The second-order asymptotic of the EKG sequence

P. Graham
https://orcid.org/0009-0007-8493-1505

October 26, 2026

Abstract

The EKG sequence $a(n)$ starts $1, 2$ and thereafter chooses the least unused positive integer not coprime to the preceding term. Jonathan Ayres discovered it, and Lagarias, Rains and Sloane conjectured its additive second-order asymptotic at almost all indices. Haver reduced the asymptotic for the least unused even integer to counting used odd values above it, excluding three times odd primes, and proved the sharp upper bound. Hofman and Pilipczuk proved the weaker ratio limits and prime pattern. Combining these estimates with an elementary sieve and the prime number theorem, we prove the lower bound and transfer to the smoothed sequence. Smoothing $p, 3p$ to $2p$ for odd primes gives $b(n) = n + n/(3\log n) + o(n/\log n)$ $(n \to \infty)$, hence the conjectured formula at almost all original indices. We prove Ordowski's nonprime inequality and density-one conjectures, and Jasinski's finiteness conjectures for the three critical lines; we also obtain the below-diagonal asymptotic and finiteness of cycles of each bounded length. Lean formalizes the main theorem.

This work has not yet been reviewed by experts. The proofs, the Lean formalization and the text were produced by AI models (from OpenAI and Anthropic), working under the author's direction.

1 Introduction

The EKG sequence starts $1, 2, 4, 6, 3, 9, 12, 8, 10, 5, 15, 18, 14, 7, 21$. Write $U_t = \{a(1), \ldots, a(t)\}$. Its defining greedy rule is

$$a(1) = 1, \quad a(2) = 2, \quad a(n) = \min\{m \geq 1 : m \notin U_{n-1},\ \gcd(m, a(n-1)) > 1\} \quad (n \geq 3). \tag{1}$$

We use Lagarias, Rains and Sloane's (LRS) smoothing convention [15, Figure 5]:

$$b(n) = s(a(n)), \qquad s(m) = \begin{cases} 2p, & m = p \text{ or } m = 3p \text{ for an odd prime } p, \\ m, & \text{otherwise.} \end{cases} \tag{2}$$

Logarithms are natural.

Theorem 1. *For the sequence defined by (1) and (2),*

$$b(n) = n + \frac{n}{3\log n} + o\left(\frac{n}{\log n}\right) \qquad (n \longrightarrow \infty).$$

Precisely, for every real $\varepsilon > 0$ there is an $N \in \mathbb{N}$, $N \geq 3$, such that for every $n \in \mathbb{N}$ with $n \geq N$,

$$\left| b(n) - n - \frac{n}{3\log n} \right| \leq \varepsilon \frac{n}{\log n}. \tag{3}$$

LRS write "discovered by Jonathan Ayres" [15, Section 1, first paragraph] and, of the three lines, "This was also observed by Ayres" [15, Section 4]. LRS's abstract says "We conjecture that almost all $a(n)$ satisfy the asymptotic formula" [15, abstract]; Hofman and Pilipczuk define "$a'_n = 2p$ if $a_n = 3p$ or $a_n = p$ for p prime, $a'_n = a_n$ otherwise" [13, Section 1]; (2) differs only at values $2, 6$. OEIS A256417 also uses the all-prime convention [17, A256417]. Hofman and Pilipczuk's Theorem 4.17 gives $a'_n = n + O(n/\log\log\log n)$ [13], the weaker ratio limits in LRS's Conjecture 2 [15].

The companion proves LRS Conjecture 3's uniform linear bounds with computer assistance [10, Theorem 1]. Haver's criterion [12, Corollary 2(e)] reduces the asymptotic for the least unused even integer to counting used odd values above it, excluding $3p$ for odd primes p; Haver's Theorem 1(d) and Corollary 12(d) give the sharp upper bound. We prove the lower half and transfer to all smoothed indices. The reduction and the estimates it uses are Haver's. Haver writes "Our spans are only $O(t/\log t)$ because of rough predecessors" and "rough missing composites" [12, remark after Corollary 12]; these have large least prime factors. The hard step in Propositions 5–6 improves the two displacement bounds from $O(x/\log x)$ to $o(x/\log x)$.

2 Preliminaries

We collect the bounds on used values and missing composites that the sieve will sharpen. Write $L(x) = x/\log x$ for $x > 1$. Let $E(t)$ be the least positive even integer outside U_t; following Haver, we call it the frontier [12, Sections 1–2]. An ordinary value is one unchanged by smoothing. A value $3p$ for an odd prime p is a *tick*, as in Haver [12, Sections 1–2]. Put

$$T(t) = \max\{a(i) : 1 \le i \le t,\ a(i) \text{ is not a tick}\}, \qquad K(t) = \#\{m \in U_t : 2 \mid m,\ m > E(t)\}.$$

LRS prove that a permutes the positive integers [15, Theorem 1]; hence $E(t)$ is nondecreasing and divergent. Haver's sharp upper bound is $E(t) - t \le (1/3 + o(1))t/\log t$ [12, Theorem 1(d), Corollary 12(d)]. Haver's reduction states that

$$E(t) = t + \frac{t}{3\log t} + o(t/\log t) \quad \Longleftrightarrow \quad \#\{w \in U_t : w > E(t),\ w \text{ odd and not a tick}\} = o(t/\log t)$$

[12, Corollary 2(e)]. The task is to establish the count on the right; the two displacement bounds below provide it and the transfer to the smoothed values.

Here $\pi(x)$ counts primes at most x, and $\mathrm{spf}(h)$ is the least prime factor of h. The three-term pattern in (9) is called a prime block; its introduction index N_p is the first occurrence of a multiple of p.

Proof strategy. Propositions 5 and 6 place used ordinary values above $E(t)$ and missing composites below it within $o(L(E(t)))$ of $E(t)$. Haver's bounds give only $O(L(E(t)))$. A displacement comparable to $L(E(t))$ forces too many distinct prime cofactors into a short interval by the greedy rule, contradicting Lemma 3's elementary sieve. The upper estimate enters the lower proof; counting gives (20), then the prime pattern transfers to all indices. In the proofs, "old rate" and "old span" mean (4) and (6).

Haver's estimates are: (5) uses Haver's Theorem 1(a), (6) is Theorem 1(b), (7) is Theorem 11, and (4) is Corollary 12(c) [12]. The prime pattern is Hofman–Pilipczuk's Theorem 4.3 [13], as

formulated in Haver [12, Section 2, (F2)]. The numerical weakenings $793, 802$ are derived in Appendix A.

Proposition 2 (Established sequence estimates)**.** *There are fixed constants $C \geq 1$ and $C_1 > 0$ such that, eventually, writing $e = E(t)$ and $k(t) = K(t) + 1$,*

$$|b(n) - n| \leq CL(n), \tag{4}$$

$$k(t) \leq K_0 \frac{\sqrt{e}}{\log e}, \qquad K_0 = 793, \tag{5}$$

$$T(t) - e \leq C_1 L(e). \tag{6}$$

If $t \geq 9$ and a composite $15 \leq h < e$ is outside U_t, then

$$e - h \leq 2\,\mathrm{spf}(h)\big(13\pi(\sqrt{h}) + 724\big). \tag{7}$$

For all sufficiently large e,

$$F(e) := 13\pi(\sqrt{e}) + 724 \leq B \frac{\sqrt{e}}{\log e}, \qquad B = 802. \tag{8}$$

Every composite below 15 is in U_{13}. For each odd prime p, its first multiple occurs at an index N_p, and

$$a(N_p) = 2p = E(N_p - 1), \quad a(N_p + 1) = p, \quad a(N_p + 2) = 3p. \tag{9}$$

Distinct introduction indices N_p are at least four apart.

3 Greedy bounds and a sieve

The greedy rule and the sieve will bound used ordinary values above the least unused even integer and missing composites below it.

If q is any prime divisor of $a(j-1)$, every positive multiple of q smaller than $a(j)$ was used before index j; otherwise the greedy choice would be smaller. Consequently, at any index t,

$$a(j) \leq E(t) + 2qk(t) \qquad (3 \leq j \leq t+1). \tag{10}$$

Indeed, all multiples of $2q$ strictly between $E(t)$ and $a(j)$ are used even values above $E(t)$, and there are at least $(a(j) - E(t))/(2q) - 1$ of them. Their number is at most $K(t)$. Notice that (10) holds for every prime divisor q, at the same index, even if j is much earlier than t.

Euler's totient $\varphi(W)$ counts the reduced residue classes modulo W, and $\omega(W)$ counts its distinct prime divisors.

We use the sieve of Eratosthenes [6].

Lemma 3. *For every $\eta > 0$ there is a constant C_η such that every closed real interval of length $z \geq 0$ contains at most $\eta z + C_\eta$ primes.*

Proof of Lemma 3. Choose a positive integer $Z \geq 2$ sufficiently large, and put $W = \prod_{p \leq Z} p$. Unique factorization and geometric series give

$$\frac{W}{\varphi(W)} = \prod_{p \leq Z} (1 - 1/p)^{-1} \geq \sum_{m=1}^{Z} \frac{1}{m} \geq \log(Z+1).$$

We may therefore choose Z with $\varphi(W)/W < \eta$. A residue class modulo W has at most $z/W + 1$ members in such an interval. Primes not dividing W lie in its $\varphi(W)$ reduced residue classes, and at most $\omega(W)$ primes divide W. Thus $C_\eta = \varphi(W) + \omega(W)$ works. Both constants are fixed before the interval varies. □

Lemma 4 (Forced common factor)**.** *Let $x, \beta > 0$, let r be prime, and let $M \geq 1$ be an integer. Suppose every prime divisor of $2M$ is smaller than $\beta\sqrt{x}$, and $(\beta\sqrt{x})^3 > 2x$. For selected positive integers $j < \beta\sqrt{x}$, put $y_j = 2rMj$. Suppose the distinct integers d_j are composite, satisfy $d_j < 2x$, have every prime divisor at least $\beta\sqrt{x}$, and satisfy $\gcd(d_j, y_j) > 1$. Then $d_j = rq_j$ for distinct primes q_j. If all d_j belong to an interval I, the q_j belong to the interval I/r.*

Proof. Choose a prime ℓ dividing both d_j and $2rMj$. Every prime divisor of j is at most $j < \beta\sqrt{x}$; every prime divisor of $2M$ is smaller than that bound by assumption. Since all prime divisors of d_j are at least $\beta\sqrt{x}$, ℓ divides neither j nor $2M$. Euclid's lemma forces $\ell \mid r$, so $\ell = r$. A composite d_j has at least two prime factors counted with multiplicity. It cannot have three, since their product would be at least $(\beta\sqrt{x})^3 > 2x > d_j$. It therefore has exactly two, one of which is r; the other is a prime q_j, with $q_j = r$ allowed. Distinct d_j give distinct q_j by cancellation of the fixed positive integer r. Division by r gives the interval assertion. □

Uniform occurrence indices. Put $e = E(t)$. For an ordinary $y = a(j)$ with $e/2 < y < 2e$, $j \to \infty$ uniformly: eventually y exceeds each fixed initial maximum. Equation (4) gives $j/2 \leq y \leq 3j/2$, hence $e/3 < j < 4e$. For an ordinary neighbor $d = a(j \pm 1)$, the same estimate gives

$$|d - y| \leq CL(j) + CL(j \pm 1) + 1 = O(L(e)). \tag{11}$$

with a uniform constant.

4 Displacement bounds

We bound both displacements. Each fixed initial segment has a bounded maximum. An occurrence index of y means j with $a(j) = y$; its pre- and post-occurrence frontiers are $E(j-1)$ and $E(j)$. Observation at t means comparison with the fixed prefix U_t in (10).

Proposition 5. *As $t \to \infty$, $\max\{T(t) - E(t), 0\} = o(L(E(t)))$.*

Proof of Proposition 5. Fix $\delta > 0$. Write $e = E(t)$, $v = T(t)$, and suppose

$$D = v - e \geq \delta L(e). \tag{12}$$

The old span implies $v < 2e$ for sufficiently large e. Choose $i \leq t$ with $a(i) = v$, and let $c = a(i-1)$. The predecessor is not even: the unused even e would then be an eligible choice smaller than v. It is not 1, which appears only at the first index, and it is not an odd prime, since v would then be a tick. If it were a tick, (10) with $q = 3$ would give $D \leq 6k(t) = o(L(e))$, a contradiction. Thus c is an odd composite non-tick, so $c \leq v$. For $r = \mathrm{spf}(c)$, (10) and (5) imply

$$\alpha\sqrt{e} \leq r \leq \sqrt{v} < \sqrt{2e}, \qquad \alpha = \frac{\delta}{2B}. \tag{13}$$

Here $B \geq K_0$, so the displayed lower bound is a weakening of $r \geq \delta\sqrt{e}/(2K_0)$.

Set $\beta = \delta/(4K_0)$, and fix an integer $M \geq 2$ with $M > 1/(\alpha\beta)$. Consider

$$Y = \{2rMj : j \in \mathbb{N},\ j \geq 1,\ e + D/2 < 2rMj < v\}.$$

Every member of Y is a multiple of r smaller than v, so it occurred before i. Counting a lattice in an open interval gives

$$|Y| \geq \frac{D}{4rM} - 2. \tag{14}$$

For $y = a(i_y) \in Y$, put $d_y = a(i_y - 1)$. For every prime $q \mid d_y$, the common observation-time inequality (10) yields

$$q > \frac{D}{4k(t)} \geq \beta\sqrt{e}.$$

For large e this excludes 2 and 3. The predecessor is not 1, a prime, or a tick: in the prime case its successor would be odd, whereas y is even. Therefore d_y is an odd composite non-tick, and $d_y \leq v < 2e$. For large e, $(\beta\sqrt{e})^3 > 2e$.

Write $y = r(2Mj)$. Equation (13) gives

$$j < \frac{\sqrt{e}}{\alpha M} < \beta\sqrt{e}.$$

For large e, every prime factor of the fixed integer $2M$ is also smaller than $\beta\sqrt{e}$. Consecutive values d_y, y share a prime, and distinct selected values have distinct predecessor indices and hence distinct predecessors. Lemma 4 therefore gives

$$d_y = rq_y \quad \text{for distinct primes } q_y. \tag{15}$$

Both y and d_y are unchanged by smoothing. The finite-prefix argument preceding (11) applies uniformly to their indices. Since $e < y < v \leq e + C_1 L(e)$, it gives, for a fixed $A > 0$,

$$e - AL(e) \leq d_y \leq e + AL(e).$$

Distinct y have distinct occurrence indices, distinct predecessor indices, and hence distinct d_y. Thus (15) injects Y into the primes in $[(e - AL(e))/r, (e + AL(e))/r]$. By Lemma 3,

$$\frac{D}{4rM} - 2 \leq \eta\frac{2AL(e)}{r} + C_\eta.$$

Choose the sieve parameters as in Appendix A. It follows that

$$D \leq \frac{\delta}{4}L(e) + 4\sqrt{2}M(C_\eta + 2)\sqrt{e}.$$

Since $\sqrt{e} = o(L(e))$, the right side is less than $\delta L(e)$ for sufficiently large e, contradicting (12). Uniform cutoffs are recorded in Appendix A. □

Unused composites

The upper bound now controls missing composites.

Proposition 6. *For every $\delta > 0$ there is t_0 such that, for all $t \geq t_0$ and every composite $h < E(t)$ outside U_t, $E(t) - h < \delta L(E(t))$.*

Proof of Proposition 6. Fix $\delta > 0$, put $e = E(t)$, and suppose such a missing h has $D = e - h \geq \delta L(e)$. Take $t \geq 13$. Then $h \geq 15$, and h is odd because every positive even below e is used. Put $s = \mathrm{spf}(h)$. Equations (7) and (8) imply

$$D \leq 2sF(e) \leq 2BL(e), \qquad \alpha\sqrt{e} \leq s < \sqrt{e}, \quad \alpha = \frac{\delta}{2B}. \tag{16}$$

Set $\beta = \delta/(6B)$ and fix $M \geq 2$ with $M > 1/(2\alpha\beta)$. Select the even multiples $y = 2sMj$ in $(e - D/2, e)$. They all belong to U_t. Discard $a(t)$ if it belongs to this set, losing at most one value in total, and call the resulting set Y. Then

$$|Y| \geq \frac{D}{4sM} - 3. \tag{17}$$

Each $y = a(i_y)$ has a successor $u_y = a(i_y + 1)$ with $i_y + 1 \leq t$. The value h is still unused and eligible after y, so $u_y \leq h$ by the greedy rule. Equality is impossible since h is missing at time t. Thus $u_y < h$.

By (16), $y > e - BL(e) > e/2$ for large e. The old rate and a finite-prefix maximum force all i_y beyond any fixed cutoff, uniformly in y; in particular they exceed the cutoff for Proposition 5 with parameter $\delta/6$. For $f_y = E(i_y) \leq e$, that proposition gives

$$f_y \geq y - \frac{\delta}{6}L(f_y) \geq y - \frac{\delta}{6}L(e).$$

We used that L is increasing for arguments at least 3. Consequently,

$$f_y - u_y > \frac{D}{2} - \frac{\delta}{6}L(e) \geq \frac{\delta}{3}L(e). \tag{18}$$

These are post-occurrence frontiers. Their divergence follows from the uniform divergence of the indices i_y.

The successor $u = u_y$ is unused at time i_y and below f_y, so it is not even. It is not an odd prime p: its predecessor would be $2p$, whereas $y/2 = sMj$ is composite. It is not a tick, whose predecessor is odd, and it is not 1. It is therefore an ordinary odd composite. Both y and u are unchanged by smoothing. Their occurrence indices satisfy $e/4 < i_y < 2e$ for large e, by the old rate. Applying that rate at $i_y, i_y + 1$ gives

$$|u - y| \leq (6C + 1)L(e), \qquad e - A_B L(e) < u < e, \quad A_B = B + 6C + 1.$$

In particular $u \geq 15$ for sufficiently large e, before we apply the missing-composite bound to it. Combining (7), now at time i_y, with (18) yields

$$\frac{\delta}{3}L(e) < f_y - u \leq 2\,\mathrm{spf}(u)F(e), \qquad \mathrm{spf}(u) > \beta\sqrt{e}.$$

Every prime factor of u is large, while

$$j < \frac{\sqrt{e}}{2\alpha M} < \beta\sqrt{e},$$

and every prime factor of the fixed integer $2M$ is smaller than $\beta\sqrt{e}$ eventually. Since $u < e < (\beta\sqrt{e})^3$ eventually, Lemma 4 gives $u = sq_y$ with q_y prime.

Distinct selected y have distinct successor indices and values, so the q_y are distinct. They lie in $((e - A_B L(e))/s, e/s)$, an interval of length $A_B L(e)/s$. By (17) and Lemma 3,

$$\frac{D}{4sM} - 3 \le \eta \frac{A_B L(e)}{s} + C_\eta.$$

Choose the sieve parameters as in Appendix A. Since $s < \sqrt{e}$,

$$D \le \frac{\delta}{4} L(e) + 4M(C_\eta + 3)\sqrt{e} < \delta L(e)$$

for sufficiently large e, a contradiction.
Uniform cutoffs are recorded in Appendix A. □

5 Proof of Theorem 1

Haver identifies LRS's heuristic [15, Section 4, (6)] as "the identity of (b)" with terms "neglected" [12, remark after Corollary 12]. We retain every term to prove Theorem 1. Fix $t \ge 9$ with $e = E(t) \ge 14$, write $K = K(t)$, and put

$$\begin{aligned}
H &= \#\{h < e : h \text{ composite},\ h \notin U_t\},\\
O &= \#\{w \in U_t : w > e,\ w \text{ odd and not a tick}\},\\
P_1 &= \#\{p \text{ prime} : e/2 \le p < e\},\\
P_2 &= \#\{p \text{ prime} : e/3 < p < e/2\},\\
\theta_1 &= \mathbf{1}_{\{\exists \text{ odd prime } p<e/2 \text{ with } N_p=t\}},\\
\theta_2 &= \#\{p \text{ prime} : e/3 < p < e/2,\ N_p \in \{t-1, t\}\}.
\end{aligned}$$

Let $\theta = \theta_1 + \theta_2$; the spacing of introduction indices gives $0 \le \theta \le 2$.

For $p < e/2$, $2p < e$ is already used, so (9) shows that p is missing precisely when $N_p = t$. For $e/2 \le p < e$, $2p$ has not been used, so p is missing. In counting missing primes, 2 is already used and excluded. For $e/2 \le p < e$, a used $2p \ge e$ would have entered at a frontier $2p = E(N_p - 1) \le e$, forcing the impossible used value e. There are therefore $H + P_1 + \theta_1$ missing positive integers below e. The used ticks above e are exactly those indexed by P_2, except for the θ_2 unfinished blocks, whose third terms have not yet occurred. No used prime lies above e. Counting the t used values below and above the unused value e gives

$$t = (e - 1) - H - P_1 - \theta_1 + K + O + P_2 - \theta_2.$$

Write $I = \mathbf{1}_{\{e/2 \text{ prime}\}}$. Since e is even and at least 14,

$$P_1 = \pi(e) - \pi(e/2) + I, \qquad P_2 = \pi(e/2) - I - \pi(e/3).$$

We obtain the exact identity

$$\begin{aligned} e - t = \pi(e) - 2\pi(e/2) + \pi(e/3) \\ + H - K - O + 1 + \theta + 2I. \end{aligned} \tag{19}$$

Haver proves this counting identity in Corollary 12(b) [12]. Equation (7) counts the missing composites. For $t \geq 13$, classify a missing composite by $s = \mathrm{spf}(h)$. By (7), it is an odd multiple of the odd prime s in $[e - 2sF(e), e)$. There are at most $F(e) + 1$ such multiples, because their spacing is $2s$. Hence

$$H \leq \pi(\sqrt{e})(13\pi(\sqrt{e}) + 725) = O(e/\log^2 e) = o(L(e)).$$

Equation (5) gives $K = o(L(e))$. Proposition 5 gives $O \leq \max\{T(t) - e, 0\}/2 + 1 = o(L(e))$. The correction $1 + \theta + 2I$ is between 1 and 5.

The prime number theorem, in the form $\pi(x) \sim x/\log x$ [11, 8], implies for each fixed $c > 0$ that $\pi(ce)/L(e) \to c$. Thus

$$\frac{\pi(e) - 2\pi(e/2) + \pi(e/3)}{L(e)} \longrightarrow 1 - 2 \cdot \frac{1}{2} + \frac{1}{3} = \frac{1}{3}.$$

Equation (19) yields $e - t = L(e)/3 + o(L(e))$. Since $L(e) = o(e)$ and $e \to \infty$, this implies $e \sim t$, and consequently $L(e) \sim L(t)$. We have proved

$$E(t) = t + \frac{1}{3}L(t) + o(L(t)). \tag{20}$$

The bound on O, with $e \sim t$, supplies the condition in Haver's equivalence [12, Corollaries 2(e), 12(d)].

Proof of Theorem 1. First suppose $x = a(n)$ is ordinary, at a sufficiently large index, and use the pre-occurrence frontier $f = E(n-1)$. If $x < f$, then x is odd, unused at time $n - 1$, and neither 1 nor a prime, so Proposition 6 gives $f - x = o(L(f))$, uniformly over these occurrences. If $x = f$, the discrepancy is zero. If $x > f$, the unused even f remains unused at time n, whence $E(n) = f$. Now $x \leq T(n)$ and Proposition 5 gives $x - f = o(L(f))$. All three cases therefore give

$$a(n) = E(n-1) + o(L(E(n-1))).$$

By (20), $E(n-1) \sim n$, $L(E(n-1)) \sim L(n)$, and $E(n-1) = n + L(n)/3 + o(L(n))$. This proves the desired formula at all ordinary positions. The use of the pre-occurrence frontier also covers a step at which the frontier jumps.

For a prime block with introduction index $m = N_p$, all three smoothed values equal $2p = E(m-1)$. For $n = m, m+1, m+2$, the index $m - 1$ is one of $n - 1, n - 2, n - 3$. Equation (20) is unchanged on the $L(n)$ scale by any of these bounded shifts, since $L(n - j)/L(n) \to 1$ and $j = o(L(n))$. The formula follows at the prime and tick positions too.
The quantifiers in (3) are verified in Appendix A. □

6 Consequences

Theorem 1 gives the three branches of the original sequence:

Corollary 7 (The three branches)**.** *Along the indices of each indicated kind, as* $n \to \infty$*,*

$$\begin{aligned} a(n) &= n + \frac{n}{3\log n} + o(n/\log n), && a(n) \neq p, 3p \quad (p \text{ odd prime}), \\ p &= \frac{n}{2} + \frac{n}{6\log n} + o(n/\log n), && a(n) = p, \\ 3p &= \frac{3n}{2} + \frac{n}{2\log n} + o(n/\log n), && a(n) = 3p. \end{aligned}$$

Proof. In the three cases $a(n) = b(n)$, $a(n) = b(n)/2$, and $a(n) = 3b(n)/2$, respectively. Multiply Theorem 1 by the fixed factor. The finitely many cases involving 2 do not matter for any limit. □

For large k with $a(k) \in \{p, 3p\}$, $b(k) = 2p \leq 2k$, so at most $2\pi(N) + O(1) = o(N)$ indices $k \leq N$ are of these kinds. Thus the first branch gives the LRS formula at almost all original indices. Figure 1 plots the initial terms; Appendix B illustrates the smoothed error.

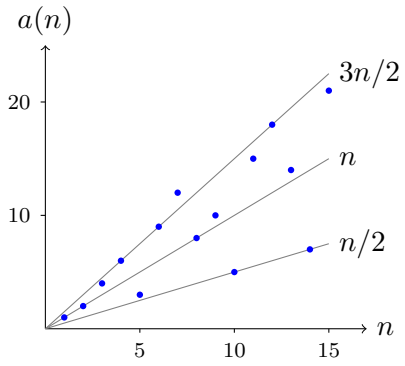


Figure 1: The first fifteen terms and the three critical lines.

The diagonal, three lines and cycles

The second-order term settles these diagonal questions.

Thomas Ordowski conjectured "If $a(n) \neq p$, then almost everywhere $a(n) > n$" and "$\lim \#(a_n > n)/n = 1$" [17, A064413, comments, January 23, 2009]. Artur Jasinski conjectured finiteness of A152458 ($a(k) = k$) and A152519 [17, A152458 and A152519, comments, December 2008]. Here p denotes a prime.

Hofman–Pilipczuk [13, Theorem 4.17] and Haver [12, Corollary 12(c)] establish $b(n)/n \to 1$. Our counting argument below already gives from this ratio $\#\{k \leq N : a(k) \text{ prime}\} \sim N/(2\log N)$. The new conclusion $b(n) > n$ for all large n identifies these prime-value indices, up to finitely many exceptions, with those below the diagonal.

Lemma 8. *For every* $\eta > 0$ *there is* $N_0 \geq 3$ *such that* $n < b(n) \leq (1+\eta)n$ *for all* $n \geq N_0$*. For every* n*, exactly one of the following holds:* $a(n)$ *is an odd prime and* $b(n) = 2a(n)$*;* $a(n) = 3p$ *for an odd prime* p *and* $2a(n) = 3b(n)$*; or* $b(n) = a(n)$*.*

The elementary inequalities are verified in Appendix A.

Corollary 9 (The diagonal)**.** *For all sufficiently large n, $a(n) < n$ if $a(n)$ is prime and $a(n) > n$ if $a(n)$ is not prime. Thus, apart from finitely many n, $a(n) > n$ exactly when $a(n)$ is composite, and $a(n) \neq n$. Moreover, as $N \to \infty$,*

$$\#\{k \le N : a(k) < k\} \sim \frac{N}{2\log N}, \qquad \#\{k \le N : a(k) > k\} = N - \Big(\frac{1}{2} + o(1)\Big)\frac{N}{\log N}.$$

In particular $\#\{k \le N : a(k) > k\}/N \to 1$.

Proof. Take N_0 from Lemma 8 with $\eta = 1/3$, and let $n \ge N_0$. If $a(n)$ is prime, it is not 2, since $b(n) > n \ge 3$; so $a(n) = b(n)/2 \le 2n/3 < n$. If $a(n)$ is not prime, it is not 1 (else $b(n) = 1 < n$), and either $a(n) = b(n) > n$ or $a(n) = 3b(n)/2 > n$.

For the counts, fix $\eta \in (0, 1/3]$ and its N_0. For $N \ge N_0$, the sets $\{k \le N : a(k) < k\}$ and $\{k \le N : a(k) \text{ prime}\}$ agree on $[N_0, N]$, and $a(k) = k$ only for $k < N_0$. Let $P(N)$ be the number of $k \in [N_0, N]$ with $a(k)$ prime. Each such k has $a(k) = b(k)/2 \le (1+\eta)N/2$, and a is injective, so $P(N) \le \pi((1+\eta)N/2)$. Conversely, let $M = \max\{a(k) : k < N_0\}$. Since a is a permutation, each prime p with $M < p \le N/2$ is $a(k)$ for one k, and $k \ge N_0$ because $p > M$. Then $k < b(k) = 2p \le N$, so $P(N) \ge \pi(N/2) - M$. By the prime number theorem $\pi(cN) \sim cN/\log N$ for each fixed $c > 0$, hence

$$1 - o(1) \le \frac{P(N)}{N/(2\log N)} \le 1 + \eta + o(1),$$

and η is arbitrary. The second count is N minus the first and minus the finitely many k with $a(k) = k$. □

Corollary 9 strengthens Ordowski's first conjecture and proves Ordowski's second.

Corollary 9 also gives finiteness of Ordowski's A155549, the nonprime values $a(n) < n$, and A064420, the k with $a(k+1) = k$. The short deductions are in Appendix A. De Vlieger [9] called finiteness of A064420 "not surprising".

Corollary 10 (Three lines)**.** *Each of the sets $\{k : 2a(k) = k\}$, $\{k : a(k) = k\}$ and $\{k : 2a(k) = 3k\}$ is finite.*

The elementary inequalities are verified in Appendix A.

In fact the ratios $a(n)/n$ eventually lie in $(1/2, 2/3] \cup (1, 4/3] \cup (3/2, 2]$, avoiding $1/2$, 1 and $3/2$. The finiteness results do not establish completeness of the OEIS lists; Amiram Eldar recorded in 2024 that neither A152458 nor A152519 has a further term below 10^5 [17, A152458 and A152519, Eldar comments, 2024].

LRS write "The experimental evidence suggests that there are infinitely many finite cycles and infinitely many infinite cycles" [15, Section 7]; we treat bounded cycle lengths:

Corollary 11 (Short cycles)**.** *For each $L \ge 1$, the permutation $n \mapsto a(n)$ of the positive integers has only finitely many cycles of length at most L.*

In the cycle proof, $(\sigma, \tau) = (0,0), (1,0), (1,1)$ correspond to unchanged values, odd primes and three times odd primes; L bounds cycle length.

Proof. By Lemma 8, for each n,

$$2^{\sigma(n)} a(n) = 3^{\tau(n)} b(n), \qquad (\sigma(n), \tau(n)) \in \{(0,0), (1,0), (1,1)\}.$$

Choose $\eta > 0$ with $(1+\eta)^L < 2^L/(2^L-1)$, and N_0 for this η. Suppose a cycle $x_1, \dots, x_\ell$, with $a(x_i) = x_{i+1}$ for $i < \ell$, $a(x_\ell) = x_1$ and $\ell \le L$, has every $x_i \ge N_0$. Since a permutes the x_i, $\prod a(x_i) = \prod x_i$; with $u = \sum \sigma(x_i)$ and $v = \sum \tau(x_i)$, so $v \le u \le \ell$,

$$2^u \prod_{i=1}^{\ell} x_i = 3^v \prod_{i=1}^{\ell} b(x_i).$$

As $b(x_i) > x_i$, this forces $3^v < 2^u$, so $u \ge 1$ and $3^v \le 2^u - 1$. As $b(x_i) \le (1+\eta)x_i$, the right side is then at most $(2^u-1)(1+\eta)^\ell \prod x_i$, which is less than $2^u \prod x_i$ because $(1+\eta)^\ell < 2^L/(2^L-1) \le 2^u/(2^u-1)$. This contradiction shows that every cycle of length at most L contains an integer below N_0. Distinct cycles are disjoint, so there are fewer than N_0 of them. □

7 Formal verification

The formal statement matches Theorem 1: Theorem `ekg_sharp_all_index` states that every real $\varepsilon > 0$ admits a natural $N \ge 3$ for which all natural $n \ge N$ satisfy

$$\left| b(n) - n - \frac{n}{3 \log n} \right| \le \varepsilon \frac{n}{\log n}.$$

The canonical sequence follows (1): it starts $1, 2$ and then chooses the least unused positive integer not coprime to its predecessor. The paper uses positive indices; the formal sequence extends to zero by $a(0) = 1$. Smoothing (2) fixes $2, 6$. Arithmetic is real, logarithms natural, and there are no additional hypotheses. The proof uses propositional extensionality, classical choice and quotient soundness. Unproved library declarations lie outside its dependencies. We use Lean [7], mathlib [16] and the prime number theorem library [14]. Our Lean proof imports Haver's formal proofs of the estimates (Proposition 2) and checks their weakenings from the canonical definition. The formal proof derives the asymptotic directly from these estimates and counting identities. Haver's upper bound and criterion enter only the comparison with Haver's results.

Haver's Lean archive already proves the general counting identities [12, Corollary 12(b)]; our proof discharges their prime-interval and tick endpoint hypotheses and derives the coefficient and all-index transfer. Haver's sharp upper bound and second-order criterion lie outside Haver's archive's verified core [12, verification note, "Claims outside the checked coverage"]. Lean also checks the sign, density, below-diagonal count, three-line and exceptional-set finiteness, and bounded-cycle finiteness; the three-branch formulas (Corollary 7) and the second count in Corollary 9 are deductions outside Lean.

Acknowledgments. The author thanks Tomer Haver for the estimates, counting identity, reduction and accompanying Lean sources.

Appendix A Routine estimates

We record the uniform parameter choices and the elementary inequalities used in the consequences.

Constants in the established estimates. The constants $793 = 12{\cdot}6{+}721$ and $802 = 13{\cdot}6{+}724$ are convenient weakenings using $\pi(\sqrt{e}) \le 6\sqrt{e}/\log e$ and $\sqrt{e}/\log e \ge 1$ eventually; $C_1 = 4 \cdot 793^2 + 2 \cdot 793$ suffices.

Parameters for Proposition 5. The constants A and η in the interval and sieve bounds may be chosen as follows. All constants in this proof are chosen before the final threshold on t. For example, $A = C_1 + 8C + 1$ suffices. Choose $\eta = \delta/(32AM)$ after fixing M. The frontier tends to infinity, so one threshold on t works for every later prefix. This proves the proposition.

Parameters for Proposition 6. Here s is the fixed factor of the missing composite, and u is a selected successor; the factorization $u = sq_y$ uses that fixed factor. This does not assert that $s = \mathrm{spf}(u)$.

Fix $\eta = \delta/(16MA_B)$. The choices have the order δ, the upper-excursion cutoff, M, η, C_η, and finally the frontier and index cutoffs. All are uniform in the missing composite h and the selected y.

Quantifiers for Theorem 1. In the following verification, the clock estimates are (20), the excursion propositions are Propositions 5–6, and finite seed exclusions mean the finitely many initial indices omitted by their eventual estimates. The upper-excursion cutoff above is the threshold in Proposition 5. For completeness, the uniform little-o conclusions here give exactly the quantifiers in (3). Given $\varepsilon > 0$, apply the two excursion propositions with $\delta = \varepsilon/8$. Eventually $L(E(n-1)) \le 2L(n)$, so their contribution is at most $\varepsilon L(n)/4$. Take a common cutoff for the clock estimates shifted by $1, 2, 3$, each with error at most $\varepsilon L(n)/2$, and for the finite seed exclusions. Their sum is at most $\varepsilon L(n)$ at every ordinary index; the shifted clock alone covers the prime blocks. Increasing the common cutoff to at least 3 proves the theorem.

Proof of Lemma 8. Apply (3) with $\varepsilon = 1/6$: for $n \ge N$, $b(n) \ge n + n/(6\log n) > n$ and $b(n) \le n + n/(2\log n)$, which is at most $(1+\eta)n$ once $\log n \ge 1/(2\eta)$. The second statement is the definition (2). □

Proof of Corollary 10. Take N_0 from Lemma 8 with $\eta = 1/3$, and let $n \ge N_0$, so $n < b(n) \le 4n/3$. If $a(n)$ is an odd prime, then $2a(n) = b(n)$, so $n < 2a(n) \le 4n/3$. If $a(n) = 3p$ for an odd prime p, then $2a(n) = 3b(n) > 3n$. Otherwise $a(n) = b(n)$ and $2n < 2a(n) \le 8n/3$. In no case is $2a(n)$ equal to n, $2n$ or $3n$, so all three sets lie in $[1, N_0)$. □

The exceptional OEIS sets. By Corollary 9 the indices n with $a(n)$ not prime and $a(n) \le n-1$ form a finite set, so its image A155549 is finite. For the second, $a(n) = n - 1$ fails for large n: a prime value has $a(n) \le 2n/3$ and any other value has $a(n) > n$.

Appendix B Numerical illustration

Table 1 displays $R(n) = (b(n) - n)\log n/n$, computed by the recursive greedy rule (1) with prime-factor sieving.

n	$b(n)$	$R(n)$
1,000	1,043	0.2970
10,000	10,361	0.3325
100,000	103,088	0.3555
1,000,000	1,024,686	0.3410
10,000,000	10,210,804	0.3398
100,000,000	101,839,957	0.3389
500,000,000	508,457,544	0.3388

Table 1: The normalized smoothed error, rounded to four decimal places; Theorem 1 gives its limit $1/3$.

Related work. Haver's Yellowstone papers [1, 2, 3, 4, 5] are unused.

Other nearby visiting rules include Applegate, Havermann, Selcoe, Shevelev, Sloane and Zumkeller's Yellowstone permutation [18], Basistha and Ionaşcu's coprime permutation [19], and Na's real-multiplier process [20]. Nothing from these works is used here.

References

[1] T. Haver, *The Yellowstone permutation, III: few large terms*, October 22, 2026. doi:10.5281/zenodo.23557280.

[2] T. Haver, *The prime frontier of the Yellowstone permutation is linear*, October 22, 2026. doi:10.5281/zenodo.23558450.

[3] T. Haver, *The prime frontier of the Yellowstone permutation falls behind time: limsup R(t)/t < 1*, October 23, 2026. doi:10.5281/zenodo.23604646.

[4] T. Haver, *The Yellowstone permutation: limsup R(t)/t ≤ 0.68605, liminf n_p/p > 1.4576, and Conjecture 7 for large n*, October 23, 2026. doi:10.5281/zenodo.23604811.

[5] T. Haver, *The Yellowstone permutation: Conjecture 7 for all n reduced to a riser law on a finite range, the normality and entry times of primes, and a run certificate*, October 23, 2026. doi:10.5281/zenodo.23604970.

[6] A. C. Cojocaru and M. R. Murty, *An Introduction to Sieve Methods and Their Applications*, LMS Student Texts **66**, Cambridge University Press, 2005, Sections 5.1–5.2. doi:10.1017/CBO9780511615993.

[7] L. de Moura and S. Ullrich, The Lean 4 theorem prover and programming language, in *Automated Deduction—CADE 28*, LNCS **12699**, Springer, 2021, pp. 625–635. doi:10.1007/978-3-030-79876-5_37.

[8] Ch.-J. de la Vallée Poussin, Recherches analytiques sur la théorie des nombres premiers. Première partie: Les nombres premiers en général, *Ann. Soc. Sci. Bruxelles* **20**, deuxième partie (1896), Section 56 (*Digitized collected offprint*, 1897, pp. 67–69). Original offprint.

[9] M. De Vlieger, Constitutive analysis of the EKG sequence, 2021, `https://vincico.com/seq/a064413.html`.

[10] P. Graham, Sharp linear bounds for the EKG sequence, version 1, October 22, 2026, manuscript and Lean sources. doi:10.5281/zenodo.23562177.

[11] J. Hadamard, Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques, *Bull. Soc. Math. France* **24** (1896), 199–220. doi:10.24033/bsmf.545.

[12] T. Haver, Explicit rates for the EKG sequence, version 3, October 4, 2026, manuscript and Lean sources. doi:10.5281/zenodo.23129632.

[13] P. Hofman and M. Pilipczuk, A few new facts about the EKG sequence, *J. Integer Sequences* **11** (2008), Article 08.4.2.

[14] A. Kontorovich, T. Tao, et al., *PrimeNumberTheoremAnd* (Prime Number Theorem and More), Lean sources, revision `c39a751132c88b6e8080b74c74023fd95b3d8be0`, 2026. https://github.com/AlexKontorovich/PrimeNumberTheoremAnd.

[15] J. C. Lagarias, E. M. Rains, and N. J. A. Sloane, The EKG sequence, *Experimental Math.* **11** (2002), 437–446. doi:10.1080/10586458.2002.10504486; arXiv:math/0204011v1.

[16] The mathlib Community, The Lean mathematical library, in *Proceedings of CPP '20*, ACM, 2020, pp. 367–381. doi:10.1145/3372885.3373824; arXiv:1910.09336.

[17] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, entries A064413, A064420, A152458, A152519, A155549 and A256417.

[18] David L. Applegate, Hans Havermann, Robert G. Selcoe, Vladimir Shevelev, N. J. A. Sloane and Reinhard Zumkeller, *The Yellowstone Permutation*, arXiv:1501.01669, first posted January 7, 2015; version 2, March 8, 2015. https://arxiv.org/abs/1501.01669v2.

[19] Amit Kumar Basistha and Eugen J. Ionaşcu, *A special sequence and primorial numbers*, arXiv:2302.02838, first posted February 6, 2023; version 5, November 22, 2023. Rev. Roumaine Math. Pures Appl. 70 (2025), 235–255. https://arxiv.org/abs/2302.02838v5. https://doi.org/10.59277/rrmpa.2025.235.255.

[20] Sungsoo Na, *An integer-visiting process with real multipliers: coverage and rates*, version 1.0.0, October 4, 2026. https://doi.org/10.5281/zenodo.23138274.

2020 *Mathematics Subject Classification*: Primary 11B83; Secondary 11A05, 68V15.
Keywords: EKG sequence, greedy permutation, prime number theorem, asymptotic formula, formal verification.
Concerned with sequences A064413, A064420, A155549, A152458, A152519 and A256417.