APERIODICITY AND SUBWORD COMPLEXITY IN THE BINARY EXPANSION OF POWERS OF THREE

RALF STEPHAN

Abstract. We prove two results on the fine structure of the binary digits of 3^m. First, for every fixed period p, the number of positions at which the binary expansion of 3^m breaks p-periodicity grows in order like $\log m/\log\log m$; equivalently, no window of the expansion deeper than a fixed power of $\log m$ is p-periodic. Second, the finite binary word formed by the low-order digits of 3^m has full low-order subword complexity: its complexity function satisfies $p_{3^m}(n) \geq n+1$ for every length n, once m is large enough.

1. Introduction

Write a positive integer N in binary, $N = \sum_{i\geq 0} \mathrm{bit}_i(N)\, 2^i$ with $\mathrm{bit}_i(N) \in \{0,1\}$, and let

$$M(N) := \lfloor \log_2 N \rfloor, \qquad s_2(N) := \sum_{i\geq 0} \mathrm{bit}_i(N),$$

so that N has $M(N)+1$ binary digits and $s_2(N)$ is its number of 1-bits. This note concerns the arrangement of the binary digits of $N = 3^m$. Throughout, "log" is the natural logarithm.

We use two further statistics of the digit string. A *digit change* of N is an index $i < M(N)$ with $\mathrm{bit}_i(N) \neq \mathrm{bit}_{i+1}(N)$; writing N as a concatenation of maximal runs of equal digits (its *blocks*), the number of digit changes is one less than the number of blocks. For a fixed period $p \geq 1$, a *period-p break* of N is an index $i < M(N)$ with $\mathrm{bit}_i(N) \neq \mathrm{bit}_{i+p}(N)$.

Definition 1.1. For $N \geq 1$, let

$$T(N) := \#\{\, i < M(N) : \mathrm{bit}_i(N) \neq \mathrm{bit}_{i+1}(N) \,\},$$
$$B_p(N) := \#\{\, i < M(N) : \mathrm{bit}_i(N) \neq \mathrm{bit}_{i+p}(N) \,\}.$$

Thus $T(N)$ is the number of digit changes of N (its number of blocks is $T(N)+1$), and $B_p(N)$ is its number of period-p breaks; note $B_1 = T$.

Finally, for a word $w = (w_0, w_1, \dots)$ over a finite alphabet, the *subword complexity* (or *factor complexity*) $p_w(n)$ is the number of distinct factors of length n occurring in w. We read the low bits of 3^m as the finite word $w = (\mathrm{bit}_0(3^m), \dots, \mathrm{bit}_{M-1}(3^m))$, $M = M(3^m)$, and write $p_{3^m}(n)$ for

Date: July 16, 2026.
Institute for Globally Distributed Open Research and Education (IGDORE).

its complexity: the number of distinct length-n factors at the positions $0, 1, \dots, M-n$.

1.1. **Known results.** The starting point is a theorem of Stewart on integers with few digits in two bases. In the notation of [Ste80], $s_2(N)$ is the number of nonzero binary digits, and Stewart's Theorem 2, applied to the recurrence $u_n = 3^n$ with base 2, gives the following.

Theorem 1.2 (Stewart [Ste80]). *There is a constant $C > 0$ such that*

$$s_2(3^m) \ \geq \ \frac{\log m}{\log\log m + C} - 1 \qquad \textit{for every } m \geq 2.$$

In particular $s_2(3^m) \to \infty$: no power of three eventually has sparse binary digits.

Stewart's proof rests on Baker's theory of linear forms in logarithms; in the present work the corresponding input is the following effective theorem of Baker and Wüstholz [BW93], stated in Section 2. For an embedding $\varphi\colon K \to \mathbb{C}$ of a number field K of degree d, write $h'(\alpha)$ for the modified height of $\alpha \in K$ and $C(n,d)$ for the Baker–Wüstholz constant; the theorem asserts that a nonzero linear form $\Lambda = \sum_i b_i \log\varphi(\alpha_i)$ with integer coefficients bounded by $B \geq 2$ obeys $\log|\Lambda| \geq -C(n,d)\max(\log B, \frac{1}{d})\prod_i h'(\alpha_i)$ (Theorem 2.1).

Counting the 1-bits says nothing about their placement, and it is natural to count instead the blocks of 3^m. That the number of blocks tends to infinity is a theorem of Blecksmith, Filaseta and Nicol [BFN93], who introduced this quantity; an effective, and faster, lower bound follows from work of Bugeaud and Kaneko. Since 3^m is an integral S-unit with $S = \{3\}$, their Corollary 1.5 and the block variant recorded in their Remark 4.4 apply.

Theorem 1.3 (Blecksmith–Filaseta–Nicol [BFN93]; Bugeaud–Kaneko [BK17]). *The number of binary blocks of 3^m tends to infinity, and grows in order at least like $\log m/\log\log m$; equivalently $T(3^m) \geq \log m/(\log\log m + C) - 1$ for a constant $C > 0$ and all $m \geq 2$.*

Here the order $\log m/\log\log m$ comes from $\log\log(3^m) = \log m + O(1)$ in [BK17, Cor. 1.5, Rem. 4.4]. On the opposite, sparse side, the powers of three with few nonzero bits are known exactly in the relevant range. The case $s_2 \leq 2$ is elementary; the case $s_2 = 3$ is a theorem of Dimitrov and Howe.

Theorem 1.4 (Dimitrov–Howe [DH23]). *The equation $3^m = 2^a + 2^b + 1$ with $1 \leq b < a$ has the unique solution $(m,b,a) = (4,4,6)$. More generally* [DH23, Thm. 1.1]*, the powers of three with at most 22 nonzero binary digits are exactly $3^0, \dots, 3^{25}$.*

Theorem 1.5 (sparse powers of three). *$s_2(3^m) \leq 2$ if and only if $m \leq 2$, and $s_2(3^m) = 3$ if and only if $m = 4$. Explicitly $3^0 = 1_2$, $3^1 = 11_2$, $3^2 = 1001_2$ have $s_2 \leq 2$, and $3^4 = 1010001_2$ is the unique power of three with $s_2 = 3$.*

The first assertion of Theorem 1.5 is elementary (peeling the least significant bit reduces it to ruling out $3^m = 2^a + 1$ for $m \geq 3$); the second is the $s_2 = 3$ reading of Theorem 1.4. The smallest powers of three breaking the sparse bound are $3^3 = 11011_2$ ($s_2 = 4$) and $3^4 = 1010001_2$ ($s_2 = 3$).

The combinatorial input to our second result is a finite form of the Morse–Hedlund theorem [MH38]: an infinite word of subword complexity $p_w(n) \leq n$ for some n is eventually periodic. We use a quantitative finite version, stated and proved in Section 4 (Lemma 4.1), in the spirit of the special-factor analysis of Carpi and de Luca [CdL00].

1.2. **Results of this paper.** Theorems 1.2 and 1.3 forbid long constant runs. We strengthen this in two directions, neither of which, to our knowledge, has been treated for the integer 3^m; the existing quantitative results on digit changes and on subword complexity in this circle concern algebraic irrationals, whose expansions are infinite, rather than the finite word of a single power.

Our first result forbids long *periodic* runs, of any fixed period. It is proved in Section 3.

Theorem A. *For every fixed period $p \geq 1$ there is a constant $C_p > 0$ such that*

$$B_p(3^m) \;\geq\; \frac{\log m}{\log\log m + C_p} - 2 \qquad \text{for every } m \geq 2;$$

in particular $B_p(3^m) \to \infty$. Equivalently, no window of the binary expansion of 3^m deeper than a fixed power of $\log m$ can be p-periodic.

The case $p = 1$ recovers Theorem 1.3. The engine is a new estimate for periodic windows (Lemma 2.2), a four-logarithm companion of the run estimates behind the classical bounds.

Our second result is a Morse–Hedlund floor for the finite binary word of a single 3^m. It is proved in Section 4.

Theorem B. *For every length $n \geq 1$, the binary word of 3^m has full low-order subword complexity,*

$$p_{3^m}(n) \;\geq\; n + 1,$$

for all sufficiently large m.

The threshold is effective: $p_{3^m}(n) \geq n+1$ once m exceeds the point where the linear growth of $M(3^m) \geq m$ overtakes a bound of order $n^2 \log m$ coming from Theorem A's engine.

Both theorems have been formally verified in the Lean 4 proof assistant; Appendix A records the Lean name and status of every statement appearing in this paper.

2. Linear forms in logarithms and the periodic-run estimate

The engine common to the classical bounds and to Theorems A and B is Stewart's: a long run, or more generally a long periodic stretch, in the binary expansion of 3^m produces a small nonzero linear form in the logarithms of 2, 3 and an auxiliary integer, which the Baker–Wüstholz theorem forbids from being too small.

Theorem 2.1 (Baker–Wüstholz [BW93])**.** *Let $\alpha_1, \dots, \alpha_n$ be nonzero elements of a number field K of degree d, and let $b_1, \dots, b_n \in \mathbb{Z}$ satisfy $|b_i| \le B$ with $B \ge 2$. If $\Lambda := \sum_i b_i \log \varphi(\alpha_i) \neq 0$ (principal branch), then*

$$\log|\Lambda| \;\ge\; -C(n,d) \cdot \max\Big(\log B, \tfrac{1}{d}\Big) \cdot \prod_i h'(\alpha_i).$$

We use Theorem 2.1 only over $K = \mathbb{Q}$ $(d = 1)$, where each α_i is a positive rational, $h'(2) = \log 2$, $h'(3) = \log 3$, and $h'(a) = \log a$ for a positive integer a. Set $M := M(3^m) = \lfloor \log_2 3^m \rfloor$.

The estimates behind Theorems 1.2 and 1.3 bound how far below the top of the expansion a run of *constant* bits can begin: a run of zeros (resp. ones) in positions $[x, y)$ forces $(M - x)\log 2 \le C(3,1)\log 3\,(M + 1 - y)\max(\log 2m, 1)$, via a three-term linear form.[1] The result of this paper needs the analogous estimate for *periodic* runs, which is new and uses a four-term form.

Lemma 2.2 (Period-p run estimate)**.** *Let $m \ge 1$, $p \ge 1$, $1 \le x < y \le M$ with $2p \le y - x$, and suppose $\mathrm{bit}_i(3^m) = \mathrm{bit}_{i+p}(3^m)$ for all $x \le i < y - p$. Then*

$$(M-x)\log 2 \;\le\; C(4,1)\,p\,\log 3\,(M+2p-y)\,\max\big(\log(2m+p), 1\big) + (p+1)\log 2.$$

Proof sketch. Trim the window to a whole number of periods, $y' = x + p\lfloor (y-x)/p \rfloor$. With $A_2 = 3^m \bmod 2^x$, $c = \lfloor 3^m/2^x \rfloor \bmod 2^p$ and $A_1' = (2^p - 1)\lfloor 3^m/2^{y'} \rfloor + c$, periodicity gives the integer identity $(2^p - 1)\,3^m = A_1'\,2^{y'} + E$, where $E = (2^p-1)A_2 - c\,2^x$. Since $(2^p-1)A_2$ is odd and $c\,2^x$ is even (here $x \ge 1$ is used), E is odd, hence nonzero—the parity nondegeneracy that replaces Stewart's use of the irrationality of $\log 2/\log 3$. The four-term form

$$\Lambda := \log(2^p - 1) + m\log 3 - y'\log 2 - \log A_1'$$

satisfies $|\Lambda| \le 2^{x+2-M}$; the hypothesis $2p \le y - x$ forces $x \le M - 2$, so the associated ratio R has $|R - 1| \le \frac{1}{2}$ and the two-sided estimate $|\log R| \le 2|R-1|$ applies. Theorem 2.1 with $\alpha = (2^p - 1, 3, 2, A_1')$, $b = (1, m, -y', -1)$, $B = 2m + p$, and the height bounds $h'(2^p - 1) \le p$ and $h'(A_1') \le M + 1 + p - y'$, gives the stated bound (after replacing $M + 1 + p - y'$ by the larger $M + 2p - y$). □

[1] We use the descriptive names *run estimate* for these lemmas and *period-p run estimate* for Lemma 2.2; the formalization calls them "gap principles". Each is an instance of Theorem 2.1.

Stewart's counting device turns such estimates into an effective lower bound. Fix $\theta \geq 2$ and consider, for $j = 0, 1, 2, \ldots$, the depth windows of bit positions

$$\big[\, M - \theta^{j+1},\ M - \theta^{j}\,\big);$$

there are $\lfloor \log_\theta M \rfloor$ disjoint such windows below the top bit. If a counting function is bounded below by the number of windows carrying a witness, an effective lower bound follows with a single constant.

Proposition 2.3. *Let $\kappa > 0$ and $Q\colon \mathbb{N} \to \mathbb{N}$. Suppose that for all $m \geq 2$ and all $\theta \geq 2$ with $\kappa \log m < \theta \log 2$ one has $\lfloor \log_\theta \lfloor \log_2 3^m \rfloor \rfloor \leq Q(m)$. Then there is a constant $C > 0$, which one may take to be* $\max\big(\log \frac{\kappa+2}{\log 2}, 1\big)$, *such that*

$$Q(m) \ \geq\ \frac{\log m}{\log\log m + C} - 1 \qquad \textit{for every } m \geq 2.$$

Proof sketch. Choose $\theta \approx C' \log m$ just above the threshold, so $\theta \leq C' \log m$ for a constant C'; then $k := \lfloor \log_\theta M \rfloor \leq Q(m)$ by hypothesis. Since $3^m \geq 2^m$ gives $M \geq m$, one has $\log m \leq (k{+}1) \log \theta$ and $\log \theta \leq \log\log m {+} C$; dividing yields the bound. Positivity of the denominator uses $\log\log 2 > -1$. □

Lemma 2.4. *If $Q\colon \mathbb{N} \to \mathbb{N}$ satisfies $Q(m) \geq \log m/(\log\log m + C) - 1$ for all $m \geq 2$ and some $C > 0$, then $Q(m) \to \infty$.*

Proof sketch. The right-hand side tends to infinity, since $\log m$ dominates $\log\log m$; the quantitative threshold uses $\log\log m \leq 2\sqrt{\log m}$. □

3. Aperiodicity: period-p breaks

We prove Theorem A: for each fixed period p, the number of places where the binary expansion of 3^m fails period p grows at the Stewart rate.

Proof of Theorem A. Fix $p \geq 1$. By the period-p run estimate (Lemma 2.2), a depth window deeper than a threshold $\theta_p \asymp p^2 \log(2m + p)$ cannot be p-periodic: it contains an index i with $\mathrm{bit}_i(3^m) \neq \mathrm{bit}_{i+p}(3^m)$, that is, a period-$p$ break. Running the window scheme above, all but one of the $\lfloor \log_\theta M \rfloor$ disjoint windows (one is dropped to keep the base position of the periodic stretch at least 1, as Lemma 2.2 requires) inject into the set of period-p breaks, so

$$\Big\lfloor \log_\theta M \Big\rfloor \ \leq\ B_p(3^m) + 1.$$

Applying Proposition 2.3 to $Q(m) = B_p(3^m) + 1$ with the coefficient

$$\kappa_p = C(4,1) \log 3{\cdot}p(1{+}2p)\, \beta_p {+} (p{+}1) {+} 2p \ \asymp\ p^2 \log p, \qquad \beta_p = 2 {+} \tfrac{\log(2+p)+1}{\log 2},$$

gives $B_p(3^m) + 1 \geq \log m/(\log\log m + C_p) - 1$, and subtracting 1 yields the stated inequality. The divergence $B_p(3^m) \to \infty$ is Lemma 2.4. □

Remark 3.1. The offset is -2, rather than the -1 of Theorems 1.2 and 1.3, because one window is sacrificed: Lemma 2.2 requires the base position of the periodic stretch to be at least 1. The bound is uniform in m for each fixed p; the growth $C_p \asymp \log p$ of the constant is carried by the coefficient $\kappa_p \asymp p^2 \log p$.

4. Subword complexity

We now prove Theorem B. The combinatorial input is the following finite form of the Morse–Hedlund theorem, in which we make no appeal to Fine–Wilf; it is proved by a self-contained determinism-propagation argument.

Lemma 4.1 (Finite Morse–Hedlund floor). *Let $u = (u_0, u_1, \dots)$ be a word over an alphabet with decidable equality, and let $n \geq 1$ and $L \geq 3n$. If u has at most n distinct factors of length n among the positions $0, \dots, L-n$, then there exist a and $1 \leq p \leq n$ such that u has period p on the factor $[a, a + (L-2n))$; that is, $u_t = u_{t+p}$ for all $a \leq t$ with $t + p < a + (L-2n)$.*

Proof sketch. On the fixed position set $[0, L-n]$ the number $c(k)$ of distinct length-k factors is non-decreasing in k: dropping the last letter of a length-$(k+1)$ factor surjects onto the length-k factors. Since $c(0) = 1$ and $c(n) \leq n$, the count cannot strictly increase n times, so there is a plateau $c(k) = c(k+1)$ with $k < n$. A plateau forces right-determinism—two positions carrying the same length-k factor carry the same next letter. Pigeonhole over the first $c(k) + 1$ positions produces two equal length-k factors at an offset $p \leq n$, and determinism propagates the period-p relation across the window. □

Feeding such a periodic factor to the period-p run estimate caps its depth, in the following effective form.

Lemma 4.2. *Let $n \geq 1$ and $M = \lfloor \log_2 3^m \rfloor \geq 4n + 1$. If $p_{3^m}(n) \leq n$, then*

$$\big(M - (3n+2)\big) \log 2 \;\leq\; 4\, C(4,1)\, n^2 \,\log 3\, \max\big(\log(2m+n), 1\big).$$

Proof sketch. By Lemma 4.1 with $L = M$, a factor complexity $\leq n$ yields a period-p factor with $1 \leq p \leq n$ of length $M - 2n$, starting at some $a \leq 2n$. Apply the period-p run estimate (Lemma 2.2) with base $x = a + 1 \geq 1$ and $y = a + (M - 2n)$: the depth factor is $M + 2p - y = 2p + 2n - a \leq 4n$, and $p(2p + 2n - a) \leq 4n^2$ collapses the right-hand side to the stated form. □

Proof of Theorem B. Fix $n \geq 1$ and suppose, for contradiction, that $p_{3^m}(n) \leq n$ for arbitrarily large m. Since $M = \lfloor \log_2 3^m \rfloor \geq m$, Lemma 4.2 would give

$$(m - (3n+2)) \log 2 \;\leq\; 4\, C(4,1)\, n^2 \log 3\, \max(\log(2m+n), 1);$$

but the left-hand side grows linearly in m while the right-hand side grows like $\log m$, a contradiction for large m. Hence $p_{3^m}(n) \geq n + 1$ eventually. The effective threshold $m_0(n)$ is where m overtakes

$$4\, C(4,1)\, n^2 \,\frac{\log 3}{\log 2}\, \log(2m+n). \qquad \square$$

Theorem B is the analogue, for the finite digit word of a single integer 3^m, of the Morse–Hedlund rung available for the infinite steering word of $(3/2)^n$ (see our [RS26]); it lies at the linear-forms optimum, superlinear per-word complexity being a separate and harder question.

5. Acknowledgements

The author utilized Claude Code as an AI coding assistant to aid in the Lean 4 formalization of the proofs presented in this paper. The author directed and reviewed all generated code and takes full responsibility for the mathematical integrity and final content of the work.

Appendix A. Formalization

Every statement above—the classical bounds reviewed in Section 1.1 as well as the new Theorems A and B—was verified in Lean 4. The files live under `TH/DigitBlocks/`, with `TH/StewartDigits.lean`, `ForMathlib/Combinatorics/SubwordComplexity.lean`, `CITED/BakerWustholz.lean` and `CITED/DimitrovHowe.lean` supplying inputs. "std3" abbreviates the ambient axioms of classical Lean (propositional extensionality, choice, quotient soundness); "[BW93]" marks additional reliance on the cited Baker–Wüstholz axiom `BakerWustholz.linearForms_logs`, and "[DH23]" on the cited Dimitrov–Howe axiom `DH.three_pow_three_binary_digits`. Every entry is fully proved in Lean except the two marked *cited axiom*.

All Lean-4 files are available from the repository `https://github.com/rwst/Aperiodicity-and-Subword-Complexity`

Reviewed results and cited inputs (§1.1, §2).

Statement	Lean identifier	File	Status
Def. 1.1	`transitionCount`	`Defs`	def
Thm. 1.2	`stewart_digitSum_three_pow`	`StewartDigits`	std3 + [BW93]
Thm. 1.3	`transitionCount_three_pow_lower`	`Transitions`	std3 + [BW93]
—	`transitionCount_three_pow_tendsto_atTop`	`Transitions`	std3 + [BW93]
Thm. 1.4	`DH.three_pow_three_binary_digits`	`DimitrovHowe`	cited axiom [DH23]
Thm. 1.5	`digitSum_three_pow_le_two`	`SparseSide`	std3
—	`digitSum_three_pow_eq_one/_two`	`SparseSide`	std3
—	`digitSum_three_pow_eq_three`	`SparseSide`	std3 + [DH23]
Thm. 2.1	`BakerWustholz.linearForms_logs`	`BakerWustholz`	cited axiom [BW93]
—	`gap_bound_ones`	`GapPrinciples`	std3 + [BW93]
—	`windowCount_lower_bound`	`GapPrinciples`	std3

New results and their tools (§2–§4).

Statement	Lean identifier	File	Status
Def. 1.1	`breakCount`	`Aperiodicity`	def
Lem. 2.2	`gap_principle`	`GapPrinciples`	std3 + [BW93]
Prop. 2.3	`windowCount_lower_bound_gen`	`GapPrinciples`	std3
Lem. 2.4	`tendsto_atTop_of_lower_bound`	`GapPrinciples`	std3
Thm. A	`breakCount_three_pow_lower`	`Aperiodicity`	std3 + [BW93]
—	`breakCount_three_pow_tendsto_atTop`	`Aperiodicity`	std3 + [BW93]
—	`subwordComplexity`	`SubwordComplexity`	def
Lem. 4.1	`finite_morse_hedlund`	`SubwordComplexity`	std3
Lem. 4.2	`log_le_of_lowComplexity`	`Complexity`	std3 + [BW93]
Thm. B	`three_pow_complexity_ge`	`Complexity`	std3 + [BW93]

The sanity witnesses $T(243) = 2$ (three blocks), $s_2(243) = 6$, $s_2(27) = 4$ and $s_2(81) = 3$ are verified by decision procedures (`transitionCount_three_pow_five`, `digitSum_three_pow_five`, `digitSum_three_pow_three`, `digitSum_three_pow_four`).

References

[BW93] A. Baker, G. Wüstholz, *Logarithmic forms and group varieties*, J. reine angew. Math. **442** (1993), 19–62.

[BFN93] R. Blecksmith, M. Filaseta, C. Nicol, *A result on the digits of a^n*, Acta Arith. **64** (1993), 331–339.

[BK17] Y. Bugeaud, H. Kaneko, *On the digital representation of smooth numbers*, arXiv:1704.00432 (2017).

[CdL00] A. Carpi, A. de Luca, *Special factors, periodicity, and an application to Sturmian words*, Acta Inform. **36** (2000), 983–1006.

[DH23] V. S. Dimitrov, E. W. Howe, *Powers of 3 with few nonzero bits and a conjecture of Erdős*, arXiv:2105.06440v4 (2023).

[MH38] M. Morse, G. A. Hedlund, *Symbolic dynamics*, Amer. J. Math. **60** (1938), 815–866.

[RS26] R. Stephan, *Superlinear complexity of the $(3/2)^n$ steering word*, arXiv:2607.11648v2 (2026).

[Ste80] C. L. Stewart, *On the representation of an integer in two different bases*, J. reine angew. Math. **319** (1980), 63–72.