The exponents for rectangular matrix multiplication are also tight

Virginia Vassilevska Williams (via chat-GPT+Claude etc!)
MIT

Abstract

We prove lower bounds that match the known upper bounds on the exponents of rectangular matrix multiplication, due to Uznański and to Amanov, van den Berg and Zuiddam. The proof uses the arguments of OpenAI's preprint showing that the square exponent equals 9/4. Its main step constructs a homogeneous spectral tripod whose shared-label price attains its sharp value at all positive arm lengths simultaneously. Optimizing the weighted wire price on this tripod determines the rectangular exponent over every field. In particular, the dual exponent equals 1/2, and for every fixed $k > 1/2$ the exponent of the shape $(1, k, 1)$ equals $1 + k + 1/(4k)$.

1 Introduction

For fixed $a, b, c \geq 0$, let $\omega_F(a, b, c)$ be the infimum of the exponents τ such that an $\lceil n^a \rceil \times \lceil n^b \rceil$ matrix can be multiplied by an $\lceil n^b \rceil \times \lceil n^c \rceil$ matrix using $O(n^\tau)$ formal arithmetic operations over the field F. Strassen's algorithm gave the first subcubic bound on the square exponent [29]. Coppersmith and Winograd later refined Strassen's laser method [30] to bound it by 2.376 [7], and a long line of subsequent work [28, 8, 33, 16, 2, 10, 34, 1] lowered this to 2.371177 [11]. Upper bounds on the rectangular exponents have a similarly long history. Coppersmith showed that an $n \times n^k$ matrix can be multiplied by an $n^k \times n$ matrix with $n^{2+o(1)}$ operations for some $k > 0$ [5], and later increased k [6]. Lotti and Romani [20], Huang and Pan [13], and Ke, Zeng, Han and Pan [14] gave bounds for other shapes, and refinements of the laser method improved these bounds further [15, 18, 17, 34, 1]. Christandl, Le Gall, Lysikov and Zuiddam proved barriers for this method [4].

In 2026, OpenAI introduced a new approach that combines degenerations of polynomial multiplication tensors with Strassen's asymptotic spectrum [31, 35]. Its September preprint already contained rectangular bounds [22], and its October preprint proved $\omega_{\mathbb{C}} \leq 9/4$ [23]. Several authors have since simplified this argument, made it constructive, or extended it to every field [21, 3, 25, 26, 12, 19, 9]. In particular, Baitian Li's exposition of $\omega \leq 9/4$ formulates the key step as an untwisting lemma for semidirect sums of tensors [19].

Uznański extended the analysis to all rectangular shapes and proved the upper bound in (1) below over fields of characteristic zero [32]. Amanov, van den Berg and Zuiddam proved the same upper bound over every field [3]; they also showed that the degeneration ingredients of the argument cannot improve it. Filmus gave another proof over every field, with a Lean formalization [12], and Zhou recovered the bound as the value of a completely positive program on weighted triangles [36]. Song obtained bounds for asymmetric rectangular multiplication over arbitrary fields [27], and Li discusses the rectangular bounds in his exposition [19].

On the lower-bound side, OpenAI's November preprint shows that $\omega_F = 9/4$ for every field F [24]. Applying this square lower bound to the tensor product of the three cyclic permutations of a shape gives $\omega_F(a, b, c) \geq 3(a + b + c)/4$, which does not capture the full dependence on the shape. We use the proof of the square lower bound, not only its statement, to recover this dependence and to show that the upper bound of [32, 3] is tight.

Theorem 1. *Let $a \geq b \geq c \geq 0$. For every field F,*

$$\omega_F(a,b,c) = \begin{cases} a+b, & c \leq ab/(a+b), \\ \dfrac{(ab+ac+bc)^2}{4abc}, & c > ab/(a+b). \end{cases} \tag{1}$$

When $a = b = 0$, the threshold in (1) is taken to be zero. Other orderings of a, b, c follow by permutation symmetry.

The upper bound in (1) is the bound of [32, 3]; our contribution is the matching lower bound.

Corollary 2. *Let $\alpha_F = \sup\{k \geq 0 : \omega_F(1,k,1) = 2\}$ be the dual exponent. Then*

$$\omega_F(1,k,1) = \begin{cases} 2, & 0 \leq k \leq 1/2, \\ 1+k+1/(4k), & k > 1/2. \end{cases} \tag{2}$$

Thus $\alpha_F = 1/2$. For every fixed $k \geq 1$, the exponent exceeds the input-size exponent $1+k$ by exactly $1/(4k)$.

For example, $\omega_F(1,2,1) = 25/8$. For comparison, Corollary 1.3 of [24] states the interval $1/2 \leq \alpha_F \leq 2/3$.

2 A spectral certificate for the one-parameter curve

The multiplication tensor of format (A, B, C) is

$$\sum_{i=1}^{A} \sum_{j=1}^{B} \sum_{\ell=1}^{C} e_{ij} \otimes e_{j\ell} \otimes e_{\ell i}.$$

Each summation index occurs at two of the three sites, so this tensor is the product of three pairwise wires.

We use Strassen's asymptotic spectrum [31, 35]. A spectral point is additive under direct sums, multiplicative under tensor products, and monotone under restrictions. A *price* is a positive multiple of the logarithm of a spectral point. A binary wire between two sites has price equal to the distance between them. Write D for the price of a binary label shared by all three sites. A shared label with q possible values then has price $D \log_2 q$, and a decomposition into q simple summands is a restriction of it. The price of a rectangular multiplication, per $\log_2 n$, is the sum W of the three wire distances weighted by their dimension exponents. Spectral monotonicity therefore gives the lower bound W/D on the rank exponent.

Lemma 3 below provides a tripod on which, for all positive arm lengths, the shared-label price is

$$D_*(r_1, r_2, r_3) = r_1 + r_2 + r_3 - (1/r_1 + 1/r_2 + 1/r_3)^{-1}. \tag{3}$$

Consider the shape $(1, k, 1)$ with $k > 1/2$. Place the two endpoints of the wire of weight k on arms of length one, and the third site on an arm of length $t = 1/(2k-1)$. Then

$$\frac{W}{D_*} = \frac{2k + 2(1+t)}{2 + t - t/(2t+1)} = 1 + k + \frac{1}{4k}.$$

For $0 \leq k \leq 1/2$, the output flattening already gives the lower bound 2. By contrast, the identity $\omega_F = 9/4$ alone, combined with cyclic products and flattenings, gives only $\max\{2, 1+k, 3(2+k)/4\}$; at $k = 2$ this is 3, whereas (2) gives $25/8$.

3 Existence of a tripod with simultaneous equality

Fix a finite field K and let $F = \overline{K}$. A *contact tripod* over F consists of three infinite rays in a spectral space over F, with homogeneous prices, whose shared-label price agrees with (3) for every positive triple of arm lengths.

Lemma 3. *For every finite field K, there is a contact tripod over $\overline{K}$.*

Proof. **1. Analytic motion produces contact.** Take a family of sites that satisfies the three conditions stated before [24, Theorem 7.1], including nonconstant motion. The proof in Sections 7 and 8.2 of [24] constructs a tangent tripod on fixed rays and obtains

$$\lim_{s\to\infty}\bigl[D(1,1,s) - D_*(1,1,s)\bigr] = 0.$$

This limit is taken on the same homogeneous tripod for every s. Proposition 8.1 of [24] propagates this contact to every positive triple of arm lengths. Its statement includes this stronger conclusion, even though the proof of Theorem 7.1 there uses only the triple $(1,1,1)$ in its last step.

2. Absence of contact implies synchronization. Suppose that no contact tripod exists over F. Then Step 1 rules out every nonconstant family satisfying the three conditions. Consider the proof of the synchronization theorem [24, Theorem 12.1, pp. 44–46]. There, the pinned families $z \mapsto x_{z,\ell}$ satisfy the invariance, submean and continuity conditions before the square exponent is used. On p. 45, the strict square saving is used to conclude that these families are constant; our hypothesis gives the same conclusion directly. The rest of that proof, which removes the pinning penalty and passes to independent parameters, then establishes the synchronization identity for all the families in that theorem.

3. Synchronization contradicts the network lower bound. The contradiction uses the following results of [24].

- Corollary 12.2 gives simultaneous pricing. Lemma 14.2 uses it to place the auxiliary sphere site, and Theorem 14.1 and Corollaries 14.3–14.4 then give the sphere price and the common comparison tensor.
- The sphere price gives the incidence inequality of Theorem 15.1. Lemma 17.1 applies it over $\overline{K}$ and transfers the finite spectral data over K. Sections 17–18 then yield the two sorting estimates of Theorem 16.1.
- In the other branch, the comparison with the sphere tensor in Section 20 (p. 79) uses Corollary 14.4 over $\overline{K}$. Sections 19–22 obtain local concentration (Lemma 19.1). The network assembly that follows uses the sorting and concentration estimates to obtain the compression conclusion of Theorem 2.1 over K.

The iteration in the proof of Proposition 2.3 of [24] now contradicts the independent list-distinctness lower bound of Theorem 2.2 there. Hence the hypothesis of Step 2 is false, which proves the lemma for every finite field K. □

4 Optimization for three independent dimensions

Write $a_1 = a$, $a_2 = b$, $a_3 = c$, and place the wire of weight a_i opposite the arm of length r_i. Put $L = a + b + c$. The weighted wire sum is

$$W(r) = \sum_i (L - a_i) r_i.$$

Suppose first that $c > ab/(a+b)$. Set $S = ab + ac + bc$, $P = abc$, $U = S^2/(4P)$, and

$$t_i = S - 2P/a_i.$$

All three t_i are positive, since the smallest of them is $c(a+b) - ab$. Direct expansion gives $\sum_i t_i = S$ and $U - (L - a_i) = t_i^2/(4P)$. It follows that

$$UD_*(r) - W(r) = \frac{1}{4P}\left(\sum_i t_i^2 r_i - \frac{S^2}{\sum_i 1/r_i}\right) \geq 0 \tag{4}$$

by the Cauchy–Schwarz inequality, with equality when $r_i = 1/t_i$. On the contact tripod of Lemma 3, these arm lengths give $W/D = U$, which proves the lower bound.

For the upper bound, write the three pairwise distances of an arbitrary spectral placement as $r_i + r_j$ with $r_i \geq 0$; the triangle inequalities make this possible. The first part of [24, Proposition 8.1], which applies to every placement, gives $D \geq D_*(r)$ for these distances. Hence $W/D \leq U$ by (4), where zero arms are handled by continuity. Maximizing the normalized wire price and applying Strassen's spectral duality [31] gives the upper bound on the asymptotic rank, which recovers the bound of [32, 3]. For $c = ab/(a+b) > 0$, we use continuity from above, since U tends to $a + b$. For smaller c, padding gives the upper bound and flattening gives the lower bound. Shapes with $b = 0$ follow directly from flattening and a linear-size algorithm.

4.1 Passage to all fields and to arithmetic exponents

For fixed integer dimensions (A, B, C), take the weights $(\log_2 A, \log_2 B, \log_2 C)$. The wire-price comparison then gives a lower bound on the logarithm of the rank at that format over $\overline{K}$, for every finite field K. A rank decomposition over an arbitrary field has its coefficients in a finitely generated subring of that field. Reduction modulo a maximal ideal with finite residue field preserves the decomposition identities, as in [24, Lemma 2.5]. Thus the same lower bound holds at every fixed format over every field, and it survives tensor powers by homogeneity. For the upper bound, spectral values of tensors over the base field transfer to an algebraic closure [24, Section 3]. The usual bilinearization and tensor-power recursion then give the arithmetic exponent, with rounding handled by continuity or padding. In the upper-bound recursion, the sizes of the linear maps are absorbed by the rank under tensor powering, since the rank dominates each flattening.

References

[1] J. Alman, R. Duan, V. Vassilevska Williams, Y. Xu, Z. Xu, and R. Zhou. *More asymmetry yields faster matrix multiplication.* In *Proc. 2025 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2005–2039, 2025.

[2] J. Alman and V. Vassilevska Williams. *A refined laser method and faster matrix multiplication.* In *Proc. 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 522–539, 2021.

[3] A. Amanov, M. van den Berg, and J. Zuiddam. *Multiplying matrices via polynomials: exposition and barriers.* arXiv:2610.30054, 2026.

[4] M. Christandl, F. Le Gall, V. Lysikov, and J. Zuiddam. *Barriers for rectangular matrix multiplication.* Comput. Complexity 34, article 4, 2025.

[5] D. Coppersmith. *Rapid multiplication of rectangular matrices.* SIAM J. Comput. 11(3):467–471, 1982.

[6] D. Coppersmith. *Rectangular matrix multiplication revisited*. J. Complexity 13(1):42–49, 1997.

[7] D. Coppersmith and S. Winograd. *Matrix multiplication via arithmetic progressions*. J. Symbolic Comput. 9(3):251–280, 1990.

[8] A. M. Davie and A. J. Stothers. *Improved bound for complexity of matrix multiplication*. Proc. Roy. Soc. Edinburgh Sect. A 143(2):351–369, 2013.

[9] M. L. Doğan, V. Lysikov, K. Sakabe, and M. Walter. *Tensor value and the 9/4 bound for matrix multiplication*. arXiv:2611.09716, 2026.

[10] R. Duan, H. Wu, and R. Zhou. *Faster matrix multiplication via asymmetric hashing*. In *Proc. 64th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 2129–2138, 2023.

[11] E. Dupont, M. Eisenberger, B. Kozlovskii, A. Mehrabian, F. J. R. Ruiz, A. See, R. Zhou, J. Alman, V. Vassilevska Williams, and M. Balog. *Improving the matrix multiplication exponent with modern optimization and AlphaEvolve*. arXiv:2608.16884, 2026.

[12] Y. Filmus. *Matrix multiplication from block identities*. Manuscript and Lean formalization, GitHub repository, 2026.

[13] X. Huang and V. Y. Pan. *Fast rectangular matrix multiplication and applications*. J. Complexity 14(2):257–299, 1998.

[14] S. Ke, B. Zeng, W. Han, and V. Y. Pan. *Fast rectangular matrix multiplication and some applications*. Sci. China Ser. A 51(3):389–406, 2008.

[15] F. Le Gall. *Faster algorithms for rectangular matrix multiplication*. In *Proc. 53rd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 514–523, 2012.

[16] F. Le Gall. *Powers of tensors and fast matrix multiplication*. In *Proc. 39th International Symposium on Symbolic and Algebraic Computation (ISSAC)*, pages 296–303, 2014.

[17] F. Le Gall. *Faster rectangular matrix multiplication by combination loss analysis*. In *Proc. 2024 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3765–3791, 2024.

[18] F. Le Gall and F. Urrutia. *Improved rectangular matrix multiplication using powers of the Coppersmith–Winograd tensor*. In *Proc. 29th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1029–1046, 2018.

[19] B. Li. *Semidirect sums of tensors and the matrix multiplication exponent*. arXiv:2611.04175, 2026.

[20] G. Lotti and F. Romani. *On the asymptotic complexity of rectangular matrix multiplication*. Theoret. Comput. Sci. 23(2):171–185, 1983.

[21] H. Lü. *Polynomial separation of a shared tensor leg and matrix multiplication in arbitrary characteristic*. arXiv:2610.18914, 2026.

[22] OpenAI. *Complex matrix multiplication below 2.258 and rectangular bounds*. Preprint, September 24, 2026. Manuscript.

[23] OpenAI. *An upper bound of 9/4 for the matrix multiplication exponent*. Preprint, October 2, 2026. Manuscript.

[24] OpenAI. *The matrix multiplication exponent is nine fourths*. Preprint, November 24, 2026. Manuscript.

[25] R. Pass. *Matrix multiplication from polynomial packings: a constructive and elementary proof of* $\omega \leq 9/4$. arXiv:2610.28627, 2026.

[26] A. Shpilka. *Exposition of OpenAI's matrix multiplication in exponent 9/4 (and an explicit algorithm)*. ECCC TR26-281, 2026.

[27] Z. Song. *Faster asymmetric rectangular matrix multiplication*. SSRN 7600178, 2026.

[28] A. J. Stothers. *On the complexity of matrix multiplication*. PhD thesis, University of Edinburgh, 2010.

[29] V. Strassen. *Gaussian elimination is not optimal*. Numer. Math. 13:354–356, 1969.

[30] V. Strassen. *Relative bilinear complexity and matrix multiplication*. J. Reine Angew. Math. 375/376:406–443, 1987.

[31] V. Strassen. *The asymptotic spectrum of tensors*. J. Reine Angew. Math. 384:102–152, 1988.

[32] P. Uznański. *Rectangular matrix multiplication from polynomial products*. arXiv:2610.10494v2, 2026.

[33] V. Vassilevska Williams. *Multiplying matrices faster than Coppersmith–Winograd*. In *Proc. 44th ACM Symposium on Theory of Computing (STOC)*, pages 887–898, 2012.

[34] V. Vassilevska Williams, Y. Xu, Z. Xu, and R. Zhou. *New bounds for matrix multiplication: from alpha to omega*. In *Proc. 2024 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3792–3835, 2024.

[35] A. Wigderson and J. Zuiddam. *Asymptotic spectra: theory, applications and extensions*. Bull. Amer. Math. Soc. 63(2):199–331, 2026.

[36] H. Zhou. *Asymptotic rank of weighted graph tensors via completely positive bounds*. arXiv:2610.34273, 2026.