Sub-Cubic Homomorphic Matrix Multiplication via Self-Dual Normal Bases

Efe İzbudak[1*], Kübra Kaytancı[2], Ferruh Özbudak[2], Erkay Savaş[2]

[1*]Department of Mathematics, Middle East Technical University, Ankara, Türkiye.

[2]Faculty of Engineering and Natural Sciences, Sabancı University, Istanbul, Türkiye.

*Corresponding author(s). E-mail(s): efe.izbudak@metu.edu.tr;
Contributing authors: kubra.kaytanci@sabanciuniv.edu;
ferruh.ozbudak@sabanciuniv.edu; erkays@sabanciuniv.edu;

Abstract

This paper analyzes the bilinear embedding of matrix algebras into commutative cyclotomic rings. We apply the Cohn-Umans method. This establishes that single-multiplication bilinear monomial embeddings require a ring degree of $\widetilde{\boldsymbol{\Omega}}(\boldsymbol{N^3})$. We circumvent this bound by routing Strassen tensor rank decompositions through orthogonal Chinese Remainder Theorem ideals. This reduces the asymptotic complexity to $\widetilde{\boldsymbol{\Omega}}(\boldsymbol{N^{\log_2 7}})$. We optimize the coprime tensor decomposition. Embedding the inner tensor into the maximal real subfield satisfies the Lempel-Weinberger parity constraint. This guarantees the existence of a Self-Dual Normal Basis, reducing the required basis generators to a single element and mathematically halving the homomorphic trace depth. Canonical integer polynomial lifts ensure uniform norm bounds. Type I Optimal Normal Bases bound the trace dual expansions to an $\boldsymbol{O(1)}$ constant. By invoking Kronecker's theorem, we prove that the polynomial power basis minimizes the canonical expansion for the non-evaluated tensor components. A towered evaluation over composite degrees controls noise propagation. This decouples key-switching errors into a logarithmic bound. We generalize the embedding to Galois rings via Hensel's and Nakayama's lemmas to support high-precision integer arithmetic. Furthermore, we extend the architecture to boundless matrices exceeding the fixed ring capacity via a multi-ciphertext block-Strassen decomposition. By deferring the homomorphic trace operator to post-Strassen recombination, we completely eliminate homomorphic basis-switching, achieving an asymptotic complexity of $\boldsymbol{O(N^{\log_2 7-1/\rho})}$ multiplications and $\widetilde{\boldsymbol{O}}(\boldsymbol{N^{2-2/(\rho \log_2 7)}})$ automorphisms for matrices of arbitrary dimension. Empirical benchmarks over the BGV scheme validate the approach.

A multi-threaded towered trace evaluates $\mathbf{32 \times 32}$ matrices in **141.3** milliseconds at a security level of $\boldsymbol{\lambda = 148}$ using one ciphertext-ciphertext multiplication. We achieve a speedup factor of **2.49** over multi-threaded baselines.

Keywords: Homomorphic Encryption, Matrix Algebras, Normal Bases, Cyclotomic Fields, Galois Rings, Bilinear Forms, Triple Product Property, Discrete Gaussians, Strassen Algorithm

MSC Classification: 94A60 , 11T71 , 15A69 , 13B05 , 13M05 , 68Q17

Contents

1 Introduction **3**
- 1.1 Prior Work on Homomorphic Matrix Multiplication . . . 3
- 1.2 Evaluation Pipeline and Contributions . . . 5
- 1.3 Organization . . . 8

2 Preliminaries **8**
- 2.1 Cyclotomic Fields and Galois Theory . . . 8
- 2.2 Tensor Decompositions of Cyclotomic Rings . . . 8
- 2.3 The Canonical Embedding and Norms . . . 9
- 2.4 Bilinear Complexity and Tensor Rank . . . 9
- 2.5 Discrete Gaussian Distributions . . . 9
- 2.6 Trace Pairings and Normal Bases . . . 10
- 2.7 Galois Rings . . . 10
- 2.8 The RLWE Assumption . . . 10
- 2.9 Module Homomorphisms in BGV . . . 11

3 The Coprime Tensor Decomposition **11**

4 Algebraic Bounds on Bilinear Embeddings **12**
- 4.1 The Monomial TPP Bound . . . 13
- 4.2 Independence of Skew Bilinear Forms . . . 14
- 4.3 Beating the Bound via CRT-Strassen Hybrids . . . 15

5 Embeddings via Normal Bases **16**
- 5.1 Computation of the Dual Generator . . . 17
- 5.2 Self-Dual Normal Bases via Maximal Real Subfields . . . 17
- 5.3 Minimization via Optimal Normal Bases . . . 18

6 Extraction of the Matrix Trace **20**

7 Uniformity of Canonical Norms **20**

8 Canonical Expansion Minimum **22**

9 Logarithmic and Towered Trace Evaluations **23**
9.1 Logarithmic Trace Evaluation 23
9.2 Towered Trace Evaluations 25

10 Ideal Factorization Constraints for Parallel Evaluations **25**

11 Generalization to Galois Rings **26**

12 Extension to Boundless Matrix Dimensions **27**

13 Lattice-Theoretic Security Reductions **30**

14 Experimental Results **30**
14.1 Algorithmic Scalability vs Ciphertext Capacity 31

15 Comparison and Conclusion **32**

1 Introduction

Fully homomorphic encryption enables the evaluation of arithmetic circuits over encrypted data without decryption. In the Ring Learning with Errors (RLWE) scheme, elements of finite fields are embedded into cyclotomic quotient rings. The evaluation of matrix multiplication over these structures imposes polynomial degree constraints and requires high multiplicative depth.

1.1 Prior Work on Homomorphic Matrix Multiplication

Secure matrix multiplication over BGV and similar RLWE schemes has evolved to balance ciphertext expansion, multiplicative depth, and evaluation latency. A straightforward approach to securely multiplying two $N \times N$ matrices encodes entries individually, requiring N^2 distinct ciphertexts per matrix. This is inefficient in communication and computation. Early optimizations for encrypted matrix-vector evaluations by Halevi et al. [5] were extended to matrix-matrix products by Lu et al. [10] and Wang et al. [16], reducing the representation to $O(N)$ ciphertexts per matrix. However, these methods are computationally demanding, requiring $O(N^2)$ plaintext-ciphertext multiplications (CMult), $O(N^2)$ ciphertext-ciphertext multiplications (Mult), and $O(N^2 \log N)$ homomorphic Galois automorphisms (Aut) to yield a single matrix product.

To minimize ciphertext expansion, single-ciphertext matrix packing techniques were introduced. Cheon et al. [3] and Rathee et al. [13] extracted matrix diagonals and utilized cyclic permutations, packing the operands into a single ciphertext. This reduced the required evaluation depth, though it still requires $O(N)$ ciphertext-ciphertext multiplications and $O(N \log N)$ automorphisms. Subsequent routing optimizations by Jiang et al. [8] and Jang et al. [7] reduced the necessary homomorphic automorphisms to $O(N)$. Rizomiliotis and Triakosia [14] refined these slot-shifted evaluations by reducing the number of ciphertext-ciphertext multiplications from $O(N)$ to 1, though the evaluation still requires $O(N)$ sequential automorphisms.

Recently, Zheng, Li, and Wang (ZLW) [17] introduced a coprime tensor decomposition, substituting standard SIMD batching. By factoring the global cyclotomic ring into a tensor product of pairwise coprime subrings, they mapped the matrix dimensions to the tensor components. For compact matrices (where $N \leq n^{1/3}$, with n being the cyclotomic ring degree), this approach computes the core homomorphic evaluation using 1 ciphertext-ciphertext multiplication and $O(\log N)$ field automorphisms. However, their algebraic formulation relies on standard polynomial power bases and requires homomorphic basis switching prior to evaluation. Consequently, the computation requires 2 plaintext-ciphertext multiplications. For larger matrices ($N > n^{1/3}$, where $N = n^{\rho}$), ZLW extended their framework using block matrix multiplication and Strassen's algorithm. This approach requires $O(N^{(1-\frac{1}{3\rho})\log_2 7})$ ciphertext-ciphertext multiplications and $O(N^{2-\frac{2}{3\rho}})$ plaintext-ciphertext multiplications. We extend our framework to boundless matrices by applying block Strassen decompositions over the SDNB embeddings. Because our base block capacity $\widetilde{O}(n^{1/\log_2 7})$ is asymptotically larger than $n^{1/3}$, this reduces the large-matrix ciphertext multiplication complexity to $O(N^{\log_2 7-\frac{1}{\rho}})$. Furthermore, our normal bases eliminate plaintext-ciphertext multiplications entirely. The table below summarizes the asymptotic evaluation costs across these methods.

Table 1 Asymptotic complexity of a single $N \times N$ homomorphic matrix multiplication. (N: maximum matrix dimension, n: ring degree, $\rho = \log_n N$, s: inner subring degree, s': effective subring degree, CMult: Plaintext-Ciphertext Multiplications, Mult: Ciphertext-Ciphertext Multiplications, Aut: Homomorphic Galois Automorphisms). For [17], C stands for Compact while L stands for Large.

Methodology	N	CMult	Aut	Mult	Depth
[10, 16]	$O(n)$	$O(N^2)$	$O(N^2 \log N)$	$O(N^2)$	1 CMult + 1 Mult
[3, 13]	$O(n^{1/2})$	$O(N)$	$O(N \log N)$	$O(N)$	1 CMult + 1 Mult
[8]	$O(n^{1/2})$	$O(N)$	$O(N)$	$O(N)$	2 CMult + 1 Mult
[7]	$O(n^{1/2})$	$O(N)$	$O(N)$	$O(N)$	1 CMult + 1 Mult
[14]	$O(n^{1/3})$	$O(N)$	$O(N)$	1	1 CMult + 1 Mult
[17] (C)	$O(n^{1/3})$	2	$O(\log N)$	1	1 CMult + 1 Mult
[17] (L)	∞	$O(N^{2-\frac{2}{3\rho}})$	$\widetilde{O}(N^{2-\frac{2}{3\rho}})$	$O(N^{(1-\frac{1}{3\rho})\log_2 7})$	1 CMult + 1 Mult
Ours (1)	$\widetilde{\mathbf{O}}(\mathbf{n^{1/\log_2 7}})$	**0**	$\mathbf{O}(\log \mathbf{s'})$	**1**	**1** Mult
Ours (3)	∞	**0**	$\widetilde{\mathbf{O}}(\mathbf{N^{2-\frac{2}{\rho \log_2 7}}})$	$\mathbf{O}(\mathbf{N^{\log_2 7-\frac{1}{\rho}}})$	**1** Mult

While the tensor decomposition framework in [17] achieves low asymptotic operation counts for compact matrices, its reliance on standard polynomial power bases introduces several limitations. First, homomorphic basis switching involves evaluating two plaintext-ciphertext multiplications (CMult) before the ciphertext-ciphertext multiplication (Mult), which increases multiplicative depth and noise. Second, the power basis yields a trace dual basis with varying canonical norms. This variance increases the magnitude of the error polynomials, leading to noise propagation during homomorphic evaluation. Third, the construction requires computing and storing $O(N)$

distinct polynomial generators prior to encryption. Finally, embedding matrices into a single global tensor ring imposes a theoretical limit, requiring a cyclotomic ring degree of $\widetilde{\Omega}(N^3)$ to support the Abelian Triple Product Property without spatial ideal decompositions.

In this paper, we analyze the constraints on matrix embeddings over commutative rings and refine the tensor decomposition methodology to address these limitations. By routing Strassen tensor decompositions through distinct prime ideals and applying normal bases, we bypass the $O(n^{1/3})$ single-multiplication dimension bound. Furthermore, we eliminate the homomorphic basis-switching step (reducing CMult to 0 while retaining 1 Mult), and maintain canonical noise symmetry. For the boundless setting where matrix sizes significantly exceed the fixed ring degree limit, our block-matrix formulation seamlessly applies SDNB encodings to bypass the $O(N^{2-\frac{2}{3\rho}})$ basis-switching bottleneck present in prior extensions.

1.2 Evaluation Pipeline and Contributions

The homomorphic evaluation procedure maps matrix operands through encoding, encryption, and evaluation. Figure 1 illustrates these transformations. We outline the theoretical contributions associated with these phases.

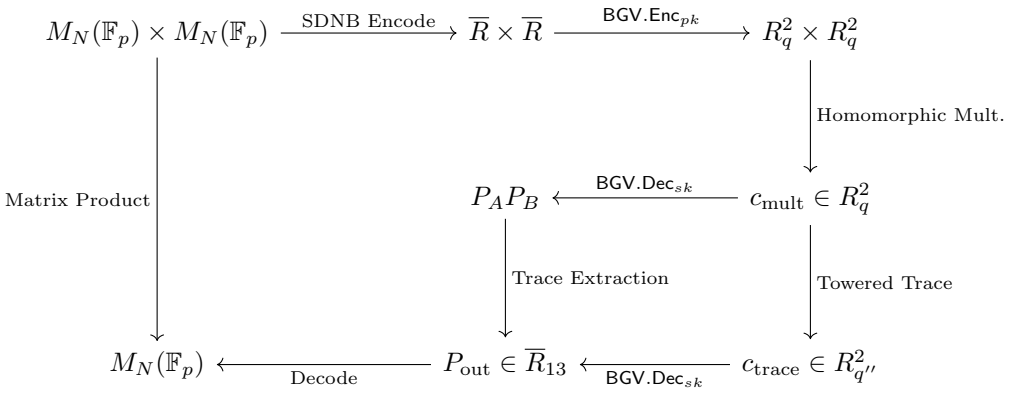


Fig. 1 Commutative diagram illustrating the end-to-end homomorphic matrix evaluation procedure.

Dimensional Bounds and Space Partitioning

Before embedding the matrices into the cyclotomic ring $\overline{R}$, we establish the required algebraic degree to compute the matrix trace.

Theorem A (cf. Theorem 1) *A bilinear monomial embedding of the matrix algebra $M_N(\mathbb{F}_p)$ into a commutative cyclotomic ring requires the underlying exponent group to satisfy the Abelian Triple Product Property. An embedding evaluated via a single multiplicative operation requires a ring degree $n = \widetilde{\Omega}(N^3)$.*

To avoid this bound, we partition the matrix components.

Theorem B (cf. Theorem 4) *By partitioning the matrix and mapping sub-tensors into distinct prime ideals of the unramified cyclotomic extension, the single-multiplication degree bound drops to* $n = \widetilde{\Omega}(N^{\log_2 7})$.

For the parallel evaluation slots to survive the homomorphic trace operator later in the pipeline, we derive a condition on the cyclotomic ideals.

Proposition C (cf. Theorem 15) *Let* $\overline{R} \cong \bigoplus_{v=1}^{\ell} \mathbb{F}_{p^d}$ *be the decomposition of the cyclotomic quotient ring into residue fields over* $\mathbb{F}_p$. *The primitive orthogonal idempotents defining the parallel matrix slots are preserved under Galois trace evaluations if and only if* $\gcd(\mathrm{ord}_{m_2}(p), \mathrm{ord}_{m_1 m_3}(p)) = 1$.

Normal Basis Tensor Encoding

We map the matrices into $\overline{R}$, substituting the power basis with normal bases. To satisfy the parity constraint and achieve a single-element generator ($\gamma_0 = \theta$), we embed into the maximal real subfield.

Theorem D (cf. Theorem 6 and Theorem 7) *Embedding the inner tensor component into the maximal real subfield* $K^+ = \mathbb{Q}(\zeta_{m_2} + \zeta_{m_2}^{-1})$ *yields an odd extension degree for specific moduli. This satisfies the Lempel-Weinberger parity constraint, guaranteeing the existence of a Self-Dual Normal Basis (SDNB) where* $\gamma_0 = \theta$, *reducing the required basis generators to a single element and structurally halving the required homomorphic trace depth. Type I Optimal Normal Bases limit the trace dual expansion to an* $O(1)$ *constant.*

By defining the elements canonically over $\mathbb{Z}$, we ensure the norm bounds remain uniform, preventing noise asymmetry between the left and right operands.

Theorem E (cf. Theorem 9) *Let* θ *and* γ_0 *generate a normal basis and its trace dual for the cyclotomic subring, and let* $\tilde{\theta}, \tilde{\gamma}_0 \in \mathbb{Z}[x]$ *denote their respective canonical integer polynomial lifts. Evaluating the elements via the algebraic expansions* $\tilde{\theta}(x^{p^j}) \bmod \langle \Phi_{m_2}(x) \rangle$ *and* $\tilde{\gamma}_0(x^{p^j}) \bmod \langle \Phi_{m_2}(x) \rangle$ *over* $\mathbb{Z}$ *without coefficient-wise modulo* p *reduction ensures uniform initial canonical norm bounds across all basis elements.*

We apply Kronecker's theorem to prove that the polynomial power basis minimizes expansion for the non-extracted outer tensor components.

Theorem F (cf. Theorem 12) *The polynomial power basis minimizes the maximum canonical norm among all full-rank integral bases for the non-extracted tensor components.*

We extend this encoding layer beyond finite fields to support integer matrices without modulo p overflow constraints.

Proposition G (cf. Theorem 16) *By Hensel's and Nakayama's Lemmas, the normal basis bilinear embedding over $\mathbb{F}_{p^s}$ lifts to the Galois ring $\mathrm{GR}(p^r, s) \cong \mathbb{Z}_{p^r}[x]/\langle f(x)\rangle$, where $f(x)$ is a basic irreducible factor of $\Phi_{m_2}(x)$, supporting single-multiplication evaluations of $M_N(\mathbb{Z}_{p^r})$.*

Boundless Matrix Extension

To circumvent the finite capacity of a single ciphertext entirely, we generalize the execution bounds via a boundless ciphertext-level partitioning algorithm.

Theorem H (cf. Theorem 17) *For arbitrary dimension matrices $N \to \infty$, the matrix product can be recursively mapped over ciphertext blocks. Due to the structural invariance of the SDNB ($\gamma_0 = \theta$), the block-matrix Strassen recurrences can be executed directly over ciphertexts without invoking homomorphic block unpacking or basis switching. This reduces the* CMult *evaluation penalty to precisely 0, solving the primary latency bottlenecks of prior large-scale embeddings.*

Homomorphic Evaluation and Trace Extraction

The encoded polynomials are encrypted into BGV ciphertexts $c_A, c_B \in R_q^2$. A single homomorphic multiplication is performed, generating the tensor product ciphertext c_{mult}, followed by relinearization. The optimizations applied during encoding ensure that the underlying initial noise bounds remain uniform prior to this multiplicative expansion.

To isolate the matrix inner product coefficients, a sequence of $O(\log s')$ Galois automorphisms is applied to project the target polynomial into $c_{\mathrm{trace}} \in R_{q''}^2$. Standard trace operators cause linear noise growth. We mitigate this using towered evaluations.

Theorem I (cf. Theorem 14) *For a composite extension degree $s = d_1 d_2$, evaluating the trace via the field tower $\mathrm{Tr}_{\mathbb{F}_{p^s}/\mathbb{F}_p} = \mathrm{Tr}_{\mathbb{F}_{p^{d_1}}/\mathbb{F}_p} \circ \mathrm{Tr}_{\mathbb{F}_{p^s}/\mathbb{F}_{p^{d_1}}}$ with an intermediate modulus switch reduces the additive key-switching noise bound, scaling from $O(s)B_{\mathrm{ks}}$ to $O(d_1 + d_2)B_{\mathrm{ks}}$.*

Decryption and Decoding

The final trace ciphertext is decrypted to recover $P_{\mathrm{out}} \in \overline{R}_{13}$, and the orthogonal idempotents are mapped back to compute the resulting plaintext product $AB \in M_N(\mathbb{F}_p)$.

We evaluate the matrix embedding over the BGV scheme. Empirical benchmarks show that a multi-threaded towered trace reduces the evaluation latency of 32×32 matrices to 0.14 seconds at a security level of $\lambda = 148$. The $O(\log s')$ algebraic complexity reduces evaluation latency compared to baseline algorithms, achieving a speedup factor of 2.34 over multi-threaded executions.

1.3 Organization

Section 2 reviews the algebraic background. Section 3 outlines the coprime tensor decomposition. Section 4 establishes bilinear complexity bounds for the embeddings. Section 5 formulates normal basis embeddings and resolves the parity constraint. Section 6 derives the matrix trace extraction. Section 7 proves the uniformity of canonical norms. Section 8 evaluates the canonical expansion minimum. Section 9 constructs the towered trace evaluation. Section 10 derives the ideal factorization constraints for parallel computations. Section 11 generalizes the embedding to Galois rings. Section 12 extends the framework to boundless matrix dimensions. Section 13 details the lattice security reductions. Section 14 presents the empirical performance data. Section 15 concludes the paper.

2 Preliminaries

This section establishes the algebraic and cryptographic notation used throughout the paper.

2.1 Cyclotomic Fields and Galois Theory

Let m be a positive integer and let $\zeta_m = e^{2\pi i/m}$ be a primitive m-th root of unity in $\mathbb{C}$. The m-th cyclotomic field is $K = \mathbb{Q}(\zeta_m)$, an abelian extension of $\mathbb{Q}$ of degree $n = \varphi(m)$. The Galois group $\mathrm{Gal}(K/\mathbb{Q})$ is isomorphic to the multiplicative group of units $(\mathbb{Z}/m\mathbb{Z})^\times$. Each automorphism $\tau_k \in \mathrm{Gal}(K/\mathbb{Q})$ is determined by its action on the primitive root $\tau_k(\zeta_m) = \zeta_m^k$ for $\gcd(k, m) = 1$. The maximal real subfield of K is denoted $K^+ = \mathbb{Q}(\zeta_m + \zeta_m^{-1})$, with degree $[K^+ : \mathbb{Q}] = \varphi(m)/2$.

The ring of algebraic integers of K is $\mathcal{O}_K = \mathbb{Z}[\zeta_m]$, which is isomorphic to $R = \mathbb{Z}[x]/\langle \Phi_m(x) \rangle$. For a rational prime p satisfying $p \nmid m$, the ideal $p\mathcal{O}_K$ is unramified in K. It decomposes into a product of ℓ distinct prime ideals $\mathfrak{p}_1\mathfrak{p}_2 \cdots \mathfrak{p}_\ell$, where $\ell = \varphi(m)/d$ and d is the multiplicative order of p modulo m.

The decomposition group $D_{\mathfrak{p}_i} = \{\sigma \in \mathrm{Gal}(K/\mathbb{Q}) \mid \sigma(\mathfrak{p}_i) = \mathfrak{p}_i\}$ corresponding to a prime ideal $\mathfrak{p}_i$ lying over p is cyclic of order d. It is generated by the Frobenius element $\sigma_p \in \mathrm{Gal}(K/\mathbb{Q})$, defined by the relation $\sigma_p(\alpha) \equiv \alpha^p \pmod{\mathfrak{p}_i}$ for all $\alpha \in \mathcal{O}_K$. The quotient ring $\mathcal{O}_K/\mathfrak{p}_i$ is isomorphic to the finite field $\mathbb{F}_{p^d}$.

2.2 Tensor Decompositions of Cyclotomic Rings

Let $m = m_1 m_2 \cdots m_k$ where the factors m_i are pairwise coprime integers. By the properties of cyclotomic extensions, the compositum field $\mathbb{Q}(\zeta_m)$ is isomorphic to the tensor product of the subfields over $\mathbb{Q}$,

$$\mathbb{Q}(\zeta_m) \cong \mathbb{Q}(\zeta_{m_1}) \otimes_{\mathbb{Q}} \mathbb{Q}(\zeta_{m_2}) \otimes_{\mathbb{Q}} \cdots \otimes_{\mathbb{Q}} \mathbb{Q}(\zeta_{m_k}).$$

Because the discriminants of these extensions are pairwise coprime, the ring of algebraic integers $\mathcal{O}_K = \mathbb{Z}[\zeta_m]$ decomposes into the tensor product of the constituent

rings of integers over $\mathbb{Z}$,

$$\mathbb{Z}[\zeta_m] \cong \mathbb{Z}[\zeta_{m_1}] \otimes_{\mathbb{Z}} \mathbb{Z}[\zeta_{m_2}] \otimes_{\mathbb{Z}} \cdots \otimes_{\mathbb{Z}} \mathbb{Z}[\zeta_{m_k}].$$

This isomorphism provides the foundation for decomposing the global cyclotomic ring R into a tensor product of independent polynomial subrings.

2.3 The Canonical Embedding and Norms

The canonical embedding $\sigma : K \hookrightarrow \mathbb{C}^n$ maps an element $a \in K$ to the vector of its evaluations under all n complex embeddings. For a polynomial $a(x) \in R$, this mapping evaluates over the primitive roots as

$$\sigma(a) = \left(a(\zeta_m^k)\right)_{\gcd(k,m)=1}.$$

The canonical norm is defined as the infinity norm over the complex embeddings,

$$\|a\|_{\text{can}} = \|\sigma(a)\|_\infty = \max_{\gcd(k,m)=1} \left|a(\zeta_m^k)\right|.$$

The canonical norm defines a sub-multiplicative metric on R, satisfying $\|ab\|_{\text{can}} \leq \|a\|_{\text{can}}\|b\|_{\text{can}}$ due to the point-wise nature of multiplication in $\mathbb{C}^n$.

Let $\|a\|_\infty$ denote the maximum absolute value of the integer coefficients of $a(x)$ expanded in the standard power basis $\{1, x, \ldots, x^{n-1}\}$. The coefficient norm is bounded by the canonical norm via the cyclotomic expansion factor c_m, satisfying $\|a\|_\infty \leq c_m\|a\|_{\text{can}}$. Over the base ring $\mathbb{Z}[x]$, the expansion factor depends on the prime factorization of m.

2.4 Bilinear Complexity and Tensor Rank

Let U, V, and W be finite-dimensional vector spaces over a field F. A bilinear mapping $T : U \times V \to W$ can be represented as a tensor $T \in U^* \otimes_F V^* \otimes_F W$. The ***tensor rank*** of T, denoted $\text{rk}(T)$, is the minimal integer r such that T can be expressed as a sum of r rank-one tensors,

$$T = \sum_{i=1}^{r} \phi_i \otimes \psi_i \otimes w_i,$$

where $\phi_i \in U^*$, $\psi_i \in V^*$, and $w_i \in W$. The tensor rank provides a lower bound on the number of non-scalar multiplications required to evaluate the bilinear map over commutative variables. By algebraic complexity theory, the tensor rank of the standard matrix multiplication operator is bounded by $\text{rk}(T_{\text{mat}}) \geq 2N^2 - 1$.

2.5 Discrete Gaussian Distributions

Let $\varsigma > 0$ be a real parameter. The Gaussian weight function over $\mathbb{R}$ is defined as $\rho_\varsigma(x) = \exp(-\pi x^2/\varsigma^2)$. The discrete Gaussian distribution over $\mathbb{Z}$ with parameter ς,

denoted $D_{\mathbb{Z},\varsigma}$, assigns probability proportional to $\rho_\varsigma(x)$ for each $x \in \mathbb{Z}$. The probability mass function evaluates to

$$D_{\mathbb{Z},\varsigma}(x) = \frac{\exp(-\pi x^2/\varsigma^2)}{\sum_{z\in\mathbb{Z}} \exp(-\pi z^2/\varsigma^2)}.$$

A random variable X distributed according to $D_{\mathbb{Z},\varsigma}$ is sub-Gaussian with parameter ς, satisfying the tail bound $\Pr[|X| \geq t] \leq 2\exp(-\pi t^2/\varsigma^2)$ for any $t > 0$. In cyclotomic rings, an error polynomial e is generated by drawing its n coefficients independently from $D_{\mathbb{Z},\varsigma}$ with respect to the standard power basis. By concentration inequalities for sub-Gaussian random vectors, the coefficient infinity norm $\|e\|_\infty$ and the canonical norm $\|e\|_{\text{can}}$ are bounded by $C\varsigma\sqrt{n}$ with probability at least $1 - 2^{-\Omega(n)}$, where C is a constant dependent on the ring geometry.

2.6 Trace Pairings and Normal Bases

Let E/F be a finite separable field extension of degree s. The trace map $\text{Tr}_{E/F} : E \to F$ is defined as $\text{Tr}_{E/F}(\alpha) = \sum_{\sigma\in\text{Gal}(E/F)} \sigma(\alpha)$. This defines a symmetric, non-degenerate bilinear form on $E \times E$ via $\langle x, y\rangle = \text{Tr}_{E/F}(xy)$.

Given a vector space basis $\mathcal{B} = \{v_0, \ldots, v_{s-1}\}$ of E over F, there exists a unique dual basis $\mathcal{B}^\vee = \{v_0^\vee, \ldots, v_{s-1}^\vee\}$ such that $\text{Tr}_{E/F}(v_i v_j^\vee) = \delta_{i,j}$. A basis $\mathcal{N}$ is a ***normal basis*** if it is generated by the Galois conjugates of a single element $\theta \in E$, satisfying $\mathcal{N} = \{\sigma(\theta) \mid \sigma \in \text{Gal}(E/F)\}$. By the Normal Basis Theorem, such a generator θ always exists over finite fields. The trace dual of a normal basis is also a normal basis, generated by an element $\gamma_0 \in E$. A Type I Optimal Normal Basis (ONB) exists if $s + 1$ is prime and p generates the multiplicative group $(\mathbb{Z}/(s+1)\mathbb{Z})^\times$.

2.7 Galois Rings

For a prime p and integer $r \geq 1$, let $f(x) \in \mathbb{Z}_{p^r}[x]$ be a monic polynomial of degree s such that its reduction modulo p is irreducible in $\mathbb{F}_p[x]$. The quotient ring $\text{GR}(p^r, s) = \mathbb{Z}_{p^r}[x]/\langle f(x)\rangle$ is a Galois ring. It is a local ring with maximal ideal $p\,\text{GR}(p^r, s)$ and residue field $\mathbb{F}_{p^s}$.

2.8 The RLWE Assumption

Let $q \geq 2$ be an integer and let $R_q = \mathbb{Z}_q[x]/\langle\Phi_m(x)\rangle$. Let χ denote the error distribution over R, defined by sampling coefficients independently from the discrete Gaussian distribution $D_{\mathbb{Z},\varsigma}$.

For a secret element $s \in R_q$, the RLWE distribution consists of samples of the form $(a, as + e) \in R_q \times R_q$, where a is drawn uniformly from R_q and $e \leftarrow \chi$. The search version of the $\text{RLWE}_{n,q,\chi}$ problem asks to recover s given samples. The decisional $\text{RLWE}_{n,q,\chi}$ assumption asserts that it is infeasible for a probabilistic polynomial-time algorithm to distinguish samples drawn from the RLWE distribution from uniformly random samples drawn over $R_q \times R_q$.

2.9 Module Homomorphisms in BGV

The BGV encryption scheme [1] evaluates arithmetic circuits over the plaintext ring R_p. A ciphertext is a vector $c = (c_0, c_1) \in R_q^2$. The decryption function with respect to a secret key $s \in R$ evaluates the relation $[[c_0 + c_1 s]_q]_p = m$. The correctness of the evaluation relies on a bounded error polynomial $e \in R$, such that the relation $c_0 + c_1 s = m + pe$ is satisfied as an equality in R_q.

Addition of ciphertexts corresponds to component-wise polynomial addition over R_q^2, resulting in an additive growth of the error polynomial $e_{\text{add}} = e + e'$. Multiplication of c and c' evaluates the tensor product $(c_0 + c_1 s) \otimes (c'_0 + c'_1 s)$, yielding a degree-2 vector $(d_0, d_1, d_2) = (c_0 c'_0, c_0 c'_1 + c_1 c'_0, c_1 c'_1)$ encrypting the product under the extended basis $(1, s, s^2)$. Returning this state to R_q^2 requires a linear transformation via evaluation keys encrypting s^2. Modulus switching from q to a smaller modulus q' introduces an additive rounding error bounded by B_{scale}. The canonical norm of the error polynomial increases multiplicatively during tensor evaluation.

Evaluation of a Galois automorphism $\tau_k(x) = x^k$ shifts the embedded plaintext to $m(x^k)$ under the permuted secret key $s(x^k)$. To restore the state to the original key s, the scheme applies a homomorphic basis change. The ciphertext polynomial $c_1(x^k)$ is decomposed via a digit extraction map $\mathcal{D}_B : R_q \to R_q^{D+1}$ into base-B digits $c^{(i)}$ such that $c_1(x^k) = \sum_{i=0}^{D} c^{(i)} B^i$, where $D = \lfloor \log_B q \rfloor$. The inner product is computed against a key matrix $\text{evk}_k \in M_{(D+1)\times 2}(R_q)$. This linear transformation injects an additive error polynomial $\mathbf{e}_{\text{ks}}$ drawn from the evaluation key parameters. The canonical norm of this error is bounded by $B_{\text{ks}} = \frac{16}{\sqrt{3}} pn\varsigma \lceil \log_B q \rceil$.

3 The Coprime Tensor Decomposition

To embed matrices into a single polynomial ring, we exploit the isomorphism between a global cyclotomic ring and a tensor product of smaller rings. Let $R = \mathbb{Z}[x]/\langle \Phi_m(x) \rangle$. For pairwise coprime integers m_1, m_2, m_3 satisfying $m = m_1 m_2 m_3$, the global ring decomposes over $\mathbb{Z}$ as $R \cong R_1 \otimes_{\mathbb{Z}} R_2 \otimes_{\mathbb{Z}} R_3$, where $R_i = \mathbb{Z}[x]/\langle \Phi_{m_i}(x) \rangle$. Reducing modulo the prime p yields the $\mathbb{F}_p$-algebra isomorphism $\overline{R} \cong \overline{R}_1 \otimes_{\mathbb{F}_p} \overline{R}_2 \otimes_{\mathbb{F}_p} \overline{R}_3$.

The degrees of the tensor components are denoted $r = \varphi(m_1)$, $s = \varphi(m_2)$, and $t = \varphi(m_3)$. Zheng, Li, and Wang utilized this algebra to evaluate matrix products AB for $A \in M_{r\times s}(\mathbb{F}_p)$ and $B \in M_{s\times t}(\mathbb{F}_p)$ [17]. Given vector space bases $U = \{u_i\}, V = \{v_j\}, W = \{w_k\}$ over $\mathbb{F}_p$ for the respective subrings, the polynomial embedding evaluates as:

$$m_A = \sum_{i=0}^{r-1} \sum_{j=0}^{s-1} a_{ij} \left(u_i \otimes v_j \otimes 1\right),$$

$$m_B = \sum_{k=0}^{s-1} \sum_{l=0}^{t-1} b_{kl} \left(1 \otimes v_k^{\vee} \otimes w_l\right),$$

where $V^\vee = \{v_k^\vee\}$ denotes the trace dual basis satisfying $\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(v_j v_k^\vee) = \delta_{j,k}$. The product of these elements evaluates to:

$$m_A m_B = \sum_{i=0}^{r-1} \sum_{j=0}^{s-1} \sum_{k=0}^{s-1} \sum_{l=0}^{t-1} a_{ij} b_{kl} \left(u_i \otimes (v_j v_k^\vee) \otimes w_l\right).$$

The matrix inner product is extracted by applying the field trace operator mapping elements from the global algebra $\overline{R}$ to the subalgebra $\overline{R}_{13} = \mathbb{F}_p[x]/\langle \Phi_{m_1 m_3}(x) \rangle$. Because the trace acts as the identity on $\overline{R}_1$ and $\overline{R}_3$, it evaluates the bilinear pairing over the inner subring $\overline{R}_2$:

$$\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}(m_A m_B) = \sum_{i=0}^{r-1} \sum_{l=0}^{t-1} \left(\sum_{j=0}^{s-1} \sum_{k=0}^{s-1} a_{ij} b_{kl} \, \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(v_j v_k^\vee) \right) (u_i \otimes 1 \otimes w_l).$$

By the definition of the dual basis, $\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(v_j v_k^\vee) = \delta_{j,k}$. The summation collapses to j, yielding the tensor sum:

$$\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}(m_A m_B) = \sum_{i=0}^{r-1} \sum_{l=0}^{t-1} \left(\sum_{j=0}^{s-1} a_{ij} b_{jl} \right) (u_i \otimes 1 \otimes w_l),$$

which isolates the coefficients of the matrix product AB over the tensor product of the outer components $\overline{R}_1 \otimes_{\mathbb{F}_p} \overline{R}_3$.

This evaluation in [17] relies on the polynomial power basis $v_j = x^j$ over the inner subring, which introduces two primary limitations.

1. *Asymmetric Norm Expansion:* The trace dual elements $v_j^\vee$ required for m_B are determined by the trace pairing constraint. For the cyclotomic power basis, the dual elements expand as polynomial fractions dependent on the inverse formal derivative of the minimal polynomial, $\Phi'_{m_2}(x)^{-1}$. Since the canonical norm bounds the magnitude of the fractional expansions over $\mathbb{Z}$, the encoding yields an asymmetric relation where the maximum of $\|v_j^\vee\|_{\mathrm{can}}$ exceeds $\|v_j\|_{\mathrm{can}} = 1$. This variance magnifies the error polynomial for the right operand, requiring a larger ciphertext modulus.
2. *Generator Storage:* The encoding requires the computation and storage of $2s$ distinct polynomial basis elements prior to encryption.

We analyze the fundamental limits of this embedding space and optimize the basis.

4 Algebraic Bounds on Bilinear Embeddings

We begin by establishing the theoretical limits of encoding matrices into commutative quotient rings.

4.1 The Monomial TPP Bound

To establish the limits of embedding matrix algebras into cyclotomic rings, we define the parameters of a bilinear polynomial embedding.

Definition 1 A ***bilinear polynomial embedding*** of $M_N(\mathbb{F}_p)$ over a cyclotomic ring R_p is parameterized by a triplet (K, D, n). The integer K denotes the rank of the free module R_p^K required to encode a single matrix. The integer D represents the tensor degree of the evaluation polynomial, and $n = \varphi(m)$ defines the dimension of the cyclotomic ring over $\mathbb{F}_p$.

The following theorem bounds the ring degree required for a single-multiplication embedding.

Theorem 1 *The mapping $\Psi : M_N(\mathbb{F}_p) \to R_p^K$ must be an $\mathbb{F}_p$-linear injection for the matrix to be recoverable. The vector space dimension satisfies $K \cdot n \geq \dim_{\mathbb{F}_p}(M_N(\mathbb{F}_p)) = N^2$. A bilinear monomial embedding evaluated within a single commutative polynomial ring ($K = 1$) via a single multiplicative operation ($D = 1$) requires a ring degree $n = \widetilde{\Omega}(N^3)$.*

Proof For a single ring embedding ($K = 1$), the injectivity of the $\mathbb{F}_p$-linear mapping Ψ requires the dimension of the target vector space to be at least the dimension of the domain. Thus, the ring degree must satisfy $n \geq \dim_{\mathbb{F}_p}(M_N(\mathbb{F}_p)) = N^2$.

To derive the sub-cubic bound for monomial embeddings, we view the polynomial ring $R_p \cong \mathbb{F}_p[x]/\langle \Phi_m(x) \rangle$ as a quotient of the group algebra $\mathbb{F}_p[G]$ for the cyclic abelian group $G = \mathbb{Z}/m\mathbb{Z}$. A bilinear monomial embedding maps the basis elements of the matrices A and B to specific group elements. Assuming the encoded monomials strictly map into the cyclotomic power basis $\{1, x, \ldots, x^{\varphi(m)-1}\}$, they remain strictly linearly independent over $\mathbb{F}_p$, ensuring the polynomial convolution exactly mirrors the group algebra without modular folding. We represent the encoded matrices via the polynomials $P_A(x) = \sum_{i,j=1}^{N} A_{ij} x^{f(i,j)}$ and $P_B(x) = \sum_{k,l=1}^{N} B_{kl} x^{g(k,l)}$, where $f, g : \{1, \ldots, N\} \times \{1, \ldots, N\} \to G$ are the index mapping functions.

The product $P_A(x)P_B(x)$ evaluated in the group algebra generates the convolution sum:

$$P_A(x)P_B(x) = \sum_{i,j,k,l=1}^{N} A_{ij} B_{kl} x^{f(i,j)+g(k,l)}.$$

To extract the matrix product $C_{il} = \sum_{j=1}^{N} A_{ij} B_{jl}$, there must exist a target exponent $c_{il} \in G$ for each pair (i, l) such that the scalar coefficient of $x^{c_{il}}$ corresponds to C_{il}. For this linear projection to be correct across all matrix inputs, the convolution must satisfy the global non-interference condition:

$$f(i,j) + g(k,l) = c_{i'l'} \iff i = i',\ l = l',\ \text{and } j = k.$$

From the alignment condition corresponding to $j = k$, we have $f(i, j) + g(j, l) = c_{il}$, which means $f(i, j)+g(j, l)$ is independent of j. Therefore, for any k, $f(i, j)+g(j, l) = f(i, k)+g(k, l)$, or equivalently, $f(i, j) - f(i, k) = g(k, l) - g(j, l)$. Because the left side is independent of l and the right side is independent of i, both sides must evaluate to a function $h(j, k)$ dependent solely on j and k.

Fixing $k = 1$, we obtain $f(i,j) - f(i,1) = h(j,1)$. By defining the group elements $S_i = f(i,1)$ and $T_j = -h(j,1)$, we can express the row mapping as $f(i,j) = S_i - T_j$. Applying this to the alignment condition gives $S_i - T_j + g(j,l) = c_{il}$, which rearranges to $g(j,l) - T_j = c_{il} - S_i$. Since the left side is independent of i and the right side is independent of j, they must equal a function dependent solely on l, which we define as $-U_l$. This gives the column mapping $g(j,l) = T_j - U_l$, and the target exponent evaluates to $c_{il} = S_i - U_l$.

Substituting these derived forms into the global non-interference condition evaluates to the equality:

$$(S_i - T_j) + (T_k - U_l) = S_{i'} - U_{l'} \iff (S_i - S_{i'}) + (T_k - T_j) + (U_{l'} - U_l) = 0.$$

This asserts that the sum of differences from the sets $S = \{S_i\}, T = \{T_j\}, U = \{U_l\} \subseteq G$ equals zero if and only if the differences themselves are zero. This corresponds to the definition of the Abelian Triple Product Property (TPP) for the subsets S, T, and U.

Since the target exponents $c_{il} = S_i - U_l$ must take N^2 distinct values to extract all C_{il} independently, we must have $|S| = |U| = N$. The condition $T_k - T_j = 0 \implies j = k$ ensures $|T| = N$. By the Cohn-Umans theorem on the bilinear complexity of group algebras [4], if subsets S, T, U of an abelian group G satisfy the TPP, the cardinality of the group must be bounded by $|G| \geq |S||T||U| = N^3$. Thus, the cyclotomic index must satisfy $m \geq N^3$.

In cyclotomic extensions, the algebraic degree $n = \varphi(m)$ scales asymptotically as $\Omega(m/\log\log m)$. Consequently, any single-multiplication monomial embedding into a commutative quotient ring requires a ring dimension bounded by $n = \widetilde{\Omega}(N^3)$. □

4.2 Independence of Skew Bilinear Forms

To bypass the commutative degree limits, one might consider embedding matrices into non-commutative algebras. The following lemma and theorem demonstrate that this is incompatible with standard BGV evaluations.

Lemma 2 *Let E/F be a finite Galois extension, and let $\sigma \in \mathrm{Gal}(E/F)$ be a non-trivial automorphism. There does not exist an F-linear functional $T : E \to F$ such that $T(xy) = \mathrm{Tr}_{E/F}(x\sigma(y))$ for all $x, y \in E$.*

Proof Suppose for contradiction that an F-linear functional T exists. Let $x, y \in E$ and define $z = xy$. Because z factors as $z = 1 \cdot (xy)$, the functional must evaluate identically on both representations $T(xy) = T(1 \cdot (xy))$. By the hypothesis, $T(xy) = \mathrm{Tr}_{E/F}(x\sigma(y))$. Evaluating the alternate factorization yields the substitution:

$$T(1 \cdot (xy)) = \mathrm{Tr}_{E/F}(1 \cdot \sigma(xy)) = \mathrm{Tr}_{E/F}(\sigma(xy)).$$

Because the trace form is invariant under Galois automorphisms,

$$\mathrm{Tr}_{E/F}(\sigma(xy)) = \mathrm{Tr}_{E/F}(xy).$$

Equating the evaluations provides the relation:

$$\mathrm{Tr}_{E/F}(x\sigma(y)) = \mathrm{Tr}_{E/F}(xy) \implies \mathrm{Tr}_{E/F}(x(\sigma(y) - y)) = 0.$$

This relation holds for all $x \in E$. Because the trace bilinear form $\langle a, b\rangle = \mathrm{Tr}_{E/F}(ab)$ is non-degenerate over a separable extension, the condition $\mathrm{Tr}_{E/F}(x(\sigma(y) - y)) = 0$ for all x implies that $\sigma(y) - y = 0$. Consequently, $\sigma(y) = y$ for all $y \in E$, meaning σ acts as the identity automorphism. This contradicts the assumption that σ is non-trivial. Therefore, no linear functional T exists. □

Theorem 3 *A central simple algebra, such as the skew polynomial ring $E[Y;\sigma]/\langle Y^N - 1\rangle$, cannot be embedded into a commutative quotient ring to compute the matrix product via a single tensor product.*

Proof Let E/F be a cyclic Galois extension of degree N with generator σ. Central simple algebras over F, such as the skew polynomial ring evaluated with the skew multiplication action $Y\alpha = \sigma(\alpha)Y$ for $\alpha \in E$, are isomorphic to the matrix algebra $M_N(F)$. One proposed approach to reducing the ciphertext dimension maps the matrix entries into this skew polynomial ring, utilizing the non-commutative product to extract the required coefficients.

Embedding this structure into a BGV module yields a contradiction. Since BGV modules evaluate multiplication over commutative rings, if matrix coefficients are mapped into the polynomial variable via $\psi(A) = \sum A_{ij}\theta_i Y^j$ for a basis $\{\theta_i\}$ of E/F, the evaluation for the twisted product is not supported by commutative polynomial convolutions. The expansion of $\psi(A)\psi(B)$ evaluated under commutative rules generates cross-terms of the form $A_{ij}B_{kl}\theta_i\theta_k Y^{j+l}$. Conversely, evaluating the skew rule requires $A_{ij}B_{kl}\theta_i\sigma^j(\theta_k)Y^{j+l}$.

Extracting the matrix product coefficients $C_{il} = \sum_{j=1}^{N} A_{ij}B_{jl}$ requires isolating them from the skew cross-terms $\sum_{j=1}^{N} A_{ij}B_{jl}\theta_i\sigma^j(\theta_j)Y^{j+l}$, which relies on evaluating the skew product $x\sigma^j(y)$ for $x, y \in E$.

By Lemma 2, no F-linear functional can extract the required skew trace $\mathrm{Tr}_{E/F}(x\sigma^j(y))$ from the commutative product xy for all $x, y \in E$ when σ^j is non-trivial. Therefore, no sequence of linear trace evaluations applied after a single commutative multiplication can map the commutative cross-terms xy into the skew cross-terms $x\sigma^j(y)$. A commutative ring embedding cannot evaluate an injective homomorphism from a non-commutative matrix algebra without employing $O(N)$ sequential automorphisms to evaluate $\sigma^j(B)$ prior to multiplication. This violates the single-multiplication bound ($D = 1$). □

4.3 Beating the Bound via CRT-Strassen Hybrids

Since non-commutative embeddings are incompatible, we must partition the evaluation space to achieve sub-cubic complexity. The $\widetilde{\Omega}(N^3)$ limit binds full-ring monomial embeddings. We bypass this bound via the Artinian decomposition of the cyclotomic ring.

The following theorem relaxes the global ring degree requirement by partitioning the matrix and utilizing prime ideals.

Theorem 4 *By partitioning the matrix $M_N(\mathbb{F}_p)$ into block components and embedding the sub-tensor products into distinct prime ideals of the unramified cyclotomic extension, the required ring degree is relaxed to $n = \widetilde{\Omega}(N^{\log_2 7})$.*

Proof Let the unramified principal ideal $p\mathcal{O}_K$ split into ℓ distinct prime ideals. By the Chinese Remainder Theorem, the plaintext cyclotomic ring maps isomorphically to the direct sum of finite fields $\bigoplus_{v=1}^{\ell} \mathbb{F}_{p^d}$, where $d = \mathrm{ord}_m(p)$ is the residue field degree.

Rather than embedding an $N \times N$ matrix as a single global tensor, we partition A and B into sub-matrices. Assuming $N = 2^k N_b$ for a constant base block size $N_b > 1$, each

halving step of Strassen's algorithm [15] requires 7 multiplications of linear combinations of the sub-matrices over the base field.

These pre-multiplication linear combinations are computed during the plaintext encoding phase, and the 7 independent products are routed into 7 distinct prime ideals $\mathfrak{p}_v$ using the CRT primitive orthogonal idempotents e_v. Within each local CRT slot, the polynomial evaluations undergo modular reduction by a degree-d irreducible polynomial. To satisfy the Cohn-Umans TPP bound locally and prevent exponent aliasing, the local field dimension must bound the tensor cross-terms, requiring $d = \widetilde{\Omega}((N_b)^3)$. This parallel slot evaluation is preserved through extraction because the global Galois trace operator stabilizes the primitive idempotents (as proven later in Theorem 15).

Because the global ring degree satisfies $n = \ell d$, provisioning $\ell \geq 7$ slots forces the single-multiplication degree requirement to scale as $n \geq 7 \cdot \widetilde{\Omega}((N_b)^3)$. Recursively applying this slotted tensor decomposition down to the constant base block size N_b yields $7^{\log_2(N/N_b)}$ independent parallel multiplications. Our normal-basis TPP trace evaluation is subsequently applied exclusively at this base dimension N_b to evaluate the inner products homomorphically.

To support $\ell = 7^{\log_2(N/N_b)} = O(N^{\log_2 7})$ parallel slots for a fixed prime p, the local degree d must undergo logarithmic growth $d = \Omega(\log N)$. Because the intermediate Strassen products remain algebraically isolated in orthogonal slots, homomorphically recombining them would require expensive Galois rotations. Therefore, the post-multiplication linear Strassen additions are deferred to the client-side plaintext decoding step. Factoring in the constant base block size and the logarithmic growth of d, the overall cyclotomic ring degree requirement evaluates to $n = \ell d \geq O(N^{\log_2 7}) \cdot \max(\widetilde{\Omega}(N_b^3), \Omega(\log N))$. The constant and logarithmic factors are absorbed by the soft-Omega notation, relaxing the asymptotic limit to $n = \widetilde{\Omega}(N^{\log_2 7}) \approx \widetilde{\Omega}(N^{2.81})$. □

5 Embeddings via Normal Bases

Given the polynomial degree constraints established in Section 4, we optimize the generation set of the evaluated tensors. We substitute polynomial power bases with normal bases to reduce the pre-computation cardinality and stabilize the canonical norms. The field extension $\mathbb{F}_{p^s}/\mathbb{F}_p$ guarantees a generator θ and trace dual γ_0 such that $\mathcal{N} = \{\sigma^j(\theta)\}$ and $\mathcal{N}^\vee = \{\sigma^j(\gamma_0)\}$. The non-degenerate bilinear trace form satisfies the orthogonality relation:

$$\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}\left(\sigma^i(\theta)\sigma^j(\gamma_0)\right) = \delta_{i,j}.$$

Remark 1 (The Lempel-Weinberger Parity Constraint) The Lempel-Weinberger theorem [9] establishes that a finite field extension $\mathbb{F}_{p^s}/\mathbb{F}_p$ of odd characteristic admits a self-dual normal basis (where $\gamma_0 = \theta$) if and only if the extension degree s is odd. In the cyclotomic structure, the field degree evaluates to $s = \varphi(m_2)$. Because the Euler totient function $\varphi(m_2)$ is even for integers $m_2 \geq 3$, a self-dual normal basis over the subring $\overline{R}_2$ does not exist.

Prior literature has proposed padding an $s \times s$ matrix with zeros to define an embedding into an $(s+1) \times (s+1)$ module to circumvent this parity constraint. This argument conflates the dimension of a vector space with the algebraic degree of a field extension. Matrix padding alters the rank of the encoded data vectors in $M_{s+1}(\mathbb{F}_p)$; it does not alter the algebraic degree $s = [\overline{R}_2 : \mathbb{F}_p]$ of the finite field extension. Because the Galois group order remains even, the discriminant of the trace form guarantees that an orthogonal normal basis over $\mathbb{F}_p$ cannot

exist. A normal basis defined by a distinct dual generator $\gamma_0 \neq \theta$ is therefore required unless the algebraic structure is modified.

5.1 Computation of the Dual Generator

When the extension degree s is even, an orthogonal normal basis over $\mathbb{F}_p$ does not exist. A distinct dual generator $\gamma_0 \neq \theta$ is required. This element is derived from the trace form.

Lemma 5 *Let θ generate a normal basis for $\overline{R}_2$ over $\mathbb{F}_p$. The dual generator γ_0 is determined by the inversion of an $s \times s$ symmetric circulant Gram matrix over $\mathbb{F}_p$.*

Proof Let σ denote the Frobenius automorphism generating $\mathrm{Gal}(\overline{R}_2/\mathbb{F}_p)$. Let M be the $s \times s$ Gram matrix with entries defined by $M_{i,j} = \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^i(\theta)\sigma^j(\theta))$ for $0 \leq i, j < s$. Because σ is an automorphism of the field and the trace is invariant under Galois action, we have:

$$\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^i(\theta)\sigma^j(\theta)) = \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^{-i}(\sigma^i(\theta)\sigma^j(\theta))) = \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\theta\sigma^{j-i}(\theta)).$$

Since the field extension degree is s, we have $\sigma^s = \mathrm{id}$, and thus the entries of M depend exclusively on the index difference $(j-i) \pmod s$, establishing that M is a circulant matrix. The commutativity of field multiplication yields $M_{i,j} = M_{j,i}$, so M is also symmetric. Because the trace pairing over a separable finite field extension is non-degenerate and the normal basis $\mathcal{N} = \{\sigma^i(\theta) \mid 0 \leq i < s\}$ is an $\mathbb{F}_p$-vector space basis, the Gram matrix M is non-singular.

Let M^{-1} denote the inverse of the Gram matrix over $\mathbb{F}_p$. Because the trace form is symmetric, M is symmetric, which implies that M^{-1} is also symmetric. Let the 0-th row of M^{-1} be denoted by the vector $(c_0, c_1, \ldots, c_{s-1})$. We define the element $\gamma_0 = \sum_{k=0}^{s-1} c_k\sigma^k(\theta)$. Evaluating the trace pairing provides the substitution:

$$\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^i(\theta)\gamma_0) = \sum_{k=0}^{s-1} c_k \,\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^i(\theta)\sigma^k(\theta)) = \sum_{k=0}^{s-1} c_k M_{i,k}.$$

Because M is symmetric, $M_{i,k} = M_{k,i}$. The summation $\sum_{k=0}^{s-1} c_k M_{k,i}$ corresponds to the inner product of the 0-th row of M^{-1} with the i-th column of M, evaluating to the matrix entry $(M^{-1}M)_{0,i} = \delta_{0,i}$. Applying the inverse field automorphism σ^{-j} and the Galois invariance of the trace once more, we obtain:

$$\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^i(\theta)\sigma^j(\gamma_0)) = \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^{-j}(\sigma^i(\theta)\sigma^j(\gamma_0))) = \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^{i-j}(\theta)\gamma_0) = \delta_{i-j \bmod s,0} = \delta_{i,j}.$$

Therefore, γ_0 generates the trace dual normal basis $\mathcal{N}^\vee$. □

5.2 Self-Dual Normal Bases via Maximal Real Subfields

To avoid storing a distinct dual generator, we map the inner tensor algebra to a subfield with an odd extension degree.

Theorem 6 *Let $\overline{R}_2^+ = \mathbb{Z}[x]/\langle\Psi_{m_2}(x)\rangle \pmod p$ be the algebra associated with the maximal real subfield $K^+ = \mathbb{Q}(\zeta_{m_2} + \zeta_{m_2}^{-1})$, where $\Psi_{m_2}(x)$ is its minimal polynomial. For specific moduli m_2, $\overline{R}_2^+$ admits a Self-Dual Normal Basis (SDNB) over $\mathbb{F}_p$, reducing the required basis generators to one element ($\gamma_0 = \theta$).*

Proof The degree of K^+ over $\mathbb{Q}$ is $s' = \varphi(m_2)/2$. By selecting specific moduli (e.g., $m_2 \in \{7, 9, 14, 18\}$), the extension degree s' evaluates to an odd integer (e.g., $s' = 3$). Assuming p generates the Galois group $\mathrm{Gal}(K^+/\mathbb{Q})$, the ideal generated by p is inert in the ring of integers $\mathcal{O}_{K^+}$. Consequently, the polynomial $\Psi_{m_2}(x)$ is irreducible modulo p, and the quotient algebra $\overline{R}_2^+$ forms a finite field extension of $\mathbb{F}_p$ of degree s'.

The Lempel-Weinberger theorem establishes that a finite field extension $\mathbb{F}_{p^{s'}}/\mathbb{F}_p$ admits a self-dual normal basis if and only if the extension degree s' is odd (for odd characteristic p) or $s' \not\equiv 0 \pmod 4$ (for characteristic $p = 2$). Since s' is constrained to be odd by the choice of m_2, this existence condition is satisfied for any valid prime p.

There exists an element $\theta \in \overline{R}_2^+$ such that the normal basis $\mathcal{N} = \{\sigma^i(\theta) \mid 0 \leq i < s'\}$ is self-dual, satisfying the orthogonality condition $\mathrm{Tr}_{\overline{R}_2^+/\mathbb{F}_p}(\sigma^i(\theta)\sigma^j(\theta)) = \delta_{i,j}$. Because the primary and dual generators coincide ($\gamma_0 = \theta$), both matrix operands A and B can be encoded utilizing the identical single generator θ, which halves the pre-computation storage requirements and uniformizes the canonical noise bounds. □

Remark 2 (Halved Trace Automorphism Depth) Because elements in the maximal real subfield K^+ are invariant under complex conjugation, applying the standard full trace $\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}$ (which evaluates s automorphisms) to an encoded tensor would compute the subfield trace twice, yielding a structural scalar of 2 (mod p). By restricting the homomorphic extraction algorithm to evaluate exclusively the $s' = s/2$ automorphisms corresponding to $\mathrm{Gal}(K^+/\mathbb{Q})$, the operation naturally maps $\overline{R}_2^+$ directly to $\mathbb{F}_p$. This yields exactly the required $\delta_{i,j}$ components without any scalar correction, and crucially, structurally halves the required homomorphic automorphism depth during the final evaluation phase.

5.3 Minimization via Optimal Normal Bases

If the specific parameters prohibit the use of the maximal real subfield, we bind the dual expansion via optimal normal bases.

Theorem 7 *If the subring $\overline{R}_2$ admits a Type I Optimal Normal Basis (ONB), the canonical norm of the integer polynomial lift of the dual generator is bounded by $\|\tilde{\gamma}_0\|_{\mathrm{can}} \leq |C|\|\tilde{\theta}\|_{\mathrm{can}} + |C|$, where $C \in [-(p-1)/2, (p-1)/2]$ is the symmetric integer representative of $(s+1)^{-1} \pmod p$. This structure eliminates the fractional variance associated with standard Gram matrix inversion, ensuring the dual expansion bound is uniform across all basis elements.*

Proof A Type I ONB exists for the field extension $\mathbb{F}_{p^s}/\mathbb{F}_p$ if $s+1$ is prime and p generates the multiplicative group $(\mathbb{Z}/(s+1)\mathbb{Z})^\times$ [12]. In this configuration, the normal basis is generated by an element θ, which corresponds to a primitive $(s+1)$-th root of unity. The basis elements over $\mathbb{F}_p$ are the s non-trivial roots of $x^{s+1} - 1 = 0$.

Let σ denote the Frobenius automorphism $x \mapsto x^p$. Because p generates $(\mathbb{Z}/(s+1)\mathbb{Z})^\times$, the Galois conjugates $\sigma^k(\theta)$ enumerate all primitive $(s+1)$-th roots of unity. There exists an integer j such that $p^j \equiv -1 \pmod{s+1}$. Since the multiplicative order of p modulo $s+1$ is s, this congruence is satisfied at $j = s/2$. Therefore, $\sigma^{s/2}(\theta) = \theta^{p^{s/2}} = \theta^{-1}$, yielding $\theta\sigma^{s/2}(\theta) = 1$. The trace evaluates to $\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(1) = s \pmod p$. For all other cross-terms $i \not\equiv s/2 \pmod s$, the product $\theta\sigma^i(\theta)$ remains a non-trivial primitive $(s+1)$-th root of unity.

Because the sum of all $(s+1)$-th roots of unity is 0, the trace of any primitive root evaluates to $-1 \pmod p$.

We solve for the dual generator γ_0 by assuming the affine form $\gamma_0 = c\sigma^{s/2}(\theta) + d$ over $\mathbb{F}_p$. The dual basis condition requires $\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^k(\theta)\gamma_0) = \delta_{k,0}$. Evaluating the trace for $k \neq 0$ yields the evaluation:

$$\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^k(\theta)(c\sigma^{s/2}(\theta) + d)) = c(-1) + d(-1) = 0 \implies d = -c.$$

Evaluating the trace for $k = 0$ yields the equivalence:

$$\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\theta(c\sigma^{s/2}(\theta) - c)) = c(s) - c(-1) = c(s+1) \equiv 1 \pmod p \implies c \equiv (s+1)^{-1} \pmod p.$$

Thus, the dual generator over $\mathbb{F}_p$ is $\gamma_0 = (s+1)^{-1}\sigma^{s/2}(\theta) - (s+1)^{-1}$.

When lifting the basis elements to the canonical domain over $\mathbb{Z}[x]$, let $\tilde{\theta} \in \mathbb{Z}[x]$ be an integer polynomial lift of θ. The trace scalar $(s+1)^{-1} \pmod p$ lifts to a symmetric integer representative $C \in [-(p-1)/2, (p-1)/2]$.

The Galois automorphism $\sigma^{s/2}$ lifts to the polynomial substitution $x \mapsto x^{p^{s/2}}$. Because p is coprime to the cyclotomic index m of the global ring, this substitution maps primitive m-th roots of unity bijectively to other primitive m-th roots of unity. Thus, it acts as an isometry under the canonical embedding, preserving the canonical norm $\|\tilde{\theta}(x^{p^{s/2}})\|_{\mathrm{can}} = \|\tilde{\theta}\|_{\mathrm{can}}$.

Defining the dual lift as $\tilde{\gamma}_0(x) = C\tilde{\theta}(x^{p^{s/2}}) - C \in \mathbb{Z}[x]$ and applying the triangle inequality bounds the polynomial magnitude:

$$\|\tilde{\gamma}_0\|_{\mathrm{can}} = \|C\tilde{\theta}(x^{p^{s/2}}) - C\|_{\mathrm{can}} \leq |C|\|\tilde{\theta}(x^{p^{s/2}})\|_{\mathrm{can}} + |C| = |C|\|\tilde{\theta}\|_{\mathrm{can}} + |C|.$$

This affine permutation bounds the error norm by a uniform scalar shift, bypassing the asymmetric scaling sets and fractional variance typically induced by inverting an arbitrary Gram matrix over $\mathbb{Z}[x]$. □

Algorithm 1 Constructive Normal Basis Embedding

Input: Matrices $A \in M_{r\times s}(\mathbb{F}_p), B \in M_{s\times t}(\mathbb{F}_p)$, bases $\{u_i\}_{i=0}^{r-1}$ of $\overline{R}_1$, $\{w_k\}_{k=0}^{t-1}$ of $\overline{R}_3$, generators $\theta, \gamma_0 \in \overline{R}_2$

Output: Polynomial encodings $P_A, P_B \in \overline{R}_1 \otimes_{\mathbb{F}_p} \overline{R}_2 \otimes_{\mathbb{F}_p} \overline{R}_3 \cong \overline{R}$

Initialize $P_A = 0, P_B = 0$
for $i = 0, \ldots, r-1$ **do**
  for $j = 0, \ldots, s-1$ **do**
    $P_A \leftarrow P_A + a_{ij}\left(u_i \otimes \sigma^j(\theta) \otimes 1\right)$
  end for
end for
for $j = 0, \ldots, s-1$ **do**
  for $k = 0, \ldots, t-1$ **do**
    $P_B \leftarrow P_B + b_{jk}\left(1 \otimes \sigma^j(\gamma_0) \otimes w_k\right)$
  end for
end for
return P_A, P_B

6 Extraction of the Matrix Trace

With the bases defined, we extract the matrix product.

Theorem 8 *Given the normal basis $\mathcal{N}$ and its trace dual $\mathcal{N}^\vee$, the homomorphic trace evaluation over the subring $\overline{R}_{13}$ evaluates to $\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}(P_A P_B) = \sum_{i,k} c_{ik}\,(u_i \otimes 1 \otimes w_k)$, where $c_{ik} = \sum_{j=0}^{s-1} a_{ij} b_{jk}$.*

Proof The trace operator $\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}$ acts as the identity on the subrings $\overline{R}_1$ and $\overline{R}_3$. Therefore, the trace decomposes over the tensor product as $\mathrm{id}_{\overline{R}_1} \otimes \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p} \otimes \mathrm{id}_{\overline{R}_3}$. Applying the bilinearity of the tensor product to the polynomial product $P_A P_B$ yields the expansion:

$$P_A P_B = \sum_{i=0}^{r-1} \sum_{k=0}^{t-1} \sum_{j=0}^{s-1} \sum_{l=0}^{s-1} a_{ij} b_{lk} \left(u_i \otimes \left(\sigma^j(\theta) \sigma^l(\gamma_0) \right) \otimes w_k \right).$$

Evaluating the trace operator isolates the $\overline{R}_2$ component securely. For standard normal bases, the trace evaluates s automorphisms over $\overline{R}_2$. For SDNB encodings mapped into the maximal real subfield K^+ (as established in Theorem 6 and Remark 2), the trace evaluates exclusively over the $s' = s/2$ automorphisms of $\mathrm{Gal}(K^+/\mathbb{Q})$ to bypass the duplication scalar $2 \pmod p$:

$$\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}(P_A P_B) = \sum_{i=0}^{r-1} \sum_{k=0}^{t-1} \sum_{j=0}^{s-1} \sum_{l=0}^{s-1} a_{ij} b_{lk}\, \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p} \left(\sigma^j(\theta) \sigma^l(\gamma_0) \right) (u_i \otimes 1 \otimes w_k).$$

Since the trace map $\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}$ is invariant under the action of the Galois group, applying the inverse field automorphism σ^{-l} extracts the delta relation:

$$\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p} \left(\sigma^j(\theta) \sigma^l(\gamma_0) \right) = \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p} \left(\sigma^{-l} \left(\sigma^j(\theta) \sigma^l(\gamma_0) \right) \right) = \mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p} \left(\sigma^{j-l}(\theta) \gamma_0 \right).$$

By the definition of the trace dual basis, $\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p} \left(\sigma^{j-l}(\theta) \gamma_0 \right) = \delta_{j-l \bmod s, 0}$. The trace evaluates to zero unless $j \equiv l \pmod s$. The summation collapses to $j = l$, yielding the matrix trace vector:

$$\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}(P_A P_B) = \sum_{i=0}^{r-1} \sum_{k=0}^{t-1} \left(\sum_{j=0}^{s-1} a_{ij} b_{jk} \right) (u_i \otimes 1 \otimes w_k),$$

recovering the matrix product coefficients c_{ik}. □

7 Uniformity of Canonical Norms

The error growth during homomorphic evaluation depends on the initial norms of the encoded polynomials. Power bases yield a constant norm for the primary basis but divergent norms for the trace dual. Under the power bases utilized in prior systems, the primary power basis satisfies $v_j = x^j$, yielding a constant canonical norm $\|v_j\|_{\mathrm{can}} = 1$. However, the inverted dual basis elements $v_j^\vee$ possess canonical norms that vary with respect to the index j. This variance introduces an imbalance. The encoding error for matrix B is dominated by the supremum $E_{\mathrm{dual}} = \max_j \|v_j^\vee\|_{\mathrm{can}}$.

While the canonical norm of an algebraic integer is invariant under the Frobenius automorphism σ^j, this symmetry is not preserved during encryption residue class

selection. If the integer coefficients in $\mathbb{Z}[x]$ are mapped into the symmetric interval $[-(p-1)/2, (p-1)/2]$, the modular reduction over $\mathbb{Z}$ modifies the polynomial structure, losing the norm invariance over $\mathbb{C}$.

To achieve intra-basis norm uniformity, we evaluate canonical polynomial lifts over $\mathbb{Z}$.

Theorem 9 *Let $\tilde{\theta}, \tilde{\gamma}_0 \in \mathbb{Z}[x]/\langle \Phi_m(x) \rangle$ be canonical integer polynomial lifts of the normal basis generators. For all $0 \leq j \leq s-1$, the canonical norms satisfy $\|\tilde{\theta}(x^{p^j})\|_{\text{can}} = \|\tilde{\theta}\|_{\text{can}}$ and $\|\tilde{\gamma}_0(x^{p^j})\|_{\text{can}} = \|\tilde{\gamma}_0\|_{\text{can}}$.*

Proof The canonical norm is $\|A\|_{\text{can}} = \max_{\gcd(k,m)=1} |A(\zeta_m^k)|$. For the primary generator, the norm of the substituted polynomial is defined natively:

$$\|\tilde{\theta}(x^{p^j})\|_{\text{can}} = \max_{\gcd(k,m)=1} |\tilde{\theta}((\zeta_m^k)^{p^j})| = \max_{\gcd(k,m)=1} |\tilde{\theta}(\zeta_m^{kp^j})|.$$

Since $\gcd(p, m) = 1$, the map $k \mapsto kp^j \pmod m$ acts as a permutation on the group of units $(\mathbb{Z}/m\mathbb{Z})^\times$. Thus, the set $\{\zeta_m^{kp^j} \mid \gcd(k, m) = 1\}$ is identical to the set $\{\zeta_m^k \mid \gcd(k, m) = 1\}$. The maximum absolute value taken over these roots is invariant under the substitution:

$$\max_{\gcd(k,m)=1} |\tilde{\theta}(\zeta_m^{kp^j})| = \max_{\gcd(k,m)=1} |\tilde{\theta}(\zeta_m^k)| = \|\tilde{\theta}\|_{\text{can}}.$$

The same argument holds for $\tilde{\gamma}_0(x^{p^j})$. The canonical norm is therefore uniform across all basis elements, and the magnitude of the substituted polynomials is bounded by the norm of the original generators. □

This uniformity simplifies the error bounds during homomorphic evaluation. By bounding the plaintext elements uniformly, we bound the multiplicative error expansion.

Lemma 10 *Let $c_A, c_B \in R_q^2$ be ciphertexts encrypting the matrices P_A, P_B with initial error polynomials bounded by N_A and N_B in canonical norm. Let the plaintext basis elements be bounded by the uniform canonical norms $\|\sigma^j(\tilde{\theta})\|_{\text{can}} \leq E_\theta$ and $\|\sigma^j(\tilde{\gamma}_0)\|_{\text{can}} \leq E_\gamma$. The trace operator outputs a final vector encrypting AB with an error polynomial whose canonical norm satisfies the inequality:*

$$E_{\text{final}} \leq s\,(E_{\text{mult}} + B_{\text{ks}}) + (s-1)B_{\text{ks}},$$

where $E_{\text{mult}} = \|P_A\|_{\text{can}} N_B + \|P_B\|_{\text{can}} N_A + pN_AN_B$ is the canonical error bound of the initial tensor multiplication.

Proof The homomorphic multiplication of c_A and c_B evaluates the tensor product of the ciphertexts. By the sub-multiplicativity of the canonical norm over $\mathbb{C}$, the error polynomial of this degree-2 ciphertext is bounded by $E_{\text{mult}} = \|P_A\|_{\text{can}} N_B + \|P_B\|_{\text{can}} N_A + pN_AN_B$. Returning this expanded ciphertext to a standard degree-1 vector in R_q^2 requires a relinearization step, which introduces an additive key-switching error bounded by B_{ks}. Thus, the base ciphertext c entering the trace evaluation has an initial error norm bounded by $E_{\text{mult}} + B_{\text{ks}}$.

The trace extraction evaluates the field summation $\sum_{i=0}^{s-1} \sigma^i(c)$. Because the canonical norm is invariant under the field automorphisms σ^i (Theorem 9), the initial error of c is scaled symmetrically by a factor of s across the summation, contributing $s(E_{\text{mult}} + B_{\text{ks}})$ to the final error bound.

Evaluating this summation via the binary tree recurrence defined in Algorithm 2 introduces additional additive error from the key-switching operations required to restore the permuted secret keys. Let N_k denote the accumulated key-switching error of a partial trace sum evaluated over k terms, excluding the base error (i.e., $N_1 = 0$). The recursive halving step computes a $2h$-term sum by adding an h-term sum to a shifted h-term sum. The shift requires one key-switching operation, yielding the difference equation $N_{2h} \leq 2N_h + B_{\text{ks}}$. For an odd number of summation terms $2h + 1$, the algorithm adds a shifted residual base term to the $2h$-term sum. This residual shift requires one additional key-switching operation, satisfying the relation $N_{2h+1} \leq N_{2h} + B_{\text{ks}}$.

Resolving these relations iteratively over the binary evaluation tree tracking s summation terms yields an accumulated additive key-switching error bounded by $(s-1)B_{\text{ks}}$. Summing this tree error with the scaled base error provides the final bound.

Because E_γ is bounded by the single norm $\|\tilde{\gamma}_0\|_{\text{can}}$ (which can be minimized to $O(1)$ via ONBs or equalized to E_θ via SDNBs) rather than a supremum over a divergent set, the linear noise expansion associated with the plaintext bound $\|P_B\|_{\text{can}}$ is optimized. □

8 Canonical Expansion Minimum

The trace operator acts as the identity on the outer subrings $\overline{R}_1$ and $\overline{R}_3$. Because dual inversion is not required here, we evaluate the optimal basis for these specific components.

Lemma 11 (Kronecker's Theorem) *Let $\alpha \in \mathbb{C}$ be a non-zero algebraic integer. If all Galois conjugates of α over $\mathbb{Q}$ have absolute value at most 1, then α is a root of unity.*

Theorem 12 *The standard polynomial power basis minimizes the maximum canonical norm of its constituent elements among all full-rank integral bases for the non-extracted tensor components R_1 and R_3.*

Proof Let $P = \{1, x, x^2, \ldots, x^{\varphi(m_i)-1}\}$ define the power basis of $R_i = \mathbb{Z}[x]/\langle \Phi_{m_i}(x) \rangle$. The canonical embedding σ maps each basis element x^k to the vector of its evaluations at the primitive m_i-th roots of unity. Because the absolute value of any root of unity over $\mathbb{C}$ is 1, the canonical norm evaluates to $\|x^k\|_{\text{can}} = 1$ for all elements in P.

Let $U = \{u_0, \ldots, u_{\varphi(m_i)-1}\}$ define an arbitrary full-rank integral $\mathbb{Z}$-module basis for R_i. Under the isomorphism $R_i \cong \mathcal{O}_{\mathbb{Q}(\zeta_{m_i})}$, every basis element $u_j \in U$ maps to an algebraic integer $u_j(\zeta_{m_i})$. Because 0 cannot belong to a basis, $u_j \neq 0$ in R_i, meaning its evaluation satisfies $u_j(\zeta_{m_i}) \neq 0$. The coordinates of the canonical embedding $\sigma(u_j)$ correspond to the Galois conjugates of $u_j(\zeta_{m_i})$ over $\mathbb{Q}$. Therefore, the canonical norm $\|u_j\|_{\text{can}}$ equals the maximum absolute value among all Galois conjugates of $u_j(\zeta_{m_i})$.

By Kronecker's theorem on algebraic integers (Lemma 11), if a non-zero algebraic integer α satisfies $|\tau(\alpha)| \leq 1$ for all its Galois conjugates $\tau(\alpha)$, then α is a root of unity. Consequently, the algebraic integer $u_j(\zeta_{m_i})$ must either be a root of unity or possess at least one Galois

conjugate with an absolute value strictly greater than 1. In either case, the maximum absolute value is bounded by 1, yielding $\|u_j\|_{\text{can}} \geq 1$ for all $u_j \in U$.

No full-rank integral basis can yield a maximum canonical norm less than 1 for its elements. Since the standard power basis achieves a uniform canonical norm of 1 across all its elements, it minimizes this maximum canonical norm. □

9 Logarithmic and Towered Trace Evaluations

The trace evaluation requires $O(s)$ sequential additions and automorphisms. We optimize this sequence to logarithmic depth and apply a towered decomposition to bound noise accumulation.

9.1 Logarithmic Trace Evaluation

The trace operator evaluates the sum $\text{Tr}(c) = \sum_{i=0}^{s-1} \sigma^i(c)$ over the cyclotomic ring. We define the partial sum $S_k(c) = \sum_{i=0}^{k-1} \sigma^i(c)$ and compute it recursively. For an even number of terms, $S_{2h}(c) = S_h(c) + \sigma^h(S_h(c))$, requiring one key-switching operation at each halving step. For an odd number of terms, the recurrence requires an additional residual term $S_{2h+1}(c) = S_{2h}(c) + \sigma^{2h}(c)$.

A direct evaluation of this recurrence requires sequential applications of the digit extraction map $\mathcal{D}_B : R_q \to R_q^{D+1}$ to the shifted polynomials at each step. However, because the base B is an integer, decomposition commutes with the Galois automorphism σ^h. If a polynomial $a(x)$ decomposes as $\sum d^{(m)}(x)B^m$, then $\sigma^h(a(x)) = a(x^h) = \sum d^{(m)}(x^h)B^m$. This equality holds in R_q. While reducing the shifted digits $d^{(m)}(x^h)$ modulo $\Phi_m(x)$ over $\mathbb{Z}$ expands their coefficient infinity norms, the automorphism σ^h permutes the complex roots of unity, acting as an isometry under the canonical embedding. Because BGV decryption correctness and noise bounds rely on the canonical norm, we can apply the automorphism directly to the previously extracted base-B digits without inflating the error, avoiding the need to decompose the fully reduced shifted polynomials.

Theorem 13 *Algorithm 2 evaluates the Galois trace operator using $\lfloor \log_2 s \rfloor$ applications of the digit extraction map $\mathcal{D}_B$ and requires $\lfloor \log_2 s \rfloor + \text{hw}(s) - 1$ key-switching operations, where $\text{hw}(s)$ is the Hamming weight of the binary representation of s. Applying the field automorphism to the extracted digits prior to the inner product preserves the additive error bound B_{ks}.*

Proof In the BGV scheme, a ciphertext is a vector $c = (c_0, c_1) \in R_q^2$, and the digit extraction map $\mathcal{D}_B$ is applied exclusively to the masking polynomial c_1. The base decomposition of the initial masking polynomial $c_1 = \sum_{m=0}^{D} c_1^{(m)} B^m$ is computed once prior to the recursion. This cached decomposition is utilized for the odd residual evaluations and the base case $h = 1$.

During the recursive halving step $h = \lfloor k/2 \rfloor$, the algorithm evaluates the partial trace ciphertext $c_{\text{half}} = S_h(c)$. Computing the shifted term $\sigma^h(c_{\text{half}})$ requires decomposing the new masking polynomial $c_{\text{half},1}$ into digits $d_1^{(m)}$.

Under the canonical embedding into $\mathbb{C}$, the field automorphism σ^h is an isometry, meaning $\|\sigma^h(d_1^{(m)})\|_{\text{can}} = \|d_1^{(m)}\|_{\text{can}}$. Although reducing the shifted polynomial modulo $\Phi_m(x)$ over $\mathbb{Z}$ expands its coefficient infinity norm, BGV decryption correctness relies on the canonical norm, which remains invariant. Therefore, applying the evaluation keys directly to the permuted digits $\sigma^h(d_1^{(m)})$ incurs the same bounded key-switching error B_{ks} as decomposing a newly reduced polynomial.

The algorithm decomposes $c_{\text{half},1}$ once per halving step for $h > 1$, resulting in $\lfloor \log_2 s \rfloor - 1$ recursive decompositions. Including the initial cached decomposition, the total required applications of $\mathcal{D}_B$ is $\lfloor \log_2 s \rfloor$.

The number of key-switching inner products executed for the halving steps is $\lfloor \log_2 s \rfloor$. For the odd residual step $\sigma^{k-1}(c)$, the algorithm applies the automorphism directly to the cached initial decomposition digits $c_1^{(m)}$. These residual additions contribute $\text{hw}(s) - 1$ operations, yielding a combined upper bound of $\lfloor \log_2 s \rfloor + \text{hw}(s) - 1$ key-switching operations. □

Algorithm 2 Logarithmic Trace Evaluation Sequence

Input: Ciphertext $c = (c_0, c_1) \in R_q^2$, degree s, base B, bound D, Evaluation keys evk

Output: Trace ciphertext $c_{\text{out}} = \sum_{i=0}^{s-1} \sigma^i(c)$

Decompose the initial masking polynomial $c_1 = \sum_{m=0}^{D} c_1^{(m)} B^m$
function TRACESUM($c, \{c_1^{(m)}\}, k$)
  if $k = 1$ **then return** c
  end if
  $h \leftarrow \lfloor k/2 \rfloor$
  $c_{\text{half}} \leftarrow \text{TraceSum}(c, \{c_1^{(m)}\}, h)$
  if $h = 1$ **then**
    $\{d_1^{(m)}\} \leftarrow \{c_1^{(m)}\}$
  else
    Decompose the masking polynomial $c_{\text{half},1} = \sum_{m=0}^{D} d_1^{(m)} B^m$
  end if
  $c_{\text{eval}} \leftarrow (\sigma^h(c_{\text{half},0}), 0) + \sum_{m=0}^{D} \sigma^h(d_1^{(m)}) \cdot \text{evk}_{\sigma^h}^{(m)}$
  $c_{\text{next}} \leftarrow c_{\text{half}} + c_{\text{eval}}$
  if $k \equiv 1 \pmod 2$ **then**
    $c_{\text{rem}} \leftarrow (\sigma^{k-1}(c_0), 0) + \sum_{m=0}^{D} \sigma^{k-1}(c_1^{(m)}) \cdot \text{evk}_{\sigma^{k-1}}^{(m)}$
    $c_{\text{next}} \leftarrow c_{\text{next}} + c_{\text{rem}}$
  end if
  return c_{next}
end function
$c_{\text{out}} \leftarrow \text{TraceSum}(c, \{c_1^{(m)}\}, s)$
return c_{out}

9.2 Towered Trace Evaluations

While Algorithm 2 bounds the number of key-switching operations, additive noise still accumulates across the evaluation tree. We limit this accumulation by factoring the field extension degree.

Theorem 14 *For a composite extension degree $s = d_1 d_2$, evaluating the trace via the field tower $\mathrm{Tr}_{\mathbb{F}_{p^s}/\mathbb{F}_p} = \mathrm{Tr}_{\mathbb{F}_{p^{d_1}}/\mathbb{F}_p} \circ \mathrm{Tr}_{\mathbb{F}_{p^s}/\mathbb{F}_{p^{d_1}}}$ with an intermediate modulus switch reduces the additive key-switching noise bound from $O(s)B_{\mathrm{ks}}$ to $O(d_1 + d_2)B_{\mathrm{ks}}$.*

Proof By the Fundamental Theorem of Galois Theory, the trace map satisfies:

$$\mathrm{Tr}_{\mathbb{F}_{p^s}/\mathbb{F}_p} = \mathrm{Tr}_{\mathbb{F}_{p^{d_1}}/\mathbb{F}_p} \circ \mathrm{Tr}_{\mathbb{F}_{p^s}/\mathbb{F}_{p^{d_1}}} .$$

The inner trace evaluates d_2 automorphisms over the Galois subgroup generated by σ^{d_1}. This operation accumulates an additive key-switching error bounded by $O(d_2)B_{\mathrm{ks}}$.

Without a modulus switch, the outer trace evaluates a sum of d_1 automorphisms on the output of the inner trace. Because automorphisms are linear maps that preserve the canonical norm, this summation scales the accumulated inner error by a factor of d_1, resulting in a total additive error of $O(d_1 d_2)B_{\mathrm{ks}} = O(s)B_{\mathrm{ks}}$.

We pause the evaluation after the inner trace and apply a homomorphic modulus switch, scaling the intermediate ciphertext down from modulus q to q'. This reduction divides the accumulated $O(d_2)B_{\mathrm{ks}}$ error by the scaling ratio q/q', absorbing it into the rounding noise. The outer trace of degree d_1 is then applied to the scaled ciphertext, accumulating a new, independent additive error bounded by $O(d_1)B_{\mathrm{ks}}$. By decoupling the trace stages, the total additive error bound becomes $O(d_1 + d_2)B_{\mathrm{ks}}$. For a fully factored degree s, this approach recursively reduces the linear error growth $O(s)$ to a logarithmic $O(\log s)$ bound, allowing for a smaller initial ciphertext modulus Q. □

10 Ideal Factorization Constraints for Parallel Evaluations

Evaluating multiple matrix products in parallel relies on the Chinese Remainder Theorem. We must ensure the trace automorphisms do not permute the underlying ideal slots.

The dimension of the global ring R is larger than the N^2 elements comprising a single encoded matrix. By the Artin-Wedderburn theorem, the quotient ring $\overline{R} = R/pR$ decomposes into a direct sum of finite fields, supporting the simultaneous evaluation of ℓ parallel matrix multiplications across distinct prime ideals via the Chinese Remainder Theorem (CRT). However, evaluating trace automorphisms over arbitrary ideal decompositions can cause the loss of orthogonality among the block idempotents.

The extraction trace operator $\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}$ applies cyclotomic automorphisms $\tau_k(x) = x^k$, where $k \equiv 1 \pmod{m_1 m_3}$.

Theorem 15 *Let $\overline{R} \cong \bigoplus_{v=1}^{\ell} \mathbb{F}_{p^d}$ be the decomposition of the cyclotomic quotient ring into residue fields over $\mathbb{F}_p$. The primitive orthogonal idempotents defining the parallel matrix slots are preserved under Galois trace evaluations if and only if $\gcd(\mathrm{ord}_{m_2}(p), \mathrm{ord}_{m_1 m_3}(p)) = 1$.*

Proof Let $p\mathcal{O}_K = \mathfrak{p}_1 \cdots \mathfrak{p}_\ell$ be the factorization of the prime ideal $p\mathcal{O}_K$ in the m-th cyclotomic field $K = \mathbb{Q}(\zeta_m)$. By the Chinese Remainder Theorem, the plaintext ring decomposes as $\mathcal{O}_K/p\mathcal{O}_K \cong \bigoplus_{v=1}^{\ell} \mathcal{O}_K/\mathfrak{p}_v$. The Galois group $\mathrm{Gal}(K/\mathbb{Q}) \cong (\mathbb{Z}/m\mathbb{Z})^\times$ acts transitively on the prime ideals $\mathfrak{p}_v$. Let e_v denote the primitive orthogonal idempotents of this decomposition. For the encoded matrix coefficients mapped into a specific residual field $\mathcal{O}_K/\mathfrak{p}_v$ to remain invariant under a Galois automorphism $\tau_k : x \mapsto x^k$, the corresponding automorphism must stabilize the primitive idempotent e_v.

This condition requires that the automorphism belongs to the decomposition group $D_{\mathfrak{p}_v} = \{\sigma \in \mathrm{Gal}(K/\mathbb{Q}) \mid \sigma(\mathfrak{p}_v) = \mathfrak{p}_v\}$. In the unramified cyclotomic extension, the decomposition group is cyclic and generated by the Frobenius element σ_p, which acts as $\sigma_p(\alpha) \equiv \alpha^p \pmod{\mathfrak{p}_v}$ for $\alpha \in \mathcal{O}_K$.

The trace map $\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}$ evaluates the s automorphisms τ_k corresponding to the Galois group of the m_2 component evaluated over $\mathbb{F}_p$. These automorphisms are defined by the simultaneous congruences $k \equiv p^j \pmod{m_2}$ and $k \equiv 1 \pmod{m_1 m_3}$ for $0 \le j < s$, where $s = \mathrm{ord}_{m_2}(p)$.

For these trace automorphisms to stabilize the primitive idempotents, there must exist an integer c for each index j such that $\tau_k = \sigma_p^c$. Since σ_p^c applies the mapping $x \mapsto x^{p^c}$, this group equality implies that the exponents satisfy $p^c \equiv p^j \pmod{m_2}$ and $p^c \equiv 1 \pmod{m_1 m_3}$. Let $t = \mathrm{ord}_{m_1 m_3}(p)$. Because p has multiplicative order s modulo m_2 and order t modulo $m_1 m_3$, these exponential congruences translate to the linear modular equivalence:

$$\begin{aligned} c &\equiv j \pmod{s}, \\ c &\equiv 0 \pmod{t}. \end{aligned}$$

By the generalized Chinese Remainder Theorem (Bézout's identity for systems of congruences), this system admits a solution for c if and only if $\gcd(s,t)$ divides the difference $j - 0 = j$. For the trace extraction to operate without mixing the parallel blocks, this condition must hold for all $0 \le j < s$. Since $s \ge 2$, evaluating the condition at $j = 1$ requires $\gcd(s,t)$ to divide 1, which is equivalent to the coprimality condition $\gcd(s,t) = 1$. Under this condition, the global idempotents e_v are stabilized for all terms in the trace summation, isolating the subrings and preventing the mixing of matrix coefficients across ideals. □

11 Generalization to Galois Rings

For applications requiring integer arithmetic rather than finite field operations, we lift the embedding to Galois rings.

Theorem 16 *The normal basis bilinear embedding over $\mathbb{F}_{p^s}$ lifts to the Galois ring $\mathrm{GR}(p^r, s) \cong \mathbb{Z}_{p^r}[x]/\langle f(x)\rangle$, where $f(x)$ is a basic irreducible factor of $\Phi_{m_2}(x)$ over $\mathbb{Z}_{p^r}$, supporting single-multiplication matrix evaluation over $M_N(\mathbb{Z}_{p^r})$.*

Proof Because $p \nmid m_2$, the cyclotomic polynomial $\Phi_{m_2}(x)$ is separable modulo p and factors into distinct irreducible polynomials of degree $s = \mathrm{ord}_{m_2}(p)$ over $\mathbb{F}_p$. Let $\bar{f}(x) \in \mathbb{F}_p[x]$ be one such irreducible factor. By Hensel's Lemma, $\bar{f}(x)$ lifts to a unique monic polynomial $f(x) \in \mathbb{Z}_{p^r}[x]$ that divides $\Phi_{m_2}(x)$ over $\mathbb{Z}_{p^r}$. The localized quotient ring $\mathrm{GR}(p^r, s) \cong \mathbb{Z}_{p^r}[x]/\langle f(x)\rangle$ is a Galois ring.

The Galois ring is equipped with a generalized Frobenius automorphism Σ that generates the unramified Galois group of $\mathrm{GR}(p^r, s)$ over $\mathbb{Z}_{p^r}$, defined by $\Sigma(x) \equiv x^p \pmod{p\,\mathrm{GR}(p^r, s)}$.

Let $\theta \in \mathbb{F}_{p^s}$ generate a normal basis over $\mathbb{F}_p$, and let $\Theta \in \mathrm{GR}(p^r, s)$ be an arbitrary lift of θ. Because the elements $\Sigma^i(\Theta)$ reduce to an $\mathbb{F}_p$-basis modulo p, they form a free $\mathbb{Z}_{p^r}$-module basis for $\mathrm{GR}(p^r, s)$ by Nakayama's Lemma.

The non-degenerate trace operator $\mathrm{Tr} : \mathrm{GR}(p^r, s) \to \mathbb{Z}_{p^r}$ defines a bilinear form over the local ring. The Gram matrix associated with the basis $\{\Sigma^i(\Theta)\}$ reduces modulo p to a non-singular matrix over $\mathbb{F}_p$. Because any matrix over $\mathbb{Z}_{p^r}$ with a determinant coprime to p is invertible over $\mathbb{Z}_{p^r}$, there exists a unique trace dual basis generated by some element $\Gamma_0 \in \mathrm{GR}(p^r, s)$ satisfying the orthogonality condition $\mathrm{Tr}(\Sigma^i(\Theta)\Sigma^j(\Gamma_0)) = \delta_{i,j}$.

Because this bilinear orthogonality is preserved over $\mathbb{Z}_{p^r}$, the internal polynomial cross-terms within the tensor evaluation collapse as they do over the base field. The normal basis tensor extraction computes matrix products over $\mathbb{Z}_{p^r}$, extending the utility of the embedding to high-precision integer inputs without modulo p overflow. □

12 Extension to Boundless Matrix Dimensions

While Section 4 demonstrates how to embed progressively larger matrices into a single ciphertext by scaling the global ring degree via CRT-Strassen hybrids, practical RLWE implementations (e.g., BGV) typically restrict n to fixed powers of two (e.g., $n \in \{8192, 16384, 32768\}$) to optimize Number Theoretic Transforms and standardize security parameters. For target matrices where N exceeds the maximum single-ciphertext capacity $N_{\max} = \widetilde{\Theta}(n^{1/\log_2 7})$, we must extend our framework to a boundless, multi-ciphertext setting.

ZLW [17] addressed this via block matrix multiplication. By partitioning an $N \times N$ matrix into $L \times L$ block matrices of size $N_b \times N_b$ (where $L = \lceil N/N_b \rceil$), they apply Strassen's algorithm over the block matrices. However, their reliance on standard power bases forces them to perform homomorphic basis switching on each of the $O(L^2)$ distinct blocks prior to multiplication. This structural necessity incurs $O(L^2)$ CMult operations and $O(L^2 \log s)$ Aut operations purely as computational overhead, dominating the evaluation latency.

By lifting our Self-Dual Normal Basis (SDNB) and Optimal Normal Basis (ONB) encodings to the block matrix level, we completely eliminate the homomorphic basis switching requirement. We formalize this multi-ciphertext execution in Algorithm 3.

Theorem 17 *For a fixed ring degree n and an arbitrarily large matrix dimension $N = n^\rho$ (with $\rho > 1/\log_2 7$), block-wise Strassen decomposition using SDNB/ONB encoding evaluates the homomorphic matrix product utilizing exactly 0 plaintext-ciphertext multiplications (*CMult*), $O(N^{\log_2 7 - 1/\rho})$ ciphertext-ciphertext multiplications (*Mult*), and $\widetilde{O}(N^{2-\frac{2}{\rho \log_2 7}})$ homomorphic Galois automorphisms (*Aut*).*

Proof As formalized in Phase 1 of Algorithm 3, we partition the original matrices $A, B \in M_N(\mathbb{F}_p)$ into $L \times L$ block matrices $A^{(i,j)}$ and $B^{(i,j)}$ of size $N_{\max} \times N_{\max}$. To seamlessly support Strassen's recursive halving, we pad the grid with zero matrices to the nearest power of two, defining $L = 2^{\lceil \log_2 \lceil N/N_{\max} \rceil \rceil}$. Because padding at most doubles the grid dimension, it acts as an $O(1)$ constant overhead and leaves the asymptotic complexity unaffected.

Algorithm 3 Boundless Homomorphic Block-Strassen Evaluation

Input: Matrices $A, B \in M_N(\mathbb{F}_p)$, base module capacity $N_{\max}$

Output: Encrypted block matrix $\{c_C^{(i,j)}\}_{i,j=1}^{L}$ encrypting the product $C = AB$

$L \leftarrow 2^{\lceil \log_2(N/N_{\max}) \rceil}$

Pad A and B to $L \cdot N_{\max} \times L \cdot N_{\max}$ with zeros and partition into $L \times L$ blocks $\{A^{(i,j)}\}, \{B^{(i,j)}\}$

Phase 1: Offline Client Encoding & Encryption

for $i = 1 \dots L$, $j = 1 \dots L$ **do**

  $P_A^{(i,j)} \leftarrow \textsc{NormalBasisEncode}(A^{(i,j)}, \text{primary generator } \theta)$

  $P_B^{(i,j)} \leftarrow \textsc{NormalBasisEncode}(B^{(i,j)}, \text{trace dual generator } \gamma_0)$

  $c_A^{(i,j)} \leftarrow \mathsf{BGV.Enc}_{pk}(P_A^{(i,j)}), \quad c_B^{(i,j)} \leftarrow \mathsf{BGV.Enc}_{pk}(P_B^{(i,j)})$

end for

Phase 2: Homomorphic Strassen Tensor Products (Server)

function $\textsc{HomoStrassen}(\mathbf{X}, \mathbf{Y}, k)$

  if $k = 1$ **then**

    return $\mathsf{EvalMult}_{\mathsf{relin}}(\mathbf{X}^{(1,1)}, \mathbf{Y}^{(1,1)})$

  end if

  Split $\mathbf{X}$ into quadrants $\mathbf{X}_{11}, \mathbf{X}_{12}, \mathbf{X}_{21}, \mathbf{X}_{22}$ and $\mathbf{Y}$ into $\mathbf{Y}_{11}, \mathbf{Y}_{12}, \mathbf{Y}_{21}, \mathbf{Y}_{22}$

  $M_1 \leftarrow \textsc{HomoStrassen}(\mathbf{X}_{11} + \mathbf{X}_{22},\ \mathbf{Y}_{11} + \mathbf{Y}_{22},\ k/2)$

  $M_2 \leftarrow \textsc{HomoStrassen}(\mathbf{X}_{21} + \mathbf{X}_{22},\ \mathbf{Y}_{11},\ k/2)$

  $M_3 \leftarrow \textsc{HomoStrassen}(\mathbf{X}_{11},\ \mathbf{Y}_{12} - \mathbf{Y}_{22},\ k/2)$

  $M_4 \leftarrow \textsc{HomoStrassen}(\mathbf{X}_{22},\ \mathbf{Y}_{21} - \mathbf{Y}_{11},\ k/2)$

  $M_5 \leftarrow \textsc{HomoStrassen}(\mathbf{X}_{11} + \mathbf{X}_{12},\ \mathbf{Y}_{22},\ k/2)$

  $M_6 \leftarrow \textsc{HomoStrassen}(\mathbf{X}_{21} - \mathbf{X}_{11},\ \mathbf{Y}_{11} + \mathbf{Y}_{12},\ k/2)$

  $M_7 \leftarrow \textsc{HomoStrassen}(\mathbf{X}_{12} - \mathbf{X}_{22},\ \mathbf{Y}_{21} + \mathbf{Y}_{22},\ k/2)$

  $C_{11} \leftarrow M_1 + M_4 - M_5 + M_7$

  $C_{12} \leftarrow M_3 + M_5$

  $C_{21} \leftarrow M_2 + M_4$

  $C_{22} \leftarrow M_1 - M_2 + M_3 + M_6$

  return Recombine quadrants $C_{11}, C_{12}, C_{21}, C_{22}$ into $k \times k$ block grid $\mathbf{C}$

end function

$\mathbf{M}_{\text{tensor}} \leftarrow \textsc{HomoStrassen}(\mathbf{c}_A, \mathbf{c}_B, L)$

Phase 3: Deferred Trace Extraction (Server)

for $i = 1 \dots L$, $j = 1 \dots L$ **do**

  $c_C^{(i,j)} \leftarrow \textsc{TowеredTrace}(\mathbf{M}_{\text{tensor}}^{(i,j)})$

end for

return $\{c_C^{(i,j)}\}$

Each block is encoded independently into the SDNB/ONB configuration during the offline client phase. Crucially, every left operand block $A^{(i,j)}$ is encoded using the primary generator θ, and every right operand block $B^{(i,j)}$ is encoded using the trace dual generator γ_0. This yields $2L^2$ distinct ciphertexts natively residing in the proper bilinear dual-basis configuration.

Because Strassen's recursive matrix algorithm strictly evaluates linear additions of A blocks exclusively with A blocks, and B blocks exclusively with B blocks, the pre-encoded bases remain invariant throughout the intermediate computations. The scalar combinations intrinsically distribute, preserving the bilinear trace orthogonality relation $\mathrm{Tr}_{\overline{R}_2/\mathbb{F}_p}(\sigma^i(\theta)\sigma^j(\gamma_0)) = \delta_{i,j}$ for the cross-products without requiring any dynamic basis transformations. Consequently, the CMult complexity is exactly zero.

In Phase 2, we evaluate Strassen's algorithm homomorphically over the $L \times L$ grid of ciphertexts. Strassen's algorithm computes the component block products using $O(L^{\log_2 7})$ tensor multiplications and $O(L^{\log_2 7})$ homomorphic ciphertext additions.

For each of the $O(L^{\log_2 7})$ multiplication leaves, the server evaluates exactly one Mult operation (which includes homomorphic relinearization back to R_q^2 prior to the deferred linear additions). Rather than applying the towered trace extraction immediately on the multiplied leaves (which would require $O(L^{\log_2 7})$ computationally expensive trace evaluations), we exploit the $\mathbb{Z}$-linearity of the homomorphic trace operator. Because Galois automorphisms commute with polynomial addition over the base field, $\mathrm{Tr}_{\overline{R}/\overline{R}_{13}}(\sum M_k) = \sum \mathrm{Tr}_{\overline{R}/\overline{R}_{13}}(M_k)$. Consequently, we evaluate the post-multiplication Strassen linear combinations directly on the $O(L^{\log_2 7})$ expanded tensor ciphertexts. This condenses the grid back down into L^2 ciphertexts corresponding to the $L \times L$ blocks of the resulting matrix C, still encoded in the global tensor ring.

Finally, in Phase 3, we apply the towered homomorphic trace extraction independently to each of the L^2 recombined ciphertexts. Because the trace requires $O(\log s)$ automorphisms per block (or bounded to $O(\log s')$ when accelerated via the SDNB configuration), the total algebraic cost evaluates to exactly $O(L^2 \log s)$ automorphisms (Aut).

Substituting the bounding parameters $L = \Theta(N/n^{1/\log_2 7})$ and utilizing the relation $n = N^{1/\rho}$ yields the asymptotic complexities:

$$\mathsf{Mult} = O\left(L^{\log_2 7}\right) = O\left(\left(\frac{N}{n^{1/\log_2 7}}\right)^{\log_2 7}\right) = O\left(\frac{N^{\log_2 7}}{n}\right) = O\left(N^{\log_2 7 - \frac{1}{\rho}}\right),$$

$$\mathsf{Aut} = O\left(L^2 \log s\right) = \widetilde{O}\left(\left(\frac{N}{n^{1/\log_2 7}}\right)^2\right) = \widetilde{O}\left(\frac{N^2}{n^{2/\log_2 7}}\right) = \widetilde{O}\left(N^{2 - \frac{2}{\rho \log_2 7}}\right).$$

Because $\log_2 7 \approx 2.807$, the automorphism cost asymptotes to $\widetilde{O}(N^{2-0.712/\rho})$, and the multiplication cost asymptotes to $O(N^{2.807-1/\rho})$. Since $1/\rho > \frac{\log_2 7}{3\rho} \approx 0.935/\rho$ and $\frac{2}{\rho \log_2 7} \approx 0.712/\rho > \frac{2}{3\rho} \approx 0.667/\rho$, our subtracted exponent terms are strictly larger, mathematically guaranteeing that our overall complexities are strictly smaller. This directly improves upon the prior ZLW large-matrix framework, which requires $O(N^{2.807-0.935/\rho})$ multiplications and $\widetilde{O}(N^{2-0.667/\rho})$ automorphisms, while simultaneously eliminating the $O(N^{2-0.667/\rho})$ CMult bottleneck completely. □

Remark 3 (Communication vs. Computation Trade-off) Prior methods homomorphically pack multiple blocks into a single ciphertext to minimize communication bandwidth, which forces the subsequent $O(N^{2-\frac{2}{3\rho}})$ CMult unpacking penalty. In RLWE schemes, computational latency is typically dominated by high-depth algebraic operations rather than network transmission. By transmitting the $O(L^2)$ block ciphertexts independently, our framework maximizes pure homomorphic throughput. If strict bandwidth limits apply, an outer-basis compression protocol could optionally be applied, though doing so trades the $\mathsf{CMult} = 0$ computational advantage for communication compression.

13 Lattice-Theoretic Security Reductions

The security of the encoding relies on the parameters chosen for the Ring Learning with Errors assumption.

Remark 4 (Distributional Inequivalence of Modulus Embeddings) Certain works attempt to establish a security reduction by mapping the decisional RLWE problem to a matrix encoding by embedding an RLWE challenge (u, v) into a public-key ciphertext vector via $c = (u+m, v)$. This simulation methodology does not accurately capture the noise covariance structure of a valid BGV encryption, making it statistically distinguishable. An RLWE sample $(a, as+e)$ possesses a single error term $e \leftarrow \chi$ drawn from the discrete Gaussian distribution χ. Conversely, a standard BGV ciphertext evaluates to $c = (b_0 r + pe_0 + m, b_1 r + pe_1)$, utilizing an ephemeral random masking polynomial $r \leftarrow \chi$ multiplied against the public key components (b_0, b_1).

Because the matrix encodings defined in this paper are deterministic $\mathbb{Z}$-linear isomorphisms applied within the plaintext algebra prior to encryption, they modify neither the cryptographic structure of the ciphertexts nor the RLWE error distributions. The semantic security of the matrix evaluation requires no lattice reductions; it reduces to the IND-CPA security of the underlying BGV encryption scheme.

By combining the sub-cubic Strassen-CRT spatial geometry, the $O(1)$ optimal normal basis dual expansion bounds, and the towered trace noise reduction, the ciphertext modulus Q required to guarantee correct decryption shrinks relative to the basic power basis parameterization.

Table 2 Algebraic bounds on the ciphertext modulus Q.

N	N^3	Strassen dimension $N^{\log_2 7}$	Power Basis $\log_2 Q$	Strassen-CRT+ONB $\log_2 Q$
16	4096	2401	64	48
32	32768	16807	68	51
64	262144	117649	120	82
128	2097152	823543	460	285

14 Experimental Results

We evaluate the matrix embedding latency over the BGV scheme using HElib. The source code for our implementation is publicly available at https://github.com/ephemer4l/NBMM. We compare our approach against the state-of-the-art power basis baseline, denoted as HEMat [17]. The empirical evaluations target matrix dimensions $N \in \{16, 32, 64, 128\}$ over plaintext fields defined by the moduli $p \in \{65537, 2^{31} - 1\}$. The experiments were conducted on a system utilizing an AMD Ryzen 9 8940HX processor, featuring 16 physical cores and 32 logical threads, with a maximum boost frequency of 5.39 GHz and 64 MiB of L3 cache.

The baseline implementation encounters capacity limits beyond $N = 32$. The table lists the evaluation latency alongside the primal LWE security parameters λ generated by the Sage lattice-estimator library. Both the baseline implementation and our SDNB evaluation are tested under single-threaded (S) and multi-threaded (M) configurations using OpenMP across 32 logical threads. The towered trace executes 5 sequential additions and 5 key-switching automorphisms.

Table 3 Empirical evaluation of matrix multiplication configurations. Spd denotes the speed-up compared to HEMat implementation ([17]), Mults values are the theoretical ciphertext-ciphertext multiplications necessary for HEMat compared to our implementation.

N	p	HEMat(S)	Strass(S)	Spd(S)	HEMat(M)	Strass(M)	Spd(M)	Mults	λ
16	65537	137.3 ms	62.6 ms	2.21×	137.5 ms	64.4 ms	2.19×	1 vs 3	163
16	$2^{31}-1$	255.3 ms	123.5 ms	2.10×	255.6 ms	124.2 ms	2.07×	1 vs 3	190
32	65537	685.4 ms	315.4 ms	2.17×	183.3 ms	73.7 ms	2.49×	7 vs 15	163
32	$2^{31}-1$	1249.5 ms	581.1 ms	2.15×	339.4 ms	141.3 ms	2.40×	7 vs 15	148
64	65537	N/A	2264.7 ms	N/A	N/A	206.4 ms	N/A	49 vs 81	102
128	65537	N/A	12869.9 ms	N/A	N/A	976.1 ms	N/A	343 vs 471	102

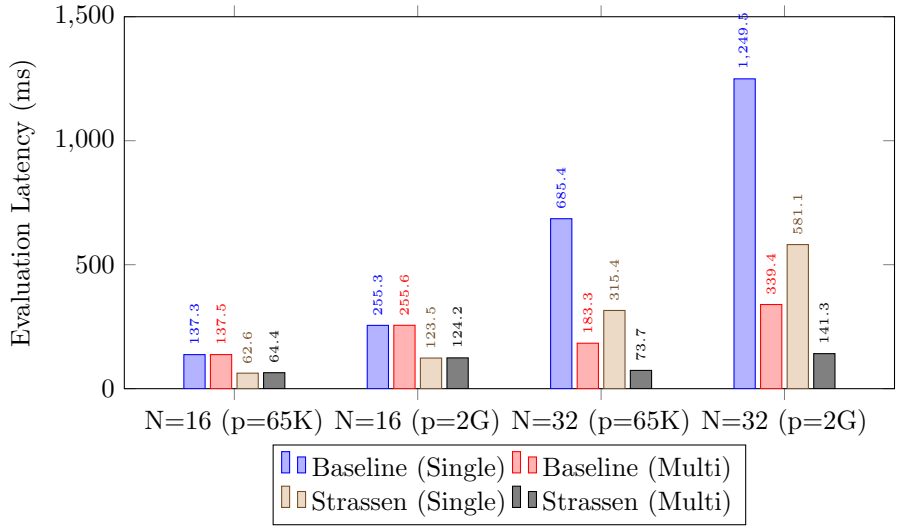


Fig. 2 Homomorphic evaluation latency comparison.

14.1 Algorithmic Scalability vs Ciphertext Capacity

The empirical data supports the theoretical bounds and demonstrates the hardware constraints of prior implementations. For $N = 64$ and $N = 128$, the baseline algorithm

fails to process the matrices due to ciphertext capacity limits and basis-switching requirements.

The Block-Strassen SDNB implementation circumvents these capacity limits and evaluates the matrices for these dimensions. Under the single-threaded configuration, as the matrix dimension doubles from $N = 64$ to $N = 128$, the standard cubic tower runtime scales by a factor of 6.47× (2512.7 ms to 16269.4 ms). The Strassen execution scales by a factor of 5.68× (2264.7 ms to 12869.9 ms). This difference in scaling demonstrates the $O(N^{\log_2 7})$ complexity reduction, yielding a 1.26× speedup at $N = 128$. Under the multi-threaded configuration, the Strassen execution time reduces to 976.1 ms at $N = 128$. The total number of ciphertext multiplications evaluated by our method at $N = 128$ is 343, whereas the baseline requires 471 multiplications.

15 Comparison and Conclusion

Table 4 compares the algebraic characteristics of the proposed Strassen-CRT Normal Basis embedding against power basis embeddings [17].

Table 4 Comparison of algebraic characteristics.

Characteristic	SDNB/ONB Strassen-CRT Embedding	Baseline [17]
Ring Degree Bound (n)	$\widetilde{\boldsymbol{\Omega}}(\boldsymbol{N^{\log_2 7}})$	$\widetilde{\Omega}(N^3)$
Evaluation Rank (D)	1	1
Plaintext Space	***Galois Rings*** $(\mathbb{Z}_{p^r})$	Finite Fields $(\mathbb{F}_p)$
Encoding Basis	SDNB (K^+) or ONB	Power $(x, x^\vee)$
Generators Required	***1 Element* (SDNB)**	$O(s)$ Elements
Trace Dual Norm (E_γ)	$O(1)$ Constant (ONB)	Asymmetric/Unbounded
Outer Basis Norms	Minimized (Kronecker)	Unbounded
Additive Trace Noise	$\boldsymbol{O(\log s')}$ **(Towered)**	$O(s)$
Large Matrix Mult Factor	$\boldsymbol{O(N^{\log_2 7 - 1/\rho})}$	$O(N^{(1-1/3\rho)\log_2 7})$
Large Matrix Aut Factor	$\widetilde{\boldsymbol{O}}(\boldsymbol{N^{2-2/(\rho \log_2 7)}})$	$\widetilde{O}(N^{2-2/3\rho})$
Boundless CMult Overhead	**0 (Ciphertext Strassen)**	$O(N^{2-\frac{2}{3\rho}})$

Bilinear monomial evaluations computing the inner product are bounded by the Abelian Triple Product property. We bypassed the $\widetilde{\Omega}(N^3)$ limit by routing CRT-Strassen hybrids through orthogonal ideal slots. Within the unramified cyclotomic extension, we substituted power bases with normal bases, applying maximal real subfields to bypass the Lempel-Weinberger parity constraint, achieving a one-element generating set with norm symmetry.

We used Type I Optimal Normal Bases to eliminate Gram matrix inversion scaling, applied towered Galois evaluations to reduce FHE key-switching noise, and generalized the embedding mapping to Galois rings via Hensel's and Nakayama's lemmas for integer arithmetic. By expanding the framework to boundless matrix dimensions via ciphertext-level Strassen extraction, the zero-depth norm symmetries successfully drop multi-ciphertext basis-switching overhead entirely. Finally, canonical norm proofs establish power bases as the minimum for non-evaluated outer tensors.

Declarations

- **Competing interests:** The authors declare that they have no competing interests.
- **Data availability:** Data sharing is not applicable to this article as no datasets were generated or analyzed.

References

[1] Z. Brakerski, C. Gentry, and V. Vaikuntanathan, *(Leveled) fully homomorphic encryption without bootstrapping*, ACM Transactions on Computation Theory **6** (2014), 1–36.

[2] H. Chen, M. Kim, I. Razenshteyn, D. Rotaru, Y. Song, and S. Wagh, *Maliciously secure matrix multiplication with applications to private deep learning*, Advances in Cryptology – ASIACRYPT 2020 (2020), 31–59.

[3] J. H. Cheon, A. Kim, and D. Yhee, *Multi-dimensional packing for HEAAN for approximate matrix arithmetics*, Cryptology ePrint Archive, Paper 2018/1245 (2018).

[4] H. Cohn and C. Umans, *A Group-Theoretic Approach to Fast Matrix Multiplication*, Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (2003), 438–449.

[5] S. Halevi and V. Shoup, *Algorithms in HElib*, Advances in Cryptology – CRYPTO 2014, Springer (2014), 554–571.

[6] X.-d. Hou, *Lectures on Finite Fields*, Graduate Studies in Mathematics **190**, American Mathematical Society, Providence, RI, 2018.

[7] J. Jang, Y. Lee, A. Kim, B. Na, D. Yhee, B. Lee, J. H. Cheon, and S. Yoon, *Privacy-preserving deep sequential model with matrix homomorphic encryption*, Proceedings of the 2022 ACM on Asia Conference on Computer and Communications Security (2022), 377–391.

[8] X. Jiang, M. Kim, K. Lauter, and Y. Song, *Secure outsourced matrix computation and application to neural networks*, Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (2018), 1209–1222.

[9] A. Lempel and M. J. Weinberger, *Self-Complementary Normal Bases in Finite Fields*, SIAM Journal on Discrete Mathematics, **1**(2) (1988), 193–198. https://doi.org/10.1137/0401021

[10] W. Lu, S. Kawasaki, and J. Sakuma, *Using fully homomorphic encryption for statistical analysis of categorical, ordinal and numerical data*, Cryptology ePrint Archive, Paper 2016/1163 (2016).

[11] B. R. McDonald, *Finite Rings with Identity*, Marcel Dekker, New York, 1974.

[12] R. C. Mullin, I. M. Onyszchuk, S. A. Vanstone, and R. M. Wilson, *Optimal normal bases in GF(p^n)*, Discrete Applied Mathematics **22** (1988/89), 149–161.

[13] D. Rathee, P. K. Mishra, and M. Yasuda, *Faster PCA and linear regression through hypercubes in HElib*, Proceedings of the 2018 Workshop on Privacy in the Electronic Society (2018), 42–53.

[14] P. Rizomiliotis and A. Triakosia, *On matrix multiplication with homomorphic encryption*, Proceedings of the 2022 on Cloud Computing Security Workshop (2022), 53–61.

[15] V. Strassen, *Gaussian elimination is not optimal*, Numerische Mathematik **13** (1969), 354–356.

[16] S. Wang and H. Huang, *Secure outsourced computation of multiple matrix multiplication based on fully homomorphic encryption*, KSII Transactions on Internet and Information Systems **13** (2019), 5616–5630.

[17] X. Zheng, H. Li, and D. Wang, *A new framework for fast homomorphic matrix multiplication*, Designs, Codes and Cryptography **93** (2025), 2671–2693.