Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 130 (2026)

Optimal Bitwise Pseudorandom Generators for Modular Sums Modulo Every Fixed Integer

Gonen Krak[†]

Abstract

We study pseudorandom bit generators for Boolean linear sums modulo a fixed integer M. A distribution $X \in \{0,1\}^n$ ε-fools these tests if, for every $a \in \mathbb{Z}_M^n$, the distribution of

$$\sum_{i=1}^{n} a_i X_i \pmod{M}$$

is ε-close in statistical distance to the corresponding distribution under independent uniform bits.

Lovett, Reingold, Trevisan, and Vadhan constructed generators with optimal seed length for every fixed prime-power modulus. We remove the restriction to prime powers: for *every* fixed modulus $M \geq 2$ we give an explicit generator with seed length

$$O_M\left(\log n + \log \frac{1}{\varepsilon}\right),$$

which is optimal up to the constant depending on M.

The proof rests on decoupling two parameters that previous arguments had tied together: the prime alphabet over which the pseudorandom source is defined, and the modulus of the test. We work with *local phase classes*, which are closed under arbitrary reparameterisation of each coordinate, and we show that the construction can move downward between prime alphabets. Two operations suffice: over the alphabet $\mathbb{F}_p$, taking p multiplicative derivatives removes one factor of p from the modulus, which lets a sequence of p small-bias masks insert that factor; and any surjection $\mathbb{F}_r \to \mathbb{F}_p$, combined with one mask, transports a generator to a smaller alphabet. Walking the prime factors of M in decreasing order of alphabet assembles the generator, ending over $\mathbb{F}_2$. The only imported pseudorandomness primitive is explicit small bias over fixed prime fields.

[†]Tel Aviv University. gonenok10@gmail.com

ISSN 1433-8092

Contents

1 Introduction

A distribution over $\{0,1\}^n$ is *pseudorandom* for a class of tests if no test in the class can distinguish it from n independent uniform bits. The smaller the class, the shorter the seed one may hope for, and for the simplest nontrivial class — linear tests modulo 2 — the problem was solved in the seminal work of Naor and Naor [NN93]: *small-bias generators* fool all parities with error ε using a seed of length $O(\log n + \log \frac{1}{\varepsilon})$, which is optimal up to constant factors [AGHP92]. Small-bias generators have since become a basic primitive of pseudorandomness, and a natural next question is what happens when the modulus 2 is replaced by an arbitrary fixed modulus M.

Concretely, for a coefficient vector $a \in \mathbb{Z}_M^n$ consider the modular sum

$$S_a^{(M)}(x) = \sum_{i=1}^{n} a_i x_i \pmod{M}, \qquad x \in \{0,1\}^n.$$

A bit distribution X *ε-fools Boolean linear sums modulo M* if, for every $a \in \mathbb{Z}_M^n$, the distribution of $S_a^{(M)}(X)$ is ε-close in statistical distance to that of $S_a^{(M)}(U_n)$. Modular sums are among the simplest tests computable in logarithmic space; they are exactly the read-once products over the cyclic group $\mathbb{Z}_M$, that is, tests of the form $\prod_i g_i^{x_i}$ with $g_i \in \mathbb{Z}_M$; and fooling them is a necessary step in several programs on pseudorandomness for group products and read-once branching programs.

Lovett, Reingold, Trevisan, and Vadhan [LRTV09] constructed generators with the optimal seed length $O_M(\log n + \log \frac{1}{\varepsilon})$ for every fixed *prime-power* modulus $M = p^k$. Their second construction works for every M, at seed length

$$O\left(\log n + \log \frac{M}{\varepsilon} \cdot \log\left(M \log \frac{1}{\varepsilon}\right)\right),$$

which for fixed M is $O_M\big(\log n + \log \frac{1}{\varepsilon} \cdot \log\log \frac{1}{\varepsilon}\big)$: optimal in n, but short of optimal by a $\log\log \frac{1}{\varepsilon}$ factor. In the years since, the composite case has resisted matching the prime-power one. The recent work of Lee and Viola [LV25], which achieves optimal seed length $c_G \log(n/\varepsilon)$ for read-once group products over every p-group, explicitly identifies the handling of composite moduli as a challenge reminiscent of a known roadblock in circuit complexity. In particular, prior to this work no generator with optimal seed length was known even for $M = 6$, the smallest modulus with two distinct prime factors.

1.1 Our result

We remove the restriction to prime powers entirely.

Theorem 1.1 (Main result; informal)**.** *For every fixed integer $M \geq 2$ there is an explicit generator that ε-fools all Boolean linear sums modulo M with seed length*

$$O_M\left(\log n + \log \frac{1}{\varepsilon}\right).$$

The formal statement is Theorem 5.4. The seed length is optimal up to the constant depending on M: any generator fooling linear sums modulo M also fools them modulo any prime divisor of M (Proposition 5.5), where the bound $\Omega(\log n + \log \frac{1}{\varepsilon})$ is standard.

The generator is built from off-the-shelf components. The only pseudorandomness primitive imported is explicit small bias over fixed prime fields [NN93], [AMN98]; everything else, including the base case, is proved here. What the construction does with those components is where the content lies.

Appendix and AI-use disclosure. The appendix records the complete construction route and explicit parameter calculations for (M=15); these calculations were performed and cross-checked with the assistance of ChatGPT and Claude, with the author retaining responsibility for their final accuracy.

1.2 Related work

Modular sums. Naor and Naor [NN93] solved the case $M = 2$ optimally. Lovett, Reingold, Trevisan, and Vadhan [LRTV09] gave the two constructions displayed above. Gopalan, Kane, and Meka [GKM18] took a different route, constructing a single generator that fools the discrete Fourier transforms of linear functions and thereby all modular tests with modulus up to M simultaneously, with seed length

$$O\left(\log \frac{Mn}{\varepsilon} \cdot \left(\log\log \frac{Mn}{\varepsilon}\right)^2\right).$$

The strength of their result is its uniformity: one generator for every modulus up to M, with near-logarithmic dependence on M. For a fixed modulus it does not improve on [LRTV09]. What Theorem 1.1 does is remove the residual $\log\log \frac{1}{\varepsilon}$ loss of the LRTV arbitrary-modulus construction, for every fixed modulus rather than for prime powers alone.

Group products and branching programs. Linear sums modulo M are read-once products over the cyclic group $\mathbb{Z}_M$, a commutative special case of read-once permutation branching programs. For group products, Meka and Zuckerman [MZ09] gave generators over S_3 with polynomially small error at seed length $O(\log n)$, handling the modulus-3 part through generators for low-degree polynomials over $\mathbb{F}_3$. Lee and Viola [LV25] recently achieved optimal seed length $c_G \log(n/\varepsilon)$ for read-once group products over every p-group, and nearly optimal seed length over every commutative group; they point to composite moduli as the boundary of current techniques. Theorem 1.1 gives the exactly optimal seed length for every cyclic group $\mathbb{Z}_M$, including those that are not p-groups. By a standard character argument it should also yield the corresponding statement for every finite abelian group, with the error adjusted by a constant depending on the group; we do not pursue that here.

The Meka–Zuckerman conjecture. For $M = 6$ our generator is not new as an object. The construction ends with three masks over $\mathbb{F}_2^n$. Grouping the first of them — the one contributed by $\mathrm{TRANSFER}_{3\to 2}$ — with the image under $\sigma_{3\to 2}$ of a sum of small-bias sources over $\mathbb{F}_3^n$ gives a

bit distribution that fools all linear sums modulo 3, which is a pseudorandom bit generator for $\mathbb{F}_3$ in the sense of [MZ09]; the remaining two masks, contributed by LIFT_2, collapse into a single small-bias mask. The generator is therefore the XOR of a small-bias space with a pseudorandom bit generator for $\mathbb{F}_3$, which is exactly the object that Meka and Zuckerman conjectured to fool width-3 read-once permutation branching programs (Conjecture 1.5 of [MZ09]). That conjecture remains open. Theorem 1.1 proves that this object fools all linear sums modulo 6. These are not a subclass of the conjectured class — S_3 has no element of order 6, so $\mathbb{Z}_6$ is not a subgroup of S_3 — but they are its untwisted abelian analogue: $\mathbb{Z}_6 \cong \mathbb{Z}_3 \times \mathbb{Z}_2$ where $S_3 \cong \mathbb{Z}_3 \rtimes \mathbb{Z}_2$. In an S_3 program the $\mathbb{Z}_2$ prefix acts on the $\mathbb{Z}_3$ component, twisting the modulus-3 coefficients by a prefix of the input, and it is exactly this twisting that our analysis does not address. We note also that the collapse of the masks describes the object but does not supply the hypothesis: the analysis needs the individual masks, because each level of the argument uses what the source below it already fools.

Relation to degree descent over $\mathbb{F}_2$. Our technique resembles a line of work on pseudorandom generators for low-degree polynomials [BV10, Lov09, Vio09], and the comparison locates what is new here. Those works descend in *degree*: Cauchy–Schwarz introduces a directional derivative, the derivative has degree $d-1$, and the induction finishes from the small-bias base case. Our Lemma 3.1 runs the same skeleton, but the descent parameter is the *modulus* rather than the degree, and — this is the difference that matters — the alphabet over which the derivative is taken is not fixed in advance. The resemblance is closest when M is a power of two: by Lucas' theorem a sum modulo 2^t is a function of $\mathbb{F}_2$-polynomials of degree at most 2^{t-1}, so on that modulus the two descents act on the same object. No analogous bounded-degree representation, with degree controlled by the modulus alone, is available once an odd prime divides M: by the Razborov–Smolensky method [Smo87], counting modulo an odd prime admits no low-degree approximation over $\mathbb{F}_2$. This is one way to see why the Boolean alphabet is the wrong place to remove an odd prime, which is the subject of the next subsection.

1.3 The new idea: matching the alphabet to the prime factor

Over the Boolean alphabet, a difference of two shifts sees divisibility by two. Explicitly, for bits z, b, b',

$$(z \oplus b) - (z \oplus b') = (b - b')(1 - 2z),$$

and the factor 2 on the right is what converts a modulo-$2m$ phase into a modulo-m phase. Iterating strips the powers of two out of the modulus one at a time. But once the modulus is odd, multiplication by 2 is invertible and nothing further happens: differencing over $\mathbb{F}_2$ removes exactly the power of two dividing M and then halts at the odd part. We prove this as Corollary 3.6; it is a special case of the sharper statement that over a fixed alphabet $\mathbb{F}_p$, differencing cannot touch the part of the modulus coprime to p, however many derivatives are taken (Proposition 3.5).

The escape is to stop insisting on the Boolean alphabet. What the identity above expresses is a general fact, and it is a fact about $\mathbb{F}_2$ only because that is the alphabet the shifts live in:

$\mathbb{F}_2$-differences reveal 2-divisibility;	$\mathbb{F}_p$-differences reveal p-divisibility.

The general form is a statement in the integral group ring $\mathbb{Z}[C_p]$, where a derivative contributes a difference of two shifts $T^u - T^v$ and p of them contribute a product lying in $p\,\mathbb{Z}[C_p]$, because $(T-1)^p \in p\,\mathbb{Z}[C_p]$ (Theorem 3.2). So to remove a factor p from the modulus, one works over the alphabet $\mathbb{F}_p$, where p derivatives produce the required factor.

This leaves the difficulty that a source over $\mathbb{F}_p^n$ is not a source of bits, and that the several primes dividing M call for several different alphabets. Both are resolved by the class of tests we work with. A *local phase* is a product $\prod_i \omega_m^{q_i(x_i)}$ of one-coordinate factors with *arbitrary* local exponent maps q_i, and this class is closed under reparameterising each coordinate by any map whatsoever. In particular, if $\sigma : \mathbb{F}_r \to \mathbb{F}_p$ is any surjection, every test over the smaller alphabet pulls back to a test over the larger one. Combined with one small-bias mask, which repairs the fact that σ cannot push the uniform distribution to the uniform distribution, this transports a generator from one alphabet down to the next (Theorem 4.6). No low-degree or algebraic structure is required of σ; a lookup table suffices.

The construction is then a walk through pairs (alphabet, modulus). Two moves are available: LIFT_p multiplies the modulus by p while fixing the alphabet at p, and $\mathrm{TRANSFER}_{r\to p}$ lowers the alphabet while fixing the modulus. Since a prime can be inserted only over its own alphabet, and a surjection only ever goes to a smaller field, the prime factors of M must be visited in decreasing order, with the walk ending over $\mathbb{F}_2$. For $M = 15$ the whole proof is the ladder

$$\begin{aligned}\mathsf{Local}_n(5,1) &\xrightarrow{\mathrm{LIFT}_5} \mathsf{Local}_n(5,5) \xrightarrow{\mathrm{TRANSFER}_{5\to 3}} \mathsf{Local}_n(3,5)\\ &\xrightarrow{\mathrm{LIFT}_3} \mathsf{Local}_n(3,15) \xrightarrow{\mathrm{TRANSFER}_{3\to 2}} \mathsf{Local}_n(2,15),\end{aligned}$$

which begins at a trivial class requiring no seed and ends at the Boolean modular sums.

Organization. Section 2 fixes the objects, states the two moves, and displays the proof plan on the factorisation of a single modulus. Section 3 proves the two engines: an analytic lemma converting control of a derivative class into control of the class itself, and the group-ring divisibility theorem behind the modulus descent. Section 4 assembles the two moves. Section 5 runs the induction through the prime factors of M, converts the resulting phase estimates into statistical distance, and accounts for the seed. Appendix A instantiates every parameter for $M = 15$.

2 The problem and the proof roadmap

This section fixes the objects, states the two construction moves, and displays the whole argument on the prime factorisation of a single modulus. Nothing here is proved; the purpose is that a reader who stops at the end of §2.3 already knows the shape of the endgame, and that a reader who continues knows at every later point which of the two moves is being performed.

2.1 Modular sums and phase functions

Throughout, n denotes the output length and U_n the uniform distribution on $\{0,1\}^n$. For random variables X, Y taking values in a finite set Ω, the statistical distance is

$$\mathrm{dist}(X,Y) := \frac{1}{2}\sum_{z\in\Omega}\big|\Pr[X=z]-\Pr[Y=z]\big|.$$

Fix an integer $M \geq 2$. For a coefficient vector $a \in \mathbb{Z}_M^n$ consider the modular sum

$$S_a^{(M)}(x) := \sum_{i=1}^{n} a_i x_i \pmod{M}, \qquad x \in \{0,1\}^n.$$

A bit distribution X *ε-fools Boolean linear sums modulo M* if

$$\mathrm{dist}\big(S_a^{(M)}(X),\, S_a^{(M)}(U_n)\big) \leq \varepsilon \qquad \text{for every } a \in \mathbb{Z}_M^n.$$

Let

$$\omega_M := e^{2\pi i/M}.$$

The map $r \mapsto \omega_M^r$ translates a residue into a point on the complex unit circle, and is useful because

$$u \equiv v \pmod{M} \quad\Longleftrightarrow\quad \omega_M^u = \omega_M^v, \qquad \omega_M^{u+v} = \omega_M^u \omega_M^v.$$

For a fixed coefficient vector a, the associated *phase function* is

$$\boxed{\phi_a^{(M)}(x) := \omega_M^{S_a^{(M)}(x)} = \omega_M^{\sum_i a_i x_i}, \qquad x \in \mathbb{F}_2^n.}$$

Object type: the phase is a function, not a distribution. The map $\phi_a^{(M)} : \mathbb{F}_2^n \to \mathbb{C}$ is a general complex-valued function. Its values have modulus 1; they need not be nonnegative and do not sum to 1. It is therefore not a probability distribution. Passing from residues to phases turns modular arithmetic into multiplicative algebra, and turns balance of the residues into cancellation of a complex expectation.

The one feature of $\phi_a^{(M)}$ that drives everything below is that it *factorises across coordinates*:

$$\phi_a^{(M)}(x) = \prod_{i=1}^{n} \omega_M^{a_i x_i}.$$

Each factor depends on a single input coordinate. The class of tests we shall actually fool is obtained by keeping this product structure and relaxing everything else: the exponent $a_i x_i$ is replaced by an arbitrary function of x_i, and the input alphabet $\{0,1\}$ is replaced by an arbitrary prime field.

2.2 Local phases, error, and derivatives

Let p be a prime and $m \geq 1$ an integer. For the degenerate modulus we use the convention

$$\mathbb{Z}_1 = \{0\}, \qquad \omega_1 = 1,$$

so that a phase of modulus 1 is the constant function 1. This convention is what makes the starting state of the construction meaningful.

Definition 2.1 (Local phase class)**.** *For a tuple* $\mathbf{q} = (q_1, \ldots, q_n)$ *of* local exponent maps

$$q_i : \mathbb{F}_p \longrightarrow \mathbb{Z}_m,$$

define the local phase

$$\Phi_{\mathbf{q}}(x) := \omega_m^{\sum_{i=1}^n q_i(x_i)}, \qquad x \in \mathbb{F}_p^n.$$

The class of all such phases, as $\mathbf{q}$ *ranges over all tuples of local exponent maps, is denoted*

$$\boxed{\mathsf{Local}_n(p, m)}$$

For $m = 1$ *the class contains only the constant function* 1.

The two parameters play genuinely different roles, and the notation is chosen to keep them visible at all times: p is the *alphabet*, the field over which the input lives, and m is the *modulus*, the order of the root of unity in which the test takes its values. No relation between p and m is assumed. They may be equal, coprime, or share some factors.

Boolean modular sums are the special case in which the alphabet is 2 and the local exponent maps are the simplest nonconstant ones.

Remark 2.2 (Modular sums are local phases)**.** *Given* $a \in \mathbb{Z}_M^n$*, set* $q_i(0) := 0$ *and* $q_i(1) := a_i$*. Then*

$$\Phi_{\mathbf{q}}(x) = \omega_M^{\sum_i a_i x_i} = \phi_a^{(M)}(x),$$

so $\phi_a^{(M)} \in \mathsf{Local}_n(2, M)$*. Fooling* $\mathsf{Local}_n(2, M)$ *therefore suffices for the main theorem, and the whole paper is devoted to constructing a source over* $\mathbb{F}_2^n$ *that does so.*

Definition 2.3 (Error against a class)**.** *For a distribution* X *on* $\mathbb{F}_p^n$ *and a class* $\mathcal{C}$ *of functions* $\mathbb{F}_p^n \to \mathbb{C}$*, set*

$$\boxed{\mathrm{Err}_X(\mathcal{C}) := \sup_{f \in \mathcal{C}} \Big| \mathbb{E} f(X) - \mathbb{E} f(U_p^n) \Big|,}$$

where U_p^n *is uniform on* $\mathbb{F}_p^n$*. We say that* X ε-fools $\mathcal{C}$ *when* $\mathrm{Err}_X(\mathcal{C}) \leq \varepsilon$.

A single functional therefore measures progress at every stage of the construction, over every alphabet and every modulus. If $\mathcal{C}' \subseteq \mathcal{C}$ then $\mathrm{Err}_X(\mathcal{C}') \leq \mathrm{Err}_X(\mathcal{C})$, a triviality that is used constantly: fooling a class is inherited by all of its subclasses.

Definition 2.4 (Active coordinates and weight)**.** *Coordinate* i *is* active *for* $\mathbf{q}$ *if the local exponent map* q_i *is nonconstant. The* weight $\mathrm{wt}(\mathbf{q})$ *is the number of active coordinates.*

For the Boolean modular sum of Remark 2.2, coordinate i is active exactly when $a_i \neq 0$, so $\mathrm{wt}(\mathbf{q})$ agrees with the usual weight of the coefficient vector. In general, however, the definition is the one above: nonconstancy of a map, not nonvanishing of a coefficient.

Finally we record the differencing operation on which the whole argument turns.

Definition 2.5 (Multiplicative derivative)**.** *For a unit-modulus $f : \mathbb{F}_p^n \to \mathbb{C}$, meaning $|f(x)| = 1$ for every x, and shifts $u, v \in \mathbb{F}_p^n$, set*

$$\boxed{D_{u,v}f(x) := f(x+u)\,\overline{f(x+v)}.}$$

For a class $\mathcal{C}$ of such functions define the derivative class

$$\partial\mathcal{C} := \big\{D_{u,v}f \;:\; f \in \mathcal{C},\; u, v \in \mathbb{F}_p^n\big\},$$

and iterate: $\partial^0\mathcal{C} := \mathcal{C}$ and $\partial^{j+1}\mathcal{C} := \partial(\partial^j\mathcal{C})$.

Since $|f| = 1$, conjugation is inversion, so $D_{u,v}$ subtracts exponents: if $f = \omega_m^P$ then $D_{u,v}f = \omega_m^{P(\cdot+u)-P(\cdot+v)}$. Applied to a local phase this acts coordinatewise,

$$D_{u,v}\Phi_{\mathbf{q}} = \Phi_{\mathbf{q}'}, \qquad q_i'(z) = q_i(z+u_i) - q_i(z+v_i),$$

so $\partial\mathsf{Local}_n(p,m) \subseteq \mathsf{Local}_n(p,m)$: the class is closed under differencing. Two features of $\mathsf{Local}_n(p,m)$ will be used repeatedly, and it is worth naming them now, because between them they are the reason the argument works at all.

The two closure properties of local phase classes. **Closure under differencing.** As just observed, a multiplicative derivative of a local phase is again a local phase over the same alphabet and the same modulus. This is what allows the derivative hierarchy $\partial^j\mathsf{Local}_n(p,m)$ to be formed and iterated.

Closure under local reparameterisation. If $\tau : \mathbb{F}_r \to \mathbb{F}_p$ is any map whatsoever and $\Phi_{\mathbf{q}} \in \mathsf{Local}_n(p,m)$, then

$$y \longmapsto \Phi_{\mathbf{q}}\big(\tau(y_1), \ldots, \tau(y_n)\big) = \omega_m^{\sum_i q_i(\tau(y_i))}$$

lies in $\mathsf{Local}_n(r,m)$, because $q_i \circ \tau$ is again a map $\mathbb{F}_r \to \mathbb{Z}_m$ and the class admits *arbitrary* such maps. No regularity, degree, or algebraic structure is required of τ. This is what will let the construction change alphabets.

2.3 States, two moves, and the prime-factor ladder

The construction is naturally described as movement through a two-dimensional state space.

The state of the construction.

State (p,m) means: we have an explicit source over $\mathbb{F}_p^n$
with seed length $O_{p,m}(\log n + \log\frac{1}{\varepsilon})$ that ε-fools $\mathsf{Local}_n(p,m)$.

The target is the state $(2, M)$, which by Remark 2.2 yields the main theorem. The starting state is $(p, 1)$ for a suitable prime p, which is free: $\mathsf{Local}_n(p, 1)$ contains only the constant function 1, and is fooled exactly by the constant-zero source, which uses no seed. Everything between is accomplished by exactly two moves.

Move 1: LIFT. Fix the alphabet, multiply the modulus by that same prime:

$$\boxed{\mathrm{LIFT}_p : \quad (p, m) \longrightarrow (p, pm).}$$

Construction. Add p independent small-bias masks over $\mathbb{F}_p^n$ to the incoming source.

Move 2: TRANSFER. Fix the modulus, lower the alphabet:

$$\boxed{\mathrm{TRANSFER}_{r\to p} : \quad (r, m) \longrightarrow (p, m), \qquad p < r.}$$

Construction. Push the incoming source through a surjection $\sigma : \mathbb{F}_r \to \mathbb{F}_p$ applied coordinatewise, and add one small-bias mask over $\mathbb{F}_p^n$.

These two moves generate everything, because a modulus is built from its prime factors one at a time, and each factor is inserted over the alphabet that matches it. The mechanisms are different in kind, and it is worth being precise about the role of differencing in each.

Differentiate downward, lift upward, transfer sideways. Differencing is an operation on *tests*, not on sources. Over the alphabet $\mathbb{F}_p$, taking p multiplicative derivatives of a modulus-pm local phase produces a modulus-m local phase:

$$\partial^p \mathsf{Local}_n(p, pm) \subseteq \mathsf{Local}_n(p, m).$$

This is an inclusion of classes, proved by an integrality argument in §3.2.

LIFT runs in the opposite direction, and is a statement about *sources*: given a source whose guarantee is known against the differenced class, adding a small-bias mask promotes that guarantee one derivative level upward, and p masks promote it across the whole hierarchy. LIFT uses the class inclusion; it does not invert it, and there is no algebraic operation on sources corresponding to ∂.

TRANSFER involves no differencing at all. It rests instead on closure under local reparameterisation: a test over the small alphabet pulls back, along σ, to a test over the large one, where the incoming source already controls it.

What each move costs

The parameters of the two moves are separate systems and are introduced where they are used, in §4.1 and §4.2 respectively. What belongs here is only the cost, stated in the direction in which it is applied: to achieve a given error after the move, how good must the incoming source be, and how many masks are added.

Cost of LIFT_p**.** To reach error ε at state (p, pm) it suffices to have error

$$\theta = \left(\frac{\varepsilon}{11}\right)^{2^p}$$

at state (p, m), and to add p independent θ-biased masks over $\mathbb{F}_p^n$.

Cost of $\mathrm{TRANSFER}_{r\to p}$**.** To reach error ε at state (p, m) it suffices to have error

$$\eta = \frac{\varepsilon}{4}$$

at state (r, m), and to add one mask over $\mathbb{F}_p^n$ of bias

$$\delta = \frac{\varepsilon}{4p^{W/2}}, \qquad W = O_{p,r,m}\left(\log\frac{1}{\varepsilon}\right), \qquad \text{so that} \qquad \log\frac{1}{\delta} = O_{p,r,m}\left(\log\frac{1}{\varepsilon}\right).$$

Both costs have the same qualitative shape, and that shape is the entire seed-length argument.

The endgame, stated in advance.

> Each move adds a constant number of masks, and requires the incoming error to be a fixed power of, or a fixed constant multiple of, the outgoing error.

Since M is fixed, the number of moves is a constant depending only on M. Composing constantly many fixed powers gives an innermost error of the form ε^{C_M}/K_M for constants $C_M, K_M > 0$, so that

$$\log\frac{K_M}{\varepsilon^{C_M}} = O_M\left(\log\frac{1}{\varepsilon}\right).$$

Every mask is an explicit small-bias distribution over a fixed prime field, whose seed length is logarithmic in n and in the reciprocal of its bias, and the seeds of constantly many independent components add. The final seed length is therefore $O_M(\log n + \log\frac{1}{\varepsilon})$, as claimed.

The ladder for $M = 15$

The smallest modulus with two distinct odd prime factors is $M = 15$. It is the smallest case in which the alphabet has to change twice, and therefore the smallest ladder that exercises both moves more than once. Its prime factors are 3 and 5; the construction visits them in decreasing order of alphabet:

$$\begin{aligned}\mathsf{Local}_n(5,1) &\xrightarrow{\mathrm{LIFT}_5} \mathsf{Local}_n(5,5) \xrightarrow{\mathrm{TRANSFER}_{5\to 3}} \mathsf{Local}_n(3,5)\\ &\xrightarrow{\mathrm{LIFT}_3} \mathsf{Local}_n(3,15) \xrightarrow{\mathrm{TRANSFER}_{3\to 2}} \mathsf{Local}_n(2,15).\end{aligned}$$

Reading the ladder left to right: start from the trivial class over the alphabet $\mathbb{F}_5$; insert the factor 5 into the modulus, which is possible because the alphabet is 5; carry the resulting modulus-5 guarantee down to the alphabet $\mathbb{F}_3$, which changes nothing about the modulus; insert the factor 3, now possible because the alphabet is 3; and carry the finished modulus-15 guarantee down to the Boolean alphabet, where by Remark 2.2 it is the statement we want.

Two features of this ladder are worth noting before the general induction in §5.2. First, each LIFT inserts a prime only when the alphabet already equals that prime, which is what forces the ordering. Second, TRANSFER only ever moves to a *smaller* alphabet, since a surjection $\mathbb{F}_r \to \mathbb{F}_p$ requires $r \geq p$; this is why the primes are visited in decreasing order and why the walk can always finish at $p = 2$. The complete construction for $M = 15$, with every intermediate source named, is given in §5.1, and its parameter ledger in Appendix A.

2.4 Fourier and small-bias preliminaries

Both moves are analysed by Fourier analysis over the input group $\mathbb{F}_p^n$, and the final conversion to statistical distance uses Fourier analysis over $\mathbb{Z}_M$. We record the common framework once and then specialise.

Let H be a finite abelian group, written additively. A *character* of H is a function $\chi : H \to \mathbb{C}$ with $\chi(x+y) = \chi(x)\chi(y)$ for all $x, y \in H$; a character converts the group operation into multiplication. Since H is finite, $|\chi(x)| = 1$ and $\chi(-x) = \overline{\chi(x)}$. The characters form a group $\widehat{H}$ under pointwise multiplication, with $|\widehat{H}| = |H|$, and they are orthonormal for the normalised inner product $\langle f, g\rangle := \mathbb{E}_{x\sim U_H}[f(x)\overline{g(x)}]$. Being $|H|$ in number, they form an orthonormal basis of all functions $H \to \mathbb{C}$.

Conceptual meaning of the Fourier transform. The characters are multiplicative with respect to the group operation, while the Fourier expansion is linear. Thus Fourier analysis represents a function as a linear combination of multiplicative basis functions. This basis is especially useful when a problem involves addition, translation, or convolution — which is exactly the situation here, since every construction step adds an independent mask.

We now fix the input group. For a prime p write $\omega_p = e^{2\pi i/p}$, and for $\xi, x \in \mathbb{F}_p^n$ let

$$\chi_\xi(x) := \omega_p^{\langle \xi, x\rangle}, \qquad \langle \xi, x\rangle = \sum_{i=1}^{n} \xi_i x_i \pmod p.$$

These are all the characters of $\mathbb{F}_p^n$. For $f : \mathbb{F}_p^n \to \mathbb{C}$ define

$$\widehat{f}(\xi) := \mathbb{E}_{x\sim U_p^n}\left[f(x)\overline{\chi_\xi(x)}\right], \qquad \text{so that} \qquad f(x) = \sum_{\xi\in\mathbb{F}_p^n} \widehat{f}(\xi)\chi_\xi(x),$$

and Parseval's identity holds:

$$\boxed{\sum_{\xi\in\mathbb{F}_p^n} |\widehat{f}(\xi)|^2 = \mathbb{E}_{x\sim U_p^n}|f(x)|^2.}$$

In particular, if $|f| = 1$ everywhere then $\sum_\xi |\widehat{f}(\xi)|^2 = 1$, a normalisation used at several points below. Note also that $\widehat{f}(0) = \mathbb{E}f(U_p^n)$: the trivial Fourier coefficient of a test is its uniform expectation, which is precisely the quantity that Definition 2.3 compares against.

Definition 2.6 (Small bias)**.** *A distribution B on $\mathbb{F}_p^n$ is* δ-biased *if*

$$|\mathbb{E}\chi_\xi(B)| \le \delta \qquad \textit{for every } \xi \neq 0.$$

Imported primitive: small bias over a fixed prime field. For every fixed prime p, every n, and every $0 < \delta \le 1$ there is an explicit δ-biased distribution on $\mathbb{F}_p^n$ with seed length

$$O_p\left(\log n + \log\frac{1}{\delta}\right)$$

[NN93], [AGHP92], [AMN98]. This is the only pseudorandomness primitive the upper bound imports; everything else is proved here.

Two consequences of the definitions will be used in both moves. First, adding an independent mask is diagonal in Fourier coordinates: if B is independent of X then, by multiplicativity of characters,

$$\mathbb{E}\chi_\xi(X+B) = \mathbb{E}\chi_\xi(X)\cdot\mathbb{E}\chi_\xi(B),$$

so in particular $X+B$ is δ-biased whenever B is, irrespective of X. Second, and equivalently, if $h(x) := \mathbb{E}_B[f(x+B)]$ denotes the smoothing of f by B, then

$$\boxed{\widehat{h}(\xi) = \widehat{f}(\xi)\cdot\mathbb{E}\chi_\xi(B).}$$

Smoothing by a small-bias mask thus attenuates every nontrivial Fourier coefficient by a factor of at most δ, and leaves the trivial coefficient untouched.

2.5 Two Fourier transforms and one difference algebra

The proof takes place in three mathematical environments. They are easy to confuse, because two of them are Fourier-analytic and all three involve the same phases, so we set out here what each is for. This subsection proves nothing; it is a map of the machinery.

The three environments and the question each answers. **Input Fourier analysis on $\mathbb{F}_p^n$ asks: how does a mask act?** The masks come with guarantees only against the characters χ_ξ. A local phase is not itself a character, so we expand it in the character basis, where adding an independent mask becomes coefficientwise multiplication and the effect of the mask is transparent. Parseval then evaluates the smoothed second moment. This is the environment of §3.1, §4.1 and §4.2.

Output Fourier analysis on $\mathbb{Z}_M$ asks: what must be proved? The conclusion of the theorem compares two probability distributions on $\mathbb{Z}_M$. Their Fourier coefficients are exactly the expectations of the phases $\phi_{ta}^{(M)}$, so uniform control of all phase expectations yields statistical distance by Parseval and Cauchy–Schwarz. This is the environment of §5.4.

The difference algebra $\mathbb{Z}[C_p]$ asks: why do derivatives simplify a phase? A local exponent map is a function on the p-cycle. Shifts act on such functions, one derivative contributes a difference of two shifts, and p derivatives contribute a product of p such differences. The content of §3.2 is that any such product is divisible by p *as an integer combination of shifts*, which is what removes a factor p from the modulus.

Concretely, let C_p be the cyclic group of order p and let

$$\mathbb{Z}[C_p] \cong \mathbb{Z}[T]/(T^p-1)$$

be its integral group ring, where T acts on functions $h : \mathbb{F}_p \to \mathbb{Z}$ by $(Th)(z) = h(z+1)$. A derivative with shifts u, v replaces a local exponent map q by $(T^u - T^v)q$, so p successive derivatives apply a product of p two-shift differences. The relation that makes the argument work is

$$\boxed{(T-1)^p \in p\,\mathbb{Z}[C_p],}$$

which holds because every intermediate binomial coefficient $\binom{p}{k}$ is divisible by p while the extreme terms combine, in $\mathbb{Z}[C_p]$, to $1 + (-1)^p$.

It is worth saying explicitly why the third environment cannot be absorbed into the first.

Why Fourier analysis does not replace the group ring. Passing to the complex character basis diagonalises the shift T, so a product of differences becomes a product of eigenvalues, and one reads off where the product vanishes. But the statement we need is not that something vanishes: it is that a certain integer combination of shifts is divisible by p. Divisibility is a statement about coefficients in the integral lattice $\mathbb{Z}[C_p]$, and the passage to $\mathbb{C}$ forgets that lattice. Fourier analysis exposes cancellation; only the integral group ring exposes divisibility.

The three environments divide the labour cleanly, and their division is essentially a summary of the proof.

One sentence for the whole argument. Fourier analysis on $\mathbb{F}_p^n$ tells us how a small-bias mask smooths a phase; the group ring $\mathbb{Z}[C_p]$ tells us why sufficiently many derivatives simplify that phase; and Fourier analysis on $\mathbb{Z}_M$ converts the resulting phase estimates into statistical distance.

3 The two engines

The move LIFT is assembled from the two results of this section: an analytic lemma, which converts control of a derivative class into control of the class itself at the cost of one small-bias mask, and an algebraic theorem, which identifies what p successive derivatives do to a local phase over the alphabet $\mathbb{F}_p$. The move TRANSFER uses neither; it is proved directly from local reparameterisation and Fourier smoothing in §4.2. The two engines are independent of each other and are proved in that order.

3.1 Analytic engine: one mask lifts one derivative level

What this subsection provides. Suppose we already know how to fool the *derivatives* of a phase. Adding one small-bias mask recovers control of the phase itself, at the cost of a square root in the error. Iterating the lemma walks a source up a derivative hierarchy, one mask per level.

Lemma 3.1 (Derivative lifting)**.** *Let p be prime and let $\mathcal{C}$ be a class of functions $\mathbb{F}_p^n \to \mathbb{C}$ of unit modulus. Let X be a distribution on $\mathbb{F}_p^n$ with*

$$\mathrm{Err}_X(\partial\mathcal{C}) \le \alpha,$$

and let B be an independent δ-biased distribution on $\mathbb{F}_p^n$. Then

$$\boxed{\mathrm{Err}_{X+B}(\mathcal{C}) \le 3\sqrt{\alpha} + \delta.}$$

Proof. Fix $f \in \mathcal{C}$ and write

$$\mu := \mathbb{E} f(U_p^n), \qquad \rho := |\mu|, \qquad W := X + B.$$

We must bound $|\mathbb{E}f(W) - \mu|$ by $3\sqrt{\alpha} + \delta$; since f is arbitrary this gives the lemma.

We first record an observation used in the second case: *W itself α-fools $\partial\mathcal{C}$*. Indeed $\partial\mathcal{C}$ is closed under shifts, because

$$D_{u,v}f(x+b) = f(x+b+u)\,\overline{f(x+b+v)} = D_{u+b,v+b}f(x).$$

So let $h \in \partial\mathcal{C}$ and fix $b \in \mathbb{F}_p^n$. The shifted function $h(\cdot + b)$ again lies in $\partial\mathcal{C}$, so the hypothesis gives $\left|\mathbb{E}h(X+b) - \mathbb{E}h(U_p^n + b)\right| \le \alpha$, and $U_p^n + b$ is uniform, so $\mathbb{E}h(U_p^n + b) = \mathbb{E}h(U_p^n)$. Averaging over $b \sim B$,

$$\left|\mathbb{E}h(W) - \mathbb{E}h(U_p^n)\right| \le \alpha \qquad \text{for every } h \in \partial\mathcal{C}. \tag{3.1}$$

Case 1: $\rho \le \sqrt{\alpha}$. By Cauchy–Schwarz,

$$|\mathbb{E}f(W)|^2 = |\mathbb{E}_X\mathbb{E}_B f(X+B)|^2 \le \mathbb{E}_X\,|\mathbb{E}_B f(X+B)|^2\,.$$

Let B' be an independent copy of B. Expanding the square,

$$\mathbb{E}_X\,|\mathbb{E}_B f(X+B)|^2 = \mathbb{E}_{B,B'}\,\mathbb{E}_X\Big[f(X+B)\overline{f(X+B')}\Big] = \mathbb{E}_{B,B'}\,\mathbb{E}_X[D_{B,B'}f(X)]\,.$$

For each fixed pair (b, b') the function $D_{b,b'}f$ lies in $\partial\mathcal{C}$, so replacing X by uniform changes the inner expectation by at most α; averaging over (b, b') and applying the triangle inequality,

$$\mathbb{E}_X\,|\mathbb{E}_B f(X+B)|^2 \le \mathbb{E}_{U_p^n}\left|\mathbb{E}_B f(U_p^n + B)\right|^2 + \alpha.$$

The smoothed function $x \mapsto \mathbb{E}_B f(x+B)$ has Fourier coefficients $\widehat{f}(\xi)\,\mathbb{E}\chi_\xi(B)$ by the smoothing identity of §2.4, so Parseval gives

$$\mathbb{E}_{U_p^n}\left|\mathbb{E}_B f(U_p^n + B)\right|^2 = \sum_{\xi\in\mathbb{F}_p^n} |\widehat{f}(\xi)|^2\,|\mathbb{E}\chi_\xi(B)|^2\,.$$

In this sum the trivial mode contributes $|\widehat{f}(0)|^2 = \rho^2$; every nontrivial mode is attenuated by $|\mathbb{E}\chi_\xi(B)|^2 \le \delta^2$; and $\sum_\xi |\widehat{f}(\xi)|^2 = 1$ because $|f| = 1$. Hence the sum is at most $\rho^2 + \delta^2$, and

$$|\mathbb{E}f(W)|^2 \le \rho^2 + \delta^2 + \alpha.$$

Using $\sqrt{s+t+u} \le \sqrt{s} + \sqrt{t} + \sqrt{u}$ and then $\rho \le \sqrt{\alpha}$,

$$|\mathbb{E}f(W) - \mu| \le |\mathbb{E}f(W)| + \rho \le \sqrt{\rho^2 + \delta^2 + \alpha} + \rho \le 2\rho + \delta + \sqrt{\alpha} \le 3\sqrt{\alpha} + \delta.$$

Case 2: $\rho > \sqrt{\alpha}$. Here the small-bias property of the mask is not needed; instead we divide by ρ. Let V be uniform on $\mathbb{F}_p^n$ and independent of everything else. Conditionally on W, the shifted variable $W + V$ is uniform, so $\mathbb{E}_V f(W+V) = \mu$ and therefore

$$\mathbb{E}_{W,V}\Big[f(W)\overline{f(W+V)}\Big] = \mathbb{E}[f(W)]\,\overline{\mu}.$$

The same computation with W replaced by a uniform input gives $\mathbb{E}_{U,V}[f(U_p^n)\overline{f(U_p^n+V)}] = \mu\overline{\mu}$. Subtracting, and recognising $f(x)\overline{f(x+v)} = D_{0,v}f(x)$,

$$(\mathbb{E}f(W) - \mu)\,\overline{\mu} = \mathbb{E}_V\left[\mathbb{E}D_{0,V}f(W) - \mathbb{E}D_{0,V}f(U_p^n)\right]. \tag{3.2}$$

For each fixed v we have $D_{0,v}f \in \partial\mathcal{C}$, so by (3.1) every bracket on the right of (3.2) has modulus at most α, and hence so does the average. Taking absolute values,

$$|\mathbb{E}f(W) - \mu|\,\rho \le \alpha, \qquad \text{so} \qquad |\mathbb{E}f(W) - \mu| \le \frac{\alpha}{\rho} < \frac{\alpha}{\sqrt{\alpha}} = \sqrt{\alpha} \le 3\sqrt{\alpha} + \delta.$$

The two cases are exhaustive, which proves the lemma. □

Two features of this lemma deserve comment before it is put to work.

The case split refines the low-weight/high-weight split. For a local phase $\Phi_{\mathbf{q}}$ the uniform expectation factorises across coordinates, and each active coordinate contributes a factor of modulus at most some $\kappa < 1$ depending only on p and m, so

$$\rho = \left|\mathbb{E}\Phi_{\mathbf{q}}(U_p^n)\right| \le \kappa^{\mathrm{wt}(\mathbf{q})}.$$

Large weight therefore forces small ρ, and Case 2 can occur only at low weight. The converse fails: small ρ does not imply large weight, since a phase with a single active coordinate may already have uniform expectation 0, or close to it, through cancellation within that one coordinate. Case 1 thus contains every sufficiently high-weight phase, but also some low-weight ones. The lemma splits on the analytically relevant quantity ρ rather than on the weight itself, which is why no threshold has to be chosen here; an explicit weight threshold reappears only in §4.2, where no derivative is available and the two regimes must be separated by hand.

No Fourier ℓ^1-mass is paid here. One might expect the low-weight regime to be settled by the mask, at a cost proportional to the Fourier ℓ^1-mass of the phase, which would force the mask bias δ to be exponentially small in the weight. Case 2 above does not use the mask's small-bias property at all: it divides the exact identity (3.2) by ρ, and the only thing it asks of B is independence, so that W inherits control of $\partial\mathcal{C}$ through translation closure. Consequently the masks in LIFT are set to the single internal parameter θ already present in the statement, with no further weight-dependent suppression, and the ℓ^1-mass of a local phase never enters this engine. It does enter TRANSFER (§4.2), where no derivative is available and the mask must do the low-weight work alone; that is the one place in the paper where a mask is required to be exponentially small in a threshold.

3.2 Algebraic engine: p derivatives remove one factor p

We now determine what p successive derivatives do to a local phase over the alphabet $\mathbb{F}_p$. As explained in §2.5, the computation takes place in the integral group ring

$$R_p := \mathbb{Z}[C_p] \cong \mathbb{Z}[T]/(T^p - 1),$$

where T acts on functions $h : \mathbb{F}_p \to \mathbb{Z}$ by $(Th)(z) = h(z+1)$, and a derivative with shifts u, v replaces a local exponent map q by $(T^u - T^v)q$.

Theorem 3.2 (Group-ring divisibility)**.** *For arbitrary shifts* $u_1, v_1, \ldots, u_p, v_p \in \mathbb{F}_p$*,*

$$\boxed{\prod_{j=1}^{p} (T^{u_j} - T^{v_j}) \in pR_p.}$$

Proof. We first show $(T-1)^p \in pR_p$. Expanding,

$$(T-1)^p = T^p + (-1)^p + \sum_{k=1}^{p-1} (-1)^{p-k} \binom{p}{k} T^k.$$

In R_p we have $T^p = 1$, so the two extreme terms combine to the constant $1 + (-1)^p$, which is 0 when p is odd and 2 when $p = 2$; in both cases it is divisible by p. Every intermediate binomial coefficient $\binom{p}{k}$ with $1 \le k \le p-1$ is divisible by p because p is prime. Hence every coefficient of $(T-1)^p$ is divisible by p, which is the claim.

Now write $h_j := u_j - v_j$, so that $T^{u_j} - T^{v_j} = T^{v_j}(T^{h_j} - 1)$. If $h_j = 0$ for some j then that factor vanishes and the entire product is $0 \in pR_p$. Otherwise choose the representative $h_j \in \{1, \ldots, p-1\}$ and use

$$T^{h_j} - 1 = (T-1)\left(1 + T + \cdots + T^{h_j - 1}\right).$$

Each of the p factors is therefore divisible by $T-1$, so the product is divisible by $(T-1)^p$ and hence lies in pR_p. □

Corollary 3.3 (Local-phase descent)**.** *For every prime* p *and every* $m \ge 1$*,*

$$\boxed{\partial^p \mathsf{Local}_n(p, pm) \subseteq \mathsf{Local}_n(p, m).}$$

Proof. Let $\Phi_{\mathbf{q}} \in \mathsf{Local}_n(p, pm)$, with local exponent maps $q_i : \mathbb{F}_p \to \mathbb{Z}_{pm}$, and choose integer lifts $\widetilde{q}_i : \mathbb{F}_p \to \mathbb{Z}$. A member of $\partial^p \mathsf{Local}_n(p, pm)$ is obtained from some such $\Phi_{\mathbf{q}}$ by p successive derivatives, with shift pairs $(u^{(1)}, v^{(1)}), \ldots, (u^{(p)}, v^{(p)}) \in (\mathbb{F}_p^n)^2$. By the coordinatewise action recorded after Definition 2.5, the exponent of the resulting phase is

$$\sum_{i=1}^{n} (D_i \widetilde{q}_i)(x_i), \qquad D_i := \prod_{j=1}^{p} \left(T^{u_i^{(j)}} - T^{v_i^{(j)}}\right) \in R_p.$$

By Theorem 3.2 we may write $D_i = pQ_i$ with $Q_i \in R_p$; the element Q_i is unique, since R_p is free as a $\mathbb{Z}$-module and therefore torsion-free. Consequently

$$\omega_{pm}^{\sum_i (D_i \widetilde{q}_i)(x_i)} = \omega_{pm}^{p \sum_i (Q_i \widetilde{q}_i)(x_i)} = \omega_m^{\sum_i (Q_i \widetilde{q}_i)(x_i)}.$$

Define $q_i'(z) := (Q_i \widetilde{q}_i)(z) \bmod m$. This does not depend on the chosen lift: altering $\widetilde{q}_i$ by a multiple of pm alters $Q_i \widetilde{q}_i$ by a multiple of pm, hence by a multiple of m. So $\mathbf{q}' = (q_1', \ldots, q_n')$ is a tuple of local exponent maps $\mathbb{F}_p \to \mathbb{Z}_m$ and the derivative phase is $\Phi_{\mathbf{q}'} \in \mathsf{Local}_n(p, m)$. □

Remark 3.4 (Why the proof uses p derivatives)**.** *In the Boolean linear-phase setting, one derivative already reduces modulus* $2m$ *to modulus* m*, but only up to a global unit scalar. Corollary 3.3 uses two derivatives when* $p = 2$ *because it asserts a literal inclusion into* $\mathsf{Local}_n(2, m)$*. More generally,* $p-1$ *derivatives suffice up to a scalar; using* p *derivatives avoids introducing scalar-closed phase classes and affects only the constant depending on* M *in the final seed length.*

3.3 Why one alphabet cannot finish the proof

Corollary 3.3 removes a factor p from the modulus using the alphabet $\mathbb{F}_p$. The next proposition shows that this matching is forced: over a fixed alphabet, differencing cannot touch the part of the modulus coprime to that alphabet, no matter how many derivatives are taken. This is why a single alphabet cannot suffice for a modulus with more than one prime factor, and it is the precise obstruction that the move TRANSFER exists to circumvent.

Write

$$\mathsf{Local}_n^0(p,m) := \Big\{ \Phi_{\mathbf{q}} \in \mathsf{Local}_n(p,m) \ : \ \textstyle\sum_{z\in\mathbb{F}_p} q_i(z) = 0 \text{ in } \mathbb{Z}_m \text{ for every } i \Big\}$$

for the subclass of local phases *admitting* a tuple of exponent maps with vanishing sums. The existential reading matters: a phase does not determine its tuple $\mathbf{q}$, since a constant may be moved between coordinates, and adding a constant c to q_i changes $\sum_z q_i(z)$ by pc. The displayed condition is a property of some representation, not of all of them, and the proof below works throughout at the level of tuples.

Proposition 3.5 (No descent past the coprime part)**.** *Let p be prime and let $m \geq 2$ satisfy $\gcd(p,m) = 1$. Then*

$$\partial^r \mathsf{Local}_n(p,m) = \mathsf{Local}_n^0(p,m) \qquad \textit{for every } r \geq 1.$$

In particular the modulus never descends: for every r the class $\partial^r \mathsf{Local}_n(p,m)$ still contains phases of exact modulus m.

Proof. Work coordinatewise; it suffices to treat the set of achievable local exponent maps. For any shifts, $\sum_z \big((T^u - T^v)q\big)(z) = \sum_z q(z) - \sum_z q(z) = 0$, so one derivative lands in $\mathsf{Local}_n^0(p,m)$, and the same computation shows $\partial\mathsf{Local}_n^0(p,m) \subseteq \mathsf{Local}_n^0(p,m)$.

For the reverse inclusion, let $r : \mathbb{F}_p \to \mathbb{Z}_m$ have vanishing sum. Define q on the representatives $0, 1, \ldots, p-1$ by

$$q(0) := c, \qquad q(z) := c + \sum_{j=0}^{z-1} r(j) \quad (1 \leq z \leq p-1),$$

for a constant c to be chosen. This is consistent as a map on $\mathbb{F}_p$ precisely because $\sum_j r(j) = 0$, and by construction $(T-1)q = r$. It remains to choose c so that q itself has vanishing sum: $\sum_z q(z)$ changes by pc as c varies, so we need $pc \equiv -\Sigma$ in $\mathbb{Z}_m$, where Σ is the sum obtained with $c = 0$. Since $\gcd(p,m) = 1$ the element p is invertible in $\mathbb{Z}_m$ and such a c exists. Hence $r \in \partial\mathsf{Local}_n^0(p,m)$, and the two classes coincide from $r = 1$ onwards.

Finally, $\mathsf{Local}_n^0(p,m)$ contains phases of exact modulus m: take a single active coordinate with $q(0) = 1$, $q(1) = -1$, and $q(z) = 0$ otherwise. Its values generate $\mathbb{Z}_m$, so the phase is not a phase of any proper divisor of m. □

Corollary 3.6 (The Boolean barrier)**.** *Let $M = 2^t m'$ with m' odd. Then*

$$\partial^{2t} \mathsf{Local}_n(2,M) \subseteq \mathsf{Local}_n(2,m'),$$

while for every r the class $\partial^r \mathsf{Local}_n(2,M)$ still contains phases of exact modulus m'. Differencing over the Boolean alphabet therefore strips the full power of two dividing M, and then stops at the odd part.

Proof. The first assertion is t applications of Corollary 3.3 with $p = 2$: each pair of derivatives replaces the modulus $2m$ by m, so $2t$ of them take $2^t m'$ down to m'.

For the second, note that $\mathsf{Local}_n(2, m') \subseteq \mathsf{Local}_n(2, M)$, since a phase $\omega_{m'}^{\sum_i q_i(x_i)}$ equals $\omega_M^{\sum_i 2^t q_i(x_i)}$ and $2^t q_i$ is again a map into $\mathbb{Z}_M$. Differencing is monotone under inclusion of classes, so

$$\partial^r \mathsf{Local}_n(2, M) \supseteq \partial^r \mathsf{Local}_n(2, m') = \mathsf{Local}_n^0(2, m')$$

for every $r \geq 1$, the last equality by Proposition 3.5, which applies because m' is odd. That class contains phases of exact modulus m', as exhibited at the end of the proposition. □

Where the two engines leave us. Algebra moves *downward* by one prime factor, and only by a prime matching the alphabet; the analytic lemma moves a *source guarantee* back upward across the resulting hierarchy, one mask per level. Together they give the move LIFT of §4.1. Proposition 3.5 shows that this is all a single alphabet can do, which is what makes the second move necessary.

4 The two moves

Each move is now assembled. The two are independent: LIFT uses both engines of §3 and no reparameterisation, while TRANSFER uses reparameterisation and no derivative at all. Each subsection opens with the state it starts from, the operation it performs, and the state it reaches, and each carries its own parameters, which are local to it.

4.1 LIFT: inserting one matching prime

The move. **Current state.** X_0 over $\mathbb{F}_p^n$ fools $\mathsf{Local}_n(p, m)$.
Operation. Add p independent small-bias masks over $\mathbb{F}_p^n$.
New state. X_p over $\mathbb{F}_p^n$ fools $\mathsf{Local}_n(p, pm)$.

symbol	meaning
θ	error of the incoming source, and bias of each mask
ε_j	error of X_j against the class at derivative level j
A_j	absolute constants of the error recursion, all below 11

Theorem 4.1 (LIFT)**.** *Let p be prime, $m \geq 1$, and $0 < \varepsilon \leq 1$. Put*

$$\theta := \left(\frac{\varepsilon}{11}\right)^{2p},$$

and suppose X_0 is a distribution on $\mathbb{F}_p^n$ with

$$\mathrm{Err}_{X_0}\big(\mathsf{Local}_n(p, m)\big) \leq \theta.$$

Let $B_1, \dots, B_p$ be independent θ-biased distributions on $\mathbb{F}_p^n$, independent also of X_0, and set

$$X_j := X_{j-1} + B_j \qquad (j = 1, \dots, p).$$

Then

$$\boxed{\mathrm{Err}_{X_p}\big(\mathsf{Local}_n(p, pm)\big) \le \varepsilon.}$$

Proof. Define the derivative hierarchy

$$\mathcal{C}_j := \partial^{p-j}\mathsf{Local}_n(p, pm) \qquad (0 \le j \le p),$$

so that $\mathcal{C}_p = \mathsf{Local}_n(p, pm)$ is the target class and, directly from the definition,

$$\partial\mathcal{C}_j = \mathcal{C}_{j-1} \qquad (1 \le j \le p).$$

Every $\mathcal{C}_j$ consists of local phases, hence of unit-modulus functions, so Lemma 3.1 applies to each.

By Corollary 3.3,

$$\mathcal{C}_0 = \partial^p\mathsf{Local}_n(p, pm) \subseteq \mathsf{Local}_n(p, m),$$

and since Err is monotone under inclusion of classes (§2.2),

$$\varepsilon_0 := \mathrm{Err}_{X_0}(\mathcal{C}_0) \le \mathrm{Err}_{X_0}\big(\mathsf{Local}_n(p, m)\big) \le \theta.$$

This is the only point at which the algebraic engine is used, and the only point at which the hypothesis on X_0 enters.

For $1 \le j \le p$ put $\varepsilon_j := \mathrm{Err}_{X_j}(\mathcal{C}_j)$. Apply Lemma 3.1 with the class $\mathcal{C}_j$, the source X_{j-1} and the mask B_j. Its hypothesis is a bound on $\mathrm{Err}_{X_{j-1}}(\partial\mathcal{C}_j) = \mathrm{Err}_{X_{j-1}}(\mathcal{C}_{j-1}) = \varepsilon_{j-1}$, and its mask parameter is $\delta = \theta$, so

$$\varepsilon_j \le 3\sqrt{\varepsilon_{j-1}} + \theta \qquad (1 \le j \le p). \tag{4.1}$$

The source climbs one derivative level per mask.

It remains to solve (4.1). Define constants

$$A_0 := 1, \qquad A_j := 3\sqrt{A_{j-1}} + 1,$$

and claim that

$$\varepsilon_j \le A_j\, \theta^{1/2^j} \qquad (0 \le j \le p).$$

For $j = 0$ this is $\varepsilon_0 \le \theta$. Assuming it at $j-1$, and using $0 < \theta \le 1$ so that $\theta \le \theta^{1/2^j}$,

$$\varepsilon_j \le 3\sqrt{A_{j-1}\theta^{1/2^{j-1}}} + \theta = 3\sqrt{A_{j-1}}\,\theta^{1/2^j} + \theta \le \left(3\sqrt{A_{j-1}} + 1\right)\theta^{1/2^j} = A_j\theta^{1/2^j}.$$

Finally $A_j < 11$ for every j: this holds at $j = 0$, and if $A_{j-1} < 11$ then $A_j < 3\sqrt{11} + 1 < 11$. Therefore

$$\varepsilon_p < 11\,\theta^{1/2^p} = 11 \cdot \frac{\varepsilon}{11} = \varepsilon,$$

which is the assertion. □

Seed length of one LIFT. Writing $R_{p,m}(n,\eta)$ for the seed length of a source over $\mathbb{F}_p^n$ that η-fools $\mathsf{Local}_n(p,m)$, the construction gives

$$R_{p,pm}(n,\varepsilon) \le R_{p,m}(n,\theta) + p \cdot O_p\left(\log n + \log\frac{1}{\theta}\right).$$

Since p is fixed,

$$\log\frac{1}{\theta} = 2^p \log\frac{11}{\varepsilon} = O_p\left(\log\frac{1}{\varepsilon}\right),$$

so the property $R = O(\log n + \log\frac{1}{(\cdot)})$ is preserved when the modulus m is replaced by pm.

Remark 4.2 (One mask per derivative level)**.** *The number of masks is exactly the number of derivatives needed by Corollary 3.3, namely p. Each application of Lemma 3.1 costs one square root, so p of them turn the incoming error θ into roughly $\theta^{1/2^p}$; solving for θ is what produces the exponent 2^p in the statement. At $p = 2$ the move reads: two masks, and a square-root loss per level, lift a modulo-m guarantee to a modulo-$2m$ one.*

4.2 TRANSFER: lowering the alphabet

The move. **Current state.** Y over $\mathbb{F}_r^n$ fools $\mathsf{Local}_n(r,m)$, where $p < r$.

Operation. Apply a surjection $\sigma : \mathbb{F}_r \to \mathbb{F}_p$ coordinatewise and add one small-bias mask over $\mathbb{F}_p^n$.

New state. $X = \sigma(Y) + B$ over $\mathbb{F}_p^n$ fools $\mathsf{Local}_n(p,m)$.

The modulus is untouched by this move; only the alphabet changes. The reason the move needs two ingredients rather than one is worth stating before the proof, because it determines the shape of the estimate.

Why a surjection alone cannot suffice. Let $\sigma : \mathbb{F}_r \to \mathbb{F}_p$ be any map. Its fibres partition a set of size r into p parts, and since $p < r$ are distinct primes we have $p \nmid r$, so the fibres cannot all have equal size. Hence

$$\sigma(U_r) \neq U_p \qquad \text{for every } \sigma,$$

and no deterministic alphabet map transports uniformity correctly. The discrepancy is repaired in two different ways depending on how many coordinates the test looks at. When few coordinates are active, the small-bias mask alone washes it out. When many are active, the discrepancy stops mattering, because the test expectation is exponentially small both under the true uniform distribution and under the pushforward $\sigma(U_r) + \beta$. The estimate below is exactly the minimum of these two.

symbol	meaning
η	error of the incoming source Y over the larger alphabet
δ	bias of the mask B over $\mathbb{F}_p^n$
w	number of active coordinates of the test, $\mathrm{wt}(\mathbf{q})$
κ_U	contraction of one active coordinate under U_p
κ_σ	contraction of one active coordinate under $\sigma(U_r) + \beta$
W	weight threshold separating the two estimates

Definition 4.3 (Contraction constants)**.** *Fix primes $p < r$, a modulus $m \geq 2$, and a surjection $\sigma : \mathbb{F}_r \to \mathbb{F}_p$. Set*

$$\kappa_U := \max_{\substack{q:\mathbb{F}_p \to \mathbb{Z}_m \\ q \text{ nonconstant}}} \left|\mathbb{E}_{z\sim U_p} \omega_m^{q(z)}\right|, \qquad \kappa_\sigma := \max_{\substack{q:\mathbb{F}_p \to \mathbb{Z}_m \\ q \text{ nonconstant} \\ \beta \in \mathbb{F}_p}} \left|\mathbb{E}_{y\sim U_r} \omega_m^{q(\sigma(y)+\beta)}\right|,$$

and $\kappa := \max\{\kappa_U, \kappa_\sigma\}$.

Lemma 4.4 (Strict contraction)**.** $\kappa < 1$.

Proof. Each of the two quantities is a maximum over a finite set, so it is enough that every individual expectation has modulus below 1. Both U_p and every translate $\sigma(U_r) + \beta$ have full support on $\mathbb{F}_p$, the latter because σ is surjective. Each expectation is therefore a convex combination, with strictly positive weights, of the points $\omega_m^{q(z)}$ for $z \in \mathbb{F}_p$. Such a combination of points on the unit circle has modulus 1 only when all of the contributing points coincide, which for a nonconstant q they do not. □

Remark 4.5 (The Boolean case)**.** *For $p = 2$ a nonconstant $q : \mathbb{F}_2 \to \mathbb{Z}_m$ has $q(1) - q(0) = c \neq 0$, so*

$$\kappa_U = \max_{c \neq 0} \frac{|1 + \omega_m^c|}{2} = \cos\frac{\pi}{m},$$

the familiar decay constant for uniform bits against a modulus-m test. The second constant κ_σ has no counterpart over a single alphabet; it is where the non-uniformity of the pushforward is absorbed.

Theorem 4.6 (TRANSFER)**.** *Let $p < r$ be primes, $m \geq 2$, and $0 < \varepsilon \leq 1$. Fix a surjection $\sigma : \mathbb{F}_r \to \mathbb{F}_p$ and let κ be as in Definition 4.3. Put*

$$\eta := \frac{\varepsilon}{4}, \qquad W := \begin{cases} 0, & \kappa = 0, \\ \left\lceil \dfrac{\log(8/\varepsilon)}{\log(1/\kappa)} \right\rceil, & \kappa > 0, \end{cases} \qquad \delta := \frac{\varepsilon}{4p^{W/2}}.$$

Let Y be a distribution on $\mathbb{F}_r^n$ with $\mathrm{Err}_Y(\mathsf{Local}_n(r, m)) \leq \eta$, and let B be an independent δ-biased distribution on $\mathbb{F}_p^n$. Then $X := \sigma(Y) + B$, with σ applied coordinatewise, satisfies

$$\boxed{\mathrm{Err}_X\left(\mathsf{Local}_n(p, m)\right) \leq \varepsilon.}$$

Proof. Fix a test $\Phi_{\mathbf{q}} \in \mathsf{Local}_n(p, m)$ and let $w = \mathrm{wt}(\mathbf{q})$ be its number of active coordinates. We prove two estimates, valid for every w, and then choose between them.

Low-weight estimate. Since B is δ-biased and independent of $\sigma(Y)$, the sum X is itself δ-biased (§2.4). Expanding $\Phi_{\mathbf{q}}$ in the character basis of $\mathbb{F}_p^n$ and using $\widehat{\Phi_{\mathbf{q}}}(0) = \mathbb{E}\Phi_{\mathbf{q}}(U_p^n)$,

$$\left|\mathbb{E}\Phi_{\mathbf{q}}(X) - \mathbb{E}\Phi_{\mathbf{q}}(U_p^n)\right| = \left|\sum_{\xi \neq 0} \widehat{\Phi_{\mathbf{q}}}(\xi)\, \mathbb{E}\chi_\xi(X)\right| \leq \delta \sum_{\xi \neq 0} \left|\widehat{\Phi_{\mathbf{q}}}(\xi)\right|.$$

It remains to bound the Fourier ℓ^1-mass. For one coordinate write $\varphi_i(z) := \omega_m^{q_i(z)}$, a unit-modulus function on $\mathbb{F}_p$. By Cauchy–Schwarz and Parseval,

$$\sum_{\alpha\in\mathbb{F}_p} |\widehat{\varphi_i}(\alpha)| \le \sqrt{p}\left(\sum_{\alpha} |\widehat{\varphi_i}(\alpha)|^2\right)^{1/2} = \sqrt{p},$$

while a constant coordinate has ℓ^1-mass exactly 1. Since $\Phi_{\mathbf{q}}$ is a product of one-coordinate functions its Fourier transform tensorises, so

$$\sum_{\xi} \left|\widehat{\Phi_{\mathbf{q}}}(\xi)\right| = \prod_{i=1}^{n} \sum_{\alpha\in\mathbb{F}_p} |\widehat{\varphi_i}(\alpha)| \le p^{w/2}.$$

Hence

$$\left|\mathbb{E}\Phi_{\mathbf{q}}(X) - \mathbb{E}\Phi_{\mathbf{q}}(U_p^n)\right| \le L(w) := \delta\, p^{w/2}. \tag{4.2}$$

Note that the incoming source Y plays no part in this estimate.

High-weight estimate. Condition on $B = \beta$. By closure under local reparameterisation (§2.2), the function

$$y \longmapsto \Phi_{\mathbf{q}}\big(\sigma(y) + \beta\big) = \omega_m^{\sum_i q_i(\sigma(y_i)+\beta_i)}, \qquad y \in \mathbb{F}_r^n,$$

is a local phase over the alphabet $\mathbb{F}_r$ with the same modulus m, that is, a member of $\mathsf{Local}_n(r, m)$: its local exponent maps are $q_i \circ (\sigma + \beta_i) : \mathbb{F}_r \to \mathbb{Z}_m$, and the class admits arbitrary such maps. The hypothesis on Y therefore applies to it, giving

$$\left|\mathbb{E}_Y \Phi_{\mathbf{q}}\big(\sigma(Y) + \beta\big) - \mathbb{E}_{U_r^n} \Phi_{\mathbf{q}}\big(\sigma(U_r^n) + \beta\big)\right| \le \eta.$$

Under U_r^n the coordinates are independent, so the second expectation factorises over coordinates. If q_i is nonconstant then so is $q_i \circ (\sigma + \beta_i)$, because σ is surjective and the two maps have the same image; such a coordinate contributes a factor of modulus at most κ_σ, while an inactive coordinate contributes a factor of modulus 1. Hence

$$\left|\mathbb{E}_{U_r^n} \Phi_{\mathbf{q}}\big(\sigma(U_r^n) + \beta\big)\right| \le \kappa_\sigma^w, \qquad \text{so} \qquad \left|\mathbb{E}\Phi_{\mathbf{q}}(X)\right| \le \eta + \kappa_\sigma^w$$

after averaging over $\beta \sim B$. The same factorisation under U_p^n gives

$$\left|\mathbb{E}\Phi_{\mathbf{q}}(U_p^n)\right| \le \kappa_U^w.$$

The two expectations are separately small, so their difference is small:

$$\left|\mathbb{E}\Phi_{\mathbf{q}}(X) - \mathbb{E}\Phi_{\mathbf{q}}(U_p^n)\right| \le H(w) := \eta + 2\kappa^w. \tag{4.3}$$

Combining. By (4.2) and (4.3),

$$\boxed{\left|\mathbb{E}\Phi_{\mathbf{q}}(X) - \mathbb{E}\Phi_{\mathbf{q}}(U_p^n)\right| \le \min\left\{\, \delta p^{w/2},\ \eta + 2\kappa^w \,\right\}.}$$

If $w \le W$ the first branch gives $L(w) \le \delta p^{W/2} = \varepsilon/4$. If $w > W$ then $\kappa^w \le \kappa^W \le \varepsilon/8$ by the choice of W, so the second branch gives $H(w) \le \varepsilon/4 + 2 \cdot \varepsilon/8 = \varepsilon/2$. When $\kappa = 0$ we have $W = 0$, and the second branch reads $H(w) \le \eta = \varepsilon/4$ for every $w \ge 1$ while $w = 0$ gives a constant test and zero error. In all cases the error is at most ε, and $\Phi_{\mathbf{q}}$ was arbitrary. □

Seed length of one TRANSFER. Since $\kappa < 1$ is a constant depending only on p, r, m and the fixed map σ, we have $W = O_{p,r,m}(\log \frac{1}{\varepsilon})$, and therefore

$$\log \frac{1}{\delta} = \log \frac{4}{\varepsilon} + \frac{W}{2} \log p = O_{p,r,m}\left(\log \frac{1}{\varepsilon} \right).$$

The total seed length is thus

$$R_{p,m}(n, \varepsilon) \le R_{r,m}\left(n, \frac{\varepsilon}{4} \right) + O_{p,r,m}\left(\log n + \log \frac{1}{\varepsilon} \right),$$

so the property $R = O(\log n + \log \frac{1}{(\cdot)})$ is preserved when the alphabet r is replaced by the smaller prime p.

How to read the estimate. The shape $\min\{L(w), H(w)\}$ is the classical low-weight/high-weight division, with the weight threshold W chosen so that the high-weight branch has taken effect by the time the low-weight branch expires, and δ set to $\varepsilon/4$ divided by the ℓ^1-mass at that threshold. Four points are specific to the present setting, and together they are what make the move work across alphabets. *Active* means that a local exponent map is nonconstant, not that a coefficient is nonzero. The low-weight cost is $p^{w/2}$, with the alphabet visible in it. The high-weight decay is required under *two* distributions, the true uniform U_p and the pushforward $\sigma(U_r) + \beta$, whence the two constants of Definition 4.3. And surjectivity of σ is what supplies both properties the high-weight estimate needs: full support, so that Lemma 4.4 applies to the pushforward, and preservation of nonconstancy, so that an active coordinate stays active after reparameterisation.

5 Assembly and completion

The two moves are now combined. We first walk a single modulus in full, then give the general induction, and then convert the resulting phase estimates into the statement about statistical distance.

Throughout this section a mask carries its alphabet in a superscript: $B^{(p)}$ denotes a small-bias distribution on $\mathbb{F}_p^n$. This is needed because the construction uses masks over several alphabets at once.

5.1 The complete construction for $M = 15$

Take $M = 15$, and let $0 < \gamma \le 1$ be the target error. Fix once and for all two surjections

$$\sigma_{5\to3} : \mathbb{F}_5 \to \mathbb{F}_3, \qquad \sigma_{3\to2} : \mathbb{F}_3 \to \mathbb{F}_2,$$

each applied coordinatewise. The construction has four steps, and it is easier to read forwards while remembering that the parameters are determined backwards, from the target γ inwards.

Base: the state $(5, 1)$. Let Y_0 be the constant source $0 \in \mathbb{F}_5^n$, which uses no seed at all. The class $\mathsf{Local}_n(5, 1)$ contains only the constant function 1, so

$$\mathrm{Err}_{Y_0}\big(\mathsf{Local}_n(5, 1)\big) = 0.$$

Step 1, LIFT_5**:** $(5,1) \to (5,5)$**.** Add five independent masks over $\mathbb{F}_5^n$,

$$Y := Y_0 + B_1^{(5)} + \cdots + B_5^{(5)},$$

and let $\varepsilon_1 := \mathrm{Err}_Y(\mathsf{Local}_n(5,5))$. Theorem 4.1 applies with $p = 5$ and $m = 1$: its hypothesis on the incoming source is satisfied with room to spare, since that source has error exactly 0.

Step 2, $\mathrm{TRANSFER}_{5\to3}$**:** $(5,5) \to (3,5)$**.** Push Y through the alphabet map and add one mask over $\mathbb{F}_3^n$,

$$Z_0 := \sigma_{5\to3}(Y) + B_0^{(3)},$$

and let $\varepsilon_2 := \mathrm{Err}_{Z_0}(\mathsf{Local}_n(3,5))$. Theorem 4.6 applies with $r = 5$, $p = 3$, $m = 5$. The modulus is unchanged by this step; only the alphabet drops, which is what makes the next step possible.

Step 3, LIFT_3**:** $(3,5) \to (3,15)$**.** Add three independent masks over $\mathbb{F}_3^n$,

$$Z := Z_0 + B_1^{(3)} + B_2^{(3)} + B_3^{(3)},$$

and let $\varepsilon_3 := \mathrm{Err}_Z(\mathsf{Local}_n(3,15))$. Theorem 4.1 applies with $p = 3$ and $m = 5$; the descent it invokes is $\partial^3\mathsf{Local}_n(3,15) \subseteq \mathsf{Local}_n(3,5)$, which is exactly the class controlled at the end of Step 2. This is the step that would be impossible over the Boolean alphabet, by Corollary 3.6.

Step 4, $\mathrm{TRANSFER}_{3\to2}$**:** $(3,15) \to (2,15)$**.** Push Z down to bits and add one mask over $\mathbb{F}_2^n$, the addition now being $\oplus$:

$$\boxed{G := \sigma_{3\to2}(Z) \oplus B_0^{(2)},}$$

and let $\varepsilon_4 := \mathrm{Err}_G(\mathsf{Local}_n(2,15))$. Theorem 4.6 applies with $r = 3$, $p = 2$, $m = 15$.

The output G is a distribution on $\{0,1\}^n$, and by Remark 2.2 the class $\mathsf{Local}_n(2,15)$ contains every phase $\phi_a^{(15)}$. So $\varepsilon_4 \le \gamma$ is exactly the character estimate we want.

Reading the parameters backwards. Fix the target $\varepsilon_4 = \gamma$. Then Theorem 4.6 at Step 4 requires $\varepsilon_3 = \gamma/4$; Theorem 4.1 at Step 3 requires $\varepsilon_2 = (\varepsilon_3/11)^{2^3}$; Theorem 4.6 at Step 2 requires $\varepsilon_1 = \varepsilon_2/4$; and Theorem 4.1 at Step 1 requires of the base source an error of at most $(\varepsilon_1/11)^{2^5}$, which the base meets since its error is 0. Each mask is taken with the bias prescribed by the theorem that introduces it. The construction therefore uses

$$5 + 1 + 3 + 1 = 10$$

masks, over three different alphabets, and the smallest bias among them is a fixed power of γ times a constant. The explicit values are collected in Appendix A.

5.2 The sorted-prime induction

The general case repeats what the $M = 15$ ladder just did, with the prime factors of the modulus visited in decreasing order of alphabet.

Theorem 5.1 (Local phases on every sufficiently small alphabet)**.** *Let $m \geq 2$ and let $p_{\min}$ be the smallest prime divisor of m. For every prime $p \leq p_{\min}$, every n, and every $0 < \varepsilon \leq 1$ there is an explicit source X on $\mathbb{F}_p^n$ with*

$$\mathrm{Err}_X\big(\mathsf{Local}_n(p,m)\big) \leq \varepsilon$$

and seed length $O_{p,m}(\log n + \log \frac{1}{\varepsilon})$. In particular this holds for the Boolean alphabet $p = 2$, since $2 \leq p_{\min}$ always.

Proof. Write the prime factorisation of m with multiplicity, in nondecreasing order,

$$m = p_1 p_2 \cdots p_k, \qquad p_1 \leq p_2 \leq \cdots \leq p_k,$$

and define the suffix moduli $M_j := p_j p_{j+1} \cdots p_k$, so that $M_1 = m$ and $M_{k+1} = 1$. We prove by descending induction on j the following invariant.

> There is a constant C_j, depending only on the suffix M_j, such that for every n and every $0 < \eta \leq 1$ one can explicitly η-fool $\mathsf{Local}_n(p_j, M_j)$ with seed length at most $C_j\left(\log n + \log \frac{1}{\eta}\right)$.

Base. At level k the class $\mathsf{Local}_n(p_k, 1)$ contains only the constant function 1 and is fooled exactly by the constant-zero source, which uses no seed. Applying Theorem 4.1 with $p = p_k$ and $m = 1$ gives the invariant for $\mathsf{Local}_n(p_k, p_k) = \mathsf{Local}_n(p_k, M_k)$.

Step. Assume the invariant at level $j+1$, that is, for $\mathsf{Local}_n(p_{j+1}, M_{j+1})$ over the alphabet $\mathbb{F}_{p_{j+1}}$. There are two cases, according to whether the next prime repeats.

Case 1: $p_j = p_{j+1}$. The alphabet is already correct, and no transfer is needed. Theorem 4.1, applied with the alphabet p_j and the modulus M_{j+1}, inserts one further copy of p_j and yields the invariant for

$$\mathsf{Local}_n(p_j, p_j M_{j+1}) = \mathsf{Local}_n(p_j, M_j).$$

Case 2: $p_j < p_{j+1}$. First apply Theorem 4.6 with $r = p_{j+1}$, $p = p_j$ and $m = M_{j+1}$, which is legitimate because $p_j < p_{j+1}$; this yields an optimal-order source for $\mathsf{Local}_n(p_j, M_{j+1})$, over the smaller alphabet and with the modulus unchanged. Then apply Theorem 4.1 as in Case 1 to reach $\mathsf{Local}_n(p_j, M_j)$.

In both cases the seed length acquires one further additive $O(\log n + \log \frac{1}{\eta})$ term and the error is replaced by a fixed power or a fixed constant multiple of itself, so the invariant propagates from level $j+1$ to level j with a new constant depending only on the primes and moduli at that level.

Running the induction down to $j = 1$ gives an explicit optimal-order source for $\mathsf{Local}_n(p_1, m)$. Finally let $p \leq p_{\min} = p_1$ be the requested alphabet. If $p = p_1$ there is nothing more to do. If $p < p_1$, one further application of Theorem 4.6, with $r = p_1$ and m unchanged, moves the source down to the alphabet $\mathbb{F}_p$. □

Why the primes are visited in this order. Two constraints fix the order, and between them they leave no choice. LIFT_p inserts the prime p only when the alphabet already equals p, so the alphabet must match each factor at the moment it is inserted. $\mathrm{TRANSFER}_{r\to p}$ requires a surjection $\mathbb{F}_r \to \mathbb{F}_p$ and therefore $p \leq r$, so the alphabet can only ever decrease. Consequently the largest prime must be handled first, over its own alphabet, and the walk descends through the prime factors, arriving at the Boolean alphabet last. This is also why the base case is always of the form $(p_k, 1)$ with the modulus trivial, and never requires an imported generator.

5.3 From local phases to Boolean modular sums

Theorem 5.2 (Character-error generator)**.** *For every fixed $M \geq 2$, every n, and every $0 < \gamma \leq 1$ there is an explicit source G on $\{0,1\}^n$ with seed length $O_M(\log n + \log \frac{1}{\gamma})$ such that, for every $a \in \mathbb{Z}_M^n$,*

$$\boxed{\left| \mathbb{E}\, \omega_M^{\sum_i a_i G_i} - \mathbb{E}\, \omega_M^{\sum_i a_i U_i} \right| \leq \gamma.}$$

Proof. Since $2 \leq p_{\min}(M)$ for every $M \geq 2$, Theorem 5.1 with $p = 2$ and $m = M$ provides an explicit source G on $\mathbb{F}_2^n = \{0,1\}^n$ with $\mathrm{Err}_G(\mathsf{Local}_n(2, M)) \leq \gamma$ and the stated seed length. Given $a \in \mathbb{Z}_M^n$, Remark 2.2 exhibits $\phi_a^{(M)}$ as a member of $\mathsf{Local}_n(2, M)$, namely the local phase with $q_i(0) = 0$ and $q_i(1) = a_i$. The displayed inequality is then the definition of Err_G applied to that member. □

5.4 Fourier conversion to statistical distance

It remains to pass from the character estimate of Theorem 5.2 to the statement about distributions. This is the output-side Fourier analysis of §2.5, and it is the one place where the transform on $\mathbb{Z}_M$ is used. For a bit distribution X and $a \in \mathbb{Z}_M^n$ write

$$\mathrm{bias}_X^{(M)}(a) := \mathbb{E}\big[\phi_a^{(M)}(X)\big] = \mathbb{E}\Big[\omega_M^{S_a^{(M)}(X)}\Big].$$

Lemma 5.3 (Fourier inversion converts character error to statistical distance)**.** *Let X be a bit distribution such that $\left|\mathrm{bias}_X^{(M)}(b) - \mathrm{bias}_{U_n}^{(M)}(b)\right| \leq \gamma$ for every $b \in \mathbb{Z}_M^n$. Then, for every $a \in \mathbb{Z}_M^n$,*

$$\mathrm{dist}\big(S_a^{(M)}(X),\, S_a^{(M)}(U_n)\big) \leq \frac{1}{2}\sqrt{M-1}\,\gamma.$$

Proof. Fix a, and let P, Q be the probability mass functions of $S_a^{(M)}(X)$ and $S_a^{(M)}(U_n)$ on $\mathbb{Z}_M$. For $t \in \mathbb{Z}_M$,

$$\widehat{P}(t) = \sum_{z \in \mathbb{Z}_M} P(z)\omega_M^{tz} = \mathbb{E}\Big[\omega_M^{tS_a^{(M)}(X)}\Big] = \mathrm{bias}_X^{(M)}(ta),$$

and similarly $\widehat{Q}(t) = \mathrm{bias}_{U_n}^{(M)}(ta)$. At $t = 0$ both equal 1, so the coefficients agree exactly. For $t \neq 0$ the vector ta is another element of $\mathbb{Z}_M^n$, so the hypothesis gives $|\widehat{P}(t) - \widehat{Q}(t)| \leq \gamma$; this is the only place where it is used that the hypothesis is uniform over all coefficient vectors, and not merely over the given a.

Parseval on $\mathbb{Z}_M$ now gives

$$\sum_{z \in \mathbb{Z}_M} |P(z) - Q(z)|^2 = \frac{1}{M} \sum_{t \in \mathbb{Z}_M} |\widehat{P}(t) - \widehat{Q}(t)|^2 \leq \frac{M-1}{M}\gamma^2,$$

and therefore, by Cauchy–Schwarz over the M residues,

$$\mathrm{dist}(P, Q) = \frac{1}{2}\sum_z |P(z) - Q(z)| \leq \frac{1}{2}\sqrt{M}\left(\sum_z |P(z) - Q(z)|^2\right)^{1/2} \leq \frac{1}{2}\sqrt{M-1}\,\gamma.$$

□

Theorem 5.4 (Main theorem). *For every fixed integer $M \geq 2$, every $n \geq 1$, and every $0 < \varepsilon \leq \frac{1}{2}$ there is an explicit generator*

$$G_{M,n,\varepsilon} : \{0,1\}^s \to \{0,1\}^n, \qquad s = O_M\left(\log n + \log \frac{1}{\varepsilon}\right),$$

whose output G satisfies, for every $a \in \mathbb{Z}_M^n$,

$$\operatorname{dist}\left(S_a^{(M)}(G), S_a^{(M)}(U_n)\right) \leq \varepsilon$$

Proof. Set

$$\gamma := \frac{2\varepsilon}{\sqrt{M-1}},$$

which satisfies $0 < \gamma \leq 1$ because $\varepsilon \leq \frac{1}{2}$ and $M \geq 2$. Let G be the source of Theorem 5.2 for this γ; its seed length is $O_M(\log n + \log \frac{1}{\gamma}) = O_M(\log n + \log \frac{1}{\varepsilon})$, since M is fixed. By construction G satisfies the hypothesis of Lemma 5.3 with parameter γ, so for every $a \in \mathbb{Z}_M^n$,

$$\operatorname{dist}\left(S_a^{(M)}(G), S_a^{(M)}(U_n)\right) \leq \frac{1}{2}\sqrt{M-1}\,\gamma = \varepsilon. \qquad \square$$

5.5 Seed length, explicitness, and optimality

Accounting

Fix M and write $m = M = p_1 \cdots p_k$ as in Theorem 5.1. The construction performs k applications of LIFT, one for each prime factor with multiplicity, and at most k applications of TRANSFER, one for each strict increase in the sorted list together with the final descent to the Boolean alphabet. Both counts depend only on M.

Each LIFT_{p_j} contributes p_j masks and replaces the error ε by $(\varepsilon/11)^{2^{p_j}}$; each TRANSFER contributes one mask and replaces ε by $\varepsilon/4$. Composing a constant number of maps of these two kinds sends the target error γ to an innermost parameter of the form

$$\frac{\gamma^{C_M}}{K_M}, \qquad C_M = \prod_{j=1}^{k} 2^{p_j},$$

with $K_M > 0$ a constant, so that

$$\log \frac{K_M}{\gamma^{C_M}} = O_M\left(\log \frac{1}{\gamma}\right).$$

Every mask is an explicit small-bias distribution over a fixed prime field, whose seed length is logarithmic in n and in the reciprocal of its bias, and the seeds of a constant number of independent components add. The total is therefore $O_M(\log n + \log \frac{1}{\varepsilon})$, as claimed.

Explicitness is immediate: each alphabet map σ is a fixed finite lookup table, each mask is sampled by an explicit small-bias construction, and the components are combined by coordinatewise addition in the relevant prime field. Given a seed, the output is computable in time polynomial in n.

The size of the hidden constant. The constant $O_M(\cdot)$ hides two separate sources of growth, and it is worth being explicit about both. The first is the error tower C_M above, which is exponential in the prime factors of M. The second, and in practice the larger, is the weight threshold W of Theorem 4.6. Already the uniform component forces this: for the Boolean alphabet $\kappa_U = \cos\frac{\pi}{m}$, so $1 - \kappa_U = \Theta(m^{-2})$ and the threshold can be of order $m^2 \log\frac{1}{\varepsilon}$. The proxy constant κ_σ may exceed κ_U and enlarge the modulus-dependent constant further; in the ledger of Appendix A it does. The mask bias $\delta = \varepsilon/(4p^{W/2})$ is correspondingly a large fixed power of ε. Neither quantity depends on n or on ε beyond a constant factor in the logarithm, so the asymptotic claim is unaffected; but the constants are large, and no attempt has been made here to optimise them.

Optimality

The dependence on both n and ε is optimal up to the constant depending on M, and this follows from the corresponding statement for a single prime divisor.

Proposition 5.5 (Reduction to a prime divisor)**.** *Let q be a prime dividing M. Every bit distribution that ε-fools Boolean linear sums modulo M also ε-fools Boolean linear sums modulo q.*

Proof. Given $b \in \mathbb{Z}_q^n$, set $a_i := \frac{M}{q} b_i \bmod M$. For every Boolean input x,

$$\sum_i a_i x_i = \frac{M}{q} \sum_i b_i x_i \pmod{M},$$

and the right-hand side depends on $\sum_i b_i x_i$ only through its residue modulo q. Hence

$$S_a^{(M)}(x) = \iota\big(S_b^{(q)}(x)\big), \qquad \iota : \mathbb{Z}_q \to \mathbb{Z}_M, \quad \iota(z) = \frac{M}{q} z.$$

The map ι is injective, so it preserves statistical distance between distributions on $\mathbb{Z}_q$ and their images. Therefore the distance between the two modulo-q sums equals the distance between the two modulo-M sums with coefficient vector a, which is at most ε by hypothesis. □

Consequently a generator meeting the conclusion of Theorem 5.4 also fools linear sums modulo any fixed prime divisor of M, and the standard lower bound for that problem applies. For even M one may take $q = 2$, where the bound $s = \Omega(\log n + \log\frac{1}{\varepsilon})$ for small-bias generators is classical [AGHP92]; for odd M the same conclusion follows from the corresponding bound for a fixed odd prime modulus, which is what makes the prime-power construction of [LRTV09] optimal. In either case

$$s = \Omega_M\left(\log n + \log\frac{1}{\varepsilon}\right),$$

so the upper bound of Theorem 5.4 is optimal up to the constant depending on M.

6 References

[AGHP92] N. Alon, O. Goldreich, J. Håstad, and R. Peralta, "Simple Constructions of Almost k-wise Independent Random Variables," *Random Structures & Algorithms*, 3(3):289–304, 1992.

[AMN98] Y. Azar, R. Motwani, and J. Naor, "Approximating Probability Distributions Using Small Sample Spaces," *Combinatorica*, 18(2):151–171, 1998. Supplies explicit small-bias distributions over non-binary alphabets, used here for every prime alphabet $p > 2$.

[BV10] A. Bogdanov and E. Viola, "Pseudorandom Bits for Polynomials," *SIAM Journal on Computing*, 39(6):2464–2486, 2010. Preliminary version in *FOCS 2007*.

[GKM18] P. Gopalan, D. M. Kane, and R. Meka, "Pseudorandomness via the Discrete Fourier Transform," *SIAM Journal on Computing*, 47(6):2451–2487, 2018. Preliminary version in *FOCS 2015*.

[LRTV09] S. Lovett, O. Reingold, L. Trevisan, and S. Vadhan, "Pseudorandom Bit Generators That Fool Modular Sums," *RANDOM 2009*. Their optimal-order construction covers every fixed prime-power modulus; the present paper imports nothing from it.

[Lov09] S. Lovett, "Unconditional Pseudorandom Generators for Low-Degree Polynomials," *Theory of Computing*, 5:69–82, 2009. Preliminary version in *STOC 2008*.

[LV25] C. H. Lee and E. Viola, "Pseudorandom Bits for Non-Commutative Programs," *40th Computational Complexity Conference (CCC 2025)*, 9:1–9:22.

[MZ09] R. Meka and D. Zuckerman, "Small-Bias Spaces for Group Products," *RANDOM 2009*.

[NN93] J. Naor and M. Naor, "Small-Bias Probability Spaces: Efficient Constructions and Applications," *SIAM Journal on Computing*, 22(4):838–856, 1993. Explicit small-bias distributions are the only pseudorandomness primitive imported by the upper bound.

[Smo87] R. Smolensky, "Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity," *19th Annual ACM Symposium on Theory of Computing (STOC 1987)*, 77–82.

[Vio09] E. Viola, "The Sum of d Small-Bias Generators Fools Polynomials of Degree d," *Computational Complexity*, 18(2):209–217, 2009. Preliminary version in *CCC 2008*.

A Parameter and seed ledger for $M = 15$

This appendix instantiates every parameter of the construction of §5.1. The purpose is to make the hidden constants of Theorem 5.4 visible in one place for a single modulus; nothing here is used elsewhere in the paper.

Throughout, γ is the target character error of Theorem 5.2. For $M = 15$ the conversion of §5.4 takes

$$\gamma = \frac{2\varepsilon}{\sqrt{14}},$$

so a statistical-distance target ε corresponds to $\gamma \approx 0.5345\,\varepsilon$.

A.1 The alphabet maps and their contraction constants

Take the two surjections to be reduction of the representative:

$$\sigma_{5\to3}(y) = y \bmod 3, \qquad \sigma_{3\to2}(y) = y \bmod 2,$$

whose pushforwards of the uniform distribution are $(\frac{2}{5}, \frac{2}{5}, \frac{1}{5})$ on $\mathbb{F}_3$ and $(\frac{2}{3}, \frac{1}{3})$ on $\mathbb{F}_2$ respectively. Both are surjective, so Lemma 4.4 applies, and each constant is a maximum over a finite set: 125 local exponent maps $\mathbb{F}_3 \to \mathbb{Z}_5$ for step 2 and 225 maps $\mathbb{F}_2 \to \mathbb{Z}_{15}$ for step 4, each with the p translates by β. Direct enumeration of those sets gives the values below. For the Boolean step the maximisation is immediate — a nonconstant $q : \mathbb{F}_2 \to \mathbb{Z}_{15}$ has $q(1) - q(0) = c \neq 0$, and both $|1 + \omega_{15}^c|/2$ and $|2 + \omega_{15}^c|/3$ decrease as the argument of ω_{15}^c moves away from 0, so $c = \pm 1$ is extremal — while for step 2 the maximising patterns concentrate as much mass as possible on one root and put the remaining mass on an adjacent one. Nothing downstream depends on these being exact maxima: Theorem 4.6 needs only some explicit value below 1, and each entry may equally be read as such a bound.

step	move	κ_U	κ_σ	κ
2	$\mathrm{TRANSFER}_{5\to3}$, $m = 5$	$\dfrac{\lvert 2+\omega_5\rvert}{3} = 0.832404$	$\dfrac{\lvert 4+\omega_5\rvert}{5} = 0.882545$	0.882545
4	$\mathrm{TRANSFER}_{3\to2}$, $m = 15$	$\cos\dfrac{\pi}{15} = 0.978148$	$\dfrac{\lvert 2+\omega_{15}\rvert}{3} = 0.980600$	0.980600

Both κ's are strictly below 1, as they must be, but the second is close to it: this is the quantitative price of the large modulus, and it is what drives the weight threshold W of Theorem 4.6.

A.2 The error ledger

The parameters are determined from the target inwards. Reading Theorem 4.6 backwards replaces the error by $\varepsilon/4$, and reading Theorem 4.1 backwards replaces it by $(\varepsilon/11)^{2^p}$. Composing the four steps of §5.1 gives the following.

step	move	state reached	required incoming error
4	$\mathrm{TRANSFER}_{3\to 2}$	$(2,15)$	$\varepsilon_3 = \gamma/4$
3	LIFT_3	$(3,15)$	$\varepsilon_2 = (\gamma/44)^8$
2	$\mathrm{TRANSFER}_{5\to 3}$	$(3,5)$	$\varepsilon_1 = (\gamma/44)^8/4$
1	LIFT_5	$(5,5)$	$\theta_1 = \gamma^{256}/44^{288}$

The constant $44 = 4 \cdot 11$ appears because a TRANSFER and a LIFT are composed at each of the two odd primes: the factor 4 comes from $\eta = \varepsilon/4$ and the factor 11 from $\theta = (\varepsilon/11)^{2^p}$. The exponent is

$$256 = 2^3 \cdot 2^5 = \prod_j 2^{p_j} \qquad \text{for } 15 = 3 \cdot 5,$$

in agreement with the general formula for C_M in §5.5. Numerically $44^{288} = 10^{473.31}$, so

$$\theta_1 = \gamma^{256} \cdot 10^{-473.31}.$$

The final line of the table is a requirement on the base source of §5.1, which has error exactly 0 and therefore meets it with room to spare. No generator is imported at the base; the requirement is vacuous.

A.3 The masks

The construction uses ten masks. Their biases are prescribed by the theorem that introduces each one: a LIFT_p contributes p masks of bias equal to its own incoming error θ, and a TRANSFER contributes one mask of bias $\delta = \varepsilon/(4p^{W/2})$ with $W = \lceil \log(8/\varepsilon)/\log(1/\kappa) \rceil$.

count	alphabet	introduced by	bias	$\log_2(1/\text{bias})$
5	$\mathbb{F}_5^n$	LIFT_5	$\theta_1 = \gamma^{256}/44^{288}$	4354.9
1	$\mathbb{F}_3^n$	$\mathrm{TRANSFER}_{5\to 3}$	$\delta_2 = \varepsilon_2/(4 \cdot 3^{W_2/2})$	720.7
3	$\mathbb{F}_3^n$	LIFT_3	$\theta_3 = \varepsilon_2 = (\gamma/44)^8$	130.6
1	$\mathbb{F}_2^n$	$\mathrm{TRANSFER}_{3\to 2}$	$\delta_4 = \gamma/(4 \cdot 2^{W_4/2})$	258.4

The final column is evaluated at $\varepsilon = 10^{-3}$, that is $\gamma = 5.345 \cdot 10^{-4}$, where the two thresholds are

$$W_2 = 742, \qquad W_4 = 491.$$

Because W_2 and W_4 are defined with ceilings, the suppression factors are not exact powers, but they are pinned within a bounded factor:

$$\left(\frac{8}{\gamma}\right)^{c_4} \le 2^{W_4/2} < \sqrt{2}\left(\frac{8}{\gamma}\right)^{c_4}, \qquad \left(\frac{8}{\varepsilon_2}\right)^{c_2} \le 3^{W_2/2} < \sqrt{3}\left(\frac{8}{\varepsilon_2}\right)^{c_2},$$

where $c_4 = \log 2/(2\log(1/\kappa_4)) = 17.69$ and $c_2 = \log 3/(2\log(1/\kappa_2)) = 4.40$. So each transfer mask has a bias that is, up to a factor below $\sqrt{3}$, a fixed power of the error it is required to achieve.

A.4 The bias ledger

Each mask is an explicit small-bias distribution over its own prime field, with seed length $O_p(\log n + \log_2 \frac{1}{\text{bias}})$, and the ten seeds are independent, so they add. The quantity that governs the total is therefore the *bias ledger*

$$\mathcal{B} := \sum_{\text{masks}} \log_2 \frac{1}{\text{bias}},$$

in terms of which the seed length is $O_M(\log n + \mathcal{B})$. We emphasise that $\mathcal{B}$ is not itself a number of seed bits: the constant relating the two is fixed by whichever small-bias construction is used, and we do not fix one here. Summing the last column with multiplicities,

$$\mathcal{B} \approx 2.31 \cdot 10^4 \qquad \text{at } \varepsilon = 10^{-3}.$$

As a function of ε,

$$\mathcal{B} = (5 \cdot 256 + 8(1 + c_2) + 3 \cdot 8 + (1 + c_4)) \log_2 \frac{1}{\gamma} + O(1) \approx 1365.9 \log_2 \frac{1}{\gamma} + 8.30 \cdot 10^3,$$

with the leading coefficient dominated by the five LIFT_5 masks. The second form is approximate rather than exact because W_2 and W_4 are defined with ceilings, so $\mathcal{B}$ is affine in $\log_2 \frac{1}{\gamma}$ only up to the $O(1)$ term already displayed.

What these numbers do and do not say. Every quantity above is a constant multiple of $\log \frac{1}{\varepsilon}$ plus a constant, with no dependence on n beyond the $\log n$ term carried by each small-bias construction. Theorem 5.4 is therefore unaffected: the seed length is $O_M(\log n + \log \frac{1}{\varepsilon})$. What the ledger shows is the size of the two parts of that statement for $M = 15$: a coefficient of order 10^3 multiplying $\log \frac{1}{\varepsilon}$, and an additive term of order 10^3 in $\mathcal{B}$. No attempt has been made to optimise either. Two reductions are visible from the ledger and would be worth pursuing. The exponent 2^p in LIFT_p comes from applying Lemma 3.1 p times, each costing a square root; any improvement of that lemma to a smaller loss would reduce C_M geometrically. And $c_4 = 17.69$ is driven by $\kappa \to 1$ as the modulus grows, so a sharper high-weight estimate in Theorem 4.6, or a better choice of σ, would reduce the transfer masks directly.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il