The Irreducible Core of Mullin's Conjecture: A Machine-Verified Reduction

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 (∼22,400 lines, **zero sorry**) that reduces MC to a single open hypothesis.

An *inductive bootstrap* eliminates the algebraic half: we prove that for any prime $p \geq 5$, some odd prime $r < p$ escapes every proper subgroup of $(\mathbb{Z}/p\mathbb{Z})^\times$, using only modular arithmetic. A *Fourier bridge* handles the analytic half: MC follows whenever certain walk character sums are $o(N)$. Together these yield **DynamicalHitting** (DH) $\Rightarrow$ MC: if a multiplicative walk whose multipliers generate $(\mathbb{Z}/q\mathbb{Z})^\times$ hits every element cofinally, MC holds.

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), proved to imply the Complex Character Sum Bound (CCSB) for all character orders, bypassing the $d \geq 3$ barrier. Twelve dead ends are documented.

The formalization comprises ∼770 theorems and ∼460 definitions across 32 files.

Contents

1 Introduction **4**
- 1.1 Mullin's Conjecture . . . 4
- 1.2 History and Computational Status . . . 4
- 1.3 Main Result and Approach . . . 5
- 1.4 Analogies and Context . . . 6

2 The Residue Walk Reformulation **8**
- 2.1 Definitions . . . 8
- 2.2 The Walk–Divisibility Bridge . . . 9
- 2.3 SubgroupEscape and the Confinement Theorem . . . 9

3 The Inductive Bootstrap **10**
- 3.1 PrimeResidueEscape . . . 11
- 3.2 The Bootstrap Mechanism . . . 12
- 3.3 The Threshold Approach . . . 13
- 3.4 The Sieve Gap . . . 13
- 3.5 The Power Residue Decomposition . . . 13
- 3.6 Quadratic Reciprocity Obstruction . . . 14
- 3.7 Concrete Verification . . . 14
- 3.8 The Reduction to DynamicalHitting . . . 14

4 The Character Sum Reduction **15**
- 4.1 Character Sums and Walk Equidistribution . . . 16
- 4.2 The Fourier Bridge . . . 16
- 4.3 The Decorrelation–PED–CCSB Chain . . . 17
- 4.4 Walk Telescoping Identities . . . 18
- 4.5 The Large Sieve Route . . . 19

5 Why It's Hard **20**
- 5.1 The Selectability Perspective . . . 20
- 5.2 The Marginal/Joint Barrier . . . 21
- 5.3 The BRE Impossibility for $d \geq 3$ . . . 22
- 5.4 The Van der Corput Barrier . . . 22
- 5.5 The Walk Bridge Falsity . . . 22
- 5.6 Dead Ends as a Roadmap . . . 22
- 5.7 The Mathematical Landscape . . . 23
- 5.8 Structural Features of the EM Walk . . . 24

6 The Lean Formalization **24**
- 6.1 Codebase Structure . . . 24
- 6.2 Axiom Usage: What's Constructive . . . 25
- 6.3 Mathlib Dependencies . . . 25
- 6.4 Verification Statistics . . . 26

7 Open Problems **26**
- 7.1 CCSB as the Precise Frontier . . . 26
- 7.2 Connection to Bombieri–Vinogradov . . . 26
- 7.3 Connection to Chebotarev . . . 27
- 7.4 The Sieve-Theoretic Approach . . . 27
- 7.5 What Would Close the Conjecture . . . 28
- 7.6 Is DynamicalHitting True? . . . 28

8 Summary of Verified Results **29**

A Additional Sieve and Spectral Routes **33**
A.1 Arithmetic Large Sieve Route 33
A.2 Analytic Large Sieve Route 33
A.3 The Spectral Energy Route 34
A.4 The Complete Hypothesis Hierarchy 35

B Methodology: Human–AI Collaboration **35**

C Glossary of Definitions and Hypotheses **36**

1 Introduction

1.1 Mullin's Conjecture

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 behaves almost randomly: 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, 51 terms are known. Remarkably, 41 and 47—the two smallest primes not yet observed—have not appeared even after 51 terms, while 31 does not show up until position 50. 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. Each step couples the next prime to the full factorization history, creating a recursive dependency that defeats both probabilistic heuristics and standard sieve methods.

The crux of the difficulty is a *selectability* problem. At each step, $\text{prod}(n)+1$ has many prime factors, and any of them could serve as the next term—but only the *smallest* is chosen. A target prime q may divide $\text{prod}(n)+1$ infinitely often (it is "selectable") yet never be selected if a smaller prime always divides $\text{prod}(n)+1$ as well. Our formalization makes this tension precise and shows that the inductive structure of MC eliminates this obstruction for the tail of the sequence.

1.2 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 $\text{gpf}(\text{prod}(n)+1)$ instead of minFac—provably misses infinitely many primes (for instance, 5 never appears). By always jumping to the largest factor of $\text{prod}(n)+1$, 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.

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 $\text{prod}(43)+1$ 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 $\text{prod}(47)+1$ 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 $a(n+1) = \text{minFac}(\text{prod}(n)+1)$ 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 our formalization provides.

1.3 Main Result and Approach

Our main result is a formally verified reduction of MC to a single hypothesis about the dynamics of a residue walk. 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 walk equidistribution—that the multipliers generate the full group—is *free*, following from the inductive hypothesis $\text{MC}(<p)$ via an elementary lemma (Section 3). This reduces MC to a single dynamical question.
3. **Diagnosis.** We develop the harmonic-analytic and sieve-theoretic infrastructure to determine *precisely* what kind of statement would close the conjecture, 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.

Theorem 1.2 (✓ dynamical_hitting_implies_mullin)**.** DynamicalHitting $\implies$ MC.

DynamicalHitting asserts: if the multiplier residues generate $(\mathbb{Z}/q\mathbb{Z})^\times$ (SubgroupEscape), then the walk hits -1 cofinally (HittingHypothesis). The proof is by strong induction on p, with PrimeResidueEscape (proved elementarily) bootstrapping SE at each step.

A parallel reduction gives a clean character-analytic statement:

Theorem 1.3 (✓ complex_csb_mc')**.** ComplexCharSumBound $\implies$ MC.

Both reductions are fully machine-verified with zero sorry. The formalization thus provides a complete "roadmap" for proving MC: any future proof need only establish one of these open hypotheses (DynamicalHitting, ComplexCharSumBound, or any of the equivalent formulations in §4 and §7), and the rest follows by machine-checked deduction.

1.4 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 3.4) 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 $\{\mathrm{prod}(n)+1\}_{n\geq 0}$. 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 self-referential coupling is what the walk–multiplier framework makes precise, and 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.

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 and establishes the algebraic prerequisites. Section 3 presents the inductive bootstrap—the core mathematical insight that makes SubgroupEscape free. Section 4 develops the character-analytic reduction (CCSB $\Rightarrow$ 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.

2 The Residue Walk Reformulation

The central idea is to replace a global number-theoretic question ("does prime q eventually appear?") with a local algebraic one ("does a certain walk on a finite group hit a specific element?"). We begin with the definitions, then explain why this reformulation is powerful.

2.1 Definitions

Let $\mathrm{minFac}(m)$ denote the smallest prime factor of a natural number $m \geq 2$. We define two sequences $\mathrm{seq}, \mathrm{prod} : \mathbb{N} \to \mathbb{N}$ by mutual recursion:

$$\mathrm{seq}(0) := 2, \qquad \mathrm{prod}(0) := 2, \tag{2}$$

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

Theorem 2.1 (✓ seq_isPrime, seq_injective)**.**

1. (Primality) *For every $n \in \mathbb{N}$, $\mathrm{seq}(n)$ is prime.*
2. (Injectivity) $\mathrm{seq}(i) = \mathrm{seq}(j)$ *implies* $i = j$*. No prime appears twice.*

Definition 2.2 (Walk and Multiplier)**.** Let q be a prime with $\mathrm{seq}(n) \neq q$ for all n (a *missing prime*).

$$\mathrm{walkZ}(q, n) := \mathrm{prod}(n) \bmod q \quad \in (\mathbb{Z}/q\mathbb{Z})^\times, \tag{4}$$

$$\mathrm{multZ}(q, n) := \mathrm{seq}(n{+}1) \bmod q \quad \in (\mathbb{Z}/q\mathbb{Z})^\times. \tag{5}$$

The restriction to missing primes is essential, not merely conventional. If $q = \mathrm{seq}(k)$ for some k, then $q \mid \mathrm{prod}(n)$ for every $n > k$, so $\mathrm{prod}(n) \bmod q = 0$, which is *not* a unit in $\mathbb{Z}/q\mathbb{Z}$; the walk and multiplier would not be well-defined. When q is missing, it never divides any sequence term (by injectivity) and never divides the running product (which is a product of primes $\neq q$), so both residues land in $(\mathbb{Z}/q\mathbb{Z})^\times$.

Mullin's Conjecture asserts that no prime is missing. The entire framework therefore works by reduction: *assuming* q is missing, we build a walk that is well-defined, and then show it must hit -1, which forces q to appear—a contradiction. Every quantifier "for every missing prime q" in the sequel should be read in this light.

Proposition 2.3 (Walk recurrence — ✓ walkZ_succ)**.** $\mathrm{walkZ}(q, n{+}1) = \mathrm{walkZ}(q, n) \cdot \mathrm{multZ}(q, n)$ *in $(\mathbb{Z}/q\mathbb{Z})^\times$. The walk is a multiplicative walk on the cyclic group $(\mathbb{Z}/q\mathbb{Z})^\times$: at each step, the position is multiplied by the next EM prime reduced mod q.*

Example 2.4 (The walk for $q = 41$)**.** The prime 41 has not appeared in the first 51 known terms of the sequence, so we may form its walk. The table below shows the first ten steps. At each step, $\mathrm{walkZ}(41, n)$ is updated by multiplying by $\mathrm{multZ}(41, n) = \mathrm{seq}(n{+}1) \bmod 41$:

n	$\mathrm{seq}(n)$	$\mathrm{seq}(n) \bmod 41$	$\mathrm{walkZ}(41, n)$	
0	2	2	2	
1	3	3	6	
2	7	7	1	
3	43	2	2	
4	13	13	26	
5	53	12	25	
6	5	5	2	
7	6221671	3	6	← pattern repeats
8	38709183810571	7	1	
9	139	16	16	

For 41 to appear in the sequence at step $n{+}1$, we need $41 \mid \mathrm{prod}(n)+1$, i.e. $\mathrm{walkZ}(41,n) = 40 \equiv -1 \pmod{41}$. Over all 51 known terms, the walk takes values

$$2,\ 6,\ 1,\ 2,\ 26,\ 25,\ 2,\ 6,\ 1,\ 16,\ 3,\ 33,\ 28,\ 12,\ 24,\ 7,\ 20,\ 2,\ 38,\ 7,\ \ldots$$

and *never reaches* 40. The group $(\mathbb{Z}/41\mathbb{Z})^\times$ has 40 elements; the walk wanders through them but has not yet found its target. Can we prove it eventually will? That is the content of Mullin's Conjecture for $q = 41$.

Why the reformulation is powerful. The walk–multiplier framework is not merely cosmetic: it makes the problem amenable to tools from group theory, harmonic analysis, and ergodic theory, and it cleanly separates two independent components:

- **Algebraic structure**: which subgroups of $(\mathbb{Z}/q\mathbb{Z})^\times$ do the multipliers $\mathrm{multZ}(q,n)$ generate? If a proper subgroup traps all multipliers, the walk is permanently confined and may never reach -1.
- **Dynamical content**: given that the multipliers generate the full group, does the walk actually visit every element? This is a question about the *ordering* of multipliers, not just their *set*.

The bootstrap (Section 3) will show that the algebraic component is free; the dynamical component is the sole remaining content of MC.

2.2 The Walk–Divisibility Bridge

The walk and multiplier become useful only when connected back to the original number-theoretic problem. The following theorem is the linchpin of the entire reduction: it converts divisibility (a number-theoretic condition) into a walk event (a group-theoretic condition), enabling all subsequent algebraic and harmonic-analytic arguments. Every route to MC in this paper—DH, CCSB, MMCSB, the large sieve, the spectral energy approach—passes through this bridge.

Theorem 2.5 (✓ walkZ_eq_neg_one_iff)**.** *For every missing prime q and every $n \in \mathbb{N}$:*

$$\mathrm{walkZ}(q,n) = -1 \ \text{ in } (\mathbb{Z}/q\mathbb{Z})^\times \quad \Longleftrightarrow \quad q \mid \big(\mathrm{prod}(n)+1\big).$$

This is the key translation: the group-theoretic event "the walk reaches -1" is the number-theoretic event "q divides $\mathrm{prod}(n)+1$." If q divides $\mathrm{prod}(n)+1$, then q could be a factor—and the minFac selection has a chance of picking q.

The bridge is formally verified because it is used in *every* reduction chain: DH $\Rightarrow$ MC, CCSB $\Rightarrow$ MC, MMCSB $\Rightarrow$ MC, and all sieve routes. An error here would invalidate the entire project.

2.3 SubgroupEscape and the Confinement Theorem

Before asking whether the walk reaches -1, we must ask whether it *can*. The key observation is that the walk is *multiplicative*: $\mathrm{walkZ}(q,n{+}1) = \mathrm{walkZ}(q,n) \cdot \mathrm{multZ}(q,n)$. Each step multiplies the current position by a new multiplier. This multiplicative structure is what makes *subgroups*—rather than arbitrary subsets—the natural obstruction.

To see why, suppose all the multipliers happen to land in some subset $S \subset (\mathbb{Z}/q\mathbb{Z})^\times$. If S is merely a subset with no algebraic structure, this tells us very little: multiplying elements of S together can produce anything, so the walk might still wander freely across the whole group. But if S happens to be a *subgroup* H, the situation changes dramatically. Products of elements of H stay in H (by closure), so $\mathrm{walkZ}(q,n)$ is trapped in the coset $\mathrm{walkZ}(q,0) \cdot H$ forever. The walk visits at most $|H|$ of the $q-1$ residues and may never reach -1.

This is not a theoretical curiosity. For instance, if every EM prime happened to be a quadratic residue mod q, then every multiplier would lie in the index-2 subgroup of squares, and the walk would be permanently confined to a coset of that subgroup—potentially one that does not contain -1.

Because $(\mathbb{Z}/q\mathbb{Z})^\times$ is cyclic of order $q-1$, its subgroup structure is particularly simple: there is exactly one subgroup of order d for each $d \mid q-1$, and the maximal proper subgroups are the index-ℓ subgroups for each prime $\ell \mid q-1$. There are relatively few of these (at most $\omega(q-1)$, the number of distinct prime factors of $q-1$), which is what makes SubgroupEscape a checkable condition.

Definition 2.6 (SubgroupEscape (**SE**))**.** For every missing prime q and every proper subgroup $H \lneq (\mathbb{Z}/q\mathbb{Z})^\times$: $\exists\, n,\ \mathrm{multZ}(q,n) \notin H$.

In other words, SE says that no proper subgroup contains all multipliers—equivalently, the multipliers *generate* the full group $(\mathbb{Z}/q\mathbb{Z})^\times$. The following theorem makes precise why SE is necessary.

Theorem 2.7 (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* $\mathrm{walkZ}(q,0) \cdot H$*. In particular, if $-1 \notin \mathrm{walkZ}(q,0) \cdot H$, the walk* never *hits -1.*

The proof is immediate from the multiplicative recurrence: if $\mathrm{multZ}(q,n) \in H$ for all n, then $\mathrm{walkZ}(q,n) = \mathrm{walkZ}(q,0) \cdot \mathrm{multZ}(q,0) \cdot \mathrm{multZ}(q,1) \cdots \mathrm{multZ}(q,n-1) \in \mathrm{walkZ}(q,0) \cdot H$. The walk cannot leave this coset, and if -1 is not in it, the walk is geometrically prevented from ever reaching -1—no amount of time will help.

SE eliminates this obstruction: once the multipliers generate the full group, the walk is not confined to any proper coset, and the possibility of hitting -1 is restored. SE does not by itself *guarantee* the walk reaches -1—that is a separate dynamical question—but without SE, the walk may be permanently locked out. Every route to MC must therefore either assume SE or prove it. (The bootstrap of Section 3 will prove it for free.)

Theorem 2.8 (✓ 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 $\ell \mid q-1$.*

3 The Inductive Bootstrap

The walk–divisibility bridge (Theorem 2.5) tells us *when* a missing prime q appears: it appears at step $n+1$ if and only if $\mathrm{walkZ}(q,n) = -1$ and q is the smallest prime factor of $\mathrm{prod}(n)+1$. SubgroupEscape (Section 2) tells us when this is *possible*: the walk can reach -1 only if the multipliers generate the full group. What remains is the dynamical question: does the walk actually reach -1?

The Hitting Hypothesis asserts that it does—not just once, but cofinally (infinitely often). Cofinality is needed because a single event $q \mid \mathrm{prod}(n)+1$ does not guarantee $\mathrm{seq}(n+1) = q$: a smaller prime might divide $\mathrm{prod}(n)+1$ as well. By asking for infinitely many hits, we ensure that eventually all smaller primes have already been selected, and q is the smallest factor available.

Definition 3.1 (HittingHypothesis (**HH**))**.** For every missing prime q: $\forall\, N,\ \exists\, n \geq N,\ q \mid (\mathrm{prod}(n)+1)$. Equivalently: the walk reaches -1 cofinally.

Theorem 3.2 (✓ hh_implies_mullin)**.** HH $\Rightarrow$ MC.

Proof sketch. By strong induction on p. Suppose every prime $< p$ is in the sequence (the IH), and assume p is missing. HH provides a cofinal sequence of indices where $p \mid \mathrm{prod}(n)+1$. At each such n, $\mathrm{seq}(n+1) = \mathrm{minFac}(\mathrm{prod}(n)+1) \leq p$. If $\mathrm{seq}(n+1) < p$, that prime is also missing—contradicting the IH. So $\mathrm{seq}(n+1) = p$. □

Proving MC therefore requires two things for each missing prime q: (i) SE, so the walk *can* reach -1, and (ii) HH, so the walk *does* reach -1 cofinally. Without further machinery, these are two independent open problems for every prime.

Original contribution. This section presents the paper's main new result: an elementary proof that SE is "free" at every step of the induction, eliminating one of the two problems entirely. Given MC($<p$) (the inductive hypothesis), SE at p follows from a single, purely algebraic lemma (PrimeResidueEscape). This reduces the entire conjecture to the dynamical question: does a walk with a generating set of multipliers hit every element cofinally?

The proof of PRE uses no analytic number theory, no Chebotarev density theorem, no character sums—only modular arithmetic. This makes the bootstrap both logically clean and Lean-friendly: the proof is short (∼100 lines) and depends on minimal Mathlib infrastructure.

3.1 PrimeResidueEscape

Definition 3.3 (PrimeResidueEscape (**PRE**))**.** For every prime $p \geq 5$ and every proper subgroup $H < (\mathbb{Z}/p\mathbb{Z})^\times$, some odd prime $r < p$ has residue $r \bmod p \notin H$.

Theorem 3.4 (✓ prime_residue_escape)**.** *PrimeResidueEscape holds.*

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:

- $p - 2 \equiv -2 \pmod{p}$ is in H (since $p - 2$ is odd and $< p$).
- $p - 4 \equiv -4 \pmod{p}$ is in H (since $p - 4$ is odd and $< p$, for $p \geq 5$).

Then $2 = (-4)(-2)^{-1} \in H$. Now every integer in $[1, p)$ is in H: even numbers are $2^k \cdot (\text{odd})$, both factors in H. So $H = (\mathbb{Z}/p\mathbb{Z})^\times$—contradicting H proper. □

The identity $2 = (-4)(-2)^{-1}$ is the only non-trivial step. No analytic number theory, no Chebotarev density theorem, no character sums—just modular arithmetic.

Remark 3.5 (Why PRE is not trivial). The problem PRE solves—showing that the natural primes below p cannot all land in a proper subgroup—looks like it should be easy but is surprisingly resistant to direct attack.

The obvious approach would be: use Dirichlet's theorem (infinitely many primes in each residue class), or Chebotarev (primes are equidistributed across cosets of any subgroup), or Linnik's theorem (the least prime in each progression is bounded). All of these work—but they are nuclear weapons for a problem that *feels* elementary. Worse, they are all absent from Mathlib, so they cannot be formalized. And they give much more than needed: one does not care about the *distribution* of primes across cosets, just that at least one prime misses H.

The other natural approach: argue directly about small primes. The first few EM primes are $2, 3, 5, 7, 13, 43, 53$—can one not just check that these escape any subgroup? One can for specific q (that is what the 30 SE instances for $q \leq 157$ do), but not universally. A proper subgroup of $(\mathbb{Z}/p\mathbb{Z})^\times$ can have index 2 and contain nearly half the elements—there is no reason a fixed finite set of primes could not all be quadratic residues mod some large p. The QR obstruction analysis shows this happens for at most 1.6% of primes, but "at most 1.6%" is not "never."

So one is stuck: the heavy theorems are not available, and direct checking does not generalize. PRE needs something else.

Why the identity is delightful. The key move is: *do not try to get any specific prime out of H. Get 2 out of H.*

If one can show $2 \in H$, one is done—because H already contains all odd numbers less than p (by the closure argument), so adding 2 gives all even numbers too, and $H = \top$.

But 2 is not odd, so the closure argument for odd numbers does not apply to 2 directly. One cannot factor 2 as a product of odd primes. Seems like a dead end.

The trick: one does not need 2 to *be* an odd number. One needs 2 to be *expressible using elements already known to be in* H. And H is a subgroup—closed under multiplication *and inversion*.

$$\begin{aligned} p-2 &\equiv -2 \pmod{p} \in H && (p{-}2 \text{ is odd and } < p), \\ p-4 &\equiv -4 \pmod{p} \in H && (p{-}4 \text{ is odd and } < p, \text{ since } p \geq 5). \end{aligned}$$

Therefore $(-4) \cdot (-2)^{-1} = 2 \in H$. One division in the group. The argument uses the specific arithmetic fact that $p-2$ and $p-4$ are both odd (which requires $p \geq 5$, hence the case split for small primes), the subgroup property (closure under products and inverses), and nothing else. No analytic number theory, no character theory, no sieve methods.

The delight is in the economy. One is trying to show that primes mod p escape every proper subgroup, and the proof does not mention primes at all after the first two sentences. It works entirely with odd numbers, uses only that H is a subgroup, and extracts 2 from the pair $(-2, -4)$ by a single division. The entire argument is five lines.

Why PRE matters for the project. PRE is the engine of the inductive bootstrap. Without it, the DH $\Rightarrow$ MC reduction would need SE as a separate hypothesis—a two-hypothesis reduction instead of a single-hypothesis one. The conjecture would be: "if the multipliers generate the full group *and* the walk hits -1 cofinally, then MC holds." That is weaker and less clean.

With PRE, SE becomes free. The inductive hypothesis MC($<p$) says all primes below p are in the sequence. PRE says some odd prime $r < p$ escapes any proper subgroup $H < (\mathbb{Z}/p\mathbb{Z})^\times$. Since r is in the sequence (by the IH), its residue mod p appears as a multiplier, and that multiplier escapes H. So the multipliers generate the full group—automatically, at every step, with no additional assumption.

This is what makes DH $\Rightarrow$ MC a *genuine reduction* rather than a *reformulation*. Without PRE, saying "DH implies MC" would be like saying "if the hard parts are true, the conjecture follows"—technically correct but not illuminating. With PRE, the algebraic part (generation) is proved for free, and the entire content of MC is concentrated in one dynamical question: does the walk hit -1 when the multipliers generate everything? The simplicity of the PRE proof—five lines, no analytic number theory—makes the reduction sharp: the hard part of MC is *purely dynamical*, not algebraic.

There is also a methodological point. The project explored Chebotarev, Linnik, Kummer theory, effective density theorems—all of which could prove SE but none of which are in Mathlib. PRE bypasses all of them with an argument that depends on nothing beyond the definition of a subgroup and the fact that $p-2$ and $p-4$ are odd. This is the kind of argument that makes formalization worthwhile: it is short enough to verify by hand, but subtle enough that one would not find it without specifically looking for a Mathlib-minimal proof of subgroup escape. The constraint of working within Lean forced a better proof.

3.2 The Bootstrap Mechanism

Theorem 3.6 (✓ mc_below_pre_implies_se)**.** MC($<p$) + PRE $\implies$ 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 MC($<p$), the prime r appears as seq(k) for some k. Then multZ$(p, k{-}1) \equiv r \pmod{p} \notin H$. So the multipliers escape H. □

The payoff of this mechanism—the full reduction from DynamicalHitting to MC—is assembled in Section 3.8, after we establish the supporting infrastructure.

3.3 The Threshold Approach

The threshold mechanism combines finite computation with the abstract bootstrap. If one can handle all primes below some bound B by direct verification, then MC reduces to DH for primes $\geq B$ only. This is formalized because it interfaces with the large sieve route (Section 4): MultiModularCSB gives DH for all primes above its threshold parameter Q_0, and the gap below Q_0 is closed by threshold verification.

Definition 3.7 (ThresholdHitting)**.** ThresholdHitting(B): DH restricted to primes $q \geq B$.

Theorem 3.8 (✓ threshold_11_implies_mullin')**.** ThresholdHitting(11) $\implies$ MC.

The computed values seq(0) = 2, seq(1) = 3, seq(2) = 7, seq(6) = 5 prove MC for every prime < 11. For $q \geq 11$, the IH + PRE give SE, and ThresholdHitting gives HH.

3.4 The Sieve Gap

The inductive structure of the proof does more than provide SE: it also resolves the selectability problem described in Section 1. Once all primes below q are in the sequence, they divide the running product and hence *cannot* divide any future Euclid number. This "sieve gap" means that past a computable stage, every prime factor of $\text{prod}(n) + 1$ is $\geq q$, so q is the *smallest* available factor whenever it divides the Euclid number. The minFac rule, which seemed like the source of difficulty, becomes an *ally*: it *must* select q the next time q divides an Euclid number.

This converts MC from "q divides $\text{prod}(n) + 1$ *and* is the smallest such factor" (hard) to simply "q divides $\text{prod}(n) + 1$ cofinally" (the walk hits -1 cofinally)—which is exactly what DynamicalHitting asserts.

Theorem 3.9 (q-ROUGHNESS — ✓ mc_below_implies_seq_ge)**.** *If* MC($< q$) *holds, then* $\exists\, N$ *such that* $\forall\, n \geq N$, $\text{seq}(n{+}1) \geq q$.

Theorem 3.10 (ONE-PRIME GAP — ✓ mc_below_cofinal_hit_implies_mc_at)**.** MC($< q$) *plus a single cofinal hitting event* ($\forall\, N, \exists\, n \geq N$, $q \mid \text{prod}(n) + 1$) *implies* MC($q$).

This is the **formal sieve gap**: the sieve at level $q-1$ is free from the IH, and extending it by one prime requires exactly one cofinal divisibility event—which is what DynamicalHitting asserts.

3.5 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 3.11 (PowerResidueEscape (**PRE**$_\ell$))**.** $\exists\, n : \text{multZ}(q, n)^{(q-1)/\ell} \neq 1$ in $(\mathbb{Z}/q\mathbb{Z})^\times$.

Theorem 3.12 (**PRE** $\Leftrightarrow$ **SE** — ✓ pre_iff_se)**.** PRE $\iff$ SE. *The forward direction uses only Lagrange's theorem; the reverse uses cyclicity of* $(\mathbb{Z}/q\mathbb{Z})^\times$.

Theorem 3.13 (✓ eight_elts_escape_order_le_seven)**.** *For* $q \geq 59$ *and* $(q{-}1)/\ell \leq 7$, $\text{PRE}_\ell(q)$ *holds automatically: the 8 known elements* $\{1, 3, 5, 7, 13, 43, 53, 21\}$ *cannot fit in any subgroup of order* ≤ 7.

3.6 Quadratic Reciprocity Obstruction

The power residue decomposition raises a natural question: for how many primes q could SE actually *fail*? The hardest case is the index-2 subgroup (quadratic residues), since it is the largest maximal subgroup. Quadratic reciprocity gives a definitive answer:

Theorem 3.14 (QR Obstruction — ✓ se_qr_obstruction)**.** *If all six multiplier primes* $\{3, 5, 7, 13, 43, 53\}$ *are quadratic residues mod* $q > 53$*, then* q *satisfies simultaneous Legendre symbol conditions. By CRT on* $\text{lcm}(12, 5, 28, 13, 43, 53) = 12{,}443{,}340$*, at most* 1.6% *of primes satisfy all conditions. For over* 98% *of primes, SE holds for the index-2 subgroup automatically.*

This means that even without the bootstrap, SE fails for at most a very sparse set of primes. With the bootstrap, SE holds for *all* primes—but the QR obstruction analysis illustrates why: the EM multiplier primes are arithmetically diverse enough to escape any single subgroup.

3.7 Concrete Verification

Theorem 3.15 (✓ se_at_11)**.** *SubgroupEscape holds for all 30 primes* $q \leq 157$ *not in the sequence, verified via power checks. For 29 of them, one of the first six multipliers is a primitive root. The exception is* $q = 131$*, where the seventh multiplier* $\text{seq}(7) = 6221671 \equiv 88 \pmod{131}$ *has full order 130.*

3.8 The Reduction to DynamicalHitting

We now have all the ingredients to assemble the main reduction. The goal is to identify the *minimal* hypothesis that, combined with the proved infrastructure, yields Mullin's Conjecture.

The Hitting Hypothesis (Definition 3.1) and its implication HH $\Rightarrow$ MC show that cofinal hitting suffices—but HH is a strong hypothesis, asserting cofinal hitting for *every* missing prime without any algebraic precondition. Can we weaken it?

Conditioning on SubgroupEscape

The confinement theorem (Theorem 2.7) tells us that if SE fails at q—meaning the multipliers are trapped in a proper subgroup H—then the walk is permanently confined to a coset of H. In this case, asking whether the walk hits -1 may be the wrong question: the walk might be structurally unable to reach -1, and no amount of dynamical analysis will help.

Conversely, when SE holds, the multipliers generate the full group $(\mathbb{Z}/q\mathbb{Z})^\times$, the walk is not confined to any proper coset, and the question of whether it hits -1 is a genuine dynamical question about the walk's long-term behavior.

DynamicalHitting captures exactly this: it asks for HH only when the algebraic precondition (SE) is satisfied.

Definition 3.16 (DynamicalHitting (**DH**))**.** For every missing prime q: $\text{SE}(q) \Rightarrow \text{HH}(q)$.

DH is *strictly weaker* than HH: it makes no claim about primes where SE fails. And DH is strictly weaker than MC: MC implies *every* prime appears, while DH only promises appearance *conditional* on SE. The remarkable fact is that DH is nevertheless *sufficient* for MC, because the bootstrap provides SE for free.

The Full Reduction

Theorem 3.17 (✓ dynamical_hitting_implies_mullin)**.** DH $\implies$ MC.

Proof. By strong induction on p. Assume $\text{MC}(< p)$ (every prime below p is in the sequence). We must show p appears.

1. **SE is free.** Theorem 3.6 gives $\mathrm{MC}(<p) + \mathrm{PRE} \Rightarrow \mathrm{SE}(p)$. Since PRE is proved unconditionally (Theorem 3.4), we obtain $\mathrm{SE}(p)$.
2. **DH gives HH.** Apply DH at the prime p: since $\mathrm{SE}(p)$ holds, DH yields $\mathrm{HH}(p)$ —the walk $\mathrm{walkZ}(p, \cdot)$ reaches -1 cofinally.
3. **The sieve gap closes.** By Theorem 3.10, $\mathrm{MC}(<p)$ combined with cofinal hitting implies $\mathrm{MC}(p)$: once all primes below p are in the sequence, the minFac rule *must* select p the next time it divides an Euclid number.

□

The Reduction Chain

Assembling all the reductions proved in this paper, we obtain the following chain. Each arrow is a formally verified implication in Lean:

$$\mathrm{DH} \xrightarrow{\text{bootstrap}} \mathrm{SE\ (free)} \xrightarrow{\mathrm{DH}} \mathrm{HH} \xrightarrow{\text{sieve gap}} \mathrm{MC}$$

The entire conjecture has been reduced to a single, cleanly stated dynamical hypothesis: *if the multipliers generate the full group, then the walk hits every element cofinally.* No algebraic hypothesis remains open; the PRE lemma and the bootstrap mechanism have eliminated SE completely.

The threshold variant (Theorem 3.8) offers a further refinement: $\mathrm{ThresholdHitting}(11) \Rightarrow \mathrm{MC}$. This combines finite computation (MC for primes < 11, from the known sequence values) with the abstract bootstrap, restricting DH to primes $q \geq 11$ only.

4 The Character Sum Reduction

The algebraic reduction (DH $\Rightarrow$ MC) leaves the dynamical content of DH untouched: *does* the walk hit -1 when the multipliers generate the full group? This section develops a parallel, character-analytic approach that precisely characterizes what is needed.

Why character sums? The algebraic route (DH $\Rightarrow$ MC) asks: does the walk hit a *single* element (-1)? Character sums answer a stronger question: does the walk visit *every* element with equal frequency? Dirichlet characters $\chi : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}^\times$ are the Fourier basis of the cyclic group; they detect imbalances in distribution. By character orthogonality, the number of times the walk visits a particular element t equals $N/(q-1)$ (the uniform share) plus correction terms involving character sums. If all non-trivial character sums are $o(N)$—negligible compared to the main term—the correction terms vanish asymptotically, and the walk is equidistributed; in particular, it visits -1 cofinally, giving HH.

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 (§4.4) are original contributions of this formalization. The large sieve infrastructure (§4.5, Appendix A.3)—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 §4.5).

The formalization develops this Fourier-analytic reduction because it provides the cleanest interface between MC 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 MC.

4.1 Character Sums and Walk Equidistribution

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

$$S_\chi(N) \;=\; \sum_{n<N} \chi\big(\mathrm{walkZ}(q,n)\big).$$

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 intuition: if χ is a non-trivial character and the walk visits every group element equally often, the sum $S_\chi(N)$ cancels out (positive and negative contributions balance). CCSB asks for exactly this cancellation.

Theorem 4.2 (✓ complex_csb_mc')**.** CCSB $\implies$ 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 walk visits every unit class with positive lower density. This follows from the character orthogonality formula (Theorem 4.3 below): the number of hits to any target t equals $N/(q-1)$ plus correction terms of size $o(N)$, so the count is eventually positive.
2. **Walk equidistribution implies DH** (✓ walk_equidist_mc): if the walk visits every unit class cofinally, it visits -1 cofinally, giving HH. SE is not needed as a separate hypothesis: equidistribution already implies the multipliers generate the full group (a walk confined to a proper coset cannot be equidistributed).
3. **DH implies MC** (Theorem 3.17): 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 $t \in (\mathbb{Z}/q\mathbb{Z})^\times$:*

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

where the sum is over all Dirichlet characters mod q.

This is a standard Fourier inversion formula on the finite group $(\mathbb{Z}/q\mathbb{Z})^\times$. To understand it, split the sum into the contribution from the trivial character χ_0 (which satisfies $\chi_0(a) = 1$ for all a) and the remaining non-trivial characters:

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

The first term is the "fair share" count: if the walk were perfectly equidistributed among the $q-1$ units, each would be visited $N/(q-1)$ times. The correction term measures the deviation from uniformity. Since $|\overline{\chi(t)}| = 1$, each summand in the correction is bounded by $|S_\chi(N)|$, and there are $q-2$ non-trivial characters. 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 for every target t, including $t = -1$.

Returning to our running example ($q = 41$): the group $(\mathbb{Z}/41\mathbb{Z})^\times$ has 40 elements and 40 characters. The Fourier bridge says the number of times the walk reaches $\mathrm{walkZ}(41, n) = 40$ (i.e. -1 mod 41) in the first N steps equals $N/40$ plus a correction bounded by the 39 non-trivial character sums. Proving CCSB for $q = 41$ would establish that 41 eventually appears.

4.3 The Decorrelation–PED–CCSB Chain

CCSB is a statement about the *walk* character sums $S_\chi(N) = \sum_{n<N} \chi(\mathrm{walkZ}(q, n))$. But the walk is built from the *multipliers*: $\mathrm{walkZ}(q, n+1) = \mathrm{walkZ}(q, n) \cdot \mathrm{multZ}(q, n)$. Can we reduce walk equidistribution to a simpler property of the multiplier sequence? This subsection formalizes a chain of progressively weaker hypotheses about the multipliers, each implying the next via proved bridges, 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 $N \geq N_0$: $|\{k < N : \chi(\mathrm{multZ}(q, k)) \neq 1\}| \geq \delta N$.

The name "escape" comes from the SubgroupEscape perspective: the kernel $\ker(\chi)$ is a proper subgroup of $(\mathbb{Z}/q\mathbb{Z})^\times$, and $\chi(\mathrm{multZ}(q, k)) \neq 1$ 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. This is a weak condition—it says nothing about cancellation or equidistribution, 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 $o(N)$: $\|\sum_{n<N} \chi(\mathrm{multZ}(q, n))\| \leq \varepsilon N$ 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 $\chi(\mathrm{walkZ}(q, n))$ 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.

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 $c \in (\mathbb{Z}/q\mathbb{Z})^\times$: $\|\sum_{\substack{n<N \\ \mathrm{walkZ}(q,n)=c}} \chi(\mathrm{multZ}(q, n))\| \leq \varepsilon N$.

CME is strictly stronger than Decorrelation: it asks for multiplier character sum cancellation *conditioned on walk position*. Decorrelation bounds the global sum $\sum_{n<N} \chi(\mathrm{multZ}(q,n)) = o(N)$; CME bounds the fiber sums $\sum_{\substack{n<N \\ \mathrm{walkZ}(q,n)=c}} \chi(\mathrm{multZ}(q,n)) = o(N)$ separately for each c. 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.

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:

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

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):

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

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

4.4 Walk Telescoping Identities

The following identities are formalized not because they solve CCSB, but because they reveal *structural constraints* that any proof of CCSB must navigate. They are "no-go" results that rule out certain proof strategies.

Theorem 4.12 (✓ 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 4.13 (✓ walk_telescope_norm_bound)**.** *The telescoping sum has norm ≤ 2 (triangle inequality on unit-norm terms).*

The ≤ 2 bound looks innocent, but it has a sharp consequence. If we write $S_N = \sum_{n<N} \chi(w(n))$ for the walk character sum, the telescope identity links S_N to the multiplier character sum $M_N = \sum_{n<N} \chi(m(n))$. Specifically, splitting the product in Theorem 4.12 gives $S_N \cdot \overline{M_N/N} - S_N = O(1)$ (after normalization), tightly coupling the walk and multiplier sums.

Theorem 4.14 (✓ 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 A.3) or (ii) establish multiplier decorrelation first.

4.5 The Large Sieve Route

Sessions 24–36 developed an extensive large sieve infrastructure across three files (LargeSieve.lean, LargeSieveHarmonic.lean, LargeSieveAnalytic.lean) totaling ∼5,870 lines. This route connects 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 4.15 (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; the threshold mechanism handles those exceptions.

Theorem 4.16 (✓ mmcsb_implies_mc)**.** MultiModularCSB $\implies$ MC.

The proof composes the per-prime Fourier bridge (Theorem 4.3) with the threshold mechanism (Theorem 3.8): for $q \geq Q_0$, MultiModularCSB gives walk equidistribution and hence HH; for $q < Q_0$, we verify MC directly. This result was previously an open hypothesis in §36; proving it (Session 24) was a key milestone that unified the large sieve and character sum approaches.

Three parallel routes to MultiModularCSB are formalized:

Bombieri–Vinogradov route. The Bombieri–Vinogradov theorem is one of the most powerful results in analytic number theory. Roughly, it says that primes are equidistributed among arithmetic progressions "on average over moduli": for most moduli $q \leq Q = \sqrt{x}/(\log x)^A$, the count of primes $\leq x$ in any progression a mod q is close to the expected $\pi(x)/\phi(q)$. If we could

apply this to the EM walk, where the multipliers are primes, we would get MMCSB and hence MC.

Theorem 4.17 (✓ 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 ⇒ EMMultCSB, where EMMultCSB bounds the *multiplier* character sums) from the dynamical content (EMMultCSB ⇒ 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 4.14) 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, Section 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 per Session 35) 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 A.

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 $\text{prod}(n)+1$ 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* $p \mid \text{prod}(n)+1$*, then* $\text{seq}(m) \neq p$ *for all* $m \leq n$*.*

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

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

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

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

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 $\text{prod}(n) + 1$ at each step. In this variant, MC follows from DH alone: whenever $p \mid \text{prod}(n) + 1$, 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 $r_n = \text{prod}(n) \bmod p$ 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., $p \mid \text{prod}(n) + 1$) occurs with probability $\to 1/(p-1)$—infinitely often with probability 1.

The lpf variant is hard. The actual Euclid–Mullin sequence uses $\text{seq}(n+1) = \text{minFac}(\text{prod}(n)+1)$, 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 $\text{prod}(n)+1$, 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: MC($<p$) ensures all primes below p are already in the sequence, hence divide $\text{prod}(n)$, hence cannot divide $\text{prod}(n)+1$. 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)**.** EuclidMinFacEscape(q) $\Leftrightarrow$ 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* $\text{multZ}(q, n) = -\text{walkZ}(q, n)^{-1}$—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 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.

5.4 The Van der Corput Barrier

The van der Corput inequality (Theorem A.10, now fully proved in the formalization) converts character sum bounds into autocorrelation bounds. Theorem 4.14 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 A.12). The telescoping identity $\sum_n \chi(w(n))(\chi(m(n)) - 1) = O(1)$ is a precise structural constraint.

5.5 The Walk Bridge Falsity

The BV route decomposes into two stages: sieve transfer (BV $\Rightarrow$ EMMultCSB, bounding *multiplier* character sums) and the walk bridge (EMMultCSB $\Rightarrow$ 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 (Section 4.8) is essential. CME uses fiber decomposition and telescoping to go directly from conditional multiplier equidistribution to CCSB, circumventing the walk bridge entirely.

5.6 Dead Ends as a Roadmap

Across 22 formalization sessions, dozens of potential approaches were explored and found to be dead ends. Each elimination is informative: it narrows the space of viable strategies. A selection of twelve, 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.

The pattern: every approach that avoids the specific EM orbit's joint distribution either reduces to CCSB or fails. Since CME implies CCSB (proved), the sharpest target is now CME: conditional equidistribution of multipliers given walk position. CME is strictly weaker than CCSB and is the *irreducible analytic content*.

5.7 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 $\mathrm{walkZ}(q,n) = \mathrm{prod}(n) \bmod q$ hits -1 cofinally.
- **CCSB**: For every non-trivial character $\chi\colon (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}^\times$, the sum $\sum_{n<N} \chi(\mathrm{walkZ}(q,n)) = o(N)$.
- 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.

5.8 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. $\text{prod}(n) \geq 2^n$ (✓ prod_lower_bound_for_sieve). The Euclid numbers $\text{prod}(n) + 1$ 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 $\text{prod}(n) + 1$ is coprime to each of $\text{seq}(0), \ldots, \text{seq}(n)$, a growing set of distinct primes. The pool of "available" small primes as factors of $\text{prod}(n) + 1$ shrinks, but the size of $\text{prod}(n) + 1$ grows so fast that it must have enormous prime factors most of the time.

Feature 2: Mutual coprimality of Euclid numbers. For $m > n$, $\text{prod}(m)$ is divisible by $\text{seq}(n + 1)$, which divides $\text{prod}(n) + 1$. So $\text{prod}(m) + 1 \equiv 1 \pmod{\text{seq}(n + 1)}$: 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 $\text{walkZ}(q, n) \neq -1$ (so that q does not divide $\text{prod}(n) + 1$ as the smallest factor), then the multiplier $\text{multZ}(q, n) = \text{minFac}(\text{prod}(n) + 1)$ satisfies $\text{multZ}(q, n) \leq (\text{prod}(n) + 1)^{1/2}$. 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 $\text{prod}(n) + 1$, not just its residue.

Feature 4: Self-correcting feedback. If the walk concentrates on certain residues mod q—say $\text{walkZ}(q, n) \equiv a \pmod{q}$ for many n—then $\text{prod}(n) + 1 \equiv a + 1 \pmod{q}$ 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.

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

6.1 Codebase Structure

The formalization uses Lean 4 with Mathlib v4.27.0 across 32 files totaling ∼22,400 lines. The dependency chain is linear, with three leaf modules:

File	Content	Lines
Euclid.lean	Constructive Euclid's theorem	425
MullinDefs.lean	seq, prod, aux, identities	527
MullinConjectures.lean	MC, Conjecture A (FALSE), HH	494
MullinDWH.lean	DivisorWalkHypothesis (leaf)	551
MullinResidueWalk.lean	WalkCoverage, residue walk, concrete MC	605
MullinGroupCore.lean	walkZ, multZ, confinement, SE	422
MullinGroupEscape.lean	Escape lemmas, 8-element argument	673
MullinGroupSEInstances.lean	30 concrete SE instances ($q \leq 157$)	364
MullinGroupPumping.lean	Gordon sequenceability (leaf)	343
MullinGroupQR.lean	QR obstruction ($\leq 1.6\%$) (leaf)	683
RotorRouter.lean	Scheduled walk coverage (standalone)	421
MullinRotorBridge.lean	EMPR + SE $\Rightarrow$ MC bridge	87
EquidistPreamble.lean	PE $\Rightarrow$ MC, bootstrapping	234
EquidistSieve.lean	Sieve, WHP $\Leftrightarrow$ HH	297
EquidistSelfAvoidance.lean	Self-avoidance, periodicity	450
EquidistCharPRE.lean	Character non-vanishing, PRE $\Leftrightarrow$ SE	811
EquidistBootstrap.lean	Inductive bootstrap, DH $\Rightarrow$ MC	522
EquidistThreshold.lean	ThresholdHitting(11) $\Rightarrow$ MC	299
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)	1114
EquidistSieveTransfer.lean	Prime density, sieve transfer, walk decomp (§38–§78)	1319
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 (§56–§65)	1438
LargeSieveSpectral.lean	Walk energy, HOD, VdC, CME, SVE (§66–§78)	1685
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, large sieve	455

6.2 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.

6.3 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.

6.4 Verification Statistics

Lines of Lean code	∼22,400
Files	32
Theorems/lemmas	∼770
Definitions	∼460
sorry occurrences	**0**
Open hypotheses (stated as def)	∼26
Concrete SE instances	30
Computed sequence terms	8
Mathlib version	v4.27.0

7 Open Problems

7.1 CCSB as the Precise Frontier

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

Are the walk character sums $\sum_{n<N} \chi(\text{walkZ}(q,n))$ *bounded* $o(N)$ *for every non-trivial* χ*?*

CCSB is a single hypothesis that implies MC with no additional conditions. It is equivalent to walk equidistribution mod q, which is the strongest "uniform" version of DH.

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 walk equidistributes; exceptional primes (finitely many) are handled by FiniteMCBelow. Combining with ThresholdHitting(11) ⇒ MC would close 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 $\text{prod}(n) + 1$, 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 $q \geq B$ 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.

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 minFac(prod(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.

Sessions 19–22 articulated a richer sieve infrastructure in two parallel routes (§38–§39 of EquidistSelfCorrecting.lean):

Cumulative route.

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

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 window equidistribution of EM multipliers 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: this applies to *generic* integers whose minFac residues are equidistributed (by CRT + Mertens), not to the specific EM orbit. Transferring from ensemble to 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 ⇒ 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 ⇒ 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 the deterministic product walk cannot maintain character-sum bias $\Theta(N)$. 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 version over q would suffice, combined with ThresholdHitting(11).

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 transferring the averaged or generic equidistribution to the specific EM walk. This is the same orbit-specificity gap as SieveTransfer, approached from a different mathematical toolkit.

7.6 Is DynamicalHitting True?

The formalization proves DH ⇒ MC but says nothing about whether DH itself holds. Intellectual honesty requires a frank assessment.

Evidence for. Three independent lines of evidence suggest DH is true. (1) *Computation*: every prime below 41 has been verified to appear in the EM sequence (51 terms computed), and the walk hits $-1 \pmod{q}$ for all tested primes q with no counterexample. (2) *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. (3) *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. DH 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 equidistribution to walk equidistribution (PED $\Rightarrow$ BRE $\Rightarrow$ CCSB) is *impossible* for character orders $d \geq 3$. This is not evidence that DH is false, but it shows that the truth of DH, if it holds, 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 DH is very likely true, primarily because the self-correcting sieve mechanism provides a concrete dynamical reason (not merely a probabilistic heuristic) for the walk to equidistribute. The strongest form of this belief: CME should hold because the EM multipliers, conditioned on walk position, have no arithmetic reason to correlate with characters of $(\mathbb{Z}/q\mathbb{Z})^\times$. 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 seq(n) is prime	Proved	seq_isPrime
No prime repeats	Proved	seq_injective
seq(0–7) computed	Proved	seq_zero..seq_seven
Main reductions to MC		
DH $\Rightarrow$ **MC** (irreducible core)	Proved	dynamical_hitting_implies_mullin
ThHit(11) $\Rightarrow$ **MC**	Proved	threshold_11_implies_mullin'
CCSB $\Rightarrow$ **MC** (single hypothesis)	Proved	complex_csb_mc'
PE $\Rightarrow$ MC	Proved	pe_implies_mullin

continued on next page

Result	Status	Lean identifier
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	mc_below_pre_implies_se
q-roughness from MC($<q$)	Proved	mc_below_implies_seq_ge
One-prime gap	Proved	mc_below_cofinal_hit_implies_mc_at
mc_below 11	Proved	concrete_mc_below_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
QR obstruction ($\leq 1.6\%$)	Proved	se_qr_obstruction
SE for 30 primes ($q \leq 157$)	Proved	se_at_11..se_at_157
Character sum chain		
Fourier bridge: CCSB ⇒ hit count lb	Proved	complex_csb_implies_hit_count_lb_proved
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

continued on next page

Result	Status	Lean identifier
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
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

continued on next page

Result	Status	Lean identifier
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		
DynamicalHitting	**Open**	DynamicalHitting
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

continued on next page

Result	Status	Lean identifier
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, §5.5)	**Dead**	MultCSBImpliesMMCSB
BlockRotationEstimate (impossible for $d \geq 3$, §5.7)	**Dead**	BlockRotationEstimate

A 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.

A.1 Arithmetic Large Sieve Route

Theorem A.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; Session 35 showed it is a **dead end**—universal coefficient bounds cannot distinguish equidistributed walks from clumped walks.

A.2 Analytic Large Sieve Route

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

Definition A.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 A.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 A.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 A.5 (✓ als_implies_prime_arith_ls)**.** AnalyticLargeSieve $\implies$ PrimeArithLS.

Theorem A.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.

A.3 The Spectral Energy Route

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

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

For the EM walk, lag-h autocorrelations involve h-step multiplier products. At $h = 1$, the autocorrelation equals the multiplier character sum (Theorem 4.14), so VdC with a single shift gives only $O(N)$. Higher lags may decorrelate:

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

Theorem A.12 (✓ hod_implies_ccsb)**.** HOD $\implies$ CCSB $\implies$ MC.

Conditional multiplier equidistribution.

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

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

Theorem A.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 $d \geq 3$ barrier. The proof works for all character orders because it uses only the fiber decomposition and telescoping—no block rotation estimate is needed.

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

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

Theorem A.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.

A.4 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 A.15).

Every hypothesis implies MC. The PED route has an open BRE bridge for $d \geq 3$ 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.

B Methodology: Human–AI Collaboration

This work was produced through a sustained collaboration between a human author and an AI system (Claude, Anthropic) across 43+ 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 43+ 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.

C 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	walkZ(q,n) = prod(n) mod q; multZ(q,n) = seq$(n{+}1)$ mod 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. 2.6
HH	HittingHypothesis	The walk reaches -1 cofinally: $\forall N, \exists n \geq N,\ q \mid \text{prod}(n)+1$	Def. 3.1
DH	DynamicalHitting	SE$(q) \Rightarrow$ HH(q) for every missing prime q	Def. 3.16
PRE	PrimeResidueEscape	Every proper subgroup of $(\mathbb{Z}/p\mathbb{Z})^\times$ is escaped by some odd prime $< p$	Thm. 3.4
PRE$_\ell$	PowerResidueEscape	Multipliers escape the index-ℓ subgroup of $(\mathbb{Z}/q\mathbb{Z})^\times$	§3
—	ThresholdHitting	DH restricted to primes $q \geq B$	§3
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 multZ$(q,n) \notin H$, for every proper H	§4
—	DecorrelationHypothesis	multZ(q,n) and multZ$(q,n{+}1)$ are asymptotically independent	§4

continued on next page

Abbr.	Name	Meaning	Ref.
BRE	BlockRotationEstimate	Cancellation in block sums of characters applied to walk	§4
SVE	SubquadraticVisitEnergy	$\sum_a \lvert\{n \le N : \mathrm{walkZ}(q,n) = a\}\rvert^2 = o(N^2/(q-1))$	§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. 2.7
—	Walk–Divisibility Bridge	$\mathrm{walkZ}(q,n) = -1 \Leftrightarrow q \mid \mathrm{prod}(n)+1$	Thm. 2.5
—	One-prime gap	$\mathrm{MC}(<q)$ + cofinal hit $\Rightarrow \mathrm{MC}(q)$	Thm. 3.10
—	QR Obstruction	SE fails for at most 1.6% of primes (index-2 subgroup)	§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 $\le x$ 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.