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

Optimal Bitwise Pseudorandom Generators for Modular Sums Modulo $2^t \cdot p^k$

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 extend this result to every fixed modulus of the form

$$M = 2^t \cdot p^k,$$

where $t, k \geq 0$ and p an odd prime. Specifically, 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.

Our main contribution is a modulus-doubling reduction that transforms any generator fooling Boolean linear sums modulo m into one fooling such sums modulo $2m$, while preserving the optimal asymptotic seed length.

[†]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 $S_a^{(M)}(U_n)$. This is a requirement on the entire residue distribution: for a fixed coefficient vector it amounts to controlling all $M-1$ nontrivial characters of the sum. Uniform control of the basic character over *all* coefficient vectors nevertheless suffices, since the tth character attached to a is the basic character attached to ta; this is the first step of the proof. 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 extend the optimal-seed regime from prime powers to every fixed modulus with at most one odd prime factor.

Theorem 1.1 (Main result; informal)**.** *Let $M = 2^t q$, where $t \geq 0$ and q is either 1 or a power p^k of an odd prime. 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 seed length is optimal up to the constant depending on M. For $t \geq 1$ this is immediate: parity tests embed into linear sums modulo any even M through coefficient vectors with entries in $\{0, M/2\}$, and already for parity a seed of length $\Omega(\log n + \log \frac{1}{\varepsilon})$ is necessary [AGHP92]. For $t = 0$ the modulus is an odd prime power and both the generator and its optimality are inherited from [LRTV09].

The generator itself is built from off-the-shelf components. Our contribution is entirely in the analysis: we prove a *modulus-doubling reduction* showing that if Z fools linear sums modulo m and B is an independent parity-small-bias vector [NN93], then $Z \oplus B$ fools linear sums modulo $2m$, with only a polynomial loss in the error and an additive $O(\log n + \log \frac{1}{\varepsilon})$ in the seed. Each doubling step therefore has the form

$$G_0 := Z, \qquad G_j := G_{j-1} \oplus B_j,$$

so that the final generator

$$G_t = Z \oplus B_1 \oplus \cdots \oplus B_t$$

is the output of the LRTV generator for the odd part q, run at polynomially smaller error, XORed with one independently parameterised small-bias mask for each factor of 2. Applying the reduction t times along the chain

$$q \longrightarrow 2q \longrightarrow \cdots \longrightarrow 2^t q$$

yields the theorem; when $q = 1$ the same chain is started from parity instead. (An XOR of independent small-bias vectors is itself small-bias, so G_t does have the two-component shape $Z \oplus B$; but the proof needs the individual masks, since level j uses that G_{j-1} fools linear sums modulo M_{j-1}.)

1.2 Related work

Modular sums. Naor and Naor [NN93] solved the case $M = 2$ optimally. Lovett, Reingold, Trevisan, and Vadhan [LRTV09] gave two constructions: one with optimal seed length for every fixed prime power, and one for arbitrary M at the seed length displayed above, which for a fixed modulus is the strongest bound known. 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 our theorem does, for the moduli $2^t q$, is remove the residual $\log\log \frac{1}{\varepsilon}$ loss of the LRTV arbitrary-modulus construction. For moduli with two distinct odd prime factors, of which $M = 15$ is the smallest, that loss remains, and the method of this paper does not apply (Section 4.4).

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. Our result gives the exactly optimal seed length for the commutative groups $\mathbb{Z}_{2^t q}$, which are not p-groups whenever $t \geq 1$ and $q > 1$. By a standard character argument it should also yield the corresponding statement for every finite abelian group whose exponent has this form, with the error adjusted by a constant depending on the group; we do not pursue that here.

The Meka–Zuckerman conjecture. Our generator is not new as an object. For $M = 6$ it coincides with the generator that Meka and Zuckerman conjectured to fool width-3 read-once permutation branching programs, namely the XOR of a small-bias space with a pseudorandom bit generator for $\mathbb{F}_3$ (Conjecture 1.5 of [MZ09]). That conjecture remains open. Our theorem 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.

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 proof runs the same skeleton with the *modulus* as the descent parameter: Cauchy–Schwarz introduces an XOR derivative, the derivative of a modulo-$2m$ phase is a modulo-m phase, and the imported lower-modulus generator finishes. The base case is foreign — an arbitrary modulo-q generator rather than small bias — and this is what lets the reduction graft onto [LRTV09]. The resemblance is more than formal 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 XOR derivative strips the powers of two and then halts (Section 4.4).

Organization. Section 2 introduces the bias framework and the proof plan. Section 3 proves the two estimates and the high-weight smoothing theorem. Section 4 assembles the one-step modulus-doubling reduction, iterates it through the full power of two, derives the main theorem from the LRTV prime-power generator, and proves that the method stops at the odd part.

2 The problem and the central idea

2.1 The problem and the modulus-doubling goal

Fix a modulus $M \geq 2$. The input space is the Boolean group

$$\mathbb{F}_2^n = \{0,1\}^n,$$

with coordinatewise XOR denoted by $\oplus$. For $a = (a_1, \ldots, a_n) \in \mathbb{Z}_M^n$ and $x = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$, define

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

Let U_n denote n independent uniform bits. For random variables Y, Y' on a finite set, write

$$\mathrm{dist}(Y, Y') := \frac{1}{2} \sum_{y} \left| \Pr[Y = y] - \Pr[Y' = y] \right|.$$

A bit distribution X ε-fools Boolean linear sums modulo M when, for every $a \in \mathbb{Z}_M^n$,

$$\mathrm{dist}\big(S_a^{(M)}(X), S_a^{(M)}(U_n)\big) \leq \varepsilon.$$

The reduction starts with a bit distribution Z that already fools all sums modulo m. It samples an independent parity-small-bias vector $B \in \mathbb{F}_2^n$ and outputs

$$\boxed{G := Z \oplus B.}$$

The theorem goal is

$$\boxed{\text{PRG modulo } m \quad + \quad \text{parity-small-bias XOR mask} \quad \Longrightarrow \quad \text{PRG modulo } 2m.}$$

Imported LRTV theorem. For every fixed prime p and integer $k \geq 1$, there is an explicit pseudorandom generator that ε-fools all Boolean linear sums modulo p^k with seed length

$$\boxed{r_{p^k}(n, \varepsilon) = O_{p^k}\left(\log n + \log \frac{1}{\varepsilon} \right).}$$

Here the hidden constant may depend on the fixed modulus p^k, but not on n or ε.

The extension theorem developed in this document aims to show that an optimal-order generator modulo m can be lifted to an optimal-order generator modulo $2m$:

$$\boxed{\text{optimal PRG modulo } m \quad \Longrightarrow \quad \text{optimal PRG modulo } 2m.}$$

Starting from the LRTV generator modulo p^k and applying this lifting theorem t times gives the chain

$$p^k \longrightarrow 2p^k \longrightarrow 2^2 p^k \longrightarrow \cdots \longrightarrow 2^t p^k.$$

Consequently, the proposed extension yields the following result.

Consequence of the modulus-doubling theorem. For every fixed modulus of the form

$$M = 2^t p^k,$$

where p is prime, $k \geq 1$, and $t \geq 0$, there is an explicit pseudorandom generator that ε-fools all Boolean linear sums modulo M with seed length

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

2.2 Roots of unity, the phase function, and modular bias

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. It 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, define the associated *phase function*

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

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.

For a bit distribution X on $\mathbb{F}_2^n$, define its modular bias against a by

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

Equivalently, grouping the expectation according to the value of the sum gives

$$\boxed{\mathrm{bias}_X^{(M)}(a) = \sum_{r \in \mathbb{Z}_M} \Pr\left[S_a^{(M)}(X) = r\right] \omega_M^r.}$$

Hence $\mathrm{bias}_X^{(M)}(a)$ is the probability-weighted average of the Mth roots of unity selected by $S_a^{(M)}(X)$. It measures the extent to which those phases fail to cancel. In particular, if $S_a^{(M)}(X)$ is fully balanced, meaning that every residue $r \in \mathbb{Z}_M$ occurs with probability $1/M$, then

$$\mathrm{bias}_X^{(M)}(a) = \frac{1}{M} \sum_{r \in \mathbb{Z}_M} \omega_M^r = 0.$$

This is the average phase selected by the random modular sum. If the residues are balanced around $\mathbb{Z}_M$, their roots of unity cancel and the bias is small. Passing to phases therefore turns modular arithmetic into multiplicative algebra and turns residue balance into a complex expectation.

2.3 Proof plan

Fix

$$M = 2m.$$

Assume that $Z \in \mathbb{F}_2^n$ is the output of a generator that already fools Boolean linear sums modulo m. Independently sample a parity-small-bias mask $B \in \mathbb{F}_2^n$, and define the candidate modulo-M generator by

$$\boxed{G := Z \oplus B.}$$

Our goal is to prove that, for every coefficient vector $a \in \mathbb{Z}_M^n$, the modular sums

$$S_a^{(M)}(G) \qquad \text{and} \qquad S_a^{(M)}(U_n)$$

are close in statistical distance.

The proof has three main steps.

Step 1: reduce statistical distance to bias control.

The first step is to replace the distributional goal by an analytic one. For a fixed coefficient vector a, the Fourier coefficients of the output distribution $S_a^{(M)}(X)$ are exactly

$$\mathrm{bias}_X^{(M)}(ta), \qquad t \in \mathbb{Z}_M.$$

Consequently, if a bit distribution X satisfies

$$\left|\mathrm{bias}_X^{(M)}(b) - \mathrm{bias}_{U_n}^{(M)}(b)\right| \le \gamma$$

for every coefficient vector $b \in \mathbb{Z}_M^n$, then Parseval and Cauchy–Schwarz imply that every modular-sum output distribution under X is close to its counterpart under U_n. In particular, bit-bias error γ yields statistical-distance error at most

$$\frac{1}{2}\sqrt{M-1}\,\gamma.$$

Thus, after this reduction, the remainder of the proof focuses entirely on uniformly bounding

$$\left|\mathrm{bias}_G^{(M)}(a) - \mathrm{bias}_{U_n}^{(M)}(a)\right|.$$

This reduction is not new. It is exactly the argument of [LRTV09], and we reproduce it in full in Section 3.1 only for completeness; the contribution of the present paper lies entirely in Steps 2 and 3.

Step 2: prove two complementary bounds on the bias error.

Let

$$w := \mathrm{wt}(a).$$

The central analytic result of the proof is that the bias error is bounded by the better of two estimates:

$$\boxed{\left|\mathrm{bias}_G^{(M)}(a) - \mathrm{bias}_{U_n}^{(M)}(a)\right| \le \min\{L(w), H(w)\},}$$

where

$$\boxed{L(w) := \delta \Lambda_M^w}$$

and

$$\boxed{H(w) := \sqrt{\kappa_M^{2w} + \delta^2 + \eta} + \kappa_M^w.}$$

Here:

- δ is the parity-bias parameter of the mask B;
- η is the modulo-m bit-bias error of Z;
- $0 \le \kappa_M < 1$ measures the exponential decay of the uniform-bit modular bias;
- $1 \le \Lambda_M \le \sqrt{2}$ controls the Fourier ℓ^1-mass contributed by one active coordinate.

The first estimate is useful when w is small, because it pays the Fourier ℓ^1-mass Λ_M^w. The second estimate is useful when w is large, because the factors κ_M^w then become small. Thus the minimum expresses the basic low-weight/high-weight division of the proof without committing in advance to one estimate for every coefficient vector.

Step 3: choose the threshold and error parameters.

After proving the combined inequality, the final step is to choose a threshold W and parameters δ, η so that one of the two bounds is small for every possible weight.

Choose W so that

$$\kappa_M^W$$

is at most a small constant multiple of the desired bias error γ. Since $\kappa_M < 1$, this requires only

$$W = O_M\left(\log \frac{1}{\gamma}\right).$$

Then choose the mask error δ so that

$$\delta \Lambda_M^W$$

is small, and choose the lower-modulus error η on the order of γ^2.

With these choices:

- when $w \le W$, use the low-weight estimate

$$L(w) = \delta \Lambda_M^w \le \delta \Lambda_M^W;$$

- when $w > W$, use the high-weight estimate, because

$$\kappa_M^w \le \kappa_M^W,$$

while δ and $\sqrt{\eta}$ have also been chosen to be small.

Thus every coefficient vector satisfies

$$\left|\mathrm{bias}_G^{(M)}(a) - \mathrm{bias}_{U_n}^{(M)}(a)\right| \le \gamma.$$

Step 1 then converts this bit-bias guarantee into statistical-distance error at most ε.

Finally, because

$$W = O_M\left(\log \frac{1}{\varepsilon}\right)$$

and the small-bias mask has seed length

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

the lifting step preserves the desired seed-length form

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

The reduction can then be iterated through the power-of-two part of the modulus.

Proof architecture. The proof first converts bias control into statistical-distance control. It then proves

$$\boxed{\text{bias error} \le \min\left\{\underbrace{\delta\Lambda_M^w}_{\text{low-weight estimate}}, \underbrace{\sqrt{\kappa_M^{2w} + \delta^2 + \eta} + \kappa_M^w}_{\text{high-weight estimate}}\right\},}$$

and finally chooses the threshold and parameters so that the appropriate term is small at every weight.

2.4 General Fourier analysis on a finite abelian group

Before introducing the two Fourier transforms used in this proof, we briefly describe the common framework behind them.

Let H be a finite abelian group, written additively. A *character* of H is a function

$$\chi : H \to \mathbb{C}$$

such that

$$\chi(x+y) = \chi(x)\chi(y) \qquad \text{for all } x, y \in H.$$

Thus a character converts the group operation into multiplication. Since H is finite, every character takes values on the complex unit circle, so

$$|\chi(x)| = 1, \qquad \chi(-x) = \overline{\chi(x)}.$$

The set of all characters is denoted by $\widehat{H}$ and is called the *character group*, or *dual group*, of H. It is itself an abelian group under pointwise multiplication, and for finite abelian H one has

$$|\widehat{H}| = |H|.$$

Characters as a multiplicative basis. The complex-valued functions $f : H \to \mathbb{C}$ form a vector space of dimension $|H|$. Under the normalized inner product

$$\langle f, g \rangle := \mathbb{E}_{x \sim U_H} \left[f(x) \overline{g(x)} \right],$$

the characters satisfy the orthogonality relation

$$\mathbb{E}_{x \sim U_H} \left[\chi(x) \overline{\psi(x)} \right] = \begin{cases} 1, & \chi = \psi, \\ 0, & \chi \neq \psi. \end{cases}$$

Because there are exactly $|H|$ characters, they form an orthonormal basis for all functions $H \to \mathbb{C}$. Consequently, every function has a unique representation as a linear combination of characters.

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, XOR, or convolution.

Fourier coefficients, inversion, and uniqueness. Using the positive-character convention of this document, define

$$\widehat{f}(\chi) := \mathbb{E}_{x \sim U_H} [f(x) \chi(x)].$$

The *Fourier transform* of f is the complete collection

$$\widehat{f} = \left(\widehat{f}(\chi) \right)_{\chi \in \widehat{H}}.$$

The inverse Fourier transform reconstructs the original function:

$$\boxed{f(x) = \sum_{\chi \in \widehat{H}} \widehat{f}(\chi) \overline{\chi(x)}.}$$

The coefficients are unique because the characters form an orthonormal basis. Therefore

$$\widehat{f} = \widehat{g} \quad \Longleftrightarrow \quad f = g.$$

The Fourier transform does not discard information; it expresses the same function in a different coordinate system. Some texts place a complex conjugate in the definition of the forward transform instead. This merely replaces each character by its inverse and does not change the theory.

Parseval's identity states that this change of coordinates preserves squared norm:

$$\boxed{\mathbb{E}_{x \sim U_H} |f(x)|^2 = \sum_{\chi \in \widehat{H}} |\widehat{f}(\chi)|^2.}$$

More generally,

$$\mathbb{E}_{x\sim U_H}|f(x)-g(x)|^2=\sum_{\chi\in\widehat{H}}|\widehat{f}(\chi)-\widehat{g}(\chi)|^2.$$

Thus controlling all Fourier coefficients controls the ℓ^2-distance between the original functions.

Why the character basis is adapted to the group law. For $b\in H$, define the translated function $(\tau_b f)(x):=f(x+b)$. Since

$$\chi(x+b)=\chi(x)\chi(b),$$

translation does not mix different Fourier components; it only multiplies each component by a scalar. Similarly, convolution and addition of independent random variables become coefficientwise multiplication in the Fourier domain. In this sense,

group operations become simple scalar operations in Fourier coordinates.

For the Boolean group $\mathbb{F}_2^n$, the group operation is XOR, so its characters form exactly the basis in which XOR shifts and XOR averaging are diagonal.

Fourier transforms of probability distributions. A probability distribution on H may be represented by its probability mass function

$$P:H\to[0,1],\qquad P(x)=\Pr[Y=x].$$

It is therefore a special kind of function on H. For probability laws, the natural normalization is

$$\boxed{\widehat{P}(\chi):=\mathbb{E}[\chi(Y)]=\sum_{x\in H}P(x)\chi(x).}$$

This coefficient is the average value of the multiplicative test χ under the distribution. At the trivial character $\mathbf{1}$,

$$\widehat{P}(\mathbf{1})=1,$$

which records that the total probability mass is 1. For a nontrivial character, a large coefficient indicates alignment with that character, while a small coefficient indicates cancellation. For the uniform distribution U_H,

$$\widehat{U_H}(\chi)=\begin{cases}1, & \chi=\mathbf{1},\\ 0, & \chi\neq\mathbf{1}.\end{cases}$$

The complete probability law can be reconstructed by Fourier inversion:

$$\boxed{P(x)=\frac{1}{|H|}\sum_{\chi\in\widehat{H}}\widehat{P}(\chi)\overline{\chi(x)}.}$$

Hence the Fourier coefficients form a complete fingerprint of the distribution. Parseval gives the quantitative identity

$$\sum_{x\in H}|P(x)-Q(x)|^2=\frac{1}{|H|}\sum_{\chi\in\widehat{H}}|\widehat{P}(\chi)-\widehat{Q}(\chi)|^2.$$

This is why controlling all Fourier coefficients of two distributions allows us to control the distance between the distributions themselves.

If X and Y are independent random variables on H, then the multiplicative property of characters gives

$$\widehat{P_{X+Y}}(\chi) = \mathbb{E}[\chi(X+Y)] = \mathbb{E}[\chi(X)]\mathbb{E}[\chi(Y)] = \widehat{P_X}(\chi)\widehat{P_Y}(\chi).$$

Thus addition of independent variables, or equivalently convolution of their laws, becomes multiplication in the Fourier domain.

The two groups used in this proof. We apply this same framework to two different finite abelian groups. On the output group $\mathbb{Z}_M$, the characters are

$$\psi_t(r) = \omega_M^{tr}, \qquad t \in \mathbb{Z}_M.$$

On the input group $\mathbb{F}_2^n$, the characters are the parity functions

$$\chi_T(x) = (-1)^{T\cdot x}, \qquad T \in \mathbb{F}_2^n.$$

The theory is identical in both cases. What changes is the object being transformed and the role of the transform in the proof: on $\mathbb{Z}_M$ we transform the probability distribution of the modular output, while on $\mathbb{F}_2^n$ we transform the general complex-valued phase function of the input bits. The next section compares these two uses directly.

2.5 Two Fourier transforms on two different groups

Before writing either transform, it is important to understand why the proof needs two of them.

The two questions answered by the two transforms. Output-side Fourier analysis asks: what must be proved? The desired conclusion compares two probability distributions on $\mathbb{Z}_M$. Their Fourier coefficients are precisely the biases $\mathrm{bias}_X^{(M)}(ta)$. Hence controlling all biases controls statistical distance by Parseval and Cauchy–Schwarz.

Input-side Fourier analysis asks: how does the XOR mask act? The mask B comes with guarantees only against parity tests $(-1)^{T\cdot x}$. The modular phase $\phi_a^{(M)}(x)$ is not itself a parity. We therefore decompose this general complex function into parity components. In that basis, XOR averaging becomes coefficientwise multiplication, so the effect of the mask is transparent.

The two transforms arise from the map

$$\mathbb{F}_2^n \xrightarrow{S_a^{(M)}} \mathbb{Z}_M.$$

On the output group $\mathbb{Z}_M$, we transform the probability distribution of the resulting residue. On the input group $\mathbb{F}_2^n$, we transform the general complex-valued phase function associated with the test.

Output-side transform: the distribution of the modular sum. Fix a bit distribution X on $\mathbb{F}_2^n$ and a coefficient vector

$$a = (a_1, \ldots, a_n) \in \mathbb{Z}_M^n.$$

The random modular sum is

$$S_a^{(M)}(X) \in \mathbb{Z}_M.$$

Let

$$P_{a,X}(r) := \Pr\left[S_a^{(M)}(X) = r\right], \qquad r \in \mathbb{Z}_M,$$

denote its probability mass function. Thus the object being transformed is explicitly the probability distribution

$$P_{a,X} : \mathbb{Z}_M \longrightarrow [0,1].$$

For every $t \in \mathbb{Z}_M$, define the character

$$\psi_t : \mathbb{Z}_M \longrightarrow \mathbb{C}, \qquad \psi_t(r) := \omega_M^{tr}.$$

The scalar t is the frequency label indexing the character ψ_t. Formally, the corresponding Fourier coefficient is attached to the character function ψ_t, and we write

$$\widehat{P_{a,X}}(\psi_t).$$

Using the positive-character convention of this document,

$$\begin{aligned}
\widehat{P_{a,X}}(\psi_t) &:= \mathbb{E}_{r\sim S_a^{(M)}(X)}[\psi_t(r)] \\
&= \mathbb{E}_{r\sim S_a^{(M)}(X)}\left[\omega_M^{tr}\right] \\
&= \sum_{r\in\mathbb{Z}_M} P_{a,X}(r)\omega_M^{tr} \\
&= \mathbb{E}_{x\sim X}\left[\omega_M^{tS_a^{(M)}(x)}\right] \\
&= \mathbb{E}_{x\sim X}\left[\omega_M^{\sum_{i=1}^n (ta_i)x_i}\right] \\
&= \boxed{\mathrm{bias}_X^{(M)}(ta)}.
\end{aligned}$$

Here

$$ta := (ta_1, \ldots, ta_n) \pmod{M}.$$

Thus,

$$\boxed{\widehat{P_{a,X}}(\psi_t) = \mathrm{bias}_X^{(M)}(ta).}$$

The complete Fourier transform of the output distribution is therefore the collection

$$\left(\widehat{P_{a,X}}(\psi_t)\right)_{t\in\mathbb{Z}_M} = \left(\mathrm{bias}_X^{(M)}(ta)\right)_{t\in\mathbb{Z}_M}.$$

For readability, one may abbreviate

$$\widehat{P_{a,X}}(\psi_t) \qquad \text{as} \qquad \widehat{P_{a,X}}(t),$$

but the formal meaning is always "the coefficient corresponding to the character ψ_t."

This output-side transform tells us *what must be controlled*. For a fixed a, if

$$\widehat{P_{a,X}}(\psi_t)$$

is close to the corresponding coefficient

$$\widehat{P_{a,U_n}}(\psi_t)$$

for every character ψ_t, then Parseval and Cauchy–Schwarz imply that the two output distributions

$$S_a^{(M)}(X) \qquad \text{and} \qquad S_a^{(M)}(U_n)$$

are close in statistical distance. Equivalently, controlling

$$\left|\mathrm{bias}_X^{(M)}(ta) - \mathrm{bias}_{U_n}^{(M)}(ta)\right|$$

for all $t \in \mathbb{Z}_M$ controls the complete distribution of the modular sum.

Input-side transform: the phase function on the Boolean cube. We now turn to a different object on a different group. For the same coefficient vector a, recall the phase function

$$\phi_a^{(M)} : \mathbb{F}_2^n \longrightarrow \mathbb{C}, \qquad \phi_a^{(M)}(x) = \omega_M^{S_a^{(M)}(x)}.$$

This is a general complex-valued function, not a probability distribution.

For every vector

$$T = (T_1, \ldots, T_n) \in \mathbb{F}_2^n,$$

define the parity character

$$\chi_T : \mathbb{F}_2^n \longrightarrow \{-1, 1\}, \qquad \chi_T(x) := (-1)^{T \cdot x},$$

where

$$T \cdot x := \sum_{i=1}^{n} T_i x_i \pmod 2.$$

The vector T indexes the character χ_T and specifies a linear parity test on the input bits.

The Fourier coefficient of the phase function corresponding to χ_T is

$$\boxed{\widehat{\phi_a^{(M)}}(\chi_T) := \mathbb{E}_{U_n}\left[\phi_a^{(M)}(U_n)\chi_T(U_n)\right].}$$

Because every parity character is real and satisfies $\overline{\chi_T} = \chi_T$, Fourier inversion gives

$$\phi_a^{(M)}(x) = \sum_{T \in \mathbb{F}_2^n} \widehat{\phi_a^{(M)}}(\chi_T)\chi_T(x).$$

For readability, the coefficient $\widehat{\phi_a^{(M)}}(\chi_T)$ may also be abbreviated as $\widehat{\phi_a^{(M)}}(T)$, with the understanding that T indexes the character χ_T.

The reason for using this basis is that parity characters interact perfectly with XOR:

$$\boxed{\chi_T(z \oplus b) = \chi_T(z)\chi_T(b).}$$

The candidate generator changes the input by

$$z \longmapsto z \oplus B.$$

Substituting the Fourier expansion of the phase gives

$$\begin{aligned}\mathbb{E}_B\big[\phi_a^{(M)}(z \oplus B)\big] &= \mathbb{E}_B\left[\sum_{T \in \mathbb{F}_2^n} \widehat{\phi_a^{(M)}}(\chi_T)\chi_T(z \oplus B)\right] \\ &= \sum_{T \in \mathbb{F}_2^n} \widehat{\phi_a^{(M)}}(\chi_T)\chi_T(z)\mathbb{E}_B[\chi_T(B)].\end{aligned}$$

Define the Fourier coefficient of the mask distribution corresponding to χ_T by

$$\widehat{B}(\chi_T) := \mathbb{E}_B[\chi_T(B)] = \mathbb{E}_B[(-1)^{T \cdot B}].$$

Then

$$\boxed{\mathbb{E}_B\big[\phi_a^{(M)}(z \oplus B)\big] = \sum_{T \in \mathbb{F}_2^n} \widehat{\phi_a^{(M)}}(\chi_T)\widehat{B}(\chi_T)\chi_T(z).}$$

Thus XOR averaging does not mix the different Fourier components. It simply multiplies the coefficient corresponding to χ_T by the scalar $\widehat{B}(\chi_T)$.

A parity-small-bias mask satisfies

$$\left|\widehat{B}(\chi_T)\right| \le \delta \qquad \text{for every } T \neq 0.$$

Therefore the constant component $T = 0$ is preserved because $\widehat{B}(\chi_0) = 1$, while every nonconstant parity component is attenuated by at most δ.

The central intuition for transforming the phase. The small-bias mask understands parities, not arbitrary modular phases. Fourier expansion supplies the bridge

$$\boxed{\text{modular phase = a linear combination of parity tests.}}$$

We transform the phase not to compare two distributions, but to express the test in the basis in which XOR smoothing is diagonal.

Equivalently, for any input distribution X on $\mathbb{F}_2^n$,

$$\begin{aligned}\mathrm{bias}_X^{(M)}(a) &= \mathbb{E}_X\big[\phi_a^{(M)}(X)\big] \\ &= \sum_{T \in \mathbb{F}_2^n} \widehat{\phi_a^{(M)}}(\chi_T)\,\underbrace{\mathbb{E}_X[\chi_T(X)]}_{\widehat{P_X}(\chi_T)}.\end{aligned}$$

The Fourier transform of the phase tells us which parity tests occur in the modular test, while the Fourier transform of the input distribution tells us how much of each parity is present in the source.

3 Proof Details

3.1 Bit-bias error

Definition 3.1 (Modulo-M bit-bias error)**.** *A bit distribution X on $\{0,1\}^n$ has modulo-M bit-bias error at most γ if*

$$\sup_{a\in\mathbb{Z}_M^n}\left|\mathrm{bias}_X^{(M)}(a)-\mathrm{bias}_{U_n}^{(M)}(a)\right|\le\gamma.$$

The word bit *is important: the reference distribution is U_n, not a uniform vector in $\mathbb{Z}_M^n$.*

Lemma 3.2 (Fourier inversion converts bit-bias to statistical distance)**.** *Let X be a bit distribution with modulo-M bit-bias error at most γ. Then, for every $a\in\mathbb{Z}_M^n$,*

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

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

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

and similarly

$$\widehat{Q}(t)=\mathrm{bias}_{U_n}^{(M)}(ta).$$

At $t=0$, the coefficients agree exactly. For $t\neq0$,

$$|\widehat{P}(t)-\widehat{Q}(t)|\le\gamma,$$

because ta is another vector in $\mathbb{Z}_M^n$.

Parseval on $\mathbb{Z}_M$ gives

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

Therefore, by Cauchy–Schwarz,

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

□

3.2 Step 2 setup: importing the XOR-smoothing identities

Step 1 reduced the statistical-distance goal to a uniform bound on

$$\left|\mathrm{bias}_{Z\oplus B}^{(2m)}(a) - \mathrm{bias}_{U_n}^{(2m)}(a)\right|.$$

We now begin Step 2 of the proof plan: proving two complementary estimates for this bias error. Retain the notation

$$M = 2m, \qquad G = Z \oplus B,$$

where $B \in \mathbb{F}_2^n$ is independent of Z and is δ-biased against every nonconstant parity.

A realized mask b specifies exactly which coordinates of Z are flipped:

$$(Z \oplus b)_i = \begin{cases} Z_i, & b_i = 0, \\ 1 - Z_i, & b_i = 1. \end{cases}$$

For a fixed coefficient vector $a \in \mathbb{Z}_{2m}^n$, define the mask-averaged phase

$$h(z) := \mathbb{E}_B\left[\phi_a^{(2m)}(z \oplus B)\right], \qquad z \in \mathbb{F}_2^n.$$

Section 2.5 proved the input-side Fourier identity that governs this averaging. In the parity basis, XOR averaging is diagonal:

$$\boxed{\widehat{h}(T) = \widehat{\phi_a^{(2m)}}(T)\widehat{B}(T) \qquad (T \in \mathbb{F}_2^n).}$$

We import this identity here rather than re-deriving it. Together with the δ-bias condition stated and discussed there, it gives:

- at $T = 0$, $\widehat{B}(0) = 1$, so the constant component is preserved;
- at every $T \neq 0$, $|\widehat{B}(T)| \leq \delta$, so the corresponding nonconstant component is attenuated by at most δ.

Section 2.5 also identified the constant component of the phase as

$$\widehat{\phi_a^{(2m)}}(0) = \mathbb{E}_{U_n}[\phi_a^{(2m)}(U_n)] = \mathrm{bias}_{U_n}^{(2m)}(a).$$

Hence XOR averaging pushes the phase toward the constant function whose value is exactly the target uniform-bit bias.

The generated bias is

$$\mathrm{bias}_{Z\oplus B}^{(2m)}(a) = \mathbb{E}_Z[h(Z)].$$

The same imported smoothing identity supports the two estimates required in Step 2. For low-weight coefficient vectors, we sum the attenuated Fourier components in ℓ^1. For high-weight vectors, that sum is too expensive, so we instead pass through the second moment $\mathbb{E}|h|^2$. The next subsection proves the low-weight estimate; the following subsection begins the high-weight argument.

3.3 Step 2: the low-weight/high-weight split

For a coefficient vector $a = (a_1, \dots, a_n) \in \mathbb{Z}_M^n$, define

$$\operatorname{supp}(a) := \{i : a_i \neq 0\}, \qquad w := \operatorname{wt}(a) := |\operatorname{supp}(a)|.$$

As announced in the proof plan, Step 2 seeks two estimates for the same bias error. The first is the low-weight bound

$$L(w) := \delta \Lambda_M^w,$$

while the second, developed through the high-weight argument, benefits from exponential decay in w. This subsection explains the split and proves the complete low-weight estimate. The following subsection begins the proof of the complementary high-weight estimate.

Low-weight calculation 1: compute the exact one-bit Fourier coefficients. Section 2.5 established the general parity-Fourier expansion of the phase. We now compute its coefficients using the product structure of the modular phase. Fix a residue $c \in \mathbb{Z}_M$ and consider the one-bit function

$$f_c(x) := \omega_M^{cx}, \qquad x \in \mathbb{F}_2 = \{0, 1\}.$$

The two Fourier basis functions on one bit are the constant function 1 and the parity function $(-1)^x$. Hence there are unique $u_c, v_c \in \mathbb{C}$ such that

$$\omega_M^{cx} = u_c + v_c(-1)^x.$$

Evaluating this identity at $x = 0$ and $x = 1$ gives

$$1 = u_c + v_c, \qquad \omega_M^c = u_c - v_c.$$

Solving the two equations yields the exact coefficients

$$\boxed{u_c = \frac{1 + \omega_M^c}{2}, \qquad v_c = \frac{1 - \omega_M^c}{2}.} \tag{3.1}$$

Therefore

$$\boxed{\omega_M^{cx} = \frac{1 + \omega_M^c}{2} + \frac{1 - \omega_M^c}{2}(-1)^x.} \tag{3.2}$$

The local Fourier ℓ^1-mass is $|u_c| + |v_c|$. Define

$$\Lambda_M := \max_{c \in \mathbb{Z}_M} (|u_c| + |v_c|) = \max_{c \in \mathbb{Z}_M} \left(\left| \frac{1 + \omega_M^c}{2} \right| + \left| \frac{1 - \omega_M^c}{2} \right| \right). \tag{3.3}$$

Since

$$|u_c| = \left| \cos \frac{\pi c}{M} \right|, \qquad |v_c| = \left| \sin \frac{\pi c}{M} \right|,$$

we have

$$1 \leq \Lambda_M \leq \sqrt{2}. \tag{3.4}$$

If $c = 0$, then $u_0 = 1$ and $v_0 = 0$, so an inactive coordinate contributes exactly one unit of Fourier mass and creates no nonconstant component.

Low-weight calculation 2: tensor the one-bit expansions over all coordinates. For $x = (x_1, \ldots, x_n) \in \mathbb{F}_2^n$, the phase factors coordinatewise:

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

Substitute the one-bit expansion (3.2) in every coordinate:

$$\phi_a^{(M)}(x) = \prod_{i=1}^{n} \left(u_{a_i} + v_{a_i}(-1)^{x_i}\right). \tag{3.5}$$

A vector $T = (T_1, \ldots, T_n) \in \mathbb{F}_2^n$ records which term is selected from each factor: $T_i = 0$ means choose u_{a_i}, while $T_i = 1$ means choose $v_{a_i}(-1)^{x_i}$. For this choice, the parity part is

$$\prod_{i:T_i=1} (-1)^{x_i} = (-1)^{\sum_i T_i x_i} = (-1)^{T \cdot x} := \chi_T(x).$$

Expanding (3.5) therefore gives

$$\phi_a^{(M)}(x) = \sum_{T \in \mathbb{F}_2^n} \left(\prod_{i:T_i=0} u_{a_i} \prod_{i:T_i=1} v_{a_i} \right) \chi_T(x). \tag{3.6}$$

By uniqueness of the Fourier expansion on $\mathbb{F}_2^n$, the coefficient at the test vector T is exactly

$$\boxed{\widehat{\phi_a^{(M)}}(T) = \prod_{i:T_i=0} u_{a_i} \prod_{i:T_i=1} v_{a_i}.} \tag{3.7}$$

Low-weight calculation 3: factor the total Fourier ℓ^1-mass. Taking absolute values in (3.7) and summing over all T gives

$$\begin{aligned} \sum_{T \in \mathbb{F}_2^n} \left| \widehat{\phi_a^{(M)}}(T) \right| &= \sum_{T \in \mathbb{F}_2^n} \prod_{i:T_i=0} |u_{a_i}| \prod_{i:T_i=1} |v_{a_i}| \\ &= \prod_{i=1}^{n} \left(|u_{a_i}| + |v_{a_i}|\right). \end{aligned} \tag{3.8}$$

The second equality is exactly the distributive law: expanding the product on the right produces one term for each choice vector $T \in \{0,1\}^n$.

If $a_i = 0$, the ith factor in (3.8) is $|u_0| + |v_0| = 1$. If $a_i \neq 0$, it is at most Λ_M. There are precisely $\mathrm{wt}(a)$ active coordinates. Hence

$$\boxed{\sum_{T \in \mathbb{F}_2^n} \left| \widehat{\phi_a^{(M)}}(T) \right| \leq \Lambda_M^{\mathrm{wt}(a)}.} \tag{3.9}$$

Strictly speaking, (3.9) says that the upper bound on the Fourier ℓ^1-mass grows with weight; the exact mass need not increase monotonically for every particular vector a.

Low-weight calculation 4: apply the XOR-smoothing identity proved in Section 2.5. Let $B \in \mathbb{F}_2^n$ be an independent δ-biased mask, so

$$\widehat{B}(T) := \mathbb{E}_B[(-1)^{T\cdot B}], \qquad |\widehat{B}(T)| \le \delta \quad \text{for every } T \neq 0. \tag{3.10}$$

At $T = 0$, one has $\widehat{B}(0) = 1$.

For fixed $z \in \mathbb{F}_2^n$, let

$$h(z) := \mathbb{E}_B[\phi_a^{(M)}(z \oplus B)].$$

The exact Fourier-diagonal identity was proved in Section 2.5. Applying it to the explicit coefficients computed above gives

$$\boxed{h(z) = \sum_{T\in\mathbb{F}_2^n} \widehat{\phi_a^{(M)}}(T)\widehat{B}(T)\chi_T(z).} \tag{3.11}$$

Thus the coefficient at T is multiplied by the single scalar $\widehat{B}(T)$; no Fourier component is mixed with another.

The constant coefficient of the phase is

$$\widehat{\phi_a^{(M)}}(0) = \mathbb{E}_{U_n}[\phi_a^{(M)}(U_n)] = \mathrm{bias}_{U_n}^{(M)}(a). \tag{3.12}$$

Low-weight calculation 5: derive the direct mask estimate. Let Z be any bit distribution independent of B. By definition,

$$\mathrm{bias}_{Z\oplus B}^{(M)}(a) = \mathbb{E}_Z[h(Z)].$$

Average (3.11) over Z and separate the term $T = 0$. Using (3.12), we get the exact identity

$$\mathrm{bias}_{Z\oplus B}^{(M)}(a) - \mathrm{bias}_{U_n}^{(M)}(a) = \sum_{T\neq 0} \widehat{\phi_a^{(M)}}(T)\widehat{B}(T)\mathbb{E}_Z[\chi_T(Z)]. \tag{3.13}$$

Now apply the triangle inequality:

$$\left|\mathrm{bias}_{Z\oplus B}^{(M)}(a) - \mathrm{bias}_{U_n}^{(M)}(a)\right| \le \sum_{T\neq 0} \left|\widehat{\phi_a^{(M)}}(T)\right| |\widehat{B}(T)||\mathbb{E}_Z[\chi_T(Z)]|. \tag{3.14}$$

For $T \neq 0$, the small-bias condition gives $|\widehat{B}(T)| \le \delta$. Moreover, $\chi_T(Z) \in \{-1, 1\}$ pointwise, so $|\mathbb{E}_Z[\chi_T(Z)]| \le 1$. Therefore, by (3.9),

$$\boxed{\left|\mathrm{bias}_{Z\oplus B}^{(M)}(a) - \mathrm{bias}_{U_n}^{(M)}(a)\right| \le \delta \sum_{T\neq 0} \left|\widehat{\phi_a^{(M)}}(T)\right| \le \delta\Lambda_M^{\mathrm{wt}(a)} = L(w).} \tag{3.15}$$

No pseudorandomness property of Z was used: the proof only invoked the trivial inequality $|\mathbb{E}_Z\chi_T(Z)| \le 1$. For low-weight tests, the mask itself supplies the entire guarantee.

Why the direct estimate is a low-weight argument. Suppose the target bit-bias error is γ, as in the proof plan, and consider every vector with $\mathrm{wt}(a) \le W$. It is enough to choose

$$\delta \le \gamma \Lambda_M^{-W},$$

because then (3.15) gives

$$\delta \Lambda_M^{\mathrm{wt}(a)} \le \delta \Lambda_M^{W} \le \gamma.$$

If

$$W = O_M\left(\log \frac{1}{\gamma}\right),$$

then

$$\log \frac{1}{\delta} = O_M\left(\log \frac{1}{\gamma}\right),$$

so the mask retains seed length

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

For weight as large as n, however, the same estimate would require

$$\delta \le \gamma \Lambda_M^{-n},$$

and therefore

$$\log \frac{1}{\delta} \ge n \log \Lambda_M + \log \frac{1}{\gamma}.$$

This can cost $\Omega(n)$ seed bits. The problem is not that the actual error must be large; rather, the triangle inequality in (3.14) treats all parity components separately and pays their full Fourier ℓ^1-mass.

At high weight, the proof changes norms. Parseval gives the weight-independent identity

$$\sum_T \left|\widehat{\phi_a^{(M)}}(T)\right|^2 = \mathbb{E}_{U_n} |\phi_a^{(M)}(U_n)|^2 = 1,$$

because every phase has absolute value 1. The second-moment argument will exploit this stable Fourier ℓ^2-mass instead of the growing ℓ^1-mass.

Regime	**Useful fact**	**Proof mechanism**
Low weight	$\Lambda_M^{\mathrm{wt}(a)}$ remains manageable	Fourier ℓ^1-mass plus parity-small-bias attenuation
High weight	$\kappa_M^{\mathrm{wt}(a)}$ makes the target bias tiny, while Fourier ℓ^2-mass is exactly 1	Cauchy–Schwarz, derivative descent, and an input-Fourier second-moment estimate

The threshold in the proof plan is

$$W = O_M\left(\log \frac{1}{\gamma}\right).$$

3.4 Step 2: the high-weight estimate via derivative descent

Lemma 3.3 (Statistical distance supplies the lower-modulus bit-bias hypothesis)**.** *Let Z be a bit distribution on $\{0,1\}^n$, and suppose that, for every coefficient vector $c \in \mathbb{Z}_m^n$,*

$$\operatorname{dist}\big(S_c^{(m)}(Z), S_c^{(m)}(U_n)\big) \le \rho.$$

Then Z has modulo-m bit-bias error at most 2ρ. In particular, if $\rho \le \eta/2$, then Z has modulo-m bit-bias error at most η, as required in the high-weight argument below.

Proof. Fix $c \in \mathbb{Z}_m^n$, and let P and Q be the probability mass functions of $S_c^{(m)}(Z)$ and $S_c^{(m)}(U_n)$ on $\mathbb{Z}_m$. The corresponding bias difference is

$$\begin{aligned}\left|\operatorname{bias}_Z^{(m)}(c) - \operatorname{bias}_{U_n}^{(m)}(c)\right| &= \left|\mathbb{E}\Big[\omega_m^{S_c^{(m)}(Z)}\Big] - \mathbb{E}\Big[\omega_m^{S_c^{(m)}(U_n)}\Big]\right| \\ &= \left|\sum_{r\in\mathbb{Z}_m} \big(P(r) - Q(r)\big)\omega_m^r\right|.\end{aligned}$$

Because $|\omega_m^r| = 1$ for every $r \in \mathbb{Z}_m$, the triangle inequality gives

$$\begin{aligned}\left|\operatorname{bias}_Z^{(m)}(c) - \operatorname{bias}_{U_n}^{(m)}(c)\right| &\le \sum_{r\in\mathbb{Z}_m} |P(r) - Q(r)| \\ &= 2\operatorname{dist}\big(S_c^{(m)}(Z), S_c^{(m)}(U_n)\big) \\ &\le 2\rho.\end{aligned}$$

This is the bounded-test inequality specialized to the root-of-unity phase $r \mapsto \omega_m^r$. Since the bound holds for every $c \in \mathbb{Z}_m^n$, taking the supremum over c proves that the modulo-m bit-bias error of Z is at most 2ρ. □

Thus a modulo-m generator run with statistical-distance error at most $\eta/2$ supplies exactly the lower-modulus bit-bias hypothesis used in this subsection. We now formulate the high-weight estimate in that bit-bias language.

Section 3.3 proved the first branch of the minimum in the proof plan:

$$\left|\operatorname{bias}_{Z\oplus B}^{(2m)}(a) - \operatorname{bias}_{U_n}^{(2m)}(a)\right| \le \delta\Lambda_{2m}^{\operatorname{wt}(a)}.$$

That estimate is effective when the weight is small, but its Fourier ℓ^1 loss grows with $\operatorname{wt}(a)$. We now prove the complementary high-weight branch

$$\boxed{H(w) := \sqrt{\kappa_{2m}^{2w} + \delta^2 + \eta} + \kappa_{2m}^{w}, \qquad w = \operatorname{wt}(a),}$$

under the additional assumption that Z has modulo-m bit-bias error at most η.

The argument follows the second branch announced in Step 2 of the proof plan. We use the same smoothed phase

$$h(z) = \mathbb{E}_B\big[\phi_a^{(2m)}(z \oplus B)\big]$$

introduced in Section 3.2, but change from Fourier ℓ^1 control to a second-moment estimate. Cauchy–Schwarz creates two shifted copies of the phase. Their product is a multiplicative XOR derivative, and the Boolean XOR identity forces every variable-dependent exponent to acquire a factor of 2. Since

$$\omega_{2m}^2 = \omega_m,$$

this derivative is a modulo-m phase. The lower-modulus guarantee for Z can then be applied inside the second moment. Finally, the Fourier-diagonal smoothing identity and Parseval proved in Section 2.5 evaluate the corresponding uniform-bit second moment.

The algebraic core: one XOR derivative halves the modulus

Fix $m \geq 1$. Because

$$\omega_m = \omega_{2m}^2,$$

a factor of 2 in the exponent converts a modulo-$2m$ phase into a modulo-m phase. The next identity is the mechanism that produces this factor.

Lemma 3.4 (Derivative descent identity)**.** *Fix*

$$a \in \mathbb{Z}_{2m}^n, \qquad b, b' \in \{0,1\}^n.$$

Define

$$\Delta_i = b_i - b_i' \in \{-1, 0, 1\},$$
$$\lambda(a; b, b') := \omega_{2m}^{\sum_i a_i \Delta_i},$$

and

$$c_i(a; b, b') := -a_i \Delta_i \pmod{m}.$$

Then, for every $z \in \{0,1\}^n$,

$$\boxed{\phi_a^{(2m)}(z \oplus b)\overline{\phi_a^{(2m)}(z \oplus b')} = \lambda(a; b, b')\phi_c^{(m)}(z).} \tag{3.16}$$

In particular, the multiplicative XOR derivative of a modulo-$2m$ phase is a unit-modulus constant times a modulo-m phase.

Proof. By definition,

$$\phi_a^{(2m)}(x) = \omega_{2m}^{\sum_i a_i x_i}, \qquad \omega_{2m} = e^{2\pi i/(2m)}.$$

Complex conjugation reverses the sign of the exponent, and therefore

$$\phi_a^{(2m)}(z \oplus b)\overline{\phi_a^{(2m)}(z \oplus b')} = \omega_{2m}^{\sum_i a_i \left((z_i \oplus b_i) - (z_i \oplus b_i')\right)}.$$

For Boolean values, XOR can be written over the integers as

$$z_i \oplus b_i = z_i + b_i - 2z_i b_i.$$

Consequently, recalling that $\Delta_i = b_i - b'_i$, we have

$$\begin{aligned}(z_i \oplus b_i) - (z_i \oplus b'_i) &= (b_i - b'_i) - 2z_i(b_i - b'_i) \\ &= \Delta_i(1 - 2z_i).\end{aligned}$$

Substituting this identity into the exponent gives

$$\sum_i a_i\Delta_i(1 - 2z_i) = \sum_i a_i\Delta_i - 2\sum_i a_i\Delta_i z_i.$$

The first term is independent of z, so it contributes only the constant phase

$$\lambda(a; b, b') = \omega_{2m}^{\sum_i a_i\Delta_i}.$$

All of the z-dependence is therefore contained in

$$\omega_{2m}^{-2\sum_i a_i\Delta_i z_i}.$$

This is the crucial point: every variable-dependent coefficient has acquired a factor of 2. For every integer r,

$$\omega_{2m}^{2r} = \left(e^{2\pi i/(2m)}\right)^{2r} = e^{2\pi ir/m} = \omega_m^r.$$

Taking

$$r = -\sum_i a_i\Delta_i z_i$$

therefore yields

$$\omega_{2m}^{-2\sum_i a_i\Delta_i z_i} = \omega_m^{-\sum_i a_i\Delta_i z_i}.$$

Now define

$$c_i \equiv -a_i\Delta_i \pmod m.$$

Since powers of ω_m depend only on their exponent modulo m,

$$\omega_m^{-\sum_i a_i\Delta_i z_i} = \omega_m^{\sum_i c_i z_i} = \phi_c^{(m)}(z).$$

Combining the constant and variable-dependent factors proves

$$\phi_a^{(2m)}(z \oplus b)\overline{\phi_a^{(2m)}(z \oplus b')} = \lambda(a; b, b')\phi_c^{(m)}(z).$$

The key distinction is that the entire expression need not itself be a pure modulo-m phase: the constant $\lambda(a; b, b')$ may still be a $2m$th root of unity. However, $\lambda(a; b, b')$ is independent of z, while all variable-dependent oscillation has descended from modulus $2m$ to modulus m. Equivalently, coordinates for which $b_i = b'_i$ disappear, whereas every coordinate for which $b_i \neq b'_i$ contributes a z_i-coefficient divisible by 2. This divisibility is the algebraic reason that one XOR derivative halves the modulus. □

Corollary 3.5 (Derivative expectations are lower-modulus biases)**.** *For every bit distribution X,*

$$\mathbb{E}_X\left[\phi_a^{(2m)}(X \oplus b)\overline{\phi_a^{(2m)}(X \oplus b')}\right] = \lambda(a; b, b')\,\mathrm{bias}_X^{(m)}(c).$$

Proof. Take the expectation of (3.16) over X. The factor $\lambda(a; b, b')$ is independent of X, and the expectation of $\phi_c^{(m)}(X)$ is $\mathrm{bias}_X^{(m)}(c)$ by definition. □

The high-weight smoothing inequality

We now insert the derivative identity into the second-moment argument and prove the second estimate in the minimum from the proof plan.

Theorem 3.6 (High-weight smoothing modulo $2m$ by a modulo-m source)**.** *Suppose:*

(i) Z *is a bit distribution with modulo-*m *bit-bias error at most* η*;*

(ii) B *is* δ*-biased in the parity sense;*

(iii) Z *and* B *are independent.*

Put

$$G := Z \oplus B.$$

Then, for every $a \in \mathbb{Z}_{2m}^n$*,*

$$\boxed{\left|\mathrm{bias}_G^{(2m)}(a)\right|^2 \le \left|\mathrm{bias}_{U_n}^{(2m)}(a)\right|^2 + \delta^2\left(1 - \left|\mathrm{bias}_{U_n}^{(2m)}(a)\right|^2\right) + \eta.} \tag{3.17}$$

Consequently,

$$\boxed{\begin{aligned}\left|\mathrm{bias}_G^{(2m)}(a) - \mathrm{bias}_{U_n}^{(2m)}(a)\right| \le \sqrt{\left|\mathrm{bias}_{U_n}^{(2m)}(a)\right|^2 + \delta^2\left(1 - \left|\mathrm{bias}_{U_n}^{(2m)}(a)\right|^2\right) + \eta} \\ + \left|\mathrm{bias}_{U_n}^{(2m)}(a)\right|.\end{aligned}} \tag{3.18}$$

Proof strategy. The proof implements the high-weight branch of Step 2 in four stages.

1. **Smooth.** Use the mask-averaged phase $h(z) = \mathbb{E}_B \phi_a^{(2m)}(z \oplus B)$ from Section 3.2, so $\mathrm{bias}_G^{(2m)}(a) = \mathbb{E}_Z h(Z)$.
2. **Square.** Cauchy–Schwarz bounds $|\mathbb{E}_Z h(Z)|^2$ by $\mathbb{E}_Z |h(Z)|^2$.
3. **Descend.** Expanding $|h|^2$ introduces two independent masks. For every fixed pair, the derivative identity (3.16) turns the integrand into a modulo-m phase. The bit-bias guarantee of Z then replaces its expectation by the corresponding uniform-bit expectation at cost η.
4. **Evaluate the uniform second moment.** Invoke the XOR-smoothing identity and Parseval proved in Section 2.5. The mask preserves the constant mode, attenuates every nonconstant mode by at most δ, and the total squared Fourier mass of the unit-modulus phase is 1.

Proof. Define, as in Section 3.2,

$$h(z) := \mathbb{E}_B[\phi_a^{(2m)}(z \oplus B)].$$

Then independence of Z and B gives

$$\mathrm{bias}_G^{(2m)}(a) = \mathbb{E}_Z h(Z).$$

Step 1: apply Cauchy–Schwarz. Since Z is distributed according to a probability measure,

$$|\operatorname{bias}_G^{(2m)}(a)|^2 = |\mathbb{E}_Z h(Z)|^2 \le \mathbb{E}_Z |h(Z)|^2. \tag{3.19}$$

This is the decisive change of norm. Instead of bounding h by the Fourier ℓ^1-mass used in Section 3.3, we pass to a quadratic quantity that can later be evaluated by Parseval.

Step 2: expand the second moment. Let B' be an independent copy of B. For each fixed z,

$$\begin{aligned} |h(z)|^2 &= \mathbb{E}_B[\phi_a^{(2m)}(z \oplus B)]\overline{\mathbb{E}_{B'}[\phi_a^{(2m)}(z \oplus B')]} \\ &= \mathbb{E}_{B,B'}\left[\phi_a^{(2m)}(z \oplus B)\overline{\phi_a^{(2m)}(z \oplus B')}\right], \end{aligned}$$

where independence permits the product of the two expectations to be written as one expectation over (B, B'). Averaging over Z gives

$$\mathbb{E}_Z |h(Z)|^2 = \mathbb{E}_{B,B'}\mathbb{E}_Z\left[\phi_a^{(2m)}(Z \oplus B)\overline{\phi_a^{(2m)}(Z \oplus B')}\right]. \tag{3.20}$$

Step 3: condition on the masks and invoke the lower-modulus guarantee. Fix realized mask values (b, b'), and let $c = c(a; b, b') \in \mathbb{Z}_m^n$ be the coefficient vector in (3.16). By the derivative-expectation corollary, the conditioned expectation under Z equals

$$\lambda(a; b, b') \operatorname{bias}_Z^{(m)}(c),$$

whereas the same conditioned expectation with Z replaced by U_n equals

$$\lambda(a; b, b') \operatorname{bias}_{U_n}^{(m)}(c).$$

Because $|\lambda(a; b, b')| = 1$ and Z has modulo-m bit-bias error at most η,

$$\left|\lambda(a; b, b') \operatorname{bias}_Z^{(m)}(c) - \lambda(a; b, b') \operatorname{bias}_{U_n}^{(m)}(c)\right| \le \eta.$$

The vector c depends on (b, b'), but the lower-modulus guarantee is uniform over every $c \in \mathbb{Z}_m^n$, so the preceding estimate holds for every fixed mask pair. Define

$$F_Z(b, b') := \mathbb{E}_Z\left[\phi_a^{(2m)}(Z \oplus b)\overline{\phi_a^{(2m)}(Z \oplus b')}\right]$$

and define $F_{U_n}(b, b')$ analogously with Z replaced by U_n. Then

$$|F_Z(b, b') - F_{U_n}(b, b')| \le \eta \qquad \text{for every } (b, b').$$

Writing

$$A := \mathbb{E}_Z |h(Z)|^2 = \mathbb{E}_{B,B'} F_Z(B, B'), \qquad A_{\text{unif}} := \mathbb{E}_{U_n} |h(U_n)|^2 = \mathbb{E}_{B,B'} F_{U_n}(B, B'),$$

linearity of expectation and the triangle inequality give

$$|A - A_{\text{unif}}| = |\mathbb{E}_{B,B'}[F_Z(B, B') - F_{U_n}(B, B')]|$$

$$\leq \mathbb{E}_{B,B'}[|F_Z(B,B') - F_{U_n}(B,B')|]$$
$$\leq \mathbb{E}_{B,B'}[\eta] = \eta.$$

Finally, A and A_{unif} are real and nonnegative, since they are expectations of squared absolute values. Hence $|A - A_{\text{unif}}| \leq \eta$ implies

$$\mathbb{E}_Z|h(Z)|^2 \leq \mathbb{E}_{U_n}|h(U_n)|^2 + \eta. \tag{3.21}$$

Step 4: evaluate the uniform second moment using the identities proved in Section 2.5. Section 2.5 proved that XOR averaging acts diagonally in the parity-Fourier basis. Applied to the present phase, that identity is

$$\widehat{h}(T) = \widehat{\phi_a^{(2m)}}(T)\widehat{B}(T).$$

Parseval on $\mathbb{F}_2^n$, also established there in the common finite-group framework, gives

$$\mathbb{E}_{U_n}|h(U_n)|^2 = \sum_{T\in\mathbb{F}_2^n} \left|\widehat{\phi_a^{(2m)}}(T)\right|^2 |\widehat{B}(T)|^2.$$

At the constant frequency $T = 0$, Section 2.5 identified

$$\widehat{\phi_a^{(2m)}}(0) = \text{bias}_{U_n}^{(2m)}(a), \qquad \widehat{B}(0) = 1.$$

For every $T \neq 0$, the δ-bias condition gives $|\widehat{B}(T)| \leq \delta$. Moreover, since the phase has unit magnitude pointwise,

$$\sum_{T\in\mathbb{F}_2^n} \left|\widehat{\phi_a^{(2m)}}(T)\right|^2 = \mathbb{E}_{U_n}|\phi_a^{(2m)}(U_n)|^2 = 1.$$

Separating the constant mode from all nonconstant modes therefore gives

$$\mathbb{E}_{U_n}|h(U_n)|^2 \leq \left|\text{bias}_{U_n}^{(2m)}(a)\right|^2 + \delta^2 \sum_{T\neq 0} \left|\widehat{\phi_a^{(2m)}}(T)\right|^2$$
$$= \left|\text{bias}_{U_n}^{(2m)}(a)\right|^2 + \delta^2\left(1 - \left|\text{bias}_{U_n}^{(2m)}(a)\right|^2\right).$$

Thus

$$\mathbb{E}_{U_n}|h(U_n)|^2 \leq \left|\text{bias}_{U_n}^{(2m)}(a)\right|^2 + \delta^2\left(1 - \left|\text{bias}_{U_n}^{(2m)}(a)\right|^2\right). \tag{3.22}$$

Combining (3.19), (3.21), and (3.22) proves (3.17).

Finally, the triangle inequality gives

$$\left|\text{bias}_G^{(2m)}(a) - \text{bias}_{U_n}^{(2m)}(a)\right| \leq |\,\text{bias}_G^{(2m)}(a)| + |\,\text{bias}_{U_n}^{(2m)}(a)|.$$

Taking the square root of (3.17) and substituting it into this inequality proves (3.18). □

Lemma 3.7 (The uniform target bias decays with weight)**.** *For every $M \geq 2$ and every $a \in \mathbb{Z}_M^n$, define*

$$\kappa_M := \cos\frac{\pi}{M} < 1.$$

Then

$$\boxed{|\operatorname{bias}_{U_n}^{(M)}(a)| \leq \kappa_M^{\operatorname{wt}(a)}.} \tag{3.23}$$

Proof. Under independent uniform bits,

$$\operatorname{bias}_{U_n}^{(M)}(a) = \mathbb{E}_{U_n}\left[\omega_M^{\sum_i a_i U_i}\right] = \prod_{i=1}^{n} \mathbb{E}_{U_i}[\omega_M^{a_i U_i}]$$
$$= \prod_{i=1}^{n} \frac{1+\omega_M^{a_i}}{2}.$$

If $a_i = 0$, the corresponding factor is 1. If $a_i \neq 0$, then

$$\left|\frac{1+\omega_M^{a_i}}{2}\right| = \left|\cos\frac{\pi a_i}{M}\right| \leq \cos\frac{\pi}{M} = \kappa_M.$$

Multiplying the bounds for the $\operatorname{wt}(a)$ nonzero coordinates proves the claim. Thus a high-weight test already has very small expectation under the target distribution. This is the decreasing quantity exploited by the complementary high-weight estimate. □

Corollary 3.8 (Completion of Step 2: the two-branch minimum)**.** *Under the assumptions of the high-weight smoothing theorem, for every $a \in \mathbb{Z}_{2m}^n$,*

$$\boxed{\begin{aligned}&\left|\operatorname{bias}_G^{(2m)}(a) - \operatorname{bias}_{U_n}^{(2m)}(a)\right| \\ &\qquad \leq \min\left\{\delta\Lambda_{2m}^{\operatorname{wt}(a)}, \sqrt{\kappa_{2m}^{2\operatorname{wt}(a)} + \delta^2 + \eta} + \kappa_{2m}^{\operatorname{wt}(a)}\right\}.\end{aligned}} \tag{3.24}$$

Proof. The first branch is exactly the direct low-weight mask estimate (3.15), proved in Section 3.3. For the second branch, apply the uniform-bit decay estimate (3.23) to (3.18):

$$\left|\operatorname{bias}_{U_n}^{(2m)}(a)\right| \leq \kappa_{2m}^{\operatorname{wt}(a)}.$$

Also use

$$1 - \left|\operatorname{bias}_{U_n}^{(2m)}(a)\right|^2 \leq 1.$$

This gives

$$\left|\operatorname{bias}_G^{(2m)}(a) - \operatorname{bias}_{U_n}^{(2m)}(a)\right| \leq \sqrt{\kappa_{2m}^{2\operatorname{wt}(a)} + \delta^2 + \eta} + \kappa_{2m}^{\operatorname{wt}(a)}.$$

Since both this estimate and (3.15) hold for the same coefficient vector a, their minimum also bounds the bias error. This proves precisely the two-part inequality stated in Step 2 of the proof plan. □

4 Proof Completion and Limitations

4.1 The low-weight/high-weight split

The two estimates in (3.24) move in opposite directions as the weight grows:

$$\delta\Lambda_M^w \text{ increases with } w, \qquad \kappa_M^w \text{ decreases with } w.$$

This is why a weight threshold is natural.

Assume $M = 2m > 2$, and fix a target bit-bias error $0 < \gamma \leq 1$. Define

$$W_M(\gamma) := \left\lceil \frac{\log(8/\gamma)}{\log(1/\kappa_M)} \right\rceil, \tag{4.1}$$

$$\delta := \frac{\gamma}{8\Lambda_M^{W_M(\gamma)}}, \tag{4.2}$$

and

$$\eta := \frac{\gamma^2}{64}. \tag{4.3}$$

For fixed M,

$$W_M(\gamma) = O_M\left(\log \frac{1}{\gamma}\right), \qquad \log \frac{1}{\delta} = O_M\left(\log \frac{1}{\gamma}\right).$$

Theorem 4.1 (One-step bit-bias theorem)**.** *Let Z have modulo-m bit-bias error at most η, and let B be δ-biased in the parity sense, with δ, η as in (4.2)–(4.3). Then*

$$G = Z \oplus B$$

has modulo-M bit-bias error at most γ.

Proof. Fix $a \in \mathbb{Z}_M^n$ and put $w = \mathrm{wt}(a)$.

Low-weight case: $w \leq W_M(\gamma)$. The low-weight mask estimate gives

$$\left|\mathrm{bias}_G^{(M)}(a) - \mathrm{bias}_{U_n}^{(M)}(a)\right| \leq \delta\Lambda_M^w \leq \delta\Lambda_M^{W_M(\gamma)} = \frac{\gamma}{8} < \gamma.$$

High-weight case: $w > W_M(\gamma)$. By the definition of $W_M(\gamma)$,

$$\kappa_M^w \leq \frac{\gamma}{8}.$$

Also,

$$\delta \leq \frac{\gamma}{8}, \qquad \sqrt{\eta} = \frac{\gamma}{8}.$$

The high-weight estimate in (3.24) yields

$$\begin{aligned}\left|\mathrm{bias}_G^{(M)}(a) - \mathrm{bias}_{U_n}^{(M)}(a)\right| &\leq \sqrt{3\left(\frac{\gamma}{8}\right)^2} + \frac{\gamma}{8} \\ &= \frac{\sqrt{3}+1}{8}\gamma < \gamma.\end{aligned}$$

Both weight regimes satisfy the required bit-bias bound. □

Remark 4.2 (The base case $M = 2$)**.** *Modulo* 2*, the nonconstant biases are parity expectations. A standard* γ*-biased distribution therefore has modulo-*2 *bit-bias error at most* γ *and seed length*

$$O\left(\log n + \log \frac{1}{\gamma}\right).$$

Equivalently, one may regard modulus 1 *as the trivial lower-modulus base.*

4.2 The one-step PRG reduction in statistical distance

Let $r_q(n, \varepsilon)$ denote the seed length of an explicit generator that ε-fools every Boolean linear sum modulo q in statistical distance.

Theorem 4.3 (Modulus-doubling PRG reduction)**.** *Let* $M = 2m \geq 4$*. For every* n *and* $0 < \varepsilon \leq 1/2$*, there is an explicit modulo-*M *generator with seed length*

$$\boxed{r_M(n, \varepsilon) \leq r_m\left(n, \frac{\varepsilon^2}{32(M-1)}\right) + O_M\left(\log n + \log \frac{1}{\varepsilon}\right).} \tag{4.4}$$

For $\varepsilon > 1/2$*, the generator for error* $1/2$ *also suffices.*

Proof. Set

$$\gamma := \frac{2\varepsilon}{\sqrt{M-1}}. \tag{4.5}$$

Since $M \geq 4$ and $\varepsilon \leq 1/2$, one has $0 < \gamma \leq 1$. Choose W, δ, η according to (4.1)–(4.3), so

$$\eta = \frac{\gamma^2}{64}.$$

Let B be an explicit δ-biased distribution. Let Z be the output of a modulo-m generator run at statistical-distance error

$$\rho := \frac{\eta}{2}.$$

For every $c \in \mathbb{Z}_m^n$, the bounded-test inequality applied to the phase $r \mapsto \omega_m^r$ gives

$$\left|\mathrm{bias}_Z^{(m)}(c) - \mathrm{bias}_{U_n}^{(m)}(c)\right| \leq 2\rho = \eta.$$

Thus Z has the lower-modulus bit-bias error required by the one-step bit-bias theorem. Use independent seeds and output

$$G_M := Z \oplus B.$$

The one-step theorem gives modulo-M bit-bias error at most γ. The Fourier conversion lemma then gives, for every $a \in \mathbb{Z}_M^n$,

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

Finally,

$$\rho = \frac{1}{2} \cdot \frac{\gamma^2}{64} = \frac{\varepsilon^2}{32(M-1)}.$$

The mask seed has length

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

which proves (4.4). □

4.3 Iterating through the complete power of two

Write

$$M = 2^t q,$$

where q is odd and $t \geq 0$. Put

$$M_j = 2^j q, \qquad 0 \leq j \leq t.$$

Starting from the desired final error $\varepsilon_t = \varepsilon \leq 1/2$, define recursively

$$\varepsilon_{j-1} := \frac{\varepsilon_j^2}{32(M_j - 1)} \qquad (j = t, t-1, \ldots, 1). \tag{4.6}$$

Lemma 4.4 (Reverse error growth)**.** *For fixed M,*

$$\log \frac{1}{\varepsilon_0} \leq 2^t \log \frac{1}{\varepsilon} + C_M \tag{4.7}$$

for a constant C_M depending only on M.

Proof. Let

$$L_j = \log \frac{1}{\varepsilon_j}.$$

The recurrence gives

$$L_{j-1} = 2L_j + \log(32(M_j - 1)).$$

Iterating,

$$L_0 = 2^t L_t + \sum_{j=1}^{t} 2^{j-1} \log(32(M_j - 1)).$$

The sum depends only on the fixed modulus M. □

Theorem 4.5 (Lifting an optimal PRG from the odd part)**.** *Assume that for the odd modulus $q \geq 3$ there is an explicit generator satisfying*

$$r_q(n, \rho) \leq C_q\left(\log n + \log \frac{1}{\rho}\right)$$

for all n, ρ, where C_q depends only on q. Then, for $M = 2^t q$,

$$\boxed{r_M(n, \varepsilon) = O_M\left(\log n + \log \frac{1}{\varepsilon}\right).} \tag{4.8}$$

Proof. Apply the one-step reduction successively along

$$q \longrightarrow 2q \longrightarrow 4q \longrightarrow \cdots \longrightarrow 2^t q.$$

The base generator is run at error ε_0. The reverse-growth lemma gives

$$\log \frac{1}{\varepsilon_0} = O_M\left(\log \frac{1}{\varepsilon}\right).$$

At level j, the additional mask seed has length

$$O_{M_j}\left(\log n + \log \frac{1}{\varepsilon_j}\right).$$

There are only $t \le \log_2 M$ levels. Summing the base seed and all mask seeds proves (4.8). □

The remaining case $q = 1$. When $M = 2^t$ has no odd part, the chain above would have to begin with the step $1 \longrightarrow 2$, which the one-step reduction does not cover, because that reduction requires $M = 2m \ge 4$. In this case the base case $M = 2$ supplies the missing bottom level directly: by the remark on the base case $M = 2$, a γ-biased distribution has modulo-2 bit-bias error at most γ and seed length $O(\log n + \log \frac{1}{\gamma})$, so an optimal-order generator modulo 2 is available outright. Starting the chain at $M_1 = 2$ and applying the one-step reduction along

$$2 \longrightarrow 4 \longrightarrow \cdots \longrightarrow 2^t$$

involves only moduli $M_j = 2^j$ with $j \ge 2$, so the hypothesis $M_j \ge 4$ holds at every step. The error recursion (4.6) is then applied for $j = t, t-1, \ldots, 2$, the modulo-2 generator is run at error ε_1, and the same telescoping as in the reverse-growth lemma gives

$$\log \frac{1}{\varepsilon_1} = O_M\left(\log \frac{1}{\varepsilon}\right).$$

Hence (4.8) holds for $M = 2^t$ as well. For $t = 0$ the modulus is $M = 1$, every modular sum is identically 0, and the statement is trivial.

Corollary 4.6 (Consequence using the LRTV prime-power generator)**.** *Let*

$$M = 2^t p^k,$$

where p is an odd prime and $t, k \ge 0$. Combining the proposed modulus-doubling theorem with the LRTV optimal-order generator for the fixed prime power p^k gives an explicit generator with seed length

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

The same conclusion holds for pure powers of 2 by starting from parity.

Remark 4.7 (Dependence on M)**.** *The proof gives optimal dependence on n and ε for each fixed supported modulus. It does not establish a universal hidden constant in a bound of the form*

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

because the imported prime-power generator has its own modulus-dependent constant.

4.4 Why the method stops at the odd part

Let

$$\phi_a^{(M)}(x) = \omega_M^{\sum_{i=1}^n a_i x_i}, \qquad a \in \mathbb{Z}_M^n.$$

For $b, b' \in \mathbb{F}_2^n$, define the *multiplicative XOR derivative*

$$\Delta^{\times}_{b,b'}\phi_a^{(M)}(z) := \phi_a^{(M)}(z \oplus b)\overline{\phi_a^{(M)}(z \oplus b')}.$$

Choose integer representatives $\widetilde{a}_i \in \mathbb{Z}$ of the coefficients a_i. We call

$$\boxed{\mathcal{E}^{(M)}_{a;b,b'}(z) := \sum_{i=1}^n \widetilde{a}_i\big((z_i \oplus b_i) - (z_i \oplus b'_i)\big)}$$

the *derivative exponent*, since

$$\Delta^{\times}_{b,b'}\phi_a^{(M)}(z) = \omega_M^{\mathcal{E}^{(M)}_{a;b,b'}(z)}.$$

The resulting phase is independent of the chosen integer representatives.

The proof of the derivative descent identity (3.16) already computed this exponent. Writing

$$\Delta_i = b_i - b'_i \in \{-1, 0, 1\},$$

that calculation gives

$$\boxed{\mathcal{E}^{(M)}_{a;b,b'}(z) = \sum_i \widetilde{a}_i\Delta_i - 2\sum_i \widetilde{a}_i\Delta_i z_i.}$$

Thus the first term is independent of z, while the variable-dependent part is

$$\boxed{\mathcal{E}^{\mathrm{var}}_{a;b,b'}(z) := -2\sum_i \widetilde{a}_i\Delta_i z_i.}$$

In particular, every variable-dependent coefficient acquires a factor of 2. This is exactly the mechanism used in Section 3 to descend from modulus $2m$ to modulus m.

Lemma 4.8 (Exact modulus reached by one XOR derivative)**.** *Let*

$$d = \gcd(M, 2), \qquad M' = \frac{M}{d}.$$

When $M' = 1$, interpret the unique modulo-1 phase as the constant function 1. For $a \in \mathbb{Z}_M^n$ and $b, b' \in \mathbb{F}_2^n$, define $c \in \mathbb{Z}_{M'}^n$ by

$$c_i \equiv -\frac{2}{d}a_i(b_i - b'_i) \pmod{M'}.$$

Then there exists a unit-modulus constant λ, independent of z, such that

$$\boxed{\Delta^{\times}_{b,b'}\phi_a^{(M)}(z) = \lambda\,\phi_c^{(M')}(z).}$$

Consequently,

$$M' = \begin{cases} M/2, & M \text{ even}, \\ M, & M \text{ odd}. \end{cases}$$

Moreover, M' is the smallest modulus that works uniformly for all multiplicative XOR derivatives of modulo-M phases.

Proof. By the derivative-exponent calculation above,

$$\Delta^{\times}_{b,b'}\phi^{(M)}_a(z) = \lambda\, \omega_M^{-2\sum_i \widetilde{a}_i \Delta_i z_i}, \qquad \lambda = \omega_M^{\sum_i \widetilde{a}_i \Delta_i}.$$

Since $M = dM'$, we have $\omega_M^d = \omega_{M'}$. Therefore

$$\begin{aligned} \omega_M^{-2\sum_i \widetilde{a}_i \Delta_i z_i} &= \left(\omega_M^d\right)^{-(2/d)\sum_i \widetilde{a}_i \Delta_i z_i} \\ &= \omega_{M'}^{-(2/d)\sum_i \widetilde{a}_i \Delta_i z_i} \\ &= \phi_c^{(M')}(z). \end{aligned}$$

The definition of c_i is independent of the representative $\widetilde{a}_i$: replacing $\widetilde{a}_i$ by $\widetilde{a}_i + kM$ changes the coefficient by

$$-\frac{2}{d} kM\Delta_i = -2kM'\Delta_i,$$

which vanishes modulo M'. This proves the factorization.

If M is even, then $d = 2$ and $M' = M/2$, recovering precisely the modulus-halving identity (3.16). If M is odd, then $d = 1$ and $M' = M$. Whenever $b_i \neq b_i'$, multiplication by

$$-2(b_i - b_i')$$

is a permutation of $\mathbb{Z}_M$. Hence the derivative coefficient on such a coordinate can still be any element of $\mathbb{Z}_M$, so no smaller modulus is obtained uniformly.

Finally, take

$$n = 1, \qquad a_1 = 1, \qquad b_1 = 1, \qquad b_1' = 0.$$

The variable-dependent derivative is

$$z \longmapsto \omega_M^{-2z}.$$

The root ω_M^{-2} has multiplicative order

$$\frac{M}{\gcd(M, 2)} = M'.$$

Thus some derivative has exact order M', proving that no smaller modulus works uniformly. □

If

$$M = 2^t q$$

with q odd, repeated application of the lemma gives

$$2^t q \longrightarrow 2^{t-1} q \longrightarrow \cdots \longrightarrow 2q \longrightarrow q.$$

Once the modulus reaches q, it is odd, so multiplication by 2 is invertible and no further descent occurs. The XOR-derivative method therefore removes exactly the full power of 2 dividing M, and then stops at the odd part.

5 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.

[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*. The proof uses their optimal-order construction for fixed prime-power moduli as the base generator.

[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. The proof imports explicit parity-small-bias spaces with seed length $O(\log n + \log(1/\delta))$.

[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*.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il