

Deterministic Integer 3SUM in $O(n^{1.999017})$ Time via a Sharper Binomial-Tail Analysis

Yuxuan Xu

School of Computer Science, Wuhan University

2024302111072@whu.edu.cn

ORCID: 0009-0001-8760-5277

October 13, 2026

Abstract

We refine the analysis of the deterministic integer 3SUM algorithm of Alman and Vassilevska Williams. Within their framework of thin matrix products and reductions through Exact Triangle, we obtain a worst-case running time of $O(n^{1.999017})$ for polynomially bounded integer inputs on a word RAM with $O(\log n)$ -bit words. The refinement retains the binomial tail in the preprocessing cost instead of replacing it immediately by a uniform geometric bound. An entropy estimate improves the preprocessing saving exponent while preserving the query algorithm. Explicit rational parameters give a 3SUM exponent of $1.999016414375 + o(1)$, leaving a positive margin below the stated fixed exponent. Retuning the previous analysis alone gives $O(n^{1.999073})$. We account for encoding, queries, prime selection, instance construction, witness scanning, and parameter rounding. The numerical inequalities admit exact rational interval certificates. This is an analysis refinement of an existing algorithm and its reductions; no complete Lean formalization of the improved bounds, global optimality, or priority claim is made.

Keywords: 3SUM; deterministic algorithms; fine-grained complexity; Exact Triangle; thin matrix products; binomial tails.

MSC 2020: 68W40, 68Q25.

1 Scope and main result

The deterministic integer 3SUM algorithm of Alman and Vassilevska Williams [1] obtains its improvement from computing a sparse set of entries of a thin matrix product. Its data structure separates the work into shared encodings, preprocessing of boxes, and queries. We revisit one estimate of the number of boxes. The underlying identity, data structure, and reductions are all those of [1]; the contribution here is a sharper tail estimate and the resulting choice of parameters.

Throughout, all logarithms are natural unless a base is displayed. The notation $\tilde{O}$ suppresses powers of logarithms, not fixed positive powers of the input size. Every decimal chosen as a parameter denotes its exact rational value. Constants may depend on fixed parameters and on the fixed exponent bounding the input integers, but not on the input instance or its size.

Theorem 1.1. *For every fixed integer $\kappa \geq 1$, there is a deterministic word-RAM algorithm deciding 3SUM on n integers of absolute value at most n^κ in $O(n^{1.999017})$ worst-case time, with word length $C_\kappa(\log n + 1)$ bits for a sufficiently large fixed constant C_κ . The decision question requires three pairwise distinct input indices; equal values at different indices are permitted.*

This is an ordinary mathematical running-time claim obtained by reanalyzing the cited algorithm. It is not a claim that a new machine-checked theorem has been produced. The formalization accompanying [1] is useful for locating the original cost estimates, but has not been rebuilt or extended here. Inputs of size less than three are rejected, and other bounded-size cases in the analysis are handled directly.

2 The inherited algorithmic ingredients

We record the precise parts of [1] used below, so that the new estimate and the inherited algorithmic claims remain separate. The entries of all matrices in this section are integers of absolute value $N^{O(1)}$.

2.1 The unrelaxed preprocessing cost

Let $m \geq 1$, $D = 4^m$, $L \geq 10m$, and $0 \leq t \leq m$. In the notation of [1, Section 4], set

$$K = \binom{L}{m}, \quad N_0 = 3^{L-m}, \quad M = KN_0^2, \quad \beta_d = \binom{L}{m-d} 9^{L-m+d}. \quad (1)$$

In particular $M = \beta_0$. A tile is a batch of block products in the $N \times D \times N$ multiplication. The box dynamic program stores sums of encoded leaf products that can be shared between queries.

Assume $N \geq \sqrt{M}$. Lemma 29 and the proof of Theorem 30 of [1] give $O(N^2/M)$ tiles and a cost of $O(L(m+1) \sum_{d=t}^m \beta_d)$ for the boxes of each tile. Before bounding the tail geometrically, the total preprocessing cost is therefore

$$O\left(\frac{N10^L}{\sqrt{M}} + \frac{N^2}{M} L(m+1) \sum_{d=t}^m \beta_d\right). \quad (2)$$

The first term also absorbs the construction of input arrays and the list of block-subset assignments, as in that proof. The time per query is

$$O\left(L \sum_{d=0}^t \binom{m}{d} 9^d\right). \quad (3)$$

These are costs of the original deterministic data structure, not a new oracle assumption. The word-size analysis of [1, Theorem 30] continues to apply when its parameters are changed within the stated range.

2.2 The original exponent formula

Define the entropy $H(x) = -x \log x - (1-x) \log(1-x)$, with $H(0) = H(1) = 0$. For $c > 10$ and $0 < \theta < 0.9$, let

$$\gamma(c, \theta) = \frac{\theta \log((c-1)/9)}{\log 4}, \quad (4)$$

$$q(\theta) = \frac{H(\theta) + \theta \log 9}{\log 4}, \quad (5)$$

$$A(c) = \frac{c \log 10 - \frac{1}{2} c H(1/c) - (c-1) \log 3}{\log 4}, \quad (6)$$

$$R(c, \gamma) = \frac{1}{A(c) + \gamma}. \quad (7)$$

Corollary 31 of [1] gives preprocessing time $O(N^2 \log^2 D / D^\gamma)$ and query time $O(D^q \log D)$ when $D \leq N^a$ and $0 < a < R(c, \gamma)$. For at most $N^2 / \sqrt{D}$ queries, it follows that every

$$0 < s < \min\{\gamma, \frac{1}{2} - q\} \quad (8)$$

is a sufficient saving exponent for total time $O(N^2 / D^s)$. These are sufficient conditions for this analysis, not necessary conditions for all possible algorithms.

2.3 The two deterministic reductions

Exact Triangle asks whether a tripartite graph with integer edge weights contains a triangle whose three weights sum to zero. We use the reduction of [1, Theorem 17 and Section 3.3] in its parameterized form. Given an Exact Triangle instance on n vertices per part, choose $16 \leq D \leq n$ and an integer $1 \leq g \leq \sqrt{D}$. The reduction produces O/ng lopsided triangle instances, each with outer dimension n , middle dimension at most D , and at most $n^2 / \sqrt{D}$ wanted pairs. Its additional costs are

$$O(n^3 \log n / g), \quad O(n^{\log_2 7} D^{3/2}), \quad O(n^2 D g) \quad (9)$$

for witness scans, deterministic prime selection, and construction, respectively. Constants can depend on the fixed input-weight exponent. A Boolean thin matrix product answers the required lopsided triangle queries.

Finally, the deterministic reduction recorded in [1, Theorem 21(a)] transforms integer 3SUM of size n into $n^{1/2+o(1)}$ Exact Triangle instances of size $n^{1/2+o(1)}$, with $n^{3/2+o(1)}$ reduction time. The weights remain polynomially bounded. We use the ordinary distinct-index version of 3SUM, as also stated in [2]. These known reductions and their correctness are inherited; we do not claim new proofs of them.

3 A sharper binomial-tail estimate

For an integer $c > 10$ and $0 < \theta < 0.9$, define

$$\gamma_*(c, \theta) = \frac{cH(1/c) - cH((1-\theta)/c) - \theta \log 9}{\log 4}. \quad (10)$$

Lemma 3.1 (Binomial entropy bounds). *For integers $r \geq 1$ and $0 \leq k \leq r$, writing $p = k/r$, one has*

$$\frac{e^{rH(p)}}{r+1} \leq \binom{r}{k} \leq e^{rH(p)}.$$

Proof. The cases $k = 0, r$ are immediate. Otherwise the probability of k successes in a binomial distribution with parameters (r, p) is $\binom{r}{k} p^k (1-p)^{r-k} = \binom{r}{k} e^{-rH(p)}$. This probability is at most one. Moreover, the ratio of the probabilities at $j+1$ and j is $(r-j)p/((j+1)(1-p))$. At $p = k/r$, this is greater than one for $j < k$ and less than one for $j \geq k$. Thus the term at k is a maximum among the $r+1$ terms, whose sum is one, and is at least $1/(r+1)$. $\square$

Lemma 3.2 (Sharp normalized tail). *Let $c > 10$ be an integer, $m \geq 1$, $L = cm$, $D = 4^m$, and $t = \lceil \theta m \rceil$, where $0 < \theta < 0.9$. With the counts in (1),*

$$\frac{1}{M} \sum_{d=t}^m \beta_d \leq \frac{cm+1}{1-9/(c-1)} D^{-\gamma_*(c,\theta)}. \quad (11)$$

Proof. For $1 \leq d \leq m$,

$$\frac{\beta_d}{\beta_{d-1}} = \frac{9(m-d+1)}{L-m+d} \leq \frac{9m}{L-m+1} < \rho_c := \frac{9}{c-1} < 1.$$

Summing a geometric majorant starting at β_t , rather than at β_0 , gives

$$\sum_{d=t}^m \beta_d \leq \frac{\beta_t}{1-\rho_c}. \quad (12)$$

This includes $t = m$, and $t \leq m$ follows from $\theta < 1$. Put $u = t/m$. Applying Lemma 3.1 to the numerator and denominator in $\beta_t/M = 9^t \binom{cm}{m-t} / \binom{cm}{m}$ gives

$$\frac{\beta_t}{M} \leq (cm+1) \exp(m [cH((1-u)/c) - cH(1/c) + u \log 9]).$$

The function $F(u) = cH((1-u)/c) + u \log 9$ has derivative

$$F'(u) = \log\left(\frac{9(1-u)}{c-1+u}\right) < 0 \quad (0 \leq u < 1),$$

and extends continuously to $u = 1$. Since $u \geq \theta$, we may replace $F(u)$ by $F(\theta)$ in this upper bound. The result is $\beta_t/M \leq (cm+1)D^{-\gamma_*}$. Combining it with (12) proves the claim, including the rounding of t . $\square$

Proposition 3.3. *For every $c > 10$ and $0 < \theta < 0.9$, $\gamma_*(c, \theta) > \gamma(c, \theta) > 0$.*

Proof. Both expressions vanish at $\theta = 0$, and

$$\frac{\partial \gamma_*}{\partial \theta} = \frac{\log((c-1+\theta)/(9(1-\theta)))}{\log 4} > \frac{\log((c-1)/9)}{\log 4} = \frac{\partial \gamma}{\partial \theta} > 0.$$

Integrate from zero to θ . $\square$

This proposition concerns the preprocessing exponent at fixed parameters. The query and encoding constraints must still be checked to improve 3SUM.

4 Returning the estimate to the data structure

Proposition 4.1. *Fix an integer $c > 10$, $0 < \theta < 0.9$, and $a > 0$ such that*

$$a(A(c) + \gamma_*(c, \theta)) < 1. \quad (13)$$

If $D = 4^m \leq N^a$, $m \geq 1$, and $|W| \leq N^2/\sqrt{D}$, then the entries in W of an $N \times D \times N$ integer matrix product can be computed deterministically in $O(N^2/D^s)$ time for every fixed

$$0 < s < \min\{\gamma_*(c, \theta), \frac{1}{2} - q(\theta)\}. \quad (14)$$

Proof. Set $L = cm$ and $t = \lceil \theta m \rceil$. Lemma 3.2 in (2) bounds the box cost by

$$O(N^2 m^3 D^{-\gamma_*}) = O(N^2 (\log D)^3 D^{-\gamma_*}). \quad (15)$$

The third power of the logarithm includes the factor $cm + 1$ from the entropy estimate.

For the encoding term, Lemma 3.1 gives

$$\frac{10^{cm}}{\sqrt{M}} \leq \sqrt{cm + 1} \frac{10^{cm}}{e^{cmH(1/c)/2} 3^{(c-1)m}} = \sqrt{cm + 1} D^{A(c)}. \quad (16)$$

Let $\eta = 1/a - A(c) - \gamma_* > 0$. Since $N \geq D^{1/a}$,

$$\frac{D^{\gamma_*} 10^{cm}}{N \sqrt{M}} \leq \sqrt{cm + 1} D^{-\eta} \longrightarrow 0 \quad (m \longrightarrow \infty).$$

For all sufficiently large m , this proves both that the encoding cost is at most $N^2 D^{-\gamma_*}$ and that the tile-fit hypothesis holds: indeed, $M \leq 10^{cm}$ by the binomial theorem, so

$$\sqrt{M} \leq \frac{10^{cm}}{\sqrt{M}} \leq \frac{D^{\gamma_*} 10^{cm}}{\sqrt{M}} \leq N.$$

This is an eventual inequality, not merely a comparison of leading exponents. For the finitely many smaller m , D is bounded and direct inner products cost $O(N^2/D^s)$ after enlarging the constant.

For queries, put $x = (1 - \theta)/\theta$, so $9x > 1$. The binomial theorem gives

$$\begin{aligned} \sum_{d=0}^t \binom{m}{d} 9^d &\leq (9x)^t \sum_{d=0}^m \binom{m}{d} x^{-d} = (9x)^t (1 + 1/x)^m \\ &\leq (9x) \exp(m[H(\theta) + \theta \log 9]) = (9x) D^{q(\theta)}. \end{aligned}$$

Here $t < \theta m + 1$ handles the ceiling. Thus all queries together cost $O(N^2 D^{q-1/2} \log D)$. The strict gaps in (14) absorb the logarithmic factors in this term and (15).

The data structure is unchanged. In particular, the integer sizes and deterministic indexing arguments of [1, Theorem 30] still apply. For example, $N \geq N_0 = 3^{L-m}$ and $L \geq 10m$ imply $10^L \leq N^{5/2}$, as in that proof. All relevant values and addresses therefore require $O(\log N)$ bits. $\square$

5 Transferring an inner saving to 3SUM

Proposition 5.1. *Suppose a deterministic inner algorithm solves the wanted-entry problem of Section 2.3 in $O(n^2/D^s)$ for powers of four $D \leq n^a$, where a, s are fixed positive rational numbers, $a < 1$, and $s/2 < 1/2$. Suppose further that*

$$\log_2 7 + \frac{3}{2}a < 3 - as/2, \quad (17)$$

$$2 + a(1 + s/2) < 3 - as/2. \quad (18)$$

Then Exact Triangle has deterministic running time $\tilde{O}(n^{3-as/2})$, and integer 3SUM has deterministic running time $n^{2-as/4+o(1)}$. In particular every fixed $r > 2 - as/4$ is an attainable 3SUM big- O exponent in this analysis.

Proof. For sufficiently large n , let D be the largest power of four not exceeding n^a , and put $b = s/2$ and $g = \lceil D^b \rceil$. Then $n^a/4 < D \leq n^a$, $D \geq 16$, and $D \leq n$. Also $1 \leq g \leq \sqrt{D}$ eventually because $b < 1/2$. The costs in Section 2.3 become those in Table 1. The first

two exponents agree when $b = s/2$, and (17)–(18) place the other two strictly below them. Rounding D and g changes only constants. The resulting time is $\tilde{O}(n^{3-as/2})$.

All parameter choices are effective integer computations. If $a = p/q$ with positive integers p, q , determine $D = 4^m$ using comparisons $(4^m)^q \leq n^p$. If $b = u/v$, g is the least integer with $g^v \geq D^u$, and may be found by integer binary search. The exponents p, q, u, v are fixed; these operations involve $O(\log n)$ -bit integers with fixed constant factors, and their polylogarithmic cost is absorbed. No real-power or logarithm instruction is required.

Applying the cited 3SUM reduction gives total time

$$n^{3/2+o(1)} + n^{1/2+o(1)}(n^{1/2+o(1)})^{3-as/2} = n^{2-as/4+o(1)}.$$

Since $a < 1$ and $s < 1$, the displayed exponent exceeds $3/2$. A fixed positive gap $r - (2 - as/4)$ absorbs the residual $o(1)$ and all logarithmic factors. Bounded-size exceptions are handled by exhaustive search. $\square$

Stage	Exponent of n , suppressing logarithms
O/ng inner instances	$3 + a(b - s)$
Witness scans	$3 - ab$
Prime selection	$\log_2 7 + 3a/2$
Instance construction	$2 + a(1 + b)$

Table 1: The complete reduction cost with $D = \Theta(n^a)$ and $g = \Theta(D^b)$.

6 Explicit parameters and certificates

6.1 Retuning the existing bound

First take

$$c = 22, \quad \theta = 0.11229, \quad a = 0.05405, \quad s = 0.06862.$$

Use γ from (4), without the sharper tail. The certificates in Table 2 imply $a < R(c, \gamma)$ and (8). Proposition 5.1 then gives the 3SUM cutoff 1.99907277225. Hence this parameter change alone gives $O(n^{1.999073})$. It is not a consequence of applying monotonicity to a previous theorem with a larger exponent; the parameterized inner theorem and the reduction must be instantiated afresh.

6.2 The sharper tail and the main theorem

Now take

$$c = 21, \quad \theta = 0.11202, \quad a = 0.05665, \quad s = 0.06945, \quad b = s/2 = 0.034725. \quad (19)$$

For these values Lemma 3.2 reads

$$\frac{1}{M} \sum_{d=t}^m \beta_d \leq \frac{20}{11} (21m + 1) D^{-\gamma*}.$$

The following displayed decimal endpoints are outward-rounded enclosures; they are not the floating-point values used by an optimizer.

Quantity	Lower endpoint	Upper endpoint
A: γ	0.068631222495934868	0.068631222495934869
A: q	0.431367880323245422	0.431367880323245423
A: R	0.054052515527303591	0.054052515527303592
B: γ_*	0.069454625736072798	0.069454625736072799
B: q	0.430536990526086701	0.430536990526086702
B: $R_* = 1/(A + \gamma_*)$	0.056657067725264770	0.056657067725264771

Table 2: Certified enclosures for the two parameter choices. A denotes retuning; B denotes the sharp-tail analysis.

Proof of Theorem 1.1. For (19), the positive margins include

$$\begin{aligned}\gamma_* - s &> 0.000004625736072798, \\ \frac{1}{2} - q - s &> 0.000013009473913298, \\ 1 - a(A + \gamma_*) &> 0.000124745694553803.\end{aligned}$$

Thus Proposition 4.1 applies. The Exact Triangle exponent is

$$3 - as/2 = 2.99803282875.$$

The prime-selection exponent is less than 2.892329922058, and the instance-construction exponent is 2.05861717125, both strictly smaller. The remaining hypotheses of Proposition 5.1 hold because $0 < a < 1$ and $0 < b < 1/2$. Its 3SUM cutoff is

$$2 - as/4 = 1.999016414375, \quad 1.999017 - (2 - as/4) = 0.000000585625 > 0.$$

The proposition now yields $O(n^{1.999017})$. $\square$

For comparison, the retuned case has Exact Triangle exponent 2.9981455445, prime-selection exponent less than 2.888429922058, and construction exponent 2.0559044555. Its target gap is $1.999073 - 1.99907277225 = 0.00000022775 > 0$.

6.3 How the interval certificates are obtained

The parameter checks use rational arithmetic only. For a positive rational x , write $x = 2^k y$ with $1 \leq y \leq 2$ and integer k . Put $z = (y - 1)/(y + 1)$, so $0 \leq z \leq 1/3$. For every positive integer K ,

$$S_K \leq \log y \leq S_K + \frac{2z^{2K+1}}{(2K+1)(1-z^2)}, \quad S_K = 2 \sum_{j=0}^{K-1} \frac{z^{2j+1}}{2j+1}. \quad (20)$$

This follows by bounding each denominator in the positive remainder series below by $2K+1$. Use the same enclosure for $\log 2$, and combine $\log x = k \log 2 + \log y$ by interval arithmetic, reversing the endpoints of the $\log 2$ interval if $k < 0$.

Taking $K = 48$ gives the certificates above after substituting into (4)–(7) and (10). Addition, subtraction, multiplication (using the minimum and maximum of the four endpoint products), and reciprocal of intervals avoiding zero preserve enclosure. Every positivity check uses the exact rational lower endpoint. The final decimal display rounds lower endpoints down and upper endpoints up. This procedure specifies finite, reproducible certificates without relying on a floating-point logarithm library.

7 Verification status and provenance

The standard-library Python verifier supplied with the research notes was rerun successfully. It checked all parameter inequalities, the two lower order reduction costs, and the strict target gaps using exact rational intervals. A separately written script, which does not import that verifier, checked the parameters using 90-digit decimal arithmetic. It also checked the entropy inequalities for all $0 \leq k \leq r \leq 100$ with $r \geq 1$ (5,150 pairs), and geometric tail bounds for $c \in \{11, 12, 21, 22, 30\}$, $1 \leq m \leq 80$, $0 \leq t \leq m$ (16,600 triples), using exact integers. Additional numerical checks of the complete sharp-tail bound covered $m = 1, \dots, 100, 200, 500, 1000$ at the parameters of the main result. These finite checks supplement the proofs; they do not establish an asymptotic theorem on their own.

The pinned upstream files consulted were `Sec4/Lemma29.lean`, `Sec4/Theorem30.lean`, and `Sec4/Corollary31.lean` in [2]. In particular, the proof of `Theorem30.bboxes_total` first invokes the exact count `lemma_29_count` before applying its geometric relaxation. The machine-level cost proof would have to be reconnected to the sharper count to certify the present result in Lean. No such extension or complete upstream build is claimed here.

The manuscript and its submission metadata were prepared using OpenAI Codex from the supplied research notes. Codex was also used to inspect the cited paper and selected upstream source files, rerun the supplied verifier, and write and execute the separate checking script. The submitting contributor reports that they have not yet checked and understood the contribution step by step; the submission therefore declares human understanding as *none*. The automated review described above is not independent human peer review. The detailed provenance of the initial research notes is not reconstructed here.

The arithmetic verification does not certify global optimality of these parameters. No exhaustive literature search establishing priority is claimed. The mathematical contribution is the sharpened analysis within the cited framework, not a new tensor, data structure, or reduction.

References

- [1] Josh Alman and Virginia Vassilevska Williams. *Truly Subquadratic 3SUM and Truly Subcubic APSP via Triangles in Sparse Lopsided Graphs*. arXiv:2610.06783v1, 2026.
<https://arxiv.org/abs/2610.06783v1>.
- [2] Anthropic. *Truly Subquadratic 3SUM and Truly Subcubic APSP: a Lean 4 formalization*. `formal-math/3sum-apsp`, commit `e1a4e6508154ea59f030480661590a9fe3018011`, 2026.
<https://github.com/anthropics/formal-math/tree/e1a4e6508154ea59f030480661590a9fe3018011/3sum-apsp>. Accessed October 13, 2026.