A formalized reduction of the Mullin's Conjecture

Marcello Paris

February 2026

Abstract

The Euclid–Mullin sequence is defined by $a(0) = 2$, $a(n+1) = \mathrm{lpf}(a(0) \cdots a(n) + 1)$, where lpf is the least prime factor. Mullin's Conjecture (MC, 1963) asserts that every prime eventually appears. We present a Lean 4 formalization (∼26,900 lines, 35 files, **zero sorry**) that reduces MC to a single open hypothesis.

An *inductive bootstrap* yields the primary reduction: the **Single Hit Theorem** shows that MC follows if, for each prime q, the multiplicative walk on $(\mathbb{Z}/q\mathbb{Z})^\times$ hits -1 at least once past a computable bound. The algebraic precondition (SubgroupEscape) is free: for any prime $p \geq 5$, some odd prime $r < p$ escapes every proper subgroup of $(\mathbb{Z}/p\mathbb{Z})^\times$, proved using only modular arithmetic. A separate *Fourier bridge* gives a parallel reduction: MC follows whenever certain walk character sums are $o(N)$.

Multiple reduction routes—algebraic, character-analytic, sieve-theoretic—all converge on the same *orbit-specificity gap*: transferring generic equidistribution to one deterministic orbit. The sharpest sufficient condition is Conditional Multiplier Equidistribution (CME)—the statement that the factoring operation in the EM construction destroys correlation between consecutive multiplier residues—proved to imply the Complex Character Sum Bound (CCSB) for all character orders, bypassing the $d \geq 3$ barrier. CME decomposes as CME = VCB + Dec, where Vanishing Conditional Bias (VCB) is a strictly weaker hypothesis permitting fiber sums to be proportional to visit counts rather than small. We prove VCB + PED ⇒ CCSB, giving nine independent routes to MC. Over a hundred dead ends are documented, precisely delineating the boundary of current methods.

Contents

1 Introduction

Euclid's proposition IX.20 of the *Elements* shows that for any finite set of primes, each prime factor of their product plus one is outside the set: to grow your set of primes, you can pick any of them. The **Euclid–Mullin sequence** (OEIS A000945), introduced by Mullin [1], makes a definite choice: always take the *smallest* prime factor.

$$a(0) = 2, \qquad a(n+1) = \text{smallest prime factor of } \big(a(0)\cdots a(n)+1\big). \tag{1}$$

The first twenty terms (0-indexed) are

$$\underbrace{2}_{a(0)},\ 3,\ 7,\ 43,\ 13,\ 53,\ \underbrace{5}_{a(6)},\ \underbrace{6221671}_{a(7)},\ 38709183810571,$$
$$139,\ 2801,\ \underbrace{11}_{a(11)},\ 17,\ 5471,\ 52662739,\ 23003,\ 30693651606209,$$
$$\underbrace{37}_{a(17)},\ 1741,\ \underbrace{1313797957}_{a(19)},\ \ldots$$

The sequence shows an erratic behavior: small primes appear out of their natural order (5 not until position 6, 11 at position 11, 37 at position 17), while enormous primes—a 7-digit number at position 7, a 14-digit number at position 8—appear early. As of 2025, only 51 terms are known and some primes like 41 and 47 are not yet observed. Computing further terms requires *complete* factorization of $\mathrm{P}(n)+1$ (where $\mathrm{P}(n) \coloneqq a(0)\cdots a(n)$ is the running product): finding any prime factor is not enough—one must certify that no smaller factor exists. Since $\mathrm{P}(n)$ grows super-exponentially, this quickly exceeds the reach of all known factoring algorithms. By construction, no prime can appear twice.

Conjecture 1.1 (Mullin, 1963)**.** *Every prime number eventually appears in the Euclid–Mullin sequence.*

The conjecture has resisted proof for over sixty years. The difficulty is showing that the deterministic minFac rule eventually *selects* each prime. That the rule matters is not idle speculation: Cox and van der Poorten [11] showed that replacing minFac with the *largest* prime factor provably misses infinitely many primes.

Each step couples the next prime to the full factorization history, creating a recursive dependency that defeats both probabilistic heuristics and standard sieve methods.

At each step, $\mathrm{P}(n)+1$ may have many prime factors, and just the *smallest* is chosen. So, a target prime q may possibly divide $\mathrm{P}(n)+1$ for many (maybe infinite) n, yet never be selected if a smaller prime always divides $\mathrm{P}(n)+1$ as well. Our formalization tries to make this tension precise.

The accumulator structure. The running product $\mathrm{P}(n)$ is an *accumulator*: a single number that commits to the entire sequence history via irreversible multiplication, analogous to the cumulative digest in a hash chain. The accumulator poses a challenge $\mathrm{P}(n)+1$; the response $\mathrm{minFac}(\mathrm{P}(n)+1)$ extends the chain; and the updated accumulator $\mathrm{P}(n{+}1) = \mathrm{P}(n)\cdot \mathrm{minFac}(\mathrm{P}(n){+}1)$ absorbs the response irreversibly—once a prime enters the product, it divides every future product and can never appear again (Theorem 5.3).

This accumulator coupling is what makes standard tools fail. Sieve methods require approximate independence between the events "$p \mid m$" for different m; here successive challenges $\mathrm{P}(n)+1$ share a cumulative history. Ergodic methods require a fixed or state-dependent map; here the map at step n depends on the full accumulator, not just the current residue. The walk reformulation (Section 2) tames this coupling by projecting the accumulator onto a finite

group: $\mathrm{P}_q(n) = \mathrm{P}(n) \bmod q$ preserves the divisibility information relevant to q while discarding the accumulator's combinatorial complexity. But the information lost in this projection is exactly the source of difficulty: the walk position determines *whether* q can divide $\mathrm{P}(n)+1$; the full accumulator determines *which* prime is actually selected. Every open hypothesis in this paper—DH, CCSB, CME—addresses this gap.

The factoring channel. The accumulator structure of the EM sequence is closely analogous to a cryptographic hash chain: each term is deterministically derived from its predecessor through an operation—integer factorization—that destroys algebraic relationships. In an iterated SHA-256 chain $x_0, H(x_0), H(H(x_0)), \ldots$, each value is uniquely determined by the seed, yet consecutive terms pass every reasonable statistical test for independence. In the EM sequence, the running product $\mathrm{P}(n)$ determines $\mathrm{P}(n)+1$, whose smallest prime factor becomes the next multiplier; but the $O(\log q)$ bits visible in the residue $\mathrm{P}(n) \bmod q$ cannot control the outcome of factoring the $\sim 2^n$-bit integer $\mathrm{P}(n)+1$. This *information bottleneck* is why the multiplier residues behave as if independent—and why the walk on $(\mathbb{Z}/q\mathbb{Z})^\times$ should visit every element, including the "death state" -1 that would put q in the sequence. The analogy is tighter than it may appear: in both cases the pseudorandomness claim concerns a fully deterministic process whose forward map is easy but whose outputs resist structural prediction—the difference being that minFac achieves this not by cryptographic design but accidentally, through the information bottleneck of projecting a $\sim 2^n$-bit integer onto $O(\log q)$ bits. Our formalization makes this intuition precise: we identify the exact mathematical content—Conditional Multiplier Equidistribution—needed to convert "the factoring channel destroys correlation" into a proof of MC, and show it implies the conjecture through a verified chain of reductions.

Our main result is a formally verified reduction of MC to a single dynamical question: for each prime q, does the walk on $(\mathbb{Z}/q\mathbb{Z})^\times$ hit -1 at least once past a computable bound? The strategy proceeds in three stages:

1. **Reformulation.** We recast "does prime q appear?" as "does a multiplicative walk on the cyclic group $(\mathbb{Z}/q\mathbb{Z})^\times$ hit the element -1?" This translation is exact (Section 2).
2. **Bootstrap.** We show that the algebraic precondition for the walk to reach -1—that the multipliers generate the full group—is *free*, following from the inductive hypothesis $\mathrm{MC}(<p)$ via an elementary lemma (Section 3). A single hit on -1 past the sieve gap suffices for MC.
3. **Diagnosis.** We develop the harmonic-analytic and sieve-theoretic infrastructure to determine *precisely* what kind of statement would produce the required hit, and why known methods fall short (Sections 4–5).

The formalization serves two purposes: (i) it guarantees that every reduction is logically sound, and (ii) it precisely delineates the boundary between what is proved and what remains open, preventing the kind of subtle gap that plagues pencil-and-paper reductions involving multiple interacting hypotheses.

The primary reduction is:

Theorem 1.2 (Single Hit Theorem — ✓ single_hit_implies_mc)**.** SingleHitHypothesis $\Longrightarrow$ MC.

SingleHitHypothesis asks: for every missing prime q, if $\mathrm{MC}(<q)$ and $\mathrm{SE}(q)$ hold, then $q \mid \mathrm{P}(n)+1$ for some n past the sieve gap. The proof is by strong induction on p: the inductive hypothesis gives $\mathrm{MC}(<p)$, PrimeResidueEscape (proved elementarily) bootstraps $\mathrm{SE}(p)$, and the single hit past the sieve gap gives $\mathrm{em}(n+1) = p$.

The algebraic precondition—that the multipliers generate $(\mathbb{Z}/q\mathbb{Z})^\times$—is free; the open problem is purely dynamical. Multiple strategies for producing the required hit are formally verified:

Theorem 1.3 (✓ dynamical_hitting_implies_mullin)**.** DynamicalHitting $\implies$ MC.

Theorem 1.4 (✓ complex_csb_mc')**.** ComplexCharSumBound $\implies$ MC.

Theorem 1.5 (✓ cme_implies_mc)**.** CME $\implies$ MC (sharpest sufficient condition).

All reductions are fully machine-verified with zero sorry. Each strategy produces at least one hit past the sieve gap, which the Single Hit Theorem converts into MC. The formalization thus provides a precise "roadmap": prove DynamicalHitting, ComplexCharSumBound, CME, VCB + PED, or any of the equivalent formulations in §4 and §7, and the rest follows by machine-checked deduction.

Notation. Theorems marked ✓ name are formally verified in Lean 4; clicking the identifier links to the source code.

Organization. The paper follows the logical structure of the reduction. Section 2 reformulates MC as a walk-hitting problem, establishes the algebraic prerequisites, and derives the first missing prime's death channel avoidance. Section 3 presents the inductive bootstrap—the core insight that SubgroupEscape is free—and proves the Single Hit Theorem, reducing MC to producing one hit at each prime. Section 4 develops the character-analytic reduction (CCSB ⇒ MC), including the large sieve infrastructure, the spectral energy bridge, and the van der Corput–autocorrelation route. Section 5 explains *why* the remaining hypothesis is difficult by analyzing dead ends and structural barriers. Section 6 describes the Lean formalization. Section 7 discusses open problems and paths forward. Appendix A collects the historical background; Appendix C discusses analogies with Artin's conjecture, multiplicative walks, Sarnak's program, and sieve theory; Appendix D presents additional sieve and spectral routes to MC; Appendix E describes the human–AI methodology; and Appendix F provides a glossary of all definitions and hypotheses.

2 The Residue Walk

We define two sequences by mutual recursion:

$$\mathrm{em}(0) := 2, \qquad \mathrm{P}(0) := 2, \tag{2}$$

$$\mathrm{em}(n+1) := \mathrm{minFac}\big(\mathrm{P}(n)+1\big), \qquad \mathrm{P}(n+1) := \mathrm{P}(n)\cdot \mathrm{em}(n+1). \tag{3}$$

Theorem 2.1 (✓ seq_isPrime, seq_injective)**.** *Every* $\mathrm{em}(n)$ *is prime, and the sequence is injective: no prime appears twice.*

Definition 2.2 (Walk and Multiplier)**.** For any prime q, define the **residue walk** and the **multiplier** by projecting the accumulator and the next prime onto $\mathbb{Z}/q\mathbb{Z}$:

$$\mathrm{P}_q(n) := \mathrm{P}(n) \bmod q$$
$$\mathrm{m}_q(n) := \mathrm{em}(n{+}1) \bmod q$$

Proposition 2.3 (Walk recurrence — ✓ walkZ_succ)**.** $\mathrm{P}_q(n{+}1) = \mathrm{P}_q(n)\cdot \mathrm{m}_q(n)$ *in* $\mathbb{Z}/q\mathbb{Z}$.

The walk can be in one of two regimes:

1. **Living regime.** If q has not yet appeared in the sequence up to step n, then $q \nmid \mathrm{P}(n)$ (since $\mathrm{P}(n)$ is a product of the primes $\mathrm{em}(0), \ldots, \mathrm{em}(n)$, none equal to q). So $\mathrm{P}_q(n) \in (\mathbb{Z}/q\mathbb{Z})^\times$—the walk lives in the group of units.

2. **Dead regime.** If $q = \mathrm{em}(k)$ for some $k \le n$, then $q \mid \mathrm{P}(n)$, so $\mathrm{P}_q(n) = 0$. From this point on, $q \mid \mathrm{P}(m)$ for all $m \ge k$, so $\mathrm{P}_q(m) = 0$ forever. The walk has *collapsed to zero* (✓ walkZ_capture_then_collapse).

The transition from living to dead can happen only through -1:

Theorem 2.4 (Walk–Divisibility Bridge — ✓ walkZ_eq_neg_one_iff)**.** *For any prime q and any n such that $\mathrm{P}_q(n) \in (\mathbb{Z}/q\mathbb{Z})^\times$:*

$$\mathrm{P}_q(n) = -1 \quad \Longleftrightarrow \quad q \mid \big(\mathrm{P}(n) + 1\big).$$

Hitting -1 is a *necessary* condition for the walk to die (for q to enter the sequence): if $q \mid \mathrm{P}(n) + 1$, then $\mathrm{P}_q(n) = -1$. But it is not sufficient by itself. When $\mathrm{P}_q(n) = -1$, the walk dies only if $q = \mathrm{minFac}(\mathrm{P}(n) + 1)$, i.e. no smaller prime divides $\mathrm{P}(n) + 1$. If a smaller prime p also divides $\mathrm{P}(n) + 1$, then $\mathrm{minFac}(\mathrm{P}(n) + 1) \le p < q$ and q is *not* selected—the walk *bounces off* -1 and continues living. The walk may therefore hit -1 multiple times without dying. Death occurs only at a hit where q is the smallest available factor:

$$\cdots \to \underbrace{\mathrm{P}_q(n) = -1}_{\text{bounce}} \to (\mathbb{Z}/q\mathbb{Z})^\times \to \cdots \to \underbrace{\mathrm{P}_q(n_0) = -1}_{\substack{\text{lethal hit:} \\ q=\mathrm{minFac}}} \to \underbrace{\mathrm{P}_q(n_0{+}1) = 0}_{\text{dead}} \to 0 \to \cdots$$

What determines whether a hit on -1 is a bounce or a lethal hit? The answer depends on which other primes divide $\mathrm{P}(n) + 1$. This is controlled by the *sieve gap*.

Definition 2.5 (MC($<q$))**.** For a prime q, write MC($<q$) for the statement that every prime smaller than q appears in the EM sequence: $\forall\, p < q,\ p \text{ prime} \ \Rightarrow\ \exists\, k,\ \mathrm{em}(k) = p$.

When MC($<q$) holds, every prime $p < q$ has entered the sequence at some stage k_p, so $p \mid \mathrm{P}(n)$ for all $n \ge k_p$. Past a uniform stage $N_0 = \max_p k_p$, every prime $p < q$ divides $\mathrm{P}(n)$, and therefore cannot divide $\mathrm{P}(n) + 1$ (since $\gcd(\mathrm{P}(n), \mathrm{P}(n) + 1) = 1$). This is the **sieve gap**: past N_0, the only primes that can divide $\mathrm{P}(n) + 1$ are $\ge q$.

Theorem 2.6 (q-roughness — ✓ mcBelow_implies_seq_ge)**.** *If* MC($<q$) *holds, then $\exists\, N_0$ such that* $\mathrm{em}(n{+}1) \ge q$ *for all $n \ge N_0$.*

The sieve gap transforms hitting -1 from a necessary condition into a sufficient one:

Theorem 2.7 (Lethal hit — ✓ mcBelow_hit_is_lethal)**.** *If* MC($<q$) *holds and* $\mathrm{P}_q(n) = -1$ *for some n past the sieve gap, then* $\mathrm{em}(n{+}1) = q$.

Proof. Since n is past the sieve gap, no prime $< q$ divides $\mathrm{P}(n) + 1$ (Theorem 2.6). But $\mathrm{P}_q(n) = -1$ gives $q \mid \mathrm{P}(n) + 1$, so $q = \mathrm{minFac}(\mathrm{P}(n) + 1)$, and $\mathrm{em}(n{+}1) = q$. □

Definition 2.8 (Missing prime)**.** A prime q is **missing** if $\mathrm{em}(n) \ne q$ for all n.

Mullin's Conjecture asserts that no prime is missing. Under MC($<q$), the first hit on -1 past the sieve gap is lethal. This immediately constrains missing primes:

Theorem 2.9 (✓ mcBelow_missing_walk_ne_neg_one)**.** *If* MC($<q$) *holds and q is missing, then the walk* never *hits -1 past the sieve gap: $\exists\, N_0$ such that $\mathrm{P}_q(n) \ne -1$ for all $n \ge N_0$.*

Proof. Any hit past the sieve gap would give $\mathrm{em}(n{+}1) = q$ (Theorem 2.7), contradicting q missing. □

For the first missing prime q, the hypothesis MC($<q$) holds by definition. So its walk avoids -1 permanently past the sieve gap, the walk $\mathrm{P}_q(\cdot)$ is infinite: it stays in $(\mathbb{Z}/q\mathbb{Z})^\times$ forever, never collapsing to zero.

2.1 The forbidden multiplier and the death channel

Suppose the walk is at position $c \in (\mathbb{Z}/q\mathbb{Z})^\times$ at step n. The walk would hit -1 at step $n{+}1$ if and only if

$$\mathrm{P}_q(n) \cdot \mathrm{m}_q(n) = -1, \quad \text{i.e.,} \quad \mathrm{m}_q(n) = -c^{-1}.$$

So at every step, there is exactly **one forbidden multiplier** $f(n) = -\mathrm{P}_q(n)^{-1}$—the unique residue class that would send the walk to -1. This is 1 element out of $q-1$ possible unit residues.

Definition 2.10 (Death channel)**.** The **death channel** at position c is the residue class $b(c) = -c^{-1} \bmod q$.

When the walk is at c, the death channel is the set of primes $p \equiv -c^{-1} \pmod q$, which by Dirichlet's theorem has density $1/(q-1)$ among all primes. The death channel *moves* with the walk: as c changes from step to step, the forbidden class changes with it. The map $c \mapsto -c^{-1}$ is a bijection on $(\mathbb{Z}/q\mathbb{Z})^\times$ (it is the composition of inversion and negation).

Theorem 2.11 (✓ walk_hits_neg_one_iff_mult_eq_forbidden)**.** *For any step n with $\mathrm{P}_q(n) \in (\mathbb{Z}/q\mathbb{Z})^\times$:*

$$\mathrm{P}_q(n{+}1) = -1 \quad \Longleftrightarrow \quad \mathrm{m}_q(n) = -\mathrm{P}_q(n)^{-1}.$$

The statement "q is missing" is therefore equivalent to:

$$\text{For all } n: \quad \mathrm{m}_q(n) \neq -\mathrm{P}_q(n)^{-1}.$$

In words: the multiplier avoids the death channel at every step, forever. This brings us to discuss multipliers.

2.2 The Confinement Theorem and SubgroupEscape

Before asking whether the walk hits -1, we must ask whether it *can*. The walk is multiplicative: $\mathrm{P}_q(n) = \mathrm{P}_q(0) \cdot \prod_{k<n} \mathrm{m}_q(k)$. The reachable set is the coset $\mathrm{P}_q(0) \cdot \langle \text{multipliers} \rangle$.

Theorem 2.12 (Confinement — ✓ confinement_forward)**.** *If every multiplier lies in a subgroup $H \leq (\mathbb{Z}/q\mathbb{Z})^\times$, then the walk is confined to the coset $\mathrm{P}_q(0) \cdot H$. In particular, if $-1 \notin \mathrm{P}_q(0) \cdot H$, the walk* never *hits -1.*

This motivates:

Definition 2.13 (SubgroupEscape (**SE**))**.** For a prime q: no proper subgroup $H < (\mathbb{Z}/q\mathbb{Z})^\times$ contains all multipliers. Equivalently, $\langle \mathrm{m}_q(n) : n \in \mathbb{N} \rangle = (\mathbb{Z}/q\mathbb{Z})^\times$.

Theorem 2.14 (✓ se_of_maximal_escape)**.** *SE holds iff multipliers escape every* maximal *proper subgroup. Since $(\mathbb{Z}/q\mathbb{Z})^\times$ is cyclic, the maximal subgroups are the index-ℓ subgroups for prime $\ell \mid q-1$.*

When SE holds, the walk is not confined to any proper coset: every element of $(\mathbb{Z}/q\mathbb{Z})^\times$, including -1, is *reachable*. But reachability does not imply hitting—that is a dynamical question.

2.3 The first missing prime

We now assemble the key argument. Suppose, toward contradiction, that Mullin's Conjecture is false. Then there exists a **first missing prime**: the smallest prime q that never appears in the EM sequence. For this q:

(a) MC($<q$) holds—by definition, every prime smaller than q is in the sequence.
(b) The walk $\mathrm{P}_q(\cdot)$ is infinite—since q is missing, $\mathrm{P}_q(n) \in (\mathbb{Z}/q\mathbb{Z})^\times$ for all n.

(c) Past the sieve gap, the walk never hits -1. Here is why. Eventually (say for $n \geq N_0$), all primes $< q$ have entered the sequence and divide $\mathrm{P}(n)$. For $n \geq N_0$, no prime $< q$ can divide $\mathrm{P}(n)+1$ (since it divides $\mathrm{P}(n)$ and $\gcd(\mathrm{P}(n), \mathrm{P}(n)+1) = 1$). So if $\mathrm{P}_q(n) = -1$ for some $n \geq N_0$, then $q \mid \mathrm{P}(n)+1$, and q is the smallest prime factor of $\mathrm{P}(n)+1$ (all smaller primes are excluded). Then $\mathrm{em}(n+1) = q$, contradicting q being missing.

Therefore: **for the first missing prime q, the walk is an infinite trajectory on $(\mathbb{Z}/q\mathbb{Z})^\times$ that avoids the death channel at every step past the sieve gap.** Equivalently: $\mathrm{m}_q(n) \neq -\mathrm{P}_q(n)^{-1}$ for all $n \geq N_0$.

3 The Inductive Bootstrap

With the first missing prime's situation established (§2), we now show that the algebraic precondition—SubgroupEscape—comes for free, and that a single hit on -1 past the sieve gap suffices for MC.

The walk avoids -1 forever. Can it at least *reach* -1? That is: does SE hold?

Theorem 3.1 (PrimeResidueEscape (**PRE**); ✓ prime_residue_escape)**.** *For every prime $p \geq 5$ and every proper subgroup $H < (\mathbb{Z}/p\mathbb{Z})^\times$, some odd prime $r < p$ has residue $r \bmod p \notin H$.*

Proof. Suppose every odd prime $r \in [3, p)$ satisfies $r \bmod p \in H$. Since H is a subgroup, every *product* of such primes is in H. Every odd number in $[1, p)$ factors into odd primes $< p$, so every odd number in $[1, p)$ maps into H. In particular, $p-2 \equiv -2$ and $p-4 \equiv -4$ are both in H (both odd and $< p$ for $p \geq 5$). Then $2 = (-4)(-2)^{-1} \in H$, so every even number in $[1, p)$ is in H as well. Hence $H = (\mathbb{Z}/p\mathbb{Z})^\times$, contradicting H proper. □

Theorem 3.2 (✓ mcBelow_pre_implies_se)**.** $\mathrm{MC}(<p) + \mathrm{PRE} \implies \mathrm{SE}(p)$.

Proof sketch. Let $H < (\mathbb{Z}/p\mathbb{Z})^\times$ be proper. By PRE, some odd prime $r < p$ has $r \bmod p \notin H$. By $\mathrm{MC}(<p)$, the prime r appears as $\mathrm{em}(k)$ for some k. Then $\mathrm{m}_p(k-1) \equiv r \pmod p \notin H$. □

Corollary 3.3. *For the first missing prime q: the multipliers generate all of $(\mathbb{Z}/q\mathbb{Z})^\times$. SubgroupEscape holds, and -1 is reachable.*

The one-prime gap (Theorem 3.14) shows that $\mathrm{MC}(<q)$ plus a single divisibility event $q \mid \mathrm{P}(n)+1$ past the sieve gap gives $\mathrm{MC}(q)$. The bootstrap (Theorem 3.2) shows that $\mathrm{MC}(<q)$ gives $\mathrm{SE}(q)$ for free. These two facts compose into the primary reduction of Mullin's Conjecture.

Definition 3.4 (SingleHitHypothesis (**SHH**))**.** For every prime q: if $\mathrm{MC}(<q)$ and $\mathrm{SE}(q)$ hold and q is missing, then there exists n past the sieve gap with $q \mid \mathrm{P}(n)+1$.

Equivalently: for every prime q, if $\mathrm{MC}(<q)$ and $\mathrm{SE}(q)$ hold, then either q already appears in the sequence, or there exists $n \geq N_0(q)$ with $\mathrm{P}_q(n) = -1$.

Theorem 3.5 (Single Hit Theorem — ✓ single_hit_implies_mc)**.** SingleHitHypothesis $\implies$ MC.

Proof. By strong induction on p. Assume $\mathrm{MC}(<p)$.

1. **Bootstrap gives SE.** $\mathrm{MC}(<p) + \mathrm{PRE} \Rightarrow \mathrm{SE}(p)$ (Theorem 3.2). Since PRE is proved unconditionally (Theorem 3.1), we obtain $\mathrm{SE}(p)$.
2. **SHH gives a hit.** Since $\mathrm{MC}(<p)$ and $\mathrm{SE}(p)$ hold, SHH provides $n \geq N_0$ (past the sieve gap) with $p \mid \mathrm{P}(n)+1$.
3. **The sieve gap closes.** Past N_0, all primes $< p$ divide $\mathrm{P}(n)$ and hence cannot divide $\mathrm{P}(n)+1$. So $p = \mathrm{minFac}(\mathrm{P}(n)+1)$, giving $\mathrm{em}(n+1) = p$.

□

3.1 The death channel avoidance paradox

We now have a precise and sharp picture of the first missing prime q:

1. The walk lives in $(\mathbb{Z}/q\mathbb{Z})^\times$ forever.
2. The multipliers generate the full group—-1 is reachable.
3. The walk never reaches -1—the death channel is avoided at every step past the sieve gap.
4. At each step, the death channel is a single residue class out of $q-1$—a "target" of density $1/(q-1)$.

Mullin's Conjecture is therefore equivalent to: **this situation is impossible.** The open problem is now:

Does the multiplicative walk on $(\mathbb{Z}/q\mathbb{Z})^\times$ defined by the EM sequence, whose multipliers generate the full group, hit -1 at least once past the sieve gap?

"At least once past the sieve gap" is the precise requirement. Not cofinally, not equidistributed—once. The intuition for impossibility rests on an information-theoretic asymmetry. At step n, the death channel $f(n) = -\mathrm{P}_q(n)^{-1}$ is determined by the walk position $\mathrm{P}_q(n) = \mathrm{P}(n) \bmod q$, which carries $O(\log q)$ bits of information. The multiplier $\mathrm{m}_q(n) = \mathrm{minFac}(\mathrm{P}(n)+1) \bmod q$ is determined by the full integer $\mathrm{P}(n)+1$, which has $\sim 2^n$ bits. The factoring operation minFac extracts global arithmetic information from all $\sim 2^n$ bits; the death channel is determined by $O(\log q)$ bits. For the multiplier to systematically avoid the death channel, the $O(\log q)$-bit residue would have to predict the outcome of an operation on a 2^n-bit integer—a "prediction" whose information content vanishes exponentially.

More precisely: for a generic q-rough integer $N \equiv a \pmod q$, the conditional distribution of $\mathrm{minFac}(N) \bmod q$ is asymptotically uniform over $(\mathbb{Z}/q\mathbb{Z})^\times$, by CRT. Knowing $N \bmod q$ does not constrain $N \bmod p$ for any other prime p, so it does not constrain which primes $\leq N^{1/2}$ divide N. The density of the death channel among all primes is $1/(q-1)$, and the conditional probability that $\mathrm{minFac}(N)$ falls in the death channel is $1/(q-1)+o(1)$ as $N \to \infty$, *independently of the residue class a.*

For the EM sequence, $\mathrm{P}(n)+1$ grows super-exponentially ($\mathrm{P}(n)+1 \geq 2^{2^n}$), so the $o(1)$ error at each step shrinks exponentially fast. The "probability" of avoiding the death channel for N consecutive steps is heuristically $(1-1/(q-1))^N \to 0$. Converting this heuristic into a proof is the content of the conjecture.

3.2 Sufficient conditions for the single hit

Several formally verified strategies produce the required single hit past the sieve gap. Each implies MC via the Single Hit Theorem.

Definition 3.6 (DynamicalHitting (**DH**))**.** For every missing prime q: $\mathrm{SE}(q) \Rightarrow \mathrm{HH}(q)$, where $\mathrm{HH}(q)$ (HittingHypothesis) asks for cofinal hitting: $\forall\, N,\ \exists\, n \geq N,\ q \mid (\mathrm{P}(n)+1)$.

Theorem 3.7 (✓ dynamical_hitting_implies_mullin)**.** DH $\implies$ MC.

DH is stronger than SHH: it does not assume $\mathrm{MC}(<q)$ and asks for infinitely many hits rather than one. But the extra strength is convenient—DH interfaces cleanly with character sum methods. In the death channel language, DH asserts: if the multipliers generate $(\mathbb{Z}/q\mathbb{Z})^\times$, then the multiplier cannot avoid the forbidden residue $-\mathrm{P}_q(n)^{-1}$ forever. Equivalently: there is no infinite walk on a cyclic group, with a generating set of multipliers, that dodges a single moving target of density $1/(q-1)$ at every step.

The key structural point: SHH is **strictly weaker** than DynamicalHitting in two ways. First, SHH assumes $\mathrm{MC}(<q)$, which DH does not (DH only assumes SE). Second, SHH asks

for one hit past the sieve gap, while DH asks for cofinal hitting. Both extra assumptions are harmless in the inductive proof—MC($<q$) is always available at the inductive step, and one hit is all the Single Hit Theorem needs—but they make SHH genuinely easier to satisfy as a mathematical statement.

Remark 3.8. The logical relationships among the hitting hypotheses are:

$$\text{HH} \implies \text{DH} \implies \text{SHH},$$

where HH on its own denotes the *unconditional* version of HH(q) from Definition 3.6: cofinal hitting asserted for every missing q without assuming SE(q). DH conditions cofinal hitting on SE, and SHH asks for a single hit past the sieve gap given both MC($<q$) and SE. All three imply MC. SHH is the weakest: it assumes the most (MC($<q$) and SE, both provided free by the inductive bootstrap) and demands the least (one hit, not infinitely many). The converses need not hold.

Definition 3.9 (ComplexCharSumBound (**CCSB**))**.** For every missing prime q and every non-trivial character χ: $\|S_\chi(N)\| = o(N)$, where $S_\chi(N) = \sum_{n<N} \chi(\mathrm{P}_q(n))$.

Theorem 3.10 (✓ complex_csb_mc')**.** CCSB $\implies$ MC.

CCSB produces the hit via Fourier inversion: if all non-trivial character sums are $o(N)$, the hit count at -1 is $N/(q-1) + o(N)$, which is eventually positive. One hit past the sieve gap is lethal (Section 4).

Definition 3.11 (ConditionalMultiplierEquidist (**CME**))**.** For every missing prime q, every non-trivial χ, and every walk position c: the multiplier characters $\chi(\mathrm{m}_q(n))$ conditioned on $\mathrm{P}_q(n) = c$ are equidistributed (their partial sums are $o(N)$).

Theorem 3.12 (✓ cme_implies_mc)**.** CME $\implies$ MC (sharpest sufficient condition).

CME is the sharpest hypothesis: it asserts that the factoring operation destroys correlation between consecutive multiplier residues conditioned on the walk position. It implies CCSB for all character orders, bypassing the $d \geq 3$ barrier that blocks the BRE route.

Theorem 3.13 (✓ vcb_ped_implies_mc)**.** VCB + PED $\implies$ MC.

VCB (Vanishing Conditional Bias) is a weakening of CME that permits fiber sums to be proportional to visit counts rather than small; combined with PED (Positive Escape Density), it reaches CCSB.

Each of these strategies produces at least one hit past the sieve gap, which the Single Hit Theorem (Theorem 3.5) converts into MC.

Why it's hard: the walk controls nothing The difficulty is entirely in the coupling between walk position and multiplier. At step n:

- The **walk position** $\mathrm{P}_q(n) = \mathrm{P}(n) \bmod q$ determines the death channel $f(n) = -\mathrm{P}_q(n)^{-1}$.
- The **multiplier** $\mathrm{m}_q(n) = \mathrm{minFac}(\mathrm{P}(n)+1) \bmod q$ is what must fall into the death channel.
- Both depend on $\mathrm{P}(n)$: the walk position depends on $\mathrm{P}(n) \bmod q$, and the multiplier depends on $\mathrm{P}(n) + 1$ as a full integer.

The walk position sees $O(\log q)$ bits; the multiplier sees all $\sim 2^n$ bits. But they share the same underlying object $\mathrm{P}(n)$. The question is whether this shared dependence creates a correlation strong enough for the multiplier to systematically avoid one residue class out of $q-1$.

For a *random* sequence of multipliers drawn uniformly from $(\mathbb{Z}/q\mathbb{Z})^\times$, the probability of avoiding the death channel for N steps is $(1-1/(q-1))^N \to 0$. The EM walk is not random—it is deterministic—but the factoring operation that determines each multiplier is, heuristically, a decorrelating "hash." This is the **factorization independence heuristic**: the function

minFac, applied to a 2^n-bit integer, produces an output that is effectively uncorrelated with the $O(\log q)$-bit residue class of the input.

Every open hypothesis in this paper—DH, CCSB, CME—is a precise formalization of this heuristic.

3.3 Supporting Infrastructure

The remainder of this section collects the supporting results that flesh out the bootstrap: the one-prime gap, the sieve gap, and the power residue decomposition.

The Sieve Gap and the One-Prime Gap The q-roughness theorem (Theorem 2.6) resolves the selectability problem described in Section 1: past the sieve gap, q is the smallest available factor whenever it divides the Euclid number. This gives the one-prime gap:

Theorem 3.14 (One-prime gap — ✓ mcBelow_cofinal_hit_implies_mc_at)**.** $\mathrm{MC}(<q)$ *plus a single hitting event* $q \mid \mathrm{P}(n)+1$ *for some* n *past the sieve gap implies* $\mathrm{MC}(q)$.

The Power Residue Decomposition Since $(\mathbb{Z}/q\mathbb{Z})^\times$ is cyclic of order $q-1$, its subgroup lattice is determined by the prime factorization of $q-1$. A multiplier set generates the full group if and only if it escapes every maximal subgroup—and the maximal subgroups correspond to the prime divisors ℓ of $q-1$. This decomposition converts SE into independent conditions, one per prime $\ell \mid q-1$.

Definition 3.15 (PowerResidueEscape (**PRE**$_\ell$))**.** $\exists\, n : \mathrm{m}_q(n)^{(q-1)/\ell} \neq 1$ in $(\mathbb{Z}/q\mathbb{Z})^\times$.

Theorem 3.16 (**PRE** $\Leftrightarrow$ **SE** — ✓ pre_iff_se)**.** PRE $\iff$ SE. *The forward direction uses only Lagrange's theorem; the reverse uses cyclicity of* $(\mathbb{Z}/q\mathbb{Z})^\times$.

Quadratic Reciprocity Obstruction The power residue decomposition raises a natural question: for how many primes q could SE actually *fail*? Since each PRE$_\ell$ condition asks only that *one* multiplier among infinitely many escapes the ℓ-th power subgroup, SE failure requires all multipliers to be ℓ-th power residues for some $\ell \mid q-1$. By CRT and quadratic reciprocity, the density of primes q for which the first k multiplier primes all fail to escape the index-2 subgroup is at most $O(2^{-k})$. Extending this to a rigorous density-zero statement for SE failure would require controlling all prime divisors $\ell \mid q-1$ simultaneously, which remains open.

4 The Character Sum Reduction

Sections 2–3 established the picture: for the first missing prime q, the walk lives in $(\mathbb{Z}/q\mathbb{Z})^\times$ forever, the multipliers generate the full group, yet the walk avoids -1 permanently past the sieve gap. The death channel $f(n) = -\mathrm{P}_q(n)^{-1}$ is dodged at every step. The Single Hit Theorem (§**??**) reduces MC to producing one hit on -1 past the sieve gap at each prime.

The question is not "does the walk equidistribute?"—that is far stronger than needed. The question is: *can the walk really avoid one class out of* $q-1$ *forever?* This section develops the harmonic-analytic tools to show it cannot: character sums detect the anomaly that permanent avoidance would create, and bounding them rules it out.

Why character sums? Permanent avoidance of -1 means the hit count $|\{n < N : \mathrm{P}_q(n) = -1\}|$ stays at zero past the sieve gap. Character orthogonality decomposes this count into a uniform share $N/(q-1)$ plus correction terms built from non-trivial character sums $S_\chi(N) = \sum_{n<N} \chi(\mathrm{P}_q(n))$. The uniform share grows linearly; for the hit count to remain zero, the correction terms must cancel this growth. That requires at least one non-trivial character sum to

be $\Omega(N/(q-1))$—a sustained asymmetry in the character spectrum. The hypothesis CCSB (all $|S_\chi(N)| = o(N)$) rules this out: it forces the hit count to be eventually positive, contradicting permanent avoidance.

What is new vs. what is infrastructure. The CCSB $\Rightarrow$ MC reduction (Definition 4.2), the Fourier bridge (Theorem 4.3), the Decorrelation–PED chain (§4.3), and the telescoping no-go results (§4.5) are original contributions of this formalization. The large sieve infrastructure (§4.6, Appendix D)—including the weak ALS, Gauss sum inversion, van der Corput, and Parseval—formalizes known results; its purpose is to identify the precise *transfer gap* between classical tools and the EM orbit, which is itself a contribution (see §4.6).

The formalization develops this Fourier-analytic reduction because character sums provide the cleanest interface between the death-channel avoidance problem and the toolkit of analytic number theory: the Bombieri–Vinogradov theorem, the large sieve inequality, and Gauss sum inversion all produce character sum bounds, and the formalization shows exactly how each connects to ruling out permanent avoidance.

4.1 Character Sums and Permanent Avoidance

For a Dirichlet character $\chi : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}^\times$, the *walk character sum* is

$$S_\chi(N) \;=\; \sum_{n<N} \chi\big(\mathrm{P}_q(n)\big).$$

Definition 4.1 (ComplexCharSumBound (**CCSB**))**.** For every missing prime q, every non-trivial character χ, and every $\varepsilon > 0$, there exists N_0 such that for all $N \geq N_0$:

$$\|S_\chi(N)\| \;\leq\; \varepsilon \cdot N.$$

In other words, the walk character sums are $o(N)$—they grow strictly slower than linearly. The contrapositive: if the walk permanently avoids -1, some non-trivial character sum must be $\Omega(N)$ (to cancel the uniform share in the Fourier bridge). CCSB rules this out.

Theorem 4.2 (✓ complex_csb_mc')**.** CCSB $\Longrightarrow$ MC.

This is a single-hypothesis reduction with no additional parameters. The proof composes three bridges:

1. **Fourier inversion** (✓ complex_csb_implies_hit_count_lb_proved): CCSB implies that the hit count at -1 satisfies $|\{n < N : \mathrm{P}_q(n) = -1\}| = N/(q-1) + o(N)$, which is eventually positive. This directly contradicts permanent avoidance past the sieve gap (the character orthogonality formula, Theorem 4.3, gives this for any target t; specializing to $t = -1$ is what matters).
2. **Cofinal hitting gives HH** (✓ walk_equidist_mc): the eventually-positive hit count at -1 gives cofinal hitting, hence HH. (In fact, the Fourier bridge yields equidistribution across all classes, but only the $t = -1$ case is needed. SE is a side effect: a walk visiting -1 cannot be confined to a proper coset.)
3. **DH implies MC** (Theorem 3.7): via the inductive bootstrap.

4.2 The Fourier Bridge

The Fourier bridge is the single most important proved result after DH $\Rightarrow$ MC itself: it converts character sum bounds into hit count lower bounds, and hence into MC.

Theorem 4.3 (✓ walk_hit_count_fourier_step)**.** *For any target $t \in (\mathbb{Z}/q\mathbb{Z})^\times$:*

$$\big|\{n < N : \mathrm{P}_q(n) = t\}\big| \;=\; \frac{1}{q-1} \sum_{\chi} \overline{\chi(t)}\, S_\chi(N),$$

where the sum is over all Dirichlet characters mod q.

This is a standard Fourier inversion formula on the finite group $(\mathbb{Z}/q\mathbb{Z})^\times$, stated for all t. Specializing to $t = -1$ and splitting the trivial character χ_0 (with $\chi_0(a) = 1$ for all a) from the non-trivial ones:

$$\big|\{n < N : \mathrm{P}_q(n) = -1\}\big| \;=\; \underbrace{\frac{N}{q-1}}_{\text{uniform share}} \;+\; \underbrace{\frac{1}{q-1}\sum_{\chi\neq\chi_0} \overline{\chi(-1)}\, S_\chi(N)}_{\text{correction}}.$$

The first term grows linearly. The correction term is bounded by $\frac{1}{q-1}\sum_{\chi\neq\chi_0}|S_\chi(N)|$, since $|\overline{\chi(-1)}| = 1$. If CCSB holds—all $|S_\chi(N)| = o(N)$—the correction is $o(N)$, and the hit count is $N/(q-1) + o(N)$, which is eventually positive.

The contrapositive. Section 3 showed that for the first missing prime q, the hit count at -1 is exactly 0 past the sieve gap. By the Fourier bridge, this forces the correction terms to cancel the entire main term $N/(q-1)$. Since there are only $q-2$ non-trivial characters, at least one must satisfy $|S_\chi(N)| = \Omega(N/(q-1)^2)$—a sustained linear-scale bias in the character spectrum. CCSB ($= o(N)$) rules this out. The Fourier bridge thus converts the geometric statement "the walk avoids -1 permanently" into the spectral statement "some character sum has sustained bias," and CCSB negates the latter.

Returning to our running example ($q = 41$): the walk must accumulate a deficit of $\sim N/40$ hits at -1 compared to uniform, which requires sustained character sum bias across the 39 non-trivial characters. Proving CCSB for $q = 41$ would rule out this bias and establish that 41 eventually appears.

4.3 The Decorrelation–PED–CCSB Chain

CCSB rules out permanent avoidance of -1. But the walk is built from *multipliers*: $\mathrm{P}_q(n+1) = \mathrm{P}_q(n) \cdot \mathrm{m}_q(n)$. Since χ is a group homomorphism, $\chi(\mathrm{P}_q(n)) = \chi(\mathrm{P}_q(0)) \cdot \prod_{k<n}\chi(\mathrm{m}_q(k))$: the walk character sum is a sum of *partial products* of the multiplier characters. The question becomes: what properties of the multiplier sequence would prevent the walk from permanently dodging the death channel?

This subsection formalizes a chain of progressively weaker hypotheses about the multipliers, each implying the next via proved bridges. The goal is to decompose CCSB into sharper conditions on the multiplier sequence, and to identify exactly where the irreducible difficulty lies.

Definition 4.4 (PositiveEscapeDensity (**PED**))**.** For every missing prime q and non-trivial χ, there exist $\delta > 0$ and N_0 such that for $N \geq N_0$: $|\{k < N : \chi(\mathrm{m}_q(k)) \neq 1\}| \geq \delta N$.

The name "escape" comes from the SubgroupEscape perspective: the kernel $\ker(\chi)$ is a proper subgroup of $(\mathbb{Z}/q\mathbb{Z})^\times$, and $\chi(\mathrm{m}_q(k)) \neq 1$ means the k-th multiplier "escapes" from $\ker(\chi)$. PED asks that a positive fraction of multipliers escape *every* proper subgroup, not just occasionally but with positive density. The connection to the death channel: since the death channel is a single class and multipliers that escape $\ker(\chi)$ "rotate" the walk character value $\chi(\mathrm{P}_q(n))$ by a non-trivial amount, enough escapes should prevent the systematic avoidance needed for permanent death-channel dodging. PED is a weak condition—it says nothing about cancellation, only that the multipliers are not asymptotically trapped in any subgroup.

Definition 4.5 (DecorrelationHypothesis)**.** For every missing prime q and non-trivial χ, the multiplier character sums are $o(N)$: $\|\sum_{n<N}\chi(\mathrm{m}_q(n))\| \leq \varepsilon N$ for large N.

Decorrelation is stronger than PED: it asks not merely that many multipliers escape $\ker(\chi)$, but that they do so with enough balance that the character values cancel. If the multipliers were independent random elements of $(\mathbb{Z}/q\mathbb{Z})^\times$, the sum would be $O(\sqrt{N})$ by the law of large numbers—far smaller than εN. Decorrelation asks for the much weaker $o(N)$.

Definition 4.6 (NoLongRuns(L))**.** For every missing prime q and non-trivial χ, past some threshold, no L consecutive multipliers all lie in $\ker(\chi)$.

NoLongRuns is a qualitative cousin of PED: if multipliers never stay inside $\ker(\chi)$ for L steps in a row, then at least $1/(2L)$ of them escape. This condition is easier to verify in practice because it only requires checking short blocks.

Definition 4.7 (BlockRotationEstimate (**BRE**))**.** If the escape count is $\geq \delta N$, then the walk character sums are $o(N)$. This encapsulates the Cauchy–Schwarz / van der Corput step in harmonic analysis.

BRE is the bridge between the multiplier-level conditions (PED/Decorrelation) and the walk-level condition (CCSB). It says: given that multipliers escape with positive density, the walk character sums must cancel. The intuition is that each escape event "rotates" the walk character value $\chi(\mathrm{P}_q(n))$ by a non-trivial amount, and sufficiently many such rotations produce cancellation in the sum. BRE is the sole unproved bridge in the PED route.

The hypotheses above—PED, Decorrelation, NoLongRuns—all treat the multiplier sequence as a single stream, ignoring what the walk is doing at the moment. But permanent death-channel avoidance is a statement about the *coupling* between walk position and multiplier: the death channel $-\mathrm{P}_q(n)^{-1}$ is a function of walk position, so avoiding it means the multiplier distribution at each walk position is biased away from one class. A more refined hypothesis should address this coupling head-on.

Definition 4.8 (ConditionalMultiplierEquidist (**CME**))**.** For every missing prime q, non-trivial χ, $\varepsilon > 0$, there exists N_0 such that for $N \geq N_0$ and every walk position $c \in (\mathbb{Z}/q\mathbb{Z})^\times$: $\| \sum_{\substack{n<N \\ \mathrm{P}_q(n)=c}} \chi(\mathrm{m}_q(n)) \| \leq \varepsilon N$.

CME says the multiplier distribution is the same at every walk position. Since the death channel is a function of walk position, CME implies the death channel has no special status—the multiplier is no more likely to avoid the forbidden class than any other. Formally, CME is strictly stronger than Decorrelation: it bounds the fiber sums $\sum_{\substack{n<N \\ \mathrm{P}_q(n)=c}} \chi(\mathrm{m}_q(n)) = o(N)$ separately for each position c, not just the global sum. Since the global sum is the sum of the fiber sums, CME implies Decorrelation by the triangle inequality (✓ cme_implies_dec). The significance of CME is that it also implies CCSB *directly*, bypassing PED and BRE entirely: the fiber decomposition and the telescoping identity together convert conditional multiplier cancellation into walk character sum cancellation, for *all* character orders, without needing the intermediate PED → BRE → CCSB chain.

Theorem 4.9 (✓ decorrelation_implies_ped)**.** *Decorrelation ⇒ PED.*

Proof sketch. Contrapositive. If few multipliers escape $\ker(\chi)$—say fewer than δN—then most contribute $\chi(m(n)) = 1$ to the sum. The at most δN exceptions contribute values of norm ≤ 1. By the reverse triangle inequality, $|\sum \chi(m(n))| \geq N - 2\delta N$, which is $\geq \varepsilon N$ for δ small enough. This contradicts Decorrelation. □

Theorem 4.10 (✓ noLongRuns_implies_ped)**.** *NoLongRuns(L) ⇒ PED with $\delta = 1/(2L)$.*

Proof sketch. Partition $\{0, \ldots, N-1\}$ into blocks of length L. Each block contains at least one escape (by assumption), so the total escape count is $\geq N/(2L)$. □

Theorem 4.11 (✓ block_rotation_implies_ped_csb)**.** *BRE ⇒ PEDImpliesComplexCSB.*

The PED route, with all proved arrows:

$$\text{Dec} \xrightarrow{\text{proved}} \text{PED} \xleftarrow{\text{proved}} \text{NoLongRuns}(L) \xrightarrow{\text{BRE, open}} \text{CCSB} \xrightarrow{\text{proved}} \text{MC}.$$

The sole open bridge in this route is BRE: converting positive escape density into walk character sum cancellation.

However, the PED route is not the only path. CME implies CCSB *directly*, bypassing PED and BRE entirely (✓ cme_implies_ccsb):

$$\text{CME} \xrightarrow{\text{proved}} \text{CCSB} \xrightarrow{\text{proved}} \text{MC}.$$

This bypass is significant: the $d \geq 3$ barrier (Remark 5.7) blocks the PED $\to$ CCSB factorization for characters of order ≥ 3, but CME $\to$ CCSB works for *all* character orders, using only the telescoping identity and fiber decomposition.

4.4 Vanishing Conditional Bias

CME is a strong hypothesis: it asks that *every* fiber character sum $F(c,\chi) = \sum_{\substack{n<N \\ w(n)=c}} \chi(m(n))$ be $o(N)$—in other words, that $\mu = 0$ in the proportionality $F(c,\chi) \approx \mu \cdot V_N(c)$. But the telescoping route to CCSB does not need $\mu = 0$; it needs only that $\mu \neq 1$. This observation motivates a strictly weaker hypothesis.

Definition 4.12 (VanishingConditionalBias (**VCB**))**.** For every missing prime q, non-trivial χ, and $\varepsilon > 0$, there exists N_0 such that for $N \geq N_0$ there is a complex number $\mu = \mu(N,\chi)$ with $|\mu| \leq 1$ satisfying, for all $c \in (\mathbb{Z}/q\mathbb{Z})^\times$:

$$\left\| \sum_{\substack{n<N \\ w(n)=c}} \chi(m(n)) \;-\; \mu \cdot V_N(c) \right\| \;\leq\; \varepsilon \cdot N,$$

where $V_N(c) = |\{n < N : w(n) = c\}|$ is the visit count.

In words: the factoring channel treats all walk positions proportionally (the formal content of the factoring channel analogy from §1)—the character statistics of multipliers are the same at every position, up to a common constant μ (which may vary with N and χ but must be common across all positions c). Combined with PED (enough multipliers escape), the proportionality constant cannot equal 1; the telescope then forces the walk character sum to be $o(N)$, ruling out the sustained bias needed for permanent avoidance.

Proposition 4.13 (✓ cme_implies_vcb)**.** CME $\Rightarrow$ VCB *with* $\mu = 0$. *Conversely,* VCB + Dec $\Rightarrow$ CME. *Thus CME decomposes as* CME = VCB + Dec. *VCB alone is strictly weaker than CME: it permits the fiber sums to be* $\Theta(N)$*, provided they are proportional to the visit counts.*

Theorem 4.14 (✓ vcb_ped_implies_ccsb)**.** VCB + PED $\implies$ CCSB.

Proof sketch. From VCB, the telescope identity gives $(\mu - 1) \cdot S_N = O(1) + O(\varepsilon N(q-1))$. If $|1-\mu|$ is bounded away from zero, then $S_N = O(\varepsilon N/|1-\mu|) = o(N)$.

To show $|1-\mu| \geq c_0 > 0$, PED provides $\delta > 0$ such that at least δN multipliers escape $\ker(\chi)$. For each escaping multiplier, $\chi(m(n))$ is a non-trivial d-th root of unity, satisfying $\mathrm{Re}(\chi(m(n))) \leq 1 - \eta_0^2/2$ where $\eta_0 = \min_{\zeta \in \mu_d \setminus \{1\}} |\zeta - 1| > 0$ (✓ norm_sub_one_sq_eq, ✓ unit_norm_re_le_of_dist). This real-part defect accumulates over δN escaping steps, forcing $|\sum \chi(m(n)) - N| \geq c_0 N$ for $c_0 = \delta\eta_0^2/2$. Combined with VCB's control $|\sum \chi(m(n)) - \mu N| \leq C\varepsilon N$, the reverse triangle inequality gives $|1-\mu| \geq c_0/2$ for small ε. □

Theorem 4.15 (✓ vcb_ped_implies_mc)**.** VCB + PED $\implies$ MC.

This decomposition splits the monolithic CME hypothesis into two independently attackable pieces:

- **VCB** ("the factoring channel preserves proportionality"): the character distribution of multipliers is the same at every walk position, up to a common rate. This captures the "factorization independence" intuition—that the factoring operation destroys the correlation between walk position and multiplier residue.
- **PED** ("enough multipliers escape every character kernel"): a positive fraction of multiplier residues fall outside any proper subgroup.

VCB is a recent contribution of the formalization.

4.5 Walk Telescoping Identities

The hypotheses above attack CCSB by decomposing it into conditions on the multiplier sequence. Before proceeding to the large sieve route, we pause to examine what the walk recurrence $\chi(w(n{+}1)) = \chi(w(n)) \cdot \chi(m(n))$ itself forces. This multiplicative structure is the *only* algebraic relation connecting walk and multiplier character values, and it constrains any possible proof of CCSB. The following identities make these constraints explicit. They are formalized not because they solve CCSB, but because they reveal the structural landscape: they show which proof strategies are compatible with the walk's algebra and which are ruled out.

Theorem 4.16 (✓ walk_telescope_identity)**.** *For any χ and N:*

$$\sum_{n<N} \chi(w(n)) \cdot \big(\chi(m(n)) - 1\big) \;=\; \chi(w(N)) - \chi(w(0)).$$

This identity follows immediately from the walk recurrence $\chi(w(n{+}1)) = \chi(w(n)) \cdot \chi(m(n))$: writing $\chi(w(n)) \cdot (\chi(m(n)) - 1) = \chi(w(n{+}1)) - \chi(w(n))$, the sum telescopes to $\chi(w(N)) - \chi(w(0))$.

Theorem 4.17 (✓ walk_telescope_norm_bound)**.** *The telescoping sum has norm ≤ 2 (triangle inequality on unit-norm terms).*

The ≤ 2 bound looks innocent, but it constrains how the walk can respond to the death channel. Each summand $\chi(w(n)) \cdot (\chi(m(n)) - 1)$ is non-zero exactly when $\chi(m(n)) \neq 1$—i.e., when the multiplier escapes $\ker(\chi)$, "rotating" the character value. The ≤ 2 bound means the net rotation over all N steps is negligible, tightly coupling the walk character sum $S_N = \sum_{n<N} \chi(w(n))$ to the multiplier character sum $M_N = \sum_{n<N} \chi(m(n))$. Splitting the product in Theorem 4.16 gives $S_N \cdot \overline{M_N/N} - S_N = O(1)$ (after normalization).

Theorem 4.18 (✓ walk_shift_one_correlation)**.** $\sum_{n<N} \chi(w(n)) \cdot \overline{\chi(w(n+1))} = \overline{\sum_{n<N} \chi(m(n))}$.

This identity says that the lag-1 autocorrelation of the walk character equals the conjugate of the multiplier character sum. It is a *no-go result* for the van der Corput method with $H = 1$: VdC bounds $|S_N|^2$ in terms of autocorrelations, but at lag $h = 1$, the autocorrelation is exactly $|M_N|$—the multiplier character sum—which need not be small. So VdC with a single shift gives only $|S_N| \leq O(\sqrt{N \cdot |M_N|})$, which is $O(N)$ in the worst case, not the $o(N)$ that CCSB requires. This means any proof of CCSB must either (i) use higher-order correlations (HOD, Appendix D) or (ii) establish multiplier decorrelation first.

4.6 The Large Sieve Route

The preceding subsections attacked the question "can the walk avoid -1 permanently?" for a single modulus q. A different strategy works with *many moduli simultaneously*: classical analytic number theory produces character sum estimates averaged over all moduli $q \leq Q$, and such averaged results are often easier to prove than pointwise ones. The large sieve inequality and the Bombieri–Vinogradov theorem are the two central tools for this.

The formalization develops this multi-modular route across three files (LargeSieve.lean, LargeSieveHarmonic.lean, LargeSieveAnalytic.lean) totaling ∼5,870 lines, connecting classical analytic number theory to MC via a multi-modular character sum bound.

Why formalize the large sieve? The analytic large sieve inequality and the Bombieri–Vinogradov theorem are among the most powerful tools in analytic number theory for controlling the distribution of primes in arithmetic progressions. If these tools could be applied to the EM walk, MC would follow. We formalize the connection—not the deep theorems themselves (which are known results, stated as open Props)—for two reasons:

1. To identify *precisely* what transfer hypothesis is needed to apply each classical result to the specific EM orbit, and
2. To verify that six apparently independent routes (BV, ArithLS, ALS, PrimeArithLS, LoD, sieve transfer) all reduce to the *same* orbit-specificity gap.

This diagnosis is itself a mathematical contribution: it shows that the difficulty of MC is not a failure of existing analytic tools but a fundamental obstacle in applying ensemble-averaged results to a single deterministic orbit.

Definition 4.19 (MultiModularCSB (**MMCSB**))**.** There exists Q_0 such that for all $q \geq Q_0$ prime, every non-trivial character χ mod q, and every $\varepsilon > 0$, there exists N_0 such that for $N \geq N_0$: $\|S_\chi(N)\| \leq \varepsilon N$.

MultiModularCSB is weaker than CCSB in that it allows finitely many exceptional primes below Q_0. This weakening is crucial because averaged results like BV naturally produce bounds that fail for finitely many moduli.

Theorem 4.20 (✓ mmcsb_implies_mc)**.** MultiModularCSB $\implies$ MC.

The proof composes the per-prime Fourier bridge (Theorem 4.3) with the inductive bootstrap: for $q \geq Q_0$, MMCSB gives hit count $\sim N/(q-1) > 0$ at -1, contradicting permanent avoidance; for the finitely many primes $q < Q_0$, the bootstrap (Section 3) handles them—once MC holds for all primes below q, the sieve gap and one-prime gap (Theorem 3.14) reduce MC(q) to a single hitting event.

Three parallel routes to MultiModularCSB are formalized:

Bombieri–Vinogradov route. BV says primes are equidistributed among arithmetic progressions "on average over moduli": for most $q \leq Q = \sqrt{x}/(\log x)^A$, the count of primes $\leq x$ in any progression a mod q is close to the expected $\pi(x)/\phi(q)$. If this applies to the EM multipliers, the death channel—a single progression—gets its fair share of multipliers, breaking the avoidance.

Theorem 4.21 (✓ bv_chain_mc)**.** BV + BVImpliesMMCSB $\implies$ MC.

The transfer hypothesis BVImpliesMMCSB is a **genuine frontier**: it requires transferring the averaged equidistribution statement of BV (valid for primes in generic progressions) to the specific EM walk orbit. The EM sequence is not a generic sample of primes—it is a deterministic sequence defined by iterated factorization—so its multipliers could exhibit special correlations that BV's averaged estimate cannot detect.

The route was decomposed into two stages:

$$\text{BV} \xrightarrow{\text{sieve transfer}} \text{EMMultCSB} \xrightarrow{\text{walk bridge}} \text{MMCSB} \xrightarrow{\text{proved}} \text{MC},$$

separating the number-theoretic content (BV $\Rightarrow$ EMMultCSB, where EMMultCSB bounds the *multiplier* character sums) from the dynamical content (EMMultCSB $\Rightarrow$ MMCSB, converting multiplier bounds to *walk* bounds). However, the walk bridge **MultCSBImpliesMMCSB is false in general** (✓ MultCSBImpliesMMCSB): the walk character sum $\sum \chi(w(n))$ is a *partial product* $\prod_{k<n} \chi(m(k))$ of the multiplier characters, and partial products of equidistributed unit complex numbers need not cancel—they perform a random walk on the unit circle whose norm grows as $\sqrt{N}$, not as $o(N)$. The telescope identity (Theorem 4.18) makes this obstruction precise: the

$h=1$ autocorrelation equals the multiplier character sum, so van der Corput with a single shift gives only $O(N)$, not $o(N)$. This is why the CME bypass (fiber decomposition + telescoping, Definition 4.8) is essential: it goes directly from conditional multiplier equidistribution to CCSB without ever requiring the walk bridge.

Additional sieve routes. Two further routes are formalized—the arithmetic large sieve (ArithLS ⇒ MC, a dead end) and the analytic large sieve (ALS ⇒ PrimeArithLS ⇒ MC), where the ALS-to-PrimeArithLS bridge via Gauss sum inversion is fully proved across eight internal lemmas. In both cases, the genuine open content is the same *orbit-specificity transfer*: applying averaged results to one deterministic orbit. The full details, including the ALS definition, weak ALS proof, Gauss sum inversion theorem, and a spectral energy route (SVE, van der Corput, HOD, CME) with its complete hypothesis hierarchy, appear in Appendix D.

5 Why It's Hard

Original contribution. The selectability analysis, oracle barrier, and CCSB-as-frontier argument below are new. They explain *why* the remaining hypothesis resists both computation and existing analytic tools.

5.1 The Selectability Perspective

The Euclid construction guarantees fresh primes at every step: every prime factor of $\mathrm{P}(n)+1$ is new (coprime to the running product). The difficulty of MC is not the *existence* of new primes but whether the minFac rule eventually *selects* each one. The formalization makes this contrast precise.

Theorem 5.1 (✓ divisor_not_yet_in_seq)**.** *If $p \mid \mathrm{P}(n)+1$, then $\mathrm{em}(m) \neq p$ for all $m \leq n$.*

Proof. Any $\mathrm{em}(m)$ with $m \leq n$ divides $\mathrm{P}(n)$. A number ≥ 2 cannot divide both a and $a+1$. □

Theorem 5.2 (✓ passed_over_persists)**.** *If $p \mid \mathrm{P}(n)+1$ but $\mathrm{em}(n+1) \neq p$ (the* minFac *rule chose a smaller prime), then $\mathrm{em}(m) \neq p$ for all $m \leq n+1$. The prime survives to potentially divide future Euclid numbers.*

Theorem 5.3 (✓ selectability_extinguished)**.** *Once $\mathrm{em}(m) = p$, we have $p \mid \mathrm{P}(n)$ for all $n \geq m$, so $p \nmid \mathrm{P}(n)+1$ ever again. Selectability is a one-shot resource.*

Definition 5.4 (InfinitelySelectable)**.** A prime p is *infinitely selectable* if $p \mid \mathrm{P}(n)+1$ for cofinally many n: $\forall N, \exists n \geq N, p \mid \mathrm{P}(n)+1$.

By Theorem 5.3, MC(p) and InfinitelySelectable(p) are mutually exclusive (✓ mc_implies_not_infinitely_selectable): a prime that enters the sequence can never be selectable again.

Theorem 5.5 (✓ dh_implies_infinitely_selectable)**.** *Under DH, every prime that never appears in the sequence (with SE satisfied) is infinitely selectable.*

The random-factor variant is easy. Consider a variant of the Euclid–Mullin construction where, instead of the smallest prime factor, one picks a *random* prime factor of $\mathrm{P}(n)+1$ at each step. In this variant, MC follows from DH alone: whenever $p \mid \mathrm{P}(n)+1$, simply choose p. Under DH with SE, this happens infinitely often (Theorem 5.5), so p eventually gets picked.

The argument is even simpler probabilistically. For any target prime p, the residue $r_n = \mathrm{P}(n) \bmod p$ performs a multiplicative walk on $(\mathbb{Z}/p\mathbb{Z})^\times$. In the random-factor variant, each multiplier is a random element of the group; once the multipliers generate the full group (which PRE guarantees), the walk is a genuine random walk with full support. By classical equidistribution on finite groups, r_n converges to uniform, so $r_n = -1$ (i.e., $p \mid \mathrm{P}(n)+1$) occurs with probability $\to 1/(p-1)$—infinitely often with probability 1.

The lpf variant is hard. The actual Euclid–Mullin sequence uses $\mathrm{em}(n{+}1) = \mathrm{minFac}(\mathrm{P}(n){+}1)$, a *deterministic* function of the walk position. This creates the exact correlation identified by the oracle analysis (§5.2): the multiplier at step n depends on the full value of $\mathrm{P}(n)+1$, coupling walk position to multiplier. The random-factor variant breaks this coupling by choosing multipliers independently of position; the minFac rule preserves it.

The difficulty of Mullin's Conjecture is entirely in the minimality of the prime selection, not in the Euclidean construction itself. The inductive bootstrap (Section 3) bridges this gap: $\mathrm{MC}(<p)$ ensures all primes below p are already in the sequence, hence divide $\mathrm{P}(n)$, hence cannot divide $\mathrm{P}(n)+1$. Past a computable stage, p is the *smallest* available factor whenever it divides the Euclid number—reducing the minFac variant to the "any-factor" variant for the tail of the sequence.

5.2 The Marginal/Joint Barrier

The verified reductions (TailSE, CofinalEscape, QuotientDH) exhaust what can be proved about the *marginal* distribution of multiplier residues.

Theorem 5.6 (✓ emfe_iff_tail_se_at)**.** $\mathrm{EuclidMinFacEscape}(q) \Leftrightarrow \mathrm{TailSE}(q)$.

Even perfect per-position equidistribution of multipliers is consistent with HH failure. DH is a *joint* statement—the (position, multiplier) pair must hit the *death curve* $\mathrm{m}_q(n) = -\mathrm{P}_q(n)^{-1}$—and no marginal statement can force this.

The orbit chain gap. The cofinal orbit analysis picks one cofinal multiplier s_x per walk position, producing a cycle $x_0 \to x_1 \to \cdots \to x_0$ in $(\mathbb{Z}/q\mathbb{Z})^\times$. Even when the cofinal multipliers generate the full group, the cycle size k can be less than $|(\mathbb{Z}/q\mathbb{Z})^\times|$. Example: in $\mathbb{Z}/6\mathbb{Z}$, the cycle $0 \to 1 \to 0$ has $\langle 1, 5\rangle = \mathbb{Z}/6\mathbb{Z}$ but misses 3.

Closing this gap requires showing that at each cofinally visited position, *multiple* multiplier classes appear—expanding the cycle until it covers -1. This is the "specific-orbit problem": transferring generic equidistribution of minFac residues to the particular EM orbit.

5.3 The BRE Impossibility for $d \geq 3$

Remark 5.7. Positive escape density (PED) alone does *not* imply CCSB for characters of order $d \geq 3$.

Counterexample: a walk on $\mathbb{Z}/3\mathbb{Z}$ that alternates between only two of the three d-th roots of unity (phase-aligned escapes) achieves positive escape density yet has walk sum $\approx N/2{\cdot}(1{+}\omega) \neq o(N)$.

For $d = 2$ this degeneracy vanishes: the only non-trivial rotation is -1, so escape frequency *is* the rotation distribution. The order-2 BRE from $\mathrm{NoLongRuns}(L)$ is proved in the formalization (✓ order2_noLongRuns_mc). But for $d \geq 3$, PED constrains how often the walk rotates without constraining the *distribution* among $d-1$ non-identity rotations. The PED $\Rightarrow$ BRE $\Rightarrow$ CCSB factorization is invalid for $d \geq 3$.

This barrier is specific to the PED route. The CME $\to$ CCSB reduction (✓ cme_implies_ccsb) bypasses PED and BRE entirely, working for all character orders d via the telescoping identity. The $d \geq 3$ problem is therefore not a barrier for the *reduction*—only for the particular factorization through PED.

The Van der Corput Barrier The van der Corput inequality (Theorem D.10, now fully proved in the formalization) converts character sum bounds into autocorrelation bounds. Theorem 4.18 gives $R_1 = o(N)$ under the Decorrelation Hypothesis. VdC with $H = 1$ yields $|S_N|^2 \leq \frac{N+1}{2}(N{+}2|R_1|) = N^2/2{+}o(N^2)$, hence $|S_N| \leq N/\sqrt{2}$. This is non-trivial but *not* $o(N)$.

To get $o(N)$, one needs higher-order correlations $R_h = o(N)$ for $h \geq 2$, which requires HigherOrderDecorrelation (Theorem D.12). The telescoping identity $\sum_n \chi(w(n))(\chi(m(n)) - 1) = O(1)$ is a precise structural constraint.

The Walk Bridge Falsity The BV route decomposes into two stages: sieve transfer (BV ⇒ EMMultCSB, bounding *multiplier* character sums) and the walk bridge (EMMultCSB ⇒ MMCSB, converting multiplier bounds to *walk* bounds). The walk bridge **MultCSBImpliesMMCSB** (✓ MultCSBImpliesMMCSB) is stated as an open Prop and is **false in general**.

The obstruction is structural: the walk character sum is a *partial product* $\chi(w(n)) = \prod_{k<n} \chi(m(k))$ of the multiplier characters. Even when the individual factors $\chi(m(k))$ are equidistributed on the unit circle (so their *sum* cancels), their *partial products* perform a multiplicative walk whose norm grows as $\sqrt{N}$, not $o(N)$. Cancellation of sums does not imply cancellation of cumulative products.

This negative result explains why the CME bypass (Definition 4.8) is essential. CME uses fiber decomposition and telescoping to go directly from conditional multiplier equidistribution to CCSB, circumventing the walk bridge entirely.

5.4 The Factorization Independence Heuristic

The preceding barriers explain why MC is hard to *prove*. This subsection explains why it should be *true*—and why the formalization's sole remaining hypothesis (CME) is the precise mathematical content of a natural intuition about factorization.

The information bottleneck. The information bottleneck of §3.2—$O(\log q)$ bits visible to the walk versus $\sim 2^n$ bits determining the multiplier—means the mutual information vanishes exponentially. This is directly analogous to the pseudorandomness of iterated hash functions: if H is a cryptographic hash, the sequence $x, H(x), H(H(x)), \ldots$ is deterministic but statistically indistinguishable from random, because the hash destroys recoverable correlations. For the EM sequence, integer factorization plays the role of the hash function. Both processes are fully deterministic (given the seed, every term is uniquely determined), one-way (computing forward is trivial; extracting structure from the output is computationally hard), and *de facto* uncorrelated (consecutive terms pass every reasonable test for independence).

The analogy is heuristic, not rigorous. SHA-256 is *designed* for pseudorandomness; minFac is not designed for anything. We do not claim that computational hardness of factoring implies MC—what MC requires is a number-theoretic statement (CME), not a complexity-theoretic one. But the analogy explains the structure of the problem: proving pseudorandomness of a deterministic process requires showing that *no exploitable structure exists*, which is harder than finding structure. This is why the conjecture resists proof despite overwhelming heuristic evidence.

Negative analogies: why the selection rule matters. The factorization independence heuristic explains why the EM sequence (minFac variant) should contain all primes, while related sequences do not.

- The **Sylvester sequence** $s(n+1) = s(0) \cdots s(n) + 1$ has density-zero prime divisors [17]. Its terms grow doubly exponentially, giving each prime only $O(1)$ chances to appear as a factor. The "hash chain" runs too fast.
- **Fermat numbers** $F_n = 2^{2^n} + 1$ have the even stronger property that $\sum 1/p$ converges over their prime divisors. Again, doubly exponential growth limits opportunities.
- The **second EM sequence** (maxFac variant) provably omits infinitely many primes [11, 5]. Here the "hash function" (maxFac instead of minFac) produces large multipliers,

causing the product to grow rapidly—analogous to using a hash function that amplifies rather than compresses.

- The **first EM sequence** (minFac variant) keeps the product growing slowly—heuristically as $\exp(n^2/2)$ (see §6). This is analogous to a hash function that compresses, giving exponentially many iterations within any fixed modulus. The conjecture is that this compression ensures coverage.

What the formalization adds. The formalization identifies three equivalent formulations of the factoring channel's decorrelation:

1. **Conditional Multiplier Equidistribution (CME)**: the distribution of $\text{minFac}(\text{P}(n)+1) \bmod q$, conditioned on $\text{P}(n) \equiv c \pmod q$, is asymptotically independent of c (Definition 4.8).
2. **Decorrelation Hypothesis**: the character sum $\sum \chi(\text{m}_q(n)) = o(N)$ for nontrivial χ—the multiplier residues have cancelling character sums, as independent random variables would.
3. **ComplexCharSumBound (CCSB)**: the walk character sums $\sum \chi(\text{P}_q(n)) = o(N)$ for nontrivial χ—ruling out permanent avoidance of any residue class.

These are related by proved implications (CME $\Rightarrow$ Dec $\Rightarrow$ PED) and direct bypass (CME $\Rightarrow$ CCSB). Each implies MC through the verified reduction chain. The irreducible mathematical content is: does the factoring operation destroy enough correlation for character sums to cancel?

5.5 Dead Ends as a Roadmap

Over a hundred potential approaches were explored and found to be dead ends (catalogued in full in the project's dead_ends.md). Each elimination is informative: it narrows the space of viable strategies. A representative selection, grouped by the type of obstruction:

Dead end	Why it fails
ENSEMBLE-TO-ORBIT TRANSFER: *tool applies to generic sequences, not the specific EM orbit*	
BV for EM subsequence	BV applies to all primes in APs, not to a greedy subsequence.
Furstenberg / ergodic theory	Standard ergodic methods assume classical multiplicativity; the EM sequence is recursive and non-multiplicative.
Diaconis–Shahshahani lemma	Requires i.i.d. random steps; inapplicable to the deterministic EM walk.
INDEPENDENCE / LINEARITY VIOLATED: *tool requires additive or independent structure the walk lacks*	
Large sieve for partial products	The large sieve handles linear sums, not multiplicative walks.
Abel summation	Converts multiplier decorrelation to walk-sum bounds, but the summation weights *amplify* rather than cancel.
Self-avoidance $\Rightarrow$ CCSB	Self-avoidance (no repeated $\hat{\mathbb{Z}}$ positions) is invisible to characters, which see only residues.
WRONG ALGEBRAIC STRUCTURE: *the group or decomposition has no room for the desired bound*	
Non-abelian / representation	$(\mathbb{Z}/q\mathbb{Z})^\times$ is cyclic; all irreps are 1-d characters. No higher-dimensional structure to exploit.
CRT product group	Reformulating on $\prod_{q\le Q}(\mathbb{Z}/q\mathbb{Z})^\times$ makes the problem harder: the product group is exponentially large.
NoLongRuns + PED $\Rightarrow$ BRE ($d\ge 3$)	Variable block lengths align adversarially with character phases.
DPED $\Rightarrow$ CCSB ($d\ge 3$)	Alternating ω, ω^2 rotations satisfy DPED yet produce $\Theta(N)$ walk sums. All PED-to-CCSB intermediates ruled out.
REDUCES TO SINGLE-MODULUS CCSB: *no genuine simplification*	
Multi-modular approaches	All variants (BV + threshold, CRT, death coupling) collapse to single-modulus CCSB.
Death set coupling across moduli	Death sets $\{m : \mathrm{minFac}(m) \equiv -c^{-1}\}$ vary per step; no uniform coupling bound exists.
Spectral gap (deterministic walk)	Spectral gap theory applies to probability measures on groups (convergence of random sampling); the EM walk is a single deterministic path.
Information-theoretic bounds	Category error: entropy/mutual-information tools assume a random variable; the EM sequence is deterministic with zero entropy.

The Four-Way Blocker. The majority of the 105 dead ends reduce to a single meta-obstacle: every known technique for proving equidistribution of sequences on finite groups requires at least *one* of (1) independence of steps, (2) multiplicativity of the generating function, (3) algebraic-geometric structure (parameter families, monodromy), or (4) ergodic stationarity. The EM walk satisfies *none* of these: the steps are deterministic and mutually dependent, the walk is not a multiplicative function, the multipliers have no known algebraic-geometric parametrization, and the non-autonomous dynamics (a different multiplier at each step) rule out stationarity. This explains why classical tools—random-walk mixing, Halász's theorem, Katz monodromy, Birkhoff averages—all fail.

The telescope exhausts the algebra. The telescoping identity $\chi(w(n+1)) = \chi(w(n)) \cdot \chi(m(n))$ is the *complete* algebraic content of the walk. Only two decomposition strategies for

the character sum $S_N = \sum_{n<N} \chi(w(n))$ exist: grouping by *value* (fiber decomposition → CME) and grouping by *lag* (autocorrelation → HOD). Every other rearrangement—Abel summation, Möbius inversion, Dirichlet series, block decomposition—either reduces to one of these two or fails outright (Abel gives $O(N^2)$ remainder in the wrong direction; Möbius/Dirichlet require multiplicativity). There is no third algebraic route to CCSB.

The pattern: every approach that avoids the specific EM orbit's joint distribution either reduces to CCSB or fails. Since CME implies CCSB (proved), and CME decomposes as VCB + Dec (§4.4), the sharpest targets are now VCB + PED or CME: conditional equidistribution of multipliers given walk position, or the weaker proportionality condition combined with escape density. CME is strictly weaker than CCSB and is the *irreducible analytic content.*

The Mathematical Landscape We need to prove one of these equivalent statements for every missing prime q:

- **DH**: If the multipliers generate $(\mathbb{Z}/q\mathbb{Z})^\times$, the walk $\mathrm{P}_q(n) = \mathrm{P}(n) \bmod q$ hits -1 cofinally.
- **CCSB**: For every non-trivial character $\chi\colon (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}^\times$, the sum $\sum_{n<N} \chi(\mathrm{P}_q(n)) = o(N)$.
- d=2 **special case** (as a stepping stone): For every quadratic character χ, the ± 1-valued walk character sum is $o(N)$.

The formalization has conclusively shown that every tool requiring independence, classical multiplicativity, or ensemble averaging fails. So we need ideas that exploit what the EM walk specifically has.

Structural Features of the EM Walk The dead ends above show what does not work. Complementarily, the EM walk has four structural features—all proved or formalized—that no dead-end approach has successfully exploited. Any proof of MC will almost certainly use at least one.

Feature 1: Super-exponential growth. $\mathrm{P}(n) \geq 2^n$ (✓ prod_lower_bound_for_sieve). The Euclid numbers $\mathrm{P}(n) + 1$ grow absurdly fast. This means the *sieve level*—the threshold below which all prime factors have been excluded—grows super-exponentially. By step n, the Euclid number $\mathrm{P}(n) + 1$ is coprime to each of $\mathrm{em}(0), \ldots, \mathrm{em}(n)$, a growing set of distinct primes. The pool of "available" small primes as factors of $\mathrm{P}(n) + 1$ shrinks, but the size of $\mathrm{P}(n) + 1$ grows so fast that it must have enormous prime factors most of the time.

Feature 2: Mutual coprimality of Euclid numbers. For $m > n$, $\mathrm{P}(m)$ is divisible by $\mathrm{em}(n+1)$, which divides $\mathrm{P}(n)+1$. So $\mathrm{P}(m)+1 \equiv 1 \pmod{\mathrm{em}(n+1)}$: successive Euclid numbers live in different residue classes modulo earlier sequence terms. This coprimality structure means the Euclid numbers cannot all "avoid" a residue class in a coordinated way—their residues are forced apart by the construction.

Feature 3: The multiplier is the smallest prime factor. This is the key constraint that everyone mentions but nobody has quantified. If $\mathrm{P}_q(n) \neq -1$ (so that q does not divide $\mathrm{P}(n)+1$ as the smallest factor), then the multiplier $\mathrm{m}_q(n) = \mathrm{minFac}(\mathrm{P}(n)+1)$ satisfies $\mathrm{m}_q(n) \leq (\mathrm{P}(n) + 1)^{1/2}$. For a number of size $\sim 2^n$, this smallest factor could be as small as 3 or as large as $\sim 2^{n/2}$. The minFac rule creates a deterministic coupling between walk position and multiplier: the multiplier at step n depends on the full value of $\mathrm{P}(n) + 1$, not just its residue.

Feature 4: Self-correcting feedback. If the walk concentrates on certain residues mod q—say $\mathrm{P}_q(n) \equiv a \pmod q$ for many n—then $\mathrm{P}(n)+1 \equiv a+1 \pmod q$ for many n. The smallest prime factor of numbers $\equiv a + 1 \pmod q$ depends on $a + 1$, creating a feedback loop: concentration in one residue class biases the multiplier distribution, which in turn pushes the walk away from that class. This self-correcting mechanism has been formalized (EquidistSelfCorrecting.lean), but

all paths from it lead to SieveTransfer—the open hypothesis that generic minFac equidistribution transfers to the specific EM orbit. We return to this feature in our assessment of whether DH is true (§7).

These four features—growth, coprimality, the minFac selection rule, and self-correcting feedback—are the raw material that any successful approach must engage with. The dead ends above fail precisely because they treat the walk generically (as a random walk, or as an arbitrary multiplicative walk) rather than exploiting the specific arithmetic of the EM construction.

6 The Lean Formalization

Having developed the mathematical reduction and identified the structural obstacles, we describe the Lean 4 formalization that certifies these results.

Codebase Structure The formalization uses Lean 4 with Mathlib v4.27.0 across 35 files totaling ∼26,900 lines. The dependency chain is linear, with three leaf modules:

File	Content	Lines
Euclid.lean	Constructive Euclid's theorem	422
MullinDefs.lean	seq, prod, aux, identities	527
MullinConjectures.lean	MC, Conjecture A (FALSE), HH	490
MullinDWH.lean	DivisorWalkHypothesis (leaf)	547
MullinResidueWalk.lean	WalkCoverage, residue walk, concrete MC	603
MullinGroupCore.lean	walkZ, multZ, confinement, SE	422
MullinGroupEscape.lean	Escape lemmas, 8-element argument	673
MullinGroupSEInstances.lean	29 concrete SE instances ($q \leq 157$)	364
MullinGroupPumping.lean	Gordon sequenceability (leaf)	343
MullinGroupQR.lean	QR obstruction ($\leq 1.6\%$) (leaf)	683
MullinCRT.lean	CRT multiplier invariance, walk recurrence	160
MullinDepartureGraph.lean	Departure graph, infinite recurrence, safe prime lattice	393
RotorRouter.lean	Scheduled walk coverage (standalone)	421
MullinRotorBridge.lean	EMPR + SE ⇒ MC bridge	87
EquidistPreamble.lean	PE ⇒ MC, bootstrapping	234
EquidistSieve.lean	Sieve, WHP ⇔ HH, forbidden multiplier, Möbius death	755
EquidistSelfAvoidance.lean	Self-avoidance, periodicity	450
EquidistCharPRE.lean	Character non-vanishing, PRE ⇔ SE	811
EquidistBootstrap.lean	Inductive bootstrap, DH ⇒ MC, first passage	617
EquidistThreshold.lean	Sieve gap, one-prime gap, cofinal pair avoidance	325
EquidistOrbitAnalysis.lean	Cofinal orbits, quotient walk, sieve, selectability	1441
EquidistFourier.lean	Character sums, Fourier bridge	1298
EquidistSelfCorrecting.lean	Decorrelation, BRE, telescoping, kernel (§31–§37, §72)	1163
EquidistSieveTransfer.lean	Sieve transfer, coprimality, neg-inv involution (§38–§78)	1457
CMEVariants.lean	CME weaker variants (CME_d, CME_avg, CME_subseq, CME_target)	113
LargeSieve.lean	BV, ALS, ArithLS, MMCSB, sieve bridge (§41–§52, §79)	1812
LargeSieveHarmonic.lean	Parseval, Gauss sums, DFT, kernel (§53–§55)	892
LargeSieveAnalytic.lean	Gauss inversion, WeakALS, GCT, dead ends (§56–§65, §81–§82)	1683
LargeSieveSpectral.lean	Walk energy, HOD, VdC, CME, VCB, SVE, transition matrix (§66–§86)	2670
IKCh1.lean	Iwaniec–Kowalski [12] Ch. 1: arithmetic functions	437
IKCh2.lean	Iwaniec–Kowalski Ch. 2: summation formulas	270
IKCh3.lean	Iwaniec–Kowalski Ch. 3: combinatorial sieve	557
IKCh4.lean	Iwaniec–Kowalski Ch. 4: summation formulas	593
IKCh5.lean	Iwaniec–Kowalski Ch. 5: Kloosterman sums	877
IKCh7.lean	Iwaniec–Kowalski Ch. 7: bilinear forms, duality, Gram, MLS, sieve	2269

6.1 Axiom Usage: What's Constructive

The core definitions (seq, prod, aux) and their basic properties (seq_isPrime, seq_injective) are **fully constructive**: they use only propext and Quot.sound (no Classical.choice, no Decidable instances beyond $\mathbb{N}$). Euclid's theorem itself (Euclid.lean) is constructive.

Classical reasoning enters at the reduction level:

- The HH $\Rightarrow$ MC proof uses well-founded induction (strong induction on $\mathbb{N}$), which in Lean 4 is constructive but relies on Classical.choice for the cofinal-implies-hit argument.
- Character theory (orthogonality, Fourier inversion) is inherently classical via open Classical.
- All open hypotheses are stated as def ... : Prop, never as sorry'd theorems. The type-checker guarantees that no proof obligation is silently assumed.

6.2 Mathlib Dependencies

The formalization draws on several Mathlib libraries:

- **Group theory**: Subgroup, QuotientGroup, orderOf, cyclic group structure, maximal subgroups (Subgroup.IsCoatom).
- **Number theory**: Nat.minFac, Legendre symbols, quadratic reciprocity, ZMod, Dirichlet characters, Gauss sums.
- **Character theory**: DirichletCharacter.Orthogonality, roots of unity in algebraically closed fields, character bounds, MulChar.sum_eq_zero_of_ne_one.
- **Analysis**: norm_sum_le, complex norms, IsOfFinOrder.norm_eq_one, Fourier analysis on $\mathbb{Z}/n\mathbb{Z}$ (ZMod.dft, discrete Fourier transform).
- **Dirichlet's theorem**: Nat.infinite_setOf_prime_and_eq_mod (primes in arithmetic progressions, via L-series).
- **Harmonic analysis**: Parseval's theorem for finite abelian groups, trigonometric exponentials, geometric series identities.

Lines of Lean code	~26,900
Files	35
Theorems/lemmas	~910
Definitions	~460
sorry occurrences	**0**
Open hypotheses (stated as def)	~26
Mathlib version	v4.27.0

7 Open Problems

The formalization certifies the reductions above and pinpoints the frontier. This section identifies the open problems whose resolution would close Mullin's Conjecture.

7.1 CCSB as the Precise Frontier

The formalization identifies **ComplexCharSumBound** as the irreducible analytic content. The question:

Are the walk character sums $\sum_{n<N} \chi(\mathrm{P}_q(n))$ *bounded* $o(N)$ *for every non-trivial* χ*?*

CCSB is a single hypothesis that implies MC with no additional conditions. As shown in Section 4, CCSB rules out permanent avoidance of -1: the Fourier bridge forces the hit count at -1 to be $N/(q-1) + o(N)$, contradicting the zero hits that the first missing prime's walk must achieve past the sieve gap.

The walk telescoping identities (Section 4) provide precise structural constraints. The identity $\sum_n \chi(w(n))(\chi(m(n)) - 1) = O(1)$ means that the walk sum S_N and the multiplier sum $M_N = \sum_n \chi(m(n))$ satisfy $S_N \approx S_N + (M_N - S_N) = M_N + O(1)$ only in the crude sense; the telescoping does *not* separate them.

7.2 Connection to Bombieri–Vinogradov

A Bombieri–Vinogradov type result for EM walk residues would give:

$$\sum_{\substack{q \le Q \\ q \text{ prime}}} \max_a \left| |\{n \le N : w(n) \equiv a \pmod q\}| - \frac{N}{q-1} \right| \ll \frac{NQ}{(\log N)^A}.$$

For non-exceptional primes, the hit count at -1 would be $\sim N/(q-1) > 0$, contradicting permanent avoidance; the finitely many exceptional primes below Q_0 are handled by the inductive bootstrap (§3), closing the conjecture.

The difficulty is that BV applies to the set of *all* primes, not to a specific subsequence. The EM walk is deterministic and self-referential: the walk at step n depends on the factorization of $\mathrm{P}(n) + 1$, which depends on all previous walk values. Standard BV does not apply.

7.3 Connection to Chebotarev

The **EffectiveKummerEscape** hypothesis asserts: for each prime ℓ, there exists B such that for $q \ge B$ with $\ell \mid q-1$, some multiplier among the first B escapes the ℓ-th power kernel. This is a Chebotarev-type statement for the Kummer extension $\mathbb{Q}(\zeta_\ell, 3^{1/\ell}, \ldots, 53^{1/\ell})$: the Frobenius at q determines which multiplier primes are ℓ-th power residues.

An effective Chebotarev density theorem for this fixed number field would give EKE for all but finitely many q (effectively bounded). Combined with finite verification for the remaining q, this would prove PRE and hence SE unconditionally—but SE is *already* proved unconditionally via the elementary PRE. The Chebotarev approach would give a stronger *effective* bound on how quickly SE kicks in, refining the density argument of §3.

7.4 The Sieve-Theoretic Approach

MertensEscape: for any prime q and proper subgroup H, infinitely many primes outside H exist (Dirichlet content). **SieveAmplification**: Mertens escape should force eventual $\mathrm{minFac}(\mathrm{P}(n)+1)$ escape from H, via super-exponential growth and mutual coprimality of successive Euclid numbers.

The formally verified chain: MertensEscape + SieveAmplification $\xrightarrow{\text{proved}}$ TailSE $\xrightarrow{\text{proved}}$ CofinalEscape $\xrightarrow{\text{proved}}$ QuotientDH.

The formalization articulates a richer sieve infrastructure in two parallel routes:

Cumulative route.

$$\mathrm{PDE} \xrightarrow{\text{Alladi}} \mathrm{GLPFE} \xrightarrow{\text{SieveTransfer}} \mathrm{SieveEquidist} \xrightarrow{\text{open}} \mathrm{NoLongRuns}$$
$$\xrightarrow{\text{proved}} \mathrm{PED} \xrightarrow{\text{open}} \mathrm{CCSB} \xrightarrow{\text{proved}} \mathrm{MC}.$$

Here PDE is PrimeDensityEquipartition (PNT in arithmetic progressions, a known theorem not yet in Mathlib), and GLPFE is GenericLPFEquidist (Alladi's theorem [9] on minFac distribution of generic integers, also known but not formalized). Both ends of the chain—from PDE to GLPFE via Alladi, and from CCSB to MC via Fourier inversion—are formally proved.

Window route.

$$\text{StrongSieveEquidist} \xrightarrow{\text{proved}} \text{NoLongRunsAt} \xrightarrow{\text{proved}} \text{PEDAt} \xrightarrow{\text{open}} \text{CCSB} \xrightarrow{\text{proved}} \text{MC}.$$

StrongSieveEquidist asserts that EM multipliers are equidistributed within sliding windows; NoLongRunsAt and PEDAt are per-prime variants proved by pigeonhole and block-counting respectively (✓ strongSieveEquidist_noLongRunsAt, ✓ noLongRunsAt_ped).

The genuine frontier. **SieveTransfer** is the critical open hypothesis: does the equidistribution of minFac residues for generic integers transfer to the specific EM orbit? Everything above SieveTransfer is known mathematics; everything below it is proved. SieveTransfer is where "known but not formalized" meets "genuinely open."

The difficulty: for *generic* q-rough integers, minFac residues are equidistributed (by CRT + Mertens)—the death channel gets its fair share of multipliers. But the EM Euclid numbers are not generic: each is determined by the entire walk history. Transferring the generic equidistribution to this specific orbit is the open step.

Sieve-to-harmonic convergence. The sieve hierarchy (§36–§39 of EquidistSelfCorrecting.lean) and the harmonic hierarchy (§30–§35) converge: both produce DecorrelationHypothesis as output. The full chain

$$\text{SieveEquidist} \xrightarrow{\text{proved}} \text{Dec} \xrightarrow{\text{proved}} \text{PED} \xrightarrow[\text{sole gap}]{\text{open}} \text{CCSB} \xrightarrow{\text{proved}} \text{MC}$$

is formalized, with the first two arrows machine-verified (✓ sieve_equidist_implies_decorrelation, ✓ decorrelation_implies_ped). The sieve route achieves SieveEquidist $\Rightarrow$ Dec via a counting-to-character-sum bridge: SieveEquidistribution produces EMMultCharSumBound with $Q_0 = 0$, meaning multiplier character sums cancel for *all* primes q, which is exactly DecorrelationHypothesis. The sole remaining gap on this route is **PEDImpliesComplexCSB** (✓ PEDImpliesComplexCSB): does positive escape density for all primes imply walk character sum cancellation? Any proof of SieveEquidistribution (e.g., from PNT in APs + Alladi's theorem) would immediately yield Dec and PED for free, isolating this single bridge as the only open step.

7.5 What Would Close the Conjecture

The cleanest paths to MC:

1. **Prove CME** (sharpest target): show that the multiplier character sum $\sum_{\substack{n<N \\ w(n)=c}} \chi(m(n))$ is $o(N)$ for each walk position c. CME is strictly weaker than CCSB, and CME $\Rightarrow$ CCSB is proved (✓ cme_implies_ccsb). CME asks only about the *conditional* distribution of multipliers given walk state—it does not require controlling the walk character sum itself. This bypasses the $d \geq 3$ barrier entirely.

2. **Prove CCSB directly**: show that no non-trivial character sum can sustain the $\Omega(N)$ bias that permanent avoidance of -1 would require. The self-correcting sieve (concentration of EM primes in a residue class is exponentially self-limiting) is the strongest heuristic argument.

3. **Prove a BV-type estimate**: even an averaged bound over q would suffice—for $q \geq Q_0$ it rules out permanent avoidance, and the finitely many $q < Q_0$ are handled by the inductive bootstrap (§3).

4. **Close the orbit chain gap**: show that at each cofinally visited walk position, at least two distinct multiplier classes appear. This would force the orbit chain to expand to the full group.

5. **Prove DH directly**: show that a multiplicative walk on a cyclic group with a generating set of multipliers must hit every element cofinally. This is a combinatorial question about deterministic walks.

6. **Prove SieveTransfer**: show that the EM orbit's minFac distribution matches that of generic integers, at least on average. The cumulative sieve route (§38) reduces this to known number theory (PNT in APs + Alladi's theorem); closing SieveTransfer gives CME (and hence CCSB) via the conditional multiplier equidistribution framework.

7. **Prove BVImpliesMMCSB or PrimeArithLSImpliesMMCSB**: the large sieve route (§41–§65) reduces MC to a transfer hypothesis. Given Bombieri–Vinogradov or the analytic large sieve, the remaining open step is showing that the EM orbit's multipliers receive their fair share of each residue class—in particular, that the death channel is not systematically avoided. This is the same orbit-specificity gap as SieveTransfer, approached from a different mathematical toolkit.

7.6 Does the Walk Hit -1?

The irreducible open question is not whether DH holds (cofinal hitting is a convenient sufficient condition) but whether the walk hits -1 *at least once* past the sieve gap. The Single Hit Theorem (Theorem 3.5) shows that a single hit at each prime suffices for MC; DH, CCSB, and CME are strategies for producing that hit.

Evidence for. Two independent lines of evidence suggest the walk does hit -1. (1) *Self-correcting feedback*: the formalized sieve analysis (EquidistSelfCorrecting.lean) shows that concentration of EM primes in a residue class is exponentially self-limiting—a walk biased toward missing -1 automatically biases the multiplier distribution toward correcting that miss. (2) *Analogy with Artin's conjecture*: Artin's conjecture (that every non-square integer is a primitive root for infinitely many primes) has the same orbit-specificity structure and is believed true; Hooley [7] proved it conditional on GRH. The walk-hitting question is the analogous statement for the EM walk and would follow from an analogous uniformity hypothesis.

Evidence against. Two features of the EM sequence give pause. (1) *Cox–van der Poorten* [11]: the "largest factor" variant of the Euclid–Mullin sequence provably misses primes. The EM sequence's completeness is not a soft consequence of the Euclid construction but depends sensitively on the minFac selection rule. This fragility means heuristic arguments ("it should work because Euclid numbers have many factors") are not reliable. (2) *The $d \geq 3$ barrier*: the formalization proves that the most natural route from multiplier escape density to walk character sum cancellation (PED $\Rightarrow$ BRE $\Rightarrow$ CCSB) is *impossible* for character orders $d \geq 3$. This is not evidence against the hit, but it shows that producing one requires mechanisms beyond the simplest equidistribution framework. The CME bypass sidesteps this barrier, but the barrier's existence means any proof must be genuinely subtle.

Assessment. We believe the walk does hit -1 at every prime, primarily because the self-correcting sieve mechanism provides a concrete dynamical reason (not merely a probabilistic heuristic) for the walk to eventually hit -1. The strongest form of this belief: CME should hold because the EM multipliers, conditioned on walk position, have no arithmetic reason to avoid the death channel systematically. But we acknowledge that no existing technique can prove this, and the $d \geq 3$ barrier shows that the proof, when found, will need to exploit the specific structure of the EM walk in ways that current analytic number theory does not.

8 Summary of Verified Results

Result	Status	Lean identifier
Sequence foundations		
Every em(n) is prime	Proved	seq_isPrime
No prime repeats	Proved	seq_injective
Main reductions to MC		
SHH ⇒ **MC** (Single Hit Theorem)	Proved	single_hit_implies_mc
DH ⇒ **MC** (cofinal hitting route)	Proved	dynamical_hitting_implies_mullin
CCSB ⇒ **MC** (single hypothesis)	Proved	complex_csb_mc'
PE ⇒ MC	Proved	pe_implies_mullin
HH ⇒ MC	Proved	hh_implies_mullin
SE + MH ⇒ MC	Proved	se_mixing_implies_mullin
WE ⇒ **MC** (single Prop)	Proved	walk_equidist_mc
Inductive bootstrap		
PrimeResidueEscape (elementary)	Proved	prime_residue_escape
MC($<p$) + PRE ⇒ SE(p)	Proved	mcBelow_pre_implies_se
q-roughness from MC($<q$)	Proved	mcBelow_implies_seq_ge
One-prime gap	Proved	mcBelow_cofinal_hit_implies_mc_at
mcBelow 11	Proved	concrete_mcBelow_11
Algebraic framework		
Confinement Theorem	Proved	confinement_forward/reverse
PRE ⇔ SE	Proved	pre_iff_se
SE ⇔ character detection	Proved	se_iff_char_detection
Maximal subgroup reduction	Proved	se_of_maximal_escape
WHP ⇔ HH	Proved	whp_iff_hh
SE density argument (CRT + QR)	Proved	se_qr_obstruction
Character sum chain		
Fourier bridge: CCSB ⇒ hit count lb	Proved	complex_csb_implies_hit_count_lb_proved

continued on next page

Result	Status	Lean identifier
Decorrelation ⇒ PED	Proved	decorrelation_implies_ped
NoLongRuns(L) ⇒ PED	Proved	noLongRuns_implies_ped
BRE ⇒ PEDImpliesCSB	Proved	block_rotation_implies_ped_csb
CME ⇒ CCSB (all d, bypasses BRE)	Proved	cme_implies_ccsb
CME ⇒ MC	Proved	cme_implies_mc
Walk char recurrence ($\mathbb{C}$-valued)	Proved	char_walk_recurrence
Telescoping identity	Proved	walk_telescope_identity
Telescoping norm ≤ 2	Proved	walk_telescope_norm_bound
Shift-one autocorrelation	Proved	walk_shift_one_correlation
Order-2 sign-flip chain	Proved	order2_noLongRuns_mc
Walk dynamics		
Walk–divisibility bridge	Proved	walkZ_eq_neg_one_iff
Products strictly monotone	Proved	prod_strictMono
Fundamental trichotomy	Proved	avoidance_contradicts_se_mixing
Self-avoidance dichotomy	Proved	self_avoidance_dichotomy
Scheduled walk coverage	Proved	scheduled_walk_covers_all
Selectability analysis		
Divisor freshness	Proved	divisor_not_yet_in_seq
Passed-over persistence	Proved	passed_over_persists
Selectability extinction	Proved	selectability_extinguished
MC ⇒ ¬InfinitelySelectable	Proved	mc_implies_not_infinitely_selectable
DH ⇒ InfinitelySelectable	Proved	dh_implies_infinitely_selectable
Sieve and orbit analysis		
EMFE ⇔ TailSE	Proved	emfe_iff_tail_se_at
TailSE ⇒ CofinalEscape ⇒ QuotientDH	Proved	tail_se_gives_sub_dh
Dirichlet: ∞ primes per residue class	Proved	dirichlet_residues_independent

continued on next page

Result	Status	Lean identifier
Minimality sieve + coupling	Proved	minimality_sieve
StrongSieveEquidist ⇒ NoLongRunsAt	Proved	strongSieveEquidist_noLongRunsAt
NoLongRunsAt ⇒ PEDAt	Proved	noLongRunsAt_ped
DPED ⇒ PED	Proved	dped_implies_ped
PDE + sieve chain ⇒ MC	Proved	primeDensity_chain_mc
GLPFE + SieveTransfer ⇒ MC	Proved	genericLPF_chain_mc
SieveEquidist ⇒ Dec	Proved	sieve_equidist_implies_decorrelation
SieveEquidist ⇒ PED	Proved	sieve_equidist_implies_ped
Large sieve route		
MultiModularCSB ⇒ MC	Proved	mmcsb_implies_mc
BV chain ⇒ MC	Proved	bv_chain_mc
ArithLS chain ⇒ MC	Proved	arith_ls_chain_mc
ALS chain ⇒ MC	Proved	als_prime_arith_ls_chain_mc
WeakALS (§58)	**Proved**	weak_als_from_card_bound
Gauss sum inversion (§57)	Proved	char_sum_to_exp_sum
ALS ⇒ PrimeArithLS (§65)	**Proved**	als_implies_prime_arith_ls
Jordan's inequality (§56)	Proved	sin_pi_ge_two_mul
Geometric sum bound (§56)	Proved	norm_eAN_geom_sum_le_inv
Parseval for ZMod.dft (§53)	Proved	zmod_dft_parseval
Gauss sum norm $\|\tau\|^2 = p$ (§54)	Proved	gaussSum_norm_sq_eq_prime
Walk autocorrelation identities (§53)	Proved	walkAutocorrelation_*
Character Parseval (§60)	Proved	char_parseval_units
All 8 GCT internal lemmas (§56–§62)	Proved	gct_nontrivial_char_sum_le
Open hypotheses — live targets		
SingleHitHypothesis (weakest sufficient)	**Open**	SingleHitHypothesis
DynamicalHitting	**Open**	DynamicalHitting

continued on next page

Result	Status	Lean identifier
ComplexCharSumBound	**Open**	ComplexCharSumBound
MultiModularCSB	**Open**	MultiModularCSB
DecorrelationHypothesis	**Open**	DecorrelationHypothesis
PositiveEscapeDensity	**Open**	PositiveEscapeDensity
PEDImpliesComplexCSB (sole sieve-route gap)	**Open**	PEDImpliesComplexCSB
NoLongRuns(L)	**Open**	NoLongRuns
SieveEquidistribution	**Open**	SieveEquidistribution
MertensEscape	**Open**	MertensEscape
SieveAmplification	**Open**	SieveAmplification
SieveTransfer (genuine frontier)	**Open**	SieveTransfer
StrongSieveEquidist	**Open**	StrongSieveEquidist
DistributionalPED	**Open**	DistributionalPED
BVImpliesMMCSB (genuine frontier)	**Open**	BVImpliesMMCSB
GaussConductorTransfer (all lemmas proved)	**Open**	GaussConductorTransfer
PrimeArithLSImpliesMMCSB	**Open**	PrimeArithLSImpliesMMCSB
Known theorems — not yet in Mathlib		
PrimeDensityEquipartition (PNT in APs)	Known	PrimeDensityEquipartition
GenericLPFEquidist (Alladi [9])	Known	GenericLPFEquidist
BombieriVinogradov	Known	BombieriVinogradov
AnalyticLargeSieve	Known	AnalyticLargeSieve
ArithmeticLargeSieve	Known	ArithmeticLargeSieve
Dead ends — false or blocked		
MultCSBImpliesMMCSB (false in general, §5.3)	**Dead**	MultCSBImpliesMMCSB
BlockRotationEstimate (impossible for $d \geq 3$, Remark 5.7)	**Dead**	BlockRotationEstimate

A History and Computational Status

Mullin posed the conjecture in 1963 [1]. In over sixty years, no proof has been found, and no theoretical approach has come close. The problem sits in an unusual position: it is elementary to state, each individual step is deterministic, yet the global behavior of the sequence appears completely intractable.

The largest-factor variant. Cox and van der Poorten [11] showed that the related sequence using the *largest* prime factor—where each term is $\mathrm{gpf}(\mathrm{P}(n)+1)$ instead of minFac—provably misses infinitely many primes (for instance, 5 never appears). By always jumping to the largest factor of $\mathrm{P}(n)+1$, the sequence leaps past small primes and can never return to them, since each Euclid number is coprime to every earlier term. This refuted the natural strengthening that surjectivity holds regardless of the factor-selection rule, and showed that the minFac rule is essential to the conjecture.

The two selection rules are also structurally different. The largest prime factor $P^+(n) = \mathrm{gpf}(n)$ has rich algebraic structure: the sets $\{n : P^+(n) \le y\}$ (smooth numbers) are controlled by the Dickman function, and for polynomial sequences $f(n) = \mathrm{P}(n)+1$ one can sometimes exploit algebraic curves and Hasse–Weil bounds. In contrast, the smallest prime factor $P^-(n) = \mathrm{minFac}(n)$ is purely a *sieve* object: controlling $P^-(n)$ requires excluding small prime divisors one at a time (CRT + Mertens), and no algebraic-geometric handle exists. This gpf/minFac asymmetry is one reason the largest-factor variant admits a short proof while the smallest-factor conjecture remains open.

Variants and surveys. Booker [3] showed that a carefully chosen variant of the Euclid–Mullin sequence *does* contain every prime: by selecting a specific (not necessarily smallest) prime factor at each step, one can steer the sequence to hit every prime. This demonstrates that the conjecture is *delicate*: the surjectivity depends on the precise rule, not just the Euclidean structure. Pollack and Treviño [5] surveyed the problem's place in the broader landscape of Euclid-inspired sequences, and studied distributional properties of primes "forgotten" by Euclid-type constructions.

Computational status. The sequence has been extended through a series of large-scale factoring efforts:

- Wagstaff (1993) computed through the 43rd term.
- In 2010, the 180-digit number $\mathrm{P}(43)+1$ was factored via GNFS (General Number Field Sieve), yielding a 68-digit prime as $a(44)$. Terms $a(45)$–$a(47)$ followed.
- In 2012, Propper factored the 256-digit number $\mathrm{P}(47)+1$ by ECM (Elliptic Curve Method), discovering a 75-digit factor and extending the sequence to 51 terms (Booker–Irvine [2]).
- Finding $a(52)$ requires factoring a 335-digit number. No factorization is known as of 2025.

After 51 terms, the smallest primes not yet observed are 41 and 47. Note that 31—a smaller prime—does not appear until position 50.

Why computation cannot resolve the conjecture. Even heroic computation is fundamentally unable to address the conjecture. Each new term requires factoring a number whose digit count grows roughly linearly with the number of terms, quickly exceeding the reach of any known factoring algorithm. But even if we could compute millions of terms, this would prove nothing: the conjecture is a $\forall$-statement over all primes, and no finite computation can rule out the possibility that some prime first appears at an astronomically large index.

More fundamentally, the sequence exhibits a sensitive dependence on its full history. Each term $a(n+1) = \mathrm{minFac}(\mathrm{P}(n)+1)$ depends on the *complete factorization* of a number that encodes all previous terms. Changing a single early term alters every subsequent one. This global coupling is what makes the sequence appear random despite being deterministic, and it means that local or statistical reasoning about "typical" behavior is unreliable. There are no known density arguments, probabilistic heuristics, or sieve-theoretic bounds that bear on the conjecture. The problem requires a structural argument about the sequence's long-term dynamics—which is precisely what the formalization in the body of this paper provides.

B The Bag-Theoretic Structure of the Euclid–Mullin Dynamics

The EM sequence as a dynamical system on bags of primes. The state of the Euclid–Mullin construction at step n is fully captured by the *bag* (finite set) of primes collected so far: $S_n = \{\mathrm{em}(0), \mathrm{em}(1), \ldots, \mathrm{em}(n)\}$. The dynamics is a deterministic map on finite subsets of primes:

$$S_{n+1} = S_n \cup \{\min \mathcal{F}(S_n)\},$$

where $\mathcal{F}(S) = \mathrm{PrimeFactors}(\prod S + 1)$ produces the *factor bag* of the Euclid number $E_n = \mathrm{P}(n) + 1$.

The trajectory is entirely determined by $S_0 = \{2\}$. The Markov property holds: the future depends only on the current bag, not on the order in which its elements were assembled. The product $\mathrm{P}(n) = \prod S_n$ is a *lossless encoding* of the bag (by unique factorization of the squarefree integer $\mathrm{P}(n)$), and the order of insertion is forgotten.

The pipeline: from bag to next prime. The selection of the next prime passes through a pipeline of operations, each with distinct information-theoretic character:

$$S_n \xrightarrow[\text{lossless}]{\prod(\cdot)} \mathrm{P}(n) \xrightarrow[+1]{\text{shift}} \mathrm{P}(n)+1 \xrightarrow[\text{factor}]{} \mathcal{F}(S_n) \xrightarrow[\text{compress}]{\min} \mathrm{em}(n+1).$$

Stage 1: Product (lossless). The map $S_n \mapsto \mathrm{P}(n)$ is an injection (on sets of distinct primes), encoding $\sim n \log n$ bits of bag information into a single integer of $\sim 2^n$ digits. This stage *expands* the representation while preserving all information.

Stage 2: The $+1$ shift (multiplicative $\to$ additive). The map $\mathrm{P}(n) \mapsto \mathrm{P}(n)+1$ is the critical bridge between two worlds. The integer $\mathrm{P}(n)$ lives in the multiplicative world: its structure (as a product of known primes) is completely understood. The integer $\mathrm{P}(n)+1$ lives in the additive world: its factorization bears no algebraic relationship to that of $\mathrm{P}(n)$, beyond the guaranteed coprimality $\gcd(\mathrm{P}(n), \mathrm{P}(n)+1) = 1$. This shift is the *source of mixing* in the dynamics.

Stage 3: Factorization (revealing the new bag). The factorization $\mathrm{P}(n)+1 = p_1^{a_1} \cdots p_k^{a_k}$ reveals the factor bag $\mathcal{F}(S_n) = \{p_1, \ldots, p_k\}$. This step is formally lossless, but computationally opaque: predicting $\mathcal{F}(S_n)$ from S_n requires factoring a number of $\sim 2^n$ digits.

Stage 4: Minimum (catastrophic compression). The map $\mathcal{F}(S_n) \mapsto \min \mathcal{F}(S_n)$ discards all but the smallest element of the factor bag. The compression ratio is exponential: the input has $\sim 2^n$ bits; the output has $O(n)$ bits. The composition of stages 2–4 is the *minFac bottleneck*: an information-destroying channel that maps $\sim 2^n$ bits to $O(n)$ bits through arithmetically chaotic operations.

The orthogonal bag property. A fundamental structural fact distinguishes the factor bag $\mathcal{F}(S_n)$ from the current bag S_n:

Proposition B.1 (Orthogonal bags)**.** *$\mathcal{F}(S_n) \cap S_n = \emptyset$. That is, the primes dividing $\mathrm{P}(n)+1$ are entirely disjoint from the primes in the current bag.*

Proof. For any $p \in S_n$: $p \mid \mathrm{P}(n)$, hence $\mathrm{P}(n)+1 \equiv 1 \pmod{p}$, so $p \nmid \mathrm{P}(n)+1$, i.e. $p \notin \mathcal{F}(S_n)$. □

This is the Euclidean coprimality at the heart of the construction, but its bag-theoretic formulation reveals its dynamical significance: at each step, the factorization produces a *fresh sample* from the complementary set $\mathbb{P} \setminus S_n$. The new bag $\mathcal{F}(S_n)$ contains no "recycled" primes—every element is drawn from the universe of primes not yet seen.

Moreover, $\mathrm{P}(n)$ is *squarefree* (a product of distinct primes), so S_n is genuinely an unordered set with no multiplicities. The accumulator $\mathrm{P}(n)$ is the unique squarefree integer with prime support exactly S_n.

The CRT perspective. The Chinese Remainder Theorem provides the natural coordinate system for the bag dynamics. The integer $\mathrm{P}(n)$ is determined by its CRT vector:

$$\mathbf{c}(n) = (\mathrm{P}(n) \bmod 2,\ \mathrm{P}(n) \bmod 3,\ \mathrm{P}(n) \bmod 5,\ \ldots)$$

indexed by all primes. For primes $p \in S_n$: $\mathrm{P}(n) \bmod p = 0$ (since $p \mid \mathrm{P}(n)$). For primes $r \notin S_n$: $\mathrm{P}(n) \bmod r = \prod_{p \in S_n} (p \bmod r) \in (\mathbb{Z}/r\mathbb{Z})^\times$.

The $+1$ shift maps $\mathbf{c}(n) \mapsto \mathbf{c}(n) + \mathbf{1}$ componentwise. The factorization of $\mathrm{P}(n) + 1$ is determined by which components satisfy $c_r(n) + 1 \equiv 0 \pmod r$, i.e., $c_r(n) = -1$. The minFac is the *winner of a race* across CRT coordinates: the smallest prime $r \notin S_n$ for which $c_r(n) = -1$. This race depends on all coordinates simultaneously, but each coordinate is determined independently (by CRT) from every other.

Fiber structure and non-reconstructibility. A natural question is whether the output of minFac carries information about the bag that produced it. The answer is: almost none.

Proposition B.2 (Massive many-to-one)**.** *For any prime q and any set size k, the number of bags T of k distinct primes satisfying* $\mathrm{minFac}(\prod T + 1) = q$ *grows exponentially in k.*

The constraints on such a bag T are purely modular: $\prod T \equiv -1 \pmod q$ (one congruence condition), plus $\prod T \not\equiv -1 \pmod p$ for each prime $p < q$ with $p \notin T$ (at most $\pi(q)$ non-congruence conditions). The total information content of the output is $O(q)$ bits, while the bag carries $\sim 2^k$ bits. The fraction of information preserved is $O(q/2^k) \to 0$ super-exponentially. Consequently, the map $S \mapsto \mathrm{minFac}(\prod S + 1)$ is an *exponentially lossy summarization* of the bag.

MinFac as a greedy algorithm. The minimum operation in Stage 4 admits a natural optimization interpretation. Among all prime factors of $\mathrm{P}(n) + 1$, the minFac rule selects the one that increases $\log \mathrm{P}$ by the least amount:

$$\mathrm{em}(n+1) = \arg \min_{p \in \mathcal{F}(S_n)} \log p.$$

This is a *greedy algorithm for slow growth*: at each step, the accumulator grows as slowly as possible (given the constraint of selecting a prime factor of $\mathrm{P}(n) + 1$).

Slow growth has a dynamical consequence. Since $\mathrm{P}(n)$ grows slowly (relative to the maxFac alternative), each target prime q receives *more trials*—more steps during which $\mathrm{P}(n) + 1$ might be divisible by q—before the accumulator becomes so large that the probability of q-divisibility becomes negligible. The minFac rule maximizes the number of opportunities each prime has to enter the sequence.

By contrast, the maxFac rule (selecting the largest prime factor) causes $\mathrm{P}(n)$ to grow so rapidly that small primes quickly lose any chance of being selected, which is why the second Euclid–Mullin sequence provably misses infinitely many primes (Cox–van der Poorten [11], Booker [3]).

The 0-accessibility of minFac. We formalize the advantage of minFac over other selection rules. Define a selection rule $\sigma \colon 2^{\mathbb{P}} \to \mathbb{P}$ mapping a nonempty set of primes to one of its elements. Say σ is 0-*accessible* for a prime q past the sieve gap if: whenever $q \in F$ and all primes $p < q$ have been screened, $\sigma(F) = q$.

MinFac is 0-accessible for every q: past the sieve gap, all primes below q are excluded from $\mathcal{F}(S_n)$, so if $q \in \mathcal{F}(S_n)$, then $q = \min \mathcal{F}(S_n)$. The event "$q$ divides $\mathrm{P}(n) + 1$" is *sufficient* for q to enter the sequence.

MaxFac is not 0-accessible: even if $q \mid \mathrm{P}(n)+1$, a larger factor is selected instead, and the probability that q is the largest factor of a number of size $\sim 2^{2^n}$ decays faster than any polynomial.

The 0-accessibility of minFac ensures that the walk mod q hitting -1 is both necessary and sufficient for q to enter the sequence (past the sieve gap). This reduces the question of whether q appears to a question about the walk alone, with no additional selection barrier.

Deterministic pseudorandomness. The EM sequence is fully deterministic, yet exhibits pseudorandom behavior. The pseudorandomness arises from the composition of two sources of arithmetic opacity:

1. *The $+1$ shift* moves from the multiplicative world (where $\mathrm{P}(n)$ is fully understood) to the additive world (where $\mathrm{P}(n)+1$ is arithmetically opaque). The relationship between the multiplicative structures of consecutive integers is the central mystery of analytic number theory.
2. *The factoring step* reveals the prime structure of $\mathrm{P}(n)+1$, but this structure is computationally hard to predict from $\mathrm{P}(n)$. The computational difficulty of factoring is, in a precise sense, the *source of pseudorandomness* for the EM sequence.

Small changes to the bag (adding one prime p) transform $\mathrm{P}(n)$ to $\mathrm{P}(n)\cdot p$ and hence $\mathrm{P}(n)+1$ to $\mathrm{P}(n)\cdot p+1$, a completely different integer with a completely different factorization. The dynamics is *sensitive to the bag contents*, analogous to sensitive dependence on initial conditions in chaotic systems, but operating through number-theoretic rather than geometric mechanisms.

The orthogonal bag property reinforces this pseudorandomness: at each step, the factor bag is drawn from $\mathbb{P} \setminus S_n$, a set that has never been "used" in the construction. Each step provides a genuinely fresh arithmetic input, preventing the buildup of systematic correlations.

C Analogies and Context

Mullin's Conjecture has no known applications: if proved tomorrow, no other theorem in number theory would follow from it. The value of the problem lies instead in what it *is an instance of*—and in the methods its resolution would require. The orbit-specificity barrier identified by this formalization appears, in recognizable form, across several active areas of mathematics.

Artin's conjecture and the orbit-specificity gap. The closest structural analogue to MC is Artin's conjecture on primitive roots [16]: for any integer $a \neq -1$ that is not a perfect square, the group $(\mathbb{Z}/p\mathbb{Z})^\times$ is generated by a for infinitely many primes p. The parallel is precise:

- **Artin** asks whether the orbit of a fixed generator a under repeated multiplication fills $(\mathbb{Z}/p\mathbb{Z})^\times$.
- **Mullin** asks whether the orbit of 2 under multiplication by successive EM primes in $(\mathbb{Z}/q\mathbb{Z})^\times$ hits the single element -1.

Both conjectures are blocked by the same fundamental obstacle: transferring *averaged* equidistribution results (which hold for most moduli or most generators) to *one specific deterministic orbit*. Hooley [7] proved Artin's conjecture conditional on GRH for Dedekind zeta functions of Kummer extensions—because GRH provides the uniformity across individual characters needed to control a single orbit. In our setting, this uniformity is exactly what CCSB demands.

The analogy is not merely structural. The Kummer extensions $\mathbb{Q}(\zeta_\ell, a^{1/\ell})$ that appear in Hooley's proof are the same extensions that arise in the EffectiveKummerEscape approach to SubgroupEscape (Section 7). The elementary PRE lemma (Theorem 3.1) sidesteps this Chebotarev machinery entirely for the algebraic component, but the dynamical component—does the walk *hit* -1, not merely *generate* the full group?—remains exactly the orbit-specificity gap that GRH closes for Artin and that no known tool closes for Mullin.

Multiplicative walks on finite groups. The walk reformulation (Section 2) places MC in the framework of random walks on finite groups, studied systematically by Diaconis and others since the 1980s [14]. The Diaconis–Shahshahani upper bound lemma [13] shows that a random walk on a finite group G driven by i.i.d. multipliers from a conjugation-invariant distribution mixes in $O(\log |G|)$ steps, with mixing measured by character sums.

The EM walk has the same algebraic structure—a multiplicative walk on the cyclic group $(\mathbb{Z}/q\mathbb{Z})^\times$—but violates every assumption of the classical mixing theory. The multipliers are deterministic, not random; they are not identically distributed; and most critically, the multiplier at step n depends on the walk position at step n, creating exactly the position-multiplier correlation that the Diaconis–Shahshahani framework assumes away. CCSB is the precise derandomization of the mixing-time bound: it asks that the Fourier coefficients of the walk's occupation measure tend to zero for every non-trivial character.

Smallest prime factor distribution. The SieveTransfer hypothesis (Section 7) connects MC to the distribution of the smallest prime factor function $P^-(n) = \min\{p : p \mid n\}$, studied by Alladi [9], Hildebrand [10], and others. For generic integers in an arithmetic progression $n \equiv a \pmod q$, the distribution of $P^-(n)$ is controlled by the Dickman function and CRT-based equidistribution results. MC asks whether this equidistribution transfers from generic integers to the specific subsequence $\{\mathrm{P}(n)+1\}_{n\geq 0}$. The same orbit-specificity transfer problem arises for Mersenne numbers $2^p - 1$, Fibonacci numbers, and polynomial iterates $f^{\circ n}(a)$ in arithmetic dynamics.

The marginal/joint barrier and Sarnak's conjecture. The formalization identifies a precise meta-obstacle (Section 5): *marginal* equidistribution of the multiplier residues is provable (the EM primes are equidistributed in residue classes, by Dirichlet's theorem); what MC requires is *joint* equidistribution of the pair (walk position, multiplier), conditioned on the walk's history. This barrier is an instance of a broader phenomenon. Sarnak's conjecture [15] asserts that the Möbius function $\mu(n)$ is orthogonal to every bounded deterministic sequence: $\frac{1}{N}\sum_{n\leq N} \mu(n)\, a_n \to 0$. CCSB is a Möbius-orthogonality-type statement for the EM sequence, placing MC squarely within the Sarnak program's conceptual framework, even though the EM sequence falls outside the technical scope of existing results (which require zero topological entropy).

Greedy sieves and orbit-hitting. MC is the simplest nontrivial instance of a broader question: does a greedy, deterministic prime-selection process eventually cover all primes? The Cox–van der Poorten result [11] shows that this determinism is fragile: choosing the *largest* factor instead provably misses primes. More generally, MC belongs to the family of *orbit-hitting problems* in arithmetic dynamics: given a map T on a space X and a target set $S \subset X$, does the orbit eventually enter S? Unlike Artin (where the map $x \mapsto ax$ is the same at every step) or Collatz (where the map depends only on the current state), the EM map varies at each step, determined by the factorization of a number that depends on the entire orbit history. This is the accumulator coupling of Section 1: the running product $\mathrm{P}(n)$ is a cumulative digest of the full orbit, and the walk–multiplier framework makes it precise. The formalization shows it is the sole source of difficulty: once the coupling is controlled (via CCSB or CME), MC follows by machine-checked deduction.

D Additional Sieve and Spectral Routes

This appendix collects the sieve and spectral-energy routes to MC that complement the three principal reductions (DH, CCSB, BV) presented in the body. All reduction arrows are machine-verified; the sole open content in each route is the orbit-specificity transfer.

Arithmetic Large Sieve Route

Theorem D.1 (✓ arith_ls_chain_mc)**.** ArithLS + ArithLSImpliesMMCSB $\implies$ MC.

The arithmetic large sieve gives character sum bounds for Dirichlet characters (a known result, not in Mathlib). The transfer ArithLSImpliesMMCSB is open and is in fact a **dead end**: universal coefficient bounds cannot distinguish equidistributed walks from clumped walks.

Analytic Large Sieve Route

The most developed route connects the analytic large sieve to MC via Gauss sum inversion.

Definition D.2 (AnalyticLargeSieve (**ALS**))**.** For well-separated points $\{\alpha_r\} \subset \mathbb{R}/\mathbb{Z}$ with $\min_{r \neq s} \|\alpha_r - \alpha_s\| \geq \delta$:

$$\sum_r \left\| \sum_{n<N} a_n\, e(n\alpha_r) \right\|^2 \;\leq\; (N - 1 + \delta^{-1}) \sum_{n<N} \|a_n\|^2.$$

Theorem D.3 (✓ weak_als_from_card_bound)**.** *A* weak *version with constant* $N \cdot (\delta^{-1}+1)$ *is proved (the optimal constant is* $N - 1 + \delta^{-1}$*, but the difference is immaterial since MMCSB requires only* $o(N)$*).*

The key bridge is Gauss sum inversion: a Gauss sum $\tau(\chi) = \sum_a \chi(a)\, e(a/p)$ intertwines multiplication and addition on $\mathbb{Z}/p\mathbb{Z}$, converting character sums to exponential sums.

Theorem D.4 (✓ char_sum_to_exp_sum)**.** *For a non-trivial character* χ *mod* p *prime:* $\sum_n f(n)\, \chi(n) = \tau^{-1} \sum_{b=1}^{p-1} \chi^{-1}(b) \sum_n f(n)\, \psi(bn)$.

The GaussConductorTransfer composes eight internal lemmas (all proved, §56–§62) into the bridge from ALS to the prime arithmetic large sieve:

Theorem D.5 (✓ als_implies_prime_arith_ls)**.** AnalyticLargeSieve $\implies$ PrimeArithLS.

Theorem D.6 (✓ als_prime_arith_ls_chain_mc)**.** ALS + PrimeArithLSImpliesMMCSB $\implies$ MC.

The remaining open content is PrimeArithLSImpliesMMCSB: transferring prime-modulus arithmetic large sieve bounds to multi-modular character sum bounds for the specific EM orbit.

The Spectral Energy Route

Instead of individual character sums, this route examines the *total energy* of the walk occupation measure $V_N(a) = |\{n < N : \mathrm{P}_q(n) = a\}|$.

Theorem D.7 (✓ walk_energy_parseval)**.** $\sum_\chi \|S_\chi(N)\|^2 = (q-1) \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^\times} V_N(a)^2$ *(Parseval).*

The *excess energy* $E(N) = \sum_{\chi \neq 1} \|S_\chi(N)\|^2$. If $E(N) = o(N^2)$, then every non-trivial character sum is individually $o(N)$, which is CCSB.

Definition D.8 (SubquadraticVisitEnergy (**SVE**))**.** For every missing prime q and $\varepsilon > 0$, there exists N_0 such that for $N \geq N_0$: $E(N) \leq \varepsilon N^2$.

Theorem D.9 (✓ sve_implies_mmcsb)**.** SVE $\implies$ MMCSB $\implies$ MC.

Van der Corput and higher-order decorrelation. The van der Corput inequality bounds $|\sum z_n|$ via autocorrelations:

Theorem D.10 (✓ van_der_corput_bound)**.** $\left\|\sum_{n\le N} z_n\right\|^2 \le \frac{N+H}{H+1}\left(N+2\sum_{h=1}^{H}|\mathrm{Re}\sum_{n\le N-h} z_n\overline{z_{n+h}}|\right)$.

For the EM walk, lag-h autocorrelations involve h-step multiplier products. At $h=1$, the autocorrelation equals the multiplier character sum (Theorem 4.18), so VdC with a single shift gives only $O(N)$. Higher lags may decorrelate:

Definition D.11 (HigherOrderDecorrelation (**HOD**))**.** For every missing prime q, non-trivial χ, and $\varepsilon > 0$: there exists H_0 such that for $H \ge H_0$, N_0 such that for $N \ge N_0$ and all $1 \le h \le H$: $\|R_h(N)\| \le \varepsilon N$.

Theorem D.12 (✓ hod_implies_ccsb)**.** HOD $\Longrightarrow$ CCSB $\Longrightarrow$ MC.

Conditional multiplier equidistribution.

Definition D.13 (ConditionalMultiplierEquidist (**CME**))**.** For every missing prime q, non-trivial χ, $\varepsilon > 0$, N_0 such that for $N \ge N_0$ and every $c \in (\mathbb{Z}/q\mathbb{Z})^\times$: $\|\sum_{\substack{n<N\\ w(n)=c}} \chi(m(n))\| \le \varepsilon N$.

Theorem D.14 (✓ cme_implies_dec)**.** CME $\Longrightarrow$ DecorrelationHypothesis.

Theorem D.15 (✓ cme_implies_ccsb)**.** CME $\Longrightarrow$ CCSB.

Proof sketch. The walk telescoping identity gives $\sum\chi(w(n)) = \sum\chi(w(n))\chi(m(n)) - (\chi(w(N)) - \chi(w(0)))$. The product sum decomposes by fiber: $\sum\chi(w(n))\chi(m(n)) = \sum_a \chi(a)\cdot\sum_{w(n)=a}\chi(m(n))$. CME bounds each fiber sum by $\varepsilon' N$; the triangle inequality sums over at most $|(\mathbb{Z}/q\mathbb{Z})^\times|$ fibers; the boundary term $\chi(w(N)) - \chi(w(0))$ has norm ≤ 2 and is absorbed for large N. □

This is the key reduction that bypasses PED, BRE, and the $d \ge 3$ barrier. The proof works for all character orders because it uses only the fiber decomposition and telescoping—no block rotation estimate is needed.

Theorem D.16 (✓ cme_implies_mc)**.** CME $\Longrightarrow$ MC.

Proof. Compose `cme_implies_ccsb` with `complex_csb_mc'`. □

Theorem D.17 (✓ cme_chain_mc)**.** CME + PEDImpliesCSB $\Longrightarrow$ MC.

This older route through the Dec $\to$ PED $\to$ CCSB chain is superseded by the direct CME $\to$ CCSB reduction above, which requires no additional hypotheses.

The Complete Hypothesis Hierarchy

$$\text{PED} \;<\; \text{Dec} \;<\; \text{CME} \xrightarrow{\text{proved}} \text{CCSB} \;\approx\; \text{HOD} \;\approx\; \text{SVE},$$

where "$<$" means strictly weaker (proved implication, known not to reverse) and "$\approx$" means equivalent. HOD $\Leftrightarrow$ CCSB via van der Corput; SVE $\Leftrightarrow$ CCSB via Parseval; CME $\Rightarrow$ CCSB via telescoping + fiber decomposition (Theorem D.15).

Every hypothesis implies MC. The PED route has an open BRE bridge for $d \ge 3$ characters, but this is now bypassed: the direct CME $\to$ CCSB arrow is proved for all character orders. CME is the *sharpest sufficient condition*—the weakest hypothesis known to imply MC.

E Methodology: Human–AI Collaboration

This work was produced through a sustained collaboration between a human author and an AI system (Claude, Anthropic) across 72+ sessions. The human author directed the mathematical strategy—proof architecture, dead-end identification, and editorial control—while the AI system handled Lean 4 formalization, Mathlib API search, literature scouting, and exploration of candidate proof strategies.

The interaction was organized at scale via an *agent swarm*: a multi-agent system built on the Claude Agent SDK. The swarm comprises seven specialized agents, each with its own system prompt, tool access, and model:

- A *coordinator* that reads the current proof state, selects the most promising action, dispatches specialists, and updates shared state files.
- A *formalizer* that writes and compiles Lean code in rapid iteration cycles.
- A *literature scout* that searches papers and Mathlib for relevant results.
- Four *attack vector specialists* focused on analytic, algebraic, combinatorial, and information-theoretic approaches.
- A *paper writer* that maintains this document.

Agent prompts are *self-evolving*: after each session the coordinator updates them to record dead ends, new Mathlib discoveries, and shifted priorities. This prevents agents from rediscovering settled territory. All agent state (progress, strategy log, findings) is stored as git-tracked markdown, making the exploration history fully reproducible.

The division of labor between human and AI was sharp:

- **Human:** mathematical direction, proof strategy, identification of dead ends, evaluation of intermediate results, architectural decisions on the reduction hierarchy, and editorial control over the final formalization and paper.
- **AI (Claude):** Lean 4 formalization using Mathlib, Mathlib API search, literature scouting, exploration of candidate proof strategies, and drafting of this paper.

The human author guided the proof effort across 72+ sessions, suggesting attack vectors (algebraic, analytic, combinatorial, sieve-theoretic), identifying when an approach had reached a dead end, and pushing toward the sharpest possible reductions. The AI agents wrote all Lean code, searched Mathlib for relevant lemmas, explored dozens of proof strategies to completion or refutation, and maintained the evolving paper.

The swarm is optimized for formalization and reduction, not mathematical discovery. The next breakthrough, if it comes, will probably be a human insight about the structure of minFac on EM products—not something an agent finds by systematic search.

F Glossary of Definitions and Hypotheses

The table below collects every named definition, hypothesis, and key theorem introduced in this paper, with abbreviations and the section where each is defined.

Abbr.	Name	Meaning	Ref.
Core sequence and walk			
—	Walk / Multiplier	$\mathrm{P}_q(n) = \mathrm{P}(n) \bmod q$; $\mathrm{m}_q(n) = \mathrm{em}(n{+}1) \bmod q$	Def. 2.2
MC	MullinConjecture	Every prime appears in the Euclid–Mullin sequence	Conj. 1.1
Algebraic hypotheses (§2–§3)			
SE	SubgroupEscape	No proper subgroup of $(\mathbb{Z}/q\mathbb{Z})^\times$ contains all multipliers	Def. 2.13

continued on next page

Abbr.	Name	Meaning	Ref.
HH	HittingHypothesis	The walk reaches -1 cofinally: $\forall N, \exists n \geq N,\ q \mid \mathrm{P}(n)+1$	Def. 3.6
DH	DynamicalHitting	$\mathrm{SE}(q) \Rightarrow \mathrm{HH}(q)$ for every missing prime q	Def. 3.6
SHH	SingleHitHypothesis	$\mathrm{MC}(<q) + \mathrm{SE}(q) + q$ missing $\Rightarrow \exists\, n \geq N_0$ with $q \mid \mathrm{P}(n)+1$	Def. 3.4
PRE	PrimeResidueEscape	Every proper subgroup of $(\mathbb{Z}/p\mathbb{Z})^\times$ is escaped by some odd prime $< p$	Thm. 3.1
PRE$_\ell$	PowerResidueEscape	Multipliers escape the index-ℓ subgroup of $(\mathbb{Z}/q\mathbb{Z})^\times$	§3
—	ThresholdHitting	DH restricted to primes $q \geq B$	§4.6
Character-analytic hypotheses (§4)			
CCSB	ComplexCharSumBound	Walk char sums $S_\chi(N) = o(N)$ for all non-trivial χ	§4
MMCSB	MultiModularCSB	CCSB simultaneously for all primes q in a range	§4
ALS	AnalyticLargeSieve	Large sieve inequality adapted to EM walk	§4
PED	PositiveEscapeDensity	Positive density of n with $\mathrm{m}_q(n) \notin H$, for every proper H	§4
—	DecorrelationHypothesis	$\mathrm{m}_q(n)$ and $\mathrm{m}_q(n{+}1)$ are asymptotically independent	§4
BRE	BlockRotationEstimate	Cancellation in block sums of characters applied to walk	§4
SVE	SubquadraticVisitEnergy	$\sum_a \lvert\{n \leq N : \mathrm{P}_q(n) = a\}\rvert^2 = o(N^2/(q{-}1))$	§4
HOD	HigherOrderDecorrelation	Higher-order correlation bounds for walk increments	§4
CME	ConditionalMultiplierEquidist	Conditional equidist. of multipliers given walk state; implies CCSB (proved)	§4
Named theorems			
—	Confinement	If SE fails, the walk is confined to a proper coset	Thm. 2.12
—	Walk–Divisibility Bridge	$\mathrm{P}_q(n) = -1 \Leftrightarrow q \mid \mathrm{P}(n)+1$	Thm. 2.4
—	One-prime gap	$\mathrm{MC}(<q)$ + cofinal hit $\Rightarrow \mathrm{MC}(q)$	Thm. 3.14
—	QR Obstruction	SE failure has density $O(2^{-k})$ by CRT + quadratic reciprocity	§3
—	Gauss sum inversion	Character sums $\leftrightarrow$ exponential sums via Gauss sums	§4

References

[1] A. A. Mullin. Recursive function theory (a modern look at a Euclidean idea). *Bull. Amer. Math. Soc.*, 69:737, 1963.

[2] A. R. Booker and S. A. Irvine. The Euclid–Mullin graph. *J. Number Theory*, 165:30–57, 2016.

[3] A. R. Booker. A variant of the Euclid–Mullin sequence containing every prime. *J. Integer Sequences*, 19:Article 16.6.4, 2016.

[4] B. Gordon. Sequences in groups with distinct partial products. *Pacific J. Math.*, 11(4):1309–1313, 1961.

[5] P. Pollack and E. Treviño. The primes that Euclid forgot. *Amer. Math. Monthly*, 121(5):433–437, 2014.

[6] M. Hardy and C. Woodgold. Prime simplicity. *Math. Intelligencer*, 31:44–52, 2009.

[7] C. Hooley. On Artin's conjecture. *J. Reine Angew. Math.*, 225:209–220, 1967.

[8] J. C. Lagarias and A. M. Odlyzko. Effective versions of the Chebotarev density theorem. In *Algebraic Number Fields (Durham Symposium)*, pages 409–464. Academic Press, 1977.

[9] K. Alladi. On the distribution of the largest prime factor. *Stud. Sci. Math. Hungar.*, 12:1–9, 1977.

[10] A. Hildebrand. On the number of positive integers $\leq x$ and free of prime factors $> y$. *J. Number Theory*, 22(3):289–307, 1986.

[11] C. D. Cox and A. J. van der Poorten. On a sequence of prime numbers. *J. Austral. Math. Soc.*, 8:571–574, 1968.

[12] H. Iwaniec and E. Kowalski. *Analytic Number Theory*. Amer. Math. Soc. Colloq. Publ., vol. 53, 2004.

[13] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57:159–179, 1981.

[14] F. R. K. Chung, P. Diaconis, and R. L. Graham. Random walks arising in random number generation. *Ann. Probab.*, 15(3):1148–1165, 1987.

[15] P. Sarnak. Three lectures on the Möbius function, randomness, and dynamics. Lecture notes, IAS, 2010. Available at `https://publications.ias.edu/sarnak/paper/512`.

[16] E. Artin. Beweis des allgemeinen Reziprozitätsgesetzes. *Abh. Math. Semin. Univ. Hambg.*, 5:353–363, 1927.

[17] R. W. K. Odoni. On the prime divisors of the sequence $w_{n+1} = 1 + w_1 \cdots w_n$. *J. London Math. Soc. (2)*, 32(1):1–11, 1985.