The XOR Distance Theorem for Twin Primes

A Binary Topological Characterization of Prime Gaps

Thiago Fernandes Motta Massensini Silva
thiago@massensini.com.br

December 2025

Abstract

We establish an exact formula relating the Hamming distance between twin primes to the 2-adic valuation of their midpoint. Specifically, for any odd prime p such that $p+2$ is also prime, we prove that $\text{xor\_dist}(p, p+2) = v_2(p+1)$, where xor_dist denotes the Hamming distance (number of differing bits) and v_2 is the 2-adic valuation. This result reveals a connection between the arithmetic structure of integers and their binary topology. We provide a complete proof based on the carry propagation mechanism in binary addition, verify the theorem computationally for all 58,980 twin prime pairs up to 10^7, and show that the distribution of xor_dist values follows a geometric distribution with parameter 1/2. Possible extensions to arbitrary prime gaps are briefly discussed.

Keywords: Twin primes, Hamming distance, 2-adic valuation, binary representation, prime gaps

MSC 2020: 11A41, 11A63, 11S80, 94B65

1 Introduction

The study of prime numbers has traditionally employed the arithmetic metric $d(a, b) = |a - b|$ on $\mathbb{Z}$. While this metric captures the "distance traveled" along the number line, it obscures structural information encoded in the binary representation of integers.

Consider two pairs of consecutive integers:

$$\begin{aligned}(126, 127): &\quad 126_{10} = 1111110_2, \quad 127_{10} = 1111111_2, \quad \text{one bit differs}\\ (127, 128): &\quad 127_{10} = 01111111_2, \quad 128_{10} = 10000000_2, \quad \text{all eight bits differ}\end{aligned}$$

Both pairs have arithmetic distance 1, yet their *binary distances* are 1 and 8 respectively. This observation motivates the study of the Hamming distance—which we call the *XOR distance*—as an alternative metric on $\mathbb{Z}_{\geq 0}$.

1.1 Main Result

Our main theorem provides an exact formula for the XOR distance between twin primes:

Theorem 1.1 (Main Theorem). *Let p be an odd prime such that $p+2$ is also prime. Then*

$$\boxed{\text{xor\_dist}(p, p+2) = v_2(p+1)}$$

where $\text{xor\_dist}(a, b) = \text{popcount}(a \oplus b)$ *is the Hamming distance and* $v_2(n) = \max\{k \geq 0 : 2^k \mid n\}$ *is the 2-adic valuation.*

1.2 Significance

This result is remarkable for several reasons:

(i) It provides an *exact* formula, not an asymptotic or probabilistic one.

(ii) It connects two seemingly unrelated concepts: binary topology and p-adic analysis.

(iii) It reveals that twin primes "know about" their midpoint's 2-adic structure.

(iv) It generalizes naturally: the formula holds for *all* odd integers, not just primes.

1.3 Organization

Section 2 establishes notation and preliminary results. Section 3 contains the complete proof of Theorem 1.1. Section 4 analyzes the distribution of XOR distances. Section 5 presents computational verification. Section 6 discusses generalizations and open problems.

1.4 Note on Originality

To our knowledge, this exact identity does not appear in the existing literature on Hamming metrics, additive combinatorics, or p-adic valuations. The argument is elementary but seems to be new.

2 Preliminaries

2.1 Binary Representation

Every non-negative integer n has a unique binary representation:

$$n = \sum_{i=0}^{k} b_i \cdot 2^i, \quad b_i \in \{0, 1\}, \quad b_k = 1 \text{ (for } n > 0)$$

We write $n = (b_k b_{k-1} \cdots b_1 b_0)_2$ and define $\mathrm{bit}_i(n) := b_i$.

Definition 2.1 (Popcount)**.** The *population count* or *Hamming weight* of $n \in \mathbb{Z}_{\geq 0}$ is

$$\mathrm{popcount}(n) := \sum_{i \geq 0} \mathrm{bit}_i(n) = |\{i : \mathrm{bit}_i(n) = 1\}|$$

2.2 XOR Distance

Definition 2.2 (XOR Distance / Hamming Distance)**.** For $a, b \in \mathbb{Z}_{\geq 0}$, the *XOR distance* is

$$\mathrm{xor\_dist}(a, b) := \mathrm{popcount}(a \oplus b)$$

where $\oplus$ denotes bitwise exclusive-or.

Proposition 2.3. *The function* $\mathrm{xor\_dist} : \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0} \to \mathbb{Z}_{\geq 0}$ *is a metric.*

Proof. We verify the metric axioms:

(a) *Identity*: $\mathrm{xor\_dist}(a, a) = \mathrm{popcount}(a \oplus a) = \mathrm{popcount}(0) = 0$.

(b) *Symmetry*: $\mathrm{xor\_dist}(a, b) = \mathrm{popcount}(a \oplus b) = \mathrm{popcount}(b \oplus a) = \mathrm{xor\_dist}(b, a)$.

(c) *Triangle inequality*: For each bit position i,

$$\mathrm{bit}_i(a \oplus c) \leq \mathrm{bit}_i(a \oplus b) + \mathrm{bit}_i(b \oplus c)$$

since XOR satisfies $(a \oplus c) = (a \oplus b) \oplus (b \oplus c)$ and at most one side can contribute to each bit position. Summing over all positions gives the result. □

2.3 2-Adic Valuation

Definition 2.4 (2-Adic Valuation)**.** For $n \in \mathbb{Z}$, $n \neq 0$, the *2-adic valuation* is

$$v_2(n) := \max\{k \in \mathbb{Z}_{\geq 0} : 2^k \mid n\}$$

Equivalently, $v_2(n)$ is the number of trailing zeros in the binary representation of $|n|$.

Example 2.5. $v_2(12) = v_2(1100_2) = 2, \quad v_2(8) = v_2(1000_2) = 3, \quad v_2(7) = v_2(111_2) = 0.$

2.4 Trailing Ones

Definition 2.6 (Trailing Ones)**.** For $n \in \mathbb{Z}_{>0}$, the number of *trailing ones* is

$$\tau(n) := \max\{k \geq 0 : \mathrm{bit}_i(n) = 1 \text{ for all } 0 \leq i < k\}$$

The following lemma is fundamental:

Lemma 2.7. *For any odd integer n, we have $\tau(n) = v_2(n+1)$.*

Proof. Let n be odd with $k = \tau(n)$ trailing ones. Since the bit at position k must be 0 (otherwise we would have more trailing ones), we can write:

$$n = (\cdots b_{k+1} 0 \underbrace{11\cdots1}_{k})_2 = N \cdot 2^{k+1} + (2^k - 1)$$

for some $N \geq 0$, where N encodes the bits at positions $k+1, k+2, \ldots$

Adding 1:

$$n + 1 = N \cdot 2^{k+1} + 2^k = (2N+1) \cdot 2^k$$

Observe that $2N+1$ is *always odd*, regardless of the value of N. Therefore, the largest power of 2 dividing $n+1$ is exactly 2^k, which means:

$$v_2(n+1) = k = \tau(n)$$

Remark: The fact that $(2N+1)$ is odd is precisely what guarantees the carry "stops" at position k. □

3 Proof of the Main Theorem

We now prove Theorem 1.1. The proof proceeds by analyzing the carry propagation in binary addition.

3.1 Structure of Odd Integers

Lemma 3.1. *Let n be an odd positive integer with $k = \tau(n)$ trailing ones. Then n has the binary form:*

$$n = (\cdots b_{k+1} 0 \underbrace{11\cdots1}_{k})_2$$

where $b_{k+1}, b_{k+2}, \ldots$ are arbitrary bits.

Proof. Since n is odd, $\mathrm{bit}_0(n) = 1$. By definition of $\tau(n) = k$, we have $\mathrm{bit}_i(n) = 1$ for $0 \leq i < k$ and $\mathrm{bit}_k(n) = 0$ (otherwise $\tau(n) \geq k+1$). □

3.2 Effect of Adding 2

Lemma 3.2. *Let n be an odd positive integer with $k = \tau(n) \geq 1$ trailing ones. Then:*

$$n = (\cdots b_{k+1} 0 \underbrace{11\cdots1}_{k})_2 \implies n+2 = (\cdots b_{k+1} 1 \underbrace{00\cdots0}_{k-1} 1)_2$$

Special case: *When $k = 1$, the expression reduces to $(\cdots b_2 1\, 1)_2$ (no intermediate zeros).*

Proof. Adding $2 = (10)_2$ to n:

Position 0: $1 + 0 = 1$, no carry. The result bit is 1.

Position 1: $1 + 1 = 10_2$, result bit is 0, carry 1.

Positions 2 to $k-1$: Each has bit 1, plus carry 1, giving $1 + 1 = 10_2$. Result bit is 0, carry propagates.

Position k: Has bit 0, plus carry 1, giving $0 + 1 = 1$. Result bit is 1, *no further carry.*

Positions $> k$: Unchanged (no carry reaches them).

Therefore:

$$n+2 = (\cdots b_{k+1} 1 \underbrace{00\cdots0}_{k-1} 1)_2$$

For $k = 1$: there are no bits between positions 0 and 1, so the result is simply $(\cdots b_2 1\, 1)_2$. □

3.3 Computing the XOR

Lemma 3.3. *Let n be an odd positive integer with $k = \tau(n)$ trailing ones. Then:*

$$n \oplus (n+2) = (0\cdots0 \underbrace{11\cdots1}_{k} 0)_2$$

and consequently $\text{xor\_dist}(n, n+2) = k$.

Proof. From Lemmas 3.1 and 3.2:

$$n = (\cdots b_{k+1} 0 \underbrace{11\cdots1}_{k})_2$$
$$n+2 = (\cdots b_{k+1} 1 \underbrace{00\cdots0}_{k-1} 1)_2$$

Computing XOR bit by bit:

- Position 0: $1 \oplus 1 = 0$ (both have bit 1)
- Positions 1 to $k-1$: $1 \oplus 0 = 1$ (contributes $k-1$ ones)
- Position k: $0 \oplus 1 = 1$ (contributes 1 one)
- Positions $> k$: $b_j \oplus b_j = 0$ (identical bits cancel)

The key observation is that bits above position k are *identical* in n and $n+2$, since no carry propagates past position k. Therefore, their XOR contribution is zero.

The total popcount is:

$$\text{popcount}(n \oplus (n+2)) = 0 + (k-1) + 1 + 0 = k$$ □

3.4 Main Proof

Proof of Theorem 1.1. Let p be an odd prime such that $p+2$ is also prime. Define $k := v_2(p+1)$.

Step 1: Determining trailing ones of p.

Since $v_2(p+1) = k$, the number $p+1$ has exactly k trailing zeros. By Lemma 2.7 (read "in reverse"), this means p has exactly k trailing ones:

$$p = (\cdots b_{k+1} 0 \underbrace{11\cdots1}_{k})_2$$

Step 2: Apply Lemma 3.3.

By Lemma 3.3:

$$\text{xor\_dist}(p, p+2) = k$$

Step 3: Conclude.

Combining Steps 1 and 2:

$$\text{xor\_dist}(p, p+2) = k = v_2(p+1)$$ □

Remark 3.4 (Independence from Primality)**.** The primality of p and $p+2$ is **never used** in the proof. The argument is purely arithmetic-binary, depending only on p being odd. Although the theorem is stated for twin primes, the primality condition plays no role in the proof; the identity holds for all odd integers. The restriction highlights a number-theoretic context of particular interest, where the XOR distance provides a new lens through which to study prime gaps.

3.5 Generalization

Corollary 3.5. *For any odd positive integer n:*

$$\text{xor\_dist}(n, n+2) = v_2(n+1)$$

Proof. The proof of Theorem 1.1 applies verbatim, as it never invokes primality. □

4 Distribution Analysis

4.1 Theoretical Distribution

Theorem 4.1. *Among all odd integers n in $[1, N]$, the proportion with $v_2(n+1) = k$ approaches 2^{-k} as $N \to \infty$.*

Proof. An odd n satisfies $v_2(n+1) = k$ if and only if $n + 1 \equiv 2^k \pmod{2^{k+1}}$, i.e., $n \equiv 2^k - 1 \pmod{2^{k+1}}$.

Among odd integers, this occurs with density $1/2^k$:

$$\lim_{N\to\infty} \frac{|\{n \le N : n \text{ odd}, v_2(n+1) = k\}|}{|\{n \le N : n \text{ odd}\}|} = \frac{1}{2^k}$$

□

Corollary 4.2. *The distribution of* $\text{xor\_dist}(n, n+2)$ *over odd n is geometric with parameter $1/2$:*

$$P(\text{xor\_dist} = k) = 2^{-k}, \quad k = 1, 2, 3, \ldots$$

4.2 Distribution for Twin Primes

For twin primes $(p, p+2)$, heuristically we expect the same geometric distribution, since:

1. The condition $v_2(p+1) = k$ is "independent" of primality.
2. There is no known correlation between 2-adic structure and primality.

Our computational results (Section 5) confirm this heuristic.

4.3 Statistical Properties

For the geometric distribution with $p = 1/2$:

$$\mathbb{E}[\text{xor\_dist}] = \sum_{k=1}^{\infty} k \cdot 2^{-k} = 2$$

$$\text{Var}[\text{xor\_dist}] = \sum_{k=1}^{\infty} k^2 \cdot 2^{-k} - 4 = 2$$

$$H[\text{xor\_dist}] = -\sum_{k=1}^{\infty} 2^{-k} \log_2(2^{-k}) = 2 \text{ bits}$$

5 Computational Verification

5.1 Methodology

We implemented the verification in Python using:

1. Sieve of Eratosthenes for prime generation
2. Direct computation of $\text{xor\_dist}(p, p+2) = \text{popcount}(p \oplus (p+2))$
3. Direct computation of $v_2(p+1)$ via bit operations

5.2 Results

Limit	Twin Pairs	Verified	Success Rate
10^5	1,224	1,224	100.0000%
10^6	8,169	8,169	100.0000%
10^7	58,980	58,980	100.0000%

Table 1: Verification results for Theorem 1.1

5.3 Distribution Fit

We performed a chi-squared goodness-of-fit test for the geometric distribution:
Chi-squared statistic: $\chi^2 = 5.15$ with 12 degrees of freedom.
Critical value at $\alpha = 0.01$: $\chi^2_{0.01,12} = 26.22$.
Conclusion: The data is consistent with Geom(1/2) at the 99% confidence level.

k	Observed	Expected	Obs. Freq.	Theo. Freq.
1	29,482	29,490	0.4999	0.5000
2	14,739	14,745	0.2499	0.2500
3	7,468	7,372	0.1266	0.1250
4	3,647	3,686	0.0618	0.0625
5	1,827	1,843	0.0310	0.0312
6	905	922	0.0153	0.0156
7	469	461	0.0080	0.0078

Table 2: Distribution of xor_dist for 58,980 twin prime pairs up to 10^7

5.4 Extreme Cases

The largest XOR distance found among twins up to 10^7:

$$p = 786431, \quad p + 2 = 786433, \quad \text{xor\_dist} = v_2(786432) = 18$$

Note that $786432 = 2^{18} \cdot 3$, confirming the formula.

6 Extensions and Open Problems

6.1 General Gap Formula

For gap $g = p' - p$ between consecutive primes, a natural question is:

Find a formula for $\text{xor\_dist}(p, p + g)$ in terms of p, g, and their arithmetic properties.

Our preliminary investigations suggest that for $g > 2$, the formula involves the interaction between the binary representations of p and g, and no simple closed form exists.

6.2 p-Adic Generalization

For odd prime q and primes $p, p+q$ (Sophie Germain configuration), there may exist a relationship between $\text{xor\_dist}(p, p + q)$ and v_q-adic properties of p.

6.3 Connection to Riemann Hypothesis

The XOR distance provides a natural weighting scheme for prime-related sums. We introduce the notation $M(s)$ for the weighted Dirichlet series:

$$M(s) := \sum_{(p,p+2) \text{ twin}} v_2(p + 1) \cdot p^{-s}$$

Preliminary numerical experiments suggest that the analytic properties of $M(s)$ may be of interest. A systematic study of this series and its potential connections to classical zeta functions is left for future work.

6.4 Cryptographic Applications

The XOR distance metric may have applications in:

1. Prime selection for cryptographic protocols
2. Analysis of prime-based pseudorandom generators
3. Side-channel analysis of primality testing algorithms

7 Conclusion

We have established an exact formula relating the Hamming distance between twin primes to the 2-adic valuation of their midpoint:

$$\text{xor\_dist}(p, p+2) = v_2(p+1)$$

This result reveals a previously unnoticed connection between binary topology and p-adic number theory. The proof is elementary, relying only on the mechanics of carry propagation in binary addition, yet the result has deep implications for our understanding of prime structure.

The theorem generalizes to all odd integers, and the distribution of XOR distances follows a geometric distribution with parameter $1/2$. Computational verification confirms the theorem for all 58,980 twin prime pairs up to 10^7 with 100% accuracy.

Future work will explore generalizations to arbitrary gaps, connections to the Riemann hypothesis, and potential applications in cryptography and coding theory.

References

[1] G.H. Hardy and J.E. Littlewood, *Some problems of 'Partitio Numerorum' III: On the expression of a number as a sum of primes*, Acta Mathematica **44** (1923), 1–70.

[2] B. Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsberichte der Berliner Akademie (1859), 671–680.

[3] N. Koblitz, *p-adic Numbers, p-adic Analysis, and Zeta-Functions*, Graduate Texts in Mathematics, Springer (1984).

[4] R.W. Hamming, *Error detecting and error correcting codes*, Bell System Technical Journal **29**(2) (1950), 147–160.

[5] J. Shallit, *Origins of the analysis of the Euclidean algorithm*, Historia Mathematica **21**(4) (1994), 401–419.