{1-j}}{1 - 2^{1-j}},$$

which is positive precisely when $j \geq 4$.

There is a useful economy in the argument. It is enough to establish ICC and property (T) directly for

$$\Lambda = D \rtimes K.$$

Its common group factor is a $\mathrm{II}_1$ factor, so every group realizing that factor is ICC. Moreover, property (T) of a countable group is equivalent to property (T) of its group von Neumann algebra [35]. Hence the same factor isomorphism transfers both properties to the carry group, rather than requiring an entirely separate relative-property-(T) argument for every later extension.

## 4.7 Shifting the carry produces an entire infinite fiber

The original two-group mechanism also admits a stronger refinement: the carry can begin at any coefficient position without changing the measured action. Define a continuous $K$-equivariant shift

$$T_n : X \longrightarrow X, \qquad (T_n \ell)(v) = \ell(t^n v).$$

It is surjective, and its kernel has size

$$Z_n \cong \mathrm{Hom}_{\mathbb{F}_2}(V/t^nV, \mathbb{F}_2), \qquad |Z_n| = 2^{4n}.$$

Pull the original carry back along this shift:

$$(\ell, q) \star_n (\ell', q') = \big(\ell + \ell', q + q' + r_{T_n\ell, T_n\ell'}\big).$$

Write $C_n = (X \times Y, \star_n)$, $E_n = \widehat{C}_n$, and

$$\Gamma_n = E_n \rtimes K.$$

All $C_n$ have the same product Haar measure and the same coordinate $K$-action. Consequently

$$L(\Gamma_n) \cong L^\infty(X \times Y, m_X \otimes m_Y) \rtimes K \cong L(\Lambda) \qquad (n \geq 0).$$

The map

$$p_n : C_n \longrightarrow C_0, \qquad p_n(\ell, q) = (T_n\ell, q),$$

is a $K$-equivariant surjective homomorphism with kernel $Z_n \times \{0\}$. Duality therefore embeds $\Gamma_0$ in $\Gamma_n$ with the exact index

$$[\Gamma_n : \Gamma_0] = 2^{4n}.$$

In particular, the resulting family is mutually commensurable.

The index alone is not an isomorphism invariant, so we need an intrinsic way to recover $n$. Because $K$ is torsion-free and $E_n$ has exponent four, the torsion subgroup of $\Gamma_n$ is precisely $E_n$. Consider the characteristic quotient

$$A_n = E_n[2]/2E_n.$$

There is a $K$-equivariant map

$$d : B \longrightarrow V, \qquad d(b(v)) = v,$$

and the shifted carry gives

$$2E_n = \iota_n(t^nV), \qquad 0 \longrightarrow V/t^nV \longrightarrow A_n \longrightarrow \ker d \longrightarrow 0.$$

Every nonzero element of $\ker d$ has infinite $K$-orbit, whereas $V/t^nV$ is finite. Therefore

$$\big|\{a \in A_n : |K \cdot a| < \infty\}\big| = |V/t^nV| = 2^{4n}.$$

This number is determined by the abstract group: an isomorphism preserves its characteristic torsion subgroup, the quotient action, and the finite-orbit locus. Hence $\Gamma_n \cong \Gamma_m$ implies $n = m$. The conclusion is an entire countably infinite fiber of the group-factor functor, strengthening the original two-group mechanism rather than merely producing indistinguishable factors [36, 37].

## 4.8 The complete rigidity mechanism

The guiding idea is now compact enough to state in one diagram:

$$D = V \oplus B, \qquad 0 \longrightarrow V \longrightarrow E \longrightarrow B \longrightarrow 0,$$
$$(\widehat{D}, m_{\widehat{D}}, K) \cong (\widehat{E}, m_{\widehat{E}}, K),$$
$$D \text{ has exponent } 2, \qquad E \text{ contains order-four torsion.}$$

The quadratic carry changes the discrete group while preserving the measurable dual action. The torsion-free universal lattice separates the groups; Boolean support and primitive-vector counting preserve property (T); infinite module orbits ensure ICC.

This construction does not contradict existing $W^*$-superrigidity theorems, which establish reconstruction for particular classes or particular groups rather than for every property-(T) ICC group [38, 39]. Nor does it settle the analogous reconstruction questions for higher-rank lattices themselves.

The algebraic and analytic pieces now fit coherently: the binary carry changes torsion without changing the measured dual, quadratic Boolean support supplies the Kazhdan gap, and finite orbits recover the shifted index. Together, these ingredients explain how a single von Neumann algebra arises from infinitely many pairwise nonisomorphic property-(T) ICC groups.

Chapter 5

# Permanent Circuits: Critical Cones, Matching Sums, and Root-of-Unity Cancellation

## 5.1 The gradient promised much more than an input-counting bound

For

$$\operatorname{per}_n(X) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i,\sigma(i)}, \qquad N = n^2,$$

We seek a circuit lower bound genuinely larger than $N$. A circuit may share intermediate values freely, so formula-specific occurrence counting is unavailable [40]. Instead, Baur–Strassen differentiation produces all first derivatives with constant-factor circuit overhead [41]:

$$\partial_{ij} \operatorname{per}_n = \operatorname{per}(X_{\widehat{i,j}}).$$

The first natural hope was that the gradient map had enormous generic degree. At the all-ones matrix,

$$\operatorname{Hess}(\operatorname{per}_n)(\mathbf{1}) = (n-2)!(J-I) \otimes (J-I)$$

is invertible, so the gradient is generically finite. But its coordinates have a large common base locus: two zero rows already give a critical component of codimension $2n$. Thus the naive Bézout count $(n-1)^N$ is unjustified. The determinant's high-degree but essentially birational adjugate map makes this failure particularly clear.

Neither coefficientwise absolute values nor separate differentiation repairs the problem. Absolute values erase intermediate cancellations, turning an arbitrary circuit into a monotone support computation; the tempting exponential matching-rectangle count then no longer applies. Differentiating all gradient coordinates separately likewise duplicates the circuit. A Hessian–vector product is inexpensive, but the entire Hessian is not.

## 5.2 Global Newton-volume arguments lost control of matching faces

On the torus, normalized logarithmic gradients are

$$q_{ij} = \frac{x_{ij}\partial_{ij} \operatorname{per}_n}{\operatorname{per}_n}, \qquad \sum_j q_{ij} = \sum_i q_{ij} = 1.$$

After removing row and column scalings, their dimension is $(n-1)^2$. The Birkhoff polytope suggests a normalized volume of order $\exp(\Theta(n^2 \log n))$, but Bernstein's theorem achieves that count only under face nondegeneracy. A disconnected matching face can factor as $f_1 f_2$, and

$$f_1 = f_2 = 0 \quad \Longrightarrow \quad \nabla(f_1 f_2) = 0.$$

The permanent's coefficients cannot be treated as independent.

Freezing $n - r$ rows produces coefficients

$$F_B(Y) = \sum_{\phi:[r]\hookrightarrow[n]} c_{\operatorname{im}\phi} \prod_i y_{i,\phi(i)}, \qquad c_I = \operatorname{per}(B_{[n-r],[n]\setminus I}).$$

All injections with the same image share one coefficient; generic choice of $B$ does not remove this dependency. Hyperplane-complement recursions suffer cancellations among Euler characteristics, and tropical exponent determinants are constrained by alternating paths of one common optimal matching. These sustained failures identified the real problem: we needed a permanent-specific geometric invariant that did not presume generic coefficients [42, 43].

A more sustained attempt computes the missing degree by eliminating one variable row at a time. Writing

$$F_r(Z, y) = \sum_j c_j(Z) y_j,$$

the torus fiber is a hyperplane complement whose Euler characteristic vanishes whenever a free torus factor remains. For two rows, a generic arrangement heuristic gives roughly $\binom{2n-2}{n-1}$, whose logarithm is only linear in $n$. Iterating the heuristic can even exceed the global Bézout upper bound, showing that singular discriminant strata must cancel substantial contributions. Tropicalization meets the same dependence: near-perfect matchings come from alternating paths of one global optimum, so their exponent vectors cannot be treated as independent.

Further variants confirm the same obstruction. Rank-one and Fourier specializations leave large singular strata; Ryser-type expansions introduce too many dependent coefficient parameters. Perturbing each gradient coordinate by an independent Fermat power costs $O(m \log d)$, exactly the multiplication budget a successful degree argument would recover. One shared power is inexpensive but leaves the original critical cone wherever its underlying linear form vanishes. The desired high-degree geometry must therefore come from the permanent itself.

## 5.3 A critical cone can concentrate exponential degree at one point

The decisive change was to consider a homogeneous degree-$d$ polynomial $P$ in $m$ variables with

$$\operatorname{codim} \operatorname{Crit}(P) \geq k, \qquad \operatorname{Crit}(P) = V(\partial_1 P, \ldots, \partial_m P).$$

A generic $k$-dimensional subspace $W$ avoids all nonzero critical directions. A second generic linear map $A$ combines the gradient outputs so that

$$F(u) = A\nabla P(Wu) : \mathbb{C}^k \longrightarrow \mathbb{C}^k, \qquad F^{-1}(0) = \{0\}.$$

The second choice is essential: merely restricting all $m$ gradient coordinates does not provide a square system.

We can check the first choice directly in the Grassmannian. The space of $k$-planes has dimension $k(m-k)$; those containing a fixed critical direction have dimension $(k-1)(m-k)$. The projectivized critical cone has dimension at most $m-k-1$, so all bad planes together have dimension at most

$$(k-1)(m-k) + (m-k-1) = k(m-k) - 1.$$

Thus some $k$-plane avoids every nonzero critical direction. A second incidence argument chooses the output combinations so their common kernel misses the entire gradient image.

The first interpretation concentrates the intersection at the origin: $k$ degree-$(d-1)$ gradient combinations form a regular sequence with Hilbert series

$$\frac{(1-z^{d-1})^k}{(1-z)^k} = (1 + z + \cdots + z^{d-2})^k.$$

Their local intersection length is $(d-1)^k$.

A further pivot expresses the same geometry through generic fibers and multiplicative complexity. Since $F$ has no nonzero homogeneous zero, homogenizing $F(u) = a$ creates no solutions at infinity. A generic fiber therefore has exactly $(d-1)^k$ distinct points. If $q$ multiplication gates compute $F$, introduce one variable and one quadratic equation per gate; the $k$ output equations are affine. The isolated-fiber Bézout bound gives [44]

$$(d-1)^k \leq 2^q.$$

Reverse-mode differentiation uses at most three multiplications per original multiplication: one forward product and its two adjoint updates. Free affine input and output maps then yield

$$\boxed{L \geq \frac{k \log_2(d-1)}{3}}.$$

This refined formulation avoids the additive input-dimension loss in the initial local-intersection accounting.

## 5.4 Columnwise inclusion–exclusion bounds the matching critical locus

The needed specialization must keep $m = \Theta(n^2)$ variables while attaining critical codimension $\Theta(m)$. Its basic piece is the rectangular matching polynomial

$$M_{t,s,d}(X) = \sum_{\substack{I \subseteq [t],\ J \subseteq [s] \\ |I| = |J| = d}} \operatorname{per}(X_{I,J}).$$

The initial square-zero encoding

$$u_i^2 = 0, \qquad \ell_a = \sum_i x_{ia} u_i$$

identified matching sums with coefficients of $e_d(\ell_1, \ldots, \ell_s)$. The essential insight was that, after conditioning on finitely many shared quantities, the gradient equations become separate constraints on each column.

To make this observation effective, introduce the shared power sums

$$p_B(X) = \sum_{a=1}^{s} \prod_{i \in B} x_{ia}, \qquad \varnothing \neq B \subseteq [t],$$

of which there are $2^t - 1$. Inclusion–exclusion over partitions enforces that the remaining $d-1$ matched rows choose distinct columns. For $q = d - 1$, the resulting column equations have leading form

$$Q_i(v; p) = (-1)^q q!\, e_q(v_1, \dots, \widehat{v_i}, \dots, v_t) + R_i(v; p), \qquad \deg_v R_i < q.$$

Record also the first $q - 1 = d - 2$ elementary symmetric functions of each column. Summing the $Q_i$ determines the missing $e_q(v)$, since

$$\sum_i e_q(v_1, \dots, \widehat{v_i}, \dots, v_t) = (t - q)e_q(v).$$

Substitution into an individual equation gives a monic reduction of $v_i^q$ to lower column degree. Consequently the incidence algebra is finite over a parameter ring with

$$2^t - 1 + s(d - 2)$$

generators. Passing from that enlarged incidence space to the actual gradient ideal can only decrease dimension, so

$$\dim \mathrm{Crit}(M_{t,s,d}) \le 2^t - 1 + s(d - 2).$$

This uniform argument controls the actual dependent permanental coefficients rather than replacing them with generic ones.

More explicitly, the leading coefficient in the reduction is $(t-q)q!$, which is nonzero because $q = d-1 < t$ and the ground field has characteristic zero. Whenever a column monomial contains $v_i^q$, we replace that factor by terms of strictly lower total column degree. Iteration terminates and leaves only monomials with every coordinate exponent below $q$. The incidence algebra is therefore a finite module over the ring generated by the $2^t - 1$ shared power sums and the $s(q-1)$ recorded column invariants. This justifies the dimension estimate even if those parameter values themselves satisfy algebraic relations.

The collision cancellation can be made explicit. For a row set $I$ and a partition $\pi$ of $I$, its Möbius coefficient is

$$\mu(\pi) = \prod_{B \in \pi} (-1)^{|B|-1}(|B| - 1)!.$$

The expression

$$\sum_\pi \mu(\pi) \prod_{B \in \pi} \left( p_B - \prod_{i \in B} v_i \right)$$

retains only injective assignments to columns other than the chosen one. A collision fiber of size greater than one has coefficient equal to the sum of signs of all permutations of that fiber, namely zero; a singleton fiber contributes one. The highest column degree counts all permutations of $I$, explaining the factor $q!$. This gives an identity for the actual matching coefficients rather than a heuristic independence assumption.

There is a subtle cancellation worth preserving. An actual derivative of the matching polynomial omits its own column, whereas the incidence polynomial contains high powers of that column. Those powers cancel after the shared parameters are specialized to their true values. Treating the parameters independently only enlarges the incidence variety, so its finite-module dimension remains a valid upper bound for the genuine critical locus.

## 5.5 Root-of-unity cancellation embeds many blocks into one permanent

A product of favorable blocks would be useless: $\nabla(f_1 f_2) = 0$ whenever both factors vanish. A sum in disjoint variables behaves differently:

$$\mathrm{Crit}\left( \sum_h \lambda_h f_h \right) = \prod_h \mathrm{Crit}(f_h), \qquad \lambda_h \neq 0.$$

Its critical codimensions add. The remaining challenge is to realize this sum by *one* permanent substitution, not by separate circuit evaluations.

A four-by-four example reveals the necessary cancellation:

$$\mathrm{per} \begin{pmatrix} u & v & 1 & 1 \\ w & z & 1 & 1 \\ p & q & 2 & -2 \\ r & s & 2 & -2 \end{pmatrix} = -8(uz + vw) + 2(ps + qr).$$

The mixed-block complementary permanent vanishes, while the two within-block permanents remain nonzero.

For the general construction, partition $r = bt$ rows into blocks $R_h$, write

$$y_h = \sum_{i \in R_h} z_i, \qquad z_i^2 = 0, \qquad \theta = (-1)^{d+1} 2^d,$$

and choose a primitive $d$th root of unity $\zeta$. The constant columns consist of $t-d$ copies of each block indicator and, for $h \geq 2$, the $d$ columns

$$\mathbf{1}_{R_1 \cup \cdots \cup R_{h-1}} + 2\zeta^j \mathbf{1}_{R_h}, \qquad 0 \leq j < d.$$

Their square-zero product is

$$\left(\prod_h y_h^{t-d}\right) \prod_{h=2}^{b} \left((y_1 + \cdots + y_{h-1})^d + \theta y_h^d\right) = \sum_{h=1}^{b} \lambda_h y_h^{t-d} \prod_{g \neq h} y_g^t,$$

where

$$\lambda_1 = \theta^{b-1}, \qquad \lambda_h = \theta^{b-h}(1+\theta)^{h-2} \neq 0.$$

Every surviving term leaves exactly one row block unsaturated. Appending variable columns and an all-ones lower block therefore gives

$$\operatorname{per} \begin{pmatrix} X & U \\ \mathbf{1} & 0 \end{pmatrix} = c \sum_{h=1}^{b} \lambda_h M_{t,s,d}(X_{R_h,[s]}), \qquad c \neq 0.$$

All cross-block matchings cancel algebraically.

The initial exploratory version used one hub block and two root-of-unity orbits for each other block. Its square-zero capacity discarded every pattern that omitted rows from multiple blocks. That was sufficient to discover the cancellation mechanism, but the cumulative-block construction uses exactly $r-d$ constant columns and explicitly tracks every surviving coefficient. The induction satisfies

$$\lambda_i^{(h+1)} = \theta \lambda_i^{(h)}, \qquad \lambda_{h+1}^{(h+1)} = \sum_{i=1}^{h} \lambda_i^{(h)}.$$

The latter sum equals $(1+\theta)^{h-1}$, and cannot vanish because $|\theta| = 2^d > 1$. Thus the final argument checks both requirements: mixed matchings cancel, and every intended matching block survives with a nonzero coefficient.

## 5.6 Parameter choice, the explicit bound, and the determinant check

We choose

$$d = \left\lfloor \frac{\log_2 n}{4} \right\rfloor, \quad t = 4d, \quad b = \left\lfloor \frac{n}{2t} \right\rfloor, \quad r = bt, \quad s = n - r + d, \quad m = rs.$$

Both $r$ and $s$ have order $n$. The incidence bound gives

$$\dim \operatorname{Crit}(P) \leq b\big(2^t - 1 + s(d-2)\big) < \frac{m}{2},$$

because each of the shared-parameter and column-parameter terms costs at most $m/4$. A slice of dimension

$$k = \left\lfloor \frac{m}{4} \right\rfloor \geq \frac{n^2}{48}$$

therefore avoids the nonzero critical cone. For $n \geq 2^{16}$, the resulting explicit bound is

$$\boxed{L \geq \frac{n^2}{144}\big(\log_2 \log_2 n - 3\big)}.$$

The determinant cannot be substituted silently. Its corresponding block matrix reduces to

$$\det \begin{pmatrix} X & U \\ V & 0 \end{pmatrix} = c \det(AXC),$$

where $AXC$ is only $d \times d$. Its critical locus is the rank-$\leq d-2$ variety, of codimension four. Thus the large critical-codimension witness is genuinely permanent-specific.

More generally, a homogeneous degree-greater-than-two affine specialization of the $n \times n$ determinant has critical codimension at most $n-1$, by the relationship between determinantal complexity and the hypersurface singular locus. Since its degree is at most $n$, the gradient method can then give at most

$$O(n \log n),$$

less than even the elementary quadratic input-dependence bound. This scope check is significant: the enormous Birkhoff-volume heuristic applies superficially to both matrix polynomials, whereas the actual critical-locus construction provably separates their behavior.

The lower bound follows by combining the permanent-specific block specialization with the critical-locus dimension estimate and the Baur–Strassen and Bézout degree bounds. Its strength comes from the interaction of these ingredients, not from any one of them separately.

Chapter 6

# Permanent Formulas: Matching Coefficients and the Division Barrier

## 6.1 Why neither multilinearity nor differentiation gave the right bound

We consider fan-in-two arithmetic formulas over $\mathbb{C}$ computing

$$\mathrm{per}_n(X) = \sum_{\sigma \in S_n} \prod_{i=1}^{n} x_{i,\sigma(i)}.$$

A formula is a tree: unlike a circuit, it cannot reuse an intermediate computation. The permanent is multilinear, but an unrestricted formula may create nonmultilinear intermediates and cancel them later, so syntactically multilinear lower bounds do not apply. Row-by-row multihomogenization needs all $2^b$ subsets of $b$ row groups; sharing those components changes a formula into a circuit, and output multidegree one does not remove the intermediate components. Baur–Strassen differentiation fails for the same reason: its efficiency comes from sharing [41].

The useful alternative was the algebraic Nechiporuk perspective associated with Kalorkoti [40]. For a selected variable block $Y$, expand

$$f(Y, Z) = \sum_{\alpha} c_\alpha(Z) Y^\alpha$$

and measure

$$d_Y(f) = \mathrm{trdeg}_{\mathbb{C}}\, \mathbb{C}(c_\alpha(Z) : \alpha).$$

The question becomes: how many independent outside parameters must enter through formula leaves labeled by variables in $Y$?

## 6.2 Pruning a formula charges coefficients to marked leaves

Mark every formula vertex whose subtree contains a $Y$-labeled leaf, and let $l_Y$ count those leaves with repetition. Along a path where only one child is marked, the other child computes $h(Z)$, so addition and multiplication transform the marked expression by

$$u \longmapsto A(Z)u + B(Z).$$

An arbitrarily long chain compresses to one affine map. Only a gate with two marked children creates a genuine branching point.

A marked tree with $l_Y$ leaves has at most $l_Y - 1$ such branching points. Charging four outside parameters at each branching point and two at the final root gives

$$d_Y(f) \le 4l_Y - 2 \le 4l_Y.$$

The complexities of entirely unmarked sibling subformulas never enter this count. Thus cancellations are harmless: they may change the resulting coefficient polynomials, but they cannot increase the transcendence degree beyond the number of parameters exposed by the marked skeleton.

Crucially, this is not a count of distinct coefficients. The polynomials $z$ and $z^2$ are different but have transcendence degree one, whereas $z_1$ and $z_2$ supply two independent parameters. We need algebraic independence rather than linear independence or a long list of superficially different complementary permanents. Passing to the field generated by all coefficients is what makes that distinction precise.

## 6.3 A short matching should control quadratically many coefficients

Take

$$\ell = \lceil \log_2 n \rceil, \qquad k = 2\ell, \qquad m = n - k,$$

and let $Y$ be a matching of $k$ matrix entries. After row and column permutations, take $Y = \{x_{ii} : i \in [k]\}$. For $S \subseteq [k]$, the coefficient

$$c_S = \left[\prod_{i \in S} x_{ii}\right] \mathrm{per}_n$$

is obtained by deleting the indicated rows and columns and setting all remaining marked diagonal entries to zero.

A matching is preferable to a full row: it has $2^k$ possible coefficients while approximately $n^2/k$ disjoint matching blocks fit into the matrix. The earlier choices explain why this balance matters: full rows provide too few blocks, tiny blocks provide too few complementary coefficients, and larger rectangular blocks cannot be packed densely. A logarithmic matching is the first shape with quadratically many possible coefficients and nearly quadratically many disjoint placements. The first useful specialization separated the outside indices into two halves and inserted one forced bridge. After differentiating in a cross entry, row–column balance forced that bridge to appear, and the Jacobian factored into independent left and right matching signals. Power-of-two weights made both signals Vandermonde.

That construction explained *why* a tensor-product Jacobian should exist, but splitting the outside indices discarded a constant fraction of the available rank. A stronger all-ones specialization retains all $m^2$ outside entries.

The bridge explains how this separation was first recognized. We split both the marked and outside indices into left and right halves, killed all cross edges, and retained one reverse bridge $b_0 \to a_0$. Differentiating in a forward edge leaves one side with an extra unmatched column and the other with an extra unmatched row. A surviving matching must use exactly one reverse edge and no additional forward edge, hence must use the unique bridge. The remaining permanent factors into independent left and right minors. The final all-ones witness retains that conceptual separation without physically dividing the outside variables.

## 6.4 The final Jacobian uses two binary encodings and all outside entries

The smallest illustration is

$$\begin{pmatrix} y_e & 0 & p_1 & p_2 & p_3 \\ 0 & y_f & 1 & 1 & 1 \\ 1 & q_1 & w_{11} & w_{12} & w_{13} \\ 1 & q_2 & w_{21} & w_{22} & w_{23} \\ 1 & q_3 & w_{31} & w_{32} & w_{33} \end{pmatrix}.$$

After differentiating the four $y_e, y_f$ coefficients in $w_{ab}$ and then setting every $w_{ab} = 1$, the resulting rows are

$$2, \qquad 2G_b, \qquad 2H_a, \qquad G_bH_a,$$

where

$$G_b = \sum_{j \neq b} p_j, \qquad H_a = \sum_{i \neq a} q_i.$$

On two outside rows and columns, the Jacobian is a diagonal scaling of a tensor product of two Vandermonde matrices. It is invertible when the $p$'s and $q$'s are distinct.

These four rows follow by direct matching counts. With both marked diagonals chosen, fixing one outside edge leaves two possible outside matchings. With exactly one marked diagonal chosen, the unmatched partner must either select an external column or receive an external row, producing $2G_b$ or $2H_a$. With no marked diagonal chosen, those choices are independent and contribute $G_bH_a$. The corresponding four-by-four determinant contains $(p_1 - p_2)^2(q_1 - q_2)^2$. Thus the tensor product is visible before any large binary-index bookkeeping appears.

For the general argument, split the marked indices into

$$E = \{e_0, \dots, e_{\ell-1}\}, \qquad F = \{f_0, \dots, f_{\ell-1}\}.$$

Set the unmarked internal block to zero and the outside $m \times m$ block $W$ to all ones. For independently distinct $p_1, \dots, p_m$ and $q_1, \dots, q_m$, choose

$$U_{e_u,b} = p_b^{2^u}, \quad U_{f_v,b} = 1, \qquad V_{a,e_u} = 1, \quad V_{a,f_v} = q_a^{2^v}.$$

Subsets $P(\alpha) \subseteq E$ and $Q(\beta) \subseteq F$ encode the binary digits of $0 \le \alpha, \beta < m$. Select the $m^2$ coefficient indices

$$S(\alpha, \beta) = [k] \setminus (P(\alpha) \cup Q(\beta)).$$

Differentiating in $w_{ab}$ fixes one external matching edge. The remaining internal rows inject into the external columns, and the internal columns independently receive external rows. Consequently

$$\left.\frac{\partial c_{S(\alpha,\beta)}}{\partial w_{ab}}\right|_{\xi} = L_{\alpha,\beta} g_{P(\alpha)}(p_b) h_{Q(\beta)}(q_a), \qquad L_{\alpha,\beta} \neq 0.$$

The avoidance polynomial $g_P$ satisfies

$$g_{\varnothing}(z) = 1, \qquad g_P(z) = H_P - \sum_{e_u \in P} z^{2^u} g_{P \setminus \{e_u\}}(z),$$

where $H_P$ is the unrestricted injection sum. Hence

$$\deg g_{P(\alpha)} = \alpha, \qquad [z^\alpha] g_{P(\alpha)} = (-1)^{|P(\alpha)|} |P(\alpha)|! \neq 0.$$

The same facts hold for $h_{Q(\beta)}$. Each evaluation matrix is therefore an invertible triangular change of basis times an $m$-point Vandermonde matrix. Their tensor product proves

$$d_Y(\mathrm{per}_n) \geq m^2.$$

The characteristic-zero Jacobian criterion validates independence before specialization [45].

The coefficient $L_{\alpha,\beta}$ is also meaningful. Once the remaining distinguished rows inject into outside columns and outside rows inject into distinguished columns, the unused all-ones block has

$$(m-1-|P(\alpha)|-|Q(\beta)|)!$$

matchings. Additional falling factorials count weight-one completions of the two injections. The condition $k \leq m-1$ ensures that all these factors exist and remain nonzero. We therefore need the numeric assumption $n \geq 32$, not merely the informal asymptotic statement that $k$ is logarithmic.

## 6.5 Packing entry-disjoint matchings gives the division-free bound

The $n$ cyclic perfect matchings partition the $n^2$ matrix entries. Divide each into blocks of $k$ entries, producing

$$\nu = n\left\lfloor \frac{n}{k} \right\rfloor \geq \frac{n^2}{2k}$$

pairwise entry-disjoint matchings. They may share rows or columns; only disjointness of the actual variables matters.

Every selected block satisfies

$$m^2 \leq d_Y(\mathrm{per}_n) \leq 4l_Y, \qquad l_Y \geq \frac{m^2}{4}.$$

A variable-labeled leaf can belong to at most one selected block, so the charges add within the *same* formula:

$$L_{\mathrm{var}}(\Phi) \geq \frac{\nu m^2}{4} \geq \frac{n^4}{128\log_2 n}, \qquad n \geq 32.$$

Since a binary formula with $L$ leaves has $L-1$ internal gates, the resulting bound also gives

$$G(\Phi) \geq \frac{n^4}{256\log_2 n}.$$

The resulting $n^4/\log n$ rate exceeds $n^{7/2}$, so it certainly supplies a fixed supercubic exponent.

Different matching blocks need not share one numerical witness specialization. Algebraic independence is a separate identity for each coefficient family of the same permanent, whereas leaf charging concerns the original fixed formula. The selected blocks have disjoint matrix entries but may meet the same row or column. Requiring disjoint row or column sets would shrink the packing and lose the desired power of $n$.

## 6.6 Division is handled by projective rather than affine wrappers

The initial objection to division was real but incomplete:

$$u \longmapsto h(Z)/u$$

is not an affine transformation. Instead, put $R = \mathbb{C}(Z)$ and observes that every unary marked chain acts by

$$u \longmapsto \frac{au+b}{cu+d}, \qquad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in R^{2\times 2}.$$

Chains compose by multiplying their matrices. The matrix need not be invertible: multiplication by an unmarked zero may produce a singular, nonzero matrix. Validity of every actual denominator does ensure that the composite matrix is not identically zero. Normalizing one nonzero entry leaves at most three field parameters.

It would be incorrect to require the representing matrix to be invertible. Multiplication by an unmarked zero is a legal formula operation even though its matrix is singular. Validity only says that the denominator of the actual rational value is nonzero in $R(Y)$. Under composition that denominator remains a product of prior valid denominators, so the composite matrix is nonzero and still permits normalization. Allowing singular wrappers is indispensable for fully unrestricted valid-division formulas.

Two marked children therefore cost at most six parameters at a branching point. An induction supplies a field $K \subseteq R$ generated by at most $6l_Y - 3$ elements, with the formula output in $K(Y)$. Because the output is also a polynomial in $Y$, the crucial intersection identity

$$K(Y) \cap R[Y] = K[Y]$$

shows that all its coefficient polynomials already belong to $K$. Thus

$$d_Y(\mathrm{per}_n) \le 6l_Y - 3 < 6l_Y.$$

The same $m^2$ Jacobian and disjoint matching blocks then give

$$\boxed{L_{\mathrm{var}}(\Phi) \ge \frac{n^4}{192\log_2 n}, \qquad G(\Phi) \ge \frac{n^4}{384\log_2 n}}$$

even for formulas with valid divisions. The Jacobian witness is evaluated only on polynomial coefficients, so formula denominators need not remain nonzero at the witness specialization.

We can verify the field-intersection step without pretending that a projection preserves arbitrary products. If $f \in K(Y) \cap R[Y]$, choose $P, Q \in K[Y]$ with $Q \neq 0$ and $Qf = P$. Extend a $K$-basis to obtain a $K$-linear projection $\pi : R \to K$, and apply it coefficientwise. Because $Q$ has coefficients in $K$,

$$Q\pi(f) = \pi(Qf) = P = Qf.$$

The domain property gives $f = \pi(f) \in K[Y]$. This modest algebraic observation is what allows the marked-skeleton argument to survive rational intermediate cancellations.

## 6.7 Why the determinant comparison matters

For a determinant with $k < n$ marked diagonal entries, the Schur complement gives

$$\det X = \det(W)\det\big(\mathrm{diag}(Y) + D_0 - UW^{-1}V\big).$$

All $Y$-coefficients are therefore generated by $\det W$ and the $k^2$ Schur-complement entries:

$$d_Y(\det X) \le k^2 + 1.$$

A logarithmic matching cannot contribute $\Theta(n^2)$ independent determinant coefficients. More generally, entry-disjoint determinant blocks have an $O(n^3)$ total coefficient-transcendence ceiling. The improvement is therefore specific to the permanent, not a hidden claim about determinant formulas.

This determinant ceiling is meaningful rather than an artifact of a poor specialization. For a full diagonal matching, all coefficients are principal minors of a zero-diagonal matrix. Diagonal conjugation preserves those minors and removes $n-1$ parameters, leaving $(n-1)^2$ independent generic degrees of freedom. The two- and three-element principal minors recover the remaining normalized entries up to algebraic ambiguity. Packing the $n$ cyclic full matchings therefore reproduces the classical cubic formula bound, including valid division, up to constants. What fails for the determinant is specifically the permanent's quadratically rich *logarithmic-size* matching block.

The permanent circuit theorem must also be distinguished from the formula theorem. Its circuit counterpart proceeds by simultaneous differentiation and critical-locus degree; that argument explicitly exploits shared subcomputations. Here every charging step instead relies on a marked binary tree with only $l_Y - 1$ branching vertices. Neither proof transfers to the other computational model by simply renaming gates.

Both the division-free and valid-division bounds follow from the same coefficient-transcendence argument, with the projection lemma ensuring that rational intermediate cancellations do not weaken the conclusion.

Chapter 7

# Quantum Parallel Repetition: From Holonomy to Resolvent Purification

## 7.1 Why an obvious reduction could not possibly work

Consider a finite entangled two-player game

$$G = (X, Y, A, B, \mu, V), \qquad \varepsilon = 1 - \omega^*(G) > 0,$$

and ask whether the unchanged repeated game satisfies an exponential upper bound. Playing independent strategies gives

$$\omega^*(G^{\otimes(m+n)}) \geq \omega^*(G^{\otimes m})\omega^*(G^{\otimes n}),$$

which points in exactly the wrong direction. Subadditivity of $-\log \omega^*(G^{\otimes n})$ permits, for instance, growth like $\sqrt{n}$; it does not force a positive linear rate.

Every coordinate marginal wins with probability at most $\omega^*(G)$, so a repeated strategy has at least $\varepsilon n$ expected failures. But expectation does not control the probability of *no* failures: the losses may be correlated across all coordinates. Raz and Holenstein resolve this issue classically [46, 47]; Yuen's established general quantum theorem provides polynomial rather than exponential decay [48].

The available exponential quantum results impose additional structure. Perturbing $\mu$ to have full support invokes [49], but if $\mu_\eta = (1-\eta)\mu + \eta\nu$, comparison yields only

$$\omega^*(G^{\otimes n}) \leq \exp\left[-\left(c(\eta) - \log\frac{1}{1-\eta}\right) n\right].$$

The known rate $c(\eta)$ deteriorates as the artificial support mass vanishes. Projection-game tensorization needs unique witnesses, and anchoring changes the game itself [50, 51]. None of these substitutions proves an exponential bound for the original arbitrary game.

Another sustained approach was to ask whether the question-support graph itself supplies the missing rigidity. If its components are rectangles, both players can recognize their component and reduce to smaller games. On a connected non-rectangular component, the classical conditional-expectation operator has a spectral gap. Tensor powers damp high-degree scalar functions of the questions, but the desired obstruction is an operator-valued rotation inside an unconstrained Schmidt space. Conditional states can have identical question statistics and different purification frames, so the classical spectral gap did not control the relevant quantum information.

Oracularization suggests another reduction: let one player see both questions, then ask the other to check one reported answer. But a prover who sees both questions can implement measurements unavailable to either original player. A repeated strategy also does not provide two simultaneously consistent copies of incompatible answer measurements. Thus the obstacle is not merely a quantitative loss; changing the question structure changes the physical strategy class.

The same implementability defect appears in sequential entropy accumulation, symmetric de Finetti decompositions, and detectability-lemma analogues. Each controls a classical marginal, an abstract extension, or formally local reject operators, but none constructs the tensor-product strategy the players must actually execute. These repeated failures shift the search away from another ordinary information bound and toward a purification gauge whose entropy control survives conditioning without a spectral cutoff.

## 7.2 Conditioning exposed both a probability loss and a quantum phase

We next condition on winning a small coordinate set $C$. Writing

$$p = \Pr(W_C), \qquad m = n - |C|,$$

the classical question-information budget is

$$\sum_{i \notin C} D(\mathcal{L}(X_i, Y_i \mid W_C) \,\|\, \mu) \leq \log(1/p).$$

Hence a typical remaining question pair has total-variation error $O(\sqrt{\log(1/p)/m})$.

The tempting quantum analogue is invalid. An estimate before conditioning controls the successful branch only with its probability attached:

$$p\, D(\rho_{W_C} \| \sigma_{W_C}) \lesssim D(\rho\|\sigma) + h(p).$$

Dividing by $p$ ruins the argument precisely when success is rare. Dimension-dependent de Finetti estimates are no substitute: entangled strategies may use arbitrarily large local dimension, and a commuting- or no-signaling relaxation may have value one even when $\omega^*(G) < 1$.

There is also a genuinely geometric obstruction. Represent a conditional edge state as

$$|\Phi_{r,x,y}\rangle = \frac{(S_{r,x} \otimes T_{r,y})|\psi\rangle}{\|(S_{r,x} \otimes T_{r,y})|\psi\rangle\|}.$$

Uhlmann's theorem can compare neighboring purifications, but their local unitaries need not factor simultaneously as

$$|\Phi_{r,x,y}\rangle \approx (U_{r,x} \otimes V_{r,y})|\Phi_r\rangle.$$

On a four-cycle, such a factorization would require

$$M_{x_0y_0} M_{x_1y_0}^{-1} M_{x_1y_1} M_{x_0y_1}^{-1} = I.$$

Equal reduced density matrices do not enforce that identity. Thus classical question information alone cannot control purification holonomy.

The holonomy obstruction also suggests a possible counterexample. A perfect commuting-operator strategy might be compressed and teleported, with one favorable teleportation outcome selected afterward. The favorable outcome has probability proportional to the inverse square of the teleported dimension. However, one compressed state generally cannot implement all repeated rounds because the same player's measurements need not commute; using one independent copy per round restores an exponential dimension cost. This exhibited a real obstruction to naive proofs without producing a finite game contradicting exponential repetition.

## 7.3 The exploratory breakthrough was to preserve the Born weight

The first promising idea was to change purification gauge without changing probabilities. For a positive effect $F$,

$$F^{1/2+it} * F^{1/2+it} = F.$$

Averaging the imaginary exponent against

$$\ell(t) = \frac{\pi}{2\cosh^2(\pi t)}$$

cancels the Stieltjes prefactor for the complex square root. The resulting Mellin–Plancherel calculation suggested

$$\mathbb{E}_t (A^{1/2-it} - D^{1/2-it})(A^{1/2+it} - D^{1/2+it}) \preceq f(A) - f(D) - df_D(A - D), \quad f(u) = u \log u.$$

When $D = \mathbb{E}A$, averaging cancels the derivative, leaving an operator-entropy Jensen gap.

The crucial detail is noncommutativity. With $R_A(s) = (A+sI)^{-1}$, the squared resolvent integrand has the ordered product

$$R_D(s)(A - D)R_A(s)^2(A - D)R_D(s).$$

The inequality $sR_A(s)^2 \preceq R_A(s)$ converts it into the ordered Bregman integrand. Replacing these products by scalar commuting expressions would reproduce the invalid square-root argument.

The phase construction identifies the decisive principle: preserve the Born weight while spreading the troublesome eigenvalue scales. A later pivot implements that same principle through a simpler finite purification.

The reason ordinary square roots fail deserves emphasis. In an eigenbasis of a mean effect $D$, the derivative of the square root has coefficients

$$\left(d\sqrt{\cdot}_D(H)\right)_{ij} = \frac{H_{ij}}{\sqrt{\lambda_i} + \sqrt{\lambda_j}}.$$

Attempting to compare this directly to an entropy Hessian introduces integrals of order $\log(1/\lambda_{\min}(D))$. A repeated strategy can have arbitrarily small Schmidt coefficients, so this is not a harmless dimension-dependent constant. Imaginary-power averaging was the first way to avoid coherently accumulating all eigenvalue scales; resolvent purification makes the same discovery more economical.

## 7.4 From oscillatory gauges to finite resolvent purification

The oscillatory gauge suggests replacing one coherent square root by a scale-resolved isometry. For a positive contraction $F$, define

$$[\Gamma(F)v](u) = F(F + uI)^{-1}v, \qquad u > 0.$$

Since

$$\int_0^\infty \left(\frac{\sigma}{\sigma + u}\right)^2 du = \sigma,$$

spectral calculus gives the exact probability identity

$$\Gamma(F)^*\Gamma(F) = F.$$

For any fixed finite strategy, only finitely many distinct positive eigenvalues occur. The span of the corresponding functions $u \mapsto \sigma/(\sigma+u)$, with a zero-eigenvalue flag, is finite-dimensional. Thus the integral does not introduce an infinite-dimensional strategy.

Set $\bar{F} = \mathbb{E}F$, $J = F - \bar{F}$, and $R_F(u) = (F + uI)^{-1}$. The resolvent identity gives

$$F(F + uI)^{-1} - \bar{F}(\bar{F} + uI)^{-1} = uR_F(u)JR_{\bar{F}}(u).$$

Using $uR_F(u)^2 \preceq R_F(u)$, averaging, and integrating proves

$$\mathbb{E}[(\Gamma(F) - \Gamma(\bar{F}))^*(\Gamma(F) - \Gamma(\bar{F}))] \preceq H_1(\bar{F}) - \mathbb{E}H_1(F), \qquad H_1(v) = -v \log v.$$

This is a singular-safe version of the operator Bregman principle [52, 53]. Its key advantage over ordinary square roots is dimension-independent control with no smallest-eigenvalue loss.

## 7.5 A live-coordinate martingale makes the rare branch pay for itself

We next need a reveal process that keeps the opposite player's effect fixed. The key organizational pivot is a size-biased oriented partition, a uniformly random order, and a uniform live cut. In one orientation, Alice's successive conditional effects form a martingale

$$F_0, F_1, \ldots, F_N, \qquad \mathbb{E}[F_{j+1} \mid F_0, \ldots, F_j] = F_j,$$

while Bob's effect $K$ remains unchanged. The resolvent inequality telescopes to

$$\mathbb{E}_j \left\| \big(\Gamma(F_{j+1}) - \Gamma(F_j)\big) \otimes \Gamma(K)|\psi\rangle \right\|^2 \le \frac{H_1(p_0)}{N}, \qquad p_0 = \langle\psi|F_0 \otimes K|\psi\rangle.$$

The reverse orientation supplies the analogous Bob-side bound.

This estimate is deliberately *probability weighted*. On a successful branch the posterior includes the factor $\|\phi_{r,x,y}\|^2/p$. Meanwhile

$$\left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\|^2 \le \frac{4\|u - v\|^2}{\|u\|^2}.$$

The branch weight cancels the dangerous denominator before averaging. Thus the accepted-event mass contributes $\log(1/p)$, not $1/p$.

There are two locally describable parents:

$$\bar{H}_{r,y} = \sum_{x'} \mu(x' \mid y) H_{r,x'}, \qquad \bar{K}_{r,x} = \sum_{y'} \mu(y' \mid x) K_{r,y'}.$$

Bob can describe the first from $y$; Alice can describe the second from $x$. Neither player is asked to apply the other player's operator. This distinction is what bypasses the earlier holonomy obstruction.

## 7.6 Classical and quantum sampling close the quantitative argument

Let

$$\eta = \frac{\log(1/p) + |C| \log(|A||B|)}{n - |C|}.$$

The resulting state-alignment estimate is

$$\mathbb{E}\|\Psi_{r,x,y} - \Psi^A_{r,x}\|^2 \le 8\eta, \qquad \mathbb{E}\|\Psi_{r,x,y} - \Psi^B_{r,y}\|^2 \le 8\eta.$$

Classical correlated sampling synchronizes the locally generated histories with mismatch $O(\sqrt{\eta})$. Quantum correlated sampling then prepares the desired state using a finite embezzlement catalyst [50]. For every accuracy parameter $\alpha > 0$, this produces an actual finite-dimensional strategy for the original game satisfying

$$\mathrm{win}(S_\alpha) \ge q - B_{\mathrm{qs}}\eta^{1/12} - 2K_{\mathrm{qs}}\alpha^{1/12},$$

where $q$ is the conditioned coordinate-winning probability.

We must distinguish the two samplers carefully. Classical correlated sampling uses Alice's history law conditioned on $x$ and Bob's history law conditioned on $y$. If both joint laws are within total variation $\kappa$ of the same ideal law, one finite shared random seed gives

$$\Pr(r_A \ne r_B) \le 4\kappa, \qquad \kappa \le \sqrt{\frac{3}{2}\eta}.$$

Quantum correlated sampling applies only on matching histories. Its embezzlement state is prepared before either question arrives, and the local unitaries depend only on the corresponding locally described parent states. The catalyst dimension may depend on the chosen finite strategy and accuracy; no argument requires or permits a uniform dimension bound.

No assumption is made that every question pair has positive probability. Conditionals are used only on positive-$\mu$ edges; defaults elsewhere do not change the experiment. The final estimate consequently contains no inverse minimum-support probability, precisely the loss that defeated full-support perturbation.

Set

$$c_{\mathrm{qs}} = \frac{1}{8(4B_{\mathrm{qs}})^{12}}, \qquad \alpha = \left(\frac{\varepsilon}{16K_{\mathrm{qs}}}\right)^{12}.$$

Greedy conditioning shows that any repeated strategy exceeding

$$\boxed{\omega^*(G^{\otimes n}) \le \exp\left(-c_{\mathrm{qs}}\frac{\varepsilon^{13}}{\varepsilon + \log(|A||B|)}n\right)}$$

would yield a one-game strategy winning more than $1-\varepsilon = \omega^*(G)$, a contradiction. The prefactor is one, the constant is universal, and no minimum question probability, support connectivity, question-alphabet size, or strategy-dimension bound is required. The exponent $13$ reflects the twelfth-root correlated-sampling loss.

The prefactor-one statement is part of the theorem, not cosmetic bookkeeping: it applies separately to every finite $n$, without discarding a small-$n$ regime or inserting a game-dependent constant. The greedy argument starts from whichever particular $n$ violates that bound.

The greedy conditioning must be performed for an actual finite strategy rather than an assumed optimizer. If the supremum defining the repeated value exceeds the exponential threshold, some finite-dimensional strategy already exceeds that same threshold. Greedy selection then produces a proper tested set $C$, with

$$|C| < \frac{\gamma n}{\varepsilon/4}, \qquad \Pr(W_C) \ge \Pr(W_{[n]}), \qquad q \ge 1 - \frac{\varepsilon}{4}.$$

The rate choice makes $|C| < n/2$ and

$$\eta \le \left(\frac{\varepsilon}{4B_{\mathrm{qs}}}\right)^{12}.$$

Substitution into the rounding inequality yields a legal one-game strategy winning more than $1 - 5\varepsilon/8$, which contradicts the definition of $\omega^*(G)$. Thus we do not require the repeated supremum to be attained, a uniformly bounded catalyst, or a positive lower bound on support probabilities.

The historical distinction matters. Anchored-game amplification can prove excellent decay for a deliberately modified game, while projection-game theorems exploit a uniqueness property absent here. The advance concerns specifically the unchanged game, arbitrary finite question support, finite-dimensional tensor-product strategies, and an explicit answer-alphabet-dependent rate. Confusing these settings would make the result sound like an immediate corollary of established work when its difficult new ingredient is precisely the postselection-stable state alignment.

The resulting exponential theorem applies to the unchanged game itself; postselection-stable state alignment is the mechanism that closes the argument.

Chapter 8

# Fixed-Polynomial GapCVP: From Signed Histograms to Binary Reconstruction

## 8.1 The obstacle was not merely a weak approximation gap

We seek a deterministic reduction distinguishing

$$\mathrm{dist}_2(t,\Lambda) \le r \quad \text{from} \quad \mathrm{dist}_2(t,\Lambda) > n^c r$$

for an absolute constant $c > 0$, with $n$ the lattice rank. The established almost-polynomial hardness factor $n^{\Omega(1/\log\log n)}$ does not provide a fixed exponent [54]. Conversely, the known lattice protocols and the NP∩coNP result near the square-root scale explain why arbitrarily large exponents are not expected [55, 56].

A binary coding gap $G$ becomes only $\sqrt{G}$ after translation to Euclidean distance:

$$\|u - v\|_2^2 = d_{\mathrm{H}}(u, v).$$

Existing fixed-gap gadgets and conjectural projection-game inputs either provide too little soundness or conceal the sought hardness assumption. Tensoring a constant-gap instance multiplies its dimension together with its gap, so the necessary logarithmic number of levels destroys polynomial reduction size. Direct quadratic SAT penalties permit intermediate integral coefficients to neutralize clause violations, while locally consistent signed pseudodistributions can telescope around the constraint graph. The common obstacle is cancellation: a short global object must be forced to encode at least one consistent satisfying assignment.

## 8.2 The first long detour: signed multivariate moment histograms

The initial candidate encoded an assignment by an interpolation polynomial $g$ over a large prime field and introduced signed histograms

$$S_X(a), \qquad X = (X_1, \dots, X_4), \quad a = (a_1, \dots, a_4).$$

For a genuine assignment,

$$S_X(a) = \mathbf{1}\{a = (g(X_1), \dots, g(X_4))\}.$$

Its moments

$$P_e(X) = \sum_a S_X(a) \prod_{i=1}^{4} a_i^{e_i}$$

have degree at most $de_i$ in coordinate $X_i$. Degree constraints, diagonal consistency, and forbidden clause patterns are linear in the histogram entries, so they can be embedded into integer congruence rows.

A heavy block $M = q^{20}$ forces those congruences to hold exactly for any sufficiently short lattice vector. Its squared norm bounds $\sum_{X,a} S_X(a)^2$, so Markov's inequality makes most fibers sparse. Hankel determinants reconstruct algebraic branches from their moments; quantitative finite-field point estimates distinguish genuine assignment graphs from moving algebraic covers [57].

The prime-field route also admits a complete construction. Four position coordinates make Frobenius-stable factor counting effective; valuations force the resulting rational branches to be low-degree polynomials, and Vandermonde inversion in the spare coordinate synchronizes their clause assignments. A different characteristic-two construction streamlines this geometry by replacing signed histograms with *binary* evaluation tables and cancellation with parity.

To see why either route needs more than bare moment reconstruction, consider a fiber represented schematically by

$$P_j(X) = \sum_{r=1}^{h} w_r(X)\alpha_r(X)^j.$$

A vanishing Hankel minor supplies an algebraic recurrence, but it does not say that the roots $\alpha_r$ are polynomial assignment graphs or that their weights remain constant. The prime-field proof closes exactly these gaps: explicit height bounds and point counts eliminate nonlinear covers, infinity valuations control moving branches, and the fourth coordinate separates all surviving assignments. In the binary version, parity, separability, and anchor valuations provide the corresponding synchronization directly.

## 8.3 Characteristic-two tables encode assignments and clauses

Let

$$\mathbb{K} = \mathbb{F}_{2^e}, \qquad a_1, \dots, a_m \in \mathbb{K}, \qquad P = \mathbb{K} \setminus \{a_1, \dots, a_m\}.$$

Characteristic two is crucial because addition of indicator tables then records precisely the parity cancellations needed by the clause equations. A Boolean assignment $\sigma$ has a unique interpolation polynomial

$$Q_\sigma(a_i) = \sigma_i, \qquad \deg Q_\sigma < m.$$

This is the Reed–Solomon viewpoint [58].

For $p \in P$, $w \in \mathbb{K}$, and a table type $\tau$, introduce a binary indicator $x_{\tau,p,w}$. Type $0$ is global; each remaining type is a clause $C$ together with a satisfying local Boolean tuple $\beta$. The intended global and selected clause tables each have the single value $w = Q_\sigma(p)$.

The central constraints are

$$\sum_w x_{0,p,w} = 1, \qquad x_{0,p,w} = \sum_{\beta \text{ satisfies } C} x_{(C,\beta),p,w} \quad \text{in } \mathbb{F}_2.$$

The first makes each global fiber odd; the second refines it into clause fibers by parity, without assuming disjointness or positivity.

For each table impose ordinary moment conditions

$$\mu_{\tau,j}(p) = \sum_w x_{\tau,p,w} w^j, \qquad \deg_X \mu_{\tau,j} \le mj.$$

For $\tau = (C, \beta)$ and $i \in C$, impose also

$$\eta_{\tau,i,j}(p) = \sum_w x_{\tau,p,w} \left(\frac{w - \beta_i}{p - a_i}\right)^j, \qquad \deg_X \eta_{\tau,i,j} \le (m-1)j.$$

The denominator never vanishes on $P$. For an honest assignment,

$$\frac{Q_\sigma(X) - \beta_i}{X - a_i} \in \mathbb{K}[X],$$

explaining the one-degree improvement. Every unknown is binary; field powers and Reed–Solomon parity checks are fixed linear coefficients. Expanding in an $\mathbb{F}_2$-basis yields one explicit binary affine system

$$Hx = b.$$

Importantly, these apparently nonlinear moment conditions remain genuinely linear in the reduction. The unknown is only $x_{\tau,p,w}$; both $w^j$ and $((w-\beta_i)/(p-a_i))^j$ are already fixed field elements. Membership of the resulting evaluation vector in a Reed–Solomon code is another linear parity-check condition. Neither an unknown assignment polynomial nor an unknown algebraic root is introduced as a nonlinear variable.

A satisfying assignment activates one global table and one table per clause, giving binary weight

$$R = (\ell + 1)|P|,$$

where $\ell$ is the number of clauses.

## 8.4 Hankel reconstruction survives characteristic two

Suppose a solution has small Hamming weight. For each type define

$$S_\tau(p) = \{w : x_{\tau,p,w} = 1\}.$$

Markov's inequality discards the small set of points with $|S_\tau(p)| > K$. Crucially, the good evaluation set may depend on $\tau$; requiring one common set would waste the parameter budget.

The ordinary moments of each remaining family determine a Hankel matrix

$$H_h(X) = (\mu_{\tau,i+j}(X))_{0 \le i,j < h}, \qquad \Delta_h(X) = \det H_h(X).$$

At a fiber with $h$ distinct selected values, this determinant is the Vandermonde square

$$\Delta_h(p) = \prod_{r<s} (w_s - w_r)^2 \neq 0.$$

Consequently, the Hankel recurrence reconstructs a monic separable polynomial

$$G_\tau(Y) \in \mathbb{K}(X)[Y]$$

whose distinct roots, in one common finite separable splitting field, form $\mathcal{R}_\tau$ and satisfy

$$\mu_{\tau,j}(X) = \sum_{\alpha \in \mathcal{R}_\tau} \alpha^j.$$

Its discriminant has controlled degree

$$\deg \Delta_h \le mh(h-1).$$

The proof does not divide by Newton coefficients that vanish in characteristic two: separability comes directly from the distinct fiber values and the nonzero Vandermonde determinant.

Reconstruction is needed throughout the entire moment budget, not merely on its initial Hankel window. Solving

$$\sum_{r=0}^{h-1} \mu_{j+r}(X) c_r = -\mu_{j+h}(X) \qquad (0 \le j < h)$$

defines the coefficients of $G$ over $\mathbb{K}(X)$. At every maximum-size nonsingular fiber, the resulting polynomial specializes to $\prod_{w \in S_\tau(p)}(Y-w)$. Bounding the discriminant degree leaves sufficiently many such fibers to extend the root-power identities across all prescribed moments. This is why the argument carefully tracks evaluation-set slack instead of assuming that an unlabeled pointwise fiber already comes with globally consistent labels.

The global zeroth moment gives

$$|\mathcal{R}_0| \equiv 1 \pmod 2,$$

so at least one global root exists. This turns the characteristic from a potential obstruction into the mechanism selecting a nonempty root set.

A further issue is preventing an exceptional-fiber union bound from silently destroying the reduction. For each table type separately,

$$|P \setminus P_\tau| \le \frac{\mathrm{wt}(x)}{K} < \frac{q}{20}.$$

We choose $T = N^{30}$, $K = N^4$, and $q \asymp N^{200}$, leaving enough points to satisfy both the Hankel-degree bound and the higher moment identities. Since the retained sets $P_\tau$ may differ, the argument does not multiply this loss by the number of clauses. That small organizational choice is essential for keeping the field size and output dimension polynomial.

## 8.5 Shifted moments make one root satisfy every clause

Fix a clause type $(C, \beta)$ and extend the valuation at $X = a_i$ to the common splitting field. The shifted moments force, for every $\alpha \in \mathcal{R}_{(C,\beta)}$,

$$v_i(\alpha - \beta_i) \ge 1, \qquad i \in C.$$

Conceptually, if $(\alpha - \beta_i)/(X - a_i)$ had negative valuation, its high powers would eventually dominate all cancellations. A controlled inverse Vandermonde matrix and the available moment budget make that valuation argument quantitative.

It would be invalid to claim informally that an algebraic root simply evaluates to $\beta_i$. Such a root can live in a separable extension, and the anchor valuation may extend in several inequivalent ways. We therefore put all table roots into one common splitting field and use the same extended valuation for every clause involving $i$. If

$$y = \frac{\alpha - \beta_i}{X - a_i}$$

had negative valuation, integrality and the root equation would quantize that value in steps of at least $1/h$. The inverse root-Vandermonde matrix has bounded valuation, whereas a sufficiently long block of shifted moments has nonnegative valuation. Taking a high enough power of $y$ contradicts the inverse bound. This turns each recovered clause bit into a property of the same global algebraic root.

For each clause, the binary refinement identity implies

$$\sum_{\alpha \in \mathcal{R}_0} \alpha^j + \sum_{\beta \text{ satisfies } C} \sum_{\alpha \in \mathcal{R}_{(C,\beta)}} \alpha^j = 0.$$

If a nonempty set of roots occurred an odd number of times, its initial power sums would give an invertible Vandermonde matrix applied to the all-ones vector. Therefore every $\alpha \in \mathcal{R}_0$ lies in at least one satisfying subtype root set for each clause.

Choose one such global $\alpha$. If two selected clause tuples disagreed about variable $i$, their valuations would give

$$v_i(\alpha - \beta_i) \geq 1, \qquad v_i(\alpha - \beta_i') \geq 1, \qquad \beta_i - \beta_i' = 1.$$

The ultrametric inequality would imply $v_i(1) \geq 1$, impossible. Hence all selected local tuples agree and form one satisfying global assignment. The decisive synthesis is *separable reconstruction, parity matching, and valuation compatibility*, not the original four-coordinate signed synchronization.

## 8.6 A parity-lift lattice gives the exact exponent

Let $M$ be the number of binary coordinates and suppose $Hx = b$ is consistent. Choose a binary solution $u$, define

$$\mathcal{C} = \ker_{\mathbb{F}_2} H, \qquad \Lambda_H = \{z \in \mathbb{Z}^M : H(z \bmod 2) = 0\},$$

and use a systematic generator of $\mathcal{C}$ to obtain the explicit square basis

$$B_H = \begin{pmatrix} I_{k_0} & 0 \\ \widetilde{P}_{\text{sys}} & 2I_{M-k_0} \end{pmatrix}.$$

The standard nearest-codeword relationship [59] becomes the exact identity

$$\min_{z \in \Lambda_H} \|u - z\|_p^p = \min\{\text{wt}(x) : Hx = b\}, \qquad p \geq 1.$$

An odd coordinate costs at least one; conversely, every binary solution lifts to a lattice-coset representative attaining its weight. Thus the square-root loss is precisely quantified, not hidden.

With $N$ denoting original input size, the eventual parameter choice is

$$q = 2^e \asymp N^{200}, \qquad K = N^4, \qquad T = N^{30}, \qquad M \leq 40N^{401}.$$

Its soundness statement is

$$\text{wt}(x) \leq 4M^{1/200}R \quad \Longrightarrow \quad \varphi \text{ is satisfiable.}$$

Taking $r = \lceil \sqrt{R} \rceil$, so $r^2 \leq 4R$, gives

$$\boxed{\text{GapCVP}^{(2)}_{M^{1/400}} \text{ is NP-hard}}$$

under the deterministic polynomial-time reduction. An integer radius is used rather than writing $\sqrt{R}$ into the promise instance: this guarantees an exact polynomial-size rational encoding, and the factor four in the soundness inequality absorbs its rounding. For an unsatisfiable formula the required NO condition is strict, not merely a weak lower bound on distance. The same construction gives factor $M^{1/200}$ for binary nearest codeword and syndrome decoding, and $M^{1/(200p)}$ for each fixed rational $\ell_p$ norm. Inconsistent systems map to a fixed one-dimensional strict NO instance.

The two dimension parameters must remain separate. The original formula has size $N$, whereas $M$ is the output coding dimension and exact lattice rank. The field extension degree is $e = O(\log N)$, so even a deterministic search for an irreducible binary polynomial of degree $e$ is polynomial in $N$. There are at most $1 + 8\ell$ table types, and therefore

$$M \leq (1 + 8\ell)|P|q \leq 40N^{401}.$$

The dense affine system and its square integral basis use at most $O(N^{802})$ encoded bits. These exponents are enormous, but they are fixed; the claim is polynomial-time computability, not practical efficiency.

The same exact polynomial-time encoding establishes all four hardness statements at their respective approximation factors.

Chapter 9

# The Sharp Ehrhart Inequality

## 9.1 The extremal simplex and the missing factorial

We begin with an apparently elementary question. Suppose that $K \subset \mathbb{R}^n$ is a convex body satisfying

$$b(K) = 0, \qquad \operatorname{int}(K) \cap \mathbb{Z}^n = \{0\}.$$

How large can its volume be? Ehrhart's conjectured answer is

$$\operatorname{vol}(K) \le C_n, \qquad C_n = \frac{(n+1)^n}{n!},$$

with equality at the centered lattice simplex

$$T_n = (n+1)\Delta_n - \mathbf{1} = \left\{ x_i \ge -1, \quad \sum_{i=1}^{n} x_i \le 1 \right\}.$$

Ehrhart's original simplex result and the equality analysis of Nill and Paffenholz make it clear that this is not merely a question of estimating a volume: one must explain why the lattice, centroid, and simplex recognize the same sharp constant [60, 61]. The sharp volume inequality is established in every dimension; classification of all equality cases remains a separate question.

The first warning was the factorial. The centroid condition implies $-K \subset nK$, hence

$$K - K \subset (n+1)K.$$

Combining this with symmetric-body arguments and Brunn–Minkowski gives at best $\operatorname{vol}(K) \le (n+1)^n$. The missing $n!$ is not a minor loss that a sharper estimate should automatically restore. It is the volume of a simplex, and its disappearance signals that symmetrization has forgotten how the $n$ independent directions fit together.

Halfspace and polar reductions briefly promised to recover the factorial, but required independent integral polar directions or large inscribed simplexes that lattice-freeness does not guarantee. We therefore need a mechanism respecting the actual extremizer.

## 9.2 A long and useful failed route: harmonic symmetrization

The first construction that really respected the extremizer was the *harmonic symmetral*. Write

$$f = p_K = h_Q, \qquad M = \left( \frac{Q - Q}{2} \right)^{\circ} = \left\{ x : \frac{f(x) + f(-x)}{2} \le 1 \right\}.$$

At least one of $f(x)$ and $f(-x)$ is at most one whenever $x \in M$, so

$$M \subset K \cup (-K).$$

Consequently $M$ is centrally symmetric and contains no nonzero interior lattice point; Minkowski therefore gives $\operatorname{vol}(M) \le 2^n$.

What made this construction compelling is that it does exactly the right thing for $T_n$. The ordinary intersection $T_n \cap (-T_n)$ loses too much volume, already in dimension two, whereas its harmonic symmetral has volume exactly $2^n$. Thus the whole conjecture would follow from the purely convex-geometric inequality

$$\operatorname{vol}\big((Q-Q)^{\circ}\big) \ge \frac{n!}{(n+1)^n} \operatorname{vol}(Q^{\circ}) \tag{9.1}$$

whenever $0$ is the Santaló point of $Q$.

For $\mathcal{A} = \int e^{-f}$ and $\mathcal{B} = \int e^{-f(x)-f(-x)}$, radial integration gives

$$\frac{\mathcal{B}}{\mathcal{A}} = \mathbb{E}[(1 + f(-U))^{-n}],$$

where $U$ has boundary cone measure. The centroid bound $f(-U) \le n$ recovers only $(n+1)^{-n}$; the missing $n!$ depends on the entire distribution of the simplex coordinates, not their maximum or average. An equivalent layer-cake argument rewrites the numerator as $(n+1)! \int_0^1 \operatorname{vol}(sK \cap -(1-s)K)\, ds$, but the centroid does not control these intersections sharply. Thus harmonic symmetrization identifies the extremizer without explaining its factorial. We need a genuinely $n$-dimensional counting mechanism.

## 9.3 Where the jet-counting idea first appeared

For lattice polytopes, toric geometry identifies interior lattice points with adjoint holomorphic sections. A finite torus cover first suggested why a unique interior point ought to produce a one-dimensional invariant section space, and why vanishing to order $j$ costs at most

$$\binom{n+j-1}{n}$$

Taylor conditions. The missing $n!$ is therefore the leading coefficient of the $n$-variable jet count, not a lost symmetrization factor. The cover, however, required rational polytopes and delicate boundary arguments, while our convex body might be irrational. We retain the jet-filtration idea and discard the compactification, guided by the earlier toric volume arguments of Berman and Berndtsson [62].

## 9.4 The decisive change of setting: an arbitrary body has a toric potential

The real breakthrough was to realize that rationality is unnecessary before passing to a potential. The real Monge–Ampère theorem of Berman and Berndtsson applies directly to an arbitrary centered convex body [63]. In the normalization relevant here, it supplies a smooth strictly convex function $\phi$ such that

$$\det D^2\phi = e^{-\phi}, \qquad \nabla\phi(\mathbb{R}^n) = \operatorname{int}(K), \qquad |\phi - h_K| \le C. \tag{9.2}$$

Translation in the logarithmic coordinate lets us assume $\nabla\phi(0) = 0$. Change of variables then gives

$$\int_{\mathbb{R}^n} e^{-\phi(x)}\,dx = \int_{\mathbb{R}^n} \det D^2\phi(x)\,dx = \operatorname{vol}(K) = V.$$

We now work on the open complex torus

$$X = (\mathbb{C}^*)^n, \qquad x_i = \log|z_i|^2,$$

with normalized angular measure. There is no fan, no boundary divisor, no denominator-clearing cover, and no hidden assumption that $K$ is a lattice polytope. The price is that all analytic arguments must genuinely work on a noncompact space.

The support-function estimate in (9.2) translates the lattice hypothesis exactly into a statement about weighted holomorphic functions. A Laurent monomial $z^m$ belongs to the level-$k$ Hilbert space precisely when

$$\int_{\mathbb{R}^n} \exp\bigl(m\cdot x - k\phi(x)\bigr)\,dx < \infty, \qquad \text{equivalently} \qquad \frac{m}{k} \in \operatorname{int}(K).$$

For an interior frequency, the support-function gap is uniformly positive in every direction. For a boundary frequency, one must be more careful: divergence on a single supporting ray proves nothing because that ray has measure zero. Thickening the ray into a cylinder of fixed transverse radius gives the required genuine divergence. Angular Fourier orthogonality then identifies the entire holomorphic space:

$$\mathcal{H}_k = \operatorname{span}\{z^m : m \in k\operatorname{int}(K)\cap\mathbb{Z}^n\}, \qquad d_k = \dim\mathcal{H}_k \sim Vk^n.$$

In particular, the original lattice assumption becomes the striking identity

$$\mathcal{H}_1 = \mathbb{C}.$$

That one-dimensionality is exactly what the finite-cover argument had been trying to manufacture.

## 9.5 How the lattice count became a sharp lower slope

Fix $p = (1,\dots,1) \in X$, and filter the actual section space by its order of vanishing at $p$:

$$F_k^j = \{s \in \mathcal{H}_k : \operatorname{ord}_p(s) \ge j\}.$$

The vector space of Taylor coefficients of total degree less than $j$ has dimension $\binom{n+j-1}{n}$. Consequently,

$$\dim F_k^j \ge \left(d_k - \binom{n+j-1}{n}\right)_+. \tag{9.3}$$

This estimate needs no interpolation theorem: failure of surjectivity can only make the kernel larger.

Choose an orthonormal basis $(s_a)$ adapted to this filtration and write $j_a = \operatorname{ord}_p(s_a)$. The basic identity

$$\sum_a j_a = \sum_{j\ge1} \dim F_k^j$$

counts each basis vector once for every vanishing condition it satisfies. Only orders up to the sharp jet-count scale matter. With

$$c = (n!V)^{1/n}, \qquad N_k = \lfloor ck \rfloor, \qquad q_a = \min\{j_a, N_k\},$$

define the truncated pointwise initial slope

$$g_k(z) = \frac{1}{k} \frac{\sum_a q_a |s_a(z)|^2}{\sum_a |s_a(z)|^2},$$

and the normalized Bergman probability

$$d\mu_k = \frac{B_k(z) e^{-k\phi(z)}}{d_k}\, d\nu(z), \qquad B_k(z) = \sum_a |s_a(z)|^2.$$

Orthonormality yields the exact mean identity

$$\int_X g_k \, d\mu_k = \frac{1}{k d_k} \sum_a q_a = \frac{1}{k d_k} \sum_{j=1}^{N_k} \dim F_k^j.$$

Since $d_k \sim V k^n$, the positive part in (9.3) persists until $j \sim ck$. Its Riemann sum therefore gives

$$\begin{aligned} \liminf_{k\to\infty} \int_X g_k \, d\mu_k &\geq \frac{1}{V} \int_0^c \left( V - \frac{t^n}{n!} \right) dt \\ &= \frac{n}{n+1} (n!V)^{1/n}. \end{aligned}$$

This was the numerical heart of the whole investigation. The factor $1/n!$, invisible in the early symmetric-body arguments, appears naturally as the leading asymptotic of the jet codimension.

Passing to the correct limiting probability requires a separate analytic idea. For $u = m/k$, write

$$N_k(u) = \int_{\mathbb{R}^n} e^{k(u\cdot x - \phi(x))}\, dx.$$

On compact subsets of $\operatorname{int}(K)$, Laplace asymptotics give

$$N_k(u) \sim \left( \frac{2\pi}{k} \right)^{n/2} \frac{e^{k\phi^*(u)}}{\sqrt{\det D^2\phi(x_u)}}, \qquad \nabla\phi(x_u) = u.$$

Only frequencies within $O(k^{-1/2})$ of $\nabla\phi(x)$ contribute appreciably to the Bergman density. The Gaussian lattice sum, together with the Monge–Ampère equation, yields

$$k^{-n} B_k(x) e^{-k\phi(x)} \longrightarrow \det D^2\phi(x) = e^{-\phi(x)}.$$

The normalization $d_k/k^n \to V$ prevents mass from escaping along the noncompact torus. Scheffé's argument therefore upgrades the local density limit to total-variation convergence:

$$\mu_k \longrightarrow \mu = \frac{e^{-\phi}}{V}\, d\nu.$$

The truncation gives the exact uniform bound $0 \leq g_k \leq c$ without any additional control of the largest vanishing order. Thus the jet estimate genuinely concerns the limiting Monge–Ampère probability, not an artificial discrete surrogate.

## 9.6 The level-one Bergman kernel closes the convexity gap

The filtered basis naturally defines the finite Bergman ray

$$u_t^k(z) = \frac{1}{k} \log \sum_a e^{t q_a} |s_a(z)|^2.$$

It is convex in $t$, and after introducing $\tau$ with $|\tau|^2 = e^t$, its summands become $|\tau^{q_a} s_a(z)|^2$. Hence the associated space–time function is plurisubharmonic. This remains true even though the filtration is imposed at one torus point and therefore breaks torus invariance. The regularized tail envelope

$$\Psi = \lim_{r\to\infty} \operatorname{usc}_{(z,\tau)} \sup_{k\geq r} \frac{1}{k} \log \sum_a |s_a(z)\tau^{q_a}|^2, \qquad \psi_t(z) = \Psi(z, e^{t/2}),$$

preserves joint plurisubharmonicity. Local Bergman asymptotics and the truncation yield the actual bounds

$$\psi_0 = \phi, \qquad \phi \leq \psi_t \leq \phi + ct, \qquad g = \partial_{t+}\psi_0.$$

Pointwise convexity, reverse Fatou, and total-variation convergence transfer the sharp jet-counting bound to this envelope:

$$\int_X g\, d\mu \geq \frac{n}{n+1}(n!V)^{1/n}. \tag{9.4}$$

The desired scalar functional is the genuine normalized negative-log partition

$$L(t) = -\log\left(\frac{1}{V}\int_X e^{-\psi_t}\, d\nu\right), \qquad L(0) = 0.$$

The bounds on $\psi_t - \phi$ and dominated convergence give $L'_+(0) = \int g\, d\mu$.

At first it is very tempting to integrate the pointwise convex rays and declare the resulting scalar function convex. Indeed,

$$A_k(t) = \int_X \left(u_t^k - u_0^k\right) d\mu_k$$

is genuinely convex, and its initial derivative is exactly the jet-rank sum. But $A_k$ is *not* the required logarithmic partition. The correct finite object is

$$P_k(t) = -\log \int_X \exp\bigl(-(u_t^k - u_0^k)\bigr)\, d\mu_k.$$

It has the same initial derivative as $A_k$, but Jensen gives

$$P_k(t) \leq A_k(t),$$

with strict inequality unless the potential increment is almost everywhere constant. Confusing these two functions would manufacture a proof out of an invalid identification.

Differentiating the actual partition exposes the exact missing statement. For its tilted probability $\rho_{k,t}$,

$$P_k''(t) = \mathbb{E}_{\rho_{k,t}}[\partial_t^2 u_t^k] - \mathrm{Var}_{\rho_{k,t}}(\partial_t u_t^k). \tag{9.5}$$

Pointwise convexity controls the first term, but it says nothing about whether that term dominates the variance. Even the two affine pointwise rays $0$ and $2t$ have the nonconvex logarithmic partition

$$-\log\left(\frac{1+e^{-2t}}{2}\right).$$

Thus neither nonnegative slopes nor restricting to $t \geq 0$ repairs the problem.

A sustained attempted repair used a noncompact weighted Brascamp–Lieb inequality. The cofactor identity

$$e^{-\phi}(D^2\phi)^{-1} = \mathrm{cof}(D^2\phi)$$

and divergence-free Newton tensors suggest integrating the Hessian against Monge–Ampère measure. Yet cutoffs on the noncompact torus, closure of the weighted gradient, and control of the tilted variance must all be justified simultaneously. Even success for one averaged ray would not prove scalar partition convexity for arbitrary filtrations. The decisive simplification is that the unique interior lattice point itself provides a *rank-one Bergman kernel at the final level*.

Indeed, $\psi_t - \phi$ is bounded for every fixed $t$, so its weighted holomorphic space has the same Laurent-monomial criterion as $\phi$:

$$\mathcal{H}(\psi_t) = \mathrm{span}\{z^m : m \in \mathrm{int}(K) \cap \mathbb{Z}^n\} = \mathbb{C}.$$

Its Bergman kernel is therefore independent of $z$ and equals

$$B_{\psi_t}(z) = \left(\int_X e^{-\psi_t}\, d\nu\right)^{-1} = \frac{e^{L(t)}}{V}.$$

Berndtsson's positivity theorem makes the logarithm of this Bergman kernel plurisubharmonic whenever its weight is jointly plurisubharmonic [64]. Passing to Euclidean measure adds only the pluriharmonic correction $\sum_i \log|z_i|^2$. Thus

$$(z,\tau) \longmapsto L(\log|\tau|^2) - \log V$$

is plurisubharmonic. Radial subharmonicity is exactly convexity in $t = \log|\tau|^2$, so $L$ is convex.

This closes the gap that had defeated the finite-level variance approach. For a multidimensional section space, positivity of $\log(v^* G(t)^{-1} v)$ need not control the particular Gram-matrix entry $-\log G_{00}(t)$. When the section space is one-dimensional, however, the Bergman kernel and the inverse partition are the same object. The lattice hypothesis supplies precisely that one-dimensionality, while the regularized higher-level jet envelope supplies the plurisubharmonic variation to which the positivity theorem applies. Filtered-ray and Okounkov-body ideas help explain the envelope construction conceptually [65, 66].

## 9.7 The upper slope comes from a shrinking complex ball

The second side of the slope comparison has an unexpectedly simple geometric origin. A section vanishing to order $j_a$ at $p$ satisfies a Schwarz-type estimate

$$|s_a(\zeta)| \lesssim e^{kM_r/2}\left(\frac{|\zeta|}{r}\right)^{j_a}$$

in a fixed logarithmic coordinate ball about $p$. In the Bergman ray, the factor $e^{tq_a}$, with $q_a \le j_a$, is therefore canceled by restricting to

$$|\zeta| \lesssim re^{-t/2}.$$

On this shrinking ball the limiting potential stays uniformly bounded above. Its real dimension is $2n$, so the ball has volume comparable to $e^{-nt}$. Consequently,

$$\int_X e^{-\psi_t}\, d\nu \gtrsim e^{-nt}, \qquad L(t) \le nt + O(1).$$

Convexity of $L$ now implies that its initial right derivative satisfies

$$L'_+(0) \le n.$$

Putting this together with (9.4) gives the sharp volume argument in one line:

$$\frac{n}{n+1}(n!V)^{1/n} \le L'_+(0) \le n \quad \Longrightarrow \quad V \le \frac{(n+1)^n}{n!}.$$

The provenance of the two constants explains why the argument works. The lower slope sees the $n$-dimensional simplex of Taylor multi-indices; the upper slope sees the real $2n$-dimensional volume of a shrinking complex ball. The conjectured numerical constant is the point at which these two geometries meet.

The essential distinction is that convexity comes from the genuine rank-one Bergman kernel of the limiting envelope, not from the easier averaged ray and not from a conjectured variance estimate for arbitrary finite partitions. This distinction establishes the sharp volume inequality in all dimensions.

## 9.8 What equality would have to mean

If $V = C_n$, the slope argument forces $L(t) = nt$. Classifying equality requires a further saturation argument: a holomorphic gradient $W = (D^2\phi)^{-1}\bar\partial g$ should produce a Hamiltonian Killing field. Jet sublevel bounds $\mu\{g \le s\} \lesssim s^n$ would force that field to have full complex rank at its fixed point.

The angular torus would then decompose the Killing algebra into integral root spaces. The affine slack relations imply that the root components are paths, hence only Dynkin types $A_{k_1}, \dots, A_{k_d}$ can occur. Their moment images are centered simplices, giving $V = \prod_a (k_a+1)^{k_a}/k_a!$ with $\sum_a k_a = n$. Because $C_aC_b < C_{a+b}$ for $a, b > 0$, equality allows only one simplex; the lattice condition would make its change of basis unimodular. This root-system route is mathematically suggestive [67], but constructing and saturating the global Killing field is separate from the established volume inequality. The global equality classification does not follow from that inequality alone.

Chapter 10

# Multicolor Ramsey Theory

## 10.1 Why fixed products could never settle the problem

Write $b_k = R_k(3) - 1$, the largest order of a complete graph admitting a triangle-free $k$-edge-coloring. Lexicographically multiplying two such colorings gives $b_{k+\ell} \geq b_k b_\ell$, so Fekete's lemma yields

$$L = \lim_{k\to\infty} R_k(3)^{1/k} = \sup_{k\geq 1} b_k^{1/k}. \tag{10.1}$$

Erdős asked whether $L$ is finite [68]. Known small-color constructions provide fixed exponential bases [69, 70], but tensoring any one construction can never make the base diverge.

An equivalent covering formulation clarifies the target. Let $q(n)$ be the fewest triangle-free graphs whose union covers $E(K_n)$. Assigning each edge to one covering graph gives

$$q(n) \leq k \quad \Longleftrightarrow \quad n \leq b_k.$$

Thus $L = \infty$ means that along arbitrarily large constructions, covering the complete graph takes $o(\log n)$ triangle-free colors. A single fixed product keeps $q(n)/\log n$ bounded away from zero, however many times it is iterated.

An initial upper-bound attempt partitions the neighbors of a vertex by their edge colors. Each color-$i$ neighborhood avoids color $i$ internally, giving $b_k - 1 \leq k b_{k-1}$ and only a factorial upper bound [71, 72]. Conversely, first-difference colorings, random codes, and fixed products all spend a constant number of colors per logarithmic increase in vertices. Cayley-graph constructions would need unexpectedly efficient product-free partitions; reformulating in terms of Shannon capacity merely transfers the difficulty to another hard problem [73]. None explains how to coordinate many colors while their effective exponential base grows.

Two seemingly plausible shortcuts consumed substantial attention. Coloring permutations by their first displaced symbol promises $k!$ vertices from $k$ colors, but three permutations can move the same earliest symbol to three distinct positions, producing precisely the monochromatic triangle the construction was meant to forbid. Refining the color by the displaced pair avoids that defect only by spending quadratically many colors. Likewise, a Cayley coloring would require partitioning a large group into inverse-closed product-free sets; for abelian groups this recovers the equally difficult Schur-type problem rather than bypassing it.

The Shannon reformulation is exact enough to show why it is not a shortcut. If $F$ is triangle-free and $J = \overline{F}$, an independent set in the strong power $J^{\boxtimes r}$ produces an $r$-coloring of its pairs: choose a coordinate where the two words form an edge of $F$. A monochromatic triangle would project to a triangle of $F$. Hence

$$\alpha(J^{\boxtimes r}) \leq b_r, \qquad \Theta(J) \leq L.$$

Finding complements with unbounded Shannon capacity would therefore solve the original question, rather than reduce it to an easier known estimate. A tempting stronger argument assigns each triangle-free color graph a bounded-rank positive-semidefinite certificate and multiplies those certificates entrywise. It would give an exponential upper bound, but its essential premise is false: triangle-free graphs can have unbounded chromatic and orthogonal-representation complexity. At this point, attempts to prove $L < \infty$ and attempts to achieve factorial growth had exposed the same missing ingredient: controlled reuse of colors across different blocks.

## 10.2 The missing invariant behind palette gluing

We therefore index recursive blocks by their palettes $P$ of missing colors. Coloring an edge between blocks $P, Q$ from $P \triangle Q$ excludes a monochromatic triangle meeting three blocks: three binary membership bits cannot differ pairwise. It does *not* exclude a triangle meeting two blocks. Two edges from the same outside vertex can still select a color that also occurs between their internal endpoints.

The remedy is stronger than triangle-freeness. At stage $j$, require

$$\chi(\Gamma_j(c)) \leq j + 1 \qquad \text{for every edge color } c, \tag{10.2}$$

where $\Gamma_j(c)$ is the color-$c$ graph. If all cross edges of color $c$ from one outside vertex land in one proper internal label class, their internal endpoints cannot form a color-$c$ edge. The problem becomes finding a globally fixed rule that forces this agreement without spending new colors.

Requiring every color graph to remain bipartite appears natural, but it would demand a globally compatible two-part labeling across every block. Reusing a color then becomes almost as expensive as introducing a fresh one. Allowing $j + 1$

labels at stage $j$ is the crucial relaxation: every active block retains its existing labels, while all blocks missing a given color share one additional label. A palette records both absent edge colors and which endpoint of a cross edge is allowed to invoke an existing label.

## 10.3 A saturated matrix and its exceptional words

The useful ingredient originates in bipartite hat-guessing constructions of Alon, Ben-Eliezer, Shangguan, and Tamo [74], whose saturated matrices in turn build on zero-error list decoding [75]. Their ingredients are not new; the new step is attaching their exceptional-set control to a triangle-free, palette-separated recursion.

For $H \geq 3$, set

$$m = \lceil 2H \log H \rceil, \qquad s = m(m+1) + 1. \tag{10.3}$$

Choose an $s \times H^m$ matrix $D$ with entries in $[H]$ independently and uniformly. For a fixed set of $m+1$ columns, a particular row fails to contain every alphabet symbol with probability

$$H\left(1 - \frac{1}{H}\right)^{m+1} < He^{-m/H} \leq H^{-1}.$$

All $s$ rows therefore fail with probability at most $H^{-s}$. There are at most $H^{m(m+1)}$ column sets, so the total failure probability is $H^{m(m+1)-s} = H^{-1} < 1$. Thus one fixed matrix works simultaneously for every column set.

These two union bounds have different roles. The first puts one row at the coupon-collector scale $m \asymp H \log H$; the second pays for every possible column set simultaneously with $s \asymp H^2 \log^2 H$ independent rows. Merely choosing a useful row after a particular cross edge is revealed would leave the desired agreement tautological. The successful matrix must be fixed first and remain valid for every word pair at every recursive stage.

For $x \in [H]^s$, let $\bar{x}$ denote its first $m$ coordinates and put $g_r(x) = D_{r,\bar{x}}$. For $y \in [H]^s$, define

$$E_y = \{z \in [H]^m : D_{r,z} \neq y_r \text{ for every } r \in [s]\}.$$

The matrix property forces $|E_y| \leq m$: otherwise $m+1$ exceptional columns contain $y_r$ in some row. Enumerate $E_y$ as $z^{(1)}, \ldots, z^{(\rho)}$, and prescribe

$$f_d(y) = z_d^{(d)} \quad (1 \leq d \leq \rho), \qquad f_d(y) = 1 \quad (d > \rho).$$

If $\bar{x} \notin E_y$, some $d$ satisfies $y_d = g_d(x)$. Otherwise $\bar{x} = z^{(d)}$ for some $d \leq m$, so $x_d = f_d(y)$. Consequently the same two deterministic maps satisfy

$$\forall x, y \in [H]^s \quad \exists d \in [s] : \quad x_d = f_d(y) \text{ or } y_d = g_d(x). \tag{10.4}$$

Crucially, the maps are fixed before any block pair or cross edge is chosen.

The apparent asymmetry is what makes this economical. The $m$ initial coordinates index the $H^m$ matrix columns, while all $s$ row coordinates are available to the map $g$. If a column is exceptional, its enumeration index $d$ never exceeds $m$, so the diagonal coordinate $z_d^{(d)}$ exists. The backward map handles all ordinary columns; the forward map spends one distinct coordinate on each exceptional column. Using all $s$ coordinates as column indices instead would enlarge the matrix and break the union bound.

## 10.4 Packing enough separated palettes

Set $t = s\lceil \log H \rceil$ and $M_j = jt$. At stage $j$, choose a maximal family $\mathcal{P}_j \subseteq \binom{[jt]}{t}$ whose distinct palettes satisfy

$$|P \setminus Q| = |Q \setminus P| \geq s.$$

A radius-$(s-1)$ Johnson ball has size $\sum_{d<s} \binom{t}{d}\binom{(j-1)t}{d}$. Maximality and the elementary binomial estimates therefore give, for $B_j = |\mathcal{P}_j|$,

$$B_j \geq \frac{\binom{jt}{t}}{\sum_{d<s} \binom{t}{d}\binom{(j-1)t}{d}} \geq \frac{j^t}{s(e^2 j \lceil \log H \rceil^2)^s}, \qquad B_j \geq 1. \tag{10.5}$$

Each stage thus supplies roughly $j^t$ blocks while preserving $s$ usable colors in each directed difference.

The denominator has a concrete interpretation: to obtain a palette at difference $d$ from $P$, remove $d$ of its $t$ colors and insert $d$ of the $(j-1)t$ colors outside it. Because $t/s = \lceil \log H \rceil \geq 2$,

$$\binom{jt}{t} \geq j^t, \qquad \sum_{d<s} \binom{t}{d}\binom{(j-1)t}{d} \leq s\binom{t}{s}\binom{jt}{s}.$$

Applying $\binom{N}{r} \leq (eN/r)^r$ gives the displayed bound. At $j = 1$ the analytic expression may be less than one, but maximality still supplies a palette, so there is no missing initial stage.

## 10.5 The recursive cross-edge rule

At stage $j$, make one block $V_P$ for each $P \in \mathcal{P}_j$. Since $|[jt] \setminus P| = (j-1)t$, put a relabeled copy of the preceding coloring inside $V_P$, using exactly its nonmissing colors. For each active $c \notin P$, the inductive invariant supplies a proper labeling $\ell_c^P : V_P \to [j]$.

For ordered palettes $P < Q$, select distinct colors

$$a_1, \dots, a_s \in Q \setminus P, \qquad b_1, \dots, b_s \in P \setminus Q.$$

The $a_d$ are active on $V_P$ and missing on $V_Q$; the $b_d$ have the reverse status. For $u \in V_P$, $v \in V_Q$, form

$$x(u) = (\ell_{a_d}^P(u))_{d=1}^s, \qquad y(v) = (\ell_{b_d}^Q(v))_{d=1}^s.$$

If $x_d(u) = f_d(y(v))$ for some $d$, give $uv$ the first corresponding color $a_d$; otherwise (10.4) supplies $d$ with $y_d(v) = g_d(x(u))$, and use $b_d$. Every cross-edge color lies in $P \triangle Q$. More importantly, its active endpoint has proper label determined entirely by the opposite endpoint.

Here actual colors and proper labels live in different alphabets: $a_d, b_d \in [jt]$, whereas $x_d, y_d \in [j] \subseteq [H]$. The coordinate-cover maps operate only on labels; the successful coordinate selects an already chosen edge color. Distinctness of the $a_d$ and $b_d$ means that one given cross-edge color identifies a unique forcing coordinate. This prevents the same outside endpoint from accidentally invoking different internal label classes for the same active color.

We now check a possible monochromatic triangle. One block is handled by induction. For two blocks, if the repeated cross-edge color is missing from the block containing two vertices, it cannot occur on their internal edge; if active, the forcing rule places both vertices in the same proper label class, again excluding that internal color. For three blocks, membership in the three palettes would have to alternate on every edge, which is impossible.

Explicitly, if $u, u' \in V_P$, $v \in V_Q$, and both $uv, u'v$ have the same active color $a_d$, then

$$\ell_{a_d}^P(u) = f_d(y(v)) = \ell_{a_d}^P(u').$$

The internal edge $uu'$ cannot have that color because its label classes are proper. If the palette order is reversed, the analogous argument uses $b_d$ and $g_d$; both orientations must be checked. For three blocks, $c \in P \triangle Q$, $c \in Q \triangle R$, and $c \in P \triangle R$ would require three pairwise distinct binary membership values.

The stronger invariant also propagates: give vertices in active blocks their old label and vertices in blocks missing $c$ the new label $j+1$. A color-$c$ cross edge always joins an active block to a missing block, so this is a proper $(j+1)$-labeling of its entire color graph. The recursion can therefore continue through all $H$ stages.

## 10.6 From palette gains to a growing Ramsey base

The resulting coloring uses

$$k_H = Ht = Hs\lceil \log H \rceil \asymp H^3 \log^3 H$$

colors and has $n_H = \prod_{j=1}^H B_j$ vertices. Multiplying (10.5) and using $\log(H!) \geq H \log H - H$ yields $n_H \geq (c_0 H)^{k_H}$. Consecutive stage counts obey $k_{H+1}/k_H = 1 + O(1/\log H)$; monotonicity interpolates between them with only a constant loss. Since $H \gg k^{1/3}/\log k$, we obtain

$$R_k(3) \geq \left(\frac{ck^{1/3}}{\log k}\right)^k, \qquad k^{(1/3-o(1))k} \leq R_k(3) \leq k^{(1+o(1))k}.$$

Thus $R_k(3) = k^{\Theta(k)}$ and (10.1) gives $L = \infty$.

The optimization explains the exponent $1/3$. One saturated row needs $m \asymp H \log H$ symbols; simultaneous coverage requires $s \asymp H^2 \log^2 H$ rows; palette separation enlarges this to $t = s\lceil \log H \rceil$ colors per stage. Across $H$ stages this totals $k_H \asymp H^3 \log^3 H$. Meanwhile the product of the leading $j^t$ palette counts contributes approximately $(H!)^t$, whose logarithm is $k_H \log H + O(k_H)$; all packing losses total only $O(k_H)$. Solving for $H$ therefore produces the growing base $k_H^{1/3}/\log k_H$, rather than a fixed exponential base.

Interpolation is genuinely necessary: a lower bound only at stage counts $k_H$ would not prove the statement for every color count. If $k_H \leq k < k_{H+1}$, monotonicity gives $R_k(3) \geq (c_0 H)^{k_H}$. Since $k_H/k = 1 - O(1/\log H)$, transferring $k_H$ into the exponent costs only an absolute multiplicative constant in the base. The finitely many smaller $k$ can then be absorbed into one universal constant.

Keeping the interpolation constants gives the explicit all-color bound

$$R_k(3) \geq \left(\frac{k^{1/3}}{6e^{38} \log k}\right)^k \qquad (k \geq 2),$$

At exact palette stages the sharper construction gives $R_{k_H}(3) \geq (H/e^4)^{k_H}$. The neighboring-stage estimate absorbs the interpolation loss into $e^{38}$, and the cube-root comparison

$$k^{1/3} \leq 6H \log k$$

produces the remaining factor six. For $H = 3$, the exact parameters are $m = 7$, $s = 57$, $\lceil \log 3 \rceil = 2$, and consequently $k_3 = 342$. For $2 \leq k < 342$, the stated right-hand side is below one, so the trivial Ramsey lower bound covers the initial range. This distinguishes a genuinely all-$k$ numerical assertion from an eventual asymptotic estimate.

Chapter 11

# A Counterexample to the Erdős–Simonovits Compactness Conjecture

## 11.1 The separation required for cyclic forbidden families

For a finite family $\mathcal{F}$ of graphs, write

$$\mathrm{ex}(n,\mathcal{F}) = \max\{e(G) : |V(G)| = n,\ H \not\subseteq G \text{ for every } H \in \mathcal{F}\}.$$

The compactness question of Erdős and Simonovits asks whether some single forbidden graph controls the entire family up to a constant factor [76]. Its genuinely open-looking version requires every member of the family to contain a cycle: do there necessarily exist $H \in \mathcal{F}$ and $C > 0$ for which

$$\mathrm{ex}(n,H) \le C\,\mathrm{ex}(n,\mathcal{F})$$

for all sufficiently large $n$?

Without the cycle restriction the answer is immediately negative. The family

$$\{K_{1,2}, 2K_2\}$$

has joint extremal number one: a graph avoiding $K_{1,2}$ is a matching, and avoiding $2K_2$ leaves at most one edge. Separately the two extremal numbers are linear. But these are forests, and merely rediscovering their incompatibility does not address the strengthened question [77, 78].

The required goal is a stronger, polynomial separation:

$$\mathrm{ex}(n,\mathcal{F}) = O(n^{21/16}), \qquad \mathrm{ex}(n,H) = \Omega(n^{4/3}) \quad (H \in \mathcal{F}).$$

Since

$$\frac{4}{3} - \frac{21}{16} = \frac{1}{48},$$

every individual graph then exceeds the family extremal number by a factor at least of order $n^{1/48}$. The construction would have to use connected, bipartite, cyclic forbidden graphs and supply the individual lower bound for *every* sufficiently large $n$.

## 11.2 Why the natural reductions and rooted-tree ideas stalled

A maximum cut reduces the difficult case to bipartite cyclic forbidden graphs. Familiar even cycles do not suffice: $\mathrm{ex}(n,\{C_4,C_6\}) = \Theta(n^{4/3})$ already matches the $C_6$ scale. Coning or subdividing the forest counterexample introduces a common biclique, again giving a simultaneous dense construction.

A related construction forbids $C_4$, $C_6$, and the one-subdivisions $K'_{3,t}$ and $J'$. In characteristic two, symplectic generalized quadrangles avoid $J'$ on both sides: Plücker duality identifies their line graph with another alternating polar space. Bukh–Conlon rooted trees supply the $K'_{3,t}$ witnesses [79], while an almost-regular shadow argument gives a joint $o(n^{4/3})$ bound. The sharper construction below replaces that simpler family by admissible quotients to obtain the explicit exponent $21/16$. Host-relative even-cycle counterexamples address a different question [80, 81].

There was a second, more persistent temptation: to prove compactness positively by combining the many copies supplied by supersaturation. Yet those copies can share a tiny set of roots, and a dense common extension need not exist. Hypergraph nonprincipality gives examples of incompatible forbidden configurations, but its positive-density phenomena do not transfer automatically to sparse graph extremal numbers. Rooted-tree powers recover prescribed family exponents, while their single members often inherit the same obstruction. Each route confused abundance of copies with the ability to coordinate their overlaps.

We therefore need one pattern controlling concentrated extensions, another converting the remaining concentration into an edge-cover obstruction, and separate dense witnesses avoiding each pattern individually. Ordinary projective planes remember field characteristic but have six-cycles and the wrong $n^{3/2}$ scale. Generalized quadrangles preserve the characteristic distinction while imposing girth eight and yielding exactly $n^{4/3}$ edges.

## 11.3 The decisive geometric setting: girth eight and the half-square

Fix $C_4$ and $C_6$ as two members of the forbidden family. A bipartite host avoiding them has girth at least eight. On either bipartition class $S$, define the common-neighbor relation

$$x \sim y \quad\Longleftrightarrow\quad x \neq y \text{ and } N(x) \cap N(y) \neq \varnothing.$$

Equivalently, this is the half-square of the incidence graph on $S$. The absence of a four-cycle implies that related vertices have a unique common neighbor. The absence of a six-cycle implies that if three vertices are pairwise related, all three relations come from the same common neighbor.

These two facts supply the necessary local geometry. If $y, z$ are unrelated, define

$$Z_{yz} = \{x \in S : x \sim y,\ x \sim z\}.$$

Then $Z_{yz}$ is independent for the common-neighbor relation. More generally, for an independent triple $T \subseteq S$, define its common centers by

$$L(T) = \{c \in S : c \sim t \text{ for all } t \in T\}, \qquad r(T) = |L(T)|.$$

The centers in $L(T)$ are themselves independent. If $r(T) \geq k$, the unique common neighbors between the three bases and $k$ centers realize an injective one-subdivision of $K_{3,k}$.

Denote this subdivision by $S_k$. The cases

$$|V(S_2)| = 11, \qquad |V(S_3)| = 15$$

are the two basic building blocks. Informally, $S_2$ detects a triple with two centers, while $S_3$ detects a triple with three centers. The new idea was to make these two thresholds interact across both sides of the same bipartite graph.

## 11.4 Why the forbidden family must contain admissible quotients

The first template $J_0$ consists of two copies of $S_2$, with base triples

$$\{x, y, z\} \quad \text{and} \quad \{x', y, z\},$$

sharing their two bases $y, z$. Add a two-edge path

$$x - \lambda - x'$$

between the remaining bases. The template has 21 vertices. It forbids two related third bases extending the same pair $y, z$ to two-center triples.

The second template $K_0$ takes two copies of $S_3$ with opposite bipartition orientations and connects one designated center of the first to one designated center of the second. It has 30 vertices. It forbids an edge whose two endpoints are both centers of three-center configurations.

A tempting simplification would be to forbid just these two literal templates. That does not work: two copies appearing in an arbitrary host may overlap outside their distinguished subdivisions. Such an overlap is not necessarily an injective copy of the full disjoint template. The remedy is to include all *admissible color-preserving quotients*. An identification is admissible when it identifies only vertices of the same bipartition color and remains injective on each of the two designated subdivisions. For $J_0$, also require

$$x \neq x'$$

in the quotient.

Let $\mathcal{J}$ and $\mathcal{K}$ be the resulting finite quotient families, and put

$$\mathcal{F} = \{C_4, C_6\} \cup \mathcal{J} \cup \mathcal{K}.$$

Finiteness follows because each template has finitely many vertices. Every quotient remains connected and bipartite, and the injective subdivision retained inside it contains a cycle.

The quotient convention makes the forcing argument robust. Given a homomorphism from a template into a bipartite host that is injective on each marked subdivision, identify exactly those template vertices having the same image. The resulting image is an injective copy of one of the allowed quotients. Thus arbitrary incidental overlaps are accounted for without silently assuming the two subdivision witnesses are disjoint.

## 11.5 The first template controls concentrated extension sets

Let $G$ be an $n$-vertex $\mathcal{F}$-free graph with $m$ edges. A maximum cut followed by minimum-degree pruning gives a nonempty bipartite subgraph $B$ with $N \leq n$ vertices and minimum degree $d$ satisfying

$$m \leq Cnd$$

for an absolute constant $C$. Because $B$ has girth at least eight, the radius-three neighborhood of a maximum-degree vertex is a tree. Therefore

$$\Delta(B)(d-1)^2 \leq N.$$

This maximum-degree estimate later converts a small exceptional vertex set into a small edge cover.

Fix one bipartition class $S$ and $u \in S$. Every non-backtracking walk

$$u - a - w - b - v$$

of length four has $v \neq u$, $v \not\sim u$, and $w \in Z_{uv}$. Conversely, uniqueness of common neighbors makes the walk determined by $(v, w)$. Hence

$$\sum_{v \not\sim u} |Z_{uv}| \geq p, \qquad p = d(d-1)^3.$$

Set the heavy-fiber threshold

$$R_0 = \frac{p}{2N}.$$

Fibers of size less than $R_0$ contribute less than $p/2$, so at least half of the four-path mass lies in the heavy fibers.

Write

$$C_u = \sum_{v \not\sim u} \binom{|Z_{uv}|}{3}.$$

The cubic moment counts independent base triples for which $u$ is one of at least two common centers. For sufficiently large $R_0$, the estimate

$$\binom{t}{3} \geq \frac{t^3}{27}$$

on heavy fibers yields

$$C_u \gg R_0^2 p \gg \frac{d^{12}}{N^2}.$$

This is where the long four-path count becomes useful: it forces every vertex to participate in many two-center configurations.

For an unrelated pair $y, z$, let

$$A_{yz} = \{x : \{x, y, z\} \text{ is independent and } r(\{x, y, z\}) \geq 2\}.$$

If two distinct elements $x, x' \in A_{yz}$ were related, their two $S_2$ witnesses together with the common neighbor of $x$ and $x'$ would give an admissible quotient of $J_0$. Thus $A_{yz}$ is independent. Its neighborhoods are disjoint, and minimum degree therefore gives

$$|A_{yz}| \leq \frac{N}{d}.$$

Consequently, the total number $\mathcal{T}_S$ of independent base triples having at least two centers satisfies

$$|\mathcal{T}_S| \leq \frac{1}{3}\binom{N}{2}\frac{N}{d} \leq \frac{N^3}{6d}.$$

The first template has converted a local overlap prohibition into a global bound on the number of eligible triples.

## 11.6 The second template turns bad vertices into an edge cover

Call a vertex *good* if it is a center of some $S_3$ configuration, and *bad* otherwise. When $u$ is bad, every triple counted in $C_u$ has exactly two centers: it already has $u$ and at least one other center, while three centers would make $u$ good. Each such triple is therefore counted by at most two bad vertices.

The preceding lower and upper counts now combine into

$$|\operatorname{Bad}(B)|\frac{d^{12}}{N^2} \ll \sum_S |\mathcal{T}_S| \ll \frac{N^3}{d}.$$

Hence

$$|\operatorname{Bad}(B)| \ll \frac{N^5}{d^{13}}.$$

This is the concentration-control role of $\mathcal{J}$.

The second family $\mathcal{K}$ supplies the complementary global step. An edge joining two good vertices would join the designated centers of two oppositely oriented $S_3$ configurations. The admissible-quotient lemma would then produce a member of $\mathcal{K}$. Therefore every edge of $B$ meets a bad vertex:

$$\operatorname{Bad}(B) \text{ is a vertex cover of } B.$$

Using the maximum-degree estimate,

$$Nd \le 2e(B) \le 2|\operatorname{Bad}(B)|\Delta(B) \ll \frac{N^5}{d^{13}}\frac{N}{d^2}.$$

Thus

$$d^{16} \ll N^5, \qquad d \ll N^{5/16},$$

and the original minimum-degree reduction gives

$$m \ll nN^{5/16} \le n^{21/16}.$$

The exponent is not a guess from heuristic density balancing. The four-path count contributes the twelfth power of $d$ after heavy-fiber cubing, the first template contributes one more power through its independent extension bound, and the maximum-degree estimate contributes the remaining two. Together they force the exact sixteenth-power inequality and the explicit saving $1/48$.

## 11.7 Why generalized quadrangles provide the individual witnesses

The upper bound matters only if every forbidden member separately admits a dense avoiding graph. We use the incidence graph $I_q$ of the symplectic generalized quadrangle $W(q)$, whose parameters are

$$n_q = 2(q+1)(q^2+1), \qquad e_q = (q+1)^2(q^2+1),$$

with girth eight and $e_q \ge 2^{-4/3}n_q^{4/3}$ [82].

For noncollinear points $y, z$, point regularity places every third base having two centers with $y, z$ on the independent hyperbolic line $\{y,z\}^{\perp\perp}$. Thus two such bases cannot be collinear, excluding the $J$ pattern on the point side. In even characteristic $W(q)$ is self-dual, so the same exclusion holds on the line side:

$$I_q \text{ avoids } \mathcal{J} \quad (q \text{ even}).$$

In odd characteristic the dual quadrangle has type $Q(4,q)$, whose triads have zero or two centers, never three. Therefore its line side has no $S_3$, and

$$I_q \text{ avoids } \mathcal{K} \quad (q \text{ odd}).$$

These standard finite-geometric facts are recorded in [83, 82].

The odd-characteristic distinction is visible directly: after writing the symplectic space as $U \oplus U^*$, a transverse isotropic line is the graph of an invertible symmetric map $A : U \to U^*$. Its common-center directions satisfy $(Au)(u) = 0$. In odd characteristic this nonzero binary quadratic has at most two projective roots, excluding three centers. Even characteristic instead supplies the complementary self-dual exclusion.

## 11.8 The essential quantifier: different members use different fields

The lower bounds use the following table:

| forbidden member $H$ | chosen witness |
|---|---|
| $C_4$, $C_6$, or $H \in \mathcal{J}$ | $I_{2^j}$ |
| $H \in \mathcal{K}$ | $I_{3^j}$. |

For every fixed $H \in \mathcal{F}$, its designated incidence graph is $H$-free and has

$$e(I_q) \gg |V(I_q)|^{4/3}.$$

It is emphatically *not* asserted that any single incidence graph avoids the entire family. In fact, such a graph at this density would contradict the upper bound just proved. The even and odd constructions are separate witnesses for separate singleton extremal problems.

A prime-power subsequence alone would not prove

$$\operatorname{ex}(n,H) = \Omega(n^{4/3})$$

for every large $n$. Use the bounded spacing of powers of 2 or 3. For $a \in \{2,3\}$,

$$n_{aq} = 2(aq+1)(a^2q^2+1) \le a^3 n_q.$$

Given a sufficiently large host order $n$, choose the largest admissible power $q = a^j$ with

$$n_q \le n.$$

Then

$$n_q \geq \frac{n}{a^3}.$$

Add $n - n_q$ isolated vertices to $I_q$. Every member of $\mathcal{F}$ is connected and has an edge, so isolated padding cannot create a new forbidden copy. Consequently,

$$\mathrm{ex}(n, H) \geq e_q \geq 2^{-4/3} n_q^{4/3} \geq 2^{-4/3} a^{-4} n^{4/3}.$$

This yields a uniform positive lower-bound constant, independent of $n$, for every individual family member.

## 11.9 What distinguishes this from the older compactness obstructions

Avoiding $\mathcal{J}$ makes two-center extensions sparse, while avoiding $\mathcal{K}$ turns exceptional vertices into an edge cover. Neither family alone gives the simultaneous upper bound; opposite field characteristics give the separate lower bounds. Every forbidden graph is connected, bipartite, and cyclic, with polynomial separation

$$\frac{\mathrm{ex}(n, H)}{\mathrm{ex}(n, \mathcal{F})} = \Omega(n^{1/48}) \qquad (H \in \mathcal{F}).$$

Thus the finite-family rational-exponent theorem [79] cannot automatically transfer to one graph. The complete counterexample retains all required family properties and a uniform positive individual lower-bound constant; we obtain the exact host inequality

$$e(G)^{16} \leq C|V(G)|^{21}$$

for every family-free graph $G$.

Chapter 12

# A Two-Degenerate Graph Beyond the Conjectured Exponent

## 12.1 The difference between degeneracy and a bounded bipartition

Erdős and Simonovits conjectured that if a fixed bipartite graph $H$ is $r$-degenerate, then

$$\mathrm{ex}(n, H) = O_H\big(n^{2-1/r}\big).$$

The first unsettled case is $r = 2$, where the predicted upper bound is $n^{3/2}$ [84, 85, 86]. The alternatives are a uniform embedding theorem or one fixed counterexample with

$$\mathrm{ex}(n, H) \geq cn^{3/2+\varepsilon}$$

for every sufficiently large $n$.

A related but logically different Erdős–Simonovits conjecture: that a bipartite graph $H$ is 2-degenerate *if and only if* $\mathrm{ex}(n, H) = O(n^{3/2})$ [87]. Janzer had already disproved the reverse implication by constructing, for every $\eta > 0$, a 3-regular bipartite graph with extremal number $O(n^{4/3+\eta})$ [88]. A 3-regular graph has minimum degree three and therefore cannot be 2-degenerate, so that counterexample does not address the forward implication. The problem here is precisely that still-distinct forward direction.

The basic trap is to confuse two different hypotheses. If every vertex in one bipartition class has degree at most two, existing results give the anticipated $n^{3/2}$ bound [89, 85]. Positive results also cover degenerate blow-ups of trees [90]. But 2-degeneracy only says that in every subgraph *some* vertex has degree at most two. The vertices removed by a degeneracy ordering can alternate between the two bipartition classes. The counterexample exploits exactly that alternation.

We initially sought a proof. In a graph with average degree about $n^{1/2}$, convexity forces many pairs to have common neighbors. A greedy degeneracy embedding therefore appears plausible: whenever a new vertex has two earlier neighbors, place it in their common neighborhood. The obstruction is that the useful pairs for one stage need not overlap with the useful pairs needed at the next stage. Ordinary edge density gives no simultaneous control of these alternating codegree constraints.

## 12.2 Why dependent random choice and incidence geometries did not settle it

Dependent random choice controls useful pairs on one bipartition side, but later degeneracy steps can demand incompatible pairs on the other. Size-biasing a deferred witness does not fix this: conditioning on earlier overlapping parent lists changes the witness law and destroys the proposed sequential estimate. At the critical density $p \asymp n^{-1/2}$, two neighborhood constraints leave only $p^2 n = O(1)$ candidates, so the reservoir cannot survive another alternating step. A one-subdivision still has a bounded-degree bipartition class.

Several more global perspectives fail for distinct reasons. Sidorenko-type homomorphism counts permit many vertices to collapse onto a single high-codegree pair, so plentiful homomorphisms need not yield an injective copy. Tensoring the host amplifies a putative extremal advantage but still projects an embedding to potentially noninjective homomorphisms in every coordinate. Spectral expansion or a large walk count controls aggregate correlations without making the particular successive parent pairs compatible. Projective-plane and norm-graph incidence structures are either too sparse or sufficiently regular to admit the desired embeddings.

This sustained failure changes the question. Rather than force every degeneracy pattern into an arbitrary dense graph, we construct one fixed alternating pattern and a dense host where every embedded layer increases a bounded entropy potential. Enough layers make an embedding impossible.

## 12.3 Why a Hamming host has the right kind of geometry

Take two disjoint copies of the binary cube $U = \{0, 1\}^m$, of size $Q = 2^m$, and join opposite-side words whenever $\mathrm{dist}(u, z) \leq \lfloor \tau m \rfloor$, where $0 < \tau < 1/2$. Each side has degree

$$D_m = \sum_{j \leq \lfloor \tau m \rfloor} \binom{m}{j} = 2^{(h(\tau)+o(1))m},$$

with $h(x) = -x \log_2 x - (1 - x) \log_2(1 - x)$. Thus the unsampled host has exponentially many vertices and an exponentially large Hamming-ball degree.

We retain each vertex independently with probability $p = 2^{-\beta m}$. The sampled graph has roughly $n \asymp pQ = 2^{(1-\beta)m}$ vertices and

$$p^2 Q D_m = 2^{(1+h(\tau)-2\beta+o(1))m}$$

edges. To beat $n^{3/2}$ requires

$$1 + h(\tau) - 2\beta > \tfrac{3}{2}(1-\beta) \quad \iff \quad \beta < 2h(\tau) - 1.$$

This is the density side of the optimization.

The avoidance side runs in the opposite direction. A large family of distinct words survives thinning only if it has sufficiently many coordinatewise possibilities. If every potential child layer of a particular embedded graph has conditional entropy less than $\beta$, a first-moment argument eliminates that layer. The right forbidden graph should force every successful layer to have entropy above $\beta$, while the Hamming distance constraint bounds that entropy from above unless a bounded potential increases.

## 12.4 The entropy inequality that turns geometry into a potential

Let $X, Y$ be independent Bernoulli variables with parameter $q$, and let $Z$ be an arbitrary binary child. Write $v = \mathbb{P}(Z = 1)$ and $d = [\mathbb{P}(X \neq Z) + \mathbb{P}(Y \neq Z)]/2$. The decisive estimate is

$$H(Z \mid X, Y) \leq \kappa + (\log_2 3)d + \frac{h(v) - h(q)}{2},$$

where $\kappa = \frac{3}{2} - \frac{3}{4}\log_2 3$. The first terms measure closeness to both parents; the entropy difference telescopes between layers. A log-sum inequality with weights $\sqrt{\mathbb{P}(Z=z)}\,3^{-c(x,y,z)}$, followed by the sharp binary optimization, gives the constant $\kappa$. Unlike merely intersecting two Hamming balls, this estimate charges extra child freedom to increased marginal entropy.

Actual parents in a complete pair layer are selected without replacement. If a layer has $L$ words, couple a uniformly chosen ordered pair of distinct indices to two independent indices; the coupling fails with probability only $1/L$. Applying the same conditional child law and controlling the change in binary entropy gives the corrected inequality

$$H(Z \mid X, Y) \leq \kappa + (\log_2 3)d + \frac{h(v) - h(q)}{2} + \eta(L),$$

where $\eta(L) = (1 + \log_2 3)/L + \frac{1}{2}h(1/L) \to 0$. The correction is essential: ignoring parent dependence would erase the small positive exponent window.

## 12.5 The small but decisive entropy window

The density and avoidance conditions require

$$A := \kappa + \tau \log_2 3 < \beta < C := 2h(\tau) - 1.$$

Optimizing this difference gives

$$\tau = \frac{1}{1+\sqrt{3}} = \frac{\sqrt{3}-1}{2} \approx 0.3660254.$$

The gap is small, but rigorously positive:

$$w = C - A = \frac{1}{4}\log_2\left(\frac{97 + 56\sqrt{3}}{192}\right) \approx 0.00372800177 > 0.$$

Choose $\beta = (A + C)/2 \approx 0.8932787$ and set

$$\delta = \frac{w}{8}, \qquad \varepsilon = \frac{w}{8(1-\beta)} \approx 0.0043665.$$

The narrowness of this window matters conceptually. A cruder entropy bound, a missing mixed-parent type, or an uncontrolled dependence correction would erase the entire separation. Conversely, once the gap is positive, its size is not an obstruction: choose an extremely large but *fixed* forbidden graph with enough layers to accumulate it.

## 12.6 The fixed pair-generated forbidden graph

Choose a sufficiently large fixed integer $L_0$, and build layers

$$V_0, V_1, \ldots, V_s, \qquad |V_0| = L_0, \qquad V_i = \binom{V_{i-1}}{2}.$$

Each vertex in $V_i$ is an unordered pair of vertices in $V_{i-1}$ and is joined to its two parents. This construction is related to complete degenerate graphs [90], but here sustains an alternating entropy obstruction. Choose the fixed depth $s$ so that $sw/2 > 1$. The smallest permissible depth is already at least 537; the construction is huge but finite, and neither $s$ nor $L_0$ depends on the eventual host dimension $m$.

Let $H$ denote this layered graph. It is connected because its first two layers form the one-subdivision of a complete graph and each later vertex has two parents. Layer parity gives a bipartition. In any nonempty subgraph, a vertex from the highest occupied layer has at most its two parents as neighbors, so $H$ is 2-degenerate.

It is important that $H$ does *not* fall into the known bounded-one-side case. A base-layer vertex has many children, while a first-layer vertex has two parents and many later children. These high-degree vertices lie on opposite bipartition sides. The graph is designed to alternate the obstruction indefinitely rather than hide it in one degree-two class. In fact the graph has the stronger invariant that for *every* proper two-coloring of $H$, each color class contains a vertex of degree greater than two. Thus no choice of bipartition orientation reduces the example to the previously known one-sided bounded-degree theorem.

Choose $L_0$ large enough that, for every later layer size $L \geq L_0$ and $M = \binom{L}{2}$, both inequalities

$$\eta(L) < \delta, \qquad L + 3\log_2(M+1) - \delta M < -1$$

hold. The first controls parent dependence; the second makes the later first-moment argument exponentially small in the cube dimension.

## 12.7 Why thinning excludes every low-entropy child array

Fix an arbitrary array of $L$ parent words and $M = \binom{L}{2}$ child words, each indexed by its unordered parent pair. At each coordinate there are exactly three parent-bit types, $00$, $11$, and $01$. The mixed type includes both orientations of the same unordered pair; treating $01$ and $10$ as independent types would insert a spurious entropy penalty.

Let $N_{g,j}$ count parent pairs of type $g$ at coordinate $j$, and let $b_{g,j}$ count those whose associated child bit is one. The conditional entropy at that coordinate is

$$H(Z_j \mid X_j, Y_j) = \sum_{g \in \{00,11,01\}} \frac{N_{g,j}}{M} h\left(\frac{b_{g,j}}{N_{g,j}}\right).$$

There are at most $(M+1)^{3m}$ complete count profiles. For one fixed profile the number of labeled child arrays, allowing repeated child words in the counting upper bound, is

$$\prod_{j,g} \binom{N_{g,j}}{b_{g,j}} \leq 2^{M\sum_{j=1}^{m} H(Z_j|X_j,Y_j)}.$$

Consequently, the number of arrays whose average conditional entropy is at most $\beta - \delta$ is at most

$$(M+1)^{3m} 2^{mM(\beta-\delta)}.$$

Now restrict to arrays of pairwise distinct child words. Exactly those could arise from an injective copy of $H$, and their retention probability is $p^M = 2^{-\beta m M}$. There are at most $2^{mL}$ possible parent arrays; the parents need not themselves have been retained for this union bound. Thus the failure probability at one side and one layer is at most

$$2^{m[L+3\log_2(M+1)-\delta M]} \leq 2^{-m}.$$

The bound holds simultaneously for all $s$ layers and both sides with failure probability at most $2s\,2^{-m}$.

The order of these steps matters. Repetitions may be allowed while counting possibilities, but the probability $p^M$ can only be applied after restricting to $M$ distinct retained children. The argument does not pretend that repeated words generate independent retention events.

## 12.8 Every hypothetical embedding would increase a bounded potential

Suppose, for contradiction, that the sampled Hamming graph contains an injective copy of $H$. Connectedness ensures that the image of each layer lies entirely on one host side, with sides alternating. All children are distinct and retained, so the low-entropy exclusion implies $E_i > \beta - \delta$ for $1 \leq i \leq s$, where $E_i$ is the mean conditional entropy of layer $i$ given its parent-pair bits.

Define the layer entropy potential

$$S_i = \frac{1}{m}\sum_{j=1}^{m} h\left(\frac{1}{|V_i|}\sum_{z \in \text{image}(V_i)} z(j)\right).$$

Each coordinate contributes a binary entropy, so $0 \leq S_i \leq 1$.

Every child is within Hamming distance at most $\tau m$ of both parents. Averaging over all pairs and all coordinates gives an average parent-child disagreement at most $\tau$. The corrected conditional-entropy inequality therefore yields

$$E_i \leq A + \frac{S_i - S_{i-1}}{2} + \eta(|V_{i-1}|).$$

Since $\eta(|V_{i-1}|) < \delta$ and $\beta - A = w/2$, comparison with the lower bound gives

$$S_i - S_{i-1} > 2(\beta - A - 2\delta) = \frac{w}{2}.$$

Summing over the fixed $s$ layers forces

$$S_s - S_0 > \frac{sw}{2} > 1,$$

contradicting the fact that both potentials lie in $[0, 1]$.

Thus, with probability tending to one, the sampled graph is $H$-free. This is the central structural argument: thinning rules out low-entropy layers, the Hamming metric bounds the entropy of high-entropy layers, and degeneracy lets one arrange enough alternating pair layers to exhaust a bounded potential.

## 12.9 Keeping enough edges despite their dependence

The original Hamming graph has $QD_m$ edges, and the expected surviving count is $\mu_m = p^2 QD_m$. Two edges sharing an endpoint have covariance $p^3 - p^4$. Accounting for all overlapping pairs on both host sides gives

$$\mathrm{Var}(e(G_m)) \le \mu_m + 2p^3 QD_m^2,$$

and hence

$$\frac{\mathrm{Var}(e(G_m))}{\mu_m^2} \le \frac{1}{\mu_m} + \frac{2}{pQ} \longrightarrow 0.$$

The retained vertex count likewise satisfies $|V(G_m)| \le 3pQ$ with probability tending to one. Therefore an $H$-free realization exists with at most

$$N_m = \left\lceil 3 \cdot 2^{(1-\beta)m} \right\rceil$$

vertices and at least $\mu_m/2$ edges.

The Hamming-ball lower bound

$$D_m \gg \frac{2^{mh(\tau)}}{m+1}$$

accounts for the floor in the radius and the polynomial binomial loss. The choice of midpoint $\beta$ gives the exact identity

$$1 + h(\tau) - 2\beta = \frac{3}{2}(1 - \beta) + \frac{w}{4}.$$

Consequently,

$$\frac{\mathrm{ex}(N_m, H)}{N_m^{3/2}} \gg \frac{2^{wm/4}}{m+1},$$

and, with $\varepsilon = w/[8(1-\beta)]$, we obtain

$$\frac{\mathrm{ex}(N_m, H)}{N_m^{3/2+\varepsilon}} \gg \frac{2^{wm/8}}{m+1} \longrightarrow \infty.$$

The gain is a fixed positive power, not merely a logarithmic improvement over the conjectured threshold.

## 12.10 Obtaining every host size and interpreting the conclusion

A construction on the subsequence $N_m$ would not establish the correct extremal statement. Since $1 < 2^{1-\beta} < 2$, eventually $N_{m+1} \le 2N_m$. For large $n$, choose $m$ with $N_m \le n < N_{m+1}$. Add $n - N_m$ isolated vertices to the $H$-free realization. The forbidden graph $H$ is connected and has no isolated vertices, so padding cannot create a copy. Therefore

$$\mathrm{ex}(n, H) \ge 2^{-3/2-\varepsilon} n^{3/2+\varepsilon}$$

for every sufficiently large integer $n$.

The decisive reversal is not an improved universal embedding theorem. The fixed alternating pattern instead turns the absence of low-entropy child arrays into a monotone potential increase, while thinning leaves more than $n^{3/2}$ edges and contradicts the conjectured threshold [86].

# Bibliography


[1] D. V. Gorbachev, *Extremal problem for entire functions of exponential spherical type, connected with the Levenshtein bound on the sphere packing density in* $\mathbb{R}^n$, Izv. Tula State Univ. Ser. Math. Mech. Inform. **6** (2000), 71–78; in Russian.

[2] H. Cohn and N. Elkies, *New upper bounds on sphere packings. I*, Ann. of Math. (2) **157** (2003), 689–714, doi:10.4007/annals.2003.157.689.

[3] J. Bourgain, L. Clozel, and J.-P. Kahane, *Principe d'Heisenberg et fonctions positives*, Ann. Inst. Fourier (Grenoble) **60** (2010), 1215–1232, doi:10.5802/aif.2552.

[4] H. Cohn and F. Gonçalves, *An optimal uncertainty principle in twelve dimensions via modular forms*, Invent. Math. **217** (2019), 799–831, doi:10.1007/s00222-019-00875-4.

[5] N. Afkhami-Jeddi, H. Cohn, T. Hartman, D. de Laat, and A. Tajdini, *High-dimensional sphere packing and the modular bootstrap*, J. High Energy Phys. **12** (2020), article 066, doi:10.1007/JHEP12(2020)066.

[6] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton Mathematical Series, vol. 32, Princeton University Press, Princeton, 1971.

[7] F. W. J. Olver et al., eds., *NIST Digital Library of Mathematical Functions*, Release 1.2.7, National Institute of Standards and Technology, 2026, dlmf.nist.gov.

[8] L. V. Ahlfors, *Complex Analysis*, 3rd ed., McGraw–Hill, New York, 1979.

[9] H. Cohn, D. de Laat, and A. Salmon, *Three-point bounds for sphere packing*, preprint, 2022, arXiv:2206.15373.

[10] R. J. McEliece, E. R. Rodemich, H. Rumsey, Jr., and L. R. Welch, *New upper bounds on the rate of a code via the Delsarte–MacWilliams inequalities*, IEEE Trans. Inform. Theory **23** (1977), 157–166, doi:10.1109/TIT.1977.1055688.

[11] G. A. Kabatianskii and V. I. Levenshtein, *On bounds for packings on a sphere and in space*, Problems Inform. Transmission **14** (1978), 1–17.

[12] P. Delsarte, *An algebraic approach to the association schemes of coding theory*, Philips Res. Rep. Suppl. **10** (1973), vi+97 pp.

[13] P. Delsarte, J. M. Goethals, and J. J. Seidel, *Spherical codes and designs*, Geom. Dedicata **6** (1977), 363–388, doi:10.1007/BF03187604.

[14] P. Feinsilver, *Representations of* $\mathfrak{sl}(2)$ *in the Boolean lattice, and the Hamming and Johnson schemes*, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **15** (2012), no. 3, article 1250019, doi:10.1142/S0219025712500191.

[15] A. Barg and D. Nogin, *Spectral approach to linear programming bounds on codes*, Problems Inform. Transmission **42** (2006), 77–89, doi:10.1134/S0032946006020025.

[16] L. A. Bassalygo, *New upper bounds for error correcting codes*, Problems Inform. Transmission **1** (1965), no. 4, 32–35.

[17] L. N. Coregliano, F. G. Jeronimo, C. Jones, N. Linial, and E. Loyfer, *Higher-order Delsarte dual LPs: lifting, constructions and completeness*, in *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, LIPIcs **362** (2026), article 44, 44:1–44:22, doi:10.4230/LIPIcs.ITCS.2026.44.

[18] C. Bachoc and F. Vallentin, *New upper bounds for kissing numbers from semidefinite programming*, J. Amer. Math. Soc. **21** (2008), 909–924, doi:10.1090/S0894-0347-07-00589-9.

[19] V. M. Sidel'nikov, *New bounds for densest packing of spheres in* $n$*-dimensional Euclidean space*, Math. USSR-Sb. **24** (1974), no. 1, 147–157, doi:10.1070/SM1974v024n01ABEH001911.

[20] M. Gromov, *Endomorphisms of symbolic algebraic varieties*, J. Eur. Math. Soc. **1** (1999), 109–197, doi:10.1007/PL00011162.

[21] G. Kun and A. Thom, *Inapproximability of actions and Kazhdan's property (T)*, arXiv:1901.03963, 2019.

[22] G. Kun, *On sofic approximations of property (T) groups*, arXiv:1606.04471v5, 2019.

[23] W. G. Leavitt, *The module type of a ring*, Trans. Amer. Math. Soc. **103** (1962), 113–130, doi:10.1090/S0002-9947-1962-0132761-4.

[24] M. Ershov and A. Jaikin-Zapirain, *Property (T) for noncommutative universal lattices*, Invent. Math. **179** (2010), 303–347, doi:10.1007/s00222-009-0218-2, arXiv:0809.4095.

[25] G. Higman, *Finitely Presented Infinite Simple Groups*, Notes on Pure Mathematics **8**, Australian National University, Canberra, 1974.

[26] J. W. Cannon, W. J. Floyd, and W. R. Parry, *Introductory notes on Richard Thompson's groups*, Enseign. Math. (2) **42** (1996), 215–256, author-hosted PDF.

[27] C. Bleak and M. Quick, *The infinite simple group* $V$ *of Richard J. Thompson: presentations by permutations*, arXiv:1511.02123.

[28] A. M. Vershik and E. I. Gordon, *Groups that are locally embeddable in the class of finite groups*, Algebra i Analiz **9** (1997), no. 1, 71–97; English transl., St. Petersburg Math. J. **9** (1998), 49–67.

[29] T. Gelander, *A view on invariant random subgroups and lattices*, in *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018*, vol. II, World Scientific, 2018, 1339–1362, arXiv:1807.06979.

[30] A. Connes, *Classification of injective factors*, Ann. of Math. (2) **104** (1976), no. 1, 73–115.

[31] M. Takesaki, *Theory of Operator Algebras II*, Encyclopaedia of Mathematical Sciences **125**, Springer, 2003.

[32] A. A. Suslin, *On the structure of the special linear group over polynomial rings*, Math. USSR-Izv. **11** (1977), no. 2, 221–238.

[33] H. Minkowski, *Ueber den arithmetischen Begriff der Aequivalenz und über die endlichen Gruppen linearer ganzzahliger Substitutionen*, J. Reine Angew. Math. **100** (1887), 449–458.

[34] Y. Cornulier and R. Tessera, *A characterization of relative Kazhdan property T for semidirect products with abelian groups*, Ergodic Theory Dynam. Systems **31** (2011), no. 3, 793–805.

[35] A. Connes and V. F. R. Jones, *Property T for von Neumann algebras*, Bull. London Math. Soc. **17** (1985), no. 1, 57–62.

[36] S. Popa, *Deformation and rigidity for group actions and von Neumann algebras*, in *Proceedings of the International Congress of Mathematicians (Madrid, 2006)*, Vol. I, European Mathematical Society, Zürich, 2007, 445–477, doi:10.4171/022-1/18.

[37] S. Popa, *Some open problems in $W^*$-rigidity*, problem list, Paris, June 2013, `https://www.math.ucla.edu/~popa/ProblemsJune2013.pdf`.

[38] A. Ioana, S. Popa, and S. Vaes, *A class of superrigid group von Neumann algebras*, Ann. of Math. (2) **178** (2013), no. 1, 231–286.

[39] I. Chifan, A. Ioana, D. Osin, and B. Sun, *Wreath-like products of groups and their von Neumann algebras I: $W^*$-superrigidity*, Ann. of Math. (2) **198** (2023), no. 3, 1261–1303.

[40] K. A. Kalorkoti, *A lower bound for the formula size of rational functions*, SIAM J. Comput. **14** (1985), 678–687, doi:10.1137/0214050.

[41] W. Baur and V. Strassen, *The complexity of partial derivatives*, Theoret. Comput. Sci. **22** (1983), 317–330, doi:10.1016/0304-3975(83)90110-X.

[42] J. A. Grochow, *Unifying and generalizing known lower bounds via geometric complexity theory*, Comput. Complexity **24** (2015), 393–475, arXiv:1304.6333.

[43] A. Boralevi, E. Carlini, M. Michałek, and E. Ventura, *On the codimension of permanental varieties*, arXiv:2402.17839.

[44] J. Heintz, *Definability and fast quantifier elimination in algebraically closed fields*, Theoret. Comput. Sci. **24** (1983), no. 3, 239–277, doi:10.1016/0304-3975(83)90002-6; corrigendum, **39** (1985), no. 2–3, 343, doi:10.1016/0304-3975(85)90150-1.

[45] M. Beecken, J. Mittmann, and N. Saxena, *Algebraic independence and blackbox identity testing*, in *Automata, Languages and Programming, Part II*, L. Aceto, M. Henzinger, and J. Sgall, eds., Lecture Notes in Comput. Sci. **6756**, Springer, 2011, 137–148, doi:10.1007/978-3-642-22012-8_10.

[46] R. Raz, *A parallel repetition theorem*, SIAM J. Comput. **27** (1998), no. 3, 763–803, doi:10.1137/S0097539795280895.

[47] T. Holenstein, *Parallel repetition: Simplification and the no-signaling case*, Theory Comput. **5** (2009), 141–172, doi:10.4086/toc.2009.v005a008.

[48] H. Yuen, *A parallel repetition theorem for all entangled games*, arXiv:1604.04340.

[49] A. Chailloux and G. Scarpa, *Parallel repetition of entangled games with exponential decay via the superposed information cost*, arXiv:1310.7787.

[50] I. Dinur, D. Steurer, and T. Vidick, *A parallel repetition theorem for entangled projection games*, arXiv:1310.4113.

[51] M. Bavarian, T. Vidick, and H. Yuen, *Anchored parallel repetition for nonlocal games*, arXiv:1509.07466.

[52] D. Petz, *Bregman divergence as relative operator entropy*, Acta Math. Hungar. **116** (2007), 127–131, doi:10.1007/s10474-007-6014-9.

[53] I. H. Kim, *Modulus of convexity for operator convex functions*, J. Math. Phys. **55** (2014), article 082201, doi:10.1063/1.4890292; arXiv:1310.0746v3.

[54] I. Dinur, G. Kindler, R. Raz, and S. Safra, *An improved lower bound for approximating-CVP*, Combinatorica **23** (2003), 205–243, doi:10.1007/s00493-003-0019-y.

[55] O. Goldreich and S. Goldwasser, *On the limits of nonapproximability of lattice problems*, J. Comput. System Sci. **60** (2000), 540–563, doi:10.1006/jcss.1999.1686.

[56] D. Aharonov and O. Regev, *Lattice problems in* NP ∩ coNP, J. ACM **52** (2005), no. 5, 749–765, doi:10.1145/1089023.1089025.

[57] A. Cafure and G. Matera, *Improved explicit estimates on the number of solutions of equations over a finite field*, Finite Fields Appl. **12** (2006), 155–185, doi:10.1016/j.ffa.2005.03.003; arXiv:math/0405302.

[58] I. S. Reed and G. Solomon, *Polynomial codes over certain finite fields*, J. Soc. Indust. Appl. Math. **8** (1960), no. 2, 300–304, doi:10.1137/0108018.

[59] E. R. Berlekamp, R. J. McEliece, and H. C. A. van Tilborg, *On the inherent intractability of certain coding problems*, IEEE Trans. Inform. Theory **24** (1978), no. 3, 384–386, doi:10.1109/TIT.1978.1055873.

[60] E. Ehrhart, *Volume réticulaire critique d'un simplexe*, J. Reine Angew. Math. **305** (1979), 218–220.

[61] B. Nill and A. Paffenholz, *On the equality case in Ehrhart's volume conjecture*, Adv. Geom. **14** (2014), no. 4, 579–586, doi:10.1515/advgeom-2014-0001; arXiv:1205.1270.

[62] R. J. Berman and B. Berndtsson, *The volume of Kähler–Einstein Fano varieties and convex bodies*, J. Reine Angew. Math. **723** (2017), 127–152, doi:10.1515/crelle-2014-0069; arXiv:1204.1308.

[63] R. J. Berman and B. Berndtsson, *Real Monge–Ampère equations and Kähler–Ricci solitons on toric log Fano varieties*, Ann. Fac. Sci. Toulouse Math. (6) **22** (2013), no. 4, 649–711, doi:10.5802/afst.1386; arXiv:1207.6128.

[64] B. Berndtsson, *Subharmonicity properties of the Bergman kernel and some other functions associated to pseudoconvex domains*, Ann. Inst. Fourier (Grenoble) **56** (2006), 1633–1662.

[65] S. Boucksom and H. Chen, *Okounkov bodies of filtered linear series*, Compos. Math. **147** (2011), 1205–1229.

[66] J. Ross and D. Witt Nyström, *Analytic test configurations and geodesic rays*, J. Symplectic Geom. **12** (2014), 125–169.

[67] J. S. Milne, *Lie Algebras, Algebraic Groups, and Lie Groups*, version 2.00, 2013, https://www.jmilne.org/math/CourseNotes/LAG.pdf.

[68] T. F. Bloom, *Erdős problem #183*, https://www.erdosproblems.com/183; see also https://www.erdosproblems.com/latex/183.

[69] F. R. K. Chung, *On the Ramsey numbers* $N(3,3,\dots,3;2)$, Discrete Math. **5** (1973), 317–321, doi:10.1016/0012-365X(73)90125-8.

[70] G. Exoo, *A lower bound for Schur numbers and multicolor Ramsey numbers*, Electron. J. Combin. **1** (1994), R8.

[71] F. R. K. Chung and C. M. Grinstead, *A survey of bounds for classical Ramsey numbers*, J. Graph Theory **7** (1983), 25–37, doi:10.1002/jgt.3190070105.

[72] S. P. Radziszowski, *Small Ramsey numbers*, Electron. J. Combin., Dynamic Survey DS1, https://www.combinatorics.org/ojs/index.php/eljc/article/view/DS1.

[73] W. T. Gowers, *Decompositions, approximate structure, transference, and the Hahn–Banach theorem*, Bull. Lond. Math. Soc. **42** (2010), 573–606, doi:10.1112/blms/bdq018; see also arXiv:0811.3103.

[74] N. Alon, O. Ben-Eliezer, C. Shangguan, and I. Tamo, *The hat guessing number of graphs*, J. Combin. Theory Ser. B **144** (2020), 119–149, doi:10.1016/j.jctb.2020.01.003; see also arXiv:1812.09752.

[75] S. Chakraborty, J. Radhakrishnan, N. Raghunathan, and P. Sasatte, *Zero error list-decoding capacity of the* $q/(q-1)$ *channel*, in *Foundations of Software Technology and Theoretical Computer Science*, Lecture Notes in Comput. Sci. **4337**, Springer, 2006, 129–138.

[76] P. Erdős and M. Simonovits, *Compactness results in extremal graph theory*, Combinatorica **2** (1982), 275–288, doi:10.1007/BF02579234.

[77] Y. Wigderson, *The Erdős–Simonovits compactness conjecture needs more assumptions*, available online.

[78] T. F. Bloom, *Erdős problem #180*, https://www.erdosproblems.com/180.

[79] B. Bukh and D. Conlon, *Rational exponents in extremal graph theory*, J. Eur. Math. Soc. **20** (2018), 1747–1757, doi:10.4171/JEMS/798.

[80] D. Conlon, E. Mulrenin, and C. Pohoata, *Two counterexamples to a conjecture about even cycles*, 2026, arXiv:2603.24515.

[81] J. Verstraëte, *Extremal problems for cycles in graphs*, in *Recent Trends in Combinatorics*, IMA Vol. Math. Appl. **159**, Springer, Cham, 2016, 83–116, doi:10.1007/978-3-319-24298-9_4.

[82] S. E. Payne and J. A. Thas, *Finite Generalized Quadrangles*, 2nd ed., EMS Series of Lectures in Mathematics **9**, European Mathematical Society, 2009, doi:10.4171/066.

[83] D. Bartoli, T. Héger, G. Kiss, and M. Takáts, *On the metric dimension of affine planes, biaffine planes and generalized quadrangles*, Australas. J. Combin. **72** (2018), 226–248, https://ajc.maths.uq.edu.au/pdf/72/ajc_v72_p226.pdf.

[84] P. Erdős and M. Simonovits, *Cube-supersaturated graphs and related problems*, in *Progress in Graph Theory*, Academic Press, 1984, 203–218.

[85] N. Alon, M. Krivelevich, and B. Sudakov, *Turán numbers of bipartite graphs and related Ramsey-type questions*, Combin. Probab. Comput. **12** (2003), 477–494, doi:10.1017/S0963548303005741.

[86] T. F. Bloom, *Erdős problem #146*, https://www.erdosproblems.com/146.

[87] T. F. Bloom, *Erdős problem #113*, https://www.erdosproblems.com/113.

[88] O. Janzer, *Disproof of a conjecture of Erdős and Simonovits on the Turán number of graphs with minimum degree 3*, Int. Math. Res. Not. **2023** (2023), 8478–8494, doi:10.1093/imrn/rnac076.

[89] Z. Füredi, *On a Turán type problem of Erdős*, Combinatorica **11** (1991), 75–79, doi:10.1007/BF01375476.

[90] A. Grzesik, O. Janzer, and Z. L. Nagy, *The Turán number of blow-ups of trees*, J. Combin. Theory Ser. B **156** (2022), 299–309, doi:10.1016/j.jctb.2022.05.004.