V1.0.0

Prime States as Intrinsic Basis States: Entanglement Bounds, Preparation Complexity, and the Limits of Classical Emulation

Rowan Brad Quni-Gudzinas

QNFO-ULA

ABSTRACT

A recurring proposal in quantum-architecture design is to treat the prime numbers not as parameters of an algorithm but as the fundamental addressable units of a register: a "Prime state" that superposes all primes below 2^n. The specific gap this paper addresses is quantitative: no compact derivation exists that a system architect can use to compare the preparation cost, entanglement content, and classical emulability of such a prime-basis register against a standard computational-basis register. Using the Prime-state construction of Latorre and Sierra, in which a Grover oracle implementing the Miller-Rabin test prepares $|\mathbb{P}_n\rangle = \pi(2^n)^{-1/2} \sum_{p<2^n} |p\rangle$, we derive: (i) the prime-state population $\pi(2^n) \approx 2^n/(n \ln 2)$, giving $75\,640$ for $n = 20$ under the PNT approximation; (ii) a Grover preparation cost of $(\pi/4)\sqrt{n \ln 2}$ oracle calls, i.e. 5.23 calls at $n = 64$, against 2^{64} classical candidates; (iii) an entanglement entropy ceiling of $\log_2 \pi(2^n)$ bits across any bipartition, 16.21 bits at $n = 20$; and (iv) a coprimality-derived collision floor of $6/\pi^2 \approx 0.6079$ for random register pairs. We further show that exact classical emulation of the full n-qubit state vector requires 2^n amplitudes (17.59 TB at $n = 40$, double precision) while the sparse prime representation needs only $\pi(2^n)$, and we identify the regime boundary this creates. The results convert a conjectural architecture idea into checkable cost and entropy bounds.

PUBLISHER CLASSIFICATION High-quality scholarly research (QNFO publisher classification for open-access, versioned research with a public revision history and reader correction channel; this is not a claim of independent peer review.)

PEER-REVIEW STATUS Not independently peer-reviewed

CITE AS Quni-Gudzinas, R. B. (n.d.). Prime States as Intrinsic Basis States: Entanglement Bounds, Preparation Complexity, and the Limits of Classical Emulation. QNFO Preprint v1.0.0.

1. Introduction

Engineers building quantum processors today face a resource problem that determines what is computable at all: every qubit is expensive, every error-correction cycle is expensive, and the choice of which states a register can natively represent decides how much of that budget survives. Surface-code error correction demands physical-to-logical qubit ratios of roughly 1000:1, and the fragility of continuously parametrized qubit states has been argued to be intrinsic to the continuous real-number framework rather than merely technical [9]. If a register could instead treat a structured, discrete set of classical values — the primes — as its intrinsic basis, the preparation, verification, and robustness budgets of the machine could change in kind, not just in degree. This is the practical stake of the "Prime state" proposal: a quantum superposition of all primes below 2^n, preparable by Grover's algorithm with a Miller-Rabin oracle, whose entanglement measures encode number-theoretic counting functions [3].

Who faces this problem? Designers of quantum algorithms for number-theoretic tasks (primality distribution, twin-prime counting, Goldbach verification), and, on the classical side, anyone who must decide whether a claimed prime-basis speedup can be replicated without quantum hardware. Today the first group has no closed-form cost model for prime-state preparation at register widths of engineering interest, and the second group has no stated boundary between register widths where a prime state is exactly simulable on a laptop and widths where it is not. If the derivations in this

paper hold, a designer can read off, for a given n, the oracle budget, the achievable entanglement ceiling, and the classical emulation memory, before committing hardware; and a classical skeptic can locate exactly the register width beyond which emulation fails.

The starting point is the construction of [3]: for $n > 1$ qubits,

$$|\mathbb{P}_n\rangle \equiv \frac{1}{\sqrt{\pi(2^n)}} \sum_{p \in \text{primes}(2^n)} |p\rangle,$$

where each prime is written in binary across the computational basis and $\pi(x)$ is the prime-counting function, with $\pi(100) = 25$ and Gauss-law asymptotics $\pi(x) \sim \text{Li}(x) \sim x/\log x$ [3]. The same work shows the state can be built efficiently with a primality circuit, generalized to Twin Prime states whose measurement probabilities count twin primes, and combined with the Quantum Fourier Transform to count primes within an error smaller than the fluctuations allowed by the Riemann hypothesis [3]. What [3] leaves open, and what this paper supplies, is the architect's arithmetic: explicit population, cost, entropy, and memory numbers as functions of n, with every step shown.

The broader context matters in two directions. First, prime structure is not an arbitrary choice of basis: a long line of number-theoretic work shows that primes carry rigid, provable distributional constraints. Bertrand-type guarantees give primes in short intervals — the interval $[x, x + x^{0.525}]$ contains a prime for sufficiently large x [4] — and the conjecture that every n consecutive integers between n^2 and $(n+1)^2$ contain a prime has been verified for every $n \leq 1\,193\,806\,023$ and is equivalent to Legendre's, Brocard's, Andrica's, and Oppermann's conjectures [4]. Structural constructions such as the matrix online characteristic number, whose properties are proved by induction (Theorem 1 of [1]) and applied to the Goldbach, Polignac, and twin prime conjectures, and the cluster number set of [7], which ties the minimum intersection count of the first n column sets to the twin prime pairs $(q, q+2)$, give combinatorial handles on exactly the subseries that the Twin Prime state of [3] measures. Second, there is a live caution: numerical patterns in prime data can be artifacts of unreachable scales. The Euler-Kronecker analysis of [6] shows that conjectures built on numerics can be blind to $\log\log\log$ terms and fail on very thin subsequences, with the term $S(q)$ bounded by $S(q) \leq 45$; and the continued fraction $u = [0; 2, 3, 5, 7, 11, 13, \ldots] = 0.432332087185902868909253793241999963705110896 88\ldots$, computed from all 1229 primes up to 10000, belongs to the measure-zero set of exceptions to the theorems of Khinchin and Lévy [5]. Any claim that prime-basis registers have "generic" or "random-like" entanglement must survive these cautions.

This paper is analytical throughout: it performs no experiment, no simulation on hardware, and no survey. Its method is derivation from stated inputs, in the cost-accounting discipline exemplified by contract-style verification of quantum simulators, where every number is either computed with shown arithmetic or labeled a projection with stated assumptions [10].

The contributions are:

- A derivation of the prime-state population $\pi(2^n)$ under the Gauss-law approximation, with numeric values at $n = 20, 40, 64$ (Section 4.1).
- A derivation of the Grover preparation cost $(\pi/4)\sqrt{n \ln 2}$ oracle calls and its comparison with the classical candidate scan, giving a cost ratio exponential in n (Section 4.2).
- An entanglement entropy ceiling $\log_2 \pi(2^n)$ bits valid for any bipartition of the register, with numeric values (Section 4.3).
- A coprimality collision floor $6/\pi^2 \approx 0.6079$ for pairs of independent register values, derived from the Euler product (Section 4.4).
- A classical emulability boundary: full state-vector memory $2^n \times 16$ bytes versus sparse memory $\pi(2^n) \times 16$ bytes, with the crossover width computed (Section 4.5).

- A discussion of which observations would falsify these bounds, and of the opposing ontological position that primes are spectral artifacts of a continuous substrate rather than fundamental discrete entities [11] (Sections 6 and 2).

The entropy and cost bounds derived here also bear, at the level of representation choice, on problems outside quantum architecture: any field that must choose between a microscopic and a coarse-grained description of a structured discrete system faces the same trade-off between fidelity and computational cost that the prime state versus computational-basis comparison exhibits, a trade-off formalized as predictive efficiency in [12].

2. Background and Related Work

The central prior work is the Prime-state construction of Latorre and Sierra [3]. That paper defines, for $n > 1$ qubits, the normalized superposition $|\mathbb{P}_n\rangle$ over all primes below 2^n, notes that all terms are orthogonal and that normalization involves $\sqrt{\pi(2^n)}$, and gives the $n = 3$ example explicitly [3]. It proposes building the state with Grover's algorithm whose oracle is a quantum implementation of the classical Miller-Rabin primality test, generalizes the construction to a Twin Prime state by adding 2 to each prime and re-applying the primality circuit so that measuring the ancilla yields the probability of finding a twin, and shows that the Prime state verifies Goldbach's conjecture and that the combination of the primality circuit with the Quantum Fourier Transform counts primes within an error smaller than the fluctuations allowed by the Riemann hypothesis [3]. It shows that the entanglement properties of the Prime state are directly related to counting functions of subseries of prime numbers, such as twin primes [3]. This paper takes up that entanglement question, but in the form an architect needs: explicit upper bounds and numeric values, not qualitative remarks.

The distributional guarantees that make the prime basis structured rather than generic come from analytic number theory. The work of [4] proves that the statement "every sequence of n consecutive integers greater than n^2 and smaller than $(n+1)^2$ contains a prime" holds for every $n \leq 1\,193\,806\,023$, and that a positive answer for all n would imply Legendre's, Brocard's, Andrica's, and Oppermann's conjectures; it also records that Baker, Harman, and Pintz proved the interval $[x, x + x^{0.525}]$ contains a prime for sufficiently large x, and that under the Riemann hypothesis the exponent improves to $1/2 + \varepsilon$ [4]. For a prime-basis register this means the basis states are guaranteed to appear with bounded gaps at every scale — a locality property no random subset of $\{0, \ldots, 2^n - 1\}$ possesses. The equivalence theorems of [4] (Theorems 4.2 and 5.10, expressing the conjecture as $\pi[n + g(n), n + f(n) + g(n)] \geq 1$ for all n) give explicit counting-function forms that the entanglement measures of a Prime state would inherit, since [3] shows entanglement measures encode counting functions of prime subseries.

Combinatorial constructions of the twin-prime structure supply the second pillar. The matrix online characteristic number framework of [1] studies a class of matrix sequences built from consecutive odd numbers and proves their online characteristic properties by mathematical induction (Theorem 1), with applications to the Goldbach, Polignac, and twin prime conjectures [1]. The cluster number set construction of [7] proves that the number of set elements in the intersection of the first n column sets decides the twin prime pairs $(q, q+2)$, addressing both the structure of twin primes and their infinitude [7]. Both constructions are relevant here because the Twin Prime state of [3] measures exactly the twin-prime counting function $\pi_2(x)$; a prime-basis architecture whose entanglement encodes π_2 can be analyzed with the inductive matrix and cluster tools of [1] and [7] rather than treated as an unstructured superposition.

A third pillar is the caution literature on numerical inference about primes. The cyclotomic analysis of [6] studies Kummer's conjecture on the first factor of the class number of $\mathbb{Q}(\zeta_q)$ and Ihara's conjecture on the positivity of the Euler-Kronecker constant γ_q, and warns that computers "can barely spot $\log\log$ terms and are certainly blind to the $\log\log\log$ terms," so that conjectures can be false on very thin subsequences while holding up to gigantic values [6]. Its equation (16) expresses γ_q through the term $S(q)$, contributions of prime ideals above rational primes with $\mathrm{ord}_p(q) \geq 2$, bounded by $S(q) \leq 45$ using estimates for linear forms in logarithms [6]. The continued-fraction study of [5] computes, from all 1229 primes up to 10000, the value

$u = 0.432332087185902868909253793241999963705110896 88\ldots$, which lies within $2.712\ldots \times 10^{-7}$ of Rényi's parking constant $m_{\mathrm{R}} = (1 - e^{-2})/2$ yet belongs to the measure-zero exception set of the Khinchin and Lévy theorems [5]. Both works teach the same lesson for prime-state engineering: apparent regularities in prime-derived quantities (including entanglement entropies computed at small n) may not persist, and any architectural claim resting on small-n numerics must be backed by a bound, as the bounds in Section 4 are.

Elementary structural facts about primes also enter. The $6m$ theorem of [2] states that $P = 6^{m+1} \cdot N - 1$ is prime for any $1 < N \leq 13$, $N \neq 8$, $N \neq i^{m+1} \bmod (6i + 1)$ with $i \in \mathbb{Z}^+$ and m odd, and extends conditionally to $N > 13$ [2]. While narrow, this illustrates that prime sets admit explicit congruence parametrizations, which is the mechanism a Miller-Rabin oracle exploits. The totient analysis of [8] proves unconditionally that for each $\ell \geq 1$ the difference $\varphi(p - \ell) - \varphi(p + \ell)$ is positive for 50% of odd primes and negative for 50%, resolving a bias that for twin primes was believed to run as high as 98% to 2% on numerical evidence [8]. This exact symmetry is the analytic prototype for the collision-balance argument in Section 4.4: prime-adjacent arithmetic quantities can be provably balanced where numerics suggest bias.

On the systems side, three works frame the engineering question. The deductive-verification study of discrete-time quantum-walk simulators [10] separates assumptions from proven results, separates its assumptions from its proven results and counts costs in elementary operations rather than wall-clock time [10]. This paper adopts the same accounting discipline for prime-state costs. The Alpha Pi project [9] motivates the search for discrete intrinsic bases by arguing that qubit fragility stems from the continuous real-number field underlying standard quantum mechanics, that surface codes require physical-to-logical ratios of about 1000:1, and that classical filtering hits a signal-processing ceiling against fundamental quantum back-action [9]; a prime basis is one concrete answer to the question that motivation raises. Against the whole enterprise stands the ontological position of [11], which argues through five mathematical facts — base-independence of primality, Gödelian incompleteness of Peano arithmetic, Tennenbaum's non-categoricity theorem, structural dependency of primality on the integer ring, and the Montgomery-Odlyzko correspondence — that primes arise by spectral projection from a continuous quantum-geometric substrate and "cannot be fundamental discrete entities" [11]. If that thesis is right, a prime basis is a derived, engineered basis rather than an intrinsic one; Section 6 discusses how the cost and entropy bounds of this paper are compatible with either reading, since they concern the information content of the state, not its ontology. Finally, the representational-scale analysis of [12] defines predictive efficiency as the ratio of predictive power to computational cost and argues that the optimal description of a system for a resource-bounded observer maximizes this ratio [12]; Section 4.5's emulability boundary is precisely such an efficiency computation for the prime-state representation.

3. Methods

This is an analytical paper. It performs no experiment, no hardware measurement, no simulation on a quantum or classical device, and no survey; every number below is either derived in Section 4 with shown arithmetic or explicitly labeled a projection with stated assumptions. The method has four components.

Model. The object of study is the n-qubit Prime state of [3],

$$|\mathbb{P}_n\rangle = \frac{1}{\sqrt{\pi(2^n)}} \sum_{p \in \mathrm{primes}(2^n)} |p\rangle,$$

with $n > 1$, primes written in binary as $p = p_0 2^0 + p_1 2^1 + \cdots + p_{n-1} 2^{n-1}$ and mapped to computational-basis states $|p_{n-1}, \ldots, p_1, p_0\rangle$ [3]. The register is ideal: no noise, no decoherence, no gate error. This idealization is deliberate; the bounds derived are information-theoretic ceilings that real hardware can only lower.

Derivation rules. Each quantitative claim is derived from stated inputs. The inputs are: the Gauss law $\pi(x) \sim \mathrm{Li}(x) \sim x/\log x$ [3]; the Grover optimality condition for counting M marked items among N; the Euler product identity $\prod_p (1 - p^{-2}) = 1/\zeta(2)$; and the bit-width memory model in which one double-precision complex amplitude occupies 16 bytes. Where a number stands in for data no source gives, it is labeled an assumption in place.

Cost model. Following the discipline of [10], costs are counted in elementary operations (oracle calls, gate-level exponentiation steps, memory words), not wall-clock time; any conversion to time is a projection with an assumed operation rate. The Miller-Rabin oracle is treated as a unit-cost primitive per Grover iteration, with its internal gate cost estimated separately as a projection.

Comparators. The prime basis is compared against two references: the full computational basis (all 2^n states, uniform superposition) and the classical sparse representation (one amplitude per prime). The comparison metric is the predictive-efficiency style ratio of representational fidelity to resource cost [12], instantiated here as memory ratio and preparation-cost ratio.

Scope. All results concern the ideal state and its classical description. Claims about physical robustness (error rates, decoherence of a prime-basis register) are outside the derivations and are flagged as open questions in Section 6.

4. Analysis

4.1 Prime-state population

The number of basis states in $|\mathbb{P}_n\rangle$ is $\pi(2^n)$. The Gauss law gives $\pi(x) \sim x/\log x$ [3]. Taking the natural logarithm, $\log(2^n) = n \ln 2$ with $\ln 2 = 0.693147\ldots$, so

$$\pi(2^n) \approx \frac{2^n}{n \ln 2}.$$

We compute this for three register widths.

- $n = 20$: $2^{20} = 1\,048\,576$; $n \ln 2 = 20 \times 0.693147 = 13.86294$; $\pi(2^{20}) \approx 1\,048\,576 / 13.86294 = 75\,640.5$. The PNT approximation gives $\approx 75\,640$; the true value is higher because $\mathrm{Li}(x) > x/\log x$ in this range, so we treat $75\,640$ as the Gauss-law estimate and note it underestimates $\pi(2^{20})$.
- $n = 40$: $2^{40} = 1\,099\,511\,627\,776$; $n \ln 2 = 27.72588$; $\pi(2^{40}) \approx 1.09951 \times 10^{12} / 27.72588 = 3.9644 \times 10^{10}$.
- $n = 64$: $2^{64} = 1.84467 \times 10^{19}$; $n \ln 2 = 44.36142$; $\pi(2^{64}) \approx 1.84467 \times 10^{19} / 44.36142 = 4.1580 \times 10^{17}$.

The fraction of basis states that are prime is

$$f_n = \frac{\pi(2^n)}{2^n} \approx \frac{1}{n \ln 2}.$$

At $n = 20$: $f_{20} \approx 1/13.86294 = 0.07213$; at $n = 64$: $f_{64} \approx 1/44.36142 = 0.02254$. The prime basis is thus a 7.2% slice of the computational basis at $n = 20$ and a 2.3% slice at $n = 64$.

4.2 Grover preparation cost

Grover's algorithm produces a marked uniform superposition of M items among $N = 2^n$ items in approximately

$$G(n) = \frac{\pi}{4}\sqrt{\frac{N}{M}} = \frac{\pi}{4}\sqrt{\frac{2^n}{\pi(2^n)}} \approx \frac{\pi}{4}\sqrt{n \ln 2}$$

oracle calls, using the Gauss-law substitution of Section 4.1. This is the regime in which [3] asserts efficient preparation via the Miller-Rabin oracle.

- $n = 20$: $\sqrt{20 \times 0.693147} = \sqrt{13.86294} = 3.7233$; $G(20) = 0.785398 \times 3.7233 = 2.924$ oracle calls.
- $n = 64$: $\sqrt{64 \times 0.693147} = \sqrt{44.36142} = 6.6604$; $G(64) = 0.785398 \times 6.6604 = 5.231$ oracle calls.

(Recomputing the abstract's headline: $0.785398 \times 6.6604 = 5.2312$; we state 5.23 oracle calls at $n = 64$.)

The classical comparator is the cost of enumerating candidates and running Miller-Rabin on each. We assume $k = 64$ Miller-Rabin rounds per candidate (an assumption; no source fixes k) and, as a projection, one modular exponentiation per round costing $O(n^3)$ bit operations with the standard schoolbook bound, i.e. $\approx c\,n^3$ with $c = 1$ as the unit convention. The classical scan cost is then

$$C_{\text{cl}}(n) \approx 2^n \cdot k \cdot n^3.$$

At $n = 64$: $2^{64} \times 64 \times 262\,144 = 1.84467 \times 10^{19} \times 1.67772 \times 10^7 = 3.0946 \times 10^{26}$ modular-exponentiation-level operations. The quantum preparation, at 5.23 oracle calls plus the oracle's internal gate cost, replaces a 10^{26}-scale classical scan with a polylogarithmic number of oracle invocations; the ratio of oracle calls to classical candidates is

$$\frac{G(64)}{2^{64}} = \frac{5.231}{1.84467 \times 10^{19}} = 2.836 \times 10^{-19}.$$

The oracle itself is not free: a quantum Miller-Rabin implementation performs modular exponentiation on n-qubit registers, and we project its gate cost as $O(k\,n^3)$ gates per call by the same schoolbook convention, i.e. $\approx 64 \times 262\,144 = 1.678 \times 10^7$ gates per oracle call at $n = 64$ under the unit convention. Total projected gate count: $5.23 \times 1.678 \times 10^7 = 8.77 \times 10^7$ gates. This is a projection with the stated assumptions ($k = 64$, schoolbook multiplication, unit-cost convention); it is not a circuit-level result.

4.3 Entanglement entropy ceiling

For any bipartition of the n qubits into A and B, the entanglement entropy S_A of a pure state $|\psi\rangle$ obeys $S_A \le \log_2 r$, where r is the rank of the reduced density matrix, which is at most the number of nonzero Schmidt coefficients, which is at most the number of computational-basis terms in $|\psi\rangle$. Since $|\mathbb{P}_n\rangle$ has exactly $\pi(2^n)$ nonzero terms,

$$S_A \le \log_2 \pi(2^n) \approx \log_2\!\left(\frac{2^n}{n \ln 2}\right) = n - \log_2(n \ln 2).$$

Numeric values:

- $n = 20$: $\log_2(75\,640.5) = \ln(75\,640.5)/\ln 2 = 11.2339/0.693147 = 16.206$ bits. Equivalently $20 - \log_2(13.86294) = 20 - 3.7935 = 16.2065$ bits.
- $n = 64$: $64 - \log_2(44.36142) = 64 - 5.4729 = 58.527$ bits.

Two comparisons calibrate this ceiling. First, a uniform superposition over all 2^n basis states is the product state $2^{-n/2}\sum_x |x\rangle$, which has $S_A = 0$ for every bipartition; restricting to primes destroys the product structure and is the sole source of entanglement. Second, the ceiling sits below the maximally mixed bound n bits by $\log_2(n \ln 2)$: at $n = 64$ the deficit is 5.47 bits, i.e. the ceiling is 91.5% of the maximal value ($58.527/64 = 0.9145$). Whether the actual entropy of $|\mathbb{P}_n\rangle$ approaches this ceiling is a property of the prime distribution that this paper leaves open; the ceiling itself is unconditional.

A caution applies, in the spirit of [6] and [5]: small-n numerics of the actual entropy could suggest a trend that fails at larger n, exactly as numerics suggested a 98%-to-2% totient bias that [8] proved is exactly 50%–50% once only p is assumed prime. The ceiling derived here is proof-based and immune to that failure mode; any claim that the entropy saturates the ceiling is not, and is flagged as such.

4.4 Coprimality collision floor

For two independent integers a, b drawn from a large range, the probability that $\gcd(a,b) = 1$ is

$$P(\gcd = 1) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

Arithmetic: $\pi^2 = 9.86960$; $6/9.86960 = 0.607927$. So $P \approx 0.6079$, and the collision-inducing event (a common prime factor) has probability $1 - 0.607927 = 0.392073$.

For a prime-basis architecture, this floor enters wherever two register values interact through a shared factor: modular hashing, gcd-based routing, or any coprimality gate. The relevant structural point is the balance theorem of [8]: for each $\ell \geq 1$, $\varphi(p-\ell) - \varphi(p+\ell)$ is positive for 50% of odd primes and negative for 50%, unconditionally [8]. Applied to a prime register interacting with shifted neighbors, this says the sign of the totient asymmetry is exactly balanced across the basis, so no systematic drift accumulates from totient asymmetries; deviations that numerics suggest (the twin-prime case, where the inequality $\varphi(p-1) \geq \varphi(p+1)$ appears to hold for an overwhelming majority of twin primes, with the Bateman–Horn-conditional bound of at least 65.13% versus 0.47% reversed, and numerical evidence near 98% to 2% [8]) are bias in subseries, not in the full prime basis. An architecture-level error model built on totient asymmetry therefore has zero mean over the prime basis, by the theorem of [8].

4.5 Classical emulability boundary

The full state vector of an n-qubit register has 2^n complex amplitudes; the sparse prime representation stores one amplitude per prime, i.e. $\pi(2^n)$ amplitudes. Under the memory model of Section 3 (one double-precision complex amplitude $= 16$ bytes), the dense and sparse memory footprints are

$$M_{\text{dense}}(n) = 2^n \times 16 \text{ bytes}, \qquad M_{\text{sparse}}(n) = \pi(2^n) \times 16 \text{ bytes} \approx \frac{2^n \times 16}{n \ln 2} \text{ bytes}.$$

The sparse-over-dense ratio is $M_{\text{sparse}}/M_{\text{dense}} \approx 1/(n \ln 2)$, so the dense representation is always larger for $n > 1$; the operative question is not where the two cross each other (they never do) but where each crosses a fixed storage budget. We take a 1 TB budget, $B = 10^{12}$ bytes, as the reference point of a large single workstation.

- Dense boundary: $2^n \times 16 > 10^{12}$ requires $2^n > 6.25 \times 10^{10}$. Since $2^{35} = 3.436 \times 10^{10}$ and $2^{36} = 6.872 \times 10^{10}$, the dense representation exceeds 1 TB from $n = 36$ onward.
- Sparse boundary: $\pi(2^n) \times 16 > 10^{12}$ requires, under the Gauss-law estimate, $2^n > 6.25 \times 10^{10} \times n \ln 2$. At $n = 40$: the right-hand side is $6.25 \times 10^{10} \times 27.72588 = 1.733 \times 10^{12} > 2^{40} = 1.0995 \times 10^{12}$, so the sparse representation still fits. At $n = 41$: $6.25 \times 10^{10} \times 41 \times 0.693147 = 6.25 \times 10^{10} \times 28.41903 = 1.776 \times 10^{12} < 2^{41} = 2.199 \times 10^{12}$, so the sparse representation exceeds 1 TB from $n = 41$ onward.

The emulability window under this budget is therefore $n = 36, 37, 38, 39, 40$: five register widths where the dense state vector no longer fits a 1 TB machine but the sparse prime representation does. Numeric footprints at the two anchor widths:

- $n = 40$: $M_{\text{dense}} = 1\,099\,511\,627\,776 \times 16 = 17\,592\,186\,044\,416$ bytes ≈ 17.59 TB; $M_{\text{sparse}} \approx 3.9644 \times 10^{10} \times 16 = 6.343 \times 10^{11}$ bytes ≈ 634 GB.
- $n = 64$: $M_{\text{dense}} = 1.84467 \times 10^{19} \times 16 = 2.951 \times 10^{20}$ bytes $\approx 2.95 \times 10^{8}$ TB; $M_{\text{sparse}} \approx 4.1580 \times 10^{17} \times 16 = 6.653 \times 10^{18}$ bytes $\approx 6.65 \times 10^{6}$ TB. Even the sparse representation is out of reach of any single machine at $n = 64$; the boundary is a window, not a permanent refuge.

Two caveats bound the usefulness of this window. First, the sparse representation is exact for the state's description but not for arbitrary operations on it: a generic n-qubit unitary maps the prime support to a dense support, so emulation of general circuits inside the window still costs dense-scale work; the window applies to state storage, measurement statistics of the prime basis, and operations that preserve or are diagonal in the prime support. Second, the window's location scales logarithmically with the budget: doubling B to 2×10^{12} bytes moves the dense boundary from $n = 36$ to $n = 37$ and the sparse boundary from $n = 41$ to $n = 42$, i.e. one width per doubling of storage, because both boundaries grow as 2^n up to the polynomial factor $n \ln 2$.

5. Results

All numbers below are computed in Section 4 with shown arithmetic, or are projections with the assumptions stated there ($k = 64$ Miller-Rabin rounds, schoolbook $O(n^3)$ modular exponentiation, unit-cost convention, 16 bytes per double-precision amplitude, 1 TB reference budget).

- Prime-state population (Section 4.1): $\pi(2^{20}) \approx 75\,640$ (Gauss-law estimate, a lower bound on the true value in this range); $\pi(2^{40}) \approx 3.9644 \times 10^{10}$; $\pi(2^{64}) \approx 4.1580 \times 10^{17}$. The prime fraction of the basis is $f_{20} \approx 0.07213$ and $f_{64} \approx 0.02254$.
- Preparation cost (Section 4.2): $G(20) = 2.924$ oracle calls; $G(64) = 5.23$ oracle calls. The ratio of oracle calls to classical candidates at $n = 64$ is 2.836×10^{-19}. The projected total gate count at $n = 64$ is 8.77×10^7 gates under the stated assumptions.
- Entanglement ceiling (Section 4.3): $S_A \le \log_2 \pi(2^n) \approx n - \log_2(n \ln 2)$ bits for any bipartition; 16.21 bits at $n = 20$ and 58.527 bits at $n = 64$, the latter being 91.5% of the maximal 64 bits.
- Collision floor (Section 4.4): $P(\gcd = 1) = 6/\pi^2 = 0.607927$ for independent register pairs; the complementary shared-factor probability is 0.392073. The totient-asymmetry sign balance over the prime basis is exactly 50%–50% by the unconditional theorem of [8].
- Emulability boundary (Section 4.5): dense memory 17.59 TB at $n = 40$ versus sparse memory ≈ 634 GB; under a 1 TB budget the exact-emulation window is $36 \le n \le 40$, closing at $n = 41$; each doubling of the budget widens the window by one register width.

6. Discussion

The results are ceilings and floors for an idealized object, and their interpretation must respect that status. The entropy bound $S_A \le \log_2 \pi(2^n)$ is unconditional: it follows from the rank bound on the reduced density matrix and the term count of the state, and no property of the primes can violate it. What the bound does not establish is saturation. Whether the actual bipartite entropy of $|\mathbb{P}_n\rangle$ approaches $\log_2 \pi(2^n)$ depends on how evenly the Schmidt weights are distributed, which is a property of the prime distribution that [3] explicitly leaves open. A prime state whose Schmidt spectrum is dominated by a few large weights could have entropy far below the ceiling while still satisfying the bound. The same asymmetry applies to the preparation cost: $G(n) \approx (\pi/4)\sqrt{n \ln 2}$ counts oracle calls in the ideal Grover regime with a unit-cost Miller-Rabin oracle, and the projected gate count of 8.77×10^7 at $n = 64$ rests on the assumed $k = 64$ rounds and schoolbook exponentiation; a circuit-level implementation with ancilla overhead, uncomputation, and error correction would raise it by factors this paper does not estimate.

The caution literature bears directly on what these numbers license. The cyclotomic analysis of [6] shows that conjectures built on numerics can be false on very thin subsequences while holding to gigantic values, because computations are blind to $\log\log\log$ terms; and the continued-fraction study of [5] shows that a prime-derived constant can sit within $2.712\ldots \times 10^{-7}$ of a famous constant (Rényi's parking constant) and still belong to a measure-zero exception set. Any claim that the entanglement of $|\mathbb{P}_n\rangle$ is "generic" or that small-n entropy values extrapolate would be exposed to exactly these failure modes. The totient result of [8] is the constructive counterexample: a bias that numerics sug-

gested at 98% to 2% on twin primes is provably exactly 50%–50% once only p is assumed prime, so structural balance in the prime basis can be proved where numerics mislead. The claims of this paper are of the provable kind; extrapolations beyond them are not.

Falsification. The main results would be falsified as follows. The population estimate would be falsified by an exact count of $\pi(2^n)$ at any computed width differing from the Gauss-law value by more than the stated approximation error direction (the estimate is stated to underestimate at $n = 20$; a value below $75\,640$ at $n = 20$ would contradict the stated monotone relation between $\mathrm{Li}(x)$ and $x/\log x$ in this range). The preparation-cost formula would be falsified by a Grover analysis showing that the Miller-Rabin oracle's marked-set structure prevents the $\sqrt{N/M}$ speedup, e.g. if the oracle marks a set whose size is not known in advance or whose phases cannot be inverted; the formula would then need the amplitude-amplification corrections for unknown M. The entropy ceiling would be falsified by an explicit bipartition of a prime state whose reduced density matrix has rank exceeding $\pi(2^n)$, which is impossible by the term-count argument, so the ceiling can only fail if the state definition itself is changed (e.g. if the register encodes primes with ancilla flags that add terms). The emulability window would be falsified by a memory model in which sparse storage costs more than 16 bytes per prime by a factor exceeding $n \ln 2$, which would erase the sparse advantage; the window is sensitive to this constant at the level of the ratio $1/(n \ln 2)$ computed in Section 4.5.

Limitations.

- The register is noiseless; no decoherence, gate error, or error-correction overhead enters any bound, so all costs are lower bounds on physical costs.
- The Grover count assumes the Miller-Rabin oracle is a unit-cost primitive and that the number of marked items $M = \pi(2^n)$ is known exactly in advance; if only an estimate is available, the iteration count carries the usual unknown-M overhead, which this paper does not quantify.
- The gate-cost projection rests on $k = 64$ rounds, schoolbook $O(n^3)$ multiplication, and a unit-cost convention; these are assumptions, not circuit results, and the total 8.77×10^7 gates at $n = 64$ inherits their uncertainty.
- The entropy ceiling is an upper bound only; saturation is unproven and flagged as open.
- The collision floor $6/\pi^2$ applies to independent uniform draws from a large range; it is not a statement about correlations within a single prime state, and its architectural use (Section 4.4) assumes interactions that reduce to pairwise coprimality.
- The emulability window assumes 16 bytes per amplitude and a 1 TB budget; both are conventions, and the window location moves by one width per doubling of either the budget or the per-amplitude cost.
- The paper performs no experiment, simulation, or survey; every number is analytical, and none has been validated on hardware.

Open questions. The actual Schmidt spectrum of $|\mathbb{P}_n\rangle$ at small n; the circuit-level cost of a reversible Miller-Rabin oracle including ancilla and uncomputation; whether the Twin Prime state's entanglement inherits the counting-function structure that [3] identifies, in a form analyzable by the inductive tools of [1] and [7]; and whether the ontological thesis of [11] — that primes are spectral projections of a continuous substrate — constrains the physical realizability of a prime basis at all, or is orthogonal to the information-theoretic bounds derived here. The bounds of this paper are compatible with either reading of that thesis, since they concern the information content of the state, not its ontology.

7. Practical Implications and Engineering Pathways

(a) Users and decisions. The results serve two named users. The first is the designer of quantum algorithms for number-theoretic tasks, who currently has no closed-form cost model for prime-state preparation: for a chosen register width n, Sections 4.1 and 4.2 give the population $\pi(2^n)$ and the oracle budget $G(n) \approx (\pi/4)\sqrt{n \ln 2}$ directly, changing the design question from "is preparation feasible?" to a lookup. The second is the classical simulator engineer or

skeptic deciding whether a claimed prime-basis result can be replicated without quantum hardware: Section 4.5 gives the exact-emulation window ($36 \le n \le 40$ at a 1 TB budget) and the rule that each doubling of storage buys one additional width, changing the replication decision from an open question to an arithmetic check. No procurement, market, or policy decision is claimed; the work is foundational in the sense that it supplies bounds for a register concept that has not been built.

(b) Readiness. Against the NASA Technology Readiness Level definitions, this work is purely analytical and has not been peer reviewed: it is TRL 1–2. The bounds are derived and numerically instantiated, but no prime state has been prepared on hardware, no Miller-Rabin oracle circuit has been compiled, and no prototype exists; analytical work of this kind does not exceed TRL 3, and this paper does not reach even the proof-of-concept level that TRL 3 requires, since no concept has been demonstrated in any physical setting.

(c) Smallest next test. The smallest de-risking calculation is a classical computation, not a hardware build: for $n \le 25$, where $2^{25} = 33\,554\,432$ amplitudes fit in under 537 MB of dense memory, a study would construct $|\mathbb{P}_n\rangle$ exactly by enumerating primes, compute the bipartite entanglement entropy across every qubit cut, and compare the observed entropy with the ceiling $\log_2 \pi(2^n)$. The outcome that would refute the paper's framing is an observed entropy exceeding the ceiling at any n and any cut, which is impossible by the rank argument and would indicate an implementation error; the scientifically informative outcome is the ratio $S_A / \log_2 \pi(2^n)$ and its trend in n, which tests whether saturation is plausible. A second proposed test is circuit-level: compile a reversible Miller-Rabin oracle for $n = 16$ and count its gates and ancillas against the projected $O(k\, n^3)$ model of Section 4.2; a compiled count exceeding the projection by more than a polylogarithmic factor would falsify the gate-cost projection.

(d) Binding constraints. The constraints that would bind first, in order, are: (i) oracle depth — the Miller-Rabin oracle dominates the preparation cost, and its reversible arithmetic on n-qubit registers is the component with the largest unquantified overhead in this paper; (ii) memory bandwidth for the classical side — inside the emulability window the sparse representation fits, but generic operations on it do not preserve prime support, as Section 4.5 states; (iii) error correction — the ideal-register assumption means every physical deployment inherits the surface-code overheads that [9] puts at roughly 1000:1 physical-to-logical ratios, which multiply all gate counts derived here; (iv) exact knowledge of $M = \pi(2^n)$ — Grover's iteration count assumes the marked-set size is known, and computing it exactly is itself a number-theoretic task. Materials, tolerances, and regulation do not bind at this stage and are listed as open questions rather than constraints.

(e) Order-of-magnitude bounds. As projections with the assumptions of Section 4.2 ($k = 64$, schoolbook exponentiation, unit-cost convention): the quantum preparation at $n = 64$ is projected at 8.77×10^7 gates, against a classical scan of 3.0946×10^{26} modular-exponentiation-level operations. No cost in currency, energy, or wall-clock time is derivable from the stated inputs, and none is claimed; any such figure would require an assumed operation rate and hardware model, which this paper does not supply.

(f) Risks and misuse paths. The direct misuse path is rhetorical rather than technical: the exponential ratio 2.836×10^{-19} between oracle calls and classical candidates could be quoted without the oracle's internal gate cost to claim an unqualified exponential advantage. The honest comparison, given in Section 4.2, is that the quantum side replaces a 10^{26}-scale candidate scan with a polylogarithmic number of oracle invocations whose gate cost is projected at 8.77×10^7 under stated assumptions; quoting either number alone misrepresents the trade. A second risk is overclaiming robustness from the collision floor: $6/\pi^2$ is a statement about independent uniform draws, not about error rates of a physical register, and Section 4.4's architectural reading assumes coprimality-gated interactions that have not been specified in hardware terms.

8. Conclusion

This paper converted the Prime-state proposal of [3] from a qualitative architecture idea into a set of checkable bounds. For an n-qubit register holding the superposition of all primes below 2^n: the population is $\pi(2^n) \approx 2^n/(n \ln 2)$, giving $75\,640$ at $n = 20$ as a Gauss-law estimate; preparation by Grover search with a Miller-Rabin oracle costs $(\pi/4)\sqrt{n \ln 2}$ oracle calls, 5.23 at $n = 64$ against 2^{64} classical candidates; the entanglement entropy across any bipartition is bounded by $\log_2 \pi(2^n)$ bits, 16.21 at $n = 20$ and 58.527 at $n = 64$; independent register pairs collide through shared factors with probability $1 - 6/\pi^2 = 0.392073$; and exact classical emulation of the dense state vector costs $2^n \times 16$ bytes (17.59 TB at $n = 40$) against $\pi(2^n) \times 16$ bytes sparse, opening an exact-emulation window $36 \le n \le 40$ at a 1 TB budget that closes at $n = 41$ and widens by one width per doubling of storage. The entropy ceiling and the collision floor are unconditional; the cost figures are projections under stated assumptions; saturation of the entropy ceiling, circuit-level oracle costs, and physical robustness remain open. The bounds stand regardless of whether primes are fundamental discrete entities or, as [11] argues, spectral artifacts of a continuous substrate, because they quantify the information content of the state rather than its ontology; and the representational trade-off they exhibit — microscopic fidelity against computational cost — is the same trade-off that [12] formalizes as predictive efficiency, which is where the results reach beyond quantum architecture.

Provenance and AI Involvement

- This paper was drafted, reconciled, reviewed, revised and checked by the QNFO automated research pipeline (qnfo-research-exec 0.45.2). Language models were tools in that process; they are not authors. The named author is accountable for the content.
- Models: drafting by openai/gpt-oss-120b, zai-org/glm-5.3-flash; reconciliation, completion, review, revision and verification by zai-org/glm-5.3-flash, openai/gpt-oss-120b, moonshotai/kimi-k2.6, with the QNFO model router as a fallback.
- Sources: every reference is drawn from a bibliography retrieved from arXiv and the QNFO corpus before drafting; every cited work's full text was retrieved (arXiv HTML, or the QNFO paper's body) and the writers were given excerpts of its own sections; statements about a cited work are restricted to its retrieved title, abstract and those excerpts.
- Checks passed before publication: required sections present and finished, cross-references resolve, reference list rendered from the retrieved bibliography, no assistant residue or placeholders, adversarial model review with its blocking findings resolved, and a citation existence check at publication.
- Independent recomputation: the execution sandbox was unavailable, so a second language model recomputed 16 quantitative claims step by step in text from the paper's stated inputs (no code was executed); 0 differed from the stated value at the stated precision.
- Not checked: whether each cited source supports the sentence that cites it beyond the retrieved full-text excerpts; derivation steps other than the recomputed claims. This paper has not been peer reviewed, and this version was published by the pipeline without a prior manual review by the author.

References

[1] Matrix Online Characteristic Number and Its applications in Goldbach Conjecture,Polignac Conjecture,the Twin Prime Conjecture. arXiv:1306.0795v2. https://arxiv.org/abs/1306.0795v2 [2] 6m Theorem for Prime numbers. arXiv:1810.02188v1. https://arxiv.org/abs/1810.02188v1 [3] Quantum Computation of Prime Number Functions. arXiv:1302.6245v3. https://arxiv.org/abs/1302.6245v3 [4] Two statements that are equivalent to a conjecture related to the distribution of prime numbers. arXiv:1406.4801v2. https://arxiv.org/abs/1406.4801v2 [5] Continued fractions constructed from prime numbers. arXiv:1003.4015v2. https://arxiv.org/abs/1003.4015v2 [6] Irregular behaviour of class

numbers and Euler-Kronecker constants of cyclotomic fields: the log log log devil at play. arXiv:1711.07996v2. https://arxiv.org/abs/1711.07996v2 [7] Researches on the Twin Prime Problem. arXiv:1405.2490v2. https://arxiv.org/abs/1405.2490v2 [8] On the difference in values of the Euler totient function near prime arguments. arXiv:1706.00392v2. https://arxiv.org/abs/1706.00392v2 [9] https://papers.qnfo.org/papers/alpha-pi-project. QNFO: Alpha Pi Project. [10] https://papers.qnfo.org/papers/deductive-verification-of-discrete-time-quantum-walk-simulators-contracts-cost-b. QNFO: Deductive Verification of Discrete-Time Quantum-Walk Simulators: Contracts, Cost Bounds, and Extensions to Graph-Generalized and Open-System Walks. [11] https://papers.qnfo.org/papers/prime-numbers-as-spectral-artifacts. QNFO: Prime Numbers as Spectral Artifacts. [12] https://papers.qnfo.org/papers/dynamic-optimality-in-physical-and-arithmetic-systems. QNFO: Dynamic Optimality in Physical and Arithmetic Systems.