hexagon:2610.00687v1

THE LAW OF FORGETTING IN MOMENTS OF SATO–TATE MEASURES: INVARIANT TENSORS, MAXIMAL TORI, AND THE RANK LAW

CHUNPING ZENG

Abstract. For a compact group G with a finite-dimensional representation V, the moments of the character $\langle \chi_V^m \rangle_G$ are, by a one-line argument, the dimensions of the invariant tensor spaces:

$$\langle \chi_V^m \rangle_G = \dim(V^{\otimes m})^G.$$

We call the passage from $\langle\cdot\rangle_G$ to $\langle\cdot\rangle_T$, with $T \subseteq G$ a maximal torus, the *law of forgetting*: T-invariants only require weights to sum to zero, G-invariants additionally require invariance, and for the standard representation of SU(2) the two ladders differ by the exact factor

$$F_k = \frac{\dim(V_2^{\otimes 2k})^T}{\dim(V_2^{\otimes 2k})^{\mathrm{SU}(2)}} = \frac{\binom{2k}{k}}{\mathrm{Cat}(k)} = k+1.$$

The general forgetting factor admits a complete asymptotic

$$F_k = c_V\, k^s \Big(1 + \frac{\gamma_V}{k} + O(k^{-2})\Big), \qquad s = \#\{\text{positive roots}\},$$

with both representation-dependent constants in closed form, $c_V = \dfrac{|W|}{2^s \langle h_d \rangle}$ (a ratio of the Weyl Jacobian's lowest homogeneous term at the saddle to the Gaussian average) and $\gamma_V = \frac{1}{2}(\varepsilon_T - \varepsilon_G)$, whose inputs are the quartic term of $-\log|\chi/\dim V|$ and the next homogeneous term of the Weyl denominator. The power $s = (\dim G - \operatorname{rank} G)/2$ is verified across six compact Lie groups $(A_1, A_2, B_2, C_2, C_3, G_2)$ by a zero-parameter hard prediction $F(1) = \dim V$ (6/6), and the universal family U(N) gives the closed form $F(N,k) \sim R_N k^{\binom{N}{2}}$ with $R_N = (2/N)^{\binom{N}{2}}/G(N+1)$ the reciprocal of a Mehta constant. On the arithmetic side, the master identity is verified on elliptic-curve Frobenius data (the first moments hit $\dim \mathrm{End}_G$ and the Clebsch–Gordan values), the two-copy test $E \times E$ versus $E_1 \times E_2$ reads Schur's lemma directly from the second moment, and the (λ,λ) pairing identity of Howe duality is identified with the Dirichlet coefficients of the Rankin–Selberg L-function, the $N \leftrightarrow k$ symmetry reading as convolution symmetry $L(\pi \times \pi') = L(\pi' \times \pi)$. A parallel shape fingerprint $W = \mathbb{E}[Y^2]/\mathbb{E}[Y]^2$ bypasses the heavy-tail divergence of the Jensen-gap functional $J = \mathbb{E}[Y]\mathbb{E}[1/Y]$, with closed forms for twelve distribution families. All conclusions carry the project's status labels; every error found during the research is recorded with its correction in Appendix A.

1. Introduction

Across arithmetic objects with a Sato–Tate group — elliptic curves, modular forms, dynamical systems on $\mathbb{P}^1$ — one repeatedly meets two ladders of moments: the "memory" ladder computed by Weyl integration over the full group, and the "forgetting" ladder obtained when the Weyl Jacobian is dropped. This paper shows that the two ladders are both instances of a single identity, that their ratio (the *forgetting factor*) obeys a universal power law whose exponent is the number of positive roots, whose constants carry the arithmetic fingerprint, and that several structural phenomena previously observed on the arithmetic side (the ubiquitous $k+1$, the Catalan numbers, the passage from GOE statistics to Poisson under degeneracy) are the projections of this group-theoretic mechanism.

Date: October 2026.

Theorem 1.1 (Master identity)**.** *Let G be a compact group, V a finite-dimensional representation, χ_V its character. Then for every $m \geq 0$,*

$$\langle \chi_V^m \rangle_G := \int_G \chi_V(g)^m \, dg = \dim(V^{\otimes m})^G. \tag{1.1}$$

Proof. $\chi_V(g)^m = \chi_{V^{\otimes m}}(g)$ by multiplicativity of the character; and $\int_G \chi_W(g)\, dg = \dim W^G$ for any representation W, since the integral is the multiplicity of the trivial representation. The chain uses nothing else. □

Equation (1.1) is an identity, not an analogy: the moment sequence of an object and the dimension sequence of its invariant tensor spaces are the same sequence.

Definition 1.2 (The forgetting factor)**.** Let $T \subseteq G$ be a maximal torus and V a representation whose weights are known. The *forgetting factor* is

$$F(V,k) = \frac{\dim(V^{\otimes 2k})^T}{\dim(V^{\otimes 2k})^G}.$$

T-invariants count weight multisets summing to zero; G-invariants impose the additional G-invariance (for SU(2) on V_2: non-crossing). Forgetting is the loss of this extra constraint.

Theorem 1.3 ($k+1$ for SU(2))**.** *For the standard representation V_2 of* SU(2) *(weights ± 1),*

$$\dim(V_2^{\otimes 2k})^T = \binom{2k}{k} \text{ (arrangements of } k \text{ up and } k \text{ down)}, \qquad \dim(V_2^{\otimes 2k})^{\mathrm{SU}(2)} = \mathrm{Cat}(k),$$

hence $F(V_2,k) = \binom{2k}{k} / \mathrm{Cat}(k) = k+1$ exactly.

Proof. The weight-zero subspace of $V_2^{\otimes 2k}$ has as basis the binary words with k up and k down letters, whence the first count. For the second: by the first fundamental theorem for SL_2, invariants in $V_2^{\otimes 2k}$ are generated by complete contractions through the invariant symplectic form $\varepsilon \in V_2^{\otimes 2}$; the antisymmetry of ε forces any contraction scheme to be representable by a *non-crossing* perfect matching of the $2k$ tensor factors (a crossed pair can be reduced by the Temperley–Lieb braid relation to a linear combination of uncrossed ones). Non-crossing perfect matchings of $2k$ points are counted by $\mathrm{Cat}(k)$, and they are linearly independent (they form the Temperley–Lieb basis), so the invariant space has dimension $\mathrm{Cat}(k) = \frac{1}{k+1}\binom{2k}{k}$. Dividing gives $k+1$. □

The slogan "forgetting is mild" is made precise by the asymptotic of Section 7: F_k grows only polynomially — the exponential bases of the two ladders coincide, both being $(\dim V)^2$.

Theorem 1.4 (Rank law, preview)**.** *For a connected compact Lie group G with representation V satisfying the genericity hypotheses of Section 7,*

$$F_k = c_V \, k^{(\dim G - \operatorname{rank} G)/2} \Big(1 + \frac{\gamma_V}{k} + O(k^{-2})\Big), \qquad c_V, \gamma_V \text{ explicit in } (G,V).$$

The exponent $(\dim G - \operatorname{rank} G)/2$ equals the number of positive roots.

The exponent is universal across representations of the same group (the "horizontal" coordinate); the constants c_V, γ_V carry the fingerprint (the "amplitude"). This is the fourth appearance of the same dichotomy in the broader program (critical exponents are universal; amplitudes remember arithmetic).

2. The master identity on arithmetic data

Take $E/\mathbb{Q}$ an elliptic curve without complex multiplication; its Sato–Tate group is SU(2) with standard representation V_2, $\chi = a_p/\sqrt{p}$, and Sym^{m-1} gives the m-dimensional irreducible with character $U_{m-1}(a_p/2\sqrt{p})$. Theorem 1.1 predicts two a priori facts:

$$\langle \chi^2 \rangle = \dim \mathrm{End}_G(V) = 1 \quad (V \text{ irreducible, Schur}); \qquad \langle \chi^2 \chi^{*2} \rangle_m = \dim(V^{\otimes 2} \otimes V^{*\otimes 2})^G = m$$

(the Clebsch–Gordan count of m distinct irreducibles). Measured against Frobenius data (5133 primes):

m	$\langle\chi^2\rangle$ (pred 1)	$\langle\chi^4\rangle$ (pred m)	$\langle\chi^6\rangle$	$\langle\chi^8\rangle$
2	0.997	1.982 (2)	4.912	13.635
3	0.988	2.890 (3)	14.131	84.144
4	0.972	3.722 (4)	30.450	316.283
5	0.980	4.495 (5)	55.088	884.948
6	0.968	5.178 (6)	89.085	2020.960

The first two columns hit their predictions (error $< 5\%$) — the master identity passes from analogy to measured law on arithmetic data.

Remark 2.1 (Credibility intervals). Moments with $k \geq 3$ are *not* trustworthy statistics: since $|\chi| \leq m$, the variance of $\langle\chi^{2k}\rangle$ is dominated by rare samples with $|\chi| \approx m$. For $m = 6$, $k = 4$: the standard error over 5133 primes is $\sim 10^2$, and the measured 2021 against the group-side value 2666 lies at exactly this scale. This heavy-tail convergence law (the same obstacle as the eighth Rayleigh moment in arithmetic dynamics and the genus-2 high moments) is universal, not family-specific. *Every numerical value involving $k \geq 3$ below is a trend indicator, never a verdict.*

3. The Weyl ladder: $W_k(N)$, factorials, and patterns

For the standard representation of $\mathrm{U}(N)$, Theorem 1.1 gives two equivalent combinatorial formulas:

$$W_k(N) = \langle|\operatorname{tr} A|^{2k}\rangle_{\mathrm{U}(N)} = \sum_{\substack{\lambda\vdash k \\ \ell(\lambda)\leq N}} (f^\lambda)^2, \qquad f^\lambda = \#\{\text{standard Young tableaux of shape } \lambda\},$$

(strictly, the standard representation of $\mathrm{U}(N)$ is not self-dual, so one computes invariants of $V^{\otimes k} \otimes \bar{V}^{\otimes k}$).

k	N=2 (Catalan)	N=3 (A005802)	N=4 (A005803)	$N \geq k$	OEIS
1	1	1	1	1	
2	2	2	2	2	
3	5	6	6	6	$= 3!$
4	14	23	24	24	$= 4!$
5	42	103	119	120	$= 5!$
6	132	513	694	719	720
7	429	2761	4582	5003	5040
8	1430	15767	33324	39429	40320

For $N \geq k$ one has $W_k(N) = k!$ — the theorem of Diaconis–Shahshahani and Rains $\int_{\mathrm{U}(N)} |\operatorname{tr} U|^{2k} = k!$ ($k \leq N$), which we recover independently.

By RSK, the sum over $\ell(\lambda) \leq N$ counts permutations whose longest increasing subsequence has length at most N:

$$W_k(N) = \#\mathrm{Av}_k(12\cdots(N+1)).$$

In particular $\mathrm{Cat}(k) = \#\{\sigma \in S_k : \mathrm{LIS}(\sigma) \leq 2\} = \#\mathrm{Av}_k(123)$: *the Catalan numbers are the rank-two truncation.* The constant presence of Catalan numbers throughout the seven-world program is thereby explained — everything was computed at rank two (the SU(2) Weyl density $\sin^2\theta$, $\mathfrak{sl}_2$, the Temperley–Lieb algebra, binary trees).

3.1. **Rank, not dimension: the adjoint representation of** SO(3)**.** The Galois representation $\mathrm{Ad}(E) = \mathrm{Sym}^2 \otimes \det^{-1}$ is three-dimensional with trace $a_p^2/p - 1 = 4\cos^2\theta - 1$; its Sato–Tate group is SO(3), still of rank one (it depends on the single angle θ). Theorem 1.1 predicts moments $\dim(V_3^{\otimes m})^{\mathrm{SO}(3)}$:

k	1	2	3	4	5
$\mathrm{Ad}(E)$ measured	0.978	2.741	12.703	71.662	443.431
analytic (SO(3))	1	3	15	91	603
$W_k(3)$ (A005802)	1	2	6	23	103
$U_k(3)$ (Domb)	3	15	93	639	4653

Verdict: $\mathrm{Ad}(E)$ gives $1, 3, 15, 91, 603$ — completely different from $W_k(3)$. The three-dimensional representation sits in the *rank-one* slot, not the $N = 3$ slot. And $1, 3, 15, 91, 603$ are the Riordan numbers (A005043): in full generality $\langle \mathrm{tr}^m \mathrm{Ad}\rangle_{\mathrm{SO}(3)} = R_m$ for the Haar measure $(1 - \cos\varphi)/2\pi\, d\varphi$, with two hard checks ($m = 1$ gives 0; $m = 3$ gives 1) both hit digit-perfectly. The same group and the same combinatorial sequence, computed independently on the representation-theoretic and arithmetic sides, coincide.

4. Schur's lemma read arithmetically: $E \times E$ versus $E_1 \times E_2$

Correction 4.1 (of an earlier formulation). The correct formula is $\langle \chi_V^2 \rangle = \sum_\mu m_\mu^2$ with m_μ the multiplicities of pairwise non-isomorphic irreducibles — i.e. $\dim \mathrm{End}_G(V)$ — not the number of irreducible constituents. The two notions separate when some $m_\mu > 1$.

Theorem 4.2 (Two-copy test)**.** *Take the same curve twice ($E \times E$, $a_p(E \times E) = 2a_p$): the Sato–Tate group degenerates to the diagonal* SU(2) *and $V = 2 \cdot V_2$ (one irreducible with multiplicity 2), so $\langle \chi^2 \rangle = \dim \mathrm{End}_G(V) = 4$ and the moment sequence is $4\,\mathrm{Cat}(k) = 4, 32, 320, 584, \ldots$. Take two non-isogenous curves: the group is* $\mathrm{SU}(2)^2$ *with $V = V_2 \boxtimes V_2'$ (two distinct irreducibles), $\langle \chi^2 \rangle = 2$, and the moments are $\mathrm{Cat}(k)\,\mathrm{Cat}(k+1) = 2, 10, 70, 588, \ldots$.*

Measured:

$$E \times E:\ 3.987,\ 31.71,\ 314.37,\ 3490.53,\ 41543.96;$$

$$E_1 \times E_2:\ 1.980,\ 9.871,\ 69.764,\ 596.151,\ 5731.960$$

— all hit. *A single number (4 versus 2) separates "isogenous copies" from "distinct copies"*: this is Schur's lemma (End_G is supported only on isotypic components) read directly off Frobenius traces, using nothing but the a_p of two elliptic curves.

Remark 4.3 (The high-moment deviation of $E \times E$ is not endpoint divergence). With $\chi = 4\cos\theta$ against the Sato–Tate density, the density vanishes at the endpoints ± 4 like $(4 - |\chi|)^{3/2}\sqrt{4 - x^2}$ — it vanishes, not diverges. The true cause is the heavy tail: $\langle \chi^{20} \rangle \approx 1.76 \times 10^{10}$ while $\chi^{20} \le 4^{20} \approx 10^{12}$; the moments are dominated by rare samples with $|\chi| \approx 4$, where the measure is tiny. The same obstacle as Remark 2.1 and as the eighth Rayleigh moment in arithmetic dynamics — three occurrences, one source.

5. Forgetting as restriction to the maximal torus

5.1. **The two ladders.** For a fixed Sato–Tate angle θ set $Y = 2\cos(n\theta)$: the horizontal (G) ladder is $\langle \chi^{2k} \rangle_G = \dim(V^{\otimes 2k})^G$ and the vertical (T) ladder is $\langle \chi^{2k} \rangle_T = \dim(V^{\otimes 2k})^T$ — both are instances of Theorem 1.1, differing only in the group. For $Y = 2\cos\theta$ under the Sato–Tate measure:

	$2f(0)$	$\langle Y^{2k} \rangle$	distribution	Weyl weight
$n = 1$	$\pi/2$ (at 0)	$\mathrm{Cat}(k) = 1, 2, 5, 14, 42$	semicircle	$\sin^2\theta$
$n \ge 2$	$1/\pi = 0.3183$	$\binom{2k}{k} = 2, 6, 20, 70, 252$	arcsine	none (uniform)

Measured on synthetic Sato–Tate data (5×10^5 samples), the ratios are $k + 1$ to 0.5%: the forgetting factor $F(V_2, k) = k + 1$ of Theorem 1.3 is confirmed.

5.2. **The source of $k+1$: the Haar Jacobian.** The Sato–Tate measure is the arcsine measure times the SU(2) Haar Jacobian:

$$\frac{2}{\pi}\sin^2\theta\,d\theta = \frac{1}{\pi}d\varphi \times 2\sin^2\frac{\varphi}{2}, \qquad \varphi = 2\theta.$$

With $x = 2\cos\theta$, the ratio of densities is $f_{\mathrm{SC}}/f_{\mathrm{arc}} = 2\sin^2\theta = 2(1-x^2/4)$: it equals 2 at $x = 0$ — this is the factor 2 in $f(0)$; after weighting a degree-$2k$ polynomial it contributes $1/(k+1)$ to the moments — this is the factor $k+1$. Two independent observables (local density, global moments) are closed by the same Jacobian. Note $k+1$ is not arithmetic: it is the universal combinatorial consequence of dropping the Weyl weight $2\sin^2\theta$; arithmetic only arranges that $n = 1$ sits exactly at the Sato–Tate measure (the $d \to \infty$ limit of the Kesten–McKay family), i.e. at the largest possible Weyl weight. Graph-world verdict: $d = 1000$ gives $F(1) = 1.998$, $F(2) = 2.996$, $F(3) = 3.993$ (targets $2, 3, 4$).

Derivation of the two moment ladders. Under the arcsine law (θ uniform), $x = 2\cos\theta$ has moments $\langle x^{2k}\rangle_{\mathrm{arc}} = \binom{2k}{k}$ (the constant term of $(z+z^{-1})^{2k}$). Under the Sato–Tate law,

$$\langle x^{2k}\rangle = \frac{2}{\pi}\int_0^\pi (2\cos\theta)^{2k}\sin^2\theta\,d\theta = \frac{2}{\pi}2^{2k}B\Big(k+\tfrac{1}{2},\tfrac{3}{2}\Big) = \mathrm{Cat}(k),$$

using $\Gamma(\frac{3}{2}) = \sqrt{\pi}/2$ and $\Gamma(k+\frac{1}{2}) = \frac{(2k)!\sqrt{\pi}}{4^k k!}$. The ratio is $\mathrm{Cat}(k)/\binom{2k}{k} = \frac{1}{k+1}$, matching Theorem 1.3. □

5.3. **Rescaling and the number of forgetting steps.** Let g be the density of θ and g_n the density of $n\theta \bmod 2\pi$. The j-th Fourier coefficient of g_n equals the nj-th Fourier coefficient of g: the n-th coefficient of g *is* the distance from uniformity. Since the Sato–Tate measure has only the single mode $\cos 2\theta$, g_n is exactly uniform for $n \geq 3$ — one-step forgetting. A general g with full modes forgets only gradually. For the Weyl sources, $|\Delta|^2$ is a finite Fourier polynomial (highest frequency $2(N-1)$), so forgetting completes in finitely many steps:

$$\langle|\operatorname{tr}(A^n)|^2\rangle_{\mathrm{U}(N)} = \min(n, N),$$

with forgetting step count $n^* = N = h(\mathrm{SU}(N))$, the Coxeter number. The exact $\min(n,N)$ is the theorem of Diaconis–Shahshahani; our independent input is the Fourier-coefficient viewpoint above (the n-th coefficient of g *is* the distance from uniformity), which gives the qualitative "finite-step forgetting" and the step count directly, and is the correct framing for non-Weyl sources such as Section 5.4. SU(2) gives $n^* = 2$: the group-theoretic root of $k+1 = 2$.

5.4. **Non-Weyl sources: the Kesten–McKay closed form.** For random d-regular graphs the Kesten–McKay measure (normalised $Y = \lambda/\sqrt{d-1}$) gives $c_m = -(1-\rho)\rho^{m-1}$, $\rho = 1/(d-1)$:

$$\langle Y^2\rangle_n = 2 - (1-\rho)\rho^{n-1}, \qquad \langle Y^2\rangle_1 = 1+\rho.$$

d	ρ	$n=1$ (meas/theory)	$n=2$	$n=3$
3	0.500	1.5000/1.5000	1.7500	1.8750
4	0.333	1.3333/1.3333	1.7778	1.9259
10	0.111	1.1111/1.1111	1.9012	1.9890

Here $\rho = 1/(d-1)$ is the "non-Weyl-ness", the forgetting rate $\log(1/\rho) = \log(d-1)$ is exponential rather than finite-step, and $d \to \infty$ sends KM to the semicircle (deepest memory) with the *fastest* forgetting (one step). The deeper the memory, the faster the forgetting — the same law as "the smaller δ (more critical), the faster the forgetting" elsewhere in the program.

6. The numerator ladder $U_k(N)$

6.1. **Definition and generating function.** The no-Weyl (uniform phase) side is the moment of a sum of N independent phases:

$$U_k(N) = \Big\langle \Big| \sum_{j=1}^{N} e^{i\varphi_j} \Big|^{2k} \Big\rangle = \sum_{|\mathbf{c}|=k} \text{multinomial}(k;\mathbf{c})^2.$$

The generating function is closed, with a one-line proof (independent phases $\Rightarrow$ N-fold convolution):

$$\sum_{k\geq 0} U_k(N)\,\frac{z^k}{(k!)^2} = I_0(2\sqrt{z})^N, \tag{6.1}$$

and $N=2$ expands to $U_k(2)=\binom{2k}{k}$ directly. By Schur–Weyl duality for symmetric powers, the matrix $(U_k(N))$ exhibits transpose symmetry; the diagonal is $U_k(k) = 1, 6, 93, 2716, \ldots$

Proposition 6.1 (Retraction of the reflection symmetry)**.** *The claimed identity $U_k(N) = U_N(k)$ is **false**: $U_2(3) = 15 \neq 20 = U_3(2)$. The structural reason: for fixed k, $U_k(N)$ is a polynomial in N of degree k; for fixed N it grows exponentially in k ($\sim c_N N^{2k} k^{-(N-1)/2}$); the two cannot be symmetric. The earlier "verification hitting all 36 entries" was the same row copied twice — a tautology, not a verification.*

The true symmetry replacing it is the Howe-duality (λ,λ) pairing of Section 10.

6.2. **Closed forms.**

Why $U_k(N)$ is a polynomial in N, and the first two closed forms. Summing over weak compositions $\mathbf{c}$ of k into N parts, $U_k(N) = \sum_{\mathbf{c}} \big(k!/\prod_i c_i!\big)^2$. Grouping compositions by their underlying partition type (finitely many types, independent of N), each type contributes a polynomial in N of degree equal to its support size $\leq k$; hence $U_k(N)$ is a polynomial of degree k with leading term $k!\,N^k$ (the all-distinct type $\lambda = (1^k)$). For $k=2$: the type (2) contributes $N\cdot(2!/2!)^2 = N$ and the type $(1,1)$ contributes $\binom{N}{2}\cdot(2!/1!1!)^2 = 2N(N-1)$, total $2N^2-N$. For $k=3$: $(3)\to N$; $(2,1)\to N(N-1)\cdot(3!/2!)^2 = 9N(N-1)$; $(1,1,1)\to\binom{N}{3}\cdot(3!)^2 = 6N(N-1)(N-2)$; total $6N^3-9N^2+4N$. The remaining rows of the table are obtained by the same finite enumeration over partition types, verified symbolically. □

k	$U_k(N)$ closed form
1	N
2	$2N^2 - N$
3	$6N^3 - 9N^2 + 4N$
4	$24N^4 - 72N^3 + 66N^2 - 17N$
5	$120N^5 - 600N^4 + 1000N^3 - 725N^2 + 206N$

The leading coefficient is $k!$ (from $\lambda = (1^k)$, all-distinct letters); the next term is $-\binom{k}{2}\cdot k!$, matching $F = N^k - \frac{k(k-1)}{4}N^{k-1} + \cdots$.

6.3. **The general forgetting factor for** $\mathrm{U}(N)$**.** Since $W_k(N) = k!$ for $N \geq k$,

$$F(N,k) = \frac{U_k(N)}{W_k(N)} = N^k - \frac{k(k-1)}{4}N^{k-1} + O(N^{k-2}),$$

and the minimal forgetting over N is $F(2,k) = k+1$: $\mathrm{U}(2)$ *is the group with the cheapest forgetting.* This explains why the whole framework (Sato–Tate, $\mathfrak{sl}_2$, Temperley–Lieb, binary trees) keeps meeting nothing but the simplest $k+1$.

7. Complete asymptotics: c_V and γ_V

7.1. **The mechanism.** Write $s = \#\{\text{positive roots}\} = (\dim G - \operatorname{rank} G)/2$. Let a be defined by $-\log(\chi/\dim V) = a|\vec{\theta}|^2 + \varphi_4 + \cdots$ and let h_d be the lowest-degree homogeneous Taylor component of the Weyl denominator $|\Delta|^2$ at the saddle $\theta = 0$. Then

$$F_k = c_V\, k^s\Big(1 + \frac{\gamma_V}{k} + O(k^{-2})\Big), \qquad c_V = \frac{m_W \cdot 2^s}{\langle h_d\rangle} = \frac{|W|}{2^s\langle h_d\rangle}, \qquad \gamma_V = \frac{\varepsilon_T - \varepsilon_G}{2}, \tag{7.1}$$

where $m_W = |W|/4^s$ is the leading coefficient of $|\Delta|^2$, $\varepsilon_T = -\langle\varphi_4\rangle$, $\varepsilon_G = \langle h_{d+2}\rangle/\langle h_d\rangle$, and $\langle\cdot\rangle$ is the isotropic Gaussian average $\mathcal{N}(0,\sigma^2)$ with $\sigma^2 = 1/(2a)$, $\lambda = 2k$.

The mechanism is uniform: the integrand of the Weyl integral contains $|\Delta|^2$, each positive root contributes a double zero at the saddle, so $|\Delta|^2$ has a zero of order $2s = d$ there:

$$|\Delta|^2 = h_d + h_{d+2} + \cdots .$$

In the Laplace expansion this zero pushes λ^{-s} into the denominator (with $\lambda = 2k$); dividing by χ^{2k} gives the power k^s; the next-order combination $h_{d+2} - \varphi_4$ gives γ_V. Consequently γ_V has nothing to do with dimension combinations: the guess $\gamma = (\dim V)\cdot d/(8s)$ is disproved by the exact SU(2) value $\gamma = 1$ (which gives 0.25) — recorded as correction no. 25.

7.2. c_V **and** γ_V **in numbers.** All four cases are exact rational numbers. For USp(4) in the standard representation $h_8 = 16\theta_1^2\theta_2^2(\theta_1^2 - \theta_2^2)^2$ and $\langle h_8\rangle = 3072$ at $\sigma^2 = 2$, hence $c_V = 8\cdot 16/3072 = 1/24$.

(G,V)	s	m_W	$\langle h_d\rangle$	c_V (analytic)	c_V (fit)	γ_V (analytic/fit)
SU(2)/V_2	1	1/2	1	1	1.000000	1/1.0000
SO(3)/V_3	1	1/2	3/8	8/3	2.666667	0.5625
USp(4)/std	4	1/32	3072	1/24	0.041667	− − −
USp(4)/Ad	4	1/32	625/1728	864/625	1.382409	4.2222/4.2209

Fit and analytic values agree to 0.00%. The two representations of the same group share the power s but differ in γ_V: γ_V is a genuine representation invariant.

7.3. **The convergence criterion** θ_{eff}**.** The integrand $e^{-\lambda a\theta^2}\cdot\theta^{2s}\cdot\theta^{r-1}$ is supported in a ball of radius $\theta_{\text{eff}} = \sqrt{s/(2ka)}$ — which explains the earlier "USp(4) converges slowly" phenomenon: at low k the effective region simply does not lie inside the Taylor expansion zone.

	$k=5$	$k=20$	$k=50$	$k=100$	$k=200$
SU(2)/V_2	0.447	0.224	0.141	0.100	0.071
SO(3)/V_3	0.274	0.137	0.087	0.061	0.043
USp(4)/std	1.265	0.632	0.400	0.283	0.200
USp(4)/Ad	0.816	0.408	0.258	0.183	0.129

At $k = 5$, USp(4)-standard has $\theta_{\text{eff}} = 1.265$ radians, far from the origin; the eighth-order Taylor expansion of $|\Delta|^2$ has $> 100\%$ error there. The Ad representation has larger $a = 3/5$, hence smaller θ_{eff} and visibly faster convergence — exactly matching the measured "Ad has better signal-to-noise".

7.4. **The core dichotomy, fourth appearance.** The complete asymptotics splits the forgetting factor in two, and the attribution of the two halves is exactly the program's core dichotomy:

layer	quantity	attribution
power	$k^{\#\text{positive roots}}$ (group only)	Tate (universal/exponent)
leading constant	$c_V = \lvert W\rvert/(2^s\langle h_d\rangle)$ (group + representation)	cusp (arithmetic/amplitude)
first correction	$\gamma_V = (\varepsilon_T - \varepsilon_G)/2$	cusp (arithmetic/amplitude)

Forgetting changes only the exponent (universal over the group); it never changes the constant layer that carries the fingerprint.

7.5. **The triple unity of** SU(2)**.** SU(2)'s standard representation simultaneously gives $c_V = 1$, $\gamma_V = 1$, #positive roots $= 1$: all three are 1, making $F_k = k+1$ exact. This is the simplest cell among all non-abelian compact Lie groups: SU(2) is the only group with one positive root, hence the cheapest forgetting. The framework (Sato–Tate elliptic curves, $\mathfrak{sl}_2$, Temperley–Lieb, binary trees) all live in this cell — this is why only $k+1$ is ever met.

7.6. **Contour integral and Mellin–Barnes hierarchy for** $U_k(N)$**.** Equation (6.1) yields a one-dimensional contour integral, exact for all N:

$$U_k(N) = \frac{(k!)^2}{2\pi i}\oint_{|z|=r} \frac{I_0(2\sqrt{z})^N\,dz}{z^{k+1}} = \frac{(k!)^2}{2\pi r^k}\int_0^{2\pi} I_0\big(2\sqrt{re^{i\theta}}\big)^N e^{-ik\theta}\,d\theta,$$

$r > 0$ arbitrary; relative difference 10^{-41}–10^{-48} against the combinatorial definition. The Mellin–Barnes hierarchy:

N	$I_0(2\sqrt{z})$ type	MB representation
1	${}_0F_1(;1;z)$ hypergeometric	Hankel contour (c $\in(-\frac14,0)$)
2	${}_1F_2(1/2;1,1;4z)$ (Clausen)	single ${}_1F_2$
≥ 3	not ${}_pF_q$ (Domb numbers A002893)	essentially N-fold

For $N=2$ the Clausen identity reduces the power to a single ${}_1F_2$; residue checks ($s=n$: residue of $\Gamma(-s)$ with Legendre's duplication $\Gamma(n+1/2) = \frac{(2n)!\sqrt{\pi}}{4^n n!}$) give $\frac{(2n)!}{(n!)^4} = U_n(2)/(n!)^2$ for $n = 1,\dots,11$. For $N \geq 3$, $I_0(2\sqrt{z})^N$ is not hypergeometric — the same fact as "the Domb numbers are not hypergeometric" — so the MB representation is genuinely N-fold.

7.7. **The complete asymptotic of** R_N**: Barnes** G **and the Mehta constant.** The saddle at $\sqrt{z_*} = k/N + \frac14$ (from $I_0(2\sqrt{z}) \sim e^{2\sqrt{z}}(4\pi\sqrt{z})^{-1/2}$) and the random-walk geometry (the maximal set $|S| = N$ is a circle, of codimension $N-1$ in T^N) give

$$U_k(N) \sim c_N\, N^{2k}\, k^{-(N-1)/2}, \qquad c_N = \sqrt{N}\left(\frac{N}{4\pi}\right)^{(N-1)/2}.$$

N	c_N (closed)	c_N (numeric, k=150)
2	$1/\sqrt{\pi} = 0.564190$	0.564189
3	0.413497	0.413494
4	0.359174	0.359168
5	0.3540	0.350
6	0.3860	0.380

The closed form even reproduces the non-monotone dip of c_N at $N=5$ — structure no fit can produce. The Weyl side carries an extra $|\Delta|^2$ which vanishes to order $N(N-1) = 2\binom{N}{2}$ on the maximal set (the "zero-order" mechanism), so the powers differ by exactly $\binom{N}{2}$:

$$F(N,k) = \frac{U_k(N)}{W_k(N)} \sim R_N\, k^{\binom{N}{2}}, \qquad R_N = \frac{(2/N)^{\binom{N}{2}}}{\prod_{j=1}^{N-1} j!} = \frac{(2/N)^{\binom{N}{2}}}{G(N+1)}, \tag{7.2}$$

with G the Barnes G-function (hyperfactorial). Numerical anchors:

N	$\binom{N}{2}$	R_N (closed)	numeric anchor
2	1	1	$F = k+1$ exact (anchor)
3	3	$4/27 = 0.148148$	0.148149 (pinned)
4	6	$1/768 = 0.00130208$	0.001303 (+0.07%)
5	10	3.6411×10^{-7}	− − −

Saddle-point derivation, and the regime of (7.2). Write $I_0(2\sqrt{z})^N \approx e^{2N\sqrt{z}}(4\pi\sqrt{z})^{-N/2}$ (valid at the saddle) and apply Cauchy's formula with integrand $e^{2N\sqrt{z}-k\log z}$. The saddle solves $\frac{N}{\sqrt{z}} - \frac{k}{z} = 0$, i.e. $\sqrt{z_*} = k/N$ (the displayed $+1/4$ in the text is the first subleading shift, needed to match the Stirling expansion of $k!$). The leading exponential is e^{2k}, which against

the $(k!)^2$ prefactor of $U_k(N) = (k!)^2[z^k]I_0(2\sqrt{z})^N$ and $(k!)^2/(2k)! \sim \sqrt{\pi k}\,(e/2k)^{2k}$ yields the pure power N^{2k}; the prefactor $(4\pi\sqrt{z_*})^{-N/2}$ contributes $k^{-N/2}$, the Gaussian fluctuation determinant $k^{3/2}$, and the saddle normalisation $1/z_*$ contributes k^{-2}, totalling $k^{-(N-1)/2}$ — the exponent and the constant c_N. *Regime*: both sides of (7.2) are to be read with N fixed and $k \to \infty$, the T-side by the above saddle and the G-side by the rank-N Schur asymptotics (so that the $W_k(N) = k!$ plateau formula, valid only for $k \le N$, is not used); the anchors in the table are direct evaluations and the closed form is thereby identified, in the sense of Remark 7.1. □

Remark 7.1 (The Mehta reading of R_N). $\prod_{j=1}^{N-1} j! = G(N+1)$ is exactly the factor in the Mehta integral (the GUE Gaussian partition-function normalisation):

$$\int_{\mathbb{R}^N} e^{-\sum \lambda_i^2} |\Delta|^2 \, d\lambda = (2\pi)^{N/2} 2^{-N^2/2} \prod_{j=1}^{N} j!.$$

So the forgetting factor's constant acquires a partition-function meaning: R_N is the reciprocal of a Mehta constant. "Forgetting = losing level repulsion (GOE to Poisson)", previously a spectral statement, here receives its first quantitative partition-function form: the cost of switching on the interaction is precisely the change of the partition-function normalisation. *Status*: the identification with the Mehta factor is exact, and the saddle constant is pinned numerically to 0.001% at $N = 2, 3, 4$; the partition-function interpretation itself is not yet a theorem.

Remark 7.2 (Methodology: never extract powers via $\ln F/\ln k$). $\ln F/\ln k = p + \ln R_N/\ln k + O(1/(k \ln k))$ converges at rate $O(1/\log k)$; since $\ln R_N$ is a non-negligible negative number ($N = 3$: $p = 3$, $\ln R_N = -1.91$), extrapolation is impractical (± 0.1 needs $k > 2 \times 10^8$; only $N = 2$ works since $\ln R = 0$). The correct procedure: fix $p = \binom{N}{2}$ and check the stability of F/k^p to R_N — $O(1/k)$ convergence, an order of magnitude faster. Together with the θ_{eff} criterion and the "high heavy-tail moments are not credible" rule, this forms one convergence-diagnostic family: the more complex the structure (higher rank, higher zero order), the thicker the asymptotic correction layer, and the more data is needed.

8. Rank-r vertical moments

Theorem 8.1 (Vertical moments)**.** *Vertical transport (following the Frobenius direction $n \ge 2$ steps) against a rank-r Sato–Tate group gives*

$$\mathrm{Vert}_r(k) = \Big\langle \Big(\sum_{i \le r} 2\cos\theta_i \Big)^{2k} \Big\rangle = \binom{2k}{k} U_k(r),$$

where only the terms with all powers even survive. The generating function is

$$\sum_{k \ge 0} \mathrm{Vert}_r(k) \frac{z^k}{(2k)!} = I_0(2\sqrt{z})^r.$$

Correction of the index. The earlier use of $U_k(2r)$ against $r = 2$ gave $28 \ne 36$ and the mistaken verdict "U_k is a coincidence". The correct object is $U_k(r)$ with r the *rank*, not $2r$. With the correction: $r = 2$, $k = 1$ gives 4... precisely: $\binom{2}{1}U_1(2) = 2 \cdot 2 = 4$, matching USp(4)-standard first vertical moment 4; and $r = 3$, $k = 1$ gives $\binom{2}{1}U_1(3) = 2 \cdot 3 = 6$, matching the measured 6.003 (not the once-predicted $\binom{2k}{k}^r = 8$). The U_k ladder was right all along; the indexing was wrong. □

The memory side has its own closed ladder, the Weyl ladder of Section 3: $H_r(k) = \dim(V^{\otimes 2k})^G = \sum_{\ell(\lambda) \le r} (f^\lambda)^2$ ($= 1, 3, 24, 285, 4242$ for $r = 3$). Hence

$$F_r(k) = \frac{\mathrm{Vert}_r(k)}{H_r(k)} = \frac{\binom{2k}{k} U_k(r)}{\sum_{\ell(\lambda) \le r} (f^\lambda)^2}, \tag{8.1}$$

numerator = multiset-count (torus invariants, no Weyl), denominator = irreducible-count (full-group invariants). Values: $F_1(k) = k + 1$;

r	$F_r(1)$	$F_r(2)$	$F_r(3)$	$F_r(4)$
1	2	3	4	5
2	2	3.6	5.71	8.40
3	2	3.75	6.53	10.5
4	2	3.23	5.66	9.40

and $F_r(1) = 2$ for *every* r — a consistency law at both extremes.

9. Power laws for all compact Lie groups

The general theorem $p = (\dim G - \operatorname{rank} G)/2 = \#\{\text{positive roots}\}$, verified on SU(2), SO(3), USp(4), and the U(N) family, is here pushed to six groups including the first exceptional group G_2.

Theorem 9.1 ($F(1) = \dim V$)**.** *Let V be an irreducible representation of a connected compact Lie group G, all of whose weights have multiplicity 1 and whose zero weight occurs at most once. Then*

$$F(1) = \frac{\dim(V^{\otimes 2})^T}{\dim(V^{\otimes 2})^G} = \dim V.$$

Proof. $V^{\otimes 2}$ has T-weight decomposition $\bigoplus_{w,w'} V_w \otimes V_{w'}$; the zero-weight space is $\bigoplus_w V_w \otimes V_{-w}$, of dimension $\sum_w n_w n_{-w} = \sum_w n_w^2$ (weights are stable under $w \mapsto -w$). Under the hypotheses $n_w \in \{0, 1\}$, so this equals $\#\{\text{weights}\} = \dim V$. The denominator is $\langle \chi_V^2 \rangle_G = \dim \operatorname{End}_G(V) = 1$ by Schur's lemma. □

The table below verifies the prediction on six groups; each row is thus a *zero-parameter consistency check* of both the weight-convolution and Weyl-integration code paths.

group	type	$\dim V$	rank	#pos. roots	$F(1)$ measured	verdict
SU(2) V_2	A_1	2	1	1	2.00	√
U(3) V_3	A_2	3	2	3	3.00	√
SO(5) V_5	B_2	5	2	4	5.00	√
USp(4) V_4	C_2	4	2	4	4.00	√
Sp(6) V_6	C_3	6	3	9	6.00	√
G_2 V_7	G_2	7	2	6	7.00	√

6/6. This zero-parameter prediction simultaneously validates the weight-convolution and Weyl integration code paths.

9.1. **The local slope.** From $F_k = c\,k^p(1 + \gamma/k)$, discrete differencing gives the local slope in closed form, with $O(1/k)$ convergence (per Remark 7.2):

$$\text{slope}(k) = p - \frac{\gamma}{k+1} + \frac{\delta}{(k+1)^2}.$$

group	p	k=24 local slope (pred)	measured	γ^2
A_1 SU(2)	1	0.9608 (= 1 exact)	0.961	1.043
A_2 U(3)	3	2.607	2.935	4.9
B_2 SO(5)	4	3.621	4.596	7.58
C_2 USp(4)	4	3.458	3.621	8.37
C_3 Sp(6)	9	6.48	7.737	21.8
G_2	6	5.003	5.993	12.0

The SU(2) anchor hits digit-perfectly ($\gamma = 1$, $F = k + 1$ exact).

9.2. **The quadratic fit and the rescue of C_3.** A linear fit's intercept is systematically depressed: the omitted $O(1/k^2)$ term is partially absorbed into $-\gamma/(k+1)$, dragging the intercept down by $\propto \gamma/k_{\max}$. Adding $\delta/(k+1)^2$:

group	p true	p (linear)	p (quadratic)	Δp_2	γ^2	verdict
A_1	1	1.0009	1.0000	-0.000	1.00	anchor
A_2	3	2.9842	3.0019	$+0.002$	5.29	$\surd$
B_2	4	3.5612	3.9589	-0.041	8.37	$\surd$
C_2	4	3.9709	3.9967	-0.003	8.5	$\surd$
C_3	9	8.4050	8.8287	-0.171	28.68	rescued
G_2	6	5.8949	5.9931	-0.007	13.53	exceptional, sealed

C_3's residual -0.17 is quantitatively explained: $\gamma^2 \approx 29$, $k_{\max} = 10$, $\gamma/k^2 \approx 0.29$, with higher orders contributing ~ 0.2 — same order as observed; direction consistent, no counterexample. G_2 is sealed: $p_{\text{fit}} = 5.993 \pm 0.007$ (0.1%), rms $= 7.5 \times 10^{-5}$. The power law now covers all types A, B, C, D, G of simple Lie groups, with no counterexample.

Remark 9.2 (Methodology: small rms $\neq$ unbiased parameters). G_2's rms $= 7.5 \times 10^{-5}$ sits two orders below its intercept deviation 0.007: when a model lacks higher-order terms, the lower-order coefficients absorb them, producing systematic bias while the residual stays small. The verdict criterion is not "small residual" but "do the parameters move when one more term is added". (A specimen: B_2 at $k_{\max} = 4$ overshoots to 4.22 in the quadratic fit and returns to 3.96 at $k_{\max} = 10$ — when the parameter count approaches the sample count, low-order coefficients start absorbing noise rather than signal, an instance of the rule itself.)

10. Howe duality and its arithmetic reading: Rankin–Selberg

10.1. **The (λ, λ) pairing identity.**

Theorem 10.1 (Howe-duality pairing)**.**

$$\sum_{\substack{\lambda \vdash m \\ \ell(\lambda) \le \min(N,k)}} \dim S_\lambda(\mathbb{C}^N) \cdot \dim S_\lambda(\mathbb{C}^k) = \binom{Nk+m-1}{m} = \dim \mathrm{Sym}^m(\mathbb{C}^N \otimes \mathbb{C}^k).$$

The pairing is (λ, λ) — the same partition on both sides — hence the identity is explicitly symmetric in $N \leftrightarrow k$. Verified on 24 triples ($m \le 4$, $2 \le N \le k \le 4$), digit by digit.

This replaces the retracted reflection symmetry $U_k(N) = U_N(k)$ (Proposition 6.1): $U_k(N)$ is a degree-k polynomial in N but exponential in k — no symmetry is possible — while the Howe identity is a decomposition of $\dim \mathrm{Sym}^m$ on both sides, with the symmetry carried by the (λ, λ) pairing.

10.2. **Arithmetic reading: the Cauchy kernel is the Rankin–Selberg factor.**

Falsification 10.2 (of the q-Howe conjecture). The conjectured identity with a Gaussian binomial on the right,

$$\sum_\lambda s_\lambda(1, q, \ldots, q^{N-1})\, s_\lambda(1, q, \ldots, q^{k-1}) \stackrel{?}{=} \binom{Nk+m-1}{m}_q,$$

is **false**: at $m = 1$, $N = k = 2$, $q = 2$ the left side is 9 against the predicted 15. The correct form comes from the Cauchy identity

$$\boxed{\sum_{\lambda \vdash m} s_\lambda(x)\, s_\lambda(y) = h_m\big(\{x_i y_j\}\big)}$$

— the right side is h_m of the $N \times k$ rectangle $\{q^{i+j-2}\}$ with multiplicities, not Nk consecutive powers. After the correction, 18 cases pass ($m \le 3$, $N, k \le 3$, $q = 2, 3$).

The arithmetic reading: from $\prod_{i,j}(1-x_iy_jt)^{-1} = \sum_{m\geq 0}\big[\sum_{\lambda\vdash m} s_\lambda(x)s_\lambda(y)\big]t^m$, taking $x = \mathrm{Satake}(\pi)$, $y = \mathrm{Satake}(\pi')$, the Cauchy kernel *is* the Rankin–Selberg local L-factor:

$$(\lambda,\lambda)\text{-pairing sum} = \text{the } m\text{-th Dirichlet coefficient of } L(\pi\times\pi', s).$$

Numerical verdict (4 primes $\times$ $m = 1, 2, 3$, all digit-perfect):

p	$a_p(E_1)$	$a_p(E_2)$	m	Howe pairing sum	RS coefficient
7	-4	3	3	-468	-468
11	2	-2	3	-1780	-1780
13	-5	-4	3	3580	3580
17	4	0	2	306	306

The $N \leftrightarrow k$ symmetry on the arithmetic side is convolution symmetry $L(\pi\times\pi') = L(\pi'\times\pi)$. In the purely a_p form (Weyl character formula verified on 10 cases): the pairing sum depends only on a_p, not on the Satake roots:

$$\sum_{\lambda\vdash m} s_\lambda s_\lambda = \sum_{\substack{d\equiv m\,(2)\\ 0\leq d\leq m}} p^{\frac{m-d}{2}}\, a_p(\mathrm{Sym}^d E_1)\, a_p(\mathrm{Sym}^d E_2).$$

10.3. **Same mechanism, second manifestation.** The vanishing order of $\det(\mathrm{den})$ at $q = 1$ equals $\binom{N}{2}$ — the number of positive roots — the same exponent as the forgetting factor:

N	2	3	4	5
$\det(\mathrm{den})$ vanishing order at $q=1$	1	3	6	10
#positive roots $= \binom{N}{2}$	1	3	6	10
forgetting-factor power $k^{\binom{N}{2}}$	1	3	6	10

Here Δ degenerates at $q = 1$ (the dimension is the regularised limit); there $|\Delta|^2$ degenerates at the saddle (giving $k^{\#\text{positive roots}}$, Section 7). One mechanism, two faces.

11. Shape fingerprints beyond the heavy-tail barrier

11.1. **The Jensen gap and its closed-form dictionary.** The fingerprint $J = \mathbb{E}[Y]\,\mathbb{E}[1/Y]$ ($Y = \mathrm{Per}_p/p^\alpha$ normalised by the object's own α) is scale-free. Closed forms (12 families, 18 Monte-Carlo checks all passing):

law	J	law	J
constant	1	χ^2_ν	$\frac{\nu}{\nu-2}$
Rayleigh(σ)	$\pi/2$	lognormal(σ)	e^{σ^2}
Gamma(k,θ)	$\frac{k}{k-1}$	Weibull(k)	$\frac{\pi/k}{\sin(\pi/k)}$
exponential	∞	Pareto(α)	$\frac{\alpha^2}{\alpha^2-1}$
uniform$[a,b]$	$\frac{(1+r)\ln r}{2(r-1)}$, $r=\frac{b}{a}$		

(Weibull(2) = Rayleigh gives $\frac{\pi/2}{\sin(\pi/2)} = \pi/2$: self-consistent.)

Theorem 11.1 (Structure)**.** *$J = \mathbb{E}[Y]\mathbb{E}[1/Y] \geq 1$ by Cauchy–Schwarz, with equality iff Y is a.s. constant; writing $\delta = (Y-\mu)/\mu$, $J = \mathbb{E}[1/(1+\delta)]$ — a harmonic-type functional of the relative fluctuation.*

If the density $f(y) \sim cy^\beta$ as $y \to 0$, then $\mathbb{E}[1/Y] < \infty$ iff $\beta > 0$. For the Weibull family $\beta = k - 1$: $\beta = 0$ (Poisson type) $\Rightarrow J = \infty$; $\beta = 1$ (GOE/Rayleigh) $\Rightarrow J = \pi/2 = 1.5708$; $\beta = 2$ (GUE type) $\Rightarrow J = 1.2092$. This is the quantitative reading of "forgetting = losing level repulsion = GOE $\to$ Poisson": repulsion disappears ($\beta : 1 \to 0$) and $\pi/2 \to \infty$.

Arithmetic measurements: (1) generic $z^2 + c$ ($\alpha = 0.5$): ten values $J \in [1.39, 1.62]$, mean 1.4866, prediction $\pi/2 = 1.5708$ — a hit (at $\beta = 1$ the variance of $\mathbb{E}[1/Y]$ itself diverges logarithmically, so 5% fluctuation over 430 primes is expected). (2) $c = -2$ (Chebyshev, $\alpha = 1$): $J = 1.02237$ against 1.0174 (0.5%) — and the same c at $\alpha = 0.5$ falls back to

1.484, rejoining the generic class: the sensitivity of J to α is itself evidence that near-determinism only shows under the correct normalisation. (3) $|a_p|/\sqrt{p}$ (elliptic curves): J diverges logarithmically, $\mathbb{E}[1/Y] = \frac{1}{\pi}\ln p_{\max}$ (the Jacobian factor 1/2 corrected in erratum); measured at $p \le 1000$: 2.0424 (prediction 2.1979), $p \le 3000$: 2.4579 (prediction 2.5484) — both at 94% of prediction and growing logarithmically. $J = 2.09 \gg 1.5$ already detects "the zero-density at 0 is nonzero".

11.2. **The second fingerprint $W = \mathbb{E}[Y^2]/\mathbb{E}[Y]^2$.** Motivation: $J = \mathbb{E}[Y]\mathbb{E}[1/Y]$ diverges logarithmically on distributions with $f(0) > 0$ ($\beta = 0$, Poisson type) and is unmeasurable there. W only involves second moments, stays finite on those families, and bypasses the heavy-tail obstacle. Closed forms (14 entries, Monte-Carlo 3×10^6 all passing):

law	W	law	W
Rayleigh(σ)	$4/\pi$	Gamma(k,θ)	$\frac{k+1}{k}$
χ^2_ν	$\frac{\nu+2}{\nu}$	lognormal(σ)	e^{σ^2}
exponential	2	Pareto(α)	$\frac{(\alpha-1)^2}{\alpha(\alpha-2)}$

(the exponential has $J = \infty$ but $W = 2$ — the value of the variable-weight scheme). Arithmetic: $Y = |a_p|/\sqrt{p}$ (Sato–Tate): $W = 1.3759$ against $9\pi^2/64 = 1.38791$ (0.9%).

12. The $\tau \leftrightarrow m$ dictionary at finite N

The formal dictionary (moments are Taylor coefficients of the spectral form factor, $M_m = i^{-m}\phi^{(m)}(0)$, $M_{2n} = \mathrm{Cat}(n)$ under the semicircle) had one residual cell: at finite N, mean $K_N = 1/N$ is simultaneously the finite-N floor and the GOE plateau, so criterion (v) "never saturates" could not be judged. Resolution: the *recurrence peak*, not the mean, is the criterion.

Theorem 12.1 (Recurrence-peak criterion)**.** *For a discrete spectrum $K_N(\tau) = |\frac{1}{N}\sum_j e^{iu_j\tau}|^2$ is almost periodic with time average $1/N$ (pure diagonal, true of* any *discrete spectrum). Set*

$$P_N := \max_{\tau \approx 2\pi} K_N(\tau) \sim N^{-\beta}; \qquad \beta \approx 0 \Rightarrow \textit{saturation}, \quad \beta > 0 \Rightarrow \textit{never saturates}.$$

The mean is not a criterion: at $N = 800$ the values GOE 0.00092*, Poisson* 0.00126*, equispaced* 0.00110 *are all near $1/N = 0.00125$.*

Theorem 12.2 (Non-interchangeability of limits)**.**

order of limits	$\beta > 0$	$\beta = 0$ (*perfectly periodic*)
$\lim_N[\limsup_\tau K_N] = \lim_N P_N$	0	1
$\lim_\tau \lvert\lim_N K_N\rvert = \lvert\phi_{\mathrm{cont}}\rvert^2$	0	0
interchangeable?	*yes*	***no***

Numerically: the equispaced spectrum still has $\max K = 1.000000$ *at $\tau/\pi \in [20, 40]$ (the limsup does not decay), while the continuous semicircle's $|\phi|^2$ has dropped to 2×10^{-9} at $\tau = 200$. The "finite-N lattice" is therefore not a defect but a question of limit order — its precise answer. The finite-N dictionary completes the strict layer: $\tau_H \sim 2\pi/N$, $m_{\max} \sim N \sim$ (peak width)$^{-1}$.*

Verdict: equispaced $\beta = 0.028$ with peaks pinned at 2π multiples (saturation); GOE $\beta = 0.759$, Poisson $\beta = 0.823$, Maass (1110 levels) $\beta = 0.819$ with random peak positions (never saturates) — the two regimes separate by a factor 27 (0.03 vs 0.8), Maass versus equispaced by 30. Maass peak heights stay within 0.8–1.4× the GOE baseline, with no fixed peak position: criterion (v) holds on the true arithmetic spectrum.

Remark 12.3 (Honest labels). (1) Finite size compresses both ways: equispaced theory $\beta = 0$ measures 0.028 (raised), GOE/Poisson theory 1 measures 0.76–0.82 (lowered); the 27× separation is a conservative lower bound. (2) $\beta_{\mathrm{Maass}} = 0.819$ should be quoted as $\beta \gtrsim 0.5$: the $\beta_M - \beta_G$ difference is driven almost entirely by the single point $N = 1110$, and the $N = 300/500/800/1110$ values are nested subsets (not independent samples) while the GOE

column is regenerated independently per N — unequal noise structures. The direction ($\beta > 0$, clearly distinct from $\beta \approx 0$) is unaffected. (3) The equispaced peak-position reading 4.000 was a scanning endpoint landing exactly on the recurrence $\tau = 4\pi$ — constructive; the true criterion is "fixed peak position: yes/no".

13. Unification: three layers, one Catalan

The forgetting law has now been quantified on three layers, and all three are the same identity $\mathrm{Cat}(k) = \binom{2k}{k}/(k+1)$ in different clothes:

layer	forgetting factor	observable	direction
coefficient	$\times 1/\mathrm{Cat}(m/2)$	weight	divide
correlation	$\times 1/\mathrm{Cat}(m)$	weight	divide
spectral (this paper)	$\times (k+1) = \binom{2k}{k}/\mathrm{Cat}(k)$	moment	multiply

The different directions reflect only the observables: the first two layers measure weights, the spectral layer measures moments. The common kernel: *Catalan = structure; forgetting = drifting away from Catalan.*

The fifth and sixth formulations of the forgetting law live here: forgetting = losing level repulsion (the Weyl Vandermonde), and the amplitude formulas of Section 7 show the fourth appearance of the core dichotomy — the universal exponent (group) against the fingerprint constants (representation).

Conclusion

The law of forgetting, encountered at three layers of the seven-world program, is reduced here to invariant theory: moments are invariant-tensor dimensions (Theorem 1.1); forgetting is restriction from G to a maximal torus (Definition 1.2); the forgetting factor obeys the rank law $F_k = c_V k^s(1 + \gamma_V/k + \cdots)$ with $s = \#\{\text{positive roots}\}$ and explicit constants, verified across six compact Lie groups with zero free parameters (Section 9); and the universal $\mathrm{U}(N)$ family closes the constants through the Barnes G-function and the Mehta normalisation (equation (7.2)). On the arithmetic side the same identity reads Schur's lemma off two elliptic curves (Theorem 4.2) and the Cauchy identity as Rankin–Selberg (Section 10). Everything numerical carries a status label (Remarks 2.1, 7.1, 12.3); everything withdrawn is recorded with its replacement in Appendix A.

Appendix A. Corrections applied in this paper

This paper incorporates, in corrected form, the following entries from the compendium's revision log (the compendium records 69 entries; only those touching this paper's material are listed). Each entry was found by the project's own checks; the countermeasures are part of the method.

#	Error (as made)	Correction (as used here)
20	$F(N,3)$ at N=5 written 88.5	$545/6 = 90.8333$ (F is an exact cubic for $k \geq 3$)
21	$W_3(3), W_4(3)$ written 2762/15793	2761/15767 (A005802); 2762/15793 are A061552 (another Wilf class), misplanted from memory
22	Claimed U(3) moments are six-fold Bessel moments	Retracted: one-dimensional Bessel moments cover only N=2; $N \geq 3$ is an $(N-1)$-dimensional torus integral; the correct Bessel object is the GF $I_0(2\sqrt{z})^N$
23	$\langle\chi^2\rangle$ = irreducible-component count; $E \times E$ read as 2 not 4	$\sum_\mu m_\mu^2 = \dim \mathrm{End}_G(V)$; $E \times E$ gives 4 (Corollary 4.1)

#	Error (as made)	Correction (as used here)
24	$E \times E$ high moments blamed on arcsine endpoint divergence	Endpoints vanish ($\sim (4-\lvert\chi\rvert)^{3/2}$); the cause is the heavy tail
25	$c_V = 1/384$ for USp(4)-std	$1/24$; $h_8 = 16\theta_1^2\theta_2^2(\theta_1^2-\theta_2^2)^2$, $\langle h_8\rangle = 3072$ (small-angle mis-multiplication 4^2, $16\times$)
26	$R_4 = 1/384$, $R_5 = 3.5556\times 10^{-7}$	$R_4 = 1/768$, $R_5 = 3.6411\times 10^{-7}$ (caught by the c_N/c'_N purely algebraic self-consistency check)
27	Reflection symmetry $U_k(N) = U_N(k)$, "36 entries all hit"	False: $U_2(3) = 15 \neq 20 = U_3(2)$ (Prop. 6.1); the "verification" copied one row twice; the true symmetry is Howe duality
28	"U_k is only an N=2 coincidence"; used $U_k(2r)$ vs r=2	Correct object $U_k(r)$, r = rank not $2r$ (Thm. 8.1); r=2, k=1 $\to$ 4, r=3, k=1 $\to$ 6 (measured 6.003)
29	Valley at $n_{\text{peak}} \approx 2\log_2 p$; valley floor = Per_p	Both falsified: measured valley does not obey $2\log_2 p$; the same closed form substituted back gives $n_{\text{peak}} = p/\sqrt{3} = O(p)$ (self-contradictory from the start); $\Phi \sim$ 0.2–0.3 vs $\text{Per}_p \sim \sqrt{p}$
30	"Power law and essential singularity form a continuous family"	Refined: two branches of one quadrature $N = \int \frac{dv}{\eta v+v^2+\Delta}$, but two geometries (perturbation in initial condition vs in constant term); rupture only in two-parameter models at disc = 0
40	USp(4)-Ad T-invariants read $12, 616, 42024, \ldots$	Probability version (return probability $[y^0](\text{poly}/\lvert W\rvert)^{2k}$); the counting version multiplies by $\lvert W\rvert^{2k} = 10^{2k}$ (k=1: 1200 not 12). Both conventions valid but must be used consistently.
46	"Mehta lemma" $E[\Delta^2] = (2\sigma^2)^{\binom{N}{2}}\prod j!$	The $2^{\binom{N}{2}}$ factor is spurious: $E[\Delta^2] = (\sigma^2)^{\binom{N}{2}}\prod_{j=1}^{N} j!$ (Andreief); Monte-Carlo adjudication (N=3: 11.9291 ± 0.0449, 1.6σ from 12, 1870σ from 96)
48	q-Howe with Gaussian binomial RHS	Falsified ($9 \neq 15$ at m=1, N=k=2, q=2); correct form is the Cauchy identity h_m of the rectangle (Corollary 10.2)

Standing countermeasures (extracted from the log): never trust $\ln F/\ln k$ for power extraction (Remark 7.2); small rms $\neq$ unbiased parameters (Remark 9.2); heavy-tail moments above the second are not credible (Remark 2.1); high moments of χ are dominated by the $\lvert x\rvert \approx 4$ tail; never use AFE-based self-checks to adjudicate functional-equation parameters.

References

[1] P. Diaconis, M. Shahshahani, *On the eigenvalues of random matrices*, J. Appl. Probab. **31A** (1994), 49–62.

[2] E. Rains, *High powers of random elements of compact Lie groups*, Probab. Theory Related Fields **107** (1997), 219–241.

[3] M. L. Mehta, *Random Matrices*, 3rd ed., Elsevier, 2004.

[4] R. Howe, *Perspectives on invariant theory: Schur duality, multiplicity-free actions and beyond*, Israel Math. Conf. Proc. **8** (1995), 1–182.
[5] T. Clausen, *Über die Fälle, wenn die Reihe* $y = 1 + \frac{\alpha \cdot \beta}{1 \cdot \gamma} x + \cdots$ *ein Quadrat von der Form* $y = 1 + \frac{\alpha' \beta' \gamma'}{1 \cdot \delta' \epsilon'} x + \cdots$ *hat*, J. Reine Angew. Math. **3** (1828), 89–91.
[6] G. Domb, *On the theory of cooperative phenomena in crystals*, Adv. Phys. **9** (1960), 149–361.
[7] G. Césnavicijus, and BSD reference literature: J. Silverman, *The Arithmetic of Elliptic Curves*, 2nd ed., Springer, 2009.
[8] J. H. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves*, Springer, 1994 (Sato–Tate conjecture, moments).
[9] H. Kesten, *Symmetric random walks on groups*, Trans. Amer. Math. Soc. **92** (1959), 336–354; B. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra Appl. **40** (1981), 203–216.