RESEARCH PROGRAMME

Goldbach as Geometry, Parity and Correlation

Visual structures and proof targets for the binary problem

Christopher Yamba
Research programme developed with computational and AI-assisted investigation

28 October 2026

Abstract

Every even integer greater than two is conjectured to be a sum of two primes. This article develops a visual and analytic research programme around that open question. Prime-pair sums form layers of a Ferrers diagram, reflection matrices encode representations as elementary blocks, and modular sieving becomes a covering problem. A prime-versus-semiprime parity statistic exposes a precise pointwise hurdle; an averaged calculation reveals a cutoff at which the leading balance reverses. The article also describes a classical Buchstab integral, a conditional Möbius-twisted distribution route, and adversarial examples that test what any proof must use. The mathematical deductions here are working results derived from established inputs and checked computationally or by independent research agents; they have not received external peer review. No proof of Goldbach or claim of priority is made.

Status and scope

Goldbach's binary conjecture remains open. All finite checks in this article are evidence about their stated ranges, not proofs at arbitrary size. Terms such as *research target* identify unproved statements. The averages in this article concern ordered pairs; in particular, $p+q$ and $q+p$ count separately when $p \neq q$.

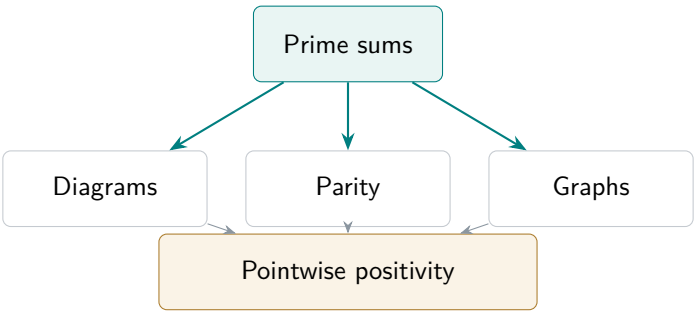


Figure 1: Different languages meet at the same difficult question: can a positive pair count be proved for *each* even target?

1 A small equation, an exacting demand

Write

$$R(N) = \#\{p \in \mathbb{P} : N - p \in \mathbb{P}\}. \tag{1}$$

For even $N \geq 4$, the strong Goldbach conjecture asks for $R(N) > 0$. For example, 30 has six ordered representations, coming from $7 + 23$, $11 + 19$ and $13 + 17$ and their reversals. There are no known counterexamples; the published verification by Oliveira e Silva, Herzog and Pardi reaches $4 \cdot 10^{18}$ [2]. Computation does not settle infinitely many cases.

The surrounding landscape is rich. Helfgott proved the ternary conjecture: every odd integer greater than five is a sum of three primes [1]. Chen proved that every sufficiently large even integer is a prime plus a number with at most two prime factors. Both achievements illuminate the binary problem, but neither automatically fixes an arbitrary two-prime slice. The difficulty is especially clear in the difference between a statement about *almost every* even integer and a pointwise statement about *every* one.

The central question

How can we turn abundant average structure into a positive lower bound at one fixed target N? Each representation below sharpens this question. None removes it by relabelling alone.

2 The geometry of prime sums

List the odd primes $p_1 < p_2 < \cdots$ and form the set

$$\mathcal{F}_2(N) = \{(i, j) : p_i + p_j \leq N\}. \tag{2}$$

It is a Ferrers diagram: if a box appears, every box weakly above and to its left appears too. The difference $\mathcal{F}_2(N) \setminus \mathcal{F}_2(N-2)$ consists exactly of the ordered representations of N by odd primes. For even $N \geq 6$ this is all of $R(N)$; the exceptional case $4 = 2 + 2$ is handled separately.

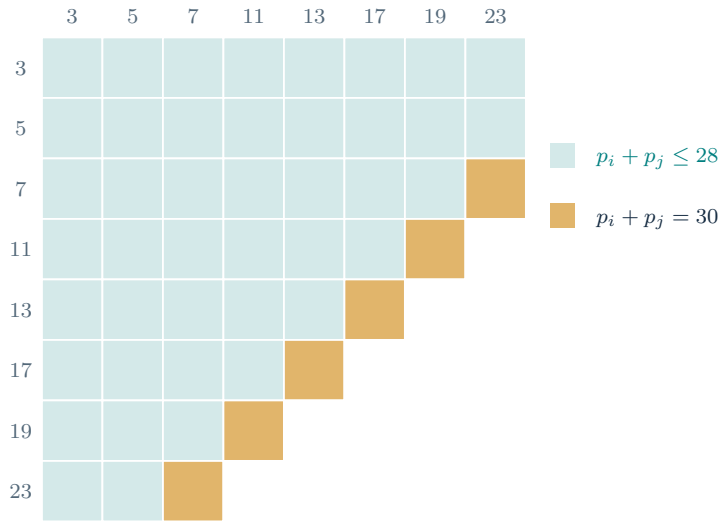


Figure 2: Prime-index Ferrers diagram at 30. Gold squares are the new layer. Row and column labels are prime values, not consecutive integers.

Every row and every column contains at most one new box. Thus the layer is a symmetric placement of nonattacking rooks. A diagonal box is the special case $N = 2p$, where p is prime.

For arbitrary Ferrers diagrams a layer may be empty; the unanswered question is what additional property of the *prime-index* family forbids an empty even layer.

The construction extends naturally to

$$\mathcal{F}_k(T) = \{(i_1, \dots, i_k) : p_{i_1} + \cdots + p_{i_k} \le T\}.$$

Each new shell is a partial Latin-hypercube pattern: an axis-parallel line contains at most one new box. The two-prime problem for N can be viewed as the slice of the three-prime shell at $N+3$ whose third coordinate is 3. A nonempty three-dimensional shell need not intersect that specific slice. This is a geometric picture of why a ternary theorem does not immediately establish the binary one.

3 Reflection, permutations and graphs

Let D be diagonal on the integers $2, \dots, N-2$, with $D_{aa} = 1$ for primes and zero otherwise. Let J_N be the permutation that reflects a to $N-a$. The symmetric matrix

$$B_N = DJ_ND$$

retains exactly those reflection edges whose endpoints are both prime. Every unequal Goldbach pair contributes a block $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$; a pair $p + p = N$ contributes the block (1). Therefore

$$R(N) = \operatorname{tr}(B_N^2) = 2k + f, \tag{3}$$

where k is the number of unequal unordered pairs and $f = 1$ if $N/2$ is prime, otherwise 0.



Figure 3: The reflection matching at $N = 30$. A loop occurs only when $N/2$ is prime.

There is also a spectral form. Take $M > 2N$ and a directed Cayley graph on $\mathbb{Z}/M\mathbb{Z}$ with an edge $x \to x + p$ for each prime $p \le N$. Its two-step path count from 0 to N equals $R(N)$. Fourier diagonalization turns this into an exponential-sum identity. But simple expander mixing is too blunt: prime steps have strong local biases, especially parity. Parseval alone forces the largest-eigenvalue bound to be too large for the most naive positivity criterion. A promising spectral argument would need to preserve the *phases* and isolate rational-frequency structure.

4 A sieve is a covering problem

For $2 \le a \le N-2$, examine the pair $(a, N-a)$. For each prime $\ell \le \sqrt{N}$, mark a if either coordinate is a nontrivial multiple of ℓ. The qualification matters: a coordinate equal to ℓ is itself prime and must not be crossed out. With that exception in place, an unmarked candidate is exactly a Goldbach representation.

If $Q_z = \prod_{\ell \le z} \ell$, the Chinese remainder theorem gives the exact density over a complete period:

$$\delta_N(z) = \prod_{\ell \le z} \left(1 - \frac{\nu_\ell(N)}{\ell}\right), \qquad \nu_\ell(N) = \begin{cases} 1, & \ell \mid N, \\ 2, & \ell \nmid N. \end{cases}$$

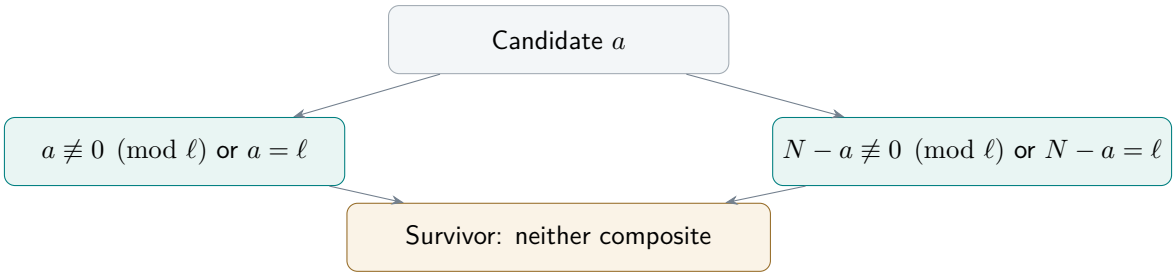


Figure 4: The sieve as a hypergraph cover. Both tests must pass for each sieving prime. A survivor avoids all nontrivial divisibility marks; prime-equals-sieving-prime exceptions are essential.

The period becomes enormous when z approaches $\sqrt{N}$. A positive density over that period does not guarantee a point in the short interval associated with one N. The team checked an average obstruction: at complete-sieve depth, the ratio of the full-period prediction to the actual survivor total tends to $4e^{-2\gamma} \approx 1.260947$. This follows from the prime number theorem, Mertens' product formula and averaging the local factors. It rules out a naive negligible-error approximation, not every possible one-sided lower bound. A closely related two-cloud formulation has appeared in a 2026 preprint [12].

5 Chen's bridge and the parity problem

Chen's theorem offers a particularly sharp distinction between what is known and what remains. Let

$$\mathcal{C}_N = \{p \in \mathbb{P} : q = N - p \geq 2,\ \Omega(q) \leq 2\}, \quad S_N = |\mathcal{C}_N|, \quad L_N = \sum_{p \in \mathcal{C}_N} (-1)^{\Omega(N-p)}.$$

A prime complement contributes -1 to L_N, while a semiprime contributes $+1$. Thus

$$\boxed{R(N) = \frac{S_N - L_N}{2}}. \tag{4}$$

Chen supplies $S_N > 0$ for sufficiently large even N. The missing step is $L_N < S_N$, a pointwise statement equivalent here to the existence of a prime complement.

A prior version of this programme suggested a fixed relative gap $L_N \leq (1 - \eta)S_N$ for every large N. That target is *false* for the unrestricted set. The prime number theorem and Landau's semiprime asymptotic give, summing over even $N \leq X$,

$$\sum R(N) \sim \frac{X^2}{2\log^2 X}, \qquad \sum S_N \sim \frac{X^2 \log\log X}{2\log^2 X}.$$

The aggregate prime share tends to zero like $1/\log\log X$; semiprime complements with a small factor are too numerous [3]. This corrects a proposed target rather than disproving Goldbach.

6 A roughness threshold and a signed statistic

Take $z_N = \lfloor N^{1/3} \rfloor$. A number $n \leq N$ whose prime factors all exceed z_N has at most two prime factors, counting multiplicity: three would have product exceeding N. For ordered pairs $a + b = N$ with both coordinates larger than z_N and free of prime factors up to z_N, define counts PP, PS, SP, SS according to whether each coordinate is prime or semiprime.

	right prime	**right semiprime**
left prime	PP	PS
left semiprime	SP	SS

Table 1: The four types of cube-root-rough pairs. Reflection exchanges PS and SP.

Let $\lambda(n) = (-1)^{\Omega(n)}$ and set

$$T(N) = \sum_{\substack{a+b=N,\ a,b>z_N \\ P^-(a),P^-(b)>z_N}} \lambda(a). \tag{5}$$

The mixed terms cancel by the map $(a, b) \mapsto (b, a)$, so

$$T(N) = SS(N) - PP(N). \tag{6}$$

In particular, $T(N) < 0$ implies Goldbach at N. This is a stronger criterion: $T(N) \geq 0$ does not disprove Goldbach.

Theorem 6.1 (A signed average; derived in this programme)**.** *As $X \to \infty$, the signed statistic in (5) obeys*

$$\sum_{\substack{N \leq X \\ 2 \mid N}} T(N) \sim \frac{(\log 2)^2 - 1}{2} \frac{X^2}{\log^2 X}.$$

Proof sketch. At a fixed cutoff $z = X^{1/3}$, the prime number theorem gives $X/\log X$ primes at scale X. The number of semiprimes with both factors above z has leading coefficient

$$\int_{1/3}^{1/2} \frac{dt}{t(1-t)} = \log 2.$$

Rescaled prime and rough-semiprime counting measures converge to constant multiples of Lebesgue measure. The triangle $u, v \geq 0$, $u + v \leq 1$ has area $1/2$; hence PP has coefficient $1/2$ and SS has coefficient $(\log 2)^2/2$. Replacing the fixed cutoff by $N^{1/3}$ changes the sums by lower-order terms; contributions from very small targets can be bounded and then discarded. Equation (6) supplies the sign. The accompanying research dossier gives the boundary argument. □

The constant is negative, about -0.25977. Using a classical almost-all Goldbach estimate and a separate finite-residue argument, the research team obtained a lower asymptotic bound of roughly 26.22% for the proportion of even targets satisfying $T(N) < 0$. This is a working deduction awaiting external review; it does not exclude an exceptional target. The classical log-weighted estimate used in that deduction is described by Tolev [4].

> **Average is not pointwise**
>
> A negative sum of many values does not make each value negative. Establishing $T(N) < 0$ for every sufficiently large even N would itself settle a strengthening of Goldbach. No such estimate is proved here.

7 The cutoff where the average changes sign

Consider a wider family with $z_N = \lfloor N^\alpha \rfloor$, $0 < \alpha < 1/2$. Let PP_α count prime-prime pairs above the cutoff and SS_α count semiprime-semiprime pairs with all prime factors above it. Their

average has a simple coefficient:

$$\sum_{\substack{N\le X\\ 2\mid N}} (SS_\alpha(N) - PP_\alpha(N)) \sim \frac{\log^2((1-\alpha)/\alpha) - 1}{2} \frac{X^2}{\log^2 X}. \tag{7}$$

The sign changes at

$$\alpha_* = \frac{1}{1+\mathrm{e}} \approx 0.268941.$$

For every fixed $\alpha < \alpha_*$ the average is positive, forcing $SS_\alpha(N) > PP_\alpha(N)$ for infinitely many even N. At the quarter-root cutoff, a scan of 27,986 selected targets through one million found no strict $SS > PP$ instance, despite the asymptotic conclusion. This is a clear example of a convincing finite pattern that cannot persist forever. At $\alpha < 1/3$, numbers with three or more factors also survive, so (7) compares the two named classes rather than the complete signed survivor set.



Figure 5: The average balance changes at $\alpha_* = 1/(1+\mathrm{e})$. Only at and above $1/3$ does the four-cell classification exhaust rough survivors.

8 Just below the cube root: three-factor contamination

For $N^{1/5} < z < N^{1/3}$, let T count prime summands p with $q = N - p > 1$ and $P^-(q) > z$. Let L sum $\lambda(q)$ over those candidates, and let H count those with $\Omega(q) = 3$. Five prime factors above z are impossible, so

$$R_z(N) = \frac{T - L}{2} - H, \tag{8}$$

where R_z counts prime complements above the cutoff. A sufficient pointwise inequality is $(T - L)/2 > H$. Obtaining it requires a bound on H at the prime-pair scale; the elementary bound obtained in the programme is a logarithm too weak.

For $z = N^\alpha$ with $1/4 < \alpha < 1/3$, the averaged three-factor count has the coefficient

$$c_3(\alpha) = \frac{1}{6} \iint_{\substack{u,v\ge\alpha\\ u+v\le 1-\alpha}} \frac{du\,dv}{uv(1-u-v)}.$$

This is a classical Buchstab rough-number term, not a new constant [5]. Writing $\alpha = 1/3 - \varepsilon$, the triangle has side length 3ε and

$$c_3(\alpha) = \frac{81}{4}\varepsilon^2 + O(\varepsilon^4).$$

Thus three-factor contamination becomes quadratically small *on average* near the cube-root boundary. The limit in the size parameter is taken first with α fixed. A pointwise inequality still requires new information. A 2026 preprint by Arneth also studies a cubic rough-prime minorant, semiprime cancellation and three-factor contributions [11]; its claims have not been independently verified here.

9 What the graph must know

Two model obstructions help distinguish a promising graph invariant from an insufficient one. First, consecutive even targets need not share a prime summand. If $C(N)$ counts common summands in representations of N and $N+2$, then for $N \equiv 2 \pmod 6$ and $N \geq 8$,

$$C(N) = \mathbf{1}_{\mathbb{P}}(N-3)\,(1+\mathbf{1}_{\mathbb{P}}(N-1))\,.$$

Indeed, for $q = N - p > 3$, the primes q and $q+2$ must have $q \equiv 2 \pmod 3$, forcing the shared prime p to be 3. The other possibility is $q = 3$. Consequently $C(30k+8) = 0$ for every $k \geq 1$. Shared-summand constructions have existing antecedents [13].

A small formal certificate

The accompanying file `GoldbachElementary.lean` was checked with the Lean 4.19.0 kernel. In its explicitly defined elementary divisor-based predicate `PrimeLike`, the theorem `no_adjacent_common_summand` verifies for all $k \geq 1$ that $30k+8$ and $30k+10$ share no prime summand. The file also checks reflection symmetry, two integer-algebra identities underlying the four-cell and parity counts, and the soundness and completeness of a finite trial-division test. That test certifies the six ordered prime summands for $N = 30$. The identities do not formalize the connection between their variables and set cardinalities. This is a certificate for those exact statements, not a verification of Goldbach, the prime-number theorem or the asymptotic deductions in this article.

Second, one can remove a zero-density subset from the primes while making the remaining set's sumset miss arbitrarily long blocks of even integers. To see the mechanism, average prime-pair counts over short windows in $[X, 2X]$, select a window with few total representations, and remove the larger prime in every pair. Repeating this at separated scales removes $o(X/\log X)$ primes but produces growing missing blocks. Hence one-point density and equidistribution in every *fixed* reduced residue class cannot alone force every even sum. Related density results for prime subsets are known [6].

10 A precise analytical bridge

The Fourier and sieve viewpoints meet in the Möbius-twisted distribution of the reflected sequence $\Lambda(n)\mu(N-n)$. Huang and Li proved that eventual binary Goldbach would follow from specified ordinary and Möbius-twisted Elliott–Halberstam estimates whose levels add to more than one. Their corollary identifies a twisted level beyond $1/2$ as a sufficient condition with the complementary ordinary input supplied by Bombieri–Vinogradov [7]. This is a conditional theorem, not an established estimate of the required strength. A 2026 preprint by Moya identifies a potential moving-endpoint omission in the original derivation and claims a repair under the full maximal-in-y hypothesis [10]; that criticism and claimed repair have not been independently verified here. Any use of this route must preserve the stated maximal uniformity.

Recent OpenAI mathematics materials and follow-up manuscripts develop multiplicative-function correlations, divisibility graphs and reflected prime-slot methods [8, 9]. These are valuable sources of ideas. Their exact hypotheses must be inspected individually: fixed-form correlation is not automatically uniform in a moving target N, and a subexponential roughness cutoff is not automatically a power cutoff. At the cube-root scale the aggregate reflected Liouville bias is nonzero, so a proposed zero relative bias would itself be false. A correctly normalized *one-sided* estimate is the appropriate kind of target.

A bridge worth pursuing

Compare the exact residue-class and modulus requirements of Huang–Li with prime-slot and weighted divisibility-graph estimates. Track the moving target, prime weights, the sieve cutoff, all maxima and the size of every error. The goal is an independently checkable lemma, not a similarity of terminology.

11 A programme for the next investigation

The following tasks are deliberately separated so that one failure does not obscure the others.

1. **Audit the deductions.** Check the signed average, moving-cutoff errors, finite-residue density calculation and three-factor integral independently; formalize the elementary identities where feasible.
2. **Search for countermodels.** Test each proposed uniform inequality against prime-like sequences with the same one-point or low-order statistics. Record the weakest assumption whose removal permits failure.
3. **Quantify three-factor contamination.** Decompose H in (8) by factor ranges, then seek a uniform sieve bound with a constant small enough to leave a positive prime margin.
4. **Investigate a softer intermediate theorem.** Study whether the rough-semiprime pair count admits an almost-all asymptotic with the appropriate singular series. A proof would require progression uniformity, bilinear minor-arc control and moving-cutoff estimates. It would still leave exceptional targets.
5. **Build the analytical bridge.** Compare the exact Möbius-twisted hypothesis in Huang–Li with the strongest applicable prime-slot estimates, including every parameter and dependency.

Christopher Yamba is also developing a complementary approach to the additive structure of the primes, intended for a future follow-up publication. Its specific claims and methods are not assumed in the results above.

References

[1] H. A. Helfgott, *The ternary Goldbach conjecture is true*, arXiv:1312.7748, 2013.

[2] T. Oliveira e Silva, S. Herzog and S. Pardi, *Empirical verification of the even Goldbach conjecture and computation of prime gaps up to* $4 \cdot 10^{18}$, Mathematics of Computation 83 (2014), 2033–2060. doi:10.1090/S0025-5718-2013-02787-1.

[3] D. Crişan and R. Erban, *On the counting function of semiprimes*, Integers 21 (2021), arXiv:2006.16491.

[4] D. I. Tolev, *The binary Goldbach problem with arithmetic weights attached to one of the variables*, Acta Arithmetica 142 (2010), 169–178. doi:10.4064/aa142-2-6.

[5] T. Tao, *Probabilistic models and heuristics for the primes*, 2015, rough-number discussion.

[6] A. Alsetri and X. Shao, *Density versions of the binary Goldbach problem*, Acta Arithmetica 218 (2025), 285–295. doi:10.4064/aa240615-19-9.

[7] J.-J. Huang and H. Li, *On the connection between the Goldbach conjecture and the Elliott–Halberstam conjecture*, arXiv:2005.03811.

[8] OpenAI, *Mathematics research repository*, `https://github.com/openai/math`. Statements and formalizations have differing scopes and verification status.

[9] J. Guo, *Follow-ups to the OpenAI mathematics release*, `https://github.com/aster2024/openai-math-followups`. Manuscript claims are used as leads with their stated assumptions, not as automatic Goldbach inputs.

[10] R. Moya, *A Moving-Cut Correction to the Huang–Li Conditional Goldbach Argument*, 2026, Zenodo record 21939906, preprint; the criticism and proposed repair are unverified here.

[11] B. Arneth, *A Cubic Prime Minorant and the Remaining Correlation in Binary Goldbach*, 2026, Zenodo record 23477014, preprint.

[12] G. D. Dura, *Binary Goldbach as a two-cloud complete-sieve survivor problem*, 2026, doi:10.5281/zenodo.20660791, preprint.

[13] OEIS Foundation, *A335045: Minimal common prime of two Goldbach partitions*, `https://oeis.org/A335045`.